In the Name of God

Journal of
Information Systems & Telecommunication
Vol. 6, No. 3, July-September 2018, Serial Number 23

Research Institute for Information and Communication Technology
Iranian Association of Information and Communication Technology
Affiliated to: Academic Center for Education, Culture and Research (ACECR)

Print ISSN: 2322-1437
Online ISSN: 2345-2773
Publication License: 91/13216
Editorial Office Address: No.5, Saeedi Alley, Kalej Intersection., Enghelab Ave., Tehran, Iran,
P.O.Box: 13145-799 Tel: (+9821) 88930150 Fax: (+9821) 88930157
E-mail: info@jist.ir , infojist@gmail.com
URL: www.jist.ir

Indexed by:
- SCOPUS www. Scopus.com
- Index Copernicus International www.indexcopernicus.com
- Islamic World Science Citation Center (ISC) www.isc.gov.ir
- Directory of open Access Journals www.Doaj.org
- Scientific Information Database (SID) www.sid.ir
- Regional Information Center for Science and Technology (RICeST) www.ricest.ac.ir
- Iranian Magazines Databases www.magiran.com

Publisher:
Regional Information Center for Science and Technology **(RICeST)**
Islamic World Science Citation Center **(ISC)**

This Journal is published under scientific support of
Advanced Information Systems (AIS) Research Group and
Digital & Signal Processing Research Group, ICTRC

Acknowledgement

JIST Editorial-Board would like to gratefully appreciate the following distinguished referees for spending their valuable time and expertise in reviewing the manuscripts and their constructive suggestions, which had a great impact on the enhancement of this issue of the JIST Journal.

(A-Z)

- Abdolvand, Neda, Azahra University, Tehran, Iran
- Alizadeh, Sasan, Qazvin Islamic Azad University, Qazvin, Iran
- Eskandari, Marzieh, Azahra University, Tehran, Iran
- Ghasemzadeh, Mohammad, Yazd University, Yazd, Iran
- Ghazalian, Reza, Babol Noshirvani University of Technology, Mazandaran, Iran
- Gheibi, Amin, Amirkabir University of Technology Tehran, Iran
- Gholamalitabarfirouzjaee, Saeed, Ball State University (BSU), Indiana, USA
- Jafari, Seyed Mohammadbagher, University of Tehran, Tehran, Iran
- Kasaei, Shohreh, Sharif University of Technology, Tehran, Iran
- Lotfi, Mohammad Mahdi, Yazd University Yazd, Iran
- Mahdieh, omid, University of Zanjan, Zanjan, Iran
- Mirroshandel, SeyedAbolghasem, University of Guilan, Guilan, Iran
- Raeesi Vanani, Iman, Allameh Tabatabai University, Tehran, Iran
- Rafighi, Masoud, Malek-Ashtar University of Technology, Tehran, Iran
- Sadeghi Bigham, Bahram, Institute for Advanced Studies in Basic Sciences, Zanjan, Iran
- Sadeghzadeh, Ramezan Ali, K. N. Toosi University of Technology, Tehran, Iran
- Sedghi, shahram, Iran University of Medical Sciences, Tehran, Iran
- Zeinali, Esmaeil, Qazvin Islamic Azad University, Qazvin, Iran

Table of Contents

Information Bottleneck and its Applications in Deep Learning

Hassan Hafez-Kolahi
Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
hafez@ce.sharif.edu
Shohreh Kasaei*
Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
kasaei@sharif.edu

Received: 24/Feb/2018 Revised: 22/Aug/2018 Accepted: 16/Sep/2018

Abstract

Information Theory (IT) has been used in Machine Learning (ML) from early days of this field. In the last decade, advances in Deep Neural Networks (DNNs) have led to surprising improvements in many applications of ML. The result has been a paradigm shift in the community toward revisiting previous ideas and applications in this new framework. Ideas from IT are no exception. One of the ideas which is being revisited by many researchers in this new era, is Information Bottleneck (IB); a formulation of information extraction based on IT. The IB is promising in both analyzing and improving DNNs. The goal of this survey is to review the IB concept and demonstrate its applications in deep learning. The information theoretic nature of IB, makes it also a good candidate in showing the more general concept of how IT can be used in ML. Two important concepts are highlighted in this narrative on the subject, i) the concise and universal view that IT provides on seemingly unrelated methods of ML, demonstrated by explaining how IB relates to minimal sufficient statistics, stochastic gradient descent, and variational auto-encoders, and ii) the common technical mistakes and problems caused by applying ideas from IT, which is discussed by a careful study of some recent methods suffering from them.

Keywords: Machine Learning; Information Theory; Information Bottleneck; Deep Learning; Variational Auto- Encoder.

1. Introduction

The area of information theory was born by Shannon's landmark paper in 1948 [1]. One of the main topics of IT is communication; which is sending the information of a source in such a way that the receiver can decipher it. Shannon's work established the basis for quantifying the bits of information and answering the basic questions faced in that communication. On the other hand, one can describe the machine learning as the science of deciphering (decoding) the parameters of a true model (source), by considering a random sample that is generated by that model. In this view, it is easy to see why these two fields usually cross path each other. This dates back to early attempts of statisticians to learn parameters from a set of observed samples; which was later found to have interesting IT counterparts [2]. Up until now, IT is used to analyze statistical properties of learning algorithms [3, 4, 5].

After the revolution of deep neural networks [6], the lack of theory that is able to explain its success [7] has motivated researchers to analyze (and improve) DNNs by using IT observations. The idea was first proposed by [8] who made some connections between the information bottleneck method [9] and DNNs. Further experiments showed evidences that support the applicability of IB in DNNs [10]. After that, many researchers tried to use those techniques to analyze DNNs [11, 12, 10, 13] and subsequently improve them [14, 15, 16].

In this survey, in order to follow current research headlines, the main needed concepts and methods to get more familiar with the IB and DNN are covered. In Section 2, the historical evolution of information extraction methods from classical statistical approaches to IB are discussed. Section 3, is devoted to the connections between IB and recent DNNs. In Section 4 another information theoretic approach for analyzing DNNs is introduced as an alternative to IB. Finally, Section 5 concludes the survey.

2. Evolution of Information Extraction Methods

A shared concept in statistics, information theory, and machine learning is defining and extracting the relevant information about a target variable from observations. This general idea, was presented from the early days of modern statistics. It then evolved ever since taking a new form in each discipline which arose through time. As is expected from such a multidisciplinary concept, a complete understanding of it re- quires a persistent pursuit of the concept in all relevant fields. This is the main objective of this section. In order to make a clear view, the methods are organized in a chronological order with the emphasis on their cause and effect; i.e., why each concept has been developed and what has it added to the big picture.

* Corresponding Author

In the reminder of this section, first the notations are defined and after that the evolution of methods from sufficient statistics to IB is explained.

2.1 Preliminaries and Notations

Consider $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ as random variables with the joint distribution function of $p(x,y)$, where $\mathcal{X}$ and $\mathcal{Y}$ are called input and output spaces, respectively. Here, the realization of each Random Variable (r.v.) is represented by the same symbol in the lower case. The conditional entropy of X, given Y, is defined as $H(X|Y) = \mathbb{E}[-\log p(X,Y)]$ and their Mutual Information (MI) is given by $I(X;Y) = \mathbb{E}[\log \frac{p(X,Y)}{p(X)p(Y)}]$. There are also more technical definitions for MI allowing it to be used in cases that the distribution function $p(x,y)$ is singular [17, 18]. An important property of MI is that it is invariant under bijective transforms f and g; i.e., $I(X;Y) = I(f(X), g(Y))$ [19].

A noisy channel is described by a conditional distribution function $p(\tilde{x}|x)$, in which $\tilde{x} \in \tilde{\mathcal{X}}$ is the noisy version of X. In the rate distortion function, the distortion function $d: \mathcal{X} \times \tilde{\mathcal{X}} \to \mathbb{R}$ is given and the minimum required bit-rate for a fixed expected distortion is studied. Then

$$R(D) = \min_{\substack{p(\tilde{x}|x) \\ s.t. \mathbb{E}[d(X,\tilde{X})] \le D}} I(\tilde{X};X). \tag{1}$$

2.2 Minimal Sufficient Statistics

A core concept in statistics is defining the relevant information about a target Y from observations X. One of the first mathematical formulations proposed for measuring the relevance, is the concept of sufficient statistic. This concept is defined below [20].

Definition 1 *(Sufficient Statistics) Let $Y \in \mathcal{Y}$ be an unknown parameter and $X \in \mathcal{X}$ be a random variable with conditional probability distribution function $p(x|y)$. Given a function $f: \mathcal{X} \to \mathcal{S}$, the random variable $S = f(X)$ is called a sufficient statistic for Y if*

$$\forall \quad x \in \mathcal{X}, y \in \mathcal{Y}: \quad P(X = x|Y = y, S = s) = P(X = x|S = s). \tag{2}$$

In other words, a sufficient statistic captures all the information about Y which is available in X. This property is stated in the following theorem [21, 2].

Theorem 1 *Let S be a probabilistic function of X. Then, S is a sufficient statistic for Y iff*

$$I(S;Y) = I(X;Y). \tag{3}$$

Note that in many classical cases that one encounters in point estimation, it is assumed that there is a family of distribution functions that is parameterized by an unknown parameter θ and furthermore N Independent and Identically Distributed (i.i.d.) samples of the target distribution function are observed. This case fits the definition by setting $Y = \theta$ and considering the high dimensional random variable $X = \{X^{(i)}\}_{i=1}^{N}$ that contains all observations.

A simple investigation shows that the sufficiency definition includes the trivial identity statistic $S = X$. Obviously, such statistics are not helpful, as copying the whole signal cannot be called "extraction" of relevant information. Consequently, one needs a way to restrict the sufficient statistic from being wasteful in using observations. To address this issue, authors of [22] introduced the notion of minimal sufficient statistics. This concept is defined below.

Definition 2 *(Minimal Sufficient Statistic) A sufficient statistic S is said to be minimal if it is a function of all other sufficient statistics*

$$\forall T; T \text{ is sufficient statistic} \Rightarrow \exists g; S = g(T). \tag{4}$$

It means that a Minimal Sufficient Statistic (MSS) has the coarsest partitioning of the input variable X. In other words, an MSS tries to group the values of $\mathcal{X}$ together in as few number of partitions as possible, while making sure that there is no information loss in the process.

The following theorem describes the relation between minimal sufficient statistics and mutual information [21].

Theorem 2 *Let X be a sample drawn from a distribution function that is determined by the random variable Y. The statistic S is an MSS for Y iff it is a solution of the optimization process*

$$\min_{T:\text{ sufficient statistic}} I(X;T). \tag{5}$$

By using Theorem 1, the constraint of this optimization problem can be written by information theory terms, as

$$\min_{T:I(T;Y)=I(X;Y)} I(X;T). \tag{6}$$

It shows that MSS is the statistic that have all the available information about Y, while retaining the minimum possible information about X. In other words, it is the best compression of X, with zero information loss about Y.

In Table 1, the components of MSS are presented in a concise way by using Markov chains. Note that these Markov chains should hold for every possible statistic S, sufficient statistic SS, and minimal sufficient tatistic MSS. By these three Markov chains and the information inequalities corresponding to each, it is easy to verify Theorems 1 and 2. By using the two first inequalities, is easily proved that $(SS;Y) = I(SS;X)$. The last inequality shows that MSS should be the SS with minimal $I(SS;X)$.

Table 1. Markov chains corresponding to conditions that form a Minimal Sufficient Statistic, along with its enforced information inequality.

	Markov Chain	Data Processing Inequality
Statistic	$Y - X - S$	$I(S;Y) \le I(X;Y)$
Sufficient	$Y - SS - X$	$I(SS;Y) \ge I(X;Y)$
Minimal	$X - SS - MSS$	$\forall\, SS : I(MSS;X) \le I(SS;X)$

In most practical problems where $X = \{(X^{(i)}\}_{i=1}^{N}$ is an N -dimensional data, one hopes to find a (minimal)

sufficient statistic S in such a way that its dimension does not depend on N. Unfortunately, it is found to be impossible for almost all distributions (except the ones belonging to the exponential family) [21, 23].

2.3 Information Bottleneck

To tackle this problem, Tishby presented the IB method to solve the Lagrange relaxation of the optimization function (6), by [9]

$$\min_{p(\tilde{x}|x)} I(\tilde{X};X) - \beta I(\tilde{X};Y) \quad (7)$$

where $\tilde{X}$ is the representation of X, and β is a positive parameter that controls the trade-off between the compression and preserved information about Y. For $\beta <= 1$, the trivial case where $\tilde{X}$ is independent of X is a solution. The reason is that the data processing inequality enforces $I(\tilde{X};X) \geq I(\tilde{X};Y) = ((1-\beta)+\beta)I(\tilde{X};Y)$. Therefore, the value $(1-\beta)I(\tilde{X};Y)$ is a lower bound for the objective function of optimization problem (7). For $\beta \geq 1$, this lower bound is minimized by setting $I(\tilde{X};Y) = 0$. It is achieved by simply choosing $I(\tilde{X};X) = 0$.

As such, the solution starts from $I(\tilde{X};X) = I(\tilde{X};Y) = 0$, and by increasing β, both $I(\tilde{X};X)$ and $I(\tilde{X};Y)$ are increased. At the limit, $\beta \to \infty$, this optimization function is equivalent to (5) [21]. Note that in IB, the optimization function is performed on conditional distribution functions $p(\tilde{x}|x)$. Therefore, the solution is no longer restricted to deterministic statistics $T = f(X)$. In general, the optimization function (7) does not necessarily have a deterministic solution. This is true even for simple cases with two binary variables [16]. The IB provides a quite general framework with many extensions (there are variations of this method for more than one variable [24]). But, since there is no evident connection between these variations and DNNs, they are not covered in this survey.

Tishby et al. showed that IB has a nice rate-distortion interpretation, using the distortion function $d(x,\tilde{x}) = KL(p(y|x)||p(y|\tilde{x}))$ [25]. It should be noted that this does not exactly conform to the classical rate-distortion settings, since here the distortion function implicitly depends on the $p(\tilde{x}|x)$ which is being optimized. They provided an algorithm similar to the well-known Blahut-Arimoto rate-distortion algorithm [26, 27] to solve the IB problem.

Till now, it was considered that the joint distribution function of X and Y is known. But, it is not the case in ML. In fact, if one knows the joint distribution function, then the problem is usually as easy as computing an expectation on the conditional distribution function; e.g., $f(x) = \mathbb{E}_{p(y|x)}[Y]$ for regression and $E_{p(y|x)}[1(Y = c)]; c \in \mathcal{Y}$ for classification. Arguably, one of the main challenges of ML is to solve the problem when one has the access to the distribution function through a finite set of samples.

Interestingly, it was found that the value of β, introduced as a Lagrange relaxation parameter in (7), can be used to control the bias-variance trade-off in cases for which the distribution function is not known and the mutual information is just estimated from a finite number of samples. It means that instead of trying to reach the MSS by setting $\beta \to \infty$, when the distribution function is unknown, one should settle for a $\beta^* < \infty$ which gives the best bias-variance trade-off [21]. The reason is that the error of estimating the mutual information from finite samples is bounded by $O(\frac{|\tilde{\mathcal{X}}|\log m}{\sqrt{m}})$, where $|\tilde{\mathcal{X}}|$ is the number of possible values that the random variable $\tilde{X}$ can take (see Theorem 1 of [21]). This has a direct relation with β: small β means more compressed $\tilde{X}$, meaning that less distinct values are required to represent $\tilde{X}$. This is in line with the general rule that simpler models generalize better. As such, there are two opposite forces in play, one trying to increase β to make the Lagrange relaxation of optimization function (7) to be more accurate, while the other tries to decrease β in order to control the finite sample estimation errors of $I(\tilde{X};X)$ and $I(\tilde{X};Y)$. The authors of [21] also tried to make some connections between the IB and the classification problem. Their main argument is that in equation (7), $I(\tilde{X}|Y)$ can be considered as a proxy for the classification error. They showed that if two conditions are met, the miss-classification error is bounded from above by $(\tilde{X}|Y)$. These conditions are: i) the classes have equal probability, and ii) each sample is composed of a lot of components (as in the document (text) classification setting). The latter is equivalent to the general technique in IT where one can neglect small probabilities when dealing with typical sets. They also argued that $I(\tilde{X};X)$ is a regularization term that controls the generalization-complexity trade-off.

The main limitation of their work is that they considered both X and Y to be discrete. This assumption is violated in many applications of ML; including image and speech processing. While there are extensions to IB allowing to work with continuous random variables [28], their finite sample analysis and the connections to ML applications are less studied.

3. Information Bottleneck and Deep Learning

After the revolution of DNNs, which started by the work of [29], in various areas of ML the state-of-the-art algorithms were beaten by DNN alternatives. While most of the ideas used in DNNs existed for decades, the recent success attracted unprecedented attention of the community. In this new paradigm, both practitioners and theoreticians found new ideas to either use DNNs to solve specific problems or use previous theoretical tools to understand DNNs.

Similarly, the interaction of IB and DNN in the literature can be divided in two main categories. The first is to use the IB theories in order to analyze DNNs and the other is to use the ideas from IB to improve the DNN-based learning algorithms. The remaining of this section is divided based on these categories.

Section 3.1 is devoted to the application of IB in analyzing the usual DNNs, which is mainly due to the conjecture that Stochastic Gradient Descent, the de facto learning algorithm used for DNNs, implicitly solves an IB

problem. In Section 3.2, the practical applications of IB for improving DNNs and developing new structures are discussed. The practical application is currently mostly limited to Variational Auto-Encoders (VAEs).

3.1 Information Bottleneck and Stochastic Gradient Descent

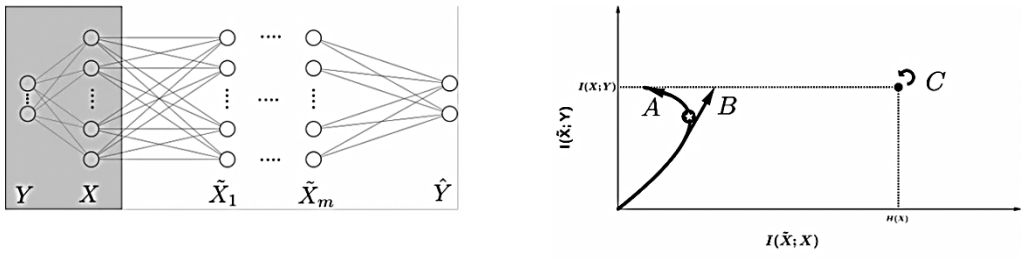
From theoretical standpoint, the success of DNNs is not completely understood. The reason is that many learning theory tools analyze models with a limited capacity and find inequalities restricting the deviation of train test statistics. But, it was shown that commonly used DNNs have huge capacities that make such theoretical results to be inapplicable [7, 4]. In recent years, there were lots of efforts to mathematically explain the generalization capability of DNNs by using variety of tools. They range from attributing it to the way that the SGD method automatically finds flat local minima (which are stable and thus can be well generalized) [30, 31, 32, 33], to efforts trying to relate the success of DNNs to the special class of hierarchical functions that they generate [34]. Each of these categories has its critics and thus the problem is still under debate (e.g., [35] argues that flatness can be changed arbitrarily by re-parametrization and the direct relation between generalization and flatness is not generally true). In this survey, the focus is on a special set of methods that try to analyze DNNs by information theory results (see [36] for a broader discussion).

Tishby et al. used ideas from IB to formulate the goal of deep learning as an information theoretic trade-off between compression and prediction [8]. In that view, an NN forms a Markov chain of representations, each trying to refine (compress) the representation while preserving the information about the target. Therefore, they argued that DNN is automatically trying to solve an IB problem and the last layer is the optimal representation $\tilde{x}$ that is to be found. Then, they used the generalization theories of IB (discussed in 2.3) to explain the success of DNNs. One of their main contributions is the idea to use the information plane diagrams showing the inside performance of a DNN (see Figure 1b). The information plane is a 2D diagram with $I(\tilde{X};X)$ and $I(\tilde{X};Y)$ as the x and y axis, respectively. In this diagram , each layer of the network is represented by a point that shows how much information it contains about the input and output.

Later, they also practically showed that in learning DNNs by a simple SGD (without regularization or batch normalization), the compression actually happens [10]. The Markov chain representation that they used and their results are shown in Figure 1. As the SGD proceeds, by tracking each layer on the information plane, they reported observing the path A in Figure 1b. In this path, a deep hidden layer starts from point $(0,0)$. The justification is that at the beginning of SGD, where all weights are chosen randomly, the hidden layer is meaningless and does not hold any information about either of X or Y . During the training phase, as the prediction loss is minimized, $I(\tilde{X};Y)$ is expected to increase (since the network uses $\tilde{X}$ to predict the label, and its success depends on how much information $\tilde{X}$has about Y). But, changes in $I(\tilde{X};X)$ are not easy to predict. The surprising phenomena that they reported is that at first $I(\tilde{X};X)$ increases (called the learning phase). But, at some point a phase transition happens (presented by a star in Figure 1b) and $I(\tilde{X};X)$ starts to decrease (called the compression phase). It is surprising because the minimized loss in deep learning does not have any compression term. By experimental investigations, they also found that compression happens in later steps of SGD when the empirical error is almost zero and the gradient vector is dominated by its noisy part (i.e., observing a small gradient mean but a high gradient variance). By this observation, they argued that after reaching a low empirical error, the noisy gradient descent forms a diffusion process which approaches the stationary distribution that maximizes the entropy of the weights, under the empirical error constraint. They also explained how deeper structures can help SGD to faster approach to the equilibrium. In summary, their results suggested that the reason behind the DNN success, is that it automatically learns short descriptions of samples, which in turn controls the capacity of models. They reported their results for both synthesis datasets (true mutual information values) and real datasets (estimated mutual information values).

Saxe et al. [13] further investigated this phenomena on more datasets and different kinds of activation functions. They observed the compression phase just in cases for which a saturating activation function is used (e.g., sigmoid or $\tanh$). They argued that the explanation of diffusion process is not adequate to explain all different cases; e.g., for Relu activation which is commonly used in the literature, they usually could not see any compression phase (path B in Figure 1b). It should be noted that their observations do not take the effect of compression completely out of picture, rather they just reject the universal existence of an explicit compression phase at the end of the training phase. As shown in Figure 1b, even though there is no compression phase in Path B, the resulting representation is still compressed compared to X. This compression effect can be attributed to the initial randomness of the network rather than an explicit compression phase. They also noticed that the way that the mutual information is estimated is crucial in the process. One of the usual methods for mutual information estimation is binning. In that approach, the bin size is the parameter to be chosen. They showed that for small enough bin sizes, if the precision error of arithmetic calculations is not involved, there will not be any information loss to begin with (Path C in Figure 1b). The reason is that when one projects a finite set of distinct points to a random lower dimensional space, the chance that any two points get mixed is zero. Even though this problem is seemingly just an estimation error caused by a low number of samples in each bin (and thus does not invalidate synthesis data results of [10]), it is actually

connected to a more fundamental problem. If one removes the binning process and deals with true values of mutual information, serious problems will arise when using IB to study common DNNs on continuous variables. The problem is that in usual DNNs, for which the hidden representation has a deterministic relation with inputs, the IB functional of optimization (7) is infinite for almost all weight matrices and thus the problem is ill-posed. This concept was further investigated in [37].



(a) (b)

Fig. 1. Information plane diagram of DNNs. (a) Markov chain representation of a DNN with m hidden layers. [Note that the predicted label $\hat{Y}$ has access to Yonly through X.] (b) Path hidden layers undergo during SGD training in information plane. Three possible paths under debate by authors are represented by A, B, and C.

Even though the problem was not explicitly addressed until recently, there are two approaches used by researchers that automatically tackle this problem. As mentioned before, the first approach, used by [8], applies binning techniques to estimate the mutual information. This is equivalent to add a (quantization) noise, making the IB functional limited. But, in this way, the noise is added just for the analysis process and does not affect the NN. As noted by [13], unfortunately some of the advertised characteristics of mutual information, namely the information inequality for layers and the invariance on reparameterization of the weights, does not hold any more.

The second approach is to explicitly add some noise to the layers and thus make the NN truly stochastic. This idea was first discussed by [10] as a way to make IB to be biased toward simpler models (as is usually desired in ML problems). It was later found that there is a direct relationship between the SGD and variational inference [38]. On the other hand, the variational inference has a "noisy computation" interpretation [16]. These results showed that the idea of using stochastic mappings in NNs has been used much earlier than the recent focus on IB interpretations. In the light of this connection, researchers tried to propose new learning algorithms based on IB in order to more explicitly take the compression into account. These ideas are strongly connected to Variational Auto-Encoders (VAEs) [39]. The denoising auto-encoders [40, 41] also use an explicit noise addition and thus can be studied in the IB framework. The next section is devoted to the relation between IB and VAE which recently has been a core concept in the field.

3.2 Information Bottleneck and Variational Auto-Encoder

Achille et al. [16] introduced the idea of information dropout in correspondence to the commonly used dropout technique [6]. Starting from the loss functional in the optimization function (7) and noting that $I(\tilde{X};Y) = H(Y) - H(Y|\tilde{X})$, one can rewrite the problem as

$$\min_{p(\tilde{x}|x)} I(X;\tilde{X}) + \beta H(Y|\tilde{X}). \tag{8}$$

Moreover, the terms can be expanded as per sample loss of

$$\begin{aligned} H(Y|\tilde{X}) &= \mathbb{E}_{p(x,y)}\left[\mathbb{E}_{p(\tilde{x}|X)}[-\log(p(Y|\tilde{X}))]\right] \\ I(X;\tilde{X}) &= \mathbb{E}_{p(x)}[KL(p(\tilde{x}|X)||(\tilde{x}))]. \end{aligned} \tag{9}$$

where KL denotes the Kullback-Leibler divergence. The expectations in these two equations can be estimated by a sampling process. For distribution functions $p(x)$ and $p(x,y)$, the training samples $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$ are already given. Therefore, the loss function of IB can be approximated as

$$\begin{aligned} \mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \quad & \mathbb{E}_{p(\tilde{x}|x^{(i)})}[-\log(p(y^{(i)}|\tilde{x}))] \\ & +\beta KL(p(\tilde{x}|x^{(i)})||p(\tilde{x})). \end{aligned} \tag{10}$$

It is worth noting that if we let $\tilde{x}$ to be the output of NN, the first term is the cross entropy (which is the loss function usually used in deep learning). The second term acts like a regularization term to prevent the conditional distribution function $p(\tilde{x}|x)$ from being too dependent to the value of x. As noted by [16], this formulation reveals interesting resemblance to Variational Auto-Encoder (VAE) presented by [39]. The VAE tries to solve the unsupervised problem of reconstruction, by modeling the process which has generated each data from a (simpler) random variable $\tilde{x}$ with a (usually fixed) prior $p_0(\tilde{x})$. The goal is to find the generative distribution function $p_{\theta(x|\tilde{x})}$ and also a variational approximation $p_\phi(\tilde{x}|x)$. This is done by minimizing the variational lower-bound of the marginal log-likelihood of the training data, given by [16]

$$\begin{aligned} \mathcal{L}_{\theta,\phi} = \frac{1}{N}\sum_{i=1}^{N} \quad & \mathbb{E}_{p_\phi(\tilde{x}|x^{(i)})}[-\log(p_\theta(x^{(i)}|\tilde{x}))] \\ & +KL(p_\phi(\tilde{x}|x^{(i)})||p_0(\tilde{x})). \end{aligned} \tag{11}$$

Comparing this with equation (10), it is evident that VAE can be considered as an estimation for a special case of IB when: i) $Y = X$, ii) $\beta = 1$, iii) the prior distribution function is fixed $p(\tilde{x}) = p_0(\tilde{x})$, and iv) the distribution

functions $p(\tilde{x}|x)$ and $p(x|\tilde{x})$ are parameterized by ϕ and θ, respectively. These parameters are optimized separately as suggested by the variational inference (note that in IB, the attention is on $p(\tilde{x}|x)$, and assuming that $p(x,y)$ is given, the values of $p(\tilde{x})$ and $p(y|\tilde{x})$ are determined from that). It is worth noting that the ii and iii restrictions are crucial. The reason is that just setting $X = Y$ and $\beta = 1$, without any other restrictions, would make the objective function (7) to be a constant, making every $p(\tilde{x}|x)$ to be a solution. Even if $\beta \neq 1$, the trivial loss function $(1-\beta)I(\tilde{X};X)$ is obtained which is minimized either for $x = \tilde{x}$ (when $\beta > 1$) or x independent of $\tilde{x}$ (when $\beta < 1$). Neither of these solutions is desired in representation learning (for another view on this matter, see the discussion of [42] on "feasible" vs "realizable" solutions).

A similar variational approach, is used to solve the IB optimization process (10), which is a more general setting with $\beta \neq 1$ and $X \neq Y$ [16].

Another concept to note is that despite the connection between IB and VAE, some of VAE issues that researchers have reported do not directly apply to IB. In fact, we think that it is helpful to use the IB interpretation to understand the VAE problems to remedy them. For example, one of the improvements over the original VAE, is β-VAE [45]. They found that having $\beta > 1$ leads to a better performance compared to the original configuration of VAE which is restricted to $\beta = 1$. This phenomena can be studied by using its counterpart results in IB. As mentioned in Section 2.3, β controls the bias-variance trade-off in case of finite training set. Therefore, one should search for β^* which practically does the best in preventing the model from over-fitting. The same argument might be applied to VAE.

Another issue in VAE, which has attracted the attention of many researchers [42, 43, 46, 47], is that when the family of decoders $p_\theta(x|\tilde{x})$ is too powerful, the loss function (11) can be minimized by just using the decoder and completely ignoring the latent variable; i.e. $p_\theta(x|\tilde{x}) = p(x)$. In this case, the optimization function (11) will be decomposed into two separate terms, where the first term just depends on θ and the second term just depends on ϕ. As a result, the second term will be minimized by setting $p_\phi(\tilde{x}|x) = p_0(\tilde{x})$. Therefore, x and $\tilde{x}$ will be independent, which is obviously not desired in a feature extraction problem. This problem does not exist in the original IB formulation, in which the focus is on $p_\phi(\tilde{x}|x)$ and $p(x|\tilde{x}) \propto p(\tilde{x}|x)p(x)$ is computed without any degrees-of-freedom (no parameter θ to optimize). It is in contrast with the VAE settings where the discussion starts from $p_\theta(x|\tilde{x})$ and later $p_\phi(\tilde{x}|x)$ is introduced in variational inference. Note that having a strong family of encoders $p_\phi(\tilde{x}|x)$, does not make any problem as long as it is adequately regularized by $KL(p_\phi(\tilde{x}|x^{(i)})||p_0(\tilde{x}))$. It should be added that even though IB does not inherently suffer from the "too strong decoder" problem, the current methods which are based on the variational distribution and optimization of both θ and ϕ are not immune to it [14, 12, 16]. This is currently an active research area and we believe the IB viewpoint will help to develop better solutions to it.

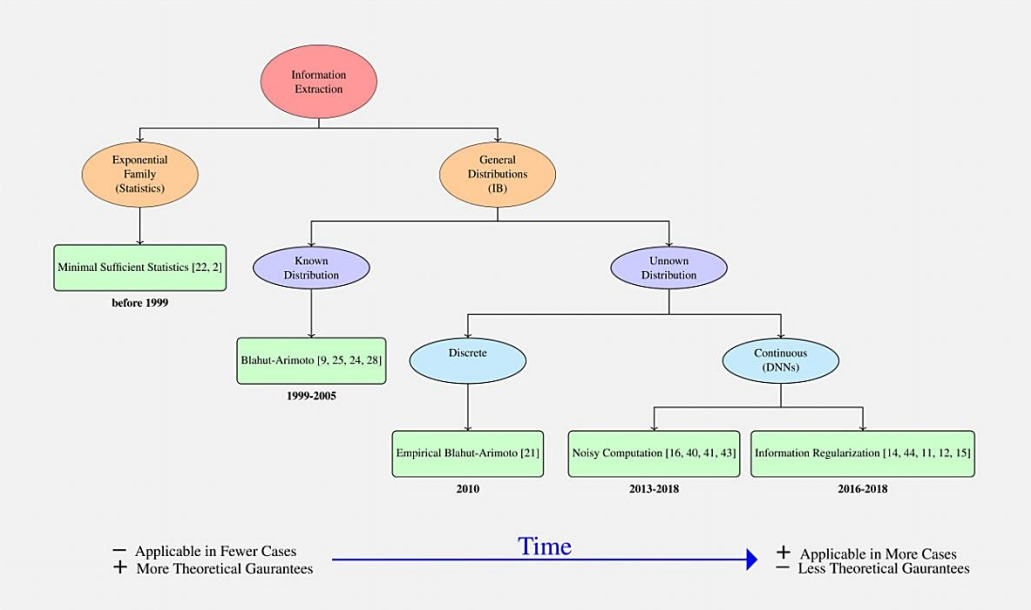


Fig. 2. Schematic review of main information extraction methods discussed in this survey, representing the evolution of algorithms through time. Moving from left to right, the methods are sorted in a chronological order. This figure shows that recent algorithms are applicable in more general cases (but usually provide less theoretical guarantees).

In Figure 2, the summary of existing methods and how they evolved trough time, is represented in a hierarchical structure. Note that the solution based on variational techniques [16] bypasses all the limitations that are faced in previous sections; i.e., meaning that it is not limited to a specific family of distributions, does not need the distribution function to be known, and also works for continuous variables. As it is represented in this figure, while the recent methods are capable of solving more general problems, the theoretical guarantees for them are more scarce.

4. Beyond Information Bottleneck

All the methods discussed till now were using IB which uses the quantity $I(X;T)$ to control the variance of the method (see Section 2.3). While this approach is used successfully in many applications, its complete theoretical analysis in the general case is difficult. In this section, a different approach based on mutual information which recently has attracted the attention of researchers is presented. In this new view, instead of looking at $I(X;T)$ as the notion of complexity, one considers $I(S;\mathcal{A}(S))$. Here S is the set of all training samples, and $\mathcal{A}$ is the learning algorithm which uses training points to calculate a hypothesis *h*.

In this approach, not only the mutual information of a single sample X and its representation is considered, but also the mutual information between all of the samples and the whole learned model is studied.

Following recent information theoretic techniques from [48, 49, 50], authors of paper [3] used the following notion to prove the interesting inequality

$$P[|\mathrm{err}_{\mathrm{test}} - \mathrm{err}_{\mathrm{train}}| > \varepsilon] < 0\left(\frac{I(S;\mathcal{A}(S))}{n\varepsilon^2}\right), \qquad (12)$$

where $\mathrm{err}_{\mathrm{test}}$ and $\mathrm{err}_{\mathrm{train}}$ are the test (true) error and the training (empirical) error of the hypothesis $\mathcal{A}(S)$, respectively, n is the training size, and $\varepsilon > 0$ is a positive real number.

The intuition behind this inequality is that, the more a learning algorithm uses bits of the training set, there is potentially more risk that it will overfit to it. The interesting property of this inequality is that the mutual information between the whole input and output of the algorithm, depends deeply on all the aspects of the learning algorithm. It is in contrast with many other approaches that use the properties of the hypotheses space $\mathcal{H}$ to bound the generalization gap, and usually the effect of final hypothesis chosen by the learning algorithm is blurred away due to the usage of a uniform convergence in proving bounds; like in the Vapnik-Chervonenkis theory [51]. In paper [52], the chaining method [53] was used to further improve the inequality (12) to also take into account the capacity of the hypotheses space.

Though the inequality (12) seems appealing as it directly bounds the generalization error by the simple-looking information theoretic term $I(S;\mathcal{A}(S))$, unfortunately the calculation/estimation of this term is even harder than $I(X;T)$ which was used in IB. This made it quite challenging to apply this technique in real world machine learning problems where the distribution is unknown and the learning algorithms is usually quite complex [54, 4].

To the best knowledge of the authors, the only attempt made to use this technique to analyze the deep learning process is the recent article [55]. In that work, authors argue that as the dataset S goes trough DNN layers 1 ... m, the intermediate sequence of datasets $(S_\ell)_{\ell=1}^{m}$ are formed and $I(S_\ell;W)$ is a decreasing function of ℓ (here W is the set of all weights in the DNN). They further argue that this can be used along the inequality (12) to show that deeper architectures have less generalization error. A major problem with their analysis is that they used the Markov assumption $W - S - S_1 - S_2 \ldots S_{m-1} - S_m$. This assumption does not generally hold in a DNN. Because for calculating the S_ℓ, a direct usage of W is needed (more precisely the weights up to layer ℓ are used). Therefore, it seems that the correct application of this technique in analyzing DNNs requires a more elaborate treatment which is hopped to be released in near future.

5. Conclusion

A survey on the interaction of IB and DNNs was given. First, the headlines of the prolong history of using the information theory in ML was presented. The focus was on how the ideas evolved over time. The discussion started from MSS which is practically restricted to distributions from exponential family. Then the IB framework and the Blahut-Arimoto algorithm were discussed which do not work for unknown continuous distributions. After that methods based on variational approximation introduced which are applicable to quite general cases. Finally, another more theoretically appealing usage of information theory was introduced, which used the mutual information between the training set and the learned model to bound the generalization error of a learning algorithm. Despite its theoretical benefits, it was shown that its application in understanding DNNs, is challenging.

During this journey, it was revealed that how some seemingly unrelated areas have hidden relations to the IB. It was also shown that how the mysterious generalization power of SGD (which is the De facto learning method of DNNs) is hypothesized to be caused by the implicit IB compression property which is hidden in SGD. Also, the recent successful unsupervised method VAE was found to be a special case of the IB when solved by employing the variational approximation.

In fact, the profound and seemingly simple tools that the information theory provides bring some traps. As the understanding of these pitfalls are as important, they were also discussed in this survey. It could be seen that how seemingly harmless information theoretic formulas can make impossible situations. Two major discussed cases were: i) using the mutual information to train continuous deterministic DNNs, which made the problem ill-posed, and ii) using variational approximations without restricting the space of solutions can easily result in meaningless situations. The important lesson learned from these revelations was how the ideas from the information theory can give a unified view to different ML concepts. We believe that this view is quite helpful to understand the shortcomings of methods and to remedy them.

Acknowledgment

We wish to thank Dr. Mahdieh Soleymani Baghshah for her beneficial discussions and comments.

References

[1] C. E. Shannon, "A mathematical theory of communication," Bell system technical journal, vol. 27, no. 3, pp. 379–423, 1948.

[2] S. Kullback and R. A. Leibler, "On information and sufficiency," The annals of mathematical statistics, vol. 22, no. 1, pp. 79–86, 1951.

[3] R. Bassily, S. Moran, I. Nachum, J. Shafer, and A. Yehudayoff, "Learners that Use Little Information," in Algorithmic Learning Theory, pp. 25–55, 2018.

[4] M. Vera, P. Piantanida, and L. R. Vega, "The Role of Information Complexity and Randomization in Representation Learning," arXiv:1802.05355 [cs, stat], Feb. 2018.

[5] I. Nachum, J. Shafer, and A. Yehudayoff, "A Direct Sum Result for the Information Complexity of Learning," arXiv:1804.05474 [cs, math, stat], Apr. 2018.

[6] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," The Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929–1958, 2014.

[7] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," International Conference on Learning Representations, 2017.

[8] N. Tishby and N. Zaslavsky, "Deep Learning and the Information Bottleneck Principle," arXiv preprint arXiv:1503.02406, 2015.

[9] N. Tishby, F. Pereira, and W. Bialek, "The information bottleneck method," in Proceedings of the 37-th Annual Allerton Conference on Communication, Control and Computing, pp. 368–377, 1999.

[10] R. Shwartz-Ziv and N. Tishby, "Opening the Black Box of Deep Neural Networks via Information," arXiv:1703.00810 [cs], Mar. 2017.

[11] P. Khadivi, R. Tandon, and N. Ramakrishnan, "Flow of information in feed-forward deep neural networks," arXiv preprint arXiv:1603.06220, 2016.

[12] A. Achille and S. Soatto, "On the Emergence of Invariance and Disentangling in Deep Representations," arXiv:1706.01350 [cs, stat], June 2017.

[13] A. M. Saxe, Y. Bansal, J. Dapello, M. Advani, A. Kolchinsky, B. D. Tracey, and D. D. Cox,"On the Information Bottleneck Theory of Deep Learning," International Conference on Learning Representations, Feb. 2018.

[14] A. A. Alemi, I. Fischer, J. V. Dillon, and K. Murphy, "Deep Variational Information Bottleneck," International Conference on Learning Representations, 2017.

[15] A. Kolchinsky, B. D. Tracey, and D. H. Wolpert, "Nonlinear Information Bottleneck," arXiv:1705.02436 [cs, math, stat], May 2017.

[16] A. Achille and S. Soatto, "Information Dropout: Learning Optimal Representations through Noisy Computation," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2018.

[17] A. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," IRE Transactions on Information Theory, vol. 2, no. 4, pp. 102–108, 1956.

[18] T. M. Cover, P. Gacs, and R. M. Gray, "Kolmogorov's Contributions to Information Theory and Algorithmic Complexity," The annals of probability, no. 3, pp. 840–865, 1989.

[19] T. M. Cover and J. A. Thomas, Elements of Information Theory. John Wiley & Sons, 2012.

[20] M. A. RA Fisher, "On the mathematical foundations of theoretical statistics," Phil. Trans. R. Soc. Lond. A, vol. 222, no. 594-604, pp. 309–368, 1922.

[21] O. Shamir, S. Sabato, and N. Tishby, "Learning and generalization with the information bottleneck," Theoretical Computer Science, vol. 411, pp. 2696–2711, June 2010.

[22] E. L. Lehmann and H. Scheffe, "Completeness, Similar Regions, and Unbiased Estimation," in Bulletin of the American Mathematical Society, vol. 54, pp. 1080–1080, Charles St, Providence, 1948.

[23] B. O. Koopman, "On distributions admitting a sufficient statistic," Transactions of the American Mathematical society, vol. 39, no. 3, pp. 399– 409, 1936.

[24] N. Friedman, O. Mosenzon, N. Slonim, and N. Tishby, "Multivariate information bottleneck," in Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, pp. 152–161, Morgan Kaufmann Publishers Inc., 2001.

[25] R. Gilad-Bachrach, A. Navot, and N. Tishby, "An information theoretic tradeoff between complexity and accuracy," in Learning Theory and Kernel Machines, pp. 595–609, Springer, 2003.

[26] R. Blahut, "Computation of channel capacity and rate-distortion functions," IEEE transactions on Information Theory, vol. 18, no. 4, pp. 460–473, 1972.

[27] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," IEEE Transactions on Information Theory, vol. 18, no. 1, pp. 14–20, 1972.

[28] G. Chechik, A. Globerson, N. Tishby, and Y. Weiss, "Information bottleneck for Gaussian variables," Journal of machine learning research, vol. 6, no. Jan, pp. 165–188, 2005.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, pp. 1097–1105, 2012.

[30] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: Generalization gap and sharp minima," arXiv preprint arXiv: 1609.04836, 2016.

[31] S. Hochreiter and J. Schmidhuber, "Flat minima," Neural Computation, vol. 9, no. 1, pp. 1–42, 1997.

[32] P. Chaudhari, A. Choromanska, S. Soatto, and Y. LeCun, "Entropy-sgd: Biasing gradient descent into wide valleys," arXiv preprint arXiv: 1611.01838, 2016.

[33] M. Hardt, B. Recht, and Y. Singer, "Train faster, generalize better: Stability of stochastic gradient descent," arXiv preprint arXiv:1509.01240, 2015.

[34] T. Poggio, H. Mhaskar, L. Rosasco, B. Miranda, and Q. Liao, "Why and When Can Deep–but Not Shallow–Networks Avoid the Curse of Dimensionality," arXiv preprint arXiv:1611.00740, 2016.

[35] L. Dinh, R. Pascanu, S. Bengio, and Y. Bengio, "Sharp minima can generalize for deep nets," arXiv preprint arXiv:1703.04933, 2017.

[36] R. Vidal, J. Bruna, R. Giryes, and S. Soatto, "Mathematics of Deep Learning," arXiv: 1712.04741 [cs], Dec. 2017.

[37] R. A. Amjad and B. C. Geiger, "How (Not) To Train Your Neural Network Using the Information Bottleneck Principle," arXiv: 1802.09766 [cs, math], Feb. 2018.

[38] P. Chaudhari and S. Soatto, "Stochastic gradient descent performs variational inference, converges to limit cycles

for deep networks," arXiv: 1710.11029 [cond-mat, stat], Oct. 2017.
[39] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," arXiv preprint arXiv:1312.6114, 2013.
[40] Y. Bengio, L. Yao, G. Alain, and P. Vincent, "Generalized denoising auto-encoders as generative models," in Advances in Neural Information Processing Systems, pp. 899–907, 2013.
[41] D. J. Im, S. Ahn, R. Memisevic, and Y. Bengio, "Denoising Criterion for Variational Auto-Encoding Framework.," in AAAI, pp. 2059– 2065, 2017.
[42] A. A. Alemi, B. Poole, I. Fischer, J. V. Dillon, R. A. Saurous, and K. Murphy, "Fixing a Broken ELBO," arXiv:1711.00464 [cs, stat], Feb. 2018.
[43] X. Chen, D. P. Kingma, T. Salimans, Y. Duan, P. Dhariwal, J. Schulman, I. Sutskever, and P. Abbeel, "Variational lossy autoencoder," arXivpreprint arXiv:1611.02731, 2016.
[44] C.-W. Huang and S. S. S. Narayanan, "Flow of Renyi information in deep neural networks," in Machine Learning for Signal Processing(MLSP), 2016 IEEE 26th International Workshop On, pp. 1–6, IEEE, 2016.
[45] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "Beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework," International Conference on Learning Representations, Nov. 2016.
[46] S. Zhao, J. Song, and S. Ermon, "Infovae: Information maximizing variational autoencoders," arXiv preprint arXiv:1706.02262, 2017.
[47] H. Zheng, J. Yao, Y. Zhang, and I. W. Tsang, "Degeneration in VAE: In the Light of Fisher Information Loss," arXiv:1802.06677 [cs, stat], Feb. 2018.
[48] D. Russo and J. Zou, "How much does your data exploration overfit? Controlling bias via information usage," arXiv:1511.05219, 2015.
[49] D. Russo and J. Zou, "Controlling bias in adaptive data analysis using information theory," in Artificial Intelligence and Statistics, pp. 1232–1240, 2016.
[50] A. Xu and M. Raginsky, "Information-theoretic analysis of generalization capability of learning algorithms," in Advances in Neural Information Processing Systems, pp. 2521–2530, 2017.
[51] V. N. Vapnik and A. Y. Chervonenkis, "On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities," Theory of Probability and its Applications, vol. 16, no. 2, p. 264, 1971.
[52] A. R. Asadi, E. Abbe, and S. Verdu, "Chaining Mutual Information and Tightening Generalization Bounds," arXiv:1806.03803, 2018.
[53] R. M. Dudley, "The sizes of compact subsets of Hilbert space and continuity of Gaussian processes," Journal of Functional Analysis, vol. 1,no. 3, pp. 290–330, 1967.
[54] M. Vera, P. Piantanida, and L. R. Vega, "The Role of the Information Bottleneck in Representation Learning," in 2018 IEEE International Symposium on Information Theory (ISIT), pp. 1580–1584, IEEE, June 2018.
[55] J. Zhang, T. Liu, and D. Tao, "An Information Theoretic View for Deep Learning," arXiv preprint arXiv:1804.09060, 2018.

