IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 9, 2019, accepted July 31, 2019, date of publication August 6, 2019, date of current version August 19, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2933505

Post-Analysis Framework for Mining Actionable Patterns Using Clustering and Genetic Algorithms

CHUN-HAO CHEN[1], JI-SYUAN HE[1], TZUNG-PEI HONG[2,3], AND SUBBAIYA RAMMOHAN KANNAN[4]

[1]Department of Computer Science and Information Engineering, Tamkang University, Taipei 251, Taiwan
[2]Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan
[3]Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung 804, Taiwan
[4]Department of Mathematics, Pondicherry University, Pondicherry 605014, India

Corresponding author: Tzung-Pei Hong (tphong@nuk.edu.tw)

This research was supported by the Ministry of Science and Technology of the Republic of China under grants MOST 107-2221-E-390-016-MY2 and 108-2221-E-032-037. This is a modified and expanded version of the paper "A two-stage multi-objective genetic-fuzzy mining algorithm," presented at the 2013 IEEE Symposium Series on Computational Intelligence [7].

ABSTRACT Mining association rules is an important technique in data analysis. Many approaches for rule analysis have been designed to address different problems. Among them, some works developed from multiobjective genetic algorithms derive a set of Pareto solutions, each of which contains a set of membership functions for fuzzy data mining from quantitative transactions with taxonomy. However, because more than one solution exists in a Pareto set, finding a method to determine the appropriate membership functions and combine them with useful knowledge for mining actionable patterns (such as fuzzy generalized association rules and fuzzy utility itemsets) is a useful research problem. Hence, this paper presents a post-analysis-based genetic-fuzzy mining (PA-GFM) framework for mining actionable patterns that involves two phases: membership-function mining and actionable pattern mining. In the first phase, an existing approach is utilized to derive the Pareto solutions with objective functions. In the second phase, a clustering technique using clustering attributes selected by the users is employed to group the Pareto solutions. The representative solution from each group is then exploited to mine actionable patterns based on the users' requirements. Experiments were conducted on both a simulated dataset and a real one to investigate the performance of the PA-GFM framework.

INDEX TERMS Clustering algorithms, domain-driven data mining, fuzzy generalized association rules, fuzzy utility itemsets, multiobjective genetic algorithms.

I. INTRODUCTION

The rapid growth of transactions in supermarkets has revealed a need to develop tools for decision-makers to derive actionable information when creating their marketing strategies. In other words, because supermarket databases store many transactions, it is difficult for supermarket owners to understand how to choose the appropriate criteria to analyze them. Selecting inappropriate evaluation criteria for mining patterns can reduce profits. In contrast, adopting appropriate evaluation criteria such as item frequency, utility, or taxonomy may result in derived patterns that are more actionable. Thus, a pattern-mining approach that supports making tradeoffs among diverse criteria is required.

The associate editor coordinating the review of this manuscript and approving it for publication was Mustafa Servet Kiran.

One of the best-known mining algorithms for analyzing transactions uses association rules [2]. Many previous rule mining approaches have been created to address various problems, and fuzzy rule mining approaches exist that can process transactions with quantitative values [14], [20], [30]. Because items have taxonomy, algorithms for mining that use generalized fuzzy rule are designed to extract implicit patterns from transactions under single or multiple supports [15], [17] [21]. Recently, some scholars have designed genetic fuzzy mining (GFM) algorithms to obtain both fuzzy association rules and membership functions [13], [24], [27]. The knowledge mined can reflect linguistic and uncertain characteristics of the databases.

In real applications, decision-makers may consider multiple criteria when determining business strategies. Thus, multiobjective optimization has become increasingly important. Many GFM approaches using multiobjective genetic algorithms (MOGA) have been presented that use different objective functions such as the number of interesting itemsets and rules and the shapes of membership functions [1], [6], [7], [18]. Although Pareto solutions derived from these algorithms can provide users with a variety of choices, they still faced the dilemma of how to generate fitting solutions that mine useful knowledge. Cao et al. proposed a domain-driven data mining (D^{3}M) method that combined domain knowledge with meta-knowledge to mine useful information. This approach was also called actionable knowledge discovery (AKD) [4], [5], where the word ''actionable'' meant that the derived knowledge patterns could reflect realities that decision-makers could use to determine appropriate business strategies. Under this goal, four logical concepts of D^{3}M were described in [4], including post-analysis-based AKD (PA-AKD), unified-interestingness-based AKD, combined-mining-based AKD, and multisource combined-mining-based AKD. Hence, actionable patterns mean the information could be derived using the four logical concepts, and the patterns could be represented in different forms. For instance, the fuzzy set theory and the attribute flexibility are considered for the enhancement of the fuzzy cost-effective action mining algorithm (F-CEAMA) to maximize the fuzzy net profit, where the actionable patterns could be derived from the fuzzy decision tree using the fuzzy post-processing [19]. As to other existing approaches, they can be divided into three types, which are domain-driven data mining, action mining and providing measures to verify discovered patterns.

Based on the D^{3}M concept, this work presents an approach that can not only mine membership functions (Pareto solutions) but also provide representative solutions that can mine actionable knowledge for decision-makers. Thus, the Post-Analysis-based Genetic-Fuzzy Mining (PA-GFM) framework is proposed for mining actionable patterns according to the PA-AKD framework. The main reason to employ the PA-AKD framework to design the PA-GFM framework is that it focuses on how to apply domain knowledge for post-processing to derive actionable patterns. Using the proposed framework, the previous approach [7] is utilized to extract membership functions (Pareto solutions) in the first stage, and actionable patterns are derived through domain knowledge in the second stage. In other words, initially, the genetic-fuzzy mining algorithm [7] is used to derive nondominated solutions, which is the goal of first stage, with the given objective functions (technique knowledge). In second stage, a clustering technique using selected clustering attributes that are extracted by the domain knowledge is then applied to divide the Pareto solutions into groups. Then, in each group, the representative solution can be employed to discover actionable patterns based on the users' requirements, where the actionable patterns are fuzzy generalized association rules (FGAR) and utility fuzzy closed itemsets (UFCI). Hence, based on the derived FGAR and UFCI, the proposed approach could belong to not only descriptive but also predictive analytic method because the derived information can be used for data description and prediction based on [33]. For example, the derived fuzzy generalized association rules could not only be used for description what happened in the historic data but also for predicting what will happen in the future. At last, experiments on two datasets were conducted to investigate the effectiveness of the framework.

The differences and improvements between this work and the proposed study in [8] includes: (1) By utilizing the D^{3}M concept, the Post-Analysis-based Genetic-Fuzzy Mining (PA-GFM) framework, which composes of two phases: non-dominated solution mining and actionable pattern mining, is proposed for deriving actionable patterns, where a non-dominated solution is a set of membership functions for items and actionable patterns include fuzzy generalized association rules and utility fuzzy closed itemsets. Note that proposed approach presented in [8] can only be used to mine fuzzy generalized association rules. (2) To make readers understand our paper easily, the related work is introduced in details and an example is also given to illustrate the proposed PA-GFM framework. (3) At last, extensive experiments carried on a simulated and a real transaction datasets show that the proposed approach can actually be utilized to derive actionable patterns, including (a) the evolution of Pareto front in the first phase, (b) the evaluation of the clustering results in the second phase, and (c) the evaluation of the derived actionable patterns. Besides, the main difference between the proposed approach and the previous approach [7] is that the proposed approach provides a more sophisticated procedure which is the actionable pattern mining procedure (Stage II) to extract actionable patterns.

The remainder of this work is organized as follows. The related work is described in Section II. The design of the PA-GFM framework is presented in Section III. The clustering attributes and objective functions used in the proposed approach are provided in Section IV. In Section V, the proposed algorithm based on the PA-GFM framework is explained, and in Section VI, an example is provided to illustrate the proposed approach. The experimental evaluation is presented in Section VII. Finally, conclusions and future work are given in Section VIII.

II. RELATED WORK

The related work is introduced in this section. In Section II.A, the preliminaries of multiobjective optimization are described. Genetic-fuzzy mining approaches are then stated in Section II.B. The utility of fuzzy itemset mining approaches is given in Section II.C.

A. PRELIMINARIES OF MULTIOBJECTIVE PROBLEMS

For optimization problems, the goals that must be satisfied are transformed into different factors. Through those factors, a fitness function is then designed such that evolutionary algorithms can solve the specific problem. Unfortunately, it is not always easy to define an appropriate fitness function for a problem in advance. Hence, multiobjective optimization problems that consider various criteria have become important in real applications. This type of problem is formally stated as follows:

$$\begin{aligned}&\text{Min/Max } z = o(x) = (o_1(x), o_2(x), \ldots, o_{\mathrm{m}}(x)),\\ &\text{subject to } a = (a_1, a_2, \ldots, a_h) \in A \quad \text{and}\\ &\qquad\qquad b = (z_1, z_2, \ldots, z_k) \in Z,\end{aligned}$$

where a and z are the decision and objective vectors and A and Z represent the objective space and decision space, respectively. Fonseca and Fleming proposed the Multi-Objective Genetic Algorithm (MOGA) to solve such problems. MOGA evaluated the fitness values of the solutions based on the extended ranks and then used them to find Pareto solutions [10]. Enhanced approaches such as NSGA-II [9] and SPEA2 [40] were later proposed to find better Pareto fronts than MOGA. The main idea behind NSGA-II is to utilize a procedure to perform a fast sort of the nondominated individuals to derive Pareto solutions more effectively [9]. SPEA2 then proposed a better strategy for assigning fitness values to solutions by estimating the density of the nondominated solutions and enhancing archive truncation [40]. In addition, Yu et al. presented the neighborhood knowledge-based evolutionary algorithm (NKEA) to solve multiobjective problems via three stages, including the direction-learning, mutual-adaptation, and self-adaptation stages [39]. It not only took the advantages of multiobjective approaches into consideration, e.g., NSGA-II [9], but also utilized the neighborhood knowledge systematically to enhance searching ability during the evolution process.

B. GENETIC-FUZZY MINING APPROACHES

The use of association rules is the best-known technique for market basket analysis [2]. Because quantitative values usually appear in transactions, many approaches that use fuzzy sets have been designed to handle these transactions when mining fuzzy association rules [14], [20], [30]. Fuzzy generalized rule mining approaches have also been presented that obtain valuable information from quantitative transactions with taxonomy under single or multiple supports [15], [17], [21]. In addition, item sales earn different profits. Thus, various methods have been presented to derive fuzzy utility itemsets [22], [34].

