

A Clustering Expert System Using Particle Swarm Optimization And K-Means++ For Journal Recommendation To Publish The Papers

Seyedeh Malihe Khatami¹, Mansoureh Maadi², Rohollah Ramezani³

^{1,3}School of Engineering, Damghan University, Damghan, Iran

²School of Mathemati and Computer Science, Damghan University, Damghan, Iran

Article Info

Article history:

Received Apr 9, 2018

Revised May 20, 2018

Accepted Jul 11, 2018

Keywords:

Fuzzy

Particle Swarm Optimization

Clustering

Journal recommendation system.

ABSTRACT

In this paper, an android expert system for recommending the suitable journal for publishing the researchers' papers has been presented. In so doing, the expectations of different journals for accepting an article and also the requests of papers' writers for choosing the journals have been examined. Language, quality, waiting time for judgment, waiting time for publication after acceptance, field, length of paper and cost are the system inputs and its output is the degree of suitability of journals for publishing a certain paper. This system includes a database of different journals and their parameters. It uses particle swarm optimization method and K-means++ algorithm for assessing and clustering the journals database and determines an index for every cluster of journals. The process for matching the paper with a cluster's index is done through fuzzy induction system. After choosing the most similar cluster, the paper is compared with all the journals of the cluster in the same way and the result including the most similar journals is presented. This system has been tested for journals and papers in the computer field indexed in Elsevier.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Seyedeh Malihe Khatami

School of Engineering,

Damghan University, Damghan, Iran.

Email: faroughi.alireza.h@gmail.com

1. INTRODUCTION

Collective intelligence and fuzzy logic generate powerful and android precise system. In recent years, a lot of attempts have been made to produce android systems to use the capabilities of both fuzzy and collective intelligence methods such as hybrid systems of fuzzy particle swarm optimization algorithm [1-2], ant colony optimization algorithm with parallel fuzzy cluster [3], modified artificial bee colony algorithm [2], [4], etc. In addition, there is a significant increase in successful use of hybrid android systems in many different issues including data clustering [5-6], data categorization, medical diagnosis systems, picture process, handwriting recognition, etc.

Expert system is a branch of artificial intelligence which was introduced and developed by artificial intelligence committee in the mid-1960s. An expert system is a computer program which aims at knowledge modeling in one or more technical fields. The main idea of an expert system is to transfer and expert's knowledge to computer so that through induction it could have some suggestions for users regarding this recorded technical knowledge and explain the logic behind each suggestion, if necessary. Expert system has been used in solving various problems such as medical and treatment [7-11], education [12], path recognition [2][13], fault recognition [14-15], prediction of gas engines' efficiency [16], business management [17], control of robot's path [18], and control [19].

One of the main applications of expert system is the recommendation systems for matching a request with a database. The expert system for matching an unemployed with a given job [20] was presented in 2004 in which fuzzy induction system (FIS) and neural network are used to do the matching between the patterns [16]. Also, an expert system for efficient portfolio recommendation through fuzzy expert system has been presented in [21]. However, the subject of making an expert system for recommending a suitable journal for publishing a paper is less investigated in the area of recommendation. Akbarzadeh & Soleimani [22] suggested an expert system for this problem which despite the suitable time, is far from the optimized precision.

Choosing a suitable journal for publishing a certain paper is a matter which every researcher or student is faced with during the education time. There are some people who have become expert in this selection due to publishing a lot of papers. These people usually choose the suitable journal by considering some of the features of the paper and searching in journals' information bank. So far, no android system has been designed and presented for knowledge modeling and selection methods of these people. Designing such a system will help people avoid confusion, repeated sending, wasting time and money, frequent rejections in publishing their papers and by having a list of suitable journals they won't send the paper to an unsuitable journal. The presented system is being tested and examined by a database including journals in computer field recorded in Elsevier database [23]. The structure of the remaining part of the paper is as follows: section 2 examines the system's structure analytically. Section 3 shows a case study using the system (matching the paper and journal). Findings are presented in section 4.

2. SUGGESTED SYSTEM

Our system has the following characteristics:

- a) A database which includes a list of journals and their features.
- b) Clusters which are resulted from database clustering using Fuzzy-PSO methods. This clustering affects the speed increase and decreasing the evaluation steps of the suitable journal.
- c) A fuzzy induction system which selects the cluster most similar to the questioned paper. This system is also used in comparison of existing journals in the selected cluster with the given paper and presenting the degree of suitability of each journal with the paper.
- d) The user mediator is used by the user to insert the features of the given paper.

