

Adaptation of a Cluster Discovery Technique to a Decision Support System

Namdar Mogharreban

**Department of Computer Science, Southern Illinois University
Carbondale, Illinois 62901**

namdar@cs.siu.edu

Abstract

This paper reports on the implementation of a cluster discovery technique to a decision support system. The model is a multi-criteria multi-alternative decision environment. The decision support reported here is a diagnostic system. Adaptive Resonance Theory 1 (ART1), which is a cluster discovery neural network, was employed to cluster related symptoms to identify a specific disorder. ART1 was then modified to determine the degree of the belongingness of a new set of values to a cluster. For the purpose of demonstrating the functionality of the system, a set of symptoms for various diseases of the eye was utilized to create different clusters. Thus, each cluster represents a sub-type of an eye disorder. The closeness of a set of values representing a user's symptoms to a particular cluster is first determined and a rank ordering of the degree of membership to each cluster is returned. The cluster with the highest rank is reported as the disorder. The proposed clustering approach avoids major problems faced by the traditional multi-criterion decision making practices. First, the clustering approach is immune to the problem of interdependence of the criteria. Second, the clustering approach avoids the problems arising from the criteria being measured along different dimensions. Finally the clustering approach is unaffected by heterogeneity of the criteria. The system is implemented as a web application and the functionality of its implementation in the teaching environment as well as use by lay persons is discussed.

Key words: Multi-criteria, Decision making, Neural Network, Cluster Discovery, Web Application, Diagnostic Systems.

Introduction

In every day decision making there are very few decisions made without a number of competing alternatives. A decision maker must choose an alternative by evaluating a number of criteria. As the number of alternatives and criteria increases, the task of selecting the best alternative becomes increasingly difficult.

Material published as part of this journal, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact Publisher@InformingScience.org to request redistribution permission.

A typical multi-criteria multi-alternative decision making model can be represented by a matrix, such as shown in Table 1, with M number of alternatives and N criterion such that $i = 1, 2, 3, \dots, m$ and $j = 1, 2, 3, \dots, n$ [14]. Each element of the matrix a_{ij} indicates the user's response to alternative A_i when it is evaluated in terms of criterion C_j .

Table 1: A Typical Decision Matrix

Criterion		C_1	C_2	C_3	...	C_n
Alternatives	Wts	W_1	W_2	W_3	...	W_n
A_1		a_{11}	a_{12}	a_{13}	...	a_{1n}
A_2		a_{21}	a_{22}	a_{23}	...	a_{2n}
A_3		a_{31}	a_{32}	a_{33}	...	a_{3n}
.				
.						.
.						.
A_m		a_{m1}	a_{m2}	a_{m3}	...	a_{mn}

To accommodate the importance of different goals or criteria a weight value, sometimes referred to as decision weight, is associated with each criterion that signifies the importance of the criterion. A numeric value is then assigned to each cell in the matrix by the decision maker that represents the level of significance a criterion may have within a particular alternative. The populated matrix is then manipulated and a rank ordering of the alternatives is obtained. The best alternative considering the inputs by the decision maker can then be selected. Depending on the number of criteria and the units upon which each is evaluated determining the best alternative becomes a complex task.

In this paper, we are proposing a modification of a clustering technique that indicates the degree of closeness of a vector to a vector representing an alternative. The alternative choices employed are different disorders of the eye and the criteria are different symptoms of the eye disorders. An alternative is determined by the cluster discovery technique ART1, an unsupervised cluster discovering technique using the set of values for 10 different symptoms. ART1 is then modified and the closeness of the new set of values representing the user's symptoms to different clusters is determined. The closest matching alternative to the given symptoms is returned by the system. This allows for determination of the best matching disorder to the set of symptoms.

This paper is organized as follow. First, some typical multi-criteria decision methods (MCDM) that are utilized to validate our design are discussed. We then present the clustering technique used and the modifications of its algorithm. The results of comparisons between the typical MCDM models and our design are then presented. Finally, the implementation of and the development of future related work we are currently engaged in are discussed.

