© 2018 IJSRST | Volume 4 | Issue 9 | Print ISSN : 2395-6011 | Online ISSN : 2395-602X
Themed Section: Science and Technology

Sales Prediction : Analysis of Time Series Data Using K-Means Based Smooth Subspace Clustering

M. Jayasri*, S.Santhoshkumar
Department of Computer Science, Alagappa University, Karaikudi, Tamil Nadu, India

ABSTRACT

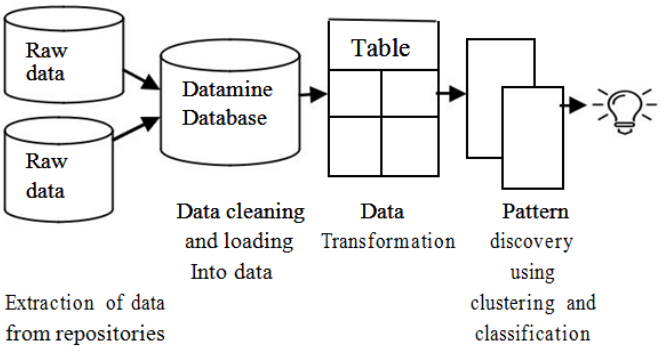
The large database to mine information that is a Data mining process and convert it into a reasonable structure for further use. Launched and order to support the organization is decision making, business planning and Data mining techniques. Data analysis, increase profitability, innovation, efficiency in resource utilization is based on important management tool in data mining. Today companies gains competitive advantage from collecting past data and using for future forecasting. Past data and information based on future estimates. In this paper, the research subject is selected as the data of a consumer electronics store company. Two year Time series sales amount data of consumer electronics was used and grouped as four quarters in a year. Next year's regression equations and naive bayes classifier methods and comprised by real sales amounts using the first quarter sales are forecasted. The real amounts and seasonal factors are really important to some product ranges that are near the sales forecasts results. In this context, various campaigns and marketing approaches have been proposed for the sales of company products by evaluating the forecast results.

Keywords: Data mining, K-means, Clustering, Time series data.

I. INTRODUCTION

The large datasets is an effort of discover patterns that is a connection of the process in data mining field (the KDD process is one of the Data mining analysis step).It use method at the relationship of statistics, database system, machine learning and artificial intelligence. The on the whole the Data mining is used to extract hidden and useful information from various databases and convert it into a reasonable structure for further use [15].The data mining tasks to verify of the preprocessing step. That is considered as the time series segmentation [12].

Regression analysis is used to the dependent variable are related to among the independent variables and the relationship between the independent and depend variables. There is different Data Mining techniques are built-up and used in Data Mining analysis work. In recent times mutually with prediction, clustering, classification, association and sequential patterns. Data mining process is to identify the data points is called prediction. It is explanation of another related data value. Generally prediction is used for the Regression analysis.



Classification based on the machine learning and that is a one of the classic Data mining technique. It is used

to classify each time in a set of data into one of predefined set of classes or groups [13]. Classification methods apply of mathematical technique such as linear programming, decision tree, neural network and statistics [13].The clustering is Different from classification [14].it is creating a meaningful and useful cluster of objects and clustering is a Data mining technique. So, in this paper proposed work a stock marketing for analyze the time series data.

II. RELATED LITERATURE

2.1 Paul Goodwin, Karima dyussekeneva and Sheik meeran "The use of analogies in forecasting the annual sales of new electronics products" 2012. One of the most popular is the bass model. The Mathematical models are often used to describe the sales and adoption patterns of products following in this year. New products are problematical because there are main time series data to evaluation the model parameters, so this model to forecast sales time series. Earlier time period to launch the sales time series of related products that is one possible solution is to fit the model and to assume that the parameter values identified for the analogy are applicable to the new product. This paper focus on the study of analogies to find the efficiency on electronic products marketing in USA for forecasting methods. It is found that all the methods have tendency to lead to forecast with high absolute percentage faults, which is consistent with other studies of new product sales forecasting. The use of published parameter values for analogies led to higher errors than the parameters we estimated from our own data [16]. When using our own data, averaging the parameter values of multiple analogies rather than relying on a single most-similar, product led to improved accuracy. However, there was little to be gained by using more than five or six analogies.

