The 2022 International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/5/20220068

Effectiveness Analysis of Stock KDJ Indicator Method based on K-means Clustering

Linhui Qin[1,a,*]

[1]*Department of Accounting and Finance, Middlesex University, London, The Burroughs, Hendon, London, NW4 4BT*
a. qlhkv98@163.com
**corresponding author*

Abstract: The application of data mining technology expands various techniques in stock investment. Among them, cluster analysis is one of the common means to study stock technical indicators. There is a problem in the current cluster analysis of stock technical indicators -- the lack of validation of large-scale stock technical indicators data sets. Most of them are suitable for the comprehensive analysis of technical indicators of a single stock or multiple stocks. Aiming at this problem, this paper takes "the validity analysis of the large-scale stock KDJ index set method based on K-means clustering" as the theme. Firstly, the k-means clustering algorithm was used to construct a deep analysis model (KDJ-k-means) for the KDJ index set of the Shenzhen Index component data group. Secondly, the K, D and J index sets of 2697 constituent stocks of Shenzhen Composite Index are analyzed experimentally. Finally, the results of integrated data mining are obtained. The KDJ-k-means model is an optimization scheme based on the KDJ index set using clustering technology, which provides an intuitive and efficient visual application for deep analysis of large-scale stock data groups.

Keywords: K-means algorithm, KDJ, cluster analysis, data mining, visualization

1. Introduction

In the traditional financial era, market information was dated and deceptive [1]. With the advent of big data, such information has been integrated into the data changes of various technical indicators in the stock market. In view of the uncertainty factors and decision errors caused by complex problems in the stock market, data mining technology has always been a hot topic in the analysis of technical indicators [2]. The early technology development until now has developed a lot of sophisticated metrics and data mining methods [3]. In the research, it is found that most mining methods are only suitable for a single stock or several stocks. Large - scale stock data set lacks the validity test of technical indicators. From the perspective of data mining, this paper focuses on the effectiveness of deep mining of large-scale stock KDJ index set based on K-means clustering algorithm. The K-means clustering algorithm was used to construct a deep analysis and test model (KDJ-k-means) for K, D, J and other index sets of all components of the Shenzhen Stock Index of China. The optimal value -k of the experimental data in this study is mined through experiments. To test whether the clustering method can provide an effective visualization application for investors in decision information in the in-depth analysis of large-scale stock data. Through the technical index set of large-scale data in this study, it can be found that the characteristics of big data are manifested in its high noise, nonlinearity and multi-dimension [4][5]. The effectiveness of investors' decision-making will also be affected by these characteristics, which requires data mining technology to optimize [6][7]. In fact, the research significance of mining technology for the prediction of China's securities market is more prominent. Stock price fluctuation of non-objective factors often occurs in the Chinese stock market, which will inevitably affect the validity of technical analysis.

2. The principle and Operation of the Algorithm Model

2.1. Related Theoretical Preliminary Work

Data mining methods in stock technical analysis are mainly divided into classification, clustering, association rules and neural networks [8]. First of all, classification technology is based on classification prediction of various technical indicators, which is characterized by fast execution efficiency [6]. Secondly, the characteristic of association rules is that useful prediction rules can be mined [9]. Finally, a neural network is a large-scale, parallel, and complex nonlinear dynamical system with high prediction accuracy [10]. Compared with the above mining methods, the characteristic of clustering is that it can automatically learn data to classify data features [11]. Among them, K-means clustering has the characteristics of scalability and efficient execution efficiency, which is suitable for the application of big data mining. However, k-means clustering also has limitations that need to be optimized [11]. Most obviously, the choice of K and starting point is not easy to determine. Secondly, it is greatly influenced by the ecological niche value.

2.2. KDJ Index Depth Mining Model Based on K-means Algorithm

This chapter will explain in detail the principle of K-means algorithm, the original KDJ index, and the principle of KDJ-k-means deep mining model.

2.2.1. K-means Clustering Algorithm

K-means algorithm is a clustering algorithm that divides clusters based on the concept that large-scale data cannot be processed [12][13][14]. The number of clusters and the cluster center are given during initialization, and the cluster center is updated iteratively to achieve the optimal solution. Objective function J is often used:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{m} dist(p_{i,} c_j) \tag{1}$$

In formula (1), k represents the number of clusters initially given, m represents the number of attributes, c_j represents the clustering center of data, and p_i represents the data object of c_j. dist represents that this algorithm uses Euclidean distance to constantly iterate the distance between the remaining elements and the target center elements according to rules [15]. The definition is as follows:

$$dist_{ij} = \left(\sum_{k=1}^{p} | x_{ik} - x_{jk} |^2\right)^{\frac{1}{2}} \tag{2}$$

In formula (2), p represents the number of attributes of the data.