Our system consists of special parts Figure 1 which should be integrated in an android system to produce the final output. Every matching between the paper and the journal is done by the following features: Language, quality, waiting time for judgment, waiting time for publication after acceptance, field, length of paper and cost Figure 2. In general, the questionnaire is made of two precise and approximate parts. Journals' clustering is done based on the approximate criteria. The questionnaire's approximate part is given to the fuzzy induction system as the input. The results of the induction system are examined by the precise criteria and those which contrasted with binary criteria are deleted from the list.

The output of the induction system is the degree of matching the paper with journals and providing some suggestions. This problem is an approximate practical problem with successive inputs. We have used the experience of expert people to determine the criteria and the degree of the importance of each criterion.

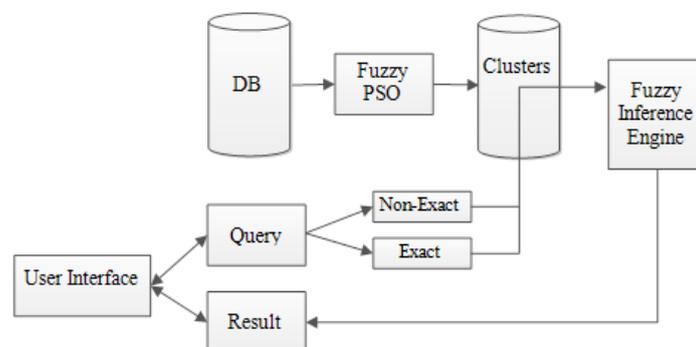


Figure 1. Simplified architecture System

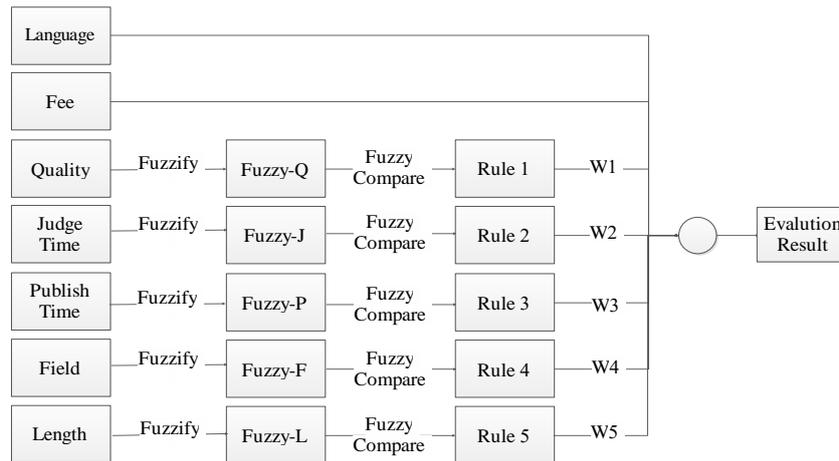


Figure 2. Fuzzy inference engine

2.1. System's Rules

There is a fuzzy rule for each criterion which should be performed for each form: The A feature of the given paper matches with A criterion. A can be one of the approximate criteria (Language, quality, waiting time for judgment, waiting time for publication after acceptance, field, and length of paper). There is a set of these rules for the index of each cluster and also for each journal. The sum of all these rules makes the system's rule Table 1.

Table 1. System's rules

Paper's quality matches quality criterion
Paper's judge time matches judge time criterion
Paper's publish time matches publish time criterion
Paper's field matches field criterion
Paper's length matches length criterion

2.2. Fuzzy Methods

In the following, the fuzzy methods in the field of questionnaire are presented. Fuzzy binary relations are fuzzy sets of A×B in which every element is drawn with membership degree between 0 and 1. We can explain a fuzzy relation using a relation matrix as follows:

$$U = \begin{bmatrix} \phi(1,1) & \phi(2,1) & \dots & \phi(n,n) \\ \phi(2,1) & \phi(2,2) & \dots & \phi(n,2) \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \phi(m,1) & \phi(m,2) & \dots & \phi(m,n) \end{bmatrix} \tag{1}$$

In which $\phi(m, n)$ is a function which shows the membership degree of m to n. if the membership degree of an element of the set equals 0, that element is completely out of the set, and if the membership degree of an element of the set equals 1, that element is completely in the set. Among the introduced features, approximate features (Language, quality, waiting time for judgment, waiting time for publication after acceptance, field, and length of paper) need fuzzy methods

2.2.1. Quality

The quality of each paper is determined by the user while filling in the form as one of the four groups of excellent, good, average, not so good. Mutually, this categorization should be done for every journal, as well. In our own investigated area, we have done the journal categorization based on the factor which ISI gives to each journal regarding its quality. This factor is called impact factor and has different ranges for journals of different fields of study. For our area of study-computer- figures more than 3 are very rare.

