Economic Analysis on Real Time Data Fusion Using Kibana & ElasticSearch

R.Augustian Isaac [1], Srijan Mehrotra [2], Arpit Agrawal [3], Mayukh Biswas [4]
[1] Assistant Professor, Department of Computer Science, SRM Institute of Science and Technology, Chennai-600089, Tamil Nadu, India.
[2, 3, 4] Department of Computer Science, SRM Institute of Science and Technology, Chennai-600089, Tamil Nadu, India.

Abstract – In a the present scenario data has become most important aspect to take decisions for any company whether it's purely based on numbers,charts or graphs,logs etc.The big data analysis helps to find fluctuating patterns, influencing factors,risk involved but one can make much use of such data if the different types of data are fused in the database such as relational , object oriented, semantical and metadata involving their issues and enhanced capabilities.This paper proposes economic analysis by fusing of different type of data and helping analysts to take decisions for increase in business,in association with big data predicting model such as neural network which can be used for rich business experience.

Index Terms – Big Data, Data Fusion, Prediction Model, Economic and Technical data, Business and marketing.

1. INTRODUCTION

Big Data is a small word covering several aspects by one term, ranging from a technology base to a set of economic models. "Big Data" is a term encompassing the use of techniques to capture, process, analyze and visualize potentially large data sets in a reasonable time frame.[1].But to extract all this information especially in real time environment is nearly infeasible and moreover current methods and tools of handling such big data are still incapable.[2].Therefore their is a need of a platform for accessing the real time data and making quickest predicting models on the basis of data analysis.These analysis and visualisation helps us to do the prediction for building the Customer Relationship Management (CRM),marketing strategy for the product development.Predicting Modeling helps in resource management and operations of the enterprise.

Elasticsearch is a search engine based on Lucene. It provides a distributed, multitenant-capable full-text search engine with an HTTP web interface and schema-free JSON documents.It is used to do analytics and visualisation platform called Kibana. The three products are designed for use as an integrated solution, referred to as the "Elastic Stack".Different APIs provides scalable search, has near real-time search, and supports multitenancy.

Such type of analysis is extensively helpful for taking right decisions and predicting the predictions[3].The synchronisation of this type of data is helpful in making visualisations. and creating the right type of graphs.[4]We do by making small clusters of data and pulling the data sets for each particular scenario and using linear regression ,the maximum expected method.

In order to improve the understanding between product and customer, to yield the growth and interpolation for any company by using the big data and the data should give accurate and unified data so the visualisation can lead to better predictions for the CRM ,marketing campaign etc for the development of the company.

2. DATA FUSION

A. Data Fusion

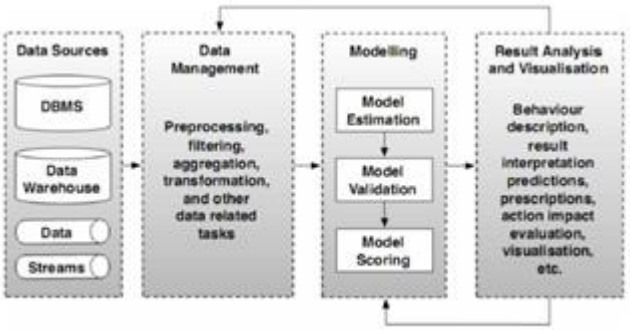
A set of techniques that integrate and analyze data from multiple sources in order to develop insights in ways that are more efficient and potentially more accurate than if they were developed by analyzing a single source of data is call data fusion.[]2To understand this first we should know traditional data and its evolution i.e Relational Database Evolution and the Relational Database and Relational Functional Data.

Generation	Time	Model	Example
First	1960-1970	File System	VMS
Second	1970s	Hierarchical	IMS
Third	Mid 1970s to present	Relational	DB2
Fourth	Mid 1980 to Present	Object Oriented Extended Relational	DFS Oracle 10
Next Gen	Present to Future	XML and Semantic	MS SQL

Table 1 Data Evolution

B. Data Fusion implementation

To fusion such big and complex data we should know the general Database Management System of big data so to know where the link has to be form to get seamless operations in order to get the accurate data[5].The table 2 is provided to get a brief overview of its working on data.