Because membership functions are utilized to handle the quantitative values, their settings have significant impacts on the mining results. In the methods mentioned above, the membership functions were assumed to be predefined; however, later approaches presented some genetic fuzzy mining algorithms to derive both membership functions and linguistic association rules [13], [24], [27]. Considering the number of large itemsets and the suitability of membership functions, Hong *et al.* proposed a GA-based approach to solve the above problem [13]. By considering the temporal aspect of transactions, Matthews *et al.* designed a method to derive temporal linguistic rules [24]. Palacios *et al.* then applied GFM to perform fuzzy mining of low-quality data [27].

Because decision-makers may adopt multiple criteria when designing application strategies, multiobjective problems have recently attracted much attention. Several GFM algorithms based on multiobjective evolution algorithms (MOEA) that use various objective functions have been proposed [1], [6], [7], [18]. Kaya presented an algorithm based on MOGA to find optimized rules under three criteria, termed strength, interestingness and comprehensibility [18]. Alhajj and kaya derived membership functions using MOGA with an automated clustering method to obtain fuzzy rules from [1]. The criteria used in that work include the number of frequent itemsets and improvement in execution time. Chen *et al.* also used MOGA to find suitable membership functions and adopted two criteria—the suitability of membership functions and the number of frequent itemsets—which they used to derive fuzzy association rules [6]. Chen *et al.* also proposed an algorithm that used a given taxonomy to derive multilevel linguistic rules using MOGA [7].

Although the derived Pareto solutions can provide decision-makers with numerous choices, it is still difficult to identify a useful solution from those choices. For instance, suppose a Pareto set contains forty solutions and that no solution is dominated by any other under the given criteria. It is then an interesting problem to select the most appropriate solutions for mining actionable patterns. In this work, the goal is to design a framework that can not only derive Pareto solutions and mine actionable knowledge but also provide more useful information to users. Note that each Pareto solution consists of a set of potential membership functions.

C. THE UTILITY OF FUZZY ITEMSET MINING APPROACHES

Using an association rule mining approach can find large itemsets and rules with high confidence values. However, those patterns may not provide actionable knowledge that meet business goals. Thus, utility mining algorithms have been presented and used widely in various applications such as retail transaction analysis. As described above, transactions contain quantitative values. Therefore, some algorithms have been proposed that consider the utility of fuzzy itemsets in recent years.

For example, to utilize quantitative transactions, Wang et al. proposed an approach named High-Utility Fuzzy Itemsets Miner (HUFI-Miner) that derived fuzzy itemsets resulting in high profits [34]. Using the given membership functions, the quantitative values were first converted into linguistic regions. Then, the utility of each fuzzy itemset was calculated based on the external utility values of the items. Finally, the utilities of the fuzzy itemsets were compared with a minimum utility threshold to determine whether they were high-utility fuzzy itemsets. To reveal the influence of fuzzy degrees and the profit of high-utility itemsets, Lai *et al.* introduced the Fuzzy High-Utility Itemsets Mining (FHUI-Mine) method to discover fuzzy high-utility itemsets [22]. In this paper, the utilities of the fuzzy itemsets are incorporated into a

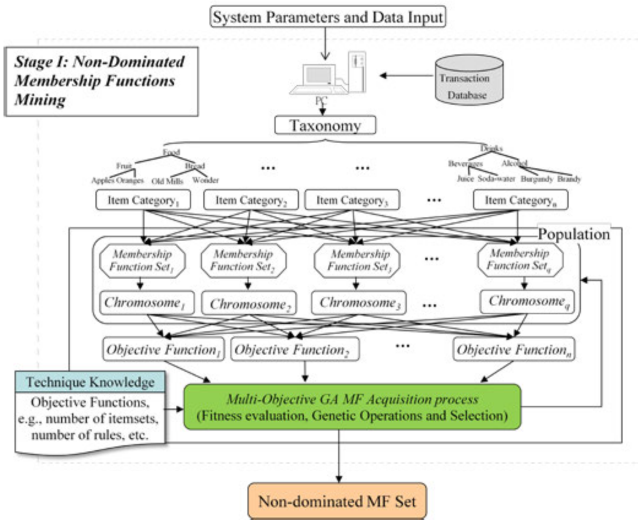

FIGURE 1. The first stage of PA-GFM framework.

second stage of the proposed framework to find representative solutions and mine actionable knowledge.

III. PROPOSED POST-ANALYSIS-BASED GENETIC-FUZZY MINING FRAMEWORK

The proposed PA-GFM framework for mining actionable patterns, which utilizes the PA-AKD framework, is described in this section. The PA-GFM framework involves two phases. The goal of the first phase is to obtain Pareto solutions from the technique knowledge (objective functions). Each Pareto solution contains a set of membership functions. Then, in the second phase, a clustering method is used to divide the Pareto solutions into several groups. The representative solutions from each group are presented to users to assist them in making decisions. The first stage of the framework is depicted in Fig. 1.

In the first stage, the generalized items are encoded into chromosomes based on the taxonomy provided as input to the first phase. Here, each generalized item uses a set of membership functions. Therefore, each individual includes the set of membership functions for all generalized items. Below, the chromosome representation presented in [7] used to for encoding generalized items is described. Assume there are two generalized items, a possible chromosome is shown in Fig. 2.

Fig. 2 shows two generalized items which are "Food" and "Drinks". Because each membership function used for an generalized item is represented by two genes that are the center abscissa c_{jk} and half the spread w_{jk} of the fuzzy region R_{jk}, the generalized item "Food" which has three fuzzy regions (Low, Middle and High) are encoded as "2, 2, 7, 4, 11, 3". In the same way, the fuzzy regions for "Drinks" can be encoded as "1, 2, 6, 3, 11, 4". As a result, the two generalized items are represented as "2, 2, 7, 4, 11, 3, 1, 2, 6, 3, 11, 4" in the chromosome C_1. Then, utilizing the multi-objective genetic-fuzzy acquisition process [7], the nondominated membership functions are derived. Note that different

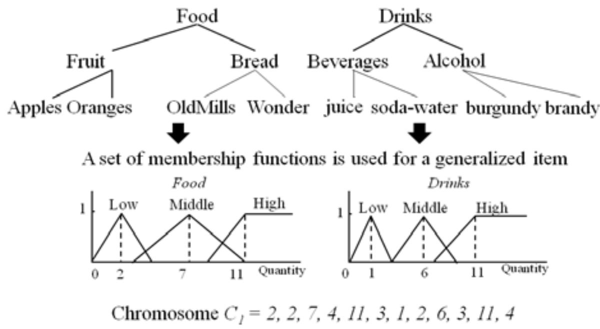

FIGURE 2. An example for chromosome representation.

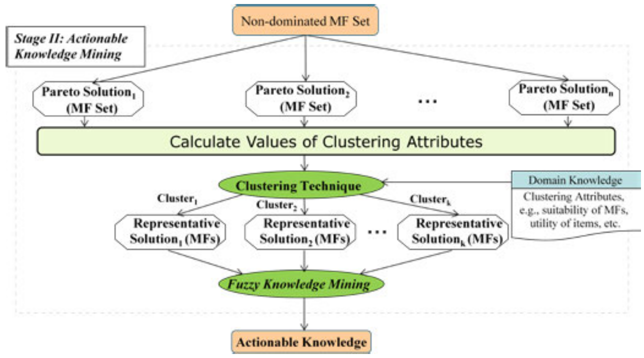

FIGURE 3. The second stage of the PA-GFM framework.

MOGA approaches can be used in this stage such as those in [9], [10], [39], [40]. Additionally, the objective functions, which are known as technique knowledge, are selected and incorporated into the evolution procedure based on the users' preferences [1], [6], [7], [18].

As to the genetic operators, the max-min-arithmetical (MMA) crossover [12] and the one-point mutation are used to form offspring solutions. The MMA crossover will generate four candidate chromosomes from the selected two chromosomes, $C_u^t = \{c_1, \ldots, c_h, \ldots, c_z\}$ and $C_u^t = \{c_1', \ldots, c_h', \ldots, c_z'\}$ that are shown as follows:

1. $C_1^{t+1} = \{c_{11}^{t+1}, \ldots, c_{1h}^{t+1}, \ldots, c_{1z}^{t+1}\}$, where $c_{1h}^{t+1} = dc_h + (1-d)c_h'$.
2. $C_2^{t+1} = \{c_{21}^{t+1}, \ldots, c_{2h}^{t+1}, \ldots, c_{2z}^{t+1}\}$, where $c_{2h}^{t+1} = (1-d)c_h + dc_h'$.
3. $C_3^{t+1} = \{c_{31}^{t+1}, \ldots, c_{3h}^{t+1}, \ldots, c_{3z}^{t+1}\}$, where $c_{3h}^{t+1} = \min\{dc_h, c_h'\}$.
4. $C_4^{t+1} = \{c_{41}^{t+1}, \ldots, c_{4h}^{t+1}, \ldots, c_{4z}^{t+1}\}$, where $c_{4h}^{t+1} = \max\{dc_h, c_h'\}$.

The best two out of the four chromosomes are kept in population. The one-point mutation operator will derive a different membership function through randomly adding a value (between $\pm w_{jk}$) to the center or spread of a fuzzy region R_{jk}. The second stage of the framework is illustrated in Fig. 3.

In the second stage, because each nondominated solution in the Pareto set involves a tradeoff between the incorporated objective functions, it is not easy for the user to choose appropriate solutions. The goal of this phase is to obtain representative solutions for mining actionable patterns. Here, the experts' domain knowledge is considered to determine the appropriate clustering attributes. For example, item utility could be considered when the decision-maker focuses on profit; alternatively, the number of frequent itemsets (rules) could be used when the decision-maker wants to know the relationships among items. Hence, the nondominated solutions are grouped into several clusters by the clustering techniques, e.g., k-means or others [35], [36], [37], [38]. Then, the representative solutions from the groups are utilized to mine actionable patterns such as fuzzy utility itemsets or fuzzy association rules [15], [17], [22], [34].

According to the references [3], [11], transductive and inductive learning algorithms are used to construct classification models. The transductive learning is utilized to solve a specific problem. It will use specific training instances to build the classification model, which is then used to predict specific testing instances. Two well-known approaches are the k-nearest neighbors (KNN) and the support vector machine (SVM) [11]. As to the inductive learning, it is employed to deal with a more general problem. The goal of inductive learning is to utilize the observed training instances to construct a general classification model (like rules). Then, the general rules are applied on the testing instances. The well-known algorithms are ID3 and C4.5 [3]. Because the proposed framework is designed to extract fuzzy generalized association rules and utility fuzzy closed itemsets, they don't match the definitions of the transductive and inductive learning. However, if we limit the itemsets in the consequence part as class labels, the proposed method can fall in the scope of inductive learning algorithms.

