Empirical Analysis of Stock Investment Based on $\beta$-KFCM

Yuxue Wang, Zongxiang Song and Xiaoping Ren*

School of Mathematics and Statistics, Northeast Petroleum University, Daqing, China

*Corresponding author email: xpren0703@stu.xjtu.edu.cn

Abstract. Cluster analysis is an analysis method that classifies data sets. It measure the similarity between different data and classifies them on the basis of similarity, but the classification is unknown. In recent years, with the popularization and improvement of information technology, a large number of diversified data have been produced in various industries. In order to extract data features and classify them, this paper, based on the study of fuzzy C-means clustering algorithm, take the systematic risk coefficient as the weight, expand the model to get the weighted fuzzy kernel clustering algorithm: $\beta$-KFCM fuzzy clustering algorithm and solve the algorithm. Based on the empirical research, we choose six Accounting Indicators to study 50 stocks in A-share market, including current ratio, cash ratio, year-on-year growth rate of operating revenue, return rate of total assets, net sales interest rate and asset liability ratio. According to the empirical research, we can get the results that the investment grade is divided into three categories, among them, Category I is the recommended high-quality investment variety, category II and category III are next to it.

1. Introduction

At present, people's investment awareness is growing. Investors no longer only choose bank deposits, also stocks, bonds and a variety of financial derivatives. It is still the most important way to choose stocks for investment, however the concept of general investors and the method of constructing portfolio are not mature, most of them rely on fundamental analysis and K-line chart to make decisions. In fact, there are stocks with similar characteristics in the stock market. By analyzing and classifying them, we can build a stock portfolio with decentralized risk to maintain stable or multi-dimensional returns. Clustering analysis is based on unsupervised learning method, which is more suitable for the application of multi-dimensional data in practice, but there are few empirical studies in financial market research. K-means algorithm[1] proposed by MAC queen in 1967 is one of the most widely used clustering algorithms. Yan Wei[2] proposed KPCA-DBEFCM algorithm, which improved FCM algorithm, innovatively used data set density information to construct equilibrium term. Yao Hong-man[3] proposed an improved artificial bee colony (ABC-SC) algorithm to adjust and optimize the fuzzy parameters of FCM and improve the accuracy of clustering. Sun Li-Juan[4] and others proposed a fuzzy clustering algorithm based on weight attenuation for data flow. Deng Lun-Bing[5], Zhang Wen-Qi[6] and Liu Yi-Ping[7] all use the traditional cluster analysis method to select different accounting indicators to cluster the stocks for empirical analysis. However, there are two deficiencies in the previous research: First, the traditional clustering analysis method maximizes the homogeneity of the objects among the Categories and the heterogeneity of the objects between the Categories. The classification conditions of this hard clustering are too harsh; Second, only accounting indicators are selected for classification in the process of classification, without considering the volatility of the stock's own rate of return.

In this paper, the contribution of portfolio risk to the overall market risk is introduced into FCM...
clustering, and a weighted algorithm is established: -KFCM fuzzy clustering algorithm. The classification condition is better than the traditional clustering method, and the risk of individual stocks is included in the classification result.

2. Risk Weighted Fuzzy Kernel Clustering Model and Its Algorithm

2.1. Risk Weighted Fuzzy Kernel Clustering Model
KFCM algorithm considers that all feature attributes play the same role in the clustering process, but in fact, different attributes have different contribution to the clustering structure. In view of this defect, after studying the characteristics of securities, this paper presents a simple improved algorithm for the stock principle after studying the characteristics of securities. The algorithm takes the contribution rate of stock portfolio risk to market variance into account, in the process of clustering, adjust the importance of each attribute to different categories according to the specific characteristics of different categories. According to this idea, -KFCM weighted fuzzy kernel clustering algorithm is improved in this paper. The following introduces the weighted fuzzy kernel clustering algorithm. This method fully considers the imbalance between attributes, and it can truly reflect the actual situation of clustering problem.

