In the Name of God

Journal of
Information Systems & Telecommunication
Vol. 3, No. 4, October-December 2015, Serial Number 12

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
Email: info@jist.ir
URL: www.jist.ir

Indexed in:
- Index Copernicus International www.indexcopernicus.com
- Journal of Information Systems and Telecommunication www.jist.ir
- Islamic World Science Citation Center (ISC) www.isc.gov.ir
- Scientific Information Database (SID) www.sid.ir
- Regional Information Center for Science and Technology (RICeST) www.ricest.ac.ir
- Magiran 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)

- Ahmadi Ali, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Akbarizadeh Gholamreza, Shahid Chamran University of Ahvaz, Ahvaz, Iran
- Amindavar Hamid Reza, Amirkabir University of Technology, Tehran, Iran
- Badie Kambiz, ITRC (Iran Telecommunication Research Center), Tehran, Iran
- Baleghi Yasser, Babol Noshirvani University of Technology, Babol, Iran
- Borna Keivan, Kharazmi University, Tehran, Iran
- Fouladi Kazem, University of Tehran, Tehran, Iran
- Falahati Abolfazl, Iran University of Science and Technology, Tehran, Iran
- Geran Gharakhili Fatemeh, Shahid Rajaee Teacher Training University, Tehran, Iran
- Ghanbari Mohammad, University of Essex, Colchester, UK
- Haji Mohammadi Zeinab, Abrar Institute of Higher Education, Tehran, Iran
- Hamidi Hojjatollah, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Hasanian Esfahani Roya, Academic Center for Education Culture and Research (ACECR), Tehran, Iran
- Heydarian Mohsen, Azarbijan Shahid Madani University, Tabriz, Iran
- Jabraeil Jamali Mohammad Ali, Islamic Azad University, Shabestar Branch, Shabestar, Iran
- Jamili Oskouei Rozita, Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan, Iran
- Kashef Seyed Sadra, Tarbiat Modares University, Tehran, Iran
- Lotfi Abbasali, Academic Center for Education Culture and Research (ACECR), Tehran, Iran
- Mahdipour Hadi, Ferdowsi University of Mashhad, Mashhad, Iran
- Maleki Marjan, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Moghavvemi Mahmoud, University of Malaya, Kuala Lumpur, Malaysia
- Mohammadzadeh Sajad, University of Birjand, Birjand, Iran
- Mohebbi Keyvan, Islamic Azad University, Mobarakeh Branch, Tehran, Iran
- Moradi Gholamreza, Amirkabir University of Technology, Tehran, Iran
- Moradi Parham, University of Kurdistan, Sanandaj, Iran
- Mosallanejad Ali, Shahid Beheshti University, Tehran, Iran
- Naghibzadeh Mahmoud, Ferdowsi University of Mashhad, Mashhad, Iran
- Parvin Hamid, Iran University of Science and Technology, Tehran, Iran
- Rahmani Mohsen, Arak University, Arak, Iran
- Ramezani Amin, Tarbiat Modares university, Tehran, Iran
- Sadeghzadeh Ramezan Ali, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Shirvani Moghaddam Shahriar, Shahid Rajaee Teacher Training University, Tehran, Iran

Table of Contents

Selecting Enterprise Resource Planning System Using Fuzzy Analytic Hierarchy Process Method

Hodjatollah Hamidi*
Department of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran
h_hamidi@kntu.ac.ir

Received: 07/Mar/2015 Revised: 08/Aug/2015 Accepted: 15/Aug/2015

Abstract

To select an enterprise resource planning (ERP) system is time consuming due to the resource constraints, the software complexity, and the different of alternatives. A comprehensively systematic selection policy for ERP system is very important to the success of ERP project. In this paper, we propose a fuzzy analytic hierarchy process (FAHP) method to evaluate the alternatives of ERP system. The selection criteria of ERP system are numerous and fuzzy, so how to select an adequate ERP system is crucial in the early phase of an ERP project. The framework decomposes ERP system selection into three main factors. The goal of this paper is to select the best alternative that meets the requirements with respect to product factors, system factors and management factors. The sub-attributes (sub-factors) related to ERP selection have been classified into twelve main categories of Functionality, Reliability, Usability, Efficiency, Maintainability, Cost, Implementation time, User friendliness, Flexibility, Vendor Reputation, Consultancy Services, and R&D Capability and arranged in a hierarchy structure. These criteria and factors are weighted and prioritized and finally a framework is provided for ERP selection with the fuzzy AHP method. Also, a real case study from Iran is also presented to demonstrate efficiency of this method in practice.

Keywords: ERP System Selection; Analytic Hierarchy Process (AHP); Fuzzy Logic; Decision Analysis; ERP Vendor.

1. Introduction

An Enterprise Resource Planning (ERP) system represents an information management system which is supposed to manage the data flow among the working modules of a company. An ERP system generally includes a shared data base and different modules and applications which are used in order to facilitate planning, production, sales, marketing, distribution, human resources, project management, inventory, data processing and information storage. ERP systems allow the company's processes to be automated thus increasing the operational efficiency [1]. The use and the importance of computing information systems and their applications to improve effectiveness and efficiency of business functions have increased significantly. Furthermore, because of the exponential increase in the competition in the globalized economy, coupled with ever so changing customer needs and wants, the complexity of the business processes has also risen. These all have led to ERP systems becoming an essential part of any modern day solution to the increasingly complex business environment [2]. ERP is increasingly important in modern business because of its ability to integrate the flow of material, finance, and information and to support organizational strategies [3-4]. A successful ERP project involves managing business process change, selecting an ERP software system and a co-operative vendor, implementing this system, and examining the practicality of the new system [5].

There are three phases that constitute ERP system life cycle. These phases are selection, implementation and use. Problem identification, requirements specification, evaluation of options and selection of system can be regarded as the activities within the ERP selection process. ERP selection is the first phase and is regarded as the most critical success factor for ERP implementation [6]. Determining the best ERP software that fits with the organizational necessity and criteria, is the first step of tedious implementation process. Hence, selecting a suitable ERP system is an extremely difficult and critical decision for managers. An unsuitable selection can significantly affect not only the success of the implementation but also performance of the company. However, many companies install their ERP systems hurriedly without fully understanding the implications for their business or the need for compatibility with overall organizational goals and strategies [7-8]. The result of this hasty approach is failed projects or weak systems whose logic conflicts with organizational goals.

ERP selection issue can be viewed as a multiple criteria decision making (MCDM) problem in the presence of many quantitative and qualitative criteria that should be considered in the selection procedure including a set of possible vendor alternatives. A decision maker is required to choose among quantifiable or non-quantifiable and multiple criteria. The decision maker's evaluations on qualitative criteria are always subjective and thus imprecise. The objectives are usually conflicting and therefore the solution is highly dependent on the

* Corresponding Author

preferences of the decision maker. Besides, it is very difficult to develop a selection criterion that can precisely describe the preference of one alternative over another. The evaluation data of ERP alternatives suitability for various subjective criteria, and the weights of the criteria are usually expressed in linguistic terms. This makes fuzzy logic a more natural approach to this kind of problems. In this paper, we used its fuzzy analytic hierarchy process (FAHP) extension to obtain more decisive judgments by prioritizing criteria and assigning weights to the alternatives.

This paper presents a comprehensive framework for selecting a suitable ERP system based on an AHP-based decision analysis process. The proposed procedure allows a company to identify the elements of ERP system selection and formulate the fundamental-objective hierarchy and means objective network. The pertinent attributes for evaluating a variety of ERP systems and vendors can be derived according to the structure of objectives.

Cebeci and Ruan investigated some quality consultants using fuzzy AHP [5]. Wei, Chien, and Wang proposed a comprehensive framework for selecting a suitable ERP system based on an AHP-based decision analysis process [7]. The AHP is one of the extensively used multi-criteria decision-making methods. One of the main advantages of this method is the relative ease with which it handles multiple criteria. In addition to this, AHP is easier to understand and it can effectively handle both qualitative and quantitative data. The literature cited herein is just an exemplary sample of what has been studied in the area of ERP system selection. The quantity and quality of the published articles in this field are a testament to both importance and the complexity of the ERP system selection problem. What differentiates our approach from the ones conducted previously is the following: first, it ensures that the structure of objectives is consistent with corporate goals and strategies. The project team can understand the relationships among different objectives and assess their influence by modeling them to the hierarchical and network structures. Second, the project team can decompose the complex ERP selection problem into simpler and more logical judgments of the attributes. Particularly, knowledge of structure of objectives can help the project team to identify the company requirements and develop appropriate system specifications. These objectives also indicate how outcomes should be measured and what key points should be considered in the decision process. Third, the approach is flexible enough to incorporate extra attributes or decision makers in the evaluation. Notably, the proposed framework can accelerate the reaching of consensus among multiple decision makers. Finally, the approach systematically assesses corporate attributes and guidance based on the company goals and strategic development. It can not only reduce costs during the selection phase, but also mitigate the resistance and invisible costs in the implementation stage.

The remaining part of the paper is organized as follows: Section 2 describes the related work. The fuzzy analytic hierarchy process algorithm is introduced in Section 3. The ERP system selection framework is presented in Section 4. The Application of FAHP in ERP System Selection using a real case study and the obtained results are discussed in section 5. Finally, Section 6 gives the conclusion of the study.

2. Related Work

Selection process is a critical success factor. The process of selecting an Enterprise Resource Planning (ERP) system is a complex problem which involves multiple actors and variables, since it is a decision-making process which is characterized as unstructured type [4,5].The number of studies have explored various selection methods of ERP system either qualitative or quantitative.

Owing to the essence of IT system, selection problem is a Multi-criteria decision-making (MCDM) process. Several papers adopted analytic hierarchy process (AHP) to be the analytical tool [6,7]. Lin [8] and Luo and Strong [9] studied the ERP evaluation models for universities. Selection criteria of ERP system is also a crucial issue in ERP project. When implementing an ERP project, price and time are both the most important factors. Besides, the vender's support is also a crucial issue [10]. Except the investment cost of ERP project, the annual maintenance cost and human resource cost are also the potential expense for organizations [11-13].There was an ERP selection model containing three categories of selection attributes including project factors, software system factors and vender factors [3].

In the Wei and Wang [3] several methods have been proposed for selecting a suitable ERP system [14-18]. The scoring method is one of the most popular. Although it is intuitively simple, it does not ensure resource feasibility. Teltumbde [14] suggested 10 criteria for evaluating ERP projects and constructed a framework based on the Nominal Group Technique (NGT) and the analytic hierarchy process (AHP) to make the final choice.

Lee and Kim [17] combined the analytic network process (ANP) and a 0–1 goal programming model to select an information system. However, these mathematical programming methods can not contain sufficient detailed attributes, above all, which are not easy to quantify, so that the attributes were restricted to some financial factors, such as costs and benefits. Furthermore, many of them involved only the consideration of internal managers, but do not offer a comprehensive process for combining evaluations of different data sources to select an ERP project objectively.

Wei and Wang [3] stated clearly that; a successful ERP project involves selecting an ERP software system and vendor, implementing this system, managing business processes, and examining the practicality of the system. However, a wrong ERP project selection would either fail

the project or weaken the system to an adverse impact on company performance [19-20] .It is obvious that one firm organization needs some metrics in order to choose the right ERP and its implementers. Thus decision needs some tools. Wei, Chien and Wang [7] introduced AHP based approach to ERP system selection.

Ayag and Ozdemir, [47], used fuzzy ANP as the methodology for the selection of ERP software and presented a case study in a firm in electronics sector and Percin, [48], also proposed ANP as a viable decision making tool for ERP selection problem. The criteria used in the study are divided into two groups: system factors (i.e., functionality, strategic fitness, flexibility, user friendliness, implementation time, total costs, and reliability) and vendor factors (i.e., market share, financial capability, implementation ability, R&D capability, and service support). With this study, they showed the utility and versatility of ANP for this complex selection problem. Similarly, Unal and Guner, [49], and Cebeci, [50], proposed a methodology based on AHP and fuzzy AHP respectively for ERP supplier selection for an organization in the textile industry. A similar application of fuzzy AHP was also performed in an automotive company for the selection of ERP outsourcing firm [51]. With another study, Sen, Baraclı, Sen, and Baslıgil [52], showed the viability of a combined decision making methodology for the ERP selection problem. Within the proposed methodology, the fuzzy set theory and random experiment based methods are combined and successfully applied to both quantitative and qualitative factors. The hybrid methodology was proposed by Kilic, Zaim, and Delen, [53], they used fuzzy AHP and TOPSIS for the selection of ERP software for an airline company.

3. Fuzzy Analytic Hierarchy Process Algorithm

Analytic Hierarchy Process: AHP was proposed by Saaty [21] to model subjective decision-making processes based on multiple attributes in a hierarchical system. Saaty introduced AHP as a powerful and flexible decision making technique that helps decision makers to set priorities and choose the best alternative [21]. From that moment on, it has been widely used in corporate planning, portfolio selection, and benefit/cost analysis by government agencies for resource allocation purposes. It should be highlighted that all decision problems are considered as a hierarchical structure in the AHP. The first level indicates the goal for the specific decision problem. In the second level, the goal is decomposed of several criteria and the lower levels can follow this principal to divide into other sub-criteria. Therefore, the general form of the AHP can be depicted as shown in Fig.1.

The four main steps of the AHP can be summarized as follows [22]:

Step 1: Set up the hierarchical system by decomposing the problem into a hierarchy of interrelated elements;

Step 2: Compare the comparative weight between the attributes of the decision elements to form the reciprocal matrix;

Step 3: Synthesize the individual subjective judgment and estimate the relative weight;

Step 4: Aggregate the relative weights of the elements to determine the best alternatives/strategies.

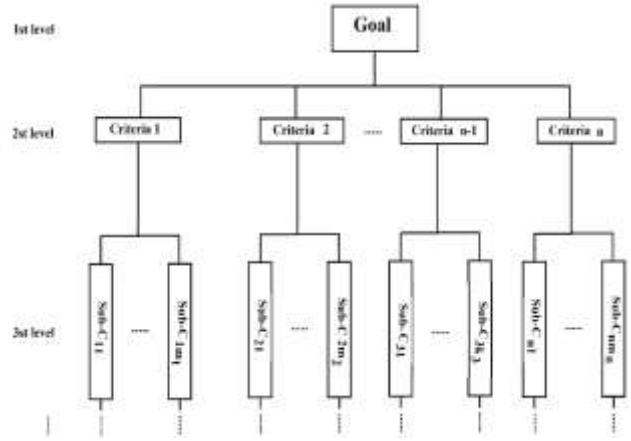


Fig. 1.The hierarchical structure of the AHP.

Fuzzy Analytic Hierarchy Process: The fuzzy AHP technique can be viewed as an advanced analytical method developed from the traditional AHP. Despite the convenience of AHP in handling both quantitative and qualitative criteria of multi-criteria decision making problems based on decision makers' judgments, fuzziness and vagueness existing in many decision-making problems may contribute to the imprecise judgments of decision makers in conventional AHP approaches [23]. So, many researchers[24-32]who have studied the fuzzy AHP which is the extension of Saaty's theory, have provided evidence that fuzzy AHP shows relatively more sufficient description of these kind of decision making processes compared to the traditional AHP methods. Yu [33] employed the property of goal programming to solve group decision making fuzzy AHP problem. Weck et al. [34] evaluated alternative production cycles using fuzzy AHP. Sheu [35] presented fuzzy-based approach to identify global logistics strategies. Kulak and Kahraman [36] used fuzzy AHP for multi-criteria selection among transportation companies. Kuo et al. [37] integrated fuzzy AHP and artificial neural network for selecting convenience store location. Cheng [27, 38] proposed a new algorithm for evaluating naval tactical missile systems by the fuzzy AHP based on grade value of membership function. Zhu et al. [39] made a discussion on the extent analysis method and applications of fuzzy AHP.

In complex systems, the experiences and judgments of humans are represented by linguistic and vague patterns. Therefore, a much better representation of these linguistics can be developed as quantitative data, this type of data set is then refined by the evaluation methods of fuzzy set theory. On the other hand, the AHP method is mainly used in nearly crisp (non-fuzzy) decision applications and creates and deals with a very unbalanced scale of judgment. Therefore, the AHP method does not take into account the uncertainty associated with the mapping [40]. The AHP's subjective judgment, selection and preference of decision-makers have great influence on the success of

the method. The conventional AHP still cannot reflect the human thinking style. Avoiding these risks on performance, the fuzzy AHP, a fuzzy extension of AHP, was developed to solve the hierarchical fuzzy problems.

In this study, Chang's [41] extent analysis on fuzzy AHP is formulated for a selection problem. Chang's extent analysis on fuzzy AHP depends on the degree of possibilities of each criterion. According to the responses on the question form, the corresponding triangular fuzzy values for the linguistic variables are placed and for a particular level on the hierarchy the pairwise comparison matrix is constructed.

The fuzzy AHP algorithm is constructed in six steps using Chang's extent analysis method [27, 41], a popular fuzzy AHP approach. The method is relatively easier than other proposed approaches and has been used in several cases [42-43]. Let $X=\{x_1,x_2,\ldots,x_n\}$ be an object set and $G=\{g_1,g_2,\ldots,g_n\}$ be a set of goals. According to the method of Chang's extent analysis, each object is taken and extent analysis for each goal is performed, respectively. Therefore, *m* extent analysis values for each object can be obtained with the following signs: $M^1_{gi},M^2_{gi},\ldots,M^m_{gi}$, i=1,2,…,*n*, where all $M^j_{gi}(j=1,2,\ldots,m)$ are triangular fuzzy numbers. Among various membership functions, the triangular fuzzy number is the most popular in the engineering applications. The triangular fuzzy number M̃ is denoted simply by *(l, m, u)* and shown in Fig. 2. The parameters l and u, respectively, represent the smallest and the largest possible values and m stands for the most promising value that describe a fuzzy event. Each triangular fuzzy number has linear representations on its left and right side such that its membership function can be defined as the following:

$$\mu(x)=\begin{cases}0 & x<l\\ (x-l)/(m-l) & l\le x\le m\\ (u-x)/(u-m) & m\le x\le u\\ 0 & x>u\end{cases} \quad (1)$$

In this study, only addition and multiplication are used. Defining two triangular fuzzy numbers M_1 and M_2 by the triplets as $M_1=(l_1,m_1,u_1)$ and $M_2=(l_2,m_2,u_2)$ the addition and multiplication operations of M1 and M2 can be expressed as follows:

Addition: if $\oplus$ denotes addition.

$$M_1\oplus M_2:(l_1,m_1,u_1)\oplus(l_2,m_2,u_2)$$
$$=(l_1+l_2,m_1+m_2,u_1+u_2) \quad (2)$$

Multiplication: if $\otimes$ denotes multiplication.

$$M_1\otimes M_2:(l_1,m_1,u_1)\otimes(l_2,m_2,u_2)$$
$$=(l_1+l_2,m_1+m_2,u_1+u_2),\quad l_1,l_2\ge 0 \quad (3)$$

The steps of Chang's analysis can be given as in the following:

Step 1: The AHP framework is composed of a goal, a set of factors and related sub-factors. The components of the framework are related to each other by different types of conjunctive arrows (unidirectional and bilateral) based on relationship types.

Step 2: The local weights of the factors and sub-factors are determined by pair-wise comparisons. In this step, the factors are compared with each other assuming that there is no dependency among them. The fuzzy synthetic extent value (S_i) with respect to the i^{th} criterion is defined as:

$$S_i=\sum_{j=1}^{m}M^j_{g_i}\otimes\left[\sum_{i=1}^{n}\sum_{j=1}^{m}M^j_{g_j}\right]^{-1},\quad i=1,2,\ldots.n \quad (4)$$

$$\sum_{j=1}^{m}M^j_{g_i}=\sum_{j=1}^{m}(l^i_j,m^i_j,u^i_j)=(\sum_{j=1}^{m}l_j,\sum_{j=1}^{m}m_j,\sum_{j=1}^{m}u_j) \quad (5)$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m}M^j_{g_i}=\sum_{i=1}^{n}(\sum_{j=1}^{m}l_j,\sum_{j=1}^{m}m_j,\sum_{j=1}^{m}u_j)=$$
$$(\sum_{i=1}^{n}\sum_{j=1}^{m}l_j,\sum_{i=1}^{n}\sum_{j=1}^{m}m_j,\sum_{i=1}^{n}\sum_{j=1}^{m}u_j) \quad (6)$$

$$\left[\sum_{i=1}^{n}\sum_{j=1}^{m}M^j_{gi}\right]^{-1}=$$
$$(1/\sum_{i=1}^{n}\sum_{j=1}^{m}u_j,1/\sum_{i=1}^{n}\sum_{j=1}^{m}m_j,1/\sum_{i=1}^{n}\sum_{j=1}^{m}l_j). \quad (7)$$

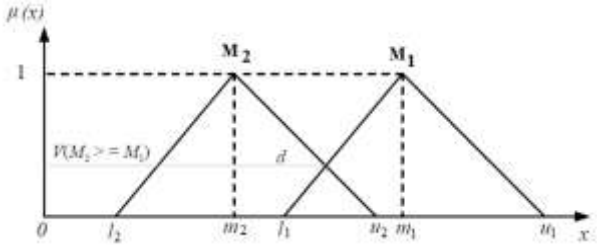


Fig. 2. Membership functions of linguistic variables

As $M_1=(l_1,m_1,u_1)$ and $M_2=(l_2,m_2,u_2)$ are two triangular fuzzy numbers, the degree of possibility of $M_2=(l_2,m_2,u_2)\ge M_1=(l_1,m_1,u_1)$ defined as (see Fig.2):

$$V(M_2\ge M_1)=\sup_{y\ge x}[\min(\mu_{M_1}(x),\mu_{M_2}(x))]=$$
$$hgt(M_2\cap M_1)=\mu_{M_2}(d)=$$
$$\begin{cases}1 & m_2\ge m_1\\ 0 & l_1\ge u_2\\ \dfrac{l_1-u_2}{(m_2-u_2)-(m_1-l_1)} & otherwise\end{cases} \quad (8)$$

were *x* and *y* are the values on the axis of membership function of each criterion and *d* is the highest intersection point μ_{M_1} and μ_{M_2}.

To compare M_1 and M_2; we need both the values of $V(M_2\ge M_1)$ and $V(M_1\ge M_2)$.

Step 3: The degree possibility for a convex fuzzy number to be greater than k convex fuzzy numbers $M_i(i = 1,2,\dots,k)$ can be defined by:

$$V(M \geq M_1, M_2, \dots, M_k) = V[(M \geq M_1) and (M \geq M_2 and \dots . and (M \geq M_k)] = \min V(M \geq M_i) \quad (9)$$

Assume that $d'(A_i) = minV(S_i \geq S_k)$, for $k=1,2,3,4,5,\dots\dots,n; k \neq i$, then the weight vector is given by:

$$W' = [d'(A_1), d'(A_2), \dots, d'(A_n)]^T \quad (10)$$

Where A_i *(i = 1, 2, 3, 4, 5, 6, …., n)* are *n* elements.

Step 4: Via normalization, the normalized weight vectors are:

$$W = [d(A_1), d(A_2), \dots, d(A_n)]^T \quad (11)$$

Where W is non-fuzzy numbers. Also, the non-fuzzy weight factor would be as $W = (minV(S_1 \geq S_j), minV(S_2 \geq S_j), minV(S_n \geq S_j))$.

The weight factor is normalized and used in the third step.

Step 5: The weights of the factors and sub-factors are determined.

Step 6: The selection for the different of alternatives is determined.

4. ERP System Selection Framework

4.1 Procedure of Selection

This research, a framework is developed using fuzzy analytic hierarchy process (FAHP) to selection of an ERP system. The methodology comprises of many steps. Every ERP project is considered as a multi-stage process. Hence, a framework proposed to better explain different dimensions of the select most suitable ERP system.

Figs. 3 illustrate the conceptual framework of the proposed methodology for the ERP selection process. The complete procedure of our proposed ERP selection model is shown in Fig. 3. The model involves four principle essentials. In this paper an ERP selection methodology is proposed. The evaluation procedure of this study consists of seven steps as follows Fig. 3:

I. Organize the committee of decision makers:

First of all, managers formed a project team which was included personnel chosen from different departments and was supported by top management to select an ERP system.

II. Identify the ERP system criteria:

The project team created a vision to define the corporate mission, objectives, and strategy.

III. Construct the structure of objectives of ERP selection project:

The project team conducted the business process reengineering with a function list which was created to define what the requirements were.

IV. Extract the attributes for selecting ERP systems:

Selecting a suitable ERP project involves various factors. Project team made a preliminary analysis of the strengths and weaknesses of each criteria. Team members expressed their opinions on the importance and the strengths of the relationships between selection criteria pair wises in the form of linguistic variables such as very strong, strong, medium, weak, and none to build the structure of comparison frame.

V. Identify ERP system alternatives:

After collecting all possible information about the current system and establishing the evaluation criteria, the project team evaluated all software vendors' characteristics in the market. Finally, they filtered out unqualified vendors and selected three software vendors.

VI. Evaluate the ERP systems by the fuzzy AHP method:

To help the project team make a decision, we offer to use fuzzy AHP decision making methodology to decide on the best vendor to select. The fuzzy AHP approaches allow team members to use their experience, values and knowledge to decompose a problem into smaller sets by solving them with their own procedures in making a decision.

VII. Make the final decision ERP system alternatives:

Discuss the results and make the final decision. The project team compared the sub-attributes with respect to main attributes in the hierarchical approach by utilizing fuzzy triangular numbers in fuzzy AHP procedure. A detailed questionnaire related with the data regarding the qualitative criteria for ERP selection model was prepared for the paired comparisons to tackle the ambiguities involved in the process of the linguistic assessment of the data. Finally, with the weights of importance we attempted to find best ERP vendor among all alternatives.

4.2 The AHP Model

The AHP hierarchy is composed of four levels, as illustrated in Fig. 4. Level 1 reveals the goal for selection the most suitable ERP system. Level 2 consists of three main objectives, namely choosing the product factors, system factors and management factors. Level 3 contains the associated attributes that are used to measure various products, systems and management, respectively. The four level consists of the alternative ERP systems.

The ERP selection critical success factors have been vastly addressed and analyzed in ERP literature by many researchers [44-46].

From this model, main critical of the selection are determined. Then, critical success factors for ERP selection are evaluated and the assessment factors are determined. The factors are grouped and the assessment framework is constructed. The fuzzy analytic hierarchy process is then used for this framework (Fig.4).

Criteria of product: Selecting a suitable ERP project involves various factors. The product criteria is derived from the international norm ISO/IEC 9126 [54]. The ISO 9126 software quality model is chosen to describe the

ERP product characteristic and we categorize it as product aspect in the model. This quality model identifies five external attributes of interest, namely functionality, reliability, efficiency, usability and maintainability.

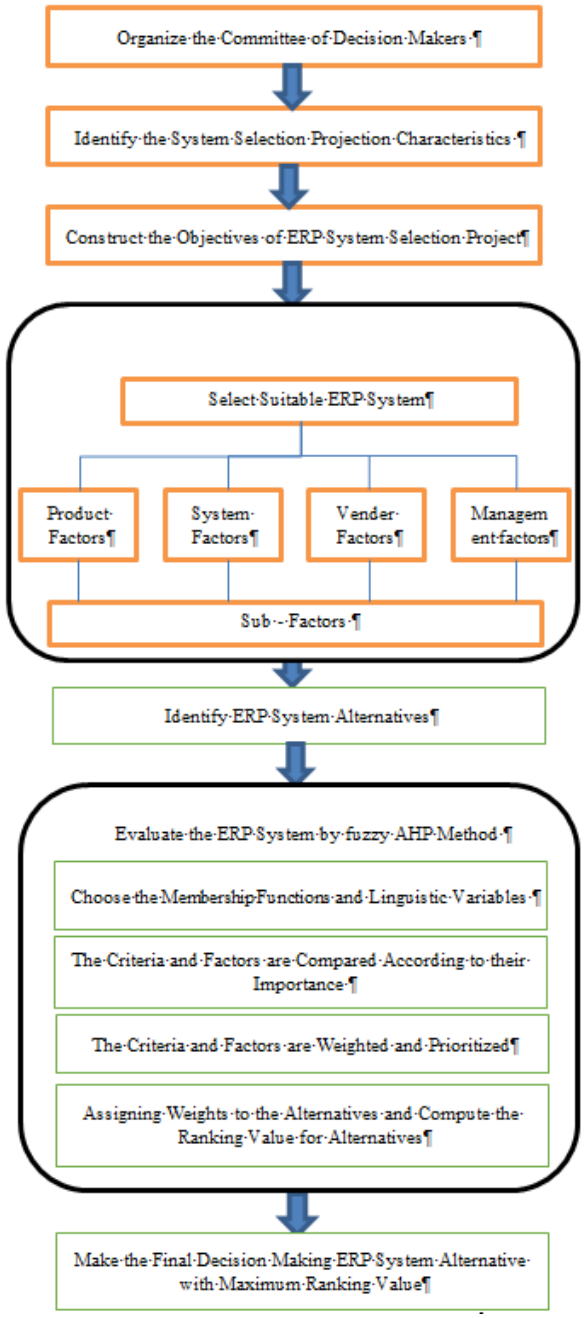


Fig. 3. The proposed methodology for the selection of ERP system.

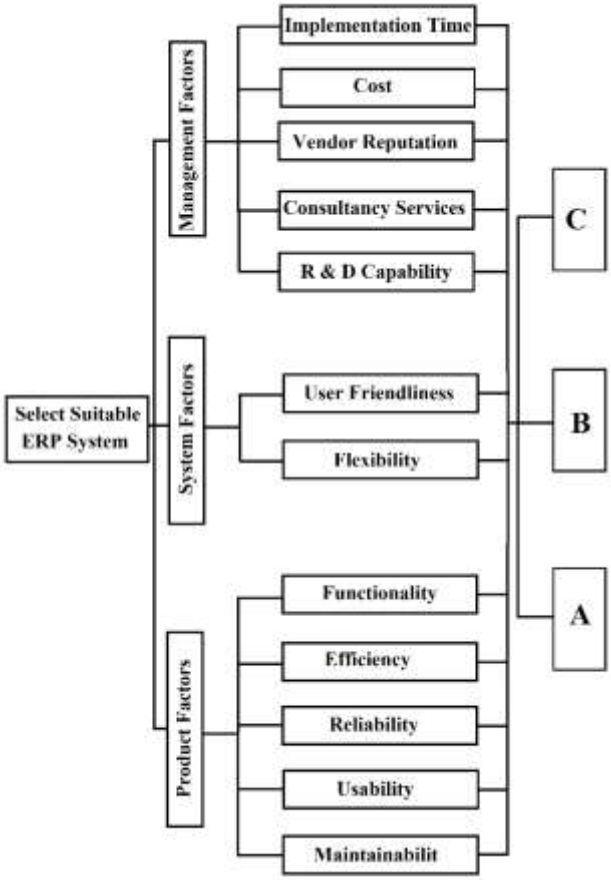


Fig. 4. AHP framework for ERP selection

The detailed characterization is presented as follows [55]:

(1) Functionality :This attribute is defined as the degree to which the software functions satisfies stated or implied needs and can be broken down into five sub-characteristics as follows: suitability, accuracy, interoperability, compliance and security.

(2) Reliability :This attribute is defined as the capability of software that could maintain its level of performance under stated conditions for a stated period of time. It can be decomposed into three sub-characteristics as follows: maturity, fault tolerance and recoverability.

(3) Usability :This attribute is defined as the degree to which the software is available for use and can be broken down into three sub-characteristics as follows: understandability, learnability and operability.

(4) Efficiency :This attribute is defined as the degree to which the software makes optimal use of system resources. It can be decomposed into two sub-characteristics as follows: efficiency of time behavior and efficiency of resource behavior.

(5) Maintainability :This attribute is defined as the ease with which repair may be made to the software and can be broken down into four sub-characteristics as follows: analyzability, changeability, stability and testability.

Criteria of management: the management criteria of ERP system contains five major criteria: vender factors,

cost factors and time factors. The detailed characterization of three factors is presented as follows:

(1) Sub criteria of vender factors: market share and reputation, industrial credential, service and support, training solution. We gathered these factors based on vendor's reputation. By vendor's ability criteria, we implied vendor's technology level, implementation and service ability, consulting service, training support. As far as vendor's condition we considered vendor's financial condition, certifications and credentials.

(2) Sub criteria of cost factors: software cost, hardware cost, annual maintenance cost, and staff training cost. This price contains licensing arrangement cost, product and technology cost and consulting cost, which involves adapting and integrating cost, supporting cost, training cost, maintenance (upgrades) cost.

(3) Sub criteria of time factors: time for planning and preparation, time for BPR and system tuning, time for testing and go-live.

In addition to management and product criteria we considered system criteria such as user friendliness and flexibility.

As shown in Fig. 4 our model includes four hierarchy levels. Finally, with the weights of importance we attempted to find best ERP vendor among all alternatives.

5. The Application of FAHP in ERP System Selection

The AHP model provides priority weights for the ERP packages, based on the ERP project team's preferences on multiple characteristics. The alternative with the highest priority weight is then selected for the company (Fig. 4). A case study in Iran belong to different industries are conducted to prove the practicality of our proposed model in this section.

The proposed model is composed of four hierarchical stages: goal, sub-goals (factors), sub-factors and alternative ERP systems which are related to each other by means of conjunctive arrows. This model has been applied to measure the firm's readiness to selection an ERP system. Firstly, the general manager of company organizes the project team including eight senior managers in different sections. Unfavorable alternatives are eliminated by thorough examination of system specifications and requirements derived from the main goals. After the preliminary elimination which is subjected to budget, time and system functions, three feasible ERP system alternatives are came out. The sub-factors are determined according to the vision and the strategies of the company. After assigning the weights to each sub-factor, the evaluation team compared all ERP alternatives. Assume that twelve sub-factors are evaluated under a fuzzy environment. For selecting best ERP, main factors the product factors, system factors and management factors are used in application, are explained in fuzzy sets and fuzzy numbers. Fig.4 shows the all main factors and sub-factors. The project team compared the sub-factors with respect to main factors in the hierarchical approach by utilizing fuzzy triangular numbers in fuzzy AHP procedure. To create pair wise comparison matrix, linguistic scale is used which is given in Table 1.

The matrix of paired comparisons for alternatives (A, B and C) is given in Tables 2-6. Tables 3-6 show the judgment matrix (Pairwise comparisons) and weight vector of each matrix.

Table 1. Linguistic scale for relative importance

Linguistic scale	Triangular fuzzy scale	Inverse Triangular fuzzy scale
Just Equal	(1,1,1)	(1,1,1)
Equal importance	(1/2,1,3/2)	(2/3,1,2)
Moderate importance	(1,3/2,2)	(1/2,2/3,1)
Strong importance	(3/2,2,5/2)	(2/5,1/2,2/3)
Very Strong importance	(2,5/2,3)	(1/3,2/5,1/2)
Absolutely importance	(5/2,3,7/2)	(2/7,1/3,2/5)

According to decision maker's preferences for main factors, pair wise comparison values are transformed into triangular fuzzy number's as in Table 2.

After forming fuzzy pairwise comparison matrix, weights of all main factors are determined by the help of FAHP. According to the FAHP method, firstly synthesis values must be calculated.

Table 2. The Fuzzy pairwise comparison matrix regarding main factors

Main Factors	Product	System	Management
Product	(1,1,1)	(2,5/2,3)	(1/2,1,3/2)
System	(1/3,2/5,1/2)	(1,1,1)	(1/2,2/3,1)
Management	(2/3,1,2)	(1,3/2,2)	(1,1,1)

Tables 3, 4 and 5 indicate the product, system, and management sub-factors' Pairwise comparisons.

Table 3. Judgment matrix (Pairwise comparisons) of product sub-factors

Product	Functionality	Reliability	Usability	Efficiency	Maintainability
Functionality	(1,1,1)	(5/2,3,7/2)	(2/7,1/3,2/5)	(1,3/2,2)	(3/2,2,5/2)
Reliability	(2/7,1/3,2/5)	(1,1,1)	(3/2,2,5/2)	(1,3/2,2)	(1/2,1,3/2)
Usability	(5/2,3,7/2)	(2/5,1/2,2/3)	(1,1,1)	(2/5,1/2,2/3)	(1,3/2,2)
Efficiency	(1/2,2/3,1)	(1/2,2/3,1)	(3/2,2,5/2)	(1,1,1)	(1/2,1,3/2)
Maintainability	(2/5,1/2,2/3)	(2/3,1,2)	(1/2,2/3,1)	(2/3,1,2)	(1,1,1)

Table 4. Pairwise comparisons of system sub-factors

System	User friendliness	Flexibility
User friendliness	(1,1,1)	(0.25,0.5,0.75)
Flexibility	(1.33,2,4)	(1,1,1)

Table 5. Pairwise comparisons of management sub-factors

Management	Cost	Implementation Time	Vendor Reputation	Consultancy Services	R&D Capability
Cost	(1,1,1)	(3/2,2,5/2)	(1/3,2/5,1/2)	(1/2,1,3/2)	(1,3/2,2)
Implementation Time	(2/5,1/2,2/3)	(1,1,1)	(3/2,2,5/2)	(1/2,1,3/2)	(2,5/2,3)
Vendor Reputation	(2,5/2,3)	(3,2,1/2,2/3)	(1,1,1)	(2/7,1/3,2/5)	(1/2,1,3/2)
Consultancy Services	(2/3,1,2)	(2/3,1,2)	(5/2,3,7/2)	(1,1,1)	(3/2,2,5/2)
R&D Capability	(1/2,2/3,1)	(1/3,2/5,1/2)	(2/3,1,2)	(2/5,1/2,2/3)	(1,1,1)

Table 6 indicate the pairwise comparisons of alternatives, the ERP vendors, (A, B and C) regarding various sub-factor of product, system, management and vendor factor.

Table 6. Pairwise comparisons of alternatives with twelve categories (sub-factors)

Functionality	**A**	**B**	**C**
A	(1,1,1)	(1/2,2/3,1)	(2/3,1,2)
B	(1,3/2,2)	(1,1,1)	(1/2,1,3/2)
C	(1/2,1,3/2)	(2/3,1,2)	(1,1,1)
Reliability	A	B	C
A	(1,1,1)	(1,1,1)	(1/2,2/3,1)
B	(1,1,1)	(1,1,1)	(2/3,1,2)
C	(1,3/2,2)	(1/2,1,3/2)	(1,1,1)
Usability	A	B	C
A	(1,1,1)	(2/3,1,2)	(1,1,1)
B	(1/2,1,3/2)	(1,1,1)	(1/2,1,3/2)
C	(1,1,1)	(2/3,1,2)	(1,1,1)
Efficiency	A	B	C
A	(1,1,1)	(1,3/2,2)	(2/3,1,2)
B	(1/2,2/3,1)	(1,1,1)	(2/3,1,2)
C	(2/3,1,2)	(1/2,1,3/2)	(1,1,1)
Maintainability	A	B	C
A	(1,1,1)	(1,1,1)	(1,3/2,2)
B	(1,1,1)	(1,1,1)	(1/2,1,3/2)
C	(1/2,2/3,1)	(2/3,1,2)	(1,1,1)
Cost	A	B	C
A	(1,1,1)	(1/2,1,3/2)	(1/2,1,3/2)
B	(2/3,1,2)	(1,1,1)	(1/2,1,3/2)
C	(2/3,1,2)	(2/3,1,2)	(1,1,1)
Implementation time	A	B	C
A	(1,1,1)	(2/3,1,2)	(2/3,1,2)
B	(1/2,1,3/2)	(1,1,1)	(1,1,1)
C	(1/2,1,3/2)	(1,1,1)	(1,1,1)
User Friendliness	A	B	C
A	(1,1,1)	(1/2,1,3/2)	(1,3/2,2)
B	(2/3,1,2)	(1,1,1)	(1/2,1,3/2)
C	(1/2,2/3,1)	(2/3,1,2)	(1,1,1)
Flexibility	A	B	C
A	(1,1,1)	(1/2,1,3/2)	(1,3/2,2)
B	(2/3,1,2)	(1,1,1)	(1/2,1,3/2)
C	(1/2,2/3,1)	(2/3,1,2)	(1,1,1)
Vendor Reputation	A	B	C
A	(1,1,1)	(1/2,2/3,1)	(2/3,1,2)
B	(1,3/2,2)	(1,1,1)	(1/2,1,3/2)
C	(1/2,1,3/2)	(2/3,1,2)	(1,1,1)
Consultancy Services	A	B	C
A	(1,1,1)	(1/2,2/3,1)	(2/3,1,2)
B	(1,3/2,2)	(1,1,1)	(1/2,1,3/2)
C	(1/2,1,3/2)	(2/3,1,2)	(1,1,1)
R&D Capability	A	B	C
A	(1,1,1)	(1,1,1)	(2/3,1,2)
B	(1,1,1)	(1,1,1)	(2/3,1,2)
C	(1/2,1,3/2)	(1/2,1,3/2)	(1,1,1)

5.1 Data Analysis

The final fuzzy weights of 12 sub-factors are calculated as shown in Table 7. Table 8 shows the final scores for the ERP vendors. As shown in Table 8, the ERP system B is the dominant solution in the final rank. The alternative with maximum weight value is the best choice in the decision-making problem.