IV. CLUSTERING CRITERIA AND ATTRIBUTES IN THE PROPOSED APPROACH

The criteria used as technique knowledge for multiobjective evolutionary processes in the literature are stated in Section IV.A. Then, the attributes used for clustering to obtain representative solutions are given in Section IV.B.

A. OBJECTIVE FUNCTIONS USED IN THE FIRST STAGE

According to the literature, a MOGA process may have different criteria. For example, Kaya presented three criteria to evaluate each solution: support threshold, rule correlation, and the average number of fuzzy sets in a rule [18]. In addition, Alhajj and Kaya used two different criteria in a MOGA-based fuzzy mining approach: runtime improvement and the number of frequent itemsets [1]. With a given taxonomy, Chen *et al.* considered the number of frequent itemsets and the suitability of membership functions as the two objective functions and designed MOGA-based approaches to mine membership functions and fuzzy association rules [6], [7]. The criteria mentioned above can be explained as follows.

(1) Fuzzy Support *FS(X)* of an itemset X is

$$FS(X) = \frac{\sum_{i=1}^{n} f_X^{(i)}}{n}, \tag{1}$$

where $f_X^{(i)} = MIN_{j=1}^{m} f_{Xj}^{(i)}$, and $f_{X_j}^{(i)}$ is the fuzzy value of the j-th item in X in the i-th transaction [6].

(2) The comprehensibility of a rule R is

$$Comprehensibility(R) = \frac{\log(1+|con|)}{\log(1+|ante \cup con|)}, \tag{2}$$

where $|\,con|$ is the attribute number in the consequent part of R and $|ante \cup con|$ is the attribute number of R [18].

(3) The fuzzy correlation of a rule "$(X, A) \rightarrow (Y, B)$" is defined as follows:

$$Corr((X,A)(Y,B)) = \frac{Cov((X,A)(Y,B))}{\sqrt{Var(X,A).Var(Y,B)}}, \tag{3}$$

where A, B, X and Y are itemsets, $Cov_{((X,A)(Y,B))} = FS_{(Z,C)} - FS^*_{(X,A)}FS_{(Y,B)}$, $Var_{(X,A)} = FS^2_{(X,A)} - (FS_{(X,A)})^2$, $FS_{(Z,C)}$ is the fuzzy support of item (Z, C), Z is the union of X and Y, and C is the union of A and B [18].

The range of this measurement is between -1 and 1, and it is utilized to evaluate the dependence of the antecedent and the consequent in a rule.

(4) The number of large (or frequent) itemsets in a chromosome C_q is defined as follows:

$$numL\ arg\ eItemSet(C_q) = \sum_{k=1}^{levels} |L_{1q}^k|, \tag{4}$$

where $|L_{1q}^k|$ represents the number of level-k large (frequent) 1-itemsets due to the membership functions represented by chromosome C_q [7].

In addition to the number of frequent 1-itemsets, Formula 4 can be extended to reflect the number of all frequent itemsets and the number of all rules generated from all levels; however, doing so involves a tradeoff between computational cost and knowledge accuracy. The rationale is that more frequent 1-itemsets will incur more frequent itemsets at a high probability, which will generate more association rules as well. A similar objective function that focuses on obtaining the maximum profit from a predefined minimum support interval could also be used [1].

(5) The suitability of the membership functions in chromosome C_q is defined as follows:

$$suitability(C_q) = \sum_{j=1}^{m} [\ overlap\_F(C_{qj}) + coverage\_F(C_{qj})], \tag{5}$$

where $overlap\_F(C_{qj})$ is used to measure the overlap degree of membership functions for item I_j and is calculated by the overlapping length of two fuzzy regions divided by half the minimum span of the two fuzzy regions. Here, $coverage\_F(C_{qj})$ is used to measure the ratio of the coverage range of the fuzzy regions over the maximum quantity of the generalized item I_j in the database [7]. These criteria are utilized to avoid fuzzy sets that are too redundant and too widely separated.

B. CLUSTERING ATTRIBUTES IN THE MINING ACTIONABLE PATTERN PHASE

After the Pareto solutions are derived from the first phase, the second phase involves clustering them to obtain the representative solutions for mining actionable patterns. The actionable knowledge discovery (AKD) and actionable pattern (AP) are stated as follows.

Definition 1 Actionable Knowledge Discovery (AKD): The definition given by Cao *et al.* [4] is given as follows:

"The AKD is an iterative optimization process toward the actionable pattern set, considering the surrounding business environment and problem states. It is a loop-closed and iterative refinement process, multiple feedbacks, iterations and refinement are involved in the understanding of data, resources, the roles and utilization of relevant intelligence, the presentation of patterns, the delivery specification, and knowledge validation".

Definition 2 Actionable Pattern (AP): The definition given by Cao *et al.* [4] is shown as follows:

"The AP satisfies both technical and business interestingness needs, and can be seamlessly taken over by business people for decision-making action-taking. AP can support business problem-solving by taking actions recommended by the pattern, and correspondingly transform the problem status from an initially non-optimal status to a greatly improved one".

Based on the definitions 1 and 2, we can understand that AKD is an iterative optimization process to derive the AP which should not only satisfy both technical and business interestingness but also can be utilized for business decision-making. In this paper, the two types of actionable patterns are defined as follows.

Definition 3 A Fuzzy Generalized Association Rule (FGAR): A FGAR is represented as "If X, Then Y ", where X and Y are sets of fuzzy itemsets, and its support and conference values are the two criteria used to assess its quality. For instance, a FGAR could be "If Food.Low, Then Drinks.Middle, sup. = 10%, conf. = 80%".

Note that an associative classification rule can be represented in the form: (A_{i1}, a_{i1}) ^…^$(A_{ik}, a_{ik}) \rightarrow c$, where the antecedent of the rule is an itemset and the consequent is a class [32]. Based on the definitions of ACR and FGAR (Definition 3), the difference between ACR and FGAR is their consequent parts. The former is a class and the latter is an itemset. In general, the purpose of ACR is to use the association-rule analysis to get classification rules from a supervised dataset. But the goal of the FGAR is to extract the relationships among items from given transactions (an unsupervised dataset).

Definition 4 A Utility Fuzzy Closed Itemset (UFCI): A UCFI is represented as a fuzzy itemset X, where X is a subset of the universal set $\{R_{11}, R_{12}, \ldots R_{jl}, \ldots, R_{nh}\}$. R_{jl} means the l-th fuzzy region of item I_j, and X has no superset with the same support value and its utility being larger than a given utility threshold. For example, a UCFI could be "(*11∗.Middle, 42∗.High, 31∗.High, 21∗.High*), utility = 87.52", where *11∗, 42∗, 31∗* and *21∗*are generalized items. For example, the code *42∗* represents the generalized item at the second branch in level 2 which is a branch of *4*-th branch in level 1.

Here, two sets of clustering attributes are designed for the mining utility and relationship of items, namely, Combination I and Combination II. The suitability of membership functions and the number of frequent fuzzy closed itemsets are the two attributes used to cluster the Pareto solutions for mining the relationships of items in Combination I. In contrast, the suitability of membership functions and the utility of closed fuzzy itemsets are the two attributes used to cluster Pareto solutions to mine the utility relationship of items in Combination II.

Users who select "Combination I" as the set of clustering attributes to obtain representative solutions want to find the relationships among items. The number of fuzzy closed large itemsets is defined as follows:

$$numLarFuzClosedItemSet(ps) = \sum_{k=1}^{levels} |LFCI^k|,$$

where $|LFCI^k|$ is the number of level-k large fuzzy closed itemsets in the Pareto solution ps, $1 \leq k \leq levels$. The large fuzzy closed itemset of each Pareto solution can be calculated by using the fuzzy mining algorithm [14]. Thus, the set $LFCI$ is defined as follows:

$$LFCI = \{X^k | count_x^k \geq \alpha, 1 \leq j \leq m^k, 1 \leq l \leq h, 1 \leq k \leq level\},$$

where X^k is a large fuzzy closed itemset in level k and $count_x^k$ is the fuzzy value of X^k.

Because items have different prices, item prices should be considered to meet business requirement. In such situations, the clustering attributes of "Combination II" can be selected. The representative solutions of clusters can then provide different ways to mine utility fuzzy itemsets. Each item's utility is then taken into consideration in the mining of the Utility of Fuzzy Itemsets (*UFI*) [34]. The formula for the utility of a utility fuzzy itemset is defined as follows:

$$UFI(X) = \sum_{i \in TID(X)} \sum_{R_{jl} \in X} f_{jl}^{(i)} \times WEU(R_{jl}), \tag{6}$$

where *TID(X)* is the set of transaction identifiers for which the cardinality of the fuzzy region of the items in fuzzy itemset X is larger than zero, $f_{jl}^{(i)}$ is the cardinality of the l-th fuzzy region of I_j in the i-th transaction, R_{jl} is the l-th fuzzy region of item I_j, and $WEU(R_{jl})$ is calculated by the maximum value of R_{ji} multiplied by the external utility of item I_j. Thus, the utility fuzzy itemsets are first derived for each chromosome. Then, according to the derived utility fuzzy itemsets, the summation of the utilities of all utility fuzzy itemsets is calculated as a clustering attribute, and incorporated into the membership suitability functions to obtain representative solutions to derive the utility of items.

TABLE 1. The pseudo code of the second stage.

Input	Given a dataset *Trans* with *n* transactions and *m* items, a predefined taxonomy *Tax*, the derived nondominated solutions *NDS*, the number of clusters *numCluster*, the external utility *EU*, the utility threshold of a chromosome *UTC*, the utility threshold *UT*, a support threshold α, and a confidence threshold λ.
Output	The derived actionable patterns *AP*.
1	*UFI* = findUtilityofFuzzyItemsets(*Trans, Tax, NDS, EU,* α);
2	*NDS'* = NormalizeChromosomeAttributes(*NDS*);
3	*clusterResult* = findClusters(*NDS', clusteringAttributes, UTC*); *//Clustering attributes could be Combination I or Combination II.*
4	*repRsult* = getRepresentativeParetoSolutions(*clusterResult*);
5	if (*Combination I*) *AP* = generalizedFuzzyRuleMining(*repRsult,* λ); else *AP* = utilityClosedFuzzyItemsetMining(*repRsult, UT*);

V. PROPOSED ACTIONABLE PATTERN MINING APPROACH

In accordance with the PA-GFM framework, the proposed actionable pattern mining approach is stated here. As mentioned above, the goal of the first stage of the proposed framework is to obtain Pareto solutions using the MOGA-based mining method presented in [7]. The goal of the second phase is to cluster the nondominated solutions using the selected clustering attributes for actionable pattern mining. Because the procedure of the first stage can be found in [7], the pseudo code of the second stage is given in Table 1.

