A Review on Big Data Analysis on Climatic Study

Yogesh Kumar Bhoi[1], Rahul Kumar Chawda[2]
[1]Student of MCA, [2]Assistant Professor
Department of Computer Science, Kalinga University, Raipur.

Abstract - Big data, the flourishing fame in recent years, exceeds the traditional data processing methods with 5 'V' characteristics. The classification of big data is the demanding challenge to be addressed among all research issues since it provides a larger business value in any analytics environment. Classification is a mechanism that labels data enabling economical and effective performance in valuable analysis. Research has indicated that the quality of the feature may cause a backlash to the classification performance. Also squeezing the classification model with entire raw features can create a bottleneck to the classification performance. Thus there is necessity for selecting appropriate features for training the classifier. This is feasible by incorporating a feature selection method with a classification model. Hence, this research work initiate a hybrid method named HCFS-Hierarchical learning for identifying appropriate feature subsets related to target class and given to classifier model to enhance the performance.

Keyword - Big Data, Climatic Study, Hierarchical learning

I. INTRODUCTION

The idiom big data has the ability to manage huge volume of information and the analytical skill overcomes the limitations in the existent data processing technologies. The frequent and expanding use of sensors, internet, heavy machines etc., in a flying ratio has made accelerated increase in data on today's digital world. Big data characteristics such as velocity and volume have made complications to the computing systems in handling the data. The data management, warehousing techniques and systems being used for analysis in the traditional days abort to analyze this variety of data. In order to overcome this complication, big data storage is handled by a distributed architecture file system.

The Hadoop is used to store and manage big data developed by Apache. As big data comprises large distributed file systems, the analysis of data is a difficult job. The IBM states that 90% of the available data in the universe has been created during the past 2 years. Daily they build 2.5 quintillion bytes of data from Climate information producing sensors, Social media posts, pictures and videos, E-business transaction, etc., All these sources produce big data. The data is all over in the pattern of text, number, images and videos. This type of data continues to grow bigger, there by organizing these data as a necessary process. The collected enormous data should produce logical use unless it would be waste of time, effort and storage. The action of grabbing or collection of big data is called data fication. Big data can be used effectively as it is data field. The organizing of data alone cannot make useful but should identify what can be performed by its use. Optimal processing power, analytical capabilities and skills are needed for squeezing essential information from big data. The big data features are shown in Figure 1.1.
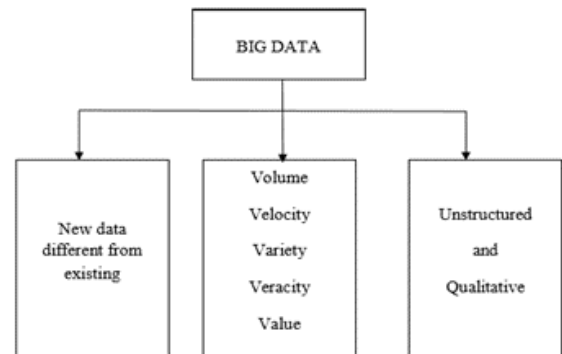

Figure 1: Big data features

Big data is of various types, formats and shapes which are brought together from various sources. Some of the common types of data and their sources .
Here are some points for thought,

- Approximately 822 tweet on twitter every one second.
- Approximately in Facebook 510 comments posted, 2, 93,000 status are updated and 1, 36,000 images uploaded every minute.
- Approximately 1 million customer transactions are made in E-business environment every hour.
- Approximately 11.5 million payments are made using Paypal for every day.

II. BIG DATA STRUCTURING

Big data can be structured by organising the existing data in a way that it should be simple to learn, analyze and obtain decisions from it. The need for structuring raises the following questions.

- By what means the huge volume of data and information are to be handled?
- Which articles among what are seen should be read?
- By what means a book should be selected in millions of available sources?
- By what means updation can be done in current events, sports, discoveries etc available in the world?