MCDM Methods

Different MCDM methods have been proposed for the numeric manipulation of a decision matrix to arrive at the best alternative (Chen, & Hwang, 1991; Hwang, & Yoon, 1981; Zimmerman, 1996). The first stage of processing the matrix is manipulation of the criteria to obtain a value that can best represent each alternative. However, depending on the type of values being considered, the operation may not be a straightforward process. For example, a manager is trying to decide between three different vendors. The criteria on which the alternatives are compared include price, delivery cost, and restocking cost for the surplus. Each criterion is evaluated based on its dollar value and the lower the values the more attractive the alternative. The criteria are said to be homogeneous and along the same dimension. They are both costs, measured in dollar

terms, and the lower the cost the better the alternative. However, if the choices are between foreign and domestic vendors one of the criterion might be the delivery time. The criteria, cost value versus time, are along different dimensions. That is, of course, to assume that time can not be measured in dollar terms. Thus it is not clear the type of operations that would result in the most accurate representation for the criteria (Hamalainen & Salo, 1997). Another criterion might be familiarity with the vendor. No monetary value can be assigned to this criterion. The criteria are heterogeneous, since the higher the familiarity, the better the alternative. Therefore, both the dimensionality and the heterogeneity of the criteria become of significant importance in avoiding arbitrary and false ranking of the alternatives (Mateu, 2002). The analysis is further complicated by potential interdependence of the criteria. Fuller and Carlsson (1996) point out that “multiple goals or objectives almost by necessity represent conflicting interests. By not allowing interdependence multi criteria problems are simplified beyond recognition...” (p. 140). In the next section, we will review three popular analytical methods in MCDM to illustrate how the problems of homogeneity and dimensionality are managed.

One of the earliest methods is the weighted sum model (WSM) (Fishburn, 1967) and its slight variation the weighted product model (WPM) (Miller, & Starr, 1969). The underlying assumption in these approaches is the additive utility assumption. The best alternative is determined by summing the product of the individual matrix entry and its weight and choosing the one with the highest sum value. Equation 1 represents the WSM calculation where n is the number of criteria and w_j is the weight of importance of the j -th criterion.

$$A = \max_i \sum_{j=1}^n a_{ij}w_j, \text{ for } i = 1, 2, 3, \dots, m \quad \text{Equation 1}$$

The alternative with the highest value of A is returned as the best alternative. The technique is robust so long as there is only one dimension and all the units for the criteria are the same. However, when multi-dimensional problems are considered with different units of measurements the additive utility assumption is violated and the technique is not valid.

A slight variation is the weighted product method in which a performance value (P) is calculated using equation 2:

$$P(A_k) = \prod_{j=1}^n (a_{kj})^{w_j} \quad \text{Equation 2}$$

where n is the number of criteria, a_{kj} is the actual value of the k -th alternative in terms of the j -th criterion, and w_j is the weight of importance of the j -th criterion. The alternative with the highest P value is considered the best alternative. The WPM is sometimes called dimensionless analysis. The calculation avoids any measurement units (Triantaphyllou, 2000). It is applied in both single- and multi-dimensional problems. However, since the weights are used as a power factor heterogeneous criteria will pose a problem with this technique. High values for different criterion tend to cancel each other and result in a false ranking.

One of the most popular analytical techniques for complex MCDM problems is the analytic hierarchy process (AHP) and its variants. Based on multiple attribute utility theory, Saaty (1980) developed AHP which decomposes a MCDM into a system of hierarchies of objectives, alternatives and criteria. An AHP hierarchy can have as many levels as needed to fully characterize a particular decision situation. A number of functional characteristics make AHP a useful methodology. These include the ability to handle decision situations involving subjective judgments, multiple decision makers, and its ability to provide measures of consistency of preference (Triantaphyllou, 2000).

In AHP the relative importance of the criteria and preferences among the alternatives is compared using a pair-wise comparison. A 9-point system ranging from 1 (the two options are equally preferred) to 9 (one choice option is extremely preferred over the other) is utilized to solicit the decision makers' opinion. A priority weight vector is then calculated for each criterion using the preference scores and then normalized, i.e. summing to 1. A similar matrix is calculated for the alternatives. The final step in the AHP process is to multiply the criteria vector by the alternative matrix to obtain a vector of normalized unit-less weighted preference scores for each of the alternatives. The alternative with the highest relative value is considered the preferred choice.

Despite the popularity of AHP as a MCDM in many areas (Zahedi, 1986) it is not without its critics. The process is criticized on the bases of lack of a theoretical foundation for the formation of hierarchies as well as for the use of a ratio scale for ranking a subjective opinion (de Steiguer, Duberstein, & Lopes, 2003). Next, we present a brief discussion of soft-computing in general and our utilization of neural net for implementation of a MCDM to remedy some of the issues confronted with the traditional methodologies just outlined.

