

Application of Data Technology in Security Market ——Phase Space Reconstruction Data Mining Method Based on Time Series

Ji XueJing^{1, a}

¹Full Northeastern University, information and computing science, 1501, Shenyang, Liaoning, 110004
China

^a2650037388@qq.com

Keywords: phase space reconstruction; time series; data mining

Abstract. this paper starts from concept and process of data mining, it illustrates mining topic and method in security analysis and it makes analysis application of data technology in security market based on data mining method of time series phase space reconstruction, which provides new research clue for data mining technology.

1. Introduction

Security market is the barometer of national economy; security traders have high requirements on correctness, real-time and safety of data. The transaction system of every trader is always listed top in technology application and innovation, meanwhile it accumulates rich data. The data produced by the whole operation system is mainly divided into 2 kinds: stock market data and customer transaction data. Stock market data is generated by exchange; it is universally distributed and real-time sharing information. Data is to discover knowledge, mine, develop and apply these data, which can make enterprises make the most proper positioning, so that they can establish competition advantage. This paper is based on phase space reconstruction data mining method of time series; it makes analysis on application of data technology in security market.

2. Application of data technology in security market

Data mining technology is divided into basic analysis method, technology analysis method and combined analysis method in the application field of security analysis. People have summarized a set of technology analysis methods on the basis of practice and research for long time, these methods are basically established on the basis of statistics, such as moving average method, point and figure method, K-line method etc.

A Time series model

In finance field, time series data is one kind of important data model, making analysis on time series is one kind of important content of financial data analysis. The traditional financial times series analysis model and method mainly use mathematical statistics model as basis to obtain mathematical model of describing financial time series and discipline through model assumption, parameter evaluation and model test etc measures and technology, but it can not support some new application and analysis tasks in the financial field. Time series model is one kind of method based on historical record series to establish model, it also includes some relevant series model or the historical record based on series transfer model function. One p-phase auto-regression AR (p) model means time series x_t , it has the following ways:

$$x_t = \sum_{i=1}^p a_i x_{t-i} + \varepsilon_t$$

Of which, ε_t is white noise. It uses lag operation L and can be indicated as follows:

$$x_t = (1 + \sum_{j=1}^q \theta_j L^j) \varepsilon_t = \theta(L) \varepsilon_t$$

One q-phase moving average MA (q) means time series x_t has the following ways:

$$x_t = \sum_{j=0}^q \theta_j \varepsilon_{t-j}$$

Of which, $\theta_0=1$, ε_t is white noise. It uses lag operator and can be indicated as follows:

$$x_t = (1 + \sum_{j=1}^q \theta_j L^j) \varepsilon_t = \theta(L) \varepsilon_t$$

Auto-regression moving platform ARMA (p,q) model is combination of AR and MA, it means time series x_t has the following ways:

$$\alpha(L)x_t = \theta(L)\varepsilon_t$$

Of which, $\alpha(L)$ and $\theta(L)$ have no common factor to guarantee the unique expression way. If time series x_t undergoes difference for d times, it has expression way of ARMA (p,q), it means meeting the following formula:

$$\alpha(L)\Delta^d x_t = \theta(L)\varepsilon_t$$

Stock pricing model can find stock price is over-estimated or underestimated by calculating and comparing stock internal value with its actual market price. The listed company financial report form provides various of data and relevant information reflecting company management condition and finance state, while according to regulations of accounting system, important financial data, report form etc must be kept for over 10 years. Except for plenty of data with long time span, security analysis dose not only need internal data of enterprise, but also external data, not only needs current data, but also needs historical and future development data, not only needs relevant data reflecting enterprise production and management activity, but also needs data in aspects of market, price, finance, investment etc. Therefore, there is massive data. In addition, security analysis should not solely focus on certain subjects, it should make comprehensive analysis together on company financial report form and macro economic surface, data mining technology is good at making process such as concentration, storage, classification, recombination, analysis, check etc on massive data, meanwhile it makes multi-perspective and multi-level analysis and observation on each kind of data, predict development trend in the future according to current data.

B. Series time model mining of stock yield rate

China Shanghai and Shenzhen Stock Exchange set price fluctuation limit on December 16th in 1996; fluctuation limit is within 10%. It refers to research by Povinelli, set phase space dimension $d=2$, dimension value change range is $[-10, 10]$, because of multi-solution characteristic of small change, firstly it can choose smaller dimension interval. Considering 2-step sampling of small wave, take dimension division $m=160$. It uses yield rate as miner to describe interest degree of event, take event characteristic function $e(x_t)=x_{t+1}$, of which, $x_t=(Y_t - Y_{t-1})/Y_{t-1}$, y_t is the closing price of stock index in t transaction day, it means using the next point of time model as event characteristics value. Option of event characteristic function has big effect on mining method, in the practice, single-step prediction effect is relatively better. It still uses single-step prediction idea and take time-delay width $\tau=1$, which means listing daily yield rates with continual change to describe time model. Considering on making transaction in stock exchange, it can be also invested into currency market without risk, set event trigger marker ε as daily yield with continuous compounding equals to bank deposit interest r yield for 1 year, which requires the unit average yield rate of clustering concentration less than ε , it uses prediction point as event point. If time model vector of prediction point gets in clustering set, after this model is still less than ε , it uses prediction point as event point. While credibility threshold value δ is used to describe credibility of clustering, δ is used as point total correlation in data space, take threshold value $\delta=0.05n$, which means category, whose density is bigger than 5%, which can be used as credible clustering result. From August 23rd of 1996 to October 29th of 2004, People's Bank of

