

Article

Traffic Flow Prediction Based on Federated Learning and Spatio-Temporal Graph Neural Networks

Jian Feng *, Cailing Du and Qi Mu

College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an 710054, China; klaus@stu.xust.edu.cn (C.D.); muqi@xust.edu.cn (Q.M.)

* Correspondence: fengjian@xust.edu.cn

Abstract: In response to the insufficient consideration of spatio-temporal dependencies and traffic pattern similarity in traffic flow prediction methods based on federated learning, as well as the neglect of model heterogeneity and objective heterogeneity, a traffic flow prediction model based on federated learning and spatio-temporal graph neural networks is proposed. The model is divided into two stages. In the road network division stage, the traffic road network is divided into subnetworks by the dynamic time warping algorithm and the K-means algorithm, to ensure the same subnetwork has the similar traffic flow pattern. The federated learning stage is divided into two sub-stages. In the local training phase, the spatio-temporal graph neural network with an attention mechanism is utilized to create personalized models and meme models to capture the spatio-temporal dependencies of each subnetwork. At the same time, deep mutual learning is utilized to address model heterogeneity and objective heterogeneity through knowledge distillation. In the global aggregation phase, a multi-factor weighted aggregation strategy is designed to measure the contribution of each local model to the global model, to enhance the fairness of aggregation. Three sets of experiments were conducted on two real datasets, and the experimental results demonstrate that the proposed model outperforms the baseline models in three common evaluation metrics.

Keywords: traffic flow prediction; federated learning; pattern similarity; model heterogeneity; objective heterogeneity



Citation: Feng, J.; Du, C.; Mu, Q. Traffic Flow Prediction Based on Federated Learning and Spatio-Temporal Graph Neural Networks. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 210. <https://doi.org/10.3390/ijgi13060210>

Academic Editors: Wolfgang Kainz, Xiao Li, Xiao Huang and Zhenlong Li

Received: 19 April 2024
Revised: 27 May 2024
Accepted: 14 June 2024
Published: 18 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the acceleration of urbanization and motorization, urban traffic is facing severe challenges in terms of carrying capacity and operational efficiency [1]. Traffic flow prediction is used to predict future traffic flow for a certain period based on the historical traffic flow. It is an important component of Intelligent Transportation Systems (ITSs). Accurate traffic flow prediction is crucial for urban road planning, traffic control and estimating planning time accurately [2].

Traffic flow has the characteristic of dynamic spatio-temporal attributes. At present, the mainstream methods for traffic flow prediction are based on deep learning (DL) [3], which captures the spatio-temporal dependencies of traffic flow by the deep feature extraction and strong nonlinear fitting ability of DL. However, existing research is focused on centralized learning, which requires the transmission of local datasets from the clients to a central server. It will bring about a huge transmission overhead, the problem of data loss [4] and the problem of data privacy exposure [5] more seriously.

To protect data privacy, federated learning (FL) is introduced to the latest research [6]. FL is a distributed machine learning framework based on privacy-preserving and secure encryption techniques [7] that enables participants to collaboratively train models without sharing raw data. In FL, local clients perform local model training on local data and upload model weights, and the central server aggregates all model weights to obtain a new global model. The process protects data privacy and reduces communication costs by keeping data localized.

However, there are still some problems when applying FL to traffic flow prediction.

- (1) Firstly, the local models cannot capture spatio-temporal dependencies effectively because allocated data to local clients do not have distinct patterns. Generally, training data allocated to local clients come from subnetworks formed by sensors. In existing FL-based methods, sensors to form subnetworks are selected randomly or based on geographical proximity, such as latitude and longitude, or node connectivity patterns, and fail to consider traffic pattern similarity. This leads to the inadequate learning of spatio-temporal dependencies on local clients. So, how to divide subnetworks to ensure that local clients obtain data with distinct patterns is the first challenge.
- (2) Secondly, there is insufficient support for personalized learning needs for local clients. To achieve simplicity of aggregation, existing research typically adopts the same learning model for local clients and the central server based on the model homogeneity assumption. However, model homogeneity limits the diversity of local models and the personalization of training objectives. To support personalized learning on local clients, an intuitive approach is to adopt different learning models for each client. However, this leads to the problem of model heterogeneity between local models and the global model. To achieve the personalization of training objectives for both local clients and the central server, the issue of objective heterogeneity must be addressed. So, how to overcome the dilemma of objective heterogeneity and model heterogeneity is the second challenge.