The implementation of data fusion happens by using the XML as an intermediate storage and connecting with RDB,Big Data and Semantic Web.It can be done with four ways of fusion with there respective algorithm are in [13] i.e

a) RDB to XML

b) RDF to XML

c) XML to RDB

d) XML to RDF

3. PREDICTION MODELLING ON TECHNICAL AND ECONOMIC DATA

The prediction on economic data for the marketing and business expansion is done by the interpolation method and geological collection of data for the expansion.It tells about the quarter expansion and mid reviews of the product.The target place of the economic data will be divide into two phases one on behalf of quantitative and other the casual prediction models[11].The input data is summarized by forecasting techniques.

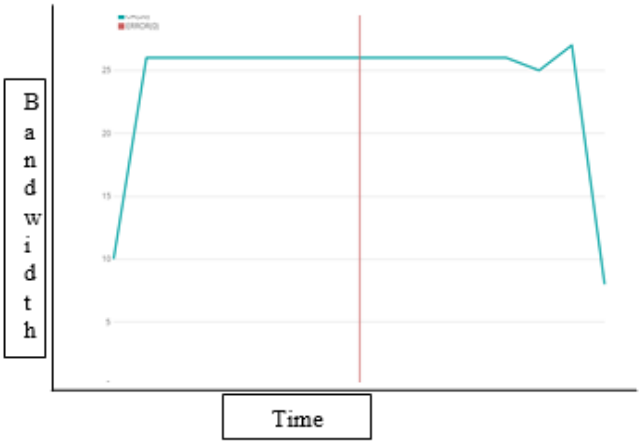
A. Prediction by neural network

We are using one method knows as forecasting technique by neural networks.Suppose for a cyclical series ,we have *x(t)* be the time series values,and let the period be fixed (assume T).So according to the time series graph the Fourier Series will be presented by giving the Inputs to the neural network.For every t,we calculate Discrete Fourier Series (DFS) of the points $x(t-T+1),\ldots,x(t)$, to obtain DFS coefficient $X_0(t)\ldots,X_{T-1}(t)$.Since DFS coefficients are complex numbers,we consider the real component $X_{nR}(t)$ and the imaginary component $X_{nI}(t)$ seperately.The network takes previous value of the DFS coefficient time series as an inputs and is trained to estimate the future coefficients $X_{nR}(t+1)$.Once all 2T forecasts from 2T neural network are available ,they are inverted to obtain the estimate for the whole i.e t-T+2 to t-1.[7,8,9]

B. The Estimation Model

- We are estimating the model by predicting the graphs as the graphs are logs of data sets ,we are using one of the software “kibana” and “elasticsearch”.
- The logs are for a daily basis and the predictions is done by t =15 minutes.Suppose the login traffic hits high at certain time so we are going to predict the rise or fall.By above formula we check on data sets such as traffic log stash API and the hits are encountered with it.[6]
- So when the coefficient of X_{nR} (t) where t=15 mins we get 2T from fourier series we can plot the graph.
- If the data is constant for first half the prediction of the model for the terminal can be found and plotting can be done in the Time Graph.[12]

C. Visualisation



Prediction for Log Stash API 1

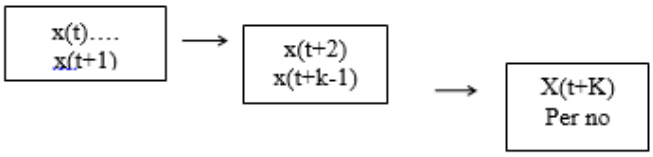
- As we can see in the above graph that after the second half there was bump and then rise in the graph.[11]
- This shows the API has a high traffic instance so we can increase our bandwith during that particular if have a neural connection.
- This can tell about the traffic and the log stash can be manged better by viewing the status and complying with network service provider.

