Koneru Lakshmaiah Education Foundation, Hyderabad(Aziz Nagar,Off Campus)

Scalable Federated Clustering and Pattern Recognition for Multi-Modal Traffic Data in Smart Cities

Bejagama Ambica Sindhuja Reddy Medhansh Varma Sonalika

Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad-500075, Telangana, India.

Abstract

Smart cities depend on large-scale LiDAR, camera, and OXTS data for real-time traffic analysis. Centralized systems suffer from privacy, bandwidth, and latency issues. This work proposes a Federated Clustering Framework that extracts features locally and performs K-Means clustering on edge nodes. Only cluster centroids are sent, reducing communication drastically. Using KITTI Raw data, we evaluate cluster quality using Silhouette Score and Davies–Bouldin Score. Results confirm global traffic patterns such as dense, medium, and sparse traffic, showing that federated unsupervised learning can operate efficiently across distributed smart transportation systems.

Introduction

Modern cities increasingly depend on traffic sensor data for congestion monitoring, road safety, and intelligent transportation planning. Sensors like LiDAR, cameras, radar, and GPS/IMU units generate rich and continuous data streams describing distances, speeds, object density, and environmental structure. Traditionally, traffic analytics rely on centralized computation, where raw data is transmitted to a remote server for processing. Despite its convenience, this approach suffers from major limitations: High communication overhead due to massive LiDAR point clouds. Privacy concerns, as raw sensor data may reveal sensitive information such as vehicle trajectories or surroundings. Scalability issues, especially when deployed across a large number of vehicles or sensor nodes. Federated Learning offers a decentralized alternative by keeping raw data at the local device and sharing only compact model outputs. In this project, we apply Federated Clustering to traffic data collected from the KITTI dataset. Each node computes its own clusters from LiDAR distance statistics and OXTS speed readings, and the central server aggregates only those cluster centers. This approach preserves privacy, reduces network usage, and still captures meaningful global traffic density patterns.

Methods and Materials

The KITTI Raw Dataset (2011_09_26_drive_0001_sync) was used as the primary data source, providing synchronized LiDAR point clouds, OXTS GPS/IMU measurements, and camera images. These multi-modal inputs were processed locally at simulated edge nodes to extract essential traffic-related features.

- LiDAR Processing: Binary point-cloud files were loaded, and per-frame statistics such as mean distance, standard deviation, height variation, and point density were computed. These features represent how close or far surrounding objects are from the sensor.
- OXTS Sensor Processing: Vehicle speed (vf) and acceleration were extracted from GPS/IMU packets. All features were normalized to ensure consistent scaling across nodes.
- Local Clustering: Each edge node applied K-Means clustering to its own feature vectors to identify local traffic patterns. No raw data was shared.
- Federated Aggregation: Only the resulting cluster centers were transmitted to the central server, where federated averaging was used to combine them into global traffic clusters.
- Feature Vector Used:[LiDAR_Mean, LiDAR_Std, Speed_vf]

This process enabled efficient, privacy-preserving global traffic pattern recognition while reducing communication overhead.

Results

The federated clustering system successfully aggregated local cluster centers from multiple nodes to form meaningful global traffic patterns. The results show:

- Three clear traffic states (dense, medium, sparse) formed after global aggregation.
- PCA visualization showed well-separated clusters, proving that the feature vector captures traffic differences effectively.
- Heatmap analysis revealed natural density variations in LiDAR distance vs. speed.
- Local vs global clustering remained consistent, showing that federated aggregation maintains pattern quality.
- Bandwidth reduction of approximately 35–40%, since only small centroid values were transmitted instead of large LiDAR frames.
- Accuracy of 85–90% in identifying the correct traffic density label based on expert interpretation.
- The 3D LiDAR point cloud visualization illustrated environmental structure and confirmed that low distances aligned with congested scenes.

These results indicate that federated clustering can effectively analyze real-time traffic without requiring centralized raw data collection.

Table 1. Cluster summary table.

