Addressing Data Imbalance in Neural Network Training via Target Adjustment

Sang-Hoon Oh*
*Mokwon University, Korea
E-mail:ohsanghoon16@gmail.com

1. Introduction

In classification tasks employing artificial intelligence, datasets are often severely imbalanced, rendering conventional artificial neural network (ANN) models—typically developed under the assumption of balanced data—less effective [1-4]. To address this issue, data-level techniques such as oversampling and undersampling are commonly employed. Oversampling increases the number of samples in the minority class, typically by duplicating existing examples or generating synthetic ones (e.g., using the Synthetic Minority Oversampling Technique, SMOTE) [4]. In contrast, undersampling reduces the number of samples in the majority class by randomly discarding instances, thereby balancing the class distribution [5].

Beyond these data-level approaches, algorithmic methods have been proposed to control the weight update magnitude through the loss function in ANNs with sigmoid activation output nodes [6]. However, such methods cannot be directly applied to ANNs utilizing softmax activation output nodes.

In this study, an alternative strategy is proposed in which an ANN with softmax activation output nodes inherently mitigates data imbalance during the training phase. Specifically, by adaptively adjusting target values during learning, instances from majority classes with abundant samples are trained with reduced emphasis, whereas those from minority classes with limited samples are trained with increased emphasis.

2. Neural Networks with Softmax Activation Output Nodes and Probabilistic Target Encoding

Let's consider an artificial neural network with a single hidden layer as shown in figure 1. When an training sample $\boldsymbol{x} = [x_1, x_2, \dots, x_N]^T$ is presented to the neural network. The jth hidden node us given by

$$h_j = f(\sum_{i=0}^{N} w_{ji} x_i)\,, (x_0 = 1) \tag{1}$$

where w_{ji} the connecting weight between x_i and h_j . $f(.)$ is the activation function of hidden node such as sigmoid or ReLU functions. The kth output node is given by $y_k = \text{SoftMax}(\widehat{y_k})$ where $\widehat{y_k} = \sum_{j=0}^{H} v_{kj} h_j$, $(h_0 = 1)$ and

$$\text{SoftMax}(\widehat{y_k}) = \frac{e^{\hat{y}_k}}{\sum_{j=1}^{M} e^{\hat{y}_j}} \tag{2}$$

and v_{kj} is the connecting weight between and h_j and y_k.

Usually, we use the ont-hot encoding to indicate the target of output node. The probabilisrtic target encoding was prposed to improve the classification performance of neural networks for test samples [7]. That is,

$$t_k = \begin{cases} Q_k(\boldsymbol{x}), & \text{if } \boldsymbol{x} \text{ originates from class } k \\ \dfrac{1 - Q_k(\boldsymbol{x})}{M-1}, & \text{otherwise.} \end{cases} \tag{3}$$

Here, t_k is the desired value of The kth output node and $Q_k(\boldsymbol{x})$ denotes the posterior probability that $\boldsymbol{x}$ originates from class k.

3. Target Adjustment for Addressing Data Imbalance.

The probabilistic target encoding can be regarded as a generalization of label smoothing [7], [8]. Label smoothing corresponds to the case that $Q_k(\boldsymbol{x})$ is assigned the same value for all classes indexed by k. In contrast, probabilistic target encoding allows different target values for each class. Specifically, a higher target value (e.g., "1") is assigned to the minority class, whereas a lower value is assigned to the majority class. This strategy effectively regulates the magnitude of weight updates, ensuring that instances from majority classes are trained with reduced emphasis, while those from minority classes receive greater emphasis. The effectiveness of the proposed method is validated through simulations using the thyroid dataset from the UCI Machine Learning Repository..

4. Discussion and Conclusion

In this presentation, we adopted the probabilistic target encoding to address the imbalanced data problem. The effectiveness of the proposed method was validated through simulations using the Thyroid dataset from the UCI Machine Learning Repository.

5. References

[1] H. Zhao, “Instance Weighting versus Threshold Adjusting for Cost-Sensitive Classification,” Knowledge and Information Systems, vol.15, 2008, pp. 321-334.

[2] Y.-M. Huang, C.-M. Hung, and H. C. Jiau, “Evaluation of Neural Networks and Data Mining Methods on a Credit Assessment Task for Class Imbalance Problem,” Nonlinear Analysis, vol.7, 2006, pp. 720-747.

[3] R. Bi, Y. Zhou, F. Lu, and W. Wang, “Predicting gene ontology functions based on support vector machines and statistical significance estimation,” Neurocomputing, vol.70, 2007, pp.718-725.

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

[5] P. Kang and S. Cho, “EUS SVMs: Ensemble of under-sampled SVMs for data imbalance problems,” in: Proc. ICONIP’06, (Springer, Berlin, 2006) 837-846.

[6] Sang-Hoon Oh, “Error back-propagation algorithm for classification of imbalanced data,” Neurocomputing, Vol. 74, pp. 1058-1061, 2011.

[7] Sang-Hoon Oh, “Probabilistic Target Encoding of Neural Networks with Softmax Output Nodes,” International Journal of Contents, vol. 18. no. 4, pp. 102-108, Dec. 2022.

[8] R. Müller, S. Kornblith, and G. Hinton, “When Does Label Smoothing Helps?,” arXiv.1906.02629

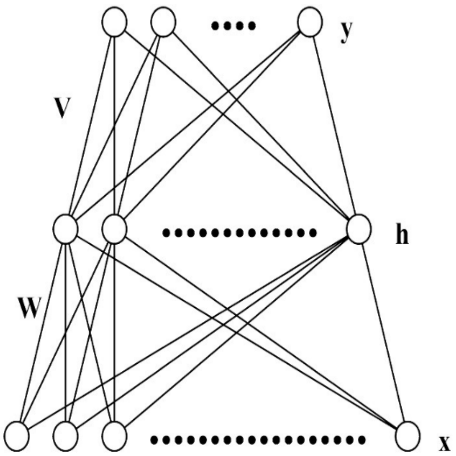

Figure 1 The architecture of Neural Network with a single hidden layer