International Journal of Enterprise Information Systems
Volume 12 • Issue 3 • July-September 2016

Hybrid Fuzzy Neural Search Retrieval System

Rawan Ghnemat, Computer Science Department, King Hussein Faculty for Computing Sciences, Princess Sumaya University for Technology, Amman, Jordan

Adnan Shaout, The Electrical and Computer Engineering Department, The University of Michigan – Dearborn, Dearborn, MI, USA

ABSTRACT

Search engines are crucial for information gathering systems (IGS). New challenges face search engines concerning automatic learning from user requests. In this paper, a new hybrid intelligent system is proposed to enhance the search process. Based on a Multilayer Fuzzy Inference System (MFIS), the first step is to implement a scalable system to relay logical rules in order to produce three classifications for search behavior, user profiles, and query characteristics from analysis of navigation log files. These three outputs from the MFIS are used as inputs for the second step, an Adaptive Neuro-Fuzzy Inference System (ANFIS). The training process of the ANFIS replaced the rules by adjusting the weights in order to find the most relevant result for the search query. This proposed system, called MFIS-ANFIS, is implemented as an experimental system. The system performance is evaluated using quantitative and comparative analysis. MFIS-ANFIS aimed to be the core of intelligent and reliable search process.

KEYWORDS

Back-Propagation Algorithm, Hybrid Intelligent System, Information Retrieval, Multilayer Fuzzy Inference Systems, Neuro-Fuzzy System, Search Engines

1. INTRODUCTION

The increasing growth of available Internet data, produced by a great number of heterogeneous providers, has great potential in various domains like IGS (Tao, Li, & Zhong, 2011). Centralized traditional search process poses a serious problem in finding the relevant information for a given query. We still need more time examining the retrieved Web sites than the time needed to retrieve the list of the Web sites (Das & Kalita, 2016). Meanwhile recent research in intelligent systems development to improve the search process uses new methods, such as Web Content Mining (Bock & Hettenhausen, 2012), Web Structure Mining (Putra & Akbar, 2013) and Web Usage Mining (Sisodia & Verma, 2012). Search engines should progress from being Web document retrieval tools to become intelligent systems that fully support user's behavior during their interactions with the Web (Klusch, Kapahnke, Schulte, Lecue, & Bernstein, 2015). This creates novel and exciting research challenges ranging from the ability to recognize tasks from the issued queries (Lucchese, Orlando, Perego, Silvestri, & Tolomei, 2013), to the design of new recommendation strategies and user-customized search for showing relevant results (Dao, Hoang, Ta, & Tho, 2013).

While the navigation search behavior can help the Web master to restructure their Web sites, we can also use the history of query text produced by users and also the user profile to predict the most

DOI: 10.4018/IJEIS.2016070105

relevant information (Yu, Ma, Hsu, & Han, 2014). Having an adaptive Web search process based on navigation search history, user profile and query text is our motivation for this work.

In this paper, we face the problem of inefficient search engine retrieval by proposing a hybrid intelligent system aimed to enhance the search result based on the user needs. This will save user's time and increase their satisfaction. By using the Web browsing behavior, user profile, and needed query, we implement an intelligent system that can predict the most relevant Web sites regardless of the rank of these Web sites. The proposed system improves the performance of the search engines by providing intelligent recommendations not only matching based on the ranking of the Websites.

The proposed system architecture is divided into two stages. The first stage is composed of three parallel fuzzy subsystems that use navigation search data, profile data, and query data independently to generate three outputs based on fuzzy rule-based inference engines. In the second stage these three outputs are used as inputs for an Adaptive Neuro-Fuzzy Inference system to generate the most relevant search results using back-propagation learning algorithm.

The rest of this paper is structured as follows. Section 2 reviews the recent research in Web mining and intelligent Web search engines. Section 3 describes the proposed system and the rationale behind our design. Section 4 explains the system implantation procedure; some experiments are explained and evaluated. The paper concludes in section 5.

2. RELATED WORK

In Web mining (Jiang, Pei, & Li, 2013) the miner tries to benefit from the data created through the sessions of surfing the Web or behavior of this surfing. Although Web content and structure mining use real and primary data on the Web. Web usage mining extracts also from the secondary data originated as a result of the user interactions during Web sessions. Information in Web usage covers data from server access logs, proxy server logs, browser logs, to user profiles, registration data, user sessions or transactions, cookies, user queries, bookmark data, mouse clicks and scrolls, and any other data as the results of interactions (Umagandhi & Kumar, 2013).

Intelligent Web mining supported by semantic based search leads to a useful pattern for better search process (Bollegala, Matsuo, & Ishizuka, 2011). This area of research addresses the optimization of the structure and the connection of the Web sites (Yang, Sun, Tang, Ma, & Li, 2015).

Soft computing is one of the emerging approaches that has been used in intelligent search engines area because of its parallels reasoning and ability to learn in an environment of uncertainty and imprecision (Szczepaniak, Segovia, & Zadeh, 2012). Neural networks and fuzzy logic are two powerful techniques in soft computing; the hybridization of both techniques would inherit all the advantages of both techniques. Problems that have dynamic nature and uncertain input are a good application for such techniques.

In particular, soft computing techniques have been used to improve Web page ranking and structuring using genetic algorithms (Yan, Gui, Du, & Guo, 2011), artificial neural networks (Mohammad, Thabtah, & McCluskey, 2014), knowledge based concepts, and semantic retrieval systems (Lin et al., 2004). Fuzzy inference systems have been used in different domain of application, like estimating software reliability (Tyagi & Sharma, 2014) using adaptive neuro-fuzzy system, medical data classification (Dennis & Muthukrishnan, 2014) using adaptive genetic fuzzy system, and for electrical energy demand prediction using clustering based genetic fuzzy expert system (Ghanbari, Ghaderi, & Azadeh, 2010).

In Arotaritei and Mitra's study (2004), a survey of using fuzzy inference systems for clustering, association rule mining and information retrieval is presented. Nikravesh, Loia, and Azvine (2002) address the research of fuzzy and intelligent search engines using user-defined queries to retrieve useful information according to certain measures based on predefined linguistic formulations and rules defined by experts or based on a set of known homepages. The authors concluded by proposing fuzzy conceptual model and search engine. In Khan, Khodke, and Bhagat's work (2015) the fuzzy system used neural network for query information filtering and then categorizing them into categories that enhance the search results. In Lee, Kim, Chung, and Kwon's work (2002), fuzzy cognitive maps have been used for association rules of navigation categories in order to optimize the Web surfing path.

