Ran TIAN, Pulun GAO, Yanxing LIU, 2024. A privacy-preserving vehicle trajectory clustering framework. *Frontiers of Information Technology & Electronic Engineering*, 25(7):988-1002.
https://doi.org/10.1631/FITEE.2300369

A privacy-preserving vehicle trajectory clustering framework

Key words: Privacy protection; Variational autoencoder; Improved *K*-means; Vehicle trajectory clustering

Corresponding author: Ran TIAN
E-mail: tianran@nwnu.edu.cn
ORCID: https://orcid.org/0000-0003-4435-580X

Motivation

- Due to the privacy concerns and data security risks, traditional vehicle trajectory clustering methods that upload original data to the server are prone to privacy leakage.
- When implementing vehicle trajectory clustering analysis, the model architecture of traditional methods does not adequately address the privacy protection needs of vehicle trajectory data.
- Therefore, in order to improve the privacy protection and clustering quality of vehicle trajectory data, it is necessary to develop a novel framework that can perform vehicle trajectory clustering while safeguarding user privacy.

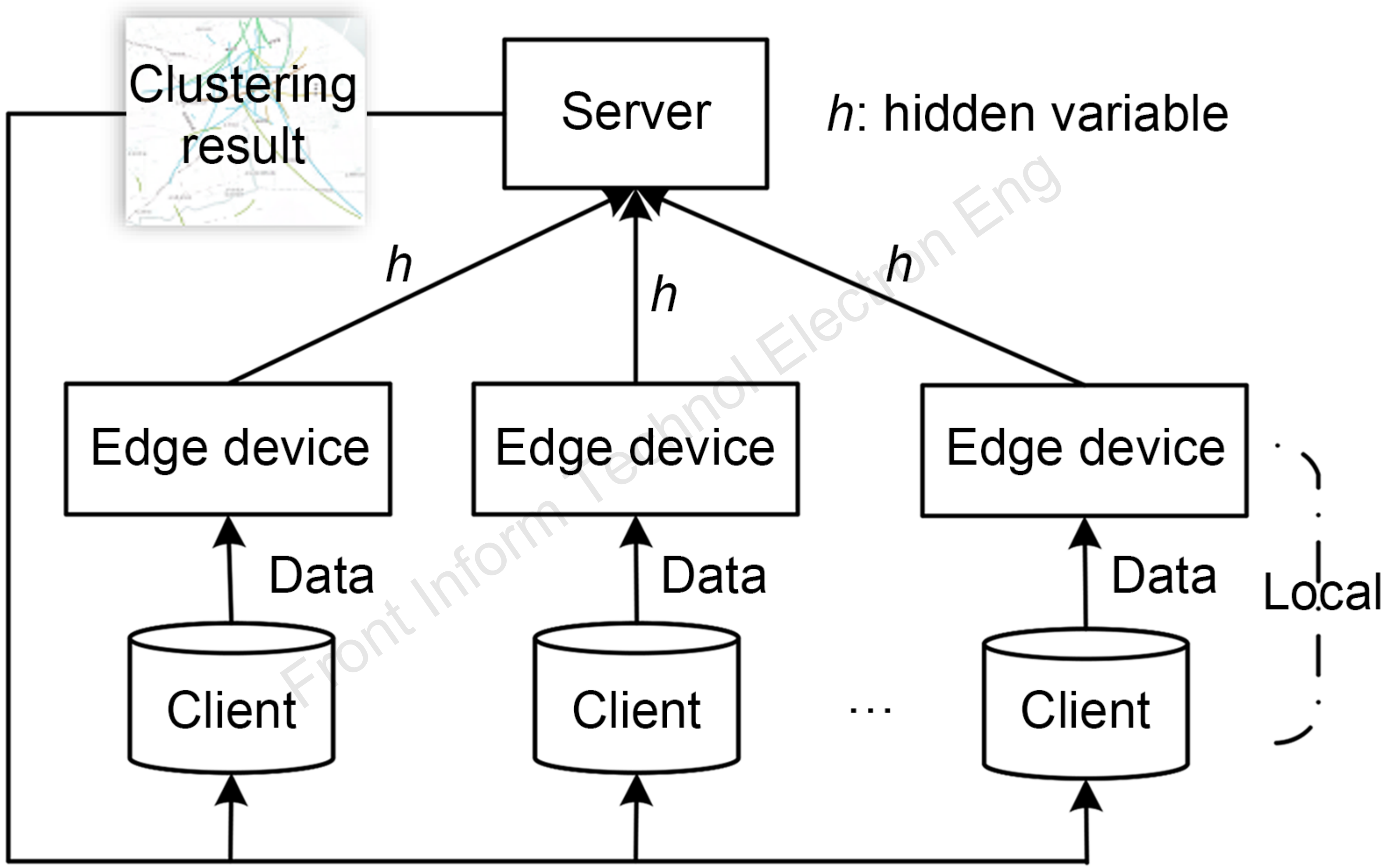
Method

- **Privacy-preserving vehicle trajectory clustering framework:** We propose a privacy-preserving vehicle trajectory clustering framework, which generates the hidden variables of the original data locally and uploads the hidden variables to the server for clustering. On one hand, the original private data on vehicle trajectory are safeguarded. On the other hand, the hidden variables can significantly improve the clustering effect.

