Fuzzy Rules Mining from Decision Tree for Stock Market Prediction

1 Rashid Hafeez Khokhar, 2 Mohd Noor Md Sap, 
Faculty of Computer Science and Information System 
University Technology of Malaysia, 81310 
Skudai, Johor, Malaysia 
Tel: (607)-5532419, Fax: (607) 5565044 
1 rashid@siswa.utm.my, 2 mohdnoor@fsksm.utm.my

Abstract

In this paper, data mining technique has been used to uncover hidden patterns and predict future trends and behaviors in financial markets. In data mining, new version of fuzzy decision tree (FDT) has been constructed for extraction of weighted fuzzy production rules (WFPR) by applying proposed method of fuzzy reasoning. Many existing fuzzy reasoning methods are based on Zadeh’s Compositional Rule of Inference (CRI), which requires setting up a fuzzy relation between the antecedent and the consequent part. There are some other fuzzy reasoning methods which do not use Zadeh’s CRI. Among them, the similarity-based fuzzy reasoning methods, which make use of the degree of similarity between a given fact and the antecedent of the rule to draw conclusion are well known. In proposed similarity-based fuzzy reasoning method weight parameter can be assign to each proposition in the antecedent of a fuzzy production rule (FPR) and assign certainty factor (CF) to each rule. Certainty factors have been calculated by using some important variables like effect of other companies, effect of other local stock market, effect of overall world situation, effect of political situation in stock market. The experimental results show that proposed FDT can generate relatively comprehensive tree without much computation effort and can generate WFP rules with high learning accuracy. These generated WFPRs can be used for stock market prediction.

Keywords: Data Mining, Fuzzy Logics, Decision Tree.

1.0 Introduction

Forecasting has long been the domain of linear statistics. The traditional approaches to time series prediction, such as the Box-Jenkins or ARIMA method (Box and Jenkins 1976) used by many researchers [1] [2] [3]. ARIMA assume that the time series under study are generated from linear processes. Linear models have advantages in that they can be understood and analyzed in great detail and they are easy to explain and implement. However they may be totally inappropriate if the underlying mechanism is nonlinear. It is unreasonable to assume a priori that a particular realization of a given time series is generated by a linear process. An ARIMA model is a linear non-stationary model that used different operator to convert non-stationary series to stationary. It does not work well in modeling nonlinear series by itself. In fact, real world systems are often nonlinear [4]. Several nonlinear time series models such as the bilinear model [5], the threshold autoregressive (TAR) model [6] [7] and the autoregressive conditional heteroscedastic (ARCH) model [8] [9] and its generalized version (GARCH) [10] have been developed. However these nonlinear models are still limited in that an explicit relationship for the data series at hand has to be hypothesized with little knowledge of the underlying law. In fact the formulation of nonlinear model to a particular data set is very difficult task since there are too many possible nonlinear patterns and a pre-specified nonlinear model may not be general enough to capture all the important.

Refenes et al., [11] compared regression models with a back-propagation network using the same data for stock prediction. Results showed that a back-propagation network was better predictor. Artificial Neural Networks (ANN) techniques that have been widely used for load forecasting are now used for price prediction [12-15]. In particular Ramsay et al [13] have proposed a hybrid approach based on neural networks and fuzzy logics with examples from the England-Wales market and daily mean errors around 10%. Also Szkuta et al [14] have proposed a three-layered ANN with backpropagation showing results from the Victorian electricity market with daily mean errors around 15%. Nicolaisen et al have presented Fourier and Hartley Transforms [15] as “filter” to the price data inputs of an ANN. Stochastic model of prices as in [16] are also competing with traditional time series models in order to predict daily or average weekly prices [17].

One possible explanation for above stated neural networks is that neural networks with one hidden layer can approximate a continuous function and achieve the desired accuracy [18]. However in practice, it may be invalid because no knowledge exists about the optimal structure for a special problem. Structure and training method are the determinant factors in achieving accurate forecasting results for neural networks. Although various numbers of hidden neurons are tested in the experiments and no significant improvements appear, it may be due to not finding the optimal architecture and available training methods. Another reasonable explanation is that such nonlinearity cannot improve the performance in stock return forecasting.

