Diversification of Fuzzy Association Rules to Improve Prediction Accuracy

Bilal I. Sowan, Keshav P. Dahal, Alamgir M. Hossain, and Mohammad S. Alam

Abstract—Reliability of the Fuzzy Association Rules (FARs) extraction is a challenging research in knowledge discovery and data mining. Reliability refers to the trade-off between the prediction accuracy and the rules diversity. In this paper, an approach called Diverse Fuzzy Rule Base (DFRB) is proposed to extract the FARs which are used later to predict the future values. This approach also aims to ensure high quality and diversity of the FARs. This is achieved through four phases: firstly, the integration of Fuzzy C-Means (FCM) and Multiple Support Apriori (MSapriori) algorithms are applied to extract the FARs. The second phase calculates the correlation values for these FARs, and performs an efficient orientation for filtering FARs as a post-processing method. In the third phase, the FARs diversity is maintained through the clustering of FARs, based on the concept of the sharing function technique used in multi-objectives optimization. After that, the best and the most diverse FARs are evaluated and then stored in the Knowledge Base (KB). Finally, these FARs, stored in the KB, are utilized within the Fuzzy Inference System (FIS) for prediction. Experimental results for two case studies have shown that the proposed DFRB approach predicted the future values effectively, thus, outperforming the existing work.

1. INTRODUCTION

Association mining remains one of the most important techniques in Knowledge Discovery in Databases (KDD), increasingly attracting the attention of researchers. Mining of the association rules can be applied to predict the future value in a particular domain. Association rules can be constructed as: \( X \rightarrow Y \) where \( X \) is the antecedent (left hand side) and \( Y \) is the consequent (right hand side). The association rules are extracted from a database which relies on two metrics (interestness measures): Minimum Support threshold (\( \text{minsupp} \)) and Minimum Confidence threshold (\( \text{minconf} \)). Since \( \text{minsupp} \) is the time frequency for particular item(s) that are divided by the number of transactions, and \( \text{minconf} \) is the percentage value of the support value for the antecedent and consequent together, divided by the support value of the antecedent part. Association rules mining can be divided into two main parts [1]:

- \( \text{minsupp} \): Frequent item sets are found if the support value of the itemset is greater than or equal to \( \text{minsupp} \).
- \( \text{minconf} \): Association rules are extracted if the confidence value of the frequent itemset is greater than or equal to \( \text{minconf} \).

The definition of the association rules is as follows: let \( I = \{i_1, i_2, \ldots, i_n\} \) be a set of distinct items (attributes), and any set of items is called an itemset. Let \( D = \{t_1, t_2, \ldots, t_m\} \) be a set of transactions database \( \text{TID} \). Each transaction \( \text{TID} \) in \( D \) consists in a set of items \( I \), strong association rules are defined in accordance to the form \( X \rightarrow Y \) where, \( X, Y \subseteq I \) and \( X \cap Y = \emptyset \) approve \( \text{minsupp} \) and \( \text{minconf} \).

Most of the common association rules algorithms are based on level-wises, such as Apriori [2], and others use the tree structure namely Frequent Pattern Growth (FP-Growth) [3]. As a concept derived from the association rules, Fuzzy Association Rule (FAR) applies a fuzzy approach to deal with quantitative attributes (quantitative data or crisp one) and to represent them in a natural and understandable manner.

Fuzzy approach is widely exploited among the intelligent systems, since it is very simple and similar to the human way of thinking [4]. Furthermore, it helps to extract abstract patterns at a higher level than just on the data one (crisp data set). These are then used to transform quantitative data into fuzzy data through the identification of membership functions. The membership functions are defined using different methods, basically, based on human experience/know-how [5], [6]. However, in many occasions, it is too difficult to obtain the information required and the human experts. The Fuzzy C-Means (FCM) is a clustering technique commonly used to provide reliable membership values for constructing fuzzy sets [5]-[7] if data sets are available.

