

Federated Learning-based Similarity Computation over Trajectory Data Federation

**Kaining ZHANG, Lirui ZHANG, Yi XU, Yuxiang WANG,
Yongxin TONG**

Frontiers of Computer Science, DOI: [10.1007/s11704-025-41349-6](https://doi.org/10.1007/s11704-025-41349-6)

Problems & Ideas

- Problems of conventional centralized approaches:
 - Existing studies have focused on trajectory similarity computation in centralized data storage scenarios. In real-world applications, trajectories are typically held by multiple distinct entities that form a data federation. Due to constraints such as commercial confidentiality and legal regulations, direct data sharing is often prohibited, which makes traditional centralized methods unsuitable for trajectory similarity computation in such federated environments.
- Ideas:
 - The framework trains a federated embedding model across data owners to encode trajectories into vectors. A federated negative sample sharing mechanism is used to address data heterogeneity. Final trajectory similarity is computed via vector distance.

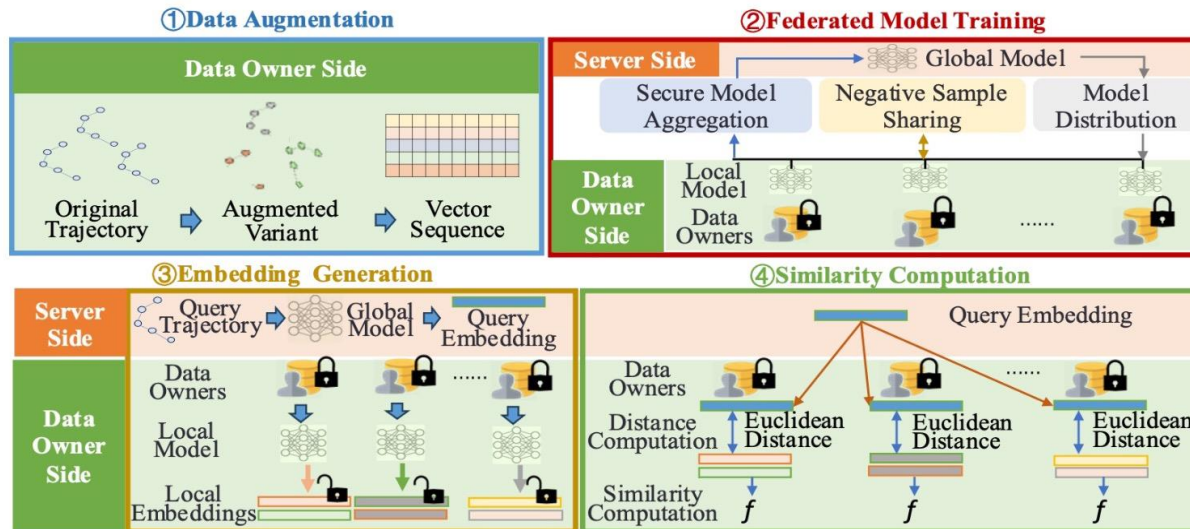
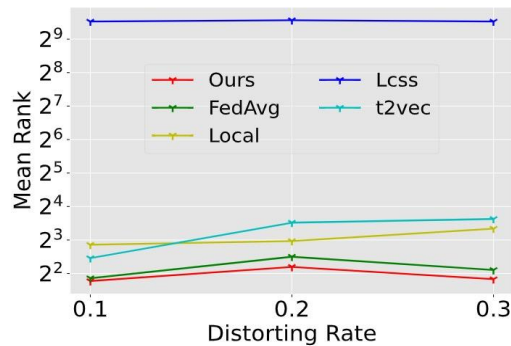


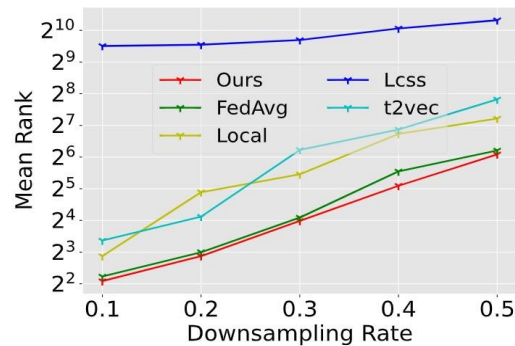
Fig. 1 The Framework of Our Method

Main Contributions

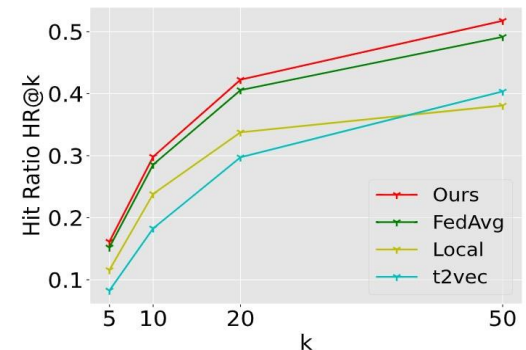
- Contributions:
 - We define the problem of similarity computation over trajectory data federation;
 - We propose a federated trajectory similarity computation framework based on federated learning, aiming to enable trajectory similarity analysis across data owners while ensuring the privacy and security of each party's data;
 - Experiments on real datasets show that our method outperforms baselines in similar trajectory search and k NN query.



(a) Results When Varying Distorting Rates



(b) Results When Varying Downsampling Rates



(c) Hit Ratio When Varying k

Fig. 2 Experimental Results of Similarity Computation over Trajectory Data Federation

We evaluated our method against four methods: LCSS and t2vec, which extend open-source centralized methods (non-learning LCSS and learning-based t2vec) to federated settings; Local, an ablation variant training model on single-client data; and FedAvg, another ablation variant excluding the negative sample sharing mechanism.