



Universiteit
Leiden

The Netherlands

Learning class-imbalanced problems from the perspective of data intrinsic characteristics

Kong, J.

Citation

Kong, J. (2023, September 27). *Learning class-imbalanced problems from the perspective of data intrinsic characteristics*. Retrieved from <https://hdl.handle.net/1887/3642254>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3642254>

Note: To cite this publication please use the final published version (if applicable).

Propositions

Accompanying the thesis

Learning Class-Imbalanced Problems from the Perspective of Data Intrinsic Characteristics

by Jiawen Kong

1. Most machine learning classification algorithms assume that (1) the classes are equally distributed; and (2) the costs of classification errors are equal. However, both assumptions do not always hold when dealing with class-imbalance problems.
2. Class imbalance is not the unique factor hindering the classification performance degradation; the data intrinsic characteristics, such as *feature overlapping* and *separability* are also main contributors to the difficulty of a supervised classification problem.
3. Hyperparameter optimisation for both classification algorithms and resampling approaches, although consumes more time, can significantly improve the performance. The trade-off between time consumption and the expected gain should be considered.
4. The anomaly detection problem can be considered as an extreme case of class imbalance problem with an extreme imbalance in terms of class distribution.
5. If one is dealing with a class-imbalance problem, the first thing he/she should try is to check if it is possible to get more data samples; applying the imbalanced-handling techniques should be the second step.
6. As recent advanced technology enables practitioners from industry and engineering to collect a large amount of data with the purpose of extracting knowledge and acquiring hidden insights. This emphasizes the importance of understanding the data.
7. As most approaches in the imbalanced learning domain are tested on benchmark datasets, fine adjustments are needed when applying these techniques to real-world application problems.
8. The core concept of solving ‘imbalanced’ problem is to use ‘balancing’ idea.
9. Imbalanced problems are everywhere, also in everyone’s life. We are using our own techniques to balance our life.

Jiawen Kong
Leiden, September 27, 2023