The 6th International Conference on Digital Arts, Media and Technology (DAMT) and
4th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (NCON)

2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering | 978-1-6654-1569-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/ECTIDAMTNCON51128.2021.9425715

Adaptive Learning Rate for Dealing with Imbalanced Data in Classification Problems

Ratanon Jantanasukon[1] and Arit Thammano[2]
Computational Intelligence Laboratory,
Faculty of Information Technology,
King Mongkut's Institute of Technology Ladkrabang,
Bangkok 10520 Thailand
Email: [1]60070079@kmitl.ac.th and [2]arit@it.kmitl.ac.th

***Abstract*—This research modified a backpropagation learning algorithm in order to increase its ability to deal with imbalanced data problems. We used the backpropagation algorithm and a concept of multiple adaptive learning rates to train the feedforward neural network. Using multiple adaptive learning rates allowed us to achieve a classification model that had fewer problems when dealing with an imbalanced dataset. The experimental results showed that the proposed method performed significantly better than the conventional backpropagation neural network in all tests.**

***Keywords*—Imbalanced data, Classification, Adaptive learning rate, Feedforward neural network, Backpropagation algorithm**

I. INTRODUCTION

Classification is a method that can identify an instance of unseen data as an instance of a class of data that have certain differentiating features. The classification model is trained with a set of training data before it can identify the instance of unseen data. Today, a classification model can be constructed based on a variety of techniques and approaches, such as Neural Network, KNN, Decision Tree. Previous research studies [1, 2] concluded that a classification model based on Neural Network provided a lower error rate and better efficiency than a model based on other techniques. However, the artificial neural network that utilizes the backpropagation algorithm in its model training often have problems when classifying imbalanced data. The distinctive characteristic of imbalanced data is that the data can be classified into more than one classes, and most of the data are in majority classes, while much fewer instances of data are in minority classes, i.e., there is an imbalance of the number of instances of data in each class. When a classification model is trained with a set of imbalanced data, it is likely to misclassify an instance of unseen data.

In this research, the classification was conducted based on a feedforward neural network model, the most widely used neural network model. In each iteration of the training process, the data is fed in the forward direction from the input layer to the output layer. The model then adjusts the weight value assigned to each connection by using the backpropagation algorithm. Backpropagation algorithm [3] is a learning algorithm that is widely used in the feedforward neural network model. The algorithm is based on the gradient descent technique. In gradient descent, choosing a suitable learning rate is crucial for the effectiveness of the algorithm. Learning rate is usually set to be a fixed constant. However, using a constant learning rate when training a model with imbalanced data generally causes the model to be biased towards the majority classes, the main issue in classification. Therefore, we had an idea that it can and should be adjusted throughout the model training to effectively minimize the error curve.

In the past, many attempts at solving imbalanced data problems have been based on 2 approaches: 1) balancing the data in the pre-processing stage and 2) improving the model training method. The first approach focuses on balancing the data before using them for model training. There are two common methods that are used in data balancing: over-sampling or under-sampling and SMOTE. Over-sampling increases the amount of data in the minority class to be close to that of the majority class, while under-sampling decreases the amount of data in the majority class to be close to that of the minority class. SMOTE [4], on the other hand, creates a new set of minority class based on the original data. In addition, there have been several recent methods of the first approach reported in the literature. One method used Uncorrelated Discriminant Analysis to select data attributes that had little relation to each other so that the means of every class were the farthest apart from each other. Another method used Fuzzy rule-based classifiers to train the model with imbalanced data. Regarding SMOTE, a SMOTEBoost [5] algorithm had been developed that combined SMOTE with a model efficiency increasing method to make the model more accurate and able to learn imbalanced data effectively. The second approach, however, focused on improving the model training method. Examples are such as Hellinger distance (HDDT) and Class Confidence Proportion Decision Tree (CCPDT) methods [6]. The authors of these methods attempted to improve the capability of Decision Tree classifier in dealing with imbalanced data. Another example was in improving KNN Algorithm to deal with imbalanced data [7]. KNN, one of the most common data mining algorithms, was modified into a Distance-weighted K-Nearest Neighbor (DWKNN) algorithm. DWKNN assigned a weight value to each data instance in the classification procedure. In the training process, a positive weight value was assigned to each data instance so that when the training process ended, the weight value would only be positive. Another approach, proposed by Amiri et al. [8], used adaptive learning rate to enhance the ability of the backpropagation neural network. Their technique adaptively adjusted the learning rate during the training process based on error curve changes. However, the authors did not focus on solving imbalanced data problems.

