Fuzzy Decision Tree for Data Mining of Time Series Stock Market Databases

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

ABSTRACT

If the given fact for an antecedent in a fuzzy production rule (FPR) does not match exactly with the antecedent of the rule, the consequent can still be drawn by technique such as 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 this paper, new Fuzzy Decision Tree (FDT) has been constructed by using weighted fuzzy production rules (WFPR). In WFPR, assign a weight parameter 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 (e.g. effect of other companies, effect of other stock exchanges, effect of overall world situation, effect of political situation etc) in dynamic stock market. Finally, our proposed approach will be able to predict stock share indices, and improve computational efficiency of data mining approaches
.

Keywords Data Mining, Fuzzy Logics, Decision Tree, Neural Networks.

1. INTRODUCTION

Time series data are of growing importance in many new database applications, such as data mining. A time series is a sequence of real numbers, each number representing a value at a time point. For example, the sequence could represent stock or commodity prices, sales, exchange rates, weather data biomedical measurements, etc. Different similarity queries on time-series have been introduces [1, 2]. For example, we may want to find stocks that behave in approximately the same way (or approximately the opposite way) for hedging; or products that had similar selling patterns during the last year; or years when the temperature patterns in two regions of the world were similar. In queries of this type, approximate, rather than exact, matching is required. However mining different queries from huge time-series data is one of the important issues for researchers. In useful data mining techniques like classification and clustering, to handle time-series data is one of the stimulating research issue.

First we convert dynamic time series data segment into equal set of pattern (sample), which are appropriate for general data mining method, and use the attributes of these static samples as the basis for exploratory fuzzy rule induction. Each static sample consists of a set of fuzzy attributes and each fuzzy attribute consist of a set of linguistic terms. By using these linguistic terms, we calculate classification attribute. 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, we have used powerful fuzzy reasoning method in our algorithm.

The general data-mining method (such as FDT) can then be used directly on the information database to uncover the rules for predicting the length of sample (i.e., the turning point of a stock market quotation). FDT's are perhaps successful methods for accurate decision making. It can provide a high level of predictive accuracy, which rarely do they facilitate human inspection or understanding. FDT can be build by using our algorithm to training these static samples. Our proposed FDT can be used to generate and mine WFPR's. The representation power of WFPR's can be enhanced by including several knowledge parameters such as weight and certainty factor. Many factors are affecting stock market directly or indirectly, in this paper we have analyzed some factors which will be presented in next section. We use these factors to evaluate certainty factor in WFPR's.

The reminder of this paper is arranged as: section 2 construction of FDT, section 3 experimental results (extract WFPR's from FDT), and finally conclusion in section 4.

2. PROCESSING (FDT CONSTRUCTION)

FDT induction can be used to generate and mine WFPR's and the representation power of WFPR's can be enhanced by including several knowledge parameters such as weight and certainty factors [3]. In this section, we present a new FDT by enhancing these parameters. We use some live parameters which can affect stock market directly or indirectly.

Our proposed FDT [4, 5] was based on minimum classification information entropy to selected expanded attributes. This concept proposed by Quinlan in 1986 [6]. In this paper, we extract WFPR's from FDT by using parameters, weight and certainty factors. FDT has four components: set patterns, Fuzzification of numerical numbers, similarity-based fuzzy reasoning method, built fuzzy decision tree to training fuzzy sets examples. The following sections will present the comprehensive study of these components of FDT.

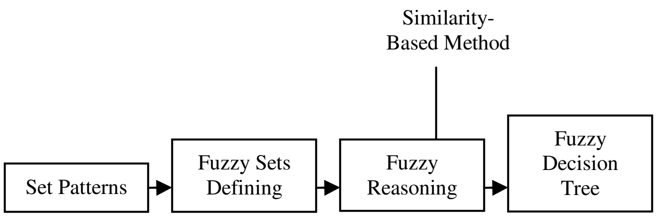

Figure 1: The Components of Fuzzy Decision Tree

In first step of FDT we identify patterns, and then fuzzy sets are used to optimize the profitability of the identified patterns. Historical data is very important for stock prediction. Using information how stock performed in the past, we can predict stock indices the best course of action to take now.

In second component of FDT to define fuzzy sets selection. For fuzzy sets defining, we use triangular fuzzy membership function corresponds to each patterns. In next fuzzy reasoning, we check that given fact for an antecedent in a fuzzy production rule does not match exactly with the antecedent of the rule, and then consequent can be drawn by using our proposed similarity-based fuzzy reasoning method. After fuzzy reasoning we construct fuzzy decision tree for training data. In the following sections these components are outlined in detail.