Big data classification is considered as a critical and challenging problem to be addressed in big data analytics. The action of classifying the data using issues and difficulties opened by the big data controllers is called big data classification. The different paces connected to classification are input data collection, data understanding,

data shaping and big data environment understanding. The success in big data classification requires the understanding about design and structure of algorithms. It demonstrates activities such as configuration of big data, management of big data and the methodology advancement related to classification. The cross validation and early stopping decision methods are applied for solving problems seen in the validation phase.
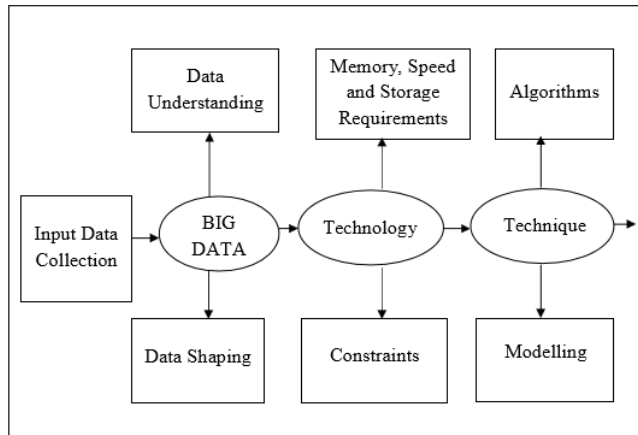

Figure 2: Big Data Classification Process

III. CONCLUSIONS

The need for privacy based deep learning algorithm for big data analytics. The idiom big data requires the ability of the developing model to deal with volume, velocity, veracity, variety and value characteristics. Keeping these characteristics in mind, the implementation of privacy based deep learning algorithm for big data analytics is performed by four works. The analytics task used in this research work is big data classification. The action of classifying data with the issues and difficulties opened up by the big data environment is critical and challenging. Analysis is carried out using four works proposed for overcoming the challenges of privacy and classification in big data environment.

IV. REFERENCES

[1]. K. Liang, W. Susilo and J.K. Liu, Privacy-preserving ciphertext Multi-Sharing Control for big data storage, IEEE Trans. Inf. Forensics Security, 10(8), (2015), 1578-1589.

[2]. D. Liu, S. Wang, D. Huang, G. Deng, F. Zeng and H. Chen, Medical image classification using spatial adjacent histogram based on adaptive local binary patterns, Comp. Bio.Med., 72, (2016), 185-200.

[3]. J. Lu, G. Wang, W. Deng, P. Moulin and J. Zhou, Multi-manifold deep metric learning for image set classification, In Proc. IEEE Conf. CVPR, (2015), 1137-1145.

[4]. J. Lu, G. Wang and P. Moulin, Localized multi feature metric learning forimage-set-based face recognition, IEEE Trans. Circ. Sys. Video Technology, 26, (2016), 529-540.

[5]. 52. R. Lu, H. Zhu, X. Liu, J. K. Liu and J. Shao, Toward efficient and privacy preserving computing in big data era, IEEE Network, 28(4), (2014), 46-50.

[6]. X. Ma, H. Wang, J. Geng and J. Wang, Hyperspectral image classification with small training set by deep network and

relative distance prior, In Proc. IEEE Symp. Geoscience and Remote Sensing, (2016), 3282-3285.

[7]. Z. Ma, F. Nie, Y. Yang, J.R.R. Uijlings and N. Sebe, Web image annotation via subspace-sparsity collaborated feature selection, Multimedia, IEEE Trans. Multimedia, 14(4), (2012), 1021–1030.

[8]. S. Maldonado, R. Weber and F. Famili, Feature selection for high dimensional class-imbalanced data sets using support vector machines, Information Sciences, 286, (2014), 228–246.

[9]. J. Martens, Deep learning via Hessian-free optimization, In Proc. Int. Conf. Mach. Learn., (2010), 735-742.

[10]. A. Mehmood, I. Natgunanathan, Y. Xiang, G. Hua and S. Guo, Protection of Big Data Privacy, IEEE Access, 4, (2016), 1821-1834.

[11]. K. Mei, J. Peng, L. Gao, N. Zheng and J. Fan, Hierarchical Classification of Large-Scale Patient Records for Automatic Treatment Stratification, IEEE Biomed. Health Inform, 19(4), (2015), 1234-1245.