Table 7. Final fuzzy weights of sub-factors

Factors	**Fuzzy sub-factors weights**
Functionality	(0.51,0.76,0.95)
Reliability	(0.49,0.77,0.92)
Usability	(0.54,0.81,0.96)
Efficiency	(0.37,0.66,0.84)
Maintainability	(0.36,0.56,0.79)
Cost	(0.51,0.76,0.96)
Implementation time	(0.56,0.79,0.99)
User Friendliness	(0.62,0.89,0.99)
Flexibility	(0.39,0.62,0.92)
Vendor Reputation	(0.51,0.71,0.84)
Consultancy Services	(0.47,0.72,0.99)
R&D Capability	(0.59,0.84,0.99)

According to Table 7, the decision makers are fairly consistent in ranking the attributes. For valuation, the consistency index of each decision maker's paired comparison matrix should be less than the threshold value 0.1 to ensure that the decision maker was consistent in assigning paired comparisons, otherwise the decision maker may need to reconsider his evaluation [21].

From Table 8, the weight of ERP vendor alternative B is 0.6572, and the weights for ERP vendor alternatives A and C are 0.2278 and 0.0898 respectively. According to fuzzy AHP method, the best ERP vendor alternative is B. Thus, the project team agrees that system B is the most suitable decision for Company.

Table 8. The ranking values of the fuzzy appropriateness indices for alternatives

Alternatives	**Fuzzy Weight**	**Non- Fuzzy Weight**	**Final Ranking**
A	(0.1622,0.2204,0.3007)	**0.2278**	2
B	(0.5080,0.6483,0.8152)	**0.6572**	1
C	(0.0573,0.0856,0.1264)	**0.0898**	3

It is obvious that the most appropriate ERP system is B. Thus, the committee can be comfortable in recommending alternative B as the most suitable ERP system for the selection project for this company.

The reason for choosing the combination of AHP and fuzzy is based on these decision modeling techniques' strengths and suitability to the current decision situation. The specific reason of combining of AHP and fuzzy in our study can be described as follows: First of all, it is an out ranking method suitable for ranking the alternatives among conflicting criteria. The second is that fuzzy is a rather simple ranking method with respect to conception and application when compared with the other MCDM methods. Third one is the popularity of it.

5.2 Comparisons

5.2.1 Comparison with Kilic Approach

In Kilic, (2014), an ERP system selection problem at a large airline company in Turkey is considered. First, based on the requirements and the demands of the

company executives, the ERP selection criteria are determined. Then, the alternative ERP firms and their offerings are investigated and determined. After determining the criteria and solution alternatives, the proposed hybrid methodology, consisting of fuzzy AHP which incorporates the vagueness of the decision making process and TOPSIS, is applied and validated. Specifically, the importance/weights of the selection criteria are obtained via fuzzy AHP based on the triangular fuzzy preference scales. Then these weights are used in the TOPSIS methodology to reach the ranking of alternative ERP system suppliers.

The use of a hybrid selection/evaluation methodology proved to produce results that are both technically sound and organizationally acceptable. Knowing that the vagueness and complexity of the decision situation are handled using the strengths of two popular decision support methods makes the decision makers confident in their final selection. They feel that by breaking the complex problem space into smaller pieces, dealing with them at that granular level, and then aggregating them at the higher decision level have a much better chance of producing optimal (or near optimal) decisions.

Weakness: It should be acknowledged that the paper of [53] is subject to some limitations. Perhaps the most serious limitation of this study is its narrow focus on a single case study in aviation industry. To generalize on the findings and the viability/validity/value of the methodology, more real-world cases need to be performed. Another limitation of the individual methods is the independent structure of the selection criteria. Since the comparisons are made in a piece-meal/pairwise fashion, reaching the true optimal may not be possible. Also, for manageability purposes, various low-level criteria are grouped in clusters, by doing so, some detailed specifications may have been lost. Finally, the methodology proposed in this study, as systematics as it may sound, is a heuristic one. That is, it does not guarantee finding the optimal solution. The "optimality" of the results is often subject to the richness (in terms of quantity and quality) of the participants; positively influenced by their knowledge, experience and dedication.

5.2.2 Comparison with Cebeci Approach

In Cebeci, (2009), presents an approach to select a suitable ERP system for textile industry. The proposed ERP selection methodology was applied successfully for a textile manufacturing company for young people as a real case study. The methodology also gives some suggestions about successful ERP implementation. The proposed methodology can be used for other sectors with some changes. Decisions are made today in increasingly complex environments. In more and more cases the use of experts in various fields is necessary, different value systems are to be taken into account, etc. In many of such decision-making settings the theory of fuzzy decision making can be of use. Fuzzy group decision-making can overcome this difficulty. In general, many concepts, tool and techniques of artificial intelligence, in particular in the field of knowledge representation and reasoning, can be used to improve human consistency and implement ability of numerous models and tools in broadly perceived decision-making and operations research. The proposed decision support system integrated with strategic management by using BSC may be an alternative to some methods for ERP selection. In this paper, ERP packages and vendors for textile companies were compared using fuzzy AHP.

The presented methodology is flexible and can be used for other sectors with some sector specific characteristics changes. Humans are often uncertain in assigning the evaluation scores in crisp AHP. Fuzzy AHP can capture this difficulty.

Weakness: In this paper, [50], Fuzzy AHP cannot support all phases of ERP selection and implementation. Hence, an intelligent decision support system or expert system can be added when gathering data for selection process.

6. Conclusions

In this paper, we present an approach to select a suitable ERP system. In order to deal with this problem appropriately, the analytic hierarchy process (AHP) method is extended into a fuzzy domain. A framework is developed to select most suitable ERP system using this fuzzy AHP. The factors and sub-factors are determined, classified, weighted and prioritized and then a framework is provided for ERP selection with the fuzzy analytic hierarchy process (FAHP) method. Then, we used fuzzy AHP to obtain pairwise comparison judgments by prioritizing criteria and assigning weights to the factors and alternatives. The framework decomposes ERP system selection into four main factors. The goal of this paper is to select the best alternative that meets the requirements with respect to "product factors", "system factors" and "management factors". The sub-attributes (sub-factors) related to ERP selection have been classified into twelve main categories of ''Functionality", ''Reliability", ''Usability", ''Efficiency", ''Maintainability", ''Cost", ''Implementation time", ''User friendliness", ''Flexibility ", ''vendor Reputation", ''Consultancy Services", and ''R&D Capability"and arranged in a hierarchy structure.

In this paper intends to show how effective is fuzzy AHP as a decision-making tool in system selection problem. Even with the complete accurate information, different decision making methods may lead to totally different results. Thus, the proposed methodology demonstrates the selection of the best ERP vendor under the cost and product quality restrictions in the presence of vagueness. It is seen that fuzzy AHP is a useful decision-making methodology to make more precise selection-decisions that may help the company to achieve a competitive edge in a complexity environment. Fuzzy AHP approach incorporates quantitative data of the criteria, which have to be evaluated by qualitative measures. The proposed selection methodology is flexible to incorporate new or extra criteria or decision making for the evaluation process.

A real case study from Iran is also presented to demonstrate efficiency of this method in practice. In the future we offer to apply other decision-making methods using fuzzy concept to capture the uncertainty in complex approaches. Also, in this topic it is possible to make the decision by using fuzzy analytic network process (ANP) model and compare with fuzzy AHP model and the expert system can be used before the ERP system selected.

Acknowledgements

We are grateful to the comments from the Mrs. M. Meshginfam that significantly improved the quality of this paper.

References

[1] S.Onut, T.Efendigil, "A theorical model design for ERP software selection process under the constraints of cost and quality: A fuzzy approach," Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, Volume 21 Issue 6, pp. 365-

[2] J.Esteves, and J.Pastor, "Towards the unification of critical success factors for ERP implementations," published in 10thAnnual Business Information Technology (BIT) 2000 Conference, Manchester, 2000.

[3] C.C. Wei, and M.J.Wang., "A comprehensive framework for selecting an ERP system". *International Journal of Project Management* 22, pp.161-169, 2004.

[4] E.Turban, and J.Aronson. *Decision Support Systems and Intelligent Systems*: Prentice Hall, 1998.

[5] U. Cebeci, & D. Ruan, "A multi-attribute comparison of Turkish quality consultants by fuzzy AHP". International Journal of Information Technology and Decision Making, 6(1), 191–207, 2007.

[6] M. J. Schniederjans, and R.L.Wilson, "Using the analytic hierarchy process and goal programming for information system project selection". *Information & Management* 20, pp.333- 342, 1991.

[7] C.C.Wei, C.F. Chien, and M.J Wang. "An AHP-based approach to ERP system selection". *International Journal of Production Economics* 96, pp.47-62, 2005.

[8] C. H. Lin, "An ERP application study on college property management system". Master Thesis, Nanhua University, Chiayi, Taiwan, R.O.C., 2002.

[9] A.Ishizaka, N.H. Nguyen, "Calibrated fuzzy AHP for current bank account selection," Expert Systems with Applications: An International Journal, v.40 n.9, p.3775-3783, July. 2013.

[10] G.A. Langenwalter, *Enterprise Resources Planning and Beyond Integrating Your Entire Organization*. CRC Press LLC, Florida, 2000.

[11] J.Butler. "Risk management skills needed in a packaged software environment". *Information System Management* 16(3), pp.15-20, 1999.

[12] P.Bingi, M.K. Sharma, and J.K.Godla, "Critical issues affecting and ERP implementation". *Information Systems Management* 16(3), pp. 30-36, 1999.

[13] S.Parthasarathy, "Research directions for enterprise resource planning (ERP) projects," International Journal of Business Information Systems, v.9 n.2, p.202-221, 2012.

[14] A.Teltumbde, "A framework of evaluating ERP projects", International Journal of Production Research, No. 38, pp. 4507–4520, 2000.

[15] CA.Ptak, "ERP tools, techniques, and applications for integrating the supply chain," St. Lucie Press, 2000.

[16] K.Chen, and N.Gorla, "Information system project selection using fuzzy logic", IEEE transactions on systems, man, and cybernetics part A: Systems and Humans, No. 28, pp.849–55, 1998.

[17] J.W. Lee, and SH.Kim, "An integrated approach for interdependent information system project selection", International Journal of Project Management, No.19, pp.111–118, 2001.

[18] MA. Badri, and D.Davis,"A comprehensive 0–1 goal programming model for project selection", International Journal of Project Management, No. 19, pp. 243–52, 2001.

[19] F.Wilson, J.Desmond, and H.Roberts, "Success and failure of MRPII implementation", British Journal of Management, No. 5, pp.221–40, 1999.

[20] J.May, G. Dhillon, M.Caldeira, "Defining value-based objectives for ERP systems planning," Decision Support Systems, v.55 n.1, p.98-109, 2013.

[21] T.L. Saaty, "The Analytic Hierarchy Process," McGraw-Hill, New York, 1980.

[22] G.H.Tzeng, J.J. Huang, "Multiple Attribute Decision Making Method and applications," CRC Press, Taylor & Francis Group, 2011.

[23] D.Bouyssou, T.Marchant, M.Pirlot, P.Perny, , A.Tsoukias, , and P.Vincke, Evaluation Models: A Critical Perspective, Kluwer, Boston, 2000.

[24] C. G. E.Boender, J. G.De Graan, and F. A.Lootsma, "Multicriteria Decision Analysis with Fuzzy Pairwise Comparisons", Fuzzy Sets and Systems, 29, 133-143, 1989.

[25] J. J.Buckley, "Ranking Alternatives Using Fuzzy Members", Fuzzy Sets and Systems, 15, 21-31, 1985.

[26] J. J.Buckley, "Fuzzy Hierarchical Analysis", Fuzzy Sets and Systems, 17, 233-247, 1985.

[27] D. Y.Chang, "Applications of the Extent Analysis Method on Fuzzy-AHP", European Journal of Operational Research, 95, 649-655, 1996.

[28] P. J. M.Laarhoven, and W.Pedrycz, "A Fuzzy Extension of Saaty's Priority Theory", Fuzzy Sets and Systems, 11, 229-241, 1983.

[29] F. Lootsma, "Fuzzy Logic for Planning and Decision-Making," Kluwer, Dordrecht, 1997.

[30] R. A.Ribeiro, "Fuzzy Multiple Criterion Decision Making: A Review and New Preference Elicitation Techniques", Fuzzy Sets and Systems, 78, 155-181, 1996.

[31] O.Duru, E.Bulut, and, S.Yoshida, "Regime switching fuzzy AHP model for choice-varying priorities problem and expert consistency prioritization: A cubic fuzzy-priority matrix design," Expert Systems with Applications: An International Journal, v.39 n.5, p.4954-4964, 2012.

[32] Lee,S., Kim,W. , Kim,Y.M , Joo Oh,K .(2012) . Using AHP to determine intangible priority factors for technology

transfer adoption, Expert Systems with Applications: An International Journal, v.39 n.7, p.6388-6395.
[33] C. S.Yu, "A GP-AHP Method for Solving Group Decision-Making Fuzzy AHP Problems", Computers and Operations Research, 29, 1969-2001, 2002.
[34] M.Weck, F.Klocke, H.Schell, and E.Rüenauver, "Production Cycles Using The Extended Fuzzy AHP Method", European Journal of Operational Research, 100, 2, 351-366, 1997.
[35] J. B., Sheu, "A Hybrid Fuzzy-Based Approach for Identifying Global Logistics Strategies", Transportation Research, 40, 39-61, 2004.
[36] O.Kulak, and C.Kahraman, "Fuzzy Multi-Criterion Selection among Transportation Companies Using Axiomatic Design and Analytic Hierarchy Process", Information Sciences, 170, pp.191-210, 2005.
[37] R. J.Kuo, S. C.Chi, and S. S.Kao, "A Decision Support System for Selecting Convenience Store Location Through Integration of Fuzzy AHP and Artificial Neural Network", Computers in Industry, in Press, 2002.
[38] K.J. Zhu, Y.Jing, D. Y.Chang,. "A discussion on Extent Analysis Method and applications of fuzzy AHP," European Journal of Operational Research, 116, Volume, Pages 450 – 456, 1999.
[39] D.Khosroanjom, M.Ahmadzade, A.Niknafs, R. Kiani Mavi, "Using fuzzy AHP for evaluating the dimensions of data quality," International Journal of Business Information Systems, v.8 n.3, p.269-285, September. 2011.
[40] C. H.Cheng, K. L.Yang, and C. L.Hwang, "Evaluating Attack Helicopters by AHP Based on Linguistic Variable Weight", European Journal of Operational Research, 116, pp.423-435, 1999.
[41] D. Y. Chang, "Extent Analysis and Synthetic Decision", Optimization Techniques and Applications, World Scientific, Singapore, 1, 352, 1992.
[42] C.Jian You, C. K. M. Lee , S. L. Chen , J. Jiao, "A real option theoretic fuzzy evaluation model for enterprise resource planning investment," Journal of Engineering and Technology Management, v.29 n.1, p.47-61,v.29 n.1, pp.47-61, 2012.
[43] G.Büyüközkan, G. Çifçi,G, S.Güleryüz, "Strategic analysis of healthcare service quality using fuzzy AHP methodology," Expert Systems with Applications: An International Journal, v.38 n.8, p.9407-9424, August,2011.
[44] H. S. Kilic, S. Zaim, D. Delen , Selecting "The Best" ERP system using a combination of ANP and PROMETHEE methods, Expert Systems with Applications, Volume 42, Issue 5, Pages 2343-2352, 1 April 2015.
[45] D.Liao, G.Sun, V.Anand, H.Yu,.Reliable Design for Stochastic Multicast Virtual Network in Data Centres .IETE Technical Review. Volume 31, Issue 5, September 2014, pages 327-341. 2014.
[46] T.Ma, Y.Chu, L.Zhao, & O.Ankhbayar. Resource Allocation and Scheduling in Cloud Computing: Policy and Algorithm, .IETE Technical ReviewVolume 31, Issue 1, pages 4-16, January 2014.
[47] Ayağ, Z., & Özdemir, R. G. An intelligent approach to ERP software selection through fuzzy ANP. International Journal of Production Research, 45(10), 2169–2194, 2007.
[48] Perçin, S. Using the ANP approach in selecting and benchmarking ERP systems. Benchmarking: An International Journal, 15(5), 630–649, 2008.
[49] Ünal, C., & Güner, M. G. Selection of ERP suppliers using AHP tools in the clothing industry. International Journal of Clothing Science and Technology, 21(4), 239–251, 2009.
[50] Cebeci, U. Fuzzy AHP-based decision support system for selecting ERP systems in textile industry by using balanced scorecard. Expert Systems with Applications, 36(5), 8900–8909, 2009.
[51] Kahraman, C., Beskese, A., & Kaya, I.. Selection among ERP outsourcing alternatives using a fuzzy multi-criteria decision making methodology. International Journal of Production Research, 48(2), 547–566, 2010.
[52] Sen, C. G., Baraçlı, H., Sen, S., & Baslıgil, H.. An integrated decision support system dealing with qualitative and quantitative objectives for enterprise software selection. Expert Systems with Applications, 36, 5272–5283, 2009.
[53] Kilic, H. S., Zaim, S., & Delen, D. Development of a hybrid methodology for ERP system selection: The case of Turkish Airlines. Decision Support Systems. http://dx.doi.org/10.1016/j.dss.2014.06.011.,2014.
[54] T. Gürbüz, S. Emre Alptekin, G. I. Alptekin , A hybrid MCDM methodology for ERP selection problem with interacting criteria, Decision Support Systems, Volume 54, Issue 1, Pages 206-214,December2012.
[55] R.Bache, and G.Bazzana, Software metrics for Product Assessment, McGraw-Hill, England, 1994.

Hodjatollah Hamidi, born 1978, in shazand Arak, Iran, He got his Ph.D. in Computer Engineering. His main research interest areas are Information Technology, Fault-Tolerant systems (fault-tolerant computing, error control in digital designs) and applications and reliable and secure distributed systems and e- commerce. Since 2013 he has been a faculty member at the IT group of K. N. Toosi University of Technology, Tehran Iran. Information Technology Engineering Group, Department of Industrial Engineering ,K. N. Toosi University of Technology.

A Fuzzy Approach for Ambiguity Reduction in Text Similarity Estimation (Case Study: Persian Web Contents)

Hamid Ahangarbahan
Department of Information Technology Engineering, Tarbiat Modares, Tehran, Iran
ahangarbahan@gmail.com
Golam Ali Montazer*
Department of Information Technology Engineering, Tarbiat Modares, Tehran, Iran
montazer@modares.ac.ir

Received: 24/Jun/2015 Revised: 24/Nov/2015 Accepted: 05/Dec/2015

Abstract

Finding similar web contents have great efficiency in academic community and software systems. There are many methods and metrics in literature to measure the extent of text similarity among various documents and some its application especially in plagiarism detection systems. However, most of them do not take ambiguity inherent in word or text pair's comparison that gained form linguistic experts as well as structural features into account. As a result, pervious methods did not have enough accuracy to deal vague information. So using structural features and considering ambiguity inherent word improve the identification of similar contents. In this paper, a new method has been proposed that taking lexical and structural features in text similarity measures into consideration. After preprocessing and removing stop words, each text was divided into general words and domain-specific knowledge words. For each part, appropriate features and measures are extracted. Then, the two lexical and structural fuzzy inference systems were designed to assess lexical and structural text similarity respectively. The proposed method has been evaluated on Persian paper abstracts of International Conference on e-Learning and e-Teaching (ICELET) Corpus. The results shows that the proposed method can achieve a rate of 75% in terms of precision and can detect 81% of the similar cases.

Keywords: Text Similarity; Similarity Metric; Fuzzy Sets; Lexical Similarity; Structural Similarity; Persian Text.

1. Introduction

At present, a vast amount of text resources can easily be accessed on the Internet. Although such high frequency of web documents provides us with rapid and immediate access to information, similar and duplicated data results in waste of time and confusion on the part of researchers who would like to detect the originality of a document. Finding the similar contents recently attracts a lot of researchers. Text or content similarity detection can be employed in paraphrasing identification, plagiarism, text summarizing, sentiment analyses, text clustering, text entailment, tracking text news flow on web, etc. For instance, accurate text similarity detection leads to better performance in paraphrasing identification and can improve text clustering [1].

Numerous studies have been conducted to detect similar documents as well as plagiarism [2-3], but most of them have not taken the types of content and domain-specific knowledge as well as style and structural of writing into account. From another perspective, in these studies the methods of text similarity measurement which exert a major influence on the accuracy of evaluation have been used for a certainty and express in crisp way. For instance the word "Process" has different importance in image processing content versus e-learning. As a result, while comparing two texts in specialized or academic domain we will face ambiguity in word or text pair's comparisons since pervious methods don't consider structural features such as author's style of writing. Such limitations in similarity measurement reduce the effectiveness of previous methods in surface and semantic level of text [4-5]. To overcome these limitations, we have proposed a new method that can deal with the ambiguity in the similarity measurement and also consider structural features of text. This method deploys fuzzy linguistic variables to express experts' knowledge about text similarity in surface level of text. Two lexical and structural fuzzy inference systems were designed to accurately figures out the lexical and structural similarity respectively. The output of these two fuzzy inference systems are combined with specific factor and finally the extent of similarity between two contents is determined.

The rest of this paper is organized as follows: the following section provides a brief review of the most related studies. In Section 3text or content similarity problem has been described. Fundamental concepts of fuzzy set theory will be represented in Section 4, the architecture of the proposed method and numerical results has been provided in Section 5. Finally, Section 6 offers the conclusions and implications of the study.

* Corresponding Author

2. Related Work

This section provides the most related research in text similarity methods. Alzahrani, et al., did a comprehensive overview of all the text similarity methods proposed for plagiarism detection. They classified these methods in two broad categories: Literal and Intelligent. With respect to literal methods, individual use simple operations such as copy and paste, which is the most common form of plagiarism and in intelligent methods they use more sophisticated methods such as paraphrasing, obfuscation, changing the structure to hide cheating (redrafting), translation from one language to another and adopting other people's ideas. They also presented taxonomy of techniques that can be utilized to detect text similarity. That taxonomy was categorized into six groups: character-based, vector- based, grammar-based, semantics-based, fuzzy-based, and structure-based. They concluded and proposed semantic- and fuzzy- based methods to be applied due to their better detection and estimation of text similarity and plagiarism detection [8].

Osman, et al., also made another research on text similarity methods. Their proposed taxonomy almost resembled pervious work and consisted of six categories. They consider cluster-based and cross language-based instead of fuzzy-based and vector-based methods. They also proposed semantic role labeling for sentence to detect similarity. The main drawback of their methods is related to numerous computations needed to make [9].

Alzahrani and Salimi used fuzzy membership function to calculate the degree of similarity between two words. They obtained an accuracy of 54.24% on PAN-PC-09 corpus. Their algorithm suffers from time complexity and also needs to improve membership function [10]. In another study, Gupta, et al., drawing on the model offered by Alzahrani and Salimi, redefined fuzzy membership function in smaller range intervals. They also used different preprocessing on PAN-PC-2012 dataset and applied it to fuzzy algorithm. The authors concluded that preprocessing on POS level and its integration with fuzzy-based methods would contribute to effective recognition similar documents [11].

Based on their study, El-Alfy, et al., proposed a framework that employs abductive neural networks. They used five simple and weak metrics and by using abductive neural networks selected the best adopted metrics and boosted them. They applied algorithm on PAN 2010 but as it was the case with aforementioned research, their proposed framework suffers from time and memory complexity. Their work also does not take semantic meaning of words into consideration [12]. In another research, El-Alfy deployed lexical similarity metrics and proposed a hybrid method that used three different types of machine learning techniques such as Bayesian learning, support vectors machine and artificial neural networks. The author applied algorithms to MSRPC and demonstrated that artificial neural networks method performed better than Bayesian learning, but it takes a long time to train it [13].

Barrón-Cedeño, et al., adopted an n-gram approach to recognizing suspicious documents. They tested different n size to generate n-gram on METER corpus and finally proposed 2, 3 for n at word level [14]. Kumar and Tripathi using continuous 3-gram that selects the longest substring in a document to detect plagiarism [15]. Zesch, et al., demonstrated that context-based measures cannot detect all forms of text similarity and proposed to apply style-based and grammar-based measures. They came to the conclusion that for more accurate detection, all types of similarity measures should be exploited together. They also stated that text features should be properly selected so that the similarity measure works to its optimum [16]. Both Brockett, et al.[17] and Rus, et al. [18] used combined lexical, semantic and grammatical features in support vector machine to detect paraphrasing.

As it is clear from the review of the literature, few researches have focused on the ambiguity of the word pair's comparison and similarity of structural and style of writing in content similarity measurement; meanwhile, there is a need to use lexical, structural and stylistic features to obtain a comprehensive evaluation of text similarity. To do this, we need to use mathematical theories that have capability to deal vague data. The fuzzy set theory can represent expert knowledge and can deal with ambiguity in real problem. We use this theory to design our new method. In the next section, fundamental concepts of fuzzy sets theory will be discussed.

3. Problem Statement

Text similarity detection can be stated as finding two similar parts of two different documents. Since users change the word order of sentences, style of writing, transpose different sections of a text or rewrite a text by changing a word to its equivalent semantic meaning (substituting hyponyms, synonyms or paraphrasing), detecting these types of text similarity is too difficult, especially when we are concerned with specific content domain texts [7] and in low resource language such as Persian. However in specific content domain texts usually it is difficult to change the style of origin author. This will help us significantly to detect similar document. So we need to clarify the problem statement and identify factors that affect similarities measurement to obtain a better performance. In this paper, we consider the following question:

Given two Persian contents in specific (scientific) content domain, how can we assess the degree of similarity between two texts with high accuracy and precision?

The proposed method uses two kinds of features to solve this problem. It uses lexical features to assessing surface similarity of text pairs and applies structural features to estimating similarity in author writing style and usage of hyponyms words.

4. Fundamental Concepts of Fuzzy Set Theory

Fuzzy sets theory was introduced by Lotfi Zadeh in 1965[19]. This theory is an appropriate framework for handling uncertain and imprecise data. This framework uses a set of "if-then" rules assigned to inference, in which any of these rules are defined by fuzzy sets. Fuzzy logic uses linguistic variables, which can be easily understood by humans and allows decision-making in spite of incomplete and uncertain information.

Fuzzy inference systems can provide appropriate and practical solutions to complex systems engineering in different situations. These systems consist of four main components that seen in figure 2. These components are described as follows [20]:

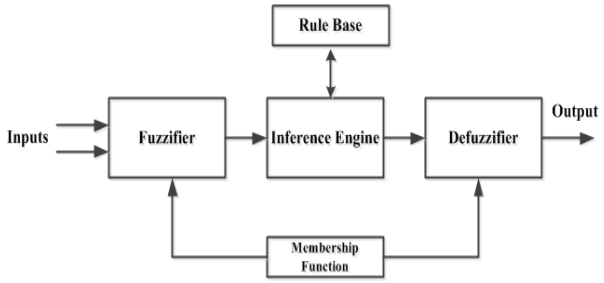


Fig. 1. Basic configuration of fuzzy inference system [20]

In the fuzzifier process, relationships between the inputs of system and linguistic variables are defined by fuzzy membership functions. In this research, each input variables model as trapezoidal fuzzy number that donated as [a,b,c,d] and fuzzy membership are being defined as follows:

$$\mu_A(x)=\begin{cases}\frac{x-a}{b-a} & a\le x\le b\\ 1 & b\le x\le c\\ \frac{x-d}{c-d} & c\le x\le d\\ 0 & o.w.\end{cases} \quad (1)$$

In knowledge base, all linguistic rules are extracted from the domain experts, that is, experts on linguistics. These rules use linguistic variables to express relationships between inputs and outputs of the system. The format of the rule is represented as follows:

If *"the input conditions are true" then "the set of outputs is inference"*.

Inference system is related to the decision part of the system and is able to infer outputs using fuzzy rules and operators. In this research, Mamdani product inference system through the following process generates outputs by using inputs based on predefined rules [21]:

$$\mu_{A^k\to B^k}(x,y)=\mu_{A^k}(x).\mu_{B^k}(y) \quad (2)$$

In which $\mu_{A^k}(x)$ signifies the membership function value of k^{th} rule in the knowledge base.

The defuzzification step performs the reverse of fuzzifier process and generates a crisp value of fuzzy output. There are many techniques to defuzzification. In this research, the center of gravity is exploited in defuzziness process as follows [21]:

$$y'=\frac{\int y_i\mu_B(y)\,dy}{\int \mu_B(y)\,dy} \quad (3)$$

5. The Proposed Method

In this study, a mixed fuzzy inference system (FIS) was proposed which accomplishes the inference through two FISs, assesses similarity between two sentences and finally detects text or document similarity. The advantages of a combination of lexical and structural features are obvious [16]. As in [16] mentioned; experiments show the drawbacks pertaining to use features alone seem complementary and therefore it is good idea to take composing a mixed system combining these two types features. Such combination don't optimize the FIS but it help to improve the accuracy and performance of overall system. Figure 2 demonstrates the conceptual structure of the proposed method. The proposed method is composed of four components: preprocessing, segmentation, features extraction and similarity measures selection and finally fuzzy inference system. These components are described in the rest of the paper.

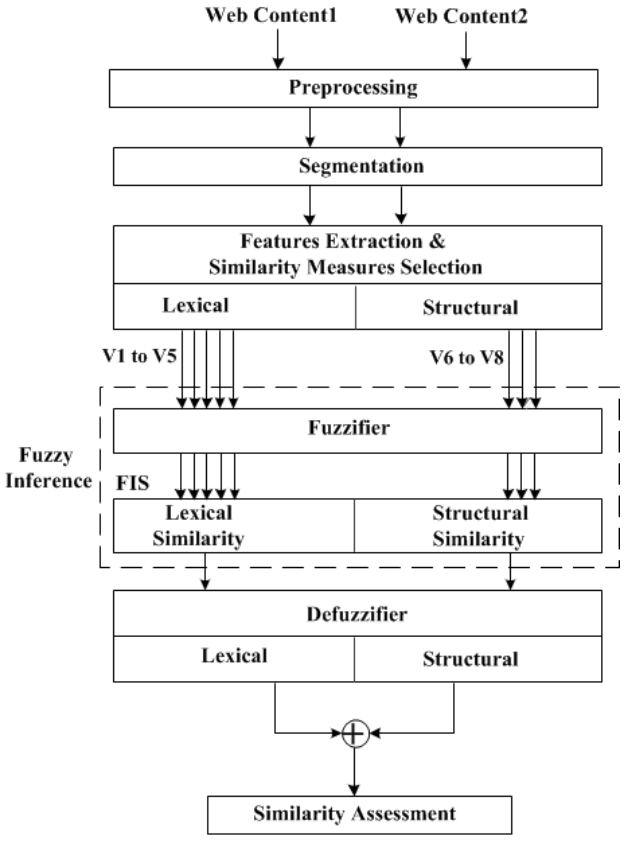


Fig. 2. Proposed mixed FIS Method

5.1 Preprocessing and Segmentation

In these two components, preprocessing operations such as stemming and stop-word removal is done in each input text. Text contains one or more words that express a special meaning in the context in which it is stated. This

meaning is usually expressed through domain-specific knowledge (DSK) terms. Therefore, we need to disambiguate the sense of a word according to its context. To gain more precision and accuracy, in the next step each text is divided into general and domain-specific knowledge parts. In the general part, words such as 'book' and in the domain-specific knowledge part words such as 'E-Learning' are included.

For instance, consider the following two sentences:

1. *"I am working in web programming and semantic web" and*
2. *"Alex has experience in business intelligent and software applications".*

Figure 3 depicts the segmentation of these sentences after preprocessing. Domain-specific knowledge terms and words are extracted using the ontology of IT technical terms. In this research, due to the lack of such domain-specific knowledge ontology in Persian, we had to compile it.

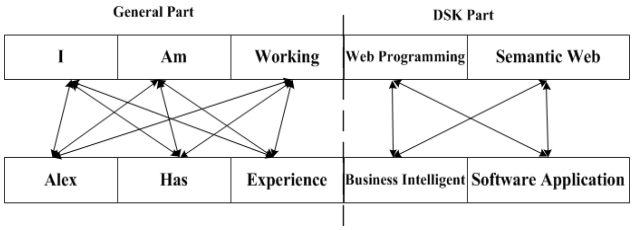


Fig. 3. Segmentation and comparing in sentence pairs

5.2 Measures and Features Extraction

After segmentation, the relevant sections are compared in a way that the general part of the first sentence is being compared to the general part of the second sentence; the same process also occurs for domain-specific knowledge parts. We used lexical- and structural- based features and metrics to assess the extent of similarity between two the texts. For this propose we classified main metrics and measures that mentioned in literature in lexical and structural category. These measures have been experimented on part of PeLeT corpus (20%). To select most appropriate measures, we define *Pr_Ti* criteria as follow:

$$\Pr\_\mathrm{Ti} = \alpha.\,\mathrm{Precision} + (1-\alpha).\,\mathrm{Time\ complexity} \quad (4)$$

Where *Precision* shows the measure accuracy in text similarity detection and *Time complexity* is its time complexity. *α* should be determine based on the problem. In the context of this paper our main goal is accuracy improvement in text similarity detection and therefore we set$\alpha = 1$. Based on the results, *overlapping ratio* and *skip gram* have been selected from lexical category and *stopword overlapping* and *the number of hyponyms words* have been selected from structural category.

The selected features and metrics have been categorized in two groups and are introduced in the following:

5.2.1 Lexical Approach Similarity

The similarity measures and features in this approach only address the surface level of words in contents and do not take the meaning and senses of them into account.

A. Overlapping Ratio.

This measure calculates the common words between two texts by using n-grams in word or character level, divided by the length of the first and second text respectively. Then, it computes the geometric mean of the two ratios. This ratio for first text computes as equation (5) [22].

$$S(T_1,T_2)=\frac{|T_1\cap T_2|}{|T_1|} \quad (5)$$

Which $|T_1 \cap T_2|$ is a number of common words between two texts and $|T_1|$ is a number of first text words.

This measure is to be handled in general and DSK parts separately and for simplicity denoted by V1, V3 respectively in this paper.

B. Skip-Gram Measure

This measure is similar to the n-gram but can skip between words as 'skip' length. According to the examination on Persian IT corpus, the best skip number was found to be 3. This measure deals with word order as well as detects the common phrases in two sentences [23].This measure also uses general and DSK parts separately and is denoted by V2, V4 respectively.

C. The Ratio of Number of DSK to General Words in Union of Two Texts

This ratio indicates the degree of importance of DSK words to general words in union of two texts and is denoted by V5.

5.2.2 Structural Approach to Similarity

In this approach that can be considered the same as the lexical approach, the structural features of each text are considered as follows:

A. The Number of Hyponyms Words

This feature has been selected for the reason that in plagiarized texts the hyponyms words such as "look and view" are often used interchangeably for secrecy. This feature is handled in general and DSK parts separately and denoted by V6, V7 respectively.

B. The Overlapping of Stopword

This feature has been chosen because if two texts are structurally similar, the same structure of word stop is being used specially in scientific texts. This feature can be considered as author style that is denoted by V8[16].

Table 1 presents the complete list of measures and features used in fuzzy systems. The linguistic variables of the proposed FIS were developed on the basis of these measures and features.

Table 1. Selected Features and Measures

		Union of Two Texts	Term Part	
			General	Domain Specific Knowledge
Similarity Approach	Lexical	The ratio of DSK words to general words	Overlapping Ratio Skip Gram	Overlapping Ratio Skip Gram
	Structural	The Overlapping of Stopword	The Number of Hyponyms Words	The Number of Hyponyms Words

5.3 The Surface level Text Similarity Fuzzy Inference System

Due to the complex nature of natural language, it is quite difficult to detect similarity between scientific and domain-specific knowledge texts. Meanwhile, the ambiguity inherent in human language prohibits us from developing efficient NLP techniques. By the same token, the same content might be worded differently in various paraphrases. That is why we need to gather information from experts to achieve a more precise assessment. To do this, we designed two lexical and structural FISs that assess text similarity from a different perspective. This is our first contribution. The output of each FIS will be combined and finally the proposed mixed method can determine whether two sentences like each other base on weighted average of two lexical and structural FISs result. We model similarity assessment as fuzzy linguistic variables to overcome the ambiguity and vagueness of the assessment. In Table 2 similarity of linguistic variables and their membership function will be expressed.

Table 2. The Similarity of Linguistic Variable and Membership Function

Linguistic Variable	Interval
Different	(-∞,0,0.1,0.35)
Semi Similar	(0.1,0.35,0.55,0.75)
Similar	(0.55,0.75, 1,+∞)

Since assessing the similarities between the two texts is defined as a fuzzy number, inputs of FIS are also modeled as a fuzzy number. For easy and quick access, Table 3 displays all types of variables and Table 4 depicts variables used in each FIS with their correspondence.

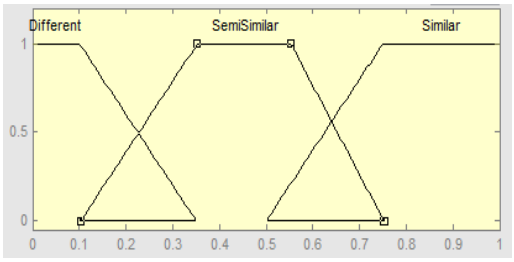
Table 3. Lingustic Variables in Simialrity Assesment

TYPE	LINGUISTIC VARIABLE	INTERVAL
T1 (Stopword overlaping)	Low	(-∞,0,0.1,0.35)
	Middle	(0.1,0.35,0.55,0.85)
	High	(0.6,0.75, 1,+∞)
T2 (Hyponymword overlaping)	Low	(-∞,0,0.1,0.3)
	Middle	(0.1,0.3,0.5,0.7)
	High	(0.6,0.7, 1,+∞)
T3 (The ratio of DSK words to general words)	Low	(-∞,0,0.1,0.25)
	Middle	(0.15,0.35,0.55,0.8)
	High	(0. 5,0.75, 1,+∞)
T4 (Similarity Assessment)	Different	(-∞,0,0.1,0.35)
	Semi Similar	(0.1,0.35,0.55,0.75)
	Similar	(0.55,0.75, 1,+∞)

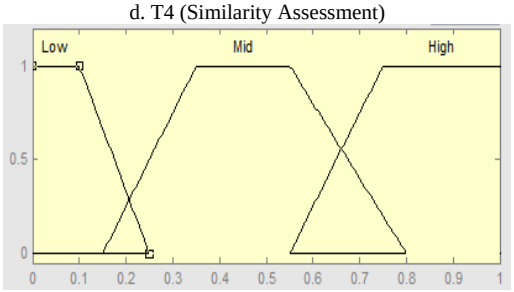
Table 4. FISs Variables

FIS	PART	VARIABLE NAME	VARIABLE TYPE
Lexical Similarity	General	V1	T4
		V2	T4
	DSK	V3	T4
		V4	T4
	Union of two Texts	V5	T3
Structural Similarity	General	V6	T2
	DSK	V7	T2
	Union of two Texts	V8	T1

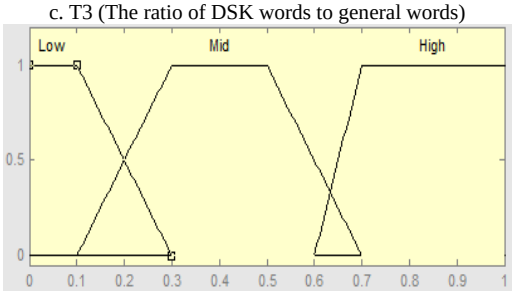
Because of wide variety of ambiguity in word or text pairs comparisons; all fuzzy variables model as trapezoidal membership function. Figure 4 also displays the membership function of variable types in FISs.



