Federated Clustering for Smart Traffic Management Using Real-Time Sensor Data

1st Bejagama Ambica
Computer Science and Engineering
KL University Hyderabad
Telangana, India
2210030039@klh.edu.in

2nd Sindhuja
Computer Science and Engineering
KL University Hyderabad
Telangana, India
2210030089@klh.edu.in

3rd Medhansh
Computer Science and Engineering
KL University Hyderabad
Telangana, India
2210030119@klh.edu.in

4th Sonalika
Computer Science and Engineering
KL University Hyderabad
Telangana, India
2210030127@klh.edu.in

Abstract

Smart traffic management is essential for reducing congestion, improving safety, and optimizing road usage. Traditional centralized systems require collecting raw vehicle data, which raises privacy and communication overhead concerns. This work proposes a privacy-preserving traffic analysis model using **Federated Clustering**, where each vehicle or node performs local clustering on LiDAR and speed data, and only cluster centers are shared with a central server. The global cluster centers provide meaningful traffic density insights without exposing raw data. Experimental results show that the proposed federated aggregation technique effectively detects traffic patterns while ensuring privacy and reducing data transfer overhead.

Keywords: Federated Learning, Clustering, Smart Traffic Management, LiDAR Data, Privacy-Preserving Analytics

1 Introduction

Traffic congestion in modern cities leads to increased travel time, fuel consumption, and accident rates. Traditional traffic monitoring approaches rely on centralized servers that collect raw sensor data from vehicles or roadside units. However, this raises issues such as data privacy, system load, and communication costs.

Federated learning provides a decentralized solution by enabling local computation and global aggregation without sharing raw data. In this work, we design a federated clustering model where vehicle nodes cluster their LiDAR distance measurements and speed data locally. A global server aggregates the local cluster centers to derive overall traffic density patterns.

2 Literature Review

Recent works emphasize decentralized learning in autonomous systems and intelligent transportation. Key findings include:

- Studies on federated learning demonstrate privacy-preserving capabilities for vehicular networks.
- Clustering algorithms such as K-Means and hierarchical clustering have been widely used for traffic density estimation.
- Ensemble and hybrid clustering approaches improve congestion interpretation in dynamic road environments.

However, limited research combines federated learning with real-time clustering for traffic analytics, motivating the proposed work.

3 Problem Definition

The goal of this work is:

- To develop a traffic density detection model without sharing raw sensor data.
- To apply clustering on local LiDAR and speed datasets at individual nodes.
- To aggregate cluster centers globally using a federated averaging mechanism.

This approach ensures privacy, reduces bandwidth usage, and improves scalability.

4 Methodology / Proposed System

The proposed system introduces a decentralized, privacy-preserving approach for analyzing real-time traffic conditions by applying federated clustering across multiple vehicular nodes. Instead of transmitting raw sensor data such as LiDAR distance readings and vehicle speeds to a centralized server, each node independently performs clustering on its local dataset. Only the computed cluster centers are shared with the federated server, ensuring that sensitive road or driver information remains private. The overall methodology consists of four major components: data collection, local clustering, global federated aggregation, and traffic density interpretation.

Sindhu KL, [14-11-2025 17:34]

4.1 System Architecture

The system architecture is composed of three main layers: local vehicle nodes, a federated server, and the visualization module. Each vehicle acts as an autonomous computational unit equipped with LiDAR sensors and speedometers. These nodes capture real-time traffic parameters including object distance, obstacle presence, and instantaneous speed. The architecture follows a cyclical process in which nodes collect data, perform clustering, and transmit only processed outputs to the server.

The federated server aggregates local model outputs to generate a global understanding of traffic flow. This hierarchical design ensures scalability, as the computational burden is distributed across participating nodes rather than relying on a single central system.

4.2 Data Collection and Preprocessing

For the purpose of developing and testing the system, synthetic datasets were generated to represent LiDAR sensor readings and vehicle speed profiles. Each dataset approximately simulates the conditions experienced by a moving vehicle, including varying distances between vehicles and fluctuating speeds depending on traffic density.

Before applying clustering, the data undergoes preprocessing steps:

- **Normalization:** All speed and distance values are normalized to ensure uniform scaling.
- **Noise Reduction:** Minor fluctuations in LiDAR data are smoothed to prevent misleading cluster formation.
- **Outlier Removal:** Abnormal readings caused by sensor misalignment or unexpected obstacles are filtered out.

These preprocessing activities help improve clustering quality and stability.

4.3 Local Clustering at Vehicle Nodes

Each participating vehicle node performs clustering locally using the K-Means algorithm. The choice of K-Means is justified by its computational efficiency, ease of implementation, and suitability for numerical sensor data. Every node applies K-Means on its normalized dataset, resulting in a set of local cluster centers that represent distinct traffic patterns observed by that vehicle.