The details of the proposed algorithm for the second stage are stated as follows:

Phase II: Actionable pattern mining approach:

STEP 1: For each chromosome in *NDS*, calculate its utility of fuzzy itemsets (*UFI*) as follows.

SUBSTEP 1.1: Calculate the fuzzy closed large itemset of each chromosome in *NDS* with the fuzzy mining algorithm [14] and collect the derived large fuzzy closed itemsets into a set *LFCI* as follows:

$$LFCI = \{X^k | count_x^k \geqq \alpha,\ 1 \leq j \leq m^k, 1 \leq l \leq h, 1 \leq k \leq level\},$$

where X^k is a large fuzzy closed itemset in level k and $count_x^k$ is the fuzzy value of X^k. Note that different fuzzy data mining algorithms can be used in this step.

SUBSTEP 1.2: Find the *UFI* value of each large fuzzy closed itemset X in the *LFCI* of each chromosome using the following formula:

$$UFI(X^k) = \sum_{i \in TID(X^k)} \sum_{R_{jl} \in X^k} f_{ijl}^{(k)} \times WEU(R_{jl}^k),$$

where R_{jl}^k is the linguistic region of item I_j with the l-th linguistic term at the k-th level and $TID(X^k)$ means the transactions containing X^k.

SUBSTEP 1.3: Set the *UFI* of each chromosome as the summation of all $UFI(X^k)$ in*LFCI*.

STEP 2: Normalize the attributes of each chromosome as follows:

$$\underset{a_i \in A}{Normalize}(a_i) = \frac{a_i - Min(A)}{Max(A) - Min(A)},$$

where A is an attribute with n values a_1 to a_n, and $Max(A)$ and $Min(A)$ are the maximum and minimum values of A, respectively. Here, A may be either the *UFI* or the *suitability* of chromosomes.

STEP 3: Use k-means to partition the chromosomes into *numCluster* groups based on the attributes in STEP 2. The adopted attributes can be either Combination I (*normalized suitability*and*normalized number of large itemsets*) or Combination II (*normalized UFI* and *normalized suitability*) of each Pareto solution. When Combination II is utilized to cluster nondominated points, the utility threshold of the Pareto solution *UTC* can be used to prune the Pareto solutions with low normalized *UFI* values.

STEP 4: Obtain representative Pareto solutions from the groups, where a representative solution is the one closest to the center of its corresponding cluster.

STEP 5: Mine and output the fuzzy association rules at all levels according to the confidence threshold λ, (or mine and output the utility closed fuzzy itemsets according to the *UT* of each representative solution).

In the following, we provide some possible solutions for estimating the number of groups for STEP 3 because different datasets may have different numbers of groups. For estimating the number of groups, Rahman and Islam presented an algorithm, called GenClust, that combined GA and k-means for dealing with clustering problems [29]. Using the GenClust, GA is utilized to search for a suitable number of groups for k-means. Hence, the number of groups does not be set by users. In addition, based on the GenClust, an enhanced algorithm, called GENCLUST++, was provided to speed up the evolution process [16]. For parameter-independent clustering approaches, Rahman *et al.* proposed two density-based clustering (PIDC) algorithms, namely PIDC-WO and PIDC-O, which are used for datasets without and with outliers, respectively [28].

VI. AN EXAMPLE OF THE ACTIONABLE PATTERN MINING PHASE

A simple example is shown below to help clarify the second phase of the framework. The six encoded transactions, the external utility of all items, and the six nondominated solutions derived using the previous approach [7] are shown in Tables 2, 3 and 4, respectively.

STEP 1: The utility of the fuzzy itemsets for each chromosome is found as follows.

TABLE 2. The six encoded transactions.

TID	Items
T_1	(*111*, 1) (*112*, 4) (*211*, 4) (*212*, 6) (*311*, 7) (*421*, 7)
T_2	(*111*, 3) (*112*, 3) (*121*, 1) (*221*, 5) (*222*, 3) (*321*, 4) (*322*, 4)
T_3	(*211*, 7) (*221*,8) (*312*, 5) (*322*, 7)
T_4	(*111*, 2) (*211*, 5) (*312*, 5)
T_5	(*211*, 5) (*222*, 4)
T_6	(*111*, 3) (*112*, 10) (*411*, 3) (*412*, 9)

TABLE 3. External utility of all items.

Item	EU	Item	EU	Item	EU
111	5	*321*	8	*11**	6.5
112	8	*322*	6	*12**	5
121	4	*411*	2	*21**	1
122	6	*412*	8	*22**	7.5
211	1	*421*	2	*31**	6
212	1	*422*	7	*32**	7
221	6	*1***	5.75	*41**	5
222	9	*2***	4.25	*42**	4.5
311	3	*3***	6.5		
312	9	*4***	4.75		

TABLE 4. Six Pareto solutions in this example.

ps_q	Membership functions of all items
ps_1	3.33, 1.00, 6.24, 2.12, 9.00, 3.89, 4.81, 1.85, 7.53, 2.82, 9.20, 1.49, 4.21, 1.23, 5.67, 2.46, 7.21, 3.35, 1.97, 1.42, 4.11, 1.67, 6.69, 2.30;
ps_2	3.44, 1.25, 6.53, 3.12, 9.21, 2.99, 3.49, 1.03, 4.74, 3.68, 9.39, 3.19, 1.74, 1.00, 3.62, 1.89, 7.82, 1.85, 2.36, 1.00, 5.54, 1.92, 8.37, 2.53;
ps_3	3.89, 1.70, 6.47, 2.52, 8.92, 3.69, 3.83, 1.70, 5.55, 3.23, 9.08, 2.75, 2.23, 1.05, 3.81, 2.05, 7.68, 2.86, 2.19, 1.09, 5.46, 2.20, 8.51, 2.79;
ps_4	3.80, 1.60, 6.43, 2.70, 9.08, 3.49, 3.58, 1.44, 5.10, 3.32, 9.06, 2.93, 1.95, 1.02, 3.66, 1.95, 7.64, 2.43, 2.51, 1.16, 5.64, 2.19, 8.61, 2.65;
ps_5	3.87, 1.67, 6.46, 2.57, 8.96, 3.64, 3.77, 1.63, 5.43, 3.25, 9.07, 2.80, 2.16, 1.04, 3.77, 2.02, 7.67, 2.75, 2.27, 1.11, 5.51, 2.19, 8.54, 2.75;
ps_6	3.82, 1.63, 6.44, 2.65, 9.04, 3.55, 3.65, 1.51, 5.21, 3.30, 9.07, 2.88, 2.03, 1.03, 3.70, 1.98, 7.65, 2.54, 2.43, 1.14, 5.59, 2.19, 8.58, 2.69;

TABLE 5. Fuzzy closed itemsets in all levels.

level	Fuzzy closed itemsets
1	(*3**.High*, *2**.High*, *4**.High*, *1**.Middle*), (*3**.Middle*, *2**.High*, *4**.High*, *1**.Middle*), (*3**.High*, *2**.Middle*, *1**.High*), (*3**.High*, *2**.Middle*, *1**.Middle*)
2	(*11*.Middle*, *42*.High*, *31*.High*, *21*.High*), (*11*.Middle*, *42*.High*, *31*.Middle*, *21*.High*), (*11*.Middle*, *32*.High*, *21*.Low*)
3	(*212.Middle*, *211.Low*, *112.Low*, *421.High*, *311.High*), (*212.Middle*, *211.Low*, *112.Low*, *421.High*, *311.Middle*), (*212.Low*, *211.Low*, *112.Low*, *421.High*, *311.High*), (*212.Low*, *211.Low*, *112.Low*, *421.High*, *311.Middle*), (*211.Low*, *112.Low*, *111.Low*, *322.Middle*), (*211.Low*, *112.Low*, *111.Low*, *322.Low*), (*412.High*, *411.Middle*, *112.High*), (*412.High*, *411.Low*, *112.High*), (*211.Middle*, *221.Middle*, *312.High*), (*211.Middle*, *221.Middle*, *312.Middle*), (*211.Middle*, *221.Middle*, *312.Low*), (*321.Middle*, *111.Low*, *322.Middle*), (*321.Middle*, *111.Low*, *322.Low*), (*321.Low*, *111.Low*, *322.Middle*), (*321.Low*, *111.Low*, *322.Low*)

SUBSTEP 1.1: First, the large fuzzy itemsets are found by the fuzzy data mining algorithm in [14]; only the large fuzzy closed itemsets are kept. Taking ps_1 as an example; the large fuzzy closed itemsets in all levels with minimum support of 0.04 are shown in Table 5.

SUBSTEP 1.2: The *UFI* of each large fuzzy closed itemset generated from each chromosome is calculated. Take the large fuzzy closed itemset "(*11∗.Middle*, *32∗.High*, *21∗.Low*)" in ps_1 as an example. Because the large fuzzy closed itemset only appears in $TIDD_1$, $the UFI(11^*.Middle, 32^*.High, 21^*.Low) = 0.88^* WEU(11*.Middle) + 0.76^* WEU(32^*.High) + 0.9^* WEU(21^*.Low) = 0.88^*(6.24^*6.5) + 0.76^*(7.21^*7) + 0.9*(4.81^*1) = 35.69 + 38.36 + 4.33 = 78.38$. The *UFI* of all the large itemsets in level 2 in ps_1 are shown in Table 6. The summary of all *UFIs* at different levels in ps_1 is depicted in Table 7.

TABLE 6. The UFI values of all large itemsets in level 2 in ps_1.

Itemset	*UFI*
(*11*.Middle*, *42*.High*, *31*.High*, *21*.High*)	87.52
(*11*.Middle*, *42*.High*, *31*.Middle*, *21*.High*)	62.48
(*11*.Middle*, *32*.High*, *21*.Low*)	78.38

TABLE 7. The UFIs at different levels in ps_1.