In this study, classification data mining technique has been used to uncover hidden rules from historical data of stock market. Among the variety of classification techniques [19] [20] [21], decision trees has been found very effect for accurate decision making. Many researchers extend the idea of decision tree by using fuzzy reasoning method and developed fuzzy decision trees (FDT) [22] [23] [24]. It can provide a high level of predictive accuracy, which rarely do they facilitate human inspection or understanding. The basic concept in FDT is that it merge fuzzy representation with its approximate reasoning capabilities and symbolic decision trees while preserving advantages of both: uncertainty handling and gradual processing of the former with the comprehensibility, popularity and ease of application of the latter. From the stock investor perspective, the simple change in the general trend (i.e., from increasing to decreasing) is very
important, since it may trigger a buying or a selling action. Therefore to analyze every simple change, powerful fuzzy reasoning method has been used in proposed algorithm.

We are using the results of our previous papers [25] [26] [27] for FDT algorithm. In this paper we extract WFPRs from FDT. Including several knowledge parameters such as weight and certainty factor can enhance the representation power of WFPR’s. Many factors are affecting stock market directly or indirectly, in this paper these factors like effect of other companies, effect of other local stock market, effect of overall world situation, effect of political situation in stock market have been analyzed and these factors use to evaluate certainty factor in WFPR’s.

The reminder of this paper is organized as follows: section 2 presents fuzzy decision tree classification algorithm by handling historical data from stock market, section 3 weighted fuzzy production rules handling, section 4 will present proposed predictive reasoning method, section 5 extract WFPR’s from FDT: experimental results and discussion, and finally conclusion in section 6.

2.0 Fuzzy Decision Tree Classification

Proposed fuzzy decision tree classification method consists of three steps: similarity-clustering, fuzzification of numerical numbers, triangular membership function, and FDT algorithm. In first step, the purpose of similarity-clustering is to compress the data set and to extract the essential concepts concerning this data set. Usually, a fuzzy attribute can take many values if the representation of the fuzzy value is given directly by a membership degree. Like numerical attributes, the range of such a fuzzy attribute can be described as the interval \([0,1]M\) where \(M\) is the dimension. For a given set of membership functions, in this research author intend to find several new fuzzy sets, which are regarded as clustering result to reasonably describe this set of membership functions. In second step, fuzzy sets have been proposed to deal with numerical and categorical attributes. In third step, triangular membership functions have been proposed to calculate fuzzy sets. By using the algorithm in our previous papers [25] [26] [27] figure 1 shows the fuzzy decision tree to train table 1.

3.0 Weighted Fuzzy Production Rules

The weighted fuzzy production rules (WFPRs) is an extended form of fuzzy production rules (FPR) proposed by Yaung et al., [28]. WFPRs defined here is similar to the conventional production rules with the exception that fuzzy values such as “fat” or “small” are allowed in the propositions. A weight is assigned to each proposition in the antecedent part, and a certainty factor is also assigned to each rule.

A WFPR is defined as: \[ R: IF a THEN c (CF = \mu), Th, \]

<table>
<thead>
<tr>
<th>Table. Training Sets with Fuzzy Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

Figure. 1. Fuzzy Decision Tree by using our proposed algorithm to train table 1
\( w \), where \( a = (a_1, a_2, \ldots, a_n) \) is the antecedent portion which comprises of one or more propositions connected by either "AND" or "OR". Each proposition \( a_i \) (\( 1 \leq i \leq n \)) can have the format "\( x \) is \( f_{ai} \)", where \( f_{ai} \) is an element of a set of fuzzy sets \( F = \{ f_1, f_2, \ldots, f_n \} \). The consequent of the rule \( c \) can be expressed, as "\( x \) is \( f_c \)" , where \( f_c \) is also an element of \( F \). The parameter \( \mu \) is the certainty factor of the rule \( R \) and it represents the strength of belief of the rule. The symbol \( Th = (\lambda_1, \lambda_2, \ldots, \lambda_n) \) represents a set of threshold values specified for the proposition in the antecedent \( a \). The set of weights assigned to the propositions \( \langle a_1, a_2, \ldots, a_n \rangle \) is given by \( w = \langle w_1, w_2, \ldots, w_n \rangle \). The weight \( w_i \) of a proposition \( a_i \) shows the degree of importance of \( a_i \) contributing to the consequent \( c \) when comparing to other proposition \( a_j \), for \( j \neq i \). It is obvious that when there is only one proposition in the antecedent of WFPR, the weight \( w_j \) is meaningless. The set of weight \( w \) assigned to each proposition in the antecedent is referred as local weights.