2.1 Set Patterns for FDT

In first step of FDT to identify patterns, then fuzzy sets are used to optimize the profitability of the identified patterns. We convert dynamic time series stock market data segment into equal set of pattern, which are appropriate for general data mining method, and use the attributes of static samples as the basis for exploratory fuzzy rule induction. These samples consider as a static samples.

2.2 Fuzzification of Numerical Numbers

Fuzzification is a process of fuzzifying numerical numbers into linguistic terms, which is often used to reduce information overload in human decision making process. The numerical salary, for example, may be perceived in linguistic terms such as high, average and low. Linguistic terms are simple forms of fuzzy values but generally their membership functions are unknown and need to be determined. One way of determining membership functions of these linguistic terms is by expert opinion or by people's perception. Yet another way is by statistical methods [7]. An algorithm for generating certain type of membership functions, which is based on self-organized learning [8], can be found in [9].

Let X is a given data set which will be sample into k linguistic terms $T_j,\ j = 1,2,\ldots,\ k$. The triangular membership function is denoted as $\mu_A(x)$ and is defined as:

$$\mu_A(x) = \begin{cases} \dfrac{x-l}{c-l} & l \le x \le c \\ \dfrac{h-x}{h-c} & c \le x \le h \\ 0 & \text{otherwise} \end{cases} \quad \ldots\ldots(1)$$

A triangular membership function can be represented by a triple numbers (l, m, h), where l is lower number, m is mid point and h is higher number in the triangular membership function. Consider Open attribute from table I has three linguistic terms $(k = 3)$, we can find membership function for these linguistic terms as:

$$T_{low} = \begin{cases} \dfrac{-(x+l)}{-2l} & l \le x \le r_1 \\ 0 & \text{otherwise} \end{cases} \quad \ldots\ldots(2)$$

$$T_{med} = \begin{cases} \dfrac{x-l}{c-l} & l \le x \le c \\ \dfrac{h-x}{h-c} & c \le x \le h \\ 0 & \text{otherwise} \end{cases} \quad \ldots.(3)$$

$$T_{high} = \begin{cases} \dfrac{-(x+h)}{-2h} & r_2 \le x \le h \\ 0 & \text{otherwise} \end{cases} \quad ..(4)$$

It is obvious that the three linguistic terms can be described as Low, Medium and High. The second column of Table II shows the membership degree of the attribute Open belonging to the three membership functions. Similarly we can find membership function for others attributes and result shows in table II.

2.3 Similarity-based Fuzzy Reasoning Method

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 [10]. In our algorithm certainty factor (CF) is the key point. In previous similarity-based method certainty factor is just assign for strength of every rule but here we first calculate certainty factor by applying some variables in stock market. Let us first calculate similarity measure, which is denoted by $E_c(a'_i, a_i)$ and defined as:

TABLE I: One sample from stock exchange databases

No	Date	Open	Close	High	Low	Volume
1	08/25/3 9:30	43.75	43.85	43.87	43.57	110,210
2	08/25/3 9:35	43.86	43.86	44.14	43.74	138,103
3	08/25/3 9:40	43.87	43.65	43.88	43.64	101,167
4	08/25/3 9:45	43.64	43.76	43.94	43.59	118,189
5	08/25/3 9:50	43.77	43.899	43.91	43.72	91,042
6	08/25/3 9:55	43.9	43.9	44.1	43.73	113,685
7	08/25/3 10:00	43.88	43.91	44.05	43.7	143,362
8	08/25/3 10:05	43.9	44.22	44.25	43.87	82,994
9	08/25/3 10:10	44.22	44.029	44.28	43.98	83,474
10	08/25/3 10:15	44.03	43.95	44.1	43.92	103,152

TABLE II: Training Set with Fuzzy Representation

No	Change in open			Change in close			Change in Volume		Optimal Signal		
	Low	Med	High	Low	Med	High	Low	High	Down	Hold	Up
1	0.00	0.53	0.00	0.00	0.94	0.00	0.00	0.50	0.00	0.53	0.00
2	0.00	0.15	0.67	0.00	0.88	0.00	0.00	0.97	0.00	0.67	0.00
3	0.00	0.49	0.00	1.00	0.00	0.00	0.81	0.00	0.49	0.00	0.00
4	1.00	0.00	0.00	0.00	0.27	0.68	0.00	0.79	0.00	0.00	0.68
5	0.00	0.08	0.70	0.00	0.09	0.72	0.72	0.00	0.00	0.00	0.70
6	0.00	0.08	0.70	0.00	0.94	0.00	0.00	0.88	0.00	0.70	0.00
7	0.54	0.59	0.00	0.00	0.88	0.00	0.00	1.00	0.00	0.59	0.00
8	0.00	0.46	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.46
9	0.00	0.00	1.00	0.95	0.00	0.00	0.00	0.51	0.95	0.00	0.00
10	0.91	0.00	0.00	0.69	0.00	0.00	0.00	0.83	0.69	0.00	0.00

