ISSN (e): 2250 – 3005 || Volume, 08 || Issue, 2|| February – 2018 ||
International Journal of Computational Engineering Research (IJCER)

Implementation of Cognition in Big Data Using Bayesian Statistics

Dr. Vikas Singh[1], Ruchi Jain[2]
[1]Head, Dept. of Computer Science, Dean, Faculty of Computronics, IGNTU, Amarkantak (M.P.)
[12]Research Scholar IGNTU, Amarkantak (M.P.)
Corresponding author: Dr. Vikas Singh

ABSTRACT

Abstract:
The plethora of data which is termed as big data introduces certain opportunities and challenges in the field of cognitive sciences. Bayesian statistics may play a vital role to mine the information from the volume, velocity and variety of the Big data. The Bayesian statistics is beneficial for big learning with the recent development of cognitive phenomena. This paper gives an idea of Bayesian inference from Big data and Cognitive Science perspectives. The advantages of Bayesian inference for the assessment of Big Data have been discussed. The probabilistic approach of Bayesian statistics from cognitive aspect can be applied at various levels. The Bayes Theorem in terms of cognition can be expressed and explained.
Keywords: Bayes Theorm, Big Data, Cognitive Science

Date of Submission: 08-03-2018 Date of acceptance: 24-03-2018

I. INTRODUCTION

Cognition can be elucidated as the usage of the mental skills and activities implemented for the accomplishment of certain tasks like understanding, reasoning, learning, recalling, reminding and many more. Some abilities of cognition even includes paying attention, remembering, processing information, solution to the problems, organization and reorganization of information, communication and act upon those information [1]. Cognitive computing is the way to form the self-learning systems which incorporates the concept of pattern recognition, data mining and natural language processing (NLP) to imitate the working of the human brain. The basic motive of cognitive computing is to develop the systems to solve the complex problems which does not want any continuous human assistance.

In current scenario, the enormous data have been flooded from various sources and in many formats is termed as big data which is not manageable. To make the data more relevant and useful, it must be continuously interpreted and ingested to be adapted. This big data cannot be accommodated in the main memory of a single machine. For the said purpose, the MapReduce framework implements the concept of Distributed File System (DFS) which segregates the data for multiple machine. The two functions are involved for the storage of the data and they are Map and Reduce. The Map function is meant to split the input data into independent blocks of data which is processed in entirely parallel manner. The next phase consolidates the relevant records obtained from the previous phase which is aggregated by the reducing phase[3].

This paper contains section1 which gives introduction of the topic, section 2 gives the literature review, section 3 explains the fundamental of Bayes Theorem in a Big data perspectives, section 4 explains the cognition in Bayes Theorem. The last section concludes the paper.

II. LITERATURE REVIEW

According to Zeta *et. al.* proposes the management of 3 V's of Big data which are Volume, Velocity and Veracity of the data creates lots of opportunities and challenges for the various new emerging technologies like Artificial Intelligence, Machine Learning and Cognitive Science. In order to mine these elementary properties of Big data, the sequential Bayesian model have been proved an effective tool. The voluminous and large variety of data requires the complex models for the synthesis process. Cognitive Science have underlying mechanisms which can be applied to meaningful terms for the purpose of extraction of important aspects from the raw behavioral data. [12]

According to Nick Chater *et. al.* states that Bayesian methods contributes a probable associations between higher level and lower level cognition. They even proposed that degree of acceptance evaluated by the Bayesian Methods is inversely proportional to level of cognitive phenomenon that is the acceptance is more at the perceptual level and acceptance declines as the cognition moves higher such as reasoning. Bayesian approach can be applied at different levels of cognition like Perception, Categorization, Learning and Causality, Language Processing, Inductive Reasoning, Deductive Reasoning and Argumentation. [9]

The scientists of Tsinghua University located in China have published an overview on the recent advances in Bayesian methods for Big Data Analytics in the Beijing which was based on National Science Review whereas the co-authors Wenbo-Hu, Jianfei Chen, Bo Zhang and Jun Zhu proposed an overview on the recent development of flexible Bayesian methods, scalable and effective algorithms and the implementations of distributed systems besides the elementary concept of Bayesian methods.