In general WFPR’s are categorized into three types, which are defined as follows:

Type 1: A Simple Fuzzy Production Rule

\[ R: \text{IF } a \text{ THEN } c \ (CF = \mu), \lambda, w, \]

For this type of rule, since there is only one proposition ‘a’ in the antecedent, the weight \( w \) is meaningless.

Type 2: A Composite Fuzzy Conjunction Rule

\[ R: \text{IF } a_1 \text{ AND } a_2 \text{ THEN } c \ (CF = \mu), \lambda_1, \lambda_2, w_1, w_2, \]

Type 3: A Composite Fuzzy Disjunction Rule

\[ R: \text{IF } a_1 \text{ OR } a_2 \text{ THEN } c \ (CF = \mu), \lambda_1, \lambda_2, w_1, w_2, \]

For both types 2 and 3, \( \lambda_i \) is the threshold value for \( a_i \) and \( w_i \) is the weight assigned to \( a_i \).

Some authors do not assign a certainty factor to a FPR while others ignore the weight and the threshold value assigned to each proposition in the antecedent. We consider that the capturing of fuzzy knowledge using fuzzy production rule with weights and threshold values plays an important role in real world applications. Hence the weight (degree of importance), the threshold value as well as the certainty factor have been taken into account.

3.1 Extraction of WFPRs from FDT

The transformation from the tree to WFPRs is described as follows.

1. Each path of branch from the root to a leaf can be converted into a rule. The antecedent of the rule represents the attributes on the passing branches from the root to the leaf and the consequent of the rule represents the cluster labeled at the leaf node.

2. The degree of importance of each attribute value (linguistic term) of the expanded attribute, \( \theta_j^{(k)} \), is regarded as the value of the weight \( w_{ij} \) of the corresponding proposition of the converted rule.

3. For the converted rule \( R_j \):

\[ IF \ A_{j1} \ and \ A_{j2} \ldots \ and \ A_{jn} \ THEN \ B_{j} \ ((CF, w_{j1}, w_{j2}, \ldots, w_{jn}) \) \]

in which the weight \( w_{ij} \) are obtained from step (2) and the certainty factor is defined as

\[ CF_j = \sum_{i=1}^{N} (\wedge_{j-1} (w_{ij} A_{ij} (k) \wedge B_{i}(k)) / \sum_{i=1}^{N} (\wedge_{j-1} (w_{ij} A_{ij} (k)))) \]

\( \wedge \) denotes the minimum.

3.2 Knowledge Parameters

Several knowledge parameters such as certainty factor, local weight, threshold value and global weight should be incorporated into the WFPRs

1. Certainty factor: One type of knowledge parameter called certainty factor was introduced in [29] to represent the degree of certainty of a proposition or the entire rule. Sometimes it is called MYCIN-linked certain factor [29]. Subsequently many researchers used this kind of knowledge parameter to investigate WFPRs with uncertainty.

2. Local weight: The method used in MYCIN to calculate the certain factor of the consequent suffers from the shortcomings that all propositions in the antecedent part are assumed to be equally important. To overcome this shortcoming another type of knowledge parameter related to WFPRs called local weight which indicates the degree of importance of each proposition in the antecedent was proposed by a number of researchers [30, 31].

3. Threshold value: For a given set of WFPRs, the observed fact may approximately match each of the set of rules with different degrees. That will result in some rules being misfired. To overcome this problem, several researchers employed a kind of knowledge parameter called threshold value which is assigned to each proposition [32] or assigned to entire rule [30] for the purpose of not only obtaining a reasonable result of an approximate reasoning method but also preventing or reducing rule miss-firing.