This research also used adaptive learning rate to solve the imbalanced data problem. In general, most methods for solving the imbalanced data problem focused on pre-processing data or modifying cost function. This research, however, focused on using multiple adaptive learning rates, of which each learning rate could be adjusted to be compatible with the model in each iteration. The efficiency of the model would be compared with that of a conventional feedforward neural network.

Authorized licensed use limited to: California State University Fresno. Downloaded on June 19,2021 at 19:25:09 UTC from IEEE Xplore. Restrictions apply.

The rest of this paper is organized as follows: Section II describes the background theories; Section III presents the proposed method; Section IV describes the datasets and the test results; and Section V concludes the paper.

II. BACKGROUND THEORIES

A. Feedforward neural network

This research focuses on solving classification problems using the feedforward neural network. In the beginning, the number of layers and the number of nodes in each layer are determined. The number of nodes in the input layer is equal to the number of data features, and the number of nodes in the output layer is equal to the number of classes in the dataset. The number of hidden layer nodes depends on the user.

B. Backpropagation algorithm

At the beginning of the training process, the weight value of each connection line is randomly generated between 0 and 1. In this research, a backpropagation algorithm [9] is used to update the weights of the model. The algorithm is divided into 2 parts: Forward propagation and Backward propagation. Forward propagation propagates the input value from the input layer to the output layer. In the input layer, the output value is equal to the input value that comes into each node. For the hidden layer and output layer, the output value of each node is calculated by evaluating the weighted sum of its inputs and passing it through an activation function, as shown in (1) – (4). The second part, Backward propagation, propagates the output value back to the previous node, i.e., it determines the derivative value of the total error of the model and send it back to each layer starting from the output layer all the way to the input layer. For the output layer, it calculates the error value of each node, as defined in (5). For the hidden layer, it calculates the error value of each node, as defined in (6).

$$I_j = \sum_i w_{ij} O_i \tag{1}$$

$$O_j = \frac{1}{1+e^{-I_j}} \tag{2}$$

$$I_k = \sum_j w_{jk} O_j \tag{3}$$

$$O_k = \frac{1}{1+e^{-I_k}} \tag{4}$$

$$Err_k = O_k(1-O_k)(T_k - O_k) \tag{5}$$

$$Err_j = O_j(1-O_j)\sum_k Err_k w_{jk} \tag{6}$$

Weight value of each connection is then updated based on (7) – (10).

$$\Delta w_{jk} = \alpha Err_k O_j \tag{7}$$

$$w_{jk} = w_{jk} + \Delta w_{jk} \tag{8}$$

$$\Delta w_{ij} = \alpha Err_j O_i \tag{9}$$

$$w_{ij} = w_{ij} + \Delta w_{ij} \tag{10}$$

where w_{ij} is the weight value of the connection between the input node i and the hidden node j; w_{ij} is the weight value of the connection between the hidden node j and the output node k. O_i is the output of the input node, and O_j is the output of the hidden node. T_k and O_k are the target output and the actual output of the output node respectively. Err_k is the error of the output node, and Err_j is the error of the hidden node.

III. PROPOSED METHOD

A. Overview of the proposed method

Imbalanced data problem is a problem that the amount of data in one class is significantly larger or smaller than the amount of data in other classes. Imbalanced data causes the backpropagation neural network to be more biased towards majority classes. Therefore, our work attempts to solve this problem by applying a concept of multiple adaptive learning rates in the backpropagation algorithm. In the proposed multiple adaptive learning rates, the number of learning rates corresponds to the number of classes. The learning rate of each class is self-adapted to reduce the bias towards majority classes.

