Neural Network Optimization Revolutionizing Big Data Analytics

Aishwarya Kaul

Department of Computer Science and Engineering, Bharati Vidyapeeth's College of Engineering

Guru Gobind Singh Indraprastha University, New Delhi, India

Abstract—**Big Data is the new buzz word in the IT industry, and rightly so. All industries today are facing issues with increasing volumes of data and consequently increasing difficulty in handling and using it for benefit of the business etc. Data mining techniques such as neural networks come to the rescue. These artificial neural networks are unique in their approach, analogous to the human nervous system and information processing. These, aid in discovering hidden information that has proved to benefit the company or organisation using them in their data handling and analysis tools. This paper is a brief into the big data era, the use of data mining using neural networks and a glimpse of the real world application and its results.**

Keywords—*big data, data mining, neural networks*

I.   INTRODUCTION

Oracle believes, "Big data is the derivation of value from traditional relational database-driven business decision making, augmented with new sources of unstructured data." According to Microsoft, "Big data is the term increasingly used to describe the process of applying serious computing power—the latest in machine learning and artificial intelligence—to seriously massive and often highly complex sets of information." [1]
The major challenge of today's fast moving economy, big corporate, hospitals, organisations and institutions is the piling up data that is increasing at an exponential rate each year, nearly doubling in volume. This data is now popularly called Big Data. Various organisations may have different kinds of Big Data. For instance a Hospital will have patient documents, lab reports whereas a university will have student information, results of examinations etc. Hence volume is not the only parameter.

The 4 Vs of Big Data are:

- Volume
- Variety
- Velocity
- Veracity [2]

Numerous attempts have been made at managing this data. Data Warehousing has become an essential process for all organisations who wish to manage data effectively and perform analytics and data mining. A few of the business objectives that can be achieved with effective use of big data are:

- Market Basket Analysis
- Analysing costumer trends
- Improving company profits
- Reduce wastage of resources
- Enhance customer satisfaction
- Improve heath and critical care
- Mine previously unknown data for making strategic decisions [3]

For coping with the growing volumes of data and need to make use of it, numerous techniques have been put forward and adopted. Performing analytics on data is one of them. Various Technology giants, such as IBM, Microsoft, Dell etc, have developed analytics softwares. One such software is the IBM PureData System for Analytics. By using ERP Solutions in combination with Analytics softwares, query response times can be greatly reduced and predictive analytics can be achieved.
One such caselet is of Harvard Medical School wherein they switched to IBM's PureData System for Analytics. Seattle Children's Hospital is one of the premier paediatric care institutes in the U.S. For about a century, the Hospital has been successful in delivering superior patient care and advancing new treatments through research. For the hospital their Big Data is patient records, number of patient visits, prescriptions, bills, staff records and other medical advances that took place during patient stay. Hence, to manage this Big Data, the Hospital required analysis and maintenance in such a way that reduced the query response time for better patient care and analysis/research. They used IBM's PureData System for Analytics. Until June 2013, the Seattle Children's Hospital was storing its data on 31 SQL servers, which were progressively breaking down. Their eight-person IT team was stuck trying to maintain the large stack of ever-growing data and was unable to make any future plans for improving patient care through a well managed database. They wanted to automate the process of data categorization and integration to decrease response time and increase accessibility. In order to solve this concern, they started working with BrightLight Consulting, a business intelligence and data warehousing consultancy, to find a better system. The hospital switched to IBM's PureData System for Analytics. Using it, they got rid of failing SQL servers and modified the existing systems with IBM technology which is scalable and easily works on existing frameworks. [4][5][6] The hospital used this analysis for better understanding of the patient's history

hence improving healthcare. It eliminated manual processing/searching and speeded up query responses by 50 to 100 percent. They have more than doubled the number of projects that they can deliver to the business. Doctors get fast and accurate results of patients that helps identify trends and develop new protocols for improved patient care and better resource allocation. Queries that took few minutes to retrieve now respond in seconds. [7][8][9]

An upcoming technique for knowledge retrieval is the use of artificial neural networks. These work on concepts analogous to that of the complex human brain to achieve difficult tasks. These can be supervised or unsupervised, just as the human brain and can classify and recognize variations, in turn predict trends and enable efficient analysis. This is data mining technique and is under further development due to its tremendous scope for the future of data and processing. Data Mining is nothing but discovering patterns that are hidden in data. Pre-processed data is acted upon by a data mining tool such as one using neural networks. This gives rise to details that were previously hidden or unknown, thus increasing revenue, business, insight etc. The concept of neural networks and how they can be used for data analysis is expanded upon in further sections. [10][11][12]