The use of single \( \text{minsupp} \) for a whole database considers and assumes that all items in the database have the same frequency. However, in real applications, the database contains some items of a high frequency, while others are of a low frequency. The human expert, based on domain knowledge, can set \( \text{minsupp} \) for a specific value in order to find the frequent items. In that case, if \( \text{minsupp} \) is set too high it will extract a low number of frequent items. On one hand, the rare items will appear and cause a dilemma (called rare item problem). On the other hand, if \( \text{minsupp} \) is set too low, it will extract a high number of frequent items, which causes combinatorial explosions. In other words, all the possible associations will be found; thus, some of the items are uninteresting/insignificant [8]. In the literature, the rare item problem can be solved through the adoption of two methods. Firstly, the data set is divided into blocks according to their occurring item frequency, and then the association rules are extracted in each block with different \( \text{minsupp} \). Secondly, the related rare items are grouped into one abstract item; therefore, this abstract item becomes a frequent one. It would not be possible for the first method to extract the association rules among these blocks, and the second method is suffering from extracting rules that involve
individual rare items and frequent items. To overcome the dilemma of rare item problem, Liu et al. [8] proposed an algorithm called MSapriori using Multiple Item Support (MIS). However, the number of generated rules depends on the control parameters used. For instance, if the parameter values are set low, then they tend to extract a large number of rules, most of which are not useful. Otherwise, a small number of rules are generated that do not adequately cover the data.

It is believed that such an approach is applied for extracting the FARs in order to store it in the knowledge Base (KB) and use it in the prediction of future values. This can be accomplished and achieved through FARs assessments, while considering the diversity within the FARs. The prediction system can therefore accept any input data to predict the output of future values. The best FARs, in some cases, do not always lead to good results; therefore, the diversified FARs can offer better results.

In this paper, an approach called Diverse Fuzzy Rule Base (DFRB) is proposed which aims to extract a robust and diverse fuzzy rule base that enables the prediction of future values effectively. These fuzzy rules base are generated from FARs, and should be sustained and proved. The proposed approach facilitates the trade-off between both the prediction accuracy and the diversity within FARs.

This paper is organized as follows. Section II presents the related works. Section III describes the proposed approach. Section IV reports the experimental results and evaluates the performance study. Finally, section V draws the conclusions.

II. RELATED WORKS

Mining association rules is a vital task in data mining research. Several algorithms and approaches are proposed for generating association rules which have support and confidence values higher than user-specified thresholds [9]. Similarly, there are several techniques that can be conducted to prune the huge number of such association rules and transform them into more representative ones [9]-[11]. Marzena [9] introduced an approach to obtain representative rules from a large set of association rules using cover operator; these rules are based on satisfying the minsupp and minconf measures. The use of such measures to generate these rules could be affected at the representative rules level. The representative rules implicates small numbers of rules, which decrease the accuracy when they are evaluated and validated. Many techniques were offered by scholars suggesting objective measures to evaluate the association rules [12], [13]. Lenca et al. [14] proposed an approach for selecting the most interesting rules based on improvement of the objective measurements. Their approach used a Multi-Criteria Decision Aid (MCDA) method to sustain these rules for a non-expert user in a specific domain.

A clustering technique, for association rules, was developed by Dechang and Xiaolin [15] based on the similarity of the antecedent part of the rules. The rules and its similarity coefficient are stored in the fuzzy simulation matrix to be understood by the user and then evaluated. However, the use of the Apriori algorithm for extracting these rules and the similarity of the antecedent parts only affects the result performance.

Various approaches have been developed in order to extract FARs from quantitative data sets. These rules are presented in the form of “IF-Then” statements using statistical significant methods as minsupp and minconf. Huang et al. [6] proposed a fuzzy data mining approach to discover rules by applying the Apriori algorithm while adapting the discovered rules for the training Adaptive Network based on Fuzzy Inference System (ANFIS). The approach is applied in the human resources department for predicting future employee performance, in either suitable projects or positions. This approach [6] was tested for a small data set with some noise. It uses human expertise/experience (knowledge) to define the membership functions. It could be adapted for a small data set, but it is not feasible for large data sets. An approach was implemented by Lu et al. [5] which compared two methods for the prediction of one output value, this was applied to Abalone data set taken from the University of California, Irvine (UCI) of Machine Learning Repository. In the first method, FCM and Apriori algorithms were used to extract FARs and then a Genetic Algorithm (GA) was applied for tuning the fuzzy sets. The second method proceeded as the first, but it used variable thresholds in the prediction. The prediction accuracy in the first approach was 72.16% for training and 70.56% for testing data; while the second approach obtained 72.55% for training and 71.06% for testing data.

Associative classification concept has also been widely investigated. Pach et al. [16] proposed a fuzzy associative classification technique. This technique considers both rules: the interpretability and classification accuracy, but it used the Apriori approach for extracting frequent items.

