iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: XI Month of publication: November 2023

DOI: https://doi.org/10.22214/ijraset.2023.56547

Bayesian Neural Networks for Enhanced Predictive Performance in Regression Problems

Manaswi Patil[1], Tanmay Zade[2], Sarthak Wankhade[3], Samruddhi Shirbhate[4]
Dept. of Computer Science & Engineering, Prof. Ram Meghe Institute of Technology & Research, Amravati, Maharashtra, India

Abstract: Bayesian statistics has a lot of influence on neural networks and deep learning for artificial intelligence (AI). The inference and learning of Bayesian statistics is based on prior, likelihood and posterior. The prior is the current belief of data field and the posterior is the updated belief after learning from observed data. By repeated learning using prior and posterior distributions, Bayesian statistics provides advanced data learning for AI. In this paper, we compare the previous Bayesian inference and learning methods for AI and propose a model based on Bayesian inference and learning for neural networks and deep learning.
Keywords: Bayesian statistics, Deep learning, Neural networks, Artificial intelligence.

I. INTRODUCTION

Bayesian and frequentist are two main approaches in statistics [1,2]. The difference between two approaches is the use of prior information about the domain to which the data belongs [3]. Bayesian statistics uses the prior distribution as probability distribution of model parameter. In addition, the prior distribution combines with the likelihood function representing observed data to produce posterior distribution. When new data is added, the current posterior distribution is used as the prior distribution and updated as new posterior distribution by combining with the likelihood function based on the new data. The updating procedure of this posterior distribution enables the model to be learned from new data. This procedure improves the intelligence of the model for data domain. So, Bayesian learning provides an efficient model for artificial intelligence (AI). In previous studies, the Bayesian AI was dependent on the Bayesian networks [4]. The Bayesian networks model is a graphical model representing the relations between random variables under uncertainty. This model consists of nodes (random variables) and arcs. The node and arc represent variable and connection between variables respectively. The connections show the causal relations between the nodes. The strength of the connection is a probability representing the belief connection. This belief is also updated by new observed data. So, this model is a popular approach to AI. In this paper, we study on a method for AI using Bayesian inference and learning.

We build the posterior distribution by combining the likelihood function with new probability distribution with prior information for Bayesian learning model. To illustrate the validity of our proposed approach, we present experimental results using simulation and existing data.

II. BAYESIAN INFERENCE AND LEARNING

A. Bayesian Inference

The Bayesian inference is based on Bayes' theorem as follow [5].

$$P(A|B) = P(A)P(B|A)/P(B) \tag{1}$$

This is the conditional probability of A, given B. Using Bayes rule, we can inference about uncertainty from data. In general, A and B represent parameter and data respectively. In Bayesian inference, P(A) is the prior belief of A and P(A|B) is the posterior belief of A after considering B. P(B|A) is the likelihood function of B for A, and P(B) represents the sum of the likelihood of B over all values of A as follow [5],

$$P(B) = \sum_A P(A)P(B|A) \tag{2}$$

So, the probability of P(A) is updated to P(A|B) by the likelihood of B. This approach to update of A is linked to Bayesian learning described in the next section.

B. Bayesian Learning

From the equation (1), we can derive the following proportional expression,

$$P(A|B) \propto P(A)P(B|A)\ ) \qquad (3)$$

Where P(A|B), P(A) and P(B|A) represent posterior, prior and likelihood respectively. That is, the posterior distribution is updated with the product of prior distribution and the likelihood function with observed data each time. This posterior update process is Bayesian learning [6]. We get an updated probability distribution of parameter from the result of Bayesian learning [1,6]. In the next section, we propose a new method for Bayesian inference and learning models for neural networks and deep learning.

III. BAYESIAN INFERENCE AND LEARNING FOR NEURAL NETWORKS

In this paper, we study on the Bayesian inference and learning to improve the performance of neural networks and deep learning models. We also compare the performance between various Bayesian neural networks and deep learning models. The final goal of Bayesian learning is the updated probability distribution for model parameters (weights). So, we first denote the prior distribution of model parameter as P(θ). This is the belief about the parameter. Next, we get the observed data (x1, x2, ..., xn), and represent the data as likelihood function P(x1, x2, ..., xn|θ). Using this likelihood of parameters in model, we update the prior distribution of θ to the posterior distribution as follow,

$$P(\theta|X) \propto P(\theta)P(X|\theta) \qquad (4)$$

Where X is the data set (x1, x2, ..., xn), and P(θ|X) is the posterior distribution of θ. When new data is observed, the previous posterior distribution becomes the current prior distribution. This is combined with the likelihood function of the observed data to be the new updated posterior distribution. Form the equation (4), we derive the predictive distribution for new data x(new) given X as follow [6],

$$P\ (X\ |\ X) = \int\ P(X\ |\theta\ )\ P\ (\theta\ |\ X)\ d\theta) \qquad (5)$$

We apply the Bayesian inference and learning with equations (4) and (5) to neural networks and deep learning models. Bayesian neural networks has input, hidden and output layers like traditional neural networks. But, each weights (parameter) of the neural networks is considered as random variable with probability distributions. In this paper, we consider Bayesian neural networks as follow [6].

$$y = f\ (b_2 + \sum \nu\ (b_1 + \sum u\ x)) \qquad (6)$$

Where x and y are input and output variables. b1 and b2 are the biases for hidden and output respectively, and f() is one of activation functions such as sigmoid or hyperbolic tangent. Also, u is the weight vector from input to hidden, and v is also weight vector from hidden to output. We present θ as the network parameters with biases and weights. In the Bayesian learning, we get the posterior distribution by equation (4) and the following likelihood function based on x and y [6,7].

