Received July 24, 2019, accepted August 2, 2019, date of publication August 5, 2019, date of current version September 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2933361

Incremental Fuzzy Association Rule Mining for Classification and Regression

LING WANG, QIAN MA, AND JIANYAO MENG
School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China
Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, University of Science and Technology Beijing, Beijing 100083, China

Corresponding author: Ling Wang (lingwang@ustb.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61572073, in part by the National Key Research and Development Program of China under Grant 2017YFB0306403, and in part by the Graduate Education Development Funds for University of Science and Technology Beijing.

ABSTRACT The aim of mining fuzzy association rules is to find both the association and the casual relationships between the itemsets. With the arrival of dynamic data, the fuzzy association rules should be updated in real time. However, most of the existing algorithms must remine the updated database and can only be applied in classification. This paper proposes an incremental fuzzy association rule mining algorithm to solve classification and regression problems. First, the sliding window is adopted to divide the fuzzy dataset. Second, the dynamic fuzzy variable selection algorithm is adopted to select variables for reducing the search space of the fuzzy association rule mining. Finally, in each sliding window, the result of variable selection is used to incrementally mine the causal fuzzy association rules with the fuzzy Eclat algorithm. When new data are added, the process judges whether concept drift occurs, and if so, the rule set is updated; otherwise, the original rule set is still applied. The weights of the rules are calculated to find the evolving relationship. The simulation result shows that this algorithm can improve accuracy and efficiency.

INDEX TERMS Fuzzy association rules, incremental, classification, regression, Eclat algorithm.

I. INTRODUCTION

In data mining, the technique of association rule mining (ARM) aims to find frequently occurring data items with minimum support and minimum confidence constraints, thus discovering the associations existing among data items without any predetermined target. Several studies in the data mining community have shown that classification based on associations rule mining (as known as associative classification mining (ACM)) is able to build accurate classifiers [1]–[3], which involves the prediction of a categorical (discrete unordered) label in the consequent of the rules and is comparable to traditional methods such as decision trees, rule induction and probabilistic approaches. Classification based on associations (CBA) [1] supplies the first algorithm that integrates association rule mining and classification and uses the apriori approach to discover the association rules. The best association rules within any of targeted rules are selected based on the confidence, support and size of antecedent. Classification based on multiple association rules (CMAR) [2] uses the FP-growth approach to find the association rules. The classification rules are stored in a prefix tree structure, known as a CR tree. The CR tree offers effective storage and rapid retrieval of rules in the classifier. Classification based on predictive association rules (CPAR) [4] is a greedy associative classification approach. The best rule is measured by the FOILgain of the rules generated among the available rules in the dataset. Associative classifiers have been found [1], [2], [5] to work better than other traditional classifiers in terms of accuracy and interpretability.

However, associative classifiers focus on a special subset of association rules whose right-hand sides are restricted to the categorical attribute, which cannot predict the continuous attribute. In fact, practitioners and researchers intend to explore these rules to predict quantitative output (future behavior of certain variables) based on selected other known variables. Thus far, the mined association rules are not used to solve such regression estimation problems by mining the relationship between variables. Considering the aim for classification and regression, we strive to acquire the association rules that clearly display the potential information in the

The associate editor coordinating the review of this article and approving it for publication was Sing Kiong Nguang.

dataset according to the causal relations between the variables, while the redundant and nonpredictive rules are pruned and removed.

To address the quantitative attributes in mining association rules, the classical association rule mining algorithms discretize the continuous attributes into several intervals as a Boolean vector, but it is possible to encounter the sharp boundary problem. Fuzzy association rules mining approaches based on the fuzzy set concept [6]–[9] are proposed to overcome such disadvantages. These approaches are based on fuzzy extensions of the classical association rules mining by defining the support and confidence of the fuzzy rule. In recent years, a lot of researchers have focused attention on the use of fuzzy association rules to realize classification. In [6], a novel associative classification model based on a fuzzy frequent pattern mining algorithm (AC-FFP) is proposed that uses the membership function to fuzzify the input variables and further mine the classification association rules based on FP-growth. An efficient mining algorithm known as fuzzy association rules for high-dimensional problems (FAR-HD) was proposed in [7], which processes frequent itemsets using a two-phased multiple-partition approach especially for large high-dimensional datasets. The FAR-HD process improves the accuracy in terms of associative soft category labels by building a framework for the fuzzy associative classifier to leverage the functionality of fuzzy association rules. These approaches all involve two basic steps:1) the generation of classifiers consisting of a set of class association rules, and 2)the prediction of new data with the classifier. However, to the best of our knowledge, fuzzy associative classifiers make decisions based on the results of a data-mining algorithm instead of based on fuzzy inference rules [10]–[14]. The question of how to address the fuzzy association rules to realize fuzzy inference prediction is a challenging issue. In addition,the performance and accuracy of most association classifiers are affected by new objects imported into the systems over time. Therefore, considering future performance, the complexity and accuracy of these systems should be our key focus.

In these above mentioned systems, the extracted rules should be updated based on newly added objects or changes that take place in the existing objects over time. Hence, it is a challenge to find a suitable methodology to efficiently identify, store and match a large number of rules that classify historical data incrementally. In addition, it could be an extra burden if we wanted to keep it up-to-date dynamically. Thus far, the incremental mining problem has been studied extensively for fuzzy association rules [15]–[17]. For example, an algorithm known as the incremental update fuzzy association rules (IUAC) is proposed in [15], which needs to scan the original transaction database once and scan the new portion of the database several times. The algorithm is highly efficient when the size of the new portion of the database is relatively smaller than that of the original database. In [16], a fuzzy association rules incremental mining algorithm is proposed that can immediately obtaion the latest fuzzy association rules. In [17], a rapid incremental mining algorithm used to generate fuzzy association rules is proposed in which the transactions or data records are instantly collected online from live packets. In other words, as one data record is collected online, the latest fuzzy rules can be obtained immediately, but all of the support counts of the itemsets need to be stored, which leads to a large space cost. Similarly, when the fuzzy association classifier is used to analyze the dynamic datasets, the time complexity in the mining process remains an important issue. Once the new data are added, the incremental mining algorithm considers the past mining results and the current dataset to generate the latest fuzzy association rule set. Minimal work has been performed on the incremental fuzzy associative classifier.

This paper presents a novel incremental fuzzy association rule mining approach for classification and regression. First, the sliding window is used to partition the data, and an incremental clustering algorithm is adopted to fuzzify the data in each sliding window to solve the sharp boundary. Second, the fuzzy variable is fetched by a dynamic fuzzy variable selection algorithm to decrease the search space for the current sliding window. Finally, an incremental fuzzy association rule mining algorithm for classification and regression is applied in the current sliding window.

The main contributions of the current paper can be summed up as follows:

- Classification and regression are implemented based on causal fuzzy association rules. If the output is a categorical attribute, the mined fuzzy association rules can be used directly to solve the classification problems. If the output is a continuous variable, quantitative prediction can be achieved by fitting with the consequent of the fuzzy rule.
- In the process of incremental mining fuzzy association rules, the expected frequent itemsets are obtained by calculating the probability of the nonfrequent itemsets in the current sliding window and judging whether an itemset is expected to be a frequent itemset in the next sliding window. Based on this information, the dataset only needs to be scanned once to obtain compact and comprehensive fuzzy association rules.
- When new data are added, the existing rules are updated by considering the previously expected frequent itemsets and the new data of the current sliding window if concept drift occurs.

The remainder of this paper is organized as follows. Section 2 outlines the basic concept and introduces the fuzzy association rule mining based on Eclat and the causal fuzzy association rules for the classification and regression algorithm. Section 3 details the incremental fuzzy association rule mining. Section 4 elaborates numerical examples in various datasets, encompassing discussions on the performances of benchmarked algorithms. Conclusions and future work are presented in Section 5.

II. FUZZY ASSOCIATION RULES

A. BASIC CONCEPT

Suppose the dataset D contains n samples, L input features and one output feature, which can be represented as:

$$D = \begin{bmatrix} x_{11} & \cdots & x_{1L} & y_1 \\ x_{21} & \cdots & x_{2L} & y_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{nL} & y_n \end{bmatrix} \tag{1}$$

The input variable set is denoted as $X = [X_1, X_2, \cdots, X_j, \cdots, X_L]$, the output feature is denoted as Y. x_{ij} represents the jth $(1 \le j \le L)$ input variable X_j of the ith $(1 \le i \le n)$ sample at the current moment, and $y_i$is the value of the output feature of the ith sample.

To improve the performance of the fuzzy association rules, fuzzification must be applied before mining the fuzzy association rules. Therefore, in this paper, an incremental clustering algorithm of Bayesian adaptive resonance theory based on local distribution [18] is adopted to automatically transform each quantitative value into a fuzzy set.

Definition 1: Feature fuzzification. Assume that the input feature X_j is discretized after clustering to form $|S_j|$ partitions, which can be fuzzified as the fuzzy features vector

$$\tilde{X}_j = \{\tilde{X}_{j1}, \tilde{X}_{j2}, \cdots, \tilde{X}_{j|S_j|}\} \tag{2}$$

where $\tilde{X}_{jq}$ is the qth $(1 \le q \le |S_j|)$ fuzzy item of the jth$(1 \le j \le L)$ input feature at the current time, which can be represented as:

$$\tilde{X}_{jq} = \frac{\mu_{\tilde{X}_{jq}}(x_{1j})}{x_{1j}} + \frac{\mu_{\tilde{X}_{jq}}(x_{2j})}{x_{2j}} + \cdots + \frac{\mu_{\tilde{X}_{jq}}(x_{ij})}{x_{ij}} + \cdots + \frac{\mu_{\tilde{X}_{jq}}(x_{nj})}{x_{nj}} \tag{3}$$

where

$$\mu_{\tilde{X}_{jq}}(x) = \exp(-\frac{(x - c_{\tilde{X}_{jq}})^2}{2\sigma^2_{\tilde{X}_{jq}}}), \quad 1 \le j \le L,\ 1 \le q \le |S_j| \tag{4}$$

and $c_{\tilde{X}_{jq}}$ and $\sigma^2_{\tilde{X}_{jq}}$ are respectively the mean and the variance of samples of the qth fuzzy item of the jth input feature at the current time. Similarly, the output features Y can also be fuzzified as the fuzzy feature vector

$$\tilde{Y} = \{\tilde{Y}_1, \cdots, \tilde{Y}_{|S_Y|}\} \tag{5}$$

where $\tilde{Y}_q$ is the qth $(1 \le q \le |S_Y|)$ fuzzy item of the output feature at this time, which can be represented as:

$$\mu_{\tilde{Y}_q}(y_i) = \exp(-\frac{(y_i - c_{\tilde{Y}_q})^2}{2\sigma^2_{\tilde{Y}_q}}) \tag{6}$$

where $c_{\tilde{Y}_q}$ and $\sigma^2_{\tilde{Y}_q}$ are respectively the mean and the variance of samples of the qth fuzzy item of the output feature at the current time.