B. Adaptive learning rate

As mentioned earlier, each data class has its own learning rate. At the beginning of a training process, every learning rate is assigned a fixed, equal value such as 0.1 or 0.05. Then, the learning rate of each class is adjusted based on the difference between its error in the current and previous iterations. If the difference is negative, the learning rate of that class will be decreased, and if the difference is positive, the learning rate of that class will be increased. If the difference is equal to 0, the learning rate of that class will be kept unchanged.

$$\alpha_c(t+1) = \begin{cases} \alpha_c(t) \times \delta_1(t), & \text{if } E_c(t-1) > E_c(t) \\ \alpha_c(t) \times \delta_2(t), & \text{if } E_c(t-1) < E_c(t) \\ \alpha_c(t), & \text{if } E_c(t-1) = E_c(t) \end{cases} \tag{11}$$

$$\delta_1(t+1) = \delta_1(t) + \sigma e^{-\frac{t}{\mu}} \tag{12}$$

$$\delta_2(t+1) = \delta_2(t) - \sigma e^{-\frac{t}{\mu}} \tag{13}$$

$$\delta_1(0) = 1 - \beta \tag{14}$$

$$\delta_2(0) = 1 + \beta \tag{15}$$

where $\alpha_c(t)$ represents a learning rate of class c at iteration t, $\delta_1(t)$ and $\delta_2(t)$ are the coefficients that are used to adjust the learning rate and will vary according to the training iteration of the model as shown in (12) and (13). The value of $\delta_1(t)$ is always less than 1 while the value of $\delta_2(t)$ is always greater

Authorized licensed use limited to: California State University Fresno. Downloaded on June 19,2021 at 19:25:09 UTC from IEEE Xplore. Restrictions apply.

than 1. Figs. 1 and 2 show the values of $\delta_1(t)$ and $\delta_2(t)$ overtime. $\delta_\bullet(0)$ represents the initial value of the coefficient.

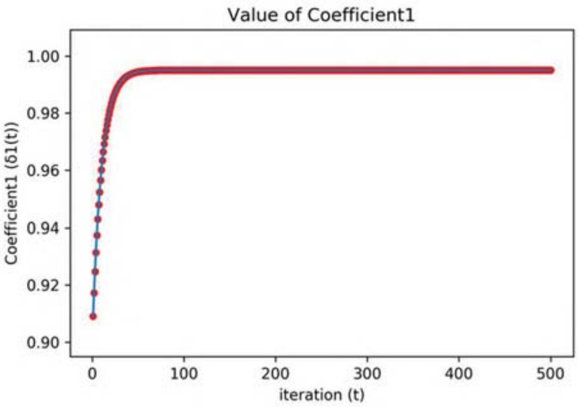

Fig. 1. Coefficient 1 ($\delta_1(t)$) in each iteration

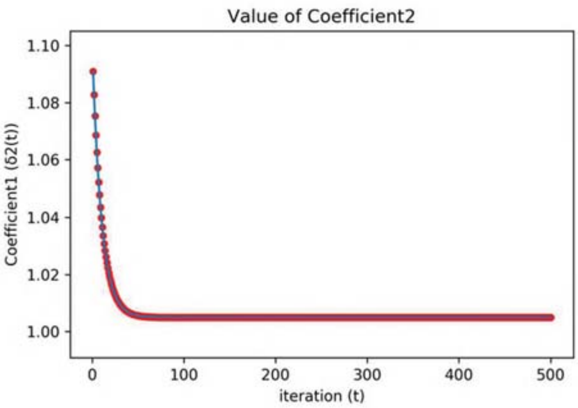

Fig. 2. Coefficient 2 ($\delta_2(t)$) in each iteration

IV. Datasets and Test Results

As mentioned in Section III, the goal of the proposed method is to improve the performance of the backpropagation algorithm when dealing with imbalanced datasets. Its performance was evaluated on four imbalanced benchmark datasets as describes in the following subsection.

A. Datasets

The four imbalanced benchmark datasets used were Breast Cancer Coimbra, Diabetes, Vertebral, and Yeast datasets from the UC Irvine Machine Learning Repository. K-fold cross validation was used to split each dataset into K sets of training and testing data. K is either 5 or 10, depending on the size of the dataset.

