Research on the intelligent investment systems based on fuzzy clustering ensemble analysis

Wei Gao[1, a *], Ruzhen Yan[2]

[1]Business School, Sichuan Agricultural University, Chengdu, China

[2]Business School, Chengdu University of Technology, Chengdu, China

[a]Email: gaowei@sicau.edu.cn

Keywords: clustering analysis; fuzzy clustering; intelligent systems; securities investments

Abstract. This paper develops a fuzzy clustering method based on the fuzzy t-norm, and uses the data of China securities market to analyze the factors influencing the portfolio selection. Furthermore, this paper estimates an investment portfolio based on fuzzy clustering. The results show that the investor can learn the feature of securities. This method is very important for the portfolio selection.

Introduction

In recent years, with the sustained and rapid development of economy, the capital market obtained the rapid development. However, affected by macroeconomic policies and the investor factors, the speculative arbitrage activities are very frequent in China capital market. The investors urgently need the guidance of the relevant investment theory, so that can make a rational investment activity. Hence, it is very important to further study the modern portfolio theory for the investor.

Portfolio selection is the study of how people should invest their wealth. It is a process of trading off risk and expected return to find the best portfolio of assets and liabilities. Cluster analysis has been widely applied in portfolio selection. However, it is an ill-posed problem and no single clustering algorithm is able to obtain satisfactory solutions for all types of data sets in customer segmentation. Cluster ensemble emerges as an alternative approach for improving the investment performance. Using the clustering analysis, investors can understand the basic features of securities, and determine the scope of the portfolio choice. The traditional clustering analysis is that all objects or characteristics need to identify and strictly divided. In this process, the division of category boundaries is very obvious. However, in the real situation, many characteristics of objects are not obviously. In order to solve this situation, this paper develops a cluster segment method based on the fuzzy t-Norm, considers the problem of how to construct a port-folio based on the bid-ask spread, depth, order imbalance, market value, profit margins, asset-liability ratio.

Portfolio selection problem has important academic value for social and economic development. Eastham and Hasting(1988) develops a portfolio selection problem with fixed and variable transactions costs. The solution of this problem is given by way of quasi-variational inequalities for the object function. Best and Hlouskova(2003) analysis the portfolio selection problem of risky assets with a diagonal covariance matrix, upper bounds on all assets and transactions costs, and develop an algorithm for its solution. Meanwhile, they present the efficient portfolios under appropriate assumptions. Lobo et al.(2007), Ren et al.(2012) consider the portfolio selection problem with transaction costs and constraints on exposure to risk. Portfolio optimization problems with transaction costs that include a fixed fee, or discount breakpoints, cannot be directly solved by convex optimization. They get a suboptimal solution and an upper bound on the optimal solution. Kim et al. (2003) used neural network clustering method to segment the customers of tourism; contrasting K-means, self-organizing map neural net-work and particle swarm optimization for three kinds of clustering algorithm, Deng et al. (2011) proposed hybrid clustering algorithm which was used for segmentation problem of catering industry customer. The above data mining customer segmentation method is constantly applied to the research of customer segmentation, and has achieved good effective in some empirical studies.

In this paper, we consider the problem of how to construct a portfolio based on the bid-ask spread, depth, order imbalance, market value, profit margins, asset-liability ratio, and develops a cluster segment method based on the fuzzy t- Norm.

Related Theoretical Knowledge

Cluster analysis has been widely applied in portfolio selection. However, it is an ill-posed problem and no single clustering algorithm is able to obtain satisfactory solutions for all types of data sets in customer segmentation. Cluster ensemble emerges as an alternative approach for improving the performance of cluster. Markowitz(1952) develops the mean-variance model that is a quantitative trade-off between risk and expected return. In this model, there is not the any transaction cost. However, the transaction cost is essential in the real securities market. Hence, when the investor constructing the portfolio selection, they must consider the bid-ask spread, depth, order imbalance, market value, profit margins, and asset-liability ratio.

