Association for Information Systems

AIS Electronic Library (AISeL)

ICEB 2004 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-5-2004

Temporal Association Rule Mining in China's Closed-end Fund Data

Ying Liu

Guoqing Chen

Temporal Association Rule Mining in China's Closed-end Fund Data

Ying Liu, Guoqing Chen

School of Economics and Management, Tsinghua University, Beijing 100084, China
{liuy6, chengq}@em.tsinghua.edu.cn

ABSTRACT

Financial market plays an important role in economy. Although funds developed only a few years in China, it has been a focal point in research and practice. The conventional methods analyzing fund data are fundamental analysis and technical analysis. Data mining can extract implicit, previously unknown and potentially useful knowledge from data. This paper presents the new technique to analyze China's closed-end fund data and temporal association rules (TAR) are discovered which reflect the relationship among open price, close price, trading volume and grail index. Experimental results show some interesting outcomes.

Keywords: data mining, temporal association rules, closed-end fund

1. INTRODUCTION

Fund is an important investment tool. During the past several years, fund market has developed rapidly in China. By the end of 2003, there were totally 50 fund management companies issuing 54 closed-end funds and 56 opened-end funds. Closed-end funds had reached 86.2 billion RMB by net assets and opened-end funds had been up to 76 billion RMB by circulation [1]. There are at least two types of methods, namely fundamental analysis and technical analysis, for analyzing fund data, which are widely recognized and applied. Fundamental analysis typically focuses on the general economy, the condition of the industry and the condition of company. Technical analysis is an approach to stock investing where the past prices are studied and using charts as the primary tool [2].

Data mining is to discover interesting patterns from large amounts of data. Data mining functionalities include the discovery of association, classification, prediction, clustering and so on. As a promising technology, which may be supplementary to conventional methods, data mining in financial data has increasingly attracted attention of academia and practitioners. This paper focuses on finding potential association from closed-end fund data which is deemed helpful in fund investment decision-making.

2. TEMPORAL ASSOCIATION RULE

Association rule is one of the most important forms of knowledge [3]. It is of the form $X \Rightarrow Y$, where X and Y are item sets in databases. Two thresholds, namely degree of support and degree of confidence, are defined to evaluate the rule. For example, rule: trading volume rises $\Rightarrow$ close price rises, reflects a sort of association relationship between price fluctuation and trading volume. Apriori algorithm is regarded a classical algorithm [4] for association rules mining. Several efforts have been made to extend the classical association rule mining to inter-transactional association rule mining [5] and certain applications have been proposed in stock movement predictions [6].

Notably, fund data is typically time-series, which results in the need of temporal association rules mining. A Temporal Association Rule (TAR) is of the form $X \overset{t}{\Rightarrow} Y \ (t \geq 0)$ where X and Y are item sets, and t is the time-lag [7]. TAR is an extension of a conventional association rule, and represents that Y will occur at t time units after X occurs. Its degree of support (Dsupp) and degree of confidence (Dconf) are:

$$D\,\mathrm{supp}(X \overset{t}{\Rightarrow} Y) = \frac{\| X \underset{t}{\cup} Y \|}{|D| - t}$$

$$D\,\mathrm{conf}(X \overset{t}{\Rightarrow} Y) = \frac{\| X \underset{t}{\cup} Y \|}{\| X \|}$$

where $\|X\|$ is the number of records containing X at time point T, $\|X \underset{t}{\cup} Y\|$ is the number of records that contain Y at time point T+ t in correspondence to the records containing X at T, and $|D|$ is the number of records in dataset D.

Chen et al [8] proposed an algorithm so as to obtain a balance between the semantics and complexity. Two types (i.e. upward and downward) of the stock returns patterns are considered, which result in Boolean data items. Furthermore, in many real world applications, discovering associations for multiple stock returns patterns (e.g., categorical with 3 values: upward, unchanged, downward) is meaningful in supporting financial decision-making. Moreover, for continuous returns (e.g., return values in continuous domains), the respective domains can usually be descretized into several intervals. This method is further investigated and applied to closed-end fund data in China in this paper.

3. TAR MINING IN CLOSED-END FUND DATA

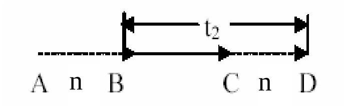
The characteristic of fund data is numerical, enormous and time-series. Our data for 2000-2003 come from the security database of TianXiang investment consulting. Every fund has its individual IPO time which results in different data volume. In order to assure the same external environment (i.e., policy, macroeconomics), selecting complete data in a same interval is necessary. Data from 2000.1.4 to 2002.12.31 were selected as the mining set and data from 2003.1.1 to 2003.12.31 will be used for test purposes. There are totally 20 funds that have complete data during this period in China market.