The decision condition of the end of iteration of k-means algorithm is mainly to determine that the similarity degree of the elements in all clusters is high enough. Given the obtained cluster set {C1, C2, ... Ck}, their center points are c_1, c_2... c_k, then the convergence discriminant E can be calculated to test whether the iteration is over. The calculation formula is as follows:

$$E = \sum_{i=1}^{k} \sum_{x \in c_i} (x - \bar{c_i})^2 \tag{3}$$

The K-means algorithm flow is as follows:

(1) Given the hypothesis cluster number K;

(2) Select K data randomly from the data sample points as the initial clustering center c_j;

(3) According to the Euclidean distance metric (this paper is based on Equation (2)), the data object (p_i) is assigned to the cluster center with the smallest distance;

(4) Recalculate the cluster center;

(5) Calculate the value of the objective function based on Equation (1);

(6) Repeat formula (3) until the number of iterations is reached or the cluster value does not change.

The setting of the initial cluster center has a great influence on the clustering results. The selection of the initial value is too large or too small, and the effective clustering results may not be obtained. In addition, K clusters and initial clustering center points should be given in advance before using K-means clustering algorithm.

2.2.2. KDJ Indicator Set

In the early days, the predecessor of the KDJ indicator KD was originally created by George Lane for use in the futures market. After KD only use the closing price calculation, lack of volatility in the stock market performance. In order to reflect a period of stock volatility. In the later period, the idea of moving average was integrated on the basis of KD, and finally KDJ technical theory was created [4]. It is an index set formed by K, D and J. It mainly uses the relationship between the high and low price and the closing price in a specific cycle, and determines the stock price trend in this cycle by calculating the real amplitude of the price fluctuations such as the highest price, the lowest price and the closing price on the same day or in recent days.

The short-and medium-term fluctuations of KDJ can be divided into different periods for analysis [4]. This study takes 10 days as a period, and the corresponding K, D and J values are calculated as follows:

$$\mathrm{RSV} = (\mathrm{C} - \mathrm{L}_{10}) \div (\mathrm{H}_{10} - \mathrm{L}_{10}) \tag{4}$$

In Formula (4), RSV stands for immature random value. C is the closing price on the 10th day, L_{10} is the lowest price and H_{10} is the highest price within the 10 days. After the RSV value is obtained, the K, J and D values can be calculated.

$$\mathrm{K} = 2/3 \times \mathrm{K_b} + 1/3 \times \mathrm{RSV_0} \tag{5}$$

$$\mathrm{D} = 2/3 \times \mathrm{D_b} + 1/3 \times \mathrm{K_0} \tag{6}$$

$$\mathrm{J} = 3 \times \mathrm{D} - 2 \times \mathrm{K} \tag{7}$$

b: the day before; 0: On the day

Investors will be affected by the sensitivity in the process of using KDJ indicators. Of course, this sensitivity also exists in other technical indicators. As investors flooded into the stock market, KDJ became more resonant. This leads to increasing sensitivity of the indicator. In the single index of K, D and J. The J value is the least reliable because it is too sensitive [4], followed by the K value and the D value, which is slightly more stable. This was demonstrated in the following experiment.

2.2.3. KDJ Deep Analysis Model Based on K-means.

In this experimental model, the three different attribute values of K, D and J are used to cluster the technical indexes of all the components of China's Shenzhen Index analyzed. Since the properties of the three attribute values are different, it is necessary to assign a smoothing factor to the attribute values. In this study, α represents the smoothing factor, and finally gives α_k, α_d, α_j. In this paper, α is obtained through experiments. Firstly, the smoothing factor is assumed to be three decimal places, and the labeled historical data are used for learning. The values of K, D and J are assigned from 0.001 to 0.008 respectively. After 500 iterations, the accuracy of each case is obtained, and the combination with the highest accuracy is selected. Finally, the smoothing factors corresponding to the values of K, D and J are 0.008,0.008 and 0.001 respectively.

In this paper, by combining k-means algorithm and KDJ technical index set, the following in-depth analysis model is designed: KDJ-k-means.