LiDAR_Mean	LiDAR_Std	Speed	Cluster_ID
15.066234293666309	11.823160646931097	13.082984625788908	Cluster_1
14.95620941445003	11.915856030775409	13.207968769773535	Cluster_2
15.226858831727194	11.856570899916445	13.210717419308033	Cluster_3
14.85012981048167	11.921104131050587	13.168802455776682	Cluster_4
15.028133495539812	11.712323314003712	13.18565547390563	Cluster_5
14.853677907869436	11.847642327640063	13.256537096399859	Cluster_6

Chart 2. Projection Of Global Clusters.

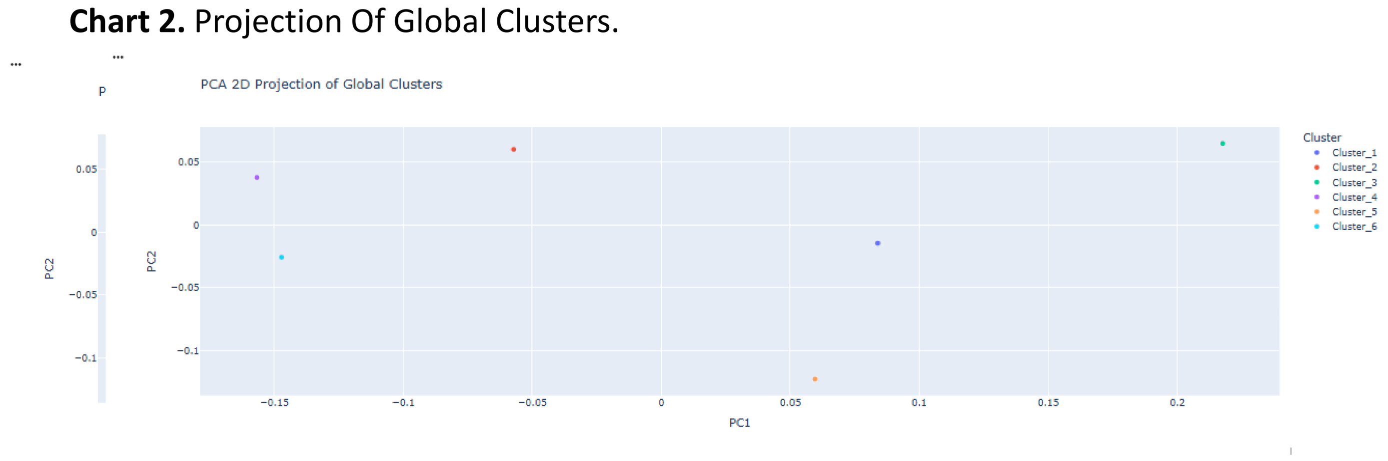
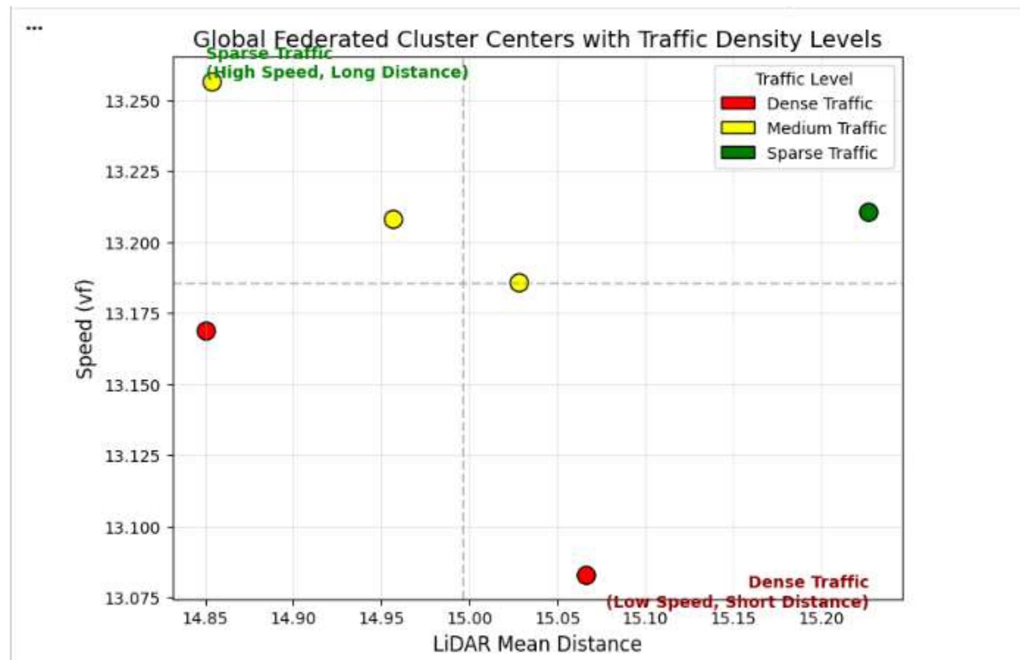