Let p(T) and p(T-n) denote the values of attribute p (i.e., p can be open price, close price, trading volume or grail index) of some fund on time points T and T-n (e.g., unit of T is day). The n-day rate of increase $P_n(T)$ is defined as in (1).

$$P_n(T) = \frac{p(T) - p(T-n)}{p(T-n)} \quad (1)$$

Theoretically the domain of $P_n(T)$ is $[-1,+\infty)$. For association rule mining purposes, numerical data need to be transformed into discrete data. Thereby what we want to discover is the temporal association rules between rates of increase for funds. Note that n is the time units during which the rate changes for a fund, and t is the time-lag for association rules. Figure 1 is an illustrative description for n and t.

Figure 1: Description of n and t



Let max(p) and min(p) respectively denote the maximum and minimum value of attribute p. K denote the number of discretized intervals. Then c denotes the interval span which is defined as in (2).

$$c = \frac{\max(p) - \min(p)}{K} \quad (2)$$

Let q_1, q_2,……,$q_{(k-1)}$ represent (K-1) points used to discretize the domain of $P_n(T)$. Then the values of $P_n(T)$ can be discretized into K intervals, where q_i (i=0, 1, 2, ……, K) is defined in (3).

$$q_i = \begin{cases} -1 & \text{i=0} \\ \min(p) + i \cdot c & \text{i=1,2,\dots,(K-1)} \\ \max(p) & \text{i=K} \end{cases} \quad (3)$$

When $P_n(T)$ belongs to $[q_i, q_{i+1})$, it is labeled as (i+1).

An example of the process is illustrated in Table 1 as follows with K=4, n=1, t=1. The values in Table 1 represent data of fund 500001 from Jan, 1st, 2000 to Jan, 10th, 2004 including open price, close price, trading volume and grail index.

Table 1: Data of fund 500001

date	open price(￥)	close price(￥)	trading volume(share)	grail index
2000-01-04	1.17	1.19	250,737	1,494.50
2000-01-05	1.2	1.17	228,309	1,498.05
2000-01-06	1.17	1.21	210,202	1,555.95
2000-01-07	1.21	1.25	628,288	1,611.81
2000-01-10	1.26	1.25	455,134	1,642.53

Each value in Table 2 represents the corresponding 1-day rate of increase of open price, close price, trading volume and grail index calculating by (1).

Table 2: Rate of increase of fund 500001

date	Rate of increase of open price	Rate of increase of close price	Rate of increase of trading volume	Rate of increase of grail index
2000-01-05	2.564103	-1.68067	-8.94483	0.237538
2000-01-06	-2.5	3.418803	-7.93092	3.865025
2000-01-07	3.418803	3.305785	198.8973	3.59009
2000-01-10	4.132231	0	-27.5597	1.905932

The discretization values of the rates are represented by integers in Table 3.

Table 3: Discretization data of fund 500001

date	Disctetiza-tion of open price	Discretiza-tion of close price	Discretiza-tion of trading volume	Discretiza-tion of grail index
2000-01-05	3	1	1	1
2000-01-06	1	4	1	4
2000-01-07	4	4	4	4
2000-01-10	4	2	1	3

The price of funds fluctuates greatly at ex-dividend data. It is usually regarded as a kind of noise which needs to be filtered out.

In Table 4, each of the first five columns represents fund 500001's discretization value of every attribute. The last two columns represent temporal data of open price and close price. Considering these temporal data as new attributes, temporal association rules can be obtained by Apriori algorithm on the new table. For example, if in this table, rule: *grail rises* ⇒ *temporal open price rises* can be discovered, then it is a rule with a time-lag of 1 day. A detailed description of the results can be seen in section 4.

Table 4: Fund 500001's discretization data added temporal attributes

ID	Open price	Close price	volume	grail	Temporal open price	Temporal close price
1	3	3	1	4	1	4
2	1	4	1	4	3	4
3	3	4	4	3	3	3
4	3	3	3	1	2	2
5	2	2	2	2		

4. RESULTS

Temporal association rule mining technique has been applied to closed-end fund data in China and a few interesting temporal association rules have discovered which reflect the potential knowledge hidden in the fund data.

Concretely, 10 rules with high level of Dsupp and Dconf are discovered. Most of the rules are consistent with those with conventional methods. However, some of the rules discovered are new that cannot be easily obtained by conventional methods including fundamental analysis and technical analysis.

4.1 Relationship among funds

According to the results of experiments, all funds support the rule *grail rises* $\overset{1}{\Rightarrow}$ *open price rises* with both high Dsupps and Dconfs, except for the funds with small sizes as 1 to 1.5 billion RMB. In addition funds issued by the same fund management company have similar Dconf with grail index.

