Financial Asset Price Forecasting Based on Intertransaction Association Rules Mining

Ping Li, Wenjing Xing
School of Economics and Management
Beihang University,
Beijing, China
liping124@buaa.edu.cn

Huang Guangdong
School of Information Engineering
China Geosciences University (Beijing),
Beijing, China
gdhuang@cugb.edu.cn

***Abstract*—It has been widely accepted that association rules mining, the task of searching for correlations between items in a database, can discover useful rules in stock analysis. Previous studies mainly emphasize on mining intratransaction associations. In this paper, we introduce the concept of intertransaction and the FITI algorithm so that we can effectively forecast the price changes in Chinese capital markets, then we compare FITI with EH-Apriori, and demonstrate the advantages of FITI over EH-Apriori. At the end of this paper, we apply the algorithm to a dataset of Chinese asset indices and the results indicate the usefulness of intertransaction association rules in price prediction.**

Keywords-Price forcasting, Intertransaction association rules, Data mining, EH-Apriori algorithm, FITI algorithm

I. Introduction

Data mining is widely used in various academic fields such as forecasting stock market or bank bankruptcies, financial risk management, credit rating, bank customer profiling and cross-selling analysis. One important problem in data mining is mining association rules, which was introduced by Agrawal [1]. Given two item sets X and Y, an association rule means X $\Rightarrow$ Y.

Association rules mining can be used in market analysis, fraud detection, medical research and process re-engineering [2]. The most often cited work on market analysis is mining interesting rules from supermarkets transaction database. Each transaction contains the items bought by one of its customer. An example of the rule is "70% of customers who bought milk also bought bread" which can be expressed:

R1: milk $\Rightarrow$ bread. (*support*: 60%, *confidence* 80%).

Association rules mining of large databases gives rise to many research directions, including Apriori-like mining methods [3][4]); quantitative association rules mining[4]; constraint-based rules mining [5]; mining association rules with multiple supports [6]; constraint-based rule mining [7]; mining sequential patterns and episodes[4][9].

In spite of all these efforts, there is a kind of association relationships of great importance and usefulness that could not be expressed using the traditional association rule mining framework---intertransaction association rule mining. This problem was first explored by Anthony [10], which introduced the notion of intertransaction association rule, defined its own measures *support* and *confidence* and developed an efficient Apriori-like algorithm---FITI algorithm. This algorithm decomposes intertransaction association rule mining problem into three subproblems. First, one finds and stores frequent intratransaction item sets; and then, he/she transforms the database into a set of encoded Frequent-Itemset Tables (Called FIT tables); finally, he/she can find frequent intertransaction item sets.

The difference between the traditional association rule and intertransaction association rule can be depicted as the following:

R2: ''When the prices of TCL and Changhong go up, the price of Hisense will increase on the same day with probability of 80%.''

However, stockjobbers may be more interested in the following rule.

R3: ''If the prices of TCL and Changhong go up on the first day, the price of Hisense will increase two days later probability of 80%.''

Classical association rules, like R2, discover the relationship among items within the same transactions, while R3 expresses association among items of different transactions along certain dimension. Relevant studies include sequential patterns mining [5] and mining episodes [9].

In this paper, we will investigate the application of intertransaction association rules mining in stock price predication and the possibility of generalizing this method to futures market.

The remainder of this paper is organized as follows. In Section 2 we explain the definitions of intertransaction and describe the intertransaction association rule mining task; in Section 3, we present the framework of intertransaction association rule mining method and express the main idea of FITI algorithm, then briefly review the FITI algorithm and show the advantage of it over traditional algorithms. Real datasets are investigated in Section 4, where we discuss the generalization and application of this method to Chinese capital market. We summarize our work in Section 5.

