Genetic Learning of Membership Functions for Mining Fuzzy Association Rules

Rafael Alcalá and Jesús Alcalá-Fdez and M.J. Gacto and Francisco Herrera

Abstract— Data mining is most commonly used in attempts to induce association rules from transaction data. Most previous studies focused on binary-valued transaction data. Transaction data in real-world applications, however, usually consists of quantitative values. In the last years, the fuzzy set theory has been applied to data mining for finding interesting association rules in quantitative transactions.

Recently, a new rule representation model was presented to perform a genetic lateral tuning of membership functions. It is based on the 2-tuples linguistic representation model allowing us to adjust the context associated to the linguistic label membership functions.

Based on the 2-tuples linguistic representation model, we present a new fuzzy data-mining algorithm for extracting both association rules and membership functions by means of an evolutionary learning of the membership functions, using a basic method for mining fuzzy association rules.

I. INTRODUCTION

Data Mining (DM) is the process for automatic discovery of high level knowledge by obtaining information from real data. Some important problems that DM deal with are: rule extraction, identification of associations, feature analysis, linguistic summarization, clustering, classifier design and novelty/anomaly detection [1].

Association rules are used to represent and identify dependencies between attributes in a database [2]. Most previous studies focused on database with binary values [3], however the data in real-world applications usually consist of quantitative values. Designing sophisticated DM algorithms, able to deal with various types of data, presents a challenge to workers in this research field.

Lately, fuzzy set theory has been used more and more frequently in intelligent systems because of its simplicity and similarity to human reasoning [4]. The use of fuzzy sets to describe association between data extends the types of relationships that may be represented, facilitates the interpretation of rules in linguistic terms, and avoids unnatural boundaries in the partitioning of the attribute domains [5], [6].

Many researchers have proposed methods for mining fuzzy association rules from quantitative data [7], [8], [9], where the membership functions (MFs) were assumed to be known in advance. The given MFs may have a critical influence on

Rafael Alcalá, Francisco Herrera and M.J. Gacto are with the Department of Computer Science and Artificial Intelligence, University of Granada, Granada, 18071, Spain (phone: +34 958 244019; email: {alcala, herrera}@decsai.ugr.es, mjgacto@ugr.es).

Jesús Alcalá-Fdez is with the Department of Computer Science, University of Jaén, Jaén, 23071, Spain (phone: +34 953 212477; email: jalcala@decsai.ugr.es).

the final mining results. For this reason, some approaches have also achieved a learning or tuning of the MFs [10], [11], [12], [13], [14].

Recently, a new linguistic rule representation model was proposed to perform a genetic lateral tuning of MFs [15]. This new approach was based on the 2-tuples linguistic representation [16], that allows the symbolic translation of a linguistic term by considering only one parameter per linguistic term. In this way, two main objectives were achieved:

- to obtain MFs with a more adequate context by maintaining a high covering degree of the data, and
- to reduce the search space respect to the classic learning [17], [18], [19] (usually considering three parameters in the case of triangular MFs), in order to easily obtain optimal models.

Based on the 2-tuples linguistic representation model, in this work, we present a new fuzzy data-mining algorithm for extracting both association rules and MFs from quantitative transactions by means of an evolutionary learning of the MFs and the use of the method presented in [8] for mining the fuzzy association rules. In this way, the search space reduction provided by the 2-tuples linguistic representation helps the evolutionary search technique to obtain more suitable MFs.

The paper is arranged as follows. The next section describes the linguistic rule representation model based on linguistic 2-tuples. Section III introduces the mining scheme considered in this work. Section IV details the evolutionary learning algorithm proposed to obtain the MFs. Section V describes the proposed mining algorithm. Section VI shows the results of the proposed mining algorithm applied over a real-world problem. Finally, Section VII points out some conclusions.