Level	*UFI*
Level 1	417.11
Level 2	228.38
Level 3	654.32

SUBSTEP 1.3: Thus, the total *UFI* of ps_1 is 1299.81 = (417.11+228.38+654.32). In the same way, the total *UFI* and the *suitability* of all Pareto solutions are shown in Table 8.

TABLE 8. Total UFI and suitability of all Pareto solutions.

ps_q	Total *UFI*	*Suitability*
ps_1	1299.81	9.07
ps_2	1364.78	5.97
ps_3	2114.83	6.80
ps_4	1750.29	6.45
ps_5	2122.74	6.78
ps_6	1762.52	6.57

STEP 2: The value of *UFI* and *suitability* of all Pareto solutions are normalized. Take the *suitability* of ps_3 as an example. The normalized *suitability* is $0.27 = ((6.80 - 5.97)/(9.07 - 5.97))$. The normalization of the *UFI* and the *suitability* of all Pareto solutions can be obtained in the same way. Table 9 shows the results.

TABLE 9. Normalized suitability and UFI of Pareto solutions.

ps_q	Normalized *UFI*	Normalized *Suitability*
ps_1	0.0	1.0
ps_2	0.08	0.0
ps_3	0.99	0.27
ps_4	0.55	0.15
ps_5	1.0	0.26
ps_6	0.56	0.19

TABLE 10. Clustering results.

Group	ps_q
G_1	ps_1, ps_2
G_2	ps_3, ps_5
G_3	ps_4, ps_6

TABLE 11. LFCI and normalized UFI in level 1 by the Pareto solution ps_3.

Fuzzy closed large itemset	Normalized *UFI*
(3**.High, 2**.High, 1**.High)	0.77
(3**.High, 2**.High, 4**.Middle, 1**.Middle)	0.53
(3**.High, 2**.High, 4**.High, 1**.Low)	0.67
(3**.High, 2**.High, 4**.High, 1**.Middle)	1.0

TABLE 12. Basic information of the used datasets.

Simulated Dataset		Foodmart dataset	
Number of Levels	3	Number of Levels	4
Number of Terminal Nodes	64	Number of Terminal Nodes	1,559
Transaction Size	10K	Transaction Size	21,557

STEP 3: Assume that the user selected Combination II in this example. The Pareto solutions are divided into groups using k-means with the normalized *UFI* and *suitability* values. If the *numCluster* parameter is set to 3 and *UTC* is set to 0; the clustering results are shown in Table 10.

STEPS 4 to 5: The representative Pareto solutions of the groups are ps_1, ps_3 and ps_6. Assume that the user's goal is to find the valuable patterns that can maximize profits. When UT is 0.3, the derived utility of the fuzzy closed itemsets in level 1 by the membership functions in the representative Pareto solution ps_3 are shown in Table 11.

VII. EXPERIMENTAL EVALUATION

Experimental results for the proposed framework and actionable pattern mining approach are presented in this section. The proposed approach was programmed in Java on a personal computer with an Intel Core i7 CPU @ 2.9 GHz and 4.0 GB of RAM. The initial population size P was 100, the crossover rate p_c was 0.8, and the mutation rate p_m was 0.001. In addition, the minimum support was 0.04 (4%) and the d value in the crossover operator was 0.35 according to [12].

A. DATASET DESCRIPTION

Two datasets, including a simulated one and the Foodmart, used in these experiments are stated in this section. The basic information of the datasets is shown in Table 12.

The simulated dataset is the same as that used in [8]. It contains 64 terminal nodes (items) on level 3, and the numbers of generalized items on levels 2 and 1 are 16 and 4, respectively. Only the terminal nodes could appear in transactions. There were four branches for each internal node in the taxonomy. The proposed algorithm was tested on different numbers of randomly generated transactions. In total, 10,000 transactions were used in the experiments.

As to the Foodmart dataset, it has 21,557 transactions, 1,559 items in level 4 (terminal nodes), 110 generalized items in level 3, 49 generalized items in level 2, and three generalized items in level 1. The generalized items in level 1 are drink, food, and non-consumables, with 7, 27, and 15 branches, respectively.

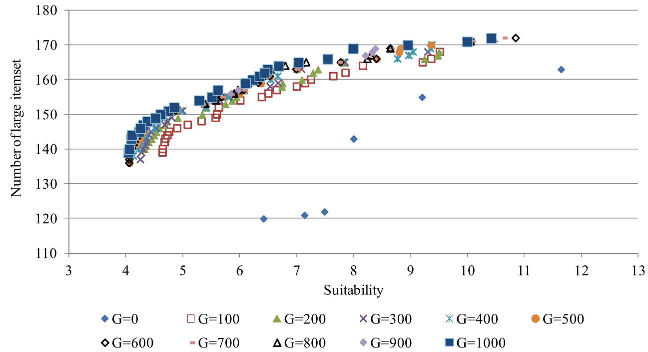

FIGURE 4. The Pareto fronts for different generations.

B. THE EVOLUTION OF PARETO FRONT IN THE FIRST PHASE

First, the experiments were constructed to show the Pareto fronts at different generations on the simulated dataset. The Pareto fronts for generations 0 to 1000 are shown in Fig. 4.

As Fig. 4 shows, the Pareto solutions improved steadily in throughout the generations. The final results at 1000 generations outperform those of earlier generations. In addition, many Pareto solutions were found—in this case, 27 Pareto solutions—which means that it might not be easy to choose an appropriate one to mine actionable patterns. Thus, the experiments presented in the next section were conducted to show the merits of the second phase of the proposed PA-GFM framework.

C. EVALUATION OF THE CLUSTERING RESULTS IN THE SECOND PHASE

As shown above, after the first phase of PA-GFM, a set of Pareto solutions is obtained, each of which represents a nondominated tradeoff. When too many Pareto solutions are found, users will have difficulty selecting the appropriate solution to extract the information they need to make decisions. Thus, the goal of the second phase is to find representative solutions so that decision-makers can employ them efficiently. The k-means clustering strategy is used to achieve this goal. To evaluate the quality of the derived clusters, the similarity of two Pareto solutions is first defined as follows:

$$Sim(ps_q, ps_h) = \frac{sameItemset(ps_q, ps_h)}{MaxNumberofItemset(ps_q, ps_h)},$$

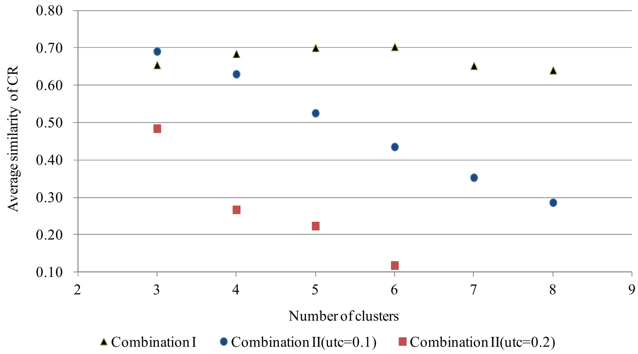

FIGURE 5. The relationships between the average similarity of clustering results and the number of clusters.

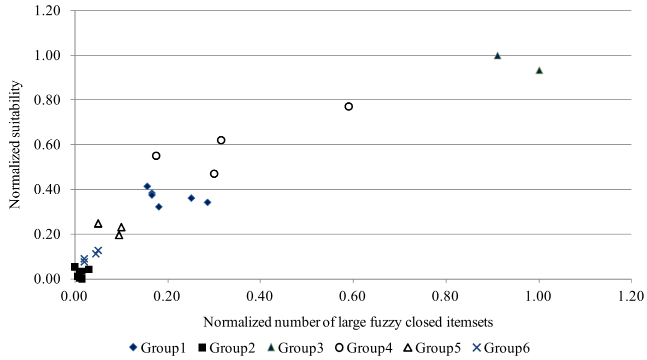

FIGURE 6. The groups formed based on Combination I ($k = 6$).

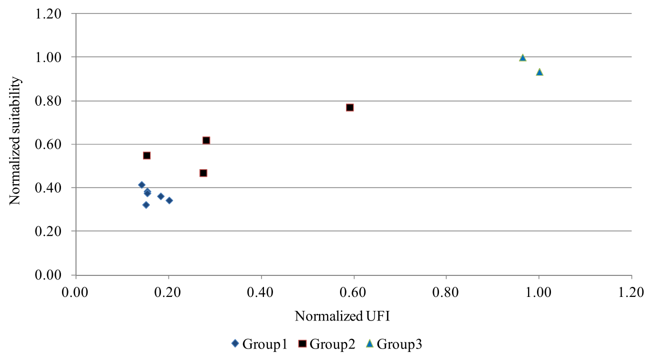

FIGURE 7. The groups formed based on Combination II ($k = 3$).

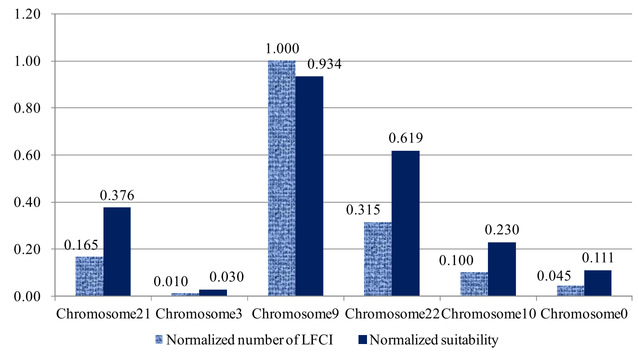

FIGURE 8. The normalized number of large fuzzy closed itemsets and the suitability of representative solutions by Combination I on the simulated dataset.

where $sameItemset(ps_q, ps_h)$ means that some number of the same large itemsets appears in both the Pareto solutions ps_q and ps_h and $MaxNumberofItemset(ps_q, ps_h)$ is the maximum number of large itemsets between chromosomes ps_q and ps_h. The average similarity of the Pareto solutions in the same cluster is defined as follows:

$$avgSimCluster(Cluster_i) = \sum_{C_q, C_h \in Cluster_i, q \neq h} \frac{2 * Sim(ps_q, ps_h)}{|Cluster_i| * |Cluster_i - 1|}.$$

Thus, given a clustering result CR, the average similarity of all clusters is defined as follows:

$$avgSimAllCluster(CR) = \sum_{i=1 \ to \ k,\ CLuster_i \in CR} \frac{avgSimCluster(Cluster_i)}{|CR|}.$$

The experiments were then conducted on the simulated dataset as follows. The Pareto solutions obtained previously were partitioned into clusters by k-means with the designed clustering attributes (Combination I or II) to obtain the representative solutions. Using the similarity measurement, the ten-run average results of the relationships between the average similarity of clustering results and the number of clusters are depicted in Fig. 5.