4. Global weight: This type of knowledge parameter was proposed in [28] to indicate that a number of rules executed in an reference path leading to a specific goal, or the same rule employed in various inference paths leading different goals have different degrees of importance.

The following knowledge parameters have been used to calculate the value of \( \mu \) by assigning different values to these parameters and these values are with in the range specified local thresholds. By considering four possible cases for every antecedent and consequent portion. Consequent places (conclusion) have been considering every possible fluctuation in dynamic stock market.
Therefore before rules firing user assign some specific values with in threshold range to these knowledge parameters. 


### 4.0 Proposed Predictive Reasoning (PR) Method

In proposed predictive reasoning (PR) method, first simple rule case has considered, where only one observation $A'$ and one simple rule in the form $R: A \rightarrow C$ is presented. The basic idea is to modify the consequent of $C$ a simple rule $R: A \rightarrow C$ according to the closeness of an observation (fact) $A'$ to the antecedent (pattern) $A$ of the rule $R$. If they are close (similar) enough in comparison to a threshold, then the rule $R$ can be fired and the consequent can be deduced by some modification technique to be shown later in the sequel.

Formally, their “closeness” is expressed as a similarity measure (SM), which in turn is obtained form a distance measure (DM). More specifically, given as SM we will present a threshold $\lambda_c$ of the SM. Once the $SM (A', A) = \lambda$ between $A'$ and $A$ exceeds the threshold $\lambda_c$, we fire the rule $R$, that is we construct a modification function (MF) based on $\lambda : MF = f(\lambda)$, and use this MF to modify the right side $C$ of the rule $R$ to deduce a consequent $C'$. The selection of $\lambda_c$ above 0.5 and closer to 1 assures better similarity $A$ and $A'$. Hence we believe specification of a $\lambda_c$ does improve the behavior of consequents. Proposed predictive reasoning method consist of five steps, these steps have been presented in the following subsections.

### 4.1 Similarity Measure

The similarity measure between linguistic terms can be defined their membership functions. In this section we propose another similarity-based fuzzy reasoning method, which calculates the similarity between $A$ and $A'$ using equality and cardinality as Yeung et al [33]. In our algorithm certainty factor (CF) is the key point. In previous similarity-based method certainty factor is just assigned for strength of every rule but here we first calculate certainty factor by applying some variables in stock market. Let us denote such similarity measure as $S_{EC} (a^{'}, a_i)$, which is defined as

$$S_{EC} (a^{'}, a_i) = 1 - \frac{M(a') \backslash M(a_i)}{M(a_i)} \quad \ldots \ldots \quad \text{(2)}$$

Where $\backslash$ is the symmetrical difference of $A$ and $A'$ viz, $\forall x \in X, \mu_{A \backslash A'} = |\mu_A(x) - \mu_{A'}(x)|$. Thus the equation can be expressed as

$$S_{EC} (a^{'}, a_i) = 1 - \frac{\sum_{x \in X} |\mu_{a^{'}}(x) - \mu_{a_i}(x)|}{\sum_{x \in X} \mu_{a_i}(x)} \quad \ldots \ldots \quad \text{(3)}$$

### 4.2 Aggregated Weighted Average

It is observed that $S_{EC} (a^{'}, a_i)$, which is the degree of equality of $a^{'}, a_i$ differs from $S_{EC} (a^{'}, a_i)$ and $0 \leq S_{EC} (a^{'}, a_i) \leq 1$, for instance, if $a^{'}, a_i$ then $S_{EC} (a^{'}, a_i) = 1$. Again if $S_{EC} (a^{'}, a_i) \geq \lambda_{a_i}$ for all $a_i$, the rule can be fired and the aggregated weighted average, $AG_w$ is defined as

$$AG_w = \sum_{i=1}^{n} AG_{w_i} = \sum_{i=1}^{n} \left[ S_{EC} (a^{'}, a_i) \cdot \frac{w_i}{\sum_{j=1}^{n} w_j} \right] \quad \ldots \ldots \quad \text{(4)}$$

After the aggregated weighted average has been calculated, two modification functions are proposed to modify the consequent $C'$. 