Suppose the sample set }{ \{x_1, x_2, \ldots, x_n \} to be clustered with n samples. After nonlinear transformation \( \varphi \), from the original input space to the high-dimensional kernel space, it is expressed as \( \varphi(x_k) \). Suppose \( \mu_{jk} \) denote the membership degree of the k-th sample to the j-th category, and satisfy that the sum of the membership degrees of each sample to each cluster is 1. That is \( \sum_{j=1}^{c} \frac{\delta_{kn}}{\delta^2} \mu_{jk} = 1 \). The fuzzy partition matrix is defined as follows:

\[
U = \frac{1}{\delta^2_m} \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1k} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{jk} & \mu_{j2} & \cdots & \mu_{jk} \end{pmatrix}_{cxn}
\] (1)

Among them, the j-th row of matrix \( U \) is the membership function of the j-th clustering subset, the k-th column of matrix \( U \) is the membership value of sample \( X_k \) for the number of c, that is, \( \mu_{jk} \) represents the membership degree of sample \( X_k \) to the j-th category. \( \frac{\delta_{kn}}{\delta^2_n} \) represents the proportion of the k-th stock in the overall risk, then the objective function of -KFCM clustering algorithm is:

\[
J_P = \sum_{j=1}^{c} \sum_{k=1}^{n} \mu_{jk} \frac{\delta_{kn}}{\delta^2_n} \| \varphi(x_k) - m_{ji}^\varphi \|^2
\] (2)

2.2. Solving Process of Objective Function
In order to obtain the optimal solution of the objective function, the following constraints can be applied: \( \sum_{j=1}^{c} \mu_{jk} = 1 \). The solution of the extreme value problem of the objective function mentioned in the previous section can be obtained by the Lagrange multiplier method, suppose the Lagrange function is:

\[
\]
\[ L_\beta = \sum_{j=1}^{c} \sum_{k=1}^{n_j} \mu_{jk} \frac{\delta b_{jk}}{\delta n} \right| \varphi(x_k) - m_j^p \right|^2 - \sum_{k=1}^{n_j} \lambda_k \left[ \sum_{j=1}^{c} \mu_{jk} - 1 \right] \quad (3) \]

let \( L_\beta \) offset \( m_j^p, \mu_{jk}, \lambda_k \) and equal to zero:

\[ \frac{\partial L_\beta}{\partial m_j^p} = 2 \sum_{k=1}^{n_j} \mu_{jk}^b (\varphi(x_k) - m_j^p) = 0 \quad (4) \]

\[ \frac{\partial L_\beta}{\partial \mu_{jk}} = \sum_{j=1}^{c} \left[ b \sum_{k=1}^{n_j} \mu_{jk}^{b-1} \right] \right| \varphi(x_k) - m_j^p \right|^2 = 0 \quad (5) \]

\[ \frac{\partial L_\beta}{\partial \lambda_k} = m \left( \sum_{j=1}^{c} \mu_{jk} - 1 \right) = 0 \quad (6) \]

The results are as follows:

\[ m_j^p = \frac{\sum_{k=1}^{b} \mu_{jk}^b \frac{\delta b_{jk}}{\delta n} \varphi(x_k)}{\sum_{k=1}^{n_j} \mu_{jk}^b \frac{\delta b_{jk}}{\delta n}} \quad (7) \]

\[ \mu_{jk} = \frac{\left[ \varphi(x_k) - m_j^p \right]^{2-b}}{\sum_{j=1}^{c} \frac{\delta b_{jk}}{\delta n} \left[ \varphi(x_k) - m_j^p \right]^{2-b}} \quad (8) \]

The implementation of fuzzy clustering algorithm is similar to KFCM algorithm until the end of convergence. At this point, the final clustering center and the fuzzy partition matrix of the sample set are obtained. This algorithm combines the systematic risk coefficient and KFCM fuzzy clustering algorithm in the capital asset pricing model, considers the return and risk of a single security and the return and risk of the market, and proposes to apply the risk as the weight to the fuzzy kernel clustering.