II.   LITERATURE REVIEW

Big Data is a term that fascinates the world today. The impact of Big Data can be envisaged by the fact that Google alone has contributed around 54 billion dollars to the U.S. Economy in 2009, and many other such companies work entirely on processing and manipulating Big Data. [13] Not just maintaining and querying data, but also, processing and transforming this data into knowledge and useful strategic information. People and organisation across the world are digitizing, connecting and expanding their opportunities. The world is heading to a highly connected and well equipped (in terms of availability of necessary data and information) age. As a result large amounts of data is, and will be generated in the future. The healthcare industry alone generates 500PetaBytes of Big Data and by 2020 it is estimated a total of approximately 0.0281 ZetaBytes of data will be available [14]

Hence, a lot of data is available and can be used for strategic decision making, profit increments, predictive analytics and so on. But one of the major problems with Big Data apart from volume is variety and incompatibility. When the data is of different forms and is obtained from different sources, it is bound to be in different format. It is essential to normalise it first, only then can it be used for analytics and knowledge retrieval effectively. [15]

In the past few years, lesser emphasis has been given to normalisation of data. As a result, many clients of Big Data Analytics Systems have faced dissatisfaction. It is important to understand the necessity of ERP Systems. In the case of Harvard Medical School stated in the introduction, the System had been more of a success because they used an ERP System along with their own tailor made algorithms for better computation.

Few steps are being taken in this direction, such as inclusion of usage of information retrieval from data for better student performance in schools. McKinsey has estimated that U.S. can save as much as 300 billion dollars by using Analytics.

III.   REAL WORLD INSTANCE

An application of neural networks has been though the software NeuralTools, which is basically an add-on for MS Excel by Palisade Corp. An article by Miroslaw W. Janusz, Ph.D., is a software engineer for Palisade Corp. in Ithaca, New York quotes the instances of two areas where this software has been used, one from medicine and the other from healthcare industry.

Taking the instance of Katharinen hospital in Stuttgart, Germany, Dr. José R. Iglesias-Rozas, associate professor at the Universität Tübingen used the software to rate the malignancy of tumours on a scale of one to four. In the 629 cases that were available for training the neural net and the 157 cases that were used for test set, the software gave 98.58% and 95% accuracy in results, respectively.
The other instance is of tracking patient load for the hospitals of Richmond, Virginia. Taking lessons from a previous instance of bed shortage, Barbara Tawney, a healthcare industry consultant started using the neural network tool for predicting patient load. This has benefited the management of hospitals and patients to a great degree and given previously unseen data such as the trend that patients are less during the Christmas-New Year Season etc. The predictions can also be made for days or months together. [16]

IV.   CONCLUSION

Usage of neural networks in data mining has been highly successful. It has been used by the Insurance industry, banking industry and even the healthcare industry. Some of the areas were not very successful primarily due the certain external reasons or the way in which the tool was used. The customer service industry has shown a 71% accuracy and is very promising. [17] Neural networks hence have immense scope and the researchers and developers have just scratched the surface.

REFERENCES

[1]. www.technologyreview.com/view/519851/the-big-data-conundrum-how-to-define-it/
[2]. www.ibmbigdatahub.com/tag/587
[3]. Data Warehousing, Fundamental for IT Professionals, Second Edititon, Paulraj Ponniah
[4]. data-informed.com/seattle-childrens-hospital-upgrades-to-ibm-puredata-to-improve-care/
[5]. www.slideshare.net/davidpittman1/seattle-childrens-hospital-turns-big-data-into-better-care
[6]. www.youtube.com/watch?v=bjGWIectvkI
[7]. IBM Publication on "Seattle Children's Hospital"

[8]. Information Builders publication on "Netezza and Information Builders"

[9]. IBM's success in aiding healthcare-Seattle Children's Hospital-A case study, Paristhiti BVIMR, Aishwarya Kaul, Dr.Broto Bhardwaj

[10]. Applying Multiple Neural Networks on Large Scale Data,Kritsanatt Boonkiatpong, Sukree Sinthupinyo

[11]. Neural networks in data analysis, Michal Kreps

[12]. Data Warehousing, Paulraj Ponniah

[13]. Challenges and Opportunities with Big Data

[14]. www.prezi.com

[15]. Big Data Analytics Using Neural networks, Chetan Sharma

[16]. http://www.qualitydigest.com/feb06/articles/03_article.shtml

[17]. Data mining and Neural Networks from a Commercial Perspective Portia A. Cerny