II. Problem Descriptions

Definition 1. Let $\Sigma = \{e_1, e_2, \ldots, e_u\}$ denote a set of items, where u is the number of transactions, and denote D a set of non-negative integers called domain attributes. A

978-0-7695-3997-3/10 $26.00 © 2010 IEEE
DOI 10.1109/ICEE.2010.361

transaction database is a set of transactions $\{T_1, T_2, \ldots, T_n\}$, where T_i (i= 1, 2,..., n) is a subset of Σ. A transaction is in the form of <d, E>, where $d \subseteq Dom(D)$ and $E \subseteq \Sigma$. d is named a dimensional attribute and E is named an itemset.

The dimensional attribute describes the properties associated with the data items, such as time and location. We assume that the domain of the dimensional attribute is ordinal and can be divided into equal length intervals. For example, time can be divided into hours, days, weeks, etc. The intervals can be represented by integers.

Definition 2. A sliding window W in a transaction database Γ is a block of w continuous intervals along domain D, starting from interval d_0 such that Γ contains a transaction at interval d_0. Each interval d_j in W is called a subwindow of W and denoted as $W[j]$, where $j= d_j - d_0$. We call j the subwindow number of d_j within W.

The concept of sliding window is introduced to describe how many intervals an intertransaction association rule spans across. In order to save resources and mine more economically, the parameter *maxspan* (w) is introduced to express the size of the sliding window.

Definition 3. A megatransaction M is defined as

$$M = \{e_i(j) \mid e_i(j) \in W[j],\ 1 \le i \le u,\ 0 \le j \le w-1\} \qquad (1)$$

Items in one megatransaction are defined as extended-items and denote the set of all extended-items as

$$\Sigma' = \{e_1(0), \ldots, e_1(w-1), e_2(0), \ldots, e_2(w-1), \ldots, e_u(0), \ldots, e_u(w-1)\}. \qquad (2)$$

We define the set of items $I_1 \in \Sigma$ as intratransaction itemset and the set of items $I_2 \in \Sigma'$ as intertransaction itemset.

Definition 4. An intertransaction association rule is written in the form of $X \Rightarrow Y$, where $X \subseteq \Sigma', Y \subseteq \Sigma'$ and $X \cap Y = \varnothing$.

Definition 5. X and Y are both extended itemsets subject to Definition 4. We define the *support* and *confidence* of an intertransaction association rule $X \Rightarrow Y$ as

$$\text{Support}(X) = \frac{|\{T \mid T \in M \text{ and } X \subseteq T\}|}{|M|} \qquad (3)$$

$$\text{Confidence}(X \Rightarrow Y) = \frac{|\{T \mid T \in D and (X \cup Y) \subseteq T\}|}{|\{T \mid T \in D and X \subseteq T\}|} \qquad (4)$$

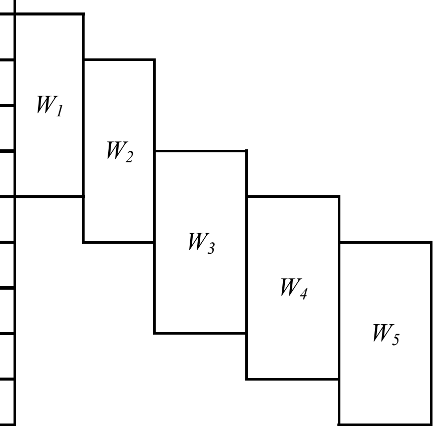
Above definitions can be explained by an example. Table 1 shows a transaction database containing five transactions. The five transactions are located at intervals 1, 2, 4, 5, 6. Let w=4, we now have five sliding windows W_1, W_2, W_3, W_4 and W_5, with addresses of 1, 2, 4, 5, 6, respectively. Each window contains 4 subwindows. For example, W_1 has four subwindows $W_1[0]$ (with items a, b), $W_1[1]$ (with items b, d), $W_1[2]$ and $W_1[3]$ (with items a, b, c, d).