Definition 2: Fuzzy support count. Let the fuzzy itemset be written as shown

$$\tilde{X} = \{\tilde{X}_{11}, \tilde{X}_{12}, \cdots \tilde{X}_{1q}, \cdots, \tilde{X}_{1|S_1|}, \tilde{X}_{21}, \tilde{X}_{22}, \cdots \tilde{X}_{2q}, \cdots, \tilde{X}_{2|S_2|}, \cdots, \tilde{X}_{j1}, \tilde{X}_{j2}, \cdots \tilde{X}_{jq}, \cdots, \tilde{X}_{j|S_j|}, \cdots, \tilde{X}_{L1}, \tilde{X}_{L2}, \cdots \tilde{X}_{Lq}, \cdots, \tilde{X}_{L|S_L|}\} \tag{7}$$

and the support count of $\tilde{X}$ is defined as follows:

$$Supc(\tilde{X}) = \sum_{i=1}^{n} (\prod_{\tilde{X}_{jq} \in \tilde{X}} \mu_{\tilde{X}_{jq}}(x_{ij})) \tag{8}$$

where n is the number of the samples collected at the current time.

Definition 3: Fuzzy support and fuzzy confidence. Suppose the hth fuzzy association rule is

$$R_h : \tilde{X}_h \to \tilde{Y}_h \tag{9}$$

where $\tilde{X}_h \in \tilde{X}$ is a conjunction of multiple items and the multiple antecedent of the hth rule. $\tilde{Y}_h \in \tilde{Y}$ is the consequent of the hth rule; here, it is the single categorical attribute, and R_h can be represented as

$$if\ X_{1,h}\ is\ \tilde{X}_{1,f_{1,h}}, \cdots, X_{j,h}\ is\ \tilde{X}_{j,f_{j,h}}, \cdots, X_{r_h,h}\ is\ \tilde{X}_{r_h,f_{r_h,h}}\ then\ Y_h\ is\ \tilde{Y}_{f_h} \tag{10}$$

where $r_h(1 \le r_h \le L)$ is the number of variables contained in the antecedent, $\tilde{X}_{j,h}$ is the jth variable in the hth fuzzy rule, $\tilde{X}_{j,f_{j,h}}$ is the $f_{j,h}$-th fuzzy item value corresponding to the variable $X_{j,h}$ is the hth fuzzy rule, and $\tilde{Y}_{f_h}$ is the f_h-th fuzzy item value corresponding to the variable y_h in the hth fuzzy rule. The support and confidence can be expressed for a fuzzy association rule as follows:

$$\sup\left(\tilde{X}_h \cup \tilde{Y}_h\right) = \frac{\sum_{i=1}^{n} (\prod_{\tilde{X}_{jq} \in (\tilde{X}_h \cup \tilde{Y})_h} \mu_{\tilde{X}_{jq}}(x_{ij}))}{n} \tag{11}$$

$$Conf(\tilde{X}_h \to \tilde{Y}_h) = \frac{Sup(\tilde{X}_h \cup \tilde{Y}_h)}{Sup(\tilde{X}_h)} \tag{12}$$

Fuzzy itemsets with at least a minimum support are known as frequent fuzzy itemsets. Fuzzy rules with at least a minimum support and confidence are known as interesting rules.

According to the principle of Bernoulli trials [19], the probability of an infrequent itemset to appearing in l transactions out of n transactions, denoted by $P(l)_{itemset}$, can be found by the following equation:

$$p(l)_{itemset} = \binom{n}{l} \cdot p^l_{itemset} \cdot (1 - p_{itemset})^{n-l} \tag{13}$$

where $P_{itemset}$ is the probability of an itemset appearing in a transaction, n is the number of transactions in current database. According to Eq.(13), $P_{itemset}$ is approximated from the support count of an itemset in a current database. Thus, if min $Supc$ is a minimum support count after inserting new transactions into an original database, the probability of an itemset to be a frequent itemset in the updated database can be obtained as the following equation:

$$p(l \geq \min Supc)_{itemset} = 1 - p(l < \min Supc)_{itemset} \quad (14)$$

According to Eq.(13), $p(l \geq \min Supc)_{itemset}$ can be found as the following equations:

$$p(l \geq \min Supc)_{itemset} = 1 - \sum_{l=0}^{\min Supc-1} \binom{n}{l} \cdot p_{itemset}^{l} \cdot (1 - p_{itemset})^{n-l} \quad (15)$$

Let l be the number of observed success in n Bernoulli trials, then the possible values of l are $0, 1, 2, \ldots, n$. If l ($l = 0, 1, 2, \ldots, n$) successes occur, then $n - l$ failures occurs. The number of ways of selecting l positions for the l success in the n trials is

$$\binom{n}{l} = \frac{n!}{l!(n-l)!} \quad (16)$$

Definition 4: Expected frequent probability. Let $\bar{F}$ be an infrequent itemset, and *minSupc* is a minimum support count for current dataset. The probability that $\bar{F}$ becomes a frequent itemset is

$$\begin{aligned} P(l \geq minSupc)_{\bar{F}} &= 1 - P(l < minSupc)_{\bar{F}} \\ &= 1 - \sum_{x=1}^{minSupc} \frac{n!}{i!(n-i)!} p_{\bar{F}}^{l} (1 - p_{\bar{F}})^{n-l} \end{aligned} \quad (17)$$

where $p_{\bar{F}}^{l}$ is the probability of $\bar{F}$ appearing in l transactions and can be calculated by:

$$p_{\bar{F}}^{l} = \frac{\sum_{i=1}^{n} \mu_{\tilde{X}_1}(x_i) \wedge \cdots \wedge \mu_{\tilde{X}_j}(x_i) \wedge \cdots \wedge \mu_{\tilde{X}_c}(x_i)}{n}, \quad (\{\tilde{X}_1, \cdots, \tilde{X}_j, \cdots, \tilde{X}_c\} \in \bar{F}) \quad (18)$$

where c is the counts of the infrequent itemset $\bar{F}$, and n is the number of all data samples collected, where $\mu_{\tilde{X}_j}(x_i)(j = 1, \cdots, c)$ is the membership of the ith sample corresponding to the fuzzy item $\tilde{X}_j$. Any infrequent itemsets that have a probability greater than the threshold can be treated as promising frequent itemsets.

Definition 5: Threshold of expected frequent itemsets [20].

Let β indicates the minimum probability threshold with which an infrequent itemset in current moment may be become a frequent itemset in the next sliding window, which is defined as:

$$\beta = 1 - \int_{-0.5}^{v-0.5} \frac{1}{\sigma_\beta \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-c_\beta}{\sigma_\beta}\right)^2} dx \quad (19)$$

where v is the number of possible successes which an infrequent itemset in current moment may be become a frequent itemset in the next sliding window.

$$c_\beta = n(\rho + 0.95\sqrt{\frac{\rho(1-\rho)}{n^{(t-1)}}}) \quad (20)$$

$$\sigma_\beta = \sqrt{n(\rho + 0.95\sqrt{\frac{\rho(1-\rho)}{n}})(1 - (\rho + 0.95\sqrt{\frac{\rho(1-\rho)}{n}}))} \quad (21)$$

$$\rho = Supc(\tilde{X}_K)\ (X_K \in \bar{F},\ K = 1, \cdots, j, \cdots, c) \quad (22)$$

$$K = \arg_K \min_{j=1}^{c} (p_{\tilde{X}_j}) \quad (23)$$

where n is the number of all data samples, and ρ is the support count of the minimum expected frequent probability of the infrequent itemset $\tilde{X}_K$.

B. FUZZY ASSOCIATION RULES BASED ON ECLAT

As a component of association rule mining, frequent item set mining is a highly popular method for discovering interesting relationships between sets of items in large databases. The Eclat algorithm [21] was proposed to generate all frequent item sets in a depth-first manner based on vertical data representation. For vertical data layout, each item is represented by a tidsets (a set of transaction IDs whose transactions contain the item). This layout could be maintained as a bit vector. Different from the a priori algorithm based on horizontal data representation (ie. each transaction consists of a set of items and the database is a set of transactions), the Eclat algorithm only reads the database twice to find the frequent items and reduces the memory used to count the support when the sample size is not particularly large.

The search space in the Eclat algorithm is divided into k-equivalence classes. After finding all frequent sets of length k, the algorithm organizes them into disjoint groups with identical $k-1$ partial prefixes. If two sets of lengths $k+1$ have a common prefix of length k, they are in the same equivalence class. Frequent itemset candidates of length $k + 1$ are generated by intersecting sets of transactions identifiers (tidsets) of every two frequent itemsets of length k from a given equivalence class. After k iterations, all equivalence classes of k size are analyzed. The algorithm stops when all the candidates can be generated.

However, the Eclat algorithm can only work with Boolean data, but in the real world, most data are numerical and need to be discretized before the mining association rules are applied. To avoid the hard boundary problem, the data need to be fuzzified. In this paper, a fuzzy association rule mining algorithm based on the Eclat algorithm is proposed.

To find all frequent itemsets with fuzzy Eclat, the intersection operation is applied to the vertical database. In this work, the dataset in current time t are viewed as a local sliding window *SW* to count the support of items and compute the frequent item pairs. Table 1 presents the vertical dataset of the current sliding window, $\widetilde{X}_j \in \widetilde{X}$ is the jth item of the dataset, and the membership set $\{\mu(\widetilde{X}_j)\}$ includes the membership of each item for each data. An item can appear in multiple samples, and a sample also contains multiple items. The purpose of mining frequent itemsets is to find the association between items based on the vertical dataset. By scanning the

TABLE 1. vertical dataset.

item	membership set $\{\mu(\tilde{X}_j)\}$
$\tilde{X}_1$	$\mu_{\tilde{X}_1}(X_1), \mu_{\tilde{X}_1}(X_2), \ldots, \mu_{\tilde{X}_1}(X_n)$
$\tilde{X}_2$	$\mu_{\tilde{X}_2}(X_1), \mu_{\tilde{X}_2}(X_2), \ldots, \mu_{\tilde{X}_2}(X_n)$
$\tilde{X}_3$	$\mu_{\tilde{X}_3}(X_1), \mu_{\tilde{X}_3}(X_2), \ldots, \mu_{\tilde{X}_3}(X_n)$
$\tilde{X}_4$	$\mu_{\tilde{X}_4}(X_1), \mu_{\tilde{X}_4}(X_2), \ldots, \mu_{\tilde{X}_4}(X_n)$
…	…

vertical dataset one time, the support count is defined as:

$$Supc(\widetilde{X}_j) = \sum_{i=1}^{n} \mu_{\tilde{X}_j}(x_i) \tag{24}$$

where $\sum_{i=1}^{n} \mu_{\tilde{X}_j}(x_i)$ is the membership of item $\widetilde{X}_j$ for the ith data sample in the current sliding window.

If $Supc(\widetilde{X}_j)$ is less than the minimum support count min$Supc$, then $\widetilde{X}_j$ is a frequent 1-itemset of the current window. After finding all frequent $k-1$ itemsets, the algorithm organizes them into disjoint groups with identical $k-2$ partial prefixes. If the two $k-1$ itemsets have a common prefix of length $k-2$, they are in the same equivalence class. The candidate frequent k-itemset are generated by intersecting every two $k-1$ itemsets from a given equivalence class, meaning

$$Supc(C_k) = \sum_{i=1}^{n} (\mu_{F_{k-1}}(x_i)) \bigcap (\mu_{F_{k-1}}(x_i)) \tag{25}$$

where $\mu_{F_{k-1}}(x_i)$ is the membership of the frequent item F_{k-1} for the ith data x_i in the current sliding window.

According to definition 4 and definition 5, the expected frequent itemsets are obtained by finding the frequent itemsets in the vertical dataset. Based on the fuzzy Eclat algorithm, the frequent itemsets and expected frequent itemsets mining process are shown in algorithm 1.