3. Empirical Research and Analysis

Based on the data of 50 stocks selected in this paper, considering the completeness of data, the length of time and the accuracy of empirical analysis, select the fourth quarter of 2015, and collect the data of six related indexes from the software of TongHuaShun, in order to eliminate the influence of different dimensions, the original data is standardized. Some results in 2015 are shown in table 1, among them, variables X1-X6 represent current ratio, cash ratio, year-on-year growth rate of operating revenue, return rate of total assets, net sales interest rate and asset liability ratio, and variable Y represents stock number.
Table 1. Partial standardized data in the fourth quarter of 2015

|    | X1   | X2   | X3   | X4   | X5   | X6   |
|----|------|------|------|------|------|------|
| Y1 | 0.0669 | 0.0056 | 0.0327 | 0.0000 | 0.1695 | 0.4441 |
| Y2 | 0.0292 | 0.0350 | 0.1086 | 0.3259 | 0.4409 | 0.5157 |
| Y3 | 0.0173 | 0.0234 | 0.3107 | 0.2460 | 0.4112 | 0.6435 |
| Y4 | 0.1082 | 0.1001 | 0.2536 | 0.2395 | 0.3996 | 0.1734 |
| Y5 | 0.0544 | 0.0762 | 0.1748 | 0.1084 | 0.2801 | 0.7759 |
| Y6 | 0.1014 | 0.0795 | 0.1422 | 0.2298 | 0.3990 | 0.4470 |
| Y7 | 0.0278 | 0.0253 | 0.2132 | 0.1429 | 0.3437 | 0.8943 |
| Y8 | 0.5342 | 0.8787 | 0.3360 | 0.2991 | 0.4346 | 0.1255 |
| Y9 | 0.3429 | 0.5362 | 0.3606 | 0.3991 | 0.4711 | 0.1058 |
| Y10| 0.0669 | 0.0056 | 0.0327 | 0.0000 | 0.1695 | 0.4441 |

Run FCM and -KFCM algorithm in MATLAB, the results are divided into three categories. The clustering center of FCM algorithm is shown in table 2. The clustering center of -KFCM algorithm is shown in table 3. The calculated objective function values are shown in table 4.

**Table 2. Clustering center of FCM algorithm**

|    | X1   | X2   | X3   | X4   | X5   | X6   |
|----|------|------|------|------|------|------|
| I category | 0.0778 | 0.0758 | 0.2623 | 0.3349 | 0.4457 | 0.3702 |
| II category | 0.0483 | 0.0372 | 0.2724 | 0.2710 | 0.4016 | 0.6936 |
| III category | 0.4864 | 0.6228 | 0.3554 | 0.3216 | 0.5179 | 0.1111 |

**Table 3. Clustering center of -KFCM algorithm**

|    | X1   | X2   | X3   | X4   | X5   | X6   |
|----|------|------|------|------|------|------|
| I category | 0.0739 | 0.0693 | 0.2901 | 0.2927 | 0.4183 | 0.5957 |
| II category | 0.1050 | 0.1073 | 0.2596 | 0.2917 | 0.4258 | 0.5101 |
| III category | 0.1369 | 0.1522 | 0.2908 | 0.3249 | 0.4468 | 0.3777 |

**Table 4. Objective function values**

| Objective function value of FCM | Iteration times | Objective function value | Iteration times | Objective function value of -KFCM | Iteration times | Objective function value |
|---------------------------------|----------------|--------------------------|----------------|-----------------------------------|----------------|--------------------------|
| The fourth quarter of 2015      |                | 3.7641                   | 1              | 0.2260                            | 2              | 0.1954                   |
| 2                              | 2.9775         | 0.1695                   | 4              | 2.5611                            | 25             | 2.5611                   |
| 3                              | 2.8894         | 2.5611                   |                |                                   |                |                          |
| 4                              | 2.8417         |                          |                |                                   |                |                          |
| ...                             |                |                          |                |                                   |                |                          |
| 23                             | 2.5612         |                          |                |                                   |                |                          |
| 24                             | 2.5611         |                          |                |                                   |                |                          |
| 25                             | 2.5611         |                          |                |                                   |                |                          |

After iteration, the minimum value of the objective function is less than the given threshold value ($10^{-6}$). At this time, stop the iteration and get the clustering result. The FCM clustering results in the fourth quarter of 2015 are shown in Figure 1, and the -KFCM clustering results are shown in...
Figure 2:

Figure 1. FCM clustering in the fourth quarter of 2015
Among them, Category I is Y8, Y9, Y13 and Y25. Category II is Y2, Y4, Y6, Y11, Y12, Y19, Y20, Y24, Y29, Y31, Y34, Y35, Y37, Y38, Y40, Y43, Y44, Y45 and Y47. Category III is the rest.