Web personalization depending on user profile extracted using fuzzy system was presented in Ansari, Sattar, Babu, and Azeem's study (2015). A fusion between conceptual graph and fuzzy formalism to handle natural fuzzy conceptual graphs (FCG) is used by Nikravesh (2002) for semantic Web and information retrieval with more intelligent and human natural way.

Fuzzy inference engines and machine learning can be combined in search process to have intelligent retrieval for the search quires. Table 1 summarizes the approaches used in intelligent search retrieval system.

In this work we focus on implementing a new intelligent system based on three categories of inputs: Web usage attributes, query understanding, and user profile information. Phase one has been implemented using three parallel fuzzy inference engines. The parallel architecture allows the system to be scalable with respect to the number of attributes and so to be efficient.

In the second phase, an adaptive neuro-fuzzy inference system integrates search usage information with query information and user profile information for effective search result based on back-propagation learning algorithm. The whole system is a hybrid architecture that combines both the power of fuzzy expert system rules and the adaptive learning of neural network to have the most relevant and efficient search result.

Table 1. The intelligent approaches for search retrieval systems

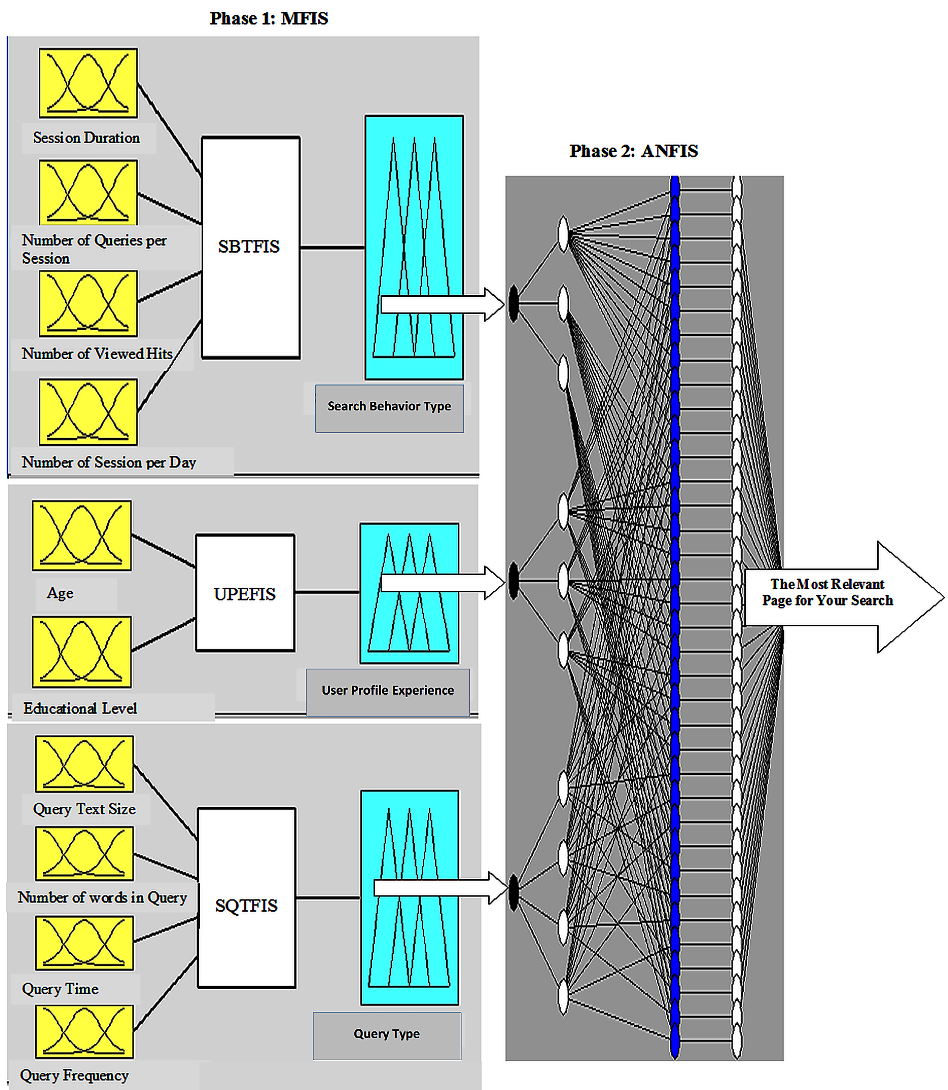
Approach	Intelligent Tool Used	Performance	Limitations
Query processing and categorizations (Vidhyapriya & Sampath, 2015).	Fuzzy C means algorithm for inferring user search goals with feedback sessions.	The complexity of the approach is low and can be used in reality easily. For each query, the running time depends on the number of feedback sessions.	The work mainly focused in comparative performance using different combinations of variables than in the comparative performance of different techniques.
Ranking algorithm enhancement (Jain, Sharma, Dixit, & Tomar, 2013).	New intelligent search method using indexing of the Web pages based on Web topology.	The algorithm is not efficient in real time. The algorithm was used in a prototype search engine called Clever for an IBM research project.	Could not be implemented in a real time search engines.
Context-based knowledge search (Smirnov, Levashova, & Shilov, 2015).	Ontology-based knowledge source decision support systems.	In real-life, the implementation of all the phases can be impossible or un-needed.	The user is not an owner of the application ontology; he/she does not have an authority to modify it.
Indexing enhancement and semantic Web (Leung, Chan, Milani, Liu, & Li, 2012).	Adaptive evolutionary computation used to capture human judgment for meaningful indexing of new media objects and new terms.	The proposed algorithm covered most of the object while maintaining a good performance in terms of total relevance to the query answers.	The lack of flexibility is a major drawback in domains where the user relevance evaluation dynamically evolves over time.
The search matching mechanism (Huang, He, Gao, Deng, Acero, & Heck, 2013).	Learning deep structured semantic models for Web search using click through data.	The method has the best performance, beating other methods by a statistically significant margin.	Click information is unavailable for many URLs, especially new URLs and tail URLs.

