Grenze International Journal of Engineering and Technology, January 2016

Big Data Mining : Weight Correction in CNN Method

Sreelekshmi A.N
Department of Computer Science, Sree Ayyappa College,Chengannur

***Abstract*—The work "Residual data based weight correction in CNN applied to big data mining" is developed on the basis of Deep Learning. The basic concept of deep learning is implemented with different algorithms. The disease prediction work utilizes deep learning by adopting the convolution neural network. The basic convolution neural network is modified by implementing an algorithm to learn the residual data in each layer to layer transformation. To overcome the difficulty of incomplete data in existing system, we are doing experiments with modified model to reconstruct the missing data. To the best of my knowledge none of the existing works focused on both data types and learning from residual data in the area of medical big data analytics.**

***Index Terms*— Big data analytics; Machine Learning; CNN; Residual data.**

I. INTRODUCTION

"Residual data based weight correction in CNN applied to big data mining" is an algorithm based web application that helps the admin,hospital and patient when proper login is provided. This work a mainly concentrates with early diseaseprediction and patients care in disease-frequent communities. The work **"**Residual data based weight correction in CNN applied to big data mining" is developed on the basis of Deep Learning. The basic concept of deep learning is implemented with different algorithms. The disease prediction work utilizes deep learning by adopting the convolution neural network. The basic convolution neural network is modified by implementing an algorithm to learn the residual data in each layer to layer transformation. To overcome the difficulty of incomplete data in existing system, here use a modified model to reconstruct the missing data.
The convolutional neural network is the idea of a "moving filter" which passes through the image. This moving filter, or convolution, applies to a certain neighbourhood of nodes.

II. RELATED WORK

The term "Big Data" has recently been applied to datasets that grow so large that they become awkward to work with using traditional database management systems. They are data sets whose size is beyond the ability of commonly used software tools and storage systems to capture, store, manage, as well as process the data within a tolerable elapsed time [1].Big data sizes are constantly increasing, currently ranging from a few dozen terabytes (TB) to many petabytes (PB) of data in a single data set. Consequently, some of the difficulties related to big data include capture, storage, search, sharing, analytics, and visualizing. Today,

Grenze ID: 01.GIJET.2.1.525

enterprises are exploring large volumes of highly detailed data so as to discover facts they didn't know before [2].

Convolutional Neural Network (CNN) is a deep learning architecture which is inspired by the structure of visual system. In 1962, Hubel and Wiesel [3] in their classic work on cat's primary visual cortex found that cells in the visual cortex are sensitive to small sub-regions of the visual field called as receptive field. These cells are responsible for detecting light in the receptive fields.Neocognitron proposed by Fukushima [4] was the first model which was simulated on a computer and was inspired from the works of Hubel and Wiesel. This network is widely considered as a predecessor of CNN and it was based on the hierarchical organization between neurons for the transformation of image. LeCun. [5, 6] established the framework of CNNs by developing a multi-layer artificial neural network called as LeNet-5. LeNet-5 was used to classify handwritten digits and could be trained with the back propagation algorithm [7] which made it possible to recognize patterns directly from raw pixels thus eliminating a separate feature extraction mechanism.

III. METHODOLOGY

The work is the effective prediction of disease by the utilization of big data features, Convolution Neural Network and enhanced Convolution Neural Network.

This is designed as a web application which can be used by hospitals as well as patients. Using this i can predict the diseases of the patients. In this web application it provides a unique Electronic Medical Record (EMR) id for all patients, which can be used in all hospital in the cluster for getting the patients whole EMR data.

Here I tend to use some data mining techniques to guess the foremost correct illness of the patients that might be related to patient's details. Based on the result, the system automatically shows the prediction.

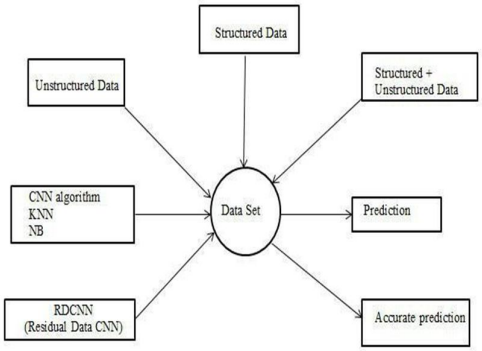

Fig.1-Proposed system

This work is developed on the basis of Deep Learning. The basic concepts of Deep Learning are implemented with different algorithms used. It utilizes deep learning by adopting the Convolution Neural Networks (CNN). The CNN is some enhanced version of neural network where layers are compared when data transfer to one layer to another. A loss function is calculated and the error is back propagated to update filter (weights). For the processing of medical text data, we utilize CNN algorithm.