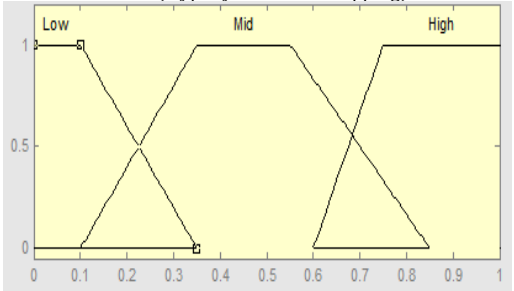
d. T4 (Similarity Assessment)



c. T3 (The ratio of DSK words to general words)



b. T1 (Hyponymword overlapping)



a. T1 (Stopword overlapping)

Fig. 4. The membership functions of FISs variables

We implemented mixed similarity fuzzy inference system using MATLAB 2010 fuzzy tool and used Mamdani type of fuzzy system with following configuration that set by try and error:

- AND method: prod
- OR method: max
- Defuzzification method:LOM (Large Of Maximum)

5.4 Fuzzy Function of Part Pairs Similarity

Since each part is comprised of several words, we need to compare word pairs and then aggregate their results as it was done for the example in Figure 2. The only difference in this section is that the comparisons will be made in a fuzzy manner. As a result, for example, the output matrix resembles equation 4. Each element of this matrix is a fuzzy number, in which we use maximum as S norm in each row and column to compute the final result of two texts; then, the values of the average of these two fuzzy numbers are obtained by Rosenfeld relationship [21].

$$\mathrm{Sim}_{\mathrm{DSK}}(\mathrm{T}_1, \mathrm{T}_2) = \begin{bmatrix} \{(\mathrm{Similar}, 0.53), (\mathrm{Semi-similar}, 0.12), (\mathrm{Dissimilar}, 0.16)\} & \{(\mathrm{Similar}, 0.73), (\mathrm{Semi-similar}, 0.4), (\mathrm{Dissimilar}, 0.08)\} \\ \{(\mathrm{Similar}, 0.81), (\mathrm{Semi-similar}, 0.32), (\mathrm{Dissimilar}, 0.11)\} & \{(\mathrm{Similar}, 0.12), (\mathrm{Semi-similar}, 0.22), (\mathrm{Dissimilar}, 0.67)\} \end{bmatrix} \quad (6)$$

5.5 Fuzzy Rules Base

After precisely defining FIS variables, system rules were deduced and developed by conducting an interview with a group of five natural language and domain experts. Table 5 and 6 demonstrates some fuzzy rules for lexical and semantic FIS respectively. In this rule base as mentioned, we consider fuzzy "AND" for t-norm and operators among fuzzy variables and the operator for rules aggregation is s-norm.

For example, rule 5 in Table 5 is as follow:

If *"Overlapping Ratio [V2] in General Part is Low"*, AND *"Overlapping Ratio [V4] in Knowledge domain Part is Low"*, THEN *"[Lexical] Similarity is Different"*.

Table 5. Some rules of FISs

FIS	RULE #	VARIABLE NAME					SIMILARITY
		V1	**V2**	**V3**	**V4**	**V5**	
Lexical similarity	1	-	High	-	Low	Low	Similar
	2	-	High	Low	Low	Middle	Semi Similar
	3	-	High	Middle	Low	Middle	Semi Similar
	4	Low	Middle	Low	Low	Middle	Different
	5	-	Low	-	Low	-	Different
Structural similarity		**V6**	**V7**	**V8**			
	1	Low	Low	Low			Different
	2	Middle	Low	Low			Different
	3	-	High	-			Similar
	4	Middle	Low	Middle			Semi Similar
	5	High	Low	Middle			Semi Similar

6. Numerical Results

In this section, the numerical results of the proposed method will be reported. The proposed method was implemented in Visual Studio Environment and used MATLAB fuzzy toolbox to simulate FISs. Because of there is no technical English corpus, we applied the method on Persian corpus to evaluate its efficacy. To create the Persian corpus, that named PeLeT, we used the papers' abstracts presented in ICELET conferences in e-learning domain knowledge. This corpus contains 810 sentence pairs totally that each *"Different"*, *"Semi Similar"* and *"Similar"* class has 270 sentence pairs. The first sentence of pairs gathered from papers' abstracts presented in ICELET and a group of expert domain generated a paired sentence in one of the randomly assigned class. The average length of sentence is 13 words. After preprocessing and segmentation, the proposed method applied to the corpus. Table 6 indicates some examples of text pairs in PeLeT corpus.

Table 6. Some examples of text pairs in PeLeT corpus

FIRST TEXT	PAIRED TEXT	CLASS
E-learning is a new type of learning using information technology tools.	The expansion of changes in information technology led to a new type of learning called e-learning.	Similar
The most important goal of e-learning is to transfer the focus of learning from teacher to learner.	One of the most important goals of e-learning is to create a learner-centered atmosphere.	Semi-similar
The intelligent educational system adapts the educational strategy to the learner characteristics.	The main part of the intelligent educational system is the learner-based model that includes information which the system holds about the learner. The educational strategy is adapted based on the information obtained from this section.	Different

In final step, two FISs outputs fused and this fusion is our other contribution. For lexical and structural similarity detection FISs, weights 0.65 and 0.35 have been considered respectively. Table 7 shows results of two FISs fusion for two sample text in corpus.

Table 7. Fusion of lexical and structural similarity detection FISs

FIS	SIMILARITY ASSESSMENT			WEIGHTS
	Similar	Semi-Similar	Different	
LEXICAL	0.73	0.16	0.09	0.65
STRUCTURAL	0.57	0.25	0.12	0.35
FUSION	0.674	0.192	0.205	

Table 8 indicates the confusion matrix of the proposed method. The first row of table indicates that our proposed method correctly detect 185 of 270 sentence pairs are similar. This method also failed for rest of sentence pair and placed 20 and 56 of sentence pairs in semi-similar and different class respectively. The other rows show the same results.

Table 8. Confusion matrix of proposed method

		Predicted Class		
		Similar	Semi Similar	Different
Actual Class	Similar	185	20	56
	Semi Similar	28	173	78
	Different	41	53	176

To evaluate the performance of the proposed method, we used recall, precision and F-measure that are widely employed in text mining. It should be noted the use of other languages datasets such as WordNet will not show the method efficiency.so we also intended to cast more light on the comparison of the efficiency of the proposed algorithm and given the fact that the proposed method has been applied to PeLeT Corpus. We experimented three methods mentioned in the literature with PeLeT Corpus to assess the efficiency of our method. Table 9 illustrates the results of proposed method implementation and its comparison with three methods. The Cosine Coefficient was considered as baseline. The results show that proposed method outperforms the other methods.

Table 9. The result performance

Method	F Measure	Percision	Recall
The Proposed Method	0.78	0.75	0.81
Gupta, et al.[11]	0.75	0.72	0.77
Alzahrani and Salimi[10]	0.65	0.63	0.67
Mihalcea and Corley[26]	0.62	0.60	0.65
Cosine Coefficient (Baseline)	0.55	0.54	0.57

7. Conclusions

In this paper, a mixed fuzzy inference system method was proposed to overcome the ambiguity and consider structural features and author style of writing in the similarity measurement in Persian texts. In this method, in the first step, after preprocessing and stop word removal, the text is divided into general and domain-specific knowledge parts then appropriate features are extracted and similarity metrics are calculated. Two lexical and structural fuzzy inference systems were designed that its rules are extracted from the experts' knowledge and finally the outputs of these two FISs are integrated through weighted combination. With regard to the fact that the proposed method was applied to PeLeT Corpus, we also carried out tests using three methods proposed in the literature to evaluate its efficiency. Such a comparison was thought to throw light on the efficiency of the proposed algorithm. The results show that the proposed method outperforms than others and gained accuracy rate of 78% which increases precision and recall measure.

References

[1] Wang, Yong, and Julia Hodges. "Document clustering with semantic analysis." System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on. Vol. 3. IEEE, 2006.

[2] Mohler. M., , and Mihalcea Rada. "Text-to-text semantic similarity for automatic short answer grading." Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2009.

[3] Hariharan. S., , et al. "Detecting plagiarism in text documents." Information Processing and Management. Springer Berlin Heidelberg, 2010. 497-500.

[4] Barrón-Cedeño, Alberto, and Paolo Rosso. "On automatic plagiarism detection based on n-grams comparison." In Advances in Information Retrieval, pp. 696-700. Springer Berlin Heidelberg, 2009.

[5] Kent, Chow Kok, and Naomie Salim. "Features based text similarity detection." arXiv preprint arXiv:1001.3487 (2010).

[6] Joy, Mike, and Michael Luck. "Plagiarism in programming assignments." Education, IEEE Transactions on 42.2 (1999): 129-133.

[7] Potthast, Martin, et al. "Cross-language plagiarism detection." Language Resources and Evaluation 45.1 (2011): 45-62.

[8] Alzahrani, Salha M., Naomie Salim, and Ajith Abraham. "Understanding plagiarism linguistic patterns, textual features, and detection methods." Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 42.2 (2012): 133-149.

[9] Osman, Ahmed Hamza, Naomie Salim, Mohammed Salem Binwahlan, Rihab Alteeb, and Albaraa Abuobieda. "An improved plagiarism detection scheme based on semantic role labeling." Applied Soft Computing 12, no. 5 (2012): 1493-1502.

[10] Alzahrani, Salha, and Naomie Salim. "Fuzzy semantic-based string similarity for extrinsic plagiarism detection." Braschler and Harman (2010).

[11] Gupta, Rohit, et al. "UoW: NLP techniques developed at the University of Wolverhampton for Semantic Similarity and Textual Entailment." SemEval 2014 (2014): 785.

[12] El-Alfy, El-Sayed M., et al. "Boosting paraphrase detection through textual similarity metrics with abductive networks." *Applied Soft Computing* 26 (2015): 444-453

[13] El-Alfy, El-Sayed M. "Statistical Analysis of ML-Based Paraphrase Detectors with Lexical Similarity Metrics." In *Information Science and Applications (ICISA), 2014 International Conference on*, pp. 1-5. IEEE, 2014.

[14] Barrón-Cedeño, Alberto, and Paolo Rosso. "On automatic plagiarism detection based on n-grams comparison." In Advances in Information Retrieval, pp. 696-700. Springer Berlin Heidelberg, 2009.
[15] Kumar, Ranjeet, and R. C. Tripathi. "A Trigram Word Selection Methodology to Detect Textual Similarity with Comparative Analysis of Similar Techniques." In *Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on*, pp. 383-387. IEEE, 2014.
[16] Zesch, Daniel Bär1 Torsten, and Iryna Gurevych. "Text reuse detection using a composition of text similarity measures." In *Proceedings of COLING*, vol. 1, pp. 167-184. 2012.
[17] Brockett, Chris, and William B. Dolan. "Support vector machines for paraphrase identification and corpus construction." In *Proceedings of the 3rd International Workshop on Paraphrasing*, pp. 1-8. 2005.
[18] Rus, Vasile, Philip M. McCarthy, Mihai C. Lintean, Danielle S. McNamara, and Arthur C. Graesser. "Paraphrase Identification with Lexico-Syntactic Graph Subsumption." In *FLAIRS conference*, pp. 201-206. 2008.
[19] Zadeh, Lotfi A. The concept of a linguistic variable and its application to approximate reasoning. Springer US, 1974.
[20] Huang, Yo-Ping, et al. "An intelligent approach to detecting the bad credit card accounts." Proceedings of the 25th conference on Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications. ACTA Press, 2007.
[21] Rutkowski, Leszek, and Krzysztof Cpalka. "Flexible neuro-fuzzy systems." *Neural Networks, IEEE Transactions on* 14.3 (2003): 554-574.
[22] Metzler, Donald, et al. "Similarity measures for tracking information flow." Proceedings of the 14th ACM international conference on Information and knowledge management. ACM, 2005.
[23] Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013).
[24] Wu, Zhibiao, and Martha Palmer. "Verbs semantics and lexical selection." Proceedings of the 32nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 1994.
[25] Resnik, Philip. "Using information content to evaluate semantic similarity in a taxonomy." *arXiv preprint cmp-lg/9511007* (1995).
[26] Mihalcea, Rada, Courtney Corley, and Carlo Strapparava. "Corpus-based and knowledge-based measures of text semantic similarity." *AAAI*. Vol. 6. 2006.

Hamid Ahangarbahan is currently a Ph.D. candidate in Information Technology Engineering in Tarbiat Modares University, Tehran, Iran. His research interests are in Soft computing, computational intelligence, Information Engineering and Data Mining Knowledge Discovery, Intelligent Methods and System Modeling.

Gholam Ali Montazer received his B.Sc. degree in Electrical Engineering from Kh. N. Toosi University of Technology, Tehran, Iran, in 1991, his M.Sc. degree in Electrical Engineering from Tarbiat Modares University, Tehran, Iran, in 1994, and his Ph.D. degree in Electrical Engineering from the same university, in 1998. He is a Professor of the Department of Information Technology Engineering in Tarbiat Modares University, Tehran, Iran. His areas of research include Information Engineering, Knowledge Discovery, Intelligent Methods, System Modeling, E-Learning and Image Mining.

Scalable Community Detection through Content and Link Analysis in Social Networks

Zahra Arefian*
Department of Computer Engineering, University of Isfahan, Isfahan, Iran
zahra_arefian@yahoo.com
Mohammad Reza Khayyam Bashi
Department of Computer Engineering, University of Isfahan, Isfahan, Iran
m.r.khayyambashi@eng.ui.ac.ir

Received: 16/May/2015 Revised: 29/Sep/2015 Accepted: 28/Oct/2015

Abstract

Social network analysis is an important problem that has been attracting a great deal of attention in recent years. Such networks provide users many different applications and features; as a result, they have been mentioned as the most important event of recent decades. Using features that are available in the social networks, first discovering a complete and comprehensive communication should be done. Many methods have been proposed to explore the community, which are community detections through link analysis and nodes content. Most of the research exploring the social communication network only focuses on the one method, while attention to only one of the methods would be a confusion and incomplete exploration. Community detections is generally associated with graph clustering, most clustering methods rely on analyzing links, and no attention to regarding the content that improves the clustering quality. In this paper, a novel algorithm for community selection is proposed. Scalable community detections, an integral algorithm is proposed to cluster graphs according to link structure and nodes content, and it aims finding clusters in the groups with similar features. To implement the Integral Algorithm, first a graph is weighted by the algorithm according to the node content, and then network graph is analyzed using Markov Clustering Algorithm, in other word, strong relationships are distinguished from weak ones. Markov Clustering Algorithm is proposed as a Multi-Level one to be scalable. Finally, we validate this approach through a variety of data sets, and the effectiveness of the proposed method is evaluated.

Keywords: Social Networks; Community Detections; Link Analysis; Clustering; Scalable.

1. Introduction

In recent years, social networks have not only been being used for creating relationships, but they are also used to share opinions, communicate, fans, activists and interact over diverse geographical regions [1]. Due to the multiple modes of communication, these networks share information and do a variety of interactions. These relationships will lead to the creation of groups like friends, colleagues, acquaintances, family and other similar groups, and that's why social networks have been popular. According to a report in 2012, internet users spent 22 percent of their online time surfing social networks.

Among the popular social networks, we can mention Facebook[1], YouTube[2], Flickr[3], twitter[4], etc [2]. Social networks are a social structure; a social network is a network of interactions and relationships that are a graph and set of nodes and edges (nodes consisting of individuals or organizations). These nodes have interactions and according to the social relations that exist in the real world, these relations can be obtained and we can analyze them (links represent the connections between users) [3]. Due to the amount information of data in these networks, the analysis of network data has become an important issue for research. Now, according to the large volume of information, we should discover the unknown relations, and the discovery can be exploited to improve opportunities. Despite the increasing significance and complexity of Social network, there has expanded of methods for detecting communities. The discovery of communication by link analysis and regarding the content of nodes is an important issue.

The importance of addressing the link analysis and the nodes content for community detection is illustrated in Fig. 1 [1]. In Fig. 1(a) presents a very small social network. The nodes indicate the number of involved members in the social activities and the edges represent the social relations and interactions among members. The weight wrote to each edge illustrates the strength of connections between the corresponding members and also each node is labeled according to its interests. Fig. 1(b) presents the result of discovered communities based on link analysis, that the discovery relates to the link analysis, they only pay attention to the network topological structure or analysis from respect data mining [4]. Fig. 1(c) presents the result of discovered communities based on nodes content, they only pay attention to the "Similarity theory" for categorizing individuals with different

[1] www.facebook.com
[2] www.youtube.com
[3] www.flickr.com
[4] www.twitter.com

* Corresponding Author

communication; but this type of grouping doesn't have enough accuracy and can not create a strong social relations. Fig. 1(d) presents the desired grouping, the groups are defined to having the same interest topics and strong relation between each cluster.

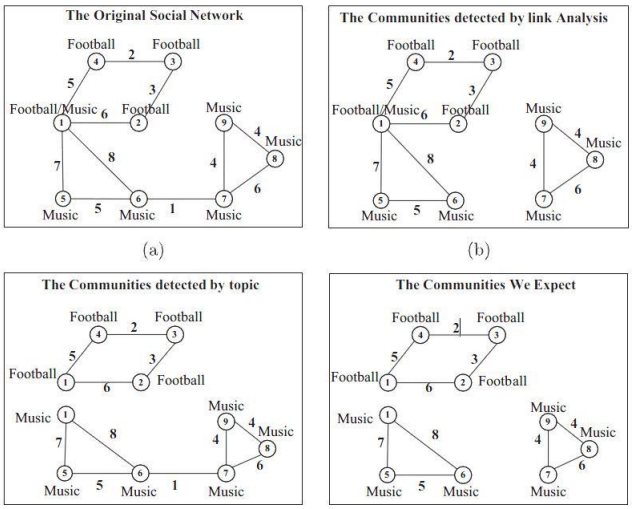


Fig. 1. An example that in part(a) presents a very small social network. Part(b) presents the result of discovered communities based on link analysis. Part(c) presents the result of discovered communities based on nodes content. Part(d) presents community detection based on node content and link analysis[1].

large social networks challenge the issue of discovering the communications, and the old methods of discovering such communications have a lot of problems.

By modeling complex social network with graphs, the community detection can be modeled by graph partitioning, and there is a clustered set of nodes which are connected by the edges. Although the number of different clustering algorithms exploring the relationship exists, but it is not easy to present a good algorithm for the above cases, and it requires careful consideration.

In this paper, an algorithm for clustering the graph topology structure is presented according to the vertices features, and a graph is formed according to the features and content which are among users, then the scalable communications will be discovered using Multi-Level Markov clustering (ML-MCL) algorithm.

The article is organized as follows; first, in part 2, reviews the related work, then will be propound the subject and the proposed approach in the part 3, finally, Performance assessment methodology and conclusions will have been done.

2. Related Work

Social networks has been a very important matter in recent decades, so a lot of fundamental and important research has been done in all fields and topics; that is because these networks are posing global communications. In Ref [5] One of the most important topic that researchers has been working on it is exploring the community in the social networks, that some of them are mentioned in the following expression.

In Ref [6] Markov Clustering (MCL) Algorithm groups nodes randomly, and clusters graphs via transition probability matrix corresponding to the graph. The MCL algorithm is an iterative process of applying two operators (expansion and inflation) in alternation, until convergence. Additionally, a prune step is performed at the end of each inflation step in order to save memory. One of the important algorithm used for community detection is KL[1] algorithm that is graph partitioning algorithm and is run in classic way and do optimization operations. In Ref [7] Another group of Algorithms for community detection are Agglomerative/Divisive Algorithms. Agglomerative algorithms at first begin with each node in its own community, and at each step communities merge each other, continuing till either the desired number of communities is obtained or the remaining communities don't have enough similarity for merging. Divisive algorithms operate in reverse. Both types of algorithms are hierarchical clustering algorithms and their output is a type of binary tree. The other way which is mainly used to for community detections is the Local Graph Clustering which is used to reduce the scalability challenges by focusing on the studying section of the network. To discover, it is started from a peak as a seed and then by adding the neighbor peaks to the community, it is resulted to increase the network and obtain a high-quality proper size in [8-10]. Another groups of Algorithms for community detection are Spectral algorithms. Generally, assign nodes to communities based on the eigenvectors of matrices, such as the adjacency matrix or other related matrices. Spectral methods aim to minimize the defined cut-function that lead to more resolution in graph clustering structure in [11-12]. Multi-level algorithms are of the other algorithms which are used to discover the communications. Multi-level methods present a framework for high-quality, fast partitioning of a graph and are used to solve many problems. The main idea is to minimize the input graph continuously and reach a smaller graph. The resulted graph is partitioned and then returned to the top to reach the main graph. Some methods to partition multi-level graphs are multi-level spectral clustering, Metis (improved KL function) and Graclus (improved normal cut and weight loss) [13-15]. In [16], at first, they develop the original similarity based on the social balance theory. Then, based on the natural contradiction between positive and negative links and the signed similarity, two functions are designed to model a multi objective problem, called MEAs -SN. In [17], Based on the Max-Flow Min-Cut theorem, they propone a novel algorithm which can output an optimal set of local communities automatically.

[1] Kernighan-Lin Algorithms

3. Community Detection Mechanism

In this section, the proposed mechanism of Community Detections is presented; the work done in this section is as follows:

First, graph topology structure is combined with node features, then edges are weighted according to vertices content, and links are analyzed by MCL algorithm according to the weighted graph; On the other hand, MCL clustering algorithm is proposed as a multi-level one to be scalable.

3.1 Social Network Data Modeling Based on Similarities

Social networks are shown in graph G= (V,E,χ), that V={v_1, v_2, ..., v_n} is the set of nodes and |V| = n illustrate the number of persons in graph. Also E ⊆ V ×V is the set of edges, where E = {(v_i, v_j): v_i, v_j ∈ V} and shows collection of interactions and communications among individuals. In this graph, matrix χ is attributes of vertices and $\chi \in R^{|V|\times d}$, where d indicates number of node attributes [4].

Similarity function C determines the similarity between each pair of vertices in an attributed graph G. all of the characteristics are binary, so we use Jaccard's coefficient as similarity criteria for attributes data that is eq. 1:

$$J(V_i, V_j) = \frac{\text{Number of common one's between i and j}}{\text{Number of attributes}} \quad (1)$$

Then based on vertices content matrix S is constituted, so if content of vertices v_i and v_j are similar, according to the number of topics that are interacted to each other S_{ij} from matrix S are calculated power of link and if they do not have any interaction with each other is placed 0. Finally weight matrix W is collection of matrix S and matrix A, that is eq. 2:

$$W=A+S \quad (2)$$

Until this step, social network graph is weighted based on vertices content. In the following we describe the proposed clustering Algorithm for grouping social topics.

3.2 The Clustering Algorithm

In this section, clustering algorithm to Community Detections in social networks through vertices content and link analysis has been proposed; following steps is required in this algorithm:

Pseudo-code of the integral clustering algorithm to discover communications in social network graphs is presented in Fig. 2. In this pseudo-code, the social network graph is used as the algorithm input and after a few steps; a clustered graph is returned as an output.

1. Input: G
2. Output: clustering

3. A ← Adj(G)
4. Compute the attributes similarity matrix, C
5. Compute matrix S
6. W ← A + S
7. clusters ← Apply Multy-Level Markov Clustering on W
8. Return clusters

Fig. 2. Clustering Algorithm based on MCL[4]

First, a similarity matrix C is developed, then matrix S is formed in the fifth line of the algorithm, next matrix W is formed by adding the matrix S and matrix A. Finally, in the seventh line of the algorithm, the formed weighted graph using the multi-level clustering algorithm of Markov (ML-MCL) is clustered. In the following, the pseudo-code related to each one is presented.

Fig.3 shows the pseudo-code of MCL algorithm, and Fig. 5 shows ML-MCL algorithm which is obtained by some changes in original pseudo-code of MCL algorithm.

1. A := A + I // Add self-loops to the graph
2. M := AD−1 // Initialize M as the canonical transition matrix
3. repeat
   a. M := Mexp := Expand(M)
   b. M := Minf := Inflate(M, r)
   c. M := Prune(M)
4. until M converges
5. Interpret M as a clustering

Fig. 3. MCL Algorithm[18]

MCL algorithm is a clustering algorithm based on graph stochastic flows simulation. The reasons to choose this algorithm for some of the clustering steps are that this algorithm has no need to specify the number of clusters at the beginning, and has resistance to noise in the number of components, also its efficacy for weighted and non-weighted graphs and oriented and non-oriented ones. First, to conclude more quickly and prevent some unexpected cases, the first line of the algorithm is done for convergence. Then, to implement it, the transition matrix should be formed from the weighted graph W [18]. To calculate the components of the transition matrix, eq. 3 is implemented:

$$M_{ij} = \frac{A_{ij}}{\sum_{i=0}^{n}(A_{ij})} \quad (3)$$

The resulted transition matrix is a type of the column-stochastic transition matrix, a matrix that the sum of each of its columns equals to 1. Such matrices can be defined as transition matrix of Markov chain in which the i^{th} column of matrix M represents the possibility of transferring the output V_i. Therefore, M_{ij} represents the possibility of transferring V_i to V_j. In the transition matrix M, the i^{th} column includes the flow of the output V_i and the i^{th} row includes the input flow to V_i. So, the sum of the elements of each column equals to 1, but this rule is not true for every row [19].

The process of the algorithm MCL includes two expand and inflate operators on the random matrix; and this is continued until the matrix is converged. In addition, there is also a prune step in end of each inflate step to save memory and increase speed, which are addressed as bellow:

Expand calculates the square of the matrix M as $M_{exp} = M * M$, which is a factor to transfer power according to Markov chain and lets the different regions in a graph to be connected.

Inflate increases each element of the matrix M to the value of the inflate parameter r (r>1), and then normalizes the columns of the matrix so that the sum of the existing

entries in each columns is 1. How inflate is calculated for each matrix element is presented in eq. 4:

$$M_{inf}(i, j) = \frac{M(i,j)^r}{\sum_{k=1}^{n} M(k,j)^r} \quad (4)$$

The parameter r is considered as 2 which causes strong flows become stronger, and weak flows become weaker. Therefore, the Inflation equation will become eq. 5:

$$M_{inf}(i, j) = \frac{M(i,j)^2}{\sum_{k=1}^{n} M(k,j)^2} \quad (5)$$

The last step of the MCL algorithm is prune so that very small values are emitted to reduce the used memory and calculation operation. To do so, a threshold in considered and the values smaller than it are considered as zero [18].

In the MCL algorithm, with the beginning of a standard flow matrix, then algorithm is an iterative process of applying two operators - expansion and inflation - on a matrix, until the output matrix reaches the steady state M^{∞} and after that applying these two operators has no effect on the output matrix.

Up to this step, the given idea, community detections based to content and link analysis, is done. By studying the MCL algorithm, it has certain features to the spectral clustering algorithm and heuristic clustering algorithm, but the algorithm speed has no proper scalability for the large networks, on the other hand, graphs have lower speed in early iterations of the MCL algorithm due to fewer zero values, and if it is done on a smaller graph, algorithm speed is considerably increased. So, in the following, the algorithm is presented in a scalable manner with some changes. So, ML-MCL algorithm is proposed. The general design of a multi-level algorithm and its illustration are presented in the next section.

3.3 Multi-Level Markov Algorithm to Scalab

At first, the general framework of the multi-level algorithm is shown in Fig. 4 for better understanding.

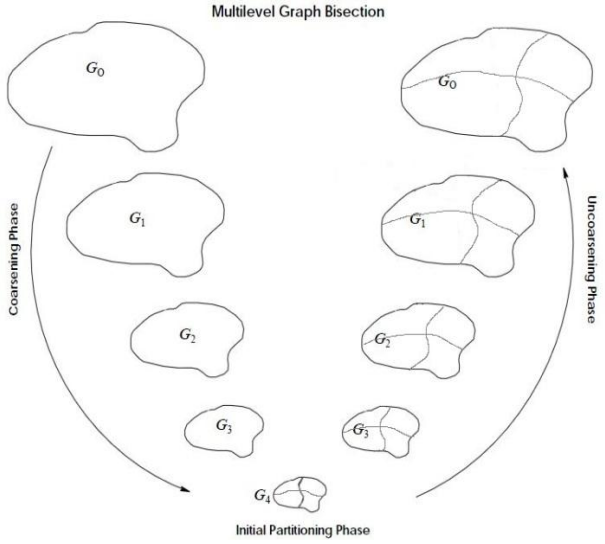


Fig. 4. doing a multi-level algorithm has three main steps; Coarsening, primary grouping, Uncoarsening. In this figure, algorithm has 4 levels, and it is divided into 4 groups in the primary grouping of the graph, then the main graph is resulted by implementing uncoarsening level[20].

As it is observed in Fig. 5, the algorithm is implemented in three levels: coarsening, primary grouping and uncoarsening, which each one is briefly described in the following:

1. Coarsening: the input of the level is the main graph G (a graph which has been weighted in the previous steps). And it is frequently divided into smaller graphs $G_1,G_2,G_3,\ldots G_1$, in which $|V_0|>|V_1|>|V_2|>\ldots>|V_1|$. This minimizing is continued until G_1 size can be controlled. In each level of this step, graph nodes are merged with each other and formed a super-node and sent to the next level. The ways to go from G_1 to G_1 are different of which different types are described in [13]. Coarsening utilizes the maximum matching to maintain the main graph properties. The time required to calculate these levels is O(log(n'/n)) in which n is the number of the peaks in graph G_0 and n' is the number of the peaks in G_1. In each level of coarsening, three steps are done to convert graph G_i to G_{i+1}: the first step is considered a subset of nodes to convert to a super-node according to coarsening method (this choice can be done according to the strengths and weaknesses of the interactions, randomly or with other factors). In this paper, the criterion, the most similarity is considered to merge the nodes. In the second step, the rules required to merge are applied, and in the third one, the edge weights are calculated according to the new nodes [20].
2. Primary grouping: in this step, the MCL algorithm is iterated on G_1 with few times (e.g. 4 or 5 iterations) by starting from the graph $G_1=(V_1;E_1)$ of the previous step. The reason to implement the algorithm for few times in this step is only controlling the graph distribution, and in this step, obtaining a balance is not considered. How the MCL algorithm works is fully described in the previous section. There is no problem according to the fact that the MCL algorithm did not have a proper scalability but because the graph size is not minimized in this step and on the other hand the algorithm work properly in small-sized graphs.

Uncoarsening: in the step, the multi-level algorithm is the goal to obtain the main graph by decomposing super-nodes and forming its primary node while the grouping done in the previous step is maintained. Finally, when the main graph is formed, the MCL algorithm is implemented to obtained convergence.

4. Experiment

In the previous section, the proposed strategy was presented to community detections using node content and link-analysis in social networks. In this section, Facebook real world datasets[1] was used to evaluate the efficacy of the proposed method, in which the number of the nodes is 4039

[1] http://snap.stanford.edu/

and the number of the edges is 88234. Generally, the datasets content is divided into three educational, work, location and sections; and each of these sections presents the trend of users to different groups, which represents difference in interests and properties of users. Consider the following four scenarios; the first scenario corresponds to the educational Facebook social network dataset, the second scenario is the location, third scenario corresponds the fields of work Facebook social network dataset and the

Input: Original graph G, Inflation parameter r, Size of coarsest graph c

// Phase 1: Coarsening: Coarsen graph successively down to at most c nodes.
$\{G_0, G_1, \ldots, G_k\}$ = CoarsenGraph(G, c)
// G_0 is the original graph and G_k is the coarsest graph
// Phase 2: Curtailed MCL along with refinement
// Starting with the coarsest graph, iterate through successively refined graphs.
// Run MCL for a small number of iterations.
for small number of iterations do
Markov Clustering
end for
// Phase 3: UnCoarsening graph successively access to original graph.
Run MCL on original graph until convergence
repeat
Markov Clustering
until M converges

Fig. 5. Details of ML-MCL Algorithm

fourth scenario considers all fields. Evaluations were done in a dual-core system with a 4GB main memory and processing speed 2.53 GHz.

4.1 Experiment Criterions

The sachan's models algorithm and the MCL algorithm were selected because of the similarity the algorithms are presented.

The first criterion to study the quality of clusters is similarity measurements that has been scaled by entropy. Well, low entropy means the high similarity between clusters and homogeneous clusters, and high entropy means there is no similarity. After checking the integral algorithm, sachan's models algorithm and the MCL algorithm, According to this criterion the results will be shown by Fig. 6, as you see, the proposed algorithm has low entropy. One of the reasons that the MCL algorithm entropy is higher than the integral algorithm entropy is only paying attention to the network structure for clustering.

Another criteria is the number of clusters that has been assessed. If the number of the cluster are much more, it causes Fragmentation in network graph in clustering, and it causes low communication discovering and high clustering. There is no paying attention to this subject in the MCL algorithm while in the integral algorithm we have done some reforms and as a result we find high coherence and thematic similarities. Evaluation results are shown in Fig. 7.

Another criterion is normalized cut or conductance that has been for evaluating of cluster quality. The normalized cut of a cluster is simply the number of edges that are "cut" when dividing this cluster from other clusters. The Normalized Cut criterion has been the quality of clusters . The normalized cut of a cluster C in the graph G is defined as eq. 6. The average normalized cut of a clustering is the average of the normalized cuts of each of the constituent clusters.

$$\text{N cut (C)} = \frac{\sum_{v_i \in C, v_i \notin C} A(i,j)}{\sum_{v_i \in C} \text{degree}(v_i)} \qquad (6)$$

Evaluation results are shown in Fig. 8, that the Integrative algorithm is presented better than MCL algorithm. Generally, the sachan's models algorithm is very close to our approach, but according to the figures proposed method is clearly effectiveness.

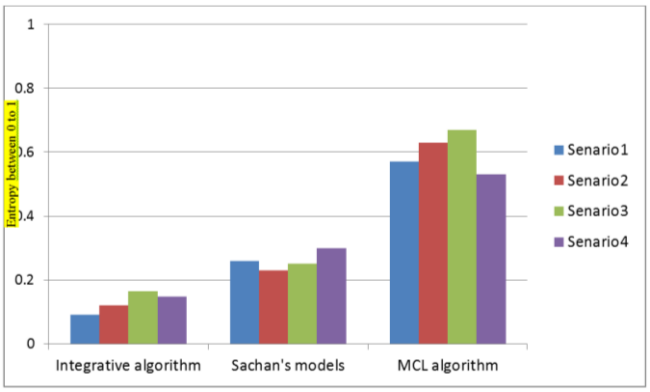


Fig. 6. Entropy

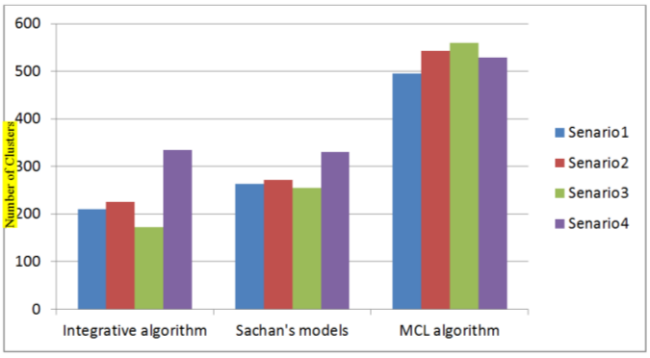


Fig. 7. Number of Clusters

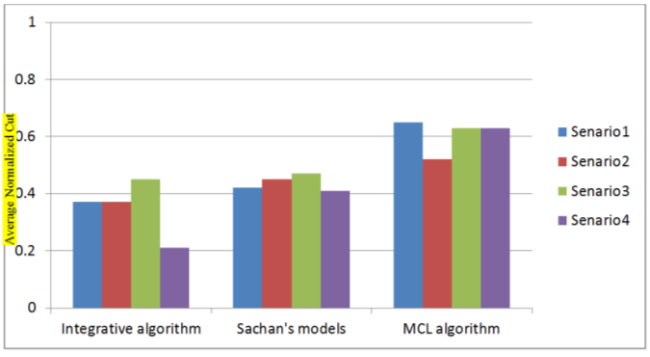


Fig. 8. Average Normalized Cut

5. Conclusions

The first obstacle to threaten many clustering algorithm is large graphs. Many clustering algorithms, including restrictions like directional network are not considered. But in the clustering algorithms, for simplicity, the direction of network graphs is not considered. In addition, combining the link and content analysis method in the same time to get a better clustering has been noted less. But all of these issues have been considered in integrative algorithm.

According to the evaluation, The clusters are homogeneous and dense , so it is clear that the integral algorithm is better than MCL algorithms and sachan's models algorithm, and it can be used to explore the communication between the social networks (weighted or non- weighted and directional or non-directional). Developing this algorithm and attention to overlapping nodes are future work.

References

[1] M. Sachan, D. Contractor, T. A. Faruquie, and L. V. Subramaniam, Using content and interactions for discovering communities in social networks, in Proceedings of the 21st international conference on World Wide Web, 2012, pp. 331–340.

[2] 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, pp. 164–173, 2012.

[3] S. Wasserman, Social network analysis: Methods and applications, Vol. 8. Cambridge university press, 1994, pp. 1-27.

[4] S. Salem, S. Banitaan, I. Aljarah, J. E. Brewer, and R. Alroobi, Discovering Communities in Social Networks Using Topology and Attributes, in Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on, 2011, Vol. 1, pp. 40–43.

[5] S. Dongen, A new cluster algorithm for graphs, Center for Mathematics and Computer Science (CWI), Amsterdam, 1998.

[6] B. W. Kernighan and S. Lin, An efficient heuristic procedure for partitioning graphs, Bell system technical journal, Vol. 49, No. 2, pp. 291–307, 1970.

[7] M. E. Newman and M. Girvan, Finding and evaluating community structure in networks, Physical review E, Vol. 69, No. 2, p. 026113, 2004.

[8] S. Papadopoulos, Y. Kompatsiaris, A. Vakali, and P. Spyridonos, Community detection in social media, Data Mining and Knowledge Discovery, Vol. 24, No. 3, pp. 515–554, 2012.

[9] A. Clauset, Finding local community structure in networks, Physical review E, Vol. 72, No. 2, p. 026132, 2005.

[10] F. Luo, J. Z. Wang, and E. Promislow, Exploring local community structures in large networks, Web Intelligence and Agent Systems, Vol. 6, No. 4, pp. 387–400, 2008.