We have determined a fuzzy membership function for the journals' quality which increases the homogeneity between these groups Figure 3. In this case, those qualified journals which are a little far from the quality limitations are not rejected. Therefore, the rule of "paper quality matches journal quality" comes true and for example, an average paper can have choices among journals with factors between 0.4 to 1.2.

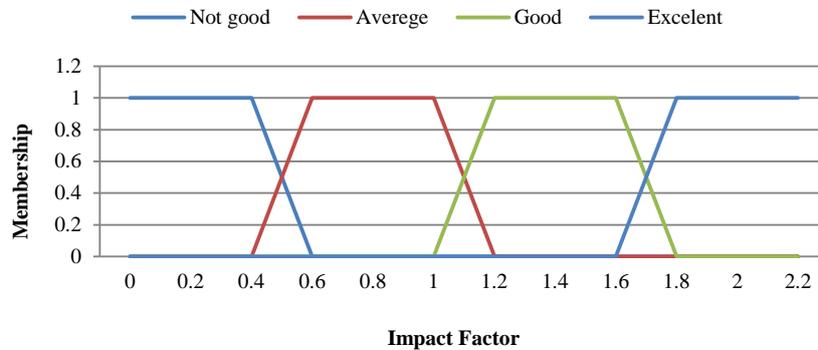


Figure 3. Fuzzy membership function for the journals' quality

2.2.2. Waiting Time for Judgment and Publication

The maximum time that a person can wait for the paper's judgment and publication is recorded in 4 areas: less than 1 month, between 1 to 2 months, between 2 to 4 month, and more than 4 months. These conditions are coded in a general way as field codes. Rules are: waiting time for the applicant's judgment matches with journal's criterion of waiting time for judgment. Waiting time for the applicant's publication matches with journal's criterion of waiting time for publication. These rules are understood as a relation between fields of waiting time and the request for waiting time Table 2. The results of these rules are the degree of matching the user's request for the paper's judgment and publication with the journal's performance.

Table 2. Fuzzy relations between field of waiting time (C) and request for waiting time (R)

C	R			
	Less than 1 m	1-2 m	2-4 m	More than 4 m
Less than 1 m	1	0.6	0.3	0
1-2 m	1	1	0.6	0.3
2-4 m	1	1	1	0.6
More than 4 m	1	1	1	1

2.2.3. Area and Field of Study

Here, the field of study is computer. This field is recorded in 4 main areas of software, hardware, artificial intelligence and network Table 3. Its related rule is "paper's field matches with field criterion" and the output of this rule shows the degree of matching the paper's field with the journal. Every journal can publish papers in several fields, so the output of maximum matching is met.

Table 3. Fuzzy relations between field of study (C) and request for field of study (R)

C	R			
	Software	Hardware	Artificial intelligence	Network
Software	1	0.5	0.8	0.5
Hardware	0.5	1	0.5	0.8
Artificial intelligence	0.8	0.5	1	0.5
Network	0.5	0.8	0.5	1

2.2.4. Paper's Length

The maximum length of published papers in a journal is grouped in 5 categories of less than 7 pages, between 7 to 9 pages, between 10 to 14 pages, between 15 to 20 pages, and more than 20 pages for journals without limitations Table 4.

Table 4. Fuzzy relations between field of paper's length (C) and request for paper's length (R)

C	R				
	Less than 7 p	7-9 p	10-14 p	15-20 p	More than 20 p
Less than 7 p	1	0.6	0.2	0	0
7-9 p	1	1	0.6	0.2	0
10-14 p	1	1	1	0.6	0.2
15-20 p	1	1	1	1	0.6
More than 20 p	1	1	1	1	1

2.2.5. Paper's Length

In this section, we examine receiving the fee for publishing the paper. This section can include 2 figures {yes=1, No=0}. It is an exact variable. All the journals which contrast with the user's request regarding the reception of fee are deleted from the final list.