Most of the above mentioned approaches suffer from one or more of the following problems:

- The single minsupp approach is not fair when using a single minsupp for the whole data. Single minsupp assumes the same frequency for all items in the data. In this manner, the real application data possesses some items with a high frequency; while others possess a low frequency. Therefore, by using a single minsupp some of the significant association rules could be missed.
- The use of such an objective measure to only assess the rules is not effective. It depends on selecting and tuning a parameter threshold to extract the best rules.
- Association rule mining techniques produce many association rules that will affect the prediction results.
- Extraction of representative rules may assist in understanding the perspective points of the user. Yet, it is not dependable and coverable in case of prediction accuracy.

III. THE PROPOSED DIVERSE FUZZY RULE BASE APPROACH

The proposed Diverse Fuzzy Rule Base (DFRB) approach considers both the trade-off between the prediction accuracy and the diversity of FARs to provide robustness in the fuzzy rules base. This approach is based on selecting the strong
FARs in order to increase prediction accuracy and the FARs diversity to maintain more representative rules. As described in Fig. 1, the proposed approach consists of four phases; each phase is detailed as follows:

**First phase:** Generating the FARs based on FCM and MSapriori algorithms. FCM is employed as an automatic system which transforms the quantitative data set into fuzzy sets (terms). The MSapriori algorithm is applied to extract the FARs by setting the control parameters, used in MSapriori, at low values. Thus, the number of generated FARs increases.

**Second phase:** Calculating the correlation coefficient value for each FAR. Correlation values are calculated for each of the FARs using equations (1) and (2). One of the correlation measures is used to evaluate the importance and strength of the association rule [16]-[18] in order to filter a large number of association rules over different objective measures.

\[
\text{corr}(X. \text{FS} \rightarrow Y. \text{FS}) = \frac{\text{supp}(X. \text{FS} \land Y. \text{FS}) - \text{supp}(X. \text{FS}) \cdot \text{supp}(Y. \text{FS})}{\sqrt{\text{supp}(X. \text{FS}) \cdot (1 - \text{supp}(X. \text{FS})) \cdot \text{supp}(Y. \text{FS}) \cdot (1 - \text{supp}(Y. \text{FS}))}}
\]

**Third phase:** The FARs with positive correlation values are subsequently considered, they are grouped according to their length size \(K\). The FAR length is determined based on the number of attributes included. For example,

- **group1**
  - \(r_1: X. \text{Low} \rightarrow Y. \text{Low}\)
  - \(r_2: Z. \text{Low} \rightarrow U. \text{Medium}\)

- **group2**
  - \(r_1: X. \text{Low} \text{ and E. Medium} \rightarrow Y. \text{Medium}\)
  - \(r_2: Z. \text{Medium and W. Medium} \rightarrow U. \text{High}\)

The correlation value is normalized for each FAR within the group as follows:

1. Determine the maximum correlation value \(\text{MaxCorr}_K\)
2. Divide each FAR correlation value by \(\text{MaxCorr}_K\)

For example, \(\text{MaxCorr}_K = 0.3\) and two FARs in the group

\[\text{MaxCorr}_K(r_1) = 0.3 = 1\]
\[\text{MaxCorr}_K(r_2) = 0.2 = 0.66\]

This normalization is applied to rescale the correlation value for each FAR in each group. The reason behind the length size is that the longer the FAR size increases the smaller the correlation value is. As a result, this assists in selecting the best FAR from each group that satisfies the minimum correlation threshold \(\text{minCorr}\).

The motivations behind using the correlation measure are:
- The tuning and setting the \(\text{minsupp} \) and \(\text{minconf} \) have some difficulties [16], [19].
The correlation measure is used as a filtration for the generated FARs after using MIS as minsupp and minconf.

The correlation measure is robust compared with the confidence measure of equation (3) (as mentioned in [20]).

\[ \text{Conf}(X, FS, Y, FS) = \frac{\text{Supp}(X, FS, Y, FS)}{\text{Supp}(X, FS)} \]  

(3)

**Third phase:** Finding the diversity FARs by calculating the distance between the FARs. A diversity of FARs is calculated for each of the FAR groups, it can be found through clustering FARs within each group as follows:

1. The FARs are clustered based on their distance using equations (6) and (7).
2. Clustering the FARs based on their distance, hence the number of FARs and their similarities can be identified within a cluster and other clusters using equations (4) and (5); these equations were applied [21] to maintain diversity within the sharing function technique in multi-objective optimization.