Hassan Hafez-Kolahi was born in Mashhad, Iran, in 1989. He received the B.Sc. degree in Computer Engineering from Ferdowsi University of Mashhad in 2011, the M.Sc. degree in 2013 from Sharif University of Technology. He is currently a Ph.D. candidate at Sharif University of Technology. His research interests are in machine learning and information theory.

Shohreh Kasaei received the B.Sc. degree from the Department of Electronics, Faculty of Electrical and Computer Engineering, Isfahan University of Technology, Iran, in 1986, the M.Sc. degree from the Graduate School of Engineering, Department of Electrical and Electronic Engineering, University of the Ryukyus, Japan, in 1994, and the Ph.D. degree from Signal Processing Research Centre, School of Electrical and Electronic Systems Engineering, Queensland University of Technology, Australia, in 1998. She joined Sharif University of Technology since 1999, where she is currently a full professor and the director of Image Processing Laboratory (IPL).

Improvement in Accuracy and Speed of Image Semantic Segmentation via Convolution Neural Network Encoder-Decoder

Hanieh Zamanian
Department of Electrical and Computer Engineering., University of Birjand, Birjand, Iran
hanieh.zamanian@birjand.ac.ir
Hassan Farsi*
Department of Electrical and Computer Engineering., University of Birjand, Birjand, Iran
hfarsi@birjand.ac.ir
Sajad Mohamadzadeh
Technical faculty of Ferdows, University of Birjand, Birjand, Iran
s.mohamadzadeh @birjand.ac.ir

Received: 24/Feb/2018 Revised: 22/Aug/2018 Accepted: 16/Sep/2018

Abstract

Recent researches on pixel-wise semantic segmentation use deep neural networks to improve accuracy and speed of these networks in order to increase the efficiency in practical applications such as automatic driving. These approaches have used deep architecture to predict pixel tags, but the obtained results seem to be undesirable. The reason for these unacceptable results is mainly due to the existence of max pooling operators, which reduces the resolution of the feature maps. In this paper, we present a convolutional neural network composed of encoder-decoder segments based on successful SegNet network. The encoder section has a depth of 2, which in the first part has 5 convolutional layers, in which each layer has 64 filters with dimensions of 3×3. In the decoding section, the dimensions of the decoding filters are adjusted according to the convolutions used at each step of the encoding. So, at each step, 64 filters with the size of 3×3 are used for coding where the weights of these filters are adjusted by network training and adapted to the educational data. Due to having the low depth of 2, and the low number of parameters in proposed network, the speed and the accuracy improve compared to the popular networks such as SegNet and DeepLab. For the CamVid dataset, after a total of 60,000 iterations, we obtain the 91% for global accuracy, which indicates improvements in the efficiency of proposed method.

Keywords: Semantic Segmentation; Convolutional Neural Networks; Encoder – Decoder; Pixelwise Semantic Interpretation.

1. Introduction

Semantic segmentation for 2D images, video and even 3D data is one of the key problems in computer vision [1]. For large images, semantic segmentation is one of the high-level tasks that makes a full scene understanding [2]. The importance of the scene understanding as a major problem in computer vision is due to the fact that a large number of applications is improved or developed by the inference of image information [3,4]. Some of these include independent driving, human-machine engagement, image search engines, and virtual reality [5]. In the past, solutions were developed by using various machine learning techniques for this problem. Despite the popularity of machine learning based methods, deep learning has revolutionized the solution of these problems, so that many computer vision problems, including semantic segmentation, with the use of deep architecture, especially the convolutional neural networks (CNN), perform with even better accuracy than other approaches [6-8]. Semantic segmentation is still challenging task today. Theoretically, semantic segmentation combines two functions [9]; one is the segmentation of the image, and the other is the classification of the objects in which eventually connects parts of the image that belongs to one object class. By semantic segmentation, we can obtain the pixelwise semantic interpretation of the image [10]. Compared to the object detection, semantic segmentation is considered to be a major improvement because the distinction between objects is mentioned based on the distinction between the pixels. However, there are several problems and challenges that are mainly summarized in the following aspects: 1) Object Level: due to differences in lighting, viewing points and distance, an object in the image may be seen in very different ways. 2) Class Level: objects in one class may be different, and objects in different classes may be similar. For example, a pedestrian in front of a car divides the visual view of the car into two parts. 3) Background: a clean background helps to split, but in practice, the background is usually complicated which may be misleading [11].

Before the development of the deep learning algorithms, there were several popular ways to segment the image. Threshold splitting is one of the most basic methods of image segmentation, in which pixels are divided according to their color or gray levels [12]. The edge segmentation is the identification of some points at the edge of the objects which extracts a segmented region by using some particular algorithms. The Snake model transforms the segmentation into an energy minimization problem to find the edges [13].

* Corresponding Author

Watershed algorithm is a regional division based on morphology. Regional growth method is also a common method for regional segmentation [14]. The main idea of this algorithm is to find the growth criterion and then to search for a pixel of grain in each region. The random forest, which has multiple decision trees, is used as a classifier [15]. In image segmentation based on the graph theory, the image is depicted as an indirect weighted graph in which pixels are considered as nodes. The weight of the edge between the nodes is related to the difference between two pixels. Cutting these edges depends on the energy function. Markov Random Fields are an indirect probability graph model used to split the image. Each pixel is assigned a random value and then each pixel is categorized by using probabilistic methods.

Following the development of deep learning, a series of semantic segmentation methods based on the convolutional neural network were proposed and resulted in great progress. One of the most popular primary learning methods was the fragmentation, in which each pixel was categorized separately by using a piece of the surrounding image. The main reason for the use of patches is that the classification networks usually have fully connected layers, which require fixed-dimensional images.

In 2014, the fully convoluted (FCN) network is introduced by Long and colleagues [16], presented the well-known CNN architecture for dense prediction without fully-connected layers. In the FCN algorithm, the size of the input image is arbitrary and is faster than the fractional classification method. Almost all of the subsequent later methods of semantic segmentation somehow try to improve this pattern.

After FCN, SegNet [17], Detailed Convolutions [18], DeepLab V1 [19], DeepLab V2 [20], RefineNet [21], PSPNet [22], Big Core Problems [23] and DeepLab v3 [24] have been consecutively proposed and improved the accuracy of pixel-wised segmentation.

However, the main problems with these methods are the size of the networks and the time of calculation which are great for using them for real-time applications. Especially for semantic segmentation applications, such as independent driving, they are undesirable and sometimes impossible.

In this paper, in order to overcome to these problems, an idea has been presented in which a new architecture for the semantic segmentation, especially for city images, is introduced with a better accuracy than the successful architecture of SegNet and provides 10 times fewer parameters than SegNet. In later sections, after presenting an overview of existing architectures by using deep learning architecture, an innovative technique that has been tested in the framework of MATLAB is described and, finally, the results on the CamVid database [25] are shown by common criteria and compared to other successful methods.

2. Related Work

In order to understand the meaning of the semantic segmentation with the modern learning architecture of deep learning, it must be noted that in fact, semantic segmentation is the achievement of the correct inference, that its base is classification, and its result is a prediction of likelihood to each object class. Therefore, a ranking list of the object similarity with the objects in the image should be provided. Localization or diagnosis is the next step in deduction, not only the classes but also additional information about the location of these classes, such as their center or boundary boxes, should be taken into account. Therefore, it is clear that semantic segmentation is a natural step for achieving accurate inferences; and its purpose is dense predictions and labeling for each pixel. In this way, each pixel is labeled with the object class or region that is most similar to it.

Training a deep neural network from the beginning is often not possible for various reasons; first, a large set of data is needed for network training (and usually not available), and achieving convergence for an acceptable result is taking a long time. Second, even if a dataset is large enough to deprive its long-term convergence, it is often useful to begin with pre-trained weight training rather than randomly selecting them [26, 27]. Pre-trained weight training means initializing the weights of the network when are learned for another dataset or task instead of initializing the weights randomly and then start training the network for the special task and dataset. Initializing the weights is referred to use a pre-trained network. The first network is pre-trained network. The second one is the network which is fine-tuning. Adjusting the weight of a pre-trained network is one of the most important learning transfer scenarios by continuing the training process.

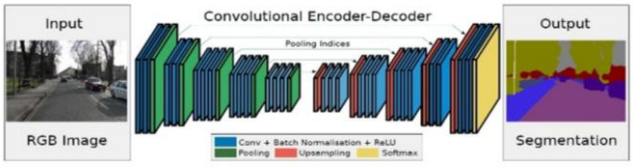


Fig. 1. SegNet architecture for pixelwise classification in semantic segmentation [17]

Yosinski et al. proved that the transfer of features, even on different issues, could be better than using a random initial value [28]. So, it is important to consider that when the difference between the previously trained problem and the goal is great, the ability to transfer features decreases.

However, the use of the transfer learning method is not simple. Using a pre-trained network contains architectural constraints that should be considered. However, since the introduction of a completely new architecture is not common, usually the architecture of the networks or their components are reused, so transfer learning is possible. On the other hand, the training process is also slightly different when the network configuration is used to be fine-tuned instead of training it from the beginning. Proper selection of layers which are usually the high levels of the network needs to be fine-tuned. Since the layers of the lower levels extract the general characteristics, lower layers also select the appropriate learning rate, which is usually a small number. Since it is expected that the pre-trained weights are relatively proper, so they do not require many changes.

Another problem for image semantic segmentation with deep networks is the learning dataset. Due to the inherent complexity of aggregation and the creation of a segmentation dataset with marked pixels, the number of the images in these datasets is not as large as the classification datasets, such as ImageNet [29,30]. This problem is even worse when dealing with colorful or 3D image datasets. So, the transfer learning and, in particular, the precise adjustment of pre-trained classified networks, is a common trend in the segmentation networks. The success of deep learning techniques in high-level issues in computer vision, in particular in supervised approaches such as CNN for image classification or object detection [31-33], induced researchers to use this technique, to check the capabilities of such networks for problems with pixel level labeling such as semantic segmentation. The key advantage of these techniques, which surpasses them from traditional methods, is the ability to learn the proper features for the desired problem. Nowadays, the most successful advanced deep learning methods for the semantic segmentation are based on a common pioneer: Fully Convolutional Network (FCN) [16]. The insight of this approach was to use existing CNNs as powerful visual models that were able to learn the hierarchy of features. They deployed existing and well-known classification models, AlexNet [31], VGG16 (16 pure layers), [32], GoogLeNet [33], and ResNet [34] into very complex networks. The result of replacing fully connected layers with a convolutional layer is achieving the network output, as a spatial map, instead of a ranking list. These maps are sampled using deconvolution [35, 36] and produce dense outputs with labeled pixels. This was considered as a milestone since it showed how CNNs can train for this problem and effectively evaluate dense predictions for the semantic segmentation for arbitrary input sizes. This method greatly improves the segmentation accuracy along with maintaining efficiency, compared to traditional methods, on different datasets such as PASCAL VOC.

For all these reasons, the FCN is the base of deep learning methods, which are applied to semantic segmentation. Despite the power and flexibility of the FCN model, this model does not have some of the necessary features which makes it difficult to use for some problems. The causes of the undesirable results of this model are inherent irregularity of its spatial form, which makes it impossible to use useful general information, and on the other hand, does not work for real-time use when the resolution is high.

Of course, FCN-based architectures are very popular and successful, but there are other alternatives that are noteworthy. In general, all of them, like the VGG-16 [32], consider a network for classification and eliminates all of its fully connected layers. This part of the new segmentation network is often called the encoder and produces a low-resolution image or a feature map. Decoding or displaying the low-resolution images is difficult to segment as pixel level predictions, and usually, the difference of these kinds of architectures is in the decoding section.

SegNet [17] is a clear example of these kinds of architectures. Figure 1 shows an overview of this architecture. The SegNet decoding part is comprised of a set of upsampling and convolution layers. The softmax classification layer which is located at the end of the network is used to predict the pixel tags of the output, and the output has the same resolution as the input image. Each layer of upsampling in the decoding section corresponds to a max-pooling in the encoder section.

These layers upsample the features by using the max-pooling indices in the encoder phase. Then upsampled maps are convolved with a set of trained filter banks to produce a map of dense features. When the feature map is returned to the original resolution, it is fed to the softmax layer to produce the final segmentation. In the SegNet encoder section, the number of convolutional layers is equal to Vgg16, only the fully connected layers in the VGG16 architecture are eliminated, which significantly reduces network dimensions and learning parameters. An important part of the SegNet architecture is its decoder section. The decoder in SegNet is hierarchically related to each step of the encoding section. Each decoder must receive max pooling indices from their respective encoders and apply non-linear upsampling to their inputs. The use of these indexes has several advantages [17]: First, it improves the boundary detections. Second, high-frequency details are maintained. Third, it reduces the training parameters. Forth, this method can be used in many encoder-encoder architectures by some modifications. Figure 2 shows how to apply the unpooling operation in SegNet architecture. As it shows, the indexes of each max pooling layer are stored in the codec section, and then in the decoding section and in the upsample layer, the unpooling operation is performed by using stored parameters.

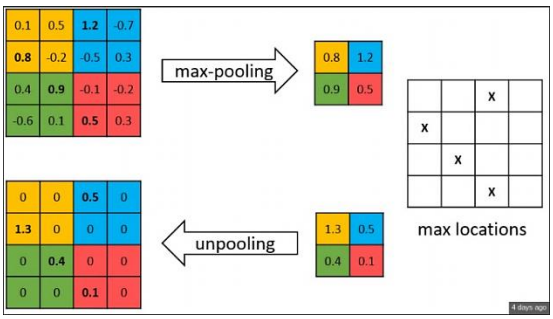


Fig. 2. Unpooling operation in SegNet architecture [17]

3. Proposed Method

In recent years, a lot of research has been done on pixel labeling for semantic segmentation of images. Some of these approaches have used deep architecture to predict pixel tags, but the results seem to be undesirable. The reason for these unacceptable results is mainly due to the existence of max pooling operators, which reduces the resolution of the feature maps. SegNet, introduced an idea, for translating the low resolution of these features to the input image resolution for pixel categorization, which results in the creation of useful features for determining the exact location of the objects

boundaries in the images. SegNet is designed for pixel-wise semantic segmentation and mainly used to understand road imagery. In a typical road image, the majority of pixels are related to large classes such as roads and buildings. A desirable semantic segmentation operator must correctly classify and isolate the boundaries of objects, due to the inequality and proportion between the numbers of pixels belonging to different classes. In addition, a segmentation operator must determine the type of object in spite of its small dimensions; therefore, it must extract the correct information from the boundary of objects, so that it can correctly decide on the type of objects.

From the computational point of view, the designed network should be efficient in terms of memory and the duration of computation at the inference level. The ability to train the network based on weighing techniques, such as Stochastic Gradient Descent (SGD), is an advantage in deep learning networks such as SegNet which speeds up the convergence of network learning.

Thus, with regard to the capabilities of SegNet, which so far has been able to improve the semantic segmentation results, proposed method is a codec method that is inspired by the Segnet algorithm to better determine the parameters and the number of layers for CamVid dataset. As noted, from a computational point of view, the designed network should be efficient in terms of memory and the duration of computations in the inference step. Certainly, with having the lower number of layers and learning parameters, the network will be more efficient from the computational point of view, and will be more useful for online uses. The goal of proposed algorithm is to reduce these parameters simultaneously with increasing precision in semantic segmentation.

In proposed method, the encoder section has a depth of two, which in the first part has five convolutional layers, in which each layer has 64 filters with dimensions of 3×3. After convolution layer, there is a batch normalization layer and then a ReLU layer. A graphical representation of the suggested network graph that is compared to the SegNet network is shown in Figure 3. In Figure 4, proposed network encoding architecture is displayed.

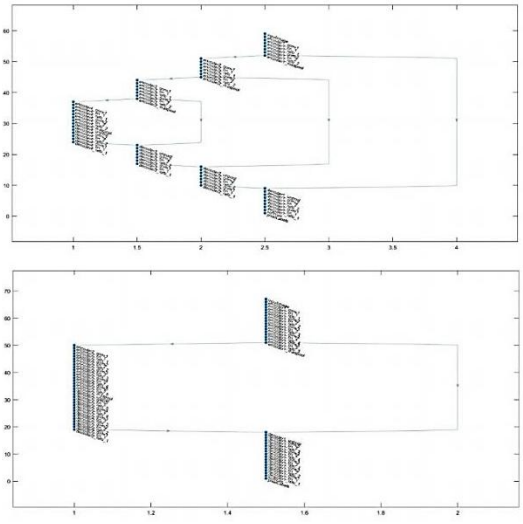


Fig. 3. SegNet architecture [17] (first diagram) vs proposed architecture (second diagram).

In the decoding section, the dimensions of the decoding filters are adjusted according to the convolutions used at each step of the encoding. So, at each step, 64 filters with the size of 3×3 are used for coding, the weights of these filters must be adjusted by network training and adapted to the training data. As shown in Figure 5, the network architecture of proposed method for decoding is displayed. At the end of the decoding section, the output is created with dimensions equal to the input image. The softmax layer performs pixel classification and the result of semantic segmentation of the input image is achieved. Due to a large size of the input image and consequently the large number of pixels that should be decided upon in the classification stage, this stage contains the most adjustable parameters.

Figure 6 shows a complete view of proposed network. One of the goals of this research is to reduce the number of parameters that need to be set during the network training. Table 1 presents a comparison between the number of parameters in different networks and proposed network. As can be seen, the number of parameters in proposed method is about 2 mega. This reduction in the number of parameters reduces the computational cost and memory needed to store network parameters as well as increasing the speed of training and even test the network.

In general, the lower numbers of parameters provide the more effective the network utilization for online and real-time applications. Table 1 shows a comparison between the numbers of adjustable parameters for several methods. As can be seen, due to the low depth of proposed network, the number of parameters that should be adjusted during the training period is much lower than the other methods, resulting in faster training and faster convergence. For example, after 100 epochs, which contain approximately 20,000 iterations, the network converges for the CamVid database.

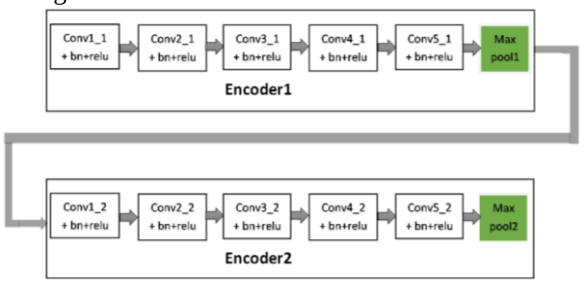


Fig. 4. Encoding block diagram of proposed method

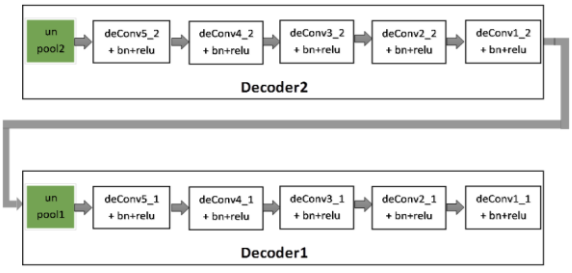


Fig. 5. Decoding block diagram of proposed method

Fig. 6. Block diagram of proposed method

Table 1. Comparison between the adjustable parameters of proposed method and some others

Network name	Type	Number of parameters(*106)
SegNet [17]	convolution	14.7
ENet[37]	residual	0.36
SqueezeNet[38]	convolution	2.7
VGG 16[32]	convolution	138
Proposed method	convolution	1.41

3.1 Database

As mentioned, in this research, CamVid road scene dataset [25] was used to evaluate the performance of the network. This dataset is small and contains 701 images, in which 421 images are used for training set and 280 images are used for testing and validation sets. These images are RGB and include scenes of day and evening, with a resolution of 360 by 480 pixels. The challenge is to separate 11 classes of roads, buildings, cars, pedestrians, signs, columns, pedestrians, sky, trees, bicycles, and fence.

3.2 Network Training

For training and testing of proposed network, the CPU with Intel Core i7-6700HQ, NVIDIA GEFORCE GTX 950M graphic card, one GPU and 12GB of memory have been used. Proposed method codes are written using the MATLAB software toolkit. For training, the stochastic gradient descent (SGD) with the initial learning rate of 0.1 and its reduction by a factor of 0.1 after every 100 epochs (20,000 iterations) and momentum of 0.5 were used. In addition, crossover entropy loss [16] has been used as a target function for network training.

When there is a large variation in the number of pixels in each class in the training images (for example, for the road, sky, and building, the number of pixels in the CamVid dataset are more abundant than other objects), there is a need to weight reduction differently according to the class of objects. This method is called the class balancing. Here, the medium frequency balancing is used [39]. This means that the larger classes in the training set have the weights less than 1 and the smallest class has the highest weight value.

3.3 Comparative Metrics

To compare the quantitative performance of different types of methods, three common metrics have been used:

- Global accuracy (GA), which measures the percentage of correctly categorized pixels in the dataset [45] which is given by:

$$GA = \frac{P_C}{N} \times 100\% \quad (1)$$

where P_c is the number of pixels correctly categorized and N is the total number of pixels in the image.

- Class Accuracy (CA), which measures the average prediction accuracy over all classes [45] which is given by:

$$CA = \frac{1}{M}\sum_{i=1}^{M}\frac{P_{c_i}}{P_{t_i}} \times 100\% \quad (2)$$

where P_{ci} is the number of correctly categorized pixels in the i^{th} class and P_{ti} is the total number of pixels in the i^{th} class of the image with M different classes.

Mean Intersection Over Union (mIoU), which is used in the Pascal VOC12 challenge [40]. If A_i shows the segmented region for i^{th} class in ground truth image and B_i shows the prediction for the segmented region for i^{th} class according to the used algorithm, mIoU for M classes is calculated by [45]:

$$mIoU = \frac{1}{M}\sum_{i=1}^{M}\frac{A_i \cap B_i}{A_i \cup B_i} \quad (3)$$

The mIoU criterion is more precise than the average class accuracy since it penalties pseudo-positive predictions.

4. Experiments and Results

As already mentioned, proposed network is being trained and tested by CamVid dataset images. Table 2 shows the comparison between the semantic segmentation accuracy for proposed method and some of the known algorithms for the CamVid dataset. As shown in Table 2, proposed algorithm has been able to perform more successfully than other methods in semantic segmentation. Improving the performance of this technique is particularly noticeable in the segmentation of objects with a small number of pixels, such as the fence, pole, bicyclist, and sign symbols. This is because of the low network depth and less use of the max-pooling layers. Because, this layer is actually a kind of drop in the map of features, which results in loss of information and image clarity.

Table 2. Comparison between different algorithms for the percentage of class accuracy for each class of CamVid dataset.

Architecture	sky	building	pole	road	pavement	tree	sign symbol	fence	car	pedestrian	bicyclist
SegNet [17]	96.1	89.6	32.1	96.4	93.3	83.4	52.7	53.45	87.7	62.2	36.5
SqueezeNet [38]	94.5	88.9	36.9	93.6	93.6	75.5	19.8	1	97.7	64.4	67.6
Super Parsing [41]	96.9	87	1.7	95.9	70	67.1	30.1	17.9	62.7	14.7	19.4
Proposed method	97.8	92.3	**70.2**	**97.7**	91.6	**91.1**	**70.6**	**79.8**	94.0	**81**	**82.3**

Table 3 also shows the comparison between proposed method and several other methods after performing 40,000 iterations for training. To demonstrate proposed network convergence rate compared to other methods, the overall performance of proposed method and other methods for the CamVid dataset after 40,000 iterations of training in all methods is shown in Table 3. As is clear from the results, proposed method has achieved a better accuracy compared to other methods. Except for proposed method, other results have been adopted from the SegNet article [17].

Table 4 shows the values of the comparison criteria in Table 3 with the maximum number of iterations for the best response, according to the SegNet article. This is while proposed method is only trained for 60,000 iterations. The results of the table represent the convergence rate and achievement of higher accuracy in all criteria for proposed method than the other methods. Obviously, if the number of training iterations increases the better results will be achieved.

5. Conclusion

Proposed method is a convolutional neural network architecture based on SegNet, successful architecture of encoder and encoder components. The purpose of this network design is to reduce the amount of computational cost and memory required to process and increase speed, while at the same time, the increase in the accuracy of the training and testing of the network. Therefore, due to a 15-times reduction in the number of parameters compared to the SegNet network and achieving higher accuracy than other methods in all criteria, after only 60,000 replications of the network training because of the low volume of the database, the efficiency of proposed method has been improved in both accuracy and speed.

Table 3. Comparison between proposed method and several other methods after performing 40,000 iterations for training

Architecture	GA	CA	mIoU
SegNet [17]	88.81	59.93	50.02
DeepLab-LargeFOV [19]	85.95	60.41	50.18
FCN [16]	81.97	54.38	46.59
FCN (learnt deconv) [16]	83.21	56.05	48.68
DeconvNet [42]	85.26	46.40	39.69
Proposed method	89.49	83.76	62.65

Table 4. Comparison between proposed method and several other methods after performing the maximum number of iterations for the best response for training

Architecture	GA	CA	mIoU	Iterations 1000×
SegNet [17]	88.81	59.93	50.02	140
DeepLab-LargeFOV [19]	85.95	60.41	50.18	140
FCN [16]	81.97	54.38	46.59	200
FCN (learnt deconv) [16]	83.21	56.05	48.68	160
DeconvNet [42]	85.26	46.40	39.69	260
FC-DenseNet67 [43]	90.8	-	65.8	-
G-FRNet [44]	90.8	-	68.0	-
Proposed method	91.18	84.64	65.94	60

References

[1] A. a. M. Ess, Tobias and Grabner, Helmut and Gool, Luc van, "Segmentation-Based Urban Traffic Scene Understanding," Proceedings of the British Machine Vision Conference, pp. 84.1-84.11, 2009.

[2] A. Geiger, "Are we ready for autonomous driving? the kitti vision benchmark suite," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3354–3361, 2012.

[3] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding." in 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3213–3223, 2016.

[4] M. Oberweger, P. Wohlhart, and V. Lepetit, "Hands Deep in Deep Learning for Hand Pose Estimation," CoRR, vol. abs/1502.06807, 2015.

[5] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. S. Kweon, "Learning a Deep Convolutional Network for Light-Field Image Super-Resolution," in IEEE International Conference on Computer Vision Workshop (ICCVW), Santiago, Chile, pp. 57-65, 2015.

[6] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep Learning for Content-Based Image Retrieval: A Comprehensive Study," in Proceedings of the 22nd ACM international conference on Multimedia, Orlando, Florida, USA, pp. 157-166, 2014.

[7] F. Ning, D. Delhomme, Y. LeCun, F. Piano, L. Bottou, and P. E. Barbano, "Toward automatic phenotyping of developing embryos from videos," Transaction of Image. Processing, vol. 14, no. 9, pp. 1360-1371, 2005.

[8] D. C. Cire, #351, an, A. Giusti, L. M. Gambardella, #252, and r. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in Proceedings of the 25th International Conference on Neural Information Processing Systems, Vol. 2, Lake Tahoe, Nevada, pp. 2843-2851, 2012.

[9] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning Hierarchical Features for Scene Labeling," IEEE Transaction of Pattern Analysis, Machine Intelligence, vol. 35, no. 8, pp. 1915-1929, 2013.

[10] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Simultaneous Detection and Segmentation," Computer Vision – ECCV 2014. pp. 297-312, 2014.

[11] S. Gupta, R. Girshick, P. Arbeláez, and J. Malik, "Learning Rich Features from RGB-D Images for Object Detection and Segmentation," Computer Vision – ECCV 2014. pp. 345-360, 2014.

[12] S. Bittel, V. Kaiser, M. Teichmann, and M. Thoma, "Pixel-wise Segmentation of Street with Neural Networks," CoRR, vol. abs/1511.00513, 2015.

[13] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," International Journal of Computer Vision, vol. 1, no. 4, pp. 321-331, January 01, 1988.

[14] M. D. Levine, and S. I. Shaheen, "A Modular Computer Vision System for Picture Segmentation and Interpretation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 3, no. 5, pp. 540-556, 1981.

[15] T. K. Ho, "Random decision forests," in Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Vol. 1, pp. 278, 1995.

[16] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence., vol. 39, no. 4, pp. 640-651, 2017.
[17] V. Badrinarayanan, A. Kendall, and RobertoCipolla, " SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, pp. 2481-2495, 2017.
[18] F. Yu, and V. Koltun, "Multi-Scale Context Aggregation by Dilated Convolutions," CoRR, vol. abs/1511.07122, 2015.
[19] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs," CoRR, vol. abs/1412.7062, 2014.
[20] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," CoRR, vol. abs/1606.00915, 2016.
[21] G. Lin, A. Milan, C. Shen, and I. D. Reid, "RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation," CoRR, vol. abs/1611.06612, 2016.
[22] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6230-6239, 2017.
[23] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large Kernel Matters — Improve Semantic Segmentation by Global Convolutional Network." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1743-1751, 2017.
[24] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," CoRR, vol. abs/1706.05587, 2017.
[25] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic object classes in video: A high-definition ground truth database," Pattern Recogn. Lett., vol. 30, no. 2, pp. 88-97, 2009.
[26] A. Ahmed, K. Yu, W. Xu, Y. Gong, and E. Xing, "Training Hierarchical Feed-Forward Visual Recognition Models Using Transfer Learning from Pseudo-Tasks," Computer Vision – ECCV 2008. pp. 69-82, 2008.
[27] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks," in Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1717-1724, 2014.
[28] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in Proceedings of the 27th International Conference on Neural Information Processing Systems, vol. 2, Montreal, Canada, pp. 3320-3328, 2014.
[29] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255, 2009.
[30] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211-252, December 01, 2015.
[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proceedings of the 25th International Conference on Neural Information Processing Systems, vol. 1, Lake Tahoe, Nevada, pp. 1097-1105, 2012.
[32] K. Simonyan, and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," CoRR, vol. abs/1409.1556, 2014.
[33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 1-9, 2015.
[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770–778, 2016.
[35] M. D. Zeiler, G. W. Taylor, and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in Proceedings of the 2011 International Conference on Computer Vision, pp. 2018-2025, 2011.
[36] M. D. Zeiler, and R. Fergus, "Visualizing and Understanding Convolutional Networks," Computer Vision – ECCV 2014. pp. 818-833, 2014.
[37] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: {A} Deep Neural Network Architecture for Real-Time Semantic Segmentation," CoRR, vol. abs/1606.02147, 2016.
[38] G. Nanfack, A. Elhassouny, and R. O. H. Thami, "Squeeze-SegNet: {A} new fast Deep Convolutional Neural Network for Semantic Segmentation," CoRR, vol. abs/1711.05491, 2017.
[39] D. Eigen, and R. Fergus, "Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-scale Convolutional Architecture," in Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), pp. 2650-2658, 2015.
[40] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes Challenge: A Retrospective," International Journal of Computer Vision, vol. 111, no. 1, pp. 98-136, January 01, 2015.
[41] J. Tighe, and S. Lazebnik, "SuperParsing: Scalable Nonparametric Image Parsing with Superpixels," Computer Vision – ECCV 2010. pp. 352-365, 2010.
[42] H. Noh, S. Hong, and B. Han, "Learning Deconvolution Network for Semantic Segmentation," in Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1520-1528, 2015.
[43] S. Jégou and M. Drozdzal and D. Vazquez and A. Romero and Y. Bengio, "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation", in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1175-1183, 2017.
[44] M. A. Islam, M. Rochan, N. D. Bruce, and Y. Wang, "Gated Feedback Refinement Network for Dense Image Labeling", in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[45] Z, Tan, B. Liu, N. Yu, "PPEDNet: Pyramid Pooling Encoder-Decoder Network for Real-Time Semantic Segmentation", in International Conference on Image and Graphics, pp. 328-339, 2017.

Hanieh Zamanian was born in Mashhad. She received the B.Sc degree in Communication engineering from Sadjad University of Technology, Mashhad, Iran in 2009. She then received M.Sc degree from University of Ferdowsi, Mashhad, Iran, in 2014. She is currently Ph.D. student in Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran. Her research interests include Image and Video Processing, Pattern Recognition, Machine Learning and Deep Learning.

Hassan Farsi received the B.Sc. and M.Sc. degrees from Sharif University of Technology, Tehran, Iran, in 1992 and 1995, respectively. Since 2000, he started his Ph.D in the Centre of Communications Systems Research (CCSR), University of Surrey, Guildford, UK, and received the Ph.D degree in 2004. He is interested in speech, image and video processing on wireless

communications. Now, he works as professor in communication engineering in department of Electrical and Computer Eng., University of Birjand, Birjand, IRAN.

Sajad Mohamadzadeh received the B.Sc. degree in communication engineering from Sistan & Baloochestan, University of Zahedan, Iran, in 2010. He received the M.Sc. and Ph.D. degree in communication engineering from South of Khorasan, University of Birjand, Birjand, Iran, in 2012 and 2016, respectively. Now, he works as assistant professor in Faculty of Technical and Engineering of Ferdows, University of Birjand, Birjand, Iran. His area research interests include Image and Video Processing, Retrieval, Pattern recognition, Digital Signal Processing, Sparse Representation, and Deep Learning.

Handwritten Digits Recognition Using an Ensemble Technique Based on the Firefly Algorithm

Hamed Agahi*
Department of Electrical Engineering, Shiraz Branch, Islamic Azad University, Shiraz, Iran
agahi@iaushiraz.ac.ir
Azar Mahmoodzadeh
Department of Electrical Engineering, Shiraz Branch, Islamic Azad University, Shiraz, Iran
mahmoodzadeh@iaushiraz.ac.ir
Marzieh Salehi
Department of Electrical Engineering, Shiraz Branch, Islamic Azad University, Shiraz, Iran
m.artista09@yahoo.com

Received: 24/Feb/2018 Revised: 22/Aug/2018 Accepted: 16/Sep/2018

Abstract

This paper develops a multi-step procedure for classifying Farsi handwritten digits using a combination of classifiers. Generally, the technique relies on extracting a set of characteristics from handwritten samples, training multiple classifiers to learn to discriminate between digits, and finally combining the classifiers to enhance the overall system performance. First, a pre-processing course is performed to prepare the images for the main steps. Then three structural and statistical characteristics are extracted which include several features, among which a multi-objective genetic algorithm selects those more effective ones in order to reduce the computational complexity of the classification step. For the base classification, a decision tree (DT), an artificial neural networks (ANN) and a k-nearest neighbor (KNN) models are employed. Finally, the outcomes of the classifiers are fed into a classifier ensemble system to make the final decision. This hybrid system assigns different weights for each class selected by each classifier. These voting weights are adjusted by a metaheuristic firefly algorithm which optimizes the accuracy of the overall system. The performance of the implemented approach on the standard HODA dataset is compared with the base classifiers and some state-of-the-art methods. Evaluation of the proposed technique demonstrates that the proposed hybrid system attains high performance indices including accuracy of 98.88% with only eleven features.

Keywords: Classifiers Ensemble; Feature Extraction; Feature Selection; Firefly Algorithm; Multi-objective Genetic Algorithm; Optical Character Recognition.

1. Introduction

Written pattern recognition is fast becoming a key instrument in document processing in various languages. Investigating this issue is a continuing concern for several researchers to obtain fast and reliable optical character recognition systems (OCRs). Such devices aim to import information in printed or scanned documents into computers [1]. Recognition of optical characters has many applications in the real world, including checking passport documents, processing bank checks, sorting mail letters and automatic identification of license plates [2-4]. Handwritten digit recognition in different languages is considered as one of the most significant current discussions in the issue of OCRs. Recent developments in human life and automated industry have heightened the need for this technology. Surveys such as that conducted by Savas and Eldén [5] reported that the main difficulty in allocating observations to ten different classes of Arabic digits is due to high inter-class similarity and intra-class variability in such problems. Several attempts have been made to automatically recognize Western Arabic digits (i.e., 0, 1, 2, ..., 9) and good results have been obtained [6-9]; however, far less research has been conducted on the recognition of Farsi digits (i.e., 0, 1, 2, …, 9) and the reported accuracies of the existing techniques are lower than those for Western Arabic methods [10-14]. In Farsi language, research has consistently shown that due to the large similarity of the digits and also the wide differences in their drawing methods, creating a recognition system with acceptable accuracy and reliability has a number of problems in practical use. Like in Western Arabic, we face ten digits in Farsi and Eastern Arabic. Meanwhile, despite in Eastern Arabic which digits are written in one type, six digits in Farsi (i.e., 0, 2, 3, 4, 5 and 6) are described in two shapes (e.g., 6: ‘6’ and ‘6’). Such a high diversity, as shown in Fig. 1, makes it more difficult to identify Farsi digits. A recognition system, in general, contains three important steps of pre-processing, feature extraction and classification. First, in the pre-processing step some operations are performed to improve the images quality and prepare them for the main steps. Noise elimination, smoothing and normalizing the input data are some examples of such operations. Extracting features is the second step in which some characteristics from the image are extracted to constitute a feature vector assigned

* Corresponding Author

to that image. Numerous methods have been proposed to extract features from handwritten digits in different languages, including pixels density functions, geometric momentums, wavelet coefficients, projections and profiles on multiple orientations, and digit contours; see [15] for a general review of such methods. In the classification step, plentiful techniques may be recruited for recognizing handwritten digits and texts including for example the k nearest neighbor (KNN), the artificial neural networks (ANN), and the decision trees (DT). The highest proportion of the research performed on this subject concentrates on adapting the features used for digit classification. The regular features testified in the literature are extracted from writing samples and a traditional classifier like ANN is used to learn to distinguish between the handwritten digits.

This paper develops a hybrid system (also called the classifiers ensemble model) which combines the classifiers to better recognize Farsi handwritten digits. The performance of the proposed technique on a standard dataset is evaluated and some comparisons with the existing methods are presented. The approach proposed in this study contains multiple steps. In the first step, the pre-processing operations are carried out which include (i) digit shape separation and frame enfolding, (ii) inversion, (iii) resizing and (iv) thinning and removing inner pixels. These operations make the images ready for the next steps. In the second step, some characteristics from the image are extracted containing the 'branch points', the 'chain codes', and the 'crossing counts', each of which contain several features. All of these features constitute a single feature vector allocated to that image. The classification is performed by discriminating the feature vectors. Using several features increases the computational complexity and the processing time of the OCR system. For this reason, selecting features with most discriminative properties is at the heart of every effective recognition system. Hence, in the third step we use a multi-objective version of the genetic algorithm called the 'non-dominated sorting genetic algorithm II' (NSGA-II) [16]. The feature selection task is typically considered as a single-objective optimization (SOO) problem. While SOO considers only one objective function to be improved, multi-objective optimization (MOO) tries to concurrently enhance multiple objective functions. In fact, MOO generates a set of trade-off answers, among which the designer may choose one answer depending on the desires of the problem. The literature demonstrates that for solving complicated problems, methods exploiting MOO commonly perform better than those make use of SOO methods [17]. For our problem, this search metaheuristic considers the cardinality (number of members) of the selected subset and the F-measure of the classification using the ANN, as the two objective functions. The goal is that the first index is minimized while the second one is concurrently maximized. To this end, features are encoded in the form of a chromosome and the NSGA-II is applied. The final outcome is a Pareto optimal front that consists of a set of answers, each of which characterizes a different set of selected features. Finally, the specific answer of the Pareto front that returns the best accuracy is chosen as the vector of ideal features subset. Once the features are selected, only these features are taken out from new images to organize the feature vector and be fed into the classifiers. In the fourth step, the classification of the images is performed solely using the ANN, the KNN and the DT classifiers. These classification models yield different performance rates.

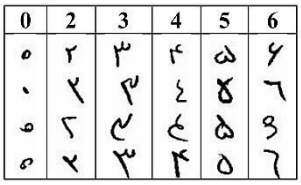


Fig. 1. Samples of Farsi handwritten digits with different shapes

Each classification system has pros and cons; thus appropriate combination of classifiers may strengthen the advantages and compensate for weaknesses of each classifier by others and provide a hybrid system with higher performance measures. The idea of the Multi-Classifier Systems (MCS), as a kind of Hybrid Intelligent Systems, was to take advantage of the individual classifiers to deliver classification systems that outperform these base classifiers. The idea of the MCS was first presented by Chow [18], who recommended conditions for ideal weighted mixture of binary classifiers. Dietterich [19] outlined the benefits of the MCS: (i) sweeping away the improper assumptions possibly caused due to small training dataset. (ii) Mixing classifiers that are trained by initiating from different starting values. This could help not to catch in local optimums. (iii) The correct decision making system may be unmanageable to be modelled by any single classifier, but mixture of classifiers may work. Consequently, we expect that a subtle mixture of the classifiers reaches higher performance measures. Subsequently, the outcomes of the mentioned base classifiers are mixed rather than considering the decision of the best classifier. Yet, the classifiers in the mixture should be selected as to be precise and diverse [20], meaning that their errors take place on different parts of the dataset. Due to its unsystematic nature and numerous neurons in hidden layers, MLP can be easily made diverse. As well, KNN yields the answers of the MCS for the patterns coming from the conflictive area of the search space. Using this classifier as a base contributor can considerably decrease the exploitation cost of the multi-classifier system [21]. The most important concern in the ensemble methodology is to find a correct method to mix the results of the classifiers. The majority voting and the weighted combination are the conventional procedures for mixing the classifiers [22]. While in the first technique, all the classifiers are mixed using the same weights; in the second technique, different weights are allocated to the classifiers. In the weighted method, the final conclusion and the overall recognition performance of the hybrid model severely depend on the weighting factors. In fact, all the donor classifiers may not be similarly capable of

perceiving all the classes. For example, in a two-class problem (call classes C1 and C2), classifier A may be good at identifying class C1, while classifier B may be skilled at discovering class C2. Hence, the weights should be varied among the diverse classes for each classifier. Allocating different weights to the conclusion of a classifier about different classes increases the performance of the mixed classification coordination.

In the latest step, similar results are linearly mixed and then the maximization is accomplished to discover the final consequence. The weights of this recognition scheme are found using an optimization method. Random optimization algorithms are one of the approaches which are able to find the appropriate combinations of the classifiers, cf. [23-27]. In the paper the 'firefly algorithm' (FA) is used to mix the individual classifiers so that the recognition accuracy of the whole system is boosted. Firefly algorithm, proposed by Yang [24], is a metaheuristic search technique for the global optimization. This method finds the optimal solution with respect to the rules inspired by the movements of fireflies due to their attractiveness. This optimization process has become a progressively significant tool of swarm intelligence with numerous applications in almost all areas of optimization, as well as engineering practice. Various problems from many areas have been effectively solved using the firefly algorithm and its variants [28, 29]. Here, the mixing method based on the FA regulates the weights of voting for each class in any classifier by maximizing the accuracy of the final recognition outcomes as the objective function. Conclusions of classifiers about classes are encoded in the form of the locations of the fireflies. Accordingly, each entry in a location vector represents the voting weight allocated to the selection of every class by each classifier. The set of ideal voting weights is characterized by the location vector of the final answer found by the FA. This answer is related to the best accuracy attained by the fusion organization.

Briefly, the main novelties of this paper are: i) selecting the most effective features to reduce the dimensionality of the feature space by removing irrelevant, redundant or misleading features. This task also decreases the computational complexity and running time of the system while increases the classification accuracy. This operation is performed by a multi-objective GA. ii) Combining the outcomes of the base classifiers such that the hybrid system attains higher performance measures. For this purpose, a weighted combination approach is taken in which a specific voting weight is assigned to each classifier selecting each class. iii) For this purpose, the firefly algorithm, as a metaheuristic, finds the optimal weights such that the accuracy of the hybrid system is enhanced. The overall structure of the study takes the form of seven sections, including this introduction. Section 2 provides an overview on related work in the field of digit recognition. Section 3 begins by laying out the steps of the proposed recognition approach, including the pre-processing operations, feature extraction and selection, individual classification, and finally classifiers combination. The details and parameters settings for the proposed algorithm are described in Section 4. Then, Section 5 presents the results of applying the recognition system to the HODA dataset and measuring the performance indices. Section 6 performs discussion and comparison with some other techniques according to the standard performance criteria. Finally, Section 7 gives a brief summary and critique of the findings and includes a discussion of the implication of the results to future research into this area.

2. Related Works

In recent years, there has been an increasing interest in the recognition of handwritten digits. A considerable amount of literature has been published on this issue, some of which are referred to in this section. Soltanzadeh and Rahmati [30] found that utilizing outer profiles, crossing counts and projection histograms as features can result in acceptable accuracy values on the test samples provided in their research. Sadri et. al. [31] proposed a method for extracting a new attribute which considered four orientations for each digit and extracted sixteen features for each orientation. Finally a vector with 64 elements was given to the SVM to classify the image. The results of applying this method to the paper dataset showed the accuracy of 94.14%. Salimi and Giveki [32] suggested an ensemble of SVD classifiers to improve the system's performance. In their study, the combination of classifiers were performed using the PSO algorithm. In a paper by Ziaratban et al. [33], based on the template matching method a system for recognizing Farsi/Arabic handwritten digits was presented. In this approach, the patterns represented the pre-determined form of numbers. Khorashadizadeh and Latif [34] proposed a new method based on a feature set including directional chain code histogram and histogram of oriented gradient. This technique also utilized local features to improve the system performance by using 164 features. For this method, the SVM was used as the classifier.

Safdari and Moin [35] introduced a new method based on two-layer sparse auto-encoder and used the weights learned from the training phase for extracting the features. Hajizadeh et al. [36] proposed a new local manifold learning called FSLL, in which the locally linear embedding (LLE) and a Stochastic Laplacian Eigenmap (SLEM) are combined. This technique reduced the dimensionality of the feature space and represented the high-dimensional data manifold in low-dimensional space. Sadeghi and Testolin [37] presented a computational model of Persian character recognition based on deep belief networks. They emerged complex visual features in an unsupervised manner by fitting a hierarchical generative model to the sensory data. Zamani et al. [38] compared the performance of the random forest (RF) and convolutional neural network (CNN) for Persian

handwritten digit recognition on the HODA dataset. They performed different experiments and finally showed that CNNs are the faster if appropriate hardware is available. It is worth mentioning that the techniques indicated above did not use the same dataset in their experiments. Although several techniques have been proposed on recognizing Farsi handwritten digits, results are not still as accurate as those achieved for Latin digits. Hence, finding a more accurate and reliable approach was the main motivation of this work.

