

---

## **On the mining of fuzzy association rule using multi-objective genetic algorithms**

---

### **Harihar Kalia**

Department of Computer Science and Engineering,  
Seemanta Engineering College,  
Jharpokharia, Mayurbhanj, 757086, India  
Email: smarahari07@gmail.com

### **Satchidananda Dehuri\***

Department of Systems Engineering,  
Ajou University,  
San 5, Woncheon-dong, Yeongtong-gu,  
Suwon 443-749, South Korea  
Email: satchi.lapa@gmail.com  
\*Corresponding author

### **Ashish Ghosh**

Center for Soft Computing Research,  
Indian Statistical Institute,  
203, B.T. Road, Kolkata, 700108, India  
Email: ash@isical.ac.in

### **Sung-Bae Cho**

Soft Computing Laboratory,  
Department of Computer Science,  
Yonsei University,  
50 Yonsei-ro, Sudaemoon-gu,  
Seoul 120-749, South Korea  
Email: sbcho@yonsei.ac.kr

**Abstract:** The discovery of association rule acquire an imperative role in data mining since its inception, which tries to find correlation among the attributes in a database. Classical algorithms/procedures meant for Boolean data and they suffer from sharp boundary problem in handling quantitative data. Thereby fuzzy association rule (i.e., association rule based on fuzzy sets) with fuzzy minimum support and confidence is introduced as an alternative tool. Besides, rule length, comprehensibility, and interestingness are also potentially used as quality metrics. Additionally, in fuzzy association rule mining, determining number fuzzy sets, tuning membership functions and automatic design of fuzzy sets are prominent objectives. Hence fuzzy association rule mining problem can be viewed as a multi-objective optimisation problem. On the other side, multi-objective genetic algorithms are established and efficient techniques to uncover Pareto front. Therefore, to bridge these two fields of research many

methods have been developed. In this paper, we present some of the popular state-of-art multi-objective fuzzy-genetic algorithms for mining association rules. In addition, their novelty, strengths, and weaknesses have been analysed properly with a comparative performance. The indicative future research direction and an extensive bibliography of this paper may be an attracting point for researchers from diversified domains to explore and exploit further.

**Keywords:** association rule; fuzzy set; fuzzy association rule; multi-objective optimisation; genetic algorithm; multi-objective genetic algorithm; support; confidence.

**Reference** to this paper should be made as follows: Kalia, H., Dehuri, S., Ghosh, A. and Cho, S-B. (2016) 'On the mining of fuzzy association rule using multi-objective genetic algorithms', *Int. J. Data Mining, Modelling and Management*, Vol. 8, No. 1, pp.1–31.

**Biographical notes:** Harihar Kalia is an Assistant Professor in the Department of Computer Science and Engineering, Seemanta Engineering College, Jharpokharia, Mayurbhanj, Odisha. He received his MSc in Mathematics from Ravenshaw University, Odisha in 1995, and MTech in Computer Science from Utkal University, Vani Vihar, Odisha in 2002. Currently, he is working towards his PhD in the area of multi-objective fuzzy rule mining in Department of Information and Communication Technology, Fakir Mohan University, Vyasa Vihar, Balasore, Odisha. His area of interest includes: data mining, multi-objective optimisation, fuzzy logic, evolutionary algorithms, and hybrid systems. He has about 14 years of experience in teaching both graduate and postgraduate students.

Satchidananda Dehuri is an Associate Professor in Department of Systems Engineering, Ajou University, South Korea. He received his PhD in Computer Science from Utkal University, Vani Vihar, Odisha in 2006. He received his Post Doctoral Research from Soft Computing Laboratory, Yonsei University, Seoul, Korea under the BOYSCAST Fellowship Programme of DST, Government of India. In 2010, he received Young Scientist Award in Engineering and Technology for the year 2008 from Odisha Vigyan Academy, Department of Science and Technology, received his of Odisha. He has already supervised three PhD scholars in the area of computer science.

Ashish Ghosh is a Professor with the Machine Intelligence Unit, Indian Statistical Institute. He has already published more than 150 research papers in internationally reputed journals and refereed conferences, and has edited eight books. He received the prestigious and most coveted Young Scientists Award in Engineering Sciences from the Indian National Science Academy in 1995, and in Computer Science from the Indian Science Congress Association in 1992. He was selected as an Associate of the Indian Academy of Sciences, Bangalore, India, in 1997. He is a member of the founding team that established the National Center for Soft Computing Research at the Indian Statistical Institute, Kolkata in 2004, with funding from the Department of Science and Technology, Government of India, and is currently the in-charge of the centre.

Sung-Bae Cho received his PhD in Computer Science from Korea Advanced Institute of Science and Technology (KAIST), Taejeon, Korea in 1993. He was an Invited Researcher of Human Information Processing Research Laboratories at Advanced Telecommunications Research (ATR) Institute, Kyoto, Japan from 1993 to 1995, and a Visiting Scholar at University of New South Wales, Canberra, Australia in 1998. He was also a Visiting Professor at University of British Columbia, Vancouver, Canada from 2005 to 2006. Since 1995, he has been a Professor in the Department of Computer Science, Yonsei University. His research interests include neural networks, pattern recognition, intelligent man-machine interfaces, evolutionary computation, and artificial life.

---

## 1 Introduction

Data mining (Han and Kaufman, 2011; Witten et al., 2011; Adamo, 2000) refers to the extraction of novel, interesting, and potentially useful patterns from large databases in association to preprocessing and post processing tasks of knowledge discovery in databases (KDD). The data mining technique aims at building an efficient predictive or descriptive model of a large amount of data, which explains it appropriately and able to generalise to the new data. Classification (Quinlan, 1993), clustering (Jain and Dubes, 1988), sequential pattern analysis (Koper and Nguyen, 2011), prediction (Ozekes and Camurcu, 2002), association rule mining (Agrawal et al., 1993), data visualisation (Keim, 2002), etc., are popularly used data mining (Chen et al., 1996; Freitas, 2002) techniques.

Association rule mining is an important data mining technique, which was initially proposed by Agrawal et al. (1993), and is widely known as market-basket problem analyser for studying consumer-purchasing patterns. This problem can be stated as follows: “A set of items and their sales records, which consist of the sale of items in a particular date, are given. The task is to find the association/relation among items of the transactions”. An efficient approach of categorisation to association rules is presented in Won and McLeod (2012).

In other words, the problem is to discover association rules in the form:

$$IF A Then B, \tag{1}$$

where  $A$  and  $B$  are set of items called frequent item sets.

Fundamentally association rule mining problem aims at mining association rules satisfying minimum support and confidence. Several other objectives like lift, coverage and conviction, etc. (Brin et al., 1997; Geng and Hamilton, 2006; Bagui et al., 2009) have been used to enumerate the superiority of the rule. These evaluation measures may incur trade-off of different degrees depending on the database used and extracted information type. In lieu of this fact, researcher tries to visualise association rule mining problem as a multi-objective problem (Ghosh and Nath, 2004; Dehuri et al., 2006; Atlas et al., 2008).

Most of the conventional association rule mining algorithms (HooshSadat et al., 2011; Chi, 2012; Srikant and Agrawal, 1996; Coenen et al., 2004; Sekhvat et al., 2010) are competent to handle Boolean or binary data. These algorithms mine rules from quantitative data as well, by first partitioning into interval and subsequently converting them into Boolean type. However, the partitioning procedures suffer from ‘sharp

boundary problem', which means, it either over-emphasise or under-estimate the attributes near the boundaries, which may cause inappropriate representation of semantics in the rules (Olufunke et al., 2010).

In order to vacate from the sharp boundary problem in association rule mining process, fuzzy association rules (Kouk et al., 1998; Gyenesei, 2001) are introduced, which uses fuzzy sets in form of linguistic variables. According to the assertion of Dubois et al. (2005), expressing associations between data using fuzzy sets widen the represented relationship types and make easy interpretation of rules in linguistic terms and avoids the unnatural boundaries in partitioning the attribute domains. More precisely, fuzzy association rules are intuitive, human understandable, and can handle vague and imprecise data.

A typical fuzzy association rule (Kouk et al., 1998; Muyebe et al., 2008) is represented in the form given in equation (2):

$$\text{If } P \text{ is } F \text{ then } Q \text{ is } G, \quad (2)$$

where  $F$  and  $G$  are set of fuzzy sets used in describing set of items  $P$  and  $Q$  respectively.

Many of the existing fuzzy association rule mining algorithms (Kalia et al., 2012; Mangalampalli and Pudi, 2009; Ouyang, 2012) ask for pre-defined membership functions, which is quite a difficult task in practice. Like (crisp) association rule mining algorithms, fuzzy association rule mining problem modelled as a multi-objective optimisation problem with additional objectives/criteria like; automatic discovery of membership functions, optimising number of fuzzy sets, and interpretability, in addition to the objectives of (crisp) association rule mining problem.

Solving multi-objective problem by single objective genetic algorithms causes problem (Coello, 1998; Tamaki et al., 1996), as separate objectives have unequal effective ranges (Bentley and Wakefield, 1997). If the used multi-objective ranking method (di Pierro et al., 2007) is not range-independent, then one or more objectives in the problem can dominate the others; give rise to poor solutions, which prompt to use multi-objective genetic algorithms in solving these problems.

The rest of the paper is organised as follows. In Section 2, we discuss association and fuzzy association rule mining problem. Introduction to multi-objective optimisation problem and genetic algorithm-based approach is discussed in Section 3. Fuzzy association rule mining using multi-objective genetic algorithms is the focus of Section 4. In Section 5, we discuss the performance comparison of the popular approaches. Applications and conclusions along with future note of research are given in Sections 6 and 7.

## 2 Association and fuzzy association rule mining problem

Association rule mining (Agrawal et al., 1993) originate, to find buying pattern of the customers through a process of market basket Analysis. Since then, association rule mining has been studied and applied in different areas like medical diagnosis (Rajendran and Madheswaran, 2010), stock market prediction (Argiddi and Apte, 2012), web mining (Chai and Li, 2010), network intrusion detection (Mao and Zhu, 2002), manufacturing (Wantanabe, 2010), recommender system (Xizheng, 2007), etc., and find patterns that associate different attributes (Tzacheva, 2012; Zhang et al., 2010).

An association rule mining problem can be defined as: “Given a dataset of transactions, each of which is a set of items, a minimum support value, and minimum confidence value. Find rules in the dataset, satisfying support-confidence thresholds”. Mathematically, Let  $I$  be the given set of  $m$ -items,  $I = \{i_1, i_2, \dots, i_m\}$  and database  $D$  be the set of  $n$  transactions,  $D = \{t_1, t_2, \dots, t_n\}$ . For a given transaction with  $k$  items  $t_j = \{i_{j_1}, i_{j_2}, \dots, i_{j_k}\}$  and  $t_{j_i} \in I$ , an association rule may be of the form  $A \Rightarrow B$ , or *If A THEN B*, where  $A, B \subset I$  are sets of items called itemsets and  $A \cap B = \varnothing$  (Kantardzic, 2003). Apriori (Agrawal, 1993) is a widely used algorithm, used for association rule generations.

Most of the association rule mining algorithms/approaches are designed to handle binary or Boolean data. In order to handle quantitative data, we have to partition these data into intervals and then transforming those data into Boolean type. This process of conversion suffers from sharp boundary problem, which cause the ignorance or over-emphasise of element near the boundary. With respect to human perception, use of sharp boundary is not intuitive. As an example, this method may classify a person as tall if height is greater than 160cm and short if it is less than 160cm. This does not match with the human perception of tall and short. Similarly these methods classify both height of 60cm and 155cm as short. This is also against the human perception.