[11] M. Fiedler, Algebraic connectivity of graphs, Czechoslovak Mathematical Journal, Vol. 23, No. 2, pp. 298–305, 1973.

[12] S. Smyth and P. Smyth, A spectral clustering approach to finding communities in graphs, in SDM, Vol. 5, 2005, pp. 76–84.

[13] G. Karypis and V. Kumar, A fast and high quality multilevel scheme for partitioning irregular graphs, SIAM Journal on scientific Computing, Vol. 20, No. 1, pp. 359–392, 1998.

[14] I. S. Dhillon, Y. Guan, and B. Kulis, Weighted Graph Cuts without Eigenvectors: A Multilevel Approach, IEEE Trans. Pattern Anal. Mach. Intell, Vol. 29, pp.1944–1957, 2007.

[15] S. T. Barnard and H. D. Simon, Fast multilevel implementation of recursive spectral bisection for partitioning unstructured problems. Concurrency: Practice and Experience, Vol. 6, No. 2, pp. 101–117, 1994.

[16] Liu, Chenlong, J. Liu, and Zh. Jiang, A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks, *Cybernetics, IEEE Transactions* vol. 44, pp. 2274-2287, 2014.

[17] Qi, Xingqin, et al, Optimal local community detection in social networks based on density drop of subgraphs, *Pattern Recognition Letters* 36, pp. 46-53, 2014.

[18] S. M. Van Dongen, Graph clustering by flow simulation, PhD Thesis, University of Utrecht, 2000.

[19] V. M. Satuluri, Scalable Clustering of Modern Networks, The Ohio State University, 2012.

[20] C. Chevalier and I. Safro, Comparison of coarsening schemes for multilevel graph partitioning, in Learning and Intelligent Optimization, Springer, 2009, pp. 191–205.

Zahra Arefian received the M.Sc. degree in Computer Architecture Engineering from Isfahan University, Esfahan, Iran, in 2014. She received the B.Sc. degree in Computer Hardware Engineering from HaMedan University of Technology, Hamedan, Iran, in 2012. Her area research interests include Social Network.

Mohammad Reza Khayyam Bashi received Ph.D. degree in Department of Computing Science, University of Newcastle Upon Tyne, Newcastle Upon Tyne , England in 2006. He received the M.Sc. degree in Computer Architecture Engineering from Sharif University of Technology, Tehran, IRAN in 1990. He received the B.Sc. degree in Computer Hardware Engineering from Tehran University, Tehran, IRAN in 1987. His area research interests include Distributed Systems, Computer Networks, Fault Tolerance.

On-road Vehicle Detection based on Hierarchical Clustering and Adaptive Vehicle Localization

Moslem Mohammadi Jenghara
Department of Electrical and Computer Engineering, University of Kashan, Kashan, Iran
m.mohammadi.14@gmail.com
Hossein Ebrahimpour Komleh*
Department of Electrical and Computer Engineering, University of Kashan, Kashan, Iran
ebrahimpour@kashanu.ac.ir

Received: 16/May/2015 Revised: 22/Sep/2015 Accepted: 04/Oct/2015

Abstract

Vehicle detection is considered to be a significant task in automatic driving which is regarded as a challenge and thorny issue for researchers in this field. The majority of commercial vehicle detection systems are based on radar. However, methods using radar suffer from problems such as the one encountered in zigzag motions. Image processing techniques can overcome these problems .This paper proposed an approach based on hierarchical clustering in which low-level image features are used to detect on-road vehicles. The approach introduced in this study is based on a new clustering method called teammate selection. In this clustering method, a new merging measure based on cluster center distances and gray scale values was introduced. Each vehicle was assumed to be a cluster. In traditional clustering methods, the threshold distance for each cluster was fixed; however, in the method proposed in this paper, the threshold distance is adaptive which varies according to the position of each cluster. The threshold measure was computed with bivariate normal distribution. Sampling and teammate selection for each cluster were carried out by cluster members based on weighted average. Unlike other methods which used only horizontal or vertical lines, a fully image edge detection algorithm was utilized in this study. Corner is an important video image feature which is commonly used in vehicle detection systems. However, Harris features were used in this paper to detect the corners. Furthermore, LISA data set was used to evaluate the proposed method. Several experiments were conducted to investigate the performance of proposed algorithm. Experimental results indicated good performance compared to other algorithms.

Keywords: Adaptive Feature Grouping; Moving Camera Image Processing; Vehicle Detection; Hierarchical Clustering; Teammate Selection Clustering.

1. Introduction

Driver assistance and traffic monitoring systems are of high significance in intelligent vehicles. Object detection and tracking which were observed by ego-vehicle on/around road such as cars, pedestrians and other obstacles are the main requirement to design and implement these systems. Camera is usually located at the center of the front bumper of ego-vehicle car. The performance of detection system on the road is a critical issue for system administrators with respect to security. Hence, systems must be robust enough for various conditions. Vehicle detection is a challenging domain for the committee of intelligent machines. Road environments vary in terms of traffic situations, number of cars, lighting conditions, weather conditions, construction of roads, tunnels and more. Consequently, a fully adaptive and parametric system is required for the existing variable conditions. Near or mid-range vehicle detection is another effective issue in the structure of environment and camera parameters. Various sensors are available which can be used in driver assistance systems such as lidar, radar, ultrasound and embedded cameras. There are four situations based on camera and object movement which are listed below:

I. Stationary camera, constant object
II. Stationary camera, moving object
III. Moving camera, stationary object
IV. Moving camera, moving object

The second situation is more practical and more common [1]. A fixed camera located on a highway for speed control is an example of the second situation. However, driver assistance systems, camera and objects are mobile which is regarded as a highly sophisticated condition. Due to background changes and inapplicability of differential techniques, the detection of a moving object with a moving camera is a challenging task.

Most commercial vehicle detection systems are based on radar. It is a sensor which has lots of limitations such as angular constraint and temporal resolution. In general, radar-based systems can detect vehicles located directly in front of the observing car but they cannot detect marginal cars which are located at different angles. In this case, any change in car line may be dangerous. Also, the use of radar-based systems in high zigzag and steep roads can be problematic. In contrast with radar-based systems, it should be noted that cameras are inexpensive, consume little

* Corresponding Author

power and can be easily managed to capture information from the environment. Visual data processing is complex but it provides valuable information about the environment.

In this paper, the researchers focused on low-level features and used them to specify the location of vehicles. The shape and structure of vehicles were not of any significance. Since the intended vehicle detection system is in a dynamic environment and inside an ego-vehicle, both the observing car and other vehicles are moving. The size and position of the devices can vary over time. Thus, an adaptive method based on coordinates can be useful. This adaptability can be utilized in different components of the system including thresholds setting which is discussed in the proposed method.

2. Related Works

Detection of vehicles using embedded camera image processing can be done through various approaches such as learning-based methods.

Carrafi et al. [4] proposed a learning-based approach using waldboost [2]. They introduced a cascading fine-grain detection system similar to the viola-jones [3] method. Waldboost algorithm is implemented to reject negative choices in the first steps which has a significant impact on system speed [4]. Deformable object model learning [5] was proposed for learning and identifying vehicle features as an object model in which latent support vector machine (LSVM) and Histogram of Oriented Gradients (HOG) were combined. Despite high detection efficiency of the above-mentioned method, its computational complexity allows for only 1 fps processing.

Jazayeri et al. [6] used Hidden Markov Model (HMM) to develop a system which could separate vehicles and background from one another; it was also able to probabilistically model motions in terms of scenes based on frame features. Low-level features such as edges and corners which are resistant to light and shape variations were used in [6].

Samadi et al. [7] proposed a multi-agent system for vehicle detection. Each agent in this multi-agent system carries out one part of the diagnostic process. The hypothesis exchange and conflict resolving are done by cooperation among agents. In this system, the following agents were used: edge detection agent, contour agent, vehicle agent, license-plate rectangle detection agent, license-plate line detection agent, wheel detection agent, plate candidate verification agent and symmetry detection agent.

In recent years, approaches based on active learning led to good results on detecting road vehicles [8]. HOG-SVM and Haar-like features with Adaboost classifier, traditional methods in active learning were investigated and compared with each other in terms of time complexity and other parameters [9]. In general, methods based on active learning produce a simple classifier with an easy supervised learning. Then, querying classifier selects informative patterns for retraining classifier again. Indeed, the query function which selects informative and complex patterns and classifier input for retraining and updating decision boundaries is the main part of active learning [10,11].

A method based on active learning through split and merge was proposed in [12] to determine various components of vehicles. The learned classifier is SVM. PCA, Gabor, Wavelet and combined Gabor-Wavelet features with NN and SVM classifiers were used by Zhang [13] for diagnosis before the occurrence of accident conditions. An augmented Gabor feature for detecting vehicles was used in [14] where Gabor filter parameters were improved and learned by SVM classifier to cover more sub windows containing pieces of vehicles.

Kim et al. [15] used a combination of sonar and vision sensors where lighting and distance conditions had no impact on system accuracy. Sonar sensors for the distance and cameras up to 10 meters were used. Features such as shadows, road lines, horizontal lines and vehicle symmetry were used in image processing. Template matching [16] and vertical symmetry detection [17] are other methods for vehicle detection. Many researchers work on computer vision-based intelligent transportation systems [18].

Most on-road vehicle detection systems detect vehicles in line with the observer. There are many horizontal lines at the back of the vehicles including shadow lines, window lines and top and bottom lines of the vehicle. This feature is relatively stable against light changing and scaling. Matthews et al [19] used edge and vertical lines detection to indicate the left and right margins of vehicle. Image edge is considered as a favorite feature for researchers in the field of vehicle detection [20, 21]. Corners are maintained in their position on vehicles and background; hence, they provide useful information about the different components of an image. Bertozzi et al. [22] proposed a corner-based method for vehicle detection.

Each vehicle is represented by local variation of gray level with a certain texture [23]. Texture regions can be processed further to make accurate detection. Calculating entropy based on neighboring pixels is the criterion for determining texture. Areas with high entropy are selected for further processing [24]. Kalinke [25] used texture to focus algorithm on the areas having lots of information. Shannon introduced local entropy to measure the information of each image patch [26]. Symmetry is an invariant feature at the rear of vehicles which is stationary and stable under different light conditions and scaling. Many researchers use this feature to detect vehicles [27,28]. A symmetry and edge-based vehicle detection method using edge oriented histogram (EOH) and support vector machine (SVM) was proposed in [29] for approximate vehicle location and improving post-processing.

Optical Flow (OF) is an informative motion-based feature which provides information about the direction and speed of moving objects in video frames [6,30,31]. Pixel-based and feature-based methods are two main approaches for optical flow computation [18]. In

computing optical flow, (u,v) feature points in (It) and (It+1) are mapped so that eq.1 is minimized [32, 33].

$$e(dx, dy) = \sum_{x=u-w}^{u+w} \sum_{y=v-w}^{v+w} \big(I(x, y) - I(x + dx, y + dy)\big) \quad (1)$$

Shadow is used as a feature for vehicle bounding. The lower part of vehicles in different lighting conditions has different shadows. This feature can be used to determine the underside of a vehicle [34]. Another feature used for separating vehicles from background is color [35,36]. RGB color system [36] and L*A*B color system [37] are more conventional. LED lighting is used for detecting and tracking vehicles at night [6].

3. The Proposed Method

Different image features which are of high importance for vehicle detection were briefly introduced and reviewed in several studies in the previous section. Indeed, multiple low-level features were used in this study. The researchers tried to group and cluster these features to identify vehicles. ROI (Region of Interest) feature is determined based on Gaussian probability distribution function. Fig.2 depicts the block diagram of the proposed method. The components of the proposed method are described consequently.

3.1 Feature Extraction

As discussed earlier, there exist many horizontal lines at the back of vehicles such as shadow lines, window lines, and top and bottom lines of a vehicle. This feature is relatively stable to the changes of light and scale. Indeed, it should be noted that, in some frames, vehicles might not be exactly in front of the camera i.e. they may be in margins; hence, horizontal or vertical lines are not clear. Thus, an algorithm used for edge detection methods such as canny [38] is regarded as more appropriate. In contrast with other methods which only used horizontal or vertical lines, a fully edge detection algorithm was utilized in this study. Fig.1.b illustrates the result of edge detection on the image.

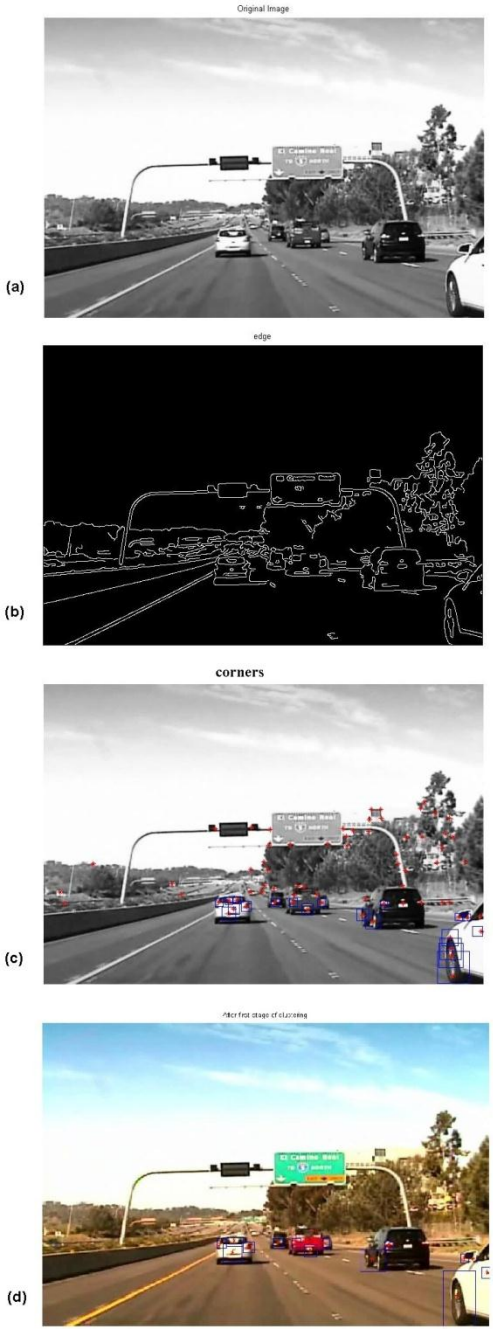


Fig.1. Original image and extracted features (a) original image (b) image edge extracted by canny algorithm (c) corners extracted by harris method (d) adaptive boundingbox

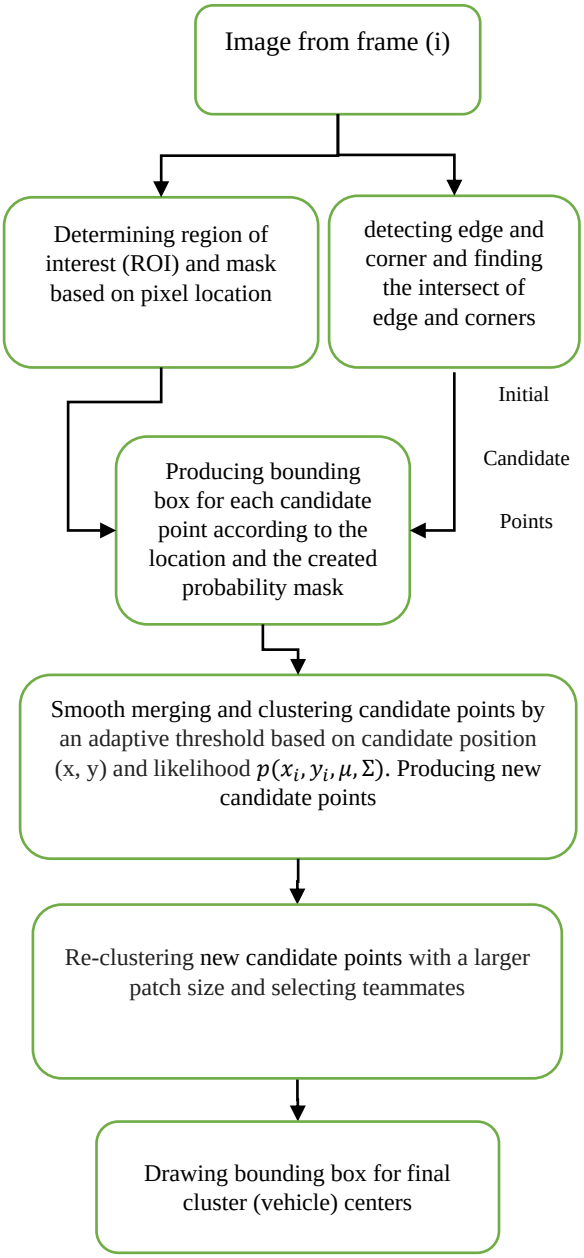


Fig. 2. Block diagram of the proposed method

Corner is considered to be an important feature of video images which is commonly used in vehicle detection systems. The most common corner extraction method is based on harris features [39]. In this study, image corners in combination with edges were used as low level features for detecting vehicle. Fig.1.c depicts corners on an image from LISA dataset.

3.2 Focus on Probability

As shown in Fig.1.a, each frame has numerous details around the road. Hence, determining ROI for further processing is essential for two aspects. Firstly, appropriate ROI placing has a direct impact on the accuracy of detection. Secondly, as a given area is determined more precisely, less time will be spent for processing in the next level. According to camera settings and its circumstance and location on the ego_vehicle, ROI will vary in size and details.

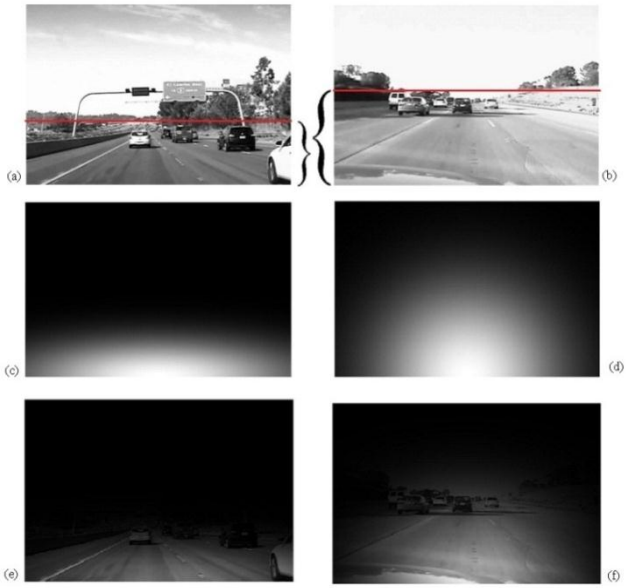


Fig. 3. Angular installed camera effect on captured image, more angular (left side); less angular (right side)

In this paper, certain parameters such as ROI determination and marginal detail parameters including camera parameters were specified. As illustrated in Fig.3.a, in case the installed camera is angular, the captured image will include more details and the sky will be the greatest part of image. However, in case camera position and angle relation to the horizon is less, the image will include more vehicles but less marginal details. Fig.3.b illustrates this fact. λ and θ refer to ego-vehicle height and camera angle with respect to the horizon. The values of the parameter are converted to interval [0,1].

The probability that each image pixel is a vehicle pixel is computed according to its position X(x, y) and the bivariate Gaussian distribution function mentioned in eq.2. For obtaining more adaptation in this distribution function, the parameters λ and θ are used to create the covariance matrix (eq.3).

$$p(X; \mu, \Sigma) = \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)\right) \quad (2)$$

$$\Sigma = \begin{bmatrix} C*\lambda & 1 \\ 1 & R*\theta \end{bmatrix} \quad (3)$$

In equation 3, C and R denote the size of image columns and rows respectively. Since the Gaussian distribution is symmetrical in relation to the mean of variables, the positive part was only used. Fig.3.c and Fig.3.d indicate the probability mask of each region as a gray level in the image with the specified parameters. Fig.3.e and Fig.3.f are produced by applying these masks on the original images. Inasmuch as bright parts of the masks are ROI, they must be processed further in the next levels.

3.3 Clustering of Initial Center Points

As discussed earlier, edge and corner features were used in the present study to detect vehicles. Consequently, candidates for further processing were selected from the intersection of dilated edges and corners according to equation 4.

$$C = \text{corners} \cap (\text{edge} \oplus \text{se}) \tag{4}$$

Indeed, edges and corners are features which were extracted in the previous stage. se stands for dilation mask which is like a circle with 5-pixel radius. Further processes were conducted on C.

$$C = \{c_i = (x_i, y_i)\ \ i = 1,2, \dots, K\}, \tag{5}$$

C includes coordinates of candidate points. It should be noted that the image sizes of vehicles near camera are large and those of distant vehicles are small. Hence, applying a uniform threshold for drawing bounding box and clustering centers are not possible. Consequently, the following adaptive threshold based on candidate position (x, y) and the likelihood $p(x_i, y_i, \mu, \Sigma)$ was used in the study.

$$AD_i = \text{patch}_{\text{size}} * \ p(x_i, y_i, \mu, \Sigma) \tag{6}$$

For drawing a bounding box around the initial centeral points and clustering them, $\text{Adaptive}_{\text{Distance}}$ (AD) was used which is shown in eq.5 threshold. Thus, the center which is probably considered to be a vehicle has an even greater margin (Fig.1.c).

3.4 Clustering of Candidate Centers

For obtaining more accurate results, centers were clustered using the Euclidean distance measure and the AD threshold was computed for each center. It should be pointed out that the number of clusters is determined while clustering.

$$D(c_i, c_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{7}$$

$$TD(i, j) = \min(AD_i, AD_j) \tag{8}$$

$$G(c_i, c_j) = \begin{cases} 1 & \text{if } D(c_i, c_j) \le TD(i, j) \\ 0 & \text{o. w} \end{cases} \tag{9}$$

In this grouping method, the number of clusters is variant which is determined while clustering according to the cluster merging. After clustering, cluster centers should be specified according to the following equation for subsequent processing.

$$\underbrace{CL_k}_{k=1..N} = \{[\text{mean}(x_i) \quad \text{mean}(y_i)] | \forall i \ \ G(c_i, c_k) == 1\} \tag{10}$$

In this equation, N denotes the number of pre-generated clusters. After initial clustering, the final step to vehicle determination is carried out by re-clustering with a larger patch size and averaging based on the number of members. Neighbor points are selected by choosing teammates based on correlation and average computation. In this stage, pairwise correlation among all central points is measured. Then, a multivariate data adjacency matrix is produced. Correlations between points are computed based on gray level differences and Euclidean distances between center coordinates as bellow:

$$CR(c_i, c_j) = \frac{1}{(|v_i - v_j|)^{\rho} + (\text{dist}(c_i, c_j))^{\gamma} + 1} \tag{11}$$

Where, v_i and c_i refer to the i'th center point gray level and coordinate respectively.

CR has a value within the interval [0, 1]. The self-adjacency value for each center point is 1. γ and ρ parameters determine the effectiveness of the Euclidean distance and gray level differences respectively. By increasing the Euclidian distance between centers and the difference between gray levels, CR value tends towards zero. After constructing multivariable adjacency matrix, the recruitment is conducted for grouping the same samples. The following question arises in this stage: how many centers with how much adjacency can form a cluster? For answer this question, an instance candidate center is selected and it is assumed to be a cluster. CR values of this center and those of other centers are computed. The closest center is added to the cluster. The mean of CR values in the cluster is computed and the average is weighted based on the number of samples it takes. The weight grows while the number of samples increases. All the processes including selecting, averaging and weighting is referred to as teammate selection which is depicted in fig.4.

In fact, sampling and teammate selection and weight increase continue as far as monotony property of averaging is not violated in recruiting. By imposing this constraint, nearby points form a cluster. Even though distant point selection increases averaging weight, it is not enough to enhance the total amount of weighted average with respect to the previous value. Hence, our objective was to recruit more so that the adjacency value of members would not be less than that of a threshold. As mentioned earlier, after applying weighted averaging and recruitment, an adaptive clustering step based on different threshold was used to obtain the final results.

```
Compute CR matrix based on eq.10;
foreach ( unselected candidate center point i in CR matrix)
  Num=1;
  Len=number of disjoint centers;
  K=1;
  raw_mean=0;
  While (K<=Len)
    Max_v=Choose maximum value CR in i'th row;
    Update raw_mean according Num and Max_v value;
    New_mean=weighting raw_mean according to Num value;
    If ( new_mean >=old_mean)
       Add Max_v to i'th cluster and remove from CR matrix;
       Num++;
    else
       Break;
    End {if}
    K++;
  End {while}
End {foreach}
```

Fig. 4. The proposed teammate selection clustering algorithm

4. Results

4.1 Dataset Description

The proposed method was evaluated by means of LISA dataset available online at http://cvrr.ucsd.edu/LISA/index.html. The evaluation data set included three different sets with different traffic congestions and different levels of identification complexity. The first dataset included 300 frames captured on March with an SUV vehicle in a curved road. Also, pedestrians moved along and across the street. The second data set also included 300 frames which were taken in different lighting conditions on April. This data set consisted of multiple vehicles which were moving regularly. The third data set including 1600 frames was taken on January. This data set was complicated because it was taken at the most crowded time and included shadows, high-motion maneuvering and five movement lines. All these data sets were hand-labeled and a box was drawn around each vehicle.

Fig. 5. True labeled vehicles shown by green boxes and blue boxes in the proposed method

There are some well-known performance metrics such as precision, recall [40], average false positive per frames, average false positive per object and average true positive per frame [8] which were used in the study as the evaluation criteria.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{12}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{13}$$

$$\text{Average FP/frames} = \frac{\text{FP}}{\text{total number of frames processed}} \tag{14}$$

$$\text{Average FP/objects} = \frac{\text{FP}}{\text{true vehicles}} \tag{15}$$

$$\text{Average TP/frames} = \frac{\text{TP}}{\text{total number of frames processed}} \tag{16}$$

Table 1. Experimental results

Criteria	Description	#1	#2	#3
ATPpF	*Average TP per Frames*	1	2.966	2.67
AFPpF	*Average FP perFrames*	2.133	1	2.22
ATPR	Average True Positive Rate (recall)	1	0.989	0.621
AFDR	False Detection Rate	0.519	0.223	0.399
AFNpF	*Average FN per Frames*	0	0.033	0.193
AVpF	*Average Vehicle number per Frame*	1	3	4.38
AFPpObj	*Average FP per Objects*	2.133	0.333	0.507
ATPpObj	*Average TP per Objects*	1	0.989	0.621
AFNpObj	*Average FN per Objects*	0	0.011	0.044
Precession		0.48	0.768	0.60

Table 2. Comparison of the proposed method with Elvis [41] and active learning method [8].

	Best value➔	Depends on #vehicles	Small	small	large
tracking system	Criteria / dataset	$\boldsymbol{\frac{TP}{frame}}$	$\boldsymbol{\frac{FP}{frame}}$	FDR	TPR
The Proposed method	#1	1	2.133	0.519	1
	#2	2.966	1	0.223	0.989
	#3	2.67	2.22	0.399	0.621
Elvis[41]	#1	1	1.13	0.531	1
	#2	2.92	1.06	0.267	0.975
	#3				
Active learning method[8]	#1	1	4	0.797	0.835
	#2	3.16	2.7	0.458	0.981
	#3				

The obtained results are given in table (1) and the results of comparing the proposed method with two other methods are given in Table (2). As shown in Table (2), the proposed method has good results with respect to all of the computed measures. The problem was that only the rears of vehicles were tagged. In other words, in case a vehicle appears with a part other than the rear part, then, the detection will fail and the algorithms detecting these vehicles will be registered as an error in terms of system evaluation. Since a correctly detected vehicle (TP) will be considered as the system error (FP), hence, the system performance will decreased. Fig.4 shows a frame of each data set in which the proposed system detected a vehicle or a pedestrian which are not labeled in the data set. As a case in point, in the first dataset, a pedestrian crossing the street was detected but it was not labeled in the data set.

The algorithm used in the present study had some parameters which depended on the capturing condition and image size. For example, in dataset#2, $\theta = 0.9$ indicates that the camera is installed angularly and the sky is included in the majority of the image. The patch size and λ and θ were selected according to the image size and the

conditions. Nevertheless, γ and ρ were selected expertly and based on trial and error. Due to the incomplete labeling of the datasets, Average FP per objects (labeled) was too high, particularly in the first dataset.

Table 3. The parameters used in the proposed method

Parameter	Brief description	#1	#2	#3
λ (normalized)	The ego-vehicle's height	0.4	0.1	0.6
θ(normalized)	The camera's angle relative to the horizon	0.7	0.9	0.4
γ	determine effectiveness of the Euclidean distance	0.7	0.5	0.5
ρ	determine effectiveness of the gray level differences	0.5	0.5	0.8
Patch size #1	Initial patch size for bounding box determination in first phase	100	100	100
Patch size #2	Initial patch size for bounding box determination in second phase	300	250	200
τ	Threshold for neighboring	10	10	10

5. Conclusion and Future Works

The study reported in this paper focused on the significant challenge and issue of detecting vehicles with a camera embedded inside a car. The underlying approach was based on a new clustering method which is referred to as teammate selection. In this clustering method, a new merging measure based on the distances of cluster centers and gray scale values was used. Several general features such as "edges" and "corners" were taken into consideration to robustly characterize vehicles. The detection was carried out hierarchically at several stages. After extracting feature, masks were used to determine ROI based on the probability of the position of each pixel. Then, feature clustering was performed according to the listed parameters. Next, a weighted average based sampling and teammate selection was measured and the final clustering was implemented by the new parameters.

It can be argued that the *significant* contribution of this study was to determine vehicles as clusters without having any information about the number of them in each frame. The utilized algorithm had some parameters which were functions of the capturing condition and image size. The results of the study indicated that the high TPR (recall) is a notable benefit of the proposed method. Due to the complexity and difficulty of dataset#3, the obtained results were not satisfying. Indeed, it can be maintained that parameters for complex environments can be improved by optimizing algorithms such as genetic algorithms.

References

[1] S. Nadimi and B. Bhanu, "Multistrategy fusion using mixture model for moving object detection," in *Multisensor Fusion and Integration for Intelligent Systems, 2001. MFI 2001. International Conference on*, 2001, pp. 317-322.

[2] J. Sochman and J. Matas, "Waldboost-learning for time constrained sequential detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 150-156.

[3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, vol. 1, pp. I-511-I-518.

[4] C. Caraffi, T. Vojir, J. Trefny, J. Sochman, and J. Matas, "A system for real-time detection and tracking of vehicles from a single car-mounted camera," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, 2012, pp. 975-982.

[5] A. Takeuchi, S. Mita, and D. McAllester, "On-road vehicle tracking using deformable object model and particle filter with integrated likelihoods," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, 2010, pp. 1014-1021.

[6] A. Jazayeri, H. Cai, J. Y. Zheng, and M. Tuceryan, "Vehicle detection and tracking in car video based on motion model," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, 2011, pp. 583-595.

[7] S. Samadi and F. M. Kazemi, "A Multi-Agent Vision-Based System for Vehicle Detection," *World Applied Sciences Journal*, vol. 15, 2011, pp. 1722-1732.

[8] S. Sivaraman and M. M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, 2010, pp. 267-276.

[9] S. Sivaraman and M. M. Trivedi, "Active learning for on-road vehicle detection: A comparative study," *Machine Vision and Applications*, 2011, pp. 1-13.

[10] D. Cohn, L. Atlas, and R. Ladner, "Improving generalization with active learning," *Machine Learning*, vol. 15, 1994, pp. 201-221.

[11] C. H. Lampert and J. Peters, "Active structured learning for high-speed object detection," in *Pattern Recognition*, ed: Springer, 2009, pp. 221-231.

[12] S. Sivaraman and M. M. Trivedi, "Real-time vehicle detection using parts at intersections," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, 2012, pp. 1519-1524.

[13] Z. Sun, G. Bebis, and R. Miller, "Monocular precrash vehicle detection: features and classifiers," *Image Processing, IEEE Transactions on*, vol. 15, 2006, pp. 2019-2034.

[14] H. Cheng, N. Zheng, and C. Sun, "Boosted Gabor features applied to vehicle detection," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, 2006, pp. 662-666.

[15] S. Kim, S.-Y. Oh, J. Kang, Y. Ryu, K. Kim, S.-C. Park, et al., "Front and rear vehicle detection and tracking in the day and

night times using vision and sonar sensor fusion," in *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, 2005, pp. 2173-2178.
[16] R. K. Nath and S. K. Deb, "On road vehicle/object detection and tracking using template," *Indian Journal of Computer Science and Engineering*, vol. 1, 2010, pp. 98-107.
[17] A. Broggi, P. Cerri, and P. C. Antonello, "Multi-resolution vehicle detection using artificial vision," in *Intelligent Vehicles Symposium, 2004 IEEE*, 2004, pp. 310-314.
[18] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, 2006, pp. 694-711.
[19] N. Matthews, P. An, D. Charnley, and C. Harris, "Vehicle detection and recognition in greyscale imagery," *Control Engineering Practice*, vol. 4, 1996, pp. 473-479.
[20] M. Betke, E. Haritaoglu, and L. S. Davis, "Real-time multiple vehicle detection and tracking from a moving vehicle," *Machine Vision and Applications*, vol. 12, 2000, pp. 69-83.
[21] N. Srinivasa, "Vision-based vehicle detection and tracking method for forward collision warning in automobiles," in *Intelligent Vehicle Symposium, 2002. IEEE*, 2002, pp. 626-631.
[22] M. Bertozzi, A. Broggi, and S. Castelluccio, "A real-time oriented system for vehicle detection," *Journal of Systems Architecture*, vol. 43, 1997, pp. 317-325.
[23] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *Acm Computing Surveys (CSUR)*, vol. 38, 2006, p. 13.
[24] T. Ten Kate, M. Van Leewen, S. Moro-Ellenberger, B. Driessen, A. Versluis, and F. Groen, "Mid-range and distant vehicle detection with a mobile camera," in *Intelligent Vehicles Symposium, 2004 IEEE*, 2004, pp. 72-77.
[25] T. Kalinke, C. Tzomakas, and W. v. Seelen, "A Texture-based Object Detection and an adaptive Model-based Classiæcation," 1998.
[26] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, 2001, pp. 3-55.
[27] A. Kuehnle, "Symmetry-based recognition of vehicle rears," *Pattern recognition letters*, vol. 12, 1991, pp. 249-258.
[28] T. Zielke, M. Brauckmann, and W. Vonseelen, "Intensity and edge-based symmetry detection with an application to car-following," *CVGIP: Image Understanding*, vol. 58, 1993, pp. 177-190.
[29] S. S. Teoh and T. Bräunl, "Symmetry-based monocular vehicle detection system," *Machine Vision and Applications*, vol. 23, 2012., pp. 831-842
[30] J. Diaz Alonso, E. Ros Vidal, A. Rotter, and M. Muhlenberg, "Lane-change decision aid system based on motion-driven vehicle tracking," *Vehicular Technology, IEEE Transactions on*, vol. 57, 2008, pp. 2736-2746.
[31] M. Perrollaz, J.-D. Yoder, A. Nègre, A. Spalanzani, and C. Laugier, "A visibility-based approach for occupancy grid computation in disparity space," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 13, 2012, pp. 1383-1393.
[32] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *IJCAI*, 1981, pp. 674-679.
[33] J. Choi, "Realtime On-Road Vehicle Detection with Optical Flows and Haar-Like Feature Detectors," *Urbana*, vol. 51, p. 61801, 2007.
[34] H. Mori and N. M. Charkari, "Shadow and rhythm as sign patterns of obstacle detection," in *Industrial Electronics, 1993. Conference Proceedings, ISIE'93-Budapest., IEEE International Symposium on*, 1993, pp. 271-277.
[35] J. D. Crisman and C. E. Thorpe, "Color vision for road following," in *1988 Robotics Conferences*, 1989, pp. 175-185.
[36] S. D. Buluswar and B. A. Draper, "Color machine vision for autonomous vehicles," *Engineering Applications of Artificial Intelligence*, vol. 11, 1998, pp. 245-256.
[37] D. Guo, T. Fraichard, M. Xie, and C. Laugier, "Color modeling by spherical influence field in sensing driving environment," in *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*, 2000, pp. 249-254.
[38] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 1986, pp. 679-698.
[39] C. Harris and M. Stephens, "A combined corner and edge detector," in *Alvey vision conference*, 1988, p. 50.
[40] D. L. Olson and D. Delen, *Advanced data mining techniques [electronic resource]*: Springer, 2008.
[41] RK. Satzoda and MM. Trivedi. "Efficient lane and vehicle detection with integrated synergies (ELVIS)." In Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on, 2014, pp. 708-713.

Moslem Mohammadi Jenghara is Instructor of Computer Engineering at Payame Noor University (PNU), Miandoab, Iran. He is currently a Ph.D. student in the Department of Electrical and Computer Engineering at the University of Kashan, Kashan, Iran. He had received his M.Sc. degrees in Computer Engineering (Artificial intelligence) in 2008, from Iran University of Science and Technology, Tehran, Iran. He has published several articles and presented several papers in international and national journals and conference about data mining, text mining and image processing. His main area of research includes Data mining, Pattern Recognition, text mining, temporal graph mining and applications of Artificial Intelligence in Bioinformatics.

HosseinEbrahimpour-Komleh is currently an Assistant Professor at the Department of Electrical and Computer Engineering at the University of Kashan, Kashan, Iran. His main area of research includes Computer vision, Image Processing, Pattern Recognition, Biometrics, Robotics, Fractals, chaos theory and applications of Artificial Intelligence in Engineering. He received his Ph.D. degree in Computer Engineering from Queensland University of Technology, Brisbane, Australia in 2006. His Ph.D. research work was on the “Fractal Techniques for face recognition”. From 2005 to 2007 and prior to joining the University of Kashan, he was working as a Post-doc researcher in the University of Newcastle, NSW, Australia and as a visiting scientist in CSRIO Sydney. Hossein Ebrahimpour- Komleh has B.Sc. and M.Sc. degrees both in Computer Engineering from Isfahan University of Technology (Isfahan, Iran) and Amirkabir University of Technology (Tehran, Iran,) respectively. He has served as the invited keynote speaker, editorial board member and reviewer of several journals and international and national conferences

A New Architecture for Intrusion-Tolerant Web Services Based on Design Diversity Techniques

Sadegh Bejani
Department of Information and Communication Technology, Imam Hossein University, Tehran, Iran
sbejani@ihu.ac.ir
Mohammad Abdollahi Azgomi*
Department of Computer Engineering, Iran University of Science and Technology, Tehran, Iran
azgomi@iust.ac.ir

Received: 28/Sep/2014 Revised: 22/Aug/2015 Accepted: 12/Sep/2015

Abstract

Web services are the realization of service-oriented architecture (SOA). Security is an important challenge of Web services. So far, several security techniques and standards based on traditional security mechanisms (i.e., encryption and digital signature) have been proposed to enhance the security of Web services. The aim of this work has been to propose an approach for securing Web services by employing the concepts and techniques of software fault tolerance (such as design diversity), which is called *intrusion tolerance*. Intrusion tolerance means the continuous delivery of services in presence of security attacks, which can be used as a fundamental approach for enhancing the security of Web services. In this paper, we propose an architecture for intrusion-tolerant Web services (ITWSs) by using both design diversity and composite Web services techniques. The proposed architecture is called *design-diverse intrusion-tolerant Web service* (abbreviated as DDITWS). For Web service composition, BPEL4WS is used. For modeling and verification of the proposed architecture, coloured Petri nets (CPNs) and the "CPN Tools" are used. We have model-checked the behavioral properties of the architecture to ensure its correctness using this tool. The reliability and security evaluation of the architecture is also performed using a stochastic Petri net (SPN) model and the "SHARPE" modeling tool. The results show that the reliability and mean-time-to-security-failure (MTTSF) in the proposed architecture are improved.