In this paper, the stock evaluation indexes include:

(1) Return on equity

Return on equity (ROE) measures the rate of return for shareholders' equity, and measures the efficiency of a firm at mean profits from each unit of shareholder equity, also known as net assets or assets minus liabilities. This index shows how well a company uses investments to generate earnings growth, and can be calculated as following:

$$ROE = \frac{NetR}{StockH} \qquad (1)$$

(2) Quick ratio

In finance, the quick ratio mainly measures the ability of a company to use its near cash or quick assets to extinguish or retire its current liabilities immediately. A company with a quick ratio of less than 1 cannot currently fully pay back its current liabilities.

(3) Net profit grow rate

Net profit, also referred to as net income, is a measure of the profitability of a venture after accounting for all costs.

$$NPGR = \frac{NetP - LNetP}{LNetP} \qquad (2)$$

where, *NetP* refers to the net profit; *LNetP* represents the net profit of last year; *NPGR* refers to the net profit grow rate.

(4) Liquidity index

Liquidity refers to the ability to facilitate the purchase or sale of stocks without causing drastic change in the stock price. In a liquid market, selling quickly will not reduce the price much. In a relatively illiquid market, selling it quickly will require cutting its price by some amount. In this paper, turnover is used to measure the liquidity.

$$TR_t = \frac{1}{n} \sum_{i=1}^{n} \frac{volume_i}{dsmvosd_i} \times 100\% \qquad (3)$$

(5)Order imbalance

A situation resulting from an excess of buy or sell orders for a specific security on a trading exchange, making it impossible to match the orders of buyer and seller. For securities that are overseen by a market maker or specialist, shares may be brought in from a specified reserve to add liquidity, temporarily clearing out excess orders from the inventory so that the trading in the security can resume at an orderly level. Extreme cases of order imbalance may cause suspension of trading until the imbalance is resolved.

$$Orderimb_t = \frac{1}{n} \sum_{i=1}^{n} \left| BV_i - SV_i \right| \qquad (4)$$

Portfolio selection based on the fuzzy clustering method

There were 30 stocks selected from Shanghai and Shenzhen stock market as samples for this study. Based on the samples, we estimate and analyze the fuzzy clustering ensemble.

The descriptive statistical results of this study are shown in table 1.

Table 1 Descriptive statistics

variable	Mean	Meddle	Variance
ROE	9.1103	7.6784	7.0333
QR	1.3222	1.246	0.7355
NPGR	46.6468	18.2576	176.2601
TR	1.1208	1.0776	0.9708
Orderimb	12899.92	10565.11	9489.38

For the data of 30 stocks, the clustering analysis and result is shown using fuzzy clustering ensemble method by the MATLAB 7.0 software. The result is shown in Table 2.

The I group includes: Wuhan Iron and Steel Company Limited(600005), Hangzhou Iron & Steel Group Company(600126), HUANENG International Power Inc.(600011), LANTAI Industry(600328), LAOFENGXIANG Co., Ltd.(600612), Lanzhou Great Wall Electrical Corporation (600192). The II group includes: Shanghai Guangdian Electric Group Co., Ltd.(601616), HUAYI Electric Co., Ltd.(600290), Guangzhou Development Group Incorporated(600098), Jiangsu Yulong Steel Pipe Co., Ltd.(601028). The III group includes: CITIC Securities(600034), SHANGHAI ZHONGJI Investment Holding Co., Ltd.(600634), Shanghai Pudong Development Bank(600000), China CNR Corporation Limited(601299), China Meheco Co., Ltd.(600056).