3. ARCHITECTURE

The proposed hybrid intelligent system should be fast to save time and accurate to retrieve the most relevant information. The system is implemented as follows:

- **Phase 1:** Consists of three parallel fuzzy expert inference systems: Search Behavior Type Fuzzy Inference System (SBTFIS), User Profile Experience Fuzzy Inference System (UPEFIS), and Search Query Type Fuzzy Inference System (SQTFIS).
- **Phase 2:** Consists of an Adaptive Neuro-Fuzzy Inference System. Use the outputs of the fuzzy sub-systems search behavior type, user profile experience, and search query type to find the most relevant search results to the user query. The architecture of the proposed system is shown in Figure 1.

Figure 1. Design of the proposed hybrid system



3.1. Parallel Fuzzy Expert Inference Systems

The most commonly used fuzzy inference technique is the Mamdani's method (Lee, 1990). Our parallel fuzzy system uses the Madmani-style fuzzy inference process and performs its function in four steps as follows:

3.1.1. Fuzzification of the Input Variables

There are three main fuzzy sub-systems. All input and output variables, universe of discourses, and linguistic values are chosen based on the system needs and are described in Table 2.

Fuzzy sets for the linguistic values for the fuzzy variables can have a variety of shapes. However, a triangle or a trapezoid can often provide an adequate representation of the expert knowledge and, at the same time, significantly simplifies the process of computation. We chose trapezoidal membership function for all variables, as shown in Figure 2.

3.1.2. Rule Evaluation

To accomplish this task, we used experts to define rules for each sub-system based on the available data, like the transaction logs, user logs of information, and query information.

Sample of each fuzzy rule base for each sub-system is shown in Figure 3.

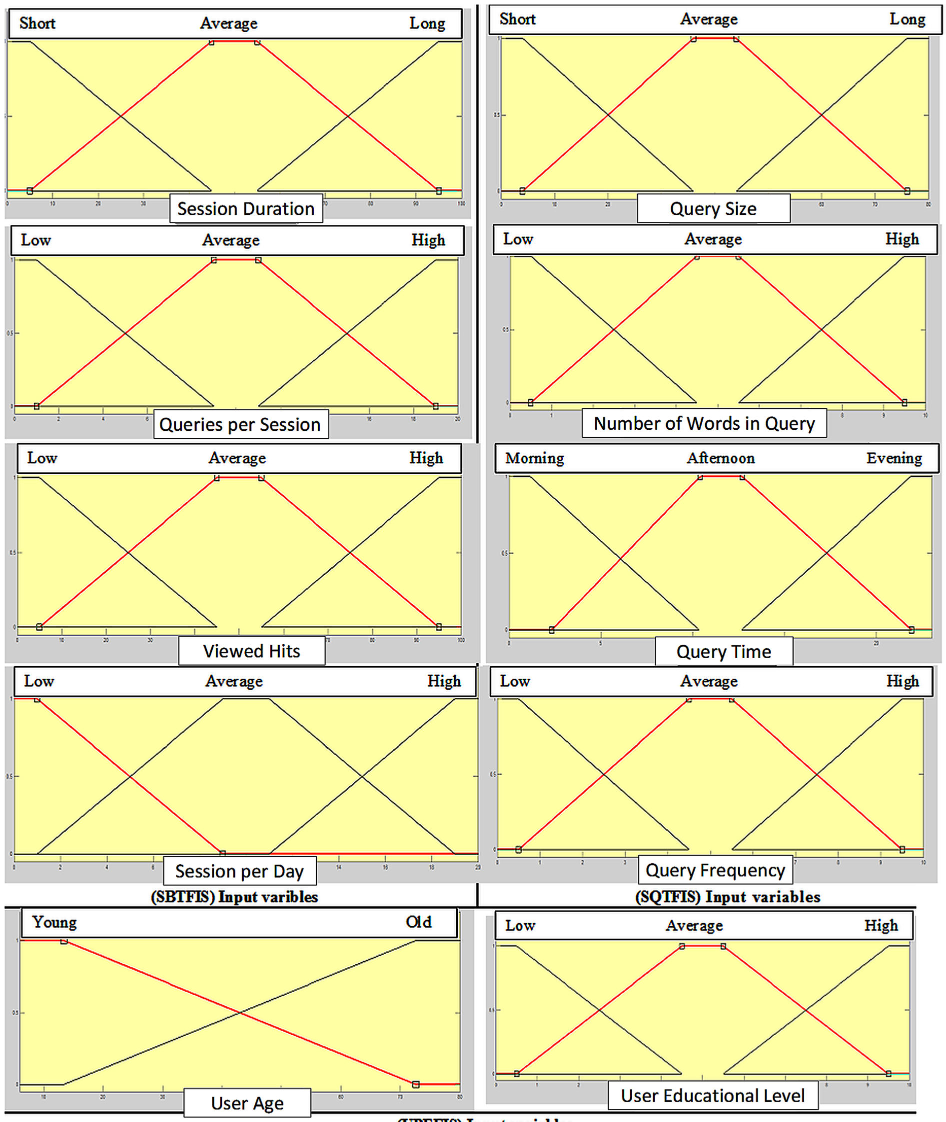
3.1.3. Aggregation of the Rules

Aggregation is the process of summation of the output of all rules. We take the membership functions of all rules consequents previously clipped and combine them into a fuzzy output. Figure 4 shows the membership function for all the linguistic values for the fuzzy output variables. As a result of this step we have values for each of the fuzzy sets in the outputs variables.

