

A Selective Under-Sampling (SUS) Method for Imbalanced Regression (Abstract Reprint)

Jovana Aleksic^{1,2}, Miguel Garca-Remesal³

¹Universidad Politecnica de Madrid, Spain

²Weill Cornell Medicine in Qatar, Qatar

³Departamento de Inteligencia Artificial, Universidad Politecnica de Madrid, Spain

Abstract Reprint. This is an abstract reprint of the journal article by Aleksic and Garca-Remesal (2025).

Abstract

Many mainstream machine learning approaches, such as neural networks, are not well suited to work with imbalanced data. Yet, this problem is frequently present in many real-world data sets. Collection methods are imperfect, and often not able to capture enough data in a specific range of the target variable. Furthermore, in certain tasks data is inherently imbalanced with many more normal events than edge cases. This problem is well studied within the classification context. However, only several methods have been proposed to deal with regression tasks. In addition, the proposed methods often not yield good performance with high-dimensional data, while imbalanced high-dimensional regression has scarcely been explored. In this paper we present a selective under-sampling (SUS) algorithm for dealing with imbalanced regression and its iterative version SUSiter. We assessed this method on 15 regression data sets from different imbalanced domains, 5 synthetic high-dimensional imbalanced data sets and 2 more complex imbalanced age estimation image data sets. Our results suggest that SUS and SUSiter typically outperform other state-of-the-art techniques like SMOGN, or random under-sampling, when used with neural networks as learners.

References

Aleksic, J.; and Garca-Remesal, M. 2025. A Selective Under-Sampling (SUS) Method for Imbalanced Regression. *Journal of Artificial Intelligence Research*, 82: 111–136.