

Data mining in web personalization using the blended deep learning model

Qusay Abdullah Abed, Osamah Mohammed Fadhil, Wathiq Laftah Al-Yaseen

Karbala Technical Institute, Al-Furat Al-Awsat Technical University, Iraq

Article Info

Article history:

Received Feb 21, 2020

Revised May 21, 2020

Accepted Jun 14, 2020

Keywords:

Data Mining

Deep Learning Model

Map Reduce

Support Vector Machine

Web Personalization

ABSTRACT

In general, multidimensional data (mobile application for example) contain a large number of unnecessary information. Web app users find it difficult to get the information needed quickly and effectively due to the sheer volume of data (big data produced per second). In this paper, we tend to study the data mining in web personalization using blended deep learning model. So, one of the effective solutions to this problem is web personalization. As well as, explore how this model helps to analyze and estimate the huge amounts of operations. Providing personalized recommendations to improve reliability depends on the web application using useful information in the web application. The results of this research are important for the training and testing of large data sets for a map of deep mixed learning based on the model of back-spread neural network. The HADOOP framework is using to perform a number of experiments in a different environment with a learning rate between -1 and +1. Also, using the number of techniques to evaluate the number of parameters, true positive cases are represent and fall into positive cases in this example to evaluate the proposed model.

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Qusay Abdullah Abed,
Karbala Technical Institute,
Al-Furat Al-Awsat Technical University, 56001, Karbala, Iraq.
Email: inkr.ks@atu.edu.iq

1. INTRODUCTION

Nowadays, the data set is continuously growing in big way. Existing machine-learning algorithms are unable to solve this in an effective way. Big data is huge in itself because it includes data in both cases organized and unorganized [1]. Since data is present in a complex way, storing and retrieving data is not an easy task in Big Data. If we have to process this big data. Big data is available in three levels and levels are speed, size, and variety [2, 3].

The speed here represents the speed of data access. As well, the size here represents the amount of data being process. Here, a variety of data sets represents the different shapes of a well-organized, unorganized, semi-structured structure. If we do this analysis, it will be useful for both predicting user behavior and decision-making [4, 5].

In this research work, they have offered the coordinating of Support Vector Machine (SVM)s are combined together in a remote group and responses to be conversion to a separate SVM [6]. The procedure continues while the remote vector group is left in the queue. Within this framework, a unique SVM is not mandatory in the implementation of the training arrangement. The traditional SVM model is completely full to have scrutiny on big scale lucrative problems. The research work has focused on the following: Specific data that go into three sections (data collection, data cleaning and data transformation) through which incorrect and hydrological noise data is to be processed. The learning machine (LM) algorithm enables both inbound and

outbound search[7]. Back Propagation (BP) neural network consists of two hidden folded layers assembled to predict level [8, 9].

Many other models are compare to the Back Propagation Network (BPN) to create an independent platform for disaggregated data and supervised text classification process [10]. Graph-primarily based semi-supervised gaining knowledge of (SSL) has attracted brilliant attention over the past decade [11]. Refining deep learning performance using random forest HTM cortical learning algorithm [12]. The proposed model using Map Reduce based Bayesian deep learning [13].

This paper presents “data mining in web personalization using the blended deep learning model” is based on a smart and dynamic algorithm that enhances performance and level of accuracy to extract user's large web usage information. This can be using for training and testing in different studies.

2. RESEARCH METHOD

Map Reduce Blended Deep Learning Based on Back Propagation Neural Network (MR-FDL-BPNN). In this section, to present exactly how one can Back Propagation learning for deep modes can be used if unique cooperatively learns a recommendation task using mapper reducer model [14, 15]. Many researchers take advantage of shallow methods. Additional importance, we have deployed these methods to the circumstance where to train a Back Propagation Neural Network entrenching together through a Back Propagation with sigmoidal deep multi-layer patterns on several layers of the network is required, and this has provided real benefits in the difficulty of tasks. We will use a similar approach to use the concepts established in Back Propagation for deep learning. It also helps to know the correct way to use the standard Back Propagation in deep positions by persevering in any layer of design. The most common and useful technique for Map Reduce Blended is Deep Learning Based on Back Propagation Neural Network [16]. A back-propagation is a supervised learning method and used in distinct diversities of the programs. Through Back-propagation learning method including gradient descent rule, we attempt to reduce open error curves. With the use of the Back propagation and multi-layer popular feeding of the neural forward, we did data to learn new things every time. One of the problems is the local minimum that to prohibit in the data training process by the steepest descent technique, in a noble manner way [17, 18].