According to these scientists " Bayesian methods are gradually getting pertinent to the era of Big data to secure the highly complex models from the overfitting and to empower the model to acculturate refurbish their capacity. Although, Bayesian model undergoes the computational bottleneck which needs to be designated with approximate inference methods while applying to the problems of Big data."

One more extension of Bayesian model came into existence to make the data manageable especially the voluminous data. From the Big Data perspective, prior distribution plays a vital role as it classifies the large dataset into smaller chunks. The prior distribution becomes more precise as the availability of information increases which results in noiseless parameter estimation. The likelihood demonstrates the mechanism of generation of assumed data. The posterior distribution can be considered as conditional distribution of the observed data or the given data. The incorporation of the posterior regularization imposes the linear constraints on the posterior distributions which optimizes the prior distribution.[4] This can be illustrated as follows:

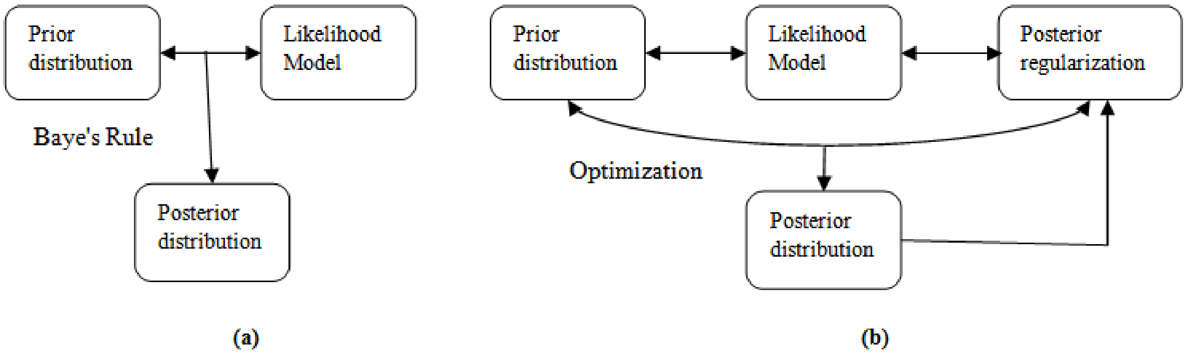

Figure 1

Figure 1 shows the Illustration of Bayesian Inference with Baye's rule whereas (b) demonstrates the Regularized Bayesian inference under an optimization framework.

The posterior regularization can be defined as the informative constraints which imposes regulation to the basic characteristics of the posterior distribution. This regularization contributes to cognition.

Baye's Theorem:

Baye's theorem states is based on the intervention of learning and experience. In order to predict, the certain sequence of steps is followed. It helps to evaluate the probability based on the prior knowledge of events occurred that would have led to the resultant event. The formula can be expressed as:[5]

$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$

Baye's Theorem in Big Data:

Baye's is considered as an effective machine learning methodology which is meant to categorize and classify data as data streams enters. This model is intended to predict the future. The data on which operation has to be performed for predicting or to query must be a trained data and can be termed as features. The resultant classification of Baye's can be specified as categories. Thus, features ascertains the categories of data in the past.[7]

More data leads to more accuracy. So it may be inferred as that the accuracy of the prediction is directly proportional to the amount of the data used or to the trained data. The prediction ascribes the probability with it. Scepticism is associated with it. Besides, being very slow it is very effective.

This model works in two modes training or prediction. When the model is meant to be taught with historical data, it is in training mode else when we provide features or the category is fed then it is in prediction mode.

