Optimising Machine Learning Technique for Irregular Sampling

Zhenyu Xu
University of Warwick, United Kingdom

Negar Riazifar
University of Warwick, United Kingdom

Abstract:

Optimising the model to lower the influence of irregular data produced by real-world business cases would help to increase the accuracy of machine learning and thereby support the operations of companies. While interpolation methods are widely used for data preprocessing, their impact on model accuracy in irregular datasets remains underexplored. This study addresses this gap by comparing the effects of Simple Linear Interpolation (SLI) and Natural-Neighbour Interpolation (NNI) on predictive performance using Tesla stock and Seoul bike-sharing data.

The methodology involves preprocessing datasets, applying SLI and NNI to handle irregularities, and evaluating the results through machine learning models, including XGBoost, Random Forest, and a Stacking model. Key methodologies are supported by Voronoi diagrams for NNI [1], stepwise feature selection methods [2], the importance of hyperparameter optimisation in model tuning [3].

Results indicate SLI consistently improves model performance (Figure 1), particularly in complex models like Stacking, while NNI shows mixed results, performing well on stable features like temperature but struggling with volatile data. In conclusion, SLI is more stable for model improving, especially in XGBoost, random forest and stacking model. These findings help to find the best model and data preprocessing strategies for irregular data, advancing predictive analytics in irregular datasets.

Keywords:

Machine Learning, Irregular Data Sampling, Linear Interpolation, Natural-Neighbour Interpolation.