<English Version>

Enhancing predictive accuracy for a minority class in imbalanced data: An integrated approach with ROSE and Tomek link[1)]

Jeong-Wook Lee[2)] · Young Eun Jeon[3)] · Jung-In Seo[3)]

Summary: This study addresses the issue of class imbalance, which commonly occurs in scenarios where the event of interest occurs far less frequently than other events. Machine learning models trained on such imbalanced data often exhibit bias toward the majority class, resulting in high overall accuracy but poor detection of minority class instances. This imbalance is problematic, especially when focusing on the minority class as the models can lead to significant misclassification, which undermines the reliability and validity of predictions. To resolve this problem, we propose an integrated sampling technique with random over-sampling examples and Tomek link. This method enhances data diversity by generating synthetic instances based on a probability distribution while simultaneously removing noisy instances, thereby improving the quality of the data. As a practical example, we use a stroke dataset with a serious imbalance ratio of 98:2 and demonstrate the validity and applicability of the proposed method by performing predictions using machine learning methods including support vector machine, elastic net regression, random forest, and extreme gradient boosting, as well as deep learning methods including deep and convolutional neural networks. Our findings reveal that the proposed method can significantly improve the prediction of minority class instances, leading to more reliable and efficient predictive modeling in critical healthcare applications.

Keywords: Imbalanced data, Machine learning, Random over-sampling examples, Tomek link

---

1) This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A8063350).
2) Institute of Data Science and Informatics, University of Missouri, Columbia, USA
3) Department of Data Science, Andong National University, Andong, Korea