- **Vehicle trajectory clustering model (IKV):** We propose a vehicle trajectory clustering model (IKV) based on the variational autoencoder and an improved *K*-means algorithm. The model uses historical trajectory data of vehicles for training on the variational autoencoder and calculating latent variables, which helps discover the intrinsic patterns in the trajectory data while protecting privacy, thereby improving the effect of clustering.

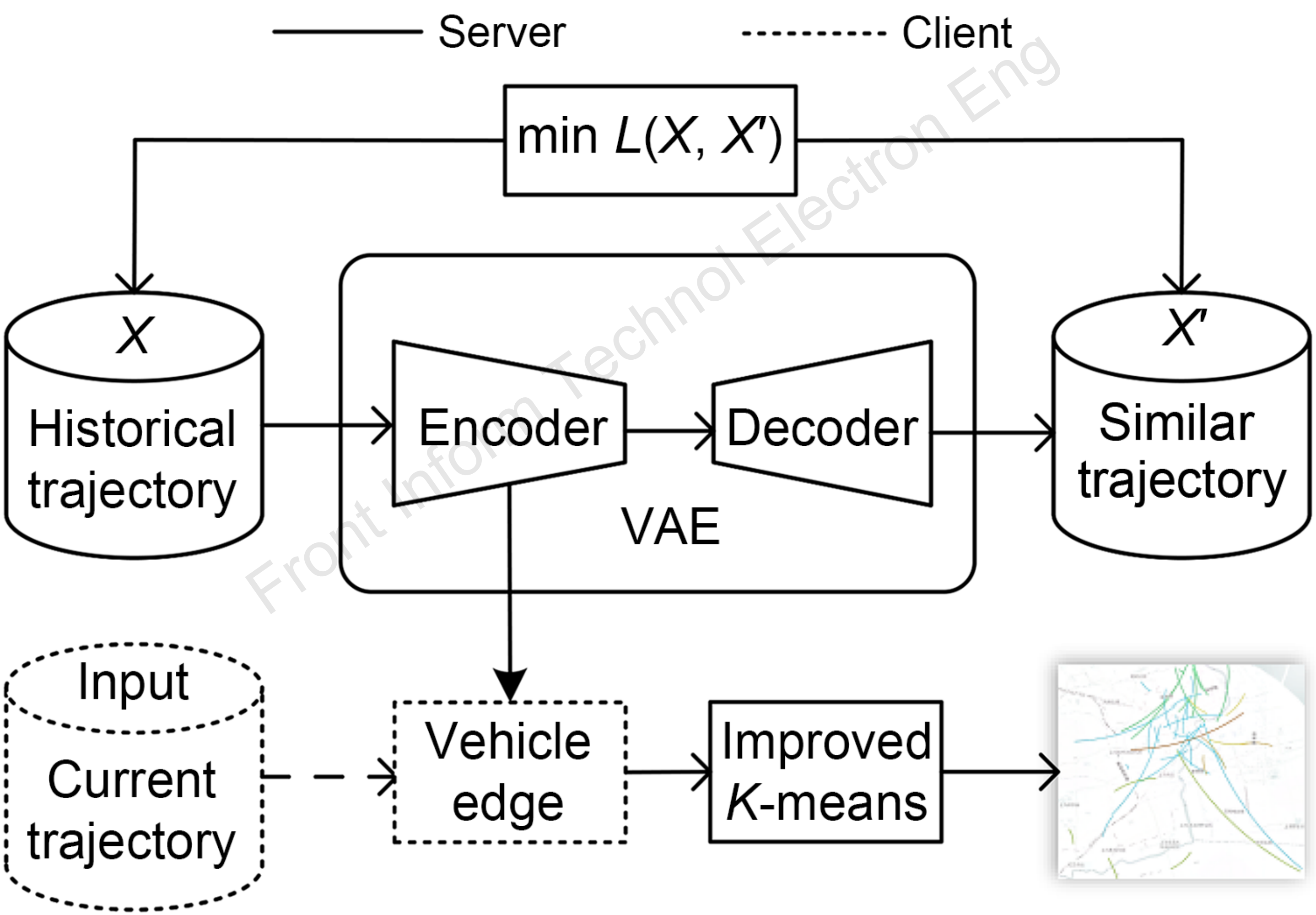
- **Improved *K*-means algorithm:** We have improved the *K*-means algorithm. In each iteration, the cluster centers are updated by minimizing the error value from the cluster center to other trajectory points. The trajectory objects that match the cluster centers screened by the IKV can be found in the real world.