\[ \text{Range}_i = \sum_{j=1}^{N_i} \text{Accum}(\text{Distance}_{ij}) \]  

(4)

where, \( \text{Range}_i \): the value that can find out whether the cluster contains only one FAR (\( \text{Range}_i \) enumerates the number of rules similar to a rules \( i \)); if the \( \text{Range}_i \) value equal to 1 or the cluster contains more than one FAR (crowded cluster) if the \( \text{Range}_i \) value greater than 1. \( \text{Accum}(\text{Distance}_{ij}) \): the accumulative distance between \( i \) and all other FAR in the same group, whereas \( i \in \text{FARs} \).

\[ \text{Accum}(\text{Distance}_{ij}) = \begin{cases} 1 - \left( \frac{\text{Distance}_{ij}}{\sigma} \right), & \text{if Distance} \leq \sigma; \\ 0, & \text{otherwise}. \end{cases} \]  

(5)

where, \( \sigma \): the threshold value that represents the cluster size (cluster radius), its range is in the interval \([0, 1]\). \( \text{Distance}_{ij} \): the distance between two individual FARs \( i \) and \( j \), whereas \( i, j \in \text{FARs} \). Measuring the distance between two FARs is the main part at this phase using equations (6) and (7); these FARs can be clustered based on their similarities. Actually, the similarity between two FARs can be found as follows:

\[ S_{ij} = \frac{|\text{RAFS}_i \cap \text{RAFS}_j|}{|\text{RAFS}_i |} \]  

(6)

where, \( S_{ij} \): the similarity between two FARs(i and j), the interval of its range is \([0, 1]\). \( \text{RAFS}_i \) (Rule Attribute Fuzzy Set): the fuzzy set concerning an attribute within the FAR\(_i\). \( \text{RAFS}_j \): the fuzzy set concerning an attribute within the FAR\(_j\). Consequently, the distance can be calculated using equation (7).

\[ \text{Distance}_{ij} = 1 - S_{ij} \]  

(7)

The following example illustrates the calculation of the distance. Assuming the two rules FAR\(_i\) and FAR\(_j\) then:

FAR\(_i\): \( X. \text{Low} \) and \( Z. \text{Medium} \rightarrow Y. \text{Low} \)

FAR\(_j\): \( X. \text{Medium} \) and \( Z. \text{Medium} \rightarrow Y. \text{Low} \).

The similarity between FAR\(_i\) and FAR\(_j\) can be calculated by using equation (6):

\[ S_{ij} = \frac{|2|}{|4|} = 0.5 \]

The numerator 2 comes from FAR\(_i\) [Z. Medium] similar to FAR\(_j\), [Z. Medium] and FAR\(_j\), [Y. Low] similar to FAR\(_i\), [Y. Low], whereas denominator 4 comes from all attributes fuzzy sets FAR\(_i\), [X. Low, X. Medium, Z. Medium, Y. Low]

Once the similarity between FAR\(_i\) and FAR\(_j\) is calculated, then the distance can be found by using equation (7), below:

\[ \text{Distance}_{ij} = 1 - 0.5 = 0.5 \]

**Fourth phase:** Selecting the best FARs of the highest correlation values and the diverse FARs; then, store these FARs in the Knowledge Base (KB) to be used by the Fuzzy Inference System (FIS) for a prediction.

The approach for extracting robust and diverse fuzzy rule base is summarized in Fig. 2:

| Input: Fuzzy Association Rule (FAR), Minimum Correlation threshold value \( \text{minCorr} \), \( \sigma \) value, Number of the Diversed Rules (DR). |
| Output: Robust and Diverse Fuzzy Rule Base. |

**Method:**

1. Calculate the correlation value \( \text{corr} \) for each FAR\(_i\).
2. Divide FAR into a group based on their length size \( K_i \).
3. Sort (rank) FAR automatically in each group \( K_i \) based on their highest correlation value.
4. Select the best FAR from each group \( K_i \)
   - For each \( K_i \)
     - IF \( \text{corr}(\text{FAR}_i) \geq \text{minCorr} \)
       - \( \text{Best}_{FAR_i} \cup \text{FAR}_i \)
     - EndIf
   - EndFor
5. Identify the diversified FARs for each group \( K_i \)
   - \( N_{\text{Range}_i} = \text{counting the number of FARs of Range}_i \) (Range, equal 1).
   - \( DR \leq N_{\text{Range}_i} \).
   - For each \( K_i \)
     - IF \( \text{maximum} (\text{Range}_i(\text{FAR}_i)) \neq 1 \)
       - IF \( \text{corr}(\text{FAR}_i) < \text{minCorr} \)
         - While \( (\text{Diversed}_{FAR_i} \leq DR) \)
           - IF \( (\text{Range}_i(\text{FAR}_i) == 1) \)
             - \( \text{Diversed}_{FAR_i} \cup \text{FAR}_i \)
           - EndIf
         - EndWhile
     - EndIf
   - EndFor