The basic CNN is modified by implementing an algorithm to learn the residual data in each layer to layer transformation. A framework will be established to review the residual data and store it in an array. The back propagation will be modified to study the residual data. Major steps involved in residual learning are:

1. Residual data will be collected from existing CNN to represent, ability and optimization ability will be measured and identity will be mapped.
2. Filter weights will be modified as per the ability feedback.
3. Iteration will be continued with updated weights.

V. EXPERIMENTAL STUDY AND EVALUATION

RWU-CNN has the design enhancement in terms of residual parameters which is reduced considerably. An experiment is conducted to perform CNN based data prediction using a disease dataset. The dataset consist of 1800 instance with 24 attributes. A class instance is assigned to the type of disease. The experiment is run on a machine with core i3 based processor with 2.4GH clock speed.

The application is built with .Net framework. CNN is adopted with WEKA .Net which is a port of the WEKA data mining tool. The application has generated better predictions when an input test set is fed to the system. The input set was on online data while the training set is processed offline. The system is also evaluated with changes on residual parameters. The RWU-CNN is evaluated with its previous system based on CNN. Same dataset is fed to both systems and found considerable change in results. The number of classification results was higher in RWU-CNN than the CNN prediction. The time complexity has to be compromised in the new system because of the extra processing time needed for residual data. The prediction accuracy is also increased in the enhanced system. A graph representing the time complexity of both systems is given on figure 2. The number of instance correctly predicted in RWU-CNN is also greater than previous model. Figure 3 shows the comparison of ordinary CNN based system to that of RWU-CNN.

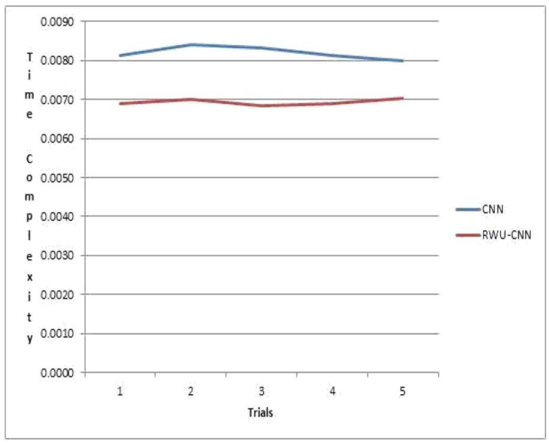

Fig:2-The time complexity of both systems

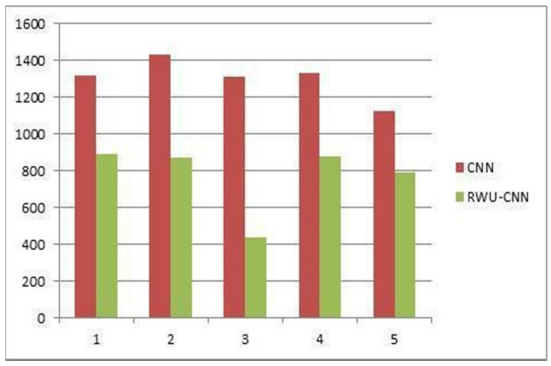

Fig:3- Comparison of ordinary CNN based system to that of RWU-CNN

VI. Conclusion

In this work, propose a new enhanced convolutional neural algorithm using data from hospital. To the best of my knowledge, none of the existing work focused on both data types in the area of medical big data analytics. Compared to several typical prediction algorithms, the prediction accuracy of proposed algorithm reaches above94.8% with a convergence speed which is faster than that of the CNN-based multimodal disease risk prediction (CNNMDRP) algorithm.

The accuracy of risk prediction depends on the diversity feature of the hospital data.This application could be enhanced with additional of more features of deep learning. With this more diseases can be predicted.

Reference

[1] Kubick, W.R.:" Big Data, Information and Meaning. In: Clinical Trial Insights", pp. 26–28(2012)
[2] Russom, P.: "Big Data Analytics. In: TDWI Best Practices Report", pp. 1–40 (2011)

[3] Hubel, David H., and Torsten N. Wiesel. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." The Journal of physiology 160.1 (1962): 106-154.
[4] Fukushima, Kunihiko. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position."Biological cybernetics 36.4 (1980): 193-202.
[5] Le Cun, B. Boser, et al. "Handwritten digit recognition with a backpropagation network." Advances in neural information processing systems. 1990.
[6] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.
[7] Hecht-Nielsen, Robert. "Theory of the backpropagation neural network."Neural Networks,IJCNN,InternationalJoin Conference on IEEE,1989.