Algorithm 1 Frequent itemsets mining algorithm

1. Scan the fuzzy horizontal dataset and transforms it into a vertical dataset;
2. By scanning the vertical dataset one time, calculate the support count of each item
    $Supc\left(\tilde{X}_u\right) = \sum_{i=1}^{n} \mu_{\tilde{X}_u}(x_i)$
3. If $Supc\left(\tilde{X}_u\right) \geq$ min $Supc$
        $\tilde{X}_u$ is a frequent 1-itemset F_1
    Otherwise
        $\tilde{X}_u$ is a frequent 1-itemset $\overline{F}_1$
    End
4. While (F_k is nonnull)
        $C_k = F_k \times F_{k-1}$
        Calculate the support count $Supc\,(C_k)$
5. If $Supc\,(C_k) \geq$ min $Supc$
        Then C_k is an infrequent itemset $\overline{F}_k = \overline{F_k} \cup C_k$
        If each of itemset C_k == frequent
            C_k is a frequent itemset
        Otherwise
            C_k is an infrequent itemset $\overline{F}_k = \overline{F_k} \cup C_k$
        end
    end
6.Calculate the expected frequent possibility $p_{\bar{F}_k}$ of the infrequent
itemset $\bar{F}_k$ and the minimum expected frequent possibility β of the current sliding window.
7 If $p_{\bar{F}_k} \geq \beta$
        F_k is the expected frequent itemset, $EF_k^{(i)} = EF_k \cup \overline{F}_k$
    end

C. CAUSAL FUZZY ASSOCIATION RULE MINING FOR CLASSIFICATION

Most existing fuzzy associative classifiers only focus on generating rules with the support-confidence framework and without considering the predictive ability of the features involved in a classification rule. Thus, extraction of a minimal set of rules with a strongly predictive capability is critical to building an efficient incremental associative classifier from the high-dimensional data that are prevalent in many real-world applications. To meet these challenges, we propose a new framework that integrates causality into fuzzy associative classification.

Traditional associative classification algorithms identify relationships between the class label and its antecedents using statistical correlation. However, correlation is not causation [22]. For example, we not only want to know whether a particular operation parameter is associated with a process fault (a typical fault class), but we also want to know definitively whether the association is due to an adverse reaction. Without knowing the true relationship, plain associations can lead to false conclusions, e.g., the operation parameter causes special process fault class label. Consequently, by detecting the causal relationships between the class label and its antecedents, we can uncover causal or consequential factors with respect to the class label. In generating the set of classification rules, the only features considered are those that belong to this causal feature space instead of the combinations of all features. Thus, this process can greatly reduce the computational cost and large resource demands in the stage of rule extraction. Furthermore, the extracted rules are not only causally interpretable but also causally informative. Thus, the causal index can be introduced into fuzzy association rule mining to solve classification problems.

Suppose the hth fuzzy association rule is R_h : $\tilde{X}_h \rightarrow \tilde{Y}_h$, the antecedent $\tilde{X}_h$ can be represented as $\{\tilde{X}_{1,f_{1,h}}, \cdots, \tilde{X}_{j,f_{j,h}}, \cdots, \tilde{X}_{k,f_{k,h}}, \cdots, \tilde{X}_{r_h,f_{r_h,h}}\}(1 \leq j < k \leq r_h)$, and $\tilde{Y}_h \in \tilde{Y}$ is the consequent of the hth rule.

When the rule $\tilde{X}_h \rightarrow \tilde{Y}_h$ has more than three antecedents, any subrules containing two antecedents must discriminate the conditional dependence relationships with the consequence.

In addition, the interestingness of rule $\tilde{X}_h \to \tilde{Y}_h$ is dependent on the weakness of its subrules, so the causal index for $\tilde{X}_h \to \tilde{Y}_h$ is defined as

$$CI(\tilde{X}_h \to \tilde{Y}_h) = \min_{(\tilde{X}_{j,f_{j,h}}, \tilde{X}_{k,f_{k,h}}) \in \tilde{X}_h} CI(\tilde{X}_{j,f_{j,h}}, \tilde{X}_{k,f_{k,h}} \to \tilde{Y}_h) \tag{26}$$

where $\forall \tilde{X}_{j,f_{j,h}}, \tilde{X}_{k,f_{k,h}} (j \neq k)$, and the causal index between two items for $CI(\tilde{X}_{j,f_{j,h}}, \tilde{X}_{k,f_{k,h}} \to \tilde{Y}_h)$ is defined as

$$\begin{aligned} CI(\tilde{X}_{j,f_{j,h}}, \tilde{X}_{k,f_{k,h}} \to \tilde{Y}_h) &= 2H(\tilde{X}_{j,f_{j,h}}) + 2H(\tilde{X}_{k,f_{k,h}}) \\ &\quad + 3H(\tilde{Y}_h) - H(\tilde{X}_{j,f_{j,h}}, \tilde{Y}_h) - H(\tilde{X}_{k,f_{k,h}}, \tilde{Y}_h) \\ &\quad - H(\tilde{X}_{j,f_{j,h}}, \tilde{X}_{k,f_{k,h}}, \tilde{Y}_h) \end{aligned} \tag{27}$$

where $\tilde{X}_{j,f_{j,h}}$ and $\tilde{X}_{k,f_{k,h}}$ are the memberships applied for variable X_j and X_k in the antecedent of hth fuzzy rule respectively, and $H(.)$ is the information entropy.

The information entropy of $\{\tilde{X}_{j,f_{j,h}}, \tilde{X}_{k,f_{k,h}}, \tilde{Y}_h\}$ is defined as:

$$H(\{\tilde{X}_{j,f_{j,h}}, \tilde{X}_{k,f_{k,h}}, \tilde{Y}_h\}) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{B_{i,h}}{n} \tag{28}$$

where $B_{i,h}$ is the cardinality of the ith sample corresponding to the hth fuzzy rule, as derived in Appendix A. The entropy of other fuzzy items can be calculated in the similar manner.

The mining process of the causal fuzzy association rule is given in algorithm 2, and the obtained rule for classification can be described as follows.

$$R_h : \mathit{if}\ X_{1,h}\ \mathit{is}\ \tilde{X}_{1,f_{1,h}} \cdots, X_{j,h}\ \mathit{is}\ \tilde{X}_{j,f_{j,h}} \cdots, X_{r_h,h}\ \mathit{is}\ \tilde{X}_{r_h,f_{r_h,h}}\quad \mathit{then}\ Y_h\ \mathit{is}\ \mathit{class}_h \tag{29}$$

Algorithm 2 Causal fuzzy association rule mining algorithm

1. For $\tilde{X} \in F_k, \tilde{Y} \in (F_k - \tilde{X})(k \geq 2)$
calculate the confidence level $conf(\tilde{X} \to \tilde{Y})$ of the rule $\tilde{X} \to \tilde{Y}$
2. If $conf(\tilde{X} \to \tilde{Y}) \geq \min conf$
Then $\tilde{X} \to \tilde{Y}$ is an interesting fuzzy association rule
3.Calculate the casual index $CI(\tilde{X} \to \tilde{Y})$ of the rule $\tilde{X} \to \tilde{Y}$
4. If $CI(\tilde{X} \to \tilde{Y}) \geq \min CI$
Then $\tilde{X} \to \tilde{Y}$ is a casual fuzzy association rule

In this work, the single categorical attribute $class_h$ is the class labels of the hth rules.

D. CAUSAL FUZZY ASSOCIATION RULE MINING FOR REGRESSION

To realize the regression prediction accuracy of the fuzzy association rules mining models, the causal fuzzy association rules are reconstructed by combining the TS fuzzy model [23] with the extracted quantitative association rules, and on this basis, the best rule is directly identified to match the new sample and predict its output. The existing fuzzy association rules are rewritten with a similar expression by combining with the TS fuzzy inference rules to predict the output values to realize fuzzy inference with the regression function (see Eq.(30))

$$\begin{aligned} R_h &: \mathit{if}\ X_{1,h}\ \mathit{is}\ \tilde{X}_{1,f_{1,h}}, \cdots, X_{r_h,h}\ \mathit{is}\ \tilde{X}_{r_h,f_{r_h,h}} \\ &\mathit{then}\ Y_h = \sum_{i=1}^{n_h} (\alpha_{i,h} - \alpha_{i,h}^*) K(\vec{x}_{i,h}, \vec{x}) + b_h, \quad \mathit{with}\ RW_h \end{aligned} \tag{30}$$

The antecedent of the hth fuzzy association rule is same as in (29), the consequent of the hth fuzzy association rule is shown in (30), Y_h is the output variable in the hth fuzzy association rules at the current moment, and the consequent is expressed as a regression function by using the support vector regression method (SVR) [24]. SVR may be used for TS structure learning [25], [26], parameter learning [27]–[30], or both [31]. The use of SVR for TS parameter learning is helpful for improving the generalizability of TS fuzzy inference model. In this work, the fuzzy rule weight is proposed, which is another important index used to measure the fuzzy association rule in [0, 1], which can reflect the evolution process of the fuzzy rules in different sliding windows. The closer the fuzzy rule weight is to 1, the more important it is; otherwise, the less important it is. The weight of fuzzy rules is defined as:

$$RW_h = \frac{\sum_{i=1}^{n_h} w_h(x_i)}{\sum_{i=1}^{n} w_h(x_i)} \tag{31}$$

where $w_h(x_i)$ is the degree of activation of the hth rule

$$w_h(x_i) = T(\mu_{\tilde{X}_{1,h}}(x_i), \cdots, \mu_{\tilde{X}_{j,h}}(x_i), \cdots, \mu_{\tilde{X}_{r_h,f_{r_h,h}}}(x_i)) \tag{32}$$

where $\mu_{\tilde{X}_{j,h}}(x_i)$ is the membership of the hth rule for the ith data x_i at the current moment, n_h is the number of samples matching with the hth rule, and T is the t-norm.

The input-output mapping becomes

$$f(\vec{x}) = \sum\nolimits_{h=1}^{m} RW_h Y_h \tag{33}$$

where $\vec{x} = [x_1, \cdots, x_i, \cdots, x_{r_h}]^T$ is the input sample.

The support vector regression algorithm [32] is applied to predict the output. As is known, the support vector regression learning mechanism is based on statistical learning theory, which has strong generalization ability and can avoid overfitting. The SVR approach is described as follows. Let n_h be the number of the data samples matching with the hth association rule, $0.7n_h$ of the data samples are used in training, and the remaining $0.3n_h$ samples are used only to verify the identified model. We construct the kernel function of SVR based on the fuzzy basis function to realize the fuzzy inference system.

The Mercer kernel is defined as

$$K(x_j, \vec{x}) = \prod_{i=1}^{r_h} \exp\left(-\frac{1}{2}\left(\frac{x_i - c_{x_{jq}}}{\sigma_{x_{jq}}}\right)^2\right) \tag{34}$$

where $c_{x_{jq}}$ and $\sigma_{x_{jq}}$ are respectively the mean and the variance of samples of the qth ($q = 1, \ldots, |s_j|$) fuzzy item of the jth ($j = 1, \ldots, r_h$) input variable at the current sliding window.