Table 2 The result of fuzzy cluster

	I	II	III	IV	V
1	600005	601616	600030	600176	600219
2	600126	600290	600634	600016	600028
3	600011	600098	600000	600824	600104
4	600328	601028	601299	600064	600068
5	600612		600056	600261	600007
6	600192				600667
7					600113
8					600019
9					600230
10					600565

The IV group includes: China Fiberglass Company Ltd.(600176), China Minsheng Banking Co.,Ltd(600016), Shanghai Yimin Commercial Group Co. Ltd(600824), NANJING XINGANG High-tech Co., Ltd.(600064), Zhejiang Yankon Group Co.,Ltd.(600261). The V group includes: SHANDONG NANSHAN Aluminum Co., Ltd.(600219), CHINA Petroleum & Chemical Corporation(600028), China World Trade Center Co. ,Ltd.(600007), SAIC Motor Corporation Limited(600104), CGGC International Ltd.(600068), WUXI TAIJI Industry Co., Ltd.(600667), ZheJiang DongRi Company Limited(600113), :Baoshan Iron & Steel Co.,Ltd.(600019), HEBEI CANGZHOU DAHUA Co., Ltd.(600230), Chongqing DIMA Industry Co., Ltd.(600565).

Conclusions

Portfolio selection is the study of how people should invest their wealth. It is a process of trading off risk and expected return to find the best portfolio of assets and liabilities. Cluster analysis is the task of grouping a set of objects in such a way that objects in the cluster are more similar to each other than to those in other clusters. It is a main task of exploratory data mining, and a common technique for

statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, portfolio selection.

This paper develops a cluster segment method based on the fuzzy t-Norm, considers the problem of how to construct a portfolio based on the bid-ask spread, depth, order imbalance, market value, profit margins, asset-liability ratio. The method from this paper provides guidance in portfolio selection for the investors.

References

[1] X.Q. Zeng, Q. Xu and D Zhang, New multi-indicator customer segmentation method based on consuming data mining, Application Research of Computers, 30 (2013), 2944-2947.

[2] P.V.Freytag, Business to business market segmentation, In-dustrial Marketing Management, 30 (2001), 473-486.

[3] L. Chen, K. Soliman, S.E. Mao, Measuring user satisfaction with data warehouse: An exploratory study, Information & Management, 37 (2000), 103-110.

[4] H.V. Shashidhar, V. Subramanian, Customer segmentation of bank based on data mining security value based heuristic ap-proach as a replacement to k-means algorithm, International Journal Computer Application, 19 (2011), 13-18.

[5] D.S. Boone, M. Roehm, Retail segmentation using artificial neural networks, International Journal of Research in Market-ing, 19 (2002), 287-301.

[6] J Kim, et al. Segmentation the market of West Australian senior tourist using artificial neural network [J]. Tourism Management, 2003, 24 (1): 25-34.

[7] X.Y. Deng, C. Jin, Q.P. Han, KSP: a hybrid clustering algo-rithm for customer segmentation in mobile E-commerce, Journal of Management Science, 24 (2011), 54-61.

[8] A. Strehl, J. Ghosh, Cluster Ensembles: a knowledge reuse framework for combination multiple partition, Journal of Ma-chine Learning Research, , 3 (2002), 583-617.

[9] E.B. Fowlkes, R. Gnanadesikan, J.R, Kettenring, Variable selection in clustering, Journal of Classification, 5 (1998), 205-228.

[10] S. Dudoit, J. Fridlyand, Bagging to improve the accuracy of a clustering procedure, Bioinformatics, 19 (2003), 1090-1099.

[11] Z.H. Zhou, W. Tang. Cluster ensemble, Knowledge-Based Systems, 19 (2006), 77-83.

[12] W. Gao,C.Z. He, X.Y. Jiang, Customer segmentation study based on fuzzy clustering ensemble, Journal of Intelligence, 30 (2011), 125- 129.

[13] Y. Yang, F. Jin, M. Kamel, Survey of clustering validity eval-uation, Application Research of Computer, 25(2008), 1630-1633.

[14] Qian W, ZHOU A. Analyzing popular clustering algorithms from different viewpoints[J]. Journal of Software, 2002, 13(8): 1382-1394.

[15] Zadeh L A. Fuzzy sets[J]. Information and Control, 1965, 8(3): 338-353.