Table 2. Input and output linguistic variable and their ranges

Linguistic Variable	Linguistic Value	Numerical Range
Search Behavior Type Fuzzy Inference System		
Input 1: Session Duration	[Short, Average, Long]	[0-100]
Input 2: Number of query per session	[Low, Average, High]	[0-20]
Input 3: Number of viewed hits	[Low, Average, High]	[0-100]
Input 4: Number of session per day	[Low, Average, High]	[0-20]
Output: Search Behavior Type	[Informational, Transactional, Navigational]	[0-100]
User Profile Experience Fuzzy Inference System		
Input 5: User age	[Young, Old]	[6-80]
Input 6: User Educational level	[Low, Average, High]	[0-10]
Output: User Profile experience	[Low, Average, High]	[0-100]
Search Query Type Fuzzy Inference system		
Input 7: Query size	[Short, Average, Long]	[0-80]
Input 8: Number of words in query	[Low, Average, High]	[0-10]
Input 9: Query time	[Morning, Afternoon, Evening]	[0-23]
Input 10: Query frequency	[Low, Average, High]	[0-10]
Output: Query Type	[Science, News, Product, Online Community]	[0-100]

Figure 2. The membership functions for the linguistic values of the fuzzy inputs variables



3.1.4. Defuzzification

The last step in the fuzzy inference process is defuzzification. Fuzziness helps us to evaluate the rules, but the final output of a fuzzy system has to be a crisp number. The input for the defuzzification process is the aggregate fuzzy output and the output of the defuzzification is a crisp number from [0-100].

3.2. Adaptive Neuro-Fuzzy Inference System Structure and Parameters Adjustments

The Adaptive Neuro-Fuzzy Inference System (ANFIS) is a multilayer neural network structure that functions as Sugeno-type fuzzy reasoning system. Based on backpropagation learning algorithm the system accomplishes a number of adaptive IF-THEN rules. ANFIS merge artificial neural network's learning ability and Fuzzy-Logic decision making capability together. The learning rules of ANFIS

Figure 3. Sample of the fuzzy rule base

```
If Session duration is long and Number of queries is low and Number of viewed hits
is high and number of session per day is low then Search behavior type is Informational.

If Session duration is short and Number of queries is low and Number of viewed hits
is low and number of session per day is low then Search behavior type is Transactional.

If Session duration is short and Number of queries is high and Number of viewed hits
is average and number of session per day is high then Search behavior type is Navigational.
```

(SBTFIS) Rules

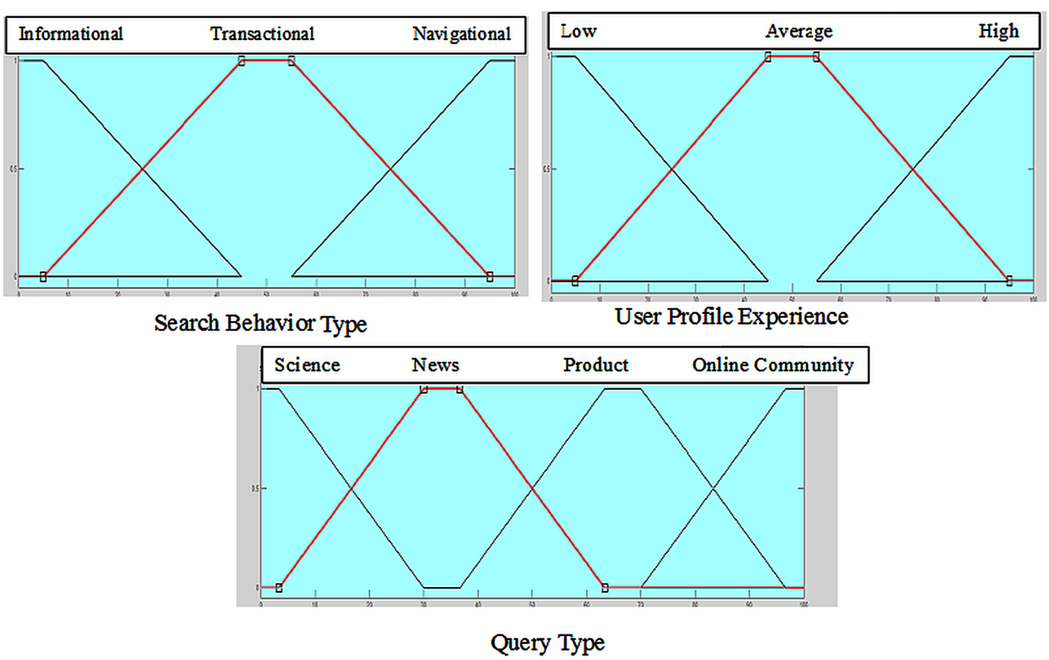
```
If Age is young and Educational level is low then User profile experience is Low.

If Age is young and Educational level is average then User profile experience is Average.

If Age is old and Educational level is high then User profile experience is High.
```

(UPEFIS) Rules

```
If Query text size is long and number of words is low and query time is morning
and Query frequency is low then Query type is Science.

If Query text size is short and number of words is average and query time is afternoon
and Query frequency is low then Query type is News.

If Query text size is short and number of words is high and query time is evening
and Query frequency is average then Query type is product.

If Query text size is long and number of words is high and query time is evening
and Query frequency is low then Query type is Online Community.
```

(SQTFIS) Rules

Figure 4. The membership functions for the linguistic values of the fuzzy outputs variables



have been described in detail in Jang's work (Jang, 1993). From the typical form for an IF-THEN rule this phase rule expressed follows:

```
IF search behavior type is InformationalAND user profile
Experience isAverage
AND query type is ProductTHENrelevant page rank= f (search
behavior type, user profile Experience, query type).
```

When x_1, x_2, x_m are input variables, A_1, $A_{2,}$ …, Am are fuzzy values for the input variables, y is output variable that can be expressed as a first-order polynomial as shown in Equation (1).

$$y = k_0 + k_1 x_1 + k_2 x_2 + \ldots + k_m x_m \tag{1}$$

Figure 5 shows the ANFIS architecture that corresponds to the first order Sugeno fuzzy model. The layers are described as follows:

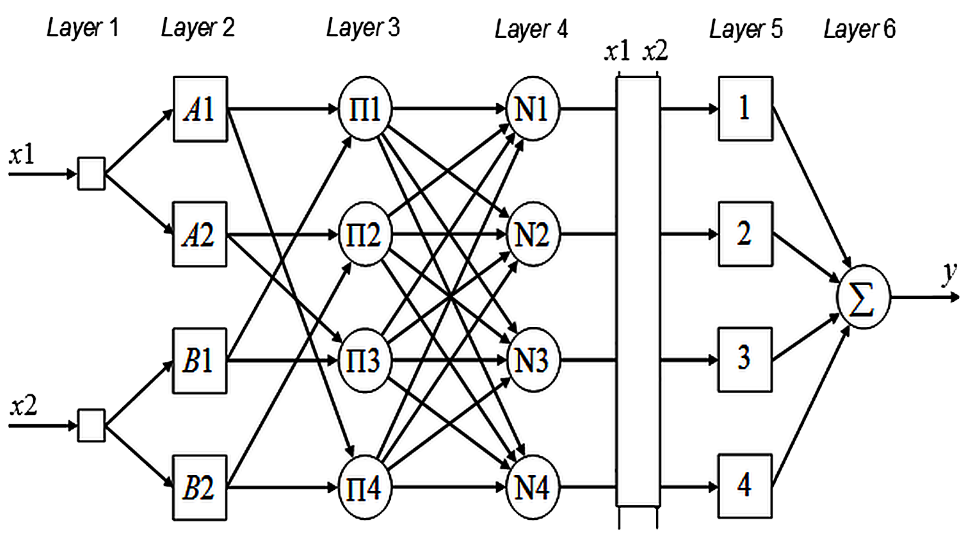
Layer 1: The input layer - Each neuron in this layer represents input variable as shown in Equation (2).

$$y_i^{(1)} = x_i^{(1)} \tag{2}$$

where $x_i^{(1)}$) is the input and $y_i^{(1)}$ is the output of input neuron i in layer 1. The range for inputs variables that have been used is: [0-100].

Layer 2: The fuzzification layer - Jang's model is used. A bell activation function specified in Equation (3) to obtain the fuzzy value for each input is used.

Figure 5. Adaptive Neuro-Fuzzy Inference System (ANFIS) Architecture



$$y_{\mathrm{i}}^{(2)} = \frac{1}{1 + \left(\frac{x_i^{(2)} - a_i}{c_i} \right)^{2b_i}} \tag{3}$$

Where $x_i^{(2)}$ is the input and $y_i^{(2)}$ is the output of neuron i in Layer 2; and a_i , b_i and c_i are parameters that control the center, width, and the slope of the bell activation function of neuron i, respectively. This step reverse the defuzzification work that has been done in the first phase based on the training data and not on the rule based. This will make the system adaptive according to the training data. For search behavior type this layer has [Informational, Transactional, Navigational] nodes, for user profile Experience it has [Low, Average, High], and the four final nodes are representing the query type [Science, News, Product].

Layer 3: The rules layer - Each neuron in this layer represents a fuzzy rule. The output of this layer is the firing strength of each rule as shown in Equation (4).

$$y_i^{(3)} = \prod_{j=1}^{k} x_{ji}^{(3)} \tag{4}$$

Where $x_{ji}^{(3)}$ represents the value of membership function for the jth input and the ith rule and $y_i^{(3)}$ is the output of neuron i in layer 3. Rule numbers can be minimized to enhance the performance of the system. Four comprehensive rules have been represented by four nodes in the implemented system.

Layer 4: The normalization layer - Each neuron in this layer receives inputs from all neurons in the rules layer and calculates the normalized firing strength of a given rule. The normalized firing strength is the ratio of the firing strength of a given rule to the sum of firing strengths of all rules. It represents the weight of a given rule to the final result. The output of neuron *i* in this layer is determined as shown in Equation (5).

$$y_i^{(4)} = \frac{x_{ji}^{(4)}}{\sum_{j=1}^{n} x_{ji}^{(4)}}_{i} \tag{5}$$

Where $x_{ji}^{(4)}$ is the input from neuron j located in Layer 3 to neuron i located in Layer 4 and n is the total number of rule neurons.

Layer 5: The defuzzification layer - Each neuron in this layer is connected to the respective normalisation neuron and also receives initial inputs, $x1$ and $x2$. A defuzzification neuron calculates the weighted consequent value of a given rule as shown in Equation (6).

$$y_i^{(5)} = x_i^{(5)} \left[k_{i0} + k_{i1}\, x1 + k_{i2}\, x2 \right] \tag{6}$$

Where $x_i^{(5)}$ is the input and $y_i^{(5)}$ is the output of defuzzification neuron *i* in *Layer* 5 and k_{i0}, k_{i1} and k_{i2} is a set of consequent parameters of rule *i*.

Layer 6 is represented by a single summation neuron. This neuron calculates the sum of outputs of all defuzzification neurons and produces the overall ANFIS y output. *y* is obtained by Equation (7).

$$y = \sum_{i=1}^{n} x_i^{(6)} \tag{7}$$

Where y is the actual output of the ANFIS which is the rank number of the most relevant page of a given search.

The ANFIS architecture has two adaptive layers: the fuzzification layer and the defuzzification layer. The task of the learning algorithm is adjusting the modifiable parameters (the membership function value for each input, and the weighted consequent value of a given rule) based on error signal.

ANFIS has been used in this stage because mapping tool for the data is needed and knowledge based cannot be obtained from an expert.

4. EXPERIMENTS AND SYSTEM EVALUATION

In this section, we describe the last and the most laborious task, which is to evaluate and tune the system. Experimental results are described in this section. Quantitative and comparative analysis has been performed to verify the performance of the proposed system.

With a system having i5 processor, main memory of 4GM RAM, and Windows 7 as operation system, the experiments has been achieved. MATLAB was used to implement the proposed system.

One hundred PSUT (Princess Sumaya University for Technology) students and employees were asked to give their search information during a four month period. The used search data set contains 7000 search records with 3264 unique queries. Figure 6 show a sample of this data. 70% of the data was used for training and 30% was used for testing. Evaluation metric was used in the evaluation process. The definitions for the evaluation metrics are given in Equation 8 through Equation 11.

$$\text{Correct rate} = \frac{\left(\text{correctly obtained sample}\right)}{\left(\text{obtained sample}\right)} \tag{8}$$