In order to eliminate these problems, fuzzy sets (Kouk et al., 1998) are introduced in quantitative association rule mining. Each attributes are associated with one or more fuzzy sets. So a person may belong to tall or short with some membership degree.

Let  $I$  be the given set of  $m$ -items,  $I = \{i_1, i_2, \dots, i_m\}$  and database  $D$  be the set of  $n$  transactions,  $D = \{t_1, t_2, \dots, t_n\}$ . Each attribute  $i_k$  is associated with a set of  $p$  fuzzy sets  $F$ , i.e.,  $F = \{f_{i_k}^1, f_{i_k}^2, \dots, f_{i_k}^p\}$ . Then a fuzzy association rule is of the form: *A is F  $\Rightarrow$  B is G* or, *If A is F then B is G*, where  $A, B \subseteq I$  are itemsets and  $A = \{a_1, a_2, \dots, a_p\}$  ( $a_i \neq a_j$ , if  $i \neq j$ ) and  $B = \{b_1, b_2, \dots, b_q\}$  ( $b_i \neq b_j$ , if  $i \neq j$ ). The fuzzy sets correspond to the attributes in  $A$  and  $B$  are  $F = \{f_{a_1}, f_{a_2}, \dots, f_{a_p}\}$  and  $G = \{g_{b_1}, g_{b_2}, \dots, g_{b_q}\}$ .

The pair  $\langle A, F \rangle$ , is defined as a fuzzy itemset, in which  $A(\subseteq I)$  is an itemset and  $F$  is a set of fuzzy sets associated with attributes in  $A$ .

The support of a fuzzy item set  $\langle A, F \rangle$  denoted as  $FS\langle A, F \rangle$  is calculated using equation (3).

$$FS(\langle A, F \rangle) = \frac{\sum_{v=1}^m \{ \alpha_{a_1}(t_v[a_1]) \otimes \alpha_{a_2}(t_v[a_2]) \otimes \dots \otimes \alpha_{a_p}(t_v[a_p]) \}}{|D|}, \quad (3)$$

where

- $A = \{a_1, a_2, \dots, a_p\}$  and  $t_v$  is the  $v^{\text{th}}$  record in  $D$
- $\otimes$  is the T-norm operator of fuzzy logic, similar to AND in traditional logic
- $\alpha_{a_u}(t_v[a_u])$  is calculated using the formula:

$$\alpha_{a_u}(t_v[a_u]) = \begin{cases} m_{a_u}(t_v[a_u]) & \text{if } m_{a_u}(t_v[a_u]) \geq \theta \\ 0 & \text{if otherwise} \end{cases},$$

where  $m_{a_u}$  the membership function of fuzzy set is  $f_{a_u}$  associated with  $a_u$  and  $\theta$  is the threshold for membership value specified by users.

$|D|$ , is the total number of transactions in the database  $D$ .

A fuzzy itemset  $\langle A, F \rangle$  may be a frequent itemset if its support is greater or equal to a fuzzy minimum support ( $fminsupp$ ) supplied by the user. That is,

$$FS(\langle A, F \rangle) \geq fminsupp. \quad (4)$$

The support of a fuzzy association rule ( $A$  is  $F \Rightarrow B$  is  $G$ ) is calculated as:

$$FS(A \text{ is } F \Rightarrow B \text{ is } G) = FS(\langle A \cup B, F \cup G \rangle). \quad (5)$$

The confidence of a fuzzy association rule ( $A$  is  $F \Rightarrow B$  is  $G$ ) is calculated by equation (6):

$$FC(A \text{ is } F \Rightarrow B \text{ is } G) = \frac{FS(\langle A \cup B, F \cup G \rangle)}{FS(\langle A, F \rangle)}. \quad (6)$$

A fuzzy association rule is accepted if its support and confidence value is greater than the predefined minimum support and minimum confidence value.

Like association rule discovery, fuzzy association rules are discovered via two phases. In first phase all possible frequent itemsets are obtained from the input database. In second phase all possible confident fuzzy association rules are mined from the frequent itemsets obtained in first phase. The quality of a fuzzy association rules are judged using many other metrics like interestingness (Freitas et al., 1999), comprehensibility (Hsu et al., 1997), interpretability (Jin, 2000), number of fuzzy sets, etc.

### 3 Multi-objective optimisation problem and approaches

Finding best result under the given conditions is known as optimisation. In an optimisation problem (Boyd and Vandenberghe, 2004) with more than one objective function, locating one or more optimum solutions is referred as multi-objective optimisation.

In a single objective optimisation method, a best solution is found out, corresponding to the minimum or maximum value of a single objective function, which knobs all objectives into one. Hence it cannot provide a set of different solutions that trade different objectives against each other. However multi-objective optimisation method (Marler and Arora, 2004) simultaneously optimises two or more conflicting objectives subject to certain constraints.

According to Osyczka (1985), multi-objective optimisation problem can be defined as:

“A vector of decision variables which satisfies constraints and optimizes a vector function whose elements represent the objective functions. These functions form a mathematical description of performance criteria which are usually in conflict with each other. Hence the term “optimizes” means finding such a solution, which would give the values of all the objective functions acceptable to the decision maker”.

Mathematically this can be stated and visualised as follows:

Find the vector  $\vec{x} = \langle x_1, x_2, \dots, x_d \rangle$ , which can optimise the vector function  $\vec{f}(\vec{x}) = \langle f_1(\vec{x}), f_2(\vec{x}), \dots, f_n(\vec{x}) \rangle$  simultaneously and satisfy  $m$  inequality constraints  $\vec{g}_i(\vec{x}) \geq 0$ ,  $i = 1, 2, \dots, m$  and  $p$  equality constraints  $h_j(\vec{x}) = 0$ ,  $j = 1, 2, \dots, p$ .

The solution of a MOP, generally not unique but can be represented by a set of solutions which are referred as good trade-offs or compromises. This concept is termed as Pareto optimality, originally proposed by Edgeworth (1881) and generalised by Pareto (1896). A vector solution  $\vec{x}$  is called Pareto optimal if there exists no feasible vector  $x$  which would increase some criterion without causing a simultaneous decrease in at least one other criterion, in case of a maximisation problem. This is accepted as the Pareto optimum solution. Pareto optimum set is generally called an efficient solution set or a non-dominated solution set.

Hence, the main goal of multi-objective optimisation algorithms is to identify solutions in Pareto optimal set. In practice, it is very difficult to identify the whole Pareto optimal set for multi-objective problems because of its size. So practically in multi-objective optimisation a set of solutions are found which are very close to the Pareto optimal set. These methods give rise to a set of compromised solutions, known as the trade-off, non-dominated, non-inferior or Pareto-optimal solutions (Cohon, 1978).

The multi-criteria decision making (MCDM) techniques (Miettinen, 1998) are conventional methods employed in solving these multi-objective optimisation problems. Now-a-days multi-objective evolutionary algorithms (MOEA) are commonly adopted techniques in solving these problems. By using a set of population and evolution process, these algorithms effectively and efficiently handle these optimisation problems having composite trait like discontinuities, multimodality, disjoint feature space and noisy features. One of the advantages of MOEAs over MCDM techniques (Miettinen, 1998) is that in a single run, a number of Pareto-optimal solutions with a wide range of objective values can be simultaneously obtained, however multiple runs are needed to find same number of Pareto optimal solution when MCDM techniques are followed.

The classical multi-objective optimisation methods such as weighted sum (Cohon, 1978), constraint method (Cohon, 1978), goal programming (Steuer, 1986), min-max approach (Koski, 1984), etc., are suffers from the following difficulties:

- to find multiple optimal solutions, repeated application is sought
- some prior knowledge is needed about the problem under consideration
- shape of the Pareto-optimal front may be affected by some method
- problems with uncertainties or stochastic may not handled properly
- problems with discrete search space may not handle efficiently.

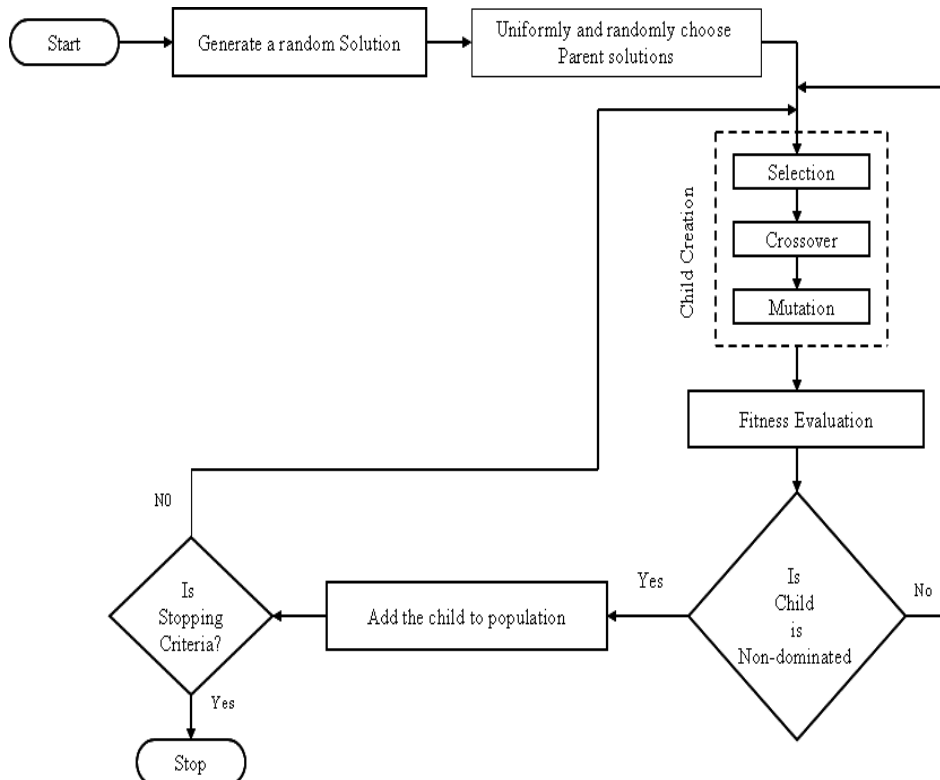
On the contrary, multi-objective genetic algorithms (Dehuri and Ghosh, 2004) have the following advantages:

- ability to handle with a set of possible solutions simultaneously
- in single run of the algorithm, many constituent of the Pareto optimal set can be obtained
- in the whole search space, solutions are searched
- shape of Pareto front is rarely disturbed.

Hence multi-objective genetic algorithm is good candidate for solving multi-objective problem and fuzzy association rule mining problems.

A few popular multi-objective genetic algorithms are: vector evaluated genetic algorithm (VEGA) (Schaffer, 1985), multiple objective genetic algorithm (MOGA) (Fonseca and Fleming, 1993), non-dominated sorting genetic algorithm (NSGA) (Srinivas and Deb, 1994), Niche-Pareto genetic algorithm (NPGA) (Horn et al., 1994), elitist non-dominated sorting genetic algorithm (NSGA-II) (Deb et al., 2002), non-dominated rank-based sorting genetic algorithm (NRSGA) (Ghosh and Das, 2008), multi-objective genetic modified algorithm (MOGMA) (Vandeva, 2012), etc. A typical multi-objective genetic algorithm may look like as in Figure 1.