3. The Proposed Approach

The hybrid system proposed in this paper aims to recognize Farsi handwritten digits in the HODA dataset [39]. Five main steps are carried out in our procedure consisting of pre-processing, feature extraction and selection, and finally individual and combined classification. The block diagram of the proposed system is shown in Fig. 2. In the following, the functioning of each block is described in details.

3.1 Pre-processing

Datasets used in the studies usually contain noises and incompatibilities due to their large size and also combination of several different resources. Using these data in the raw form leads to systems with unreliable results. Pre-processing is a step which plans to enhance the image quality and prepare it for the next actions. This phase significantly affects the performance of the main recognition steps. There are numerous pre-processing techniques for this purpose including blurring, histogram equalization and normalization [40]. This paper performs a particular course of operations to improve the efficiency of the recognition system including (i) segmentation and framing, (ii) image binary inversion, (iii) resizing, and finally (iv) thinning operation and inner pixels removal. In the following, the pre-processing stages are briefly outlined.

***i.* Image Segmentation and Framing:** The image of each digit in the HODA dataset is a binary image with no particular boundary who separates it from the rest of the images. Thus, first each digit image should be separated and then placed in a black frame [41]. An example of such image, which is manually performed, is shown in Fig. 3a which is surrounded by a rectangular black frame in Fig. 3b.

***ii.* Image Inversion**: In each binary image of the dataset, the digit shape is represented in white while the background is shown in black. An inversion operates by converting black to white and vice versa. This task is necessary for the next stages of the pre-processing [42]. The image inversion is shown in Fig 3c.

***iii.* Image Resizing**: Images in the HODA dataset have different sizes since they were taken from various resources. Having the same size is crucial when extracting the characteristics considered in this paper. For this purpose, the images of digits are resized according to a pre-specified size [42]. In this paper, the sizes of each image are changed to 46×46 as represented in Fig. 3d.

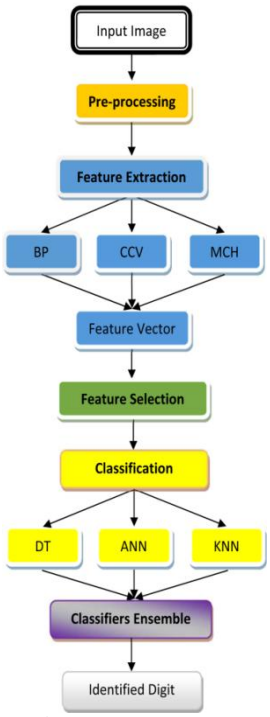


Fig. 2 Block diagram of the proposed hybrid system for recognizing Farsi handwritten digits.

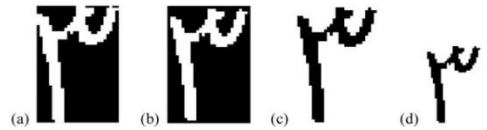


Fig 3 (a) A digit image from the HODA dataset [39], (b) framing, (c) inversion, (d) resizing

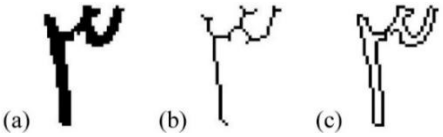


Fig. 4 (a) The original image, (b) the thinned image, and (c) the inner-pixel-removed image

Fig. 5 Skeleton of a sample digit '3' with its branch points represented by red spots; #BP = 7 [43].

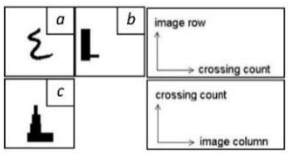


Fig. 6 (a) The original image, (b) VCCV=[1 1 1 1 1 1 2 1], and (c) HCCV=[1 2 3 4 3 2 1 1] [30].

3	2	1
4	**P**	0
5	6	7

Fig. 7. The mask centered at a pixel 'P' [41].

***iv.* Thinning Operation and the Inner pixels Removal:** To extract characteristics such as 'branch points' and 'crossing counts', the thinned image of the

digits should be used. In the thinning operation, the skeleton of the digit image is extracted. To attenuate the noise effects, a median filter is applied and then the segments that are separate from the biggest connected segment are considered as disturbances and thus they are eliminated [30]. In this paper, the pixel-wise method is used for thinning the images. A sample image and its skeleton are illustrated in Fig. 4a,b.

Generating the 'chain codes' needs the digit shape boundaries to be found. Here, we use the technique of inner pixels removal to detect the edge points of the digit shape. For each pixel, four neighbor pixels are considered. If all four neighbors are black, the intended pixel is considered as an inner one and it is converted to white; otherwise, that pixel is an edge point and remains black [42]. This procedure for the image in Fig. 4a is shown in Fig. 4c.

3.2 Feature Extraction

In the second step, we investigate some characteristics of the images from which several features are extracted. These characteristics include 'the branch points' (BP), 'the crossing count vectors' (CCV) and 'the masked code histogram' (MCH). As the first feature, the 'number of the branch points characteristic' (#BP) in the skeleton of the digit image is found as a structural feature. A pixel is referred to as a *branch point* if at least three pixels in its 3×3 neighbor window (without considering the center pixel itself) are black. A digit image with its branch points are shown in Fig. 5 [43].

The next characteristics to be found are crossing counts which contain statistical features [30]. To find the horizontal crossing counts vector (HCCV), consider the first and the last non-empty columns of the image and the columns located within this interval. The HCCV is a vector whose length is equal to the number of columns in the mentioned interval and it is formed by setting any of its element as the number of the segments in the associated column of the digit image. In fact, each element represents the number of the connected segments in one column of the interval. The same approach is taken for the vertical crossing counts vector (VCCV) by considering rows instead of columns. The crossing counts vector in each orientation is normalized into a vector of size eight by carrying out the simple averaging or by up-sampling, when necessary. This normalization makes the features robust to the image stretch in the orientation of the crossing counts. Each element of the normalized HCCV and VCCV represents a single feature. The results for a sample digit is illustrated in Fig. 6.

Since we used only the skeleton characteristics, the outlier shape information may be lost. That is why we added a complementary characteristic; *i.e.*, the 'chain code' which consists of statistical features. This characteristic takes the boundary shape information into account. To do this, a 'mask' (see Fig. 7) is applied to each pixel on the boundary of the inner-pixels-removed image [41]. For any edge pixel, the black neighbor pixels in the 3×3 mask are weighted and summed up; the result is called a 'code'. When this computation is done for all the edge points, the histogram of the codes is determined. This feature vector is called the 'masked code histogram' (MCH). Finally, this vector is resampled into a vector of size 8, similarly to what happened to the crossing counts vectors.

Once the mentioned features are computed, the complete feature vector for each digit image will be generated by appending these feature vectors into one single vector containing #BP, HCCV, VCCV and MCH. Thus, a feature vector of size 25 will be obtained for each sample image (1 feature for the #BP, 8 features for the HCCV, 8 features for the VCCV, and 8 features for the MCH). It is important to note that the length of this vector is high. Accordingly dealing with such a vector is difficult and the computational burden will be high. Hence, in next step, by selecting more effective features, the feature vector length is reduced.

3.3 Feature Selection

Use of several features makes it more challenging to develop accurate classification models. From the practical viewpoint, using a large number of features leads to high computational complexity and large running time along with high risk of over-fitting and worsening the classification performance. Feature selection (FS) is a worthy approach to address these challenges. FS is the procedure of choosing a subset of significant features in order to make straightforward the model production and understanding and also to improve the model generality. Let m be the total number of features existing to pick from; then there exist 2^m possible feature subsets. Therefore, for large values of m (here, m=25), exhaustive search is impracticable. For the FS problem, many algorithms are proposed in the literature; see [44] for a review of the commonly used FS techniques. In this paper the genetic algorithm (GA) is used to find an optimal feature subset with large discrimination power. In fact, GA tries to remove redundant or irrelevant features. GAs are optimization methods based on the Darwin's principle inspired from the genetic reproductions. In this metaheuristic method, better populations among different species are developed during evolution. The GA presents an operative methodology for problems like FS, due to its capability of fast global search of large, non-linear and poorly understood spaces and also direct operations and low computational load [45].

In the process of selecting features by means of the GA, a population of chromosomes is considered each of which represents a candidate solution for this problem. Any chromosome is represented by a binary vector of length m. In the initial population, the genes of each chromosome (i.e., the vector elements) are randomly initialized to either 1 or 0. The value of '1' means that the corresponding feature is selected, while the value of '0' indicates that that feature is not chosen [46, 47]. The fitness of a chromosome is determined by evaluating some objective functions when an ANN classifier is applied to the test dataset. The input patterns of this classifier are represented by only the selected subset of features. If a chromosome has n ($n \leq m$) bits turned on, the

associated ANN has n inputs. In this paper, a multi-objective genetic algorithm based on the 'non-dominated sorting genetic algorithm II' (NSGA-II) [16] is recruited for the FS purpose. The NSGA-II is a fast non-dominated sorting approach (NSA) with low computational complexity and an elitism approach. This technique is capable of dealing with multi-objective optimization problems [16]. For such tasks, the NSGA-II generates a set of sub-optimal solutions, among which one solution is nominated as the final chromosome depending on the desires of the problem. Accordingly, the designer has a large degree of freedom in selecting the final answer. For our FS problem, two objective functions (OF) are considered. OF1: The cardinality of the selected feature subset; OF2: The *F*-measure of the classification task using an ANN. The *F*-measure (also known as F_1 score) is a degree of a test's accuracy, defined as the harmonic mean of the precision and recall indices of that test. Indeed, the precision and recall are united to form a single index by which the system performance could be generally evaluated [48]. The precise mathematical definitions of these measures are presented in Section 4. The NSGA-II plans to minimize OF1 while simultaneously OF2 is maximized. In each generation, numerous solutions are produced which NSA ranks them with respect to the concept of domination and non-domination relations in the objective function space, as shown in Fig. 8. Accordingly, a number of non-dominated solutions may be found on the final Pareto front due to multi-objective optimization. None of these results completely dominate the others. Some results have smaller subsets; while some are better in regard to the *F*-measure. To select the most suitable features for the FS problem, the system chooses, among the solutions in the final Pareto front, the solution with the best accuracy value assessed on the training set. The optimal subset found by the GA contains selected features that will be given to the classifiers of the next step. The procedure of the feature selection task by means of the GA is shown in Fig. 9.

3.4 Classification Models

Classification is the main step in every recognition system. This task aims to determine each new pattern belongs to which of the known classes. Several algorithms are proposed for the classification of handwritten digits [15, 30, 31, 49]. In this paper, three classification models are individually used; *i.e.*, decision tree, *k*-nearest neighbor and the artificial neural network. Decision tree (DT) is a data mining technique which is widely used in classification and regression problems. As the name implies, this tree is composed of a number of nodes and branches. In a classification application, the leaf nodes represent the classes among which, one should be assigned to a query. To classify a query, the tree is traversed along a path from its root toward a leaf node. The path is decided by the subtrees chosen via the test answers in the internal nodes. No particular knowledge or parameter setting is needed to extract trees. Thus learning, deduction and decision making are straightforward and fast [50]. Depending on the application and the type of criteria, different decision trees may be utilized; the most famous ones to be noticed are ID3, C4.5, CART and CHAID [51]. This paper uses the CART decision tree as the first base classifier.

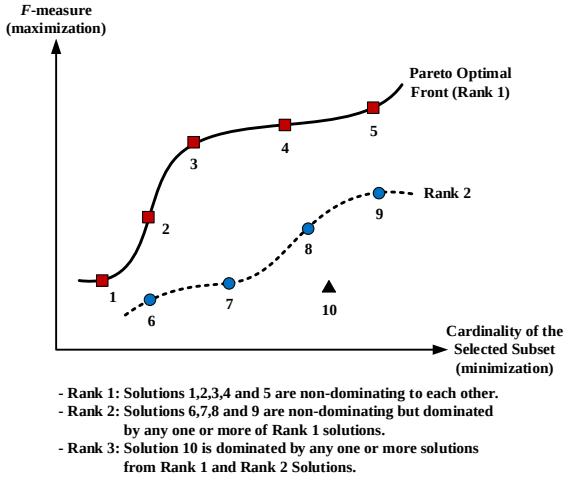


Fig. 8. Representation of dominated and non-dominated solutions.

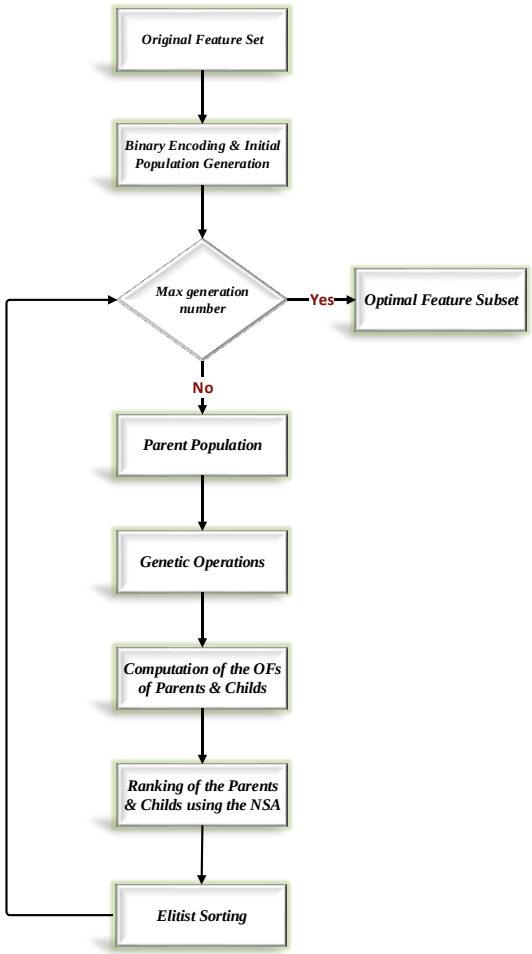


Fig. 9. Block diagram of the optimal feature selection by GA

While most classification methods need precise models to recognize the input pattern, in some approaches it is possible to find the class of a query without creating models and just by comparing the query and samples in a dataset according to some similarity indices [52]. One of

the most famous model-free methods is the '*k*-nearest neighbor algorithm' (KNN) which is frequently used for the pattern classification purposes. To assign a class label to a new input pattern, the algorithm first looks for a subset containing a *k* number of samples with the least distances to the query sample. Following this, the majority voting method determines the class with most number of samples in the nearest subset and allocates it to the input pattern. Commonly, the parameter *k* is adjusted after several experiments. The Euclidean and the Manhattan distances are mostly used as the similarity criteria in the nearest neighbor algorithms [52]. The KNN is the second base classifier used in this paper.

Artificial neural networks (ANNs) are computing systems vaguely inspired by the biological neural networks of human or animal brains. Such systems "learn" tasks from some examples, generally without any a priori knowledge about the patterns. The learning is realized only by evolving the set of characteristics of the ANNs from the learning material that they process. Some of the most appreciated neural networks are the multi-layer perceptron (MLP), the Hopfield network and the radial basis functions [53]. One of the simplest yet most efficient neural networks is the MLP network. MLP consists of an input layer, one or more hidden layers and one output layer. For the classification applications, the network should be "trained" using supervised methods. Backpropagation is a well-known algorithm for supervised learning of MLP networks. Given a MLP and an error criterion, this learning method computes the gradient of the error function backwards through the network [54]. This process continues for a certain number of iterations or it stops when an acceptable accuracy criterion is achieved. Once the MLP learned to identify the samples in the training set, it is ready to classify new patterns of the test dataset. The ANN (MLP) is the third classification model for the problem of recognizing digits of this paper.

It should be noticed that the main reasons for selecting these three classifiers is their high performance and simplicity when dealing with the classification problems. In addition, KNN is suitable for complex search spaces while DT has high speed in the classification problems. Moreover, ANN is capable of making diversity in the classifiers combination.

3.5 Classifiers Mixture

The idea of the classifiers mixture is to weigh several distinct classifiers, and mix them to acquire a fusion classification model that outperforms each of the base contributors. Given the potential advantages of the mixture methods, it is not unexpected that several methods are now accessible for theoretical researches and industrial applications [22]. Multi-Classifier Systems (MCS) try to mix some distinct classifiers and generate ensemble systems that give the final results. The following benefits of the MCS are commonly approved: (*i*) MCS act well both in the cases the data samples for learning are very limited and when an enormous number of them exist. (*ii*) The classifier mixture may outdo the best distinct base classifier. (*iii*) Several classifiers act on the basis of the heuristic search algorithms. Such procedures are not assured to find optimal answers. Accordingly, the mixture method, possibly initialing from different start points in the search space, might be considered as a multi-start local random search. This method may boost the probability of determining the optimal answer. (*iv*) MCS can be easily realized in parallel computer architectures or on distributed computing systems (*e.g*, Cloud computing). When a dataset is partitioned and the partial answer is determined on each partition, the final conclusion is made by mixing the networked consequences. (*v*) As stated by Wolpert [55], any classifier has its own competence domain; on which it exceeds other competing classifiers. As a result, a single classifier cannot be found that it beats every other one for all recognition problems. MCS attempt to generate an optimal hybrid model from the trained classifiers. The main concerns in the MCS design are: (*a*) The topology of interconnecting distinct classifiers; (*b*) Choosing valuable classifiers; and (*c*) Constructing the suitable decision fusion model (fuser) [22]. In our paper, we chose the evolutionary firefly algorithm as the fuser.

3.5.1 The Firefly Algorithm

Firefly Algorithm (FA), proposed by Yang [24], is a metaheuristic for finding the global solution in an optimization problem. This algorithm is inspired by flashing behavior of firefly insects for attracting other fireflies. Attractiveness of a firefly is relative to the brightness of the light it emits; the brighter one will be more attractive towards which the less bright ones are moved. Motivated by this process of bioluminescence, the FA updates the attractiveness and movement of any firefly in the population on every iteration. After running per-determined iterations, the firefly in the last population with the best fitness function yields the optimal solution. Decentralized decision-making structures in fireflies behavior and other natural species (Like ants, bees and birds), as examples of natural swarm intelligence, were inspiring to design of plentiful algorithms for solving complex issues such as optimization, multi-agent decision-making and robotics. The randomness characteristics of these algorithms avoid getting stuck in local optimums and helps to find the global solution for the problem [56]. Applying to several standard optimization problems validated that FA is more efficient than other meta-heuristic algorithms such as genetic algorithms (GA), simulated annealing (SA), particle swarm optimization (PSO) and differential evolution (DE) [57]. For this reason, we choose the FA for the problem of optimizing the voting weights of the MCS of this paper.

The FA is a parallel direct search technique which explores complex spaces to determine optimal answers for an optimization task. In this technique, a number of *d* parameters, whose optimal values are requested to be found, are encoded using some vectors called the

'locations of fireflies'. Each location is a nominee for the optimization problem, which is represented by a *d*-dimensional vector $\boldsymbol{x}=[x_1,x_2,\cdots,x_d]$. The collection of such vectors is called a *population*. The initial population of fireflies is generated at random locations in the search space. In the FA, a cost function $f(\boldsymbol{x})$ should be enhanced; the fitness of each firefly is characterized by this index. Three main assumptions are made in this algorithm: (*i*) all fireflies have the same sex; (*ii*) the attractiveness of each firefly is proportional to its brightness; and (*iii*) the brightness of each firefly at every location $I(\boldsymbol{x})$ is determined by the objective function of the problem at that location; *i.e.*, $I(\boldsymbol{x})\propto f(\boldsymbol{x})$. Yet, the attractiveness β is relative and should be judged by other fireflies. Here β is adjusted as a proportion to the Euclidean distance between the i^{th} and the j^{th} fireflies (represented by r_{ij}). In the simplest form, the light intensity $I(r)$ is approximated using the following Gaussian form of the distance [58].

$$I(r)=I_0.\mathrm{e}^{-\gamma r^2} \quad (1)$$

where $I(r)$ is the light intensity emitted by a firefly received to another firefly at the distance r, I_0 is the original light intensity and γ is the absorption coefficient. Since the attractiveness of a firefly is relative to the light intensity seen by adjacent fireflies, it can be defined as follows with similar definition in (1):

$$\beta(r)=\beta_0.\mathrm{e}^{-\gamma r^2} \quad (2)$$

A firefly *i* located at $\boldsymbol{x}_i$ moves toward a more attractive firefly *j* at $\boldsymbol{x}_j$ ($I_j>I_i$) as described below:

$$\boldsymbol{x}_i^{new}=\boldsymbol{x}_i+\beta_0.e^{-\gamma.r_{ij}^2}(\boldsymbol{x}_j-\boldsymbol{x}_i)+\alpha.\boldsymbol{\varepsilon}_i \quad (3)$$

Where the second term shows the attraction of the i^{th} firefly to the j^{th} firefly and the third one is a random term with a constant parameter α and a random vector $\boldsymbol{\varepsilon}_i$ with Gaussian or uniform distribution. The Pseudo code of the FA [58, 59] is brought here in order to the paper be self-contained:

Start
Determine the fitness function $f(\boldsymbol{x})$.
Generate the initial population of fireflies $\boldsymbol{x}_i$, $i=1,\ldots,n$
Define the brightness intensity I_i at $\boldsymbol{x}_i$ by $f(\boldsymbol{x}_i)$
Determine the absorption coefficient γ.
Iterate the following steps to exceed the termination criteria:
  For all fireflies: (*i*= 1 to *n*)
    For all fireflies: (*j*= 1 to *n*)
      If $I_j>I_i$:
        Move firefly *i* toward firefly *j* using (3).
      End If.
      Evaluate new solutions and update the brightness using (1).
    End For *j*.
  End For *i*.
  Rank fireflies to find the current best one.
End Iteration.
Return the best result
End.

3.5.2 The Classifiers Ensemble Using the Firefly Algorithm

Although the three classification models mentioned in Section 3.4 are solely able to perform the classification task, combination of their decisions may improve the overall performance of the recognition system. The decision making based on the classifiers mixture is executed according to some weights allocated to the classifiers. In fact, for each classifier picking each class, a specific voting weight is assigned. A large weight shows that the classifier choice about that class is more assured and reliable. These weights are taken to mix the outcomes of classifiers to attain the final judgement.

Assume that the number of classifiers and classes are *N* and *M* respectively. The aim is to find the voting weights so that an objective function –*i.e.*, the accuracy of the overall system- is maximized. The weights are enclosed in a real matrix *V* of size $N\times M$; in which *V*(*n*,*m*) is the weight of the n^{th} classifier for the m^{th} class. In this paper, the FA considers vectors of length $D=N.M$ as the locations of the fireflies. The entries of each location in the population are randomly initialized to cover the search space. Here, we selected the *F-measure* as the index of the reliability of each classifier; the higher the *F*-measure, the more reliable the outcome of that classifier. Symbolize the *F*-measure of the classifiers for the training set by F_n, n=1,…,*N*. Consider each sample in the training set. The mixture result about the class of this sample is found using the weights of the classes of the classifiers. The weight for the n^{th} classifier is equal to F_n. The score of a specific class for a sample '*s*' is:

$$g(c_m)=\sum_{n=1}^{N}F_n*Q(n,m),\ s.t.\ op(s,n)=c_m,\ m=1,\cdots,M \quad (4)$$

Here, $Q(n,m)$ symbolizes an entry of the firefly location associated with the voting weight of the n^{th} classifier for the m^{th} class. Furthermore, $op(s,n)$ characterizes the output class allocated by the n^{th} classifier to the sample *s*. In fact, only those classifiers that pick the m^{th} class are integrated in calculating $g(c_m)$. Finding m^* as $m^*=\arg Max_m\ g(c_m)$ yields the final conclusion for the class allocated to that sample in the training dataset. Then the accuracy of the classifier mixture for the training dataset is computed to be used as the cost function. FA attempts to determine an optimal location vector consisting of all entries $Q(n,m)$ that maximizes this cost function. Final results on the test dataset are reported using the classifier mixture associated with this best answer of the voting weights.

For each query pattern, its feature vector is entered to the three base classifiers to determine the class label. Then, the results are given to the ensemble system to make the final decision by combining outcomes according to the optimal voting weights. To clarify the ensemble approach, a simple artificial example is brought here with three base classifiers and two classes; thus the length of the location vector is 3*2=6. It is assumed that the optimal voting weights are found by the FA (shown in the third row of Table 1) along with the *F*-measure of each classifier on the training dataset. Now consider a query whose class label is needed to be determined. Suppose

that the results for the first, second and third classifiers are the class labels A, B and B respectively. Then the score of class A is computed via multiplying the *F*-measure of the first classifier (0.98) in the weight of the first classifier/class_A (0.9). Similarly the score of class B is found by summing up two values: *i*) the second classifier *F*-measure (0.96) in the weight of the second classifier/class_B (0.2); and *ii*) the *F*-measure of the third classifier (0.9) in the weight of the third classifier/class_ B (0.3). Finally, the score of class B (0.466) is compared to that of class A (0.882). The class with the higher score (hear, the class A) wins, see Table 2.

Table 1. Voting weights and F-measure for an illustrative example

Classifiers	**First classifier**		**Second classifier**		**Third classifier**	
Classes	Class A	Class B	Class A	Class B	Class A	Class B
Class/classifier voting weights	0.9	0.3	0.7	0.2	0.8	0.3
F-measure of classifiers	0.98		0.96		0.90	

Table 2. Computation of the scores of the classes

First Classifier	Decision: Class A	Classifier score: 0.98*0.9=0.882
Second Classifier	Decision: Class B	Classifier score: 0.96*0.2=0.196
Third Classifier	Decision: Class B	Classifier score: 0.90*0.3=0.270
Score of the classes	Class A: **0.882**	Class B: 0.270+0.196=0.466

4. Simulation Details

Appropriate tests must be performed to answer to this question that "whether using the mentioned feature selection and classifiers combination methods can improve the efficiency of handwritten digits recognition system, or not?". In this paper, the HODA dataset [39] is used to evaluate the proposed system. This dataset contains 80000 Farsi handwritten digit images with the resolution of around 200 dpi (dots per inch). Fig. 10 shows some examples of the HODA dataset. The proposed algorithm is run under the MATLAB R2014a programming environment on a PC equipped with 3.2 GHZ CPU and 8 GB RAM memory.

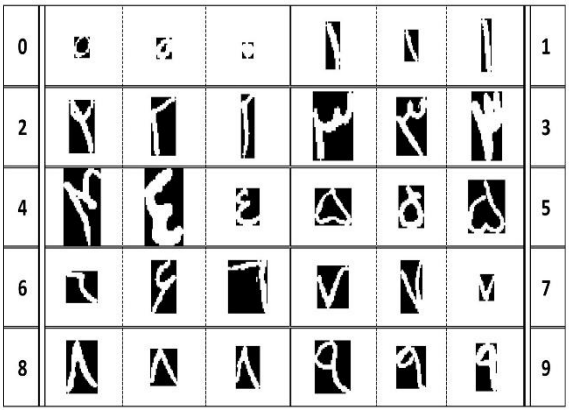


Fig. 10. Examples of the HODA dataset. Three example for each digit [39].

To assess the performance of our proposed method we outline the following experiments:

– Experiment 1: all features being fed into the distinct classifiers (no feature selection, no classifier mixture);
– Experiment 2: selected features entered into the individual classifiers (i.e., feature selection, yet no classifier mixture);
– Experiment 3: the classifiers mixture is applied to patterns with all features (no feature selection, but classifiers mixture).
– Experiment 4: The classifiers trained by the chosen features are mixed with the help of a voting weight methodology. These weights are determined by solving an optimization problem using the FA. The final conclusion is made based on the maximum score. (i.e., feature selection and classifier mixture). This experiment characterizes our proposed method.

The GA is employed to select the most discriminative features. The selection is carried out by the Roulette wheel method, and single-point crossover with the probability of 0.7, mutation rate of 0.2 and penalty coefficient of 0.5 are used in this paper. The population size is 30 and the number of generations is 50. The base classifiers in this paper are (1) the MLP with 20 neurons in one hidden layer, (2) the CART decision tree, and (3) the *k*- nearest neighbor by setting $k = 3$. In addition, the values of the parameters of the FA algorithm for finding optimal weights of classifiers combination are $\alpha = 0.02$, $\beta_0 = 2$, $\gamma = 1$; the values of these parameters are generally taken from [57]. Moreover, the population size is 20 and the number of iterations is 50.

To evaluate the system performance, the accuracy, precision, recall and the *F*-measure are used. These indices are defined according to the *TP*, *TN*, *FP* and *FN* values [60], as follows:

$$\mathrm{Pr}ecision(\mathrm{Pr}e.) = TP/(TP+FP)*100 \quad (5)$$

$$\mathrm{Re}call(\mathrm{Re}c.) = TP/(TP+FN)*100 \quad (6)$$

$$F-measure = (2*\mathrm{Pr}e.*\mathrm{Re}c.)/(\mathrm{Pr}e.+\mathrm{Re}c.)*100 \quad (7)$$

$$Accuracy(Acc.) = (TP+TN)/(TP+TN+FP+FN)*100 \quad (8)$$

Where *TN* is the number of negative truly recognized as negative; *TP* is the number of positive truly recognized as positive; *FN*, positive falsely recognized as negative; and *FP*, negative falsely recognized as positive. *F*-measure, defined in the interval [0,1], is the harmonic mean of the precision and recall and considers both rates in a single index. Values close to 1 is desired for the *F*-measure of a classification system. Furthermore, accuracy is the proportion of correctly classified samples from the total number of samples. To evaluate the results, the *k*-fold cross validation [61] is carried out. In this scheme, the total number of data is divided into *k* subsets. In each round, one subset is left out for the test and the classifier is trained using the rest. This process is repeated so that each subset is left out once. Lastly, the average of the cost function of all rounds is calculated to develop a more accurate estimate of the system prediction performance. In this paper *k* is set to 4 and the

dataset is divided into four subsets, each containing 20000 samples. Accordingly in each round, 60000 samples are devoted to the training and the rest are left for the testing.

5. Simulation Results

Table 3 shows the performance measures for four mentioned experiments. The GA selected eleven most dominant features among the 25 features extracted from each image. The optimal feature subset includes 1 feature for #BP, 4 features from HCCV, 2 features from VCCV and 4 features from 8DCC. As shown in Table 3, almost all performance measures of the proposed method are higher than those of other experiments which demonstrate the advantage of selecting the most dominant features and also combining the classifiers results. Fig. 11 shows the confusion matrix for 10 digits when the proposed approach is applied to the HODA dataset. It is obvious from this figure that there are some digits more frequently misclassified. The major mistakes occurred for discriminating the digits '2', '3', '4' and for discriminating the digits '0' and '5'. This is caused by the fact that they are fairly similar in shape.

Table 4 shows some of the misclassifications of the proposed system, which are mainly due to poor quality of images or bad handwriting. Table 5 compares the proposed system with some other methods applied to the Farsi handwritten digit recognition problem with respect to the accuracy measure. The high performance of our method is due to selecting the most dominant features and utilizing diverse classifiers in the combination, and also because of applying the FA for finding the best voting weights in the classifiers ensemble.

6. Discussion

The method proposed in this paper presents a hybrid multi-procedure system for recognizing Farsi handwritten digits. Once the image pre-processing is performed, 25 features are extracted among which eleven ones selected by the two-objective GA are given to three base classifiers (*i.e.*, DT, ANN and KNN). These classifiers are widely used in recognition applications. Nevertheless, the existing literature shows that each classification model may outperform the others in different situations; which points to this fact that each technique has its shortcomings. This is the main motivation of this paper to integrate different classifiers in order to improve the accuracy and other performance indices. This combination is accomplished by assigning some voting weights to the contributor classifiers. The appropriate selection of these weights is critical to attain a more accurate recognition organization. The mixture with different weight values might return very different consequences. An unsuitable setting may lead to poor and erroneous classification algorithm, even worse than the distinct classifiers. The FA approach finds the weights of the classifiers according to their effectiveness so that a classification model with higher performance indices has a larger weight and thus a greater role to play and more discriminative information. Therefore, the inadequacy of each classifier is compensated by adequacy of other classifiers to obtain better classification measures. Our results show that the combination of classifiers through the FA ensemble technique is accurate and satisfactory and yields the classification accuracy of 98.88%. This rate is higher that the base classifiers acting solo.

From the results of Experiments 1 and 2 in Table 3, in which each classifier individually works, it can be seen that the ANN, outperforms the other two classifiers on both experiments with a best classification accuracy of 97.88%. The accuracy results of the KNN, as a lazy learner, are slightly smaller than those of the ANN. However, this classifier has an advantage of being computationally less expensive than ANN since it has no training phase. Although the DT performed worse than the other two classifiers, as shown in Table 3, still it is used to help other contributors in the hybrid system to increase the overall accuracy. It should be noticed that the accuracy values in Experiments 2 and 4 are greater than those of Experiments 1 and 3. The reason is that in the former experiments the most discriminative features are selected; while in the latter experiments all features are used which may reduce the generalization characteristics of the classification system or some features might be misleading. The feature selection technique reduces the computational cost and concurrently increase the classification accuracy. It is noteworthy to point out that simultaneously considering two performance objective functions -i.e., the *F*-measure and the cardinality of the selected features subset- for the feature selection problem is of great use to benefit from informative data.

Table 3. Comparison of performance measures for different experiments using 4-fold cross validation. The average values are shown in the table. Here, the following abbreviations are used: 'Acc.': Accuracy, 'Pre.': Precision, 'Rec.': Recall, 'Fmea.': F-measure. 'TrT': training time and 'TsT': testing time (Second). Also, 'PrM': The proposed method.

Experiment\Measure		Acc.	Pre.	Rec.	*F*mea.	TRT	TST
Experiment 1	*ANN*	97.63	96.61	89.15	92.73	0.136	0.0035
	DT	92.55	90.38	88.31	89.33	0.117	0.0031
	KNN	93.91	91.77	90.36	91.06	0.058	0.0010
Experiment 2	*ANN*	97.88	96.13	92.15	94.10	0.063	0.0023
	DT	93.18	91.18	89.36	90.26	0.049	0.0018
	KNN	94.78	90.78	91.19	90.98	0.021	0.0016
Experiment 3		98.02	97.11	94.23	95.65	2.835	0.0367
Experiment 4 (PrM)		**98.88**	97.52	93.75	95.60	0.424	0.0218

	1	2	3	4	5	6	7	8	9	0
1	98.43	0.64	0	0.07	0	0.16	0.29	0.18	0.04	0.3
2	0.42	98.48	0.66	0.29	0	0	0.05	0.06	0.09	0.28
3	0.06	0.52	98.62	1.21	0.13	0	0.05	0.04	0	0.06
4	0	0.36	0.47	98.14	0.16	0	0	0.05	0	0
5	0	0	0.16	0.17	98.59	0.04	0.05	0	0	0.57
6	0.21	0	0	0.08	0	99.57	0.1	0	0.29	0.09
7	0.13	0	0	0	0	0.16	99.31	0	0	0.22
8	0.16	0	0.03	0.04	0	0	0	99.67	0	0
9	0	0	0	0	0	0.07	0	0	99.51	0
0	0.59	0	0.06	0	1.12	0	0.15	0	0.07	98.48

Fig. 11 Confusion matrix for the 10-class problem of the proposed method on the HODA dataset (%). Columns show the input digits, while rows present the recognition results.

Table 4. Some examples of misclassifications of the proposed system

Handwritten digits										
True digit	0	1	2	3	4	5	6	7	8	9
Recognized digit	1	0	1	2	3	0	7	1	1	6

Table 5. Comparison of the proposed approach with some related methods based on the accuracy measure of the test data. Here, the following abbreviations are used: 'Mtd.': Method, 'Acc.': Accuracy, 'TrD': training data, 'TsD': testing data. Also, 'PrM': The proposed method.

Mtd.	TrD	TsD	Acc.	Mtd.	TrD	TsD	Acc.
[31]	7390	3035	94.14	[12]	6000	2000	97.10
[30]	4979	3939	99.57	[14]	60000	20000	98.84
[33]	6000	4000	97.01	[32]	1000	5000	97.02
[10]	60000	20000	98.71	[34]	60000	20000	99.31
[13]	6000	2000	95.30	**PrM**	60000	20000	**98.88**

In Experiment 1, ANN has the highest accuracy while its recall is less than that of the KNN. Similar conditions exist in some other experiments and models in Table 3. The main performance measure for evaluating the recognition systems of the paper is the accuracy. Handwritten digit recognition is a nonlinear and complicated problem. Thus, it should not expect that the behaviors and rates have a constant harmony. Another example contains the recall and *F*-measure of Experiment 4 which are less than those of Experiment 3, while converse condition holds for their accuracies. This is not unconnected to the fact that the mathematical relation of the *F*-measure contains the recall index. Hence when recall is small, the *F*-measure will also be small. Nonetheless, the system in Experiment 4 achieves greater accuracy with smaller number of features (eleven) while the system of Experiment 3 uses 25 features to obtain the performance indices mentioned in Table 3.

The running time of the experiments is also stated in Table 3. Experiments were executed on a PC (3.2 GHZ CPU, 8 GB RAM memory). The running time rests on the size of the training dataset, number of features to be given to the classifiers, number of the classes, etc. The testing time is very less in comparison to the training time. It can be seen from Table 3 that the systems based on the feature selection (Experiments 2 & 4) are faster than those use all features (Experiment 1 & 3), when examined in similar circumstances. It should be noted that when several classifiers are mixed using any weighted mixture procedure, the training time complication rises because of several runs wanted for discovering ideal voting weights. However, when these weights are established, the testing time comprises the time needed for any base classifier to deliver its outcome accompanied by the time for a simple decision making based on the weighted results. This testing time is trivial as shown in the last column of Table 3. Hence, using this collaborative attitude does not have much time complexity. On the other hand, the performance measures of the mixture methods are higher. The results in Table 3 validate the advantage of selecting the most dominant features in conjunction with reasonable mixing the classifiers results.

A number of researches have been reported on the recognition of Farsi handwritten digits, some of them are reported in Table 5. It deserves to be noted that the methods in Table 5 were assessed on different datasets with different image sizes and resolutions. The method of Soltanzadeh and Rahmati [30] is evaluated on their own dataset including 8918 high resolution (300 dpi) samples with a feature vector of length 257. They removed the incorrectly or unusually written digits to obtain a dataset with well-written numerals. In this paper, the HODA dataset is used which contains 80,000 samples with the resolution of 200 dpi. In work of Alaei et al. [10], Rashnodi et al. [14] and Khorashadizadeh and Latif [34] with the dataset same as that of this paper respectively 196, 154 and 164 features are used for the classification. Although our method achieved the accuracy of 98.88% which is a little smaller than some of the results mentioned in Table 5, it uses only eleven features while others utilize many more ones.

7. Conclusion

This paper proposed a hybrid multi-step procedure for the recognition of Farsi handwritten digits. First a set of pre-processing operations were performed on each digit image to make it prepared for the next steps. Following this, multiple structural and statistical features were extracted leading to a feature space with large dimensionality. For this reason, the multi-objective GA selected the most discriminative features to being fed to a decision tree, an artificial neural network and a *k*-nearest neighbor classifier. At the last step, the final decision about the digit class label was made by a classifiers ensemble system whose voting weights were found by the firefly evolutionary algorithm. The performance of the individual and combined classifiers were evaluated on the standard HODA dataset and compared with other existing methods from the literature. The proposed approach achieved Farsi handwritten digit classification with acceptable accuracy. The results of this research support the idea that the best classification performance could be obtained when the results of individual classifiers are combined into a single decision made by an ensemble classifier. Considering that different approaches suggested for this pattern recognition problem did not use the similar dataset, the precise comparison of the presented method with others is not possible. Though, due to the high recognition rate that is touched by the technique proposed in this paper, we can say that, to best of our knowledge, this system is at least one of the best procedures proposed until now for the recognition of Farsi handwritten digits.

For future work, the proposed method can be applied to the recognition of handwritten digits and characters in different languages and styles. In addition, using other features and different feature selection techniques (e.g., student's t-test or PCA) coupled with other base classifiers (such as SVM) can be considered for generating and selecting dominant features and classifying the handwritten digits. More broadly, research is also needed to determine the effectiveness of other ensemble techniques (for example multi-objective PSO or Cuckoo Search) when dealing with different digit images. The main imperfection of this technique is that the computational complexity of the ensemble technique is

naturally higher than each sole classifier since it needs all classifiers to be run and give their results to the ensemble classifier to make the final decision. Due to the complexity and high applicability of handwritten recognition, the truthful classification of digit patterns is crucial and of great importance in several technical and non-technical tasks.

References

[1] C.-L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: benchmarking of state-of-the-art techniques," Pattern recognition, vol. 36, pp. 2271-2285, 2003.

[2] R. Al-Jawfi, "Handwriting Arabic character recognition LeNet using neural network," Int. Arab J. Inf. Technol., vol. 6, pp. 304-309, 2009.

[3] M. N. Ayyaz, I. Javed, and W. Mahmood, "Handwritten character recognition using multiclass svm classification with hybrid feature extraction," Pakistan Journal of Engineering and Applied Sciences, 2016.

[4] A. K. A. Hassan, "Arabic (Indian) Handwritten Digits Recognition Using Multi feature and KNN Classifier," Journal of University of Babylon, vol. 26, pp. 10-17, 2018.

[5] B. Savas and L. Eldén, "Handwritten digit classification using higher order singular value decomposition," Pattern recognition, vol. 40, pp. 993-1003, 2007.

[6] Y. Chen, Z. Xu, S. Cai, Y. Lang, and C.-C. J. Kuo, "A Saak Transform Approach to Efficient, Scalable and Robust Handwritten Digits Recognition," in 2018 Picture Coding Symposium (PCS), 2018, pp. 174-178.

[7] W.-S. Lu, "Handwritten digits recognition using PCA of histogram of oriented gradient," in Communications, Computers and Signal Processing (PACRIM), 2017 IEEE Pacific Rim Conference on, 2017, pp. 1-5.

[8] A. Boukharouba and A. Bennia, "Novel feature extraction technique for the recognition of handwritten digits," Applied Computing and Informatics, vol. 13, pp. 19-26, 2017.

[9] J. Qiao, G. Wang, W. Li, and M. Chen, "An adaptive deep Q-learning strategy for handwritten digit recognition," Neural Networks, 2018.

[10] A. Alaei, U. Pal, and P. Nagabhushan, "Using modified contour features and SVM based classifier for the recognition of Persian/Arabic handwritten numerals," in Advances in Pattern Recognition, 2009. ICAPR'09. Seventh International Conference on, 2009, pp. 391-394.

[11] M. Nahvi, K. Kiaee, and R. Ebrahimpour, "improvement the feature extraction method of Gradient based on the discrete cosine transform for recognizing Farsi handwritten digits," presented at the 18th Iranian Conference on Electrical Engineering, Isfahan, Iran, 2010.

[12] M. J. Abdi and H. Salimi, "Farsi handwriting recognition with mixture of RBF experts based on particle swarm optimization," International Journal of Information Science and Computer Mathematics, vol. 2, pp. 129-136, 2010.

[13] R. Ebrahimpour, A. Esmkhani, and S. Faridi, "Farsi handwritten digit recognition based on mixture of RBF experts," IEICE Electronics Express, vol. 7, pp. 1014-1019, 2010.

[14] O. Rashnodi, H. Sajedi, M. Abadeh, A. Elci, M. Munot, M. Joshi, et al., "Persian Handwritten Digit Recognition Using Support Vector Machines," International Journal of Computer Applications, vol. 29, pp. 1-6, 2011.

[15] D. Ghosh, T. Dube, and A. Shivaprasad, "Script recognition—a review," IEEE Transactions on pattern analysis and machine intelligence, vol. 32, pp. 2142-2161, 2010.

[16] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE transactions on evolutionary computation, vol. 6, pp. 182-197, 2002.

[17] U. K. Sikdar, A. Ekbal, and S. Saha, "MODE: multiobjective differential evolution for feature selection and classifier ensemble," Soft Computing, vol. 19, pp. 3529-3549, 2015.

[18] C. Chow, "Statistical independence and threshold functions," IEEE Transactions on Electronic Computers, pp. 66-68, 1965.

[19] T. G. Dietterich, "Ensemble methods in machine learning," in International workshop on multiple classifier systems, 2000, pp. 1-15.

[20] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," Machine learning, vol. 51, pp. 181-207, 2003.

[21] B. V. Dasarathy and B. V. Sheela, "A composite classifier system design: Concepts and methodology," Proceedings of the IEEE, vol. 67, pp. 708-713, 1979.

[22] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," Information Fusion, vol. 16, pp. 3-17, 2014.

[23] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on, 1995, pp. 39-43.

[24] X.-S. Yang, "Firefly algorithm, stochastic test functions and design optimisation," arXiv preprint arXiv:1003.1409, 2010.

[25] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "GSA: a gravitational search algorithm," Information sciences, vol. 179, pp. 2232-2248, 2009.

[26] H. Duan and P. Qiao, "Pigeon-inspired optimization: a new swarm intelligence optimizer for air robot path planning," International Journal of Intelligent Computing and Cybernetics, vol. 7, pp. 24-37, 2014.

[27] S. Mirjalili, S. M. Mirjalili, and A. Hatamlou, "Multi-verse optimizer: a nature-inspired algorithm for global optimization," Neural Computing and Applications, vol. 27, pp. 495-513, 2016.

[28] I. Fister, I. Fister Jr, X.-S. Yang, and J. Brest, "A comprehensive review of firefly algorithms," Swarm and Evolutionary Computation, vol. 13, pp. 34-46, 2013.

[29] X.-S. Yang and X. He, "Firefly algorithm: recent advances and applications," arXiv preprint arXiv:1308.3898, 2013.

[30] H. Soltanzadeh and M. Rahmati, "Recognition of Persian handwritten digits using image profiles of multiple orientations," Pattern Recognition Letters, vol. 25, pp. 1569-1576, 2004.

[31] J. Sadri, C. Y. Suen, and T. D. Bui, "Application of support vector machines for recognition of handwritten Arabic/Persian digits," in Proceedings of Second Iranian Conference on Machine Vision and Image Processing, 2003, pp. 300-307.

[32] H. Salimi and D. Giveki, "Farsi/Arabic handwritten digit recognition based on ensemble of SVD classifiers and reliable multi-phase PSO combination rule," International Journal on Document Analysis and Recognition (IJDAR), vol. 16, pp. 371-386, 2013.