One simple thing is to rearrange the weights for random numbers variance and try deep training yet again and again. The next solution is to change the weight and frequency that can be depended not only on the current error, but also on the previous changes as well. There are many ways to extract policies, where the maximum schemes commonly used are deep learning, back propagation neural network. The establishment of the rules of binary arrives and output strategies using blended level deep learning based back propagation neural network [19]. Discovery of the determined suitable collection of mining rules it finds high-quality consequences that can distinguish the set of specific information and assess the accuracy of the strategies extracted [20].

It is appropriate to explain the problems of the information that is mining according to the details. The deep learning mix is based on the neural network model of posterior diffusion that can improve the performance of the recommendation system suggestively [21], as shown in Figure 1. It has been widely used to improve performance based on deep learning. Back propagation neural network is one of the most utilized technique for data mining prediction that have been proposed [22, 23]. BPNN is one of the very different methods for forecasting algorithm very often number of times.

The neural process of posterior diffusion [10] consists of three layers as shown in Figure 2, with good illustration of the presence of an input layer, an output layer and a hidden layer of data. It is training in two distinct ways: forward propagation deep learning method and back deep learning process [24]. Through back propagation, we correct errors that can be raise while entering data into the output layer. We are able to enable search techniques as well as reduce error rate with this processing. In the hidden layer, the $2N + 1$ configuration run on continuously. Where, Where N represents the nerve numbers of comprehension. With the result from the hidden layer, it can be treated as an input to the output layer and the result is observed in its evaluation whether it is correct or not.

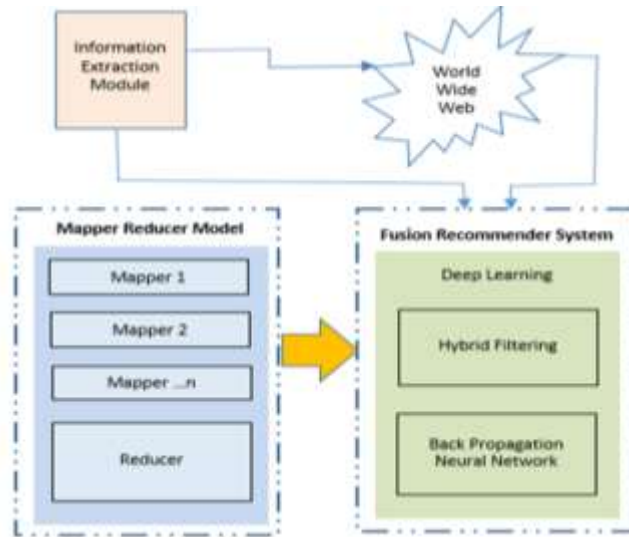


Figure 1. Map Reduce blended deep learning based on back propagation neural network

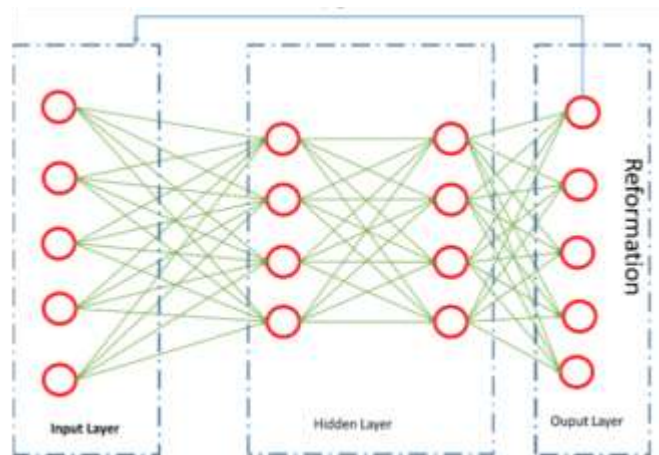


Figure 2. Back Propagation Neural Network

$N = [2, 3, \dots, 10]$ represents the discovery the greatest forecasting outcome.