Figure 6. Sample of the used data

User ID	Age	Eduaction Level	Query text	Qeury date and time	Session Duration(M)	number cliked throw sites	selected site Ranke
1	20	Bachelor	Facebook	15/oct-14:22	4	1	1
2	19	Bachelor	psut regestration	15/oct-11:23	10	3	2
3	22	Bachelor	artificial neural network	15/oct-1:24	30	6	7
4	29	Ph.D	chat bot	15/oct-13:25	50	5	6
5	20	Bachelor	Chess Online	15/oct-20:26	55	3	4
6	19	Bachelor	android	15/oct-22:27	40	2	3
7	20	Bachelor	Cars	15/oct-16:28	15	4	4
8	22	Bachelor	Music	15/oct-14:29	10	2	2
9	33	Bachelor	pray time	16/oct-04:30	5	1	1
10	34	Ph.D	Youtube	17/oct-13:31	10	3	3
11	44	Ph.D	weather	17/oct-00:32	15	3	4
12	38	Ph.D	birthday cake	17/oct-01:33	20	4	5
13	34	Master	Hair Style	20/oct-22:34	40	2	3
14	48	Ph.D	best smart phones	20/oct-14:35	25	6	7
15	55	Ph.D	ministry of eduaction	20/oct-14:36	25	1	2

$$\text{Error rate} = \frac{\left(\text{incorrectly obtained sample}\right)}{\left(\text{obtained sample}\right)} \tag{9}$$

$$\text{Obtained rate} = \frac{\left(\text{obtained sample}\right)}{\left(\text{total number sample}\right)} \tag{10}$$

$$\text{Inconclusive rate} = \frac{\left(\text{non-obtained sample}\right)}{\left(\text{total number sample}\right)} \tag{11}$$

The quantitative values for the correct rate of the system for the used testing sample is 92.15%, which was obtained after 30 iteration. The other evaluation factors for the testing samples are shown in Table 3.

This result is quite satisfactory for such problem, while the error in the training data is not far from the error in the testing data. Figure 7 shows the RMSE (Root Mean Square Error) that was obtained after 30 epochs with the training data.

In order to prove that the implemented two-phase system is the most efficient tool to optimize the result, we used the same data set with two other existing intelligent systems and we compared the results. The two systems are:

1. Single Fuzzy Inference System (SFIS): all the ten inputs listed in Table 2 used in one Mamdani's inference system with the same membership functions used in the proposed fuzzy multi-layer system. The architecture of this system is shown in Figure 8.
2. Multilayer neural network (MNN): ten inputs as listed in Table 2 with four hidden layers with five neurons in each. The architecture of the feedforward neural network is shown in Figure 9. Backpropagation learning algorithm was used in the training process as was used in the ANFIS model (Jang, 1993).

The comparative analysis of the three systems with the same dataset is specified in Table 4. Table 4 shows that the new proposed system is the best in term of correctness.

Another comparison criterion is tested, which is the performance; comparing the time needed to obtain the result in both training data and the testing data for the three systems. The result in Table 5 shows that the proposed system gave intermediate result comparing with the two other systems, while SFIS is the fastest and MNN is the slowest. MFIS-ANFIS uses the backpropagation partially

Table 3. The Quantitative evaluation for the implemented system

Evaluation Factor	Obtained value
Correct rate	0.9215
Error rate	0.0785
Obtained rate	1
Inconclusive rate	0

Figure 7. Training data error

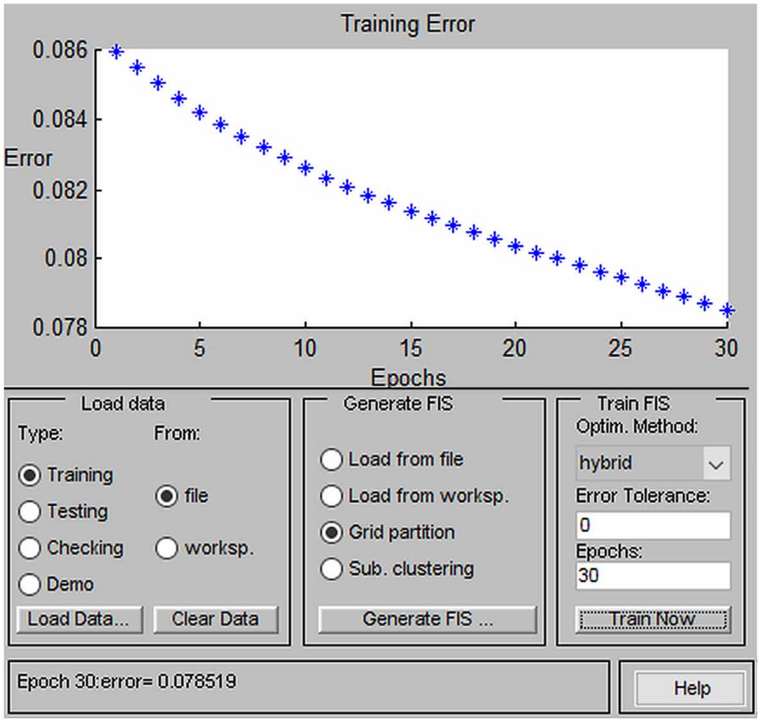
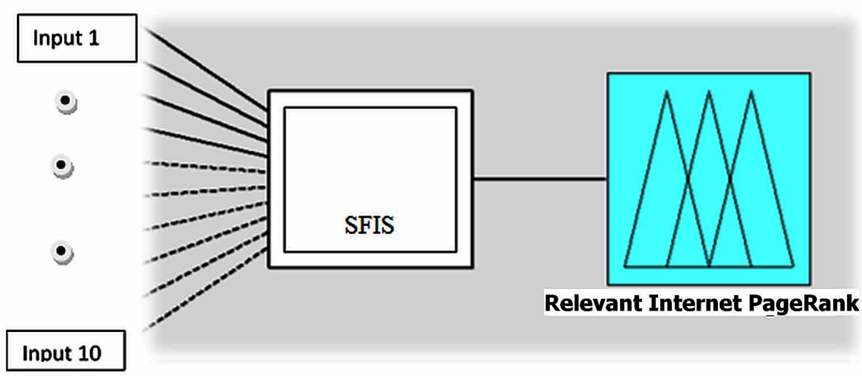


Figure 8. Single Fuzzy Inference System



in the second part, but with less number of inputs than the MNN. On the other hand, the correctness of the fastest system, which is SFIS, is the lowest, which makes the MFIS-ANFIS the best in term of correctness and time combined.