**Figure 1** A simple multi-objective genetic algorithm





#### 4 Multi-objective genetic algorithms for crisp/fuzzy association rule mining

Crisp and fuzzy association rule mining problem can be viewed as a multi-objective problem rather than a single objective problem. These multi-objective optimisation models are solved using genetic algorithm (Coello, 1998), by converting the objectives into a single objective function by forming a linear combination of objectives, where setting the weight of the objectives is a vital problem and hence desired goal may not be achieved. Instead multi-objective GA can handle multiple objective functions simultaneously and generate a good approximation of the Pareto front and hence are suitable tool for solving these problems. Hence researchers deliberately exercised the existing MOGAs or attempted to propose new MOGA for crisp and fuzzy association rule mining problem (Ghosh et al., 2008).

##### 4.1 Multi-objective association rule and approaches

Association rule mining in its basic form is to mine all the association rules that are satisfied minimum support and confidence (Agrawal et al., 1993; Agrawal and Srikant, 1994). This support-confidence framework is followed by all most all popular rule mining techniques such as apriori (Agrawal, 1996), AIS (Agrawal et al., 1993), DIC (Brin et al., 1997), estMax (Woo and Lee, 2009), etc. Beyond these two, several rule evaluation criteria used in judging the interestingness or goodness of the rules (Bayardo and Agrawal, 1999; Freitas, 1999; Tan et al., 2002) are variance, gain, entropy gain, gini, lift, conviction and chi-square value, J-measure, surprise, Laplace. Another rule measure is comprehensibility, which measure the number of attributes involved in the rule and tries to compute the understandability of the rule. According to Bayardo and Agrawal (1999), the association rule extracted using any of the metrics described above is a Pareto-optimal rule with respect to support and confidence. This study motivates the researchers to incorporate the evolutionary multi-objective algorithms (Del et al., 2011; Schaffer, 1984; Coello, 1999) in rule mining.

As pointed out by Ghosh and Nath (2004) that some of basic difficulties faced by the popular association rule mining algorithm based on Agrawal et al.'s approach work for binary databases are:

- 1 In case of a database having huge number of attributes, where each attribute has many distinct values, the number of items (number of field) will be high. Storing these huge databases is very difficult by existing algorithms in order to use for rule mining algorithms.
- 2 Frequent item-set generation is one of two phases used by the existing algorithm, where frequent item-sets are generated from the all possible item-sets, measuring support count of the item-set and using minimum support, a user-defined threshold. Less number of frequent item-sets will be generated if minimum support is too big, on the other hand most of the item sets will become frequent if minimum support is too small. Hence better rule selection will be effected.

- 3 These algorithms use separate symbols for each possible value of an attribute in order to encode. Although these methods are well suited for categorical attributes but are not suitable for numerical attributes. As a remedy some range of values may be defined and for each range of values an item is defined. But this method is not fit in all situations, because range of different attributes are different, which create problem in defining range.
- 4 In generating rule by these algorithms, the ordering of the items are always maintained, hence the rule without sequence are not mined, although they appear in the database.

In view of the above limitations, Ghosh and Nath (2004) modelled the association rule mining problem as a multi-objective problem considering the measures like support count, comprehensibility, and interestingness and then solve it by using genetic algorithms.

Dehuri et al. (2006) presents a fast and scalable multi-objective association rule mining technique using genetic algorithm from large database using the objective functions such as confidence factor (Imielinski et al., 1993), comprehensibility (Hsu et al., 1997) and interestingness (Freitas et al., 1999). As the multi-objective genetic algorithm (MOGA) tends to be slow in comparison with most classical rule mining methods, in order to overcome these difficulties they propose a fast and scalable MOGA using the inherent parallel processing nature of genetic algorithms in a homogeneous dedicated network of workstations.

Hu and Yang-Li (2007) have modelled the association rule mining problem as a multi-objective optimisation problem by considering three objectives; namely statistical correlation, comprehensibility, and confidence. Authors solves the said problem in an intelligent data mining way by implementing multi-objective co-evolutionary algorithm which does not necessitate the user supplied threshold. The proposed process reduces majority of the weak and negative rules and improve the comprehensibility of the generated rules.

Martin et al. (2011) have designed an approach by extending the NSGA-II (Deb et al., 2002), a popular multi-objective genetic algorithm for mining a set of quantitative association rules by means of a good trade-off between interpretability and accuracy. For this, an evolutionary learning of the intervals of the attributes and maximisation of three objectives; the interestingness, comprehensibility, and performance (product between support and confidence) are performed.

Wakabi-Waiswa and Baryamureeba (2011) have proposed a multi-objective genetic algorithm, to generate association rules satisfying the five quality rule measures, such as: confidence, support, interestingness, lift, and J-measure referred as multi-objective genetic algorithm for mining association rules (MOGAMAR).

Hadian et al. (2010) have proposed a clustering-based multi-objective genetic algorithm (CBMOGA) for generating association rules, considering four objectives; support, confidence, comprehensibility, and interestingness. The proposed algorithm optimises the support counting phase by clustering the database.

Qodmanan et al. (2011) have proposed an approach to find multi-objective association rule with genetic algorithm without taking the minimum support and confidence into account. Authors use two measures out of four measures support, confidence, comprehensibility and interestingness as objectives in modelling the problem.

Besides genetic algorithm, ant colony algorithm (Blum, 2005), differential evolution (Das and Suganthan, 2011), simulated annealing (Aarts and Korst, 1989) and some hybrid approaches are also employed in multi-objective association rule mining.

Atlas et al. (2008) have designed a Pareto-based multi-objective differential evolution algorithm for mining accurate and comprehensible numeric association rules by formulating the association rules mining problem as a four-objective optimisation problem using support, confidence value, and the comprehensibility of the rule as maximisation objectives and the amplitude of the intervals which conforms the itemsets and rule as minimisation objective. Nasiri et al. (2011) have modelled association rule mining problem as a multi-objective problem using three objectives such as: confidence, support, and interestingness and solve it using simulating annealing algorithm. Moslehi et al. (2011) have designed a multi-objective numeric association rules mining approach using ant colony optimisation, which mines rules form continuous domains without specifying minimum support and minimum confidence considering support, confidence, and interestingness as objectives.

#### *4.2 Genetic-based multi-objective fuzzy association rule*

Fuzzy association rule-based systems consists of a set of logical fuzzy rules. The task of automatic design of fuzzy association rule is handled as an optimisation task or a search problem. The capability to deal with a huge search spaces, ability in determining a very close optimal solution and easy adoption of a priori knowledge, genetic algorithms are preferably engaged in performing this task.

As different objectives like confidence support, interpretability, and minimisation of number of rules are handled in fuzzy association rule mining and there is trade-off among these objectives, hence no single solution serve the purpose, so fuzzy association rule mining problem is modelled as a multi-objective optimisation problem. Multi-objective genetic algorithms are well suited for these problems, as a single run of MOGA, produce a set of non-dominated solution; however the single objective GA give a single solution.

Basically in association rule mining, we tries to mined association rules with a pre-specified minimum support and confidence values called strongness of the rules. In case of fuzzy association rules, strongness of the rule includes finding of number of appropriate fuzzy sets and their membership functions. The rule's support corresponding to a membership function decreases, if the interval of that membership function is reduces in size. The other objectives of fuzzy association rule mining are comprehensibility, interestingness, minimise number of fuzzy sets, maximise the number of large itemsets, automatic generation of membership functions, and minimise the number of fuzzy rules, etc. The interestingness of a rule is evaluated by the metrics like gain, variance, chi-squared value, entropy, gain, gini, Laplace, lift, and conviction. Bayardo and Agrawal (1999) shown that the best rule according to any of the listed metrics is a Pareto-optimal rule with respect to support and confidence.

Furthermore, in fuzzy association rule mining, there is a trade-off between number of rules and time complexity and also between number of rules and suitability of membership functions. Hence fuzzy association rule mining can be visualised as a multi-objective problem.

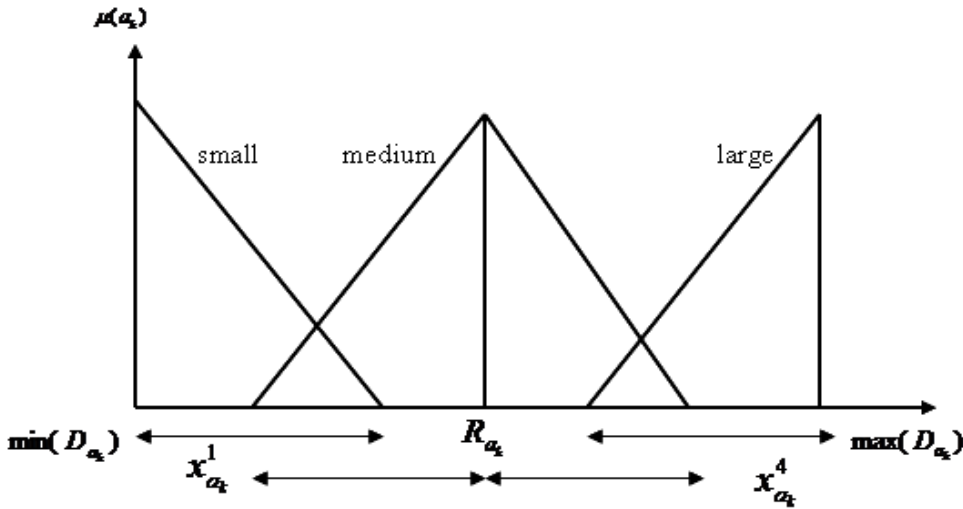
In literature, fuzzy association rule mining problems are modelled as multi-objective optimisation problem considering two or more objectives out of the candidate objectives; support, confidence, maximising number of large item sets, minimising the time required to determine fuzzy sets, interestingness, strongness, comprehensibility, number of fuzzy sets, etc.

Mehmet Kaya and his group furnish different approaches for multi-objective objective fuzzy association rule mining algorithms using different objectives. We discuss some of their approaches.

Kaya and Alhaji (2004c) designed a clustering method, which uses multi-objective genetic algorithm in deciding number of fuzzy sets and discover both membership function and fuzzy association rules automatically. This process automatically adjusts the fuzzy sets and finds many large item sets in small time, by tuning jointly the number of fuzzy sets and the base values of membership functions for each quantitative attributes. The number of large item sets and the time required for determining fuzzy sets (inverse of time required to find large item sets) are regarded as two objectives in modelling the problem as multi-objective. The Pareto GA (Zitzler and Thiele, 1999), a multi-objective algorithm is used for the purpose.

The base values of the membership functions for each quantitative attribute is represented as an individual. Triangular membership functions are used for the purpose. Figure 2 shows a quantitative attribute  $a_k$  with three fuzzy sets, the corresponding membership functions and their base variables.

**Figure 2** Membership functions with their base value of attribute  $a_k$



The base variables  $x_{a_k}^1, x_{a_k}^2, x_{a_k}^3, x_{a_k}^4$  and the intersection point  $R_{a_k}$ , of the attribute  $a_k$  have finite value lies between the intervals  $[\min(D_{a_k}), \max(D_{a_k})]$ ,  $[\min(D_{a_k}), R_{a_k}]$ ,  $[R_{a_k}, \max(D_{a_k})]$  and  $[\min(D_{a_k}), \max(D_{a_k})]$ . If  $n$  quantitative attributes are used, then each attribute can have at most  $(n-1)$  fuzzy sets. Following chromosome representation used to represent eight quantitative attributes (seven fuzzy sets), consists of base lengths and the intersecting points of the used fuzzy sets.

$$y_{a_1} x_{a_1}^1 x_{a_1}^{12} R_{a_1}^1 x_{a_1}^2 x_{a_1}^3 R_{a_1}^2 x_{a_1}^4 x_{a_1}^5 \dots R_{a_1}^5 x_{a_1}^{10} x_{a_1}^{11}$$

$$\dots y_{a_8} x_{a_8}^1 x_{a_8}^{12} \dots R_{a_8}^5 x_{a_8}^{10} x_{a_8}^{11},$$

where gene  $y_{a_j}$  represent number fuzzy sets used for the attribute  $a_j$ . In decoding, if the number of used fuzzy sets is 2, then only first two variables are consider and discarding others. If three fuzzy sets are used, then subsequent three variables are also considered. Hence the use of base variables will increase when the number of fuzzy sets increases.

