Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 3, Issue 4:95-101, DOI: 10.25236/AJCIS.030411

Analysis of Apple Stock based on Centroid clustering Analysis algorithm

Yu Zhao[1], Tong Bai[2]

[1]*Institute of Date Science, City University of Macao, Macao 999078, Macao, China*
[2]*Institute of Business Adminstration, Krirk University, Bangkok 10220,Tailand*

ABSTRACT. *Stock price analysis is an important part of financial assets analysis, which plays an important role in economics. By predicting the price of the stock, we can predict the trend of the stock and the general trend of the recent economic operation, which is one of the important goals of the development of econometrics. The purpose of this paper is to use the method of centroid cluster analysis to analyze the trend of stocks-- take Apple stock as an example to help investors make decisions, avoid investment risks and increase the rate of return on investment.*

KEYWORDS: *clustering analysis, Stock, var*

1. Introduction

With the continuous development of China's economy, the income of residents is also increasing accordingly. In recent years, venture capital has also become an important part of people's management of personal assets. In venture capital, because of the high rate of return on stock investment, more and more people put their assets into the stock market. In addition, the risk of stock investment is high, and this risk has become the most worried and concerned issue for investment, so people hope to be able to predict the trend and price of stocks, so that they can get a higher rate of return on investment. People hope that they can properly avoid risks and avoid major economic losses. Nowadays, with the rapid development of big data and artificial intelligence, econometrics and statistics have been further updated and developed. Through the establishment of modeling and the use of intelligent algorithms, stock trends and prices are analyzed and predicted. Greatly reduce people's investment risk. Classifying things is not only an important way for people to understand things, but also the most commonly used method. Therefore, taxonomy is a basic discipline for people to understand the world. Clustering analysis refers to the analysis process of grouping a collection of physical or abstract objects into multiple classes composed of similar objects. It is an important human behavior. The goal of cluster analysis is to collect data on the basis of similarity for classification. In different application fields, many clustering techniques have been developed. these techniques and methods are used to describe data, measure the

ISSN 2616-5775 Vol. 3, Issue 4:95-101, DOI: 10.25236/AJCIS.030411

similarity between different data sources, and classify data sources into different clusters.

K-means clustering algorithm is an iterative clustering analysis algorithm. Its step is to divide the data into K groups, then randomly select K objects as the initial clustering centers, then calculate the distance between each object and each seed clustering center, and assign each object to the nearest clustering center. Cluster centers and the objects assigned to them represent a cluster. For each sample assigned, the cluster center of the cluster will be recalculated according to the existing objects in the cluster. This process will be repeated until a termination condition is met. The termination condition can be that no (or minimum number of) objects are reassigned to different clusters, no (or minimum number of) clustering centers change again, and the sum of square errors is locally minimum.

2. Model Establishment

2.1 Direct clustering method

First regards each classified object as a class separately, and then selects a pair of classified objects in turn and forms a new class according to the principle of minimum distance. If one of the classified objects has been classified into one category, the other is also classified into that category; if a pair of classified objects happen to belong to the two categories that have already been classified, the two categories are combined into one category. With each merge, the row in which the object resides in the same column as the column order is crossed out. All the classified objects can be classified into-- classes after one mMel, so that the clustering pedigree diagram can be made according to the order of merging.

2.2 Shortest distance clustering method

The shortest distance clustering method is to find out from the non-diagonal elements of the original MIMO distance matrix, merge the classified objects Gp and Gq into a new class Gr, and then calculate the distance between the original class and the new class according to the calculation formula, so as to get a new distance matrix of order 1, and then select the smallest dij, from the new distance matrix to merge Gi and Gj into a new class. Then calculate the distance between each category and the new class, and go on like this until the classified objects are classified into class.

2.3 The furthest distance clustering method

The difference between the furthest distance clustering method and the shortest distance clustering method is that the formula used to calculate the distance between

ISSN 2616-5775 Vol. 3, Issue 4:95-101, DOI: 10.25236/AJCIS.030411

the original class and the new class is different. The furthest distance clustering method uses the furthest distance to measure the distance between samples.

3. Model Calculation

3.1 To calculate the distance between two points

The most important thing is to calculate the distance between two points. Here, the Euclidean distance is directly used as the distance function to realize the code:

```
function distEclud(vecA,vecB){
    var sum = 0;
    for(var i = 0;i < vecA.length;i++){
        var deta = vecA[i] - vecB[i];
        sum = sum + deta * deta;
    }
    return Math.sqrt(sum);
}
```

3.2 Randomly generate K centroids

```
function randCent(dataSet,k){
    //var n = dataSet[0].length;
    //var minJ = $M.min(dataSet,1);
    //var rangeJ = $V.sub($M.max(dataSet,1),minJ);
    var centroids = []
    for(var i = 0;i < k;i++){
        //centroids.push($V.add(minJ,$V.mul(rangeJ,$V.rand(n))));
        centroids.push(dataSet[i]);
    }
    return centroids;
}
```

Take the K data points of the dataset as the central data, the $M.min function calculates the minimum value of the matrix (two-dimensional array) column, the $M.max calculates the maximum value, the $V.addther calculates the addition and multiplication of the array, and $V.rand generates an n-dimensional random array.