Experimental process of model KDJ-k-means:

Step 1: Calculate the K, D and J values of all the stock data sets analyzed within 10 days by formula (4) to (7), and obtain the three-dimensional database A;

Step 2: Use algorithm k-means to cluster database A and get cluster set$\{C_i\}$;

Step 3: Analyze each cluster set$\{C_i\}$, give the accuracy of each cluster, and mine the optimal K value of this period;

Step 4: Apply clustering and analysis results to test the effectiveness of the method;

3. Experimental Results and Analysis

In this chapter, a series of experiments are conducted to verify the effectiveness of the proposed model using the data of all components of China's Shenzhen Index. The main work includes: 1) Comparative analysis of KDJ-k-means model and KDJ index analysis results on the experimental set; 2) Optimization model.

3.1. Data Sample

The experimental data are from the financial terminal of Oriental Fortune, and 2697 Shenzhen A-shares are selected as the mining objects in the experiment. September 19, 2022 was used as the mining date, and the previous data was used for model training. The data on September 23, 2022 is used as the test results of the formation accuracy of the test data.

Table 1 shows the calculation results of random index KDJ value for part of the Shenzhen A-share data collected (time: September 19, 2022).

Table 1: KDJ index values of some stocks.

Some stock K, D, J index values				
Stock Code	Name of The Securities	KDJ-K	KDJ-D	KDJ-J
003027.SZ	(TXHB)	32.43345312	23.7318	49.8368
002713.SZ	(DYRS)	34.2953	38.2132	26.4595
002188.SZ	(ZTFW)	23.7338	26.3670	18.4675
002622.SZ	(RY)	67.8800	57.6221	88.3959
002336.SZ	(RRL)	67.1912	66.0090	69.5557
000995.SZ	*(HT)	15.5220	15.6359	15.2942
000759.SZ	(ZB)	83.9068	81.8567	88.0070
000616.SZ	ST(HT)	14.9912	18.9181	7.1375
002316.SZ	*ST(YL)	7.4586	17.7591	-13.1424
002717.SZ	(LN)	54.2043	50.8520	60.9089
300813.SZ	(TLSW)	12.9128	18.6827	1.3732
002514.SZ	(BXKJ)	19.0160	20.8064	15.4352
…	…	…		

3.2. Experimental Results

Table2 shows that when K=3, it can be concluded that the accuracy of KDJ clustering analysis method of 2697 stock clusters in Shenzhen index is higher than that of single index analysis

The accuracy of KDJ-k-means model is also different with different k values. Since the value of k is too small or too large, it will lose the significance of inspection, so there is an optimization problem

DOI: 10.54254/2754-1169/5/20220068

for k. Obviously, the k values for different applications need to be determined by trial and error. Figure 1 shows the clustering results and effectiveness comparison results of 2697 stocks above when k is 3,4,5,6,7,8.

Table 2: The results of index analysis alone and cluster analysis were compared.

	Accuracy Rate
K	31%
D	18%
J	43%
KDJ (clustering)	76.92%

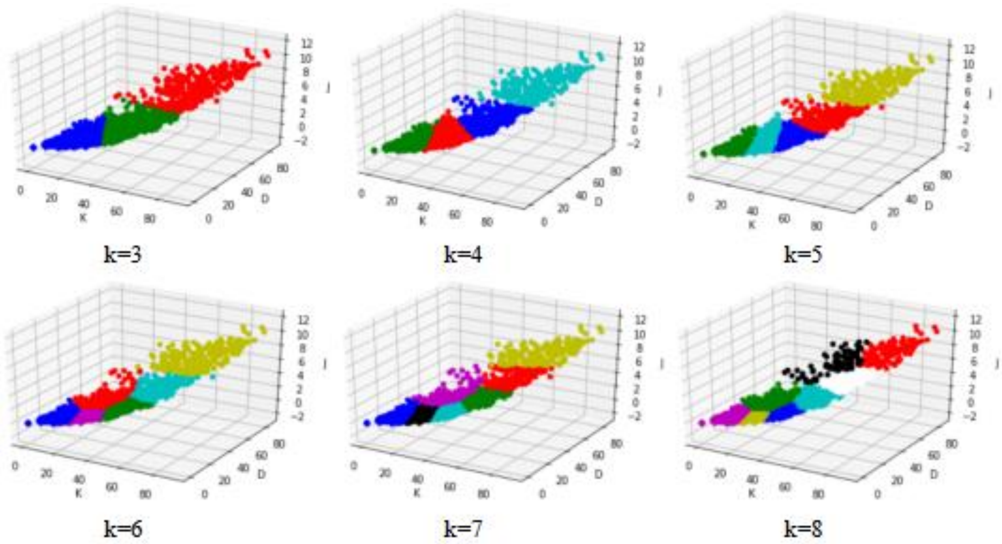

Figure 1: Results of visual cluster analysis with different k values.