Author adopted real-valued coding scheme, where chromosomes represented as floating point numbers and their genes as real parameters. The value of each gene is calculated using the following formula:

$$x_{a_k}^j = \min(x_{a_k}^j) + \frac{z}{z_{\max}} (\max(x_{a_k}^j) - \min(x_{a_k}^j)),$$

where  $z$  is the maximum value of the gene in search,  $z_{\max}$  is the value that gene  $z$  can take,  $\max(x_{a_k}^j)$  and  $\min(x_{a_k}^j)$  are maximum and minimum values of the reflected area respectively. Pareto-based ranking procedure is incorporated as parent selection policy. Elitism is also adopted. Finally crossover and mutation are applied to the individuals.

In fuzzy association rule generation, the fuzzy support of the itemset  $P$  with its corresponding fuzzy set  $F$ , denoted as  $FS_{\langle P, F \rangle}$  and defined as:

$$FS_{\langle P, F \rangle} = \frac{\sum_{t_j \in D} \prod_{p_k \in P} \mu_{p_k}(f_k \in F, t_j [p_k])}{|D|},$$

where  $|D|$  is the total number of transactions in the database  $D$ .

Every large itemset  $P$  is utilised in the derivation of association rules:  $(P - Q) \Rightarrow Q$ , for each  $Q \subset P$ . The rules with confidence over the user specified minimum confidence value infer as a strong rule. The interesting rules, from these strong rules are obtained using some objective measures as well as subjective measures. The noticeable advantages of this approach are:

- 1 automatic determination clusters for each attribute
- 2 for a given minimum support, membership functions of the attributes are optimised, which prompt to get more appropriate solutions in preferred track by varying the minimum support value\
- 3 more number of large item sets and interesting fuzzy association rules are extracted.

However, the minimum support value and confidence value are supplied by the user, which needs to be automatic.

Kaya and Alhaji (2004a, 2004b) designed a multi-objective genetic algorithm-based method for finding optimised fuzzy association rules including instantiated and uninstantiated attributes, considering support, confidence and number of fuzzy sets as objectives. According to the authors claim that fuzzy association rules may have an arbitrary number of uninstantiated attributes. Two types of encoding schemes are used in rule mining. First scheme handles the rules with instantiated attributes, where each

individual represents the base values of membership functions of a quantitative attribute in database. This scheme is same as the scheme followed by Kaya and Alhaji (2004c) explained above. The second scheme, meant for handling uninstantiated rule, where two extra bits are associated with each attribute as an indicator for member in antecedent or consequent of the rule. The attribute, belongs to the antecedent if two bits are 00, belongs to consequent if these bits are 11 and not belongs to any part if these bits are 01 or 10. This coding scheme necessitate  $2m$  extra bits in each chromosome, where  $m$  is the number of attributes in the database.

The value of a gene/variable is calculated using the formula:

$$x_{a_k}^j = \min(x_{a_k}^j) + \frac{d}{2^L - 1} (\max(x_{a_k}^j) - \min(x_{a_k}^j)),$$

where  $d$  denotes the decimal value of the variable in search,  $L$  denotes the number of bits used to represent a variable in the encoding scheme,  $\min(x_{a_k}^j)$  and  $\max(x_{a_k}^j)$  respectively are minimum and maximum values of the reflected area.

The rank-based fitness assignment scheme is used in the approach. According to this scheme if any chromosomes  $y$  dominate an individual, then its rank is  $y+1$ . All individuals are ranked by repeating this process. The individuals with smallest rank represent the highest fitness. The elitism policy adapted as the selection scheme. Then a new population is generated applying the genetic operators such as replacement, crossover and mutation. Figure 3 illustrate this approach.

Kaya and Alhaji (2004a) proposed an approach, based on multi-objective genetic algorithm for mining optimised fuzzy association rule, considering support, confidence and amplitude of fuzzy sets as three objectives. In a in a new fashion, they employ the multi-objective genetic algorithm for acquiring a pre specified number of fuzzy sets in fuzzy association rule mining, which produce optimised support and confidence satisfied rules. The fuzzy set's number can vary according to user choice, between 2 to 5. The third objective, amplitude of fuzzy sets, used in the method is defined as follows:

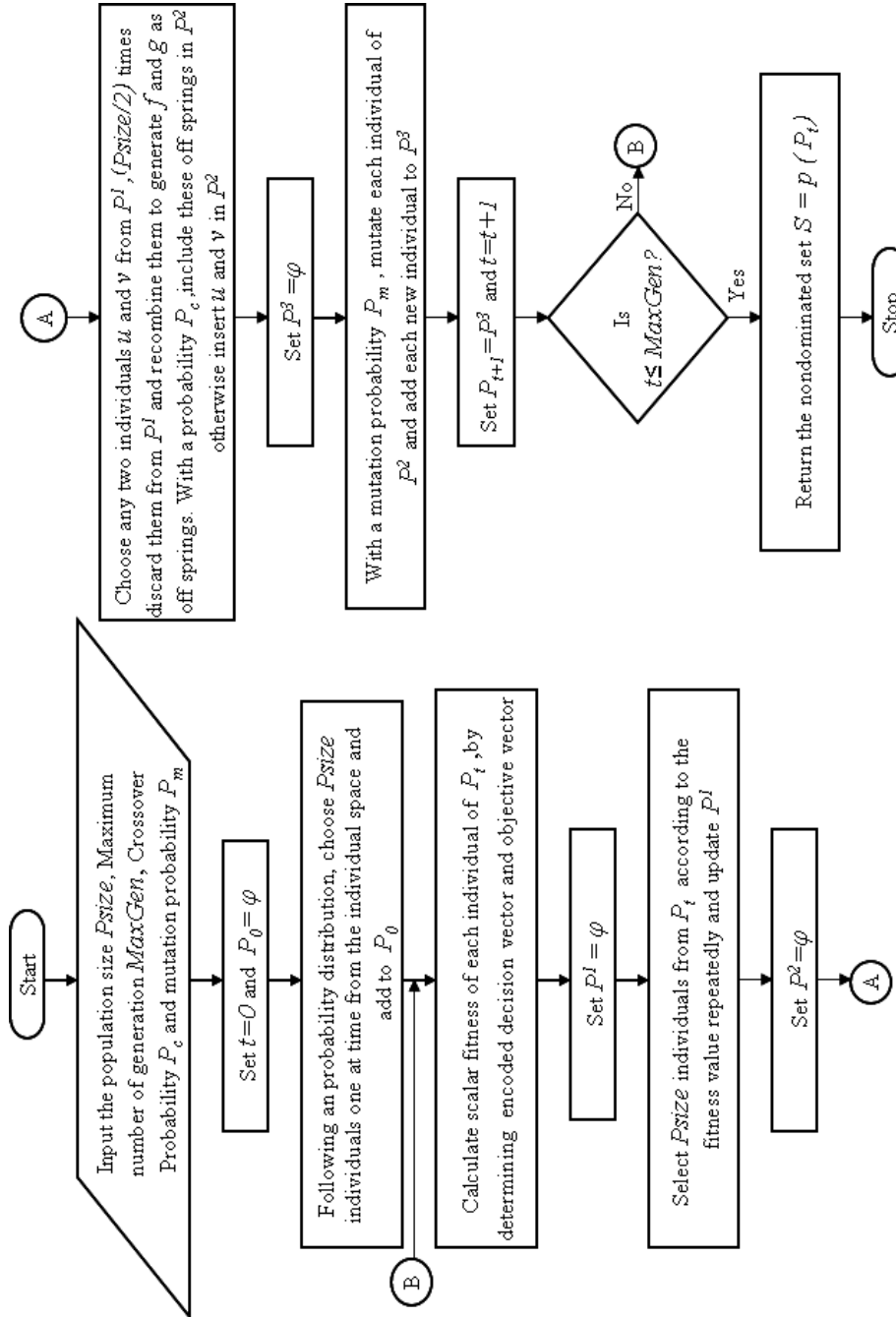
$$\text{Fuzzy set amplitude} = \frac{\text{Sum of maximum amplitudes of itemsets} - \text{sum of amplitudes of itemsets}}{\text{Sum of maximum amplitudes of itemsets}},$$

where

$$\begin{aligned} & \text{Sum of maximum amplitudes of itemsets} \\ &= \sum_{i=1}^k \max(D_i) - \min(D_i), \\ & \text{Sum of amplitudes of itemset} = \sum_{i=1}^k y_i - x_i \end{aligned}$$

$k$  represent the number of attributes in item sets and  $x_i, y_i$  are variables represents the parameter of the fuzzy sets corresponding to attribute  $i$ .

Figure 3 Kaya's MOGA for optimised fuzzy rule mining

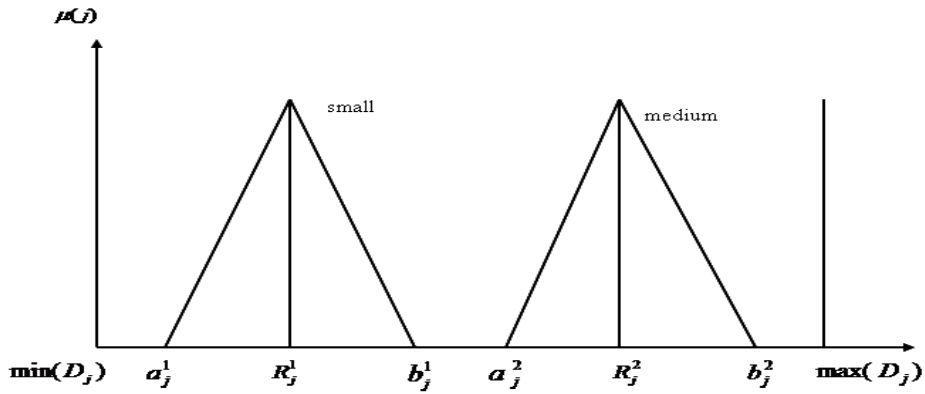


Their method worked in the following manner; every individual in the initial population represents the base values of the membership functions used for a quantitative attribute in the database. Triangular membership functions are used for the purpose. Figure 4, shows the two triangular fuzzy sets used to fuzzify the attribute  $i_k$  with their base values. Basing on the two fuzzy sets, a chromosome is represented using base lengths and intersection points as:

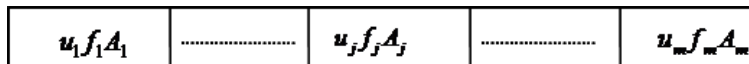
$$x_{i_1}^1 R_{i_1}^1 y_{i_1}^1 x_{i_2}^2 R_{i_2}^2 y_{i_2}^2 \dots x_{i_m}^1 R_{i_m}^1 y_{i_m}^1 x_{i_m}^2 R_{i_m}^2 y_{i_m}^2$$

In practice, for good fuzzy association rule mining, it is not possible to know the how many fuzzy sets are needed apriori, so genetic algorithm automatically adjust this number in mining process. Hence an individual is represented as given in Figure 5, where  $f_j$  denotes the number of fuzzy sets for attribute  $j$ ,  $u_j$  show the part in which attribute  $A_j$  appear by using two bits: 00 represent antecedent part and 11 represent the consequent part and 10 and 01 represent the absence of the attribute.