[33] M. Ziaratban, K. Faez, and F. Faradji, "Language-based feature extraction using template-matching in Farsi/Arabic

handwritten numeral recognition," in Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on, 2007, pp. 297-301.
[34] S. Khorashadizadeh and A. Latif, "Arabic/Farsi Handwritten Digit Recognition usin Histogra of Oriented Gradient and Chain Code Histogram," International Arab Journal of Information Technology (IAJIT), vol. 13, 2016.
[35] R. Safdari and M.-S. Moin, "A hierarchical feature learning for isolated Farsi handwritten digit recognition using sparse autoencoder," in Artificial Intelligence and Robotics (IRANOPEN), 2016, 2016, pp. 67-71.
[36] R. Hajizadeh, A. Aghagolzadeh, and M. Ezoji, "Fusion of LLE and stochastic LEM for Persian handwritten digits recognition," International Journal on Document Analysis and Recognition (IJDAR), vol. 21, pp. 109-122, 2018.
[37] Z. Sadeghi and A. Testolin, "Learning representation hierarchies by sharing visual features: a computational investigation of Persian character recognition with unsupervised deep learning," Cognitive processing, vol. 18, pp. 273-284, 2017.
[38] Y. Zamani, Y. Souri, H. Rashidi, and S. Kasaei, "Persian handwritten digit recognition by random forest and convolutional neural networks," in Machine Vision and Image Processing (MVIP), 2015 9th Iranian Conference on, 2015, pp. 37-40.
[39] H. Khosravi and E. Kabir, "Introducing a very large dataset of handwritten Farsi digits and a study on their varieties," Pattern recognition letters, vol. 28, pp. 1133-1141, 2007.
[40] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern classification: John Wiley & Sons, 2012.
[41] S. Askari, M. KHarashadizadeh, and J. Sadri, "A new method for Recognizing Farsi handwritten numbers based on Pre-Classification," presented at the First Conference on Pattern Recognition and Image Analysis, Birjand, Iran, 2012.
[42] D. Deodhare, N. R. Suri, and R. Amit, "Preprocessing and Image Enhancement Algorithms for a Form-based Intelligent Character Recognition System," IJCSA, vol. 2, pp. 131-144, 2005.
[43] F. K. Zeyaratban M, Mozzafari S, Azvaji M. , "Presenting a New Structural Method Based on Partitioning Thinned Image for Recognition of Handwritten Farsi-Arabic Digits," in Third Conference on Machine Vision, Image Processing and Applications, 2005, pp. 76-82.
[44] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on, 2015, pp. 1200-1205.
[45] C. Gunavathi and K. Premalatha, "Performance analysis of genetic algorithm with kNN and SVM for feature selection in tumor classification," Int J Comput Electr Autom Control Inf Eng, vol. 8, pp. 1490-7, 2014.
[46] A. Padma and R. Sukanesh, "A wavelet based automatic segmentation of brain tumor in CT images using optimal statistical texture features," International Journal of Image Processing, vol. 5, pp. 552-563, 2011.
[47] Y. Marinakis, G. Dounias, and J. Jantzen, "Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification," Computers in Biology and Medicine, vol. 39, pp. 69-78, 2009.
[48] O. Vechtomova, "Introduction to Information Retrieval Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze (Stanford University, Yahoo! Research, and University of Stuttgart) Cambridge: Cambridge University Press, 2008, xxi+ 482 pp; hardbound, ISBN 978-0-521-86571-5, $60.00," ed: MIT Press, 2009.
[49] M. Razavi and E. Kabir, "On-line Recognition of Farsi separate letters using the neural network. ," presented at the Third conference on machine vision and image processing, Tehran, Iran, 2004.
[50] L. A. Breslow and D. W. Aha, "Simplifying decision trees: A survey," The Knowledge Engineering Review, vol. 12, pp. 1-40, 1997.
[51] S. Singh and P. Gupta, "Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey," International Journal of Advanced Information Science and Technology (IJAIST), vol. 27, pp. 97-103, 2014.
[52] C. C. Aggarwal and S. Y. Philip, "A general survey of privacy-preserving data mining models and algorithms," in Privacy-preserving data mining, ed: Springer, 2008, pp. 11-52.
[53] M. van Gerven and S. Bohte, Artificial neural networks as models of neural information processing: Frontiers Media SA, 2018.
[54] J. W. Shavlik, R. J. Mooney, and G. G. Towell, "Symbolic and neural learning algorithms: An experimental comparison," Machine learning, vol. 6, pp. 111-143, 1991.
[55] D. H. Wolpert, "The supervised learning no-free-lunch theorems," in Soft computing and industry, ed: Springer, 2002, pp. 25-42.
[56] X. Zhang, Y. Tian, and Y. Jin, "A knee point-driven evolutionary algorithm for many-objective optimization," IEEE Transactions on Evolutionary Computation, vol. 19, pp. 761-776, 2015.
[57] A. H. Gandomi, X.-S. Yang, and A. H. Alavi, "Mixed variable structural optimization using firefly algorithm," Computers & Structures, vol. 89, pp. 2325-2336, 2011.
[58] O. Bozorg-Haddad, M. Solgi, and H. A. Loáiciga, Meta-heuristic and evolutionary algorithms for engineering optimization vol. 294: John Wiley & Sons, 2017.
[59] X.-S. Yang, "Firefly algorithms for multimodal optimization," in International symposium on stochastic algorithms, 2009, pp. 169-178.
[60] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical machine learning tools and techniques: Morgan Kaufmann, 2016.
[61] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in Ijcai, 1995, pp. 1137-1145.

Hamed Agahi received B.Sc. M.Sc. and Ph.D. degrees in Electrical Engineering from University of Shiraz, Amirkabir University of Technology and University of Tehran, Iran, in 2005, 2008 and 2013, respectively. From 2009, he was with the Islamic Azad University, Shiraz Branch, Shiraz, Iran. His research interests include pattern recognition, image and signal processing, and fault detection and diagnosis applications.

Azar Mahmoodzadeh received B.Sc., M.Sc. and Ph.D. degrees in Electrical Engineering from University of Shiraz, University of Shahed and University of Yazd, Iran, in 2005, 2008 and 2013, respectively. From 2009, she was with the Islamic Azad University, Shiraz Branch, Shiraz, Iran. Her research interests include pattern recognition and image and signal processing.

Marzieh Salehi received her MSc in telecommunications engineering from Islamic Azad University, Shiraz Branch, Shiraz, Iran. Her research interests are pattern recognition and Image processing for practical applications.

The Influence of ERP Usage on Organizational Learning: An Empirical Investigation

Faisal Aburub*
MIS Department, University of Petra, Amman, Jordan
faburub@uop.edu.jo

Received: 24/Feb/2018 Revised: 22/Aug/2018 Accepted: 16/Sep/2018

Abstract

A number of different hotels have been seen to direct significant investment towards Enterprise Recourse Planning (ERP) systems with the aim of securing sound levels of organizational learning. As a strategic instrument, organizational learning has been recommended in the modern management arena as potentially able to achieve a competitive edge and as stabilizing the success of businesses. Learning, as an aim, is not only able to improve the skillset and knowledge of employees, but also achieving organizational growth and development, whilst also helping to build a dynamic learning organization. Organizational learning is especially important in modern-day firms, when staff might choose to leave or change their role owing to the view that knowledge-sharing could be detrimental to their own success. The present work seeks to examine the impact of ERP usage on organizational learning. A new research model has been presented, this model has been empirically investigated in the Jordanian hotel industry. 350 questionnaires were distributed across a total of 350 hotels. 317 questionnaires were returned. Structural equation modeling (AMOS 18) was used to analyze the data. The findings from the empirical findings emphasize that ERP usage has significant impact on organizational learning. In line with the study findings, various aspects of organizational learning, such as continuous learning, system perspective, openness and experimentation and transfer and integration are recognized as able to best encourage the use of ERP. Suggestions for future work and discussion on research limitations are also discussed.

Keywords: ERP Usage; Organizational Learning; Organizational Performance.

1. Introduction

With the significant developments being witnessed in IT, it has become a very important part of day-to-day life [1], [2]. In the present competitive and fast-paced competitive business setting, there is a need for organizations to ensure they are able to efficiently exploit present IT infrastructure [3]. With businesses demonstrating continued investment in ERP systems, such systems are expected to encourage and drive performance, whilst creating value in what is becoming a more and more competitive and aggressive business setting. Knowledge-sharing has been improved as a result of ERP, with the flow of information and communication made quicker. Furthermore, ERP systems are continuing to develop, thus presenting a number of challenges for people, enabling them to improve, learn and adapt. Moreover, ERP systems have also affected businesses and their operations, with Ağaoğlu et al. [1] recognizing that the majority of works carried out in the field of ERP have not been successful in presenting a clear overview of the situation as it stands, predominantly owing to the fact that studies in this arena focus on a limited number of advanced countries, including the UK and the USA.

A number of works suggest that the adoption of an ERP system achieve improvements across the operational performance of organizations [4], [5]. In this regard, BolíVar et al. [5] highlight ERP systems as facilitating organizations in the streamlining, integration and standardization of their process flows and data. Organizations make well-considered changes to their installations over time and leverage ERP information so as to achieve improvements in various arenas, including order management and inventory management. It is common for organizations to incorporate modules extending ERP system use beyond the organization, with the inclusion of customers and suppliers. This continuous approach to fine-tuning, developing and extending, and stabilizing ERP systems has been recognized as further enhancing the performance of organizations.

As a strategic instrument, organizational learning has been recommended in the modern management arena as potentially able to achieve a competitive edge and as stabilizing the success of businesses. Learning, as an aim, is not only able to improve the skillset and knowledge of employees, but also achieving organizational growth and development, whilst also helping to build a dynamic learning organization [6]. In this vein, it is noted by Brown and Suzan [7] that learning is an important factor in any firm owing to the fact it facilitates competitive edge to be both achieved and extend-ed. Essentially, learning may act as a means of creating and developing a number of different capabilities in the organization, thus encouraging businesses to achieve continuous improvement as opposed to focusing on different types of knowledge. Organizational learning needs to be sufficient in generating, acquiring, transferring and

* Corresponding Author

integrating new knowledge, in addition to making changes to existing behaviors so as to emphasize new knowledge in mind of enhancing performance [4].

A number of works have been carried out in order to examine the effects of the adoption of ERP in line with business performance [8], [9], [10], [11]. Furthermore, organizational learning is recognized as a fundamental approach to enhancing the business in various fields [11]. In the literature, there is a few studies investigate the role of ERP usage on achieving organizational learning. As such, this work seeks to analyze the link between ERP use and organizational learning in the specific arena of the hotel industry in the Middle East, and specifically in Jordan. This investigation is focused on exploring and accordingly providing insight into how an empirical study combining the two paradigms is able to relate and extend theories in such arenas.

Accordingly, the main purpose of this research is to explore the impact of ERP usage on organizational learning. This research has been investigated empirically using hotel sector in Jordan. The paper is structured as follows: the next section provides literature review, followed by research hypotheses. Fourth section presents research model followed by hypotheses testing. Last section presents discussion and conclusion section.

2. Literature Review

The ERP system may be recognized as a number of different integrated software modules and a central database, which encourage a firm to manage the effective and efficient utilization of resources, i.e. financial, human resources and materials, through the automation and incorporation of business processes, data-sharing across the firm, and facilitating information access in a real-time setting [11], [12]. ERP aims to provide a number of advantages across a spectrum, including enhance quality, enhanced efficiency, profitability through enhanced capability, and productivity, with the addition of accurate communication and timely information. Businesses direct investment towards ERP systems in mind of attaining a number of advantages, which might be witnessed through enhanced organizational productivity, including lower costs, efficient communication across functional areas, and shortened lead times [13]. Previous literature has sought to garner an understanding into the motivational factors underpinning ERP benefits. In this regard, five dimensions of ERP benefits have been presented in the work of Jerez-Gomez et al. [14], including operational, managerial, strategic, IT infrastructure and organizational, with ERP benefit concluded as an ongoing process with benefits achieved at different rates across various core processes. Moreover, it was determined in the work of Hair et al. [13] that overall ERP advantages was mediated as a result of various business-related factors. A number of other ERP system-related benefits include its complete integration across all organizational processes, decreases in data entry volumes, the potential of technology to be upgraded, portability to other systems, the application of best practices, and adaptability [15].

Furthermore, ERP is viewed as being a strategic instrument with the capacity to integrate, synchronize and streamline business processes and data into one individual system so as to achieve a competitive advantage in what may be an uncertain business environment [16].

ERP is recognized as being one of the most recent technologies of which businesses may be able to take advantage [17]. ERP systems may be adapted to a certain limit in line with the particular requirements of the firm. In this vein, ERP was recognized as one of the most fundamental of developments in the corporate utilization of technology during the 1990s [18], [19]. Nonetheless, a number of different ERP projects have proven unsuccessful and therefore unable to attain the outcomes sought. Due to the costs of ERP implementation projects being significant, it is fundamental that firms achieve project success and use it efficiently so as to attain benefits as quickly as possible.

It is clear that ERP systems provide a number of organizational advantages. However, the effects of ERP usage on organizational learning are not clear. This study aims to examine this field.

Organizational learning is especially important in modern-day firms, when staff might choose to leave or change their role owing to the view that knowledge-sharing could be detrimental to their own success. In the view of Chao [20], businesses commonly expect knowledge-creation and learning to be natural for individuals, with knowledge shared in ways to encourage and motivate learning in groups and across the firm. Organizational learning is a process by which organizations create routines and knowledge that guide its future behavior and behavior through encoding inferences from experience [20],[21]. Organizational learning is fundamental across the process of garnering a competitive edge [22]. Accordingly, there is a need to analyze the effects on organizational learning and achieving a competitive edge, particularly in regards IT on organizational learning. Furthermore, the capacity of a firm to learn faster than the competition can be the only sustainable competitive edge; at the same time, ICT are known to facilitate improved performance across business activities whilst also enhancing organizational learning and the quality of such.

In the view of Ağaoğlu et al. [1], the business demonstrates learning in two different ways, namely through existing firm members or otherwise through new members who possess knowledge unknown to the firm. Organizational learning can be improved upon through the development of existing skills or new ones. In this regard, it is viewed as fundamental that attributes with the potential to enable firms to understand, possess and use knowledge are actively sought out [23] owing to the fact that organizational learning is not a fundamental cognitive activity.

Organizational learning may be recognized as the potential of a firm to implement sound management practices, procedures, policies and structure that enable and encourage learning. Learning is viewed as critical in a

firm owing to its ability to facilitate the generation and development of a sustainable competitive edge. Essentially, learning can also act as a way of generating and extending a wide range of business capabilities, thus encouraging organizations to achieve ongoing improvement as opposed to focusing on particular forms of knowledge [3]. In the view of Levitt et al. [24], organizational learning may be recognized as the process through which organizations learn, creating chance so as to ensure outcomes can be maintained and improved. In this regard, five different factors of organizational learning are presented by Madanhire et al. [25], including experimentation, risk acceptance, interaction with the environment, dialogue, and participation in decision-making. Experimentation may be viewed as the degree to which suggestions and new ideas are taken into account. Risk-taking is considered to be the degree of tolerance to errors and uncertainty. Interaction with the external setting relates to the links with the external environment. Dialogue is the collective analysis of assumptions, certainties and processes. Participative decision-making may be linked with the extent of power possessed by employees in regards decision-making.

In the view of Madapusi and Derrick [26], organizational learning encompasses continuous changes in the behavior and cognition of employees and management. In an organization, individual members facilitate learning, with the individual processes then embedded in business-related functions. Accordingly, organizational learning occurs through individuals' social processes, encompassing the creation, retention and transference of knowledge. As a whole, individuals improve the ability of the firm to learn, meaning the firm needs to be open to their efforts and accordingly applying the most suitable mechanisms so as to facilitate, support and reward learning [27]. Lee [23] present the view that learning may be witnessed across three levels, namely group, individual and firm. The concept suggest that change is witnessed across all levels of learning, with change witnessed through the form of new routines and practices that facilitate and further support the capacity to utilize learning so as to enhance performance. As such, organizational learning may be seen to encompass seven individual but nonetheless linked aspects at the individual, group and organizational levels, namely continuous learning, team-based learning, inquiry and dialogue, embedded system, empowerment, system connection, and strategic leadership [28].

A number of actions have been presented by McGill et al. [29] as ensuring learning capability, including experimentation, continuous improvement, teamwork and group problem-solving. Further, Bhatti [4] have devised a tool centered on the measurement of organizational learning, including various elements, namely managerial commitment, systems perspective, openness and experimentation and transfer and integration. Managerial commitment may be recognized as the potential of the organization to develop and facilitate support, and leadership commitment to create and build knowledge across the firm. Dedication across learning suggests that management are able to provide additional re-sources, garner new options, and apply the changes required in order to facilitate learning across the firm. By demonstrating such behavior, management are able to successfully create and support a learning setting that facilitates the firm in surviving and achieving success.

The system perspective is focused on ensuring that everyone in the firm adopts a shared vision and a mutual identity, and further involves building relatives and linking members with one another through the exchange and sharing of information and knowledge [30]. Experimentation and openness in this regard relates to the extent to which a firm may be open to implementing new suggestions and ideas [30]. This involves devising a structure that promotes the presentation of new ideas and innovativeness. Transfer and integration relates to the extent to which ideas, innovations and knowledge may be transferred, on an internal bases, through communication channels in a firm [30]. The ability of distributing new ideas and knowledge across various departments and functions is fundamental to any firm's success. Furthermore, continuous learning includes efforts of an organization to provide learning opportunities for its staff [28].

3. The Effects of ERP Usage on Organizational Learning

There is a need for businesses to enhance their performance on an ongoing basis not only to ensure they survive but also so that they achieve success across the competitive field. Organizational performance has various meanings to different groups; therefore, conceptual difficulties and a lack of clarity exist in regards its measurement. Organisational performance may be recognised as the overall capacity to gather and process resources—both human and physical—so as to satisfy business goals. More specifically, organisational performance is seen to stem from businesses and is therefore measures in line with objectives and goals. At the present time, as a result of enhanced competition across firms, in addition to the focus on organisational transformation and change, all firms seek to achieve effective performance. Understanding those issues associated with organisations results in greater efficiency and performance. Overall, organizational performance is seen to encompass both financial and non-financial considerations [27].

Organizational performance can be recognised as the extent to which organisations have been successful in satisfying their objectives, with organizational performance able to be measured in regards business learning, profitability, or other financial benefits in the management of knowledge. Without the ability to measure success, management and employee enthusiasm will be non-existent [28]. Accordingly, various works have suggested different perspectives in regards the measurement of performance. Organisational performance may be seen to consider how well businesses attain their financial and market aims, with a number of different scholars considering the subjective views of

management to measure the positive results of organisations [31]. Others, in contrast, utilise objective data, such as return on assets [32].

Organisational learning is defined as the capability within an organization to maintain or improve performance based on experience [28]. According to Ojha et al. [33], the process of improving organization's actions through better understanding and knowledge. Learning is considered to be an essential aspect in any business owing to the fact it facilitates the generation and development of a sustainable competitive edge. In essence, learning may act as an approach to generating and extending upon a number of organisational capabilities, thus encouraging organisations to demonstrate ongoing improved as opposed to focusing on particular types of knowledge. Organisational learning has been recognised as a strategic component when aiming to secure a competitive edge that can be maintained over time and in enhancing performance across firms. Various works have demonstrated organisational learning as having a clear, positive and direct effect on the performance of firms; on the other hand, others emphasise that, as a result of their effect on different aspects, this learning influences organisational performance in an indirect manner.

A number of studies have been conducted in mind of exploring the effects of organisational learning in line with various elements of a firm, as in the cases of [1], [3], [5] and [6]. As an example, Argote [3] analysed the effects of organisational learning on the utilisation of ERP systems, and further considered user satisfaction, positing that organisational learning indirectly influences user satisfaction, whilst also directly influencing the adoption of ERP system. The work of Peddler et al. [34] provides the suggestion that organisational learning adopts a mediatory role in the link between knowledge engagement and management.

Organisational learning has been defined in the work of Gomez [15] as a process as opposed to an outcome, with the process witnessing firms drawing lessons from history and accordingly completing their own interpretation and assigning them into organisational routines. Organisational learning depends on the way in which ideas, information and knowledge are utilised by members of a firm [35]. In this vein, Pedller et al. [36] posit the view that those organisational members with a larger number of networks and relationships might demonstrate a better degree of acquisition when it comes to information and knowledge. Moreover, the scholars further emphasise that those members of the firm with greater communication skills have a greater degree of access to various resources. As discussed earlier, ERP systems are well positioned to attain valuable outcomes. In the view of Lara et al. [18], however, these advantages are not always clear for those organisations adopting ERP. Although prior works have analysed the effects of the use of ERP on firms' various elements, especially in the case of firms operating in the manufacturing industry, it remains that very little focus has been directed towards the effects of the use of ERP on organisational learning within the hotel industry. According to Siniša et al. [28], information technologies can be used to support employees to get learning continuously. Therefore, it is reasonable to develop the following hypothesis

H1: ERP Usage Positively Influences Continues Learning

The use of ERP gives a greater degree of access to information, and further enables firms to organize their data [28]. Moreover, ERP usage facilitates access to information in a time-efficient manner, with the system integration enabling access to rich information from a number of different portals [30]. This may support staff of an organization to adopt shared vison and mutual identity. Accordingly, the following hypothesis is proposed

H2: ERP Usage Positively Influences System Perspective

According to Madanhire et al. [25], those staff members with ERP usage are recognized as having a greater degree of access to the ideas, insights and resources of other organizational members from varying departments. Accordingly, ERP usage, when successful, may help employees to present new ideas and innovativeness. As such, the following hypothesis is devised:

H3: ERP Usage Positively Influences Openness and Experimentation

The adoption of ERP encourages the sharing of information amongst customers, other organizational partners, and suppliers [28]. Such usage encourages cross-functional coordination [30], with shared information across various firm departments potentially resulting in organizational learning. This assists a firm in the application of sound management practices, procedure, policies and structures that can enable and encourage learning. Thus, the following hypothesis is formulated:

H4: ERP Usage Positively Influences Transfer and Integration.

4. Research Model

In consideration to the literature discussed, the research model presented below is introduced. The hypotheses devised in this work are also discussed below.

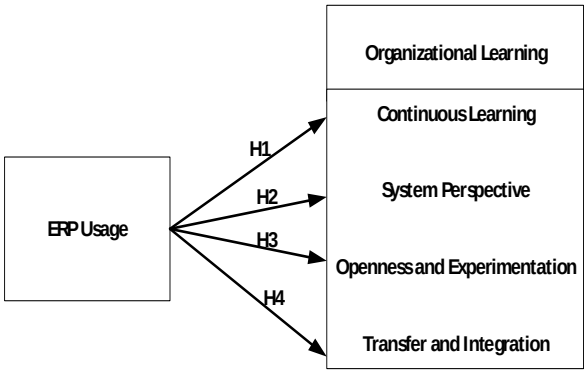


Fig. 1. Research Model

4.1 Research Methodology

In mind of exploring the impact of ERP usage on organizational learning, a research model has been devised. The model encompasses a total of five different elements, including ERP usage, continuous learning, systems perspective, openness and experimentation and transfer and integration, with each of these undergoing measurement and validation as follows: ERP usage measures were adapted from the work of Nwankpa and Yaman [30] and Fornell and Larcker [38], encompassing three items. The measures 'Systems Perspective', 'Openness and Experimentation' and 'Transfer and Integration' were incorporate in line with the work of Nwankpa and Yaman [30], whilst 'Continuous Learning' was adopted from Siniša et al. [28]. A total of four items were included in each of these measures, with the exception of 'System Perspective', which comprised 3 items. However, a number of measures underwent modification in line with the study context. All items were assessed using a five-point Likert scale. Appendix shows the items used to measure ERP usage and organizational learning.

This work warranted input from end users adopting the ERP systems across their hotels' activities, routine tasks and business processes. The present work gathered survey data through the application of a self-administered questionnaire, utilising a purposeful sample. Accordingly, Jordanian hotels that have adopted an ERP system were contacted, with data gathered from those business members utilising ERP systems within their work-related tasks and activities (hotels' managers). In total, 350 questionnaires were distributed across a total of 350 hotels; only a very small number (33) were excluded as a result of incomplete data. As such, a large number (317) provided the end volume of data for data analysis. The gender of the subjects were relatively evenly split: 56.6% male and 43.4% female. Participants' qualification was: 77% BCs, 19% MA, 4% PhD. Participants' experience: 22% less than 5 years, 61% 5-10 years, 17% greater than 10 years.

Across all of the constructs, factorial analysis was conducted. As suggested by Hair et al. [37], all item loadings were seen to exceed 0.60. Accordingly, the items are viewed as being representative of their constructs. When the Cronbach's Alpha is more than 0.70, reliability is seen to be achieved [37]. As can be seen in the Table 1, all constructs' reliabilities were seen to exceed 0.70; this means that all measures have an acceptable reliability level. Furthermore, through the use of average variance extracted (AVE), convergent validity was evaluated. The AVE of all constructs was seen to be greater than 0.50, as detailed in the table. Through the application of the Fornell-Larcker criterion, discriminant validly was assessed [38]. Moreover, as shown in the table below, the square root of each construct's AVE was found to exceed that of the correlations between the construct and all others. Accordingly, satisfactory levels of discriminant validity were demonstrated by the measurements. A relatively good fit was achieved through this conceptualization: Normed CMIN (CMIN/DF) (2.601), Root Mean Square Error of Approximation (RMSEA) (0.07), Incremental Fit Index (IFI) (0.959), (Tucker-Lewis Index) TLI (0.954) and Competitive Fit Index (CFI) (0.959).

Table 1. Measurements Validity and Reliability

	CR	AVE	TI	ERPU	CL	SP	OE
TI	0.931	0.694	**0.833**				
ERPU	0.817	0.528	0.312	**0.726**			
CL	0.914	0.605	0.530	0.390	**0.778**		
SP	0.934	0.741	0.230	0.426	0.402	**0.861**	
OE	0.884	0.656	0.504	0.316	0.347	0.229	**0.810**

TI: Transfer and Integration; ERPU: ERP Usage; CL: Continuous Learning; SP: System Perspective; OE: Openness and Experimentation;

5. Hypotheses Testing

In order to investigate the impact of use of ERP on organizational learning, AMOS 18 software was used. Fig 2 shows AMOS 18 proposed model.

It was found that ERP use has a positive influence on continuous learning, system perspective, openness and experimentation, transfer and integration. The four hypotheses are further sup-ported by the results: H1a (β=0.454, R2 = 0.388, p = 0.001); H1b (β=0.425, R2 = 0.411, p = 0.001); H1c (β=0.389, R2 = 0.344, p = 0.001); H1d ((β=0.402, R2 = 0.388, p = 0.001). Therefore, all hypotheses are accepted, as can be seen in Table 2.

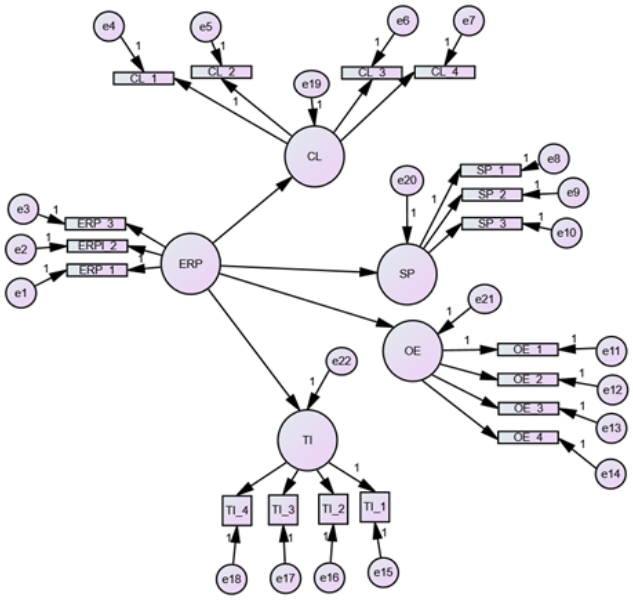


Fig. 2. Amos 18.0 proposed model

Table 2. Hypotheses Examining

Hypotheses	Path Coefficients	Significant Level	R2
H1: ERPU → CL	0.454	0.001	0.388
H1: ERPU → SP	0.425	0.001	0.411
H1: ERPU → OE	0.389	0.001	0.344
H1: ERPU → TI	0.402	0.001	0.388

6. Discussion and Conclusion

When considering prior works, this study analysed the relation of the use of ERP and organizational learning. The study model has undergone empirical analysis with

consideration to the Jordanian hotel sector, with the findings providing a number of valuable managerial and theoretical implications for both business professionals and researching academics.

First and foremost, the use of ERP has positive effect on continuous learning. This is explained, primarily, by the fact that the use of ERP provides integrated data [36] from various firm-related resources, whilst further enhancing access to data for improved learning processes. This, in the view of Lee et al. [22], further assists in enhancing ongoing learning opportunities for individuals. Furthermore, should employees be given continuous, sufficient information from various resources to allow them to complete their activities and tasks, they are then better positioned to enjoy their work and accordingly enhance their performance. This could potentially result in enhanced levels of organizational learning [15].

Secondly, the findings suggest that ERP usage influences system perspective. In the view of Argote [3] and Rajan et al. [39], the use of ERP has the capacity to link members with one another through exchanging information and knowledge. This results in firms progressing beyond the individual goals of employees and focuses more on adopting a shared, universal vision. Moreover, securing a system perspective provides an agreed upon action and language amongst those involved in the learning process, which subsequently leads to enhanced organizational learning [40].

Third, the findings have suggested that the use of ERP has both a positive and direct influence on experimentation and openness, with the researcher considering that the use of ERP might add to creating an environment that facilitates existing knowledge being called into question, thus enabling users to continuous renew, improve upon and widen organizational knowledge [40]. In this way, openness and experimentation are supported by the environment. In the view of Sadrzadehrafiei [41], these results in organizations being better positioned to be more flexibility in uncertain work settings and to demonstrate a greater degree of openness in regards learning best practices.

Lastly, as shown in the study findings, the use of ERP has a significant influence on transfer and integration, potentially owing to the use of ERP enhancing the capacity of firms to garner more information in greater detail, and to do so in real time; nonetheless, this has induced a significant distribution of information throughout the firm [15]. Moreover, transfer and integration is seen to be notably influential on organizational performance; when firms are able to distribute new ideas and knowledge across various boundaries and departments, they then might be able to achieve improvement in organizational performance [28].

This work sought to explore the impact of ERP usage on organizational learning. A new research model has been presented, this model has been empirically investigated in the Jordanian hotel industry. The findings from the empirical findings emphasise that ERP usage has significant impact on organizational learning.

In line with the study findings, various aspects of organizational learning, such as continuous learning, system perspective, openness and experimentation and transfer and integration are recognised as able to best encourage the use of ERP.

The findings have shown that the positive influence of the use of ERP on organizational learning in the Jordanian hotel industry. Therefore, hotels that already use ERP may provide members with support in the continuous process of learning, build an agreed vision and adopt a mutual identity, and build a structure that motivates and drives new ideas whilst embracing innovation and distributing ideas and knowledge across various hotel departments. Accordingly, hotel management needs to encourage organizational learning efforts across hotels to ensure the best outcomes of the use of ERP.

This study makes important theoretical and practical implications. This study shows the influence of ERP usage on organizational learning in Jordanian hotel sector, which has been mostly ignored by previous research. Although prior studies presented the importance of adoption of ERP on different areas of organizations, less is known about the impact of ERP usage on organizational learning. The empirical evidence reveals that ERP usage has positive influence on organizational learning. Therefore, this study pave the way for researches to conduct further investigation and understanding of IT role on achieving organizational learning.

Moreover, this study has key practical implications particularly for hotels' managers and executives who are seeking for utilizing ERP to achieve organization learning with hotel sector. Based on results of this study, mangers can understand that by increasing use of ERP in a hotel, this will lead to achieve organizational learning. In addition, practitioners should be aware that investing in training employees on ERP usage will contribute to increase opportunity of achieving organizational learning.

Despite the fact that this work provides various contributions, there are some limitations. Primarily, the suggested research model, with the inclusion of all relationships, has been examined in the Jordanian context, meaning subsequent works could focus on improving the generalizability of such findings through analyzing the hypothesized links with a sample based in other countries. Secondly, a limited number of factors with the potential to influence the use of ERP on organizational learning were considered. Although these aspects adopt a key role in terms of investigating the impact of ERP usage on organizational learning, other elements, including strategic leadership and empowerment, could also influence the impact of ERP usage on organizational learning. Accordingly, subsequent works should be carried out in mind of examining the influence of such elements. Fourth, this study is limited to hotel sector. More investigation could be performed on other sectors such as public sector. Furthermore, new study could be conducted to investigate the influence of using other

technologies and information systems such as e-commerce and web services on organizational learning.

Appendix

Items used to measure ERP usage construct

- Our hotel uses ERP system very intensively
- Our hotel uses ERP system very frequently
- Overall, our hotel uses ERP system a lot

Items used to measure Continuous Learning construct

- Our hotel enables employees to get required information at any time easily and quickly.
- Our hotel keeps an up-to-date database of employee skills.
- Our hotel creates systems to evaluate differences between current and expected performance
- Our hotel uses two-way communication on a regular basis, such as open meetings and suggestion systems

Items used to measure System Perspective construct

- All trained employees have information regarding hotel's objectives and goals.
- All units that make up our hotel (sections, departments, individuals and divisions work team) are aware of how they contribute to perform the all objectives and goals
- All sections that make up our hotel are unified working together in a coordinated method

Items used to measure Openness and Experimentation construct

- Our hotel encourages innovations and experimentation as an approach for business processes.
- Our hotel follows up activities of other hotels and is willing to adopt those techniques and activities that it may be interesting and useful
- Ideas and experience provided by external sources (training companies, consultants etc.) are vital tools for our hotel learning
- The culture of our hotel encourages opinion and expression as well as suggestions regarding the methods and procedures for activity performance

Items used to measure Transfer and Integration construct

- Failure and errors are always analyzed and discussed in our hotel at all level
- In our hotel, there are procedures and processes that offer employees the opportunity to talk about new activities, programs and ideas that may be beneficial to the hotel
- Our hotel has an instrument that allows what has been learnt in past situation to stay accessible to all employees
- Our hotel encourages cooperation, information distribution and teamwork

References

[1] Ağaoğlu, Mustafa, E. Serra Yurtkoru, and Aslı Küçükaslan Ekmekçi. "The effect of ERP implementation CSFs on business performance: an empirical study on users' perception." Procedia-Social and Behavioral Sciences, vol. 210, no. 22, pp. 35-42, 2015.

[2] Ahmad, M. Munir, and Ruben Pinedo Cuenca. "Critical success factors for ERP implementation in SMEs." Robotics and computer-integrated manufacturing, vol. 29, no.3, pp. 104-111, 2013.

[3] Argote, Linda. "Organizational learning research: Past, present and future." Management learning, vol.42, no. 4, pp. 439-446, 2011.

[4] Bhatti, "Critical success factors for the implementation of enterprise resource planning (ERP): empirical validation." The second international conference on innovation in information technology, 2005.

[5] BolíVar-Ramos, MaríA Teresa, VíCtor J. GarcíA-Morales, and EncarnacióN GarcíA-SáNchez. "Technological distinctive competencies and organizational learning: Effects on organizational innovation to improve firm performance," Journal of Engineering and Technology Management, vol. 29, no. 3, pp. 331-357, 2012.

[6] Bowen, L. Gary, Roderick A. Rose, and William B. Ware. "The reliability and validity of the school success profile learning organization measure." Evaluation and program planning, vol. 29, no.1, pp. 97-104, 2006.

[7] Brown, David H., and Susan He. "Patterns of ERP adoption and implementation in China and some implications." Electronic Markets, vol. 17, no.2, pp. 132-141, 2007.

[8] Chiva, Ricardo, Joaquin Alegre, and Rafael Lapiedra. "Measuring organisational learning capability among the workforce," International Journal of Manpower, vol. 28, no.3, pp. 224-242, 2007.

[9] Cook, Scott DN, and Dvora Yanow. "Culture and organizational learning," Journal of management inquiry, vol. 2, no. 4, pp. 373-390, 1993.

[10] Dong, John Qi, and Chia-Han Yang. "Information technology and organizational learning in knowledge alliances and networks: Evidence from US pharmaceutical industry," Information & Management, vol. 52, no. 1, pp. 111-122, 2015.

[11] Farhanghi, Ali Akbar, Abbas Abbaspour, and Reza Abachian Ghassemi. "The effect of information technology on organizational structure and firm performance: An analysis of Consultant Engineers Firms (CEF) in Iran," Procedia-Social and Behavioral Sciences, vol. 81, pp. 644-649, 2013.

[12] Goh, Swee C. "Improving organizational learning capability: lessons from two case studies," The learning organization, vol. 10, no. 4, pp. 216-227, 2003.

[13] Hair, R. E. Anderson, R. L. Tatham, and W. C. Black, Multivariate data analysis (5th ed.). New Jersey: Prentice-Hall, 2012.

[14] Jerez-Gomez, Pilar, José Céspedes-Lorente, and Ramón Valle-Cabrera. "Organizational learning capability: a proposal of measurement." Journal of business research, vol. 58, no. 6, pp. 715-725, 2005.

[15] Jerez-Gomez, Pilar, José Céspedes-Lorente, and Ramón Valle-Cabrera. "Organizational learning capability: a proposal of measurement." Journal of business research, vol. 58, no. 6, pp. 715-725, 2005.

[16] Kumar, Vinod, Bharat Maheshwari, and Uma Kumar. "An investigation of critical management issues in ERP implementation: emperical evidence from Canadian organizations." Technovation, vol. 23, no.10, pp. 793-807, 2003.

[17] Kwahk, Kee-Young, and Hyunchul Ahn. "Moderating effects of localization differences on ERP use: A socio-

technical systems perspective." Computers in Human Behavior, vol. 26, no. 2, pp. 186-198, 2010.
[18] Lara, Francisco J., and Andres Salas-Vallina. "Managerial competencies, innovation and engagement in SMEs: The mediating role of organisational learning." Journal of Business Research, vol. 79, no. 15, pp. 152-160, 2017.
[19] Laudon, Kenneth C., and Jane Price Laudon. Management information systems. Vol. 8. Prentice Hall, 2015.
[20] Chao, Y.C. Organizational learning and acquirer performance: How do serial acquirers learn from acquisition experience?. Asia Pacific Management Review, 23(3), pp.161-168, 2018.
[21] Wan, S. and Niu, Z. An e-learning recommendation approach based on the self-organization of learning resource. Knowledge-Based Systems, 160, pp.71-87,2018.
[22] Lee, Heeseok, and Byounggu Choi. "Knowledge management enablers, processes, and organizational performance: An integrative view and empirical examination." Journal of management information systems, vol. 20, no. 1, pp.179-228, 2013.
[23] Lee, "Relationship between the use of information technology (IT) and performances of human resources management (HRM)," Degree of Docotor of Business Administration, Alliant International University, San Diego, CA, 2008.
[24] Levitt, Barbara, and James G. March. "Organizational learning." Annual review of sociology vol. 14, no. 1, pp. 319-338, 1988.
[25] Madanhire, Ignatio, and Charles Mbohwa. "Enterprise resource planning (ERP) in improving operational efficiency: Case study." Procedia CIRP40, pp. 225-229, 2016.
[26] Madapusi, Arun, and Derrick D'Souza. "The influence of ERP system implementation on the operational performance of an organization." International Journal of Information Management, vol. 32, no.1, pp. 24-34, 2012.
[27] Marsick, Victoria J., and Karen E. Watkins. "Demonstrating the value of an organization's learning culture: the dimensions of the learning organization questionnaire." Advances in developing human resources, vol. 5, no .2, pp.132-151, 2003.
[28] Siniša Mitić, Milan Nikolić, Jelena Jankov, Jelena Vukonjanski, Edit Terek. "The impact of information technologies on communication satisfaction and organizational learning in companies in Serbia," Computers in Human Behavior, vol. 76, pp. 87-101, 2017.
[29] McGill, Michael E., John W. Slocum Jr, and David Lei. "Management practices in learning organizations." Organizational dynamics, vol. 21, no.1, pp. 5-17, 1992.
[30] Nwankpa, Joseph, and Yaman Roumani. "Understanding the link between organizational learning capability and ERP system usage: An empirical examination." Computers in Human Behavior, vol. 33, pp. 224-234, 2014.
[31] Motwani, Jaideep, Asli Yagmur Akbulut, and Vijay Nidumolu. "Successful implementation of ERP systems: a case study of an international automotive manufacturer," International Journal of Automotive Technology and Management, vol. 5, no.4, pp. 375-386, 2005.
[32] Motwani, Jaideep, et al. "Successful implementation of ERP projects: Evidence from two case studies." International Journal of Production Economics, vol. 75, no. 2, pp. 83-96, 2002.
[33] Ojha, D., Struckell, E., Acharya, C. and Patel, P.C. Supply chain organizational learning, exploration, exploitation, and firm performance: A creation-dispersion perspective. International Journal of Production Economics, 204, pp.70-82, 2018.
[34] Nwankpa, Joseph K., and Yaman Roumani. "The Influence of Organizational Trust and Organizational Mindfulness on ERP Systems Usage." CAIS, vol. 34, pp. 85-98, 2014.
[35] Nwankpa, Joseph K. "ERP system usage and benefit: A model of antecedents and outcomes." Computers in Human Behavior, vol. 45, pp. 335-344, 2015.
[36] Peddler, J. Burgoyne, and T. Boydell, The learning company, a strategy for sustainable development. London: McGraw-Hill, 1998.
[37] Hair Jr, J. F., W. C. Black, B. J. Babin, and R. E. Anderson. "Multivariate data analysis (Seventh, Pearson new international ed.)." Harlow: Pearson Education Limited 2014.
[38] Fornell, C., & Larcker, D. F. Evaluating structural equation models with unobservable variables and measurement error. Journal of marketing research, 39-50,
[39] Rajan, Christy Angeline, and Rupashree Baral. "Adoption of ERP system: An empirical study of factors influencing the usage of ERP and its impact on end user." IIMB Management Review, vol. 27, no. 2, pp. 105-117, 2015.
[40] Saadat, Vajiheh, and Zeynab Saadat. "Organizational learning as a key role of organizational success." Procedia-Social and Behavioral Sciences, vol. 230, pp. 219-225, 2016.
[41] Sadrzadehrafiei, Samira, et al. "The benefits of enterprise resource planning (ERP) system implementation in dry food packaging industry." Procedia Technology, vol. 11, pp. 220-226, 2013.

Faisal Aburub is currently an associative professor at MIS department in University of Petra, Amman, Jordan. He holds PhD degree in Information Systems from University of the West of England, UK. Aburub has more than 10 years of experience in application business process modelling concepts on different business environments. His research interests lie in ERP systems, organizational agility, and bridging the gap between business and system models.

Polar Split Tree as a Search Tool in Telecommunication

Farzad Bayat
Department of Electrical and Computer Engineering, Faculty of Engineering, Kharazmi University
std_farzadbayat@khu.ac.ir
Zahra Nilforoushan*
Department of Electrical and Computer Engineering, Faculty of Engineering, Kharazmi University
nilforoushan@khu.ac.ir

Received: 24/Feb/2018 Revised: 22/Aug/2018 Accepted: 16/Sep/2018

Abstract

Tree search algorithms are vital for the search methods in structured data. Such algorithms deal with nodes which can be taken from a data structure. One famous tree data structure is split tree. In this paper, to compute the split tree in polar coordinates, a method has been introduced. Assuming that the algorithm inputs (in form of points) have been distributed in the form of a circle or part of a circle, polar split tree can be used. For instance, we can use these types of trees to transmit radio and telecommunication waves from host stations to the receivers and to search the receivers. Since we are dealing with data points that are approximately circular distributed, it is suggested to use polar coordinates. Furthermore, there are several researches by search algorithms for the central anchor which leads to the assignment of a virtual polar coordinate system. In this paper, the structure of Cartesian split tree will be explained and the polar split tree will be implemented. Then, by doing nearest neighbor search experiments, we will compare the polar split tree and polar quad tree in terms of searching time and amount of distance to the closest neighbor and in the end, better results will be achieved.

Keywords: Split Tree; Polar Split Tree; Quad Tree; Polar Quad Tree; Nearest Neighbor Search.

1. Introduction

As it is apparent, nearest neighbor search can be used in a lot of cases relating to distance and it can also be used in classifying and clustering the data as a similarity criterion. Depending on the case, various algorithms can be used for nearest neighbor search. For instance we can find the nearest neighbor by searching trees including quad tree and split tree which will be briefly explained later. In some cases, depending on the case and the space in which the data are located, it is better to use polar coordinates. That's the reason why, we presented a method for the polar split tree. This algorithm has been compared with polar quad tree and the obtained results are compared in terms of searching time and the amount of distance to optimal response.

In the first section of this paper, searching trees and finding the nearest neighbor have been reviewed. In Section 2, quad tree and the method of computing it will be explained . Section 3 is devoted to the split tree . In Section 4, polar coordinates and some of its applications will be mentioned . In Section 5 we will indicate the way of construction and searching in polar quad trees. In Section 6, polar split tree is introduced and how to search the nearest neighbor by using this tree is defined . Section 7 is devoted to the investigating the efficiency of our proposed method and comparison of polar quad tree with polar split tree. Finally in Section 8, the conclusion and summing-up of the proposed algorithm are mentioned.

2. Quad Tree

The quad tree is a rooted tree whose internal node contains four children and each node represents one square. For a set P of points, the primary square of them is a square that contains all points of P and is the root of the tree that corresponds with partitions. If node *v* contains a child, the squares relating to its children will be the four areas of square *v*. That's why this tree is called quad tree. In other words, the leaf squares form a subset of root square and this subset is the subset of the quad tree. Fig. 1 indicates the quad tree and its subsets.

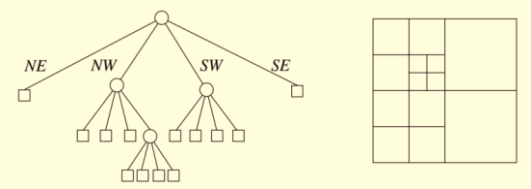


Fig. 1. A sample of quad tree and its subsets

The root children are called NE, NW, SW and SE that specify in which square they belong to. For example, NE belongs to northeast square, NW belongs to northwest square, etc.

Quad tree is used to save different kinds of data. For saving the a set of points in a plane, the square division will continue recursively as long as there is more than one point in a square. Thus the quad-tree for P inside the

* Corresponding Author

square $\sigma := [x_{min}: x_{max}] \times [y_{min}: y_{max}]$ is defined as follows:

- If the number of points of P are less than or equal to 1, the quad tree will just contain one leaf.
- Otherwise $\sigma_{NE}.\sigma_{NW}.\sigma_{SW}$ and σ_{SE} show four areas of σ as follows:

$$x_{mid} = (x_{min} + x_{max})/2$$
$$y_{mid} = (y_{min} + y_{max})/2$$

$P_{NE} := \{p \in P: p_x > x_{mid} \,.\, p_y > y_{mid}\}$
$P_{NW} := \{p \in P: p_x \leq x_{mid} \,.\, p_y > y_{mid}\}$
$P_{SW} := \{p \in P: p_x \leq x_{mid} \,.\, p_y \leq y_{mid}\}$
$P_{SE} := \{p \in P: p_x > x_{mid} \,.\, p_y \leq y_{mid}\}$

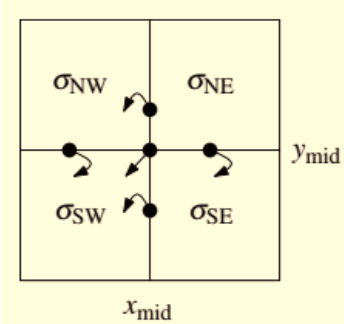


Fig. 2. A simple quad tree

Lemma 2.1. The depth of quad tree for the set P of points in the plane is at most *log(s/c)+3/2* in which *c* is the smallest distance between every two points in P and s is the lateral length of the primary square containing points of P . (For a proof refer to [3]).

