A study on the GEV activation function for classification of class imbalance data

Hyebin Park, Juyoung Hong, Yonggwan Shin and Jeong-Soo Park[[1]](#footnote-1)

**Abstract** The classification problems for the imbalance data occur frequently in our lives, and it is important to solve them well. Therefore, we propose a method combining the generalized extreme value (GEV) activation function and the cost-sensitive learning method and over-sampling in a simple neural network model. In order to check the performance of the proposed method, 100 data sets were employed and 5 evaluation metrics were considered. The one-way analysis of variance (ANOVA) and post-hoc tests were performed under the 5% significance level. 162 out of 500 combinations of data sets and 5 evaluation metrics, that is, 32.4% of total showed a significant difference under the 5% significance level. The optimal sampling ratio is judged to be 20:1 and when we compared the results of proposed method and SOTA model, excellent results were obtained in all five data sets.

**Key words:** Activation function,Class imbalance, Over-sampling, Sigmoid function

1. Introduction

Nowadays, the development of Internet technologies such as social network service (SNS) and the Internet of things (IOT) has enabled us to collect a lot of data. Various types of data are being generated rapidly, and the amount of data is also increasing. Statistical models using such big data help us to make quick and accurate decisions. For this reason, machine learning has become an essential tool in all industries to understand data and improve productivity.

Among them, classification refers to classifying each instance into a given class by deriving a meaningful relationship between input variable and target variable. Classification problem is a very important problem that occurs very often. Traditional classification algorithms assume that the number of samples between classes is approximately equal. But in reality, that is rarely the case. Such a case in which a specific class appears more frequently than other classes is said to be a class imbalance problem, and it exists in real life such as medical diagnosis, fire detection and fraudulent transaction detection.

Previously, this problem was solved through re-sampling method or cost-sensitive learning method, but recently, there are some trials using the GEV activation function to solve class imbalance problem. Wang et al.[1] used GEV as the link function of GLM, and Lkhagvadori et al.[2] improved classification performance by using a neural network model that has Gumbel distribution as an activation function. Most recently, J. Bridge et al.[3] used GEV activation function in a convolution neural network (CNN) model that diagnoses COVID-19.

1. Methods

When solving a classification problem using a multi-layer perceptron, we often use sigmoid as activation function. The sigmoid function is calculated as in Equation 1 and has a symmetrical structure as shown in Figure 1.

|  |  |
| --- | --- |
|  | Equation 1 |

Our proposed method is to use the CDF of the GEV distribution as the activation function instead of the sigmoid function, because it makes all real inputs to a value between 0 and 1. The GEV distribution has three parameters such as , and in this study, each parameters were estimated using back propagation method with the weights of the neural network model. The GEV activation function is calculated as in Equation 2, and has an asymmetric structure as shown in Figure 2.

|  |  |
| --- | --- |
| ,  Defined on {s : 1 + > 0 }  Where | Equation 2 |

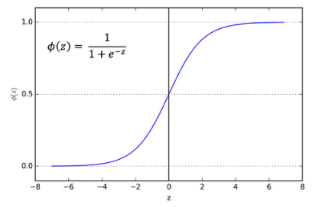
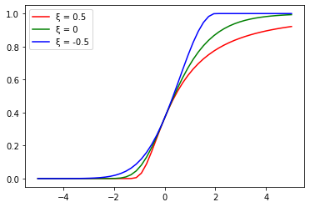
 

Figure 1: Sigmoid function Figure 2: Cumulative distribution function of GEVD

To compare the performance of the proposed method, we considered the following 5 cases for 100 KEEL imbalanced data sets.

1. (Baseline) MLP using sigmoid activation
2. MLP using GEV activation function
3. MLP using GEV activation and Thresholding
4. MLP using GEV activation, Thresholding and Focal Loss
5. MLP using GEV activation, Thresholding, Focal Loss and Over-Sampling

The data used in this experiment are shown in Table 1. The asymmetry ratio was calculated by dividing the number of majority class samples by the number of minority class samples, larger this value means the more severe asymmetry.

**Table 1:** The data used in this experiment (5 out of 100 KEEL imbalance data sets)

|  |  |  |  |
| --- | --- | --- | --- |
| ***Data name*** | ***# of samples*** | ***# of input variables*** | ***Imbalance ratio*** |
| abalone19 | 4,174 | 8 | 129.44 |
| abalone20\_vs\_8-9-10 | 1,916 | 8 | 72.69 |
| kr-vs-k-zero\_vs\_fifteen | 2,193 | 19 | 80.22 |
| pocker-8\_vs\_6 | 1,477 | 9 | 85.88 |
| pocker-8-9\_vs\_5 | 2,075 | 9 | 82.00 |

For a more reliable result, the average of the results obtained by changing the seed (30 times) was compared, and for each data, 5 evaluation indicators [4] (Equation3) suitable for unbalanced data were evaluated. The structure of the neural network model used in the experiment is shown in Figure 3, and the 5 evaluation indicators are shown in Equation 3. For the first four indicators, the higher the value, the better, and the last one, the lower the value, the better. For the reliability of comparison, all hyper parameters such as batch size were made the same. We used one-way ANOVA and post-hoc tests using the results.