2.2.6. Language

Every journal can publish papers written in one or several languages. All the journals which don't publish the given paper's language are deleted from the list. In our field, we examine 2 languages of English and Persian.

2.3. Clustering

Regarding the recommending the suitable journal, we are faced with a large number of journals and consequently high volume of information. One of the vital ways of controlling and managing this data is the categorization or grouping the similar data within a set of clusters. The aim of clustering is to organizing the data into reasonable groups to understand the differences and similarities between data. After clustering, it is possible to examine only the index of each cluster which can be the center of cluster, instead of examination and comparison of all the journals.

As mentioned before, in the present paper Fuzzy-PSO algorithm is used for clustering the journals. PSO algorithm was first introduced by Kennedy & Eberhart [24] based on the collective behavior of birds and fish. Every particle in an N-dimensional search space has a speed vector and a place vector. How to update the velocity vector and place vector in the repetition of t+1 is based on Eqs. 2 and 3.

$$v_i^{t+1} = wv_i^{t+1} + c_1r_1(pbest_i^t - x_i^t) + c_2r_2(gbest_i^t - x_i^t) \quad (2)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad i = 1, 2, \dots, n \quad (3)$$

c_1 and c_2 are constants called speed coefficient, and r_1 and r_2 are independent random numbers in the domain of [0, 1]. $pbest_i^t$ is the best place which has ever been seen by the particle i. $gbest^t$ is the best place among all the particles [25]. To select $gbest$ and $pbest$, adaptive distance dynamic clusters' criterion is used. This criterion determines the quality of each cluster based on the Eq. 4.

$$ADDC = \frac{\sum_{i=1}^k \sum_{j=1}^{N_i} D(c_i + x_{ij})}{K} \quad (4)$$

In which K is the number of clusters, N_i is the number of cluster's data, D is a distance function and d_{ij} is the j th data from i th cluster. Each data is related to the cluster which to its center it has the least distance.

In the original version of PSO, the center of preliminary clusters is randomly figured. In this paper, k-means++ algorithm is used for choosing the center of clusters. This choice leads to the increase in final precision of the method and decrease in the program's performance time [26]. This algorithm was first introduced by Arthur & Vassivitskii [27] for choosing the centroids. The steps for performing this algorithm to choose the centroids are as follows:

- The first centroid (c_1) is randomly chosen among X main data.
- The centroid c_i among X main data is chosen with the probability of Equation 5.
- Step b is repeated until the number of centroids reaches

$$P = \frac{SD(x)^2}{\sum_{x \in X} SD(x)^2} \quad (5)$$

$SD(x)$ is the shortest distance of x from the nearest existing centroid. To calculate the distance, all five fuzzy rules for each journal are performed for comparison with the cluster's center and then the result of each rule is chosen with a weight to determine the final result. These weights are the same chosen weights in the final evaluation. Eq. 3 is used for continuous data. Therefore, it is not suitable for data related to field of study. So, binary coding and binary version of PSO algorithm are used. Binary version of PSO algorithm determines the particle's place using Eqs. 6 and 7 [18].

$$s(v_i^{t+1}) = 1 / (1 + \exp(s(v_i^{t+1}))) \tag{6}$$

$$x_i^{t+1} = \begin{cases} 1 & \text{if } rand \leq s(v_i^{t+1}) \\ 0 & \text{if } rand > s(v_i^{t+1}) \end{cases} \tag{7}$$

2.4. Results and Evaluation

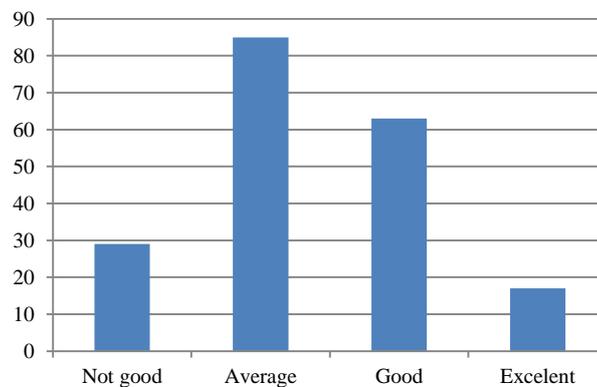
As mentioned in the introduction, for final selection of journals first the suggested paper is compared with the index of each cluster so that two clusters are chosen as winners. The number of clusters is achieved regarding the number of journals and through trial and error. For our given questionnaire, five clusters are considered. After choosing two superior clusters, the given paper is compared with all journals of those clusters. Table 5 shows the value of each rule. Those journals which are contradicted regarding the amount of exact variables (language and cost) are deleted from the comparison list. In the end, at most 10 journals with the most similarity are shown.