Theorem 2.2. A quad tree with the depth of *d* is for saving *n* points containing *O((d+1)n)* nodes and can be constructed with time complexity of *O((d+1)n).* (For a proof refer to [3]).

3. Split tree

The split tree for a set of points *P* is a rooted binary tree data structure containing the points of *P* in its leaves. Split tree can be used as a searching tree in finding the nearest neighbor [10]. In split tree algorithm, each time the bounding box of inputs is considered and then the longest edge of it will be halved and the bounding box will be drawn for the left and right sub tree as long as the points inside the box is more than one point. In other words, if there is one point inside the bounding box, then the split tree consists of one single node that stores that point and the algorithm will stop.

Theorem 3.1. A split tree is constructed for saving *n* points. Its time complexity in the worst case and in the best case are $\theta(n^2)$ and *O(nlogn)* respectively and its height is $\theta(n)$. [10]

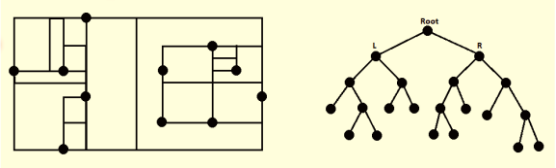


Fig. 3. A sample of split tree algorithm

For improving the complexity of split tree, partial split tree algorithm has been introduced whose threshold is *n/2*. It means that in each step a bounding box will be drawn for left and right children only if the numbers of their points are more than *n/2*. Thus, the number of children in the last level of tree can be between *1* and *n/2*.

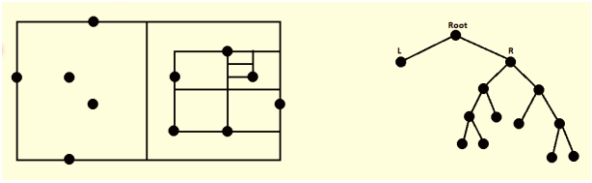


Fig. 4. A sample of partial split tree algorithm

4. Polar Coordinate and its Applications

polar coordinates is a two-dimensional coordinate system where each point p is represented by $(r.\theta)$. The distance of each point to the center of the coordinate is r and θ is the angle between *x*-axis and the line connecting p and the center of the coordinate [1,7 20].

A point can be converted from Cartesian coordinate to polar coordinate (and vice versa) as follows:

$x = rcos\theta \qquad y = rsin\theta$
$r^2 = x^2 + y^2 \qquad \theta = tan^{-1}\, y/x$

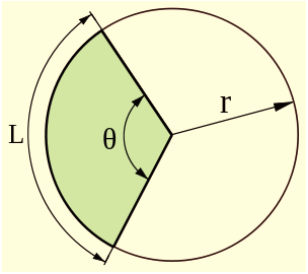


Fig. 5. A sector of the circle with angle θ and arc length L

For computing the length of arc L relating to angle θ in a circle with the radius *r,* the formula $L = r\theta$ must be used.

Polar coordinate can be used in most of physics equations such as circular movement of central force and planets rotation. Furthermore, in cases in which the data have been distributed in the form of a circle or part of a circle, polar coordinate can be used. For instance, when the data are in form of radio or telecommunication waves, as we are dealing with shapes which are approximately circular, using polar coordinate is recommended [2,17,19,21].

In the following parts, polar quad tree and polar split tree will be investigated.

5. Polar Quad Tree

A polar quad tree is a tree data structure in two-dimensional polar metric space in which each internal node has exactly four children. It is most often used to partition a two-dimensional space by recursively subdividing it into four subsectors or regions. A polar quad tree with k levels has 4^k leaf cells, defined by 2^k angular and 2^k radial divisions. There are many applications for polar quad trees [4,8,13,18,22]. For drawing quad trees, the square covering input points is drawn and as long as the number of points in each square is more than 1, each square will be recursively divided into four parts. For drawing polar quad tree (PQT), the same steps must be followed as well but we are dealing with arc and part of radius instead of square sides.

5.1 Construction of Polar Quad Tree

The following steps shall be done for constructing the polar quad tree for input points *P*:

1. Draw the smallest circle containing the points.
2. Divide the circle containing input points into four equal parts by two perpendicular diameters.
3. Divide each quarter of the circle into four parts recursively. For dividing one quarter of the circle into four parts, we first halve the radius and then halve the related arc. The numbering could be done similar to Fig. 6.

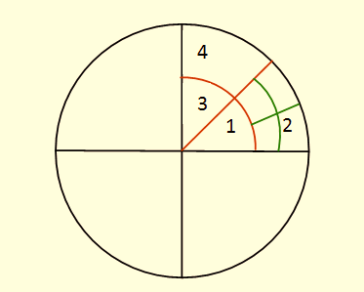


Fig. 6. Numbering polar quad tree

4. In Fig. 6, for dividing the children similar to areas corresponding 1 and 3, follow the previous steps. But for dividing the children similar to 2 and 4, half both the related arc and the part of radius located in that cell.
5. Repeat steps (1) to (4) as long as there is more than one point.

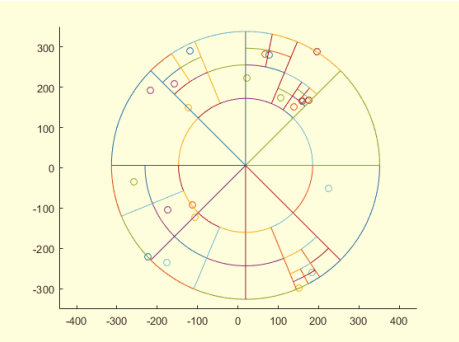


Fig. 7. A sample of polar quad tree

The pseudo code of the polar quad tree construction is as following:

Polar_Quad_Tree_Construction (DataPoints P):

1. **Compute** smallest circle C enclosing P with *Nimrod Mojido Algorithm* [9].
2. **Divide** the circle into 4 parts: Part1, Part2, Part3, Part4; Fig. 6.
3. **If** part1 has more than one data, then call **Polar_Quad_Tree_Construction (Part1)** and add it as first child of C.
4. **If** part2 has more than one data, then call **Polar_Quad_Tree_Construction (Part2)** and add it as second child of C.
5. **If** part3 has more than one data, then call **Polar_Quad_Tree_Construction (Part3)** and add it as third child of C.
6. **If** part4 has more than one data, then call **Polar_Quad_Tree_Construction (Part4)** and add it as forth child of C.
7. **Return** C.
8. **End**.

Note that the smallest enclosing circle for a set of n points in the plane can be computed in O(n) expected time [9].

Hence the time complexity of constructing polar quad tree is:

$T(n)=\max\{O(n), 4T(n/4)+O(n)\}=O(n\log n)$.

For similar tree structure see[6,11,14,16].

5.2 The Nearest Neighbor Search in Polar Quad Tree

After constructing the polar quad tree (dividing each cell into four parts) next job is to find ourthe nearest neighbor for each query point in the polar quad tree.To this end, we average the x and y coordinates of the points inside quarter 1 and 4. Let the average of the points inside quarter 1 be (x_1', y_1') and the average of the points inside quarter 4 be (x_4', y'_4). If we want to search a query point (x_q, y_q), the coordinate of it is compared using the following formula:

$$x_m = (x_1' + x_4')/2 \, , y_m = (y_1' + y_4')/2$$

Then we decide which subset of the tree should be chosen to continue the search.

- If $x_m > x_q$ and $y_m > y_q$, quarter 4 will be chosen for searching,
- If $x_m > x_q$ and $y_m \leq y_q$, quarter 3 will be chosen for searching,
- If $x_m \leq x_q$ and $y_m > y_q$, quarter 2 will be chosen for searching,
- If $x_m \leq x_q$ and $y_m \leq y_q$, quarter 1 will be chosen for searching.

By itterating this searching method, we find a leaf in the polar quad tree that the query point belongs to.

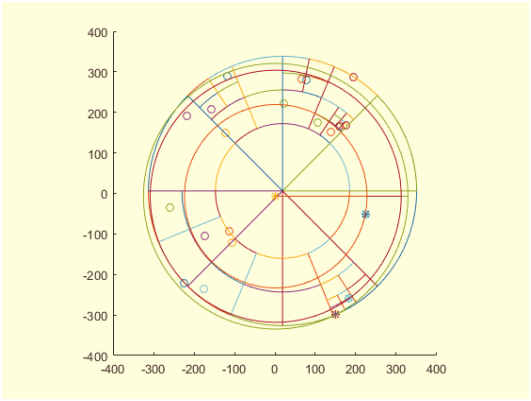


Fig. 8. Nearest neighbor search using polar quad tree

The correctness of this algorithm follow from the fact that for a given query point $Q = (r_q, \theta_q)$ inside the smallest enclosing circle C of data, according to the values of radius r_q and angle θ_q of Q, the point Q is located in one of the four division regions. If the region containing Q does not have more than one data, it will no longer be divided into four parts, and corresponding to it in the tree is a leaf indicating the region containing the Q. In this case, the search will terminate. If the region containing Q contains more than one data, it is divided into four parts again (in the same way that each side of it is halved), and Q is placed in one of the four regions according to the two values r_q and θ_q.

This process is recursively repeated until we reach to a leaf on the corresponding tree; that leaf has labeled with the name of region containing Q.

In particular, in Fig. 9, the point Q with the * sign is contained in the gray region which corresponds to the gray color leaf of the corresponding tree.

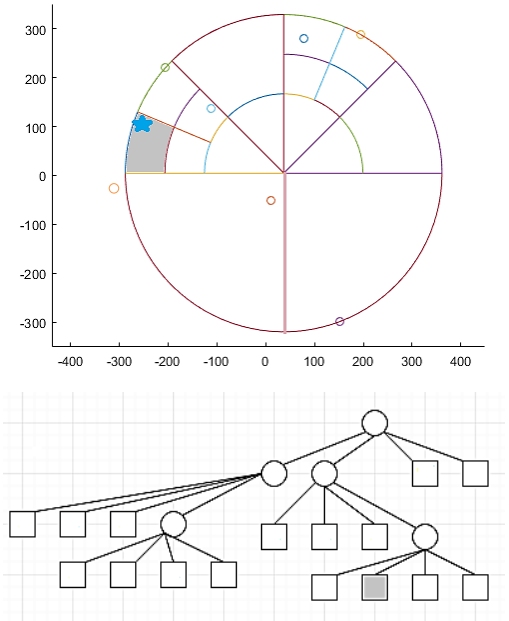


Fig. 9. A sample of nearest neighbor search for query point * using polar quad tree and corresponding search tree

As a result, finding the nearest neighbor in a polar quad tree can be computed in average case

S(n)=T(n)+logn=O(nlogn),

where T(n) is the average time complexity of polar quad tree construction.

6. Polar Split Tree

The polar split tree is a hierarchical rooted binary tree data structure in two- dimensional polar metric space which can be used as a searching tree in finding the nearest neighbor in polar metric spaces. Using polar split trees will usually improve the results obtained by the polar quad trees [5,15]. There are two steps for investigating polar split tree (PST), first the polar split tree is calculated and then the procedure of finding the nearest neighbor search in polar split tree is explained. In the following these two steps are explained.

6.1 Construction of Polar Split Tree

For construction the polar split tree, it is assumed that input data have been distributed in a circular shape. At first, based on Nimrod Megiddo algorithm, the smallest circle containing all points must be drawn [12] hich takes $O(n)$ time. Then similar to split tree in Cartesian coordinates, the longest edge must be halved. In order to do that, two components of angle and arc of the circle perimeter are compared and the one which is bigger, will be halved. For starters, we halve the perimeter of the circle. Then, we draw one of the circle's diameters and as a result the circle containing the points will be halved. Therefore the covering circle will be drawn for the left and right sub trees (part of a circle which is located at the center of the primary circle) and the length of radius sector will be compared with the angle sector and the one which is greater will be halved. The steps will continue recursively as long as there is just one point in the covering circle. For drawing the tree, we use the convention that, initially the horizontal diameter parallel to x-axis is drawn and the upper part and lower part of the circle will be the right and left child respectively.

In next steps, the right upper part will be the right child, the left upper part will be the left child, the right lower part will be the right child and the left lower part will be the left child. Finally, there will be a polar split tree. Fig. 10 indicates a sample of a polar split tree for set of 20 random input points. The pseudo-code for polar split tree construction is as follows:

Polar_Split_Tree_Construction (DataPoints P)

1 Compute the smallest circle C containing P using *Nimrod Mojido Algorithm* **[9].**

2 Divide the circle into two parts: Part1 and Part2.

3 Compute the smallest subsector covering points of Part1 and Part2, say S1 and S2.

4 Compare the length of arc and radius of S1 (and similarly for S2) and half the bigger one into Part 3 and Part4.

5 If Part 3 has more than one data, then go to line 4.

6 If Part 4 has more than one data, then go to line 4.

7 End

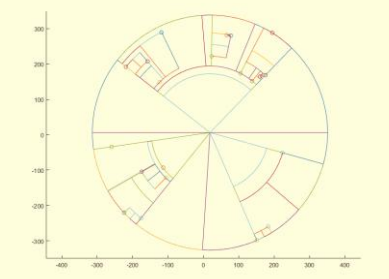


Fig. 10. A sample of polar split tree

6.2 Complexity of Polar Split Tree

As a polar split tree can be obtained by drawing a circle and we can draw a circle at least by three points, there will be two situations. Two points will be in one quarter and the other one will be in another quarter or every of these three points will be located in individual quarters. Therefore, considering the location of the points, the maximum depth of the tree will be $O(n)$. On the other hand, since this tree is a binary tree, the minimum depth of the tree will be the depth of a complete tree, i.e. $O(logn)$, and we have $O(logn) \leq h_{PST} \leq O(n)$. If there are points and the worst case of the tree happens, the depth of the tree will be $O(n^2)$ and the time constructing tree in the worst case will be and in the best case the height will be $O(logn)$ and the time of constructing the tree will be $O(nlogn)$.

6.3 The Nearest Neighbor Search in the Polar Split Tree

After construction the polar split tree for the set of input points, some random query points are tested to see whether this tree specifies the nearest neighbor to the point correctly or not. If for finding the nearest neighbor to the query point we just suffice to the search in the tree, in some cases the answer will be right and in some cases it will be wrong. The wrong answer will happen when the nearest point to the query point is in a cell to which the query point doesn't belong. But since the recursive functions use a depth first search mechanism, the point which is in the same cell as the query point will be chosen as the nearest neighbor and this answer is wrong. As it can be seen in Fig. 11, the query point which has been shown by * is nearer to point A but the tree recognizes point B as the nearest point because the query point is located in the cell belonging to B.

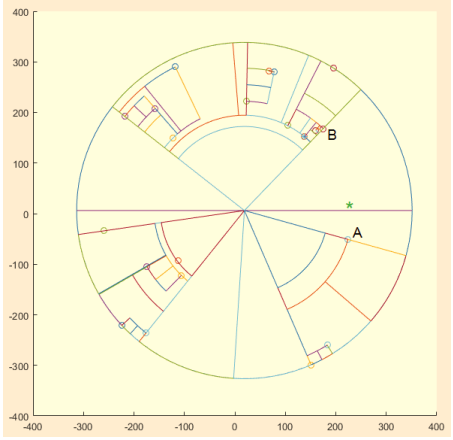


Fig. 11. Wrong output in the nearest neighbor search

For solving this problem, we will perform the search in kd-tree algorithm described in [23]. In kd-tree algorithm, when point x is found as the nearest point to the query point p, the distance between x and p is calculated and called r, i.e., d(x,p)=r. Then a circle with center p and radius r is drawn and the cells which are completely or partially inside this circle are investigated. Then the distance between the points inside every of these cells and p is calculated. If this distance is less than r, we update r and the nearest neighbor point. A circle is drawn with radius r and center p. Then the previous steps will be done recursively as long as there won't be any change in r and the nearest neighbor point. In the last step, the calculated point x is the nearest neighbor to the query point p. [12]

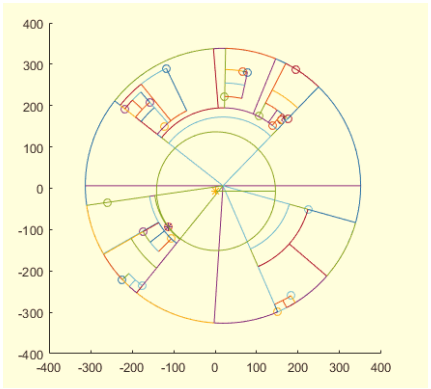


Fig. 12. Nearest Neighbor Search by polar split tree

The nearest neighbor searching pseudo code in polar split tree will be as following :

Spt_NNS (Root root, Query QPoint)

1 CurNode = root
2 distance = biggest integer number
3 Start from root.
4 If root has no children, then **return** null
5 If root has a child C which is a leaf and its value is lower than distance, then distance = value of C, and **return** distance
6 If QPointX<=CurNodeX
7 **If** QPointY-distance<=CurNodeY
8 **If** it has a left child, then **return** Spt_NNS with left child
9 **If** QPointY-distance>CurNodeY
10 **If** it has a right child, then **return** Spt_NNS with right child
11 Else
12 **If** QPointY-distance>CurNodeY
13 **If** it has a right child, then **return** Spt_NNS with right child
14 **If** QPointY-distance<=CurNodeY
15 **If** it has a left child, then **return** Spt_NNS with left child
16 End

In order to see the correctness of this algorithm note that for a given query point Q inside the smallest enclosing circle C of data, if Q is in the northern or southern semicircle of C, in the corresponding tree T we

move from root to the left or right respectively. In each semicircle that Q is located, we calculate the bounding box B for the data of that semicircle according to the Polar_Splite_Tree_Construction algorithm and split the largest side into two halves. If the line that divides B into two halves is an arc line, B is divided into two upper and lower halves, and if Q is in the upper or lower part of B, in T we move to the left or right. If the line that divides B into two halves is a radial line, B is divided into two halves left and right, and if Q is in the left or right part of B, in T we move to the left or right.

These steps are recursively repeated until we reach to a leaf of T which has labeled by the name of area containing Q.

In particular, in the Fig. 13, the point Q is displayed by * and the area containing it (the gray zone) is corresponds to the gray color leaf of T.

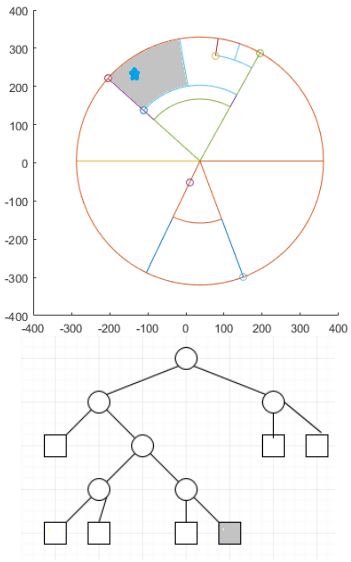


Fig. 13. A sample of nearest neighbor search for query point * using polar split tree and corresponding search tree

				Polar Split Tree (100)			Polar Quad Tree (100)		
Row#	DataSet#	QuerySet#	QueryPoint	1NN	Distance	Time(ns)	1NN	Distance	Time(ns)
1	1	1	(2,-7)	(4,-25)	18.11077028	0.09375	(6,-22)	15.5241747	0.046875
2	1	2	(-9,8)	(-7,-27)	35.05709629	0.015625	(-31,-22)	37.20215048	0.046875
3	1	3	(0,-5)	(4,-25)	20.39607805	0	(6,-22)	18.02775638	0.03125
4	1	4	(9,1)	(4,-25)	26.47640459	0.03125	(-31,-22)	46.14108798	0
5	1	5	(2,-4)	(4,-25)	21.09502311	0.03125	(6,-22)	18.43908891	0.03125
6	1	6	(5,8)	(-6,33)	27.31300057	0.015625	(-31,-22)	46.86149806	0
7	1	7	(-7,-6)	(-7,-27)	21	0.015625	(6,-22)	20.61552813	0
8	1	8	(-6,7)	(4,-25)	33.52610923	0	(-31,-22)	38.28837944	0
9	1	9	(-4,0)	(4,-25)	26.2488095	0.015625	(6,-22)	24.16609195	0
10	1	10	(5,9)	(-6,33)	26.40075756	0.03125	(-31,-22)	47.50789408	0.015625
11	2	1	(2,-7)	(-14,0)	17.4642492	0.03125	(-7,-3)	9.848857802	0.03125
12	2	2	(-9,8)	(-14,0)	9.433981132	0	(-32,1)	24.04163056	0
13	2	3	(0,-5)	(-14,0)	14.86606875	0.015625	(-7,-3)	7.280109889	0
14	2	4	(9,1)	(-14,0)	23.02172887	0.03125	(-32,1)	41	0
15	2	5	(2,-4)	(-14,0)	16.4924225	0.03125	(-7,-3)	9.055385138	0
16	2	6	(5,8)	(-14,0)	20.61552813	0.03125	(-32,1)	37.65634077	0.03125
17	2	7	(-7,-6)	(-14,0)	9.219544457	0.015625	(-7,-3)	3	0
18	2	8	(-6,7)	(-14,0)	10.63014581	0	(-32,1)	26.68332813	0.015625
19	2	9	(-4,0)	(-14,0)	10	0	(-32,1)	28.01785145	0.03125
20	2	10	(5,9)	(-14,0)	21.02379604	0.03125	(-32,1)	37.85498646	0.03125
21	3	1	(2,-7)	(-7,-3)	9.848857802	0.015625	(14,-17)	15.62049935	0.015625
22	3	2	(-9,8)	(-8,5)	3.16227766	0.015625	(-46,-21)	47.01063709	0
23	3	3	(0,-5)	(-7,-3)	7.280109889	0	(-46,-21)	48.70318265	0.015625
24	3	4	(9,1)	(-7,-3)	16.4924225	0.015625	(-46,-21)	59.23681288	0.03125
25	3	5	(2,-4)	(-7,-3)	9.055385138	0	(-46,-21)	50.92150823	0.046875
26	3	6	(5,8)	(-7,-3)	16.2728206	0	(-46,-21)	58.66856058	0
27	3	7	(-7,-6)	(-7,-3)	3	0	(14,-17)	23.70653918	0.03125
28	3	8	(-6,7)	(-8,5)	2.828427125	0.015625	(-46,-21)	48.82622246	0
29	3	9	(-4,0)	(-7,-3)	4.242640687	0	(-46,-21)	46.95742753	0.03125
30	3	10	(5,9)	(-7,-3)	16.97056275	0.015625	(-46,-21)	59.16924877	0
31	4	1	(2,-7)	(14,-17)	15.62049935	0.015625	(18,-35)	32.24903099	0
32	4	2	(-9,8)	(14,-17)	33.9705755	0.03125	(-7,-14)	22.09072203	0.03125
33	4	3	(0,-5)	(14,-17)	18.43908891	0	(18,-35)	34.98571137	0
34	4	4	(9,1)	(14,-17)	18.68154169	0.03125	(-7,-14)	21.9317122	0.03125
35	4	5	(2,-4)	(14,-17)	17.69180601	0.015625	(18,-35)	34.88552709	0.03125
36	4	6	(5,8)	(14,-17)	26.57066051	0.03125	(-7,-14)	25.05992817	0.03125
37	4	7	(-7,-6)	(14,-17)	23.70653918	0	(18,-35)	38.28837944	0.03125
38	4	8	(-6,7)	(14,-17)	31.2409987	0.015625	(-7,-14)	21.02379604	0.03125
39	4	9	(-4,0)	(14,-17)	24.75883681	0	(18,-35)	41.34005322	0
40	4	10	(5,9)	(14,-17)	27.51363298	0.015625	(-7,-14)	25.94224354	0.03125
41	5	1	(2,-7)	(-7,-14)	11.40175425	0.015625	(18,-35)	32.24903099	0
42	5	2	(-9,8)	(-7,-14)	22.09072203	0.03125	(-7,-14)	22.09072203	0.0625
43	5	3	(0,-5)	(-7,-14)	11.40175425	0.015625	(18,-35)	34.98571137	0
44	5	4	(9,1)	(3,24)	23.76972865	0.015625	(-7,-14)	21.9317122	0.03125
45	5	5	(2,-4)	(-7,-14)	13.45362405	0.015625	(18,-35)	34.88552709	0
46	5	6	(5,8)	(3,24)	16.1245155	0.015625	(-7,-14)	25.05992817	0.015625
47	5	7	(-7,-6)	(-7,-14)	8	0	(18,-35)	38.28837944	0
48	5	8	(-6,7)	(-7,-14)	21.02379604	0.03125	(-7,-14)	21.02379604	0.03125
49	5	9	(-4,0)	(-7,-14)	14.31782106	0.015625	(18,-35)	41.34005322	0.03125
50	5	10	(5,9)	(3,24)	15.13274595	0.015625	(-7,-14)	25.94224354	0.046875
51	6	1	(2,-7)	(-7,-14)	11.40175425	0	(14,-30)	25.94224354	0.09375
52	6	2	(-9,8)	(-7,-14)	22.09072203	0.015625	(-23,-16)	27.78488798	0
53	6	3	(0,-5)	(-7,-14)	11.40175425	0	(14,-30)	28.65309756	0.03125
54	6	4	(9,1)	(3,24)	23.76972865	0.015625	(-23,-16)	36.23534186	0
55	6	5	(2,-4)	(-7,-14)	13.45362405	0	(14,-30)	28.63564213	0.03125
56	6	6	(5,8)	(3,24)	16.1245155	0.015625	(-23,-16)	36.87817783	0
57	6	7	(-7,-6)	(-7,-14)	8	0	(1,-36)	31.04834939	0
58	6	8	(-6,7)	(-7,-14)	21.02379604	0.015625	(-23,-16)	28.60069929	0
59	6	9	(-4,0)	(-7,-14)	14.31782106	0.03125	(-23,-16)	24.8394847	0
60	6	10	(5,9)	(3,24)	15.13274595	0.015625	(-23,-16)	37.53664876	0
61	7	1	(2,-7)	(7,-11)	6.403124237	0	(9,-15)	10.63014581	0.03125
62	7	2	(-9,8)	(7,-11)	24.8394847	0	(-9,-6)	14	0
63	7	3	(0,-5)	(-6,-1)	7.211102551	0.03125	(9,-15)	13.45362405	0.015625
64	7	4	(9,1)	(7,-11)	12.16552506	0	(-2,-6)	13.03840481	0
65	7	5	(2,-4)	(-6,-1)	8.544003745	0.03125	(9,-15)	13.03840481	0.03125
66	7	6	(5,8)	(-6,-1)	14.2126704	0	(-2,-6)	15.65247584	0
67	7	7	(-7,-6)	(7,-11)	14.86606875	0.03125	(9,-15)	18.35755975	0.03125
68	7	8	(-6,7)	(7,-11)	22.20360331	0	(-9,-6)	13.34166406	0
69	7	9	(-4,0)	(-6,-1)	2.236067977	0	(-2,-6)	6.32455532	0.03125
70	7	10	(5,9)	(-6,-1)	14.86606875	0	(-2,-6)	16.55294536	0.03125
71	8	1	(2,-7)	(9,-15)	10.63014581	0	(-2,-33)	26.30589288	0
72	8	2	(-9,8)	(-9,-6)	14	0	(-23,-18)	29.52964612	0
73	8	3	(0,-5)	(9,-15)	13.45362405	0.03125	(-2,-33)	28.0713377	0
74	8	4	(9,1)	(22,16)	19.84943324	0.015625	(10,-14)	15.03329638	0
75	8	5	(2,-4)	(9,-15)	13.03840481	0	(-2,-33)	29.27456234	0.03125
76	8	6	(5,8)	(-6,17)	14.2126704	0.015625	(-23,-18)	38.20994635	0
77	8	7	(-7,-6)	(-9,-6)	2	0	(-2,-33)	27.45906044	0.015625
78	8	8	(-6,7)	(-9,-6)	13.34166406	0.03125	(-23,-18)	30.23243292	0
79	8	9	(-4,0)	(-9,12)	13	0	(-2,-33)	33.06055051	0.015625
80	8	10	(5,9)	(-6,17)	13.60147051	0.03125	(-23,-18)	38.89730068	0.015625
81	9	1	(2,-7)	(-9,-23)	19.41648784	0	(-2,-33)	26.30589288	0.015625
82	9	2	(-9,8)	(-23,-18)	29.52964612	0	(-23,-18)	29.52964612	0.03125
83	9	3	(0,-5)	(-9,-23)	20.1246118	0	(-2,-33)	28.0713377	0
84	9	4	(9,1)	(-9,6)	18.68154169	0.015625	(10,-14)	15.03329638	0.03125
85	9	5	(2,-4)	(-9,-23)	21.9544984	0	(-2,-33)	29.27456234	0
86	9	6	(5,8)	(-9,6)	14.14213562	0.015625	(-23,-18)	38.20994635	0.03125
87	9	7	(-7,-6)	(-9,-23)	17.11724277	0	(-2,-33)	27.45906044	0
88	9	8	(-6,7)	(-9,-23)	30.14962686	0.03125	(-23,-18)	30.23243292	0.046875
89	9	9	(-4,0)	(-9,-23)	23.53720459	0	(-2,-33)	33.06055051	0
90	9	10	(5,9)	(-9,6)	14.31782106	0.015625	(-23,-18)	38.89730068	0.03125
91	10	1	(2,-7)	(-9,-23)	19.41648784	0.03125	(14,-37)	32.31098884	0.03125
92	10	2	(-9,8)	(-23,-18)	29.52964612	0	(-38,-12)	35.22782991	0
93	10	3	(0,-5)	(-9,-23)	20.1246118	0	(14,-37)	34.92849839	0.03125
94	10	4	(9,1)	(-9,6)	18.68154169	0.015625	(-3,-5)	13.41640786	0
95	10	5	(2,-4)	(-9,-23)	21.9544984	0	(-3,-5)	5.099019514	0.03125
96	10	6	(5,8)	(-9,6)	14.14213562	0.015625	(-3,-5)	15.26433752	0.015625
97	10	7	(-7,-6)	(-9,-23)	17.11724277	0	(14,-37)	37.44329045	0.03125
98	10	8	(-6,7)	(-9,-23)	30.14962686	0.015625	(-38,-12)	37.21558813	0
99	10	9	(-4,0)	(-9,-23)	23.53720459	0	(-3,-5)	5.099019514	0.015625
100	10	10	(5,9)	(-9,6)	14.31782106	0.015625	(-3,-5)	16.1245155	0.03125

Fig. 14. Comparing PST and PQT in terms of time and distance for 100 input points

				Polar Split Tree (200)			Polar Quad Tree (200)		
Row#	DataSet#	QuerySet#	QueryPoint	1NN	Distance	Time(ns)	1NN	Distance	Time(ns)
1	1	1	(2,-7)	(0,-11)	4.472135955	0.015625	(6,-43)	36.22154055	0.03125
2	1	2	(-9,8)	(0,-11)	21.02379604	0	(-29,-10)	26.90724809	0
3	1	3	(0,-5)	(0,-11)	6	0.015625	(6,-43)	38.47076812	0.046875
4	1	4	(9,1)	(-4,6)	13.92838828	0	(-29,-10)	39.56008089	0
5	1	5	(2,-4)	(0,-11)	7.280109889	0.015625	(6,-43)	39.20459157	0.046875
6	1	6	(5,8)	(0,-11)	19.6468827	0	(-29,-10)	38.47076812	0
7	1	7	(-7,-6)	(0,-11)	8.602325267	0.015625	(6,-43)	39.2173431	0.03125
8	1	8	(-6,7)	(0,-11)	18.97366596	0.015625	(-29,-10)	28.60069929	0
9	1	9	(-4,0)	(0,-11)	11.70469991	0	(-29,-10)	26.92582404	0
10	1	10	(5,9)	(0,-11)	20.61552813	0	(-29,-10)	38.94868419	0
11	2	1	(2,-7)	(-5,-15)	10.63014581	0	(9,-40)	33.73425559	0
12	2	2	(-9,8)	(-5,-15)	23.34523506	0.015625	(-74,-27)	73.8241153	0.03125
13	2	3	(0,-5)	(-5,-15)	11.18033989	0	(9,-40)	36.138622	0.046875
14	2	4	(9,1)	(-5,-15)	21.26029163	0	(-74,-27)	87.59566199	0
15	2	5	(2,-4)	(-5,-15)	13.03840481	0.03125	(9,-40)	36.67424164	0.046875
16	2	6	(5,8)	(-17,15)	23.08679276	0.015625	(-74,-27)	86.40601831	0
17	2	7	(-7,-6)	(-5,-15)	9.219544457	0.015625	(9,-40)	37.57658846	0.03125
18	2	8	(-6,7)	(-5,-15)	22.02271555	0.015625	(-74,-27)	76.02631123	0
19	2	9	(-4,0)	(-5,-15)	15.03329638	0	(-74,-27)	75.02666193	0
20	2	10	(5,9)	(-17,15)	22.8035085	0	(-74,-27)	86.81589716	0.015625
21	3	1	(2,-7)	(-14,-27)	25.61249695	0.046875	(9,-40)	33.73425559	0
22	3	2	(-9,8)	(-14,-27)	35.35533906	0	(-74,-27)	73.8241153	0.015625
23	3	3	(0,-5)	(-14,-27)	26.07680962	0	(9,-40)	36.138622	0
24	3	4	(9,1)	(9,-40)	41	0.03125	(-74,-27)	87.59566199	0.03125
25	3	5	(2,-4)	(-14,-27)	28.01785145	0	(9,-40)	36.67424164	0
26	3	6	(5,8)	(-22,32)	36.12478374	0.03125	(-74,-27)	86.40601831	0
27	3	7	(-7,-6)	(-14,-27)	22.13594362	0	(9,-40)	37.57658846	0
28	3	8	(-6,7)	(-14,-27)	34.92849839	0	(-74,-27)	76.02631123	0
29	3	9	(-4,0)	(-14,-27)	28.7923601	0	(-74,-27)	75.02666193	0.03125
30	3	10	(5,9)	(-22,32)	35.4682957	0.03125	(-74,-27)	86.81589716	0
31	4	1	(2,-7)	(-14,-27)	25.61249695	0	(13,-28)	23.70653918	0
32	4	2	(-9,8)	(-14,-27)	35.35533906	0	(-38,5)	29.15475947	0
33	4	3	(0,-5)	(-14,-27)	26.07680962	0.015625	(13,-28)	26.41968963	0
34	4	4	(9,1)	(9,-40)	41	0.03125	(-38,5)	47.16990566	0
35	4	5	(2,-4)	(-14,-27)	28.01785145	0.046875	(13,-28)	26.40075756	0.03125
36	4	6	(5,8)	(-22,32)	36.12478374	0.015625	(-38,5)	43.10452412	0
37	4	7	(-7,-6)	(-14,-27)	22.13594362	0	(-2,-32)	26.47640459	0.03125
38	4	8	(-6,7)	(-14,-27)	34.92849839	0	(-38,5)	32.06243908	0
39	4	9	(-4,0)	(-14,-27)	28.7923601	0	(-2,-32)	32.06243908	0.015625
40	4	10	(5,9)	(-22,32)	35.4682957	0.03125	(-38,5)	43.18564576	0
41	5	1	(2,-7)	(-2,-32)	25.3179778	0.015625	(16,-14)	15.65247584	0.03125
42	5	2	(-9,8)	(-9,1)	7	0	(-49,-11)	44.28317965	0
43	5	3	(0,-5)	(-2,-32)	27.07397274	0.03125	(16,-14)	18.35755975	0.03125
44	5	4	(9,1)	(-2,-32)	34.78505426	0.03125	(-49,-11)	59.22837158	0
45	5	5	(2,-4)	(-2,-32)	28.28427125	0.015625	(16,-14)	17.20465053	0.03125
46	5	6	(5,8)	(-2,-32)	40.60788101	0.015625	(-49,-11)	57.24508713	0.03125
47	5	7	(-7,-6)	(-2,-32)	26.47640459	0.015625	(16,-14)	24.35159132	0.03125
48	5	8	(-6,7)	(-2,-32)	39.20459157	0.015625	(-49,-11)	46.61544808	0
49	5	9	(-4,0)	(-2,-32)	32.06243908	0.015625	(-49,-11)	46.32493929	0.015625
50	5	10	(5,9)	(-2,-32)	41.59326869	0.015625	(-49,-11)	57.5847202	0.03125

				Polar Split Tree (200)			Polar Quad Tree (200)		
Row#	DataSet#	QuerySet#	QueryPoint	1NN	Distance	Time(ns)	1NN	Distance	Time(ns)
51	6	1	(2,-7)	(16,-14)	15.65247584	0.015625	(16,-14)	15.65247584	0.03125
52	6	2	(-9,8)	(16,-14)	33.30165161	0.015625	(-49,-11)	44.28317965	0.03125
53	6	3	(0,-5)	(16,-14)	18.35755975	0	(16,-14)	18.35755975	0.03125
54	6	4	(9,1)	(16,-14)	16.55294536	0.015625	(-49,-11)	59.22837158	0.03125
55	6	5	(2,-4)	(16,-14)	17.20465053	0.015625	(16,-14)	17.20465053	0.03125
56	6	6	(5,8)	(-12,15)	18.38477631	0	(-49,-11)	57.24508713	0
57	6	7	(-7,-6)	(16,-14)	24.35159132	0.015625	(16,-14)	24.35159132	0.03125
58	6	8	(-6,7)	(16,-14)	30.41381265	0	(-49,-11)	46.61544808	0
59	6	9	(-4,0)	(16,-14)	24.41311123	0.015625	(-49,-11)	46.32493929	0.03125
60	6	10	(5,9)	(-12,15)	18.02775638	0.015625	(-49,-11)	57.5847202	0
61	7	1	(2,-7)	(16,-14)	15.65247584	0.015625	(30,-45)	47.20169488	0.015625
62	7	2	(-9,8)	(16,-14)	33.30165161	0.03125	(-51,-2)	43.17406629	0
63	7	3	(0,-5)	(16,-14)	18.35755975	0.015625	(30,-45)	50	0.03125
64	7	4	(9,1)	(16,-14)	16.55294536	0.03125	(30,-45)	50.56678752	0
65	7	5	(2,-4)	(16,-14)	17.20465053	0.015625	(30,-45)	49.64876635	0.03125
66	7	6	(5,8)	(-12,15)	18.38477631	0.015625	(-51,-2)	56.88585061	0
67	7	7	(-7,-6)	(16,-14)	24.35159132	0	(30,-45)	53.75872022	0.03125
68	7	8	(-6,7)	(16,-14)	30.41381265	0.015625	(-51,-2)	45.89117562	0
69	7	9	(-4,0)	(16,-14)	24.41311123	0.015625	(30,-45)	56.40035461	0.046875
70	7	10	(5,9)	(-12,15)	18.02775638	0	(-51,-2)	57.07013229	0
71	8	1	(2,-7)	(-16,-33)	31.6227766	0.015625	(1,-15)	8.062257748	0.03125
72	8	2	(-9,8)	(-16,-33)	41.59326869	0	(-92,-18)	86.97700846	0
73	8	3	(0,-5)	(-16,-33)	32.24903099	0	(1,-15)	10.04987562	0.03125
74	8	4	(9,1)	(-16,-33)	42.20189569	0.03125	(-42,-29)	59.16924877	0
75	8	5	(2,-4)	(-16,-33)	34.13209633	0.03125	(1,-15)	11.04536102	0.046875
76	8	6	(5,8)	(-42,5)	47.09564736	0.015625	(-92,-18)	100.4241007	0
77	8	7	(-7,-6)	(-16,-33)	28.46049894	0	(1,-15)	12.04159458	0.03125
78	8	8	(-6,7)	(-16,-33)	41.23105626	0.015625	(-92,-18)	89.56003573	0
79	8	9	(-4,0)	(-16,-33)	35.11409973	0	(-92,-18)	89.82204629	0
80	8	10	(5,9)	(-42,5)	47.16990566	0.015625	(-92,-18)	100.6876358	0
81	9	1	(2,-7)	(1,-15)	8.062257748	0.03125	(1,-15)	8.062257748	0.046875
82	9	2	(-9,8)	(1,-15)	25.07987241	0	(-92,-18)	86.97700846	0
83	9	3	(0,-5)	(1,-15)	10.04987562	0	(1,-15)	10.04987562	0.03125
84	9	4	(9,1)	(1,11)	12.80624847	0.046875	(-42,-29)	59.16924877	0.03125
85	9	5	(2,-4)	(1,-15)	11.04536102	0	(1,-15)	11.04536102	0.03125
86	9	6	(5,8)	(1,-15)	23.34523506	0	(-92,-18)	100.4241007	0.03125
87	9	7	(-7,-6)	(1,-15)	12.04159458	0.015625	(1,-15)	12.04159458	0
88	9	8	(-6,7)	(1,-15)	23.08679276	0.015625	(-92,-18)	89.56003573	0.015625
89	9	9	(-4,0)	(1,-15)	15.8113883	0	(-92,-18)	89.82204629	0
90	9	10	(5,9)	(1,-15)	24.33105012	0.015625	(-92,-18)	100.6876358	0.03125
91	10	1	(2,-7)	(1,-15)	8.062257748	0	(9,-16)	11.40175425	0.03125
92	10	2	(-9,8)	(1,-15)	25.07987241	0.03125	(-41,-49)	65.36818798	0.03125
93	10	3	(0,-5)	(1,-15)	10.04987562	0.015625	(9,-16)	14.2126704	0
94	10	4	(9,1)	(1,11)	12.80624847	0.046875	(9,-16)	17	0.03125
95	10	5	(2,-4)	(1,-15)	11.04536102	0	(9,-16)	13.89244399	0.015625
96	10	6	(5,8)	(1,-15)	23.34523506	0.015625	(-41,-49)	73.24616031	0.03125
97	10	7	(-7,-6)	(1,-15)	12.04159458	0.015625	(9,-16)	18.86796226	0.015625
98	10	8	(-6,7)	(1,-15)	23.08679276	0	(-41,-49)	66.03786792	0.015625
99	10	9	(-4,0)	(1,-15)	15.8113883	0.03125	(9,-16)	20.61552813	0
100	10	10	(5,9)	(1,-15)	24.33105012	0	(-41,-49)	74.02702209	0.03125

Fig. 15. Comparing PST and PQT in terms of time and distance for 200 input points

				Polar Split Tree (500)			Polar Quad Tree (500)		
Row#	DataSet#	QuerySet#	QueryPoint	1NN	Distance	Time(ns)	1NN	Distance	Time(ns)
1	1	1	(2,-7)	(-48,3)	50.99019514	0.03125	(-63,-12)	65.19202405	0.03125
2	1	2	(-9,8)	(-48,3)	39.3192065	0.03125	(-63,-12)	57.5847202	0
3	1	3	(0,-5)	(-48,3)	48.66210024	0	(-63,-12)	63.38769597	0
4	1	4	(9,1)	(-23,50)	58.52349955	0.046875	(-63,-12)	73.1641989	0.0625
5	1	5	(2,-4)	(-48,3)	50.48762225	0.015625	(-63,-12)	65.49045732	0
6	1	6	(5,8)	(-48,3)	53.23532662	0.015625	(-63,-12)	70.88018059	0
7	1	7	(-7,-6)	(-48,3)	41.97618372	0	(-63,-12)	56.32051136	0
8	1	8	(-6,7)	(-48,3)	42.19004622	0	(-63,-12)	60.08327554	0
9	1	9	(-4,0)	(-48,3)	44.10215414	0	(-63,-12)	60.20797289	0
10	1	10	(5,9)	(-48,3)	53.33854141	0.015625	(-63,-12)	71.16881339	0
11	2	1	(2,-7)	(-63,-12)	65.19202405	0	(-4,-62)	55.32630477	0.03125
12	2	2	(-9,8)	(-63,-12)	57.5847202	0	(-4,-62)	70.17834424	0
13	2	3	(0,-5)	(-63,-12)	63.38769597	0.03125	(-4,-62)	57.14017851	0
14	2	4	(9,1)	(-63,-12)	73.1641989	0	(-4,-62)	64.32728814	0.03125
15	2	5	(2,-4)	(-63,-12)	65.49045732	0	(-4,-62)	58.30951895	0
16	2	6	(5,8)	(-63,-12)	70.88018059	0	(-4,-62)	70.57619995	0
17	2	7	(-7,-6)	(-63,-12)	56.32051136	0	(-4,-62)	56.08029957	0
18	2	8	(-6,7)	(-63,-12)	60.08327554	0	(-4,-62)	69.02897942	0
19	2	9	(-4,0)	(-63,-12)	60.20797289	0	(-4,-62)	62	0
20	2	10	(5,9)	(-63,-12)	71.16881339	0	(-4,-62)	71.56814934	0
21	3	1	(2,-7)	(-48,-60)	72.86288493	0.015625	(-11,-25)	22.20360331	0.03125
22	3	2	(-9,8)	(-48,-60)	78.39005039	0.015625	(-11,-25)	33.06055051	0
23	3	3	(0,-5)	(-48,-60)	73	0	(-11,-25)	22.82542442	0.015625
24	3	4	(9,1)	(-48,-60)	83.48652586	0.015625	(-11,-25)	32.80243893	0.046875
25	3	5	(2,-4)	(-48,-60)	75.07329752	0.015625	(-11,-25)	24.69817807	0.03125
26	3	6	(5,8)	(-48,-60)	86.21484791	0	(-11,-25)	36.67424164	0.03125
27	3	7	(-7,-6)	(-48,-60)	67.80117993	0.015625	(-11,-25)	19.41648784	0.03125
28	3	8	(-6,7)	(-48,-60)	79.07591289	0.015625	(-11,-25)	32.38826948	0
29	3	9	(-4,0)	(-48,-60)	74.40430095	0.015625	(-11,-25)	25.96150997	0.03125
30	3	10	(5,9)	(-48,-60)	87.00574694	0.015625	(-11,-25)	37.57658846	0
31	4	1	(2,-7)	(-11,-25)	22.20360331	0.015625	(24,-5)	22.09072203	0.03125
32	4	2	(-9,8)	(-11,-25)	33.06055051	0.015625	(24,-5)	35.4682957	0.03125
33	4	3	(0,-5)	(-11,-25)	22.82542442	0	(24,-5)	24	0.015625
34	4	4	(9,1)	(-22,33)	44.55333882	0	(24,-5)	16.15549442	0
35	4	5	(2,-4)	(-11,-25)	24.69817807	0	(24,-5)	22.02271555	0.03125
36	4	6	(5,8)	(-22,33)	36.79673899	0.015625	(24,-5)	23.02172887	0.046875
37	4	7	(-7,-6)	(-11,-25)	19.41648784	0.015625	(24,-5)	31.01612484	0.03125
38	4	8	(-6,7)	(-11,-25)	32.38826948	0.015625	(24,-5)	32.31098884	0.03125
39	4	9	(-4,0)	(-11,-25)	25.96150997	0.015625	(24,-5)	28.44292531	0.03125
40	4	10	(5,9)	(-22,33)	36.12478374	0	(24,-5)	23.60084744	0.03125
41	5	1	(2,-7)	(24,-5)	22.09072203	0	(-16,-18)	21.09502311	0
42	5	2	(-9,8)	(-43,11)	34.13209633	0.046875	(-16,-18)	26.92582404	0
43	5	3	(0,-5)	(24,-5)	24	0.015625	(-16,-18)	20.61552813	0.03125
44	5	4	(9,1)	(24,-5)	16.15549442	0.015625	(-16,-18)	31.40063694	0.03125
45	5	5	(2,-4)	(24,-5)	22.02271555	0.015625	(-16,-18)	22.8035085	0.015625
46	5	6	(5,8)	(24,-5)	23.02172887	0.015625	(-16,-18)	33.42154993	0.015625
47	5	7	(-7,-6)	(24,-5)	31.01612484	0.015625	(-16,-18)	15	0.03125
48	5	8	(-6,7)	(24,-5)	32.31098884	0.015625	(-16,-18)	26.92582404	0.03125
49	5	9	(-4,0)	(24,-5)	28.44292531	0.015625	(-16,-18)	21.63330765	0.015625
50	5	10	(5,9)	(24,-5)	23.60084744	0	(-16,-18)	34.20526275	0