First, we display the association between fund price and grail index. Over half of the rules are involved in 80 percent of the twenty funds, and parts of them are consistent with those with conventional methods. Rules are listed in Table 5. The result indicates that funds are very similar with each other. It is supposedly relevant to great similarity in stock-pick, means of manipulation, stock centralized degree, stock position noted in quarterly reports of each funds. All of these may lead to similar rules among funds and the extremely trend-similarity between funds with grail index.

Especially, it also can be noticed that small funds with1 to 1.5 billion RMB did not follow the same pattern *grail rises* $\overset{1}{\Rightarrow}$ *open price rises*, which indicates that small funds operating with the limited capital have to keep centralized stock investment, and thus behave less dependent on grail index.

Table 5: Temporal association rules

Temporal association rules: (minsupp=0.8 minconf=0.8)
grail rises $\overset{1}{\Rightarrow}$ open price rises
open price rises $\overset{1}{\Rightarrow}$ grail rises
volume falls $\overset{1}{\Rightarrow}$ close price falls
grail falls $\overset{1}{\Rightarrow}$ close price falls
volume falls, grail rises $\overset{1}{\Rightarrow}$ open price rises
grail rises, open price rises $\overset{1}{\Rightarrow}$ open price rises
close price rises, grail rises $\overset{1}{\Rightarrow}$ open price rises
volume falls, grail falls $\overset{1}{\Rightarrow}$ close price falls
grail rises, close price rises $\overset{1}{\Rightarrow}$ volume rises
volume falls, open price falls $\overset{1}{\Rightarrow}$ close price falls

Thus, the investment strategy of the small-sized funds is different from that of larger funds. Small funds may be more independent and their stock-picks are of more concern. When it comes to large funds, the focus should be on the present price and the volume of grail index should be focused rather than on a certain stock, since there exist strong association between larger funds and grail index.

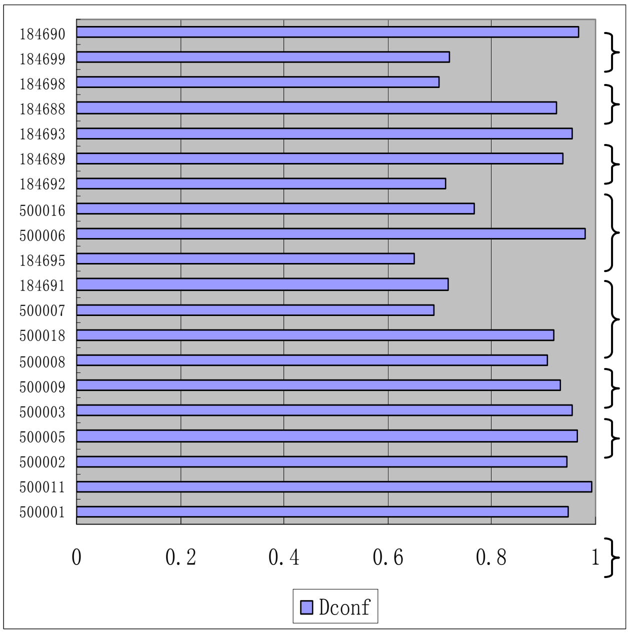


Figure 2: Dconfs of rule: *grail rises* $\overset{1}{\Rightarrow}$ *open price rises*

Second, Dconfs among 20 funds are compared. Figure 2 shows the Dconfs of rule: *grail rises* $\overset{1}{\Rightarrow}$ *open price rises* of all funds grouped by the fund management company. Funds which pertain to the same fund management company are indicated by parenthesis. It is obvious that most funds under the fund management company have similar values of Dconf. It implies that each fund management company carries out similar operation modes and manipulation ideas on its funds. The Dconfs of the funds in a certain fund management company are similar in most situations despite of different fund managers, which show that the characteristic of the manager is not as remarkable as that of the company. Figure 2 also reveals that it did not seem helpful to reduce the risk by purchasing different funds of the same company because of the similarity of the funds.

4.2 Price trend of funds

If both minsupp and minconf are set to 0.9, the rule: *close price rises* $\overset{1}{\Rightarrow}$ *close price rises* can be found in 8 funds which are listed in Table 6. The rule indicates that the higher the Dconf is, the longer that price keeps rising. Notably, six of the eight funds are also ranked very high based on the rate of increase of net asset.

Table 6 Rule: *close price rises* $\overset{1}{\Rightarrow}$ *close price rises*

Fund code	Dsupp	Dconf	Rank by rate ofincrease of net asset	Rate of increase of net asset
184690	0.942	0.956	1	29.99%
500003	0.915	0.932	2	29.93%
500002	0.904	0.921	3	28.66%
500011	0.972	0.977	5	21.44%

184693	0.929	0.939	7	19.17%
500006	0.957	0.967	8	17.82%
184689	0.905	0.913	12	14.63%
500001	0.915	0.935	13	14.45%

Results also show that more than 90% funds remain the same in content of association rules as time-lag changes, although Dconf changes in a small range. Experiments were carried out by setting time-lag as 1, 2, 3 days respectively.