5. CONCLUSION

The combination of the parallel fuzzy systems with the adaptive neuro-fuzzy Inference system using multiple layers architecture produced a new automatic intelligent optimization system. This hybrid system can be used in any kind of problem that has both expert knowledge and experience data to

Figure 9. Multilayer neural network (MNN) system

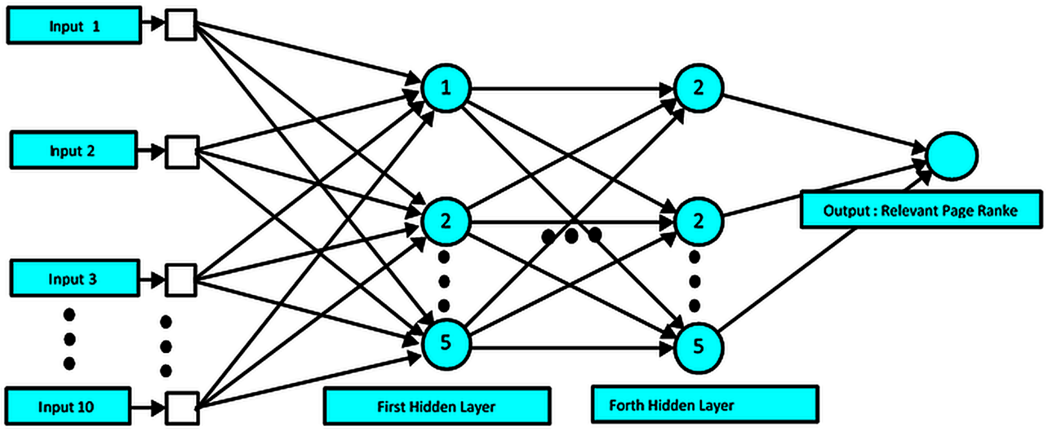


Table 4. The quantitative evaluation comparison with the two other systems

Evaluation System Factor	SFIS	MNN	MFIS-ANFIS
Correct rate	0.2651	0.4347	**0.9215**
Error rate	0.7349	0.5653	**0.0785**
Obtained rate	1	1	1
Inconclusive rate	0	0	0

Table 5. The performance comparison with the two other systems

System	Training time (second)	Retrieval time (second)
SFIS	0	$0.2\text{x } 10^{-9}$
MNN	534	$1.42\text{x } 10^{-3}$
MFIS-ANFIS	238	$0.72\text{x } 10^{-6}$

solve. In phase one, the rules reasoning save a lot of time and was capable of having huge number of inputs. In the second phase, the ANFIS has the ability to converge rapidly and this is important in on-line learning especially in adaptive control.

The dynamic nature of the search problem needs adaptive control strategy which has been applied in the second phase of the proposed hybrid system model. In the future, the model could be generalized to be used for other optimization problems. Quantitative and comparative analysis for the proposed system was evaluated. From the results, the new proposed system, MFIS-ANFIS, has generated 92.51% accuracy with an average retrieval time equal to $0.72\text{x } 10^{-6}$ second.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and the editor for their constructive comments and suggestions.

REFERENCES

Ansari, Z., Sattar, S. A., Babu, A. V., & Azeem, M. F. (2015). Mountain density-based fuzzy approach for discovering Web usage clusters from Web log data. *Fuzzy Sets and Systems*, *279*, 40–63. doi:10.1016/j.fss.2015.01.021

Arotaritei, D., & Mitra, S. (2004). Web mining: A survey in the fuzzy framework. *Fuzzy Sets and Systems*, *148*(1), 5–19. doi:10.1016/j.fss.2004.03.003

Bock, J., & Hettenhausen, J. (2012). Discrete particle swarm optimization for ontology alignment. *Information Sciences*, *1*(6), 152–173. doi:10.1016/j.ins.2010.08.013

Bollegala, D., Matsuo, Y., & Ishizuka, M. (2011). A Web search engine-based approach to measure semantic similarity between words. *IEEE Transactions on Knowledge and Data Engineering*, *23*(7), 977–990. doi:10.1109/TKDE.2010.172

Dao, T. T., Hoang, T. N., Ta, X. H., & Tho, M. C. (2013). Knowledge-based personalized search engine for the Web-based Human Musculoskeletal System Resources (HMSR) in biomechanics. *Journal of Biomedical Informatics*, *46*(1), 160–173. doi:10.1016/j.jbi.2012.11.001 PMID:23149160

Das, S., & Kalita, H. K. (2016). Semantic Model for Web-Based Big Data Using Ontology and Fuzzy Rule Mining.*Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems* (Vol. 2, pp. 431-438). Springer International Publishing. doi:10.1007/978-3-319-30927-9_42

Dennis, B., & Muthukrishnan, S. (2014). AGFS: Adaptive Genetic Fuzzy System for medical data classification. *Applied Soft Computing*, *25*, 242–252. doi:10.1016/j.asoc.2014.09.032

Ghanbari, A., Ghaderi, S. F., & Azadeh, M. A. (2010, February 26). A clustering based genetic fuzzy expert system for electrical energy demand prediction. *Proceedings of Computer and Automation Engineering (ICCAE), 2010 2nd International Conference* 2010 (Vol. 5, pp. 407-411). doi:10.1109/ICCAE.2010.5451228

Huang, P. S., He, X., Gao, J., Deng, L., Acero, A., & Heck, L. (2013, October). Learning deep structured semantic models for web search using clickthrough data.*Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management* (pp. 2333-2338). ACM. doi:10.1145/2505515.2505665

Jain, A., Sharma, R., Dixit, G., & Tomar, V. (2013, April 6). Page ranking algorithms in Web mining, limitations of existing methods and a new method for indexing Web pages.*Proceedings of International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 640-645). IEEE. doi:10.1109/CSNT.2013.137

Jang, J. S. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, *23*(3), 665–685. doi:10.1109/21.256541

Jiang, D., Pei, J., & Li, H. (2013). Mining search and browse logs for web search: *A survey.*[TIST]. *ACM Transactions on Intelligent Systems and Technology*, *4*(4), 57. doi:10.1145/2508037.2508038

Khan, S. A., Khodke, P. A., & Bhagat, A. P. (2015, March). Neuro Fuzzy based user queries categorization. *Proceedings of the2015 International Conference* on *Innovations in Information, Embedded and Communication Systems (ICIIECS)* (pp. 1-7). IEEE.