Keywords: Software Security; Intrusion Tolerance; Composite Web Service; Reliability; Petri nets.

1. Introduction

The occurrence of faults in a system is a deviation from correctness or accuracy in the system computations. A system failure means a cessation in the execution of the operation that was expected in a due time [1]. The causes of errors in software systems can be deliberate or unintentional (accidental). The occurrence of any faults or defects in software development process causes error in the software system. The cause of fault in system is unintentional and the incidents due to malicious attacks are rooted out of system. Intrusions are aimed to affect the system integrity, confidentiality or availability (CIA). Intrusion effect realizes on different aspects or incorrect system behaviors. [1]

Architectural evolution of software systems development indicates the widespread use of distributed software systems. [2] In process of software development from 2010 onwards, service-oriented architecture (SOA) has replaced the existing architectures. Web services are the main solution for the realization of service-oriented architecture. [2] Web services have specific features such as interoperability, self-description and self-containing. They use UUDI, HTTP, WSDL and SOAP interaction protocols. [3] Wide range of Web services execution environment, unknown users of Web services and challenges of communication security protocols in Web services interactions make Web services more susceptible to intrusion and attack than traditional software. [2]

Since several security techniques and standards based on traditional security mechanisms (i.e., encryption and digital signature, etc.) have been used to enhance the security of the Web services. The approach of these standards is based on "vulnerability avoidance" and "reducing system vulnerability", which are effective for known attacks. In this standards authentication mechanisms, access control, encryption, firewalls, reconfiguration management and data redundancy technique are used.

A second category of mechanisms is the usage of intrusion tolerance techniques. These techniques are effective for increasing system's tolerance against unknown attacks. In these circumstances, intrusion tolerance means the continuous delivery of services in presence of security attacks, which is a fundamental approach for increasing the security of Web services.

A software system is an intrusion-tolerant system (ITS), if after penetration, its basic services continue their performance and the system prevents from the creation of failure in its security features [3].

Web service technology enables the creation of complex services and provides composition services using simple services. There are two type of Web services: (1) based on simple object access protocol (SOAP), and (2)

* Corresponding Author

RESTful. In this research, we concentrate on SOAP-based Web services.

Composite SOAP-based Web services are composed of several Web services, in order to accomplish common work. [4] BPLE4WS is a standard software for Web service composition. Web service composition process in BPEL4WS makes it possible the realization and implementation of Web service composition. [4] The proposed architecture for ITWS is in the form of a composed Web service that can be implemented in BPEL4WS.

In this paper, we propose a new architecture for intrusion-tolerant Web services (ITWSs). The main approach of the proposed architecture is based on using intrusion tolerance concepts, design diversity techniques and composite Web service techniques. Creating efficient mechanisms for Web service intrusion detection, intrusion containment, intrusion recovery, providing data integrity, confidentiality, availability and neutralizing the influence of intrusions are special architectural considerations in the proposed architecture, which is called *design-diverse intrusion-tolerant web service* (DDITWS).

It is expected that by the realization of the proposed architecture, the developed Web service can continue its operation in the presence of intrusions and can provide the continuity of services without security failures.

The remainder of this paper is organized as follows. In Section 2, related works are reviewed. Section 3 gives an overview of the proposed DDITWS architecture. The security behavior of DDITWS is explained and investigated using coloured Petri nets (CPN) model of the architecture by using CPNs Tools is presented in Section 4. The results of the reliability and security evaluation of the proposed architecture are also given in this section. For this purpose, a stochastic Petri net (SPN) model and the SHARPE tool is used. The results show that the reliability and mean-time-to-security-failure (MTTSF) are improved. The paper will be concluded in Section 5.

2. Related Work

In the following, we briefly review the existing standards, techniques and so on for the security of Web services:

- *Web service security standards:* According to [5], various specifications discussed about Web service security. The WWW Consortium has developed various specifications, such as WS-Security (WSS), WS-Federation, WS-Authorization, WS-Policy, WS-Trust, WS-Authentication and WS-Privacy for Web service security. These standards do not protect Web services totally. For example, WS-Security specifies how integrity and confidentiality can be enforced on messages, allows the communication of security tokens and provide end-to-end security.
- *Vulnerability detection techniques:* There are best practices of software testing and a lot of tools, languages and techniques in order to analyze and detect vulnerabilities in software systems. [5] But, an evaluation of several commercial versions of vulnerabilities scanners showed that these tools are primarily limited to low coverage of existing vulnerabilities and high percentage of false positives. Few techniques and tools (such as Netsparker) exist for vulnerability scanning of SOAP-based web services.
- *Intrusion/prevention techniques:* There are intrusion prevention an intrusion detection techniques, which are not effective against new or unknown attacks. [7]
- *Dependable computing techniques:* The existing solutions for dependable Web services are divided into two categories: fault tolerance techniques (such as active and passive replications), and the use of design diversity. A dependable architecture for Web services that uses multi-version techniques is introduced in [7]. In [3], by using design diversity technique and Web services business process execution language (WS-BPEL), the authors have proposed a useful and flexible architecture for dependable Web services.
- *Software fault-tolerance techniques:* There are two types of software fault-tolerance techniques: single-version and multi-version that are used for security improvement. [8] Fault tolerance techniques, including replication, check-pointing and message logging, in addition to reliable messaging and transaction management for which Web services specifications exist. The authors of [8] have discussed how those techniques can be applied to the components of Web services involved in the business activities to make them dependable.
- *Architectures for intrusion-tolerant systems:* There are important architectures such as self-cleansing intrusion tolerance (SCIT), scalable intrusion-tolerant architecture (SITAR) for distributed services and malicious- and accidental-fault tolerance for Internet applications (MAFTIA) for intrusion-tolerant systems [9]. The assumption in the SCIT architecture is that the intrusion detection mechanism is not able to detect unknown attacks and Web services cleansing is necessary. The SITAR architecture is used for the intrusion tolerance of commercial off-the-shelf (COTS) systems. Fault tolerance techniques, such as redundancy and design diversity, are used in the SITAR architecture.
- *Fault tolerance architecture for Web services:* In [10], authors have proposed a new fault-tolerant architecture for Web services named FTWeb. In [12], authors have explained a multi-layer architecture for ITWSs. The specific goal of the architecture is to use single-version software fault-tolerance concepts in the case of malicious failures.

3. The Proposed Architecture

In this section we introduce an architecture for intrusion-tolerant Web services, which is called *design-diverse intrusion-tolerant web service* (DDITWS). The aim of the proposed architecture is to construct and strengthen the capabilities of Web services against both known and unknown security attacks. The proposed architecture is based on the following concepts and techniques:

I. Theoretical concepts of intrusion tolerance approach: In these concepts, the main indicator of intrusion tolerance and their requirements are expressed. In [12], the main indicators of an intrusion-tolerant system are defined as follows:

- Maintaining the integrity of the system's operational environment,
- Detecting intrusions,
- No failure in the security features of system, such as confidentiality, integrity and availability, and
- Stability in the system's operations.

II. Composite Web service technology: In the composite Web service technology used in the proposed architecture, the aggregation and composition of the main Web service with the supplementary Web services that provide the abilities of intrusion tolerance is performed. The demand of ITWS is a complex request. Composite Web service technology provides the possibility to implement the proposed architecture and the ability to meet complex demands.

III. Classical fault tolerance techniques: Intrusion tolerance and fault tolerance are common principles. Both of them focus on service continuity in abnormal conditions. Fault-tolerant techniques can provide appropriate policy to create a conceptual framework, that theories are developed during intrusion tolerance. To establish appropriate mechanisms for intrusion tolerance in Web services, fault tolerance techniques are used [14]. In the proposed architecture, redundancy, design diversity and replication techniques are used.

3.1 Components of the Architecture

The main motivation of the proposed DDITWS architecture is based on the facts that software systems development and maintenance cannot be without vulnerabilities. Behavior-based intrusion detection systems are not able to provide intrusion-tolerant system. To provide the continuity of services, it is necessary that the impact of attacks be managed. The overall view of the proposed architecture is shown in ("Fig. 1").

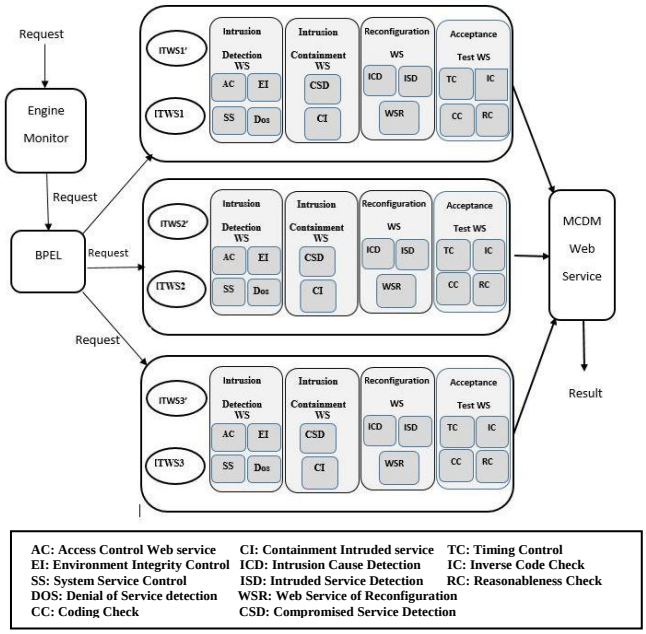


Fig. 1. Overall view of the DDITWS architecture

Intrusion tolerance capabilities through Web services in the DDITWS architecture can be constructed in the forms of features for "intrusion tolerance", "intrusion containment", "reconfiguration", and "decision maker". Appropriate intrusion-tolerant mechanisms causes the continuation of services and preventing security failure in system. Each intrusion tolerance feature in the DDITWS architecture is embodied in the form of a Web service. The composition of these Web services is achieved in the DDITWS architecture. The components of the composite Web service in the DDITWS architecture are explained in the following.

3.2 Intrusion Detection Composite Web Service

The use of resilient mechanisms is necessary for providing the service availability in systems [9]. Resilient mechanisms are meant to have a facility in the event of penetration. Intrusion-tolerant and reconfiguration mechanisms are resilient mechanisms. Intrusion detection methods are possible in knowledge-based or behavior-based. [8]. In the DDITWS architecture, intrusion detection is done by a component of the composite Web service that is based on behavior changes. Intrusion-detection Web service has a main role in providing intrusion tolerance in the DDITWS architecture. Using intrusion detection composition Web service is appropriate for new or unknown attacks.

In the DDITWS architecture, the tasks of the intrusion-detection Web service are as follows:

- Detecting a variety of attacks on Web services among either known or unknown.
- Updating the records of attacks and intrusion patterns.
- Provisioning of the analysis and detection of attacks and system failure causes.
- Acceptance testing on the response of Web service request.

- Identification of the denial-of-service (DoS) attacks on Web service.

According to ("Fig. 1"), intrusion detection composite Web service to perform its tasks includes multiple Web services as follows:

- The Web service of access control to resources (AC): This Web service controls accesses to resources and checks whether it is as expected or not.
- The Web service of environment data integrity control (EI): This Web service by comparing the data files in the operating environment while providing services with the information of data files before providing services, determines whether or not an attack is occurred.
- The Web service of system services controller (SS): This Web service checks certain system services that all of them had already determined, ordered and fully executed.
- The Web service of detect DoS attack: In traditional confronting techniques to DoS attacks, there are two basic steps as: (1) detecting real or fake IP addresses, and (2) detection of traffic conditions for DoS attacks [8]. In the DDITWS architecture, Web service DoS has a task of detecting of the traffic conditions of DoS attacks. The DoS attack occurs in each of the following two modes:

1- *(required-time to respond to a previous request) + (last request-time) > input during Web service request*
2- *Threshold number of requests > [(request-time – arrival time of the first request) / number of requests]*

3.3 Intrusion Containment Composite Web Service

The aim of intrusion containment is the encapsulated area of intrusion and preventing intrusion. This will result in reduce the service level of the system. Intrusion graph is an efficient tool of intrusion containment [15], which is used in DDITWS. Intrusion graph is a directed graph that shows intrusion propagation paths from a service to another service [15]. In intrusion graph, each node represents an intrusion target and any edge that is used to show dependencies between intrusion targets. Intrusion alerts sent by intrusion detection components, are mapped onto the intrusion graph. Intrusion containment components use intrusion graph and breaks through a communication channel section and other sections and prevent the spread of intrusion.

In DDITWS, containment composed Web service has two functions as locating and stopping the spread of intrusion through. To limit the intrusion, it is necessary to prepare the intrusion-graph data. Intrusion containment feature and limiting the intrusion is an important attribute in intrusion-tolerant systems. According to ("Fig. 1"), intrusion containment composite Web service is composed of multiple Web services as follows:

1- The Web service of *compromised service detection* (CSD): This Web service after getting information about intrusion to Web service identifies the compromised service in the attacked Web service.
2- The Web service of *containment intruded service* (CI): After the detection of the compromised service, this Web service gives the information of the compromised service and breaks their communications.

3.4 Recovery and Reconfiguration Composite Web Service

In intrusion-tolerant systems, after intrusion detection and containment, it is necessary that the compromised sections be inactive and the compromised components be reconfigured. In DDITWS, it is the responsibility of the recovery and reconfiguration composite Web service.

In DDITWS, for recovery from intrusion, several techniques, such as fragmentation, scattering and data redundancy can be used. The fragmentation and scattering techniques makes it possible if there is an unauthorized access to the data, all the valuable data should not be available. Using data redundancy technique makes it possible that after intrusion detection, intrusion masking is possible and the system returns to an optimal state. In DDITWS, for reconfigure a compromise component, the level of active service on the Web service is checked. If the service level is not satisfactory, the redundant service will be replaced and the compromised service will be reconfigured. Until replacing, service will be in the *graceful degradation* state. Intrusion recovery Web service and the compromised component reconfiguration are among intrusion-tolerant system attributes. Intrusion containment composite Web service is very important in DDITWS. It is necessary that reconfiguration and intrusion recovery be performed automatically. If reconfiguration is not automatic, the occurrence of distributed DoS (DDoS) attacks is prohibited. The recovery and reconfiguration composite Web service is a main component in DDITWS that is a key component of intrusion tolerance.

According to ("Fig. 1"), the reconfiguration composite Web service is composed of multiple Web services as follows:

- The Web service of *intrusion cause detection* (ICD): This Web service determines the main cause of the intrusion to the Web service.
- The Web service of *intruded service detection* (ISD): This Web service gets the information about intrusion and determines the intruded service.
- The Web service of *reconfiguration* (WSR): This Web service is responsible for an important task, that is, after intrusion detection and containment, it is necessary that the intrusion should be covered and the compromised service is managed by using replication technique.

3.5 Acceptance Test Composite Web Service

The acceptance test composite Web service checks the response of Web service requests. According to ("Fig. 1"), the acceptance test composite Web service is composed of multiple Web services as follows:

- The Web service of *timing control* (TC): This Web service checks whether a Web service deadline is expired or not.
- The Web service of *inverse code check* (IC): This Web service checks the correctness of the response to the Web service's request. If the answer is incorrect, then attack to the Web service is announced.
- The Web service of *control results* (RC): This Web service checks whether the results of the Web service is within the acceptance range or not.
- The Web service of *coding check* (CC): This Web service checks the validity of data transfer operations. This Web service uses encoding data technique.

3.6 Multi-Criteria Decision Maker Web Service

Based on the DDITWS architecture, for each request, three redundant Web services are provided and in this case, to determine the final outcome, using a decision maker Web service is necessary. ("Fig. 2") shows the structure of the multi-criteria decision maker composite Web service.

In multi-criteria decision maker composite Web service, for organizing redundant voters and acceptance monitors, N-self checking technique is used. Based on N-self checking technique, at any time, only one voter is enabled. There are three voters in decision maker composite Web service. All voters in decision maker composite Web service have an equal number of inputs, output type and input types. In multi-criteria decision maker composite Web service, in addition to the input values, also Web service trust value are part of inputs and the end result is effective.

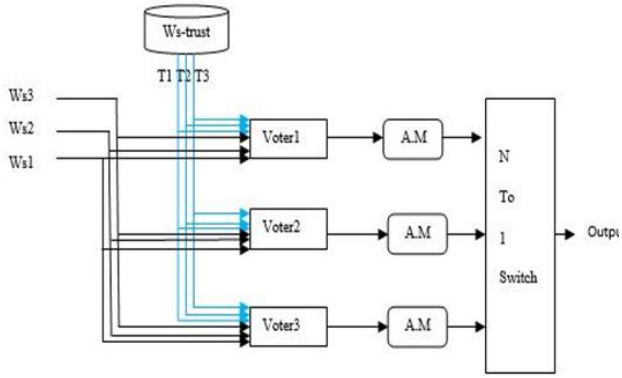


Fig. 2. The structure of the multi-criteria decision maker

3.7 Structural Features of DDITWS

The structural features of the DDITWS architecture are as follows:

1- Using redundancy technique in system causes the enhancement of the system reliability. In the proposed architecture to respond to any requests, three redundant Web services are used.
2- In order to reduce the probability of similar vulnerabilities in redundant Web services (i.e., *ITWS1*, *ITWS2* and *ITWS3*) and increasing the tolerance against similar attacks, the design diversity technology is used.
3- Any of the main Web services (i.e., *ITWS1*, *ITWS2* and *ITWS3*) along with several other Web services that provide the intrusion tolerance, are constructed as composite Web services.
4- Using replicated Web service technique, recovery from critical Web services is possible.
5- The recovery strategy is used in the proposed architecture based on using intrusion masking and replication techniques.
6- In each of the main Web services (i.e., ITWS1, ITWS2 and ITWS3), intrusion containment and reconfiguration composite Web service are used. Their main tasks include intrusion containment and the influence of recovery and reconfiguration of compromised component.
7- Using redundant Web services in the proposed architecture, shows the necessity of using the decision maker Web service within it.

3.7.1 Relationship between Intrusion Tolerance and Design Diversity Technique in DDITWS

Intrusion tolerance means that service continues in the presence of attacks. Intrusion tolerance is a non-functional requirement in systems. Having an intrusion tolerant Web service is a complicated demand. The intrusion tolerance capability of the DDITWS architecture, as several subsidiary Web services is combined with a main Web service. To achieve complete intrusion tolerance, it should be combined the intrusion avoidance and intrusion tolerance capabilities. The DDITWS architecture of intrusion prevention capabilities uses redundancy and design diversity techniques. The use of design diversity technique in the development of a variety of components reduces the same vulnerabilities in the components. Reducing the same vulnerability of the redundant components, similar attacks are successfully reduced DDITWS. Similarly, using design diversity technique in decision-maker Web service reduces the vulnerability of voters. The fundamental role of using design diversity technique is to increase the intrusion tolerance in the DDITWS-based Web services.

3.7.2 Structure of the DDITWS in BPEL

The structure of composite Web service involves all internal Web services, the order of the execution of the internal Web services and data transferring between them [16]. For each composite Web service, defining specific rules on the application level means the determination of composite Web service structure [16]. The BPEL4WS by defining specific rules on the application level specifies Web services participating in composite Web service, the order of the execution of them and data transferring between internal Web services.

There are different tools in design area of BPEL that they may make the use of the composition and execution of Web services in composite Web services of the DDITWS

architecture. Designing of ITWS can be done in the BPEL environment. (“Fig. 3”) shows the structure of ITWS based on the DDITWS architecture in the BPEL form.

As in the DDITWS architecture, each composite Web service involves several internal Web services as shown in (“Fig. 4”). Internal Web services are organized by the structures such as “sequence”, “flow”, “scope” and so on of BPEL.

4. Modeling and Evaluation

4.1 The Security Behavior of DDITWS

For modeling the security behavior of the DDITWS architecture, modeling the behavior of the “attacker” and “system response to attack” are necessary. ITWS’s security behavior can be modeled by using the definition of security states, the interaction of them and the state-transition diagram (STD). (“Fig.5”) shows the STD of Web service’s behavior in the DDITWS architecture. Successful exploitation of security holes by attacker is a main factor causing active attacks occurs.

In the DDITWS architecture, several strategies are intended to create different level of security. According to (“Fig. 5”), in the security behavior model of the Web service in DDITWS architecture, at the beginning, the Web service is in “good” state. Create new conditions such as change of Web service information, change user permissions, prolonged duration of services, change in accounting rules and not properly performed system services, causes Web service state change into “vulnerable” state.

Fig. 3. The structure of the ITWS in BPEL

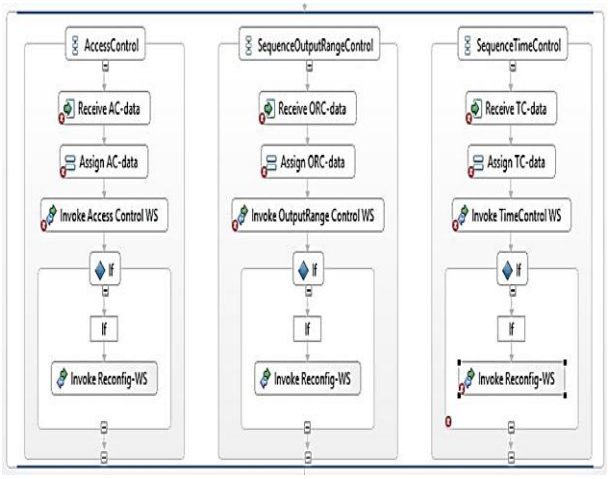


Fig. 4. Instance of internal Web services in DDITWS

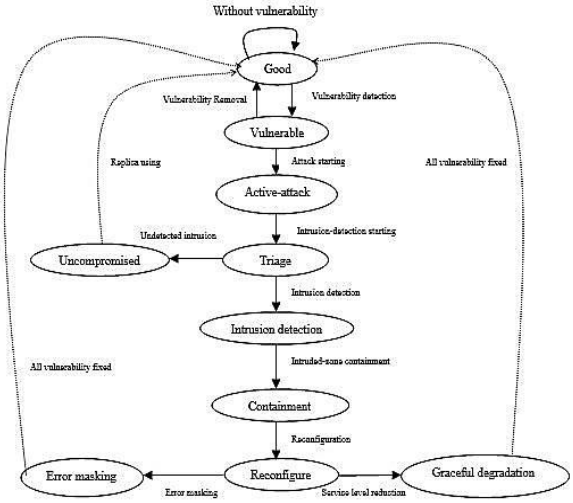


Fig. 5. The state-transition diagram of DDITWS

The vulnerability can be identified as a security hole. Make any sense of vulnerability in the Web service means a violation of Web service’s security policy. In fact, after vulnerability identification, it may be exploited be attackers.

The realization of the possibility of exploiting security holes; it will put Web service in “active attack” state. With each entry into active attack mode, the Web services become influential. In vulnerable mode, using tools such as firewall, may lead to identifying and eliminating the vulnerabilities and Web services can still be in “good” mode. If the error has not been covered in Web service and the compromised component fails, the Web service will be in “uncompromised” mode.

In the DDITWS architecture, intrusion recovery strategy uses intrusion masking and replication techniques. Based on the reinforced intrusion recovery strategy, the data used intrusion masking technique involves data fragmentation, data scattering and redundancy against intrusion. Critical services’ recovery is mostly done through replication technique. Using data redundancy technique causes Web service to be in “good” mode and the service delivery of the Web service will be continued.

According to (“Fig. 5”), if the intrusion is detected in the use mode, Web service goes to “triage” mode. In this context, the following two conditions are most likely:

1- A compromization has taken place and the intrusion detection mechanism cannot detect that, so it is in “uncompromised” state.
2- The intrusion detection component successfully detects intrusion, so Web service is in ‘‘intrusion detection” mode.

After intrusion detection, appropriate message will be sent to intrusion recovery and compromised component reconfiguration, messages received through intrusion detection component, Web service to deliver ‘‘intrusion containment” mode for limiting the restricted area and prevent intrusion expansion.

Another security state is the “reconfiguration” state. In the reconfiguration state, reconfiguration process is based on the defined policies. Also, by using data redundancy and replication techniques, Web services are in “masked-error” state. The reconfiguration mode may lead to new security mode that is named “graceful degradation”. In this case, only essential services will continue and other services will be stopped.

Essential services are the services that continue and will not stop in system, even in intrusion mode. If all strategies are predicted to fail, Web service mode will be the “failed” mode. This condition should not occur in ITWS. The Web service that returns to normal mode after a successful attack is shown in Web service security behavior model as dashed-line.

4.2 The DDITWS Architecture Modeling and Formal Analysis Using CPN Tools

For the analysis of the functionality of the proposed architecture, we have modeled it using coloured Petri nets (CPNs or CP-nets) and then, the behavioral characteristics are analyzed. For this purpose, we have used CPN Tools.

4.3 Modeling and Analysis of the Architecture

The main model (i.e., the home page in CPN Tools terminology) of the DDITWS architecture is shown in (“Fig. 6). Because coloured Petri nets are graphical models, this feature provides the opportunity to review the changes and it can help to investigate how each of the sections in the system works. [17]

The home page of the model consists the units named *intrusion detection, intrusion containment, intrusion recovery, reconfiguration, and decision maker* that each of these units are modeled by substitution transitions in the model.

Places in the model are ports for inputs and outputs. In home page, the *start* place is a driver for input Web service request. There is a token inside it, causes a *Get_WS_Access_Archive* transition be enabled.

The incoming Web service request is transmitted to the place named WS_*Access_Archive*. By placing a token in the *WS_Access_Archive* place, the *dispatcher* transition sets *WS1*, *WS2* and *WS3* places simultaneously.

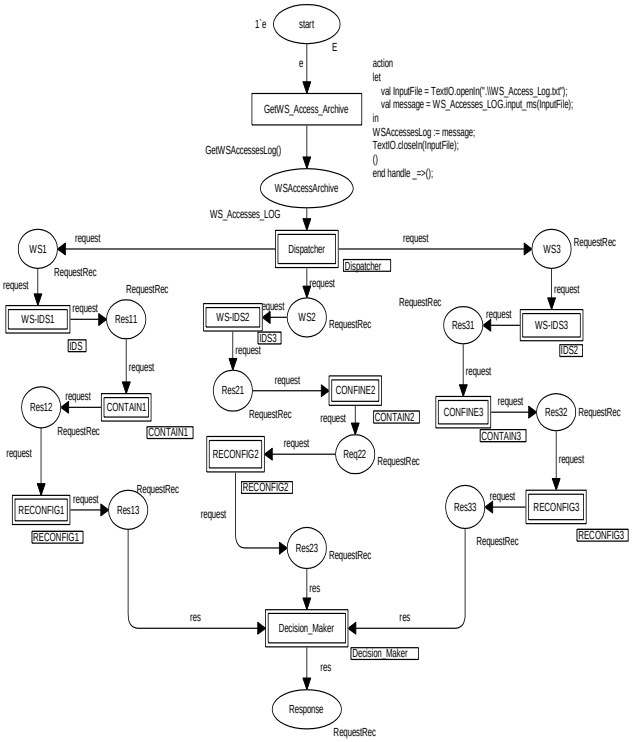


Fig. 6. The home page of the DDITWS CPN model

By submitting to three redundant of Web services at the same time, each of the *IDS1*, *IDS2* and *IDS3* transitions has the responsibility for the result of the corresponding Web service. The *IDS1*, *IDS2* and *IDS3* transitions have the duty to compare the behavior of the system in service with the expected behavior. Each of the *IDS1*, *IDS2* and *IDS3* transitions have equivalent intrusion detection units in the DDITWS model. In case of intrusion detection, the corresponding intrusion containment units (i.e., *CONFINE1*, *CONFINE2* and *CONFINE3*) are active.

Each of the intrusion containment units by limiting the infected area, prevents the spread of influence and sends the intrusion messages to correspond reconfiguration transition (i.e., *CONFIG1*, *CONFIG2* or *CONFIG3*).

Based on the determined tasks for the reconfiguration unit, intrusion recovery and compromised component reconfiguration are done. At the end, the results of Web services, in order to determine the final result, will be sent to the decision maker unit. The decision maker unit is an intrusion-tolerant composite Web service that determines the final result of the redundant Web services.

4.4 Properties of the Architecture

In the CPN Tools environment, it is possible that measuring the behavioral characteristics of the specified model. [19] The results of formal analysis of the behavioral characteristics in DDITWS are shown in (“Fig. 7), which are as follows:

- “None” value for the *Dead Marking* attribute shows that each transition in the proposed model

are live and the system of the DDITWS architecture is deadlock-free.
- "None" value for the *Dead Transition Instances* attribute shows that the system based on DDITWS is non-terminating in all states.
- The value "136" for the *Live Transition Instances* attribute shows that the modeled system is live.
- The value "No infinite occurrence sequences" for *the Fairness Properties* shows that all transitions are executed fairly.
- The value "No infinite occurrence sequences" for the *Fairness* attribute in system model shows that the system will execute in all possible states.

The behavioral characteristics of the model ensure the correctness of the functionality of the architecture.

State Space	**Liveness Properties**
Nodes: 3672	Dead Markings
Arcs: 15840	*None*
Secs: 9	Dead Transition Instances
Status: *Full*	*None*
SCC Graph	Live Transition Instances
Nodes: 3672	136
Arcs: 15840	Fairness Properties
Secs: 0	*No infinite occurrence sequences*.

Fig. 7 The results of the analysis of the characteristics of DDITWS

4.5 Evaluation of the Measures

Two measures, i.e., "reliability" and "mean-time-to-security-failure", are important in the Web services based on the DDITWS architecture. We examine these measures in this section.

4.5.1 Evaluation of the Reliability Measure

The reliability of a Web service based on the DDITWS architecture may be evaluated without their implementation details using a stochastic Petri net (SPN) model and the SHARPE modeling tool. For the evaluation of DDITWS's reliability block diagram is shown in ("Fig. 8"), which includes the following:
- The proposed architecture consists of three redundant Web services.
- In construction of redundant Web services in DDITWS architecture, design diversity technique is used, so the reliability of the replications is different. Using design diversity technique reduces the same vulnerabilities in the replicated Web services.
- To determine the final result of the redundant Web services in the DDITWS architecture, a decision maker unit is used. The decision maker unit uses three redundant voters on an N-self checking structure. In the construction of redundant voters, design diversity technique is also used, so the reliability of these redundant voters will also be different.

By Eq. (1), the reliability of the DDITWS architecture can be calculated as follows:

$$R_{Web\ services} = R_{ws1} * R_{ws2} * R_{ws3} - (R_{ws1}.R_{ws2}) - (R_{ws1}.R_{ws3}) - (R_{ws2}.R_{ws3}) + R_{ws1} + R_{ws2} + R_{ws3}$$
$$R_{voters} = R_{voter1} * R_{voter2} * R_{voter3} - (R_{voter1}.\ R_{voter2}) - (R_{voter1}\ .R_{voter3}) - (R_{voter2}.\ R_{voter3}) + Rvoter1 + R_{voter2} + R_{voter3} \quad (1)$$
$$R_{system} = R_{Web\ services} * R_{voters}$$

For evaluating the reliability of DDITWS, experimental values of the reliability of each component is given in Table 1.

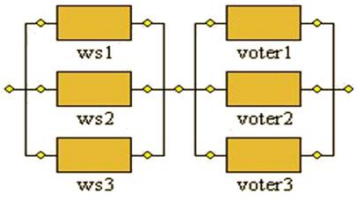


Fig. 8. Block diagram of the DDITWS architecture

Table 1. Experimental values of the reliability of each components

Bock-name	**Reliability**
Web service1	0.81
Web service2	0.78
Web service3	0.82
Voter1	0.91
Voter2	0.93
Voter3	0.95

As shown in ("Fig. 9), the reliability of each Web service is less than the reliability of ITWS. The reason is due to using the design diversity technique in the DDITWS architecture. By repeating the experiment with new values for Web service's reliability, the overall result will not change. The overall result is the reliability of ITWS greater than the reliability of each Web service. As expected, this reliability evaluation has assured that the security of the DDITWS architecture is increased.

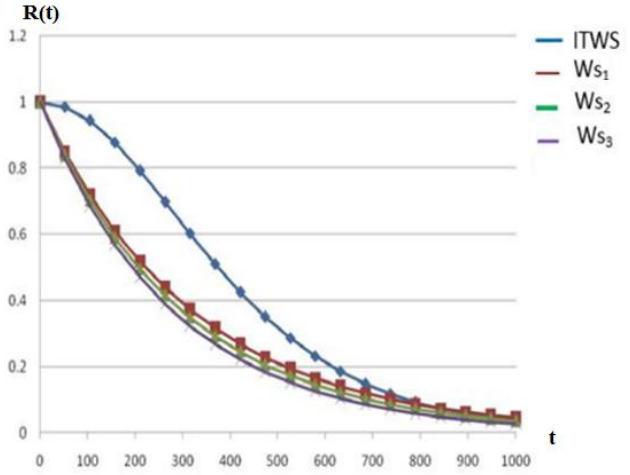


Fig. 9. The diagram comparing the reliability of three Web services with the ITWS

4.5.2 Evaluation of the MTTSF Measure

Mean-time-to-security-failure (MTTSF) is an important measure in the survivability evaluation of intrusion-tolerant systems. The SPN model of the

DDITWS is shown in (“Fig. 10”). The corresponding Markov model is also shown in (“Fig. 11”).

The evaluation of the system performance is often related to their behavior after long time, until a system steady state is achieved. In system steady state, the impact of initial conditions and system behavior into a state regulated system of compensation. In (“Fig. 11), the transfer rate from the BPEL state to the execute state of each of the Web services (i.e., WS1, WS2 and WS3) are equal.

(“Table 2”), shows the results of the solution of the model using SHARPE tool. This table shows the selected transition rates used in the model, too.

In (“Fig. 12”), the MTTSF of a traditional Web service is compared with DDITWS.

In explaining the attributes of the DDITWS architecture, it was said that using design diversity technique in developing redundant components causes reducing the Web service common vulnerabilities and increasing the Web service intrusion tolerance. The higher level of design diversity used, the lower the failure rate of the redundant Web services. This case causes increasing the Web service MTTSF. Increasing the MTTSF in Web services based on the proposed architecture is a factor to increase the availability and intrusion tolerance of the Web services based on the architecture.

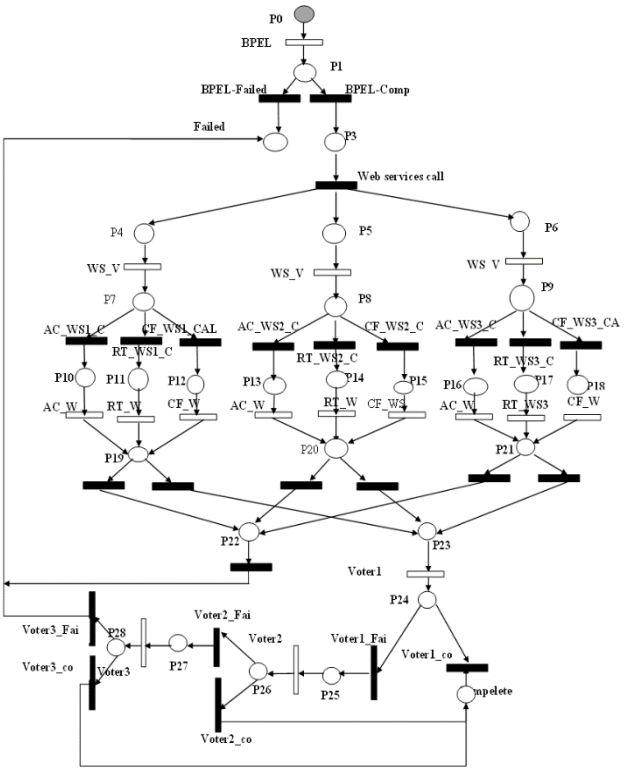


Fig. 10. The SPN model of the ITWS

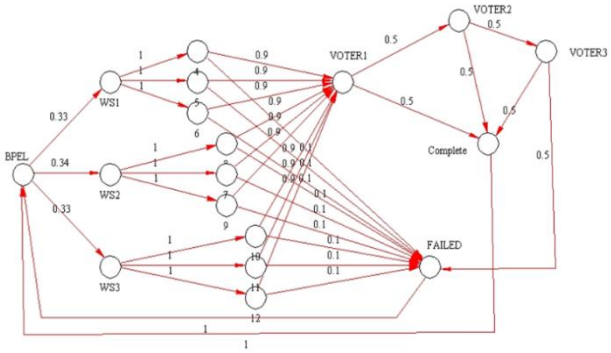


Fig. 11. The Markov chain model corresponding to the SPN model of Fig. 10.

Table 2. Transition rates and calculated MTTSF measure

IN1	IN2	MTTSF in DDITWS	MTTSF in DDITWS
0.1	0.9	39.1	29.0
0.2	0.8	20.1	14.0
0.3	0.7	13.7	9.0
0.4	0.6	10.5	6.5
0.5	0.5	8.6	5.0
0.6	0.4	7.4	4.0
0.7	0.3	6.5	3.3
0.8	0.2	5.8	2.8
0.9	0.1	5.2	2.3

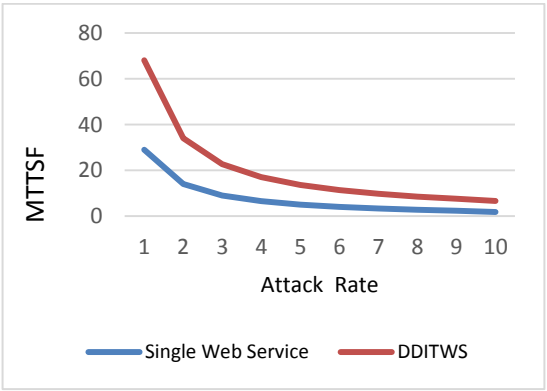


Fig. 12. The MTTSF values of the DDITWS compared with a traditional Web service

5. Conclusions

In this article, we presented a new architecture for intrusion-tolerant Web services (ITWSs). The proposed architecture, abbreviated by DDITWS, uses classical fault-tolerance techniques, such as design diversity, redundancy, N-self checking and acceptance testing. Also, the approach uses the theoretical concepts of intrusion-tolerant systems, which are used in the proposed architecture. In the proposed architecture, composite Web service technique is used, with several Web services. Composite Web service structure consists all internal Web services, their execution order and how to the data is transferred between them in the application level.

In order to understand the security behavior of Web services in the DDITWS architecture, the interaction is evaluated against the attempts of attackers.

To study the components functionality, Web services in the DDITWS architecture are modeled using coloured Petri nets. Behavioral characteristics of the proposed architecture are also analyzed. The results show that the functionality of all components is correct.

The mean-time-to-security-failure (MTTSF) and the reliability measure for the proposed architecture are evaluated using stochastic Petri nets and the SHARPE tool. Evaluation results show that the reliability and MTTSF of the proposed architecture has also increased.

The proposed architecture is a complex one and is dedicated to SOAP-based Web services. There are multiple components, i.e., subsidiary Web service, in the proposed architecture. These multiple Web services should be designed based on design diversity rules and techniques. In practice, achieving diverse versions for the same software is quite difficult. Therefore, this is the main disadvantage of the proposed architecture.

The proposed architecture may be used for other types of Web services, such as RESTful Web services. It can also be used to devise intrusion-tolerant architecture for other types of software systems.

In future, we intend to implement a prototype of a real Web service, such as an electronic commerce application. This prototype implementation can be used to evaluate other aspects of the proposed architecture.

References

[1] E. Dunrova, Fault Tolerant Design: An Introduction, Department of Microelectronics and Information Technology, Royal Institute of Technology, Stockholm, Sweden, 2008.

[2] E. Cerami, Web Services Essentials, First Edition ed., United States of America: O'Reilly, 2002.

[3] D. Gorton, Extending Intrusion Detection with Alert Correlation and Intrusion Tolerance, M.S.Thesis, Chalers Universityof Technology, Sweden, 2003.