$[x_1, x_2, \dots, x_n]$ represents the input variables.

$[w_1, w_2, \dots, w_n]$ represents the weights among the hidden layer and input layer

$[v_1, v_2, \dots, v_n]$ are weights assign to the hidden layers and output layers.

$Y(t)$ is the predictable output, in adding, the common Sigmoid function for transfer function to grip the nonlinear problematic, which can be articulated as:

$$Y = f(x) = 1/1 + e^{-x}$$

Hidden layer Z is a output of the canister be expressed as,

$$Z = f_1(\sum w_i x_i)$$

Correspondingly, the out layer of the container be expressed as,

$$U = f_2(\sum u_k z_k)$$

Thus, the progression of forward propagation is accomplished. Then, the error signal is form by the task of u and $Y(t)$, which is intend as,

$$U = 1/2 (\sum y_i(t)_u - z_i)^2$$

The runtime error is implement and some additional layer changes are providing, and the Gradient Ratio technology is applied. This method was used to advance the w_i and v_i weights and continue to do the same until the final result is not error-free, so we do this method until the final result as an output of the hidden layer. As a result matching of the standards of u and $Y(t)$ are matching by this method to generate results. The weights very much required in this are enhance through methodology that calculated some of the time.

3. APPLYING ALGORITHM

BPNN-Blended-Deep Learning it designed to classify to the effort as a common decision class input to the algorithm: complete data set attributes. Therefore, input data should be separate into two parts. Deep back propagation Training set and deep back propagation test sets. As well, the deep back propagation training ought to be generate to comprise an identical diversity of patterns. The deep back propagation training must be accessible as arrive to the BPNN. The set of output rules is the sophistication label be expect for specified enter sample and perform the blended operation for classified the proper results.

The steps to be following in the proposed approach are given below:

The first step is setting parameters for BPNN-Blended-Deep Learning algorithm. Followings parameters are to be set.

- a. Initial weight diversity (w): It is the diversity characteristically amongst $[-w, w]$. Here lowermost initial range $[-1, 1]$ is taken.
- b. Number of hidden layers: Up to five hidden layers can larger be understood the indication segment for better aspect on layers in a neural network (input, hidden and output). Occupancy us require the quantity to be 1.
- c. Number of Nodes in Hidden Layer: the required number of nodes in each hidden layer. Therefore, choosing the extensive variability of hidden layers and the number of nodes essentially relies of provisional outcome and error.
- d. Learning Rate for Gradient Descent: The values for the stage length frequently diversity varies from 0.1 to 0.9.
- e. Error tolerance: The characteristically error tolerance is a minor cost, confidential to the diversity zero to one.
- f. The production of each hidden node permits via a signed function; the variation is between 0 and 1.
- g. The anticipated mistakes (MSE) for avoiding the network training. If the real errors are similar to or less than subsequence error, the training could be ending.

The second step is applying Deep Back Propagation Training Pattern to BPNN. The provide input values to the deep back propagation from training patterns.

- a. Activation function: function to select the activation utility for hidden layer neurons and perform the operation of the output layer neurons.
- b. Learning rate: Keep confidentiality of the suggested technique in the start discovery out must be occupied insignificant, in the range of $[0.5 - 0.8]$
- c. Number of epochs: It is the unique sweep through altogether of the documents in the deep back propagation train set.

The last step is compute error among all the attained outputs and chosen output once the output layer is calculate in discovery error. Then, regulate the weights expending error gradient subsequently and to create a blended level calculation. This blended level should signify the weights of BPNN at the previous step. Therefore, must create a blended input indiscriminately related $(n-1)$ blended level in the variety of $[-1, 1]$, for blended input size of n . Also, apply deep training pattern and acquire output. At last, perform re deep back propagation training and increment new release count.