**Figure 4** Membership functions of the attribute  $j$  and the corresponding parameters



**Figure 5** Representation of an individual



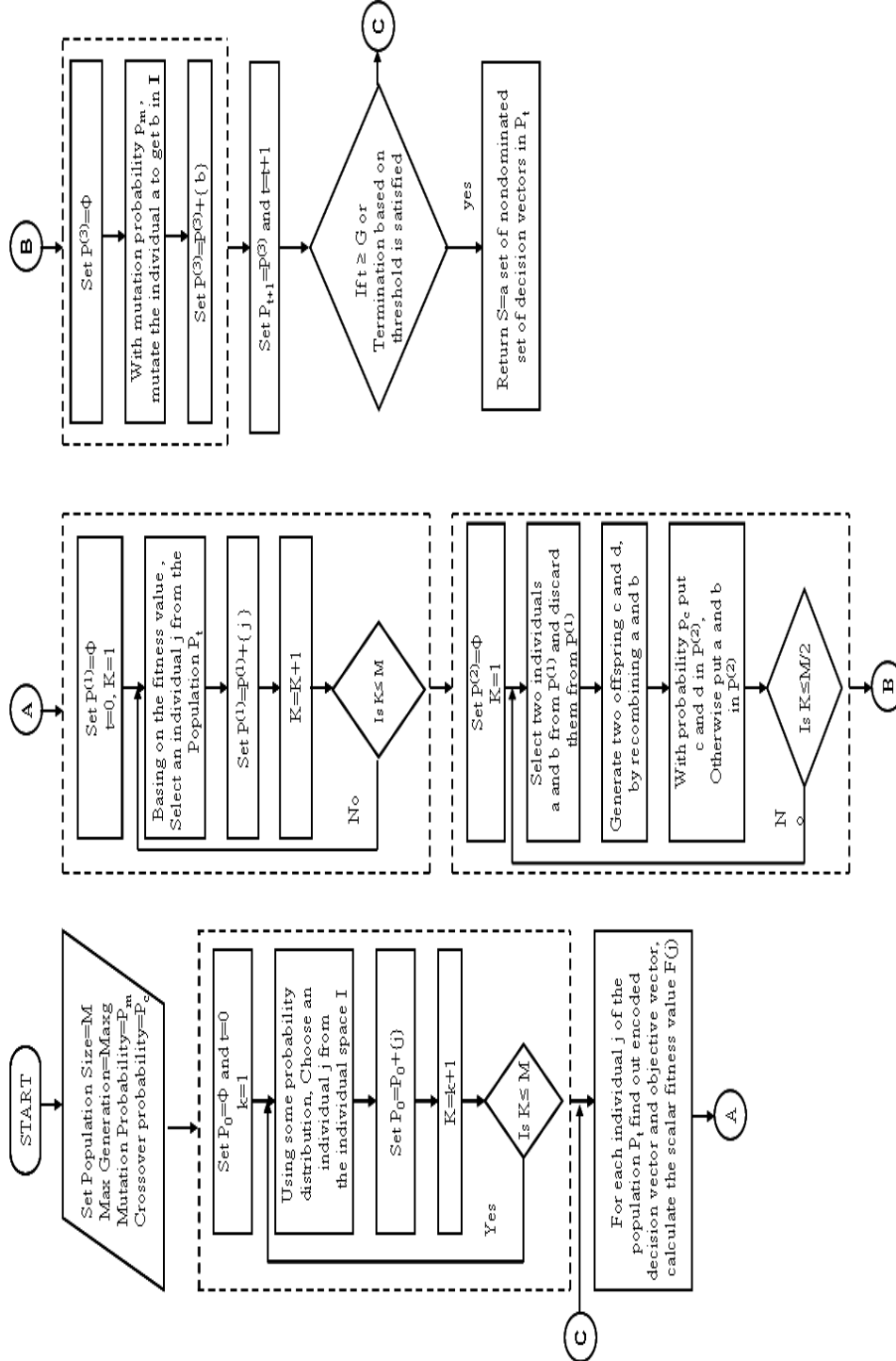
The value of a gene which is reflected under its search interval is calculated using equation (7).

$$b_{i_j}^k = \min(b_{i_j}^k) + \frac{d}{2^L - 1} (\max(b_{i_j}^k) - \min(b_{i_j}^k)), \tag{7}$$

where  $L$  denotes the numbers of bits used to represent a variable in encoding scheme,  $d$  represent the decimal value of the variable in search and in the area reflected the minimum and maximum values are respectively denoted by  $\min(b_{i_j}^k)$  and  $\max(b_{i_j}^k)$ .



Figure 6 Flow diagram of Kaya's approach



The authors directly use the concept of Pareto dominance for fitness assignment. In this scheme, using the ranks, calculated from the non-dominance property of chromosomes, the fitness value is figured. Non-dominated solutions are obtained in the raking step. Finally new populations are formed using selection, replacement, crossover and mutation operators, as in standard genetic algorithm. The flow diagram of this method is described in Figure 6.

Kaya (2006) has proposed a multi-objective genetic algorithm-based approach for mining optimised fuzzy association rules using three criteria namely strongness, interestingness and comprehensibility. Author makes an effort to derive optimised fuzzy association rules with each of the three objectives, by modelling as a separate multi-objective problem and solved using a multi-objective GA. Basing on the given criteria, the optimisation technique may be in two different forms; suitable fuzzy sets of quantitative attributes in a prespecified rule is attempt to find in first form, also referred as certain rule. Uncertain rules as well as their appropriate fuzzy sets are tries to find in second form.

The strongness of an optimised fuzzy association rule can be determined with three objectives; support, confidence and maximum number of fuzzy sets used in the rule. Hence strong optimised fuzzy association rule mining problem is modelled as a multi-objective optimisation problem with these three objectives. The interestingness of an optimised fuzzy association rule can be evaluated using support, average number of fuzzy sets in the rule or maximum number of fuzzy sets and correlation. Hence interesting optimised fuzzy association rule mining problem can be modelled as a multi-objective optimisation problem with these three objectives. The comprehensiveness of a fuzzy association rule is evaluated using support, confidence, and comprehensibility. Therefore, using these objectives a comprehensive multi-objective fuzzy association rule mining problem can be modelled.

Author employed a Pareto-based genetic algorithm in solving all three models. There are two different encoding schemes; one for certain rules and another for uncertain rules are proposed. Suitable fuzzy sets for a certain rule are attempted to optimise by the first encoding scheme. Triangular fuzzy sets are utilised in the approach. The used encoding scheme is similar to the scheme illustrated above in Kaya and Alhaji (2004c). Two extra bits are associated with each attribute in case of uncertain rule. The attribute appear in the antecedent part if these two bits are 00 and appear in the consequent part if these two bits are 11. In all other cases the attribute declared as absent.

Thilagam and Anathanarayanna (2008) have designed an approach for extracting and optimising fuzzy association rules from computer activity dataset and network audit data, using multi-objective genetic algorithm considering fuzzy support, fuzzy confidence and rule length as three objectives. Their method consists of a five phases as described below:

- *Selection of dataset:* A database is selected and if necessary, cleans it by converting data into required form to make it suitable for rule mining.
- *Preprocessing:* This phase select the appropriate attributes, define fuzzy sets, and corresponding membership functions. The fuzzy k-means clustering algorithm (Ng and Han, 1994) is used for generating membership functions.

- *Fuzzy association rule generation*: First order fuzzy association rules are generated and their confidence and support are calculated. The adjusted difference method is used in deciding the interesting association between the linguistic terms.
- *Fuzzy association rule optimisation*: Based on the defined fitness function, genetic algorithms generate fuzzy association rules and optimise during its evolution.
- *Output*: The set of fuzzy association rules of different order are produced.

In practice, the proposed method generate fuzzy association rule as follows: Following adjusted difference method (Wai-Ho and Keith, 1999), all first order rules, rules with one linguistic term in the antecedent are generated. From which second order rules, rules with two linguistic terms in antecedent are generated and similarly higher order rules are generated.

Initial population is generated randomly. Then fitness of the chromosome is calculated in term of fuzzy support and fuzzy support. The chromosome length should be ‘m’ in generating any order rule.

$$L_{x_1q_1} \quad L_{x_2q_2} \quad L_{x_3q_3} \quad L_{x_1q_1} \quad L_{x_2q_2} \quad L_{x_3q_3} \quad \dots \quad L_{x_1q_1} \quad L_{x_2q_2} \quad L_{x_3q_3}$$

#### 4.2.1 Structure of an individual

The chromosome contains  $m$  number of rule, where each rule consists of three linguistic terms. A linguistic term  $L_{x_i p_i}$  represents a set of continuous attributes and the associated terms of the attribute. The fuzzy support of a linguistic term  $L_{x_i p_i}$  is calculated by equation (8).

$$f \text{ support}(L_{x_i p_i}) = \frac{\sum_{p \in D} \lambda_{L_{x_i p_i}}(p)}{\sum_{p \in D} \sum_{j=1}^k \lambda_{L_{x_i p_j}}(p)}, \quad (8)$$

where  $\lambda_{L_{ip}}(p)$  is the degree to which  $p$  is characterised by  $L_{ip}$  is calculated using the formula defined in Qodmanan et al. (2011).

The support of an association rule  $L_{x_1 p_1} \Rightarrow L_{x_2 p_2}$  is calculated by equation (9).

$$f \text{ support}(L_{x_1 p_1} \Rightarrow L_{x_2 p_2}) = \frac{\sum_{p \in D} \min(\lambda_{L_{x_1 p_1}}(p), \lambda_{L_{x_2 p_2}}(p))}{\sum_{p \in D} \sum_{j=1}^m \sum_{k=1}^n \min(\lambda_{L_{x_1 p_j}}(p), \lambda_{L_{x_2 p_k}}(p))} \quad (9)$$

The fuzzy confidence of the association rule  $L_{x_1 p_1} \Rightarrow L_{x_2 p_2}$  is calculated by equation (10).

$$fconfidence(L_{x_1 p_1} \Rightarrow L_{x_2 p_2}) = \frac{f support(L_{x_1 p_1} \Rightarrow L_{x_2 p_2})}{f support(L_{x_1 p_1})}. \quad (10)$$

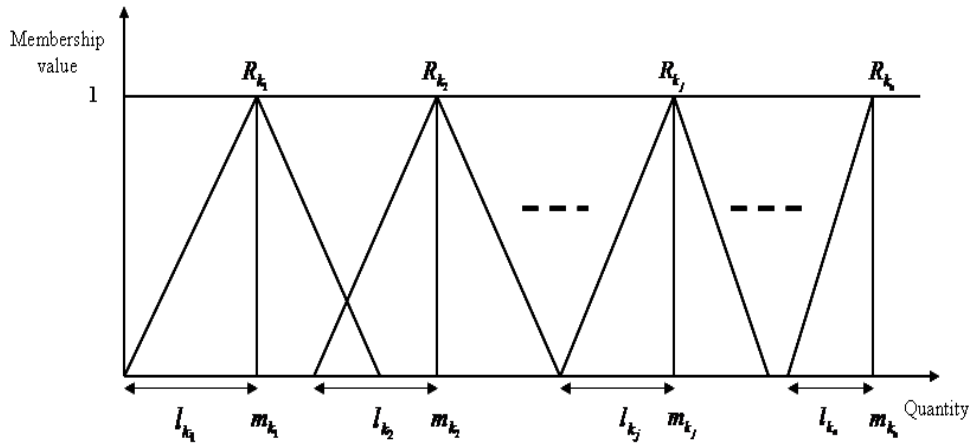
The calculated fuzzy support and confidence value of all the rules in a chromosome are used in fitness calculation. For this, rank of all the chromosomes are calculated, which defined as the number of chromosomes, that are which are dominates the given chromosome in both support and confidence. Fitness of a chromosome  $C_s$  is calculated as:

$$fitness(C_s) = population \ size - \sum_{k=1}^{r_{C_s}-1} f(k) - 0.5(f(r_{C_s}) - 1), \quad (11)$$

where  $f(k)$  denoted the number chromosomes with rank  $k$  and  $r_{C_s}$  denotes the rank of the given chromosome  $C_s$ . The diversity of the solution is maintained using niching method (Fonseca and Fleming, 1993). Roulette wheel selection method, two point crossover and mutation operator used as genetic operators in the approach.