2.2. Michael Schaidnagel_, Christian Abele_, Fritz Lauxy, Ilia Petrov “Sales Prediction with Parameterized Time Series Analysis “2013.When forecasting sales figures, not only the sales history but also the future price of a product will influence the sales quantity. At first sight, multivariate time series seem to be the appropriate model for this task. Nonetheless, in real life history is not always repeatable, i.e. in the case of sales history there is only one price for a product at a given time. This complicates the design of a multivariate time series. However, for some seasonal or perishable products the price is rather a function of the expiration date than of the sales history. This additional information can help to design a more accurate and causal time series model. The proposed solution uses a univariate time series model but takes the price of a product as a parameter that influences systematically the prediction [20]. The price influence is computed based on historical sales data using correlation analysis and adjustable price ranges to identify products with comparable history. Compared to other techniques this novel approach is easy to compute and allows to preset the price parameter for predictions and simulations. Tests with data from the Data Mining Cup 2012 demonstrate better results than established complicated time series methods.

2.3 Nirav Shah, Mayank Solanki, Aditya Tambe, Dnyaneshwar Dhangar “Sales Prediction Using Effective Mining Techniques" 2015.Mining many item sets from the large transactional database is a very serious and main task. Applications requiring large amount of data processing, consists of two huge problems, one is high storage and its management and the other one is the processing time, due to increase in data. Distributed databases do the work of solving the first problem to a tremendous extent but second problem boosts. As current era is of communication and association and people are interested in storing large data on networks, and hence, researchers are introducing various algorithms to boost the throughput of output data over distributed databases. In our research, we are introducing a algorithm to

practice large amount of data at the various servers of same company that lie at different locations and collecting the practiced data on main server machine as much as admin is requiring[19]. The local copy of found data is provided to the users if he/she needs it again, this allows causing a proxy server where constantly searched items can be saved with the density of their access. This not only grants affording fast access to the data but will also afford to maintain list of recurrently accessed data. There are several approaches for accessing the data from the various servers, such as direct networked access, mobile agents, client-server techniques and LAN etc. We have used multi-threaded environment to calculate various distributed servers to gather data. For processing of data at the server end, the use of Apriori Algorithm has been done to get the outputs, which are then addressed to the client. At client data from various servers is assembled and then disciplined into data format. As an association rule mining is defined as the relation between various item sets. Association rule mining takes part in pattern discovery techniques in knowledge discovery and data mining (KDD). The frequent item sets mining, is depends of association rule mining as performance. To mine frequent item set efficiently and necessary. Thus is necessary to mine frequent item set efficiently. Association rules arrange information of this type in the form of "if-then" statements. These rules are count from the data and, inconsistent the if-then order of logic, association rules are probabilistic simultaneously, the antecedent (the "if" part) and the consequent (the "then" part), an association rule has two numbers that explicit the degree of ambiguity about the rule.

2.4 Xin Xu Lei Tang Venkat Rangan " Hitting your number or not? A Robust & Intelligent Sales Forecast System" 2017.Enterprise sales forecasting is the use of historical data to determine the future outcome (typically, sales revenue) of target month, quarter or year. Accurate forecast helps companies efficiently allocate workforce or funds, pinpoint the revenue growth bottleneck, and strategically pivot the company. Furthermore, reporting accurate projections to the board of directors or Wall Street is a key factor in stock price stability and the overall health of the business. In this work, we present one system to provide accurate, robust forecast for enterprise sales, enabling real-time insights for business decisions. Before we present the system, we will first introduce background of enterprise sales and terms commonly used in sales domain, then discuss about requirements and challenges associated with the forecast problem [18]. Enterprise selling, also known as business-to-business (B2B) selling, typically is much more complex than B2C (business-to-consumer) sales. The B2B sales normally go through a funnel, from prospect or lead, to opportunity, and then to customer. Prospects are identified through many marketing campaigns. Events like website visit, eBook download, dropping business cards at some event imply a propensity in seller's product or service. Leads are normally generated by identifying similar companies or organizations to existing clients or prospects. Similarity can be defined in a variety of aspects. Common ones are based on firm graphic information like company size, target vertical, technology stack, and so on.