Each sliding window forms a megatransaction, which is the itemset of all the items in one sliding window. In our case, the megatransaction in W_1 is {a[0], b[0], b[1], d[1], a[3],b[3],c[3],d[3]}. Thus, we have Σ'={ $a_1[0]$, $b_1[0]$, $b_1[1]$, $d_1[1]$, $a_1[3]$, $b_1[3]$, $c_1[3]$, $d_1[3]$, $b_2[0]$, $d_2[0]$, $a_2[2]$, $b_2[2]$, $c_2[2]$, $d_2[2]$, $b_2[3]$, c_2 [3], $a_3[0]$, $b_3[0]$, $c_3[0]$, $d_3[0]$, $b_3[1]$, $c_3[1]$, $a_3[2]$, $b_4[0]$, $c_4[0]$, $a_4[1]$, $a_5[0]$}.

Then, after we set the two essential parameters *minsup* (denotes minimum *support* level) and *minconf* (denotes minimum *confidence* level), we can mine intertransaction association rules from the transaction database. For instance, setting *minsup*=0.4 and *minconf*=0.8, we can get one rule mined from Table 1, {b[0], d[0]} $\Rightarrow$ a[2] (*support*=0.4, *confidence*=1).

TABLE I. **TRANSACTION DATABASE**

D	E
1	a, b
2	b, d
3	
4	a, b, c, d
5	b, c
6	a
7	
8	
9	



III. MINING INTERTRANSACTION ASSOCIATION RULES

In this section, we give an overview of intertransaction association rules mining. Like the traditional association rule mining, the novel one consists of two steps:

(1) Find frequent intertransaction itemsets whose support is higher than minsup.

(2) For every frequent intertransaction itemset L, output an intertransaction association rule $S \Rightarrow L - S$, if the following conditions are satisfied:

i. $\exists\ e_i(j) \in S,\ 1 \le i \le u,$

ii. $\exists\ e_i(j) \in L - S,\ 1 \le i \le u,\ j \ne 0,$

iii. Confidence of $S \Rightarrow L - S$ is higher than minconf.

According to Anthony [10], first intratransaction then intertransaction (FITI), consists of three steps:

(1) Mining and Storing frequent intratransaction itemsets;

(2) Transforming the database into a set of encoded Frequent-Itemset Tables (Called FIT tables);

(3) Mining frequent intertransaction itemsets.

The outline of algorithms is as follows.

ALGORITHM I DATABASE TRANSFORMATION ALGORITHM.

```
Void Transform () {
  while (!feof (T)) {
    read next transaction Ti;
    write di to all Fj;
    Subset (Ti, 1, 1, 0) }}
void Subset (Ti, index, k, and NodeID) {
  if (k= =1) {
```

```
    for each item e_j in T_i {
      Search Itemtable for ID of {e_j}
      If (found) {
        Let nowID be ID found;
        Write nowID to F_1;
        For each item em (m>j) in T_i {
          Search childs of nowID for an itemset I that
          Contains e_m;
          If (found) {
            Let nextNode be the ID of I;
            Subset (Ti, m+1, k+1, nextNode); }}}}
            Return;}
  else {
    Write NodeID to F_k;
    For each item em (m>= index) in T_i {
      Search childs of NodeID for an itemset I that
      Contains e_m;
      If (found) {
        Let nextNode be the ID of I;
        Subset (Ti, m+1, k+1, nextNode); }}
    Return ;}}
```

ALGORITHM II LEVEL-WISE MINING OF FREQUENT INTERTRANSACTION ITEMSETS.

Input: A set of FIT tables: $F_1, \ldots, F_{max}$, and the minimum support threshold: minsup.

