International Journal of Engineering & Technology, 7 (3.34) (2018) 68-71

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Predicting the Stock Market by Using the Clustering Algorithm in Big Data

Veeramalai Sankaradass[1], T.Praveen[2], S.Padmavathy[3], R.Bharathi[4]

Professor[1], Assistant Professor[2], UG Scholar[3, 4]
Department of Computer Science and Engineering[1, 2, 3, 4]
Vel Tech High Tech Dr.RangarajanDr.Sakunthala Engineering College,
Avadi, Tamilnadu
**Corresponding author E-mail: veera2000uk@gmail.com[1] , praveen@velhightech.com [2] , padma4897@gmail.com [3],*
bharu111296@gmail.com [4]

Abstract

As one of the essential approach in record mining and pattern popularity, the Possibilistic C-Means (PCM) algorithm has been widely utilized in evaluation and understanding discovery. It is highly difficult for PCM to provide an awesome end result for clustering huge amount of data particularly for heterogeneous data due to the fact that it is designed for smally established dataset. To address this trouble, we suggest a High-Order PCM (HOPCM). It is a set of rules for massive statistics clustering. The main aim of our proposed system is to find the profit or loss for the clients share based on clustering approach for the specific tickers.

Keywords: *Clustering, Fuzzy set, Stock market, soft clustering, tickers.*

1. Introduction

Moreover, in the real-world application, transaction data are usually composed of quantitative values. Design a sophisticated data mining algorithm to deal worth different types of data turns a challenge in this research topic. Recently, fuzzy set theory is more frequently used in intelligent system, because of its simplicity and similarity to human reasoning. This theory has been successfully applied to many fields in engineering and in other technology. C-means is overlapping or soft clustering algorithm, hierarchical clustering is obvious and mixture of Gaussian is a probabilistic clustering algorithm. Fuzzy clustering is a new way called soft clustering in which each data point can belong to more than one cluster, in which the items in the same cluster are as similar as possible. Fuzzy c means is a method of clustering which allows one piece of data belongs to two or more clustering. This method is frequently used in pattern recognition. It is based on the minimization of the objective function. Clustering analysis is a primary tool for discovering previous hidden structure in the form of unordered objects, where we assume that a natural grouping exists in the data. Clustering analysis is used for classifying data to divide a given set of objects into a set of classes or cluster based on similarities. It is approach towards unsupervised learning as well as one of the major techniques in pattern recognition. The hard clustering method restricts each point of data set to exactly one cluster. These methods yield exhaustive partition of the example set into nonempty subset. Fuzzy cluster analysis, therefore allows gradual membership of data point to cluster.

2. Literature Survey

[1] In this paper, computer technology is a tremendous increase in the rise of data. There is a big problem in various fields which has been encountered is decision making for large scale data. Map Reduce concept is used in case of different parallelization techniques. Map reduce was developed for the data analysis model to retrieve the information. It has been adopted by many top IT companies like Google, Instagram, and Amazon. Recent analytical approaches are map reduce, hadoop, hive. Map reduce is the most popular paradigm for batch-style processing. It is a ubiquitous process and underlying on a distributed file system

[2] In this paper, they presented stock market trend is becoming very hot focus in our economic society. Logistic Regression is used to predict the stock market. It is fit for those dependent on variables for binominal classifications. In logistic regression, complexity is lower and it provides better accuracy in prediction. To predict stock market, we must choose the financial index. The chosen financial index must be sufficient for our model. Moreover, this model exists when a feature index variable fails and meanwhile it predicts only current monthly financial data.

[3] In this paper, they presented cluster analysis is one of the important methods to find similar and dissimilar objects within same groups in different range of clusters. Fuzzy C-means (FCM) clustering algorithm was first proposed by Dunn. It is the best and well-known method among clustering techniques. It may be inaccurate in some environments and it has a tendency to provide coincident clusters. FCM can also exhibit the robustness, analyzing was made easy. FCM clusters the data by using the data points or data centres in the group of objects. Experimental results show that it is a good aspect and parameter-free robust clustering algorithm.

[4] In this paper, one of the important tasks in text mining is the market analysis. News on the web plays an important role to predict the stock market prices. Nowadays, stock trading is increasing dramatically on the web. Moreover 90% of the newcomers withdraw their shares within a year because of the difficulty in trading. In this model, they used texts and machine learning techniques to predict the market such as ‘raise’ or ‘drop’. To predict the market price, evaluation method is used here. They have proposed several methods and strategies to predict the Nikkei Stock Average.