Fig. 2. Extraction of robust and diverse fuzzy rule base algorithm.
Fig. 3 shows an example of the selection of the fuzzy rule base with robustness and diversity.

<table>
<thead>
<tr>
<th>Group 1 of FARs, length size</th>
<th>corr value</th>
<th>Range_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. Medium → D. High</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A. Low → B. Low</td>
<td>0.9</td>
<td>1</td>
</tr>
<tr>
<td>D. High → E. High</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>A. Low → C. Medium</td>
<td>0.7</td>
<td>1</td>
</tr>
<tr>
<td>C. Medium → D. Low</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>B. High → E. Low</td>
<td>0.4</td>
<td>1</td>
</tr>
</tbody>
</table>

Assuming that \( \text{minCorr}=0.7 \), and \( \text{DR}=1 \)

\[
\begin{align*}
\text{Group 2 of FARs, length size } & \text{ K_i = 3} & \text{corr value} & \text{Range_i} \\
C. \text{High and D. Medium } & \rightarrow \text{ E. High} & 1 & 1 \\
B. \text{Low and C. Medium } & \rightarrow \text{ D. Low} & 0.8 & 1.7 \\
C. \text{Medium and D. Low } & \rightarrow \text{ E. Low} & 0.6 & 2 \\
A. \text{Low and B. Low } & \rightarrow \text{ C. Medium} & 0.5 & 1 \\
B. \text{High and D. High } & \rightarrow \text{ E. Medium} & 0.4 & 1.8 \\
A. \text{Low and B. High } & \rightarrow \text{ C. Medium} & 0.4 & 1 \\
\end{align*}
\]

Fig 3. Selection of the fuzzy rule base with robustness and diversity.

IV. EXPERIMENTAL RESULTS AND PERFORMANCE STUDY

The proposed DFRB approach has been applied in two case studies in order to evaluate its effectiveness. The first case study is a quantitative road traffic data set, which was considered in our previous work [22]. The second case is a quantitative data set called Abalone. The road traffic data has been generated using a traffic simulation model (the METANET macroscopic flow model) [23]. Each record in the data set, as shown in Table II, consists of:

- Traffic state, represented by: traffic demands of road 1 (the number of vehicles that use road 1), traffic demands of road 2 (the number of vehicles that use road 2), traffic density of road 1 (the number of vehicles using road 1 per km), and traffic density of road 2 (the number of vehicles using road 2 per km).

- The Predicted Average Total Time (ATT) required for a vehicle to cross the traffic network.

Fig. 4 shows information about the input of the road traffic data set, whereas the statistical information regarding this data set is shown in Table III. While Figs. 5 and 6 show the analyses of the data set, which clarify the distribution and consistency.

<table>
<thead>
<tr>
<th>Case No.</th>
<th>Demand 1</th>
<th>Demand 2</th>
<th>Density 1</th>
<th>Density 2</th>
<th>ATT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6030</td>
<td>316</td>
<td>5</td>
<td>87</td>
<td>629.7</td>
</tr>
<tr>
<td>2</td>
<td>1147</td>
<td>1638</td>
<td>62</td>
<td>50</td>
<td>414.4</td>
</tr>
<tr>
<td>3</td>
<td>1277</td>
<td>797</td>
<td>10</td>
<td>14</td>
<td>233.5</td>
</tr>
<tr>
<td>4</td>
<td>4061</td>
<td>1198</td>
<td>38</td>
<td>75</td>
<td>581.2</td>
</tr>
</tbody>
</table>

Fig. 5 shows the analysis of the road traffic data set. A box plot (also known as a box-and-whisker diagram) is a suitable method which graphically describes groups of numerical data, such as the: minimum value, lower quartile (Q1), median (Q2), upper quartile (Q3), and the maximum value. Therefore, a box plot is used for detecting outlier (noisy) data. The data set is found to be consistent and without any outlier data. Also all fields (attributes) were distributed in different ranges of value; however, our aim is to find out if there were any outlier cases for each separate field.
We have divided the road traffic data into 75 records for training and 25 records for testing.