Soft Computing and Problem Solving

A more intuitive approach to MCDM analysis is soft-computing which is more elastic and in line with human decision making. Soft Computing is the umbrella concept that has as its constituents fuzzy logic (FL), probabilistic reasoning (PR), and neural network theory (NN). Zadeh (1994) points out that soft computing is not a mélange of FL, PR, and NN, rather it is a partnership in which each partner contributes a distinct methodology for addressing problems in its domain. The particular techniques of Fuzzy Logic and Fuzzy Sets were developed by Zadeh (1965) as a means of solving problems in the soft sciences, particularly those that involve interactions between humans and/or humans and machines. Briefly stated, soft computing attempts to understand and incorporate the imprecision of the real world in research and in the interpretation of research. To the extent that alternative choices may not have clear and exact boundaries, soft computing seems to be well suited for this domain of human problem solving practices. For a more complete review of the recent development in MCDM and soft computing refer to (Fuller, & Carlsson, 1996).

MCDM and Neural Network

To illustrate the usefulness and take advantage of soft computing as an approach in MCDM, a medical diagnostic system is implemented as a prototype. Adaptive Resonance Theory 1 (ART1), developed by Carpenter and Grossberg (1987, 1995), was modified and utilized to determine the closeness of the match between a set of symptoms representing a user's symptoms to a cluster of symptoms representing a disease of the eye. Alternative disorders are rank ordered by the closeness value. Below ART1 is briefly discussed followed by a discussion of modifications of the algorithm for determining a closeness value. The proposed approach eliminates the problems of dimensionality and heterogeneity of the criteria. Since the evaluation is the closeness between a new set of values and an ideal set of values, the dimension along which each criterion is evaluated is irrelevant. And since each criterion is evaluated independently, the homogeneity of the criteria is not necessary.

3.1 Cluster Discovery Network (ART1)

Adaptive Resonance Theory 1 (ART1) identifies a cluster by calculating a similarity measure between a new vector V_n to the pattern for a previous vector V_p equation 3.

$$S = \frac{\|V_n \wedge V_p\|}{\|V_n\|} \quad \text{Equation 3}$$

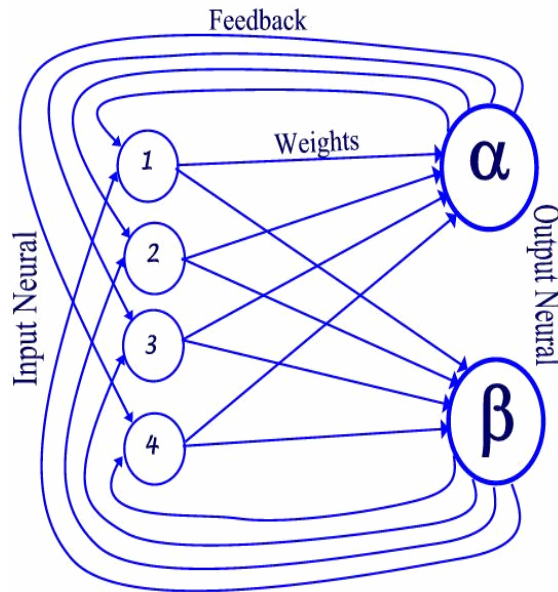


Figure 1. The conceptual view of ART1

The similarity value is then compared to a threshold value called the vigilance value ρ . If the similarity measure is less than the vigilance value, $S \leq \rho$, a new cluster is created; otherwise, the pattern is classified as belonging to an existing cluster. Thus, the number of clusters created is a function of the ρ . At a high vigilance value ($\rho \rightarrow 1$), a large number of specific clusters is created. Conversely, low vigilance ($\rho \rightarrow 0$) value will generate fewer and more general categories (Massey, 2002). Weight and feedback values are generated and the process repeats for the next entry. The conceptual model of ART1 is depicted in Figure 1. Variations on ART have been successfully implemented in many areas including document clustering (Kondadadi, & Kozma, 2002).