[5] In this paper, prediction of stock market has been attracted by businesses, economics, finance, etc. Can the stock market really predicted? Is the biggest question among the investors? Some of the recent researches suggest that it may be unpredictable but that for early indicators can extracted from social media. However, it is our goal to study how the public influences the stock market values and its related news.

[6] In this digital world, technologies provide such a big platform and development of the internet; we also face a large volume of information and data retrieved day by day from various resources which were not available to humans decades ago. Clustering algorithm is a powerful meta-learning tool to analyze the massive sets of data generated by the modern applications. The main goal of clustering is to categorize the data into clusters such that each object is grouped with respective to their metrics. These clustering techniques aim to produce a good quality for clusters. The important characteristic for big data is the volume, velocity and variety. These three V’s are the core characteristics for big data. Clustering algorithm is classified into Fuzzy C-Means, BIRCH algorithm, DENCLUE algorithm, Optimal Grid, Expectation-Maximization. In this paper, it provides a comprehensive study of the various clustering algorithms.

[7]In this paper, big data is a resource for wide range of organizations and has been used in medical, researches, business, educational institutions, etc. The important concept in big data is the map reduce framework. It was built for parallel distributed programming model to process large data set efficiently. The main advantage of using map reduce is that its scalability, data processing over multiple nodes, etc. It contains two important tasks. One is map () function and another is reduce () function. Reduce task is performed once the mapping is done.

3. Existing System

The process of partitioning a group of data points into a small number of clusters is known as a Clustering. The new way of clustering called K-Means clustering (KMC) algorithm, it analysis the Churn data in Map Reduce function. At first, we cluster the churn data using **K-means algorithm**. While using k-means, an accurate share values cannot be predicted and it finds difficult to cluster heterogeneous amount of data.

3.1. Disadvantages In Existing System

1) Churn prediction models cannot work very well
2) Time taken for the process is larger
3) We have to move to separate window to view the price for each tickers.

4. Proposed System

In our proposed system, the user will be able to take decision about buying stock shares by providing exact information of feature share values in share market. In this paper, we created a database using stack details of several companies. So using this information, the system can provide accurate share value of particular company .The algorithm used is **possibilistic c-means algorithm (PCM)**. The **High-Order** PCM algorithm (**HOPCM**) is proposed for big data clustering by optimizing the objective function in the tensor space.

4.1. Scope of Proposed System

1) Clusters a large number of heterogeneous data effectively
2) User able to know the predicted share values so, they can by shares more securely
3) Can able to view the price of each tickers by selecting them in the same window

4.2. System Architecture

The diagrammatic representation of our proposed system is as follows:

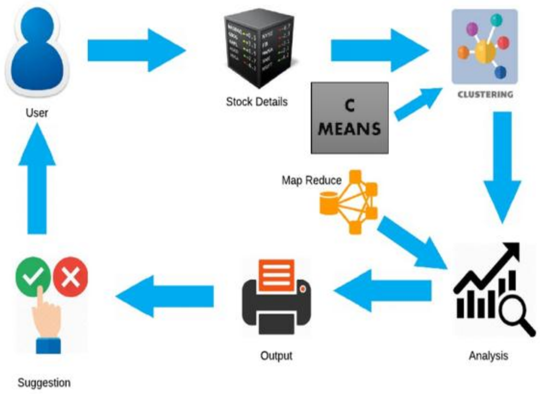


Figure (4.2): Architecture of the System

The Figure (4.2) represents thatthrough this system, the user can able to take decision about buying stock shares by providing exact information of feature share values in share market. We created a database using stack details of several companies. Using this information, the system can provide accurate share value of particular company based on clustering technique. Analysis is carried out whether the share is going to hit or flop in the market. At last, user gets clear suggestions of buying the respective shares.

4.3. Modules Description

Our proposed system is categorized into four modules. They are represened as follows:

4.3.1. Authentication And Authorization

- In this module the User have to register first, then only he/she has to access the account
- While registration the user can select the captcha and Image co-ordinates as they want
- The authorization and authentication process facilitates the system to protect itself and besides it protects the whole mechanism from unauthorized usage
- The Registration involves in getting the details of the users who wants to use this application

4.3.2. Data Clustering Using C-Means

- After Successful login, User can select the Company ticket to cluster the data
- We use C-Means Cluster algorithm to cluster the data.
- Cluster analysis groups the data objects based only on information found in the data that describes the objects and their relationships

