Cover Letter

Dear Editor,

Hello. This paper was jointly completed by the teachers and students in our research group. Our research group mainly focuses on the fields of data mining and mathematical statistics.

Synthetic data presents an effective solution to the challenges of inadequate or low-quality sample, particularly in the era of big data. For some actual tasks, original data often contains a multitude of features and unknown noise. Since synthetic method involves generating new data based on existing data, the quality of the new points depends on the quality and quantity of the original data. If the quality of the original data is poor or the quantity is insufficient, then the synthetic data may have limitations in terms of its quality. However, there exists a lack of effective synthetic methods for datasets with a restricted size and complex noise to expand the size of data set. Based on the purpose of improving the quality and quantity of the dataset, and motivated by piecewise linear interpolation and spline interpolation, we propose a robust and stable data synthesis method named Adaptive Subspace Interpolation for Sample Optimization (ASISO), which aims to adaptively adjust the sample size and structure of the original dataset containing unknown noise. The idea is to divide the original feature space into several subspaces with an equal number of samples, and then perform linear interpolation for the samples in the adjacent subspaces. This method achieves sample optimization in two aspects. First, it can adaptively adjust the size of the dataset, and the expanded data typically contains minimal error with actual. Second, it adjusts the structure of the samples, which can significantly reduce the proportion of samples with large errors, thereby minimizing the impact of noise for model generalization.

As we know, Computational Statistics and Data Analysis is a bi-monthly refereed journal that publishes papers covering the interface between the statistical and computing sciences. Our method includes a significant amount of statistical thinking and can be applied to the fields of data mining and machine learning. Moreover, we have provided theoretical proofs for the effectiveness of our method.

In conclusion, we believe that this manuscript meets the requirements for inclusion in Computational Statistics and Data Analysis.

Best regards.