Modifications to the Algorithm

For obtaining the closeness value for a vector to a cluster, the Adaptive Resonance Theory1 was modified in three respects. ART1 will create a new cluster if an input vector can not be classified into an existing cluster based on the vigilance value. In this implementation the numbers of disorders, i.e., the number of clusters, are fixed by the number of disorders. No new disorder can be created. Therefore, the network is designed to report probability values regardless of the relationship between the similarity value and the vigilance value. The probability values returned indicate the likelihood that a given user's input belongs to each disorder.

Our second modification of ART1 involves the updating process for the weight and feedback values. In our implementation once the original clusters, i.e. disorders, have been identified, the weight and feedback values do not get updated based on the input entries supplied by the user. The rationale is to ensure that classification of a new input entry into a cluster is independent of previous input entries. Once the clusters/disorders have been identified, no new clusters are formed and the new set of values is identified as belonging to an existing cluster.

The third modification involves the classification of an input vector into a cluster. ART1 sorts the clusters based on the matching scores after every input entry. A new entry is assigned to the first cluster for which its similarity score is greater than the vigilance value. Once a cluster is identified and input entry is classified, the process stops and moves to the next input entry. In our implementation, however, we calculate the similarity scores for all the clusters. The similarity scores are placed first in the similarity array and the matching scores are placed in a matching array. Based on the vigilance value and the highest matching value, the highest similarity cluster is selected and the input array is classified as belonging to that cluster. This means that the disorder that has the highest matching value and similarity to the user's input is reported rather than simply the first match found.

Validation Results

In order to validate our proposed Neural Net (NN) we compared its performance with two algorithmically implemented methods. We implemented the algorithms for the Weighted Sum and Weighted Product methods in Microsoft Excel. There were six alternative disorders, A1 – A6, each associated with a set of 10 symptoms, S1 – S10. We generated 100 sets of symptoms assigning random values within the range of the values for a symptom to S1 – S10. We compared the rank order of the alternatives that were returned in the WSM and WPM and our Neural Net. We obtained 61% match of the returned alternatives between the NN and WSM and 65% match between the returned alternatives of NN and WPM. Although a higher rate of the match was expected between the methods, the results are consistent with what is reported in the literature that there exists variation between different methods on reporting the best alternative (Triantaphyllou & Baig, 2005).

When we considered the top two alternatives returned regardless of the rank order, the match between the methods approached 80%. The results are depicted in Figure 2. Although a higher matching rate might have been obtained with real patients' data, due to the difficulty of obtaining such a data set, the author feels that the present comparison supports the validity and robustness of the proposed system.

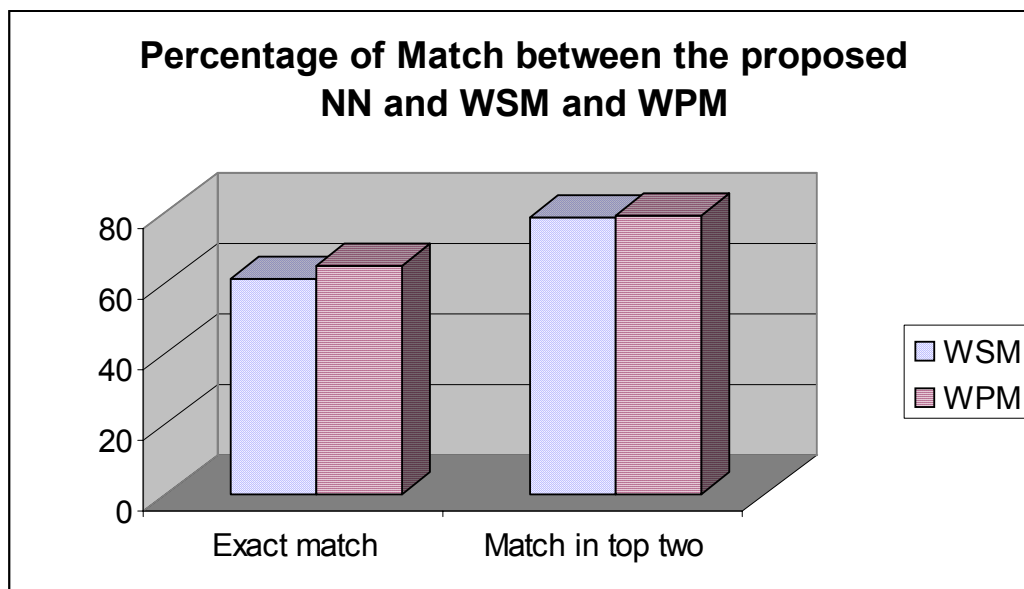


Figure 2. The comparison results between Neural Network and the two algorithmically implemented techniques.