$$L\ (\theta\ |\ x,\ y) = \prod{}^{n}{}_{i\,=\,1}\ P\ (y_i\ |\ x_i\ ,\ \theta\ ) \qquad (7)$$

We can also consider the Markov Chain Monte Carlo (MCMC) algorithm for the implementation of Bayesian learning [7]. In this paper we study on the Bayesian neural networks for regression rather than classification. On the data set D = (*X*, *y*), the output Y is predicted by the following prediction distribution.

$$P\ (ynew|D,\ Xnew,\tau)=\int P(\ ynew\ |\ Xnew,\theta\ )P(\theta\ |\ D,\tau)\ d\theta \qquad (8)$$

The value of "**θ**" represents a hyperparameter that is used in the distribution, for "λ".

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

IV. EXPERIMENTAL RESULTS

TABLE -1: Variable Description

Variable	Description	Role
CRIM	Crime Rate	Input
ZN	Residential land	
INDUS	Non-retail business acres	
CHAS	Charles river dummy variable	
NOX	Nitric oxides concentration	
RM	Average number of rooms	
AGE	Owner-occupied units	
DIS	Weighted distances	
RAD	Index of accessibility	
TAX	Full-value property-tax rate	
PTRATIO	Pupil-teacher ratio	
B	Proportion of blacks	
LSTAT	Lower status of population	
MEDV	Median value of homes	Output

For the performance comparison between Bayesian neural networks and deep learning models, we used the Boston housing data set [8]. This data consists of the input variables representing the characteristics of the house and an output variable representing the house price as shown above.

Using this data set, we carried out a simple experiment. Equally in both models, we set the number of neurons to 10 and the number of repetitions (epochs) to be equal to 1000. We show the experimental results of two comparative models in Table II.

In Table-2, we used 10 different random seed values to perform weight training on both comparative models. For each experiment, we estimated the mean squared error (MSE) and measured the calculation time [1]. Surprisingly, we found very minor changes in calculating time between the two models. However, a notable difference in prediction performance appeared, with Bayesian neural networks surpassing deep learning models by a significant margin in the MSE. This result clearly supports Bayesian neural networks' better prediction powers over standard deep learning approaches.

In summary, our studies demonstrated that, while there was minimal difference in computing time, Bayesian neural networks outperformed deep learning models in terms of predictive performance.

Table -2: Experimental Results

Seed	Comparative Models			
	Deep Neural Networks		*Bayesian Neural Networks*	
	MSE	*Time*	*MSE*	*Time*
1005	99.15	0.9945	12.41	1.4762
2020	101.58	1.0303	12.25	1.1859
3334	88.29	1.0443	12.83	0.8518
4501	90.19	0.9905	10.34	1.0971
5678	71.05	0.9615	18.73	1.2737
6329	98.06	1.0254	14.93	1.6098
7813	84.86	0.9765	9.95	0.9675
8516	114.84	1.0413	11.13	0.9785
9003	101.22	0.9862	11.23	1.0144
0255	90.68	1.0553	12.28	0.8618
Mean	93.97	1.0107	12.61	1.1317

V. CONCLUSIONS

In this paper, we showed the possibility of Bayesian inference and learning in neural networks and deep learning. We confirmed the advantages of Bayesian neural networks that provides better performance compared to the deep learning in our experimental result. In classification, the deep learning has dominated other machine learning algorithms. But, in regression problem, the Bayesian neural networks model provides competitive result over deep learning in this paper. The Bayesian neural networks model is constructed by learning the network weights using the parameters of the posterior distribution, which are updated by the prior distribution and likelihood function. However, it can be seen that the computing time of Bayesian neural networks is more than that of existing neural network models.

But, in Bayesian neural networks, we have to carry out the Markov chain simulation for constructing and sampling from the target posterior distribution. This process requires a lot of computation time. As the data size increases, the computation time of the Bayesian neural networks increases. So, in our future works, we will study on the new method of Bayesian learning to reduce the computing time of Bayesian neural networks. We will add a new distribution to Bayesian learning in addition to prior distribution, likelihood function, and posterior distribution for a faster Bayesian update process. In addition, we will study on the thinking machine like human using Bayesian learning for neural networks.

REFERENCES

[1] S. Theodoridis, Machine Learning A Bayesian and Optimization Perspective. London UK: Elsevier, 2015.
[2] S. Jun, “Frequentist and Bayesian Learning Approaches to Artificial Intelligence,” International Journal of Fuzzy logic and Intelligent Systems, vol. 16, no. 2, pp. 111–118, 2016.
[3] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, D. B. Bayesian Data Analysis, Third Edition. Boca Raton, FL: Chapman & Hall/CRC Press, 2013.
[4] K. B. Korb, and A. E. Nicholson, Bayesian artificial intelligence, second edition. London, UK: CRC press, 2011.
[5] T. M. Donovan, and R. M. Mickey, Bayesian Statistics for Beginners, Oxford, UK: Oxford University Press, 2019.
[6] R. M. Neal, Bayesian learning for neural networks. New York: Springer, 1996.
[7] H. M. Koduvely, Learning Bayesian Models with R, Birmingham, UK, Packt, 2015.
[8] The Boston house-price data, http://lib.stat.cmu.edu/datasets/boston, 2019.