II. PRELIMINARIES: THE 2-TUPLES LINGUISTIC REPRESENTATION

The 2-tuples linguistic representation scheme presented in [16], introduces a new model for rule representation based on the concept of symbolic translation (the lateral displacement of a linguistic term).
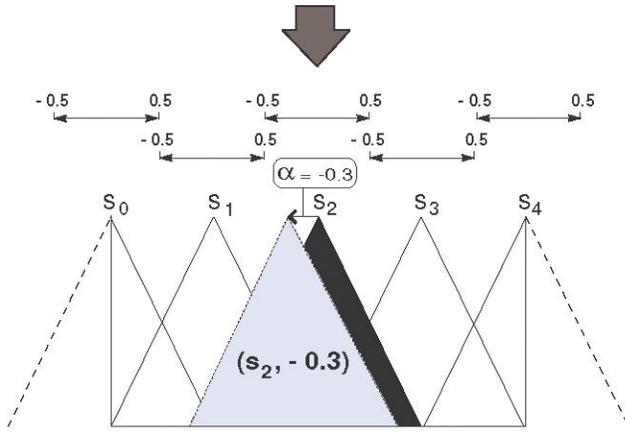
The symbolic translation of a linguistic term is a number within the interval [-0.5, 0.5) that expresses the domain of a linguistic term when it is moving between its two lateral linguistic term. Let us consider a set of linguistic terms S representing a fuzzy partition. Formally, we have the pair,

$$(s_i, \alpha_i), \quad s_i \in S, \quad \alpha_i \in [-0.5, 0.5).$$

Figure 1 depicts the symbolic translation of a linguistic term represented by the pair $(S_2, -0.3)$, considering a set S with five linguistic terms represented by their ordinal values ($\{0, 1, 2, 3, 4\}$).
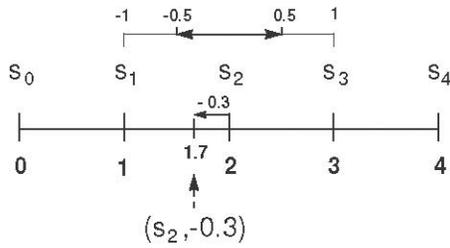
a) Simbolic Translation of a Linguistic Term

b) Lateral Displacement of a Membership Function

Fig. 1. Symbolic translation of a linguistic term and lateral displacement of the involved MF

In [16], both the 2-tuples linguistic representation model and the needed elements for linguistic information comparison and aggregation are presented and applied to the Decision Making framework. In [15], a new rule representation model has been presented based on these concepts to perform a tuning of complex linguistic fuzzy models. Now, we are going to extend its use for association rule representation. Below we present this approach considering a simple mining problem.

Let us consider a simple problem with two items (age and weight) and tree linguistic terms with their associated MFs (see Figure 2).
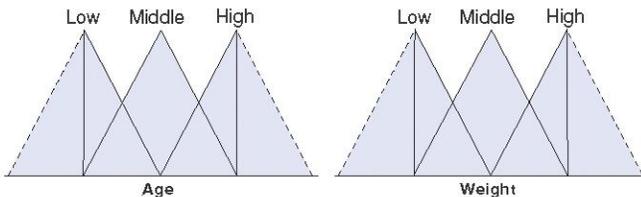

Fig. 2. Items and linguistic terms in a simple problem

Based on this definition, an example of classic association rule and 2-tuples linguistic represented rule is:

Classic Association Rule,
If the **Age** is Middle then the **Weight** is High.

Rule with 2-Tuples Representation,
If the **Age** is (Middle,0.3) then the **Weight** is (High,-0.1).

Analyzed from the rule interpretability point of view, we could interpret the 2-tuples linguistic represented rule in the following way:

If the **Age** is "higher than Middle"
then the **Weight** is "a bit smaller than High".

This proposal decreases the tuning problem complexity, since the three parameters usually considered per linguistic term [17], [18], [19] are reduced to only one symbolic translation parameter. Moreover, from the point of view of interpretability:

- the original shapes of the MFs are maintained (in our case triangular and symmetrical), by laterally changing the location of their supports,
- the lateral variation of the involved MFs is restricted to a short interval, *ensuring overlapping between two adjacent MFs* to some degree but preventing their vertex points from crossing, and
- the 2-tuples represented linguistic terms can be interpreted with respect to the initial ones.