Table 5. Rules' value in final evaluation

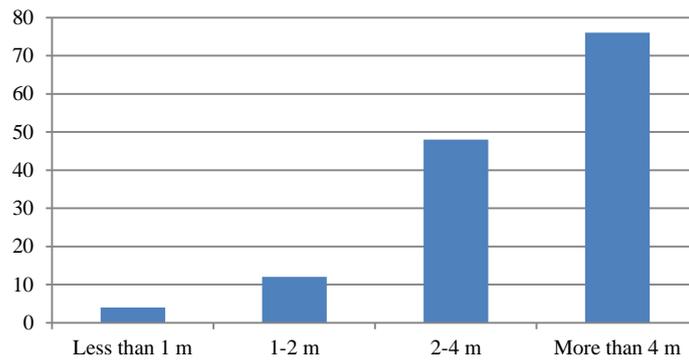
Rule	1	2	3	4	5
W_i	0.4	0.1	0.1	0.2	0.2

3. SAMPLE

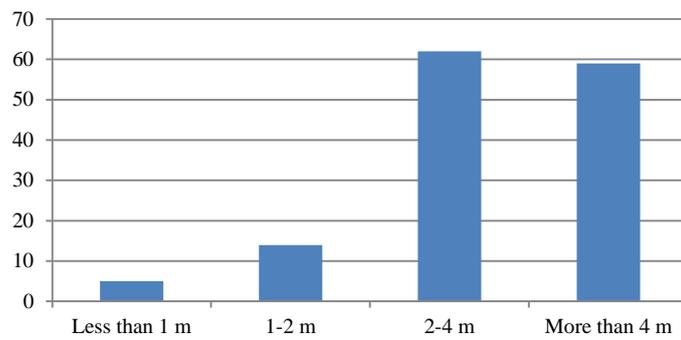
So far, 2826 journals are registered in Elsevier among which 140 titles are in the field of computer science. The statistical analysis of the database based on the evaluation criterion, i.e. data entered to the android system are presented in Graphs 1-5. This overview of database gives a better understanding about the papers' profiles and the findings of android system. The quality distribution is presented in Graph 1 which shows that most of journals are placed in average and good groups. Some journals can be placed in more than one group, so the sum of all the journals is more than 140 in the Figure. Graphs 2 and 3 illustrate that waiting time for judgment and publication of papers are generally more than two months. Graph 4 demonstrates that the number of journals in different areas is almost the same. Graph 5 showed that most of journals don't have any restrictions regarding the number of pages for the sent papers.