Implementation Platform

The proposed diagnostic system is implemented in a client server model where the user can interact with the knowledge base via a thin client interface. The server is an Apache HTTP server version 1.3 that communicates with the application server via Apache Tomcat 4.1 which is based on the Java platform that supports the servlet and Java Server Page (JSP) technology.

The knowledge base is built using Microsoft Access database with a number of tables consisting of data on diagnoses and symptoms. MS-Access supports many types of Java programs. The main part of the application is written with JSP. It is connected with database tables by JDBC through the SQL statements and special methods. The java files also perform the main computa-

tional tasks such as the ranking procedure. The technology utilized is readily available and modifiable. The objective is to deliver a flexible and inexpensive system that can easily be expanded and modified.

Summary and Discussion

In our prototype implementation, a set of 10 values representing a criteria set has been associated with an alternative. The criteria are symptoms and the alternatives are disorders of the eye. The similarity of a new set of values for the same symptoms would indicate the possibility of a particular disease. Our proposed system assumes that there is an ideal set of values for a cluster and determines if a new set of values belongs to a particular cluster. Indeed, there are many decision making environments such as manufacturing, managerial and medical diagnoses where an optimum set of values for the criteria exist (Joerin, & Musy, 2000; Lahdelma, Salminen, & Hokkanen, 2000; Simon, 1977). Thus the best decision/alternative is one in which its criteria vector closely matches the criteria vector of an alternative choice. Further, using this model the heterogeneity, dimensionality, and interdependence of criteria pose no problem for the manipulation of the criteria values and determination of the best alternative.

Since the implementation solicits responses from the user with terms such as severe or often, etc. (see Figure 3), and is developed as a web application, the use of the system is simplified and can potentially be extended for use by people without medical expertise. As Saade (2003) points out 'perceived ease of use' has a strong influence on 'perceived usefulness'. Given the appropriate

The image shows a web application interface titled "Eye Irritation". It features a table with 10 rows, each representing a symptom and its corresponding value. The values are selected from dropdown menus. At the bottom of the form, there are two buttons: "Submit for Diagnosis" and "Reset Selections".

<i>Eye Irritation</i>	
Conjunctival infection	Absent
Cornea	Clear
Discharge	Watery
Intraocular pressure	Normal
Pain	Mild
Pupil size	Small & Constricted
Light sensitivity	Severe
Smear	None
Vision	Blurred
Irritation	Please choose from below

Submit for Diagnosis Reset Selections

Figure 3. Web screen image for obtaining user's input

disclaimers, this model can be used for the initial screening of some common disorders such as back pain, sore throat or even severe acute respiratory syndrome (SARS) by lay individuals. By increasing the number of criteria and improving the explanations of the alternatives, the proposed application could be used as an expert system for training nursing or medical students.

We are continuing the improvement of different aspects of the system. The First step is to increase the number of criteria used for the creation of a disease cluster and expand it to include other potential influential factors such as gender, age, or weight (see Figure 4).

Second, we expect that training the neural network and testing the system with actual patients' data will enhance the system and result in better alternative selection. Although the clusters were identified using values from clinical sources, the test data were randomly generated. It is difficult to predict if this practice contributed to lower rates of matching between the various methods. Indeed, a high degree of interdependency is expected to exist between the various symptoms of a particular disorder. However, such interdependency can not be expected from a set of random

Disorders

Patient Information	Patient Complaint
Gender <input checked="" type="radio"/> Male <input type="radio"/> Female	<input type="radio"/> Back Pain
	<input checked="" type="radio"/> Eye Irritation
	<input type="radio"/> Leg Ache
	<input type="radio"/> Chest Pain
Age Group	
<input type="radio"/> Child (below 11 years)	
<input checked="" type="radio"/> Young (Adult 11 - 24)	
<input type="radio"/> Adult (25 - 55)	
<input type="radio"/> Senior (above 55)	
Weight	
<input type="radio"/> Under Weight <input checked="" type="radio"/> Normal <input type="radio"/> Over Weight	
<input type="button" value="Submit Form"/> <input type="button" value="Reset Form"/>	

Figure 4. Web screen image for selecting a disorder

numbers. Finally, the implementation could be enhanced by incorporating fuzzy logic in the development of the input vector for the neural network. In the present implementation, the responses are mapped to a crisp value for each symptom. Generating a membership value for each symptom and then creating a vector as the input for the neural net enhances flexibility and the ability to handle vagueness of the responses and potentially improve the diagnosis.