The decision function is defined as

$$Y_h(\vec{x}) = \sum_{i=1}^{n_h} (\alpha_{i,h} - \alpha_{i,h}^*) K(\vec{x}_{i,h}, \vec{x}) + b_h \tag{35}$$

To obtain the decision function, the SMO algorithm is adopted to solve the quadratic program and obtain the Lagrange multipliers α_i,α_i^*.

$$\begin{cases} \min\{-\frac{1}{2}\sum_{i,j=1}^{0.7n_h}(\alpha_{i,h} - \alpha_{i,h}^*)(\alpha_{j,h} - \alpha_{j,h}^*) < \vec{x}_{i,h}, \vec{x}_{j,h} > \\ \quad + \varepsilon \sum_{i=1}^{0.7n_h}(\alpha_{i,h} + \alpha_{i,h}^*) - \sum_{i=1}^{0.7n_h}(\alpha_{i,h} - \alpha_{i,h}^*) Y_{i,h}\} \\ s.t. \sum_{i=1}^{n_h}(\alpha_{i,h} - \alpha_{i,h}^*) = 0, 0 \le \alpha_{i,h}, \alpha_{i,h}^* \le C_h \end{cases} \tag{36}$$

$$b_h = \frac{1}{n_h} \begin{cases} \sum_{0<\alpha_{i,h}<c_h} [Y_h - \sum_{x_{i,h} \in sv} (\alpha_{i,h} - \alpha_{i,h}^*) k(\vec{x}_{i,h}, \vec{x}) - \varepsilon_h] \\ + \sum_{0<\alpha_{i,h}^*<c_h} [Y_h - \sum_{x_{i,h} \in sv} (\alpha_{i,h} - \alpha_{i,h}^*) k(\vec{x}_{i,h}, \vec{x}) + \varepsilon_h] \end{cases} \tag{37}$$

The algorithm does not stop the training until the error and SVs are satisfied with the given conditions.

III. INCREMENTAL FUZZY ASSOCIATION RULE MINING

The fuzzy association rule approach can combine the data mining results with human expertise and background knowledge, in the form of rules, to attain labeled classes for classification of data streams. Another advantage of the fuzzy logic approach is that it gives classification results that include a degree of probability.

Traditional fuzzy association rules are designed in batch mode, i.e., by using the complete training data all at once. For stationary processes, this approach is sufficient, but for time-based and complex nonstationary processes, efficient techniques for updating the induced models are required. To avoid starting from scratch every time, fuzzy association rule mining techniques must be able to learn online and incrementally by adapting the current model using only the new data and without referring to the old model.

To improve the efficiency and rapidity of incremental fuzzy association rules mining, the incremental fuzzy association rules based on Eclat (IFARE) is proposed. The basic framework of this algorithm is shown in Fig. 1. First, the sliding window strategy presented in Appendix B is adopted to divide the data, and the data are fuzzified using an incremental clustering algorithm [18] in each sliding window. Based on all of the fuzzy variables, the dynamic fuzzy variable selection algorithm [33] is adopted to select the fuzzy variable to reduce the searching space. Based on the fuzzy variable selection results of each sliding window, incremental fuzzy association rules mining is conducted. In the first sliding window SW^0, the causal fuzzy Eclat algorithm is adopted to obtain the frequent itemsets and the expected frequent itemsets in the next sliding window. In the next sliding window $SW^{(t)}(t \geq 1)$, if the concept drift occurs, the frequent 1-itemset and expected frequent 1-itemset obtained in the current window are combined with the expected frequent k-itemsets in the previous sliding window to renew the frequent k-itemsets and the expected frequent k-itemsets of the current sliding window. According to the updated association rules, a new fitting function is constructed to predict the output; otherwise, the current existing rules are adopted to predict the output until the next window is available.

The difficultly of incremental learning lies in of course the inability to accurately estimate the statistical characteristics of the incoming data in the future. In nonstationary changing environments, the challenge is daunting because the rule system may change drastically over time due to concept drift [34].

A. CONCEPT DRIFT DETECTION

Concept drift [34] means that the structure or the distribution of the data changes over time in unforeseen ways. Concept drift detection must find the similarities and differences between the current data and previous data and help us to judge whether the association rules in the rule set are matched with the current data.

Two criteria used to detect the concept drift are given as follows:

1) Rule matching with outliers.

Because the idea of rule matching is to make use of the existing rules to classify the incoming data points [17], outliers that are not able to match to any rule may be generated based on new concepts. Therefore, if too many outliers are detected by the rule matching step, the drifting concept may occur in the current sliding window. As a result, a threshold known as the outlier threshold θ is set in this step, which is adopted to judge the ratio of outliers in the current sliding window. In this paper, we set this value as [0.1,0.3]. If the ratio of outliers in the current sliding window is larger than the outlier threshold, concept drift occurs.

The ratio of outliers in the current sliding window is:

$$OR^{(t)} = \frac{\#outliiers^{(t)}}{n^{(t)}} \tag{38}$$

where $\#outliers^{(t)}$ is the number of the outliers in the current sliding window.

2) Clustering with the ratio of data points.

Moreover, another type of drift concept is also detected in the data fuzzification step in which each variable of the

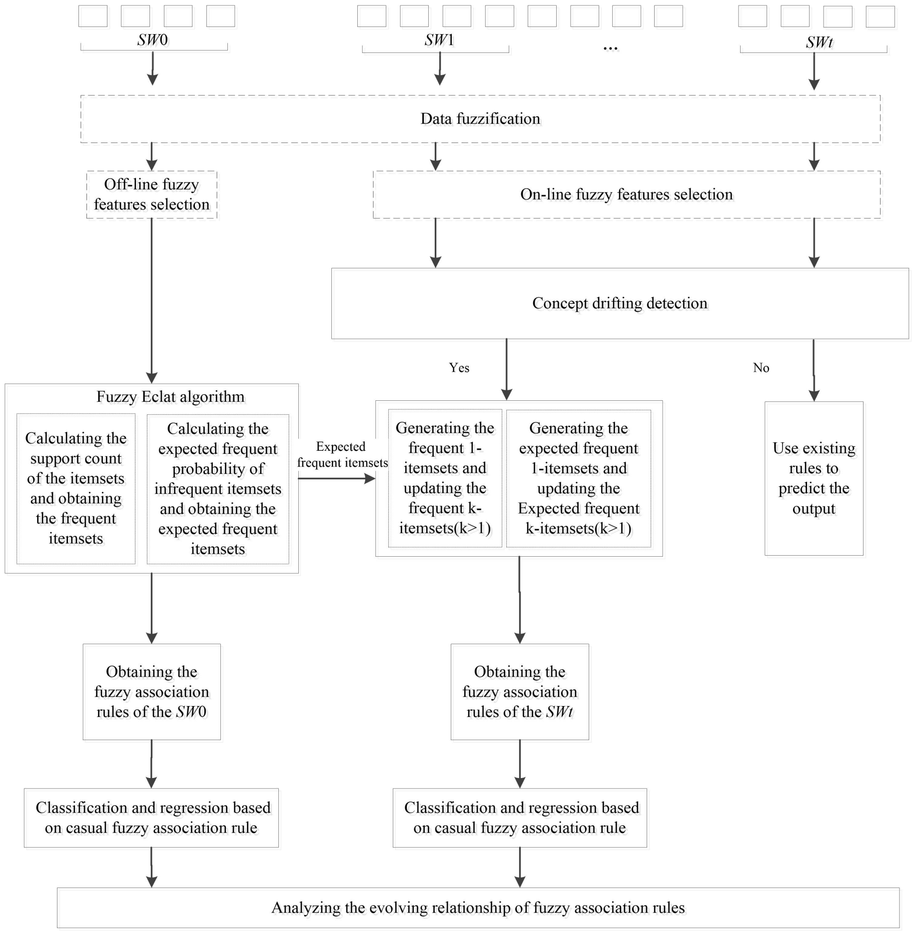


FIGURE 1. The framework of the algorithm.

original dataset is clustered separately. The ratio of data points in a cluster may be changed dramatically by a drifting concept, e.g., the current cluster only contains a small portion of the data points within the last cluster. To detect the change, we adopt a double-threshold method. One threshold is known as the cluster variation threshold γ, which is adopted to determine whether the variation of the ratio of data points in the current cluster is sufficient large. In this paper, we set $\gamma = [0.7, 1]$. The cluster that exceeds the cluster variation threshold is viewed as a new and different cluster.

The variation of the ratio of data points for the qth clustering of the jth variables in the current sliding window is denoted as:

$$DR^{[t-1,t]}_{\tilde{X}^{(t)}_{jq}} = \left| \frac{n^{(t-1)}_{jq}}{n^{(t-1)}} - \frac{n^{(t)}_{jq}}{n^{(t)}} \right| \tag{39}$$

where $n^{(t-1)}_{jq}$ is the number of data points of the qth clustering that belong to the jth variable in the sliding window $SW^{(t-1)}$, and $n^{(t)}_{jq}$ is the number of samples of the qth clustering that belong to the jth variable in the sliding window $SW^{(t)}$.

This threshold is referred to as the cluster distribution threshold, which is adopted to determine whether the variation of the ratio of the entire cluster distribution in the current sliding window is sufficiently large. In this paper, we set $\eta = [0.7, 1]$. If the variation of the ratio of the entire cluster distribution in the current sliding window is larger than the distribution threshold, then concept drift occurs in the current sliding window.

The variation of the ratio of the entire cluster distribution in the current sliding window is shown as follows:

$$CR^{(t)} = \frac{\sum_{j=1}^{L} \sum_{q=1}^{\left|S^{(t-1)}_j\right|} d(\tilde{X}^{(t-1)}_{jq}, \tilde{X}^{(t)}_{jq})}{\sum_{j=1}^{L} \left|S^{(t-1)}_j\right|} \tag{40}$$

where $\left|S^{(t-1)}_j\right|$ is the number of clusters of the jth variable in the sliding window $SW^{(t-1)}$, and the variation of the ratio of data points for the qth cluster of the jth variable between the previous sliding window $SW^{(t-1)}$ and the current sliding window $SW^{(t)}$ is calculated as:

$$d(\tilde{X}^{(t-1)}_{jq}, \tilde{X}^{(t)}_{jq}) = \begin{cases} 1, & DR^{[t-1,t]}_{\tilde{X}_{jq}} \geq \gamma \\ 0, & otherwise \end{cases} \tag{41}$$

If the concept drift occurs in the current sliding window $SW^{(t)}$, the data points in the current sliding window are subjected to remining. In constrast, the rule set is updated by adding the current mining results into the last mining result.

According to the above two criteria, as long as any criteria are satisfied, concept drift occurs. The entire concept drift detection process is shown as follows:

$$concept \quad drift = \begin{cases} yes, & if\, OR^{(t)} \geq \theta \\ yes, & if\, CR^{(t)} \geq \eta \\ no, & otherwise \end{cases} \tag{42}$$

B. INCREMENTAL FUZZY ASSOCIATION RULE MINING FOR CLASSIFICATION AND REGRESSION

To enhance the effectiveness and real-time nature of incremental rule mining, the algorithm is proposed with partitioned sliding windows. In the first sliding window, frequent itemsets are obtained to mine the causal fuzzy association rules, and the expected frequent itemsets are generated for the frequent itemsets of the next sliding window. In the following sliding window, concept drift detection is determined with two measurement criteria. If concept drift occurs, the expected frequent itemsets in the previous sliding window and the frequent itemsets in the current sliding window are considered simultaneously to update the causal fuzzy association rules and the implemented fuzzy inference by reconstructing the consequent of the fuzzy rule; otherwise, the current existing rules are adopted to predict the output until the next window is available. Finally, the evolving relationship of the fuzzy rules is analyzed according to the weights of the rules.