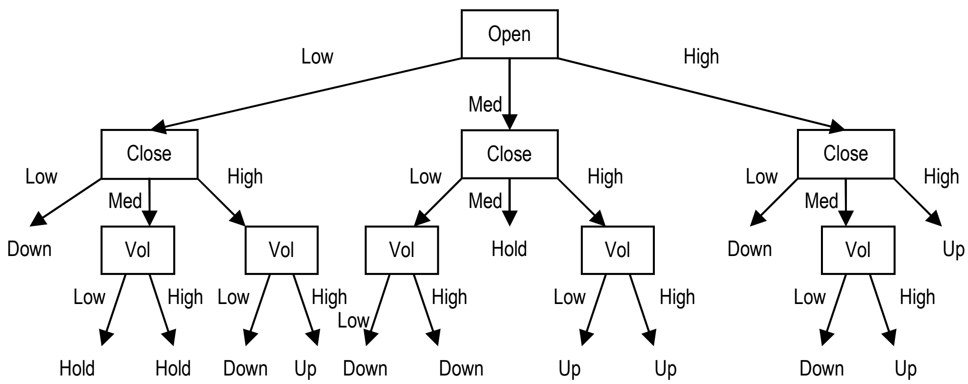

Figure 2: Fuzzy Decision Tree by using our proposed algorithm to train table II

$$E_c(a'_i, a_i) = 1 - \frac{F(A' \cap A)}{F(A)} \quad ......(5)$$

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

$$E_c(a'_i, a_i) = 1 - \frac{\sum_{x \in X} | \mu_{a'_i}(x) - \mu_{a_i}(x) |}{\sum_{x \in X} \mu_{a_i}(x)} \quad .....(6)$$

It is observed that $E_c(A', A)$, which is the degree of equality of a'_i in a_i, differs from $E_c(a'_i, a_i)$ and $0 \le E_c(a'_i, a_i) \le 1$, for instance, if $a'_i = a_i$, then $E_c(a'_i, a_i) = 1$. Again if $E_c(a'_i, a_i) \ge \lambda_{ai}$ for all a'_i, the rule can be fired and the aggregated weighted average, AG_w is defined as

$$AW_w = \sum_{i=1}^{n} AW_{wi} = \sum_{i=1}^{n} \left[E_c(a'_i, a_i) * \frac{w_i}{\sum_{i=1}^{n} w_i} \right] \quad ....(7)$$

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

$$C' = \min \{1, C / (AW_w * \mu)^{0.5}\} \quad ...(8)$$

where μ is defined as: $\mu = \left| \frac{\gamma + \delta}{2} - \rho - o \right|^{0.5}$

where γ, δ, ρ, o are active variables that can affect stock market.

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

Case 1: The antecedent A has only one proposition

$$AW_w = E_c(a'_1, a_1) = E_c(A', A) \quad ..(9)$$

Since $w_1 / \sum_{i=1}^{n} w_i = 1$ with n=1 If $E_c(a'_i, a_i) \ge \lambda_{a1}$

$$C' = \min\{1, C / (AW_w * \mu)^{0.5}\} \quad ...(10)$$

then

Depending on whether we want to restrict or dilate the membership value of C.

Case2: The antecedent A has two or more proposition connected by "AND"

$$AW_w = \sum_{i=1}^{n} (E_C(a_i', a_i) * \frac{w_i}{\sum_{j=1}^{n} w_j} ...(11)$$

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

$$C' = \min\{1, C/(AW_w * \mu)^{0.5}\}....(12)$$

Case3: The antecedent A has two or more propositions connected by "OR" This rule can be split into n simple rules as shown in Case1, i.e., $AW_w = E_C(a'_i, a_i)$ Because for each $i = 1,....,n, \sum_{j=1}^{n} w_j$ is reduced to a single term w_i, and becomes $w_i / w_i = 1$