The Abalone data is taken from the University of California, Irvine (UCI) of Machine Learning Repository [24]. We have divided the Abalone data into 3133 records for training and 1044 records for testing.

The Proposed approach has been applied to predict the ATT in road traffic data and the Abalone ring that represents the Abalone age in Abalone data.

For the purpose of evaluation and validation, prediction quality is assessed using one of the statistical measures called Mean Absolute Percentage Error (MAPE), highlighted in equation (8). It is defined as follows:

\[
\text{MAPE} = \frac{1}{N} \sum \left( \left| \frac{PV_i - RV_i}{RV_i} \right| \times 100 \right)
\]  

(8)

where,\(PV\): the predicted output value, \(RV\): the real output value, \(N\): Total number of the comparison record.

Table IV shows the sensitivity of MAPE adapting \(\text{minCorr}\) over different numbers of FARs. On the other hand, Table V shows the sensitivity of MAPE adapting \(\text{minCorr}\) over different numbers of FARs and the diverse FARs, where \(\sigma = 0.3\) in both tables. The results shown in Tables IV and V clearly demonstrate that MAPE is not affected when the diverse FARs are added. This is explained through the usage of an efficient and effective technique to select the diverse FARs.

![Table IV](image)

<table>
<thead>
<tr>
<th>minCorr</th>
<th>Number of FARs</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>152</td>
<td>22.2%</td>
</tr>
<tr>
<td>0.5</td>
<td>88</td>
<td>9.3%</td>
</tr>
<tr>
<td>0.6</td>
<td>53</td>
<td>8.1%</td>
</tr>
<tr>
<td>0.7</td>
<td>29</td>
<td>10.6%</td>
</tr>
<tr>
<td>0.8</td>
<td>16</td>
<td>14.2%</td>
</tr>
<tr>
<td>0.9</td>
<td>8</td>
<td>23.6%</td>
</tr>
</tbody>
</table>

![Table V](image)

<table>
<thead>
<tr>
<th>minCorr</th>
<th>Number of diversified FARs</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>1</td>
<td>8.1%</td>
</tr>
<tr>
<td>0.5</td>
<td>4</td>
<td>9.9%</td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>9.9%</td>
</tr>
<tr>
<td>0.7</td>
<td>-</td>
<td>8.1%</td>
</tr>
<tr>
<td>0.8</td>
<td>-</td>
<td>8.1%</td>
</tr>
<tr>
<td>0.9</td>
<td>-</td>
<td>8.1%</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>8.1%</td>
</tr>
</tbody>
</table>

![Fig. 6](image)

Fig. 6 shows the sensitivity of the MAPE and \(\text{minCorr}\) for the brute-force approach (the diversity is employed in all FARs together of different \(K_i\) (length size of FAR)). The graph shows that when \(\text{minCorr}\) is 0.49, the minimum MAPE is 11.2%, and it contains rules that cover most cases. It is noted that as \(\text{minCorr}\) decreases from 0.49, the MAPE increases. This is explained by producing a large number of rules (decrease in \(\text{minCorr}\) is accompanied by an increase in the uninteresting FARs to cause noise for the FIS). Moreover, it is observed that when the correlation value increases beyond 0.49, the MAPE also increases. Again this is explained by producing a small number of FARs, which do not give robust results for the FIS (the increase in \(\text{minCorr}\) implies a decrease in the number of relevant rules).

Table VII depicts the sensitivity of MAPE over different numbers of FARs and diverse FARs, where \(\sigma = 0.3\) and \(\text{minCorr} = 0.49\), which illustrates the brute-force approach. This approach produces a good result; its MAPE is acceptable as compared with the employed diversity in FARs with different individual \(K_i\) (length size of FAR), as shown in Table V. The diversity is applied in FARs with different \(K_i\); they produce 8.1% MAPE and generate a slightly high number of FARs when compared with the brute-force approach, which produces 11.2% MAPE and less numbers of FARs.