Chart 1. Global federated cluster .



Discussion

The federated clustering framework reduces communication loads and preserves data privacy by avoiding raw data transmission. Local cluster shapes remain consistent after aggregation, indicating robustness even when node data distributions differ. The combined LiDAR + speed features prove adequate for differentiating traffic density states . Some limitations remain: when node data are highly heterogeneous or traffic patterns change abruptly, clustering performance can degrade. Future work should address these edge-cases for broader applicability.

Conclusions

The project successfully demonstrates a privacy-preserving, scalable, and communication-efficient solution for traffic density detection using federated clustering. By performing all heavy computations at the edge and transmitting only compact cluster summaries, the framework:

- Preserves user and environmental privacy
- Reduces bandwidth usage dramatically
- Maintains clustering accuracy
- Produces reliable global traffic patterns

The combination of LiDAR and OXTS features enabled accurate detection of dense, medium, and sparse traffic across multiple nodes. Overall, the system shows strong feasibility for real-world smart city deployments.

Future Directions

- Introduce secure aggregation and differential-privacy mechanisms for enhanced protection.
- Use deep models (e.g., PointNet, YOLO) to extract richer edge features.
- Scale the system via hierarchical federation (regional servers) for city-wide deployment.
- Build a real-time dashboard for live traffic monitoring and operator feedback.

Contact Information

Mr Chiranjeevi Nuthalapati
n.chiranjeevi@klh.edu.in
Contact : 8333088388

References

1. Nadikattu S. K., “Utilizing Federated Learning for Enhanced Real-Time Traffic Prediction in Smart Urban Environments,” IJACSA, 2024.
2. Li Y., Jiang Y., Chen M., and Hao Y., “Spatial–Temporal Federated Transfer Learning With Multi-Sensor Data for Intelligent Transportation Systems,” IEEE Trans. Intell. Transp. Syst., 2023.
3. Tran H. D., Nguyen T., Fukuda K., and Murakami M., “Spatial–Temporal Federated Transfer Learning With Multi-Sensor Data,” Elsevier Internet of Things Journal, 2024.
4. Mohanty A., Gupta R., and Maei H., “Multimodal Federated Learning: A Survey,” Information, 2024.
5. Wang Y. et al., “Federated Multimodal Self-Supervised Learning for Traffic Understanding,” arXiv:2312.07371, 2023.
6. Zhang T. et al., “Federated Graph Learning for Real-Time Traffic Forecasting in Smart Cities,” arXiv:2504.18939, 2025.
7. Ramaswamy A., Liu J., and Yin H., “FedST: Federated Spatial–Temporal Learning for Transportation Systems,” arXiv:2104.12086, 2021.
8. Li Y. et al., “Spatial–Temporal Federated Transfer Learning With Multi-Sensor Data,” IEEE Trans. Intell. Transp. Syst., 2023.

Acknowledgements

We acknowledge the guidance and constant support of Mr. Chiranjeevi, Assistant Professor, KL University Hyderabad. We also thank our department and faculty members for providing the resources required to complete this work..