From Fig. 5, for Combination I, we can see that the average similarities of clusters increase as the number of clusters increases until the number of clusters reaches six. This indicates that the derived clustering results achieve the best similarity when the number of clusters is set to six. For Combination II with different utility thresholds, the best average cluster similarities for both versions appear with three clusters. This phenomenon occurs because some solutions with low normalized UFI were deleted by the utility threshold; thus degeneracy may occur as the number of clusters increases. The clustering results with the highest average similarity for Combination I and Combination II ($utc = 0.1$) are shown in Figs. 6 and 7, respectively.

Fig. 6 shows that six clusters were formed from the non-dominated solutions. The 27 Pareto solutions derived from the first stage are now reduced to only six representative solutions that provide a more effective way for users to mine actionable patterns involving relationships among items and to make decisions.

Fig. 7 shows that the 27 Pareto solutions obtained from the first phase were clustered into 3 groups when using the combination II as clustering attributes. The three representative solutions in the clusters are best suited to helping decision-makers analyze the profits among items and make actionable promotions efficiently.

To show more detailed results from the second phase, the normalized number of large fuzzy closed itemsets (LFCI) and suitability and the normalized UFI and suitability of the representative chromosomes derived by

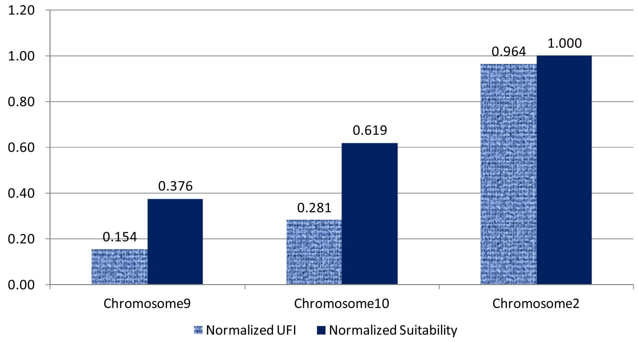

FIGURE 9. The normalized UFI and the suitability of representative solutions by Combination II on the simulated dataset.

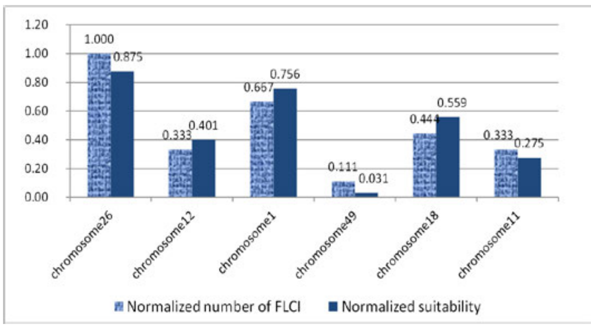

FIGURE 10. The representative chromosomes by Combination I on the Foodmart dataset.

Combinations I and II are shown in Figs. 8 and 9, respectively. The values of the two attributes in Combination I for the six representative points are shown in Fig. 8.

Fig. 8 shows that, when the derived representative solutions have large normalized suitability values, their normalized numbers of LFCI are also large (e.g., the representative solution of chromosome 9). Alternatively, although representative solution 3 has the best normalized suitability value, it has the smallest normalized number of LFCI. From these results, clearly, the six derived representative solutions represent trade-offs between objective functions and make it easier for users to mine actionable fuzzy generalized association rules.

Fig. 9 shows that the larger the suitability values of the representative chromosomes are, the larger the UFI values are. Thus, when users need to obtain more actionable patterns, representative solution 2 (chromosome 2) could be suggested to achieve that goal. Experiments were then made on different data sizes to show the execution time of the proposed approach. The results are shown in Table 13.

From Table 13, we can observe that the execution time of the first stage was larger than that of the second stage when the population size was set at 50. Because the first stage is used to derive the Pareto solutions by an evolutionary algorithm, the results are reasonable. Although the execution is a little time-consuming, it is increasing linearly along with the increasing of transaction sizes.

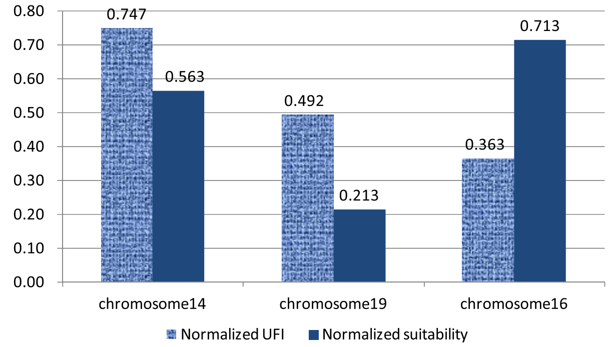

FIGURE 11. The representative chromosomes by Combination II on the Foodmart dataset.

TABLE 13. Execution time on simulated datasets with different transaction sizes.

	Execution Time (sec.)		
Size	Stage I	Stage II	Total
10k	31608.9	8318.9	39927.8
30k	87422.3	28403.5	115825.8
50k	147289.8	40778.2	188068.0

In the same process, the representative chromosomes derived by the proposed approach with Combinations I and II on the Foodmart dataset are shown in Fig. 10 and Fig. 11.

From Fig. 10, we can observe that the larger the suitability values of the representative chromosomes are, the better the number of LFCI of chromosomes are.

From Fig. 11, when Combination II was used in the second stage, large suitability values basically could generate large utility values excepting C_{16}. Thus, if user wants to mine more fuzzy rules or utility fuzzy itemsets, the chromosomes C_{26} and C_{14} could be provided for the actionable pattern mining tasks.

Notice that because the proposed PA-GFM framework consists of two phases, named nondominated membership functions mining and actionable pattern mining phases, in order to mine more useful information, the objective functions used in the first phase and clustering attributes employed in the second phase should be related. Otherwise, some derived good solutions by the first phase may be removed in the second phase. In other words, the utility of the representative chromosomes could be decreased. A possible solution for avoiding this problem is using predefined conditions to follow between the objective functions and the clustering attributes in the two phases. For instance, a condition could be "If objective functions fun_A and fun_B are selected in the first phase, then clustering attributes $Attri_C$ and $Attri_D$ should be used to find the representative solution for mining actionable patterns".

D. THE EVALUATION OF THE DERIVED ACTIONABLE PATTERNS

Finally, the representative solutions derived by the actionable pattern mining approach on the simulated dataset are

TABLE 14. The derived number of fuzzy rules at different levels.

ps_q	Min. sup. = 0.04 Level 1	Level 2	Level 3
ps_3	11 (avg. sup. = 0.208) (avg. conf. = 0.306)	85 (avg. sup. = 0.068) (avg. conf. = 0.248)	5 (avg. sup. = 0.043) (avg. conf. = 0.249)
ps_9	14 (avg. sup. = 0.223) (avg. conf. = 0.354)	842 (avg. sup. = 0.064) (avg. conf. = 0.355)	89 (avg. sup. = 0.048) (avg. conf. = 0.267)
ps_{22}	15 (avg. sup. = 0.207) (avg. conf. = 0.372)	312 (avg. sup. = 0.064) (avg. conf. = 0.309)	31 (avg. sup. = 0.051) (avg. conf. = 0.289)

TABLE 15. The rules with the highest confidence values derived by the three Pareto solutions at different levels.

(a) Pareto Solution ps_3 (Solution with the Lowest Suitability)		
	Rule	Conference
Level 1	2**.3 => 3**.3	0.798
Level 2	11*.3, 14*.3 => 32*.3	0.550
Level 3	111.1 => 114.1	0.252

(b) Pareto Solution ps_{22} (Intermediate Solution)		
	Rule	Conference
Level 1	2**.3 => 3**.3	0.875
Level 2	11*.3, 14*.3, 24*.3 => 32*.3	0.767
Level 3	322.2 => 323.2	0.375

(c) Pareto Solution ps_9 (Solution with the Highest Itemsets Number)		
	Rule	Conference
Level 1	2**.3 => 3**.3	0.907
Level 2	11*.3, 14*.3, 31*.2 => 32*.3	0.892
Level 3	322.2 => 323.3	0.357

provided to reveal the deeper information. For Combination I, three representative solutions, ps_3, ps_9 and ps_{22}, are selected for mining actionable patterns in terms of the number of fuzzy generalized association rules. Note that Pareto solution ps_3 has the best normalized suitability, ps_9 has the largest normalized number of LFCI, and the normalized number of LFCI and suitability of ps_{22} are between those of ps_3 and ps_9. The results are listed in Table 14.

Table 14 shows the number of rules derived by three representative solutions at different levels. At Level 1, the numbers of generalized rules derived the three solutions are similar. At Level 2, using ps_9, the actionable pattern mining approach mines the largest number of rules and achieves higher confidence values than when using ps_3 and ps_{22}. In contrast, the number of rules derived by the actionable pattern mining approach using ps_3 is smaller than ps_9 and ps_{22}. Thus, when users need more information about item relationships, Pareto solution ps_9 could be suggested to mine actionable knowledge for further analysis. Alternatively, Pareto solution ps_3 could be suggested to mine actionable knowledge when users do not want too much information. Below, the rules with the highest confidence values derived by the three Pareto solutions at different levels are illustrated in Table 15.

From Table 15, it is verified again that the derived rules basically have the high confidence values when the Pareto solution ps_9 is used. In addition, the three Pareto solutions ps_3, ps_{22} and ps_9 at Level 2 are shown below:

$Rule_{ps3}$: 11∗.3, 14∗.3 => 32∗.3 (conf. = 0.55),
$Rule_{ps22}$: 11∗.3, 14∗.3, 24∗.3 => 32∗.3 (conf. = 0.767),
$Rule_{ps9}$: 11∗.3, 14∗.3, 31∗.2 => 32∗.3 (conf. = 0.892).

According to the rules derived, it can be observed that no matter which Pareto solution is employed to mine rules, they all can get the rule "11∗.3, 14∗.3 => 32∗.3". Their difference is extra items could be found when the Pareto solutions ps_{22} and ps_9 are utilized. This observation also indicates that the Pareto solutions with a high normalized number of LFCI can be chosen when users want to get more item relationships. Otherwise, the Pareto solutions with a good normalized suitability values can be employed to derive few but valuable rules.