If $\exists_i$ s.t. $E_c(a'_i, a_i) \geq \lambda_{ai}$, for all $(1 \leq i \leq n)$, then

$$C' = \min\{1, C/\max(AW_{w1}, AW_{w2}, ...AW_{wi}) * \mu)^{0.5}\}...(13)$$

, OR, If $E_c(a'_i, a_i) \geq \lambda_{ai}$ for all $(1 \leq i \leq n)$, then

$$C' = \min\{1, C/\max(AW_{w1}, AW_{w2}, ...AW_{wi}) * \mu)^{0.5}\}...(14)$$

In this paper, we are interested to predict dynamic stock exchange databases. Therefore, we are considering four possible cases for every consequent portion. Consequent places (conclusion) have been considering every possible fluctuation in dynamic stock market. Here is the simple example that we have used in our algorithm:

IF a_1 is fa_1 AND a_2 is fa_2 THEN C is fc, ($CF = \mu$), $Th = \{\lambda_{a1}, \lambda_{a2}\}$, $W = \{w_1, w_2\}$ e.g., IF Open is low AND Close is low THEN Signal is low

2.4 FDT Algorithm

We first formulate a problem of learning from examples with fuzzy representations. Consider a set of examples $\{p_1, p_2, ...p_N\}$ which is defined as the universe of discourse X, where X is denoted as $\{1,2,.....,N\}$. Let $A^{(1)}, A^{(2)}, ..., A^{(n)}$ and $A^{(n+1)}$ be a set of fuzzy attributes where $A^{(n+1)}$ denotes a classification attribute. Each fuzzy attribute $A^{(j)}$ consist of a set of linguistic term $L(T^{(j)}) = \{L_1^{(j)}, L_2^{(j)}, ..., L_{mj}^{(j)}\}$ $(j = 1,2,...., n+1)$. All linguistic terms are defined on the same universe of discourse X. The value of the i^{th} example p_i with respect to the j^{th} attribute, denoted by μ_{ij}, is a fuzzy set defined on $L(T^{(j)}) (i = 1,..., N, j = 1,2,...n+1)$. In other words, fuzzy set μ_{ij} has a form of

$$\mu_{ij}^{(1)} / L_1^{(j)} + \mu_{ij}^{(2)} / L_2^{(j)} + ... + \mu_{ij}^{(mj)} / L_{mj}^{(j)}$$

where $\mu_{ij}^{(k)}$ denotes the corresponding membership degree $k = 1,....., m_j$. To illustrate these notations, we consider an example shown in table I which describes a small training set of learning from fuzzy.

Assume that these linguistic terms is consider as a set of data D, where each data has l numerical values for n attributes $A^{(1)}, A^{(2)}, ..., A^{(n)}$.

We take the reasonable sample size from n attributes $A^{(1)}, A^{(2)}, ..., A^{(n)}$.

1. Generate the root node that has a set of all data, i.e., a fuzzy set of all data with the membership value 1.

2. If a node O with a fuzzy set of data D satisfies the following conditions:

2.1 the proportion of a data set of a class C_k is greater than or equal to a threshold θ_r, that is

$$\frac{|D^{C_K}|}{|D|} \geq \theta_r, ...(15)$$

2.2 there are no attribute for more classifications, then it is a lead node and assigned by the class name.

3. If it does not satisfy the above conditions, it is not a leaf and the test node is generated as follows:

Consider a test node S having n attributes to be selected. For each $k(1 \leq k \leq n)$, the attribute $T^{(k)}$ takes m_k fuzzy subsets (linguistic terms), $L_1^{(k)}, L_2^{(k)}, ..., L_{mk}^{(k)}$. $A^{(n+1)}$ denotes the classification attribute, taking values $L_1^{(n+1)}, L_2^{(n+1)}, ..., L_m^{(n+1)}$. For each attribute value (fuzzy subset), $L_i^{(k)} (1 \leq k \leq n, 1 \leq i \leq m_k)$, its relative frequencies concerning the j^{th} fuzzy class $L_i^{(n+1)} (1 \leq j \leq m)$ at the considered nonleaf node S is defined as

$$p_{ij}^{(k)} = M(L_i^{(k)} \cap L_j^{(n+1)} \cap S) / M(L_i^{(k)} \cap S)...(16)$$

At the considered nonleaf node S, the fuzzy classification entropy of $L_i^{(k)} (1 \leq k \leq n, 1 \leq i \leq m_k)$ is defined as

$$Entr_i^{(k)} = -\sum_{j=1}^{m} p_{ij}^{(k)} \log_2 p_{ij}^{(k)} ....(17)$$

The averaged fuzzy classification entropy of the k^{th} attribute is defined as

$$E_k = \sum_{i=1}^{mk} w_i Entr_i^{(k)} ....(18)$$

in which w_i denotes the weight of the i^{th} value $L_i^{(k)}$ and is defined as

$$w_i = M(S \cap L_i^{(k)}) / \sum_{j=1}^{mk} M(S \cap L_j^{(k)})...(19)$$

The above FDT aims to search for an attribute such that its average fuzzy classification entropy attains minimum, i.e., selecting such an integer k_o (the $k_o th$ attribute) that $E_{ko} = Min_{1 \leq k \leq n} E_k$.