2.5 Vicky Chrystian Sugiarto1, Riyanarto Sarno,

Dwi Sunaryono " Sales Forecasting Using Holt-Winters in Enterprise Resource Planning At Sales and Distribution Module " 2016.Enterprise Resource Planning is a system used for managing all of the resources owned by the company, business activities, and information used to make a good business process. One of the modules in the ERP is sales and distribution. Sales and distribution is a module that handles the sale and delivery goods to customers to achieve their business objective The use sales and distribution module is intended to simplify the process of selling to customers in accordance with the interests of customers for goods and services, makes it

easy to check the sale of goods and delivery of goods, and facilitate the collection of customers who make a purchase of goods or services provided by the company, determines the appropriate services to customers, and forecast the amount of demand for goods desired by the customer[17]. Sales forecasting is calculating the expected sales of a specific product and predicting future sales of the product [17]. It helps in making informed business decisions. In this article some several of sales forecasting methods used, Holt-Winters Additive Method, and Holt-Winters Multiplicative Method applied in ERP. The implementation of sales forecasting using Holt-Winters method in the ERP can make the system more accurate and efficient in the determination of the amount of goods demanded by customers.

III. EXISTING WORK

Classification similarity based the stock prediction problem can be mapped. A set of vectors mapped into the historical stock data and the test data. Every vector denote N dimension for each features. The parallel metric such as Euclidean distance is computed to take a choice. In this part, an explanation of KNN is provided, the highest similarity to the test into closest k records of the training data set. (I.e. query record) then a maximum vote is performed among the selected k records to verify the class label and then assigned it to the query record.

Limitations

- ✓ Need to determine value of parameter K (number of Nearest neighbors)
- ✓ Distance based learning is not understandable which type of distance to apply and which attribute to use to produce the top result.
- ✓ Computational Cost is high because we want to compute distance of every query instance to all training samples.
- ✓ Can be used for both binary and multiclass classification problems. Regression analysis,

IV. PROPOSED WORK

This paper proposed the K-means clustering algorithm with regression for handling time series resources or data efficiency. The initial step for this research work is to omit the noisy data and inconsistency of data from the Cylinder-Bell-Funnel (CBF) dataset. The next step, the input data is grouped in the form is called clustering based on same time along with the usage of K-means algorithm, the result of clustered data is a sales marketing data. The experimental result of clusters constructs. In order to invent the closest level of sales time analyzing points, used for the regression using statistical tool for regression analysis that is used to investigate the association through the variables. The analyzer seeks to taken the impact of one variable upon then another is focusing on the association ship between a dependency variable and one or more independency variables. Usage of previous sales records the K-means algorithm makes clusters and attributes such as volume of sales, price or cost of sales, profit and volume of seasonal sales volume data are invented from the sales numbers of same quantization.

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\left\| x_i - v_j \right\| \right)^2$$

Advantages

- Very simple, easy to implement and fast.
- Need less training data.
- Highly scalable. It scales linearly with the number of predictors and data points.

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\left\| x_i - v_j \right\| \right)^2$$

Advantages

1. Very simple, easy to implement and fast.
2. Need less training data.
3. Highly scalable. It scales linearly with the number of predictors and data points.
4. Can be used for both binary and multiclass classification problems.
5. Can make probabilistic predictions

V.RESULTS AND DISCUSSION

Today data's are most factors in the real world. The k means algorithm is based on data analysis the accurate data. It is used in sales marketing and stock marketing. The correct and efficient data calculate based on these figures.

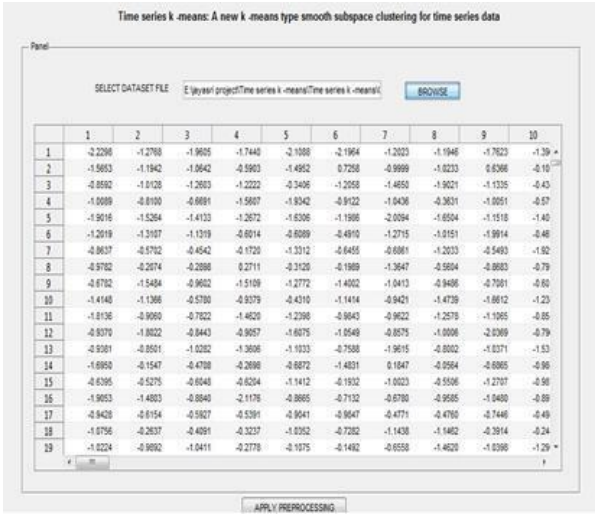


Figure 1 : Select data and Preprocessing

The Fig (1) is put to select the CBF data and it is utilize the initiate the process of the data