III. THE MINING SCHEME

An efficient way to generate the fuzzy association rules consists of learning the MFs *a priori* and then use the final best set of MFs to mine fuzzy association rules [11], [13], [14]. This way to work allows us to learn the most adequate context [20] for each fuzzy partition, which is necessary in different contextual situations (different applications).

The scheme considered for discovering both useful fuzzy association rules and suitable MFs from quantitative values is comprised of two main stages (see Figure 3):

1) An evolutionary process to learn the MFs.
2) A method to mine fuzzy association rules. The method presented in [8] will be considered for this task as a first approach.
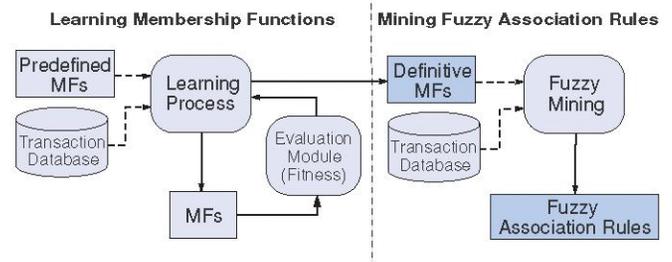

Fig. 3. Scheme for discovering both useful fuzzy association rules and suitable MFs

IV. Evolutionary Learning Algorithm

The automatic definition of fuzzy systems can be considered as an optimization or search process and nowadays Evolutionary Algorithms, particularly Genetic Algorithms (GAs), are considered as the more known and used global search technique. Moreover, the genetic coding that they use allow them to include prior knowledge and to use it leading the search up. For this reason, Evolutionary Algorithms have been successfully applied to learn and to tune fuzzy systems in the last years.

In this work, we will consider the use of GAs to design the proposed learning method of the MFs. A good evolution model is the CHC evolutionary model [21]. The CHC algorithm is a GA that presents a good trade-off between exploration and exploitation, being a good choice in problems with complex search spaces.

In the following, the components needed to design this process are explained. They are:

- CHC Evolutionary model.
- MFs codification and initial gene pool.
- Chromosome evaluation.
- Genetic operators.

A. CHC Evolutionary Model

We will consider a population-based selection approach, by using the CHC evolutionary model [21] in order to perform an adequate global search. The genetic model of CHC makes use of a "Population-based Selection" approach. N parents and their corresponding offspring are combined to select the best N individuals to take part of the next population. The CHC approach makes use of an incest prevention mechanism and a restarting process to provoke diversity in the population, instead of the well known mutation operator.

This incest prevention mechanism will be considered in order to apply the crossover operator, i.e., two parents are crossed if their hamming distance divided by 2 is over a predetermined threshold, L_T. Since, we will consider a real coding scheme, we have to transform each gene considering a Gray Code with a fixed number of bits per gene ($BITSGENE$) determined by the system expert. In this way, the threshold value is initialized as:

$$L_T = (\#Genes * BITSGENE)/4.0,$$

where $\#Genes$ is the number of genes in the chromosome. Following the original CHC scheme, L_T is decremented by one when there is no new individuals in the population in one generation. In order to make this procedure independent of $\#Genes$ and $BITSGENE$, in our case, L_T will be decremented by a $\varphi\%$ of its initial value (being φ determined by the user, usually 10%). The algorithm restarts when L_T is below zero.

A scheme of this algorithm is shown in Figure 4.