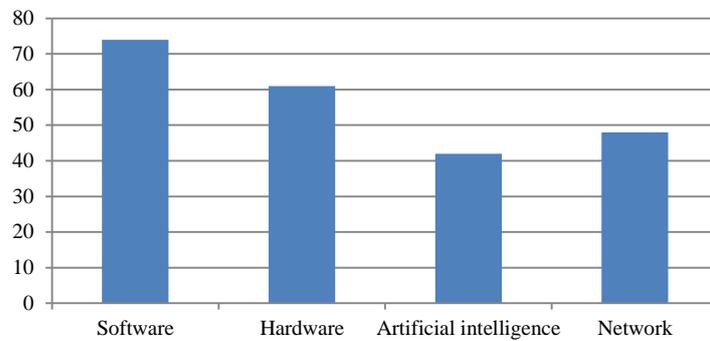
Graph 1. Quality



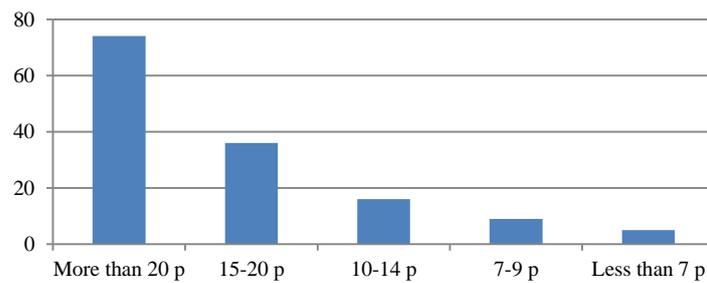
Graph 2. Waiting time for judgment



Graph 3. Waiting time for publication



Graph 4. Different areas



Graph 5. Number of pages for the sent papers

So far, students and researchers have chosen the suitable journal to publish their papers based on their personal experience or guidance of an expert, and no android system was designed to solve this problem. With unavailability of an expert, this process was time-consuming and sometimes some mistakes were happened such as publishing a rich paper in a weak journal, several sending of an average or weak paper to rich journals, etc.

Our designed system provides a list of suggested journals for a given paper. The system manager has inserted the profile of journal in the database Figure 4. Questionnaire's parameters are made of the input data to the android system. The check box next to some options allows their ineffectiveness. In the present use of the case study, just language and cost are closely examined. By clicking on the "register" button, the entered journal is registered in the database.

By pressing the "performing the questionnaire" button, the present questionnaire for the android system is run so that journals get matched with the given paper Figure 5. Then, the results of the android system by examining the questionnaire along with the journals' evaluation sign will be shown Figure 6.



Figure 4. Profile of journal



Figure 5. User interfaces in the query of the expert system

	Journal	Evaluation
1	Information and Computation	85%
2	Fuzzy Sets and Systems	76%
3	Journal of Algorithms	63%
4	Information Sciences - Applications	49%
5	Neural Networks	41%
6	Mathematics and Computation	23%
7		
8		
9		
10		

Figure 6. Query results

To test and evaluate the given system, 200 papers are randomly chosen and given to it. For every paper, system has recommended 5 to 10 journals. The recommendation list is given to 10 experts including computer teachers to evaluate so that each of them evaluates the system's recommendations for 20 papers. Each expert has given a 0 to 5 mark to every suggestion and all in all 1000 marks for 200 suggestions. For a more evaluation, the results of this evaluation have been compared with the results of a fuzzy expert system similar to what Arias-Aaranda et al. [17] and Akbarzadeh & Soleimani [22] have presented. Table 6 which has been achieved after 5 times testing process, shows that the precision of all the 3 methods is very similar, but the time needed for our suggested method is much less.

Table 6. The average precision and time for 5 times performing for 200 papers

	Precision (%)	time (seconds)
Expert system Arias-Aaranda et al. (2010) [17]	92.7	352
Expert system Akbarzadeh & Soleimani (2014) [22]	92.1	156
Suggested expert system	92.5	132

4. CONCLUSION

In this paper, a hybrid android system for recommending the suitable journal to publish a given paper has been developed. The suggested journals are chosen from among registered journals in the computer field in Elsevier's journals. The system's performance for journal recommendation with 5 approximate criteria (quality, waiting time for judgment, waiting time for publication after acceptance, field, length of paper) with the importance presented in Table 5 and 2 precise criteria (language and cost) is examined. The preliminary clustering of database leads to the decrease in the number of comparisons and increase in the speed of the response process to the user.