Take Fund TongYi (184690) as an example. The results are listed in Table 7.

Table 7: Results of different temporal time

rule	t=1	t =2	t=3
grail rises $\stackrel{t}{\Rightarrow}$ open price rises	0.925	0.900	0.899
Close price rises, grail rises $\stackrel{t}{\Rightarrow}$ open price rises	0.939	0.910	0.898
volume falls, grail rises $\stackrel{t}{\Rightarrow}$ open price rises	0.928	0.899	0.896
grail rises, open price rises $\stackrel{t}{\Rightarrow}$ open price rise	0.935	0.908	0.909
grail rises, close price rises $\stackrel{t}{\Rightarrow}$ volume rises	0.956	0.953	0.9595
volume falls, grail falls $\stackrel{t}{\Rightarrow}$ close price falls	0.889	0.887	0.886
grail falls $\stackrel{t}{\Rightarrow}$ close price falls	0.891	0.892	0.891
volume falls, grail falls $\stackrel{t}{\Rightarrow}$ close price falls	0.893	0.889	0.890
volume falls, open price falls $\stackrel{t}{\Rightarrow}$ close price falls	0.895	0.895	0.899

Although values of Dconf change a little, they are still high, which implies that price trend is stable.

Investment suggestions may be summarized from above analysis: synthetically considering macroeconomic situation and trend of grail index, buying in and holding funds once the price of one of the eight funds rises, because the price of these funds will continue rising.

4.3 On test data

Data from Jan, 1st, 2003 to Dec, 31st, 2003 are used for test purposes. Because of space limitation, only two of the rules supported most in 20 funds are tested and listed in Table 8.

Table 8: Test results

Fund code	Dconfs of rule1	Dconfs of rule2
184688	0.85	0.86
184689	0.8	0.81
184690	0.71	0.74
184691	0.89	0.89
184693	0.87	0.88
184695	0.83	0.87
184698	0.83	0.83
184699	0.82	0.81
500001	0.79	0.8
500002	0.8	0.79
500005	0.81	0.82
500006	0.89	0.89
500007	0.79	0.8
500008	0.95	0.94
500009	0.8	0.73
500011	0.9	0.9
500016	0.71	0.73
500018	0.95	0.95
Average	0.83	0.83

These two rules are *grail rises* $\stackrel{t}{\Rightarrow}$ *open price rises* and *volume falls, grail rises* $\stackrel{t}{\Rightarrow}$ *open price rises*. The values of Dconfs are both averaged above 0.80. It demonstrates that funds followed the same patterns in long term, as the rules derived from the first three years data were still supported by the forth year data.

5. CONCLUSION

Nowadays, fund market plays a more and more important role in China economy. This paper introduced a new technique, temporal association rule mining, into the research of predicting fund return. Different from conventional methods, temporal association rule mining provides a data-driven method to discover interesting patterns in large-scaled fund data, which are previously unknown and potentially useful. The experiments on real closed-end fund data have shown that the method of mining temporal association rules in fund data was effective and could be used to support investment decisions.

ACKNOWLEDGEMENT

The work was partly supported by the National Natural Science Foundation of China (70231010, 70321001)

REFERENCES

[1] http://finance.sina.com.cn/ychd/20040105/1054589797.shtml

[2] R. Edwards and J. Magee. Technical analysis of stock trends. *American Management Assoiation*, 1997

[3] Jiawei Han, Micheline Kambera. Data Mining Concepts and Techniques. *Higher Education Press*,2000

[4] Rakesh Agrawal, Tomasz Imiclinski, Arun Swami, Mining Association Rules between Sets of Items in Large Databases, *Proceedings of the 1993 ACM SIGMOD Conference* Washington DC, USA, May 1993

[5] Tung, K.H., Lu, H., Han, J., Feng, L., Breaking the Barrier of Transactions: Mining Inter-Transaction Association Rules, *the AcM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD99)*, San Diego, CA, Auguest 15-18, 1999

[6] Lu, H., Han, J., Feng, L., Stock Price Movement Prediction and N-Dimensional Inter-Transaction Association Rules, *1998 ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD98)*, Seattle, Washington, June 5, 1998

[7] G. Das, King-Ip Lin, H. Mannila. Rule Discovery from Time Series. *KDD98*,1998

[8] Chen G Q, Ai J, Yu W., Discovering Temporal Association Rules For Time-Lag Data. *In Proceedings of International Conference on e-Business (ICEB2002)*, 2002. 324