![Table VI](image)
### TABLE VII

<table>
<thead>
<tr>
<th>Number of FARS</th>
<th>Number of diversified FAR</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>0</td>
<td>11.2%</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>11.2%</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>11.2%</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>18.3%</td>
</tr>
</tbody>
</table>

Other criteria measures are applied in this study to evaluate the predicted accuracy, as follows [25], [26]:

- **Mean Absolute Error (MAE)**

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |PV_i - RV_i| \tag{8}
\]

- **Normalized Mean Absolute Error (NMAE)**

\[
NMAE = \frac{\frac{1}{N} \sum_{i=1}^{N} |PV_i - RV_i|}{\frac{1}{N} \sum_{i=1}^{N} RV_i} \times 100\% \tag{9}
\]

- **Normalized Root Mean Square Error (NRMSE)**

\[
NRMSE = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{i=1}^{N} (PV_i - RV_i)^2} \times 100\% \tag{10}
\]

- **Pearson Product-Moment Correlation Coefficient r (Pearson r)**

\[
r = \frac{\sum_{i=1}^{N} (RV_i - \bar{RV})(PV_i - \bar{PV})}{\sqrt{\sum_{i=1}^{N} (RV_i - \bar{RV})^2} \sum_{i=1}^{N} (PV_i - \bar{PV})^2} \tag{11}
\]

where

- \(PV\): The predicted output value.
- \(RV\): The real output value.
- \(N\): Total number of the comparison record.
- \(\bar{RV}\): The mean of real output.
- \(\bar{PV}\): The mean of predicted output.
- \(Ucorr\): Uncorrelated = 1 - \(r\).

These essential measures are concerned with the prediction model. **MAPE, NMAE, NRMSE, NRMSE** and Pearson correlation coefficient (Pearson r) are the most widely used and important measures to evaluate the prediction method [25], [26]. The Pearson r is used to measure the relationship between two variables (the predicted and the real output value). The range of Pearson r correlation varies from -1 to +1. When r equals 0, the predicted and real output values are uncorrelated. When r equals 1, the predicted and real output values are approximately the same. And when r equals -1, the predicted and real output values are approximately the same with opposite directions. The values of such criteria are shown in Table VIII.

Table VIII shows the impact of using different criteria measurements. It is noted that different criteria yield different values. This is attributed to the different ranges of values in each data set.

The results are compared to those of the integrated Fuzzy C-Means (FCM) and Apriori approach. Prediction quality is assessed using the MAPE in equation (8). The experimental results on the Abalone and road traffic data are summarized in Table IX.

Table IX presents the results from using two different methodologies: first, it concerns the integrated FCM and Apriori; while, the second deals with FCM and DFRB. In each methodology, the MAPE is calculated, which is the result of the sensitivity improvement of future value predictions by minimizing the MAPE using the proposed DFRB approach. The result was compared with reported work, and it achieved a better result for the DFRB than the results in [5], [22]. It is noted that the result in [5] was generated by two approaches. The first approach applied FCM and Apriori algorithms, its MAPE was equal to 29.4% (its equivalent MAE was 3.77%}. While the value of MAPE using the proposed DFRB is equal to 25.5% (its equivalent MAE was 2.78%). The second approach applied FCM and Apriori algorithms, and then a Genetic Algorithm (GA) was employed in order to optimize and minimize the MAE. The MAE, based on this approach, is equal to 1.77 (it was not mentioned how many runs were experimented for the GA to optimize the MAE in [5]).

The proposed DFRB produces a good result and can be used with a wider data set (even when the data set contains noisy data). This provides more generalized prediction method. The large differences in MAPE, between the two data sets used in the experiment, refer to the high number of noisy data and high number of attributes existing in the Abalone data set. However, the results are still better than the previous work, shown in Table IX.

### V. CONCLUSION

This paper proposed the Diverse Fuzzy Rule Base approach which focused on the Fuzzy Association Rules extraction and effective selection to predict the future values. The approach concerns both the best and diverse FARs through its capability in filtering them in order to extract the
best FARs, thus reflecting an increase in the prediction accuracy. It generates the FARs, and then calculates the correlation coefficient value for each FAR. After that, it clusters the FARs to determine its diversity while selecting the best FARs (with a higher correlation value) and diverse FARs for prediction using FIS. The proposed approach expresses and maintains the trade-off between both the prediction accuracy and the FARs diversity. The approach was employed for two case studies: real data set concerning the road traffic domain, and the Abalone data set. It is noted that the proposed DFRB approach offers higher prediction accuracy as compared to other approaches reported in the literature. In the future, the associative classification techniques will be investigated to explore further levels of accuracy.

REFERENCES