4. RESULTS AND DISCUSSION

To perform the experiment in different computing environments, the computational environment has a HADOOP 2.7.3 multi-node group to handle large data amount, i3-CPU -2.8GHz was used with 8.00 GB RAM. A different journal dataset is using to compare different classification methods forecast allocation news. Blended level Deep Learning based on back propagation neural network is assign to assess classification accuracy [25]. Correspondingly, Resampling binary neural network and back propagation neural network algorithms were using for classification. The HADOOP framework was used to perform a number of experiments in a different environment with a learning rate between -1 and $+1$ [26, 27]. To use a number of

techniques to evaluate the number of parameters, true positive cases are represent and fall into positive cases to calculate in this example to evaluate the proposed approach. To find out the exact value used to obtain a number of positive results that can show a positive response.

The metric of accuracy can be dependent on four features. These features are as follows:

- a. The real pluses are the appropriate documents that were appropriate identified as the right category.
- b. The real negatives are inappropriate documents that are appropriate identified as an inappropriate category.
- c. False positives are inappropriate documents that are incorrect identified as an appropriate category. Therefore, it is one kind of error.
- d. Erroneous negatives are appropriate documents that are imperfect identified as inappropriate category. Therefore, it is contrast another type of error.

After these four features, the two different parameters used for to find the precision scale are F.

The following two parameter Precision parameter is the parameter specifies how many documents determine properly:

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

Recall parameter is the parameter that specifies how many appropriate documents are determined. Recall is a value found by the true positive value and divided by the total rate of true positive and false negative values:

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

These are best methods to find out result in the form of measurement but this approach is not simple for getting the proper decision. Consequently, the F-measure acquires a single measure for evaluating results. The F-measure approach is the harmonic mean of precision and recall:

$$\text{F-measure} = \frac{2\text{True positives}}{2\text{True positives} + \text{False negatives} + \text{False positives}}$$

SVM: Support Vector Machine algorithm.

BPNN: Back Propagation Neural Network algorithm.

S3VM: Semi-Supervised Support Vector Machines algorithm.

FDL- BPNN : Blended Deep Learning- Back Propagation Neural Network algorithm.

Table 1 illustrates the concert metrics of the proposed algorithm with 10-fold sampling algorithms, cross-validation, and changing the learning rate among 0.1 and 0.8. The maximum technique with FDL-back propagation neural network learning through learning rate 0.3 agrees to the table.

Table 1. Blended level deep leaning (multilayer perceptron) perceptron with resample

Name Algorithm	Precision	Recall	F-Measure
SVM	0.900	0.856	0.977
BPNN	0.910	0.884	0.977
S3VM	0.910	0.844	0.977
FDL- BPNN	0.920	0.876	0.977

5. CONCLUSION

This paper presents “data mining in web personalization using blended deep learning model” is base on a smart and dynamic algorithm that enhances performance and level of accuracy to extract user's large web usage information. This can be using for training and testing in different studies. This is really a great evaluation process in the field of data mining. Our proposed model FDL-BPNN includes blended level deep learning Back Propagation Network. Blended level (means that the combination of repetition and machine learning approach). In this work to combine the back propagation network five kernel methods with deep learning concept Blended deep learning, Deep Back Propagation, operation of sigmoidal Back Propagation, gradient descent data selection technique. Using the new SSL and S3VM, developing a new Map reduce model to improve efficiency and reduce the complexity of assignment and recommendation information.