[4] D. Gorton, "Using WS-BPEL to Implement Software Fault Tolerance for Web services," in Proceedings of the 32nd EUROMICRO Conference on Software Engineering and Advanced Applications, 2006, pp. 126-133.

[5] E. Martin and M. Salas, "Security Testing Methedology for Vulnerabilities of XSS in Web Services and WS-Security," *Electronic Notes in Theoritical Computer Science*, Vol. 302, 2014, pp. 133-154.

[6] J. Reynolds, "The Design and Implementation of an Intrusion Tolerant System," in Proceedings of the International Conference on Dependable Systems and Networks, 2002, pp. 285-290.

[7] M. Abdollahi Azgomi and E. Nourani, "A Dependable Web Service Architecture Based on Design Diversity Techniques and WS-BPEL," Iranian Journal of Electrical and Computer Engineering (IJECE), Vol. 11, No. 1, 2013, pp. 1-4.

[8] P. Veríssimo, N. Neves and M. Pupo Correia, "Intrusion-Tolerant Architectures: Concepts and Design," Lecture Notes in Computer Science, Vol. 2677, 2003, pp. 3-36.

[9] L. Quyen, S. Nguyen and S. Aurn, "Comparative Analysis of Intrusion-Tolerant System Architectures," IEEE Security and Privacy, Vol. 9, No. 4, 2011, pp. 24-31.

[10] T. Giuliana Santos, L. Cheuk Lung and C. Monetez, "FTWeb: A Fault Tolerant Infrastructure for Web Services," in Proceedings of the Ninth IEEE International EDOC Enterprise Computing Conference, 2005, pp. 95-105.

[11] Z. Aghajani and M. Abdollahi Azgomi, "A Multi-Layer Architecture for Intrusion-Tolerant Web Services," International Journal of u- and e-Service, Science and Technology, Vol. 1, No. 1, 2008, pp. 73-80. .

[12] A. Sood, "Securing Web Servers Using Intrusion Tolerance (SCIT)," in Proceedings of the Second International Conference in Dependability, 2009, pp. 60-65.

[13] W. Barry Johnson, Design and Analysis of Fault-Tolerant Digital Systems, University of Virginia: Charlottesville: Addison-Wesley Publishing Company, 1989.

[14] W. Yu-Sung, F. Bingrui, Y.-C. Mao, B. Saurabh and S. Eugene," Automated Adaptive Intrusion Containment in Systems of Interactive Systems," Computer Networks, Vol. 51, 2007, pp. 1334-1360.

[15] L. Chen, "A Method for Analyzing and Predicting Reliability of BPEL Process," Journal of Software, Vol. 4, No. 1, 2009, pp. 11-18.

[16] D. Mukherjec, P. Jalote and M. Gowri Nada, "Determing QoS of WS-BPEL compositions," Lecture Notes In Computer Science, Vol. 5364, 2008, pp. 378-393.

[17] "CPN Tools," CPN Group, University of Aarhus, [Online]. Available: http://wiki.daimi.ac.dk/cpntools.

[18] K. Jensen and L. M. Kristensen, Coloured Petri Nets, Modeling Validation of Concurrent Systems, Springer, 2009.

Sadegh Bejani received B.Sc. in Computer Engineering (Hardware) (1985) from Tehran University and M.Sc. and Ph.D. degrees in Computer Engineering (Software) (1996 and 2015, respectively) from Imam Hossein (a.s.) University. His research interests include software security, intrusion-tolerant systems, analytical modeling and computer simulation. He is an Assistant Professor at School of Information and Communication Technology, Imam Hossein (a.s.) University, Tehran, Iran.

Mohammad Abdollahi Azgomi received B.Sc., M.Sc. and Ph.D. degrees in Computer Engineering (Software) (1991, 1996 and 2005, respectively) from Department of Computer Engineering, Sharif University of Technology, Tehran, Iran. His research interests include modelling and evaluation of security, privacy and trust, and dependable and secure software development. He is an Associate Professor at School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran.

Automatic Construction of Domain Ontology Using Wikipedia and Enhancing it by Google Search Engine

Sedigheh Khalatbari
Department of Computer Engineering, University of Guilan, Rasht, Iran
khalatbari@msc.guilan.ac.ir
Seyed Abolghasem Mirroshandel*
Department of Computer Engineering, University of Guilan, Rasht, Iran
asedghasem@yahoo.com

Received: 04/May/2015 Revised: 20/Nov/2015 Accepted: 08/Dec/2015

Abstract

Information and resources available on the Web are growing increasingly and web users need to have a common understanding of them. The Semantic Web whose most important role is to help machine to understand and analyze the existing data on the Web, has not been used commonly, yet. The foundation of the Semantic Web are ontologies. Ontologies play the main role in the exchange of information and development of the Lexical Web to the Semantic Web. Manual construction of ontologies is time-consuming, expensive, and dependent on the knowledge of domain engineers. Also, Ontologies that have been extracted automatically from corpus on the Web might have incomplete information. The main objective of this study is describing a method to improve and expand the information of the ontologies. Therefore, this study first discusses the automatic construction of prototype ontology in animals' domain from Wikipedia and then a method is presented to improve the built ontology. The proposed method of improving ontology expands ontology concepts through Bootstrapping methods using a set of concepts and relations in initial ontology and with the help of the Google search engine. A confidence measure was considered to choose the best option from the returned results by Google. Finally, the experiments showed the information that was obtained using the proposed method is twice more accurate than the information that was obtained at the stage of automatic construction of ontology from Wikipedia.

Keywords: Ontology; Improvement and Development of Ontology; Bootstrapping Method; Google Search Engine; Wikipedia.

1. Introduction

Nowadays, the Web is considered a live entity that is growing and evolving fast over time. The amount of content stored and shared on the web is increasing quickly and continuously. Problems and difficulties such as finding and properly managing all the existing amount of information, arise as a consequence of this extensive development. To overcome such limitations the only possible way is to promote the use of Semantic Web techniques (Cantador et al. 2007). Ontologies are the basis and foundation of the Semantic Web. Ontology is a conceptual model which formally and explicitly simulates actual entities and the relations among them in a particular domain (Gruber 1993; Staab and Studer 2004).

Ontologies have been useful in lots of applications such as knowledge management, information retrieval, and question answering systems. They are considered as the basis and foundation of many new intelligent systems. Manual ontology construction is very costly, tedious, and error-prone. They also suffer from rapid aging and low coverage. The manual construction of ontology needs a lot of experts in particular domain and many annotators must work together for a long time (ShamsFard and AbdollahZade 2002). A few ontologies have been built manually the most famous of which are WordNet (Fellbaum 1998), Cyc (Lenat 1995) and Gene Ontology (GOC[1] 2000). Consequently, in recent years, one of the main challenges for researchers has been the automatic construction of Ontology. One of the main problems of automatic ontology construction is the incompleteness of information required to construct that ontology; as the web corpora from which ontology is extracted, do not contain all information related to the given domain of ontology. In addition, during the automatically ontology construction process, certainly not all information can be fully extracted from web corpora.

The aim of this study is to provide a strategy for development of the ontologies. Therefore, this study first discusses the automatic construction of a prototype ontology in animal domain with the help of articles in Wikipedia. Hence, consequently an ontology is generated automatically with the use of semantic relations obtained in the structure of Wikipedia template pages, Infoboxes, and their hierarchical categories. Next, a Bootstrapping method is proposed to improve the constructed ontology and complete its information using the extracted information in ontology. Our method can automatically extract new information and extend the initial ontology with the help of Google search engine.

The rest of this paper is as follows: in section 2, the related work is described. In section 3 and 4, the technique

[1] Gene Ontology Consortium

* Corresponding Author

for the automatic construction of prototype in Persian ontology will be discussed and also the proposed solution to improve the ontology constructed with the use of Bootstrapping techniques will be explained. In addition, in these sections, experiments carried out to evaluate the proposed method will be described. In section 5, evaluation of proposed method has been presented and finally, section 6 draws conclusions and offers some solutions for future work.

2. Related Work

In this section, researches on automatic ontology construction, extraction concepts based on predefined patterns, ways of developing concepts of a collection based on Bootstrapping techniques, and semi-supervised solutions are briefly presented.

2.1 Automatic Ontology Construction

Kylin system (Wu and Weld 2007) is a self-supervised learning system whose main idea is automatic construction of ontology using Infoboxes in Wikipedia pages and then creating Infoboxes for all Wikipedia articles. Another purpose of Kylin is automatic production of links to Wikipedia articles.

KOG system (Wu and Weld 2008) is an autonomous system for creating a rich ontology from Wikipedia pages. It uses statistical-relational learning techniques for combining Wikipedia Infoboxes with WordNet. It also uses Markov Logic Network (MLN) (Richardson and Domingos 2006) and the proposed solution "joint-inference" to predict Subsumption relationships between Infobox classes; while simultaneously mapped the classes to WordNet nodes. As a result, the constructed ontology contains Subsumption relations and mappings between Wikipedia's Infobox classes to WordNet.

YAGO (Suchanek et al. 2008) is a high quality ontology with a high coverage that consists of 1 million entities and 5 million facts. YAGO system combines category labels and Infoboxes in Wikipedia pages with WordNet nodes and in this way, a wide ontology is created automatically by using heuristic methods and rule-based techniques.

Another automatic ontology extension method were proposed based on supervised learning and text clustering. This method uses the K-means clustering algorithm to separate the domain knowledge, and to guide the creation of training set for Naïve Bayes classifier (Song and et al 2014).

Sanabila et al. automatically built a wayang ontology from free text. The information or knowledge that is contained within the text is extracted by employing relation extraction. This method was extracted instance candidates that were subsequently clustered using relation clustering (Sanabila and Manurung, 2014).

In other paper, an automatic approach was proposed based on Ontology Learning and Natural Language Processing for automatic construction of expressive Ontologies, specifically in OWL DL with ALC (Horrocks et al., 2007) expressivity, from a natural language text. The viability of their approach is demonstrated through the generation of complex axioms descriptions from concepts defined by users and glossaries found at Wikipedia (Azevedo et al. 2014).

2.2 Extracting Concepts Based on Pattern

Marti A. Hearst used Lexico-Syntactic patterns to extract Hyponyms relationships in natural language (Heast 1992). In this method, first, some pre-defined patterns by humans were considered. Next, by matching these patterns, the concepts and relations among them were extracted.

In another approach, a category system with large scales was created from category labels in Wikipedia pages. In order to find the "is-a" relations among category labels, methods based on connectivity in the network and Lexico-Syntactic Matching (Ponzetto and Strube 2007) were used.

Rion Snow and his colleagues proposed a new algorithm for automatic learning of hyponym (is-a) relations from text (Snow et al. 2005). Their main goal was automatic detection of Lexico-Syntactic patterns. First, they extracted concepts in the text using a small collection of manually defined patterns with regular phrases. Then, using dependency path feature obtained from parse tree they presented a public and all-purpose formula for these patterns. The proposed algorithm can automatically extracts the useful dependency paths and use them for other texts as well as for detection of new hyponym pairs.

In Sprat (Maynard et al. 2009) and SOFIE (Suchanek et al. 2009), a collection of concepts and relations was extracted from Wikipedia texts using rule-based techniques and with the help of some pre-defined patterns.

In another study, a semi-automatic approach is presented to build an ontology for the domain of wind energy which is an important type of renewable energy with a growing share in electricity generation all over the world. Related Wikipedia articles are first processed in an automated manner to determine the basic concepts of the domain together with their properties. Next the concepts, properties, and relationships are organized to arrive at the ultimate ontology (Küçük and Arslan, 2014).

Xiong et al. (Xiong et al. 2014) presented a semi-automatic ontology building method to build marine organism ontology used the role theory to describe the relations among marine organisms. After the realization of the ontology concept and relation extraction using ontology learning technology, a manual review, screening and proofing, then the ontology editor by using Hozo is required (Kozaki et al. 2002).

2.3 Extending the Concepts of a Collection

DIPRE system is a bootstrapping system that uses (Author, Book) pairs to extract structured relations from a large collection of web documents about books and authors (Brin 1998). In this approach, using five initial data as seeds, requests are sent to Google search engine and then the results are examined. The patterns consist of author and book pairs. Using the detected patterns and sending new requests to Google search engine, more information is extracted. Finally, the process of search and finding patterns are repeated and the data set is extended in this way.

Snowball is another system that includes a new strategy for producing patterns and extracting multi entities from plain-text documents whose main idea is similar to DIRPE (Agichtein and Gravano 2000). Actually, by developing key factors of DIRPE solution, the quality of obtained patterns without intervention of humans is calculated by the Snowball system in each iteration of the extraction process and better patterns are used in the next iterations.

SRES has a more complex model than DIRPE and Snowball (Rozenfeld and Feldman 2008). SRES system, in addition to using simple patterns for extracting relations, can also use more general patterns which have been defined in KnowItAll (Etzioni et al. 2005).

In another research, a distant-supervision system was proposed that uses the large semantic web database Freebase (Boolacker et al. 2008) as the seed and extracts new entities (Mintz et al. 2009). Each sentence used as the seed consists of a pair of entities which have participated in a relation in Freebase.

Carlson and his colleagues proposed a semi-supervised learning method for extracting information (Carlson et al. 2010). The main purpose of this method to extract new instances from the concept category and the relations among them using an initial ontology.

Another semi-supervised bootstrapping categorization method were used for retrieving the images related to medical terms from web documents (Chen et al. 2012). This method starts with a positive image for each term as a seed and continues the search process in an iterative way. New images extracted are also used in the next search process as the seed.

Yao et al. converted web data into semantic web descriptions that uses key-value pairs in JSON objects (Crockford 2006). Meanwhile, it builds semantic models for data instances, which can be applied to further semantic reasoning applications. Their used this method to extract schemaless JSON data automatically, including concepts, properties, constrains and values, and build semantic ontology to describe the metadata and instances (Yao et al. 2014).

3. Proposed Method

In this paper, first a method is presented for the automatic construction of prototype ontology using the structures of Persian Wikipedia pages. Since the ontology may contain incomplete information, another method is presented for solving this problem which can generally be used for improving and extending all types of ontologies. Figure 1 shows an overview of the proposed method for the ontology construction. The part above the dotted line shows information extraction process from Wikipedia and the automatic construction of prototype ontology using the existing structures in Wikipedia pages. The part below the dotted line shows the process for improving the constructed ontology using Google search engine.

The extraction method from Wikipedia and the automatic construction of prototype ontology are explained in the following subsections in more detail:

3.1 Proposed Ontology Construction Method

In this study, in order to construct the prototype ontology, information in Infoboxes and Navboxes in Wikipedia pages is used to extract the triple of facts (Extracting concepts). The information in the Navbox is used to extract category hierarchy between the given entities (Extracting relations). The various parts of Wikipedia are displayed in Figure 2. Wikipedia is a Web encyclopedia whose updated information can be accessed by users in different languages. Wikipedia articles are graphical in which related pages are linked to each other.

3.1.1 First Step: Wikipedia Pages Crawler

Pages relating to the fauna from the Persian part of Wikipedia were collected by this unit according to the predefined domain. The crawler acts in a way that at the first step receives an address as the starting page, then through the links on the page, collects further pages. At this stage, approximately 3,200 pages have been identified and saved by the crawler. This collection also consisted of unrelated pages, too.

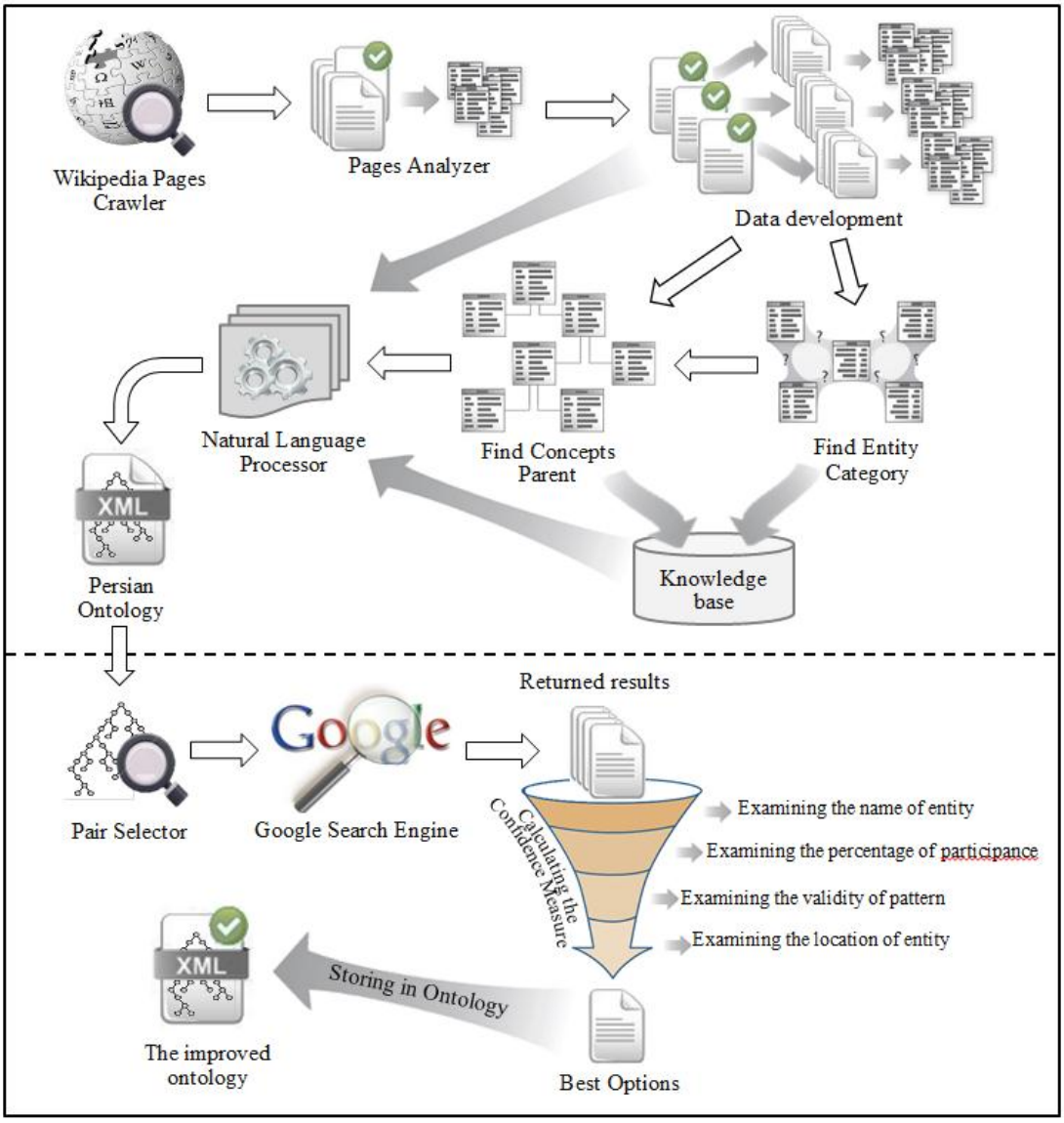


Fig. 1. The process of domain ontology construction using the existing structure in Wikipedia pages

(a) (b)

Fig. 2. (a) Navbox and (b) Infobox sample

3.1.2 Second Step: Wikipedia Pages Analyzer

This unit first examines the content of collected pages by Wikipedia crawler and then separates the template pages among them. Template pages are pages whose titles start with the word "olgô: / template:" and contain Navbox. To identify template pages relating to the fauna, page categories was also used. The page analyzer then extracts existing data in Navboxes. At this point, the proposed system from 3,200 extracted pages by the crawler have identified 11 template pages tailored to specify different conditions. Among the 11 existing Navboxes on these pages, 1,039 entities were also extracted.

Moreover, due to the existence of duplicate entities in different Navboxes and in order to achieve the best results and avoid ambiguity in the next steps, the entity tooltips were also considered in a way that if the entity include more clear, they can be used. For example in the Persian Wikipedia pages, there is the word "râh-râh / stripes" in both Navboxes relating to the sub-families of the "h æ vâsiliân / Ardeidae" [1] and "joqd-hâ / Owls". The tooltips related to these creatures show their full description, "h æ vâsil-e râh-râh / striped Heron" and "joqd-e bigôsh-e râh-râh / without ear striped Owl".

3.1.3 Third Step: Data Development

Each link in the Navbox refers to a Wikipedia page that has an article. These pages may also have Navboxes related to animals. Therefore, all these pages are also investigated and if new Navboxes exist, their contents will be extracted. There may be no pages available for some data. In this situation, those data will be marked for applying certain procedures in the next steps. Finally, after completing this step, 44 Navboxes with 2,346 unique entities were extracted from attributes collection.

3.1.4 Fourth Step: Find the Category of Entities

In this step, the extracted entity categories are found. Each entity in the classification hierarchy of animals has a category of its own; for example "Mohredârân yek zirshâxe az T æ nâbdârân æ st / Vertebrata [2] is a Subphylum[3] of Chordata[4]". This relation is actually the same as is-a relation in the classification of animals. First, In order to find the category of extracted entities, a series of relations were achieved using a simple statistics of the information in the Infoboxes and by choosing the category with the most frequency for the entity.

Next, to find the remaining entity categories (entities whose categories have not been yet found and entities that were marked in the previous step due to not having a relevant page) the location of word in Navbox will be used. To do this, the neighbors of an entity are examined by traversing Navbox and the category of entities will be guessed by using its neighbor's categories.

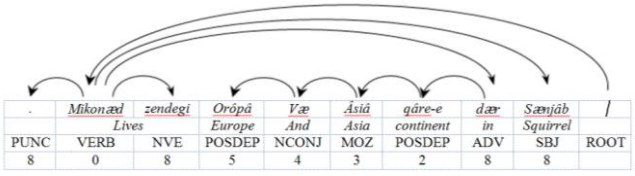
3.1.5 Fifth Step: Find Concept's Parent

In order to create a classification hierarchy of the extracted entities, finding the parent of each entity is necessary. This step is also performed in two phases: First, the concept parents are extracted through existing classification in Navboxes. Then to find the remaining concepts parent (concepts whose parents have not been found yet), the existing hierarchy in Infoboxes will be used to extract an integrated hierarchy. Finally, the information or the extracted metadata is stored in the Knowledge base (KB).

3.1.6 Sixth Step: Natural Language Processor

This unit, extracts the features associated with each entity using the existing texts in the Wikipedia pages and the metadata contained in KB. In order to extract features for each entity, five features including living location, food, size, weight, and longevity were considered. These features will be extracted using the rule-based approach (Maynard et al. 2009; Suchanek et al. 2009; Miháltz 2010). The defined patterns are given in Table 1.

At this point, after preparing the input file, the MateParser (Bohnet 2010) was applied. Below is an example to acquire the living location attribute of a Squirrel entity by using the dependency tree of sentence "Sænjâb dær qâre-e âsiâ væ orôpâ zendegi mikonæd / Squirrel lives in Asia and Europe continent".



.	Mikonæd	zendegi	Orôpâ	Væ	Âsiâ	qâre-e	dær	Sænjâb	/
	Lives		Europe	And	Asia	continent	in	Squirrel	
PUNC	VERB	NVE	POSDEP	NCONJ	MOZ	POSDEP	ADV	SBJ	ROOT
8	0	8	5	4	3	2	8	8	

In this case, the living location attribute of Squirrel is obtained through pattern 1 in Table 1 and the rules contained in the parse tree, is achieved as "qâre-e âsiâ væ orôpâ / Asia and Europe continent" noun phrase, given that the adverbial preposition "d æ r / in" comes to "zendegi mikonæd / Lives" verb.

3.1.7 Seventh Step: Ontology Producer

Finally, the obtained collection of entities and their relations were stored in an XML file. The resulting ontology includes a hierarchy of animal classification and five attributes related to each entity.

[1] Herons
[2] Vertebrates
[3] Sub-branch
[4] Chordates

Table 1. Defined patterns for extracted features

Pattern No.	Defined patterns
----------	---------- ***Living Location Pattern*** ----------
1	dær NP [zendegi mikonæd | pærâkænde æst | yâft mishævæd | sâken æst | sokônæt dâræd] [lives | scatters | finds | dwells | resides] in NP
2	[bômi | zistgâh | sâken] [dær | ""] NP æst is [native | habitat | residing] [in | to | ""] NP
3	mæhæle [sokônæt | zendegi] NP æst [life | residence] location is NP
4	dær NP [miziæd | ziste | mizist] [living|lived] in NP
----------	---------- ***Feed Pattern*** ----------
5	æz NP tæqzie mikonæd feeds of NP
6	[qæzây-e | xôrâk] ânhâ [shâmel | bishtær | ""] NP æst their [food | feed] is [included | more | ""] NP
7	rejim-e qæzâie [æz | shâmel] NP [tæshkil mishævæd | æst] diet [consists of | is] NP
8	[æz | ""] NP [râ | ""] mixoræd eats NP | NP to eat
----------	---------- ***Size (Length) Pattern*** ----------
9	tôl | derâzâ | qæd | qâmæt | bolændi | ændâze] NP [milimetr | sântimetr | metr | s.m | s | m] [æst | dâræd] [length | long | stature | height | size] [is | have] NP [millimeter | centimeter | meter | mm | cm | m]
10	NP [milimetr|sântimetr|metr|s.m|s|m] [tôl|derâzâ|qæd|qâmæt|bolændi|ændâze] dâræd have [length|long|stature|height|size] NP [millimeter|centimeter|meter|mm|cm|m]
11	tâ NP [milimetr | sântimetr | metr | s.m | s | m] roshd mikonæd grows up to NP [millimeter | centimeter | meter | mm | m | cm]
----------	---------- ***Weight Pattern*** ----------
12	[væzn | josse] NP [miligæræm | gæræm | kilogæræm | k.g | k | g | mg] [weight | body] is NP [milligram | gram | kilogram | mg | m | kg]
13	NP [miligæræm | gæræm | kilogæræm | k.g | k | g | mg] væzn dâræd weights NP [milligram | gram | kilogram | mg | m | kg]
----------	---------- ***Longevity Pattern*** ----------
14	[miângin | motevæset | ""] tôl omr NP [rôz | mâh | sâl] [average | mean] life time is NP [day | month | year]
15	NP [rôz | mâh | sâl] omr [dâræd | mikonæd] have life time NP [day | month | year]

3.2 Experiments of Ontology Construction Method

In order to evaluate the automatic construction of ontology, the experiments were performed on 3,200 Wikipedia articles saved by crawler in 30/1/2014.

To perform the experiments, 100 instances of Wikipedia articles were randomly selected and manually annotated. The evaluation of the automatic ontology construction system was done separately for calculating the accuracy of the extracted rank of the entity, accuracy of the extracted parent of the entity, accuracy of the extracted hierarchy and the accuracy of the extracted attributes for each entity. Table 2 shows the accuracy measure in the subsections evaluated. The results of the proposed method were compared with the structure of Carol Linnaeu's classification (Swedish botanist, physician and zoologist), introduced in his famous book "Systema Naturae" (Linnaeus 1735). It should be noted that this accuracy measure is one of the most popular measures in evaluation of algorithms. In some cases (concept parent extraction and hierarchy extraction), we have changed the classic accuracy measure in order to better evaluate our proposed method. The detail of this customized measure is described in more detail.

In order to evaluate the accuracy of the extracted rank of entities (row 1 of the table 2), if the rank of the entity has been extracted correctly, 1 and otherwise 0 will be considered as the score. In order to evaluate the accuracy of extracted parents of entities (row 2 of table 2), if the parent of the entity has been extracted correctly, 1 will considered as the score and otherwise for each generation distance with the parent, 0.2 will be subtracted from the score; this mean that if the proposed system, wrongly tags the grandfather of an entity as the parent of the entity instead of his father, its score will be 0.6. In addition, for evaluating the accuracy of the extracted hierarchy for each entity (row 3 of table 2), the location of each entity in the hierarchy of category of animals will be examined. If categorized correctly, 1 and otherwise 0 will be considered as the score.

Table 2. The results of experiments (accuracy criterion)

Subsection	Accuracy (%)
Data category extraction accuracy	93
Concepts parent extraction accuracy	89.8
Hierarchy extraction accuracy	99
Defined patterns accuracy to extract features	91

In the end, in order to evaluate the accuracy of the extraction of attributes, it is clear that with having 100 articles selected randomly and five attributes defined for each attribute (Location, Nutrition, Length, Weight and longevity), 500 attributes are evaluated. If an attribute is correctly extracted, its score will be 1 otherwise 0. Since most of the Persian Wikipedia articles in animal fields do not contain all five attributes, the constructed ontology in the section of attributes extractions has a very little information. According to experiments, from 500 attributes evaluated, only 67 attributes were in the pages and 91% of these were extracted correctly (Row 3 of Table 2); it means 61 attributes were extracted correctly and 6 were extracted incorrectly or they existed in the page or were not extracted by the pattern. In other words, only 13.4% of the attributes existed in the Wikipedia articles and of this, 12.2% of attributes were correctly extracted because of selecting the proper patterns. Even though, 12.2% is not an acceptable value for the attribute extraction subsection. The reason is the insufficient information in the Wikipedia articles, therefore 91% for the accuracy of the patterns defined in attribute extraction section is a considerable value which indicates the proper performance of the proposed algorithm in the attribute extraction section. As a result, in order to solve the problem of insufficient information in Wikipedia articles, it is necessary to improve the constructed ontology.

4. Improvement Method

As stated previously Wikipedia does not have complete information about all of the extracted data (Either there is no sentence related about a given attribute in the page or no pages have been defined for a given data). On the other hand, the proposed system will definitely not be able to extract whole attributes in the page during the ontology construction. Therefore, in order to improve the ontology and complete its information, a solution based on bootstrapping method were suggested so that new information can be extracted using the exiting information and with the help of Google search engine.

4.1 Proposed Ontology Improvement Method

To do this, whenever any of the attributes related to an entity, does not exist in the initial ontology (i.e., it was not extracted from Wikipedia pages in the step of automatic construction of ontology). In such cases, one or more commands are sent to Google search engine to extract the value of that attribute. In order to send the search words to Google, the (entity, attribute title) pair is used. Table 3 shows the words sent to Google search engine in case that any of the five attributes in the initial ontology is missing. For example, if the living location of an animal has not been extracted from Wikipedia pages in ontology construction stage, four commands in the pattern available in Table 3 will be sent to Google search engine. In the end, the first ten responses from Google will be evaluated and analyzed. The process of analysis and evaluation will be done on the ***snippet part*** of the returned results. Part of the return results from Google search engine is shown in Figure 3. The red rectangles in this figure are two samples of snippet part.

Since there might be different responses (relevant or irrelevant) in the results from Google search engine, a confidence measure is considered for selecting the best option and measuring the accuracy of the result which will be calculated from the sum of the following criteria and using the proposed algorithm discussed in Figure 4.

- *Name examining measure:* This measure has been in fact considered for pre-processing; it means that after receiving the results from Google, each snippet without the complete name of the data is removed so that it will not be processed in the next step.
- *Participation percentage measure:* This measure is considered for the percentage of participation of the entity in the returned results from Google. If the name of the data appears completely in more results, it can be said that the extracted result is correct with a higher confidence. Assuming that, TF_{Entity}, is frequency of the snippets in which the name of the entity exists and, $NR_{GoogleSearch}$, equivalent to DF measure, is the number of the returned results from Google (here the first ten results are evaluated), the participation percentage measure, $P_{Participation}$, is calculated as follows, which is similar to TF.IDF measure (Manning et. al. 2008):

$$P_{Participation} = \frac{TF_{Entity}}{NR_{GoogleSearch}} \times 100 \qquad (1)$$

- *Pattern accuracy measure:* A score is assigned to each pattern in Table 1 based on the percentage of their participation and the amount of correct results extracted in the automatic ontology construction stage; it means that the pattern is more accurate if it has higher participation and the number of its incorrect obtained results is smaller. Scores and ranks for the patterns in Table 1 are assigned in the following way: assuming that, $TF_{Pattern}$, is the frequency of a given pattern for extracting the attribute in the construction stage, $P_{Correct}$, is the percentage of correct results and, $P_{Incorrect}$, is the percentage of incorrect results extracted from the same patterns, the score of each pattern, P_{Score}, is calculated based on equation 2 (Han et. al. 2011).

$$P_{Score} = TF_{Pattern} \times (P_{Correct} - P_{Incorrect}) \qquad (2)$$

It is important to note that when comparing two patterns, the higher is the, P_{Score}, for a pattern, the more valid is the pattern. Therefore, the patterns in Table 1 are ranked based on this measure and one round of scores normalization (normal distribution) in this way, pattern measure, $M_{Pattern}$,

will be obtained. The reason for the ranking is that whenever the value of an attribute is extracted based on a pattern; the amount of accuracy of the pattern can be measured and used in calculating the confidence measure.

Table 3. The words sent to Google search engine for the five attributes in ontology

	Attribute title	Query words
1	Mækân zendegi mojôdiyæt Entity living location	1- Mojôdiyæt + "Zendegi Mikonæd" (Entity + "Lives") 2- "Zistgâh" + Mojôdiyæt ("Habitat" + Entity) 3- "Mækân zendegi" + Mojôdiyæt ("Living Location" + Entity) 4- Mojôdiyæt + "Bômi" (Entity + "Native")
2	Khôrâk mojôdiyæt Entity nutrition	1- "Khôrâk" + Mojôdiyæt ("Nutrition" + Entity) 2- "Ghæzâye" + Mojôdiyæt ("Food" + Entity) 3- Mojôdiyæt + "Mikhoræd" (Entity + "Eats")
3	Ændâze Mojôdiyæt Entity size	1- "Ændâze" + Mojôdiyæt ("Size" + Entity) 2- "Tôl" + Mojôdiyæt ("Length" + Entity)
4	Væzn Mojôdiyæt Entity weight	1- "Væzn" + Mojôdiyæt ("Weight" + Entity) 2- "Josse" + Mojôdiyæt ("Bulk" + Entity)
5	Tôl Omr Mojôdiyæt Entity longevity	1- "Tôl Omr" + Mojôdiyæt ("Longevity" + Entity)

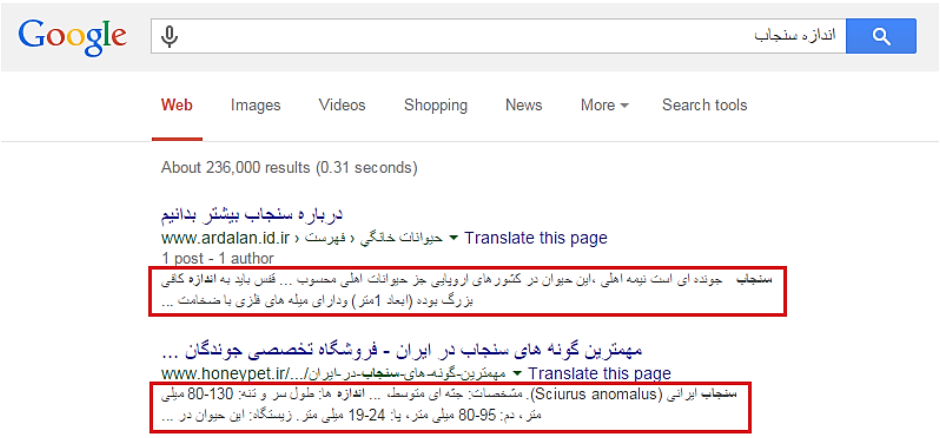


Fig 3. Part of the return results from Google search engine to query "Ændâze Sænjâb / Size of squirrel"

- *Location of presence measure:* If both pairs (the entity and the title of the attribute) appear in a sentence simultaneously and match the pattern, the confidence measure is considered 100%. The more is the distance, the less will be the confidence. Here, for each sentence distance between the pair, 20% is decreased from its score and this will be done for two sentences before and after the sentence matched with the patterns. This decay factor has been achieved by experimental results.

In the end, the best option will be obtained by the algorithm proposed in the Figure 2. Using the proposed algorithm, whenever there are multiple different options for the selection of an attribute in the results returned by Google search engine, the best option can be selected by the confidence measure and its accuracy can be measured, too. Here, based on our empirical evaluations, the results whose confidence measure is below 30% ($M_{Confidence} < 30\%$), will not be considered due to the lack of confidence in the accuracy of the result.

Table of pattern, TP, in algorithm 1 includes columns for defined pattern, $Pattern_i$, the score of the pattern, $Score_i$, the average of scores for similar patterns, Avg_{Score}, and the amount of standard deviation in proportion to the similar patterns, Var_{Score}. Similar patterns refer to the patterns defined for extracting an attribute. These values are calculated in advance and placed in the columns of the table. As stated above, the score of each pattern, $Score_i$, is calculated by equation 2 and the value of, Avg_{Score}, is calculated by calculating the average of scores of similar patterns. Similarly, the value of, Var_{Score}, is calculated via equation 3, which is the classic equation for computing variance (Han et. al. 2011):

$$Var_{Score} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(Score_i - Avg_{Score})^2} \qquad (3)$$

The search command with two words (entity e and the title of the attribute a) is first sent by Algorithm 1 in Figure 4 to Google search engine. Then results returned from Google are examined. If any of the snippets does not have the name of the entity completely, further processing on the snippet will be ignored and the operation will go on for the next

snippet. In the next step, if each sentence in the snippet matches a pattern in the table of patterns, the Confidence Measure function will be called with the sent results. This will be done for all sentences from the extracted snippets. In the end any sentence with the highest confidence measure will be selected as the best option for the given attribute.

Algorithm 2 first finds out if both search words are present in a sentence. If this is the case, the confidence measure is 100%; because it can be definitely said that the returned result is completely true. If both of the search words are not present in a sentence, the average of participation percentage of entities in the returned result of, Google $P_{Presence}$, (Percentage of Participation function), the amount of accuracy of the pattern, $M_{Pattern}$, (Measure of pattern function) and the distance of search words, $L_{Presence}$, (Location of presence function) will be calculated as the confidence measure.

PercentOfParticipation function using relation 1, calculates the percentage of participation of the entity in the results returned by Google search engine. MeasureOfPattern function using equation 4, Normal standard distribution, calculates the accuracy of the pattern (Han et. al. 2011).

$$z = \frac{Score_i - Avg_{score}}{Var_{score}} \quad (4)$$

Then the value of z is obtained from the table of normal standard distribution and the percentage of the accuracy of the pattern is calculated.

LocationOfPresence function, as mentioned previously, considers the confidence measure of the sentence, 100% if both pairs (The entity and the title of the attribute) appear in a sentence at the same time. This value is calculated separately in the first line of the ConfidenceMeasure function (Algorithm 2 in Figure 4); because in this case, There is no need for calculating the other measures. Definitely the larger is the distance of the search pairs, the less will be the confidence percentage. Here, for each sentence distance between pairs, the scores will be decreased by 20%; it means that if the distance between a given pair is one sentence, the score will be 80% and if the distance is two sentences, the score will be 60% and so on. This will be done until two sentences before and after the sentence matched with the pattern and if the distance between pairs is more than two sentences, the score will be 30%.

4.2 Experiments of Ontology Improvement

The proposed method for improving the ontology focuses on extracting information using Bootstrapping methods for extending and developing the parts of the ontology which do not exist in the Wikipedia corpus; it means those 433 attributes out of the 500 attributes evaluated that were not extracted due to the incompleteness of texts in Wikipedia articles.