B. MFs Codification and Initial Gene Pool

A real coding scheme is considered, i.e., the real parameters are the GA representation units (genes). Each chromosome is a vector of real numbers with size $n*m$ (n items with
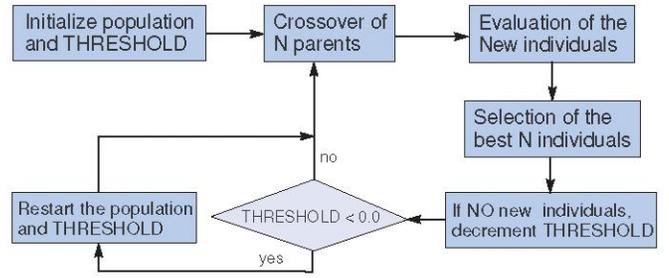

Fig. 4. Scheme of CHC

m linguistic terms per item) in which the displacements of the different linguistic terms are coded for each item. Then, a chromosome has the following form (where each gene is the tuning value of the corresponding linguistic term),

$$(c_{11}, \ldots, c_{1m}, c_{21}, \ldots, c_{2m}, \ldots, c_{n1}, \ldots, c_{nm})$$

Figure 5 graphically depicts an example of correspondence between a chromosome and its associated MFs. Notice that, the three parameters usually considered per linguistic term (in the case of triangular MFs) are reduced to only one parameter.
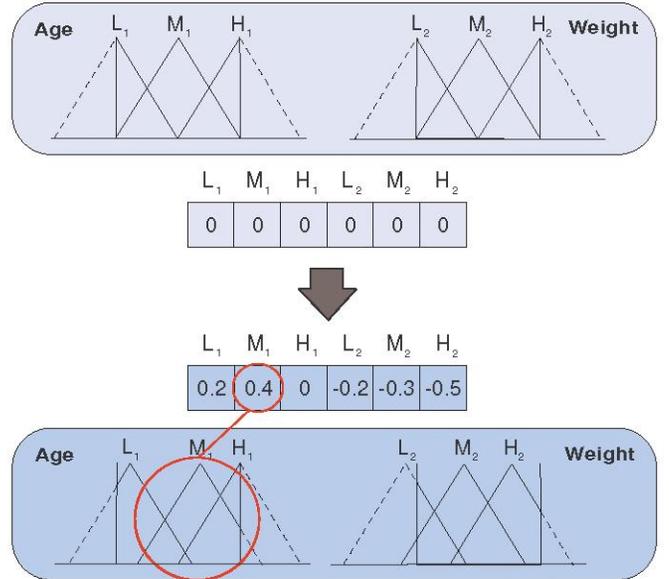

Fig. 5. Example of coding scheme

To make use of the available information, the initial MFs obtained from expert knowledge are included in the population as an initial solution. To do so, the initial pool is obtained with the first individual having all genes with value '0.0', and the remaining individuals generated at random in [-0.5, 0.5).

C. Chromosome Evaluation

To evaluate a determined chromosome we will use the fitness functions defined in [14]. Before the fitness of each set of MFs is formally described, several related terms are explained below.

The overlap ratio of two MFs R_i and R_j $(i < j)$ is defined as the overlap length divided by half the minimum span of the two functions. If the overlap length is larger than half the span, then these two MFs are thought of as a little redundant. Appropriate punishment must then be considered in this case. Thus, the overlap factor of the MFs for an item I_k in the chromosome C_q is defined as

$$\sum_{i=1}^{m}\sum_{j=i+1}^{m}[max(\frac{overlap(R_i, R_j)}{min(spanR_i, spanR_j)}, 1) - 1]$$

where $overlap(R_i, R_j)$ is the overlap length of R_i and R_j, and m is the number of MFs for I_k. Notice that the span of the two MFs is the same because in our case the original shapes of the MFs are maintained (triangular and symmetrical).

The coverage ratio of MFs for an item I_k is defined as the coverage range of the functions divided by the maximum quantity of that item in the transactions. The more the coverage ratio is, the better the derived MFs are. Thus, the coverage factor of the MFs for an item I_k in the chromosome C_q is defined as:

$$\frac{1}{\frac{range(R_1, ..., R_m)}{max(I_k)}}$$

where $range(R_1, R_2, ..., R_m)$ is the coverage range of the MFs and $max(I_k)$ is the maximum quantity of I_k in the transactions. Notice that the coverage factor is always 1 because in our case *the 2-tuples linguistic representation ensures the coverage in all the dominion*, reducing the computation time. The suitability of the set of MFs in a chromosome C_q, is therefore defined as:

$$\sum_{i=1}^{n}[overlap\_factor(C_{qi}) + 1]$$

where n is number of items. The fitness value of a chromosome C_q is then defined as

$$fitness(C_q) = \frac{|L_1|}{suitability(C_q)}$$

where $|L_1|$ is the number of large 1-itemsets obtained by using the set of MFs in C_q. The suitability factor can reduce the occurrence of the two bad kinds of MFs shown in Figure 6, where the first one is too redundant, and the second one is too separate. The overlap factor in $suitable(C_q)$ is used for avoiding the first bad case, and the 2-tuples linguistic representation prevents the second one.
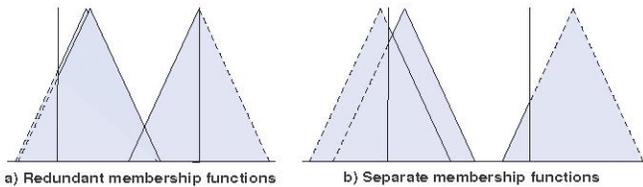

Fig. 6. Two bad kinds of membership functions

D. Genetic Operator

The genetic operators considered in CHC are crossover and restart approach (no mutation is considered). A description of these operators is presented in the following:

- Crossover. The crossover operator is based on the concept of neighborhood. These kinds of operators show a good behavior as is shown in [22]. Particularly, we consider the PBLX operator (an operator based on the BLX-α). This operator presents a good cooperation when it is introduced within models forcing the convergence by pressure on the offspring. Figure 7 shows the performance of these kinds of operators, which allow the offspring genes to be around the genes of one parent (PCBLX) or around a wide zone determined by both parent genes (BLX).
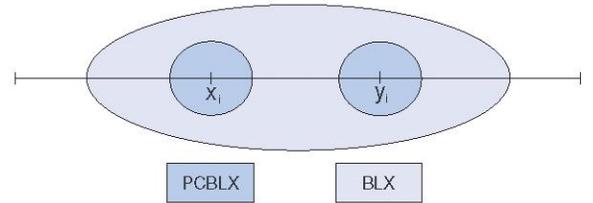

Fig. 7. Diagram of the performance of the crossover operators based on environments

The PCBLX is described as follows. Let us assume that $X = (x_1 \cdots x_n)$ and $Y = (y_1 \cdots y_n)$, $(x_i, y_i \in [a_i, b_i] \subset \Re, i = 1 \cdots n)$, are two real-coded chromosomes that are going to be crossed. The PCBLX operator generates the two following offspring:

- $O_1 = (o_{11} \cdots o_{1n})$, where o_{1i} is a randomly (uniformly) chosen number from the interval $[l_i^1, u_i^1]$, with $l_i^1 = max\{a_i, x_i - I_i\}$, $u_i^1 = min\{b_i, x_i + I_i\}$, and $I_i = |x_i - y_i|$.
- $O_2 = (o_{21} \cdots o_{2n})$, where o_{2i} is a randomly (uniformly) chosen number from the interval $[l_i^2, u_i^2]$, with $l_i^2 = max\{a_i, y_i - I_i\}$ and $u_i^2 = min\{b_i, y_i + I_i\}$.

- Restart approach. To get away from local optima, this algorithm uses a restart approach [21]. In this case, the best chromosome is maintained and the remaining are generated at random within the corresponding variation intervals [-0.5, 0.5]. It follows the principles of CHC [21], performing the restart procedure when a threshold value is reached or all the individuals coexisting in the population are very similar.

V. PROPOSED MINING ALGORITHM

According to the above description, the proposed algorithm for mining both MFs and fuzzy association rules is described below.

INPUT: T quantitative transaction data, a set of n items, each with m predefined linguistic terms, a support threshold α, a confidence threshold λ and a population size N.

OUTPUT: A set of fuzzy association rules with its associated set of MFs.