The detailed steps of the algorithm are described as follows:

Step 1. Transform the horizontal dataset into a vertical dataset.

Step 2. Generate the frequent itemset $F^{(0)}$ and expected frequent itemset $EF^{(0)}$ of the sliding window SW^0.

Step 3. Generate the causal fuzzy association rules of the sliding window SW^0.

Step 4. Calculate the ratio of outliers $OR^{(t)}$ and the ratio of the entire cluster distribution $CR^{(t)}$ of the current sliding window $SW^{(t)}(t \geq 1)$. If $OR^{(t)} \geq \theta$ or $CR^{(t)} \geq \eta$ then concept drift occurs; go to Step5. Otherwise, the current existing rules continue to be adopted.

Step 5. Generate the frequent itemset of the tth sliding window $SW^{(t)}$.

Step 5.1. Generate the frequent 1-itemset and expected frequent 1-itemset of the tth sliding window $SW^{(t)}$.

Step 5.2. Update the frequent k-itemset and expected frequent k-itemset of the tth sliding window.

Step 5.2.1. The frequent k-itemset candidates are generated by intersecting every two $(k-1)$-itemsets from a given equivalence class.

$$C_k^{(t)} = (F_{k-1}^{(t)} \cup EF_{k-1}^{(t)}) \times (F_{k-1}^{(t)} \cup EF_{k-1}^{(t)})$$

TABLE 2. The details of the dataset.

dataset	the number of samples	original dimension	dimension after fuzzification
BCH	1372	4	8
Glass	214	9	33
Wine	178	13	72
SEA	50000	3	10
Weather	18159	8	25
Airfoil Self-Noise	1503	5	17
Energy efficiency	768	8	26
white wine quality	4898	11	25
Beijing PM2.5 data	43800	6	21
PPPTS	45730	9	31

Step 5.2.2. $\forall C_k^{(t)} \in F_k^{(t-1)} \cup EF_k^{(t-1)}$, and if $Supc(C_k^{(t)}) \geq minSupc$ then $C_k^{(t)}$ becomes a frequent k-itemset $F_k^{(t)}$ of the current sliding window $SW^{(t)}$, Otherwise, the expected frequent possibility $p_{C_k^{(t)}}$ of the $C_k^{(t)}$ is calculated, and if $p_{C_k^{(t)}} \geq \beta^{(t)}$,then $C_k^{(t)}$ becomes an expected frequent k-itemset $EF_k^{(t)}$ of the current sliding window $SW^{(t)}$.

Step 5.2.3. $\forall C_k^{(t)} \notin F_k^{(t-1)} \cup EF_k^{(t-1)}$, and if $1-\beta^{(t)} \cdot \left|n^{(t)}\right| \leq 0$, then the support count of $C_k^{(t)}$ is calculated. If $Supc(C_k^{(t)}) \geq minSupc$ then $C_k^{(t)}$ becomes a frequent k-itemset $F_k^{(t)}$ of the sliding window $SW^{(t)}$. Otherwise the expected frequent possibility $p_{C_k^{(t)}}$ is calculated, and if $p_{C_k^{(t)}} \geq \beta^{(t)}$, then $C_k^{(t)}$ becomes an expected frequent k-itemset $EF_k^{(t)}$ of the current sliding window $SW^{(t)}$.

Step 6. Generate the causal fuzzy association rules of the sliding window $SW^{(t)}$.

Step 7. Realize the classification and regression for new samples based on the causal fuzzy association rules.

C. TIME COMPLEXITY ANALYSIS

To verify the efficiency of the incremental fuzzy association rule mining algorithm, we analyzed the time complexity. The time complexity of first scanning the dataset of the sliding window $SW^{(t)}(t = 0)$ to obtain the frequent 1-itemsets is $O(m \times n^{(t)})$. The time complexity for calculating the expected frequent possibility of the infrequent 1-itemset $\bar{F}_1^{(t)}$ is $O(\left|\bar{F}_1^{(t)}\right|)$, where $\left|\bar{F}_1^{(t)}\right|$ represents the number of the infrequent 1-itemset. The time complexity for generating the candidate k-itemsets by intersecting every two $(k-1)$-itemsets from a given equivalence class is $O(\sum_{k\geq 1} \left|F_k^{(t)}\right| \times \left|F_k^{(t)}\right|)$. The time complexities for calculating the support count of the candidate $(k+1)$-itemset $C_{k+1}^{(t)}$ and the expected frequent possibility of the infrequent $(k+1)$-itemset are $O(\sum_{k\geq 1} \left|C_{k+1}^{(t)}\right|)$ and $\left|\bar{F}_{k+1}^{(t)}\right|$, respectively. Where $\left|\bar{F}_{k+1}^{(t)}\right|$ is the number of the infrequent $(k+1)$-itemset. The time complexity needed to obtain the interesting rules is $O(\left|R^{(t)}\right|)$, where $\left|R^{(t)}\right|$ denotes the number of the rules. The time complexity needed to obtain the causal association rules is $O(\left|R^{(t)}\right|)$. The time complexity needed to reconstruct the rules consequence is $O(\left|R^{(t)}\right|)$. If concept drift occurs, the current sliding window

TABLE 3. The comparison of the classification performance (Acc/RI) on the different window size.

Dataset	Window size accounts for dataset proportion					The proposed
	10%	20%	30%	40%	50%	method
BCH	0.821/0.811	0.844/0.836	0.849/0.853	0.816/0.821	0.799/0.786	0.932/0.924
Glass	0.814/0.802	0.838/0.823	0.875/0.854	0.869/0.844	0.841/0.824	0.969/0.942
Wine	0.811/0.824	0.844/0.831	0.858/0.847	0.844/0.826	0.809/0.795	0.934/0.913
SEA	0.876/0.861	0.841/0.833	0.817/0.821	0.801/0.812	0.784/0.784	0.922/0.924
Weather	0.823/0.814	0.842/0.833	0.870/0.864	0.851/0.842	0.822/0.809	0.946/0.911
Mean	0.829/0.824	0.842/0.831	0.854/0.848	0.836/0.829	0.811/0.800	0.941/0.923

TABLE 4. The comparison of the prediction performance (MAPE/RMSE) on the different window size.

Dataset	Window size accounts for dataset proportion					The proposed
	10%	20%	30%	40%	50%	method
Airfoil Self-Noise	0.131/7.564	0.091/6.523	0.043/4.695	0.096/6.946	0.113/7.146	0.006/2.846
Energy effciency	0.164/8.469	0.111/7.264	0.082/6.329	0.065/5.326	0.098/5.792	0.003/3.061
white wine quality	0.188/12.364	0.126/10.498	0.084/8.429	0.166/11.263	0.241/12.036	0.009/5.239
Beijing PM2.5 data	0.121/9.562	0.063/6.239	0.143/7.126	0.163/11.364	0.221/13.945	0.007/5.236
PPPTS	0.143/8.369	0.066/7.261	0.093/8.369	0.121/10.274	0.211/13.451	0.008/3.269
Mean	0.149/9.265	0.091/7.554	0.089/6.988	0.122/7.035	0.177/10.474	0.007/3.930

is processed in the same way with the first sliding window, and the only difference is the method that generates the candidate $(k+1)$-itemset $C_{k+1}^{(t)}$ by intersecting every frequent k-itemset $F_k^{(t)}$ and the expected frequent k-itemset $EF_k^{(t)}$. From K, a given equivalence class, the time complexity is $O(\sum_{k\geq 1}(\left|F_k^{(t)}\right|+\left|EF_k^{(t)}\right|)\times(\left|F_k^{(t)}\right|+\left|EF_k^{(t)}\right|))$. In the worst case, the time complexity of the algorithm for the tth sliding window is $O(m\times n^{(t)}+\sum_{k\geq 1}((\left|F_k^t\right|+\left|EF_k^{(t)}\right|)\times(\left|F_k^t\right|+\left|EF_k^{(t)}\right|)+\left|C_{k+1}^{(t)}\right|+\left|\bar{F}_{k+1}^{(t)}\right|)+3\left|R^{(t)}\right|)$.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the IFARE algorithm is verified in terms of execution efficiency, interpretability and accuracy by comparison with other algorithms such as the IUAC algorithm [21], AC-FFP algorithm [6], FPS algorithm [35], and QART algorithm [36]. In addition, all experiments were performed on the datasets in Table 2, in which the datasets [37] such as BCH/SEA/Weather/Glass/Wine are viewed as binary classification datasets, and datasets [38] such as Airfoil Self-Noise/Energy Efficiency/White Wine Quality/Beijing PM2.5 data and PPPTS are viewed as regression datasets. As noted, the dimensions become higher than those of the original datasets after fuzzification.

In this paper, four criteria are adopted to evaluate the incremental fuzzy association rules mining algorithm as follows:

1) Classification accuracy (Acc)

$$Acc=\frac{\sum_{k=1}^{|C|}a_k}{n} \tag{43}$$

where a_k represents the number of data that have the correct classification results corresponding to the kth category, $|C|$ is the number of categories, and n is the number of data in the dataset.

2) Rand index (RI)

$$RI=\frac{a_d+a_k}{n(n-1)/2} \tag{44}$$

where a_d represents the number of data that have false classification results corresponding to the actual class label, a_k represents the number of data that have the correct classification results corresponding to the correct class label, and n is the number of data contained in the dataset.

3)Mean absolute percentage error (MAPE)

$$MAPE=\frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{y}_i-y_i\right|}{y_i} \tag{45}$$

where y_i is the actual value of the ith data, and $\hat{y}_i$ is the prediction output of ith data.

4) Root mean square error (RMSE)

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i-y_i)^2}{n}} \tag{46}$$

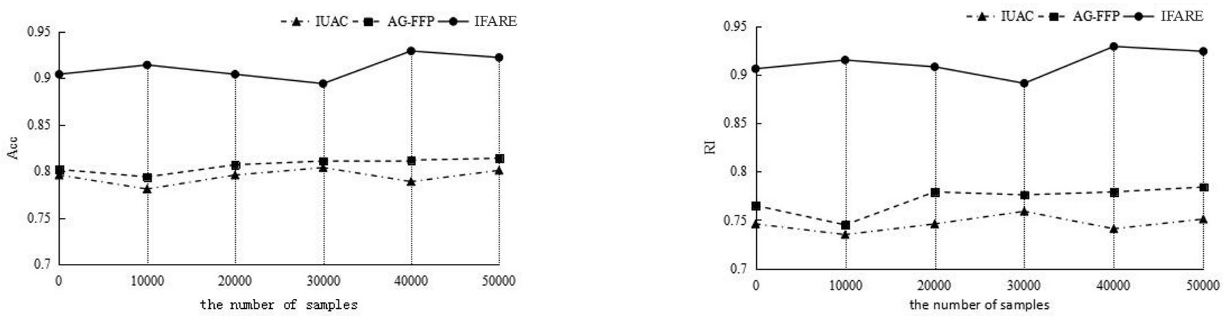
where y_i is the actual value of the ith data, and $\hat{y}_i$ is the prediction output of ith data.