Algorithm1: AttributeExtractionUsingGoogleSearchEngine
Input: Entity *e*, Attribute *a*, Table of Patterns *TP*
Output: Attribute Value $Attrib_{val}$, Best Confidence Measure $BM_{Confidence}$

```
Results ← GoogleSearchEngine(e, a)
For each (Snippet in Results)
{
    If (Snippet not contains e) then Continue
    For each (Snippet.Sentence Matched to TP.Pattern_i)
    {
        Array[i].Confidence ←
            ConfidenceMeasure(e, a, TP.Score_i, TP.Avg_Score, TP.Var_Score)
        Array[i].AttributeValue ← Snippet.Sentence
    }
}
index ← Index of Maximum(Array.Confidence)
Return Attrib_val ← Array[index].AttributeValue and
       BM_Confidence ← Array[index].Confidence
```

Algorithm2: ConfidenceMeasure
Input: Entity *e*, Attribute *a*, Score of Pattern $Score_i$, Average of Pattern Scores Avg_{Score}, Variance of Pattern Scores Var_{Score}
Output: Confidence Measure $M_{Confidence}$

```
If ("e and a" Located in One Sentence) then Return M_Confidence ← 100%
Else
    {
        P_Participation ← PercentOfParticipation (e, a)
        M_Pattern ← MeasureOfPattern(e, a, Score_i, Avg_Score, Var_Score)
        L_Presence ← LocationOfPresence(e, a)
        Return M_Confidence ← ( ∑ P_Participation, M_Pattern, L_Presence)/3
    }
```

Fig 4. Algorithm for extracting information using Google search engine and selecting the best result (Algorithm1) and calculating the confidence measure (Algorithm2).

Finally, in order to calculate the percentage of improvement of the ontology, a new attribute was examined that was extracted by the proposed algorithm and did not exist in the initial ontology. Assuming that, $N_{Correct}$, is the sum of the number of new and correct relations extracted by the Google search engine and, N_{Total}, is the total number of relations that did not exist in these 100 random samples in the initial ontology, the

percentage ontology improvement, $P_{Improvement}$, is calculated by equation 5, which is standard accuracy measure for not covered samples in initial ontology:

$$P_{Improvement} = \frac{N_{Correct}}{N_{Total}} \quad (5)$$

5. Evaluation

Experiments were performed separately in two subsections for the automatic construction of prototype ontology and the improvement of the initial ontology. As a result, the percentage of the proposed method improvement can be evaluated by comparing the initial ontology and the improved one.

According to the experiments performed, out of the 433 attributes that did not exist in the initial ontology (due to the incompleteness of Wikipedia articles), 152 attributes were extracted by the proposed method for improving the ontology. 138 of these attributes were extracted, correctly. Table 4 shows the details of the calculation of the defined patterns accuracy.

The number indicates two interesting points; first, even with extending the domain of information collection in the web using Google search engine, the patterns defined in the section of extraction attributes from texts, still gain the 91% accuracy score. Second, the number of newly extracted attributes using the proposed algorithm is twice more than (2.26 times) the number of attributes that were extracted using the information available in texts of Wikipedia articles. This value can still be extended by increasing the number of retrieved documents from Google search engine. Due to the limitation in sending requests to Google search engine (a regular user can get only 10 results per order while only 100 requests are permitted per day), we were forced to evaluate only the top 10 returned results. By increasing this value, the domain of the extracted attributes can be extended considerably.

Another important issue to note is about the comparison of our method with other existing algorithm. Our method is the first study in ontology extraction for animal's domain in Persian, and due to this fact, we are unable to compare our work with similar researches.

Conclusions and Future Work

In this study, first a prototype ontology was automatically extracted from the existing structures in Wikipedia. Since the information in the constructed ontology was not complete (due the incompleteness of Wikipedia articles), a solution for improving and extending the information in the initial ontology was proposed using Google search engine and Bootstrapping method.

Considering the complexity of processing Persian language (due to the freedom in the order of words and the lack of strong rules), by selecting correct patterns for extracting attributes in the end, good results were achieved in the section of automatic construction of ontology. In fact, the resulting ontology is a domain ontology particular for animals which in addition to having a hierarchy of different animal categories, has attributes of living location, nutrition, size, weight and longevity for each of them. This ontology can be used for educational and training purposes. Also, considering the increasing growth of Web and the fact that many up-to-date information resources are being added to the Web, the method for improving ontology can be used to extend and update the constructed ontologies.

It is predicted that in the future, machine learning methods will be tried. Furthermore, a pattern bank will be defined so that in case of detecting new patterns in the process of extracting entities, they can be processed, examined and registered in the pattern bank. This is done so that the new patterns can be used in the process of extracting entities in the next steps. These steps are done iteratively so that each time the collection can be extended with new patterns.

Table 4. Details of the calculation of defined patterns accuracy to extract features

Methods	Total	Existing attributes	Correct	Incorrect
Automatic Construction of Ontology	500	67	61	6
Improvement of Ontology	433	152	138	14

References

[1] Azevedo, R. R., Freitas, F., Rocha, R. G. C., Menezes, J. A. A., Oliveira Rodrigues, C. M., and F. P. Silva, G. (2014). An Approach for Learning and Construction of Expressive Ontology from Text in Natural Language. Proceedings of 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 149-156.

[2] Bohnet, B. Top accuracy and fast dependency parsing is not a contradiction. Proceedings of Coling 2010, 89–97.

[3] Crockford , D. JSON: The fat-free alternative to XML. Proceedings of XML 2006.

[4] Gruber, T. Ontolingua: a translation approach to providing portable ontology specs. Knowledge Acquisition, 5(2), (1993). 199-220.

[5] Han, J., Kamber, M., & Pei, J. Data mining: concepts and techniques: concepts and techniques. Elsevier. (2011).

[6] Horrocks, I. et al. OWL: a Description-Logic-Based Ontology Language for the Semantic Web. In: The Description Logic Handbook: Theory, Implementation and Applications, (2007). 458-486.

[7] Kozaki, K., Kitamura, Y., Ikeda, M. et al. Hozo: an environment for building/using ontologies based on a

fundamental consideration of 'Role' and 'Relationship'. Proceedings of Computer Science, (2002). 2473: 213-218.
[8] Küçük, D., and Arslan Y. Semi-automatic construction of a domain ontology for wind energy using Wikipedia articles. Proceedings of Renewable Energy 62, (2014). 484-489.
[9] Lenat, D. B., & Guha, R.V. Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project. Boston: Addison-Wesley Publishing Co. (1990).
[10] Linnaeus, C. Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Sweden. (1735).
[11] Malo, P., Siitari, P., Ahlgren, O., Wallenius, J., & Korhonen, P. Semantic Content Filtering with Wikipedia and Ontologies. Proceedings of Third International Workshop on Semantic Aspects in Data Mining (SADM'10) in conjunction with the 2010 IEEE International Conference on Data Mining, (2010). 518–526.
[12] Manning, C. D., Raghavan, P., & Schütze, H. Scoring, term weighting and the vector space model. Introduction to IR, 100. (2008).
[13] Maynard, D., Funk, A., & Peters, W. SPRAT: a tool for automatic semantic pattern-based ontology population. Proceedings of the International Conference for Digital Libraries and the Semantic Web. Italy: Trento. (2009).
[14] Miháltz, M. Information Extraction from Wikipedia Using Pattern Learning. Journal of Acta Cybern. 19(4), (2010). 677-694.
[15] Ponzetto, S. P., & Strube, M. Deriving a Large Scale Taxonomy from Wikipedia. Proceedings of Association for the Advancement of Artificial Intelligence (AAAI07). (2007).
[16] Richardson, M., & Domingos, P. Markov Logic Networks. Machine Learning. (2006).
[17] Sanabila, h., and Manurung, R. Towards Automatic Wayang Ontology Construction using Relation Extraction from Free Text. Proceedings of the 8th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), (2014). 128-136.
[18] ShamsFard, M. Persian text processing: past achievements, challenges ahead. Proceedings of the second workshop on research in English and computers, University of Tehran, (2006). 172-189.
[19] ShamsFard, M., & Barforoush, A. A. Extracting conceptual knowledge from text using linguistic and semantic patterns. Journal of Cognitive Science, 4(1), (2002). 48-66.
[20] Shibaki, Y., Nagata M., & Yamamoto, K. Constructing Large-Scale Person Ontology from Wikipedia. Proceedings of the 2nd Workshop on Collaboratively Constructed Semantic Resources, Beijing. (2010).
[21] Song, Q., Liu, J., & Wang, X. A Novel Automatic Ontology Construction Method Based on Web Data. Proceedings of Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, (2014). 762-765.
[22] Staab, S., & Studer, R. Handbook on Ontologies. (pp. 1-17). Springer: International Handbooks on Information Systems. (2004).
[23] Suchanek, F. M., Kasneci, G., & Weikum, G. YAGO: a large ontology from Wikipedia and WordNet. Journal of Web Semantics: Sci. Serv. Agents World Wide Web6, (2008). 203–217.
[24] Suchanek, F. M., Sozio, M., & Weikum, G. SOFIE: a self-organizing framework for information extraction. Proceedings of the 18th International conference on World Wide Web, New York, NY, USA. ACM, (2009). 631-640.
[25] Syed, Z., Finin, T., & Joshi, A. Wikitology: Using Wikipedia as ontology. Proceeding of the second international conference on weblogs and Social Media. (2008).
[26] Wu, F., & Weld, D. Autonomously Semantifying Wikipedia. Proceedings of CIKM07. Portugal: Lisbon. (2007).
[27] Wu, F., & Weld, D. Automatically Refining the Wikipedia Infobox ontology. Proceedings of WWW08. (2008).
[28] Xiong, J., Liu, Y., Wang, J., and Lan, Y., Research of Marine Organism Ontology Semi-Automatic Construction. Proceedings of the Open Cybernetics & Systemics Journal, (2014). 984-989.
[29] Yao, Y., Liu, H, Yi, J., Chen H., & Zhao X. An Automatic Semantic Extraction Method for Web Data Interchange. Proceeding of 6th International Conference on CSIT, (2014). 148-152.
[30] Yu, C., Cuadrado, J., Ceglowski M., & Payne J. S. Patterns in Unstructured Data: Discovery, Aggregation, and Visualization. Proceeding of NITLE. (2002).
[31] Zhou, G., & Zhang, M. Extracting relation information from text documents by exploring various types of knowledge. Proceedings of Information Processing and Management, (2007). 43: 969-982.

Sedigheh Khalatbari received the B.Sc. and M.Sc. degree in Software Engineering from University of Guilan, Rasht, Iran in 2012 and 2015 respectively. Her main research interests include Natural Language Processing, Text Mining and Semantic Analysis.

Seyed Abolghasem Mirroshandel received his B.Sc. degree from University of Tehran in 2005 and the M.Sc. and Ph.D. degree from Sharif University of Technology, Tehran, Iran in 2007 and 2012 respectively. Since 2012, he has been with Faculty of Engineering at University of Guilan in Rasht, Iran, where he is an Assistant Professor of Computer Engineering. Dr. Mirroshandel has published more than 30 technical papers in peer-reviewed journals and conference proceedings. His current research interests focus on Data Mining, Machine Learning, and Natural Language Processing.

A Linear Model for Energy-Aware Scheduling Problem Considering Interference in Real-time Wireless Sensor Networks

Maryam Hamidanvar*
Department of Computer Engineering, Arak University, Arak, Iran
maryam.hamidanvar@yahoo.com
Reza Rafeh
Department of Computer Engineering, Arak University, Arak, Iran
r-rafeh@araku.ac.ir

Received: 27/Mar/2015 Revised: 01/Nov/2015 Accepted: 23/Nov/2015

Abstract

An important factor in increasing quality of service in real-time wireless networks is minimizing energy consumption, which contradicts with increasing message delivery rate because of associating a time deadline to each message. In these networks, every message has a time deadline constraint and when the message is not delivered to its destination before its deadline, it will drop. Therefore, scheduling methods that simultaneously consider both energy consumption and time deadline constraint are needed. An effective method for reducing energy consumption is multi-hop transmission of packets. However, this method takes longer time for transmission as compared to single-hop transmission. Parallel transmission is another approach which on one hand reduces the transmission time and on the other hand increases the network throughput. However, a main issue with parallel transmission is the presence of interference among nearby nodes. In this paper, we propose a linear model (ILP formulation) for energy aware scheduling problem in real-time wireless sensor networks using parallel transmission. The main objective of the model is to reduce energy consumption and packet loss using multi-hop routing and parallel transmission. Simulation results show that the proposed model finds the optimum solution for the problem and outperforms the sequential scheduling based on the TDMA protocol.

Keywords: Energy Consumption; Parallel Transmission; Scheduling; Optimization; Routing; Interference.

1. Introduction

Wireless networks consist of nodes that communicate with each other by radio waves. Each node in the network has a limited amount of energy stored in a battery that is not rechargeable. Thus the end of the battery life denotes the end of the node's lifetime [1]. In real-time wireless networks, in addition to limitation of energy resources, the duration of the packet delivery has also a time-deadline constraint.

According to the quality of service metrics in real-time wireless networks, messages must be delivered within the specified time, otherwise they will become useless. Thus, real-time networks need to send the messages timely in the network. However, factors such as limited power supply, interference, network congestion and loss of links reduces the ability of the network to achieve the desired objectives [2].

A useful strategy to reduce energy consumption and thereby increasing the lifetime of the network is to use multiple smaller hops instead of a single long hop between source and destination. Because energy used to send a message is directly proportional to the square of the hop length [1,3,4]. Therefore, if the number of hops is increased and distance between the hops is decreased, the energy consumption will also decrease. However, using intermediate hops will increase message transmission time. So the number of intermediate hops should be set in such away as to minimize energy consumption while satisfying the time deadline constraint.

Parallel transmission of packets increases the number of packets transmitted per unit time and hence increases the efficiency of the real-time wireless network. In sequential scheduling based on TDMA protocol packets are transmitted by the source nodes sequentially [5]. Considering the possibility of parallel transmission, several source nodes can send the message simultaneously thereby increasing the efficiency of the network. Parallel transmission prevents the violation of time constraint as far as possible and in addition to that reduces the amount of energy required to transmit the message. Due to the increase in number of messages sent per unit time, more time units will be available and the message will be sent via route having more hops with less distance in between them. But parallel transmission is restricted by the phenomenon of interference in wireless networks. Because by increasing parallel transmission, probability of interference also increases. The phenomenon of interference is due to the collision of signals sent from nodes and causes loss of packets [6]. Several interference models have been proposed to model the interference in wireless networks and the presence or

* Corresponding Author

absence of interference between two links depends on the interference model used [7]. In [8] *Protocol Model* has been introduced for modeling the interference between communication links which is widely used in research. We have used this model in this paper and it will be explained in section 2.

There are two other types of interferences named *Primary Interference* and *Secondary Interference* that also cause packet loss. Primary interference occurs when a node transmits and receives messages simultaneously. Secondary interference occurs when a node receives more than one message at the same time [7].

The contrast between these limitations has put the problem of Energy-aware scheduling in the category of NP-Hard problems [3,9]. Hence, it is necessary that the conflicting objectives in the problem be addressed simultaneously as an optimization problem. Optimization problems are a group of combinational and optimization problems in which the aim is to find the best solution that satisfies all the constraints and maximize or minimize an objective function.

Recent approaches to solve the combinational and optimization problems consist of two steps: The first step is the modeling of the problem and the second step is solving the model [10]. There are three basic techniques for solving these problems: Mathematical Methods (MM), Constraint Programming (CP) and Local Search (LS). Each of these techniques has its own advantages and disadvantages. Among these methods, mathematical methods have particular importance due to their high efficiency. However, a linear model is required for the problem but formulation of the combinational and optimization problems as linear model is difficult [11].

The works related to scheduling and routing of wireless networks can be grouped according to their objective function. In some studies the objective is to increase efficiency and reduce end to end delay regardless of the energy consumption and in others it is to reduce energy consumption. In [9] an Integer Linear Program (ILP) formulation has been presented for the problem of energy-aware scheduling in real- time wireless networks, without taking into consideration the interference phenomenon. In this method the nodes transmit messages in a sequential manner and there is no possibility of parallel transmission. In [1] to create a balance between energy consumption and time delay a non-linear model based on Concentric Circular Bands (CCBs) has been proposed. In this study, the effect of the interference and parallel transmission has not been taken into consideration. In [12], a nonlinear model has been proposed which aims at reducing overall energy consumption in clustered WSNs and then based on results obtained by solving the model an algorithm has been suggested to obtain minimum latency. In [5] a scheduling method for parallel transmission to multiple destinations is provided that improves the network performance as compared to the sequential TDMA method. In [13], to maximize throughput, authors have considered joint routing, channel assignment and collision free scheduling of links. In [7], solutions based on the idea of graph coloring for the problem of scheduling and routing of wireless networks have been proposed in order to increase efficiency but without considering the energy consumption. In [14] authors have considered a routing tree and proposed two delay efficient algorithms to create interference-free scheduling for data aggregation.

Most models and methods that have been proposed for the problem of scheduling real-time wireless networks have considered either reduction in energy consumption or increase in the number of delivered packets separately. Moreover, most of the existing methods are not suitable because of not considering time constraint specific to each message in real-time networks. Because in these methods usually the reduction in end to end delay or increase in number of packets transmitted is considered for all messages in a specific time interval. While in real-time wireless networks, each message has its own time constraint that may be different or identical to that of the other messages. As a result, the priority of data transmission may vary according to the various time deadlines. Another problem that can be seen in most of the previous research is ignoring the remaining energy of nodes in routing. This results in nodes lacking sufficient energy to be used in the selected route.

In this paper, linear modeling (ILP formulation) of the problem by considering both the energy consumption and time constraints has been performed. The proposed model aims to minimize energy consumption and reduce the number of lost packets by utilizing the strategy of parallel transmission and multi-hop routing.

The model is solved using Simplex method. Implementation results show the improved efficiency of scheduling by the proposed model as compared with sequential scheduling approach based on TDMA protocol. The model is able to determine the optimal route for each message and parallel scheduling keeping in view the limitations of the network.

Rest of the paper is organized in this way: In Section 2, we review some important concepts in real-time wireless networks. Section 3 describes the problem and its constraints. The section 4 of the paper describes the assumed network model. Section 5 describes the new proposed model. In section 6 we will evaluate the proposed model and discus the results obtained by solving the model. Finally Section 7 concludes the paper.

2. Energy-Aware Scheduling Problem

The problem of energy-aware scheduling with parallel transmission capability can be described as following.

1. Determine the optimal route for each message using multiple hops to reduce energy consumption while keeping in view the time constraint of each message.
2. Allocating units of time according to time constraint of messages and capability of parallel

transmission to increase the network throughput. While allocating units of time, priority should be given to messages according to their time deadlines.

Constraints of the problem are [7,15]:

1. The receiver node should be in the communication range (radio range) of the transmitter node.
2. A node cannot receive more than one message at the same time.
3. A node cannot transmit more than one message at the same time.
4. A node cannot transmit and receive messages simultaneously.
5. The nodes located in the interference range of each other should not be active at the same time.
6. Scheduling and routing of a message should be done before the time deadline of that message because every message is valid only until its time deadline. Thus continuing scheduling and routing after the time deadline of the message is useless.
7. Only nodes having enough energy are able to transmit and receive the messages.

Objectives of the problem are:

1. Minimizing the energy consumption of the network to transmit messages.
2. Maximizing the number of successfully delivered messages in specified time.

Satisfaction of the objective function reflects the quality of scheduling. This means that among the existing solutions, a solution that has the lowest energy consumption as well as the lowest number of packet loss offers optimal scheduling.

3. The Network Model

In the assumed network nodes are static and distributed randomly in a two-dimensional plane and geographic coordinates of each node is known. Each node has its own specific initial energy and transmission range that may be the same or different from those of the other nodes. Energy used by every node in each transmission is calculated by the following equation [9]:

$$E_{tr} = Cd^2 \quad (1)$$

Where d is the distance between transmitter and receiver node and C is a constant coefficient that depends on the message length and physical condition of the network. In addition to the energy consumed for transmission, trace amount of energy is used by the receiving node while receiving the message and by idle nodes. But this amount is negligible as compared to energy used for transmission, so we have ignored it in this research.

In the assumed model all the nodes use shared wireless medium that is divided into equal time slots. Transmitter nodes that do not interfere with each other can access the shared wireless medium simultaneously. In the networks where the nodes use a single channel, each node can transmit message to all the neighboring nodes in a time slot [15].

Simultaneous activity of nodes that are located in the interference range of each other result in interference. According to the interference model, transmission and interference range of nodes may be equal or different from each other. In this paper, protocol model is used for modeling the interference between communication links. In this model transmission power of the node varies dynamically according to its distance from the receiver node. As a result interference range also varies according to the transmitter-receiver distance. According to this model if a node v_i transmits to a node v_j, transmission will be successfully received by node v_j if:

$$|v_k - v_j| \geq (1 + \Delta)|v_i - v_j| \quad (2)$$

Where $|v_i - v_j|$ is the Euclidean distance between node v_i and node v_j. The constant parameter $\Delta > 0$ denotes guarded zone specified by the protocol to prevent collision from neighboring node. v_k denotes every node simultaneously

transmitting at the same time in the same channel [8].

In the protocol model, transmission from node i to node j is successful only if no other node within the interference range of node j is transmitting at the same time [16].

All communication links in the network are assumed reliable and without noise.

All of the transmitted messages in the network have the same length with the message size of L bits. Time deadline and the source-destination of each message may be the same or different from other messages. Transmission of a message in each hop takes a unit time and is not dependent on the distance between sending and receiving node. It is assumed that the duration of each time slot is equal to the maximum time needed for transmission of a message between the two farthest nodes in the environment. Therefore, each message will be allocated a full time slot even if the time taken by the transmission of the message is less than the allocated time slot.

4. ILP Formulation

In this section the problem is formulated as integer linear program (ILP) keeping in view the characteristics and limitations of the network.

In energy-aware scheduling problems the total energy consumption and the number of packets delivered are calculated in a time window. In this model, size of the time window is equal to the longest time-deadline of existing messages in the network.

Parameters

Msg: Set of messages

Each message is shown by a tuple as:

(m, src, des, dln) where

m: id of message.

src: Source of message.

des: Destination of message.

dln: Time deadline of message.

Time: Size of the time window.
Nodes: Set of the nodes existing in the network.
R_i: Transmission range of node *i*.
E_i: The initial energy of node *i*.
$d_{i,j}$: Euclidean distance between nodes *i* and *j*.

Variables

Solution of the problem is presented as a four dimensional binary matrix.

$$X_{\text{Time, Nodes,Nodes,Msg}}$$

$$X_{t,s,d,m} = \begin{cases} 1 & \text{If node s transmits message m to node d} \\ & \text{during timeslot t} \\ 0 & \text{Otherwise} \end{cases}$$

The first dimension is time, second dimension is transmitter nodes, third dimension is receiver nodes and the fourth dimension is messages.

Constraints for ILP formulation are as follows:

- Equation (3) imposes limitation of transmission range for each node and ensures that receiver node is located within the transmission range of the sender node.

$$\forall i \in Nodes, \forall t \in Time, \forall k \in Msg, \forall j \in Nodes \quad X_{t,i,j,k} * d_{i,j} \le R_i \tag{3}$$

- Equation (4) ensures that a node cannot transmit and receive simultaneously. It also restricts sending and receiving more than one message by the node at a time.

$$\forall t \in Time, \forall i \in Nodes \quad \sum_{j \in Nodes} \sum_{k \in Msg} X_{t,i,j,k} + \sum_{s \in Nodes} \sum_{k \in Msg} X_{t,s,i,k} \le 1 \tag{4}$$

- According to the protocol model, equation (5) imposes that if simultaneous transmission of two sender nodes (such as node *i* and node *s*) is interfering only one of them is allowed to transmit. As stated in equation(2) about protocol model, interference occurs when the distance between the sender node *i* and the receiver node *j* multiplied by a constant 1+Δ is greater than the distance between the receiver node *j* and another transmitting node say node *s*. For example if node *i* is the sender node and node *j* is the corresponding receiver node, and another node s is also transmitting simultaneously and the distance between the node *s* and the node *j* is less than the product of distance between the node *i* and node *j* multiplied by a constant(1+Δ), interference will occur. Equation (5) prevents such interferences.

$$\forall t \in Time, \forall i \in Nodes, \forall s \in Nodes, \forall j \in Nodes \quad where\ i \ne s \ne j \ and\ d_{s,j} < (1+\Delta) * d_{i,j} \quad \sum_{k \in Msg} X_{t,i,j,k} + \sum_{m \in Msg} \sum_{d \in Nodes} X_{t,s,d,m} \le 1 \tag{5}$$

- Constraint (6) ensures that scheduling and routing of the messages is done before its time deadline. It means that the total unit times elapsed for each message must be less than its allocated time deadline.

$$\forall k \in Msg, \forall t \in Time_{where\ t > dln_k} \quad \sum_{i \in Nodes} \sum_{j \in Nodes} X_{t,i,j,k} = 0 \tag{6}$$

- Equation (7) determines energy source limitation for each node to transmit.

$$\forall i \in Nodes \quad \sum_{t \in Time} \sum_{k \in Msg} \sum_{j \in Nodes} X_{t,i,j,k} * C * (d_{i,j})^2 \le E_i \tag{7}$$

- Constraint (8) ensures that each node either sends a message once only or not at all. This means that if node *s* sends a message *k* to node *d* in time *ti*, then the node *s* will not send the message *k* again to node *d* or any other node in time *tj*.

$$\forall i \in Nodes, \forall k \in Msg \quad \sum_{t \in Time} \sum_{j \in Nodes} X_{t,i,j,k} \le 1 \tag{8}$$

- Equation (9) imposes that for each intermediate node the total number of incoming messages minus the total number of outgoing messages before their time deadline is zero i.e. the number of input messages and the number of output messages is equal. In this way equation (9) ensures that a continuous path is selected for each message. This means that each message that enters an intermediate node before deadline must be forwarded by the same intermediate node.

$$\forall k \in \text{Msg}, \forall i \in Nodes_{\ where\ src_k \ne i \ne des_k} \quad \sum_{s \in Nodes} \sum_{\substack{t_1 \in Time \\ where\ t1 < dln_k}} X_{t_1,s,i,k} - \sum_{d \in Nodes} \sum_{\substack{t2 \in Time \\ whete\ t2 \le dln_k}} X_{t_2,i,d,k} = 0 \tag{9}$$

- Constraint (10) ensures that the sending time of a message by the intermediate node is after the receiving time of that message.

$$\forall k \in \text{Msg}, \forall i \in Nodes_{\ where\ src_k \ne i \ne des_k} \quad \sum_{\substack{t_2 \in Time \\ where\ t_2 \le dln_k}} \sum_{d \in Nodes} X_{t_2,i,d,k} * t_2 - \sum_{\substack{t_1 \in Time \\ where\ t_1 < dln_k}} \sum_{s \in Nodes} X_{t_1,s,i,k} * t1 \ge 0 \tag{10}$$

- Equations (11) and (12) impose that the message does not enter the source node and does not leave the destination node, respectively.

$$\forall k \in \text{Msg}, \forall i \in Nodes \quad \sum_{t \in Time} X_{t,i,src_k,k} = 0 \tag{11}$$

$$\forall k \in \text{Msg}, \forall i \in Nodes \quad \sum_{t \in Time} X_{t,des_k,i,k} = 0 \tag{12}$$

Objective function of the problem:

The objective function in equation (13) aims to minimize energy consumption and packet loss. The first part of the equation (before minus sign) calculates energy consumption for transmission of messages and the second part of the equation calculates number of delivered messages.

Minimize

$$\sum_{t\in \text{Time}} \sum_{i\in \text{Nodes}} \sum_{j\in \text{Nodes}} \sum_{k\in \text{Msg}} X_{t,i,j,k} * C * \left(d_{i,j}\right)^2 - M * \sum_{k\in \text{Msg}} \sum_{i\in Nodes} \sum_{t\in \text{Time}} X_{t,i,des_k,k} \quad \text{where } t \le dln_k \tag{13}$$

In this equation M is a big positive number that is used as weighted coefficient for the delivered messages. The reason to use such a big coefficient is that the energy consumed by the network to send messages is much greater than the number of messages delivered, so routing causes the objective function to be always positive and the minimum value for the objective function is possible when no message is routed to its destination so that no energy is used in the network. Therefore, in the absence of the weighted coefficient for the number of delivered messages no routing is done so that the objective function remains zero. Use of big coefficient results in negative objective function value and thus maximizing the number of messages delivered without time violation.

5. Results

In this section results obtained by solving the proposed model are evaluated in different scenarios. The linear model is solved by using the ILOG CPLEX 12.2 software. The presented model is evaluated in terms of energy consumption and the number of messages delivered as compared to sequential TDMA based scheduling method. Improvement of scheduling by the proposed model because of using parallel transmission is demonstrated as compared to sequential scheduling. The reason for not comparing the results obtained by solving the model with other relevant work is different scenario of the problem because of considering more constraints as explained in section 1.

In the main scenario the nodes are distributed randomly in a two-dimensional region measuring $200 \times 200m^2$ Transmission range of each node is assumed to be 100m. Values of the parameter Δ in protocol model and M in objective function are assumed to be 10^{-9} and 10000 respectively. Each of the messages in the network can have the same or different source, destination and time deadline as compared to other messages.

Equation (1) is used to compare the energy consumption and the value of the parameter C is considered equal to 0.2. The number of nodes varies between 10 and 40, and the effect of number of nodes is evaluated on the quality of scheduling. The reason to limit the number of nodes to 40 is lengthy execution time of the proposed model and memory usage complexities. The number of messages is also decreased because of the same reason. This is due to the fact that by increasing the number of nodes and messages, the number of variables of the problem increases that results in lengthy runtime and high memory usage. When the number of nodes is 10, the solver can solve the model for routing more than 200 messages in reasonable time (less than 10 minutes). But when the number of nodes is increased, the solver will find the solution of the problem with less number of messages. For 20, 30 and 40 set of nodes, the solver can solve the model with about 65, 15 and 5 messages respectively. For evaluating the effect of number of nodes, we have limited the number of messages to 5 so that the solver can solve the model with all the four sets of 10, 20, 30 and 40 nodes.

For each of the effective factors in evaluating the model, 20 different sets are generated randomly and the average values are inserted in evaluation charts as final result.

In Figure 1 the percentage of messages successfully delivered by two methods are compared with each other. As can be seen the percentage of messages delivered by linear model is greater than the percentage of messages delivered by sequential scheduling in all conditions. The reason for this difference is the parallel transmission and the increase in number of transmitted messages per unit time by the proposed model. By the solution provided by the suggested model, all the messages that are routed will definitely reach destination within time deadline. It means that the messages with no possibility of delivery in time will not be routed at all. Because the routing of such messages will only serve to waste energy and time of the network and ultimately the message will not be delivered to its destination.

In the worst case, when parallel transmission is not possible, the number of messages delivered by proposed model is equal to the number of messages delivered by sequential scheduling and never less than that.

Increase in number of nodes has resulted in the increase of the delivered messages. Because increase in the number of nodes has resulted in increase in the density of nodes in the region, hence additional nodes are available in the transmission range of the nodes. As a result the messages that could not be transmitted due to limited radio range of the nodes are now sent by multi-hop transmission.

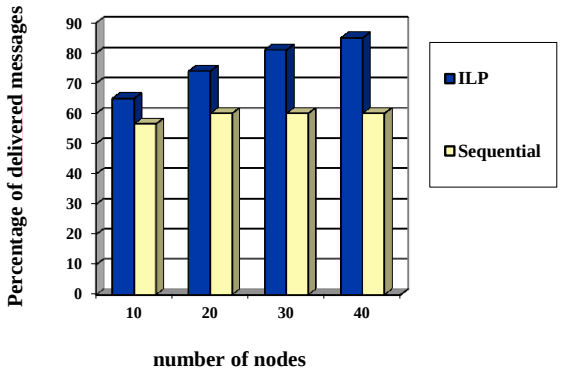


Fig. 1. Percentage of delivered messages

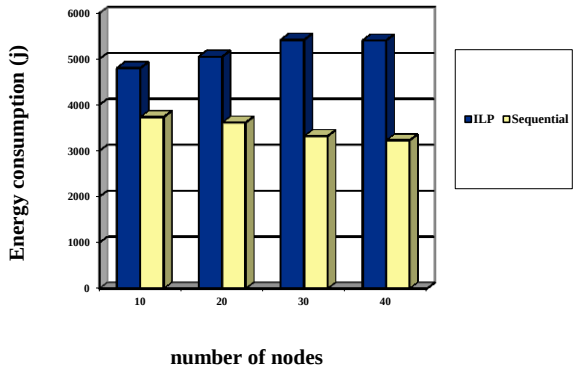


Fig. 2. Energy consumption

In Figure 2 comparison of average energy consumption by proposed model (ILP) and sequential scheduling is done.

In this figure increase in consumption of energy by the proposed model is due to the increase in percentage of delivered messages. It is obvious that more energy will be consumed to send more messages. In the sequential method due to lack of parallel transmission some packets do not meet the time deadline and are dropped. So no energy is consumed for transmission of such messages.

In certain conditions it is possible that in spite of the increase in number of transmitted messages in the proposed model, energy consumption is less than the sequential method. Due to the use of parallel transmission in the proposed model, more time slots will be available and the message will be sent via route having more hops with less distance in between them that results in decrease in energy consumption. For example there are two messages with time deadline of 2 seconds in the network. In sequential scheduling, to avoid violation of the time deadline, each message must be transmitted by a single hop. But in the proposed model, if there is no interference, each of the messages can be transmitted simultaneously using two intermediate hops and will result in reduced energy consumption.

As explained earlier, by increasing the number of nodes in the network, percentage of delivered messages also increases. Naturally, transmission of these additional messages consumes energy that increases the total energy consumption.

However, in some cases it is possible that by increasing the number of nodes, total energy consumption decreases. The reason for this reduction in energy consumption is that the messages previously routed have possibility of using more intermediate hops due to the increased number of nodes. Sometimes the reduction in energy consumption is of so considerable amount that even after compensating for the additional energy required for the transmission of additional messages, total energy consumption is less than that of the scenario when the number of nodes and messages were fewer.

6. Conclusions

In this paper, energy-aware scheduling problem is modeled as linear while keeping in consideration the phenomenon of interference. The proposed model is capable of determining the optimal route for each message by using parallel transmission and multi-hop routing. Therefore, scheduling provided by solving the proposed model can be used as a criterion for evaluating the quality of solutions provided by other methods and algorithms for the problem. But solving the model has high complexities in terms of runtime and memory resources. For this reason, in future we intend to solve this problem by using other existing method for combinational and optimization problems such as meta heuristic methods.

References

[1] H.M. Ammari, "On the energy-delay trade-off in geographic forwarding in always-on wireless sensor networks: A multi-objective optimization problem". Computer Networks, 2013. 57(9): pp. 1913-1935.

[2] A. Mahapatra, , K. Anand, and D.P. Agrawal, "QoS and energy aware routing for real-time traffic in wireless sensor networks", Computer Communications, 2006, 29(4): pp. 437-445.

[3] G.S.A. Kumar, "System Level Energy Management of Networked Real-Time Embedded Systems", Ph.D. thesis, Computer Engineering Department, Iowa State University, Ames, 2009.

[4] G.S.A.Kumar, G. Manimaran, and Z. Wang, "Energy-aware scheduling with deadline and reliability constraints in wireless networks", Broadband Communications, Fourth International Conference on IEEE, BROADNETS 2007.

[5] G.D.Nguyen, et al., "Parallel TDMA Scheduling for Multiple-Destination Wireless Networks". IEEE Transactions on Wireless Communications, 2011. 10(11): pp. 3843-3851.
[6] M. Nazarzadeh, "Proposing an Energy-Aware Scheduling Method with Parallel Transmission capability in Real-time Wireless Network ", M.S.thesis, Islamic Azad University, Malayer, 2013.
[7] Y. Wang, et al., "Interference-aware joint routing and TDMA link scheduling for static wireless networks", IEEE Transactions on Parallel and Distributed Systems, 2008, 19(12): pp. 1709-1726.
[8] P. Gupta, and P.R. Kumar, "The capacity of wireless networks", IEEE Transactions on Information Theory, 2000, 46(2): pp. 388-404.
[9] G.S.A. Kumar, G. Manimaran, and Z. Wang, "Energy-aware scheduling with probabilistic deadline constraints in wireless networks". Ad Hoc Networks, 2009. 7(7): pp. 1400-1413.
[10] M.Davoudzadeh, "Using Bee Colony Optimization Algorithm for Solving the University Timetabling Problem", M.S. thesis, Computer Engineering Department, Islamic Azad University, Arak, 2011.
[11] R. Rafeh, et al., "Towards the new modelling language Zinc", in open source developers: monash university, melbourne, 2005.
[12] L. Shi, A. O.Fapojuwo, "TDMA Scheduling with Optimized Energy Efficiency and Minimum Delay in Clustered Wireless Sensor Networks", IEEE Transactions on Mobile Computing 9, no.7, 2010, pp. 927-940.
[13] M. Al-Ayyoub, H. Gupta, "Joint Routing, Channel Assignment, and Scheduling for Throughput Maximization in General Interference Models", models. IEEE Transactions on Mobile Computing 9, no.4, 2010, pp. 553-565.
[14] X. Xi, X-Y. Li, M. Song "Efficient Aggregation Scheduling in Multihop Wireless Sensor Networks with SINR Constraints", IEEE Transactions on Mobile Computing 12, no. 12, 2013, pp. 2518-2528.
[15] C.J.L. Arachchige, et al., "Minimal time broadcasting in cognitive radio networks", Distributed Computing and Networking, 2011, pp. 364-375.
[16] S. Fan, L. Zhang and Y. Ren, "Optimization-Based Design of Wireless Link Scheduling With Physical Interference Model", IEEE Transactions On Vehicular Technology61, no. 8, 2012, pp. 3705-3717.

Maryam Hamidanvar received her B.Sc. degree in Software Engineering from Bu-Ali Sina University (BASU), Hamedan, Iran, in 2012, and her M.Sc. degree in Software Engineering from Arak University, Arak, Iran, in 2014. Her research interests include modeling of the combinational and optimization problems, modeling languages, constraint programming, wireless networks and application programming.

Reza Rafeh is a faculty member in the Department of Computer Engineering at Arak University (Iran). He got his B.Sc. and M.Sc. in Software Engineering from Sharif University of Technology (Iran). He received his Ph.D. in computer science from Monash University (Australia). His interesting areas are compiler design, constraint programming and theorem proving. He is the chief editor of Software Engineering Journal as well as an editorial board of Soft Computing Journal.

A Hybrid Object Tracking for Hand Gesture Approach based on MS-MD and its Application

Amir Hooshang Mazinan*
Department of Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran
ahmazinan@gmail.com
Jalal Hassanian
Department of Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran
jalalhasanian@gmail.com

Received: 12/Dec/2014 Revised: 16/Aug/2015 Accepted: 19/Aug/2015

Abstract

In the research presented here, the hand gesture recognition is considered in American Sign Language to propose a hybrid approach, which is now organized based on the integration of the mean shift (MS), and the motion information (MD), entitled MS-MD approach. This is in fact proposed to track and recognize the hand gesture, in an effective manner, along with the corresponding potential benchmarks. The investigated results are synchronously acquired in line with these well-known techniques in the area of object tracking to modify those obtained from the traditional ones. The MS scheme is capable of tracking the objects based on its detailed objects, so we have to specify ones, as long as the MD scheme is not realized. In the proposed approach, the advantages of two algorithms are efficiently used, in a synchronous manner, to outperform the hand tracking performance. In the first step, the MD scheme is applied to remove a number of parts without area motion, and subsequently the MS scheme is accurately realized to deal with hand tracking. Subsequently, the present approach is considered to eliminate the weakness of the traditional methods, which are only organized in association with the MS scheme. The results are all carried out on Boston-104 database, where the hand gesture is tracked, in a better form, with respect to the previous existing ones.

Keywords: Hand Tracking; Motion Detection; Mean Shift; Hand Gesture Recognition; American Sign Language.