Stage 1. *Evolutionary learning of the MFs.*

Step 1: Generate the initial population with N chromosomes.
Step 2: Evaluate the population. For each chromosome:

- For each transaction datum D_i, $i=1$ to T, and for each item I_j, $j=1$ to n, transfer the quantitative value $v_j^{(i)}$ ($D_i = (v_1^{(i)}, ..., v_n^{(i)})$) into a fuzzy set $f_j^{(i)}$ represented as:

$$f_j^{(i)} = \left(\frac{f_{j1}^{(i)}}{R_{j1}} + ... + \frac{f_{jm}^{(i)}}{R_{jm}} \right)$$

using the corresponding MFs represented by the chromosome, where R_{jk} is the $k-th$ linguistic term of item I_j, $f_{jk}^{(i)}$ is $v_j^{(i)}$'s fuzzy membership value in region R_{jk}, and m is the number of linguistic terms for I_j.

- For each linguistic term R_{jk}, calculate its count on the transactions as follows:

$$count_{jk} = \sum_{i=1}^{T} f_{jk}^{(i)}$$

- For each R_{jk}, $1 < j < n$ and $1 < k < m$, check whether its $count_{jk}$ larger than or equal to the minimum support threshold α. If R_{jk} satisfies the above condition, put it in the set of large 1-itemsets (L_1). That is:

$$L_1 = \{ R_{jk} \mid count_{jk} \geq \alpha, 1 \leq j \leq n \text{ and } 1 \leq k \leq m \}$$

- Set the fitness value of the chromosome as the number of linguistic terms in L_1 divided by $suitability(C_q)$. That is:

$$fitness(C_q) = \frac{|L_1|}{suitability(C_q)}$$

Step 3: Initialize the threshold value L.
Step 4: Generate the next population:

- Shuffle the population.
- Select the parents two by two. Each pair is crossed (PCBLX) if the hamming distance between the parent Gray codings divided by 2 is over L.
- Evaluate the new individuals.
- Join the parents with their offspring and select the best N individuals to take part of the next population.

Step 5: If the best chromosome does not change or there are no new individuals in the population, $L = L - (L_{initial} \cdot 0.1)$.
Step 6: If $L < 0$, restart the population.
Step 7: If the maximum number of evaluations is not reached, go to Step 4.

Stage 2. *Basic method for mining fuzzy association rules.*

Step 8: The set of the best MFs is then used to mine fuzzy association rules from the given quantitative database. The fuzzy mining algorithm proposed in [8] is then adopted to achieve this purpose.

VI. EXPERIMENTAL RESULTS

To illustrate the proposed mining algorithm, a study of the atherosclerosis is considered. It is a study of the risk factors of atherosclerosis in a population of 1417 middle-aged men [1]. Here, we extract five quantitative attributes out of a total of 64. The selected attributes are height, weight, systolic blood pressure, diastolic blood pressure, and cholesterol level.

The values considered for the parameters of the proposed method are:

- Evolutionary process: 50 individuals, 10,000 evaluations and 0.6 as crossover probability for PBLX.
- Method for mining fuzzy association rules: 0.7 for the confidence threshold.

Figure 8 shows the relationship between the numbers of large 1-itemsets and several values for the minimum support with one uniform fuzzy partition and the proposed method. We can see that the learning of the MFs achieves larger a number of large 1-itemsets than the one with uniform fuzzy partitions considering different values of minimum support.
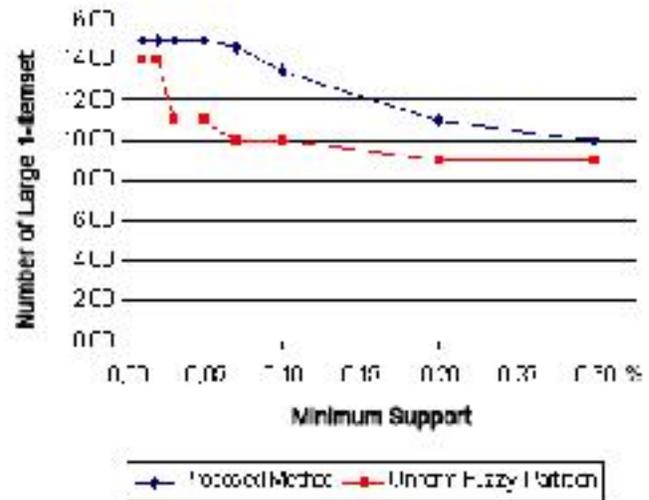