The derived numbers of utility fuzzy closed itemsets at different levels of representative solutions are shown in Table 16 for Combination II when the utility threshold was set to 0.1.

TABLE 16. The derived number of utility fuzzy closed itemsets at different levels.

ps_q	ut = 0.1 Level 1	Level 2	Level 3
ps_9	2 (avg. nor. ufi. = 0.935)	43 (avg. nor. ufi. = 0.446)	0 (avg. nor. ufi. = 0)
ps_{10}	2 (avg. nor. ufi. = 0.828)	38 (avg. nor. ufi. = 0.449)	19 (avg. nor. ufi. = 0.536)
ps_2	2 (avg. nor. ufi. = 0.817)	109 (avg. nor. ufi. = 0.404)	58 (avg. nor. ufi. = 0.389)

Table 16 shows that when Pareto solution ps_2 is utilized to derive actionable knowledge, it can mine a larger number of utility fuzzy closed itemsets than ps_9 and ps_{10} at Levels 1, 2 and 3. Additionally, when Pareto solution ps_9 is utilized to derive actionable knowledge, the number of utility fuzzy closed itemsets is smaller than that of the other two Pareto solutions at Level 3. In other words, when users want to know more actionable utility itemsets, Pareto solution ps_2 could be utilized to derive utility itemsets and help users make actionable promotions. Alternatively, Pareto solution ps_9 or ps_{10} could be used to mine the utility itemsets.

At last, the representative chromosomes obtained by the proposed approach with Combination I on the Foodmart dataset were used for mining fuzzy rules. Three Pareto solutions, ps_{26}, ps_{49} and ps_{12}, are selected from Fig. 10 for deriving fuzzy rules at different levels, where ps_{49} has the best normalized suitability, ps_{26} has the largest normalized number of LFCI, and the normalized number of LFCI and suitability of chromosome ps_{12} is between those of ps_{26} and ps_{49}. The results are shown in Table 17.

From Table 17, we can observe that the number of rules at levels 2 and 3 derived by ps_{26} are larger than the other two chromosomes. We can also find that the average support and confidence values of the rules derived by ps_{26} are bigger than or almost near the same as those of the other two

TABLE 17. The derived numbers of fuzzy rules at different levels on the foodmart dataset.

	Min. sup. = 0.001		
ps_q	Level 1	Level 2	Level 3
ps_{49}	33 (avg. sup. = 0.015) (avg. conf. =0.154)	562 (avg. sup. = 0.003) (avg. conf. = 0.078)	357 (avg. sup. = 0.002) (avg. conf. = 0.052)
ps_{26}	15 (avg. sup. = 0.03) (avg. conf. = 0.201)	1654 (avg. sup. = 0.03) (avg. conf. = 0.103)	586 (avg. sup. = 0.002) (avg. conf. = 0.06)
ps_{12}	27 (avg. sup. = 0.021) (avg. conf. = 0.182)	728 (avg. sup. = 0.003) (avg. conf. = 0.093)	284 (avg. sup. = 0.002) (avg. conf. = 0.056)

chromosomes. On the contrary, ps_{49} has a smaller number than or almost near the same number of rules as the other two. Thus, when the users need more information about the relationship of items, ps_{26} could be suggested to mine rules and provide results to them for further analysis. Besides, ps_{12} could be suggested to discover rules when users want concise information.

From these experimental results, we can conclude that the proposed PA-GFM framework can provide an effective and useful way for users to derive Pareto solutions to discover actionable patterns such as fuzzy generalized association rules and fuzzy utility itemsets for decision-makers to make appropriate business plans.

E. DISCUSSIONS

In this section, the time complexity, space complexity and limitations of the proposed framework are discussed. As mentioned in a previous section, the PA-GFM framework consists of two phase that are Pareto solution mining and actionable pattern mining.

In the first stage, the genetic-fuzzy approach is utilized to derive appropriate MFs. The most time-consuming part is the process of chromosome evolution which is calculated by the number of large generalized items and the suitability of MFs, and the time needed for finding the number of large generalized items is larger than calculating the suitability of MFs. Thus, when the number of transactions is n, the number of items at each certain level is m, the number of levels in the taxonomy is t, and the number of linguistic terms of each item is l, the worst time complexity for calculating the number of large generalized items is $O(n^*m^*t^*l)$. The actual execution time, of course, will depend on the given dataset. Let the time complexity for evaluating a chromosome $O(n^*m^*t^*l)$ as *chromoTime*, the population size as *pSize* and number of generations as *numGeneration*, the time complexity of the first stage is thus $O(numGeneration^*pSize^*chromoTime)$. In the second stage, the algorithm utilizes a clustering approach to find representative solutions, and the fuzzy rule or utility mining algorithm is then applied to the dataset to obtain actionable patterns. For the k-means clustering technique used in the proposed framework, assume the number of instances is *numIns* and the number of iterations is *numIter*. The time complexity for assigning instances into groups is $O(numIter^*k^*numIns)$. From the time complexities of the two stages, we can know that the time complexity of the proposed approach is $O(numGeneration^*pSize^*chromoTime)$.

Let the length of each transaction is *len*, the space complexity of the first stage is $O(n^*len + pSize^*m^*t^*l)$, where n^*len and $pSize^*m^*t^*l$ are spaces needed for a given transaction dataset and a population. In the second stage, since the k-means algorithm is used to divide solutions into clusters, the space complexity of the second stage is $O(numInstances+k)$, where *numInstances*and $k$are the number of instances and number of clusters, respectively. Thus, from the space complexities of the two stages, we can also know that the space complexity of the proposed approach is $O(n^*len + pSize^*m^*t^*l)$.

As to the two mining algorithms, we can know that the most time-consuming part is to derive fuzzy large itemsets. To deal with this problem, MapReduce-based algorithms can be employed to improve the efficiency [25], [31]. For example, Martín *et al.* presented a generic MapReduce framework for rule discovery [25], and Singh *et al.* proposed a MapReduce-based Apriori algorithm for performance optimization on a Hadoop cluster [31].

Based on the above analysis, the limitations of the proposed framework are listed below: (1) The memory needs to be enough to load a given dataset for analysis. When the memory is not enough, some approaches could also be utilized to handle it. For example, Nguyen and Orlowska presented a partition-based approach, named the PartitionSP algorithm, for performance improvement when dataset is very large [26]. (2) Users should make sure there are no missing values in given transactions. When the missing values are appeared in the dataset, those values should be removed or recovered by missing data recovery approaches. For instance, Liu and Dai proposed an information decomposition imputation approach for missing value recovery using fuzzy membership functions [23].

VIII. CONCLUSIONS AND FUTURE WORK

This paper proposes a Post-Analysis-based Genetic-Fuzzy Mining (PA-GFM) framework by combining domain-driven data mining, fuzzy data mining and genetic algorithms to find actionable patterns. PA-GFM involves two phases: Pareto solution mining and actionable pattern mining. In the first phase, the nondominated solutions (NDS) with given objective functions are found by a MOGA-based approach in which each chromosome represents a potential set of membership functions. Then, in the second phase, representative solutions are derived from the NDS using the two sets of designed clustering attributes, named Combinations I and II. Then, the two types of actionable patterns (fuzzy generalized association rules and fuzzy utility itemsets) are mined from the representative solutions. The performance of the PA-GFM framework is verified through experimental results on the simulated and the Foodmart datasets. The experimental results show that the proposed framework is useful because it adopts only the representative solutions to mine actionable knowledge based on the users' preferences. In addition, we have made experiments on the Foodmart dataset to show a practical application of the propose approach.

Future work can adopt the following directions: (1) In the first phase, the effectiveness of the proposed framework may be improved by utilizing other multiobjective genetic algorithms; (2) In the second phase—the actionable pattern mining phase—different mining strategies (e.g., level-by-level or cross-level mining strategies) could be utilized to derive various types of actionable knowledge; (3) In addition, the PA-GFM framework could be improved by designing a fine-tuning approach using an iterative learning algorithm; (4) Besides, the other ADK frameworks could also be employed to mine actionable patterns. For instance, the multisource combined-mining-based AKD can be utilized when the proposed approach takes transactions from multiple sources into consideration for deriving actionable patterns; (5) Combining other various mining algorithms to the proposed framework to derive more actionable knowledge. For example, classification mode can be constructed using association-rule analysis; (6) We will try to find possible ways to collaborate with retail stores and design the fair criteria to compare the proposed framework with other existing approaches in the future.

ACKNOWLEDGMENT

This is a modified and expanded version of the paper "A two-stage multi-objective genetic-fuzzy mining algorithm," presented at *the 2013 IEEE Symposium Series on Computational Intelligence* [7].

REFERENCES

[1] R. Alhajj and M. Kaya, "Multi-objective genetic algorithms based automated clustering for fuzzy association rules mining," *J. Intell. Inf. Syst.*, vol. 31, no. 3, pp. 243–264, Dec. 2008.

[2] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *Proc. Int. Conf. Very Large Databases*, Sep. 1994, pp. 487–499.

[3] A. M. AlMana and M. S. Aksoy, "An overview of inductive learning algorithms," *Int. J. Comput. Appl.*, vol. 88, no. 4, pp. 20–28, Feb. 2014.

[4] L. Cao, P. S. Yu, C. Zhang, and Y. Zhao, *Domain Driven Data Mining*. Boston, MA, USA: Springer, 2010.

[5] L. Cao, "Actionable knowledge discovery and delivery," in *WIREs Data Mining Knowledge Discovery*, vol. 2, no. 2. London, U.K.: Springer, 2012, pp. 149–163.

[6] C.-H. Chen, T.-P. Hong, and V. S. Tseng, "Finding Pareto-front membership functions in fuzzy data mining," *Int. J. Comput. Intell. Syst.*, vol. 5, no. 2, pp. 343–354, Apr. 2012.

[7] C.-H. Chen, J.-S. He, and T.-P. Hong, "MOGA-based fuzzy data mining with taxonomy," *Knowl.-Based Syst.*, vol. 54, pp. 53–65, Dec. 2013.

[8] C.-H. Chen, J.-S. He, and T.-P. Hong, "A two-stage multi-objective genetic-fuzzy mining algorithm," in *Proc. IEEE Int. Workshop Genetic Evol. Fuzzy Syst.*, Apr. 2013, pp. 16–20.