### 4.3 Modification Functions (MF’s)

In the proposed PR method, a rule $R_j: A_j \rightarrow C_j$ is to be fired with the use of an MF that modifies the consequent $C_j$ of the rule $R_j$. The MF is dependent on SM and its construction is subjective. That is one would if required adjust the form of the MF based on experts experience or historical data so that the system can function as close to the real situation as possible. However this enables us to bypass the matrix operations of CRI to deduce a consequent. There could be many forms of modification functions; we present two as possible examples here.

After the aggregated weighted average has been calculated, two modification functions are proposed to modify the consequent $C'$. 

![Figure. 2. More or Less form.](attachment:image.png)
More or Less Form:

This is an analogy to the definition of linguistic hedge “more or less” [24]. The induced consequent $C'_j$ is obtained through

$$C'_j = \min\{ 1, C_j (AG_w * \mu)^{0.5} \} \ldots \ldots \ldots (5)$$

Using the similarity transformation $SM = (1 + DM)^{-1}$ the “more or less” form of the modification function can be simplified as follows

$$C'_j = \min\{ 1, C_j * (1 + DM) \} \ldots \ldots \ldots (6)$$

Where $*$ represents multiplication.

Fig 2 shows the effect of this form of MF. We observe that additional uncertainty is introduced with such a modification function, i.e., large value of DM increase the values of membership. Obviously, at the extreme case, i.e., when $A_j' = A_j$ for the rule $R_j$, they are indeed exactly matched, i.e., DM=0 and hence the consequent $C_j$ will not be altered based on the proposed schema. This is one of the advantages of similarity based over CRI; recall that the CRI will generally produce a different result other than $C_j$, i.e.,

$$A_j' \circ (A_j \rightarrow C_j) = C'_j \neq C_j.$$ 

Membership Value Reduction Form:

In some sense this is an imitation of CRI. We simple multiply the consequent $C_j$ by SM $(A_j', A_j)$ for a given observation $A_j'$, i.e.,

$$C'_j = C_j * (AG_w * \mu)^{0.5}$$

Fig 3. shows the effect of this form of modification function. It gradually reduces the membership value of $C_j$ according to SM. In other words the smaller the SM greater the reduction of membership value of $C_j$, which imitates the min operator’s contraction effect of CRI. However, it is avoids the max operator’s expansion. At the extreme case that is when $A_j' = A_j$,

we have $SM (A_j', A_j)=1$ and hence $C'_j = C_j$.

Again this shows an advantage of the proposed PR method over CRI.

4.4 Rules Propositions

The following shows what $C'_j$ can be drawn for each of the three cases:

Case 1: The antecedent $A_j$ has only one proposition

$$AG_w = S_{EC} (a'_j, a_j) = S_{EC} (A'_j, A_j) \ldots \ldots \ldots (7)$$

Since $\sum_{i=1}^{n} w_i = 1$ with $n=1$ If $S_{DS} (a'_j, a_j) \geq \lambda_{ai}$ then

$$C'_j = \min\{ 1, C_j / (AG_w * \mu)^{0.5} \} \text{ or }$$

$$C'_j = C_j * (AG_w * \mu)^{0.5} \text{ Depending on whether we want to restrict or dilate the membership value of } C_j.$$ 

Case 2: The antecedent $A_j$ has two or more proposition connected by “AND”

$$AG_w = \sum_{i=1}^{n} (S_{EC} (a'_j, a_j) * \frac{w_i}{\sum_{j=1}^{n} w_j}) \ldots \ldots \ldots (8)$$

If $S_{DS} (a'_j, a_j) \geq \lambda_{ai}$ for all $(1 \leq i \leq n)$, then

$$C'_j = \min\{ 1, C_j / (AG_w * \mu)^{0.5} \} \text{ or }$$

$$C'_j = C_j * (AG_w * \mu)^{0.5}$$

Case 3: The antecedent $A_j$ has two or more propositions connected by “OR”