Mathematically, the Baye's Theorem in a Big Data environment can be expressed as follows:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} \quad ....Eq(1)$$

where P(H|D) - The probability of our hypothesis being true given the data collected. (Posterior)
P(D|H) - The probability of collecting this data when our hypothesis is true. (Likelihood)
P(H) - The probability of the hypothesis being true before collecting data. (Prior)
P(D) - The probability of collecting this data under all hypothesis. (Marginal)
The advantages of Baye`s Theorem are as follows:[4]

i. Unpredictability: The problems in the environment posses lots of vagueness, contradictions, physical randomness and incomplete knowledge. To solve certain kind of problems, the principled theory of Baye's has been proved very beneficial for deriving certain effective inference of unrevealed factors and predictions from the prior knowledge and ambivalent evidence.

ii. Flexibility: This model has been proved very simple and easily implementable to various others problems. One more category of Bayes Theorem have been characterized for evaluating uncertainty, latent structure, missing values etc. A posterior regularization term has been introduced to ameliorate the regularized Bayesian inference to enhance the learning objective or to assimilate domain knowledge. Ultimately, Markov Chain Monte Carlo has been instigated which is very flexible in nature and meant to perform posterior inference.

iii. Adaptability: This model is capable to handle the dynamic and unpredictable environment of big data. The data is being collected continuously so as to deal with this dynamically changing scenario, Non parametric Bayesian method have been proved effective. In order to handle the big data streams, different variants and updating rule of Baye's have been introduced.

iv. Overfitting: Currently, the amount of information produced enormously high and the powerful computers which has incorporated deep network with high number of parameters had anchored it. So as a result, the problem of overfitting exists because the ability of computer is extensively higher than the information originated.

The amalgamation of Baye's with the concept of Markov Chain Monte Carlo (MCMC) model generates the miraculous results for the purpose of data analysis. MCMC model is meant to sample the data from the whole of distribution whereas the Baye's model has to interpret the observed data. It is a general purpose model which connote deduced values of latent variables, determines learning model structure and parameters, capable of handling wide range of probabilistic model. This model basically relies on the distribution proposed provided by the user. The sample values with high probability are more likely to be used as distribution for different computations. As the dimensions of the big data is very high so the sampling becomes crucial at this stage[8].

Cognition in Baye's theorem

The every event in this world is very uncertain. Thus, cognition helps to develop the process of characterizing the information and the ambiguity in that information. One of the important models available to deal with this problem is probabilistic theory. The probabilistic approach can be acquired at three different levels: Computational Level, Algorithmic Level, Implementational Level. Computational level focuses on identification of the nature of the problem faced by the human brain and the inherent characteristics of the problem like the desired result and the environment in which problem persists. Algorithmic level constitutes the representation of the specification and computational operations considering set of certain heuristics to solve the problem. Implementational level projects that the human brain is probabilistic in nature at every stage distinct neurons imparts probabilistic information, neural populations may encapsulate the probabilistic distribution and neural processes may exhibit fundamental probabilistic inference.[9]

To understand the concept of cognition, the Bayesian model have played a prominent role. It addressed several aspects like human inductive generalization and learning, semantic memory, motor control, visual scene perception, casual learning inference, language processing, language acquisition, semantic memory, social cognition and symbolic reasoning. It generally provides the framework to solve the problems of induction. Bayesian model can facilitate the human cognition in many ways: firstly, Bayesian principles mandates the process of updating the beliefs of rational agents with the help of prior knowledge and the data pertaining the set of assumptions concerned with the problem considered. Secondly, the contribution of fields like artificial intelligence, statistics and machine learning acquiring complex probabilistic model to the principles of learning and inference. Finally, in hierarchal probabilistic model, Bayesian inference may help to deduce the abstract prior knowledge from the new data and can be utilized for manage the learning of new environment and subsequent tasks.[10]

The above equation1 can be interpreted in terms of human cognition as,

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Where H is the state of the world
D is the sensory input