4. PREDICTION MODELLING ON CRM AND MARKETING

A. THE PREDICTION MODEL

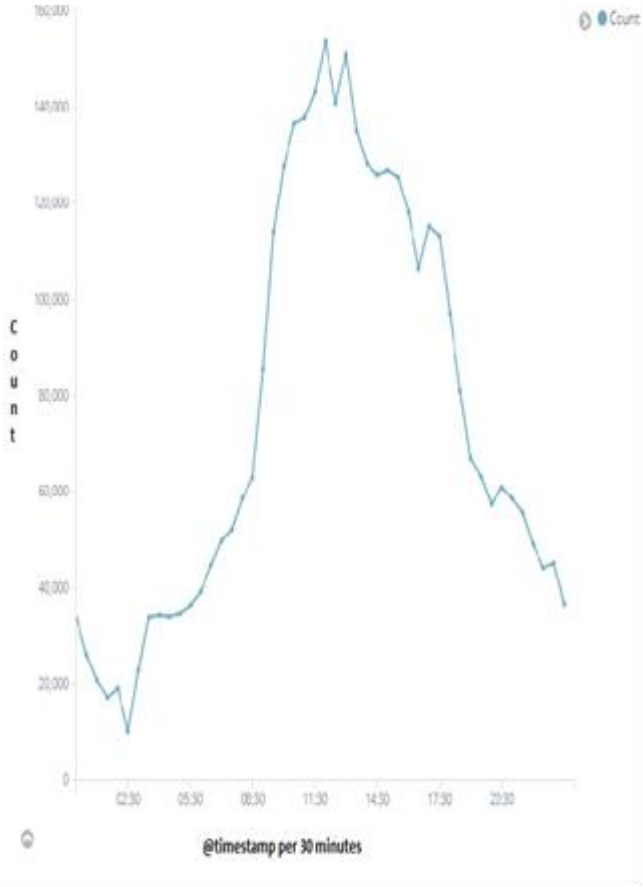
We can use the same series but if we change the approach from the linear to the the recursive approach for the forecasting.

[7,8]We can predict the CRM management.Suppose we apply this network recursively to forecast k step ahead .Thus,at any intermediate step the network will use some of the forecasts it obtained at previous steps.A model is made for understanding that we can check the relationship at each step with customer and can know about the CRM at any instance.