This rule can be split into n simple rules as shown in Case 1, i.e.,

$$AG_w = S_{EC} (a'_j, a_j)$$

Because for each

$$i = 1, \ldots, n, \sum_{j=1}^{n} w_j \text{ is reduced to a single term } w_i, \text{ and becomes } w_i / w'_i = 1$$

If $\exists \text{ s.t. } S_{DS} (a'_i, a_j) \geq \lambda_{ai} \text{ for all } (1 \leq i \leq n)$, then

$$C'_j = \min\{ 1, C_j / (AG_w * \mu)^{0.5} \} \text{ or }$$

$$C'_j = C_j * (AG_w * \mu)^{0.5}$$

OR,

If $S_{DS} (a'_i, a_j) \geq \lambda_{ai} \text{ for all } (1 \leq i \leq n)$, then

$$C'_j = \min\{ 1, C_j / (AG_w * \mu)^{0.5} \} \text{ or }$$

$$C'_j = C_j * (AG_w * \mu)^{0.5}$$

where $\mu$ is defined as:

\[ \mu(x) = \begin{cases} 
1.0 & \text{if } x = -1 \\
0.5 & \text{if } x = 0 \\
0.0 & \text{if } x = 1 \\
-0.5 & \text{if } x = 2 \\
0.0 & \text{if } x = 3 \\
0.5 & \text{if } x = 4 \\
1.0 & \text{if } x = 5 
\end{cases} \]

![Figure 3. Membership value reduction form](image-url)
\[ \mu = \left\lfloor \frac{\gamma + \delta - \rho - \alpha}{2} \right\rfloor^{0.5} \]  

where \( \gamma, \delta, \rho, \alpha \) are active variables that can affect stock market.

If \( a_1 \) is \( f_1 \) AND \( a_2 \) is \( f_2 \) THEN \( C \) is \( fC \). \( (CF = \mu) \), 

\[ Th = \{\lambda_{a_1}, \lambda_{a_2}\}, W = \{w_1, w_2\} \] e.g., IF Open is low AND Close is low THEN Signal is low

### 4.5 Fuzzy Reasoning Mechanism

Let \( C = \{\lambda_1, \lambda_2, \ldots, \lambda_n\} \) be an object to be classified, \( F \) be a group of WFPRs extracted from the tree, and there are \( k \) samples. The initial state of \( C \) with respect to classification is set to be \((x_1, x_2, \ldots, x_k) = (0,0,\ldots, 0)\).

Step 1. From the group \( F \), select a rule \( R \): IF \( A \) THEN \( B \) \( \{CF, Th, Lw\} \) where the antecedent \( A \) is supposed to be \( (A_1, A_2, \ldots, A_n) \); the consequent \( B \) to be \( (b_1, b_2, \ldots, b_n) \); and the local weight \( Lw \) to be \( (Lw_1, Lw_2, \ldots, Lw_n) \).

Step 2. Compute the membership degree of \( \lambda_j \) belonging to \( A_j \): \( SM_j = A_j(\lambda_j) \) where the membership function of the fuzzy set \( A_j \) is denoted by itself.

Step 3. Let \( Th = \{Th_1, Th_2, \ldots, Th_n\} \) be the threshold. If for each proposition \( A_j \) the inequality \( SM_j \geq Th_j \) holds, then the rules are executed. Computed the overall weighted average \( SM_w \) of similarity measures as

\[ SM_w = \sum_{j \in T} (Lw_j / \sum_{i \in T} Lw_i) SM_j \]  

where \( T = \{ j : SM_j \geq Th_j \} \).

Step 4. Modify the matching-consequent according to one of the beforehand given modification strategies:

4a) more or less form: \( B^* = \min\{1, B / SM_w\} \);

4b) membership-value reduction form: \( B^* = B \cdot SM_w \);

4c) keeping the consequent of the rule unchanged (no modification): \( B^* = B \).

Step 5. Compute the certainty degree of \( B^* \) as \( CF_{B^*} = CF \cdot SM_w \).