A. PERFORMANCE WITH DIFFERENT SLIDING WINDOW

To evaluate the effect on the IFARE algorithm with different sliding window sizes, different dataset are partitioned to form different sliding windows, where 70 percent of the data are used in model training and 30 percent of the data are selected to test the classification and prediction performance of the model. Table 3 and Table 4 respectively show the classification performance and the prediction performance of the fuzzy association rules obtained by the IFARE algorithm with

TABLE 5. The evolution performance for classification with different dataset.

dataset	the number of evolution	the number of rules			running time/s			Acc/RI		
		IUAC	AC-FFP	IFARE	IUAC	AG-FFP	IFARE	IUAC	AG-FFP	IFARE
BCH	0	32	25	17	14.326	11.785	5.792	0.861/0.844	0.879/0.853	0.942/0.935
	1	54	31	21	22.654	15.662	6.236	0.856/0.834	0.862/0.842	0.941/0.931
	2	86	45	26	32.494	20.167	8.241	0.843/0.829	0.884/0.864	0.932/0.924
Glass	0	119	57	38	15.321	11.893	7.086	0.853/0.836	0.871/0.852	0.956/0.952
	1	143	68	42	22.654	16.869	7.789	0.856/0.832	0.863/0.843	0.952/0.949
	2	164	79	49	27.986	22.875	8.224	0.849/0.814	0.852/0.835	0.963/0.932
	3	194	92	58	39.451	29.347	9.642	0.826/0.804	0.846/0.827	0.969/0.942
Wine	0	209	157	69	37.896	29.554	11.889	826/0.817	0.855/0.826	0.951/0.933
	1	296	196	86	44.794	36.579	14.623	0.825/0.806	0.842/0.816	0.944/0.926
	2	389	247	106	54.168	40.267	16.786	0.814/0.774	0.839/0.809	0.934/0.913
SEA	0	57	25	20	59.224	43.851	12.097	0.835/0.822	0.839/0.812	0.941/0.921
	1	69	31	24	65.882	49.021	13.891	0.831/0.815	0.841/0.808	0.946/0.919
	2	81	38	22	71.562	53.241	15.832	0.824/0.802	0.822/0.794	0.936/0.924
	3	92	51	25	77.842	58.124	19.226	0.813/0.779	0.821/0.792	0.923/0.931
	4	114	66	30	81.462	63.478	22.264	0.801/0.751	0.814/0.784	0.922/0.924
Weather	0	71	51	24	66.894	60.146	20.146	0.852/0.812	0.852/0.831	0.945/0.915
	1	104	68	36	75.421	65.223	25.679	0.846/0.805	0.841/0.824	0.935/0.902
	2	148	76	44	89.267	70.364	30.163	0.813/0.763	0.824/0.812	0.946/0.911



(a) Acc

(b) RI

FIGURE 2. Classification Accuracy for dataset SEA.

different sliding window sizes. As observed, the classification and prediction performances of the IFARE algorithm are better than those of the predefined window size because our proposed method can adaptively determine the window size according to the Hoeffding boundary, which improves the classification and prediction performance of the fuzzy data.

B. EVALUATION ON EFFICIENCY AND ACCURACY

To verify the effectiveness of IFARE algorithm, the classification and prediction accuracy are compared with those of the IUAC algorithm, AC-FFP algorithm, FPS algorithm and QART algorithm with multiple datasets. As shown in Table 5, compared with the IUAC algorithm and AC-FFP algorithm, the IFARE algorithm has the best classification performance with the maximum Acc and RI for the same dataset. First, the size of the sliding window is automatically determined by our algorithm, and based on this information, the number of evolution times of the five datasets are determined to be 2 times, 3 times, 2 times, 4 times and 2 times, respectively. Because the IFARE algorithm considers the concept drift of the data and updates the ruleset according to the variation of the data distribution, the classification accuracy of the rules is improved.

Similarly, as shown in Table 6, compared with the FPS algorithm and QART algorithm, the IFARE algorithm has the best regression prediction performance, producing the minimum MAPE and RMSE with different supports and confidences for the same dataset. First, the size of the sliding window is automatically determined by our algorithm, and based on this information, the number of evolution times of the five datasets are determined as 2 times, 2 times, 3 times, 7 times and 6 times, respectively. The QART algorithm is a

TABLE 6. The evolution performance for regression with different dataset.

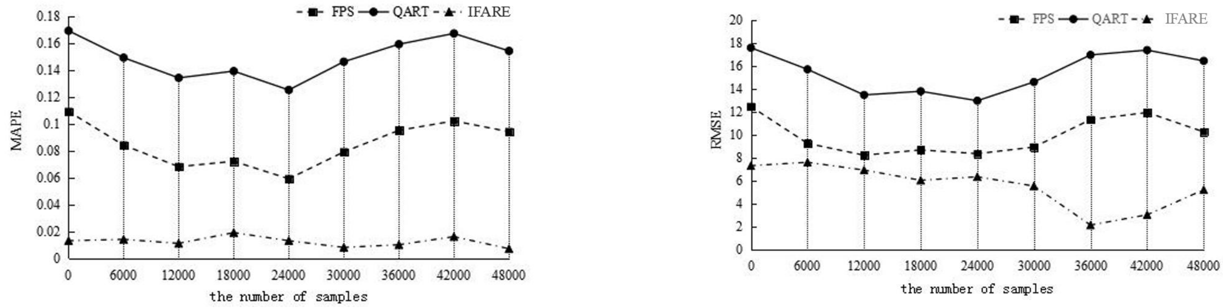
dataset	the number of evolution	the number of rules			running time/s			MAPE/RMSE		
		QART	FPS	IFARE	QART	FPS	IFARE	QART	FPS	IFARE
Airfoil Self-Noise	0	89	47	25	22.789	18.167	8.653	0.093/7.995	0.054/5.793	0.005/2.765
	1	106	59	29	29.863	22.931	10.896	0.104/8.953	0.069/6.874	0.009/3.846
	2	127	70	34	35.419	26.459	12.346	0.146/10.694	0.079/7.267	0.006/2.846
Energy effciency	0	101	59	32	30.883	22.894	10.943	0.078/8.956	0.024/6.894	0.004/3.453
	1	119	74	41	35.894	24.789	11.562	0.094/10.793	0.032/7.932	0.005/4.569
	2	131	84	51	41.269	30.0264	13.469	0.104/11.478	0.049/8.246	0.003/3.061
white wine quality	0	69	50	39	26.894	29.654	11.842	0.083/10.795	0.049/6.893	0.008/5.108
	1	86	56	39	33.456	32.894	13.897	0.095/11.658	0.058/7.943	0.007/5.089
	2	109	69	43	40.896	36.894	15.894	0.106/12.589	0.064/8.563	0.006/4.893
	3	126	78	47	50.364	42.164	17.562	0.117/14.256	0.078/9.241	0.009/5.239
Beijing PM2.5 data	0	47	22	25	44.986	29.716	13.224	0.074/10.968	0.044/6.562	0.008/5.794
	1	52	30	24	48.943	33.846	14.267	0.085/11.798	0.049/7.079	0.007/5.149
	2	57	38	23	52.894	36.793	15.095	0.096/12.097	0.053/7.946	0.006/4.956
	3	63	45	24	58.953	40.984	15.896	0.105/12.894	0.061/8,448	0.007/5.362
	4	75	52	26	63.987	44.867	16.574	0.115/13.946	0.066/8.964	0.008/5.895
	5	89	58	33	68.794	48.946	18.769	0.129/14.874	0.075/9.045	0.008/5.694
	6	104	64	37	74.892	53.896	20.562	0.136/15.896	0.089/9.674	0.006/4.892
	7	120	74	42	79.423	58.462	21.639	0.154/16.429	0.094/10.264	0.007/5.236
PPPTS	0	146	41	22	34.562	32.189	23.793	0.126/7.756	0.053/5.019	0.005/2.876
	1	152	49	25	40.925	37.964	23.896	0.132/8.046	0.063/6.946	0.005/2.894
	2	159	53	31	49/673	42.985	24.795	0.138/8.953	0.069/7.062	0.005/2.763
	3	164	65	38	55.896	48.779	25.963	0.145/9.689	0.076/7.462	0.006/2.845
	4	175	74	46	64.864	53.156	26.963	0.158/10.563	0.084/7.953	0.007/3.146
	5	196	86	53	76.894	58.146	28.953	0.164/11.876	0.097/8.263	0.007/3.056
	6	215	91	66	81.794	62.986	30.634	0.174/12.895	0.106/8.632	0.006/2.964
	7	236	104	79	88.462	66.214	33.156	0.177/12.479	0.114/9.014	0.008/3.269

quantitative association rule mining algorithm based on the hard partition for discretization, which reduces the prediction accuracy for the interval boundary. The FPS algorithm is a quantitative fuzzy association rule mining algorithm, but it does not address the concept drift problems. In contrast, concept drift detection is integrated into the IFARE algorithm, which improves the prediction accuracy.

In addition, the number of rules derived from every algorithm corresponding to the best classification and prediction performance and the running time are compared with a different algorithm in different datasets. As noted in Table 5 and 6, the IFARE algorithm obtains the least number of rules because the causal indicators ensure that the causal fuzzy association rule mining generates fewer rules with sufficient accuracy. It can be observed that the time cost of the IFARE algorithm is less than that of the other algorithms because the IFARE algorithm adopts concept drift detection to reduce the time cost for remining the rules.

To verify the effectiveness of IFARE in the mining association rules, the classification accuracy, prediction error and number of rules are compared with those of other algorithms with datasets SEA and Beijing PM2.5 for each sliding window. It can be observed from Fig. 2 that compared with IUAC and AG-FFP, the IFARE algorithm has a higher classification accuracy with the best Acc and RI for dataset SEA. As noted from Fig. 3, when compared with the FPS and QART algorithms, the IFARE has lower prediction error with the lowest MAPE and RMSE for dataset Beijing PM2.5. From Fig. 4, it can be observed that the number of rules derived from the IFARE algorithm is lower than those of the other classification and regression prediction algorithms. For the dataset SEA, the number of the rules in the sliding windows SW^2 is the same as that in SW^3. Similarly, for the dataset (Beijing PM2.5 data), the number of rules in the sliding windows SW^6 is the same as that in SW^7 because concept drift does not occur in the current sliding window during the mining process, and classification and regression prediction can be implemented with the previous rules.

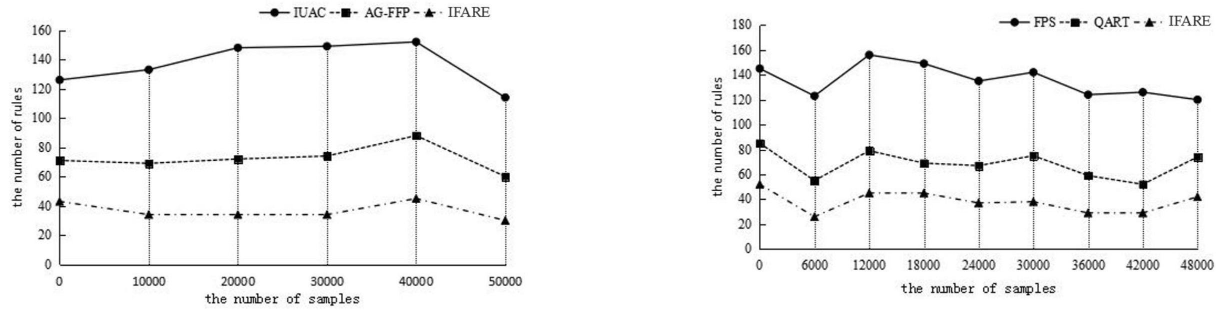
Further, to verify the validity of the rules proposed in this paper, the mined partial fuzzy association rules derived from the IFARE algorithm are compared with those from the other algorithms. The datasets Wine and Beijing PM2.5 are taken



(a) MAPE

(b) RMSE

FIGURE 3. The prediction errors of dataset Beijing PM2.5.