				Polar Split Tree (500)			Polar Quad Tree (500)		
Row#	DataSet#	QuerySet#	QueryPoint	1NN	Distance	Time(ns)	1NN	Distance	Time(ns)
51	6	1	(2,-7)	(1,-8)	1.414213562	0.015625	(21,-70)	65.80273551	0
52	6	2	(-9,8)	(1,-8)	18.86796226	0.015625	(-58,-49)	75.16648189	0.03125
53	6	3	(0,-5)	(1,-8)	3.16227766	0.015625	(-58,-49)	72.80109889	0.03125
54	6	4	(9,1)	(1,-8)	12.04159458	0.015625	(-58,-49)	83.60023923	0.03125
55	6	5	(2,-4)	(1,-8)	4.123105626	0.015625	(-58,-49)	75	0
56	6	6	(5,8)	(1,-8)	16.4924225	0.03125	(-58,-49)	84.95881355	0.03125
57	6	7	(-7,-6)	(1,-8)	8.246211251	0	(21,-70)	69.85699679	0.03125
58	6	8	(-6,7)	(1,-8)	16.55294536	0.015625	(-58,-49)	76.4198927	0.03125
59	6	9	(-4,0)	(1,-8)	9.433981132	0.015625	(-58,-49)	72.9177619	0.015625
60	6	10	(5,9)	(1,-8)	17.4642492	0.03125	(-58,-49)	85.63293759	0
61	7	1	(2,-7)	(17,2)	17.49285568	0.03125	(-102,-15)	104.3072385	0
62	7	2	(-9,8)	(17,2)	26.68332813	0	(-102,-15)	95.8018789	0
63	7	3	(0,-5)	(17,2)	18.38477631	0	(-102,-15)	102.4890238	0.03125
64	7	4	(9,1)	(17,2)	8.062257748	0.046875	(-102,-15)	112.1472247	0.015625
65	7	5	(2,-4)	(17,2)	16.15549442	0.015625	(-102,-15)	104.5801128	0
66	7	6	(5,8)	(17,2)	13.41640786	0.015625	(-102,-15)	109.4440496	0
67	7	7	(-7,-6)	(17,2)	25.29822128	0.03125	(-102,-15)	95.4253635	0.046875
68	7	8	(-6,7)	(17,2)	23.53720459	0.015625	(-102,-15)	98.48857802	0.03125
69	7	9	(-4,0)	(17,2)	21.09502311	0	(-102,-15)	99.14131329	0.03125
70	7	10	(5,9)	(17,2)	13.89244399	0.015625	(-102,-15)	109.658561	0
71	8	1	(2,-7)	(-24,-39)	41.23105626	0.015625	(-19,-38)	37.44329045	0
72	8	2	(-9,8)	(-24,-39)	49.33558553	0.015625	(-19,-38)	47.07440918	0
73	8	3	(0,-5)	(-24,-39)	41.61730409	0.015625	(-19,-38)	38.07886553	0
74	8	4	(9,1)	(-24,-39)	51.85556865	0.015625	(-19,-38)	48.01041554	0.03125
75	8	5	(2,-4)	(-24,-39)	43.60045871	0.015625	(-19,-38)	39.96248241	0.03125
76	8	6	(5,8)	(-24,-39)	55.22680509	0.03125	(-19,-38)	51.88448708	0
77	8	7	(-7,-6)	(-24,-39)	37.12142239	0.015625	(-19,-38)	34.17601498	0
78	8	8	(-6,7)	(-24,-39)	49.39635614	0.015625	(-19,-38)	46.84015371	0
79	8	9	(-4,0)	(-24,-39)	43.829214	0.015625	(-19,-38)	40.85339643	0.015625
80	8	10	(5,9)	(-24,-39)	56.08029957	0.015625	(-19,-38)	52.77309921	0.015625
81	9	1	(2,-7)	(5,-26)	19.23538406	0.015625	(8,-52)	45.39823785	0
82	9	2	(-9,8)	(5,-26)	36.76955262	0.015625	(-67,-15)	62.39390996	0
83	9	3	(0,-5)	(5,-26)	21.58703314	0.015625	(-67,-15)	67.74215822	0
84	9	4	(9,1)	(-32,35)	53.26349594	0.015625	(-67,-15)	77.66595136	0.03125
85	9	5	(2,-4)	(5,-26)	22.20360331	0.015625	(-67,-15)	69.87131028	0.03125
86	9	6	(5,8)	(-32,35)	45.80392996	0.015625	(-67,-15)	75.58438992	0.046875
87	9	7	(-7,-6)	(5,-26)	23.32380758	0.015625	(8,-52)	48.38388161	0
88	9	8	(-6,7)	(5,-26)	34.78505426	0.015625	(-67,-15)	64.84597135	0
89	9	9	(-4,0)	(5,-26)	27.51363298	0	(-67,-15)	64.76109943	0
90	9	10	(5,9)	(-32,35)	45.22167622	0.015625	(-67,-15)	75.89466384	0.03125
91	10	1	(2,-7)	(25,-22)	27.45906044	0	(-121,-43)	128.1600562	0
92	10	2	(-9,8)	(25,-22)	45.3431362	0.046875	(-121,-43)	123.0650235	0
93	10	3	(0,-5)	(25,-22)	30.23243292	0	(-121,-43)	126.8266534	0
94	10	4	(9,1)	(25,-22)	28.01785145	0	(-121,-43)	137.2443077	0
95	10	5	(2,-4)	(25,-22)	29.20616373	0	(-121,-43)	129.034879	0
96	10	6	(5,8)	(25,-22)	36.05551275	0	(-121,-43)	135.9301291	0.03125
97	10	7	(-7,-6)	(25,-22)	35.77708764	0.03125	(-121,-43)	119.8540779	0.03125
98	10	8	(-6,7)	(25,-22)	42.44997055	0	(-121,-43)	125.399362	0
99	10	9	(-4,0)	(25,-22)	36.40054945	0.015625	(-121,-43)	124.6515142	0
100	10	10	(5,9)	(25,-22)	36.89173349	0	(-121,-43)	136.3084737	0

Fig. 16. Comparing PST and PQT in terms of time and distance for 500 input points

				Polar Split Tree (1000)			Polar Quad Tree (1000)		
Row#	DataSet#	QuerySet#	QueryPoint	1NN	Distance	Time(ns)	1NN	Distance	Time(ns)
1	1	1	(2,-7)	(10,-28)	22.47220505	0.046875	(18,-74)	68.8839604	0.046875
2	1	2	(-9,8)	(10,-28)	40.70626487	0	(-147,4)	138.0579588	0
3	1	3	(0,-5)	(10,-28)	25.07987241	0.015625	(18,-74)	71.30918594	0.03125
4	1	4	(9,1)	(10,-28)	29.01723626	0	(-147,4)	156.0288435	0
5	1	5	(2,-4)	(10,-28)	25.29822128	0	(18,-74)	71.80529228	0.03125
6	1	6	(5,8)	(10,-28)	36.34556369	0	(-147,4)	152.0526225	0.03125
7	1	7	(-7,-6)	(10,-28)	27.80287755	0.03125	(18,-74)	72.44998275	0.03125
8	1	8	(-6,7)	(10,-28)	38.48376281	0.015625	(-147,4)	141.0319113	0
9	1	9	(-4,0)	(10,-28)	31.30495168	0	(-147,4)	143.0559331	0.015625
10	1	10	(5,9)	(10,-28)	37.33630941	0	(-147,4)	152.0822146	0
11	2	1	(2,-7)	(18,-23)	22.627417	0.046875	(21,-129)	123.4706443	0.015625
12	2	2	(-9,8)	(18,-23)	41.10960958	0.015625	(-213,-78)	221.3865398	0
13	2	3	(0,-5)	(18,-23)	25.45584412	0	(21,-129)	125.7656551	0.03125
14	2	4	(9,1)	(18,-23)	25.63201124	0.015625	(21,-129)	130.5526714	0
15	2	5	(2,-4)	(18,-23)	24.8394847	0.015625	(21,-129)	126.4357544	0
16	2	6	(5,8)	(18,-23)	33.61547263	0.015625	(-213,-78)	234.3501654	0.015625
17	2	7	(-7,-6)	(18,-23)	30.23243292	0.03125	(21,-129)	126.1467399	0
18	2	8	(-6,7)	(18,-23)	38.41874542	0.015625	(-213,-78)	223.7722056	0.03125
19	2	9	(-4,0)	(18,-23)	31.82766093	0.03125	(21,-129)	131.4001522	0
20	2	10	(5,9)	(18,-23)	34.53983208	0.015625	(-213,-78)	234.7189809	0
21	3	1	(2,-7)	(-17,-57)	53.48831648	0.015625	(-100,-70)	119.8874472	0
22	3	2	(-9,8)	(-64,-3)	56.08921465	0	(-100,-70)	119.8540779	0.03125
23	3	3	(0,-5)	(-17,-57)	54.70831747	0	(-100,-70)	119.2686044	0
24	3	4	(9,1)	(-17,-57)	63.56099433	0	(-100,-70)	130.0845879	0
25	3	5	(2,-4)	(-17,-57)	56.30275304	0.015625	(-100,-70)	121.4907404	0
26	3	6	(5,8)	(-17,-57)	68.62215386	0	(-100,-70)	130.8013761	0
27	3	7	(-7,-6)	(-17,-57)	51.97114584	0	(-100,-70)	112.8937554	0
28	3	8	(-6,7)	(-64,-3)	58.85575588	0.015625	(-100,-70)	121.5113163	0
29	3	9	(-4,0)	(-17,-57)	58.46366393	0.015625	(-100,-70)	118.8107739	0
30	3	10	(5,9)	(-17,-57)	69.57010852	0.015625	(-100,-70)	131.4001522	0
31	4	1	(2,-7)	(11,-33)	27.51363298	0	(31,-43)	46.22769733	0
32	4	2	(-9,8)	(11,-33)	45.61797891	0.015625	(-137,-35)	135.0296264	0
33	4	3	(0,-5)	(11,-33)	30.08321791	0.03125	(-137,-35)	140.2462121	0.03125
34	4	4	(9,1)	(11,-33)	34.05877273	0	(-137,-35)	150.3728699	0
35	4	5	(2,-4)	(11,-33)	30.3644529	0.015625	(-137,-35)	142.4148869	0
36	4	6	(5,8)	(11,-33)	41.43669871	0	(-137,-35)	148.3677863	0
37	4	7	(-7,-6)	(11,-33)	32.44996148	0.015625	(31,-43)	53.03772242	0
38	4	8	(-6,7)	(11,-33)	43.46262762	0	(-137,-35)	137.5681649	0
39	4	9	(-4,0)	(11,-33)	36.24913792	0.015625	(-137,-35)	137.5281789	0
40	4	10	(5,9)	(11,-33)	42.42640687	0.015625	(-137,-35)	148.6606875	0
41	5	1	(2,-7)	(31,-43)	46.22769733	0.046875	(-11,-6)	13.03840481	0.015625
42	5	2	(-9,8)	(31,-43)	64.81512169	0.015625	(-11,-6)	14.14213562	0
43	5	3	(0,-5)	(31,-43)	49.04079934	0.03125	(-11,-6)	11.04536102	0.03125
44	5	4	(9,1)	(31,-43)	49.1934955	0.0625	(-11,-6)	21.1896201	0.015625
45	5	5	(2,-4)	(31,-43)	48.60041152	0.03125	(-11,-6)	13.15294644	0.015625
46	5	6	(5,8)	(31,-43)	57.24508713	0.03125	(-11,-6)	21.26029163	0.015625
47	5	7	(-7,-6)	(31,-43)	53.03772242	0.046875	(-11,-6)	4	0
48	5	8	(-6,7)	(31,-43)	62.20128616	0.015625	(-11,-6)	13.92838828	0.03125
49	5	9	(-4,0)	(31,-43)	55.4436651	0.015625	(-11,-6)	9.219544457	0.03125
50	5	10	(5,9)	(31,-43)	58.13776741	0.03125	(-11,-6)	21.9317122	0.03125

				Polar Split Tree (1000)			Polar Quad Tree (1000)		
Row#	DataSet#	QuerySet#	QueryPoint	1NN	Distance	Time(ns)	1NN	Distance	Time(ns)
51	6	1	(2,-7)	(-11,-6)	13.03840481	0.015625	(38,-140)	137.7860661	0.03125
52	6	2	(-9,8)	(-11,-6)	14.14213562	0	(-192,-53)	192.8989373	0.015625
53	6	3	(0,-5)	(-11,-6)	11.04536102	0	(38,-140)	140.2462121	0.03125
54	6	4	(9,1)	(-11,-6)	21.1896201	0	(-192,-53)	208.1273649	0.03125
55	6	5	(2,-4)	(-11,-6)	13.15294644	0.015625	(38,-140)	140.6840432	0.015625
56	6	6	(5,8)	(-11,-6)	21.26029163	0	(-192,-53)	206.2280291	0.015625
57	6	7	(-7,-6)	(-11,-6)	4	0	(38,-140)	141.3541651	0.015625
58	6	8	(-6,7)	(-11,-6)	13.92838828	0	(-192,-53)	195.4379697	0
59	6	9	(-4,0)	(-11,-6)	9.219544457	0	(-192,-53)	195.3279294	0
60	6	10	(5,9)	(-11,-6)	21.9317122	0.015625	(-192,-53)	206.5260274	0.046875
61	7	1	(2,-7)	(-46,-104)	108.2266141	0.015625	(13,-180)	173.3493582	0.03125
62	7	2	(-9,8)	(-46,-104)	117.9533806	0	(13,-180)	189.2828571	0
63	7	3	(0,-5)	(-46,-104)	109.1650127	0.03125	(13,-180)	175.4821928	0
64	7	4	(9,1)	(-56,93)	112.6454615	0.03125	(13,-180)	181.0441935	0
65	7	5	(2,-4)	(-46,-104)	110.923397	0.03125	(13,-180)	176.343415	0
66	7	6	(5,8)	(-56,93)	104.6231332	0.015625	(13,-180)	188.1701358	0
67	7	7	(-7,-6)	(-46,-104)	105.4751155	0	(13,-180)	175.1456537	0.03125
68	7	8	(-6,7)	(-56,93)	99.47864092	0.015625	(13,-180)	187.9627623	0.03125
69	7	9	(-4,0)	(-56,93)	106.5504575	0	(13,-180)	180.8009956	0
70	7	10	(5,9)	(-56,93)	103.8123307	0.015625	(13,-180)	189.1692364	0
71	8	1	(2,-7)	(-32,-70)	71.58910532	0.03125	(47,-82)	87.46427842	0
72	8	2	(-9,8)	(-86,5)	77.0584194	0.015625	(-109,-39)	110.4943437	0
73	8	3	(0,-5)	(-86,5)	86.57944329	0	(47,-82)	90.21086409	0.03125
74	8	4	(9,1)	(-32,-70)	81.98780397	0	(-109,-39)	124.595345	0.03125
75	8	5	(2,-4)	(-32,-70)	74.24284477	0	(47,-82)	90.04998612	0.03125
76	8	6	(5,8)	(-32,-70)	86.33075929	0.015625	(-109,-39)	123.3085561	0
77	8	7	(-7,-6)	(-86,5)	79.76214641	0	(47,-82)	93.23089617	0.03125
78	8	8	(-6,7)	(-86,5)	80.02499609	0.03125	(-109,-39)	112.8051417	0.03125
79	8	9	(-4,0)	(-86,5)	82.15229759	0.015625	(-109,-39)	112.0089282	0
80	8	10	(5,9)	(-32,-70)	87.23531395	0	(-109,-39)	123.6931688	0
81	9	1	(2,-7)	(-57,49)	81.34494453	0.046875	(34,-128)	125.1598977	0
82	9	2	(-9,8)	(-57,49)	63.12685641	0	(1,-4)	15.62049935	0.015625
83	9	3	(0,-5)	(-57,49)	78.51751397	0	(1,-4)	1.414213562	0
84	9	4	(9,1)	(-57,49)	81.60882305	0.015625	(1,-4)	9.433981132	0.015625
85	9	5	(2,-4)	(-57,49)	79.30952024	0.015625	(1,-4)	1	0.015625
86	9	6	(5,8)	(-57,49)	74.33034374	0.015625	(1,-4)	12.64911064	0
87	9	7	(-7,-6)	(-57,49)	74.33034374	0.015625	(1,-4)	8.246211251	0.03125
88	9	8	(-6,7)	(-57,49)	66.06814664	0	(1,-4)	13.03840481	0.03125
89	9	9	(-4,0)	(-57,49)	72.18032973	0	(1,-4)	6.403124237	0.03125
90	9	10	(5,9)	(-57,49)	73.78346698	0.015625	(1,-4)	13.60147051	0
91	10	1	(2,-7)	(-45,35)	63.03173804	0	(-152,-101)	180.4217282	0.03125
92	10	2	(-9,8)	(-45,35)	45	0	(-256,-8)	247.5176761	0
93	10	3	(0,-5)	(-45,35)	60.20797289	0	(-152,-101)	179.7776404	0.03125
94	10	4	(9,1)	(64,-25)	60.8358447	0.015625	(-152,-101)	190.5911855	0
95	10	5	(2,-4)	(-45,35)	61.07372594	0	(-152,-101)	182.0027472	0
96	10	6	(5,8)	(-45,35)	56.82429058	0.015625	(-152,-101)	191.1282292	0.03125
97	10	7	(-7,-6)	(-45,35)	55.90169944	0	(-152,-101)	173.3493582	0.03125
98	10	8	(-6,7)	(-45,35)	48.01041554	0	(-256,-8)	250.4495957	0
99	10	9	(-4,0)	(-45,35)	53.90732789	0	(-152,-101)	179.1786818	0.03125
100	10	10	(5,9)	(-45,35)	56.35601121	0.046875	(-152,-101)	191.7002869	0.03125

Fig. 17. Comparing PST and PQT in terms of time and distance for 1000 input points

Since we traverse from top to down in the polar split tree, we have at most logn levels. Thus the time complexity of this code is O(logn). If we add the preprocessing steps, the total complexity is O(nlogn).

7. Evaluation of the Proposed Method

Using Matlab programming language 2015b, simulation of this algorithm has been done in a machine with Windows 10 Operating System with CPU Intel core i5 2.5 GHZ, RAM:6GB. The input contains random coordinates between -1000 and 1000 in a 2-dimensional plane. In this paper, searching time and the distance to the nearest neighbor for polar quad tree and polar split tree are compared. The number of random query points change from 100 to 1000. Four types of input with the number of points of 100, 200, 500 and 1000 are tested. This test has been done for 10 types of random input and each random input has been repeated 10 times and for each random input, one random query point is chosen and the nearest neighbor to the query point is calculated for both trees. The results of the distance and time comparison for both algorithms are presented in the following tables of figures 14, 15, 16 and 17.

The graph of comparing two PST and PQT algorithms in terms of time and distance is in shown in figures 18, 19 and 20.

	Polar Split Tree		Polar Quad Tree	
DataSet Size	Time(ns)	Distance	Time(ns)	Distance
100	0.013438	17.2179489	0.017031	28.5644889
200	0.012969	23.7880872	0.01625	47.3654892
500	0.013125	38.9709372	0.014844	62.4753168
1000	0.013438	54.5585853	0.013125	122.750356
10000	0.026563	312.666854	0.045313	762.142584

Fig. 18. Searching Time and Distance to the Nearest Neighbor for PST and PQT

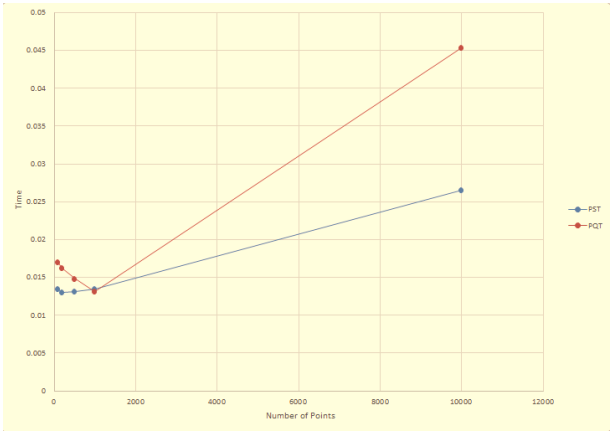


Fig. 19. Comparing PQT and PST in terms of searching Time NN1

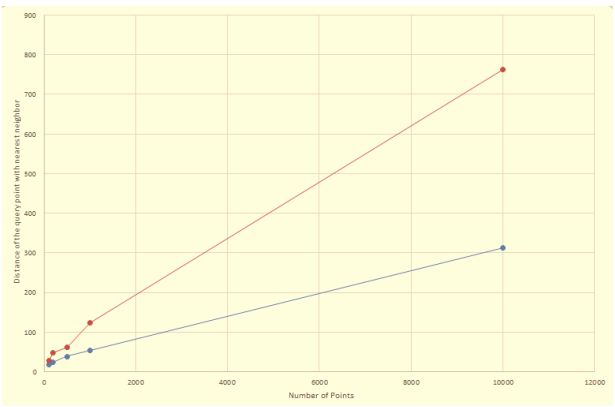


Fig. 20. Comparing PQT and PST in terms of Distance to NN1

8. Conclusion

In this paper, it is assumed that the input data have been distributed in a circle form, the smallest circle containing the input points is drawn and using the mentioned method, the polar split tree is constructed. As it was observed, the results of the nearest neighbor search in terms of time and distance in polar split tree have been better than polar quad tree. This matter indicates the optimality of the polar split tree in nearest neighbor search. This method has several applications including transmitting radio and telecommunication waves form host stations to receivers and searching the receivers. As in these areas we are dealing with circular shapes, it is better to use polar coordinate. B y entering one newly query data using NN1, polar split tree for the set of input points enables us to detect which cell of the tree belongs to this point.

Acknowledgment

We acknowledge the contribution of Marzieh Nazari for the sincere help that went through the initial writing of the paper. We also appreciate the referees for many useful comments and suggestions.

References

[1] Robert Adams and Essex Christopher, Calculus: a complete course (Eighth ed.). Pearson Canada Inc., 2013.#
[2] Pablo Alonso Gonzalez, Optimization of antenna coverage in telecommunication systems, AGH University of Science and Technology, 2018.#
[3] Mark de Berg, Otfried Cheong, Marc van Kreveld, and Mark Overmars, Computational Geometry: Algorithms and Applications. Springer, pp. 309-312, 2008.#
[4] Cristina Costa, Francesco G.B. De Natale and Fabrizio Granelli, Quality Evaluation and Nonuniform Compression of Geometrically Distorted Images Using the Quadtree Distortion Map, EURASIP Journal on Applied Signal Processing, pp. 1899–1911, 2004.#
[5] Ameneh Eskandari, Zahra Nilforoushan and javad ranjbar, A Novel Method to Improve Query in Big Databases Using a Geometric tree Base Algorithm: International Journal of computer & Information Technologies, 5(1): 53-58, 2017.#
[6] Clara Grima and Alberto Marquez, Computational Geometry on Surfaces: Performing Computational Geometry on the Cylinder, the Sphere, the Torus, and the Cone, Kluwer Academic Publishers, 2001.#
[7] Anton Howard, Irl Bivens and Stephen Davis, Calculus (Seventh ed.). Anton Textbooks, Inc., 2002.#
[8] Moritz von Looz, Christian L. Staudt, Henning Meyerhenke and Roman Prutkin, Fast generation of dynamic complex networks with underlying hyperbolic geometry, Karlsruhe Reports in Informatics, 2014.#
[9] Nimrod Megiddo, Linear-time algorithms for Linear programing in R and related problems .: SIAM Journal on Computing, 12(4): 759-776,1983.#
[10] Giri Narasimhan and Michiel Smid, Geometric Spanner Networks.: Cambridge University Press, 2007.#
[11] Joseph O'Rourke, Computational Geometry in C, Cambridge University Press,1998.#
[12] Rina Panigrahy, An Improved Algorithm Finding Nearest Neighbor Using kd-trees, Latin American Symposium on Theoretical Informatics, pp. 387--398, 2008.#
[13] Sudhir Porwal, Quad tree-based level-of-details representation of digital Globe, Defence Science Journal, Vol. 63, No. 1, pp. 89-92, 2013.#
[14] Franco P. Preparata and Michael Ian Shamos, Computational Geometry - An Introduction, Springer, 1985.#
[15] Javad Ranjbar, Zahra Nilforoushan and Ameneh Eskandari, Improved Fingerprint Matching Speed in Large Databases Using Split Tree, 3rd National Conference on Distributed Computing and Big Data Processing, 2017.#
[16] J. R. Sack and J. Urrutia, Handbook of Computational Geometry, Elsevier, 1999.#
[17] Bahram Sadegh Bigham, Ali Mohades and Lidia Ortega, Dynamic Polar Diagram, Information Processing Letters, 109.2, pp. 142-146, 2008.#
[18] Hanan Samet, Foundations of Multidimensional and Metric Data Structures, Elsevier, 2006.#
[19] S. Saric , Z. Bozanic and R. Svalina: Automation in Developing Technical Documentation of Telecommunication Networks, Promet- Traffic- Traffico, Vol. 16, No. 5, pp. 257-262, 2004.#
[20] Ian Stewart and David Tall, Complex Analysis (the Hitchhiker's Guide to the Plane). Cambridge University Press, 1983.#
[21] A. M. Sukhov and D. Yu. Chemodanov, The Neighborhoods Method and Virtual Polar coordinates in Wireless Sensor Networks, Network and Communication Technologies, Vol. 2, No. 1, pp. 19-27, 2013.#
[22] Waldo Tobler and Zitan Chen, A Quad tree for Global Information Storage, Geographical Analysis, Volume18, Issue4, 1986.#
[23] https://www.coursera.org/learn/ml-clustering-and-retrieval/lecture/6eTzw/nn-search-with-kd-rees #

Farzad Bayat received his B.Sc. degree in Software Engineering from Buali Sina University of Hamedan,Hamedan,Iran in 2016. He received the M.Sc. degree in Knowledge Engineering from Kharazmi University, Tehran,Iran, in 2019. His area research interests include Software Development like Android App, Web Site and etc. He studied and worked in CRM (Customer Relationship Management) from Microsoft for two years.

Zahra Nilforoushan received her M.Sc. degree in Pure Mathematics (Algebraic Geometry) and Ph.D in Computer Science (Computational Geometry) from Amir kabir University of Technology, Tehran, Iran in 2003 and 2009 respectively. She served as a lecturer at Dept. of Computer Science, Faculty of Mathematical Sciences and Computer, Kharazmi University, Tehran, Iran from 2009 to 2011. Since 2012 she is with Dept. of Electrical and Computer Engineering, Faculty of Engineering at Kharazmi University as an Assistant Professor. Her research interests are Computational Geometry, Computer Graphics, Computer Vision, Analysis Algorithm and Robotics.

A Multi-objective Multi-agent Optimization Algorithm for the Community Detection Problem

Amir Hossein Hosseinian
Department of Industrial Engineering, University of Islamic Azad University, Tehran North branch, Tehran, Iran
ah_hosseinian@iau-tnb.ac.ir
Vahid Baradaran*
Department of Industrial Engineering, University of Islamic Azad University, Tehran North branch, Tehran, Iran
V_Baradaran@iau-tnb.ac.ir

Received: 24/Feb/2018 Revised: 22/Aug/2018 Accepted: 16/Sep/2018

Abstract

This paper addresses the community detection problem as one of the significant problems in the field of social network analysis. The goal of the community detection problem is to find sub-graphs of a network where they have high density of within-group connections, while they have a lower density of between-group connections. Due to high practical usage of community detection in scientific fields, many researchers developed different algorithms to meet various scientific requirements. However, single-objective optimization algorithms may fail to detect high quality communities of complex networks. In this paper, a novel multi-objective Multi-agent Optimization Algorithm, named the MAOA is proposed to detect communities of complex networks. The MAOA aims to optimize modularity and community score as objective functions, simultaneously. In the proposed algorithm, each feasible solution is considered as an agent and the MAOA organizes agents in multiple groups. The MAOA uses new search operators based on social, autonomous and self-learning behaviors of agents. Moreover, the MAOA uses the weighted sum method (WSM) in finding the global best agent and leader agent of each group. The Pareto solutions obtained by the MAOA is evaluated in terms of several performance measures. The results of the proposed method are compared with the outputs of three meta-heuristics. Experiments results based on five real-world networks show that the MAOA is more efficient in finding better communities than other methods.

Keywords: Community Detection Problem; Complex Networks; Multi-agent Systems; Social Networks.

1. Introduction

Networks are usually used to model complex systems in various fields, such as computer science, physics, biology and sociology [1]. Many complex systems can be structured as networks, such as computer networks, technological networks, collaboration networks, e-mail networks, biological networks, political election networks, etc. A network is defined as a graph where the nodes represent network objects and edges show the relations between them. Each node represents a network member, while an edge between two nodes indicates that there is a relation between two members of a network. For each complex system, there is a structural property called community structure. A community is defined as a group of nodes within a network that have a high density of within-group connections, while they have a lower density of between-group connections [2]. Discovering informative and hidden structures of networks is known as community detection problem. The real number of communities is not known in real-world networks. Therefore, an automatic clustering method is needed to identify the real number of communities in a network. Since the network objects in the same community may usually have similarities, the identified communities can be used in product recommendations, dimensionality reduction, information spreading, link prediction, knowledge sharing and other beneficial applications [3]. Many community detection methods have been developed in the literature which can be used in many practical cases such as product recommendations, reduction of dimensionality in pattern recognition, prediction of links and detection of cancers [4].

Detecting communities of a network can be modeled as an optimization problem. The aim of an optimization-based algorithm is to find an optimal solution with respect to a predefined objective function. Community detection problem is an NP-hard optimization problem [2]. The solutions obtained by single-objective approaches are limited to a particular community structure property. Therefore, if an improper objective function is chosen, these algorithms may fail to find high-quality communities. Besides, in hierarchical networks where there are multiple potential structures, a fixed community structure detected by single-objective approaches may not be appropriate. In this respect, it is desirable to optimize multiple objectives simultaneously so as to explore different potential network structures [4].

To solve the community detection problem which belongs to the set of NP-hard problems, we propose a multi-objective multi-agent optimization algorithm (MAOA) based on multi-agent systems (MAS). In the

* Corresponding Author

proposed algorithm, each agent is treated as a feasible solution for the problem. Agents work together in a grouped environment. The MAOA uses the Pareto dominance concept to find non-dominated agents and to approximate Pareto optimal front. This algorithm maximizes modularity and community score as objective functions. To find promising solutions for the community detection problem, new search operators based on social, autonomous and self-learning behaviors of agents have been designed for the proposed method. As another contribution, the weighted sum method (WSM), as a multi-attribute decision making (MADM) approach, has been utilized in various stages of the proposed algorithm such as finding the best global agent and the leader agent of each group.

To evaluate the performance of the MAOA, several numerical experiments are conducted on five real-world networks. The results demonstrate that the MAOA has been more successful in terms of several performance measures compared to three other meta-heuristics.

The rest of the paper is organized as follows: Section 2 reviews the literature of the community detection problem. Section 3 explains the community detection problem. Section 4 describes the proposed algorithm. Section 5 provides experimental results over five real-world networks. Ultimately, Section 6 concludes the paper and gives some suggestions for future studies.

2. Literature Review

There are many real-world applications for the community detection problem. Detecting fraud movements in telecommunication networks, prediction of connections in dynamic social networks, discovering terrorist groups in social networks and recommending products to customers in online shopping websites are some of the examples.

Many algorithms have been developed for the community detection problem. These algorithms have different strategies to find the most homogenous communities. One of the most important strategies is to treat a community detection problem as a combinatorial optimization problem. In this respect, the community structure is identified by optimizing a predefined criterion such as modularity, modularity density, community score, etc. Pizzuti [5] developed a genetic algorithm (GA) known as GA-Net to detect communities in social networks. Gong et al. [6] proposed a memetic algorithm called the Meme-Net to optimize modularity density as a quality function for estimating the quality of detected communities. Pizzuti [7] proposed a multi-objective community detection algorithm (MOGA-Net) for complex networks. The proposed algorithm uses the Non-dominated Sorting Genetic Algorithm II (NSGA-II) as the optimization procedure for maximizing community score and minimizing community fitness, simultaneously. Community score and community fitness indicate the intra-connections within communities and inter-connections between communities, respectively. Shi et al., [8] developed a multi-objective evolutionary algorithm known as the MOCD to detect community structure. The proposed method optimizes two terms of negatively correlated modularity, concurrently. Gong et al., [9] proposed an evolutionary algorithm with decomposition to optimize ratio cut and negative ratio association, simultaneously. An extended compact genetic algorithm was developed by Li and Song [10] for community detection problem. Amiri et al., [11] proposed a firefly algorithm to discover communities by using fuzzy-based grouping and mutation operators. Cai et al., [12] developed a discrete particle swarm optimization (PSO) algorithm for detecting communities in signed social networks. Gong et al., [13] developed a multi-objective particle swarm optimization (PSO) algorithm that utilizes search strategies of the PSO so as to discover communities of complex networks. The proposed algorithm minimizes two objective functions known as Kernel K-Means and the ratio cut. Cai et al., [14] developed a greedy discrete particle swarm optimization (PSO) method to tackle large-scale social networks. A multi-objective community detection method called the MOLS-Net has been proposed by Zhou et al., [15] that aims to optimize the Kernel K-means and Ratio Cut, concurrently. To optimize each objective function, a local search method has been embedded in the MOLS-Net. A meta-heuristic based on affinity propagation has been proposed by Shang et al., [16] to decompose networks. The network is decomposed by optimizing the Ratio Association and Ratio Cut.

Cheraghchi and Zakerolhosseini [17] proposed a novel dynamic community detection algorithm inspired by social theories. Bilal and Abdelouahab [18] developed an evolutionary algorithm to find community structures by maximizing modularity. Li et al., [19] developed two algorithms for the community detection problem. One of the proposed algorithms is a quantum-mechanism-based PSO algorithm, which is a parallel method. The other algorithm uses the non-dominated sorting procedure instead of the quantum mechanism. Based on the studies reviewed in this section, none of the previous researches have developed a multi-objective multi-agent algorithm for the multi-objective community detection problem to optimize modularity and community score, simultaneously.

3. Problem Description

3.1 Multi-objective Optimization Problem

The aim of a multi-objective optimization problem (MOP) is to find a vector of decision variables that meets constraints and optimizes a vector function. A vector function is a mathematical description of performance criteria formed by objective functions which are usually in conflict with each other. The MOP tries to optimize (minimize or maximize) conflicting objective functions $(f_1(X), \dots, f_m(X))$ when the decision variables $(X = x_1, \dots, x_n)$ can take their values within a feasible region. Typically, there is not a single solution that concurrently optimizes all objectives.

Considering a minimization problem, the MOP can be modeled as follows [4]:

$$\underset{\mathbf{x}\in\mathbf{S}}{\text{Min}}\ \mathbf{F}(\mathbf{x}) = \underset{\mathbf{x}\in\mathbf{S}}{\text{Min}}\ (f_1(\mathbf{x}),\ldots,f_m(\mathbf{x})) \quad (1)$$

Where, $F(X): S \rightarrow R^m$ consists of m real-valued continuous functions that should be minimized, concurrently. $f_i(X)$ is the i^{th} objective function and $x = (x_1, \ldots, x_k) \in S$ is the decision vector. S denotes the set of feasible solutions. In a minimization problem, a decision vector $x_a \in S$ can dominate another decision vector $x_b \in S$ if and only if [4]:

$$f_i(x_a) \le f_i(x_b) \ \wedge \ f_j(x_a) < f_j(x_b) \quad \begin{matrix}\forall i = 1,\ldots,n \\ \exists j = 1,\ldots,n\end{matrix} \quad (2)$$

The solutions of a MOP is a set of Pareto points. A solution $x^* \in S$ is a Pareto optimal if there is no solution (x) in the feasible solution space such that x dominates x^*. A set of solutions that dominate other solutions, while they cannot dominate themselves are called non-dominated solutions.

3.2 Community Definition

A network can be represented as an undirected graph denoted as *G* (**V**, **E**), where **V** and **E** are the **sets** of nodes and edges, respectively. A community is defined as a partition of nodes in the network that have more intra-links than inter-links. It is possible to represent a graph by the adjacency matrix. Let assume that **A** is the adjacency matrix of the graph *G*. Considering the adjacency matrix, if the element in row *i* and column *j* is equal to 1, there is an edge between nodes *i* and *j* in the graph. The degree of node *i* is computed as $k_i = \sum_j A_{ij}$. Suppose that node *i* belongs to a sub-graph *S* ($S \subset G$). In this respect, the degree of node *i* is defined as $k_i(s) = k_i^{in}(s) + k_i^{out}(s)$. $k_i^{in}(s)$ denotes the number of edges connecting node *i* to other nodes in sub-graph *S*, while $k_i^{out}(s)$ represents the number of edges connecting node *i* to the rest of the graph ($G \backslash S$). A sub-graph like *S* is considered as a strong community if $k_i^{in}(s) > k_i^{out}(\forall i \in S)$. To be more specific, a strong community is defined as a group of nodes which have higher intra connections comparing to the rest of the graph [4].

3.3 Fitness Functions

Modularity and community scores are two of the most important objective functions considered in the literature. Both objective functions need to be maximized. The proposed algorithm optimizes these objective functions to detect community structures of networks. These criteria are described as follows:

Modularity:

Modularity (*Q*) a quantitative criterion that measures the quality of network partitions. Modularity has been designed to quantify the strength of partitioning a network into modules. Modularity takes a value in the range of 0 and 1. The modularity value close to 1 implies that a community has the best possible strength, while the modularity value close to 0 indicates that the fraction of edges connecting nodes within a community is not better than the fraction of edges connecting a random gathering of nodes. Modularity is computed as follows [2]:

$$Q = \frac{1}{2L}\sum_{ij}\left[A_{ij} - \frac{k_i.k_j}{2L}\right]d(C_i, C_j) \quad (3)$$

Where, **A** is the adjacency matrix and *L* represents the number of edges in the network. A_{ij} is equal to 1 if nodes *i* and *j* are connected to each other. Otherwise, A_{ij} is equal to 0. k_i and k_j denote the degree of nodes *i* and *j*, respectively. $\delta(C_i, C_j)$ is equal to 1 if the nodes *i* and *j* belong to the same community.

Community score:

Suppose that node *i* belongs to a sub-graph *S* ($S \subset G$). μ_i represents the fraction of edges connecting node *i* to other nodes in community *S*. μ_i is calculated as follows [4]:

$$m_i = \frac{1}{|S|}k_i^{in}(S) \quad (4)$$

Where, $|S|$ is the cardinality of community *S*. *PM*(*S*) is the power mean of *S* in order of *p* that can be defined as:

$$PM(S) = \frac{\sum_{i\in S}(m_i)^p}{|S|} \quad (5)$$

The volume of community *S* denoted as v_S is defined as the number of edges connecting nodes within the community *S*. v_S is computed by Eq. (6) and the score of community *S* ($SC(S)$) is obtained by Eq. (7):

$$v_S = \sum_{i,j\in S} A_{ij} \quad (6)$$

$$SC(S) = PM(S)' \ v_S \quad (7)$$

The community score (*CS*) of a clustering $\{S_1, S_2, \ldots, S_k\}$ of a network is computed as follows:

$$CS = \sum_{i=1}^{k} SC(S_i) \quad (8)$$

Community score sums up the local scores of detected communities so as to provide a global measure of the network division.

4. Proposed Algorithm

4.1 Solution Representation

In this paper, the proposed algorithm employs the locus-based adjacency representation (LAR) [20]. According to the locus-based adjacency representation, each solution is considered as an array of *N* genes. Each gene represents a node in the graph and it is randomly connected to one of its neighbors. Therefore, each gene takes a value on the interval $[1, N]$. Each solution is decoded as a graph in which the value of *j* assigned to the gene *i* is interpreted as a link between node *i* and node *j*.

Thus, connected nodes of each solution are recognized as communities. In this representation, there is no need to know the number of communities in advance. This implies that the number of communities is determined during the decoding procedure. Moreover, the decoding process of a solution is performed in a linear time. The graph structure of a network with 16 nodes is illustrated in Figure 1. A feasible solution and its translation to a graph is depicted in Figure 2. As shown in Figure 2, each gene takes a value on the interval $[1, 16]$ that is randomly chosen from one of its neighbors. According to Figure 2, for instance, the third node has taken the value of "4". This means that there is a link between node 3 and node 4 in the corresponding graph. Therefore, these two nodes are placed in a same community. Communities are depicted by dashed circles in Figure 2.

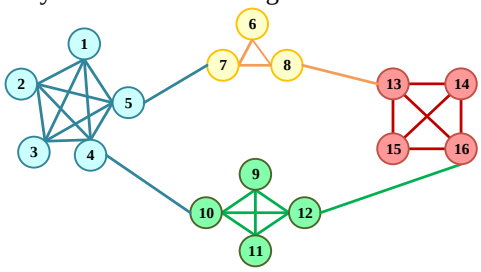


Fig. 1. A sample network

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Solution	5	5	4	2	7	7	6	6	11	4	12	10	14	15	14	13

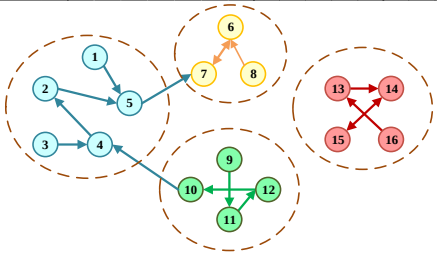


Fig. 2. A locus-based adjacency representation

4.2 Multi-agent System

An agent is defined as a computer system placed in a particular environment. Agents are able to receive information from the environment by means of sensors. An agent analyzes the information and takes consequent actions to affect the environment [21]. Agents have social behavior which makes it possible for them to interact with each other. A group of independent agents can form a multi-agent system (MAS). In a MAS, agents interact with each other and perform their tasks in an environment to achieve common goals. Each multi-agent system has three elements: (1) a set of independent agents $A = \{A_1, A_2, \dots, A_n\}$, (2) an environment where the agents carry out their duties and communicate with each other, and (3) a set of reactive rules that control the interactions between agents and environment [22]. In this paper, the agents are arranged as multiple groups. Figure 3 illustrates the group organization of agents.

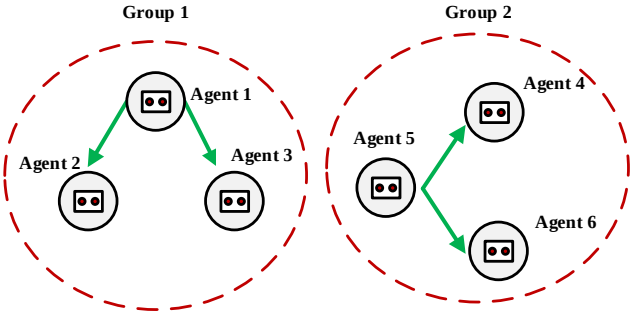


Fig. 3. Group-organized agents

Each multi-agent system has three main features: (1) behaviors of agents, (2) environment, and (3) interactions between agents. These features are explained as follows [23]:

Adjustment of Environment:

In this paper, each agent is considered as a solution. An environment is formed by multiple agents and their corresponding interactions. As mentioned earlier, we have used the grouped structure introduced by Zheng and Wang [23]. In this structure, there are G $(g = 1, \dots, G)$ groups in the environment. Each group contains N_g agents, where N_g represents the number of agents in the group g. The best agent in each group is chosen as the "leader". The group that has the best leader agent is known as the elite group. The second best agent in each group is called the "active" agent. Figure 4 shows a leader-group organization.

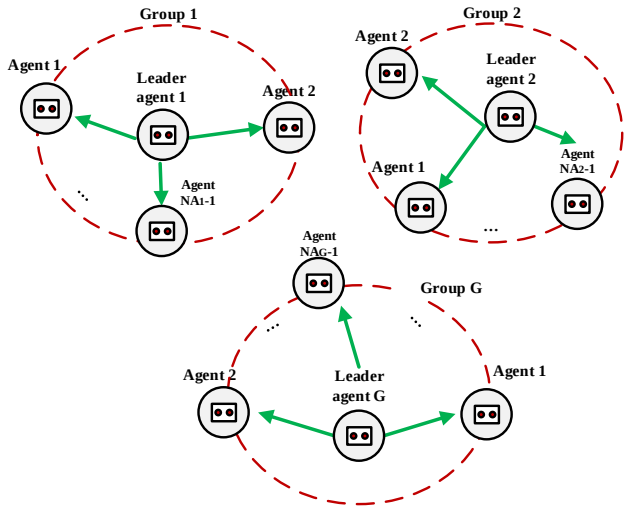


Fig. 4. Leader-group organization

Agents are re-grouped for adjustment of environment. By adjusting the environment, the agents are able to search the solution space, accurately. To adjust the environment, the active agent of each group is replaced with the worst agent of the elite group [22].

Behaviors of Agents:

There are three types of behaviors for each agent: (1) social behavior (local and global behaviors), (2) autonomous behavior, and (3) self-learning behavior. These behaviors are described as follows:

- Social behavior

Social behavior includes local and global behaviors. Local behavior shows the interactions between the leader agent of a group and other agents within the same group. Figure 5 illustrates the social behavior of agents.