KDJ-k-means model was used for the experiment. When k took different values, the prediction efficiency was compared with ordinary K, D and J values, and the results were obtained. Then experiment with different values of k to get the optimal value of k. According to Figure 2, it can be intuitively concluded that when K=7, this is the optimal scheme for the stock data observed in this study.

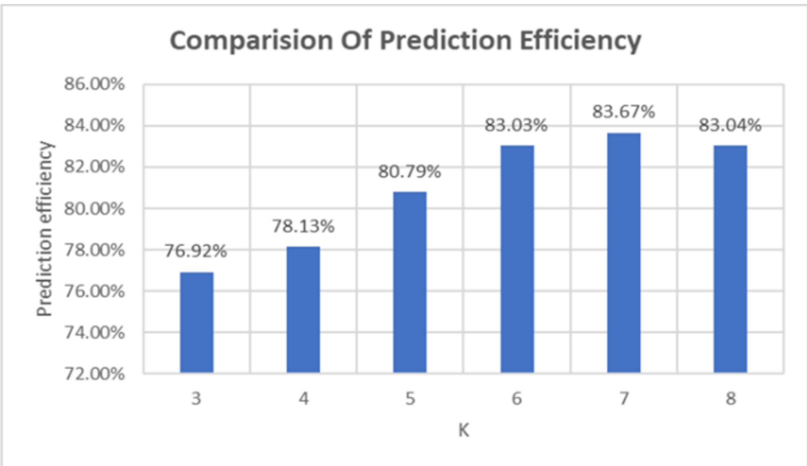

Figure 2: Effectiveness comparison of cluster analysis results with different k values in the experiment.

DOI: 10.54254/2754-1169/5/20220068

3.3. Experimental Aanalysis

The experiment is mainly analyzed according to the situation when K = 3. Investors can combine KDJ's judgment rule by clustering Figure 2.

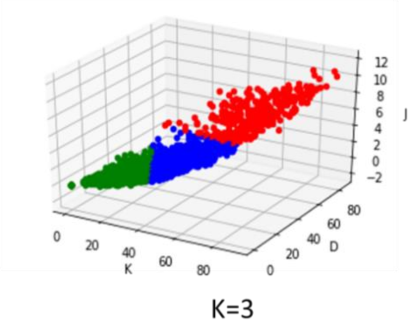

Figure 3: Results of cluster analysis of Shenzhen A-shares in China.

The green area is designated cluster 1, the blue area is designated cluster 2, and the red area is designated cluster 3. Investors can intuitively see that cluster 1 is an oversold area where K, D and J are all less than 20, which is also a short-term buy signal and will rise in the short term. Cluster 2 shows that K, D and J are greater than 20 and less than 80, which belongs to the stationary region and is also a short-term wait-and-see signal. Cluster 3 is K, D, J greater than 80, which is an overbought area, also a short-term sell signal, a short-term decline. We can apply this rule to the data in this study.

Table 3: For example verification.

Some stock K, D, J index values				
Stock Code	Name of The Securities	KDJ-K	KDJ-D	KDJ-J
003027.SZ	(TXHB)	32.43345312	23.7318	49.8368
002713.SZ	(DYRS)	34.2953	38.2132	26.4595
002188.SZ	(ZTFW)	23.7338	26.3670	18.4675
002622.SZ	(RY)	67.8800	57.6221	88.3959
002336.SZ	(RRL)	67.1912	66.0090	69.5557
000995.SZ	*(HT)	15.5220	15.6359	15.2942
000759.SZ	(ZB)	83.9068	81.8567	88.0070
000616.SZ	ST(HT)	14.9912	18.9181	7.1375
002316.SZ	*ST(YL)	7.4586	17.7591	-13.1424
002717.SZ	(LN)	54.2043	50.8520	60.9089
300813.SZ	(TLSW)	12.9128	18.6827	1.3732
002514.SZ	(BXKJ)	19.0160	20.8064	15.4352
…	…	…		

According to the partially selected KDJ index data in Table 3, the index value of stock ZB belongs to cluster 3, the index value of stock LN belongs to cluster 2, and the index value of stock TLSW belongs to cluster 1.

