Classification of Base Stations Location based on Traffic Profiles

Prof. Luis M.Correia

Iram Fatima Aulakh, Kidist Beshie, S M Rakib Hasan

Técnico Lisboa | Aalto University | Universitat Politècnica de Catalunya

Motivation

- Mobile traffic varies across environments, requiring tailored resource allocation and network planning.
- The growing scale and complexity of cellular networks make manual classification impractical.
- Dynamic traffic patterns driven by urbanization, work habits, and user behavior changes demand adaptive classification frameworks.

Problem Statement

Telecom traffic data is complex, high-dimensional, and time-varying, making it challenging to profile and cluster base stations. Since traffic patterns vary with surrounding environments, the absence of automated classification methods for identifying base station types (residential, commercial, or mixed-use) hinders effective network optimization and resource allocation.

Methodology

Dataset & Preprocessing:

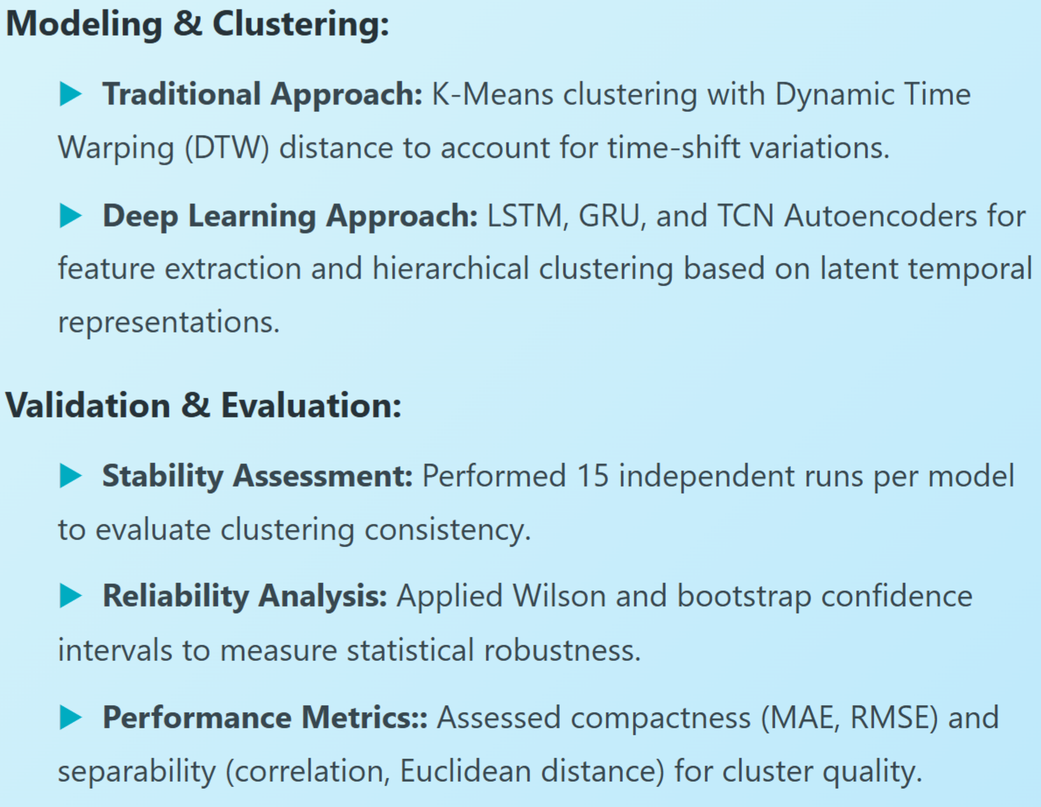
- Hourly power emission data from Portuguese mobile network base stations (24-hour profiles) were used to capture traffic behavior patterns.
- Removed low-activity hours (02:00–06:00) to reduce noise.
- Applied Min–Max normalization to ensure comparability across stations.

Modeling & Clustering:

- **Traditional Approach:** K-Means clustering with Dynamic Time Warping (DTW) distance to account for time-shift variations.
- **Deep Learning Approach:** LSTM, GRU, and TCN Autoencoders for feature extraction and hierarchical clustering based on latent temporal representations.

Validation & Evaluation:

- **Stability Assessment:** Performed 15 independent runs per model to evaluate clustering consistency.
- **Reliability Analysis:** Applied Wilson and bootstrap confidence intervals to measure statistical robustness.
- **Performance Metrics::** Assessed compactness (MAE, RMSE) and separability (correlation, Euclidean distance) for cluster quality.