1. Introduction

Due to the fact that the recognition and tracking of hand gesture are so applicable in both academic and real environments, the new, constructive, impressive and efficient insights in this area can always be appreciated. It means that state-of-the-art in the area of object tracking with a focus on hand tracking is a challenging issue among the experts of the present field. It is known as a crucial and basic ingredient computer vision. Humans can simply recognize and track an object, in an immediate manner, even in the presence of high clutter, occlusion and non-linear variations in the background as well as in the shape, direction or even the size of the object. However, hand tracking can be taken into real consideration as a difficult task for an intelligence-based machine. Tracking for a machine can be defined to find the object states. It is included by position, scale, velocity, feature selecting and many other important parameters that are obtained through a series of images, so object tracking is processed, at each incoming frame, to obtain the weighting coefficients of the entire image. Therefore, it is necessary to specify the object that is recognized in the hand gesture tracking, while a specific one is considered, as the desired object. Many solutions are now proposed to deal with hands motion that they are all used some features of objects to be obtained such as colors, area motion, texture and so on. Firstly, their method converted color frames into gray level images, and then a

kernel function is employed. Furthermore, weights of pixels are obtained for each frame. Their proposed method offers several advantages. For instance, it can be known, as a so resistant approach, against environmental difficulties such as partial occlusion, blurring via camera shaking, deformation of object position and any other sorts of translation. This is due to employing color information as feature vectors in the proposed technique.

Literature's survey in this area could now be presented by considering the work of Nguyen et al., at first, which propose tracking and recognition for facial expressions in American Sign Language; ASL. This work describes towards recognizing facial expressions that are used in sign language communication [1]. Munib et al. suggest the recognition based Hough transform and neural networks. The outcome investigated in this research aims to develop a system for automatic translation of static gestures regarding the alphabets and signs [2]. Coleca et al. describes self-organizing maps for hand and full body tracking [3], where Lee et al. present hand rotation and various grasping gestures tracking from the IR camera via extended cylindrical manifold embedding [4]. Moreover, Morshidi et al. suggest hand tracking through gravity optimized particle filter [5], while Huang et al. research is given in kinematic property of object motion conditions and eye-hand synergy during manual tracking [6]. Also,

* Corresponding Author

there is another novel research, presented by Inoue et al. [7], while robust surface tracking in range image sequences is investigated by Husain et al. [8]. Moreover,

electronics control is proposed by Premaratne et al. This work is based on the well-known wave controller technology [9]. Hereinafter, Ge et al. research is in hand gesture recognition and tracking via the distributed linear embedding [10], once González-Ortega work is to realize real-time hands, facial features detection and tracking with its application to cognitive rehabilitation tests monitoring [11]. Cui et al. propose model-based visual hand posture tracking for guiding a dexterous robotic hand [12], while Kodagoda et al. present simultaneous tracking and motion pattern learning [13]. Moreover, Guazzini et al. research is to present cognitive dissonance and social influence effects on preference judgments for eye tracking based system for their automatic assessment, as another work in this area [14]. Kılıboz et al. present a hand gesture recognition technique for human–computer interaction [15], where Li et al. propose feature learning based on SAE–PCA network for human gesture recognition in RGBD images [16]. Rautaray et al. suggest vision based hand gesture recognition for human computer interaction [17], while K. Li et al. exploit object tracking method based on mean shift and particle filter [18]. Zhang et al. introduce real-time hand gesture tracking based on particle filtering [19], where Hsia et al. present moving target tracking based on camshaft approach [20]. Shan et al. discuss real-time hand tracking through a mean shift embedded particle filter [21]. Finally, Jacob et al. exploit context-based hand gesture recognition for the operating room [22], where Kong et al. focus on independent continuous sign language recognition, as well [23].

The key goal of the approach proposed here is to present a tracking system to be applicable in the area of American Sign Language with acceptable accuracy and the appropriate time consumed, where the outcomes could be competitive with respect to the potential benchmark. Moreover, the bottlenecks of the present approach are that the limitations in choosing the objects to be tracked, the appropriate thresholds to be initialized and finally the hardware performance are existed.

The rest of the present manuscript is organized as follows: in Section 2, the proposed approach is presented. The experiments and results are shown in Section 3 and finally Section 4 draws the conclusions.

2. The Proposed Approach

The proposed approach, as an integration of two well-known algorithms including the mean shift (MS) scheme and the motion detection (MD) scheme is now presented, as a hybrid hand gesture tracking, which is working in the presence of occlusion problem, as soon as hand may be put on the face, especially. In this approach, the hand model is first determined by an authorized user, as a main tracking object, in the first frame. Then, motion range is obtained via the MD scheme to eliminate the move less parts, so having the less pixels results in minimum time for hand tracking. Hand color feature is extracted from this specific space by the MS scheme and a number of pixels are determined in the same space. Subsequently, pixels that having less intensity as compared to the hand color could be eliminated. Remained pixels are weighted in accordance with the distance from the hand center and, in the next frames, the pixels weight of the same hand center are tracked by the MS scheme. The schematic diagram of the proposed approach is sketched in Fig. 1.

Regarding the present approach, it should be noted that the main contribution of this research is to realize a new hybrid of the well-known MS scheme in association with the MD scheme, as long as its application is efficient in the specific application of object tracking. Furthermore, it is needed to note that the performance of this algorithm depends on predefined thresholds. It means that the appropriate value for this parameter needs to be first initiated, in its correct form, prior to running the program in the present research. It should be noted that the number of identified objects may correspondingly be increased, while the value of this parameter is initiated lower than its nominal one. Subsequently, by choosing the same parameter higher than its nominal one, the number of identified objects may correspondingly be decreased. In both cases, the approach performance may be poor by reaching object tracking errors. In fact, the contribution of the present research has to be twofold. First of all, the main approach of the research is recently considered in the area of paper's topic. Second, the results are somehow competitive w. r. t. the other related potential ones, as considered in the proceeding subsections.

2.1 The MD Scheme Realization

The MD scheme is a well-known approach in the area of object tracking, which is in fact faster with respect to the MS scheme-based tracker. Also, the present MD scheme is the simplest way in case of the three tasks including detection, estimation and segmentation, respectively. It should be noted that its goal is to identify which image points, or more generally which regions of the image have moved between two instants of time. The MD scheme is realized to compare the current frame with the previous one. The results are useful in video compression, as long as it is needed to estimate changes and to write these changes instead of frames. This algorithm presents an image with white pixels (motion level) in the place, where the current frame is different from the previous one. It is already possible to count these pixels and if the amount of these pixels may be greater than a predefined threshold level, this is produced as a motion event. The motion detection is calculated though

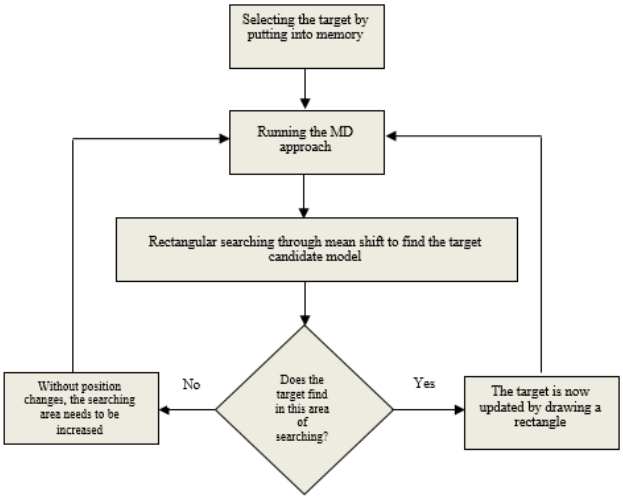


Fig.1. The schematic of the proposed approach.

the distance in the luminance space between the current image $I_k(x)$ and the last aligned image $\hat{I}_{k-1}(x)$, by obtaining the difference image $DM_k(x)$, defined as

$$DM_k(x) = \begin{cases} mIncrement; & |I_k(x) - \hat{I}_{k-1}(x)| > T_m \\ 0; & Otherwise \end{cases} \quad (1)$$

where mIncrement corresponds to a factor of incrementing in the motion and T_m corresponds to a motion threshold. $DM_k(x)$ contains the initial set of points that are taken as a candidate to be belonged to the moving visual object. In order to consolidate the blobs to be detected, a 3×3 morphological closing is applied to $DM_k(x)$. Isolated detected moving pixels are discarded, when applying to a 3×3 morphological opening. The representation of the motion history image $MH_k(x)$ is then updated by multiplying the last motion history representation $MH_{k-1}(x)$ with a decay factor and by adding the difference image $DM_k(x)$, where it could be written as $MH_k(x) = MH_{k-1}(x)\,.\,\text{decay factor} + DM_k(x)$. Finally, all pixels of $MH_k(x)$, whose luminance is over the MD scheme threshold are considered as pixels in motion. These pixels generate the detection image $DH_k(x)$, defined as

$$DH_k(x) = \begin{cases} 1; & MH_k(x) > T_k \\ 0; & Otherwise \end{cases} \quad (2)$$

A 3×3 morphological closing is now applied to the detection image $DH_k(x)$, followed by a 3×3 morphological opening.

2.2 The MS Scheme Realization

2.2.1 Target Representation

The MS scheme is realized to characterize the object, at first, while a feature space needs to be chosen. The reference object model is represented by its pdf q in the feature space. In the subsequent frame, an object candidate must be defined in location y and is characterized by the pdf $\hat{p}(y)$. Both pdfs are to be estimated from the data. To satisfy the low computational cost, it is imposed by real-time processing discrete densities, i.e., m -bin histograms. Thus, there are the object model and its candidate as

$$\begin{cases} \hat{q} = \{\hat{q}_u\}_{u=1,2,\cdots,m};\ \sum_{u=1}^{m} \hat{q}_u = 1 \\ \hat{p} = \{\hat{p}_u(y)\}_{u=1,2,\cdots,m};\ \sum_{u=1}^{m} \hat{p}_u = 1 \end{cases} \quad (3)$$

Now, the histogram is not the best nonparametric density estimation, but it suffices for our purposes. Other discrete density estimation can also be employed, where it denotes by $\hat{p} = \rho[\hat{p}(y), \hat{q}]$as a similarity function between $\hat{p}(y)$ and $\hat{q}$. The function $\hat{p}(y)$ plays the important role of likelihood and its local maximum in the image indicating the presence of objects in the second frame that having representations, which are similar to $\hat{q}$, defined in the first frame. To find the maximum of such functions, gradient-based optimization procedures are difficult to apply and only an expensive exhaustive search can be used. We regularize the similarity function by masking the objects with an isotropic kernel in the spatial domain. As long as the kernel weights, carrying continuous spatial information, are used in defining the feature space representations, $\hat{p}(y)$ becomes a smooth function in y.

2.2.2 Target Model

An object is represented by an ellipsoidal or rectangular region in the image. To eliminate the influence of different object dimensions, all the objects are first normalized to be taken in a unit circle. This is achieved by independently rescaling with the row and column dimensions along with h_x and h_y, respectively. Let us note $\{x_i^*\}_{u=1,2,\cdots,n}$ as the normalized pixel locations in the region, namely the object model. The region is centered at zero. An isotropic kernel, with a convex and monotonic decreasing kernel profile $k(x)$, assigns smaller weights to pixels, which are farther from the center. The function $b: R^2 \to \{1,2,\cdots,m\}$ associates to be the pixel at location x_i^* and also the index $b(x_i^*)$ of its bin in the quantized feature space. The probability of the feature $u = 1,2,\cdots,m$ in the object model is then calculated as $\hat{q}_u = C\sum_{i=1}^{n} k(\|x_i^*\|^2)\delta[b(x_i^*) - u]$, where $k(x)$ is the Kronecker delta function. The normalization constant $C = \frac{1}{\sum_{i=1}^{n} k(\|x_i^*\|^2)}$ is derived by imposing the condition $\sum_{u=1}^{m} \hat{q}_u = 1$, where the summation of delta functions for $u = 1,2,\cdots,m$ is equal to be one.

2.2.3 Target Candidates

Let us note $\{x_i\}_{u=1,2,\cdots,n_k}$ as the normalized pixel locations of the object candidate, centered at y in the current frame. Using the same kernel profile $k(x)$, with bandwidth h, the probability of the feature $u = 1,2,\cdots,m$ in the object candidate could be given by $\hat{p}_u(y) = C_h \sum_{i=1}^{n_k} k(\left\|\frac{y-x_i}{h}\right\|^2)\delta[b(x_i) - u]$ (4)

where $C = \frac{1}{\sum_{i=1}^{n_k} k(\left\|\frac{y-x_i}{h}\right\|^2)}$ is taken. Note that C_h cannot depend on y, as long as the pixel locations x_i are organized in a regular lattice and y is taken as one of the lattice nodes. Therefore, C_h can be recalculated for a given kernel and different values of h, while the bandwidth h defines the scale of the object candidate, i.e., the number of pixels, considered in the localization process.

2.2.4 Similarity Function Smoothness

The similarity function inherits the properties of the kernel profile k(x), while the object model and candidate are represented. The employed object representations are not restricting the way, which similarity is measured and various functions can be used for ρ.

2.2.5 Metric based on Bhattacharyya Coefficient

The similarity function illustrates a distance among object model and its candidates. To accommodate comparisons in the various objects, this should have a metric structure. The distance between two discrete distributions is now given by $d(y) = \sqrt{1 - \rho[\hat{p}(y), \hat{q}]}$ where $\hat{p}(y) = \rho[\hat{p}(y), \hat{q}] = \sum_{u=1}^{m} \sqrt{\hat{p}(y)\hat{q}}$ is taken. It could be noted that the cosine of the angle between the m-dimensional unit vectors $(\sqrt{\hat{p}_1}, \sqrt{\hat{p}_2}, \cdots, \sqrt{\hat{p}_m})^T$ and also $(\sqrt{\hat{q}_1}, \sqrt{\hat{q}_2}, \cdots, \sqrt{\hat{q}_m})^T$ are given. Note that the L_P histogram metrics including histogram intersection cannot satisfy the conditions $\sum_{u=1}^{m} \hat{p}_u = 1$ and $\sum_{u=1}^{m} \hat{q}_u = 1$.

2.2.6 Target Localization

In order to find the location, corresponding to the object, in the current frame, the distance should be minimized as a function of y. The localization procedure starts from the position of the object in the previous frame, i.e. the model and by searching in the neighborhood. Since the distance function is smooth, the procedure uses gradient information, provided by the MS scheme vector. More involved optimizations can be applied. Color information could be chosen as the object feature; however, the same framework may be used for texture and edges or any combination of them. In the sequel, it is assumed that the following information is available, i.e. (1) detection and localization in the initial frame of the objects to track (object models); (2) periodic analysis of each object to account for possible updates in case of the object models, due to significant changes in color.

2.2.7 Distance Minimization

Minimizing the distance is equivalent to maximize the Bhattacharyya coefficient $\hat{p}(y)$. It should be noted that the search for the new object location, in the current frame, starts at the location $\hat{y}_0$ of the object in the previous frame. It is obvious that the probabilities $\{\hat{p}_u(\hat{y}_0)\}_{u=1,2,\cdots,m}$ of the object candidate at location $\hat{y}_0$, in the current frame, need to be computed, firstly. Using Taylor expansion around the values $\hat{p}_u(\hat{y}_0)$, the linear approximation of the Bhattacharyya coefficient could be acquired after some manipulations as

$$\rho[\hat{p}(y), \hat{q}] \approx \frac{1}{2}\sum_{u=1}^{m} \frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)} + \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{y}_0)\hat{q}_u} \quad (5)$$

The approximation is satisfactory, as long as the object candidate $\{\hat{p}_u(\hat{y}_0)\}_{u=1,2,\cdots,m}$ cannot change, drastically, from the initial $\{\hat{p}_u(\hat{y}_0)\}_{u=1,2,\cdots,m}$. The condition $\hat{p}_u(\hat{y}_0) > 0$ or some small threshold for all $u = 1,2,\cdots,m$ can always be enforced to avoid using the feature values in violation. Recalling the outcomes could be used in

$$\rho[\hat{p}(y), \hat{q}] \approx \frac{C_h}{2}\sum_{u=1}^{n_k} w_i\, k(\left\|\frac{y-x_i}{h}\right\|^2) + \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{y}_0)\hat{q}_u} \quad (6)$$

where $w_i = \sum_{u=1}^{m} \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)}}\delta[b(x_i) - u]$. The mode of this density in the local neighborhood is the sought maximum which can be found by employing the MS scheme. In this procedure, the kernel is recursively moved from the current location $\hat{y}_0$ to the new location $\hat{y}_1$ in accordance with the following relation

$$\hat{y}_1 = \frac{\sum_{u=1}^{n_k} x_i w_i\, g(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2}{\sum_{u=1}^{n_k} w_i\, g(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2} \quad (7)$$

Here, $g(x) = -k'(x)$ is taken as the derivative of k(x) for all $x \in [0, \infty)$, except for a finite set of points. The complete object localization algorithm is now presented in the proceeding steps: (1) Present the object model $\{\hat{q}_u\}_{u=1,2,\cdots,m}$ and its location $\hat{y}_0$ in the previous frame, (2) Initialize the location of the object in the current frame with $\hat{y}_0$, compute $\{\hat{p}_u(\hat{y}_0)\}_{u=1,2,\cdots,m}$, and finally evaluate $\rho[\hat{p}(y_0), \hat{q}] = \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{y}_0)\hat{q}_u}$. (3) Derive the weights $\{w_i\}_{i=1,2,\cdots,n_k}$, (4) Find the next location of the object candidate, (5) Compute $\hat{p}_u(\hat{y}_1)_{u=1,2,\cdots,m}$ and evaluate $\rho[\hat{p}(y_1), \hat{q}] = \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{y}_1)\hat{q}_u}$, (6) While $\rho[\hat{p}(y_1), \hat{q}] < \rho[\hat{p}(y_0), \hat{q}]$ is satisfied, there is $\hat{y}_1 = \frac{1}{2}(\hat{y}_0 + \hat{y}_1)$ toevaluate $\rho[\hat{p}(y_1), \hat{q}]$, (7) If $\|\hat{y}_1 - \hat{y}_0\| < \varepsilon$stop processing, otherwise $\hat{y}_1 = \hat{y}_0$is taken and go to step (3).

2.2.8 Implementation of the Algorithm

Implementation of the proposed tracking algorithm can be much simpler than the presented above. The stopping criterion threshold is initiated, which is derived by constraining the vectors $\hat{y}_0$ and $\hat{y}_1$ to be within the same pixel, in the original image coordinates. A lower threshold may induce sub pixel accuracy. From real-time constraints, i.e., the traditional CPU performance in time, the number of the MS scheme iterations is bounded toN_{max}, typically, taken to be 20. In practice, the average number of iterations is much smaller, about 4. The role of step 5 is only to avoid potential numerical problems in the MS scheme based maximization. These problems can appear due to the linear approximation of the Bhattacharyya coefficient.

However, a large set of experiments in tracking the different objects in lengthy periods of time has shown that the Bhattacharyya coefficients are calculated at the new location $\hat{y}_1$, failed to increase only 0.1% of the cases. Therefore, the Step 5 is not used in practice, and as a result, there is no need to evaluate the Bhattacharyya coefficient in Steps 1 and 4. In the practical algorithm, the weights in Step 2 are iterated, deriving the new location in Step 3, and testing the size of the kernel shift in Step 6. The Bhattacharyya coefficient is resulted only after completing the algorithm to evaluate the similarity between the object model and its chosen candidate. The kernels with Epanechnikov profile is described as

$$k(x) = \begin{cases} \frac{1}{2} C_d^{-1} (d+2)(1-x); & x \le 1 \\ 0; & \text{Otherwise} \end{cases} \quad (8)$$

In this case, the derivative of the profile $g(x)$ is constant and the result reduces to $\hat{y}_1 = \frac{\sum_{u=1}^{n_k} x_i w_i}{\sum_{u=1}^{n_k} w_i}$, i.e. a simple weighted average. The maximum of the Bhattacharyya coefficient can also be interpreted, as a matched filtering procedure. The MS scheme procedure finds the local maximum of the scalar field of correlation coefficients; due to the use of kernels.

3. The Experiments and Results

The proposed approach performance is now considered through a sequence of different images. At first, the whole of experiments regarding the approach presented here have been entirely implemented via 2013 MATLAB programming language. In one such case, the proposed approach is implemented on a 1.6GHz T2050 under 1GB of RAM, while the frame dimensions are first taken as 466×654 pixels.

In this case, the inputs and the related outputs regarding the present simulation, in its application point of view, are simply related to the number of objects to be chosen and also the number of the same objects to be correspondingly tracked, respectively. Moreover, in the algorithm point of view, the inputs are taken as target model and target candidates, as long as the outputs are taken as the outcomes regarding the similarity function smoothness through Bhattacharyya coefficient and target localization, as well. Figure 2 depicts the hand tracking in the usual image through realizing the MS scheme. Here, four frames including 21, 38, 45 and finally 76 regarding a total of 100 frames have been illustrated. In this case, hand movements are slow and also its resolutions and the corresponding image qualities are somehow acceptable. As is seen, this algorithm could not track the right hand, as an object, properly, in all the simulation time.

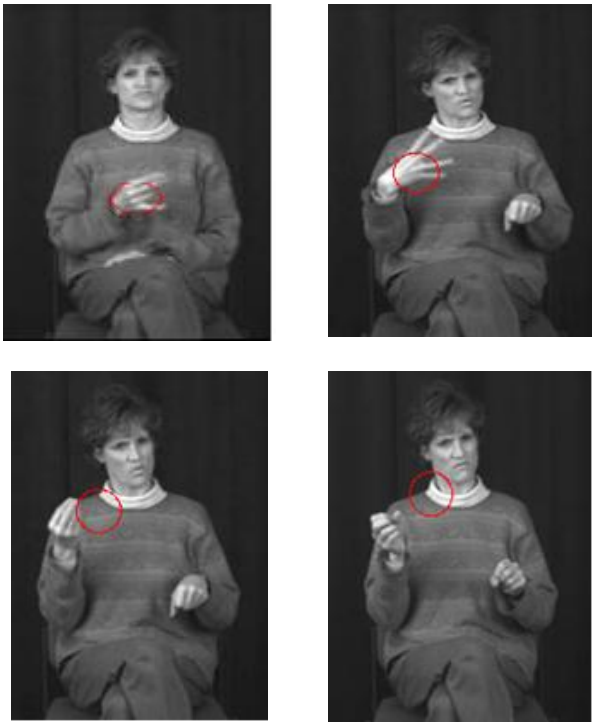


Fig. 2. The hand tracking in a sequence of images though the MS scheme algorithm.

Figure 3 shows the hand tracking in a sequence of usual images through the MD scheme. According to the results, this algorithm could not track just right hand, as an object, accurately. As is obvious, in a number of frames, the face and the left hand have been chosen, as an object. Figure 4 shows the hand tracking, in a sequence of usual images, through the PF scheme tracking the hand but needed lengthy processing time.

Figure 5 illustrates the hand tracking, in the usual image, though the integration of the MD scheme and the MS scheme, as the proposed approach. This one could track just right hand, as an object, in an accurate manner, w. r. t. the mentioned algorithms. Moreover, Fig. 6 illustrates the tracking error in the whole of three considered algorithms. In the MS scheme, object cannot be tracked in all frames because, after several frames, the occlusion happens and its error increases, as well. In the MD scheme, the error does not increase. Not only the right hand is tracked, but also all the moving things are tracked, as is clear to point out.

Fig. 3. The hand tracking in a sequence of images through the MD scheme algorithm.

Finally, in the proposed combination approach; CMP, it can track the hand, so the corresponding errors of all the frames are negligible. Table 1 tabulates the details of the hand tracking in the present above-mentioned algorithms. In all ones, the tracking error has been obtained by subtracting the center of hand and its object.

Fig. 4. The hand tracking in a sequence of images through the PF scheme algorithm

Fig. 5. The hand tracking in a sequence of images through the proposed approach

As is now considered, the processing time of the proposed approach is less than the MD scheme about half. Also, the mean position error in the MD scheme is more than the proposed approach. It is notable that the MS scheme cannot track the hand, continuously, and lose it in the middle of the way.

In one such case, the Bhattacharyya coefficients regarding the experiments, presented in Fig. 5, are illustrated in Fig. 7. It is to evaluate the similarity between the object model and the chosen candidate. As are obvious, this figure indicates that the process of tracking of the chosen object is well behaved. In making an effort to make in considering the effectiveness of the approach proposed here, Fig. 8 is presented in comparison with the MS scheme in the present complicated images sequence. The resolution and its quality regarding the images in this figure are not so desirable and also the processed images are noisy and blurring.

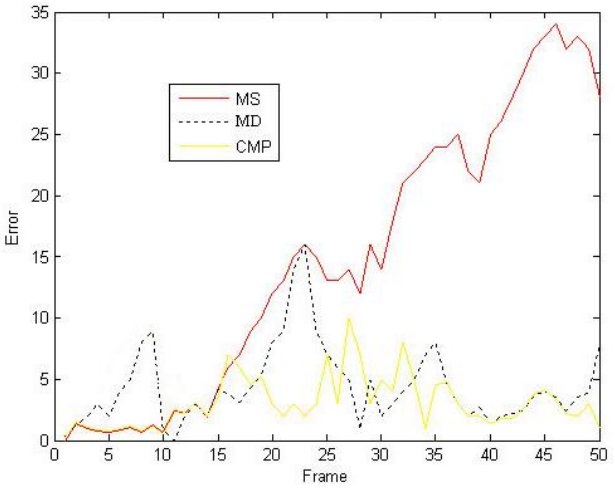


Fig. 6. The tracking error in three algorithms: the MS scheme (red), the MD scheme (dash), the proposed combination (yellow).

And hand movements are faster than the old ones, but, as is seen, this could track the hand, as an object, efficiently. These images have been taken from the forum Iran's site, although the old pictures are related to the main database, i.e. RWTH-BOSTON-104.

In order to evaluate and compare the proposed approach with a number of conventional methods, a number of tracking techniques including the MD scheme, the MS scheme, the PF scheme and finally the integration of the PF scheme along with the MS scheme are all presented in Table 1 [20]-[22].

Concerning the proposed approach, the integration of the MS scheme along with the MD scheme is realized to be applicable to remove fixed parts. It aims us to be able to track the hand, accurately and efficiently, after choosing the target. Moreover, regarding the specification of the PF scheme, a number of pixels need to be processed, in its lengthy processing time, for each one of incoming frames to realize a tracking system. Hereinafter, regarding the specification of the MS scheme, it is noted that the speed of the present tracking system run time is higher than the mentioned algorithm.

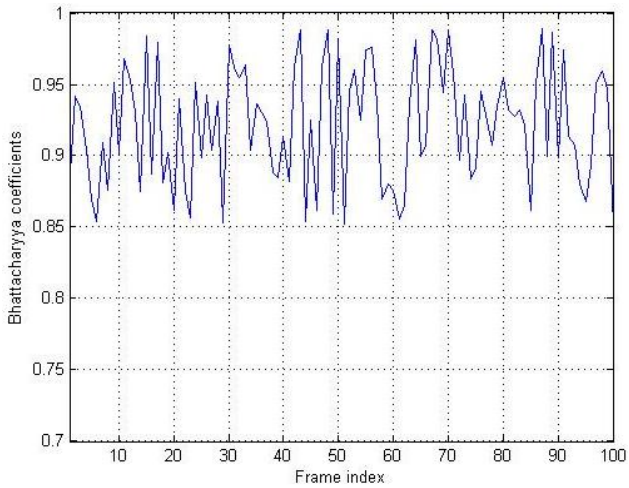


Fig. 7. The Bhattacharyya coefficients in the experiments regarding Fig. 6.

In fact, there are some potential approaches to be considered as comparable methods, while the integration of the PF scheme along with the MS scheme is one of the hybrid solutions, presented.

Now, as considered, both speed of processing time and the tracking accuracy are better than others. Moreover, it should be noted that, the proposed approach has lower complexity, compared to the same conventional tracking methods. Moreover, to consider the effectiveness of the approach proposed w. r. t. the potential benchmarks in this area, at first, four of them are chosen to be analyzed with respect to state-of-the-art. It is apparent that with a focus on the whole of methodologies, investigated in the benchmarks, as well as the potential results, illustrated in [20], the topic of the present research can be explained. It should be noted that the aforementioned reference is a deep collection of the approaches that are directly related to the proposed research topic that is published, in recent year. It aims us to see the new investigations in this field. With regard to the results illustrated there, at first, the idea of the proposed research can be of novelty and therefore its performance is then needed to consider, carefully. It means that the overall performance of the approach proposed here is to be resulted by considering

Fig. 8. The comparison between the MS scheme and the proposed approach in a sequence of the complicated images.

processing time, the accuracy criteria and also the type of environments that were experimentally processed.

Table 1. The comparable outcomes of the proposed approach along with four related tracking schemes.

	Algorithms	Needed pixels	Mean time per frame (msec.)	Mean position error (pixels)
1	The MD scheme	variable	2.5	8
2	The MS scheme	20	1.0	10
3	The PF scheme	150	3.5	8
4	The PF+MS scheme	50	2.5	7
5	The proposed approach	20	1.8	5

Now, the definition of the basic concept for the accuracy criteria (η), and its tracking errors (ℓ), are now given, respectively, by Eq. (9)

$$\begin{cases} \eta = \dfrac{TP_p}{TP_P + FN_P} \\ \ell = \dfrac{\overline{TP_p}}{TP_p + FN_p} \end{cases} \quad (9)$$

Here TP_p is taken as the objects in the frames that have correctly verified, while $\overline{TP_p}$ is taken as the objects in the frames that have mistakenly verified. And finally $TP_p + FN_p$ is the total of entities in the frames that have considered in the process of surveying, as well.

With this purpose, the performance results of the four potential benchmarks along with the proposed approach are now tabulated in Table 2. Now, with regard to the present outcomes, it is wise to note that the superior results with respect to the accuracy outcomes are directly related to the Li's experiments, while the tracking performances of the proposed approach and the Jacob's experiments are located in the second ranking position. The performance of the Kong's experiments and the Kiliboz's experiments are in the third and the forth ranking positions, respectively. Of course, another heuristic factor aims us to finalize the present comparison process. It is also considered to illustrate the complication of the environments that are experimentally processed. It means that the percentage of a very complicated frame is now scored as 100, while other related frames are correspondingly scored below versus this percentage value in order. This is obvious that the more score happens; the more environmental complexity acquires, in the corresponding matter. Now, by considering the whole of factors in one such case, in a careful manner, it is clear to note that the superiority of the proposed approach could somehow be verified in the present cases with respect to other related benchmarks. In making an effort to be realized, in the same way, the processing time regarding the approach presented here in comparison with the whole of mentioned benchmarks are considered. The results in one such case are coherently related to the hardware and the corresponding software performances. Due to the fact that the potential benchmarks are appeared, in recent years, it can be supposed to take the proposed approach hardware specification to be the same as other related ones and therefore the comparisons can be somehow meaningful. In this way, the average processing time consumed per each frame is illustrated in Fig. 9 (in msec.). The present frames processing time is only reported in the Kiliboz's benchmark as 1500 msec. in the about 750 successful processed frames, while this factor in the proposed approach is about 2 msec. The outcomes regarding the time consumed indicate that the frames processing time is somehow competitive w. r. t. the corresponding one.

4. Conclusions

A hybrid MS-MD approach has been now investigated in the present research to organize an insight in the area of hand tracking in high robust performance in the presence of complicated images. The proposed approach uses color features, which can be obtained of the frames, while the MD scheme is realized in partial. The present MD is in fact applied to remove points in association with

the MS scheme to track the hand gesture. The approach proposed here is carried out to track different kind of hands, as the objects, in the real and complicated environments. As is obvious to us, the proposed approach can track the hand gesture that is somehow better than the traditional algorithms, in general. Furthermore, this is in line with some other potential algorithms, realized in the area of object tracking, such as dynamic programming tracking, which is now desirable in real-time domains. Due to the fact that using the MS-MD scheme is somehow efficient and applicable, the less computation is coherently needed to track the object and it is faster than other corresponding algorithms, as it is two times faster than the dynamic programming tracking in the same hardware performance. In addition, less memory is needed to track the object and therefore the outcomes can be useful to implement in some practical environments.

Table 2. The accuracy criteria; η, the tracking error; ℓ, and the frame complication; ρ, considering the benchmarks as well as the proposed approach.

		$\%\eta$	$\%\ell$	$\%\rho$
1	The Kılıboz'sexperiments [15]	73.0	27.0	80
2	The Li's experiments [16]	99.0	1.0	80
3	The Jacob's experiments [22]	92.3	7.7	90
4	The Kong's experiments [23]	89.9	10.1	100
5	The proposed approach	90.4	9.6	90

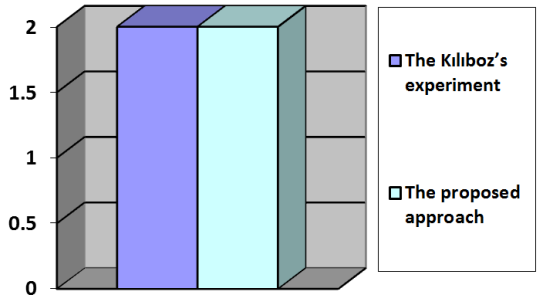


Fig. 9.The average time consumed per each frame (msec.).

Acknowledgements

The corresponding Author would like to express the warmest and best regards to respected Editors of JIST Journal. Moreover, special thanks to all the potential reviewers for the technical, valuable and constructive comments in the process of paper modifications.

References

[1] Tan Dat Nguyen, Surendra Ranganath, Facial expressions in American Sign Language: Tracking and recognition, *Pattern Recognition*, Volume 45, Issue 5, May 2012, Pages 1877-1891.

[2] Qutaishat Munib, Moussa Habeeb, Bayan Takruri, Hiba Abed Al-Malik, American sign language (ASL) recognition based on Hough transform and neural networks, *Expert Systems with Applications*, Volume 32, Issue 1, January 2007, Pages 24-37.

[3] Foti Coleca, Andreea State, Sascha Klement, Erhardt Barth, Thomas Martinetz, Self-organizing maps for hand and full body tracking, *Neurocomputing*, Volume 147, 5 January 2015, Pages 174-184.

[4] Chan-Su Lee, Sung Yong Chun, Shin Won Park, Tracking hand rotation and various grasping gestures from an IR camera using extended cylindrical manifold embedding, *Computer Vision and Image Understanding*, Volume 117, Issue 12, December 2013, Pages 1711-1723.

[5] Malik Morshidi, TardiTjahjadi, Gravity optimised particle filter for hand tracking, *Pattern Recognition*, Volume 47, Issue 1, January 2014, Pages 194-207.

[6] Chien-Ting Huang, Ing-Shiou Hwang, Kinematic property of object motion conditions gaze behavior and eye-hand synergy during manual tracking, *Human Movement Science*, Volume 32, Issue 6, December 2013, Pages 1253-1269.

[7] Yasuyuki Inoue, Yutaka Sakaguchi, Periodic change in phase relationship between object and hand motion during visuo-manual tracking task: Behavioral evidence for intermittent control, *Human Movement Science*, Volume 33, February 2014, Pages 211-226.

[8] Farzad Husain, Babette Dellen, Carme Torras, Robust surface tracking in range image sequences, *Digital Signal Processing*, Volume 35, December 2014, Pages 37-44.

[9] Prashan Premaratne, Sabooh Ajaz, Malin Premaratne, Hand gesture tracking and recognition system using Lucas–Kanade algorithms for control of consumer electronics, *Neurocomputing*, Volume 116, 20 September 2013, Pages 242-249.

[10] S.S. Ge, Y. Yang, T.H. Lee, S.S. Ge, Y. Yang, T.H. Lee, Hand gesture recognition and tracking based on distributed locally linear embedding, *Image and Vision Computing*, Volume 26, Issue 12, 1 December 2008, Pages 1607-1620.

[11] D. González-Ortega, F.J. Díaz-Pernas, M. Martínez-Zarzuela, M. Antón-Rodríguez, J.F. Díez-Higuera, D. Boto-Giralda, Real-time hands, face and facial features detection and tracking: Application to cognitive rehabilitation tests monitoring, *Journal of Network and Computer Applications*, Volume 33, Issue 4, July 2010, Pages 447-466.

[12] Jinshi Cui, Zengqi Sun, Model-based visual hand posture tracking for guiding a dexterous robotic hand, *Optics Communications*, Volume 235, Issues 4–6, 15 May 2004, Pages 311-318.

[13] Sarath Kodagoda, Stephan Sehestedt, Simultaneous people tracking and motion pattern learning, *Expert Systems with Applications*, Volume 41, Issue 16, 15 November 2014, Pages 7272-7280.

[14] Andrea Guazzini, Eiko Yoneki, Giorgio Gronchi, Cognitive dissonance and social influence effects on preference judgments: An eye tracking based system for their automatic assessment, *International Journal of Human-Computer Studies*, Volume 73, January 2015, Pages 12-18.

[15] Nurettin Çağrı Kılıboz, Uğur Güdükbay, A hand gesture recognition technique for human–computer interaction, *Journal of Visual Communication and Image Representation*, in press,2015.

[16] Shao-Zi Li[a, b]Author Vitae,
- Bin Yu[a, b, c]Author Vitae,
- Wei Wu[a, b]Author Vitae,
- Song-Zhi Su[a, b]Author Vitae,
- Rong-Rong Ji[a, b, ,] Author Vitae

[17] Shao-Zi Li, BinYu, WeiWu, Song-ZhiSu, Rong-Rong Ji, Feature learning based on SAE–PCA network for human gesture recognition in RGBD images, *Neurocomputing*, 2015, Pages 565–573.

[18] Siddharth S. Rautaray, Anupam Agrawal, Vision based hand gesture recognition for human computer interaction: a survey, *Artificial Intelligence Review*, Volume 43, Issue 1, January 2015, Pages 1-54.

[19] K. Li, Xu KH, Huang DS (2012) Improved object tracking method based on mean shift and particle filter. J Comput Appl 32(2):504–506.

[20] Qiuyu Zhang, Jingman Liu, Hongxiang Duan, Daodong Wang, Hand gesture tracking based on particle filtering aiming at real-time performance. J Computational Science 10(4):1149–1157.

[21] Kuo-hsien Hsia, S.F. Lien, J. P> Su, Moving target tracking based on camshaft approach and kalman filter, Applied Mathematics Information Sciences, 7(1), 2012, 193-200.

[22] C. F. Shan, T. N. Tan, Y. C. Wei, Real-time hand tracking using a mean shift embedded particle filter, Pattern Recognition, 40(7), 2007, 1958-1970.

[23] Mithun George Jacob, Juan Pablo Wachs, Context-based hand gesture recognition for the operating room, *Pattern Recognition Letters*, Volume 36, 15 January 2014, Pages 196–203.

[24] W.W. Kong, Author VitaeSurendra Ranganath, Author VitaeTowards subject independent continuous sign language recognition: A segment and merge approach, *Pattern Recognition*, Volume 47, Issue 3, March 2014, Pages 1294–1308.

Amir Hooshang Mazinan received the Ph.D. degree in 2009 in Control Engineering. Dr. Mazinan is the Associate Professor and also the Director of Control Engineering Department at Islamic Azad University (IAU), South Tehran Branch, Tehran, Iran. He is now acting as Associate Editor in TIMC Journal (SAGE publisher) and Guest Editor in CAEE Journal (Elsevier Publisher), as well. Moreover, he is a member of Editorial Board in three international journals and also a member of programming committee in four international conferences. Dr. Mazinan has more than 80 journal and conference papers in so many reputable publishers. His current research interests include intelligent systems, model based predictive control, over actuated mechanical system modeling and control, time–frequency representation, filter banks, wavelet theory and image-video processing.

Jalal Hassanian received the M.Sc. degree in 2014 in Electrical Engineering at the Islamic Azad University (IAU), South Tehran Branch, Tehran, Iran. His current research interests include signal, image and video processing.