We examine its accuracy by comparing the rise and fall test sets of stock 23 in Table 4. In Table 4, we can find that all three stocks have reached the expected up and down state. The validity of the model is verified. Of course, there are judgment errors in KDJ mining, such as in Table 4, where the stock ZTFW falls but does not rise in cluster 1. But it does not affect the test results of the experiment.

As can be seen from Figure 1, when the value of K is close to the optimal value, the error will also decrease.

Table 4: Partial data verification results.

Partial data verification results			
Stock Code	Name of The Securities	Cluster	Up(1) and Down(-1) -Results
003027.SZ	(TXHB)	2	1
002713.SZ	(DYRS)	2	1
002188.SZ	(ZTFW)	1	-1
002622.SZ	(RY)	3	1
002336.SZ	(RRL)	2	-1
000995.SZ	*ST (HT)	1	1
000759.SZ	(ZB)	3	-1
000616.SZ	ST (HT)	1	1
002316.SZ	*ST (YL)	1	-1
002717.SZ	(LN)	2	1
300813.SZ	(TLSW)	1	1
002514.SZ	(BXKJ)	1	1
…	…	…	…

4. Conclusion

In this study, k-means clustering combined with the large-scale KDJ index set was used to conduct in-depth effectiveness analysis on stock groups, which verified that K-menas clustering could optimize big data and provide efficient visual applications for stock selection by investors. But there are also limitations. The validity of the test is limited by the same algorithm and index. In future research, the next step is to try to comprehensively analyze and optimize the high-dimensional data clustering based on time series clustering according to the limitations of K-means-KDJ through research and experiment.

References

[1] *FANG J X, LIU Y D. Public stock listing and the dependence of corporate investment on peer stock price information. Journal of Nanjing University of Finance and Economics,2020, P85-97.*

[2] *LI N, MAO G J, DENG K Comprehensive analysis method of stock KDJ index based on k-means clustering. With modern computer, 2018-21. Doi: CNKI: SUN: JYXH. 0.2018-10-005.*

[3] *CHEN W W. Data Warehouse and Data Mining Tutorial [M]. Beijing: Tsinghua University Press, 2006:83-137.*

[4] *GAO G K. Theory and Practice of Securities Investment [M]. Shanghai: Shanghai University of Finance and Economics Press, 2002:143-151.*

[5] *MAO G J. The Concept, System Structure and Method of Data Mining [J]. Computer Engineering and Design, 2002,23 (8):13-17.*

[6] *LU R R. Application of Markov Process Based on K-means Clustering in Stock price trend Prediction [D]. Wuhan: Huazhong University of Science and Technology, 2009.*

[7] *WANG L, HU Y. Stock Data Mining Based on C4.5 Decision Tree [J]. Computer and Modernization, 2015(10):21-24*

[8] *MEGIDDO N, SRIKANT R. Discovering predictive associ-ation rules [C]/ /Proceedings of the 4th International Con-ference on Knowledge Discovery and Data Mining.1999, 7(4):274-278.*

[9] *ZHENG Z X, LIU T J. Application of Association Rules in Stock Analysis [J]. Journal of Chengdu University (Natural Science Edition) 2002(4):46-49*

[10] *HUANG Q P, ZHOU X, GAN Y J, WEI Y.' Research on the Application of SVM and Neural Network Model in Stock Prediction.' Microcomputer and Application. (2015):92-94Print.*

DOI: 10.54254/2754-1169/5/20220068

[11] YI Y F. Data analysis based on K-Means clustering. Modern manufacturing technology and equipment, Doi: 10.3969/j.issn.1673-5587.2017.04.007.

[12] CHINRUNGRUENG C, SEQUIN C H. Optimal adaptive K-means algorithm with dynamic adjustment of learning rate[J]. IEEE Neural Networks,6(1)1995:157-169.

[13] DARKEN C, MOODY J.Fast adaptive k-means cluste-ring: Some empirical results[C]// International Joint Conference on Neural Networks,||,1990:233-238.

[14] JIA R Y, LI Z. Hierarchical K-means clustering algorithm based on minimum spanning tree. Microelectronics and Computers,2016, P86-88.

[15] JAIN A K, MURTY M N, FLYNN P J. Data clustering: A review [J]. ACM Computing Surveys, 1999,31(3):264-323