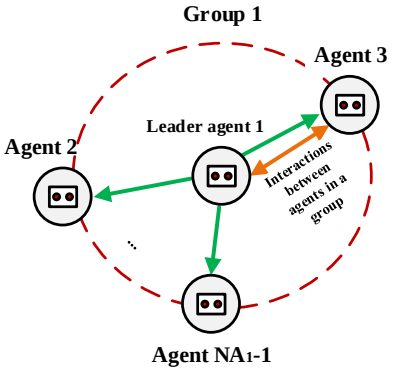


Fig. 5. Local social behavior

Global social behavior indicates the cooperation between the leader agent of the elite group and the leader agents of other groups. Figure 6 illustrates the global social behavior of agents.

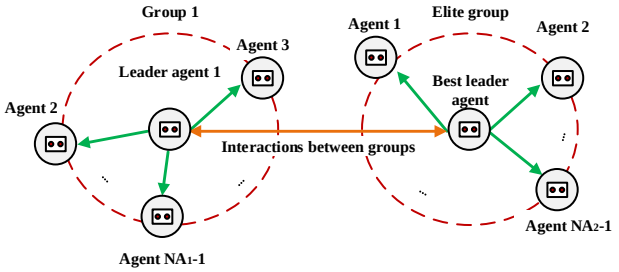


Fig. 6. Global social behavior

- Autonomous Behavior

According to the autonomous behavior, agents can act independently without external interference. Based on this behavior, each agent searches its neighborhood so as to find better solutions [23].

- Self-learning Behavior

It is possible for each agent to improve itself by learning from the obtained knowledge. Self-learning behavior of agents is a problem-dependent local search procedure used in the multi-agent optimization methods [24].

4.3 Multi-objective Multi-agent Optimization Algorithm

Due to the multi-objective essence of the problem tackled in this research, we propose a multi-objective multi-agent optimization algorithm called the MAOA. In the MAOA, the agents are initially divided into multiple groups to form the environment. Agents move among groups to share information in order to adjust the environment. The procedures of the MAOA are described in the following sub-sections.

Finding the Leader Agents of Groups:

To determine the leader agent of a group, the non-dominated sorting method proposed by Deb et al. [25] is used to find the non-dominated solutions (agents). The non-dominated sorting method makes it possible to approximate the Pareto optimal front. In case of having a single non-dominated agent, this agent is selected as the leader agent of the group. On the other side, if there are more than one non-dominated agent in a group, the weighted sum method (WSM) [26] as a multi-attribute decision making (MADM) technique is used to rank agents. The WSM provides the overall scores of non-dominated agents by computing the weighted sum average of all the criteria values. To utilize the WSM, a decision matrix (*DM*) is created, where its rows and columns represent non-dominated agents and criteria, respectively. These criteria are modularity and community score with the same importance. According to the WSM, the relative importance weights are multiplied with the normalized value of the criteria for each agent. Then, the obtained product value is summed up. The non-dominated agent with the highest score is selected as the leader agent of the group. Eq. (9) is used to calculate the overall score of agent i (OS_i) ($i = 1, \dots, N$) with respect to M criteria [26]:

$$OS_i = \sum_{j=1}^{M} w_j . n_{ij} \qquad \forall\ i = 1, \dots, N \qquad (9)$$

Where, n_{ij} is the normalized rating of i^{th} non-dominated agent with respect to j^{th} criterion. w_j denotes the importance of j^{th} criterion. Normalized elements (n_{ij}) ($i = 1, \dots, N$, $j = 1, \dots, M$) for benefit and cost criteria are calculated by the Eqs. (10) and (11), respectively [26].

$$n_{ij} = \frac{r_{ij}}{\max_i (r_{ij})} \qquad \forall\ i = 1, \dots, N,\ \forall\ j = 1, \dots, M \qquad (10)$$

$$n_{ij} = \frac{\min_i (r_{ij})}{r_{ij}} \qquad \forall\ i = 1, \dots, N,\ \forall\ j = 1, \dots, M \qquad (11)$$

Where, r_{ij} is the original rating of i^{th} non-dominated agent with respect to j^{th} criterion.

Finding the Global Best Agent:

Once again the non-dominated sorting method is used to detect the non-dominated agents among leader agents of all groups. In case of having a single non-dominated agent, this agent is selected as the global best leader agent. Otherwise, the WSM is employed to rank the leader agents in order to find the global best agent.

Social Behaviors in the MAOA:

In this paper, an operator is developed for the MAOA to generate new offspring agents from the global best agent and the leader agent. The proposed operator is considered as the global social behavior since it performs the interactions between the global best leader agent and the leader agents of each group. The best offspring agent

is substituted with the leader agent if the best offspring agent dominates the leader agent. This operator initiates by generating a random binary $(1 \times N)$ string (RBS), where N is the number nodes in the network. If the $RBS_i = 1$ $(i = 1, \dots, N)$, the value on the i^{th} gene of the global best agent is assigned to the i^{th} gene on the offspring agent. Otherwise, if the $RBS_i = 0$, the value on the i^{th} gene of the leader agent is assigned to the i^{th} gene of the offspring agent. This procedure always generates feasible solutions. Based on the network depicted in Figure 1, an example of generating a new offspring agent by the proposed operator is illustrated in Figure 7.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Best global agent	2	4	2	5	7	8	6	7	10	12	12	10	14	16	16	13
Leader agent	4	3	5	2	1	7	5	6	12	11	10	16	16	15	13	12
Random binary string (RBS)	1	0	0	0	1	1	0	1	0	1	1	0	1	0	1	1
Offspring agent	2	3	5	2	7	8	5	7	12	12	12	16	14	15	16	13

Fig. 7. Generating an offspring agent

In the following of this section, another procedure is proposed as the local social behavior to improve each agent with guidance of the leader agent within its own group. Therefore, for each agent, a random binary $(1 \times N)$ string (RBS) is generated. If the $RBS_i = 1$, the value on the i^{th} gene of the leader agent of the group will replace the value on the i^{th} gene of the agent. Otherwise, if the $RBS_i = 0$, the value on the i^{th} gene of the agent remains without any change. Figure 8 shows an example of changing an agent by the proposed procedure.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Leader agent	4	3	5	2	1	7	5	6	12	11	10	16	16	15	13	12
An agent	3	1	5	10	7	8	5	13	11	4	12	10	8	16	13	14
Random binary string (RBS)	0	0	1	1	1	0	0	0	1	0	1	0	0	1	0	1
Modified agent	3	1	5	2	1	8	5	13	12	4	10	10	8	15	13	12

Fig. 8. An operator based on local social behavior of agents

Autonomous Behavior in the MAOA:

A simple procedure is presented for the autonomous behavior of each agent. In this procedure, a random integer number (*RIN*) is generated in range of 1 to *N*. The *RIN* determines the number of genes that should be changed. In this procedure, the *RIN* number of genes are randomly selected and these genes take values based on the network.

Self-learning in the MAOA:

In this paper, a self-learning process has been designed for the MAOA to adjust a solution, iteratively. In each iteration, some genes of the best leader agent are changed to optimize modularity and community score.

Adjustment of Environment in the MAOA:

As mentioned earlier, the adjustment of environment is required to share information among groups of agents. Therefore, the environment is adjusted every 20 generations. Let assume that the global best agent belongs to group l. For each group g $(g \neq l)$, the WSM is used to find the active agent (AG_g). If the AG_g dominates the worst agent of group l (WG_l), the AG_g transfers to group l and the WG_l moves to group g.

Elitism in the MAOA:

We consider an archive of non-dominated agents for the MAOA. In each iteration, the non-dominated agents produced by social behavior, autonomous behavior, self-learning and adjustment of environment are combined. Each newly generated agent is compared with the agents existing in the archive. If a new agent dominates any of the agents in the archive, the new agent is substituted with the dominated agent. The maximum number of iterations (*MaxIt*) has been considered as the stopping criterion for the proposed algorithm. Figure 9 shows the flowchart of the MAOA.

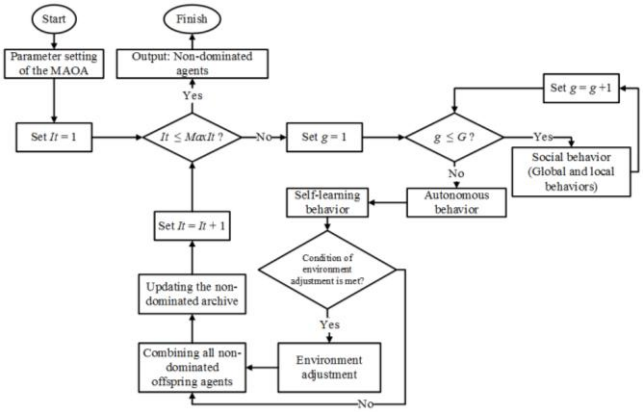


Fig. 9. Flowchart of the MAOA

5. Computational Experiments

In this section, the performance of the proposed algorithm is evaluated by comparing its results with the outputs of three multi-objective evolutionary algorithms, i.e. Multi-objective particle swarm optimization (MOPSO) algorithm [27], Non-dominated Sorting Genetic Algorithm II [25], and Strength Pareto Evolutionary Algorithm II [28]. All algorithms have been coded in the Matlab R2017b software. The codes have been run on a personal computer (PC) with Intel Core 2 Quad processor Q8200 (4M Cache, 2.33 GHz, 1333 MHz FSB) and 4GB memory.

5.1 Real-life Datasets

All algorithms are applied to five real-life networks, which can be downloaded from http://www-personal.umich.edu/~mejn/netdata. These networks are described in Table 1.

Table 1. Description of real-life networks

Network	*Description*
Karate Club	The Karate Club network consists of 34 nodes and 78 edges. The nodes of this network are divided into two groups [29].
Les Misérables	This network shows the relationships between characters of the Victor Hugo's novel "Les Misérables". The graph consists of 77 nodes and 257 edges. The network is divided into five groups [30].
Bernard	The Bernard technical dataset contains five sets of data on human interactions. There are 34

Network	*Description*
	nodes and 350 edges in this network [31].
Grevy's zebra	This network indicates the tendency of Grevy's zebras to appear together. The Grevy's zebra network consists of 28 nodes that are divided into three groups. There are 111 edges between nodes in this network [32].
Facebook	This network shows the interactions between users of the famous social network known as Facebook. This network includes 4039 nodes and 88234 edges [2].

5.2 Performance Measures

In this paper, five performance measures are used to evaluate the performances of algorithms. These performance measures are described as follows:

Normalized Mutual Information (*NMI*):

The *NMI* is a metric that measures the similarity between the real community structure of a network and the community structure found by an algorithm. Let assume that real partitioning of a network is $\alpha = \{\alpha_1, \dots, \alpha_R\}$ and the partitioning identified by an algorithm is $\beta = \{\beta_1, \dots, \beta_K\}$. R and K denote the number of communities in the partitioning α and β, respectively. Confusion matrix (C) is created at the first step to compute the *NMI*. Each element C_{ij} is the number of common nodes in communities $\alpha_i \in \alpha$ and $\beta_j \in \beta$. Then, $NMI(\alpha, \beta)$ is calculated using Eq. (12) [4]:

$$NMI(a,b) = \frac{-2\sum_{i=1}^{R}\sum_{j=1}^{K} C_{ij}\log\left(\frac{C_{ij}N}{C_{i.}C_{.j}}\right)}{\sum_{i=1}^{R} C_{i.}\log\left(\frac{C_{i.}}{N}\right)+\sum_{j=1}^{K} C_{.j}\log\left(\frac{C_{.j}}{N}\right)} \quad (12)$$

Where, $C_{i.}$ and $C_{.j}$ are the sums of the elements in the confusion matrix over row *i* and column *j*, respectively. *N* is the number of nodes existing in the network. The *NMI* takes a value in the range of [0,1]. $NMI = 1$ implies that α and β are exactly equal. On the other hand, $NMI = 0$ indicates that α and β are completely different.

Mean ideal Distance (*MID*):

This metric measures the closeness between the solutions of the approximation front and the ideal point. The mean ideal distance metric is computed by Eq. (13) [33]:

$$MID = \frac{\sum_{i=1}^{NDS}\sqrt{\left(O_{i1}-O_1^*\right)^2+\left(O_{i2}-O_2^*\right)^2}}{NDS} \quad (13)$$

Where, *NDS* represents the number of non-dominated solutions found by an algorithm. O_{i1} and O_{i2} denote the first and the second objective function values of i^{th} solution on the approximation front, respectively. O_1^* and O_2^* are the ideal solutions regarding the first and the second objective functions, respectively. Lower values of *MID* metric show better performance of an algorithm.

Diversification Metric (*DM*):

Diversification metric estimates the extension of the approximation front. Higher values of this metric mean higher diversity of solutions obtained by an algorithm. Diversification metric is computed as follows [33]:

$$DM = \sqrt{\left(\max_{i=1:NDS} O_{i1} - \min_{i=1:NDS} O_{i1}\right)^2 + \left(\max_{i=1:NDS} O_{i2} - \min_{i=1:NDS} O_{i2}\right)^2} \quad (14)$$

Set Coverage (*C*-metric):

Suppose that there are two Pareto fronts denoted as F_1 and F_2 obtained by two different algorithms. $C(F_1, F_2)$ shows the percentage of solutions on the F_2 dominated by at least a single solution of F_1. $C(F_1, F_2)$ is calculated using the Eq. (15) [33]:

$$C(F_1,F_2) = \frac{\left|\{j' \in F_2 \mid \exists j \in F_1 : j \quad Dom \quad j'\right|}{\left|F_2\right|} \quad (15)$$

Where, j and j' are the solutions on the F_1 and F_2, respectively. $|F_2|$ represents the number of solutions on the F_2.

Computation Time (CPU Time):

Another criterion that differentiates algorithms is their required computation time (CPU time) to find optimal or near optimal solutions [34].

5.3 Parameter Setting

In this study, the Taguchi method has been used to set the parameters of algorithms. To obtain optimal values for parameters, the Taguchi method provides a statistic known as signal to noise (*S*/*N*) ratio [35]. Three levels have been considered for parameters of algorithms. Table 2 shows the levels defined for each parameter.

Table 2. Parameters of algorithms

Algorithm	*Parameter*	*Symbol*	*Levels*		
			Level 1	*Level 2*	*Level 3*
MAOA	Number of groups	*G*	3	4	5
	Number of agents in each group	*NA*	30	40	50
	Maximum number of iterations	*MaxIt*	50	100	200
MOPSO	Population size	*Npop*	50	100	200
	Maximum number of iterations	*MaxIt*	50	100	200
	Social factor	C_1	1	1.25	1.5
	Cognitive factor	C_2	1	1.25	1.5
	Inertia weight	*INW*	0.70	0.75	0.80
NSGA-II	Population size	*Npop*	50	100	200
	Maximum number of iterations	*MaxIt*	50	100	200
	Crossover rate	p_c	0.75	0.80	0.85
	Mutation rate	p_m	0.05	0.10	0.15
SPEA-II	Population size	*Npop*	50	100	200
	Maximum number of iterations	*MaxIt*	50	100	200
	Crossover rate	p_c	0.75	0.80	0.85
	Mutation rate	p_m	0.05	0.10	0.15
	Archive size	*ARS*	5	10	15

A response variable known as the multi-objective coefficient of variation (*MOCV*) was proposed by Rahmati

et al. [36] for the Pareto-based algorithms to provide diverse solutions with proper convergence. The *MOCV* incorporates the *MID* and *DM* metrics, simultaneously. The MID metric estimates the convergence rates of algorithms, while the *DM* metric evaluates the diversity of solutions. The *MOCV* can be computed as follows [36]:

$$MOCV = \frac{MID}{DM} \tag{16}$$

Three real-world networks have been used to conduct the Taguchi method. Each algorithm has been run for 10 times to obtain reliable results. To calculate the *MOCV*, the values of the *MID* and *DM* metrics are converted to the relative percentage difference (*RPD*). The *RPD* is obtained as follows [37]:

$$RPD = \frac{\left|\lambda - \lambda^*\right|}{\lambda^*} \times 100 \tag{17}$$

Where, λ denotes the value of performance measure acquired by an algorithm, while λ^* is the best value of performance measure among all values. The average of *RPD*s ($\overline{RPD}$) are calculated for all experiments. Afterwards, $MOCV = \overline{RPD}(MID)/\overline{RPD}(DM)$ is obtained for all experiments. Figures 10 to 13 show the *S/N* ratio plots for the parameters of the MAOA, MOPSO, NSGA-II and SPEA-II, respectively. Table 3 reports optimal values of parameters.

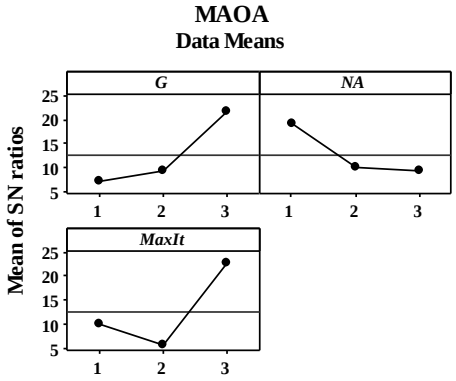


Fig. 10. The mean S/N ratio plot for the MAOA

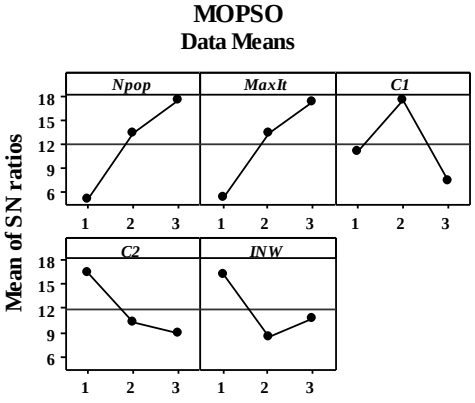


Fig. 11. The mean S/N ratio plot for the MOPSO

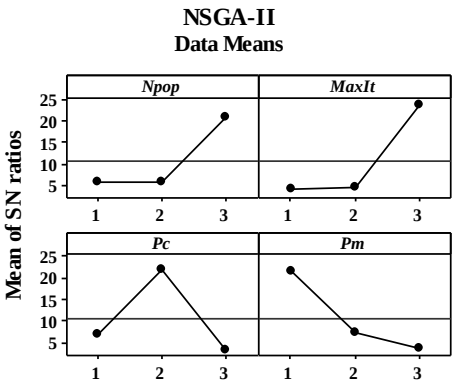


Fig. 12. The mean S/N ratio plot for the NSGA-II

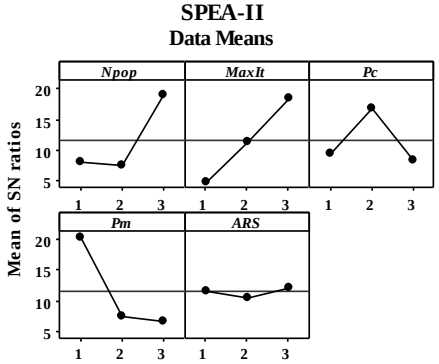


Fig. 13. The mean S/N ratio plot for the SPEA-II

Table 3. Optimal values of parameters

Algorithm	*Parameter*	*Optimal value*
MAOA	*G*	5
	NA	30
	MaxIt	200
MOPSO	*Npop*	200
	MaxIt	200
	C_1	1.25
	C_2	1
	INW	0.70
NSGA-II	*Npop*	200
	MaxIt	200
	p_c	0.80
	p_m	0.05
SPEA-II	*Npop*	200
	MaxIt	200
	p_c	0.80
	p_m	0.05
	ARS	15

5.4 Results and Discussion

The performance of the MAOA is compared with the MOPSO, NSGA-II and SPEA-II in terms of metrics described in Section 5.2. Each method has been run for twenty time and Table 4 reports the average values of performance measures. According to the results summarized in Table 4, the following outlines have been achieved:

1. The MAOA has the best *NMI* values in all networks. This means that the communities found by the MAOA have the most similarity to the real communities of networks.
2. The MAOA has been more successful in providing lower

values of the *MID* metric. This implies that the MAOA has better convergence in comparison with other methods.
3. The MAOA provides more diverse solutions than other algorithms.
4. The MOPSO is the fastest algorithm in detecting communities of all complex networks. Hence, one of the main disadvantages of the MAOA is that it requires more computation time comparing to other methods used in this study.
5. As the number of nodes and edges in networks increase, the values of most of the metrics increase.
6. In terms of the *NMI* metric, the improvement which has been made by the MAOA is nearly 59.4%.
7. For the *MID* metric, the outputs of the MAOA are 62.8% better than other algorithms.
8. In terms of the *DM*, the superiority of the results obtained by the MAOA is approximately 28%.

Table 4. Comparison of algorithms

Network	***Algorithms***	***Performance measures***			
		NMI	***MID***	***DM***	***CPU time***
Karate Club	MAOA	1.00	6.34	6759.61	5.06
	MOPSO	0.95	13.92	5815.48	3.82
	NSGA-II	0.91	24.26	5390.93	5.97
	SPEA-II	0.87	32.78	5015.48	4.94
Les Misérables	MAOA	1.00	7.09	5968.64	13.17
	MOPSO	0.87	21.08	4875.37	11.75
	NSGA-II	0.84	42.45	4322.47	15.25
	SPEA-II	0.80	46.69	4035.76	12.90
Bernard	MAOA	1.00	1.78	6790.40	5.77
	MOPSO	0.74	9.34	5671.26	3.04
	NSGA-II	0.63	13.80	5138.87	6.13
	SPEA-II	0.38	17.01	5019.57	4.48
Grevy's zebra	MAOA	1.00	3.79	8566.41	4.02
	MOPSO	0.71	5.53	7989.95	2.28
	NSGA-II	0.65	6.16	7398.58	4.91
	SPEA-II	0.49	9.19	7016.98	3.19
Facebook	MAOA	0.73	26.42	4763.95	18.57
	MOPSO	0.52	37.40	4468.02	15.09
	NSGA-II	0.36	41.29	3359.63	19.81
	SPEA-II	0.24	49.80	3101.66	17.93

Tables 5 to 9 report the average values of *C*-metric obtained by comparing the non-dominated solutions of algorithms. It is obvious from these tables that the MAOA obtained higher *C* (MAOA, MOPSO), *C* (MAOA, NSGA-II) and *C* (MAOA, SPEA-II) values in all networks. This means that the solutions of the MAOA prevailed the solutions obtained by other algorithms.

Table 5. C-metric values for Karate Club network

C(MAOA, MOPSO)	*C*(MAOA, NSGA-II)	*C*(MAOA, SPEA-II)
0.83	0.97	1.00
C(MOPSO, MAOA)	*C*(MOPSO, NSGA-II)	*C*(MOPSO, SPEA-II)
0.24	0.61	0.77
C(NSGA-II, MAOA)	*C*(NSGA-II, MOPSO)	*C*(NSGA-II, SPEA-II)
0.13	0.25	0.28
C(SPEA-II, MAOA)	*C*(SPEA-II, MOPSO)	*C*(SPEA-II, NSGA-II)
0.07	0.11	0.18

Table 6. C-metric values for Les Misérables network

C(MAOA, MOPSO)	*C*(MAOA, NSGA-II)	*C*(MAOA, SPEA-II)
0.82	1.00	1.00
C(MOPSO, MAOA)	*C*(MOPSO, NSGA-II)	*C*(MOPSO, SPEA-II)
0.26	0.73	0.81
C(NSGA-II, MAOA)	*C*(NSGA-II, MOPSO)	*C*(NSGA-II, SPEA-II)
0.11	0.49	0.76
C(SPEA-II, MAOA)	*C*(SPEA-II, MOPSO)	*C*(SPEA-II, NSGA-II)
0.08	0.14	0.18

Table 7. C-metric values for Bernard network

C(MAOA, MOPSO)	*C*(MAOA, NSGA-II)	*C*(MAOA, SPEA-II)
0.90	0.96	1.00
C(MOPSO, MAOA)	*C*(MOPSO, NSGA-II)	*C*(MOPSO, SPEA-II)
0.35	0.54	0.84
C(NSGA-II, MAOA)	*C*(NSGA-II, MOPSO)	*C*(NSGA-II, SPEA-II)
0.19	0.40	0.88
C(SPEA-II, MAOA)	*C*(SPEA-II, MOPSO)	*C*(SPEA-II, NSGA-II)
0.02	0.17	0.22

Table 8. C-metric values for Grevy's zebra network

C(MAOA, MOPSO)	*C*(MAOA, NSGA-II)	*C*(MAOA, SPEA-II)
0.85	0.92	0.98
C(MOPSO, MAOA)	*C*(MOPSO, NSGA-II)	*C*(MOPSO, SPEA-II)
0.30	0.58	0.69
C(NSGA-II, MAOA)	*C*(NSGA-II, MOPSO)	*C*(NSGA-II, SPEA-II)
0.26	0.52	0.73
C(SPEA-II, MAOA)	*C*(SPEA-II, MOPSO)	*C*(SPEA-II, NSGA-II)
0.03	0.09	0.12

Table 9. C-metric values for Facebook network

C(MAOA, MOPSO)	*C*(MAOA, NSGA-II)	*C*(MAOA, SPEA-II)
0.78	0.81	0.94
C(MOPSO, MAOA)	*C*(MOPSO, NSGA-II)	*C*(MOPSO, SPEA-II)
0.15	0.33	0.37
C(NSGA-II, MAOA)	*C*(NSGA-II, MOPSO)	*C*(NSGA-II, SPEA-II)
0.17	0.25	0.64
C(SPEA-II, MAOA)	*C*(SPEA-II, MOPSO)	*C*(SPEA-II, NSGA-II)
0.00	0.05	0.10

Figures 14 and 15 show the comparisons between algorithms in terms of modularity and community score, respectively. These figures show the average values of objective functions obtained by ten runs. As shown in Figures 14 and 15, the MAOA has achieved better results comparing to other methods.

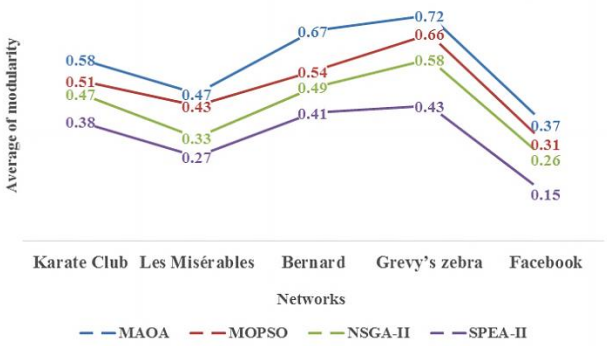


Fig. 14. Comparison of algorithms in terms of modularity

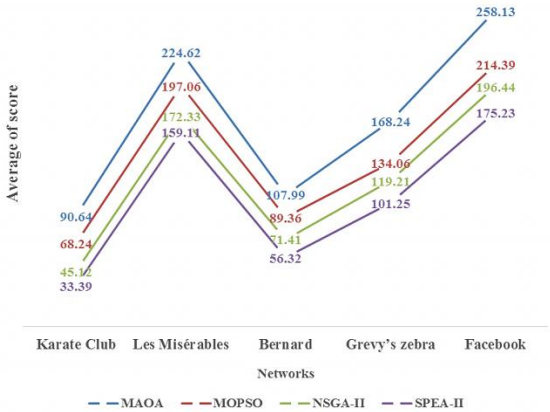


Fig. 15. Comparison of algorithms in terms of community score

6. Conclusions

This paper studied the community detection problem as a multi-objective optimization problem. A multi-objective multi-agent optimization algorithm called the MAOA was proposed to find appropriate partitions of networks by optimizing modularity and community score, concurrently. The MAOA has been inspired by the multi-agent system and swarm intelligence. For the proposed algorithm, new search operators based on social, autonomous and self-learning behaviors of agents were designed. Moreover, a new procedure was developed for adjusting the environment containing agents. Besides, the MAOA uses the weighted sum method (WSM) in finding the global best agent and leader agent of each group. The performance of the proposed algorithm was evaluated and validated by comparing its results to the outputs of three other meta-heuristics. All algorithms were tuned by the Taguchi method. Comparisons were made based on five real-world networks in terms of several metrics including normalized mutual information (*NMI*), mean ideal distance (*MID*), diversification metric (*DM*), computation time (CPU time) and *C*-metric. Results demonstrate that the MAOA has been more successful in providing better results in terms of most of the performance measures. To extend the current study, it is possible to test the effectiveness of the proposed algorithm in larger networks. For another study, the agents can have the ability to choose their social or autonomous behaviors. The agents can also have the right to refuse the role of leadership in groups. Furthermore, other multi-criteria decision making methods can be embedded in multi-objective multi-agent algorithms. Development of other procedures for finding the global best agent and the leader agent of each group is another interesting topic for further studies. In another study, the proposed algorithm can be compared with more recent algorithms to evaluate its performance comparing to state-of-the-art methods.

References

[1] K.R. Zalik, and B. Zalik, "Multi-objective evolutionary algorithm using problem-specific genetic operators for community detection in networks", Neural Computing and Applications, Vol. 1, No. 9, 2017, pp. 1-14.

[2] S. Fortunato, "Community detection in graphs", Physics Reports, Vol. 486, No. 3, 2010, pp. 1-100.

[3] Z. Zhao, S. Feng, Q. Wang, J.Z. Huang, G.J. Williams, and J. Fan, "Topic oriented community detection through social objects and link analysis in social networks", Knowledge-Based Systems, Vol. 26, 2012, pp. 164-173, DOI: https://doi.org/10.1016/j.knosys.2011.07.017.

[4] S. Rahimi, A. Abdollahpouri, and P. Moradi, "A multi-objective particle swarm optimization algorithm for community detection in complex networks", Swarm and Evolutionary Computation, Vol. 39, 2018, pp. 297-309, DOI: https://doi.org/10.1016/j.swevo.2017.10.009.

[5] C. Pizzuti, "GA-Net: A genetic algorithm for community detection in social networks", In Proc. Parallel Problem Solving from Nature–PPSN X, Springer, 2008, pp. 1081-1090, DOI: https://doi.org/10.1007/978-3-540-87700-4_107.

[6] M. Gong, B. Fu, L. Jiao, and H. Du, "Memetic algorithm for community detection in networks", Physical Review E, Vol. 84, 2011, pp. 1-9, DOI: 10.1103/PhysRevE.84.056101.

[7] C. Pizzuti, "A multi-objective genetic algorithm to find communities in complex networks", IEEE Transactions on Evolutionary Computation, Vol. 16, 2012, pp. 418-430, DOI: 10.1109/TEVC.2011.2161090.

[8] C. Shi, Z. Yan, Y. Cai, and B. Wu, "Multi-objective community detection in complex networks", Applied Soft Computing, Vol. 12, No. 2, 2012, pp. 850-859.

[9] M. Gong, L. Ma, Q. Zhang, and L. Jiao, "Community detection in networks by using multi-objective evolutionary algorithm with decomposition", Physica A: Statistical Mechanics and its Applications, Vol. 391, No. 15, 2012, pp. 4050-4060.

[10] J. Li, and Y. Song, "Community detection in complex networks using extended compact genetic algorithm", Soft computing, Vol. 17, No.6, 2013, pp. 925-937.

[11] B. Amiri, L. Hossain, J.W. Crawford, and R.T. Wigand, "Community detection in complex networks: Multi–objective enhanced firefly algorithm", Knowledge-Based Systems, Vol. 46, 2013, pp. 1-11, DOI: https://doi.org/10.1016/j.knosys.2013.01.004.

[12] Q. Cai, M. Gong, B. Shen, L. Ma, and L. Jiao, "Discrete particle swarm optimization for identifying community structures in signed social networks", Neural Networks, Vol. 58, 2014, pp. 4-13, DOI: https://doi.org/10.1016/j.neunet.2014.04.006.

[13] M. Gong, Q. Cai, X. Chen, and L. Ma, "Complex network clustering by multi-objective discrete particle swarm optimization based on decomposition", IEEE Transactions on Evolutionary Computation, Vol. 18, No.1, 2014, pp. 82-97.

[14] Q. Cai, M. Gong, L. Ma, S. Ruan, F. Yuan, and L. Jiao, "Greedy discrete particle swarm optimization for large-scale social network clustering", Information Sciences, Vol. 316, 2015, pp. 503-516, DOI: https://doi.org/10.1016/j.ins.2014.09.041.

[15] Y. Zhou, J. Wang, N. Luo, and Z. Zhang, "Multi-objective local search for community detection in networks", Soft Computing, Vol. 20, No.8, 2016, pp. 3273–3282.

[16] R. Shang, S. Luo, W. Zhang, R. Stolkin, and L. Jiao, "A multi-objective evolutionary algorithm to find community structures based on affinity propagation", Physica A: Statistical Mechanics and its Applications, Vol. 453, 2016, pp. 203-227, DOI: https://doi.org/10.1016/j.physa.2016.02.020.

[17] H.S. Cheraghchi, and A. Zakerolhosseini, "COGNISON: A Novel Dynamic Community Detection Algorithm in Social Network", Journal of Information Systems and Telecommunication, Vol. 4, No. 2, 2016, pp. 78-84.

[18] S. Bilal, and M. Abdelouahab, "Evolutionary algorithm and modularity for detecting communities in networks", Physica A: Statistical Mechanics and its Applications, Vol. 473, 2017, pp. 89-96, DOI: https://doi.org/10.1016/j.physa.2017.01.018.

[19] L. Li, L. Jiao, J. Zhao, R. Shang, and M. Gong, "Quantum-behaved discrete multi-objective particle swarm optimization for complex network clustering", Pattern Recognition, Vol. 63, 2017, pp. 1-14, DOI: https://doi.org/10.1016/j.patcog.2016.09.013.

[20] J. Handl, and J. Knowles, “An evolutionary approach to multi-objective clustering”, IEEE transactions on Evolutionary Computation, Vol. 11, No.1, 2007, pp. 56-76.
[21] N.R. Jennings, K. Sycara, and M. Wooldridge, “A roadmap of agent research and development”, Autonomous Agents and Multi-Agent Systems, Vol. 1, No. 1, 1998, pp. 7–38.
[22] J. Li, H. Jing, and Y.Y. Tang, “Multi-agent oriented constraint satisfaction”, Artificial Intelligence, Vol. 136, No. 1, 2002, pp. 101-144.
[23] X.L. Zheng, and L. Wang, “A multi-agent optimization algorithm for resource constrained project scheduling problem”, Expert Systems with Application, Vol. 42, No. 15-16, 2015; pp. 6039-6049.
[24] W.C. Zhong, J. Liu, M.Z. Xue, and L.C. Jiao, “A multi-agent genetic algorithm for global numerical optimization”, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 34, No. 2, 2004, pp. 229-244.
[25] K. Deb, A. Pratap, S. Agrawal, and T. Meyarivan, “A Fast and Elitist Multi-objective Genetic Algorithm: NSGA-II”, IEEE Transactions on Evolutionary Computation, Vol. 6, No. 2, 2000, pp. 182-197.
[26] R.T. Marler, and J.S. Arora, “The weighted sum method for multi-objective optimization: new insights”, Structural and Multidisciplinary Optimization, Vol. 41, No. 6, 2010, pp. 853-862.
[27] C. A. Coello Coello, and M. S. Lechuga, "MOPSO: a proposal for multiple objective particle swarm optimization", In Proc. Congress on Evolutionary Computation (CEC'02), Honolulu, HI, USA, Vol. 2, 2002, pp. 1051-1056, DOI: 10.1109/CEC.2002.1004388.
[28] W. Sheng, Y. Liu, X. Meng, and T. Zhang, “An Improved Strength Pareto Evolutionary Algorithm 2 with application to the optimization of distributed generations”, Computers & Mathematics with Applications, Vol. 64, No. 5, 2012, pp. 944-955.
[29] W. W. Zachary, "An information flow model for conflict and fission in small groups", Journal of anthropological research, Vol. 33, No.4, 1977, pp. 452-473.
[30] D.E. Knuth, “The Stanford Graph Base: A Platform for Combinatorial Computing”, Addison-Wesley, Reading, MA (1993), ISBN: 0-201-54275-7.
[31] H. R. Bernard, P. D. Killworth, and L. Sailer, "Informant accuracy in social network data IV: A comparison of clique-level structure in behavioral and cognitive network data", Social Networks, Vol. 2, No.3, 1980, pp. 191-218.
[32] S. R. Sundaresan, I. R. Fischhoff, J. Dushoff, and D. I. Rubenstein, "Network metrics reveal differences in social organization between two fission–fusion species, Grevy’s zebra and onager", Oecologia, Vol. 151, No. 1, 2007, pp. 140-149.
[33] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach", IEEE Transactions on Evolutionary Computation, Vol. 3, No. 4, 1999, pp. 257-271.
[34] V. Hajipour, E. Mehdizadeh, and R. Tavakkoli-Moghaddam, “A novel Pareto-based multi-objective vibration damping optimization algorithm to solve multi-objective optimization problems”, Scientia Iranica, Vol. 21, No. 6, 2014, pp. 2368–2378.
[35] B. Golpalsamy, B. Mondal, and S. Ghosh, “Taguchi method and ANOVA: An approach for process parameters optimization of hard machining while machining hardened steel”, Journal of Scientific & Industrial Research, Vol. 68, 2009, pp. 686-695.
[36] S.H.A. Rahmati, V. Hajipour, and S.T.A. Niaki, “A soft-computing Pareto-based meta-heuristic algorithm for a multi-objective multi-server facility location problem”, Applied Soft Computing, Vol. 13, No. 4, 2013, pp. 1728-1740.
[37] J. Gao, R. Chen, W. Deng, “An efficient tabu search algorithm for the distributed permutation flowshop scheduling problem”, International Journal of Production Research, Vol. 51, No. 3, 2013, pp. 641-651.

Amir Hossein Hosseinian obtained his BSc in industrial engineering from Ershad University, Damavand, Iran, and accomplished MSc in industrial Engineering at Buali-sina University, Hamedan, Iran. Currently, he is a PhD candidate in industrial engineering at Islamic Azad University, Tehran North Branch, Tehran, Iran. His research interests are project scheduling, applied operations research, social network analysis and meta-heuristics.

Vahid Baradaran received his Ph.D. degree from Tarbiat Modares University in 2010. He is assistant professor in Islamic Azad University-Tehran North Branch. He thought some courses such as Multivariate Statistical Analysis, Data Mining and Data Analysis in Master Science and Ph.D. courses.

Security Enhancement of Wireless Sensor Networks: A Hybrid Efficient Encryption Algorithm Approach

Omid Mahdi Ebadati E*
Faculty of Mathematics and Computer Science, Kharazmi University, Tehran, Iran
ebadati@khu.ac.ir
Farshad Eshghi
Faculty of Electrical & Computer Engineering, Kharazmi University, Tehran, Iran
farshade@khu.ac.ir
Amin Zamani
Faculty of Knowledge Engineering and Decision Science, Kharazmi University, Tehran, Iran
amin.zamani.cert@khu.ac.ir

Received: 24/Feb/2018 Revised: 22/Aug/2018 Accepted: 16/Sep/2018

Abstract

Wireless sensor networks are new technologies that are used for various purposes such as environmental monitoring, home security, industrial process monitoring, healthcare programs and etc. Wireless sensor networks are vulnerable to various attacks. Cryptography is one of the methods for secure transmission of information between sensors in wireless sensor networks. A complete and secure encryption system must establish three principles of confidentiality, authentication and integrity. An encryption algorithm alone cannot provide all the principles of encryption. A hybrid encryption algorithm, consisting of symmetric and asymmetric encryption algorithms, provides complete security for a cryptographic system. The papers presented in this area over the last few years, and a new secure algorithm present with regard to the limitations of wireless sensor networks, which establishes three principles of cryptography. The details of the algorithm and basic concepts are presented in such a way that the algorithm can be operational and showed a very high efficiency in compare to the current proposed methods.

Keywords: Wireless Sensor Network; Cryptography Algorithm; Hybrid Cryptography; Confidentiality Integration Authentication

1. Introduction

A wireless sensor network is a collection of sensor nodes connected to each other by wireless communication channels. Each sensor node is a small device that can collect data from the surrounding area, perform simple calculations and communicate with other nodes or with the main station. Such networks have been developed with the help of recent advances in micro-electromechanical systems and are expected to be widely used in applications such as environmental monitoring, home security, industrial process monitoring, healthcare programs, etc.

A comprehensive and optimized solution for reliable data sensing and secure, fast and timely data transmission is the use of wireless sensor networks. Wireless sensor network technology to monitor structural accuracy with regard to the advancement of technology related to that, such as the ability to measure, communicate protocols, processor speeds, embedded systems, etc. have been of great importance over the years. These developments gradually affected oil and gas industries in order to automate the processes, system capability and control.

Security in wireless sensor networks depends on what it protects. Three security goals in wireless sensor networks include confidentiality, integrity and authentication. Confidentiality in sensor networks is the ability to hide messages from an attacker. Integrity is the ability to detect unread or unrecognized messages. Authentication is the reliability of the message's origin. Furthermore, other security objectives are defined in accordance with these three principles. Availability is to confirm the ability to use the node's network resources to send the message. The freshness feature is to ensure that the recipient receives new data so that the attacker cannot send the old data recently.

Security can be established in each layer of application, network, data link, and physical layer. Cryptographic algorithms play a significant role in information security systems, and can also meet security goals in wireless sensor networks. Encryption is the process of changing message or information so that only authorized people can read that information. Encryption does not prevent the attack, but the content of the message is protected by the attacker. In an encryption scheme, the information or message set to be transmitted using an encryption algorithm change so, that it can only be read by decrypting. For technical reasons, an encryption scheme usually uses a pseudo-random encryption key generated by an algorithm. For a good design encryption scheme, computational resources and much more skills are required. An authorized recipient of the message can easily decrypt

* Corresponding Author

the message with a key and an encryption algorithm. Cryptography is a knowledge of hiding information and verification, and includes protocols, algorithms, and secure strategies to prevent unauthorized access to critical information. Encryption provides a mechanism for verifying each component of a communication.

An encryption algorithm is a component for secure electronic data transfer. Operational and mathematical stages develop cryptographic algorithms. Cryptographic algorithms prevent data frauds and unauthorized access to electronic information. Some cryptographic algorithms are faster than others. The designers and developers of the algorithms make the math background more complicated by the algorithms so that the attackers cannot penetrate. The power of an encryption algorithm usually depends on the length of the key.

In the past, organizations and companies that needed encryption or cryptographic services designed their own cryptographic algorithms. Over time, major security weaknesses appeared in these algorithms, which made it easier to decrypt. For this reason, cryptographic encryption is now outdated, and in new encryption methods, it is assumed that the full information of the cryptographic algorithm has been published, and what's hidden is just the password key. Therefore, all the security derived from standard encryption algorithms and protocols relies on the security and secret key encryption, and the full details of these algorithms and protocols are released to the public. Cryptographic algorithms and functions used in cryptography are divided into symmetric, asymmetric, hashing, key exchange, key derivation and hybrid.

Symmetric and asymmetric cryptography each has advantages and disadvantages. Symmetric cryptography is significantly faster than asymmetric cryptography, but requires the exchange of a common key. Asymmetric algorithms do not require key exchange systems, but run fast in terms of speed. The combination of asymmetric and symmetric cryptographic systems creates a hybrid encryption system. In asymmetric encryption, the sender and receiver need not to subscribe to a shared key in order to communicate securely, and often rely on complex mathematical calculations. To transmit a hidden random key from a symmetric encryption, the corresponding key can be encrypted using the public key and then sent. The receiver uses its private key to decrypt the key and use it and a symmetric encryption algorithm to send and receive encrypted messages. This protocol is a simple example of a hybrid encryption system. In many applications, the cost of encrypting long message in an asymmetric cryptographic system is very high, which is why hybrid encryption is used. All cryptographic algorithms cannot achieve the three main cryptographic goals: confidentiality, integrity, and authentication. In a hybrid cryptographic system, the weaknesses of an algorithm are covered with the strengths of the other algorithm, so that in one system, all the main purposes of the encryption are met. PGP (Pretty Good Privacy) and TLS (Transport Layer Security) are examples of hybrid cryptographic systems.

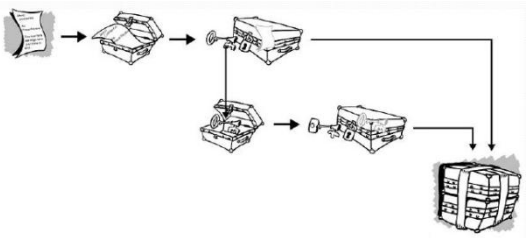


Fig. 1. Hybrid cryptography

2. Literature Review

In the Madhumita Panda [1] paper, two public-key RSA and ECC algorithms have been investigated. The ECC algorithm has significant advantages over the RSA algorithm, which reduces the computation time as well as the amount of data transmitted. The RSA algorithm is a method for implementing a public key encryption system whose security depends on the complexity of the large primes' number factoring. The RSA is made up of the first letter of last names of Ron Rivest, Adi Shamir, and Leonard Adleman in 1977. This method is suitable for data encryption and digital signature creation. Today, the RSA algorithm is a public key encryption that is widely used in the world. The ECC algorithm is related to the algebraic structure of elliptic curves and its difficulty in elliptical curve size. The key advantage of this algorithm is the smaller key size, which reduces storage and transmission. For example, an elliptic curve algorithm can provide the same level of security in an RSA based system with smaller modules and keys. For current cryptographic purposes, an elliptic curve is a curved surface that contains points of the equation $y^2 = x^3 + ax + b$. Compared to the RSA, ECC has a smaller key and uses less memory, which is highly regarded by wireless sensor networks.

The symmetric key algorithm has a weak point in the key distribution, and the asymmetric algorithm requires much computation. Therefore, it is more appropriate to use an algorithm that combines both asymmetric and symmetric algorithms, so that the benefits of both algorithms are more appropriate. A hybrid encryption system is a combination of different types of encryption protocols and is best served by this. A common method is to generate a random hidden key for a symmetric encryption, and then asymmetric encryption of this key using the public key of the receiver. The message is also encrypted using symmetric encryption and hidden key. Hidden encryption key and encrypted message are sent to the recipient. The receiver decrypts the hidden key using its private key, and then decrypts the message key using it. The main approach in the PGP algorithm is the same. Another combination algorithm can be DHA + ECC (Diffie-Hellman algorithm + Elliptic curve cryptography). Public key cryptographic designs were introduced based on the elimination of problems with symmetric cryptographic approaches. Two designs were compared in the Madhumita

Panda paper. The ECC algorithm uses less memory, less processing, and a shorter key than the RSA algorithm.

A new method for ECC cryptography has been proposed by Laiphrakpam Dolendro Singh and Khumanthem Manglem Singh [2] that eliminated the classical methods of mapping characters to offsets in ECC. ASCII values match pairs of texts, and pairings are used as ECC encryption inputs. This proposed new method reduces the cost of the mapping operation and requires the sharing of the table between the sender and the receiver. The algorithm is designed to be used to encrypt or decrypt any type of text with the values of the ASCII. In the Laiphrakpam Dolendro Singh and Khumanthem Manglem Singh article, some ECC encryption concepts are fully described, and the remainder of the fractional calculation is described in brief with the use of the Euclidean algorithm developed.