ISSN 2616-5775 Vol. 3, Issue 4:95-101, DOI: 10.25236/AJCIS.030411

3.3 Clustering

The next step is to cluster the data set

```
function kMeans(dataSet,k){
    var m = dataSet.length;
    var clusterAssment0 = [];
    var clusterAssment1 = [];
    for(var i = 0;i < m;i++){
        clusterAssment0.push(0);
        clusterAssment1.push(1);
    }
    var centroids = randCent(dataSet,k);
    var clusterChanged = true;
    while(clusterChanged){
        clusterChanged = false;
        for(var i = 0;i < m;i++){
            var minDist = 10000000;minIndex = -1;
            for(var j = 0;j < k;j++){
                var distJI = distEclud(centroids[j],dataSet[i]);

                if(distJI < minDist){
                    minDist = distJI;
                    minIndex = j;
                }
            }
            if(clusterAssment0[i] != minIndex){
                clusterChanged = true;
            }
            clusterAssment0[i] = minIndex;
            clusterAssment1[i] = minDist * minDist;
        }
```

ISSN 2616-5775 Vol. 3, Issue 4:95-101, DOI: 10.25236/AJCIS.030411

```
            for(var i = 0;i < k;i++){
            var ptsInClust = $M.subm(dataSet,$V.where(clusterAssment0,"==",i),0);
                if(ptsInClust.length == 0){
                    continue;
                }
                centroids[i] = $M.mean(ptsInClust,1);
            }
        }
        return {
            centroids:centroids,
            cluster:clusterAssment0
        };
    }
```

The results return the clustering center and the clustering results. The above K-means clustering code has been implemented.

3.4 Stock clustering to achieve stock data

We use Apple data

```
var ev_data = daily_data_17.split("\n");
var open = [];
var high = [], low = [],close = [],volume = [],date = [];bar = [];
for(var i = 1;i < ev_data.length - 1;i++){
    var es = ev_data[i].split(" ");
    date.push(es[0]);
    open.push(es[1]);
    close.push(es[2]);
    high.push(es[3]);
    low.push(es[4]);
    volume.push(es[5]);
    bar.push([es[1],es[2],es[4],es[3],es[5]]);
```

ISSN 2616-5775 Vol. 3, Issue 4:95-101, DOI: 10.25236/AJCIS.030411

```
}
```

In order to calculate the shape of the K line, the size of the stock price cannot be used as the eigenvalue. There is a correlation between opening and closing, so we need to sort out each data. Here is how I deal with the data:

```
var data = [];
for(var i = 1;i < close.length;i++){
    var mean = (parseFloat(close[i]) + parseFloat(open[i]) + parseFloat(high[i]) + parseFloat(low[i]))/4;
    var tmp = [((parseFloat(high[i]) - mean) == 0)?1:(parseFloat(close[i]) - mean)/(parseFloat(high[i]) - mean),
                ((parseFloat(high[i]) - mean) == 0)?1:(parseFloat(low[i]) - mean)/(parseFloat(high[i]) - mean),
                ((parseFloat(high[i]) - mean) == 0)?1:(parseFloat(open[i]) - mean)/(parseFloat(high[i]) - mean),
                (volume[i] - volume[i - 1])/volume[i - 1]
            ]
    data.push(tmp);
}
var max = $M.max(data,1);
var min = $M.min(data,1);
var tz = $M.div_vector($M.sub_vector(data,min,1),$V.sub(max,min),1);
```

Input tz as a dataset into the K-means clustering model, and draw a K-means diagram and classification diagram:

```
var result = kMeans(tz,12);
```

The clustering results are as follows:

ISSN 2616-5775 Vol. 3, Issue 4:95-101, DOI: 10.25236/AJCIS.030411

4. Conclusion

In this paper, by using the closing price and opening price of the stock, as well as their difference, the information of the category to which the stock belongs is calculated by using the K-means clustering algorithm. it seems that clustering can indeed be used to classify the internal structure of the stock.

To sum up, when investors choose stocks according to their risk aversion preferences, they should strengthen the research on the basic information of stocks and long-term data analysis. As China's securities market is not perfect and the transparency of asset restructuring is not high, it cannot be ruled out that some stocks rely on financial skills of financial statements, such as asset replacement, sale of main business assets or profits transferred by major shareholders to reverse losses. The actual production and operation of these stocks has not been changed, and investors should try their best to be cautious in their operation.

An in-depth understanding of the profitability of the stocks of listed companies will help investors to narrow the scope of investment, determine the value of investment, reduce the risk of investment, and improve the accuracy of investment profitability. Investors can focus on stable and growth companies with growth and profitability advantages. At the same time, the cluster analysis model also plays a positive role in the daily supervision work. at present, the application of cluster analysis in the stock market is very effective, and the clustering method is constantly improving. in the future, we will choose a more suitable and perfect model for stock analysis, further improve and improve the application of cluster analysis model in the stock market, and stabilize the stock market. Promoting the development of real economy will become the mainstream research direction. The stability and prosperity of the stock market is not only an important guarantee for economic stability and prosperity, but also an important guarantee for the improvement of people's living standards.

References

[1] Li Qingdong. Application of Cluster Analysis in Stock Analysis [J]. Journal of Liaoning University of Petroleum and Chemical Technology, 2005,25 (3): 94-96.
[2] Li Hui. Application of Cluster Analysis in Stock Investment Analysis [J]. Shang, 2015 (27): 1999.
[3] Wu Manqi. Research on ST Stock Classification and Investment Strategy based on K-means clustering [J]. China's Urban economy, 2010 (8X): 26-26.
[4] Hong Yuehua. Research on the improvement of bee colony K-means clustering algorithm [J]. Science and Technology Bulletin, 201632 (4): 170-173.
[5] Wang Qiang. Application of Cluster Analysis Model in Stock Market [J]. Economic sector, 2016 (5): 101. Wang Junbo, Yang Yiqun, Cluster Analysis in Securities Investment. Business research. Di Mingming, Sun Deshan. Application of Cluster Analysis and support Vector Machine in Stock Research, computer Technology and Development, 2008 (7): 181-246.