

Research and implementation of the Dynamic Time Warping algorithm in time series clustering problems

Lieu Xuan Nghia

Abstract: In-depth research document on the Dynamic Time Warping (DTW) algorithm and its application in clustering time series data, a common data type in fields such as meteorology and economics. The author presents the theoretical foundation of time series, methods for measuring similarity, and focuses on the DTW algorithm to address the issue of phase and speed differences between sequences. The study implements experimental clustering, performs data normalization, and evaluates effectiveness through metrics such as Silhouette. The results demonstrate DTW's ability to identify similar patterns independent of time synchronization.

Keywords: Dynamic Time Warping, Time Series Clustering, Machine Learning

References

- [1] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007.
- [2] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, 2011.
- [3] W. McKinney, "Data Structures for Statistical Computing in Python," in *Proc. 9th Python Sci. Conf.*, 2010, pp. 56–61.
- [4] C. R. Harris et al., "Array programming with NumPy," *Nature*, vol. 585, pp. 357–362, 2020.
- [5] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Stat.*, vol. 3, no. 1, pp. 1–27, 1974.
- [6] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, no. 2, pp. 224–227, 1979.
- [7] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987.