Chen et al. (2012) have proposed a MOGA-based, multi-objective genetic-fuzzy algorithm for mining membership functions and association rules from quantitative data. The suitability of membership functions and total number of large 1-itemsets from a given set of minimum support values, are two objectives used by the algorithm in finding Pareto front. Authors encode each chromosome as a fixed length real-number string as illustrated below. The set of membership functions for an item  $a_k$  is encoded as the concatenation of all  $(m, l)$ 's pairs, where individual pair  $(m, l)$  represents a membership function with centre abscissa  $m$  and  $l$  represent half span as shown in Figure 7.

**Figure 7** Membership function for an item  $a_k$



As shown in Figure 7, a triangular membership functions with triangle as isosceles type are used for the purpose. The membership function of the  $j$ -th linguistic term of item  $a_k$  is represented as  $R_{k_j}$ . Hence the chromosome for the item  $a_k$ , may look like:

$$l_{k_1} m_{k_1} l_{k_2} m_{k_2} \dots \dots \dots l_{k_m} m_{k_n}$$

For forming feasible membership functions, the initial set of chromosomes is randomly generated having some conditions.

Two objective functions used in the approach are defined as follows: The objective function for the first objective, ‘suitability of membership function’ is defined as:

$$f_1 = suitability(C_s), \tag{12}$$

where  $suitability(C_s)$  signify the shape suitability of the membership function with  $C_s$  and is calculated by equation (13).

$$\sum_{k=1}^p \left[ overlap\_factor(C_{s_k}) + coverage\_factor(C_{s_k}) \right], \tag{13}$$

where  $p$  represents the number of items.

The  $overlap\_factor(C_{s_k})$  denote the overlap factor of membership functions for an item  $a_k$  in the chromosome  $C_s$  and is calculated using equation (14).

$$overlap\_factor(C_{s_k}) = \sum_{u \neq v} \left[ \max \left( \frac{overlap(R_{k_u}, R_{k_v})}{\min(m_{k_u}, m_{k_v})}, 1 \right), -1 \right], \tag{14}$$

where  $overlap(R_{k_u}, R_{k_v})$  is the overlap length of  $R_{k_u}$  and  $R_{k_v}$ .

The  $coverage\_factor(C_{s_k})$ , corresponds to the coverage ratio of a set of membership functions for an item  $a_k$  in the chromosome  $C_s$  and is defined as:

$$coverage\_factor(C_{s_k}) = \frac{1}{\frac{range(R_{k_1}, \dots, R_{k_n})}{\max(a_k)}}, \tag{15}$$

where  $range(R_{k_1}, \dots, R_{k_n})$  is the coverage range of the membership functions,  $n$  denote number of membership functions for  $a_k$  and  $\max(a_k)$  is the maximum quantity of  $a_k$  in the transactions.

The second objective function, number of large 1-itemsets is denoted as  $f_2$  and is defined as:

$$f_2(C_s) = \text{totalNumL1}(C_s) = \sum_{h=1}^r |L1_s^{ms_h}|, \quad (16)$$

where  $|L1_s^{ms_h}|$  is the number of large 1-itemsets with minimum support value  $ms_h$  from a given set of minimum support values  $\{ms_1, ms_2, \dots, ms_r\}$ .

Author follows the fitness assignment scheme similar to scheme in MOGA (Fonseca and Fleming, 1993), which is characterise by three basic steps, viz. ranking chromosomes, assigning fitness and averaging of fitness values of the same rank individuals. The chromosomes are ranked according to their values based on two objective functions. Then fitness value of a chromosome is assigned based on its rank value.

For any chromosome  $C_s$  with ranking value 1, the fitness value is calculated by the formula:

$$\text{fitness}(C_q) = \frac{\text{Dominatedby}(C_s)}{P+1}, \quad (17)$$

where  $P$  is the population size and  $\text{dominatedby}(C_s)$  is the number of chromosomes dominated by  $C_s$ . For any chromosome  $C_s$  with ranking value greater than 1, its fitness value is calculated by the formula:

$$\text{fitness}(C_s) = 1 + \sum_{C_s \in P \text{ and } C_s \text{ dominated by } C_q} f(C_s), \quad (18)$$

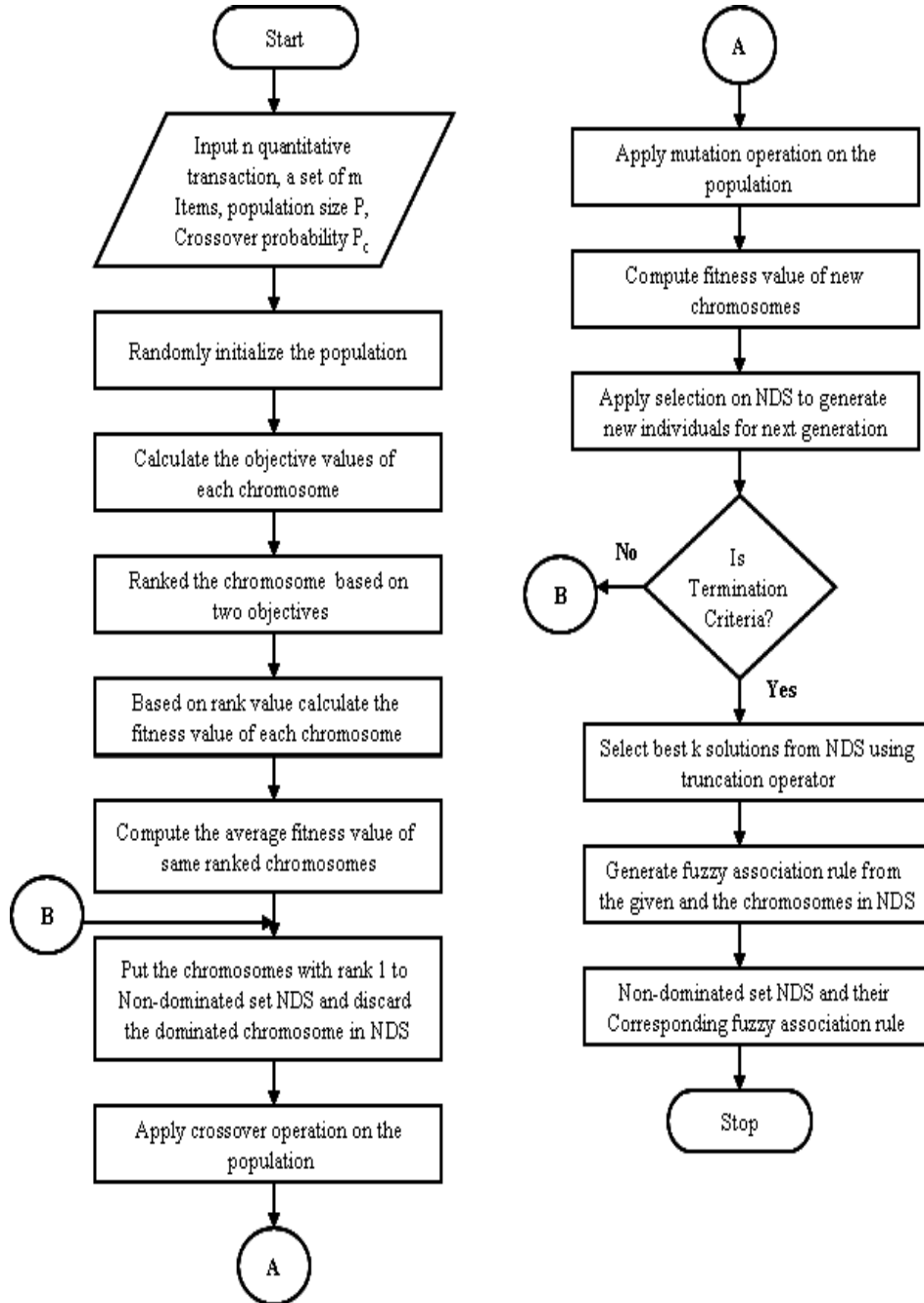
where  $\text{fitness}(C_s)$  is the fitness value of the chromosome  $C_s$  which dominate chromosome  $C_q$  and to ensure that the fitness value of a dominated chromosome is larger than a non-dominated chromosome, the value '1' is used. The max-min-arithmetical crossover and one-point mutation are used as the genetic operators in this approach. The flow diagram of the algorithm is illustrated in Figure 8.

The proposed approach efficiently finds Pareto-front solutions and resolves the limitations of the author's previous paper (Hong et al., 2012). However, the approach seeks a set of user defined support value as well as the confidence threshold.

## 5 Performance comparison

The performances of the algorithms studied in Section 4 have been examined in Table 1. In column 2, the dataset and the membership function used to validate the concerned approach is given. Our study reveals that triangular membership function is still a dominating one in multi-objective fuzzy association rule mining problem. Kaya and Alhaji (2004a) and Chen et al. (2012) have used their own developed MOGA for FARM but Thilagam and Anathanarayanna (2008) have used NSGA-II. Similarly, Kaya and Alhaji (2004c) used the method proposed in Zitzler and Thiele (1999) to generate fuzzy association rule from the adult dataset of US census in 2000.