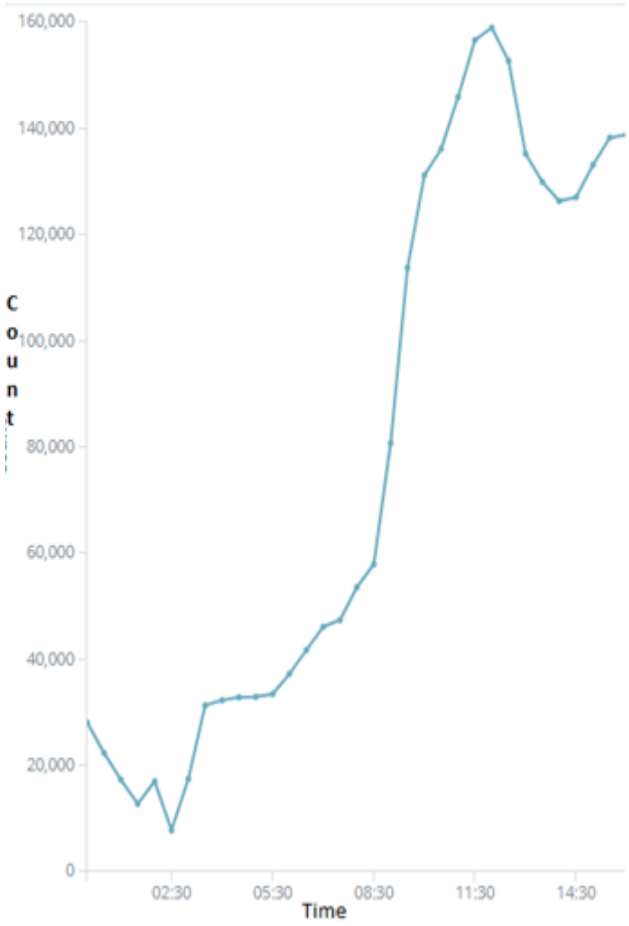


1. So by using recursive method we will be putting the prediction model with the help of elasticsearch and log stash of CRM as Casual Graph and then we actual Graph which is resulted model without predictions.[10]
2. According the predicted the graph will CRM will go high as data set of logs is in increasing with increase in time and then it will go down when the factors such as queries ,traffic ,product updates will change.

B. Visualisation



Predicted CRM 1



Actual Graph 1

5. CONCLUSION

a) Based on the economic and technical analysis of data,the prediction of the missing node can be found to increases gross profit for the enterprise.

b) Based on the graphs market response and the traffic in the API can be analysed so the the service can be improved

c) Based on the prediction neural network we can see the Predicted CRM 1 and Actual CRM are similar and this can happen due to change in customer demand , queries resolved ,product instance changed or the n factors which support in the building customer relationship with company.[14]

d) So we can say prediction models can not only increase the profit in terms of business environment but can also increase profit and reduce risk.

REFERENCES

[1] IGI Global book series Advances in Computational Intelligence and Robotics (ACIR) (ISSN: 2327-0411; eISSN: 2327-042X..)

[2] K. Höffner, "Survey on challenges of question answering in the semantic Web," Semantic Web, vol. 8, no. 6, pp. 895920, 2017.
[3] Ming Jian, Hu Nailian, "The operations plan management system of metallurgical mining enterprise based on Business Intelligence", 2nd International Conference on Information Science and Engineering,ICISE2010, Vol.I, pp.742-747, 2010.
[4] Beniger J (1986) The Control Revolution: Technological and Economic Origins of the Information Society. Boston, MA: Harvard University Press.
[5] M. Nentwig, M. Hartung, A.-C. N. Ngomo, and E. Rahm, ``A survey of current link discovery frameworks," Semantic Web, vol. 8, no. 3,Pp. 419 436, 2017
[6] J. Montgomery, M. D. Reid, and B. Drake, ``Protocols and structures for inference: A RESTful API for machine learning," in Proc. 2nd Int. Conf. Predictive APIs Apps, 2016, pp. 1-14.
[7] Y. Kamarianakis et al., "Forecasting traffic flow conditions in an urban network: comparison of multivariate and univariate approaches," Transp. Res. Record, no. 1857, pp. 74–84, 2003.
[8] Wu Qi, Yan Hongsen, "Forecasting Model of Product Sales Based on the Chaotie v-Support Vector Machine". Journal of mechanical engineering, Vol. 46, No.7, pp.102-103, April 2010
[9] Zhang Lingling,Li Guoqing, Kang Kuangsong, Li Wei, et al., "A BP neural network based method for geological missing data processing",Gold Science and Technology,Vol.23, No.5, pp.53-58, 20
[10] Elmer G (2002) Consumption in the network age: solicitation, automation, and networking. Convergence: The International Journal of Research into New Media Technologies 8(2): 86–99.
[11] Harms K (1983) Database basics: an overview. Antic 2(3): 40–44.
[12] Albers, S¨onke (2012), "Optimizable and Implementable Aggregate Response Modeling for Marketing Decision Support," International Journal of Research in Marketing, 29 (2), 111–22
[13] M. Bevilacqua, A. Tsourdos, A. Starr, and I. Durazo-Cardenas, ``Data fusion strategy for precise vehicle location for intelligent self-aware maintenance systems," in Proc. IEEE 6th Int. Conf. Intell. Syst., Modelling, Simulation (ISMS), Feb. 2015, pp. 76 81
[14] Anderson J, Lehnardt J and Slater N (2010) CouchDB: A Definitive Guide. Cambridge: O'Reilly