Time Series Clustering for Data-Driven Deep Learning Models in Dengue Case Forecasting

Juan Vicente Bogado Machuca

Universidad Nacional de Caaguazú, Paraguay

The accurate prediction of Dengue cases is a crucial challenge in public health, particularly in tropical and subtropical regions where outbreaks can overwhelm healthcare systems. Deep learning models, while powerful, often face limitations in epidemiological forecasting due to the complexity and variability of time series data. To address these challenges, time series clustering is introduced as both a preprocessing step and a data augmentation method to enhance the performance of deep learning models in Dengue case forecasting. By clustering time series data from regions with similar temporal dynamics, it becomes possible to extract shared patterns and trends, reduce noise, and improve the model's ability to generalize across different datasets.This approach was applied to Dengue case data from Paraguay, a small country with 281 cities reporting weekly case counts over a period of five years. The clustering process grouped cities based on the similarity of their temporal case trajectories, enabling the creation of more accurate models. These clustered time series were then used to train advanced deep learning architectures, such as Long Short-Term Memory (LSTM) networks, which are well-suited for capturing temporal dependencies. Time series clustering served as a form of data augmentation, as it allowed the model to learn from aggregated patterns that might otherwise remain hidden in individual city-level data.The results demonstrate that the integration of time series clustering significantly improves the accuracy of Dengue case predictions compared to models trained on unclustered data. In the case of Paraguay, this approach yielded robust predictions despite the relatively small size of the dataset, highlighting the effectiveness of clustering in enhancing the predictive power of deep learning models for smaller-scale epidemiological data.

MCA 2025 internal code: SS54-T08

Additionally, the clustering process provided valuable insights into the temporal dynamics of Dengue outbreaks, revealing patterns that could inform targeted public health interventions.This methodology underscores the potential of combining time series clustering with deep learning to tackle complex epidemiological forecasting problems. By leveraging the temporal structure of the data, the approach not only boosts prediction accuracy. These findings are particularly relevant for countries with limited data resources, where clustering can serve as a powerful tool to amplify the utility of existing datasets.

In Special Session 54 on Emergence, Spread, and Control of Mosquito-borne Diseases: Insights from Mathematical Modeling at the Mathematical Congress of the Americas 2025.