Privacy-preserving vehicle trajectory clustering framework



Overall process of IKV

We assemble VAE and improved *K*-means (IK) to the frame to obtain the IKV



Major results

- To verify the advantages of IKV, we compared IKV with seven different baselines, including ***K*-means** (MacQueen, 1967), **mini batch *K*-means** (MBK-means) (Sculley, 2010), ***K*-medoids** (Park and Jun, 2009), spectral clustering (**SPC**) (Ng et al., 2001), spectral bi-clustering (**SPBC**) (Kluger et al., 2003), BIRCH clustering (**BIC**) (Zhang et al., 1996), and **FCM** (Dunn,1973; Bezdek, 1981).

Table 1 Overall comparison results of IKV and other clustering algorithms

Algorithm	SC				DBI				CHI			
	S1	S2	E	LA	S1	S2	E	LA	S1	S2	E	LA
K-means	8.95*	7.19*	1.35	2.59	21.47*	28.71	68.83	63.93	2438.09*	1684.97*	320.01	575.35
MBK-means	7.71	0.23	1.28	2.79	26.49	71.31	63.77	61.24	1895.08	997.66	366.45	616.26
K-medoids	5.04	6.42	1.28	2.54	33.21	31.39	68.10	61.10	1326.37	1429.00	349.27	584.54
SPC	–	6.64	**3.74**	2.21	–	**17.07**	**24.26**	**23.20**	–	96.67	81.00	52.52
SPBC	6.69	0.35	1.35	1.71	32.91	101.03	69.84	75.05	1379.74	1150.01	360.17	512.03
BIC	8.57	0.99	1.96*	3.37*	22.08	94.51	60.84*	56.21	2376.16	1220.71	387.62*	655.82
FCM	5.32	−0.68	0.06	**4.37**	37.98	110.12	67.92	46.95*	1308.82	746.25	345.28	**1035.02**
IKV	**10.88**	**8.00**	1.27	1.85	**20.08**	26.64*	63.64	67.03	**3734.77**	**1739.76**	**601.91**	819.19*

The best results are in bold and the second best results are marked with asterisks. SC: silhouette coefficient; DBI: Davies–Bouldin index; CHI: Calinski–Harabaz index; S1: ShangHai1; S2: ShangHai2; E: Emeryville; LA: LosAngele

Major results (Cont'd)

- Table 2 shows that for the dataset ShangHai2, VAE performs better than the other dimensionality reduction algorithms on SC and DBI, and has a sub-optimal performance on CHI. For the dataset LosAngele, SVD and IPCA perform well on SC and DBI, while VAE performs at an average level, but VAE performs the best on CHI.

Table 2 Overall comparison results of VAE with other dimensionality reduction algorithms

Algorithm	SC		DBI		CHI	
	ShangHai2	LosAngele	ShangHai2	LosAngele	ShangHai2	LosAngele
PCA	7.43	2.62	28.28	62.29	1732.22	546.01
AE	2.77	2.18	74.66	64.55	**2534.16**	665.48
DictionaryLearning	−0.93	1.13	38.65	70.85	43.61	112.97
SVD	5.99	**2.91**	32.46	**60.60**	1372.60	644.80
IPCA	6.15	**2.91**	30.85	61.07	1341.46	647.30
VAE	**8.00**	1.85	**26.64**	67.03	1739.76	**819.19**

The best results are in bold. SC: silhouette coefficient; DBI: Davies–Bouldin index; CHI: Calinski–Harabaz index

Conclusions

- Compared to traditional vehicle trajectory clustering methods, the framework we proposed integrates a variational autoencoder (VAE) and an improved *K*-means (IK) algorithm, demonstrating superior performance in balancing privacy and utility. This framework effectively captures the inherent patterns in vehicle trajectory data while safeguarding user privacy, thereby enhancing the accuracy of clustering.
- By integrating the IKV model into our framework, we have successfully overcome the limitations of existing methods, offering a more reliable and efficient approach to vehicle trajectory clustering that respects user privacy and delivers accurate clustering outcomes. This advancement has achieved certain progress in the field of intelligent transportation systems, and it also provides new references for data mining with higher privacy protection requirements in the era of big data.

Ran TIAN obtained his Ph.D. degree in Computer Application Technology from the School of Computer Science, Southwest Jiaotong University in 2015. Since July 2018, he has served as an associate professor at the College of Computer Science & Engineering at Northwest Normal University. His main research interests include digital supply chain, intelligent transportation, and intelligent decision-making methods.

Pulun GAO received a Bachelor's degree in Computer Science from Northwest Normal University in 2024. His research focuses on deep learning and network traffic measurement.

Yanxing LIU received a Master's degree in Computer Science from Northwest Normal University in 2017. His research interests include time series data analysis, knowledge mining, and application of sensor networks, blockchain, and information systems.