Figure 2. -KFCM clustering in the fourth quarter of 2015
Among them, Category I is Y1, Y2, y6, Y19, Y20, Y35 and Y44. Category II is Y4, Y8, Y9, Y11, Y12, Y13, Y24, Y25, Y29, Y31, Y34, Y37, Y38, Y40, Y43, Y45 and Y47. Category III is the rest.

Analysis results, according to the classification results of the fourth quarter of 2015. Based on the analysis of Category I, all four stocks have positive and high net sales interest rate, which shows that they have strong profitability, and the current ratio and cash ratio are relatively high, reflecting the current stable operation of the company, which can cope with financial expenditure. In addition, the asset liability ratio is relatively low, so the growth is relatively good. When the economic crisis occurs, the risk of the stock is very low. In a comprehensive view, the four stocks can achieve high and low risk returns by constructing a portfolio. Therefore, this kind of result is better and suitable for investors to choose investment.

Based on the analysis of Category II, the current ratio and cash ratio are relatively low, which reflects that the company is unable to cope with the financial expenditure well at present. In addition, the year-on-year growth rate of operating revenue is relatively low, which reflects that the company's long-term growth in the future is poor, the asset liability ratio is relatively high, and the risk ratio of stocks is relatively high. Compared with Category I, it is more suitable for short-term investment.

Based on the analysis of Category III, the net sales interest rate is low or negative growth, while the debt ratio is very high, which is not suitable for investors to build a portfolio for investment.

4. Conclusion
This paper studies how to use fuzzy clustering analysis to cluster stocks. Different investors have different investment styles and trends. How to cluster to determine the most suitable investment
category is the focus of many investors. Based on the study of FCM algorithm and -KFCM algorithm, this paper fully considers the impact of the contribution degree of the volatility risk of stock return to the volatility risk of the overall market return on the clustering analysis, weighted and improved it to get -KFCM model. Compared with the two clustering effects, it is considered that -KFCM model can better cluster the stocks. -KFCM model fully considers the volatility of the stock's own rate of return, takes into account the contribution of portfolio risk to the overall risk of the market and obtains the weighted algorithm. The classification conditions are better than the traditional clustering method, and the classification results include the risk of individual stocks.

The result shows that fuzzy clustering has certain practicability and reference value in stock research, and provides a reference for investors to build a portfolio.

Acknowledgments
Scientific research expenses for high-level young teachers: Theoretical and empirical Analysis of Financial risk Measurement and Derivatives risk hedge under the background of Financial Science and Technology Innovation. Northeast Petroleum University No. ky12180619.

References
[1] L.A.Zadeh. Fuzzy sets[J]. Information and Control, 1965, 8(3): 338-353.
[2] Yan Wei. FCM algorithm improvement with KPCA as the core [D]. Chengdu University of technology, 2015
[3] Yao Hongman. Fuzzy Clustering Research Based on improved artificial bee colony algorithm [D]. Guangxi University, 2016
[4] Sun Lijuan, Chen Xiaodong, Han Chong, Guo Jian. A new fuzzy clustering method for data flow [J]. Journal of electronics and information, 2015, (07): 1620-1625
[5] Deng lunbing. Value analysis of SME board stock investment based on cluster analysis and discriminant analysis.[J]. Finance and finance, 2014 (7): 134-134
[6] Zhang Wenqi. Application of cluster analysis and factor analysis in the real estate stock market municipal analysis report.[J]. Times Finance (middle), 2014 (8); 155-157
[7] Liu Yiping, Wang Qiongyao, Ji Chaoyang, et al. Application of grey clustering method in the evaluation of the investment value of gem stock. [J]. Economist, 2010 (8): 107-109