REFERENCES

- [1] A. H. Al-Hamami and A. A. Flayyih, "Enhancing Big Data Analysis by using Map-reduce Technique," *Bulletin of Electrical Engineering and Informatics*, vol. 7, pp. 113-116, 2018.
- [2] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustainable Energy, Grids and Networks*, vol. 6, pp. 91-99, 2016.
- [3] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data" *IEEE Journal of Selected topics in applied earth observations and remote sensing*, vol. 7, pp. 2094-2107, 2014.
- [4] K. Zhang and X.-W. Chen, "Large-scale deep belief nets with mapreduce" *IEEE Access*, vol. 2, pp. 395-403, 2014.
- [5] D. J. Hand, "Data Mining," *Encyclopedia of Environmetrics*, vol. 2, 2006.
- [6] X. Ke, H. Jin, X. Xie, and J. Cao, "A distributed SVM method based on the iterative MapReduce," in Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015), 2015, pp. 116-119.
- [7] I. Salamah and M. A. Ganiardi, "Development of e-learning software based multiplatform components," *Bulletin of Electrical Engineering and Informatics*, vol. 6, pp. 228-234, 2017.
- [8] D. Wan, Y. Xiao, P. Zhang, and H. Leung, "Hydrological big data prediction based on similarity search and improved BP neural network," in *2015 IEEE International Congress on Big Data*, 2015, pp. 343-350.
- [9] A. Maselena, N. Sabani, M. Huda, R. Ahmad, K. A. Jasmi, and B. Basiron, "Demystifying learning analytics in personalised learning," *International Journal of Engineering & Technology*, vol. 7, pp. 1124-1129, 2018.
- [10] S. Z. Mishu and S. Rafiuddin, "Performance analysis of supervised machine learning algorithms for text classification," in 2016 19th International Conference on Computer and Information Technology (ICCIT), 2016, pp. 409-413.
- [11] Y. Lyu, X. Fan, and K. Liu, "An optimized strategy for small files storing and accessing in hdfs," in 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), 2017, pp. 611-614.
- [12] M. A. Abbas, "Improving deep learning performance using random forest HTM cortical learning algorithm," in 2018 First International Workshop on Deep and Representation Learning (IWDRL), 2018, pp. 13-18.
- [13] M. O. Shafiq and E. Torunski, "Towards MapReduce based Bayesian deep learning network for monitoring big data applications," in 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 2112-2121.
- [14] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial examples: Attacks and defenses for deep learning," *IEEE transactions on neural networks and learning systems*, 2019.
- [15] S. K. Cheung, L. Kwok, W. Ma, L. Lee, and H. Yang, *Blended Learning*: Springer, 2017.
- [16] K. Modarresi and J. Diner, "An efficient deep learning model for recommender systems," in International Conference on Computational Science, 2018, pp. 221-233.
- [17] A. Saichand and P. K. Baruah, "Blended Learning-Assimilating Authentic Data Into Deep Learning Models," in 2018 IEEE 25th International Conference on High Performance Computing Workshops (HiPCW), 2018, pp. 75-80.
- [18] L. R. Halverson and C. R. Graham, "Learner Engagement in Blended Learning Environments: A Conceptual Framework," *Online Learning*, vol. 23, pp. 145-178, 2019.
- [19] Y. Wang, G. Sun, and H. Shi, "Research on Blended Learning Activity and Application," in International Conference on E-Learning, E-Education, and Online Training, 2018, pp. 266-272.
- [20] B. A. Ajayi and H. Hussin, "Conceptualizing Information Technology Governance Model for Higher Education: An Absorptive Capacity Approach," *Bulletin of Electrical Engineering and Informatics*, vol. 7, pp. 117-124, 2018.
- [21] Q. A. Abed, M. T. Abdullah, and H. J. Dikhil, "Machine learning algorithms for distributed operations in internet of things IoT," *Periodicals of Engineering and Natural Sciences*, vol. 7, pp. 1638-1648, 2019.
- [22] X. Ding, Y. Li, and Y. Zhao, "A framework of user model based on Semi-supervised techniques," in 2008 IEEE International Conference on e-Business Engineering, 2008, pp. 396-401.
- [23] N. S. A. A. Bakar and N. Yahya, "Test First Model for Application in the Academic Setting," *Bulletin of Electrical Engineering and Informatics*, vol. 7, pp. 80-89, 2018.
- [24] K. Neshatpour, M. Malik, and H. Homayoun, "Accelerating machine learning kernel in hadoop using fpgas," in 2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2015, pp. 1151-1154.
- [25] P. De Bra, "Challenges in user modeling and personalization," *IEEE Intelligent Systems*, vol. 32, pp. 76-80, 2017.
- [26] Y. S. Tan, J. Tan, E. S. Chng, B. S. Lee, J. Li, S. Date, *et al.*, "Hadoop framework: impact of data organization on performance," *Software: Practice and Experience*, vol. 43, pp. 1241-1260, 2013.
- [27] C. Lam, *Hadoop in action*: Manning Publications Co., 2010.