B. Testing, Results, and Discussion

The proposed method and a conventional backpropagation algorithm were tested on all four imbalanced benchmark datasets. It is convenient to describe the test on the diabetic dataset as a typical example. The test was a 10-fold cross validation on the Diabetes dataset. It contained two classes of data. At the start of the run, each class was assigned the same initial learning rate of 0.1; β, σ and μ (the parameters in (12) – (15)) were assigned the value of 0.1, 0.01 and 10, respectively. The test was run 500 iterations. For each training iteration, an accuracy of classifying the training data was recorded. Fig. 3 shows the accuracy of classifying the training data of each class in all 500 iterations. From Figs. 3 and 4, it can be seen that the learning rate of the minority class continuously increased overtime as a result of its low accuracy. This relationship was exhibited that the adaptive learning rate strategy was successful in dealing with imbalanced data. At the end of 500 iterations, the classification accuracy achieved by the proposed method was much better than that achieved by the conventional backpropagation method.

The results of the experiments are shown in Tables II – V. Table II shows that the mean accuracy of the proposed method in Breast Cancer classification increased over the conventional method by 11.80%. Table III and IV show that the mean accuracy of the proposed method in Diabetes and Vertebral classification increased over the conventional method by 11.30% and 32.58%, respectively. Finally, Table V shows that the mean accuracy of the proposed method in Yeast classification increased over the conventional method by 18.57%. It can, hence, be concluded that the proposed method could handle imbalanced data problem more efficiently and more reliably than a conventional backpropagation neural network.

TABLE I. Learning parameters of the proposed method

Maximum number of iterations	500
$\alpha_c(0)$	0.1
σ	0.01
β	0.1
μ	10
$\delta_1(0)$	0.9
$\delta_2(0)$	1.1

TABLE II. Results of the proposed method versus those of the conventional method on Breast Cancer dataset

	Accuracy						
	Fold 1	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Mean**	**S.D.**
BPNN	64.64	59.09	59.09	81.81	54.54	63.65	10.658
Proposed Method	86.36	63.63	86.36	77.27	63.63	75.45	11.410

TABLE III. Results of the proposed method versus those of the conventional method on Diabetes dataset

	Accuracy											
	Fold 1	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Fold 6**	**Fold 7**	**Fold 8**	**Fold 9**	**Fold 10**	**Mean**	**S.D.**
BPNN	65.79	65.79	65.79	65.79	65.79	65.79	65.79	65.79	65.79	65.79	65.79	0
Proposed Method	72.37	80.26	78.79	73.68	72.36	82.89	78.95	76.31	73.68	81.58	77.09	3.932

Authorized licensed use limited to: California State University Fresno. Downloaded on June 19,2021 at 19:25:09 UTC from IEEE Xplore. Restrictions apply.

TABLE IV. Results of the proposed method versus those of the conventional method on Vertebral dataset

	Accuracy											
	Fold 1	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Fold 6**	**Fold 7**	**Fold 8**	**Fold 9**	**Fold 10**	**Mean**	**S.D.**
BPNN	48.39	48.39	48.39	48.39	48.39	48.39	48.39	48.39	48.39	48.39	48.39	0
Proposed Method	70.97	74.19	83.87	70.97	90.32	93.55	87.1	80.65	77.42	80.65	80.97	7.820

TABLE V. Results of the proposed method versus those of the conventional method on Yeast dataset

	Accuracy						
	Fold 1	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Mean**	**S.D.**
BPNN	31.21	28.28	30.3	31.19	28.77	29.95	1.362
Proposed Method	51.34	49.83	53.53	44.74	43.15	48.52	4.412