Step 6. Put \( x_j = \max(CF_{B^*}, x_j) \) where \( j \) is a number corresponding to the sample of the consequent \( B^* \).

Repeat the above six steps until each rule within the group \( F \) has been applied to the object \( C \).

### 5.0 EXPERIMENTAL RESULTS

We demonstrate our approach on the stock market database. Dynamic time series stock market includes date, open, low, high, close and volume attributes. In this paper we have taken open, close, and volume attribute because by these attributes investors pay more attention on movements of stock shares to every single unit of time. In this section we extract WFPR’s from FDT.

#### 5.1 WFPR’s Extraction from FDT

We can extract WFPR’s from FDT shown in figure 1:

- Rule1: IF (Open=Low) AND (Close=Low) AND (Volume=Low) THEN (Optimal Signal=Low), \( (CF_1 = 0.95, W_{j_1} = 5.65, W_{j_2} = 7.35, W_{j_3} = 0.95) \)
- Rule2: IF (Open=Low) AND (Close=Med) AND (Volume=High) THEN (Optimal Signal=Med), \( (CF_2 = 0.85, W_{j_1} = 1.65, W_{j_2} = 5.56, W_{j_3} = 1.45) \)
- Rule3: IF (Open=High) AND (Close=Low) AND (Volume=Med) THEN (Optimal Signal=Low), \( (CF_3 = 0.75, W_{j_1} = 1.45, W_{j_2} = 3.15, W_{j_3} = 7.45) \)
- Rule4: IF (Open=Med) AND (Close=High) AND (Volume=Low) THEN (Optimal Signal=High), \( (CF_4 = 0.65, W_{j_1} = 5.25, W_{j_2} = 3.58, W_{j_3} = 2.23) \)
- Rule5: IF (Open=Low) AND (Close=High) AND (Volume=High) THEN (Optimal Signal=High), \( (CF_5 = 0.60, W_{j_1} = 5.98, W_{j_2} = 2.05, W_{j_3} = 4.45) \)
- Rule6: IF (Open=High) AND (Close=High) AND (Volume=High) THEN (Optimal Signal=High), \( (CF_6 = 0.55, W_{j_1} = 0.65, W_{j_2} = 2.53, W_{j_3} = 2.35) \)

In experiment, the number of linguistic terms for open, close and volume attribute from historical databases is taken to be three, the parameter specified in previous sections for reducing the fuzziness in training process is set to 0.25 and the leaf criterion is taken to be 0.75. The learning accuracy computational complexity and the robustness can be used to compare other methods. The learning accuracy and testing accuracy, the complexity is regarded as the numbers of nodes and leaves and the robustness refers to the prediction accuracy by dropping an attribute (with smallest importance to classification) from the extracted rules. For historical database, 50% if cases are held for testing. This procedure is repeated four times for the given cut standard=0.25 and the leaf standard=0.75. The number of nodes, the number of leaves, the training accuracy, and the testing accuracy are regarded as the average of the four.
6.0 Conclusions

In this paper, data mining techniques have been used to uncover hidden patterns and predict future trends and behaviors in financial markets. In data mining, fuzzy decision tree classification has been developed to extract weighted fuzzy production rules for stock market prediction. Proposed FDT is based on minimum classification information entropy to select expanded attributes. In similarity-based fuzzy reasoning method we analyze WFPR’s which are extracted from FDT. The analysis is based on the result of consequent drawn for different given facts (e.g. variables that can affect stock market) of the antecedent. Proposed method has some advantages such as accurate stock market prediction, efficiency and comprehensibility of the generated WFPR’s rules, which are important to data mining. These WFPR’s allow us effectively classify patterns of non-axis-parallel decision boundaries using membership functions properly, which is difficult to do using attribute-based classification methods.

References

[19] Jian Pei; Jiawei Han; Hongjun Lu; Shojiro Nishio; Shiwei Tang; Dongqing Yang; “H-mine: hyper-structure mining of frequent patterns in large databases” Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on , 29 Nov.-2 Dec. 2001 Pages:441 - 448