Fig 8. Relationship between numbers of large 1-itemsets and the minimum supports

Figure 9 depicts the MFs obtained by the proposed method from one of the runs performed in this problem with 3 linguistic terms by attribute. This figure shows how small displacements in the MFs lead to important improvements in the number of obtained rules. The MFs are more or less well distributed which allows us to easily give a meaning to the corresponding linguistic terms.

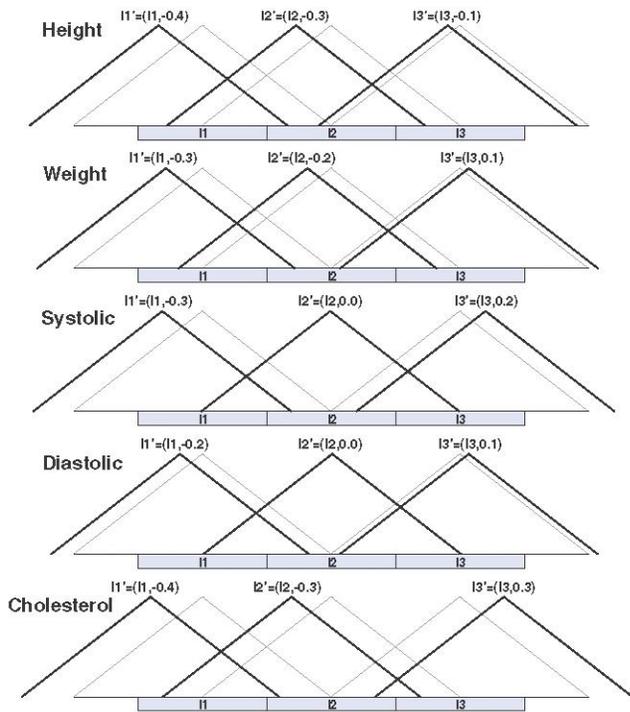One rule mined out with factor of confidence 0.94 is shown as an example:

Fig. 9. MFs with/without lateral displacements (black/gray)

If the **systolic blood pressure** is (Low,-0.12) and
the **diastolic blood** is (Low,-0.05)
then the **cholesterol level** is (Middle,0.03)

VII. Conclusions

In this work, a new rule representation scheme by using the 2-tuples linguistic representation model has been considered to extracting both MFs and fuzzy association rules from quantitative transactions by means of an evolutionary learning of the MFs and a later basic method to mine fuzzy association rules. Here, we present our conclusions and further considerations:

- The 2-tuples linguistic representation model allows an important reduction of the search space from the optimization point of view.
- The used learning scheme together with the 2-tuples linguistic representation model and the used fitness function offers a good mechanism to obtain interpretable association rules, since the shapes of the initial MFs are preserved, the MFs are maintained symmetrical and more or less well distributed, ensuring full coverage and the bad kinds of MFs are avoided in the mining process.

In the future, we will attempt to enhance the proposed method for more complex problems.

Acknowledgment

References

[1] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Second Edition, Morgan Kaufmann, 2006.

[2] Ch. Zhang and S. Zhang, *Association Rule Mining: Models and Algorithms*, Lecture Notes in Computer Science, LNAI 2307, 2002.

[3] R. Agrawal, T. Imielinski and A. Swami. "Mining association rules between sets of items in large databases", *SIGMOD*, Washington D.C., May 1993, pp. 207–216.

[4] H. Ishibuchi, T. Nakashima and M. Nii, *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining*, Springer, Berlin, 2004.

[5] M. Delgado, N. Marín, D. Sánchez and M.-A. Vila, "Fuzzy association rules: General model and applications", *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 2, pp. 214-225, 2003.