References

- Carpenter, G.A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37, 54-115.
- Carpenter, G.A., & Grossberg, S. (1995). Adaptive Resonance Theory (ART). In M. A. Arbib, *Handbook of brain theory and neural networks*. MIT Press.
- Chen, S.J., & Hwang, C.L. (1991). Fuzzy multiple attribute decision making: Methods and applications. *Lecture notes in economics and mathematical systems*, No. 375, Springer-Verlag, Berlin, Germany.
- de Steiguer, J. E. Duberstein, J. & Lopes, V. (2003). *Proceeding of the First Interagency Conference on Research in the Watersheds*, October 27-30 Benson Arizona, pp. 736-740.
- Fishburn, P. C. (1967). *Additive utilities with incomplete product set: Applications to priorities and assignments*. Baltimore, MD: Operations Research Society of America (ORSA).
- Fuller, R., & Carlsson, C. (1996). Fuzzy multiple criteria decision making: Recent developments. *Fuzzy Sets and Systems*, 78, 139-153
- Hamalainen, R. P., & Salo A. A. (1997). Rejoinder: The issue of understanding the weights. *Journal of Multi-Criteria Decision Analysis*, 6, 340-343.
- Hwang C.L., & Yoon, K. (1981). *Multiple attribute decision making: Methods and applications*. New York, NY: Springer-Verlag.
- Joerin F., & Musy, A. (2000). Land management with GIS and multicriteria analysis. *International Transactions in Operational Research*, 7(1), 67-78.
- Kondadadi, R., & Kozma, R. (2002). A modified fuzzy art for soft document clustering. In *Proceedings of the International Joint Conference on Neural Networks*, Honolulu, HI.
- Lahdelma R., Salminen, P., & Hokkanen, J. (2000). Using multicriteria methods in environmental planning and management. *Environmental Management*, 26(6), 595 – 605
- Massey, L. (2002). Determination of clustering tendency with ART neural networks. *Proceedings of the 4th Intl. Conf. on Recent Advances in Soft Computing*, Nottingham, U.K., 12, & 13 December.
- Mateu, A. V. (2002). ClusDM: A multiple criteria decision making method for heterogeneous data sets. Doctoral Thesis available at <http://www.tdx.cesca.es/TDX-0206103-205841>
- Miller, D. W., & Starr M. K. (1969). *Executive decisions and operations research*. Englewood Cliffs, NJ: Prentice-Hall.
- Simon, H. (1977). *The new science of management decision*. Englewood Cliffs, NJ: Prentice Hall.
- Saade, R. G. (2003). Web-based educational information system for enhanced learning, EISEL: Student assessment. *Journal of Information Technology Education*, 2, 267-277. Available at <http://jite.org/documents/Vol2/v2p267-277-26.pdf>
- Saaty, T. L. (1980). *The analytic hierarchy process*. New York, NY: McGraw Hill.
- Triantaphyllou, E. (2000). *Multi-criteria decision making methods: A comparative study*. Netherlands: Kluwer Academic Publishers.
- Triantaphyllou, E., & Baig, K. (2005). The impact of aggregating benefit and cost criteria in four MCDA methods. *IEEE Transactions on Engineering Management*, 52(2), 213-226.
- Zadeh L.A. (1965). Fuzzy Algorithm. *Information & control*, 8, 338-353.

Adaptation of a Cluster Discovery Technique

- Zadeh L.A. (1994). Fuzzy logic, neural networks, and soft computing. *Communications of the ACM*. New York.
- Zahedi, F. (1986). The analytic hierarchy process: A survey of the method and its application. *Interfaces*, 16, 96-108.
- Zimmerman, H. J. (1996). *Fuzzy set theory and its applications* (Third Revised Edition). Boston, MA: Kluwer Academic Publishers.

Biography



Namdar Mogharreban received his PhD in Computer Based Education in 1989. He has taught computing courses in Management Department, Technical School and most recently in a Department of Computer Science. The primary area of research interest is end-user computing, development of intelligent applications, and soft computing.