[9] K. Deb, S. Agarwal, A. Pratap, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[10] C. M. Fonseca and P. J. Fleming, "Genetic algorithms for multiobjective optimization: Formulation discussion and generalization," in *Proc. Int. Confidence Genetic Algorithms*, Jun. 1993, pp. 416–423.

[11] A. Gammerman, V. Vovk, and V. Vapnik, "Learning by transduction," in *Proc. 14th Conf. Uncertainty Artif. Intell.*, Jul. 1998, pp. 148–155.

[12] F. Herrera, M. Lozano, and J. L. Verdegay, "Fuzzy connectives based crossover operators to model genetic algorithms population diversity," *Fuzzy Sets Syst.*, vol. 92, no. 1, pp. 21–30, Nov. 1997.

[13] 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 Comput.*, vol. 10, no. 11, pp. 1091–1101, Sep. 2006.

[14] T.-P. Hong, C.-S. Kuo, and S.-C. Chi, "Mining association rules from quantitative data," *Intell. Data Anal.*, vol. 3, no. 5, pp. 363–376, Nov. 1999.

[15] T.-P. Hong, K.-Y. Lin, and B.-C. Chien, "Mining fuzzy multiple-level association rules from quantitative data," *Appl. Intell.*, vol. 18, no. 1, pp. 79–90, Jan. 2003.

[16] M. Z. Islam, V. Estivill-Castro, M. A. Rahman, and T. Bossomaier, "Combining k-means and a genetic algorithm through a novel arrangement of genetic operators for high quality clustering," *Expert Syst. Appl.*, vol. 91, pp. 402–417, Jan. 2018.

[17] M. Kaya and R. Alhajj, *Effective Mining of Fuzzy Multi-Cross-Level Weighted Association Rules* (Lecture Notes in Computer Science), vol. 4203. Berlin, Germany: Springer, 2006, pp. 399–408.

[18] M. Kaya, "Multi-objective genetic algorithm based approaches for mining optimized fuzzy association rules," *Soft Comput.*, vol. 10, no. 7, pp. 578–586, May 2006.

[19] N. Kalanat and B. Minaei-Bidgoli, "An optimized fuzzy method for finding actions," *J. Intell. Fuzzy Syst.*, vol. 30, no. 1, pp. 257–265, Sep. 2015.

[20] A. Ghosh and R. K. De, "Fuzzy correlated association mining: Selecting altered associations among the genes, and some possible marker genes mediating certain cancers," *Appl. Soft Comput.*, vol. 38, pp. 587–605, Jan. 2016.

[21] Y.-C. Lee, T.-P. Hong, and T.-C. Wang, "Multi-level fuzzy mining with multiple minimum supports," *Expert Syst. Appl.*, vol. 34, no. 1, pp. 459–468, Jan. 2008.

[22] C. Lai, P. Chung, and V. S. Tseng, "A novel algorithm for mining fuzzy high utility itemsets," *Int. J. Innov. Comput. Inf. Control*, vol. 6, no. 10, pp. 736–741, Oct. 2010.

[23] S. Liu and H. Dai, "Examination of reliability of missing value recovery in data mining," in *Proc. IEEE Int. Conf. Data Mining Workshop*, Dec. 2014, pp. 306–313.

[24] S. G. Matthews, M. A. Gongora, A. A. Hopgood, and S. Ahmadi, "Web usage mining with evolutionary extraction of temporal fuzzy association rules," *Knowl.-Based Syst.*, vol. 54, pp. 66–72, Dec. 2013.

[25] D. Martín, M. Martínez-Ballesteros, D. García-Gil, J. Alcalá-Fdez, F. Herrera, and J. C. Riquelme-Santos, "MRQAR: A generic MapReduce framework to discover quantitative association rules in big data problems," *Knowl.-Based Syst.*, vol. 153, pp. 176–192, Aug. 2018.

[26] S. N. Nguyen and M. E. Orlowska, "A partition-based approach for sequential patterns mining," in *Proc. IEEE Int. Conf. Res., Innov. Vis. Future*, Mar. 2007, pp. 200–205.

[27] A. M. Palacios, J. L. Palacios, L. Sánchez, and J. Alcalá-Fdez, "Genetic learning of the membership functions for mining fuzzy association rules from low quality data," *Inf. Sci.*, vol. 295, pp. 358–378, Feb. 2015.

[28] M. A. Rahman, K. L.-M. Ang, and K. P. Seng, "Unique neighborhood set parameter independent density-based clustering with outlier detection," *IEEE Access*, vol. 6, pp. 44707–44717, 2018.

[29] M. A. Rahman and M. Z. Islam, "A hybrid clustering technique combining a novel genetic algorithm with K-means," *Knowl.-Based Syst.*, vol. 71, pp. 345–365, Nov. 2014.

[30] M. Štěpnička, M. Burda, and L. Štěpničková, "Fuzzy rule base ensemble generated from data by linguistic associations mining," *Fuzzy Sets Syst.*, vol. 285, no. 15, pp. 140–161, Feb. 2016.

[31] S. Singh, R. Garg, and P. K. Mishra, "Performance optimization of MapReduce-based Apriori algorithm on Hadoop cluster," *Comput. Electr. Eng.*, vol. 67, pp. 348–364, Apr. 2018.

[32] F. Thabtah, "A review of associative classification mining," *Knowl. Eng. Rev.*, vol. 22, no. 1, pp. 37–65, Mar. 2007.

[33] G. Williams, "Descriptive and predictive analytics," in *Data Mining with Rattle R: The Art Excavating Data for Knowledge Discovery*. New York, NY, USA: Springer, 2011, pp. 171–177.

[34] C.-M. Wang, S.-H. Chen, and Y.-F. Huang, "A fuzzy approach for mining high utility quantitative itemsets," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Aug. 2009, pp. 1909–1913.

[35] Z. Wang, Z. Yu, C. L. P. Chen, J. You, T. Gu, H.-S. Wong, and J. Zhang, "Clustering by local gravitation," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1383–1396, May 2018.

[36] Z. Yu, H. Chen, J. You, J. Liu, H.-S. Wong, G. Han, and L. Li, "Adaptive fuzzy consensus clustering framework for clustering analysis of cancer data," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 12, no. 4, pp. 887–901, Jul./Aug. 2015.

[37] Z. Yu, H. Chen, J. You, G. Han, and L. Li, "Hybrid fuzzy cluster ensemble framework for tumor clustering from biomolecular data," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 10, no. 3, pp. 657–670, May 2013.

[38] Z. Yu, L. Li, J. Liu, J. Zhang, and G. Han, "Adaptive noise immune cluster ensemble using affinity propagation," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 12, pp. 3176–3189, Dec. 2015.

[39] Z. Yu, H.-S. Wong, D. Wang, and M. Wei, "Neighborhood knowledge-based evolutionary algorithm for multiobjective optimization problems," *IEEE Trans. Evol. Comput.*, vol. 15, no. 6, pp. 812–831, Dec. 2011.

[40] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm for multiobjective optimization," in *Proc. Evol. Methods Design, Optim. Control App. Ind. Problems*, 2001, pp. 95–100.

CHUN-HAO CHEN received the Ph.D. degree in computer science and information engineering from National Cheng Kung University, Taiwan, in 2008. After that, he joined the Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung, Taiwan, as a Postdoctoral Fellow, in 2009. From 2010 to 2013 and 2013 to 2017, he served as an Assistant and Associate Professor of the Department of Computer Science and Information Engineering, Tankang University, respectively. He is currently a Professor with the Department of Computer Science and Information Engineering, Tamkang University, Taiwan. He has published more than 100 research papers in refereed journals and international conferences. He has a wide variety of research interests including data mining, time series, machine learning, evolutionary algorithms, and fuzzy theory. His research topics include portfolio selection, trading strategy, business data analysis, and time series pattern discovery.

JI-SYUAN HE received the B.S. and M.S. degrees from the Department of Computer Science and Information Engineering, Aletheia University and Tamkang University, Taiwan, in 2010 and 2013. He has been a Product Developer with Titansoft, since 2013. His research interests include genetic algorithms, data mining, and fuzzy theory.

TZUNG-PEI HONG received the B.S. degree in chemical engineering from National Taiwan University, in 1985, and the Ph.D. degree in computer science and information engineering from National Chiao-Tung University, in 1992. He served at the Department of Computer Science, Chung-Hua Polytechnic Institute, from 1992 to 1994, and at the Department of Information Management, I-Shou University, from 1994 to 2001. He was in charge of the whole computerization and library planning for the National University of Kaohsiung in Preparation, from 1997 to 2000, and served as the First Director of the Library and Computer Center, National University of Kaohsiung, from 2000 to 2001, the Dean of Academic Affairs, from 2003 to 2006, the Administrative Vice President, from 2007 to 2008, and the Academic Vice President, in 2010. He is currently a Distinguished and Chair Professor with the Department of Computer Science and Information Engineering and the Department of Electrical Engineering, and the Director of the AI Research Center, National University of Kaohsiung, Taiwan. He is also a joint Professor at the Department of Computer Science and Engineering, National Sun Yat-sen University, Taiwan. He has published more than 600 research papers in international/national journals and conferences and has planned more than 50 information systems. He is also the board member of more than 40 journals and the program committee member of more than 500 conferences. His current research interests include knowledge engineering, data mining, soft computing, management information systems, and www applications. He received the First National Flexible Wage Award from the Ministry of Education, Taiwan

SUBBAIYA RAMMOHAN KANNAN received the Ph.D. degree from the IIT (www.iitm.ac.in) Madras, India, and PDF at DISI (www.disi.unige.it), University of Genova, Genova, Italy. He has received postdoctoral fellowship from National Cheng Kung University (web.ncku.edu.tw), Taiwan. He is currently a Professor with the Department of Mathematics, Pondicherry University (A Central University of India), India. He had been awarded a grant in the framework of a joint agreement between the Direzione Generale per la Cooperazione allo Sviluppo of the Italian Ministry of Foreign Affairs and the ICTP Programme for Training and Research in Italian Laboratories (www.ictp.it). He had been invited by the Director General, National Agriculture Research Center, Tsukuba, Japan, for the joint research work on remote sensing data to estimate rice yield (http://narc.naro.affrc.go.jp/narc-e/index.html). He has joint collaborative research projects with NSC Taiwan and MOST Israel. He has received number of research grants from Indian funding agencies, UGC India, CSIR India, and DST India and research grants from foreign funding agencies, NSC Taiwan, MOST Israel, ICTP Italy, External Affair Ministry Italy, and NARC Japan. He has been served as an editorial board member for many scientific journals.