[6] D. Dubois, H. Prade and T. Sudamp, "On the representation, measurement, and discovery of fuzzy associations", *IEEE Transactions on Fuzzy Systems*, vol. 13, pp. 250–262, 2005.

[7] S. Yue, E. Tsang, D. Yeung and D. Shi, "Mining fuzzy association rules with weighted items", *IEEE International Conference on Systems, Man and Cybernetics*, Nashville, Tennessee, October 2000, pp. 1906–1911.

[8] T.P. Hong, C.S. Kuo and S.C. Chi, "Trade-off between time complexity and number of rules for fuzzy mining from quantitative data", *International Journal Uncertain Fuzziness Knowledge-Based Systems*, vol. 9, no. 5, pp. 587-604, 2001.

[9] Y.C. Lee, T.P. Hong and W.Y. Lin, "Mining fuzzy association rules with multiple minimum supports using maximum constraints", *Lecture Notes in Computer Science*, vol. 3214, pp. 1283–1290, 2004.

[10] W. Wang and S.M. Bridges, "Genetic Algorithm Optimization of Membership Functions for Mining Fuzzy Association Rules", *International Joint Conference on Information Systems, Fuzzy Theory and Technology Conference*, Atlantic City, N.Y., March 2000, pp. 1–4.

[11] M. Kaya and R. Alhajj, "A clustering algorithm with genetically optimized membership functions for fuzzy association rules mining", *IEEE International Conference on Fuzzy Systems*, St. Louis, Missouri, 2003, pp. 881–886.

[12] M. Kaya and R. Alhajj, "Integrating multi-objective genetic algorithms into clustering for fuzzy association rules mining", *Fourth IEEE International Conference on Data Mining*, Brighton, UK, November 2004, pp. 431–434.

[13] T.P. Hong, C.H. Chen, Y.L. Wu and Y.C. Lee, "Using Divide-and-Conquer GA Strategy in Fuzzy Data Mining", *Proc. IEEE International Symposium on Fuzzy Systems*, Budapest, Hungary, July 2004, pp. 116–121.

[14] T.P. Hong, C.H. Chen, Y.L. Wu and Y.C. Lee, "A GA-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership functions", *Soft Computing*, vol. 10, pp. 1091-1101, 2006.

[15] R. Alcalá, J. Alcalá-Fdez and F. Herrera, "A Proposal for the Genetic Lateral Tuning of Linguistic Fuzzy Systems and its Interaction with Rule Selection", *IEEE Transactions on Fuzzy Systems*, (2007) In press.

[16] F. Herrera and L. Martíez, "A 2-tuple fuzzy linguistic representation model for computing with words", *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 6, pp. 746–752, 2000.

[17] R. Babuška, J. Oosterhoff, A. Oudshoorn and P. M. Bruijn, "Fuzzy self-tuning PI control of pH in fermentation", *Engineering Applications of Artificial Intelligence*, vol. 15, no. 1, pp. 463–484, 2002.

[18] O. Cordón, F. Herrera, F. Hoffmann and L. Magdalena, *GENETIC FUZZY SYSTEMS. Evolutionary tuning and learning of fuzzy knowledge bases. Advances in Fuzzy Systems - Applications and Theory*, World Scientific, vol. 19, 2001.

[19] C. Karr, "Genetic algorithms for fuzzy controllers", *AI Expert*, vol. 6, no. 2, pp. 26–33, 1991.

[20] O. Cordón, F. Herrera, L. Magdalena and P. Villar, "A genetic learning process for the scaling factors, granularity and contexts of the fuzzy rule-based system data base", *Information Sciences*, vol. 136, pp. 85–107, 2001.

[21] L.J. Eshelman, "The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination", in *Foundations of genetic Algorithms*, G. Rawlin, Ed. Morgan Kaufman, 1991, vol. 1, pp. 265-283.

[22] F. Herrera, M. Lozano and A.M. Sánchez, "A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study", *International Journal of Intelligent Systems*, vol. 18, pp. 309–338, 2003.