VI International Conference on Mathematics and its Applications in Science and Engineering (ICMASE 2025)

ICMASE 2025

DIFFERENTIAL OPERATOR-BASED IMPUTATION: A ROBUST METHOD FOR INCOMPLETE DATA ANALYSIS

Gokceana Karadag[1,*], Burcu Tunga[1], Ahmet Kiris[1]

[1]*Mathematics Engineering Department, Istanbul Technical University, 34469, Maslak, Istanbul, Türkiye*

ABSTRACT

When only some data points are available in a dataset, imputation of the approximate values at unknown points often becomes essential to maintain the integrity and usability of the data. This need becomes especially critical in real-world applications where unobserved or missing data are prevalent and can severely impact the performance of statistical analyses and machine learning models. To address this issue, this study introduces a novel imputation technique that reconstructs missing data points using only the known entries, without requiring any prior assumption about the underlying data distribution or functional form.

The method is based on three fundamental equations derived using approximate differential operators. These equations consist of the central difference and the three-point forward and backward approximate differential operators. The value at an unknown point is computed using these equations by incorporating the neighboring data points to its left and right.

To rigorously assess the accuracy, stability, and generalization capacity of the proposed approach, synthetic datasets are generated from various analytical functions that exhibit different behaviors within a fixed domain. These datasets simulate distinct imputation scenarios under controlled conditions, providing a robust testing environment. Comparative analyses with state-of-the-art imputation methods are performed using standard error metrics, including Mean Absolute Error (MAE), Mean Relative Error (MRE), and Root Mean Square Error (RMSE).

Experimental results show that the proposed method consistently delivers high-precision imputations, even when the proportion of missing data is considerable. The local nature of the computation, which is based solely on the nearest neighbors, enhances both efficiency and interpretability. Furthermore, the use of approximate differential operators contributes to the numerical reliability and transparency of the method, making it particularly well suited for critical applications involving incomplete or noisy data.

Keywords Missing Data Imputation · Approximate Differential Operators · Local Estimation

References

[1] Batista G.E.A.P.A., and Monard M.C., An analysis of four missing data treatment methods for supervised learning, Applied Artificial Intelligence, 17(5-6): 519–533, 2003.

[2] Burden R.L., Faires J.D., and Burden A.M., Numerical Analysis, 10th Edition, Cengage Learning, 2015.

[3] Gander W., Gander M.J., and Kwok F., Scientific Computing - An Introduction using Maple and MATLAB, 1st Edition, Springer Cham, 2014.

[4] Hastie T., Tibshirani R., and Friedman J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition, Springer New York, NY, 2009.

[*]*Corresponding Author's E-mail: karadagg23@itu.edu.tr*

[5] Little R.J.A., and Rubin D.B., Statistical Analysis with Missing Data, 3rd Edition, Wiley, 2019.

[6] Schafer J.L., and Graham J.W., Missing data: our view of the state of the art, Psychol Methods, 7(2): 147-77, 2002.

[7] Troyanskaya O., Cantor M., Sherlock G., Brown P., Hastie T., Tibshirani R., Botstein D., and Altman R.B., Missing value estimation methods for DNA microarrays, Bioinformatics, 17(6): 520–525, 2001.

[8] Van Buuren S., Multiple imputation of discrete and continuous data by fully conditional specification, Stat Methods Med Res., 16(3): 219-42, 2007.

[9] Van Buuren S., Flexible Imputation of Missing Data, 2nd Edition, CRC Press, Taylor & Francis Group, 2018.

[10] Willmott C.J., and Matsuura K., Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Clim Res, 30: 79-82, 2005.