REFERENCES

- [1] J. Chia-Feng, H. Che-Meng, H. Chia-Hung, "Hierarchical cluster-based multispecies particle-swarm optimization for fuzzy-system optimization", *IEEE Transactions on Fuzzy Systems*, 18(2010)14.
- [2] D. Sedighzadeh, H. Mazaheripour, "Optimization of multi objective vehicle routing problem using a new hybrid algorithm based on particle swarm optimization and artificial bee colony algorithm considering Precedence constraints". *Alexandria Engineering Journal*, (2017), (Article in Press).
- [3] K. Sankar, K. Krishnamoorthy, "Ant based rule mining with parallel fuzzy cluster", *Advances in Information Mining*, 2(2010)13.
- [4] A. Khorsandia, S. Hosseinian, A. Ghazanfari, "Modified artificial bee colony algorithm based on fuzzy multi-objective technique for optimal power flow problem", *Electric Power Systems Research*, 95(2013) 206.
- [5] D.H. Tran, M.Y. Cheng, A.D. Pham, "Using Fuzzy Clustering Chaotic-based Differential Evolution to solve multiple resources leveling in the multiple projects scheduling problem", *Alexandria Engineering Journal*, 55(4) (2016) 1541.
- [6] E.E. Nithila, S.S. Kumar, "Segmentation of lung nodule in CT data using active contour model and Fuzzy C-mean clustering", *Alexandria Engineering Journal*, 55(3) (2016) 2583.
- [7] V. Robles-Bykbaev, M. López-Nores, J. Pazos-Arias, D. Arévalo-Lucero, "SPELTA: An expert system to generate therapy plans for speech and language disorders", *Expert Systems with Applications*, 42(21)(2015) 7641.
- [8] P. Silva, P. Gago, J. Carlos Bregieiro Rib, M. Santos, F. Portela, A. Abelha, J. Machado, F. Pinto, "An Expert System for Supporting Traditional Chinese Medicine Diagnosis and Treatment", *Procedia Technology*, 16(2014) 1487.
- [9] T. Padmaa, P. Balasubramanie, "Domain experts knowledge-based intelligent decision support system in occupational shoulder and neck pain therapy", *Applied Soft Computing*, 11(2)(2011) 1762.
- [10] H.A. Elkayal, N.E. Ismail, M. Lotfy, "Microwaves for breast cancer treatments", *Alexandria Engineering Journal*, 54(4) (2015) 1105.
- [11] R. Ramezani, M. Maadi, S. M. Khatami, "A novel hybrid intelligent system with missing value imputation for diabetes diagnosis", *Alexandria Engineering Journal*, (2017), (Article in Press).
- [12] R. Tucho, J. Sierra, J. Fernández, R. Vijande, G. Moris, "Expert tutoring system for teaching mechanical engineering", *Expert Systems with Applications*, 24, (4)(2003) 415.
- [13] Akashdeep, K. Kahlon, "An embedded fuzzy expert system for adaptive WFQ scheduling of IEEE 802.16 networks", *Expert Systems with Applications*, 41(16)(2014) 7621.
- [14] D. Ma, Y. Liang, X. Zhao, R. Guan, X. Shi, "Multi-BP expert system for fault diagnosis of power system", *Engineering Applications of Artificial Intelligence*, 26(3)(2013) 937.
- [15] N. R. Sakthivel, V. Sugumaran, B. B. Nair, "Comparison of decision tree fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump", *Mechanical Systems and Signal Processing*, (2010).

- [16] S. Tasdemir , I. Saritas, M. Cinivi, “Artificial neural network and fuzzy expert system comparison for prediction”, *Expert Systems with Applications*, 38(2011) 13912.
- [17] D. Arias-Aranda, J. Castro , M. Navarro , J. Sánchez , J. Zurita, “A fuzzy expert system for business management”, *Expert Systems with Applications*, 37(2010) 7570.
- [18] Z. Bingül, O. Karahan, “A Fuzzy Logic Controller tuned with PSO for 2 DOF robot trajectory control”, *Expert Systems with Applications*, 38(2011) 1017.
- [19] F. Olusegun , A. Olusegun, “A fuzzy expert system to Trust-Based Access Control in crowdsourcing environments”, *Applied Computing and Informatics*, 11 (2)(2015) 116.
- [20] A. Drigas, S. Kouremenos, S. Vrettos , J. Vrettaro, “An expert system for job matching of the unemployed”, *Expert Systems with Applications*, 26(2004) 217.
- [21] M. Fasanghari , G. Montazer, “Design and implementation of fuzzy expert system for Tehran Stock Exchange portfolio recommendation”, *Expert Systems with Applications*, vol. 37(9)(2010) 6138.
- [22] F. Akbarzadeh, A. Soleimani, “An expert system for journal recommendation”, in persia, in *12th Iranian Conference on Intelligent Systems*, Iran, Bam, 2014.
- [23] "elsevier," [Online]. Available: <http://www.elsevier.com/journals/subjects/computer-science>. [Accessed 09 02 2013].
- [24] J. Kennedy , R. Eberhart, “Particle Swarm Optimization”, in *Neural Networks*, Perth, WA, 1995.
- [25] X. Yang, J. Yuan, J. Yuan , M. Mao, “A modified particle swarm optimizer with dynamic adaptation”, *Mathematics and Computation*, (2007) 1205.
- [26] M. Maruf Öztürk, U. Cavusoglu, A. Zen, “A novel defect prediction method for web pages using k-means++”, *Expert Systems with Applications*, 42, (19)(2015) 6496.
- [27] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," *In Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms*, p. 1027–1035, 2007.