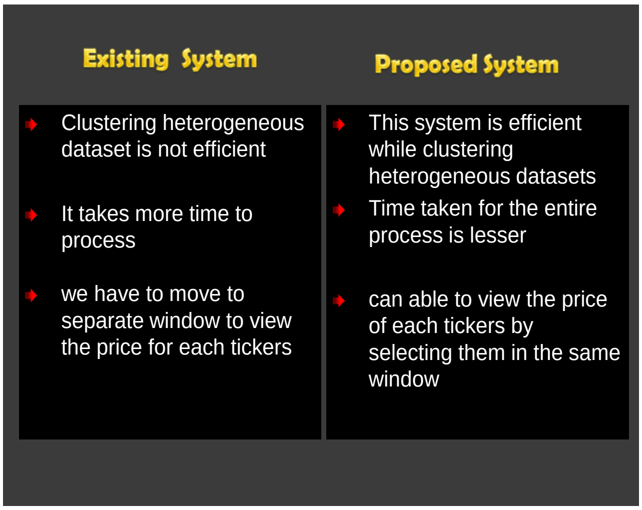
4.3.3. Analysis Using Map Reduce

- After Clustering, the next step is analysis using Map Reduce
- MapReduce is a functional programming paradigm that is well suited to handling parallel processing of huge data sets distributed across a large number of computers, or in other words
- **Map**: The map step essentially solves a small problem: It divides the problem into small workable subsets and assigns those to map processes to solve
- **Reduce**: The reducer combines the results of the mapping processes and forms the output of the MapReduce operation

4.3.4. Getting Suggestions From Admin

- After Analysis, the user will get the suggestion to reduce the churn based on the results of analysis
- The suggestions for the research results is based on the information-theoretic approaches to data analysis and inference compared to traditionally used methods
- Based on the suggestions the user can be able to reduce the Customer Churn and make a profitable business

5. Performance Analysis

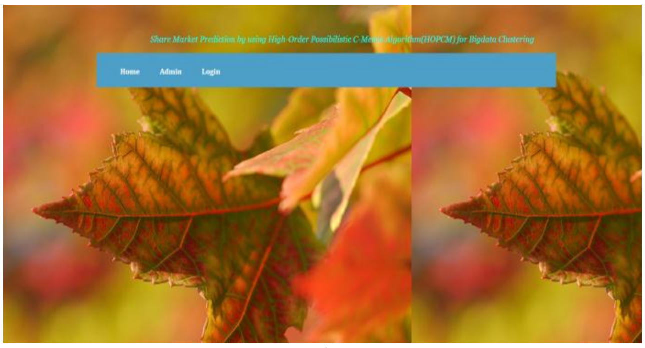


Existing System	Proposed System
Clustering heterogeneous dataset is not efficient	This system is efficient while clustering heterogeneous datasets
It takes more time to process	Time taken for the entire process is lesser
we have to move to separate window to view the price for each tickers	can able to view the price of each tickers by selecting them in the same window

Figure(5.1): Table comparison between existing and proposed system

The above table (5.1) represents the comparison between the existing and proposed system.

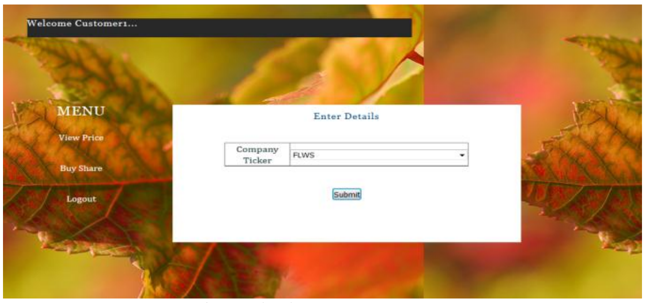
6. Experimental Results



Share Market Prediction by using High-Order Possibilistic C-Means Algorithm(HOPCM) for Bigdata Clustering

Figure (6.1): Home Page for User and Admin Login

In the above Figure (6.1),theuser and admin can register by using his name, mail id and assign a password.



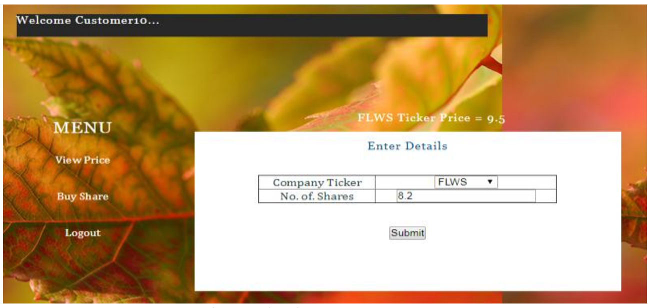
Welcome Customer1...