(a) dataset SEA

(b) Beijing PM2.5

FIGURE 4. The number of rules in different sliding windows.

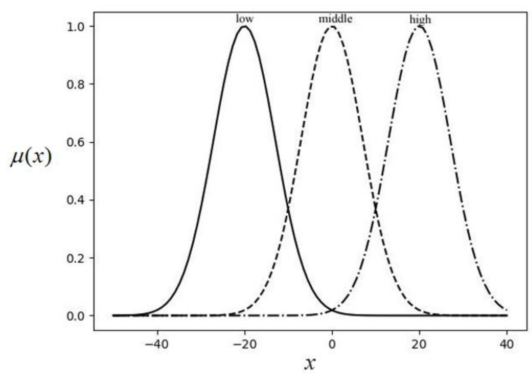


FIGURE 5. The membership function of the variable DEWP.

as examples. As shown in Table 7, the fuzzy association rules obtained by the IFARE algorithm are shorter than those obtained by the AG-FFP algorithm and are more interpretable and easier to match with the data to improve the classification accuracy of the rules. Similarly, as shown in Table 8, the antecedent of the rules obtained by IFARE is shorter than those of FPS and QART. According to the consequent of the rules, the fitting error obtained by IFARE is less than that of the other algorithms. The generation of the antecedent of the fuzzy rule is described as follows. For example, in Fig. 5, after discretization, the variable DEWP is clustered into three categories, which can be represented by the three states, *low*, *middle*, *high*, by a membership function. For Beijing PM2.5, the fuzzy association rule obtained in Table 6 can be described as follows:

$$\begin{aligned} &DEWP \text{ is } middle,\ TEMP \text{ is } low \\ &\qquad \rightarrow pm2.5 = 0.89 + 0.59 \times DEWP + 0.75 \times TEMP \end{aligned}$$

If the medium dew point (DEWP) is accompanied by low temperature (TEMP), then the pollutant concentration (PM2.5) can be predicted with the consequent function of the rules.

C. THE RULE WEIGHT ANALYSIS

The weighting concept of incremental rules is an attempt to reduce the curse of rules. Continuous weights are assigned in [0, 1], which represents the importance levels of the rules in the dynamic systems. The weights of rules close to 0 means that their importance level is relatively lower than others that have weights near 1. The main advantage of rules weighting over rules selection becomes clear within an incremental learning scenario.

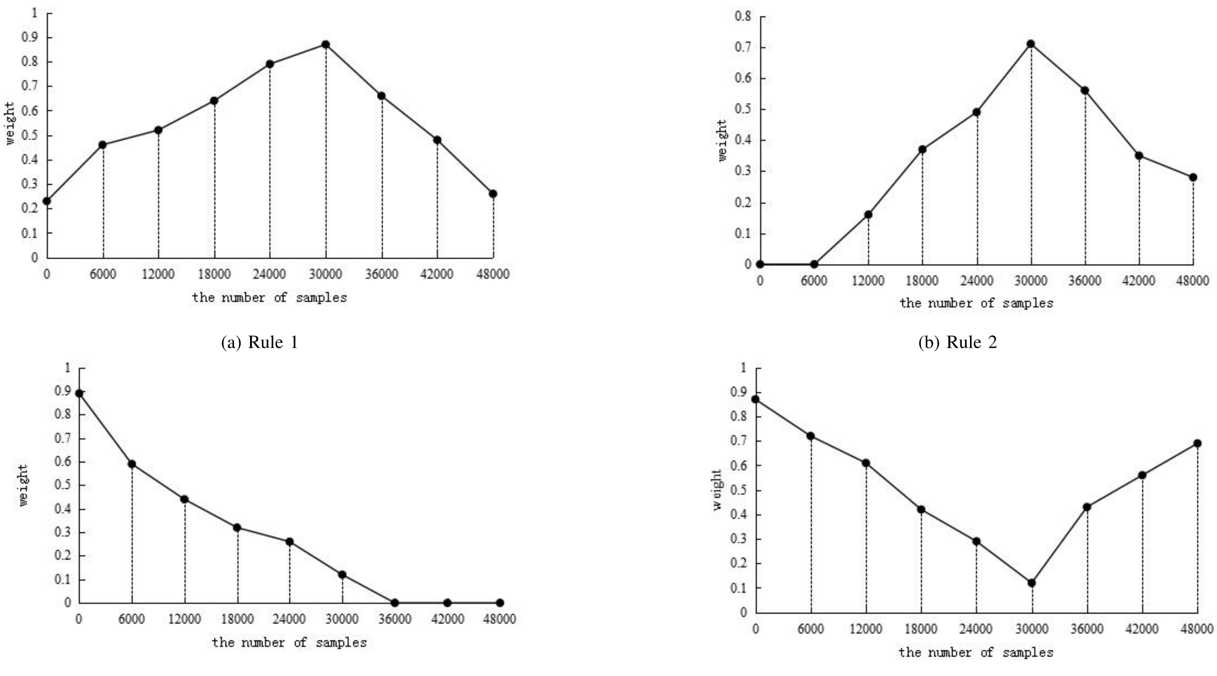
Here, we analyze the effect of the variable weight of different rules. The dataset PPPTS is partitioned by eight sliding windows, and each window consists of 6000 samples. As shown in Fig. 6, the weight of rule 1 increases at the beginning and decreases gradually. Rule 2 does not appear in the sliding window SW^1, and with data updating, the weight

TABLE 7. Comparison of rules for different algorithms in dataset Wine.

algorithm	rule
IUAC	$11.03 < Alcohol < 12.46, 2.12 < Ash < 2.59, 0.77 < Hue < 1.21 \rightarrow class = 1$
AG-FFP	$Alcohol\ \ is\ \ low,\ \ Ash\ \ is\ \ middle,\ \ Hue\ \ is\ \ middle \rightarrow class = 1$
IFARE	$Alcohol\ \ is\ \ low,\ \ Hue\ \ is\ \ middle \rightarrow class = 1$

TABLE 8. Comparison of rules for different algorithms in dataset Beijing PM2.5.

algorithm	rule
FPS	$DEWP\ \ is\ \ middle,\ \ TEMP\ \ is\ \ low,\ \ Iws\ \ is\ \ middle \rightarrow$ $pm2.5 = 0.64 + 0.79 \times DEWP + 0.58 \times TEMP + 0.64 \times Iws$
QART	$-20.6 < DEWP < 0.46, -19 < TEMP < 2.64islow, 146.39 < Iws< 332.16 \rightarrow$ $pm2.5 = 0.33 + 0.55 \times DEWP + 0.78 \times TEMP + 0.69 \times Iws$
IFARE	$DEWP\ \ is\ \ middle,\ \ TEMP\ \ is\ \ low \rightarrow$ $pm2.5 = 0.89 + 0.59 \times DEWP + 0.75 \times TEMP$



(a) Rule 1

(b) Rule 2

(c) Rule 3

(d) Rule 4

FIGURE 6. Evolution of weights of partial fuzzy association rules in dataset PPPTS.

of rule 2 increases at the beginning and decreases gradually from SW^2. The weight of rule 3 shows a downward trend and does not appear from SW^6 to SW^8. The weight of rule 4 decreases at the beginning and increases gradually.

V. CONCLUSION

Our proposed algorithm can significantly improve the classification and prediction performance based on the fuzzy association rule. In the process of the incremental mining fuzzy association rules, the expected frequent itemsets are obtained by calculating the probability of the non-frequent itemsets in the current sliding window and are assessed as to whether they are expected to be frequent itemsets in the next sliding window. Based on this information, the dataset only needs to be scanned once to obtain compact and comprehensive fuzzy association rules. In particular, the fuzzy association rules can be incrementally mined and reconstructed to realize classification and regression according to the causal index and concept drift detection, which improves the interpretability of the mining rules. The experimental results show that the IFARE algorithm can significantly improve the classification and prediction performance, and the execution time of the algorithm is greatly shortened.

The main contribution of this paper is that the fuzzy association rules can be incrementally mined and reconstructed to realize classification and regression according to causal index and concept drift detection. However, certain problems remain to be solved. First, the optimal minimum support confidence and causal relationship threshold should be studied. Second, the threshold of concept drift needs to be set by experience. Third, we plan to apply our proposed algorithm to other domains to solve more complex problems.

APPENDIX A

Let D_i be the i-th sample in the dataset D, and h be the subset $\{\tilde{X}_{j,f_{j,h}}, \tilde{X}_{k,f_{k,h}}, \tilde{Y}_h\}$. The fuzzy equivalence class generated by D_i, and h is represented as $[D_i]_h$.

$$[D_i]_h = \frac{s(D_1, D_i)_h}{D_1} + \frac{s(D_2, D_i)_h}{D_2} + \ldots + \frac{s(D_k, D_i)_h}{D_k} + \ldots + \frac{s(D_n, D_i)_h}{D_n} \quad (47)$$

where $s(D_k, D_i)_h$ is the relation value between $D_k$and D_i generated by h:

$$s(D_k, D_i)_h = \frac{1}{3}\sum_{j=1}^{L}\sum_{q=1}^{C_j}\left(1 - \left|\mu_{\tilde{X}_{jq}}(x_{ij}) - \mu_{\tilde{X}_{jq}}(x_{kj})\right|\right), \left(\tilde{X}_{jq} \in h, k \neq i\right) \quad (48)$$

where $\tilde{X}_{jq}$ is the qth fuzzy feature of the jth feature, $\mu_{\tilde{X}_{jq}}(x_{ij})$ is the membership value of the jth feature value of the ith sample corresponding to the fuzzy feature $\tilde{X}_{jq}$, $\mu_{\tilde{X}_{jq}}(x_{kj})$ is the membership value of the jth feature value of the kth sample corresponding to the fuzzy feature $\tilde{X}_{jq}$, and B_h^i is the cardinality of subset h corresponding to the ith sample:

$$B_h^i = |[D_i]_h| = \sum_{k=1}^{n} s(D_k, D_i)_h = \sum_{k=1}^{n}\frac{1}{3}\sum_{j=1}^{L}\sum_{j=1}^{C_j}\left(1 - \left|\mu_{\tilde{X}_{jq}}(x_{ij}) - \mu_{\tilde{X}_{jq}}(x_{kj})\right|\right), \left(\tilde{X}_{jq} \in h, k \neq i\right) \quad (49)$$

where n is the number of the samples collected, $\tilde{X}_{jq}$ is the qth fuzzy item of the jth variable, $\mu_{\tilde{X}_{jq}}(x_{ij})$ is the membership value of the jth variable of the ith sample corresponding to the fuzzy item $\tilde{X}_{jq}$, and $\mu_{\tilde{X}_{jq}}(x_{ij})$ is the membership value of the jth variable of the ith sample corresponding to the fuzzy item $\tilde{X}_{jq}$. In a similar manner, the entropy of other fuzzy items can be calculated.