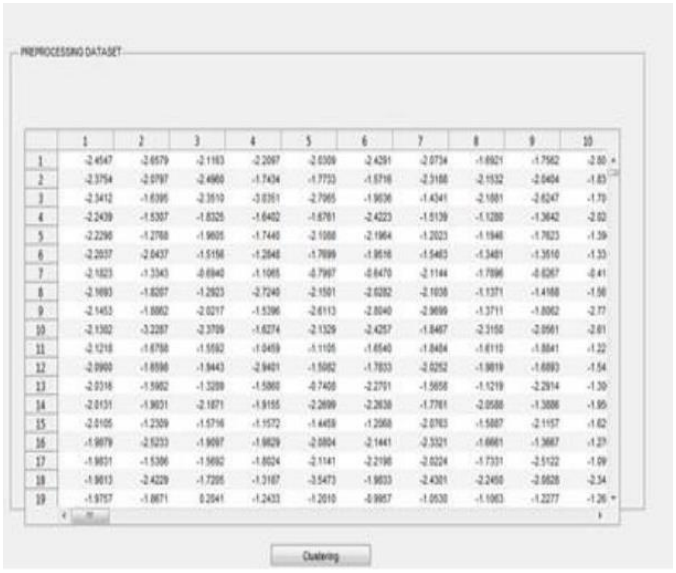


Figure 2 : Clustering the data

The fig (2) is discussed of Selected data's are applied for preprocessing.

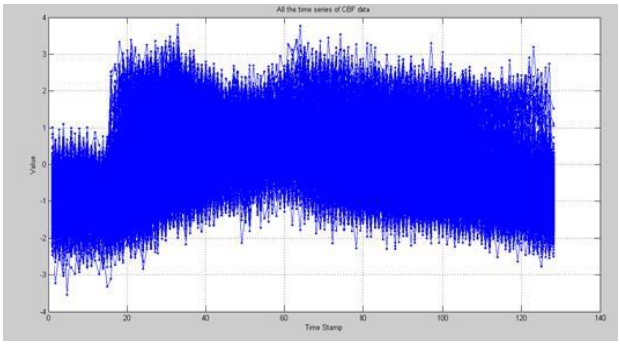


Figure 3:Timeseries of cluster1

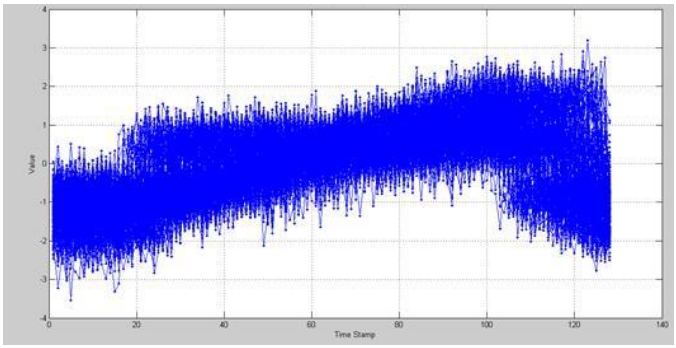


Figure 4 :Timeseries of cluster 2

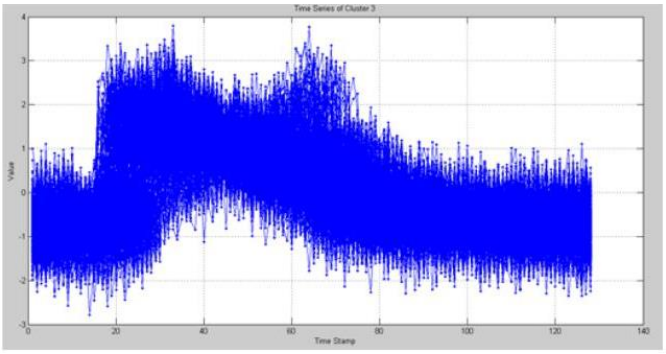


Figure 5 :Timeseries of cluster3

The cluster data's are all time series CBF file data clustering for three times above the figures 3, 4, 5. And finally figure 6 is explained and nalyzes the CBF file data. Finally to produce the cluster and analyze the CBF file data.