Figure 8 Flow diagram of Chen’s approach



**Table 1** Algorithms and their performance for discovering multi-objective fuzzy association rule using genetic algorithm

<i>Authors</i>	<i>Membership functions</i>	<i>Dataset</i>	<i>Objectives</i>	<i>Genetic algorithm used</i>	<i>Result/conclusion</i>
Kaya and Alhaji (2004a)	Triangular	Letter recognition data	Three objectives: 1 Support 2 Confidence 3 Amplitude	Own MOGA	1 When number of attributes in the rule increases, the support value of the rule decreases exponentially. 2 The run time of the algorithm increases linearly with the increase of fuzzy sets.
Chen et al. (2012)	Triangular	Own simulated dataset and food mart database	Two objectives: 1 Maximising number of large one item sets 2 Minimising suitability	Own defined MOGA	The number of rules derived with highest total number of large 1-itemsets is larger than the number of rules derived with best suitability value.
Thilagam and Anathanarayanna (2008)	Trapezoidal	1 Computer activity dataset 2 Network audit dataset	Three objectives: 1 Maximising support 2 Maximising confidence 3 Minimising rule length	NSGA-II	When order of rules increases, support decreases. So domain expert should choose the requisite rule based on support and confidence values.
Kaya and Alhaji (2004a, 2004b)	Triangular	100 K transactions from the adult data of US census in 2000	1 Number of large item sets 2 The time required for determining fuzzy sets	Pareto GA (Zitzler and Thiele, 1999)	Maximum number of large itemsets and interesting association rules are obtained in this method



## 6 Applications

- *Electrical load balancing*: Li and Jia-ju (2005) have proposed a fuzzy rule-based classifier for electrical load pattern classification by using multi-objective genetic algorithm and fuzzy association rule mining. Multi-objective genetic algorithm is used to automatically select the rules with better classification accuracy and interpretability, and the key concepts of fuzzy association rule mining are the bases of heuristic rule selection for improving the performance of genetic algorithm searching. Through computational experiments on a real power system, it is shown that the generated fuzzy rule-based classifier leads to high classification performance, and can supply more sufficient historical data for load forecasting of anomalous days, better performance of load forecasting is gained accordingly.
- *Marketing*: Albert et al. (2009) have applied multi-objective fuzzy association rules-based genetic algorithm to marketing-oriented firms, which are specially concerned with modelling consumer behaviour to improve their information and aid their decision processes on markets.
- *Network intrusion*: Mahajan and Reshamwala (2011) have applied a fuzzy association-based multi-objective genetic approach in discovering optimised sequences in the network traffic data to classify and detect intrusion. Thilagam and Ananthanarayan (2008) applied multi-objective fuzzy association rule mining approach based on genetic algorithm for intrusion detection.
- *Medical diagnosis*: Carmona et al. (2011) have proposed the MOEA MESDIF for the extraction of fuzzy association rules to harvest interesting information regarding the rate of admission to the psychiatric emergency department.
- *Stock mining*: Ghazi and Abadeh (2012) have proposed an evolutionary approach for finding fuzzy association rule with 2-tuple linguistic representation model based on multi objective genetic algorithm for identifying fuzzy association rules without specifying minimum support and minimum confidence and applied in stock data mining.

## 7 Conclusions

Mining fuzzy association rule, using multi-objective genetic-fuzzy algorithms/procedures produces interesting, comprehensible, and interpretable rules. Some of the problems like: defining membership functions, numbers of fuzzy sets, tuning membership functions, etc., are resolved by these approaches. However, some of the methods need user-specified minimum support and confidence value. Hence, multi-objective fuzzy-genetic associations rule mining methods need to be improved, so that thresholds for support and confidence can be automatically fixed. Further, KDD is undoubtedly recognised as a key ‘technology’ in business, industry, medical, etc. However, on mining fuzzy association rules by considering multiple objectives from these domains are very scarce. Along side other meta-heuristic techniques like: ant colony optimisation, particle swarm optimisation, bee colony optimisation, etc., are quite established for uncovering Pareto front in many objective optimisations. Hence, uncovering fuzzy association rules

by harnessing the potentiality of these techniques may be noted as the future line of research.

## References

- Aarts, E. and Korst, J. (1989) *Simulated Annealing and Boltzmann Machines*, John Wiley and Sons, Chichester, UK.
- Adamo, J.-M. (2000) *Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms*, Springer-Verlag, Berlin, Germany.
- Agrawal, R. and Srikant, R. (1994) 'Fast algorithms for mining association rules', in *Proceedings of the 20th International Conference on Very Large Databases*, Santiago, Chile.
- Agrawal, R., Imielinski, T. and Swami, A. (1993) 'Mining association rules between sets of items in large databases', in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Washington DC, USA, pp.207–216.
- Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. and Verkamo, A.I. (1996) 'Fast discovery of association rules', in Fayyad U.M. et al. (Eds.): *Advances in Knowledge Discovery and Data Mining*, pp.307–328, AAAI Press, Menlo Park.
- Albert, O., Jorge, C. and Francisco J.M. (2009) 'Unsupervised learning of fuzzy association rules for consumer behavior modeling', *Math-ware and Soft Computing*, Vol. 16, No. 1, pp.29–43.
- Argiddi, R.V. and Apte, S.S. (2012) 'Future trend prediction of Indian IT stock market using association rule mining of transaction data', *International Journal of Computer Applications*, Vol. 39, No. 10, pp.30–34.
- Atlas, B., Akin, E. and Karci, A. (2008) 'MODENAR: multi-objective differential evolution algorithm for mining numeric association rules', *Applied Soft Computing*, Vol. 8, No. 1, pp.646–656.
- Au, W.-H. and Chan, K.C.C. (2003) 'Mining fuzzy association rules in a bank-account database', *IEEE Transactions on Fuzzy Systems*, Vol. 11, No. 2, pp.238–248.
- Bagui, S., Just, J. and Bagui, S.C. (2009) 'Deriving strong association mining rules using a dependency criterion, the lift measure', *International Journal of Data Analysis Techniques and Strategies*, Vol. 1, No. 3, pp.297–312.
- Bayardo Jr., R.J. and Agrawal, R. (1999) 'Mining the most interesting rules', in *Proceedings of KDD*, pp.145–153.
- Bayardo, R.J. and Agrawal, R. (1999) 'Mining the most interesting rules', in *Proceedings of 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.145–153.
- Bentley, P.J. and Wakefield, J.P. (1997) 'Find an acceptable Pareto-optimal solution using multi-objective genetic algorithms', in Chawdhry, P.K. et al. (Eds.): *Soft Computing in Engineering Design and Manufacturing*, pp.231–240, Springer-Verlag, Berlin/Heidelberg, Germany.
- Blum, C. (2005) 'Ant colony optimization: introduction and recent trends', *Physics of Life Review*, Vol. 2, No. 4, pp.353–373.
- Boyd, S.P. and Vandenberghe, L. (2004) *Convex Optimization*, Cambridge University Press, New York, NY, USA.
- Brin, S., Motwani, R., Ullman, J. and Tsur, S. (1997) 'Dynamic itemsets counting and implication rules for market basket data', in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp.255–264.
- Carmona, C.J., Gonzalez, P., Del Jesus, M.J., Navio-Acosta, M. and Jimenez-Trevino, T. (2011) 'Evolutionary fuzzy rule extraction for subgroup discovery in a psychiatric emergency department', *Soft Computing*, Vol. 15, No. 12, pp.2435–2448.
- Chai, C. and Li, B. (2010) 'A novel association rules method based on genetic algorithm and fuzzy set strategy for web mining', *Journal of Computers*, Vol. 5, No. 9, pp.1448–1455.

- Chen, C-H., Hong, T-P., Tseng, V.S. and Chen, L-C. (2012) 'Multi-objective genetic-fuzzy data mining', *International Journal of Innovative Computing, Information and Control*, Vol. 8, No. 10(A), pp.6551–6568.
- Chen, M.S., Han, J. and Yu, P.S. (1996) 'Data mining an overview from database perspective', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, No. 6, pp.866–883.
- Chi, X. (2012) 'A new matrix-based association rules mining algorithm', in *Proceedings of the 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, Sichuan, China, pp.633–636.
- Coello, C.A.C. (1998) 'An updated survey of GA-based multi-objective optimization techniques', *ACM Computing Surveys*, Vol. 32, No. 2, pp.109–143.
- Coello, C.A.C. (1999) 'A comprehensive survey of evolutionary-based multi-objective optimization', *Knowledge and Information Systems*, Vol. 1, No. 3, pp.269–308.
- Coenen, F., Leng, P. and Goulbourne, G. (2004) 'Tree structures for mining association rules', *Data Mining and Knowledge Discovery*, Vol. 8, No. 1, pp.25–51.
- Cohen, J.L. (1978) *Multi-Objective Programming and Planning*, Academic Press, New York.
- Cohon, J.L. (1978) *Multi-objective Programming and Planning*, Academic Press, New York.
- Cordon, O., Herrera, F. and Villar, P. (2001) 'Generating the knowledge base of fuzzy rule-based system by the genetic learning of the data', *IEEE Transactions on Fuzzy Systems*, Vol. 9, No. 4, pp.667–674.
- Das, S. and Suganthan, P.N. (2011) 'Differential evolution: a survey of the state-of-the-art', *IEEE Transactions on Evolutionary Computation*, Vol. 15, No. 1, pp.4–31.
- Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. (2002) 'A fast and elitist multi-objective genetic algorithm: NSGA-II', *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 2, pp.182–197.
- Dehuri, S. and Ghosh, A. (2004) 'Evolutionary algorithms for multi-criterion optimization: a survey', *International Journal of Computing and Information Sciences*, Vol. 2, No. 1, pp.38–57.
- Dehuri, S.N., Jagadev, A.K., Ghosh, A. and Mall, R. (2006) 'Multi-objective genetic algorithm for association rule mining using a homogeneous dedicated cluster of workstations', *American Journal of Applied Sciences*, Vol. 3, pp.2086–2095.
- Del, J.M., Pedro, Jose, A.G., Pedro, G. and Jose, M.P. (2011) 'On the discovery of association rules by means of evolutionary algorithms', *WIREs Data Mining Knowledge Discovery*, Vol. 1, No. 5, pp.397–415.
- di Pierro, F., Soon-Thiam, K. and Savic, D.A. (2007) 'An investigation on preference order ranking scheme for multi-objective optimization', *IEEE Transactions on Evolutionary Computation*, Vol. 11, No. 1, pp.17–45.
- Dubois, D., Prade, H. and Sudkamp, T. (2005) 'On the representation, measurement and discovery of fuzzy associations', *IEEE Transactions on Fuzzy Systems*, Vol. 13, No. 2, pp.250–262.
- Edgeworth, F.Y. (1881) *Mathematical Physics*, P. Keagan, London, England.
- Fonseca, C.M. and Fleming, P.J. (1993) 'Genetic algorithms for multi-objective optimization: formulation, discussion and generalization', in *Proceedings of 5th International Conference on Genetic Algorithms*, pp.416–423.
- Freitas, A. (1999) 'On rule interestingness measures', *Knowledge-Based System*, Vol. 12, Nos. 5–6, pp.309–315.
- Freitas, A.A. (2002): *Data Mining and Knowledge Discovery with Evolutionary Algorithms*, Springer-Verlag, New York.
- Freitas, A.A., Noda, E. and Lopes, H.S. (1999) 'Discovering interesting prediction rules with a genetic algorithm', in *Proceedings of the Conference on Evolutionary Computation*, Washington DC, USA, pp.1322–1329.
- Geng, L. and Hamilton, H.J. (2006) 'Interestingness measures for data mining: a survey', *ACM Computing Surveys*, Vol. 38, No. 3, pp.1–32.