This equation can be explained as where H is the estimation of the system like possibility of placement of the limbs, likelihood of bluffing of opponent etc whereas D can be defined as the input to the system by the means of sensor. Thus P(H|D) is the posterior which can be expressed as the probability as the estimation of various states provided with certain sensory inputs. P(H) can be explained as the presumption or also subjected to the prior belief before receiving the sensory input. This belief is generally gained by the experience from the world. Now P(D|H) is the assessment of the probability of sensory input when hypothesized state is known that is to predict the input received like whether the input is polar bear or snow hill. The product of prior belief with the likelihood can be normalized with sensory input results in the posterior of the state which is meant to evaluates the probability of the each state when the sensory input is known. As this is an iterative process, the new estimated probability can further be used for evaluating next probability with new sensory inputs. [11]

III. CONCLUSION

Bayes Theorem have been proved very effective in implementing classification of data specifically when the lot of data has been generated i.e. Big Data. It is meant to make future investigations with the consideration of certain aspects like for the optimum learning, prior knowledge must be involved when the certainty is inadequate. One more consideration must be met that the domain knowledge must be gathered effectively to formulate the appropriate model and inference processes. Another aspect is that the Bayesian model is one of the most flexible models and capable of inferring the complexity automatically. Also , Statistical modelling anticipated by Bayes ameliorates to interpret results in user friendly way. Even, Bayesian model incorporates the probabilistic programming which can be effortlessly implemented by general purpose machines.

Bayesian model is quite flexible in nature as it can be extensive wide in terms of usage and representations. In various categories, Baye's is capable of formulating set of presumptions for the representation of prior beliefs, observations and desired results for the tasks. Cognitive scientists have adopted the Bayesian model for inferring the rational behavior of the problem. It is even beneficial for the intricate cognition. The three operations are common in Bayesian models and human cognition like parameter learning, structure learning and inference. Different forms of Baye's model have been used for different types of assumptions and observation of the certain problems.

As Cognition can be beneficial to manage the complications of Big Data. So both these technologies which incorporates Bayesian modelling. may be effective for considering the probabilistic approach of cognition to handle the Big data.

REFERENCES:

[1]. Medalia Alice,"Dealing with Cognitive Dysfunction Associated with psychiatric disabilities", A handbook for families and friends of individuals with psychiatric disorders

[2]. Chen Ying, Argentinis Eleene, Weber Griff, "IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research", Volume 38, Issue 4, April 2016

[3]. Shim Kyuseok," MapReduce Algorithms for Big Data Analysis", Springer Verlag-Heidelberg 2013, pp 44-48 2013.

[4]. Zhu Jun, Chen Jianfei, Hu Wenbo, Zhang Bo,"Big Learning with Bayesian Methods", arXiv:1411.6370v2 [cs.LG] 1 Mar 2017.

[5]. http://www.statisticshowto.com/bayes-theorem-problems

[6]. Gupta Dishashree," Introduction To Conditional Probability And Bayes Theorem For Data Science Professionals", Analytics Vidhya, March 14, 2017.

[7]. Catherine Bernardone, "Applying Bayes Theorem to a Big Data World", April 11, 2017

[8]. Sharma Sanjib," Markov Chain Monte Carlo Methods for Bayesian Data Analysis in Astronomy", arXiv:1706.01629v1 6 Jun 2017.

[9]. Chater Nick, Oaksford Mike, Hahn Ulrike and Heit Evan, "Bayesian Models of Cognition", Volume 1, November/December 2010 , John Wiley & Sons, Ltd.

[10]. Griffiths Thomas L., Kemp Charles and Tenenbaum Joshua B.,"Bayesian models of cognition", repository.cmu.edu , Dietrich College

[11]. Wolpert Daniel M. and Ghahramani Zoubin, " Bayes rule in perception, action and cognition", citeseerx.ist.psu.edu

[12]. Oravecz Zita, Huentelman Matt, Vandekerckhove Joachim, " Sequential Bayesian updating for Big Data", Chapter 2 (pp. 13–33) of Big Data in Cognitive Science.

Dr. Vikas Singh "Implementation of Cognition in Big Data Using Bayesian Statistics." International Journal of Computational Engineering Research (IJCER), vol. 08, no. 02, 2018, pp. 46-49.