In the Laiphrakpam Dolendro Singh and Khumanthem Manglem Singh article, a real example presented and key size, key sensitivity to change, time complexity, resistance to some attacks checked out. In this paper, a new method for encrypting text using ECC is presented. ASCII values are divided into groups whose group size is calculated using the P-value of the ECC parameters with a base that is smaller than the maximum amount of ASCII. Large numbers with a specific procedure are divided into P_m pairs. This process helps eliminate the cost of mapping characters to the corresponding elliptic curve. The proposed algorithm is applicable to any text with ASCII values. According to the efficiency comparison table, we can say that the proposed algorithm has positive aspects. It even comes with a lot of words as input, decryption and encryption. The smaller size of encrypted text contributes greatly to bandwidth saving compared to other methods.

Many images are moving through the network daily. Most of these images are confidential and should be transmitted securely. Cryptography plays a significant role in the safe transfer of images. The exponential problem solving a discrete logarithm of an elliptic curve proportional to the size of the ECC key provides a high level of security in comparison with other smaller size key encryption methods, which depends on the correct factoring or discrete logarithmic problem. In another article on Laiphrakpam Dolendro Singh and Khumanthem Manglem Singh [3], an ECC algorithm for encryption, decryption, and digital signature of an image has been implemented. In this paper, a real case study was presented and histogram analysis, key size, key sensitivity to change, correlation analysis, entropy analysis, and resistance to some attacks. In Laiphrakpam Dolendro Singh and Khumanthem Manglem Singh paper, a cryptographic method was provided to match the encrypted photo. The algorithm was presented by grouping the pixels according to ECC parameters. The grouped pixel values are paired together instead of the values mapped to elliptic curve coordinates. This helps to override the use of the mapping table for encryption and decryption. The algorithm produces a cryptographic image with low correlation, even with a picture taken from similar pixels.

Data encryption is required to prevent unwanted access to information by individuals. In Gaurav Patel and Krunal Panchal [4] paper, hybrid methods are investigated by combining two important RSA algorithms and Diffie-Hellman's algorithm. This hybrid cryptographic algorithm provides more security than the RSA algorithm. RSA is the foundation of many encrypted applications. Great progress is for public key encryption and is also used for digital signing. The algorithm process consists of three steps: key generation, encryption and decryption. Whitfield Diffie and Martin Hellman developed the Diffie-Hellman algorithm in 1976. The Diffie-Hellman key exchange protocol is a special method for exchanging cryptographic keys and is one of the first practical examples of cryptographic key exchanging. The Diffie-Hellman key exchange method allows two entities that have no prior knowledge to share a common key through an unsecured connection channel. Then use this key to encrypt the next communication. Diffie-Hellman has been involved in multi-protocol development, including SSL (Secure Sockets Layer), SSH (Secure Shell), and IPS (Internet Protocol Security).

In the proposed Gaurav Patel and Krunal Panchal, the XOR (Exclusive or) Bit Operator is used to enhance the message's complexity. This operation is performed after the message is converted into a cipher text. In this method, two prime number is selected first, and exponential keys of the encryption and decryption process are made. Two numbers A and B are selected for Diffie-Hellman algorithm. R is a randomized number generated by the system automatically. A generic number is generated by Diffie-Hellman's algorithm. Using this common number, KA and KB hidden keys are used in the XOR operator. On the sender side, encryption is performed using the algorithm. When the encryption process is complete, the XOR operation takes place between the secret text and the first hidden key. After the operation, the message is sent securely through the communication channel. On the receiver side, the XOR action is performed again between the second secret key, and the message sent by the sender. Using this operation, the original text is obtained. We can get the main message through the decoding algorithm with the inverse of the same process.

The RSA algorithm is used as one of the most efficient encryption algorithms and provides confidential, information integrity and privacy. In Meenakshi Shankar and Akshaya P [5], RSA algorithm has been integrated with Round-Robin priority scheduling to enhance security and reduce the effectiveness of infiltration. The minimum overhead, increased throughput and privacy are its benefits. In this method, the user uses RSA algorithm and generates encrypted messages that are categorized according to priority and then sent. The receiver decoded messages using the RSA algorithm and according to their priority. This method reduces the risk of man-in-the-middle and timing attacks so that encrypted and decrypted messages are more based on priority. Moreover, if there is

little information exchanged, it also reduces the risk of a power monitoring attack. This approach expands the standards of information security by ensuring greater productivity. In Meenakshi Shankar and Akshaya P research, a brief description of cryptography, steganography, cryptanalysis and cryptology was presented. Then, cryptographic types such as secret key cryptography, public key, and hash function were described. Some changes were made to the RSA algorithm. The first numbers p and q were received from the user. N and Φ(n) were calculated. All the prime numbers are listed between 1 and Φ(n) and allow the user to select value e from the given values. The private key was calculated using d. Priority messages were divided into various priority sections, medium priority, and high priority. The message was encrypted using $C = M^e \bmod n$ and sent to the recipient based on the Round-Robin method.

In Meenakshi Shankar and Akshaya P article, an effective way is presented, which is a combination of successful ways to communicate secretly in a network. The proposed algorithm also reduces brute-force attack effect even if the attacker intercepts and decrypts the message. Furthermore, decrypting a part of a message is not so easy. The RSA algorithm is an efficient and common encryption algorithm and reduces the impact of attacks. The side channel attack is not very effective in this way as it uses a different RSA algorithm. If an attacker is detected, the sender can be stopped; thus, the attacker will not receive the entire message. As a result of this, the impact of a man-in-the-middle attack is reduced. Therefore, if effective methods are combined to prevent side channel and man-in-the-middle attacks, they will be much more productive. This function is effectively performed by an application that encrypts and decrypts messages.

Symmetric and asymmetric hybrid encryption algorithms provide integrated data transfer and concealment at higher speeds and play an important role in virtual private networks. Shi-hai Zhu [6] article deals with the implementation of a hybrid algorithm. System performance analysis and practical tests show that an AES and ECC algorithm provides higher security than the DES and RSA algorithms, especially in virtual private networks that require secure transport.

The proposed hybrid encryption algorithm by Shi-hai Zhu has many advantages. By encrypting AES and ECC keys, we no longer need to send a secret key before we can communicate the secret key management with ECC. The speed of the encryption and decryption process is as large as AES, and the time consumed for ECC is only for AES key. If there are a lot of transfer data, then the use of ECC will be negligible. We send the keys to the ECC and use a digital signature.

The proposed encryption algorithm by Shi-hai Zhu is a combination of public key encryption features that easily distributes the key and is calculated at a high speed and provides a good and fast way to transmit information. In general, hybrid encryption algorithms have many benefits. For example, simple rules, high security, comprehensibility and the ability to execute with hardware and software are largely the criteria for designing a hybrid encryption algorithm. In addition, key security is a tight principle to ensure privacy and file security. In the short term, the hybrid algorithm used in virtual private networks can help to achieve the goal of fast and secure data transmission.

The Internet in the world today is widely used to access information, which is why there is a need to send secure information. The main goal of K. Brindha and others [7] paper is to examine the encryption methods, to improve some of the current algorithms, to create a way to increase the security and implementation of information encryption so that it is impossible to read the resources sent to the attackers on the web. AES and ECC are methods used for encryption. In the proposed algorithm by K.Brindha and others, the file containing the text is encrypted using AES algorithm and its key using ECC algorithm, and the encrypted text is decrypted on the recipient's side. The AES and ECC algorithms are implemented together to provide hybrid encryption. The text is encrypted using AES. The key is encrypted using ECC. The text and the encrypted key are sent to the recipient. The text and the encrypted key are received. The key is decrypted using ECC. Encrypted text is decrypted using decrypt keys and AES.

Secure data transmission on wireless networks is provided using hybrid encryption. For better communication, advanced algorithms are used to break them hard. K.Brindha and others suggest for future that they choose the right encryption so that they can use all the network resources effectively and take into account all network constraints.

The application of the network and the internet is growing at a high rate, thus increasing the need to protect such applications. Rashmi Singh and others [8] highlight this problem by providing two cryptographic methods. The first way is to compress data in half, and the second method focuses on producing characters of encrypted text differently for the same text characters than the different events of the character in the text. The combined effect of using symmetric algorithms with the proposed algorithm creates a hybrid encryption scheme that makes it difficult for an attacker to learn from messages transmitted in an unsecured transmission environment. In Rashmi Singh and others, encryption and decryption are described in detail in four sections along with the code. In the last section, the key generated is different for each character, which means that a single character in the text may have a different cipher character corresponding to the character position. A hybrid encryption scheme is a good combination of data compression and encryption to enhance data security. In this way, the data size is reduced by 50% and security increases. The characters produce the cipher text according to their position. There is currently a demand for reducing space and data security during the transfer. Rashmi Singh and others try to cover all these needs.

Prakash Kuppuswamy and Saeed Al-Khalidi [9] suggest a hybrid cryptographic system using a new public-key algorithm and a private key algorithm. A hybrid encryption system combines the convenience of a public-key system with the efficiency of a symmetric key. In the article by Kuppuswamy and Saeed Al-Khalidi, two secure data encryption methods are provided that are important for confidential. The system uses two different encryption algorithms for the encryption and decryption process; one is public key encryption based on a linear block cipher, and the other is private key encryption based on a symmetric simple algorithm. This encryption algorithm provides more security is better than other existing hybrid algorithms. Cryptography and decryption of any information require a secure key. For this purpose, in Kuppuswamy and Saeed Al-Khalidi paper, the asymmetric key is used, and the linear block cipher algorithm is used for data security. The linear block cipher algorithm is more efficient than the symmetric encryption method. Research results show that its processing time is more efficient than other algorithms. Therefore, AES algorithm, coupled with the RSA algorithm for key management, is an efficient way to ensure the security of transmission data. The security of the combination of RSA and AES is better than the combination of RSA and DES, and the proposed algorithm by Kuppuswamy and Saeed Al-Khalidi is more efficient than the combination of RSA and AES in data transfer. In the article by Kuppuswamy and Saeed Al-Khalidi, a new procedure for future research is also outlined.

A computer network is an interconnected group of independent computing nodes that interact with each other by using a proper definition and a set of agreed rules and conventions known as protocols, and permitting the sharing of resources preferably in a way that can be Predictable and controllable. Today, communications have a major impact on businesses, and data transfer with high security is demanded. Attacks have compromised security; hence, various symmetric and asymmetric encryption algorithms are provided to achieve security services such as authentication, confidentiality, integrity and availability. At the moment, various types of cryptographic algorithms provide high security for information in a controlled network. These algorithms need to specify the data security and authenticity of the user. In order to improve the strength of these security algorithms, a new security protocol for online transactions has been designed using the combination of symmetric and asymmetric encryption methods in the S. Subasree and N. K. Sakthivel [10]. This protocol meets the three basic principles of encryption: integrity, confidentiality, and authentication. These basic principles can be met by using elliptic curve cryptography, Dual RSA and MD5. ECC for encryption, Dual RSA for authentication and MD5 for integrity. This new security protocol uses a combination of symmetric and asymmetric methods for better security and integrity. The text is encrypted using ECC. At the same time, the hash value is calculated using MD5. The resulting hash value is then encrypted with the Dual RSA algorithm. The process of decrypting is an inverse process of encryption.

A computer network is a set of computing nodes that can exchange data with meaningful interaction with each other and allow resource sharing in the appropriate manner. A set of connected computers using communication channels requires security for the exchange of information. This field of work involves a specialist in network security with the network administrator that prevents and monitors unauthorized access, modifies, and disables the use of the network. To combat this growing problem, security professionals are looking for better protection. Attacks have endangered security; hence, various symmetric and asymmetric encryption algorithms are provided to achieve the appropriate security services such as identity, confidentiality, integrity and availability. These algorithms are designed to provide security and authenticate users. To improve the strength of these security algorithms, Manali J Dubai and others [11] have developed a new security algorithm using both symmetric and asymmetric encryption methods. This algorithm provides three principles of cryptography: integrity, confidentiality, and authentication. This algorithm is derived from the combination of ECDH, ECDSA, DUAL RSA algorithms and MD5 hash algorithms. This new security algorithm has been used for better security and integrity of the combination of symmetric and asymmetric encryption methods. In the paper by Manali J Dubal and others, a brief description of encryption, crypto analyzer, cryptosystem and ECDH algorithms, DUAL RSA and MD5 are presented. The text is encrypted using DUAL RSA encryption algorithm, and the key generated with ECDH. Encrypted text for identifying with the ECDSA algorithm is combined. The encrypted text is hashed by the MD5 algorithm. To decrypt, the hash value is first calculated. The computed value is compared with the signature to confirm the message. Decrypting cipher text is done with DUAL RSA.

In Manali J Dubal and others paper, a strong and lightweight protocol is being used that uses ECC pattern. The proposed protocol addresses several problems, such as operational implementation, short response time, efficient computing and cryptographic power. The ECC charm compared to RSA is to provide better security with a smaller key, which reduces processing overhead. Advantage of using it, higher speed, lower power consumption, bandwidth saving, storage efficiency and smaller certificate.

Wireless sensor networks consist of hundreds or thousands of low-cost, low-power and self-organized nodes that are highly distributed [12]. Wireless sensor networks are growing and require effective security mechanisms, because sensor networks may interact with sensitive data. Cryptographic algorithms have a good role in information security systems. Currently, various types of cryptographic algorithms provide security in wireless sensor networks, but there are still some problems. At present, symmetric and asymmetric encryption methods

can provide a level of security with some constraints. In a paper by Bhupinder Singh Dhaliwal and Vivek Soi [13], a new hybrid encryption algorithm is proposed to improve the power of these algorithms. The algorithm is designed using a combination of two symmetric and asymmetric encryption methods. The proposed algorithm is a cryptographic method composed of ECC and AES algorithms. RSA and Blowfish are used for authentication and MD5 for integrity. The results show that the proposed encryption algorithm performs better in terms of computation time and encrypted text size. Bhupinder Singh Dhaliwal and Vivek Soi, an attempt to make a fair comparison between the new protocol and the four existing protocols. The comparison is done in different ways, such as the size of the data block and the speed of encryption and decryption. The empirical results in the article determine the effect of each. Then ECC, ECDSA, ECDH, MD5, RSA and Blowfish algorithms were briefly described. Previous work in this area included Subasree, Kumar, Kady and Zhu. By changing the perception of a sensor, the performance and progress of the sensor network can be enhanced. ECC encryption supports various encryption methods and privacy by generating keystrokes. This scheme will lead to network management in an agile manner.

One of the goals of wireless sensor networks is to transmit trusted information from one node to another in the network. In the paper by Bhave and Jajoo [14], an encryption scheme of improved AES and ECC algorithms has been used to increase the security of wireless sensor networks. This paper analyzes AES algorithm and the S-Box structure and provides an improved AES encryption algorithm. Using the AES algorithm, the message sent by the sender changes to be completely new encrypted text so that the attacker cannot guess the recipient's original message.

AES is rightly recommended as the most suitable symmetric encryption algorithm for wireless sensor networks. The AES algorithm has complex mathematical computations on the text, such as the transformation of the primary key and the XOR operator with a polynomial matrix. This algorithm is 10 rounds to convert text to encrypted text. Using the S-Box Replacement makes inverse easy, but makes it harder for attackers. Moreover, in this algorithm, the time required for encryption is very low. In Bhave and Jajoo article for exploring purposes, plain text is given with 16 bytes, and a key is considered and the algorithm is implemented.

Many key management schemes are provided in a wireless sensor network. Sensor nodes are provided with insufficient battery power, low memory, limited computing, and communication constraints. Energy in safe and efficient routing is a major issue for wireless sensor networks [15]. In the article by R. Sharmila and V. Vijayalakshmi [16], the key management scheme consists of a public key encryption scheme and a symmetric schema. The symmetric key is generated using the genetic algorithm. The initial entry for the genetic algorithm is the key generated by the HECC (Hyperelliptic CurveCryptography) encryption. The design offers energy efficiency, flexibility against node capture attacks, and key refreshment between cluster heads and cluster nodes. The simulation results show that this hybrid scheme has more robustness, more energy-efficient, and smaller-sized keys. A key management pattern consists of four steps before the key distribution, deployment of the key, adding and deleting the node. Based on cryptographic methods, the key management schema is categorized into three types of symmetric management, asymmetric management, and hybrid key management techniques. The key is deployed using the HECC. The proposed new key management plan combines the benefits of using elliptical curve and symmetric key generation using the genetic algorithm to secure wireless sensor networks. The proposed new hybrid algorithm describes the combination of the genetic algorithm with public key cryptography and is applicable to encryption of text and images and is suitable for wireless sensor networks. The algorithm is strengthened by a predetermined permutation factor for the cluster head node and member nodes; thus breaking it is very difficult. The proposed method is divided into two stages of key deployment and symmetric key generation. The key deployment step is divided into three stages of key generation, before key distribution and key agreement. The server generates a key in a key repository using the HECC. Whenever a key is generated, a key pool is generated for a key using a specified process. Each time the keys and corresponding key pools are generated, they are distributed in the sensor nodes. Sensing nodes attempt to establish a secure connection by connecting the key pool. If they are not able to establish a secure connection, they will use an interface node to establish a secure connection to their neighbor's node.

Wireless sensor networks include a different set of communication levels and types of routing protocols. In another research [17], they shared the key between cluster head and member nodes using the asymmetric key distribution and genetic algorithm. Secure the internal cluster of communication and it's effective in key reconstruction for hierarchical sensor networks. The proposed method is energy-efficient use, efficient authentication and high flexibility against cluster-head compromise attacks.

Some sensors use Bluetooth technology to communicate in wireless sensor networks. In a paper by Wuling Ren and Zhiqian Miao [18], a communication encryption algorithm based on DES and RSA is provided to enhance the security of data transmission in Bluetooth communication. Currently, the encryption algorithm used in Bluetooth protects the confidentiality of data during transmission between two or more devices, and a 128-bit symmetric encryption called E0. This encryption is broken down under certain conditions with the complexity of time O (2^{64}). In the proposed hybrid algorithm instead of E0 encryption, DES algorithm is used for data transmission because it has higher efficiency in block encryption and uses the RSA algorithm to decrypt the DES key because it is superior to cryptographic key management. Under the

protection of DES and RSA algorithms, the Bluetooth system is safer. It is clear that the whole encryption method is simple and efficient, and in addition, the confidentiality of the algorithm is high. Bluetooth features include a wireless, short-range and low-power. In Wuling Ren and Zhiqian Miao, the current Bluetooth cryptographic structure is described using pin, E2, E3 and link key. The process of decrypting is an inverse process of encryption. The proposed algorithm has many advantages. The use of RSA algorithm and DES key for transmission have been used. Key management is done by RSA. The use of RSA algorithm allows the use of digital signatures. The encryption and decryption speed is similar to DES, and RSA usage time is only for the DES keys. Hybrid cryptographic algorithm is safer than the safety of the two RSA and DES algorithms.

Bluetooth technology is a new technology that has changed the way it is transmitted. However, Bluetooth technology does not fully address security issues in the standardization process. Bluetooth is used as a wireless communication channel in the transmission environment and more vulnerable to fixed networks. For programs with priority security, achieving a high level of security is essential. Currently, the E0 encryption is used with all the shortcomings in the Bluetooth standard, while the DES and RSA combination encryption algorithm is relatively safer and easier. In this way, the security of data transfer between Bluetooth devices is guaranteed in real time.

In Komal Rege and others [19], a hybrid encryption algorithm is based on AES and RSA to increase the security of data transfer in Bluetooth communications. At present, the E0 algorithm is used to transfer information between two devices or more via Bluetooth. The combination of the encryption algorithm instead of E0 uses AES algorithm, which has a great effect on block encryption, and uses RSA algorithm to encrypt AES key to exploit the key management benefits. Therefore, the use of AES and RSA algorithms makes it safer to transfer information in Bluetooth. In addition, the hybrid encryption algorithm is a convenient and easy way to encrypt data and enhance confidential. In the Komal Rege and others, the current structure of Bluetooth encryption is described using pin, E2, E3 and the link key, and the weaknesses of the E0 algorithm, such as address spoofing, LFSR (Linear Feedback Shift Register) constraint, PIN reliability, and low credibility of link key.

The process of decrypting is an inverse process of encryption. For private key encryption, there are many algorithms such as DES, 3DES, AES, and Blowfish. The DES algorithm was developed and popular in 1970. However, today, for many applications, it is insecure and weak because the 56-bit key length is too small. Many of the attacks exploited DES's shortcomings. 3DES offers an improvement on DES, in which the DES algorithm is used three times, but it is very slow. Blowfish algorithm is used in public domain for applications and suffers from poor key issues. The AES algorithm has the most priority and is considered as the best encryption standard. Brute-force attack is the only known attack against the AES algorithm. RC4, 128-bit and a fast encryption, preventing many kinds of attacks. Due to the weaknesses of other algorithms, AES is the best cryptographic standard and is preceded by other standards. Studies have shown that Blowfish is the fastest in terms of processing time, but security is a concern. In this respect, AES works best. Therefore, AES algorithm, coupled with RSA algorithm for key management, is an efficient way to ensure the security of data transmitted using Bluetooth.

Bluetooth technology is widely used to transmit information over short distances. Bluetooth as a wireless technology is more prone to attacks than other networks. Therefore, data security is important during transfer. E0 is an encryption algorithm that is currently used in Bluetooth for encryption, which has many shortcomings and can be easily broken. The AES algorithm with a low number of published attacks is very secure. Moreover, the difficulty of factoring large integers guarantees the security of RSA algorithm. Accordingly, the proposed combination of encryption algorithms using AES and RSA provides a safer and easier way to transfer data between Bluetooth devices compared to the E0 algorithm.

Security is one of the most important and fundamental issues for transmitting data in wireless sensor networks. Hence, innovative hybrid encryption algorithms for security have been developed. DNA (Deoxyribonucleic Acid) cryptography plays a vital role in the fields of communication and data transmission. In DNA encryption, the biological concept of DNA is used not only to store data and information carriers, but also to perform computations. In the paper by Monikaa and Shuchita Upadhyaya [20], computing security is provided using DNA-based encryption. This paper presents an innovative algorithm that uses a DNA encryption and SSL protocol to provide a safer channel for the exchange of information in wireless sensor networks. In Monikaa and Shuchita Upadhyaya paper described the definition of a wireless sensor network and shared cryptographic keys between sensors as a problem. Three patterns of key subscription (release by a secure server, agreement on a specific contract for the distribution and distribution of keys before network development) and their disadvantages were identified. The third pattern was chosen, and based on this, a combination of DNA encryption and the SSL protocol was presented after providing a brief description of the DNA molecule and the SSL protocol.

The process of decrypting is an inverse process of encryption. The data in the encryption is hidden using DNA-related methods. In Monikaa and Shuchita Upadhyaya paper, the concept of DNA is used in cryptography and SSL protocol, which meets three levels of security in wireless sensor networks. In the proposed system, the power consumption problem for generating key pairs and producing certificates for sensors has been raised, to a certain extent, by assigning keypads and digital certificates prior to deploying sensors in the environment. The public key and digital certificate are shared using SSL protocol. Therefore, the calculation overhead for key generation may be reduced and ultimately leads to energy efficiency in the sensors. It is anticipated that the proposed solution may provide promising results.

Encryption plays an important role in securing wireless sensor networks. In the article by Rawya Rizk and Yasmin Alkady [21], a new algorithm is comprised of symmetric and asymmetric encryption methods with a small security key. This algorithm guarantees three principles of cryptography: integrity, confidentiality and identity. The combination of ECC and AES encryption algorithms is provided. The XOR and DUAL RSA algorithms are used to identify and MD5 for integrity. The results show that the hybrid algorithm presented in computational time, cipher text size, and energy consumption. This algorithm is resistant to a variety of attacks.

In the article by Rawya Rizk and Yasmin Alkady, past activities on hybrid encryption have been investigated, including Subasree, Dubal, Kumar, Ren and Zhu algorithms. This article is one of the most complete articles in the field of cryptography for wireless sensor networks. The process of decrypting is the inverse of the encryption process. The reason for the superiority and power of the proposed algorithm is a set of features of RSA, ECC, XOR, AES and MD5. In the paper by Rawya Rizk and Yasmin Alkady, a hybrid algorithm for the security of wireless sensor networks is presented. This algorithm is designed to solve several problems, including operational implementation, short response time, efficient computing and power of the cryptographic system. An algorithm called THCA (Two phase Hybrid Cryptography Algorithm) is suggested. The THCA tries to divide the text and then apply two different methods. First of all, the advantages of using a combination of both symmetric and asymmetric encryption methods are AES and ECC. Then it uses a custom RSA algorithm that is robust and cannot be easily attacked. In addition, the hash has been used with MD5 to control the data to ensure that the original text does not change during the transmission. Also, the performance of THCA is compared to other algorithms. This algorithm provides better security with less encryption and decryption time and a shorter cipher text size. As a result, reducing the overhead of data processing has achieved lower energy consumption, which is suitable for all applications of wireless sensor networks. In the article by Rawya Rizk and Yasmin Alkady, the proposed THCA algorithm was used to encrypt an image and its resistance to various types of attacks were investigated.

3. Proposed Method

The purpose of this research is to develop it as an applied research. The general scheme of the proposed algorithm is as follows:

The process of the proposed algorithm is as follows:

1. The information string is read from a text file.

2. The information string is converted to binary string using Huffman's coding.

2.1. Create a leaf node for each character and add it to the priority queue.

2.2. While there is more than one node in the queue:

2.2.1. Remove the two nodes of highest priority (lowest probability) from the queue

2.2.2. Create a new internal node with these two nodes as children and with probability equal to the sum of the two nodes' probabilities.

2.2.3. Add the new node to the queue.

2.3. The remaining node is the root node, and the binary string constructed.

3. The binary string is divided into 256-bit blocks. If the final block is not a multiplier of 256, it will be padded with 0.

4. The 256-bit strings are divided into two categories. If the number of blocks is odd, the dividing point of the relation $\frac{n(B)+1}{2}$ and if the number of blocks is even, the dividing point is obtained from the relation $\frac{n(B)}{2}$.

5.1. In the Rijndael algorithm, 14 rounds are used to convert each block information.

5.2. The initial block or state is added to the extended key.

5.3. Round processes include S-Box, shift and mix columns operations. These processes are performed at all times, except for the column mix that are not done in the last round. The result state is added to the extended key for each round. The final result is the cipher block.

6.1. The cryptographic key required to Rijndael encryption for the first group of blocks is generated by ECDH algorithm.

6.2. For generating the secret encryption key between two nodes, A and B using ECDH, both sides of the relationship must agree on the elliptic curve domain parameters. On both sides of communication, a pair of keys contains a private key d (a random integer less than n) and a public key Q = d × G. G is the generator point for cryptographic operators. Suppose (d_A, Q_A) the private and public key pair of node A and (d_B, Q_B) are the private and public key pair of node B.

6.3. The node A, $K = (x_K, y_K) = d_A \times Q_B$, calculates.

6.4. The node B, $L = (x_L, y_L) = d_B \times Q_A$, calculates.

6.5. Since $d_A \times Q_B = d_A \times d_B \times G = d_B \times d_A \times G = d_B \times Q_A$, then K = L and hence $x_K = x_L$ is established.

6.6. The hidden encryption key is x_K.

7.1. The cryptographic key required to Rijndael encryption for the second group of blocks is generated by RSA algorithm.

7.2. The two prime numbers p and q are chosen. Being prime of the numbers should be checked through the tests.

7.3. n = pq is calculated. n is a base Modular in public and private key.

7.4. $\lambda(n) = LCM(\lambda(p), \lambda(q)) = LCM(p-1, q-1)$ is calculated.

7.5. An integer e is chosen so that $1<e<\lambda(n)$ and $GCD(e, \lambda(n)) = 1$ are established. In fact, e and $\lambda(n)$ are relative to prime.

7.6. $d = e^{-1} \bmod \lambda(n)$ is computed.

7.7. If node B wants to send node key of Rijndael algorithm, node A sends the public key (n, e) to node B. After the node B takes the public key node A, it can send the message M to node A. First, the message M must be converted to the integer m such that $0<m<n$.

7.8. The node B generates a ciphered message using the public key e and $c = m^e \bmod n$. The node B sends message c to node A. The node A retrieves the number m from message c using the private key power d and $c^d = (m^e)^d = m \bmod n$.

4. Pseudocode

Algorithm 1. Loading information string from a text file

```
1. string plaintext = read (string TextFilePath);
```

Algorithm 2. Encoding

```
1.   binary[] encodedText;
2.   for (int i=1; i<= (plaintext.Length)-1; i++)
2.1.      {
2.2.      Assume new node z;
2.3.      Z_left = X = Min(plaintext);
2.4.      Z_right = Y = Min(plaintext);
2.5.      Z_prop = X_prop+Y_prop;
2.6.      Insert Z into encodedText;
2.7.      };
```

Algorithm 3. Blocking and split blocks

```
1.   binary blockedText[] = encodedText;
2.   int padding = (encodedText.Length) mod 256;
3.   if (padding != 0)
3.1.      {
3.2.      binary paddedBits[256-padding] = 0;
3.3.      Append paddedBits to blockedText;
3.4.      };
4.   int blockCount = (blockedText.Length) / 256;
5.   binary blockedText1[(blockCount / 2) * 256];
6.   binary blockedText2[(blockCount - (blockCount / 2)) *
256];
7.   Copy blockedText from 0 to blockedText1.Length index
into blockedText1;
     8.   Copy blockedText from blockedText1.Length+1 to
blockedText.Length index into blockedText2;
```

Algorithm 4. Rijndael encryption

```
1.   binary[] resultBlock = inputBlock;
2.   for(n=1; n<=14; n++)
2.1.{
2.2.      binary[] extendedKey;
2.3.      binary block[];
2.4.      resultBlock=extendedKey[n]+resultBlock;
2.5.      S-Box_n(resultBlock);
2.6.      Shift_n(resultBlock);
2.7.      if(n != 14)
2.7.1.            {
2.7.2.            MixColumn_n(resultBlock);
2.7.3.            };
2.8.      };
```

Algorithm 5. ECDH

```
1. int d_A,Q_A,d_B,Q_B; // domain parameters of an elliptic curve
2. int K = (x_K,y_K) = d_A×Q_B;
3. int L = (x_L,y_L) = d_B×Q_A;
4. int keyForRijndael = x_K;
```

Algorithm 6. RSA

```
1. int p,q; //p and q are prime numbers
2. int n = p×q;
3. int lmbda(n) = LCM(p-1, q-1);
4. find e where GCD(e, lmbda(n)) =1;
5. int d = e^-1 mod lmbda(n);
6. (n, e) is private Key;
7. (n, d) is public Key;
```

5. Comparison

In the proposed algorithm, the advantages of symmetric and asymmetric encryption algorithms are combined to establish the three principles of confidentiality, authentication and integrity. Furthermore, the limits of wireless sensor networks are considered in the exchange and transfer of information. The cryptographic process is performed using the Rijndael algorithm with 256-bit blocks. Rijndael algorithm is a high speed symmetric encryption algorithm and variation in implementation. Based on Kerckhoffs's principle, the security of a cryptographic system depends on maintaining the confidentiality of the encryption key. In the proposed algorithm, the encryption key required by the Rijndael algorithm is provided in two ways. The reason for doing this is to increase the security of the key exchange in the wireless network. If the encryption key is identified for some of the blocks of information, since the key exchange process of the other information blocks is different, it will not be possible for the attacker to recover all information.

6. Time Complexity

The time complexity of an algorithm is the quantity that represents the amount of time consumed to run that algorithm as a function of the input string size. The time complexity of an algorithm is usually expressed by a large O that overlooks the lower-order coefficients and phrases. This display method is an asymptotic description of the complexity of time. For example, if the time needed to run an algorithm for all inputs $n > n_0$ is equal to $an^3 + bn$, the asymptotic time complexity is equal to $O(n^3)$. a, b and n_0 are constant values. The complexity of the time is estimated by counting the number of main operations of the algorithm.

The proposed hybrid encryption algorithm consists of an encoding algorithm, a symmetric encryption algorithm, two asymmetric encryption algorithms for key exchange. The time complexity of Huffman's algorithm is O(nlogn). A stack is used to store the weight of each node. The time complexity is to determine the lowest weight and add new weight O(logn) and the time complexity of cycles O(n). Rijndael's algorithm consists of 14 rounds and four distinct operations per round. Operations are all of a mapping type by a predefined table or a linear operation, resulting in a time complexity of O(14n). The time complexity of the elliptic curve encryption algorithm is equal to $O(\sqrt{n})$ and the time complexity of the RSA algorithm is $O(\log n^2)$. The remaining operations are linear and do not affect time complexity, because their grades are in the degree of other operations affecting the algorithm is less. Other operations, such as dividing the information string into two categories, blocking, padding with 0, etc., are all constant; therefore, their complexity is equal to O(k), so that k is the fixed number of the input string function. In general, the time complexity of the entire algorithm is equal to O(nlogn). The proposed algorithm has a good time complexity compared with

other proposed algorithms in recent years. While the proposed algorithm is more secure than other hybrid algorithms provided on the communication platform of wireless sensor networks, it fully enforces the three principles of confidentiality, authentication and integrity.

7. Memory Consumption

A hypothetical information string has 446 characters, and an ASCII encoding system needs to store 446 bytes or 3568 bits of space in the physical memory. However, in the proposed algorithm, the final file needs 285 bytes or 2280 bits of space for storage in physical memory. The proposed algorithm has a variable-length encoding system, and in this string of information, about 36% of the memory consumption is saved.

In the table below, the number of information blocks for transmission in a wireless sensor network is compared in several blocking methods for the hypotheses' string:

Table 1. Different blocking methods for hypothetical information

Blocking Method	Block size	Padding bits	Number of Blocks
64-bit	64	48	56
128-bit	128	112	28
256-bit	256	240	14
Proposed method	256	232	9

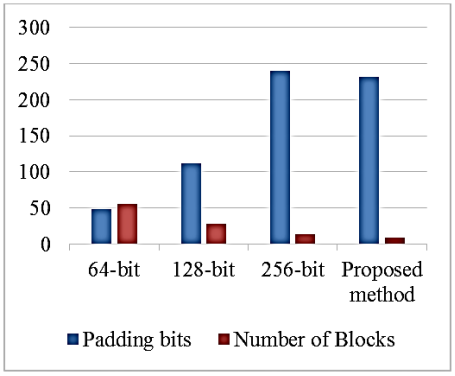


Fig. 2. Different blocking method chart for hypothetical information

The number of transition blocks in the proposed algorithm is lower than in other blocking methods. Typically, the DES encryption algorithm contains 64-bit blocks, the AES encryption algorithm has 128-bit blocks, and the Rijndael encryption algorithm has 256-bit blocks. By increasing the size of blocks, the probability of increasing the number of layers of bits in the final block increases, which only affects the final block, and its number is entirely dependent on the length of information string. The table above is achieved without regard for integrity, and it only considers the enclosed blocks of information to be transmitted. The hash algorithms also create additional information blocks to encrypted information blocks and transfer them to the wireless communication platform for integrity and authentication purposes. In the following table, several blocking methods in combination with the hash algorithms and the number of final blocks for transmission of the encrypted information strings are specified:

Table 2. Different blocking methods with hash algorithms for hypothetical information

Blocking Method	Hash Algorithm	Number of Final Blocks
64-bit	MD5	112
64-bit	SHA1	196
64-bit	SHA-256	280
128-bit	MD5	42
128-bit	SHA1	63
128-bit	SHA-256	84
256-bit	MD5	18
256-bit	SHA1	23
256-bit	SHA-256	28
Proposed method	SHA-256	18

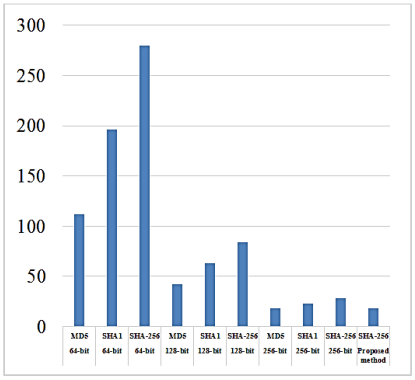


Fig. 3. Different blocking method chart with hash algorithm for hypothetical information

Because the length of final information string in the hybrid method for authentication and integrity is only dependent on the hash algorithm, only the hash algorithms are defined in the table above. The number of blocks in the 256-bit blocking methods with the MD5 and SHA1 hash algorithms is lower than the proposed algorithm. The MD5 algorithm is currently broken, and the SHA1 algorithm is broken in certain conditions. As a result, they will not be a good choice in terms of security.

The proposed algorithm provides three principles of cryptography: confidentiality, integrity, and authentication with the optimal number of blocks of information.

8. Time Consumed

The proposed algorithm consists of the steps for reading the information file, encoding, blocking, dividing the information blocks into two groups, encrypting the first group of information blocks and generating the corresponding digital signature, and encrypting the second group of information blocks and generating the corresponding digital signature. In ten times the implementation of the algorithm, the measured consumption times are shown in the table below:

Table 3. Results of ten times implementation of the proposed algorithm

Implementation	Reading information	Encoding	Blocking	Split blocks	First encryption	Second encryption	Total
1	0.2135	1.5627	0.0347	0.0033	441.9149	103.4469	547.176
2	0.1961	1.8367	0.036	0.0043	413.9018	103.8879	519.8628
3	0.2078	2.9879	0.0347	0.0033	427.2974	102.6035	533.1346
4	0.2071	1.6052	0.0277	0.0033	416.0835	120.1658	538.0926
5	0.1948	2.6196	0.0417	0.003	415.6013	99.2333	517.6937
6	0.1851	1.51	0.0307	0.0026	393.9851	99.4391	495.1526
7	0.2522	1.9788	0.0414	0.003	491.8431	103.231	597.3495
8	0.2071	2.0696	0.0357	0.003	399.3358	105.4186	507.0698
9	0.1921	1.8698	0.033	0.004	590.7126	101.8583	694.6698
10	0.1917	1.5561	0.04	0.003	405.2739	97.8342	504.8989
Average	0.2047	1.9596	0.03556	0.00328	439.5949	103.7119	545.51003

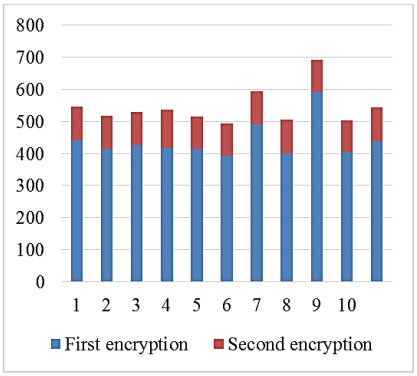


Fig. 4. Chart of results of ten times implementation of the proposed algorithm

The unit of measurement for the time consumed in each step is based on milliseconds. The first and second encryption steps take most time to execute the algorithm. As a result, only two of most influential factors are considered in the above diagram. The RSA algorithm uses longer time to encrypt longer strings than ECC algorithm. Naturally, there is a direct relationship between the increased security of data transmission and the time consumed. In the proposed algorithm, there is a balance between the time consumed and the increased security of data transmission, so that the cryptographic key for half of the information is rapidly distributed, and half of the information is distributed securely. The distribution of the encryption key in two ways has also helped to increase the security of the proposed algorithm.

9. Conclusion

In this research, a comprehensive and optimized solution for reliable data and secure, fast and timely data transmission is the use of wireless sensor networks. Wireless sensor network technology has been gaining great importance over the years, due to the advancement of technologies associated with this technology such as the ability to measure, communicate protocols, processor speeds, embedded systems, and more. Wireless sensor networks are new technologies that are used for various purposes. Wireless sensor networks are vulnerable to various attacks in different layers. Cryptography is one of the methods for secure transmission of information between sensors in wireless sensor networks. A complete and secure encryption system must establish three principles of confidentiality, authentication and integrity. Cryptography is a knowledge of hiding information and verification, and includes protocols, algorithms and secure strategies to prevent unauthorized access to critical information. Encryption provides a mechanism for verifying the components of a connection. An encryption algorithm alone cannot provide all of the principles of encryption. A hybrid encryption algorithm, consisting of symmetric and asymmetric encryption algorithms, provides complete security for a cryptographic system. The papers presented in this area over the past few years, and a new secure algorithm is presented that provide three cryptographic principles. The proposed algorithm is presented with regard to the limitations of wireless sensor networks. The proposed algorithm has optimum time complexity, time and memory usage. The details of the algorithm and basic concepts are presented in such a way that it is possible to implement the algorithm operationally.

Many advantages of the proposed algorithm are presented. Computer science is very broad and covers a variety of topics. There are many ideas to improve the performance of hybrid encryption algorithms in securing the transmission of data between sensors in wireless sensor networks. Some of them can be the basis for future research.

The use of compression algorithms reduces the amount of memory usage and the number of transition blocks in the communication platform of the wireless sensor network. Compression algorithms have complex operations that use sensor processing resources. Hence, the use of a compression algorithm commensurate with the limits of wireless sensor networks helps to improve the performance of a hybrid encryption system.

A high-security encryption system is also used for wireless sensor networks in other networks. As a result, modifying and improving algorithms helps to provide more security in monitoring systems.

References

[1] M. Panda, "Security in wireless sensor networks using cryptographic techniques," American Journal of Engineering Research (AJER), vol. 3, pp. 50-56, 2014.

[2] L. D. Singh and K. M. Singh, "Implementation of Text Encryption using Elliptic Curve Cryptography," Procedia Computer Science, vol. 54, pp. 73-82, 2015/01/01/ 2015.

[3] L. D. Singh and K. M. Singh, "Image Encryption using Elliptic Curve Cryptography," Procedia Computer Science, vol. 54, pp. 472-481, 2015/01/01/ 2015.

[4] G. R. Patel and K. Panchal, "Hybrid Encryption Algorithm," International Journal of Engineering Development and Research (IJEDR), vol. 2, pp. 2064-2070, 2014.
[5] M. Shankar and P. Akshaya, "Hybrid Cryptographic Technique Using RSA Algorithm and Scheduling Concepts," International Journal of Network Security & Its Applications, vol. 6, p. 39, 2014.
[6] S.-h. Zhu, "Research of hybrid cipher algorithm application to hydraulic information transmission," in Electronics, Communications and Control (ICECC), 2011 International Conference on, 2011, pp. 3873-3876.
[7] K. Brindha, G. Ramya, and R. A. Jayantila, "Secured Data Transfer in Wireless Networks Using Hybrid Cryptography," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, 2013.
[8] R. Singh, I. Panchbhaiya, A. Pandey, and R. H. Goudar, "Hybrid Encryption Scheme (HES): An Approach for Transmitting Secure Data over Internet," Procedia Computer Science, vol. 48, pp. 51-57, 2015/01/01/ 2015.
[9] P. Kuppuswamy and S. Q. Al-Khalidi, "Hybrid encryption/decryption technique using new public key and symmetric key algorithm," International Journal of Information and Computer Security, vol. 6, pp. 372-382, 2014.
[10] S. Subasree and N. Sakthivel, "Design of a new security protocol using hybrid cryptography algorithms," IJRRAS, vol. 2, pp. 95-103, 2010.
[11] M. J. Dubai, T. Mahesh, and P. A. Ghosh, "Design of new security algorithm: Using hybrid Cryptography architecture," in Electronics Computer Technology (ICECT), 2011 3rd International Conference on, 2011, pp. 99-101.
[12] K. Yaeghoobi SB, M. Soni, S. Tyagi, and O. M. Ebadati E, "SAERP: An energy efficiency Real-time Routing protocol in WSNs," in 2014 International Conference on Reliability Optimization and Information Technology (ICROIT), 2014, pp. 249-254.
[13] B. S. Dhaliwal and V. Soi, "Reprogramming of Wireless Sensor Network Securely with New Hybrid Encryption Scheme," International Journal of Engineering Technology, Management and Applied Sciences (IJETMAS), vol. 3, pp. 258-263, 2015.
[14] A. Bhave and S. Jajoo, "Secure Communication in Wireless Sensor Network using Symmetric and Asymmetric hybrid Encryption Scheme," International Journal of Innovative Science, Engineering & Technology (IJISET), vol. 1, pp. 382-385, 2014.
[15] K. Yaeghoobi SB, M. Soni, S. Tyagi, and O. Ebadati E, "Impact of NP-complete in triangle segments tree energy efficiency model in wireless sensor networks," J. Basic Appl. Sci. Result, vol. 3, pp. 808-817, 2013.
[16] R. Sharmila and V. Vijayalakshmi, "Hybrid Key Management Scheme for Wireless Sensor Networks," International Journal of Security and Its Applications, vol. 9, pp. 125-132, 2015.
[17] M. Yazdinejad, F. Nayyeri, and N. Afshari, "Secure Distributed Group Rekeying Scheme for Cluster Based Wireless Sensor Networks Using Multilevel Encryption," in Internet of Things: Novel Advances and Envisioned Applications, ed: Springer, 2017, pp. 127-147.
[18] W. Ren and Z. Miao, "A hybrid encryption algorithm based on DES and RSA in Bluetooth communication," in 2010 Second International Conference on Modeling, Simulation and Visualization Methods (WMSVM 2010), 2010, pp. 221-225.
[19] K. Rege, N. Goenka, P. Bhutada, and S. Mane, "Bluetooth communication using hybrid encryption algorithm based on AES and RSA," International Journal of Computer Applications, vol. 71, pp. 10-13, 2013.
[20] Monika and S. Upadhyaya, "Secure Communication Using DNA Cryptography with Secure Socket Layer (SSL) Protocol in Wireless Sensor Networks," Procedia Computer Science, vol. 70, pp. 808-813, 2015/01/01/ 2015.
[21] R. Rizk and Y. Alkady, "Two-phase hybrid cryptography algorithm for wireless sensor networks," Journal of Electrical Systems and Information Technology, vol. 2, pp. 296-313, 2015.

Omid Mahdi Ebadati E. earned his Ph.D. degree in computer science with network security expertise from Hamdard University, New Delhi, India. He is currently an assistant professor and head of information and communication technology center at Kharazmi University, Tehran. He is a senior member of IEEE and the Computer Science Association of America and India. He has published numerous research papers in peer-reviewed international journals and conferences and published books and chapter books in the field of Computer Networks, Internet of Things, Cloud Computing and Machine Learning.

Farshad Eshghi earned his Ph.D. degree in telecommunications engineering from Concordia University, Montreal, QC, Canada. He is a member of IEEE, IEEE communications society, and IEEE computer society and is currently an assistant professor at Kharazmi University. He has published dozens of research papers and books in international journals and conferences on MAC/Routing Protocols in Ad Hoc WLANs/WSNs Cooperative Transmission in Ad Hoc WLANs, Localization in Wireless Networks, Intelligent Building Management Systems and Intelligent Transportation Systems.

Amin Zamani earned a Master degree in computer science from Kharazmi University, Tehran, Iran. His current research interests are in the areas of wireless sensor networks, information security and cryptography.