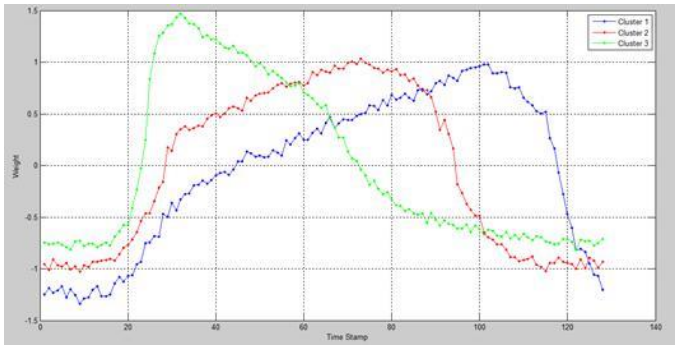


Figure 6: Timeseries of CBF Data

II. CONCLUSION

Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions. Authors are strongly encouraged not to call out multiple figures or tables in the conclusion—these should be referenced in the body of the paper.

III. REFERENCES

[1]. Karnani A. (2007), The Mirage of Marketing to the Bottom of the Pyramid, Global Competition, The William Davidson Institute at the Michigan State University, p. 90-113

[2]. Kress, G. (1985). Linguistic processes in sociocultural practice. Victoria: Deakin University Press, p. 76-77

[3]. Mentzer J.T. and Kahn K.B, (1995), Journal of Forecasting, Volume 14, Issue 5, p. 465–476

[4]. Greene K.C. and Armstrong J. S. (2007), The Ombudsman: Value of Expertise for Forecasting Decisions in Conflicts. 1–12.

[5]. Delurgio S.A. and Bhame C.D. (1991), Forecasting systems for operations management, p. 648 -649,

[6]. Freedman D.A. (2005), Statistical Models: Theory and Practice, Cambridge University Press, p. 36

[7]. Armstrong, J. S. (2012), Illusions in Regression Analysis, International Journal of Forecasting, p.689-696

[8]. Seal H. L. (1967), The historical development of the Gauss linear model, Biometrika, volume 54, p.1–24

[9]. Stuart R.,Norvig P. (2003), Artificial Intelligence: A Modern Approach, Prentice Hall, p. 126-132

[10]. Murty N., Susheela D. (2011), Pattern Recognition: An Algorithmic Approach, Springer Publishing, p. 86 -102

[11]. Bednarz T. F., (2011): Sales Forecasting: Pinpoint Sales Management Skill Development Training Series, Majorium Busine

[12]. Tai.Yu, Chiu, ting-Chie Hsu and Jia-Shung Wang (2015) ,Interpolation based consensus clustering for gene expression time series.

[13]. Paul cotofrei, kilian Stoffel (2000), Classification time=Temporal Rules.

[14]. N.Karthikeyeni Visalakshi, J.Suguna 28th North American Fuzzy Information Processing Society Annual Conference, Cincinnati, ohio,USA (June 14-17,2009).

[15]. Data mining clustering Lecturers jery stefanowski, Institute of Computer Science Poznan University of Technology Poznan, Poland (2008/2009).

[16]. Paul goodwin, Karima dyussekeneva and Sheik meeran "The use of analogies in forecasting the annual sales of new electronics products" 2012

[17]. Vicky Chrystian Sugiarto1, Riyanarto Sarno, Dwi Sunaryono " Sales Forecasting Using Holt-

Winters in Enterprise Resource Planning At Sales and Distribution Module " 2016

[18]. Xin Xu Lei Tang Venkat Rangan " Hitting your number or not? A Robust & Intelligent Sales Forecast System" 2017

[19]. Nirav Shah, Mayank Solanki, Aditya Tambe, Dnyaneshwar Dhangar "Sales Prediction Using Effective Mining Techniques" 2015.

[20]. Michael Schaidnagel_, Christian Abele_, Fritz Lauxy, Ilia Petrov "Sales Prediction with Parameterized Time Series Analysis "2013.R.M.Sahu, Akshay Godase, Pramod CONTROL ENGINEERING, Vol. 4.

[21]. Kanchan Mahajan, Proff.J.S.Chitode, "Waste Bin Monitoring