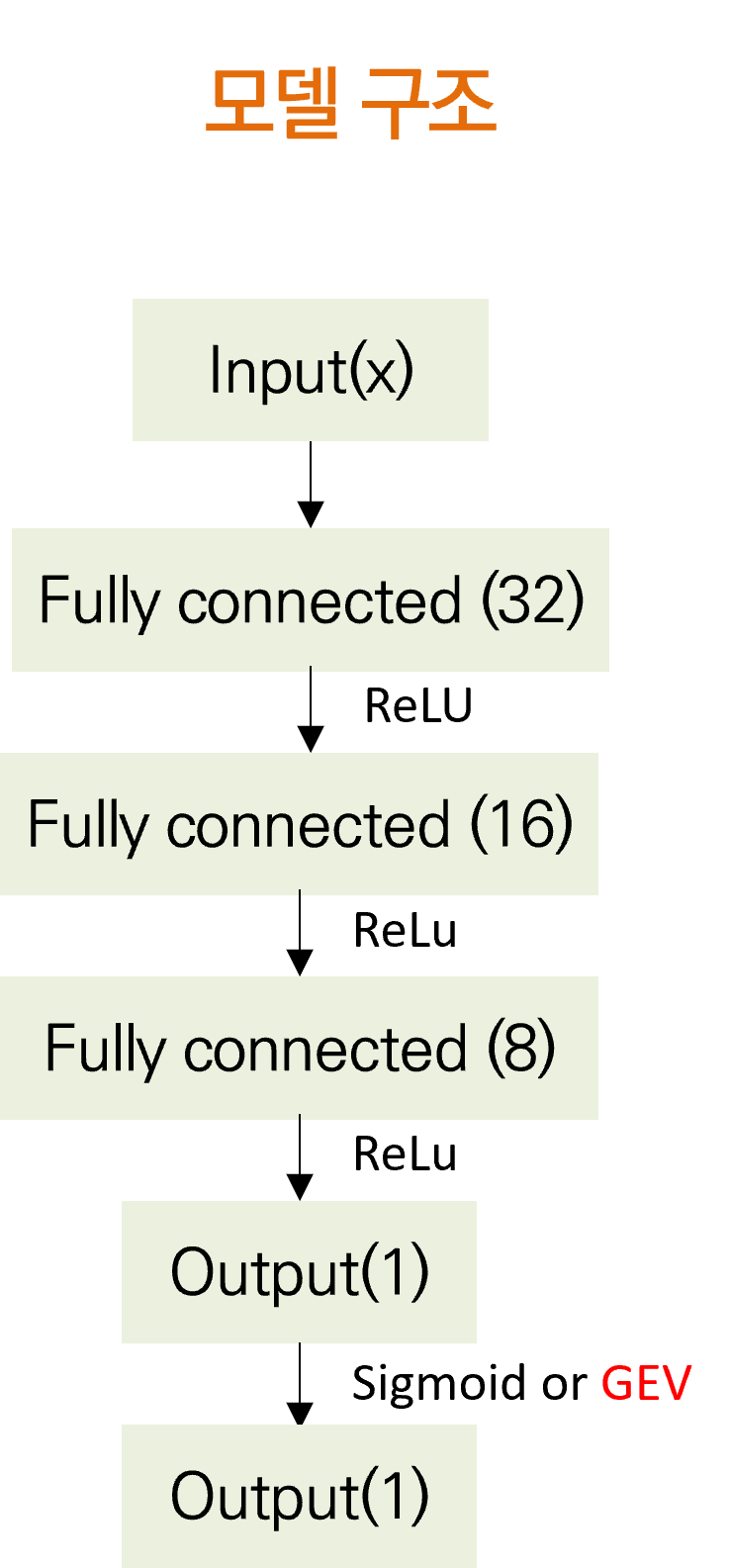


Figure 3: Structure of neural network for this study

|  |  |
| --- | --- |
| F1-score =  Geometric-Mean =  Balanced Accuracy = (TPR + TNR)  Area Under the ROC Curve (AUC)  Brier Inaccuracy = | Equation 3 |

1. Results

A summary of the experimental results is shown in Table 2. If there’s significant differences of metrics under 5%, we counted the number of better results when we compared method 1 and 5. Better values are marked in red for better visibility. It’s never been nice if the GEV activation function was used alone. In particular, it is interesting to note that as you move down the table, the results look better.