APPENDIX B

The sliding window is a technique that is widely applied in dynamic circumstances. Each sliding window reads the new samples and discards the past data. In this work, the range of features R is considered with the Hoeffding bound ς [39] to adjust the mean value of samples in the sliding window and determine the size of the window. Suppose that there are $n^{(t)}$ samples in the current sliding window, and the conference is $1 - \delta$(δ is generally set to 0.05). The difference between the calculated mean value and actual mean value of a feature does not exceed ς, which is computed by the following:

$$\varsigma = \sqrt{\frac{R \cdot R^T \mathrm{In}(1/\delta)}{2n^{(t)}}} \quad (50)$$

where R is the range of the characteristic:

$$R = [x_{1.\max} - x_{1.\min}, x_{2.\max} - x_{2,\min}, \cdots, x_{j,\max} - x_{j,min}, \cdots, x_{L,\max} - x_{L,\max}] \quad (51)$$

where $x_{j,\max}$ is the max value of the variable j, $x_{j,\min}$ is the minimum value of the variable j, and L is the number of variables in the data set.

From the Hoeffding bound, the minimum window size N_H [36] can be determined by:

$$N_H = \frac{R^2 \ln(1/\delta)}{2\zeta^2} \quad (52)$$

As shown, the Hoeffding boundary ζ is the key factor. To obtain ζ, we assume that $SW^{(t-1)}$ and $SW^{(t)}$ are two adjacent sliding windows, and that $\overline{SW}^{(t-1)}$ and $\overline{SW}^{(t)}$ are their respective means, with probabilities of $1-\delta$ respectively. If $\left|\overline{SW}^{(t-1)} - \overline{SW}^{(t)}\right| \leq 2\zeta$, from (51) and (52), we have

$$N_H = \frac{2R^2 \ln(1/\delta)}{\left(\overline{SW}^{(t-1)} - \overline{SW}^{(t)}\right)^2} \quad (53)$$

If the number of samples contained in the current sliding window $n^{(t)}$ is not less than N_H, $n^{(t)}$ is referred to as a fixed sliding window size. During the process of incremental fuzzy association rule mining, the sliding window manages the input data and stores the information in an efficient manner.

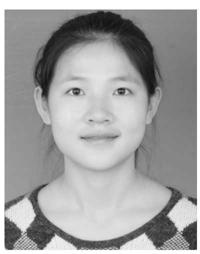
REFERENCES

[1] B. Liu, W. Hsu, and Y. Ma, "Integrating classification and association rule mining," in *Proc. Int. Conf. Knowl. Discovery Data Mining*, Aug. 1998, pp. 80–86.
[2] W. Li, J. Han, and J. Pei, "CMAR: Accurate and efficient classification based on multiple class-association rules," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2001, pp. 369–376.
[3] F. Thabtah, "A review of associative classification mining," *Knowl. Eng. Rev.*, vol. 22, no. 1, pp. 37–65, Mar. 2007.
[4] J. Han and X. Yin, "CPAR: Classification based on predictive association rules," in *Proc. SIAM Int. Conf. Data Mining. Soc. Ind. Appl. Math.*, vol. 24, 2003, pp. 236–255.
[5] H. Wang, J. Klinginsmith, X. Dong, A. C. Lee, R. Guha, Y. Wu, G. M. Crippen, and D. J. Wild, "Chemical data mining of the NCI human tumor cell line database," *J. Chem. Inf. Model.*, vol. 47, no. 6, pp. 2063–2076, Nov. 2008.
[6] M. Antonelli, P. Ducange, F. Marcelloni, and A. Segatori, "A novel associative classification model based on a fuzzy frequent pattern mining algorithm," *Expert Syst. Appl.*, vol. 42, no. 4, pp. 2086–2097, Mar. 2015.
[7] S. Naresh, M. Vijaya Bharathi, and S. Rodda, "An efficient framework for building fuzzy associative classifier using high-dimensional dataset," in *Proc. 49th Annu. Conv. Comput. Soc. India*, 2015, pp. 641–651.
[8] M. Elkano, M. Galar, J. A. Sanz, A. Fernandez, E. Barrenechea, F. Herrera, and H. Bustince, "Enhancing multiclass classification in FARC-HD fuzzy classifier: On the synergy between$n$-dimensional overlap functions and decomposition strategies," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 5, pp. 1562–1580, Oct. 2015.
[9] L. Wang, J. Y. Dong, and S. L. Li, "Fuzzy inference algorithm based on quantitative association rules," *Procedia Comput. Sci.*, vol. 61, pp. 388–394, Jul. 2015.
[10] S. Kar and D. D. Majumder, "An investigative study on early diagnosis of prostate cancer using neuro-fuzzy classification system for pattern recognition," *Int. J. Fuzzy Syst.*, vol. 19, no. 2, pp. 423–439, Apr. 2017.
[11] I. Norhayati and M. Rashid, "Adaptive neuro-fuzzy prediction of carbon monoxide emission from a clinical waste incineration plant," *Neural Comput. Appl.*, vol. 30, no. 10, pp. 3049–3061, Nov. 2017.
[12] I. Gaied, F. Jemili, and O. Korbaa, "A genetic-fuzzy classification approach to improve high-dimensional intrusion detection system," in *Proc. Int. Conf. Intell. Syst. Design Appl.*, 2016, pp. 319–329.
[13] O. Cordón, F. Herrera, J de la Montana, A. M Sánchez, and P. Villar, "A prediction system for cardiovascularity diseases using genetic fuzzy rule-based systems," in *Proc. Ibero-Amer. Conf. Artif. Intell.*, 2002, pp. 381–391.

[14] E. Ozkan, M. Iphar, and A. Konuk, "Fuzzy logic approach in resource classification," *Int. J. Mining, Reclamation Environ.*, vol. 33, no. 3, pp. 183–205, 2017.
[15] M.-Y. Su, G.-J. Yu, and C.-Y. Lin, "A real-time network intrusion detection system for large-scale attacks based on an incremental mining approach," *Comput. Secur.*, vol. 28, no. 5, pp. 301–309, Jul. 2009.
[16] M.-Y. Su, S.-C. Yeh, K.-C. Chang, and H.-F. Wei, "Using incremental mining to generate fuzzy rules for real-time network intrusion detection systems," in *Proc. Int. Conf. Adv. Inf. Netw. Appl. Workshops*, Mar. 2008, pp. 50–55.
[17] J.-J. Lu and B.-W. Xu, "Incremental update for fuzzy association rules," in *Proc. Int. Conf. Mach. Learn. Cybern.*, vol. 4, Nov. 2003, pp. 2591–2595.
[18] L. Wang and J. Meng, "incremental clustering algorithm of adaptive resonance theory based on local distribution," *Control Decision*, vol. 33, no. 3, pp. 471–478, 2018.
[19] R. Amornchewin and W. Kreesuradej, "Mining dynamic data-bases using probability-based incremental association rule discovery algorithm," *Univers Comput. Sci*, vol. 15, no. 12, pp. 2409–2428, 2009.
[20] A. Ariya and W. Kreesuradej, "An enhanced incremental association rule discovery with a lower minimum support," *Univers Comput. Sci*, vol. 21, no. 4, pp. 466–477, Dec. 2016.
[21] Z. Ling, "The research and application of incremental updating algorithm for fuzzy association rules," M.S. thesis, School Comput. Appl., Southeast Univ., Nanjing, China, 2007.
[22] C. Silverstein, S. Brin, R. Motwani, and J. Ullman, "Scalable techniques for mining causal structures," *Data Mining Knowl. Discovery*, vol. 4, nos. 2–3, pp. 163–192, 2000.
[23] X. Zhang, L. Zhao, J. Zang, H. Fan, and L. Cheng, "Flatness intelligent control based on TS cloud inference neural network," *ISIJ Int.*, vol. 54, no. 11, pp. 2608–2617, Nov. 2014.
[24] Drucker H, Burges C J C, Kaufman L, and, "Support vector regression machines," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 2409–2428.
[25] L. Wang, Z. C. Mu, and H. Guo, "Fuzzy rule-based support vector regression system," *J. Control Theory Appl.*, vol. 3, no. 3, pp. 230–234, Aug. 2005.
[26] C. T. Lin, S. F. Liang, C. M. Yeh, and K. W. Fan, "Fuzzy neural network design using support vector regression for function approximation with outliers," in *Proc. IEEE Int. Conf. Syst., Man, Cyber*, Oct. 2005, pp. 2763–2768.
[27] J. M. Leski, "TSK-fuzzy modeling based on $\sum$-insensitive learning," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 2, pp. 181–193, 2005.
[28] C. F. Juang and C. D. Hsieh, "TS-fuzzy system-based support vector regression," *Fuzzy Sets Syst.*, vol. 160, no. 17, pp. 2486–2504, May 2009.
[29] C. F. Juang and C. D. Hsieh, "A locally recurrent fuzzy neural network with support vector regression for dynamic-system modeling," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 2, pp. 261–273, Apr. 2010.
[30] C. F. Juang, R. B. Huang, and W. Y. Cheng, "TSK-fuzzy modeling based on σ-insensitive learning," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 4, pp. 689–699, Apr. 2010.
[31] J. H. Chiang and P. Y. Hao, "Support vector learning mechanism for fuzzy rule-based modeling: A new approach," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 1, pp. 1–12, Feb. 2004.
[32] D. I. Kim, "Response modeling with semi-supervised support vector regression," *J. Korea Soc. Comput. Inf.*, vol. 19, no. 9, pp. 125–139, 2014.
[33] L. Wang and J. Meng, "Dynamic fuzzy features selection based on variable weight," *J. Comput. Res. Develop.*, vol. 55, no. 5, pp. 893–907, 2018.
[34] H.-L. Chen, M.-S. Chen, and S.-C. Lin, "Catching the trend: A framework for clustering concept-drifting categorical data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 5, pp. 652–665, Apr. 2009.
[35] R. B. V. Subramanyam and A. Goswami, "Mining fuzzy quantitative association rules," *Expert Syst.*, vol. 23, no. 4, pp. 212–225, May 2010.
[36] L. Wang, S. Li, and L. Wu, "Classification and regression prediction algorithm based on quantitative association rule tree," *J. Univ. Sci. Technol. Beijing*, vol. 38, no. 6, pp. 886–892, Apr. 2016.
[37] A. Asuncion and D. J. Newman, "UCI machine learning repository irvine," School Inf. Comput. Sci., Univ. California, Berkeley, Berkeley, CA, USA, 2007.
[38] A. Ghazikhani, R. Monsefi, and H. S. Yazdi, "Online neural network model for non-stationary and imbalanced data stream classification," *Int. J. Mach. Learn. Cybern.*, vol. 5, no. 1, pp. 51–62, 2014.
[39] L. Du, Q. Song, and X. Jia, "Detecting concept drift: An information entropy based method using an adaptive sliding window," *Intell. Data Anal.*, vol. 18, no. 3, pp. 337–364, 2014.

LING WANG received the B.E. degree in electric automation and the M.E. degree in control theory and engineering from the Inner Mongolia University of Science and Technology, China, in 1996 and 2002, respectively, and the Ph.D. degree in control theory and engineering from the University of Science and Technology Beijing, Beijing, China, in 2007, where she is currently an Associate Professor with the School of Automation and Electrical Engineering. Her research interests include computational intelligence, and other machine learning techniques and their applications.

QIAN MA received the M.E. degree with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China. Her current research interest includes data mining.



JIANYAO MENG received the M.E. degree with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China. Her current research interest includes data mining.