Results

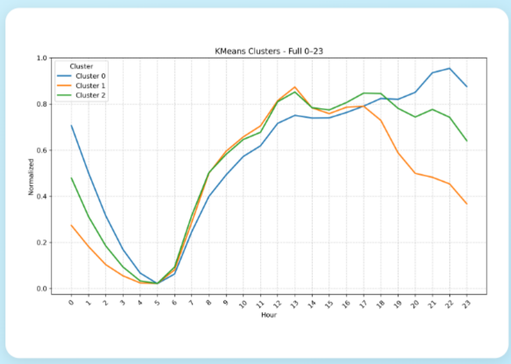
Model Performance:

- K-Means with DTW provided excellent cluster separation, making it ideal for distinguishing distinct base station categories.
- Among deep learning models (LSTM, GRU, TCN), LSTM had superior compactness, capturing nuanced temporal patterns in traffic data.
- The framework successfully classified base stations into residential, commercial, and mixed-use areas based on traffic profiles, enabling targeted network optimization strategies
- Removing inactive hours (02:00–06:00) enhanced interpretability by eliminating non-informative data segments.

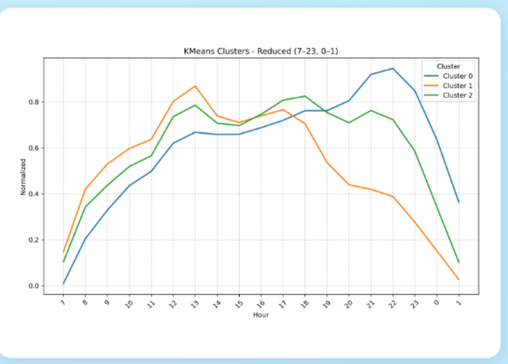
Telecom Analysis:

Each cluster was mapped to a type of area based on hourly power consumption trends:

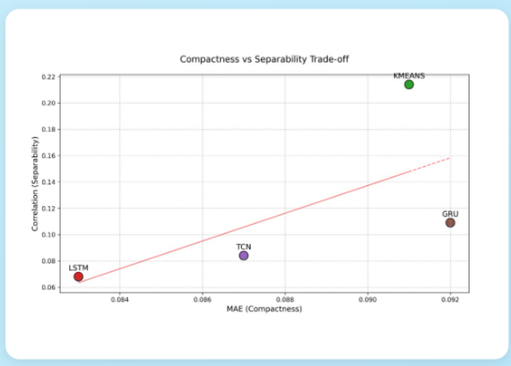
- **Residential Cluster:** Peak Hours: 19:00–23:00. High variability during late evening hours, reflecting heterogeneous lifestyles of household subscribers.
- **Commercial Cluster:** Peak Hours: 09:00–17:00. Stable daytime traffic with low variance, consistent with office and industrial activity patterns.
- **Mixed Cluster:** Moderate and balanced usage throughout the day, representing overlapping residential and commercial zones.



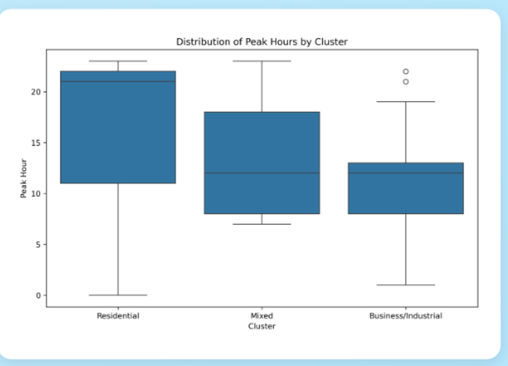
Clusters with full 24-hour transmit power profile



Clusters with reduced-hour transmit power profile



Performance Trade-off Across Models



Peak Hour Distribution by Area Type

Conclusion

- The results indicate that when interpretability and clear cluster distinction are crucial (e.g., in operational planning), traditional methods like K-Means + DTW are preferred. Conversely, for profile compression or personalized modeling (e.g., load forecasting within a known tower type), autoencoders are a strong choice. Future work may explore hybrid models to balance both aspects.
- The dataset was geographically limited to base stations in Portugal, which may not fully capture the diversity of usage patterns in other regions.
- The classification framework relies solely on transmit power data, and incorporating additional features such as user density or mobility patterns could further refine the results.