**Table 2:** Example of experiment result (Data : abalone19)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Data*** | ***method*** | ***F1-score*** | ***Geometric-Mean*** | ***Area Under Curve*** | ***Balanced Accuracy*** | ***Brier Inaccuracy*** |
|  | (1) | 0.0  (0.0) | 0.0  (0.0) | 0.794  (0.084) | 0.5  (0.0) | 0.016  (0.0) |
|  | (2) | 0.0  (0.0) | 0.0  (0.0) | 0.659  (0.143) | 0.5  (0.0) | 0.019  (0.013) |
| abalone19 | (3) | 0.033  (0.021) | 0.662  (0.16) | 0.659  (0.143) | 0.687  (0.105) | 0.019  (0.013) |
|  | (4) | 0.045  (0.023) | 0.733  (0.165) | 0.760  (0.119) | 0.757  (0.095) | 0.037  (0.012) |
|  | (5) | 0.044  (0.015) | 0.762  (0.057) | 0.781  (0.068) | 0.770  (0.056) | 0.039  (0.008) |

**Table 3: Comparing with SOTA model (GEV-NN)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Data*** | ***Geometric-***  ***Mean***  ***(Proposed)*** | ***Geometric-***  ***Mean***  ***(GEV-NN)*** | ***Area Under the ROC Curve***  ***(Proposed)*** | ***Area Under the ROC Curve***  ***(GEV-NN)*** |
| abalone19 | 0.762 | 0.7247 | 0.781 | 0.7419 |
| abalone20\_vs\_8-9-10 | 0.908 | 0.884 | 0.935 | 0.9009 |
| kr-vs-k-zero\_vs\_fifteen | 1 | 1 | 1 | 1 |
| pocker-8\_vs\_6 | 0.998 | 0.9714 | 0.999 | 0.966 |
| pocker-8-9\_vs\_5 | 0.752 | 0.5165 | 0.719 | 0.408 |

1. Summary and Discussion

In this experiment, a toy model experiment using the KEEL imbalance dataset was conducted, but it needs to be applied to real-world data set such as rainfall, financial data, etc. Since the superiority of the model is shown differently depending on the evaluation index, a comparison method that considers the characteristics of the model and data is needed. Better results can be expected if the two hyperparameters for Focal Loss are adjusted. The optimal ratio of Over-Sampling is about 20:1, but it is difficult to apply OS depending on the sample size or imbalance ratio.

In addition, it is currently applied only to the case of binary classification, but it can be extended to multi-class classification in the future.

**Acknowledgments**

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT)(No.2020R1I1A3069260) and BK21 FOUR (Fostering Outstanding Universities for Research, NO.5120200913674) funded by the Ministry of Education (MOE, Korea) and National Research Foundation of Korea (NRF).

References

1. Wang, X., Dey, D.K.: Generalized extreme value regression for binary response data: An application to B2B electronic payments system adoption. Ann. Appl. Stat. **4(4)**, 2000--2023 (2010) doi: 10.1214/10-AOAS354
2. Munkhdalai, L., Munkhdalai, T., Ryu, K.H.: GEV-NN: A deep neural network architecture for class imbalance problem in binary classification. Knowledge-Based Systems. **194**, 105534 (2020) doi: 10.1016/j.knosys.2020.105534
3. J. Bridge, et al.: Introducing the GEV Activation Function for Highly Unbalanced Data to Develop

COVID-19 Diagnostic Models. IEEE Journal of Biomedical and Health Informatics. **24(10)**,

2776--2786 (2020) doi: 10.1109/JBHI.2020.3012383

1. Johnson, J.M., Khoshgoftaar, T.M.: Survey on deep learning with class imbalance. Journal of Big Data. **6(27)**, (2019) doi: 10.1186/s40537-019-0192-5

1. Hyebin Park, Department of Mathematics and Statistics, Chonnam National University, Gwangju 61186, Korea; email: central\_\_@naver.com

   Juyoung Hong, Department of Mathematics and Statistics, Chonnam National University, Gwangju 61186, Korea; email: hjy\_stat@naver.com

   Yonggwan Shin, Department of Mathematics and Statistics, Chonnam National University, Gwangju 61186, Korea; email: syg.stat@gmail.com

   Jeong-Soo Park, Department of Mathematics and Statistics, Chonnam National University, Gwangju 61186, Korea; email: jspark@jnu.ac.kr [↑](#footnote-ref-1)