To this end, a model for traffic flow prediction called FedTFP is proposed, which is based on FL and spatio-temporal graph neural networks (STGNNs). Firstly, road network division is achieved according to the pattern similarity of traffic flow, making the same subnetwork exhibit similar traffic flow changes. Secondly, FL is used to achieve distributed training across subnetworks. In the local training phase, personalized models and meme models on local clients are created based on STGNNs with an attention mechanism (AF-STGCN) to achieve personalized learning and global aggregation separately. At the same time, deep mutual learning (DML) is utilized to achieve knowledge distillation between two kinds of models. And then, in the global aggregation phase, the multi-factor weighted aggregation (MFWA) strategy is proposed to comprehensively evaluate the contribution of the local models to the global model. The main contributions are as follows:

- (1) A new model FedTFP for traffic flow prediction based on FL and AFSTGCN is proposed to protect data privacy as well as learn the spatio-temporal dependencies of traffic flow comprehensively.
- (2) A method for dividing the road network based on the pattern similarity of traffic flow is adopted, which can enhance the learning ability for the spatio-temporal dependencies of the model.
- (3) An intermediate layer between local clients and the central server is introduced to address model heterogeneity and objective heterogeneity.
- (4) MFWA is proposed to evaluate the contribution of local models comprehensively to address the data imbalance problem of the FedAVG algorithm.
- (5) Experiments on two real-world datasets (PeMS04 and PeMS08) were conducted, and the experimental results demonstrate that FedTFP has better prediction performance than baselines.

The rest of the paper is organized as follows. Section 2 reviews the related work on traffic flow prediction and FL. Section 3 gives the definition of traffic flow prediction. Section 4 describes the FedTFP model in detail. Section 5 discusses the experimental results, and Section 6 gives the conclusions.

2. Related Work

This section outlines the related work of traffic flow prediction and FL for traffic flow prediction.

2.1. Traffic Flow Prediction

The methods for traffic flow prediction evolve in three stages: statistical methods, machine learning (ML) methods and DL-based methods.

Statistical methods predict future traffic flow by statistical analysis on historical data. Typical methods include the historical average (HA) [8] and the autoregressive integrated moving average model (ARIMA) [9]. These methods assume that the time series are stable and regular, so they are insufficient to capture nonlinear relationships in traffic flow.

To capture nonlinear relationships in traffic flow, ML is introduced. Large amounts of traffic flow can be modeled by ML methods, such as support vector regression (SVR) [10], Bayesian networks [11] and so on. Although these methods can capture the nonlinear relationships in traffic data, they are limited by manual feature selection and to shallow model structure.

Currently, DL-based methods have become the mainly used methods to deal with large-scale and high-dimensional traffic flow based on their powerful nonlinear learning capabilities. At first, recurrent neural networks (RNNs) and its variants [12,13] are adopted to capture temporal dependency. However, these methods overlook spatial dependency. To capture the spatial dependency of the road network, convolutional neural networks (CNNs) [14] are introduced, such as ST-ResNet [15]. These methods consider the road network as the grid structure and overlook its non-Euclidean structural characteristics. To address this problem, graph neural networks (GNNs) [16] are introduced to learn complex non-Euclidean relationships in the road network. At present, mainstream research combines GNNs and RNNs to capture the spatio-temporal dependencies of traffic flow. For example, the STGCN [17] uses the GCN to capture spatial dependency and 1D-conv to capture temporal dependency. To simultaneously capture spatio-temporal dependencies, T-GCN [18] is proposed. T-GCN uses the GCN to capture spatial dependency, while it employs the Gated Recurrent Unit (GRU) to capture temporal dependency. Inspired by Graph Attention Networks (GATs) [19], subsequent studies introduce attention mechanisms to capture spatio-temporal dependencies better. For example, the ASTGCN [20] adds an attention mechanism to the spatio-temporal convolution and models three temporal characteristics of traffic flow.

These methods have made significant progress in capturing the spatio-temporal dependencies of traffic flow but overlook the privacy of traffic data.

2.2. FL for Traffic Flow Prediction

To address privacy concerns, researchers have applied FL to traffic flow prediction. For example, FedGRU [21] combines the GRU and FL to achieve accurate traffic flow prediction while preserving the privacy of traffic data. The FCGCN [22] divides the traffic road network into multiple subnetworks by community detection, uses FL to protect the privacy of traffic data and uses the GCN to capture spatial dependency. DST-GCN [23] uses an enhanced FL framework and captures spatio-temporal dependencies by incorporating the GRU and GCN. Moreover, some studies focus on potential security issues in FL-based traffic flow prediction. For example, the FASTGNN [24] employs an adjacency matrix preservation method based on differential privacy to protect the topological information of the road network.