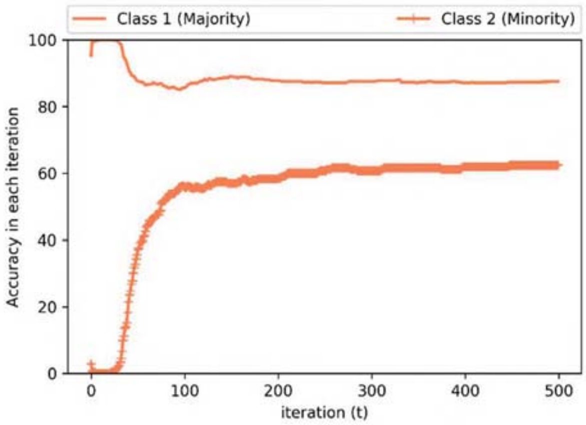

Fig. 3. Accuracy at different iterations

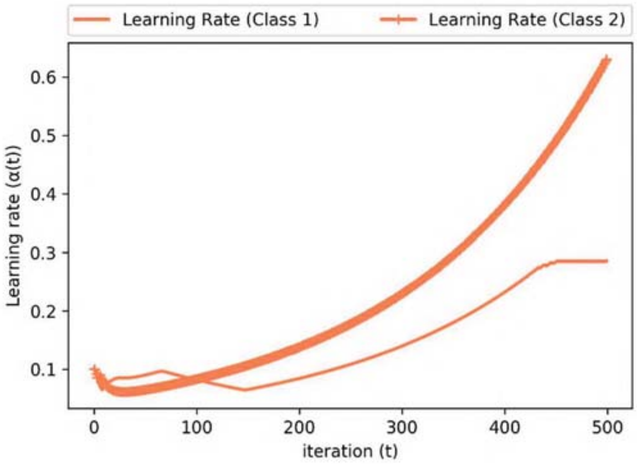

Fig. 4. Learning rate at different iterations

V. Conclusion

This paper proposes a method that used a multiple adaptive learning rate strategy to solve imbalanced data problem in classification. In the training process, the learning rate of each class is adjusted based on the difference between its error in the current and previous iterations. From results of 5- and 10-fold cross-validation runs of the proposed method versus a conventional backpropagation method, the proposed method was able to improve the classification accuracy considerably, ranging from 11.30 to 32.58% across all data classes. The findings from this study can directly benefit researchers in the field of data classification in their effort to deal with the problem of imbalanced data.

References

[1] R. Anand, K. G. Mehrotra, C. K. Mohan, and S. Ranka, “An improved algorithm for neural network classification of imbalanced training sets,” IEEE Transactions on Neural Networks, vol. 4, pp. 962–969, 1993.

[2] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, “Deep supervised learning for hyperspectral data classification through convolution neural networks,” Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium, Milan, Italy, July 26-31, 2015.

[3] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd ed., CA: Morgan Kaufmann Publishers, 2006, pp. 327-336.

[4] N. V. Chawla, K. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique” Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.

[5] N. V. Chawla, A. Lazarevic, and K. Bowyer, “SMOTEBoost: Improving Prediction of the Minority Class in Boosting,” Proceedings of the 7th European Conference on Principles and Practice of Knowledge Discovery in Databases, Cavtat-Dubrovnik, Croatia, September 22-26, 2003.

[6] W. Liu, S. Chawla, D. A. Cieslak, and N. V. Chawla, “A Robust Decision Tree Algorithm for Imbalanced Data Sets” Proceedings of the 2010 SIAM International Conference on Data Mining, Columbus, OH, USA, April 29-May 1, 2010.

[7] J. Gou, L. Du, Y. Zhang, and T. Xiong, “A New Distance-weighted k-nearest Neighbor Classifier,” Journal of Information & Computational Science, vol. 9, no. 6, pp. 1429-1436, 2012.

[8] Z. Amiri, H. Hassanpour, N. M. Khan, and M. H. M. Khan, “Improving the performance of multilayer backpropagation neural networks with adaptive leaning rate,” Proceedings of the 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems, Durban, South Africa, August 6-7, 2018.

[9] M. K. S. Alsmadi, K. Omar, and A. S. Noah, "Back Propagation Algorithm: The Best Algorithm Among the Multi-layer Perceptron Algorithm," International Journal of Computer Science and Network Security, pp. 378-383, 2009.

Authorized licensed use limited to: California State University Fresno. Downloaded on June 19,2021 at 19:25:09 UTC from IEEE Xplore. Restrictions apply.