Klusch, M., Kapahnke, P., Schulte, S., Lecue, F., & Bernstein, A. (2015). Semantic Web Service Search: A Brief Survey. *KI-Künstliche Intelligenz.*

Lee, C. C. (1990). Fuzzy logic in control systems: Fuzzy logic controller. *IEEE Transactions on Systems, Man, and Cybernetics*, *20*(2), 419–435. doi:10.1109/21.52552

Lee, K. C., Kim, J. S., Chung, N. H., & Kwon, S. J. (2002). Fuzzy cognitive map approach to Web-mining inference amplification. *Expert Systems with Applications*, *22*(3), 197–211. doi:10.1016/S0957-4174(01)00054-9

Leung, C. H., Chan, A. W., Milani, A., Liu, J., & Li, Y. (2012). Intelligent social media indexing and sharing using an adaptive indexing search engine. *ACM Transactions on Intelligent Systems and Technology*, *3*(3), 47. doi:10.1145/2168752.2168761

Lin, A. D., Graydon, P. J., Busch, J. E., Caudill, M., Chinchor, N. A., Tseng, J. C., . . . Tijerino, Y. A. (2004). Concept-based search and retrieval system. U.S. Patent No. 6,675,159. Washington, DC: U.S. Patent and Trademark Office.

Lucchese, C., Orlando, S., Perego, R., Silvestri, F., & Tolomei, G. (2013). Discovering tasks from search engine query logs. *ACM Transactions on Information Systems*, *31*(3), 14. doi:10.1145/2493175.2493179

Mohammad, R. M., Thabtah, F., & McCluskey, L. (2014). Predicting phishing Websites based on self-structuring neural network. *Neural Computing & Applications*, *25*(2), 443–458. doi:10.1007/s00521-013-1490-z

Nikravesh, M. (2002). Fuzzy conceptual-based search engine using conceptual semantic indexing.*Proceedings of NAFIPS 2002 Annual Meeting of the North American Fuzzy Information Processing Society* (pp. 146-151). IEEE. doi:10.1109/NAFIPS.2002.1018045

Nikravesh, M., Loia, V., & Azvine, B. (2002). Fuzzy logic and the Internet (FLINT): Internet, World Wide Web, and search engines. *Soft Computing*, *6*(5), 287–299. doi:10.1007/s00500-002-0182-x

Putra, W. E., & Akbar, S. (2013). Focused Crawling using Dictionary Algorithm with Breadth First and by Page Length Methods for Javanese and Sundanese Corpus Construction. *Proceedings of the4th International Conference on Electrical Engineering and Informatics, ICEEI 2013* (*Vol. 11*, pp. 870-876). doi:10.1016/j.protcy.2013.12.270

Sisodia, D. S., & Verma, S. (2012, May 30). Web usage pattern analysis through Web logs: A review. *Proceedings of the2012 International Joint Conference onComputer Science and Software Engineering (JCSSE)* (pp. 49-53). IEEE. doi:10.1109/JCSSE.2012.6261924

Smirnov, A., Levashova, T., & Shilov, N. (2015). Patterns for context-based knowledge fusion in decision support systems. *Information Fusion*, *21*, 114–129. doi:10.1016/j.inffus.2013.10.010

Szczepaniak, P. S., Segovia, J., & Zadeh, L. A. (Eds.). (2012). *Intelligent exploration of the web* (Vol. 111). Physica.

Tao, X., Li, Y., & Zhong, N. (2011). A personalized ontology model for Web information gathering. *IEEE Transactions on Knowledge and Data Engineering*, *23*(4), 496–511. doi:10.1109/TKDE.2010.145

Tyagi, K., & Sharma, A. (2014). An adaptive neuro fuzzy model for estimating the reliability of component-based software systems. *Applied Computing and informatics, 10*(1), 38-51.

Umagandhi, R., & Kumar, A. S. (2013). Search Query Recommendations using Hybrid User Profile with Query Logs. *International Journal of Computers and Applications*, *80*(10).

Vidhyapriya, N., & Sampath, S. (2016). Fuzzy C Means Algorithm for Inferring User Search Goals with Feedback Sessions. *International Journal of Computer Science and Network Security*, *16*(1), 93.

Yan, L., Gui, Z., Du, W., & Guo, Q. (2011). An improved PageRank method based on genetic algorithm for Web search. *Procedia Engineering*, *15*, 2983–2987. doi:10.1016/j.proeng.2011.08.561

Yang, Y., Sun, Y., Tang, J., Ma, B., & Li, J. (2015, August). Entity Matching across Heterogeneous Sources. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1395-1404). ACM. doi:10.1145/2783258.2783353

Yu, X., Ma, H., Hsu, B. J., & Han, J. (2014, February). On building entity recommender systems using user click log and freebase knowledge.*Proceedings of the 7th ACM International Conference on Web Search and Data Mining* (pp. 263-272).ACM. doi:10.1145/2556195.2556233

Rawan Ghnemat received her PhD in Computer Science, with distinction, from Le Havre University, France in 2009. Her study was funded by French government scholarship. In 2009, she worked as Assistant Professor in the department of Computer Science at German-Jordanian University. Since September 2011 she has worked as Assistant Professor in the department of Computer Science at Princess Sumaya University for Technology. She has several scientific publications in artificial intelligence and data mining. Rawan Ghnemat has published on journal such as: International Journal of Bifurcation and Chaos, International Journal of Interactive Mobile Technologies, Journal of Software Engineering and Applications, and International Journal of Computers & Technology. She is a strategic steering committee member in an artificial intelligence powered search engines firm (GETJo Inc.), http://gaset-gbset.com/Team.html.

Adnan Shaout received his BS, MS and PhD degrees in 1982, 1983, and 1987 from Syracuse University respectively. Since 1987, he has been with the University of Michigan - Dearborn, where he is currently a full professor in the Electrical and Computer Engineering Department. His current research interests are in the areas of applications of fuzzy engineering applications, software engineering, computer design (hardware & software), intelligent Systems, and embedded systems. He was on sabbatical at Princess Sumaya University for the academic year of 2015/2016. He has published more than 190 papers. He has published in journals such as: the IEE – Part E, the IEE electronic letters, ACM Transaction on Software Engineering Methodology, International Journal of Advanced Research in Computer and Communication Engineering, the International Journal of Internet and Web Technology, the journal of intelligent & fuzzy systems – Applications in Engineering and Technology, The journal of Expert Systems with Applications.