China has carried out interest reduction for 7 times and interest increase for 1 time. Considering high correlation in financial market, interest change may cause step change in time model, so this chapter chooses the following 4 without any interest adjustment and continual 340 transaction days as data set: SZ1: from June 6th of 1999 to November 10th of 2000, RMB deposit interest of 1 year is 2.25%. SZ2: from September 12th of 2000 to February 8th of 2002, RMB deposit interest of 1 year is 2.25%. SZ3: from March 12th of 2002 to August 7th of 2003, RMB deposit interest of 1 year is 1.98%. SZ4: from November 19th of 2002 to April 19th of 2004, RMB interest deposit of 1 year is 1.98%.

Table 1 Mining result comparison of different small wave clustering

	Haar		Bior1.3		Bfor1.5		Carry benefit%
	Event	Yield rate%	event	Yield rate%	event	Yield rate%	
SZ1	59	29.04	56	34.53	57	35.67	30.55
SZ2	64	-16.88	61	-21.71	60	-17.99	-30.46
SZ3	70	9.05	73	12.61	68	6.25	-2.41
SZ4	64	28.69	63	36.88	64	35.22	16.96
平均	64.25	12.475	63.25	15.5775	62.25	14.7875	3.66

In order to distinguish whether predicted event has obvious difference with non-event or not, 2 samples test t statistics under positive overall sample average comparison on constructing different variance:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_x^2 / m + S_y^2 / n}}$$

When 2 overall average difference is not obvious, this statistics amount conforms to t distribution of free degree as m+n-2. There is assumption test: HO: event series sample has no difference with non-event series sample. H1: event series sample has obvious difference with non-event series sample. Every group correlation statistics and credibility calculation result is indicated by table 2.

Table 2 Statistics result list table of event series and non-event series

	Event point E	Average (E)	Standard deviation (E)	Non-event point (\hat{E})	Average (\hat{E})	Standard deviation (\hat{E})	Credibility α
Harr_sz1	59	-0.1155	0.9708	179	0.0667	1.5958	0.0114
Harr_sz2	64	-0.4966	1.5183	174	-0.0536	1.5654	0.3819
Harr_sz3	70	-0.1209	1.2963	168	0.0752	1.1828	0.2067
Harr_sz4	64	-0.1063	1.2405	174	0.1214	1.1918	0.1872
Bior1.3_sz1	56	-0.1600	0.9635	182	0.0774	1.5871	0.0119
Bior1.3_sz2	73	-0.5015	1.5392	177	-0.0595	1.5582	0.5437
Bior1.3_sz3	63	-0.081s	1.2026	165	0.0579	1.2260	0.0577
Bior1.3_sz4	57	-0.1783	1.2283	175	0.1460	1.1907	4.5X10 ⁻¹³
Bior1.5_sz1	60	-1.304	0.9845	181	0.0694	1.5871	0.0284
Bior1.5_sz2	68	-0.4628	1.5538	178	-0.0750	1.5571	0.2771
Bior1.5_sz3	59	-0.1155	0.9708	168	0.0667	1.5958	0.0418
Bior1.5_sz4	64	-0.4966	1.5183	174	-0.0536	1.5654	0.0424

C. Series time mining model of stock yield rate

Research object distributed in the above-mentioned yield rate is stock index yield rate; it chooses Shanghai Composite Index and Shenzhen Composite Index as sample to construct yield series.

$$X = \{x_n, x_n = (y_{n+1} - y_n) / y_n\}$$

Of which, Y_n and y_{n+1} is the closing index of that day and the next day. Take Shenzhen Composite Index for example, data segment ranges from August 1st of 2002 to September 30th of 2004, and the total length of data segment is 526 transaction days. This paper adopts original BP neural network and EM clustering fuzzy neural network to make prediction. BP network prediction result of Shenzhen Composite Index is indicated by figure 1. EM clustering fuzzy neural network prediction result of Shenzhen Composite Index is indicated by figure 6.2. Predicted absolute error of Shenzhen Composite Index neural network is 3-43%, predicted absolute error of EM clustering network is 1.57%.

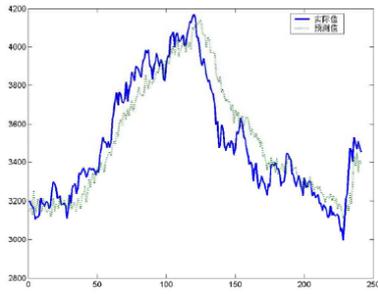


Figure 1 BP network prediction curve of Shenzhen Composite Index

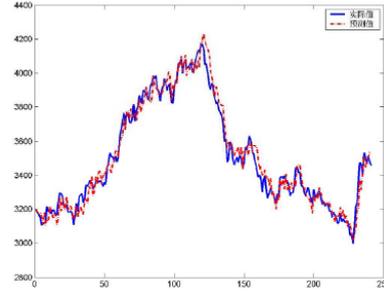


Figure 2 EM clustering fuzzy neural network prediction curve of Shenzhen Composite Index

Take data segment of the same length, Shanghai Stock Exchange ranges from December 5th of 2000 to February 21st of 2003, Training set and test set division is the same with the above-mentioned. Predicted range is from February 25th of 2002 to February 21st of 2003.