Formally, given a dataset D_i at node i:

$$C_{local,i} = KMeans(D_i)$$

These cluster centers encapsulate essential patterns, such as high-speed movement, moderate traffic flow, or dense congestion scenarios.

4.4 Federated Aggregation

Once each node completes the local clustering step, the cluster centers are transmitted to a federated aggregation server. No raw sensor readings leave the node, thereby maintaining full data privacy.

The server combines the cluster centers from all participating nodes to compute global cluster centers using a modified Federated Averaging (FedAvg) technique:

$$C_{global} = \frac{1}{N} \sum_{i=1}^{N} C_{local,i}$$

This aggregation process effectively merges local insights into a unified, global representation of traffic conditions. Because only a small set of parameters (cluster centers) are communicated, network bandwidth usage remains minimal, enabling timely updates even in large-scale deployments.

4.5 Traffic Density Interpretation

The final stage of the system involves interpreting the global cluster centers to infer overall traffic density. The clusters are visually analyzed based on their speed and LiDAR distance characteristics.

A rule-based interpretation mechanism is applied:

- **Red Cluster: Dense Traffic** — Very short distances combined with low speeds indicate heavy congestion.
- **Yellow Cluster: Medium Traffic** — Moderate distances and speeds represent average traffic conditions.
- **Green Cluster: Sparse Traffic** — Longer distances and higher speeds suggest smooth traffic flow.

These interpretations are visualized using scatter plots, allowing analysts to compare local and global traffic patterns. The system not only detects density variations but also highlights transitions between traffic states, enabling early congestion detection.

Sindhu KL, [14-11-2025 17:34]

4.6 Workflow Summary

The complete workflow can be summarized as:

$$DataCollection \rightarrow LocalClustering \rightarrow FederatedAggregation \rightarrow GlobalTrafficInterpretation$$

This methodology ensures privacy, scalability, and efficient real-time traffic analysis, making it suitable for deployment in intelligent transportation systems.

5 Implementation

The system was implemented using:

- Python for data processing
- NumPy and Scikit-learn for K-Means and aggregation

- Matplotlib for visualizing global cluster centers

A three-week development cycle included:

1. Methodology design and literature review
2. Architecture creation and partial implementation
3. Full implementation, visualization, and documentation

6 Results and Discussion

The federated clustering model successfully aggregated cluster centers from multiple nodes. Visualizations clearly highlighted traffic density differences across regions.

Key observations:

- Federated aggregation reduced data transfer by sharing only cluster centers.
- Traffic density patterns remained consistent before and after aggregation.
- Privacy was preserved as no raw LiDAR or speed data was transmitted.

7 Conclusion

The proposed federated clustering system provides an efficient, privacy-preserving method for analyzing traffic conditions using distributed sensor data. It demonstrates scalability and strong potential for real-time intelligent transportation systems.

8 Future Work

Possible extensions include:

- Integration with edge computing for faster updates
- Incorporating deep learning for advanced traffic predictions
- Developing a mobile or cloud-based dashboard for real-time monitoring

References

[1] S. K. Nadikattu, "Utilizing Federated Learning for Enhanced Real-Time Traffic Prediction in Smart Urban Environments," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 2, pp. 538–547, 2024.

[2] Y. Li, Y. Jiang, M. Chen, and Y. Hao, "Spatial–Temporal Federated Transfer Learning With Multi-Sensor Data for Intelligent Transportation Systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 4, pp. 4452–4464, 2023.

[3] H. D. Tran, T. Nguyen, K. Fukuda, and M. Murakami, "Spatial–Temporal Federated Transfer Learning With Multi-Sensor Data," *Elsevier Internet of Things Journal*, 2024.

[4] A. Mohanty, R. Gupta, and H. Maei, "Multimodal Federated Learning: A Survey," *Information*, vol. 15, no. 9, p. 550, 2024.

[5] Y. Wang et al., "Federated Multimodal Self-Supervised Learning for Traffic Understanding," arXiv:2312.07371, 2023.

[6] T. Zhang et al., "Federated Graph Learning for Real-Time Traffic Forecasting in Smart Cities," arXiv:2504.18939, 2025.

[7] A. Ramaswamy, J. Liu, and H. Yin, "FedST: Federated Spatial–Temporal Learning for Transportation Systems," arXiv:2104.12086, 2021.

[8] Y. Li et al., "Spatial–Temporal Federated Transfer Learning With Multi-Sensor Data," *IEEE Trans. Intell. Transp. Syst.*, 2023.