Enter Details

Company Ticker: FLWS

Figure (6.2): User Home Page

In the above Figure (6.2), the user can sign in by using its user name and password and can view the current market price for the particular ticker.



Welcome Customer10...

FLWS Ticker Price = 9.5

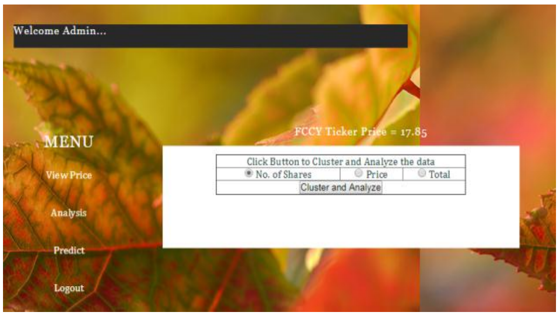
Enter Details

Company Ticker: FLWS

No. of. Shares: 8.2

Figure (6.3): User view Price

In the above Figure (6.3), the user selects the company whom they need to invest and view the current market price and buys the respective shares.



Welcome Admin...

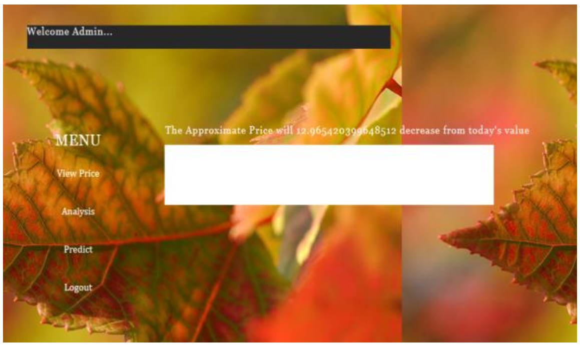
FCCY Ticker Price = 17.85

Click Button to Cluster and Analyze the data

● No. of Shares ○ Price ○ Total

Figure (6.4): Admin Cluster and Analyze

In the above Figure (6.4), the admin login page will similar as the user login in addition to the analysis and predict phase. In this, the stock market expert's will cluster and analyze the particular shares to which the investors are acquiring.



Welcome Admin...

The Approximate Price will 12.965420399648512 decrease from today's value

Figure (6.5): Admin prediction

In the above Figure (6.5), the admin will predict the share values based upon the clustering and analyzing report and finally give suggestions to the user.

7. Conclusion

This paper is employed to search out the gain or loss for the purchasers share supported agglomerative analysis technique for the actual tickers. While the existing system, the big quantity of knowledge cannot be clustered at a parallel time. It takes longer as a result of the agglomeration and analysis method is completed double for the single ticker. The ultimate result that was foretold wasn't correct. To beat this drawback, we have a tendency to project a replacement technique within which agglomeration and analysis for an oversized quantity of knowledge are often foretold at the individual method. The result obtained was conjointly accurate too. It takes only less time for its processes and analysis of these methods happens just for a single occasion.

References

[1] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding,"Data Mining with Big Data,"IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1,pp. 97-107, Jan. 2014.

[2] Q. Zhang, L. T. Yang, and Z. Chen,"Deep Computation Model forUnsupervised Feature Learning on Big Data," IEEE Transactions onServices Computing, vol. 9, no. 1, pp. 161-171, Jan. 2016.

[3] Z. Xie, S. Wang, and F. L. Chung, "An EnhancedPossibilistic c-MeansClustering Algorithm EPCM," Soft Computing, vol. 12, no. 6, pp. 593-611,2008

[4] R. Krishnapuram and J. M. Keller, "ThePossibilistic c-Means Algorithm:Insights and Recommendations," IEEE Transactions on Fuzzy Systems,vol. 4, no. 3, pp. 385-393, Aug. 1996.

[5] Q. Zhang and Z. Chen, "A Weighted Kernel Possibilistic c-Means Algorithm Based on Cloud Computing For Clustering Big Data," International Journal of Communication Systems, vol. 27, no. 9, pp. 1378-1391, 2014.

[6] Q. Zhang, L. T. Yang, and Z. Chen, "Privacy Preserving Deep Computation Model on Cloud for Big Data Feature Learning," IEEE Transactions on Computers, vol. 65, no. 5, pp. 1351-1362, May 2016.

[7] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek,"A Possibilistic Fuzzyc-Means Clustering Algorithm," IEEE Transactions on Fuzzy Systems, vol.13, no. 4, pp. 517-530, Aug. 2005.