It applies the most adjacent point method and mutual information method to calculate Shanghai Stock Exchange embedded dimension $D=13$, embed $\tau=5$, the maximum cut apunov index $L=0.0048>0$. Take predicted step length $p=1$, predicted result of Shanghai Stock Exchange is indicated by figure 3. Predicted result of Shanghai Stock Exchange EM clustering fuzzy neural network is indicated by figure 4. The predicted average error of Shanghai Stock Exchange BP neural network is 3.92%. The predicted absolute error of EM clustering network is 5.87%. Demonstration result indicates that when stock price trend is obvious, predicted result of BP neural network can make better description on the overall index change, when stock price trend presents big fluctuation, its predicted result will appear big fluctuation. Predicted error of EM clustering neural network is low, which has higher fitness with actual curve.

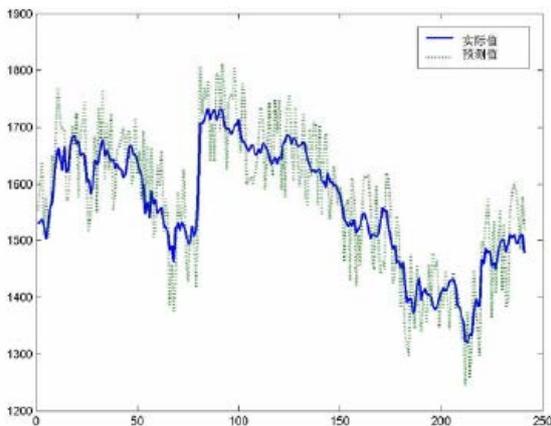


Figure 3 BP neural network prediction diagram of Shanghai Stock Exchange

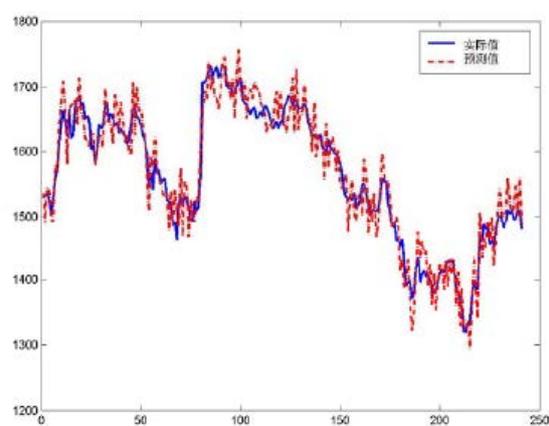


Figure 4 EM clustering fuzzy neural network prediction diagram of Shanghai Stock Exchange

It applies the most adjacent point method and mutual information method to calculate Shanghai Stock Exchange embedded dimension $D=13$, embed $\tau=5$, the maximum cut apunov index $L=0.0048>0$. Take predicted step length $p=1$, predicted result of Shanghai Stock Exchange is indicated by figure 3. Predicted result of Shanghai Stock Exchange EM clustering fuzzy neural network is indicated by figure 4. The predicted average error of Shanghai Stock Exchange BP neural network is

3.92%. The predicted absolute error of EM clustering network is 1.87%. Demonstration result indicates that when stock price trend is obvious, predicted result of BP neural network can make better description on the overall index change, when stock price trend presents big fluctuation, its predicted result will appear big fluctuation. Relatively speaking, predicted error of EM clustering neural network is low, which has higher fitness with actual curve.

3 Conclusion

Data mining highly automatically analyzes data and makes inductive reasoning; it mines the potential model and has wide application prospect in security analysis. Whether stocker understand trend and make correct investment decision-making, which has important reference meanings for economist to analyze relations between user investment behavior of user with different levels and various of stocks, timely find abnormal behavior in stock market, adjust market strategy for listed company, detect risk, for government to issue new plan etc.

References

- [1] Fang Rong. Analysis and Application on Data Mining Algorithm Based on Association Rule [J]. Electronic Test, 2016, 01:36-38+16.
- [2] Ren Yanfei, Qi Chuanwei. Simulation Research on Correct Prediction of Security Trend [J]. Computer Simulation, 2013, 03:217-219+388.
- [3] Xu junfang, Cui Jing. Design and Analysis on Security Company and Customer Value Based on Data Mining [J]. Value Engineering, 2011, 21:141-142
- [4] Xiong Yuliang, Gao Jian. Discussion on Application of Data Mining in Security Customer Service [J]. Manager and Administrator, 2015, 24:30.
- [5] Liu Bin, Lei Yong. Application Research on Data Mining in CRM System of Security Company [J]. Information Technology and informatization, 2015, 07:55-56.