Output: The complete set of frequent intertransaction itemsets

```
Generate frequent intertransaction 2-itemsets, L_2;
k=3;
While (L_{k-1} ≠ Φ ) {
    Generate candidate intertransaction k-itemsets, C_k;
    Scan transformed database to update the count for Ck;
    Let L_k = { c ∈ C_k support(c) ≥ minsup};
    k++;}
```

ALGORITHM III GENERATION OF ALL THE (K-ITEMSET) SUBSETS OF AN INTERTRANSACTION (K+1)-ITEMSET.

```
Let S be the set of k-subsets of I;
S= {};
for (p=0; p < w; p++) {
    if (Ip! = 0) {
      If (Ip is an intratransaction one-itemsets) {
        If (p!=0)
          Add {I_0,…, I_{p-1}, 0,…, I_{w-1}} to S
        Else
          Add {I_0,…, I_{w-1},0} to S
      } else
      {Let Ip be an intratransaction h-itemsets, h > 1
        For each (h-1)-subset of I_p
        {Let t be the ID of the (h-1)-subset
          add {I_0,…, I_{p-1}, t,…, I_{w-1}} to S}}}}
Return S;
```

Then we compare the FITI algorithm with EH-Apriori as the way it is made in Anthony [10]. We adjust the value of maxspan, minsup, minconf and the size of transaction of synthetic data, respectively. The result shows that although FITI requires larger computer memory space than EH-Apriori, it outperforms the traditional algorithm, since FITI could save considerable running time.

IV. APPLICATIONS IN CHINESE REAL FINANCIAL DATA

In this section, we apply the FITI algorithm to various financial datasets to demonstrate the usefulness of intertransaction association rules in price prediction. All the datasets are obtained from the Wind Database. (http://www.wind.com.cn/).

A. Dataset of Chinese capital market

We first consider the dataset of Chinese treasury bond (T-bond) daily price index traded in Shanghai Stock Exchange (TB), Shanghai Composite Index (CI) daily price, daily exchange rate of Euro against RMB (ER) and US dollar against RMB (UR) traded in the inter-bank foreign exchange market from January 23, 2007 to June 23, 2007.

We use these data for several reasons. First, the capital market comprises of foreign exchange market, stock market, bonds and loan market. Most existing studies investigate the relation between stock market and bond market. Here we apply our method to the capital market and try to find intertransaction association rules from them.

We separate the dataset into two portions: one is UP, and the other is DOWN. The UP consists of the stock indices, bond indices and exchange rates that rise on the experiment day, and the latter group contains their decrease counterparts.

We set *maxspan* = 5 so that we could detect the association rules happening within five days. The results are quite interesting. One rule of the UP part is "ER (1), TB (2) $\Rightarrow$ CI (4)" with *support*=0.19 and *confidence*=0.86. This means that if Euro against RMB rises on the first day followed by the increase of Chinese Treasury bond on the second day, then the possibility that Shanghai Composite Index rises on the fourth day is 86%. For the Down part, a rule is "CI (1), ER (3) $\Rightarrow$ TB (5)" with *support* = 0.10 and *confidence* = 0.81, which indicates that if Shanghai Composite Index decreases on the first day and the exchange rate of Euro against RMB decreases on the third day , the possibility that the Chinese treasury bond index decreases is 81%. With these rules, one speculator can buy the stock index when the events in the first rule happen, and sell the bond when the events in the second rule happen.

From this example we can see the usefulness of intertransaction association rule mining in finding rules and stock price prediction.

B. Dataset of stock indices

The second dataset consists of Shanghai Composite Index (SCI)、Hong Kong Hang Seng Index (HSI)、Taiwan Weighted Stock Index (WTI) for 129 trading days from January 1, 2007 to June 29, 2007. Like the above method, the data is first separated into two groups. From the UP group, one interesting rule we find is "HSI (1), TWI (2) $\Rightarrow$ SCI (5)" with *support*=0.26 and *confidence*=0.81, which means that if the HIS and TWI increase on the first and second day, respectively, the possibility that the SCI increases on the fifth day is 81%.

This rule among the three indices may reveal a potential effect of the stock markets of Hong Kong and Taiwan on the stock market of China mainland. Such unaware relationship can be discovered using the intertransaction mining method introduced.