3. 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.

3.1 WFPR's Extraction from FDT

We can extract WFPR's from FDT shown in figure 3:

Rule1: IF (Open=Low) AND (Close=Low) AND (Volume=Low) THEN (Optimal Signal=Low), (CF_1 =0.95, w_{j1}=5.65, w_{j4}=7.35, w_{j7})

Rule2: IF (Open=Low) AND (Close=Med) AND (Volume=High) THEN (Optimal Signal=Med), (CF_2 =0.85, w_{j1}=1.65, w_{j5}=5.56, w_{j9} =1.45)

Rule3: IF (Open=High) AND (Close=Low) AND (Volume=Med) THEN (Optimal Signal=Low), (CF_3 =0.75, w_{j3}=1.45, w_{j4} =3.15, w_{j8}=7.45)

Rule4: IF (Open=Med) AND (Close=High) AND (Volume=Low) THEN (Optimal Signal=High), (CF_4 =0.65, w_{j2} =5.25, w_{j6} =3.58, w_{j7} =2.23)

Rule5: IF (Open=Low) AND (Close=High) AND (Volume=Low) THEN (Optimal Signal=High), (CF_5 =0.60, w_{j1}=5.98, w_{j6} =2.05, w_{j7} =4.45)

Rule6: IF (Open=High) AND (Close=High) AND (Volume=High) THEN (Optimal Signal=High), (CF_6 =0.55, w_{j3}=0.65, w_{j6} =2.53, w_{j9} =2.35,)

4. CONCLUSION

This paper presents a new FDT for unpredictable dynamic stock exchange databases. Most of the existing data mining techniques are not so efficient to dynamic time series databases. FDT has been constructed with power of WFPR's. It 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 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.

We are now applying our algorithm for generating WFPR's from online stock market data. For experimentation, we are using historical data. By using information how stock performed in the past, you can use our system to predict the best course of action to take now.

5. REFERENCE

[1] R. Agrawal, K.-I. Lin, H.S. Sawhney, and K. Shim, ªFast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases,º Proc. 21st Int'l Conf. Very Large Data Bases (VLDB '95), pp. 490±501, Sept. 1995.

[2] R. Agrawal, G. Psaila, E.L. Wimmers, and M. Zait, ªQuerying Shapes of Histories,º Proc. 21st Int'l Conf. Very Large Data Bases (VLDB '95), pp. 502±514, Sept. 1995.

[3] Yeung, D.S.; Tsang, E.C.C.; Xizhao Wang; "Fuzzy rule mining by fuzzy decision tree induction based on fuzzy feature subset" Systems, Man and Cybernetics, 2002 IEEE International Conference on , Volume: 4 , 6-9 Oct.2002 Pages:6 pp. vol.4

[4] R. H. Khokhar, Mohd Noor Md Sap. " Design and Development of Neural Bayesian Approach for Unpredictable Stock Exchange Databases" CW 2003, Marina Mandarin Singapore, 3-5 Dec 2003 Pages:364 – 371 IEEE.

[5] R. H. Khokhar, Mohd Noor Md Sap. "Neuro-Pruning Method for Pruning Decision Tree in Large Databases", CIRAS 2003, Pan Pacific Hotel, Singapore, 15-18 Dec 2003.

[6] J.R. Quinlan (1986) : "Induction of Decision Tree" Machine Learning, Vol, pp.81-106.

[7] M. R. Civanlar and H. J. Trussell, "Constructing membership functions using statistical data," Fuzzy Sets and Systems, vol. 18, 1986, pp. 1-14.

[8] T. Kohonen, "Self-Organization and Association Memory " Springer, Berlin, 1988

[9] Y. Yuan and M. J. Shaw, "Induction of fuzzy decision trees," Fuzzy Sets and Systems vol. 69, 1995, pp. 125-139

[10] D. S. Yeung and E. C. C. Tsang, "A Comparative Study on Similarity-Based Fuzzy reasoning Methods" IEEE Trans. Syst., Man and Cybernetics-part A: System and Human, vol., 27, No. 2, April 1997.