- Ghazi, H.L. and Abadeh, M.S. (2012) 'Mining fuzzy association rules with 2-tuple linguistic terms in stock market data by using genetic algorithm', in *Proceedings of the 16th CSI International Symposium on Artificial Intelligence and Signal Processing*, pp.354–359.
- Ghosh, A. and Das, M.K. (2008) 'Non-dominated rank based sorting genetic algorithms', *Fundamental Informaticae*, Vol. 83, No. 3, pp.231–252.
- Ghosh, A. and Nath, B. (2004) 'Multi-objective association rule mining using genetic algorithm', *Information Sciences*, Vol. 163, Nos. 1–3, pp.123–133.
- Ghosh, A. et al. (Eds.) (2008) *Multi-objective Evolutionary Algorithms for Knowledge Discovery in Databases*, Springer-Verlag, Berlin/Heidelberg, Germany.
- Gyenesei, A. (2001) 'A fuzzy approach for mining quantitative association rules', *Acta Cybernetica*, Vol. 15, No. 2, pp.305–320.
- Hadian, A., Nasiri, M. and Minaei-Bidgoli, B. (2010) 'Clustering based multi-objective rule mining using genetic algorithm', *International Journal of Digital Content Technology and its Applications*, Vol. 4, No. 1, pp.37–42.
- Han, J. and Kaufman, M. (2011) *Data Mining: Concepts and Techniques*, Morgan Kaufman Publishers, San Francisco, CA, USA.
- Hong, T.P., Chen, C.H., Wu, Y.L. and Lee, Y.C. (2006) 'A GA-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership functions', *Soft Computing*, Vol. 10, No. 11, pp.1091–1101.
- HooshSadat, M., Samuel, H.W., Patel, S. and Zaiane, O.R. (2011) 'Fastest association rule mining algorithm predictor (FARM-AP)', in *Proceedings of the Fourth International C\* Conference on Computer Science and Software Engineering*, NY, USA, pp.43–50.
- Horn, J., Nafplotis, N. and Goldberg, D.E. (1994) 'A Niche Pareto genetic algorithm for multi-objective optimization', in *Proceedings of the first IEEE Conference on Evolutionary Computation*, pp.82–87.
- Hsu, W., Liu, B. and Chen, S. (1997) 'Using general impressions to analyze discovered classification rules', in *Proceedings of 3rd International conference on Knowledge Discovery and Data Mining*, AAAI Press, pp.31–36.
- Hu, J. and Yang-Li, X. (2007) 'Association rule mining using multi-objective co-evolutionary algorithm', in *Proceedings of the IEEE International Conference on Computational Intelligence and Security Workshops*, pp.405–408.
- Imielinski, T., Agrawal, R. and Swami, A. (1993) 'Mining association rules between sets of items in large databases', in *Proceedings of ACM SIGMOD Conference on Management of Data*, pp.207–216.
- Jain, A.K. and Dubes, R.C. (1988) *Algorithms for Clustering Data*, Prentice-Hall, New Jersey.
- Jin, Y. (2000) 'Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement', *IEEE Transactions on Fuzzy systems*, Vol. 8, No. 2, pp.212–221.
- Kalia, H., Dehuri, S.N. and Ghosh, A. (2012) 'A survey on fuzzy association rule mining algorithm', *International Journal of Data Warehousing and Mining* (accepted).
- Kantardzic, M. (2003) *Data Mining: Concepts, Models, Methods, and Algorithms*, Wiley-Interscience, Hoboken, NJ.
- Kaya, M. (2006) 'Multi-objective genetic algorithm based approaches for mining optimized fuzzy association rules', *Soft computing*, Vol. 10, No. 7, pp.578–586.
- Kaya, M. and Alhaji, R. (2004a) 'Mining optimized fuzzy association rules using multi-objective genetic algorithms', in *Proceedings of 8th IEEE Conference on Intelligent Engineering Systems, Romania*, pp.38–43.
- Kaya, M. and Alhaji, R. (2004a) 'Multi-objective genetic algorithm based method for mining optimized fuzzy association rules', in Yang et al. (Eds.): *IDEAL, LNCS*, Vol. 3177, pp.758–764.

- Kaya, M. and Alhaji, R. (2004c) 'Integrating multi-objective genetic algorithm into clustering for fuzzy association rule mining', in *The Proceedings of IEEE International Conference on Data Mining*, pp.431–434.
- Keim, D.A. (2002) 'Information visualization and visual data mining', *IEEE Transaction on Visualization and Computer Graphics*, Vol. 7, No. 1, pp.100–107.
- Koper, A. and Nguyen, H.S. (2011) 'Sequential pattern mining from stream data', in *Proceedings of 7th International Conference on Advanced Data Mining and Applications*, Beijing, China. pp.278–291.
- Koski, J. (1984) 'Multicriterion optimization in structural design', in Artek, E. et al. (Eds.): *New Directions in Optimum Structural Design*, pp.483–503.
- Kouk, C.M., Fu, A. and Wong, M.H. (1998) 'Fuzzy association rules in databases', *ACM SIGMOD Record*, Vol. 27, No. 1, pp.41–46.
- Li, F. and Jia-ju, Q. (2005) 'Short-term load forecasting for anomalous days based on fuzzy multi-objective genetic optimization algorithm', in *Proceedings of the CSEE*, Vol. 25, No. 10, pp.29–34.
- Mahajan, S. and Reshamwala, A. (2011) 'An approach to optimize fuzzy time-interval sequential patterns using multi-objective genetic algorithm', in *Proceedings of the Technology Systems and Management*, Springer Berlin, Heidelberg, pp.115–120.
- Mangalampalli, A. and Pudi, V. (2009) 'Fuzzy association rule mining algorithm for fast and efficient performance on very large datasets', in *Proceedings of IEEE International Conference on Fuzzy Systems*, Jeju Island, Korea, pp.1163–1168.
- Mao, P. and Zhu, Q. (2002) 'Association rules applied to intrusion detection', *Wuhan University Journal of Natural Sciences*, Vol. 7, pp.426–430.
- Marler, M.T. and Arora, J.S. (2004) 'Survey of multi-objective optimization methods for engineering', *Struct Multidisc Optim*, Vol. 26, No. 6, pp.369–395.
- Martin, D., Rosette, A., Alcalá-Fdez, J. and Herrera, F. (2011) 'A multi-objective evolutionary algorithm for quantitative association rules', in *Proceedings of 11th International Conference on Intelligent Systems Design and Applications*, pp.1397–1402.
- Miettinen, K. (1998) *Nonlinear Multi-objective Optimization*, Kluwer, Boston.
- Moslehi, P., Bidgoli, B.M., Nasiri, M. and Salajegheh, A. (2011) 'Multi-objective numeric association rules mining via ant colony optimization for continuous domains without specifying minimum support and minimum confidence', *International Journal of Computer Science Issues*, Vol. 8, Nos. 1, 5, pp.34–40.
- Muyeba, M., Khan, S.M. and Coenen, F. (2008) 'A framework for mining fuzzy association rules from composite items', in *Proceedings of Algorithms for Large Scale Information Processing in Knowledge Discovery*, pp.65–77.
- Nasiri, M., Sadat Taghavi, L. and Minaee, B. (2011) 'Numeric multi-objective rule mining using simulated annealing algorithm', *International Journal of Applied Operation Research*, Vol. 1, No. 2, pp.37–48.
- Ng, R.T. and Han, J. (1994) 'Efficient and effective clustering methods for spatial data mining', in *Proceedings of the 20th VLDB Conference*, pp.144–155.
- Olufunke, O.O., Charles, U.O. and Charles, A.K. (2010) 'On sharp boundary problem in rule based expert systems in the medical domain', *International Journal of Healthcare Information Systems and Informatics*, Vol. 5, No. 3, pp.14–26.
- Osyczka, A. (1985) 'Multi-criteria optimization for engineering design', in Gero, J.S. (Eds.): *Design Optimization*, pp.193–227, Academic Press.
- Ouyang, W. (2012) 'Mining positive and negative fuzzy association rules with multiple minimum supports', in *Proceedings of the International Conference on Systems and Informatics*, Yantai, Shandong, China, pp.2242–2246.
- Ozokes, S. and Camurcu, A.Y. (2002) 'Classification and prediction in data mining application', *Journal of Marmara for Pure and Applied Sciences*, Vol. 18, pp.159–174.

- Pareto, V. (1896) *Cours d'Economie Politique*, Volume I and II, F. Rouge, Lausanne.
- Qodmanan, H.R., Nasiri, M. and Minaei-Bidgoli, B. (2011) 'Multi-objective association rule mining with genetic algorithm without specifying minimum support and confidence', *Expert Systems with Applications*, Vol. 38, No. 1, pp.288–298.
- Quinlan, J.R. (1993) *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, CA.
- Rajendran, P. and Madheswaran, M. (2010) 'Hybrid medical image classification using association rule mining with decision tree algorithm', *Journal of Computing*, Vol. 2, No. 1, pp.127–136.
- Schaffer, J.D. (1985) 'Multiple objective optimization with vector evaluated genetic algorithms', in *Proceedings of International Conference on Genetic algorithms and their Applications*, pp.93–100.
- Schaffer, J.D. (1984) *Multiple Objective Optimization with Vector Evaluated Genetic Algorithms*, Unpublished PhD thesis, Vanderbilt University, Nashville, Tennessee.
- Sekhavat, Y.A., Fathian, M., Gholamian, M.R. and Alizadeh S. (2010) 'Mining important association rules based on the RFMD technique', *International Journal of Data Analysis Techniques and Strategies*, Vol. 2, No. 1, pp.1–21.
- Srikant, R. and Agrawal, R. (1996) 'Mining quantitative association rules in large relational tables', in *Proceedings of the ACM Conference on Management of the Data*, pp.1–12.
- Srinivas, N. and Deb, K. (1994) 'Multi-objective optimization using non-dominated sorting in genetic algorithms', *Evolutionary Computation*, Vol. 2, No. 3, pp.221–230.
- Steuer, R.E. (1986) *Multiple Criteria Optimization: Theory, Computations, and Application*, John Wiley & Sons, New York.
- Tamaki, H., Kita, H. and Kobayashi, S. (1996) 'Multi-objective optimization by genetic algorithms: a review', in *Proceedings of the ICE*, pp.517–522.
- Tan, P-N., Kumar, V. and Srivastava, J. (2002) 'Selecting the right interestingness measure for association patterns', in *Proceedings of the 8th ACM SIGKDD Conference*, Edmonton, Canada.
- Thilagam, P.S. and Anathanarayanna, V.S. (2008) 'Extraction and optimization of fuzzy association rules using multi-objective genetic algorithm', *Pattern Analysis and Applications*, Vol. 11, No. 2, pp.159–168.
- Tzacheva, A.A. (2012) 'Rule schemas and interesting association action rules mining. *International Journal of Data Mining, Modeling and Management*, Vol. 4, No. 3, pp.244–254.
- Vandeva, E. (2012) 'Multi-objective genetic modified algorithm', *Cybernetics and Information Technologies*, Vol. 12, No. 2, pp.23–33.
- Wai-Ho, A. and Keith, C.C. (1999) 'FARM: A data mining system for discovering fuzzy association rules', in *Proceedings of 8th International conference on fuzzy systems*, Seoul, Korea, pp.1217–1222.
- Wakabi-Waiswa, P.P. and Baryamureeba, V. (2011) 'Mining high quality association rules using genetic algorithms', in *Proceedings of the 22nd Midwest Artificial Intelligence and Cognitive Science Conference*, Cincinnati, USA.
- Wantanabe, T. (2010) 'A fast fuzzy association rules mining algorithm utilizing output field specification', *International Journal of Biomedical Soft Computing and Human Sciences*, Vol. 10, No. 2, pp.69–76.
- Witten, I.H., Frank, E., Hall, M.A. and Kaufman, M. (2011) *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kauffman Publishers, San Francisco, CA, USA.
- Won, D. and McLeod, D. (2012) 'An efficient approach to categorizing association rules', *International Journal of Data Mining, Modeling and Management*, Vol. 4, No. 4, pp.309–333.
- Woo, H.J. and Lee, W.S. (2009) 'estMax: tracing maximal frequent item sets instantly over online transactional data streams', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21, No. 10, pp.1418–1431.

- Xizheng, Z. (2007) 'Building personalized recommendation system in e-commerce using association rule-based mining and classification', in *Proceedings of the 8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, Vol. 3, pp.853–857.
- Zhang, Y., Zhang, L., Liu, Y. and Shi, Y. (2010) 'Mining intelligent knowledge from a two-phase association rules mining', *International Journal of Data Mining, Modeling and Management*, Vol. 2, No. 4, pp.403–419.
- Zitzler, E. and Thiele, L. (1999) 'Multi-objective evolutionary algorithms: a comparative case study and the strength Pareto approach', *IEEE Transactions on Evolutionary Computation*, Vol. 3, No. 4, pp.257–271.