C. Dataset of futures market

The third dataset considers SHFE and LME copper (Cu) futures, NYMEX crude oil (CL) futures, exchange rate of US dollar against RMB and the exchange rate of Euro against US dollar for 334 trading days from July 3, 2006 to June 1, 2007.

We use futures market data for several reasons. According to Chan et al. [11], Jiang and Zhou [12] and Andersen et al. [13], there may be certain relations among the LME Cu, SHFE Cu, NYMEX CL and foreign exchange rate. The mining result shows that there is a certain relation among SHFE Cu, LEM Cu, and NYMEX HL. Let *maxspan*=4, one of the rules is "SHFE Cu (1), LEM Cu (3) $\Rightarrow$ NYMEX HL (3)" with *support*=0.11 and *confidence*=0.79. But we can't find any relation among them and foreign exchange rate.

V. Conclusions and Possible Future Work

There is no doubt that traditional association rule mining is a useful method for price prediction. However, the classical method is limited within individual transaction. In this paper, we demonstrate that the mysterious association rules among different transactions could be mined by the intertransaction association rule mining and FITI algorithm, and apply the intertransaction association rule mining method into Chinese financial market trend prediction. From our dataset analysis, we find the effectiveness of intertransaction association rule mining in mining rules among different economic indicators.

In the future, we will extend our study further to mine more advisable and interesting rules that could explain Chinese financial market better.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (No. 70971006, 70831001, 70501003) and Aeronautic Science Foundation (No. 2006ZG51079).

References

[1] R. Agrawal, T. Imielinski and A Swami. Mining association rules between sets of items in large databases, Proc. of the ACM SIGMOD Conf on Management of Data, 1993, pp. 207-216.

[2] R. Srikant, and R. Agrawal. Mining generalized association rules, Future generation computer systems, 13, 1997, pp. 161-180.

[3] M.J. Zaki, S. Parthasarathy, M. Ogihara, and W. Li. New algorithms for fast discovery of association rules, Proc. of the 3rd Intl. Conf. on Knowledge Discovery and Data Mining, 1997, pp. 283–286.

[4] R. Agrawal and R. Srikant. Fast algorithms for mining association rules, Research Report IBM, 1994.

[5] R. Srikant, and R. Agrawal. Mining generalized association rules, Proc. of the 21st Intl. Conf. on Very Large Data Bases, 1995, pp. 409–419.

[6] R. Miller, and Y. Yang. Association rules over interval data, In Proc. of the ACM SIGMOD Intl. Conf. on Management of Data, 1997, pp. 452–461.

[7] E. Baralis, and G. Psaila. Designing templates for mining association rules, Journal of Intelligent Information Systems, 9 (1), 1997, pp. 7–32.

[8] K. Wang, Y. He, and J. Han. Mining frequent itemsets using support constraints, Proc. 26th Intl. Conf. Very Large Data Bases, 2000, pp. 43–52.

[9] H. Mannila, H. Toivonen and A.I. Verkamo. Discovery of Frequent Episodes in Event Sequences, Data Mining and Knowledge Discovery, 1997, pp. 259-289.

[10] A. K.H. Tung, H.J. Lu, J.W. Han and L.Feng. Breaking the barrier of transactions: mining inter-transaction association rules, Proc. ACM SIGKDD Intl. Conf. Knowledge Discovery and Data Mining, 1999.

[11] K. Chan and G.A. Karolyi. Intrada volatility in the stock index and stock index future markets, Review of Financial Studies, 1991. pp. 657-684.

[12] X. B. Jiang, and Z.M. Zhou. An empirical study on causal relations of Cu in futures markets LME and SHFE, System Engineering, 2004, pp. 9-22.

[13] T. G. Andersen, T. Bollerslev, F. X. Diebold, and C. Vega. Real-time price discovery in global stock, bond and foreign exchange markets, Journal of International Economics, 2007.