However, the research on FL-based traffic flow prediction is still in an early stage, and it pays more attention to data privacy protection but learns spatio-temporal dependencies insufficiently. Further, it ignores the pattern similarity of traffic flow as well as model heterogeneity and objective heterogeneity. To solve the above problems, FedTFP is proposed in this paper.

3. Problem Definition

The traffic road network is defined as an undirected graph $G = (V, E, A)$, where $V = \{v_1, v_2, \dots, v_N\}$ denotes the set of nodes, namely the sensors deployed in the road network, N is the number of nodes and E is the set of edges, which reflects the connectivity

between nodes. $A \in \mathbf{R}^{N \times N}$ represents the adjacency matrix of G . $\forall v_i, v_j \in V$; if v_i and v_j are connected, then $A_{ij} = A_{ji} = 1$; otherwise, $A_{ij} = A_{ji} = 0$. $\mathbf{X}_t = (x_t^1, x_t^2, x_t^3, \dots, x_t^N)$ represents the values of the features of all nodes at moment t , where $x_t^i \in \mathbf{R}^D$ represents the values of the features of node v_i at moment t , while D represents the number of features.

Give the traffic network G and historical traffic data, the problem of traffic flow prediction can be defined as learning a function f to predict the future traffic of the next T' moments, as shown in Equation (1):

$$\left[\mathbf{X}_{(t-n+1):t}, G \right] \xrightarrow{f} \left[\mathbf{X}_{(t+1):(t+T')} \right] \quad (1)$$

4. Method

FedTFP is divided into two phases, including the road network division phase (RND) and the FL phase (FLP), as shown in Figure 1. RND realizes the road network division. FLP accomplishes local model training and global aggregation.

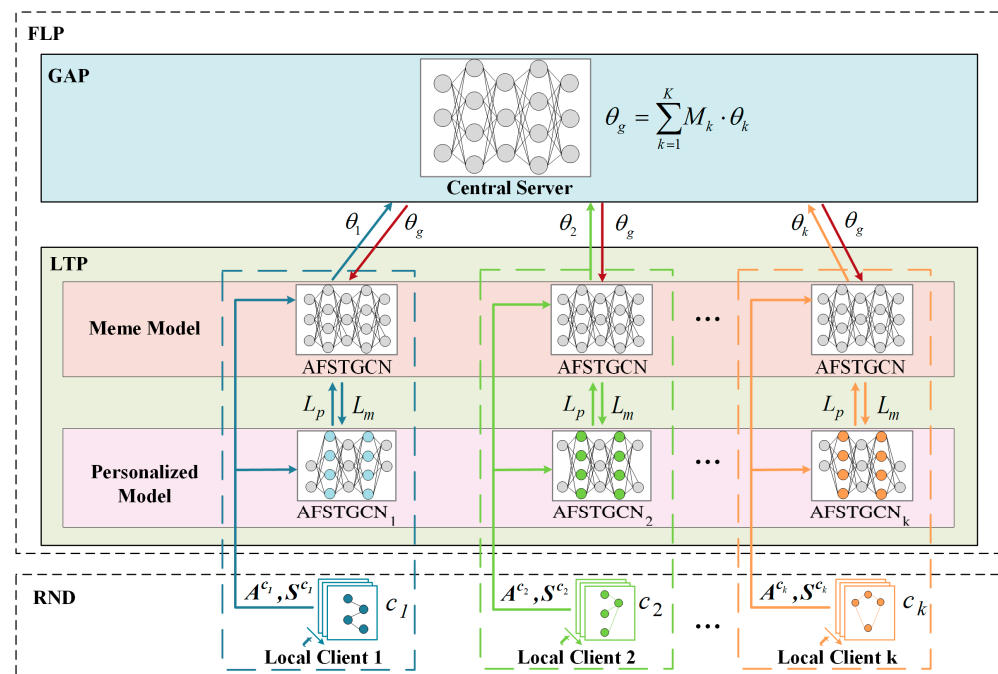


Figure 1. The framework of FedTFP.

4.1. RND

This phase implements the road network division, and the subnetworks divided will be processed by local clients in the FLP.

Due to the distinct regional characteristics of traffic flow, existing research often divides the nodes into subnetworks based on spatial geographic attributes such as latitude and longitude, to ensure nodes with geographic proximity within the same subnetwork. However, geographic proximity is not always equal to pattern similarity, and on the contrary, two nodes in different regions may have the similar pattern of changes. Intuitively, if the nodes can be divided based on the pattern similarity, a subnetwork will exhibit similar traffic flow changes, and then the local client will capture the spatio-temporal dependencies better in the FLP. In summary, the method of dividing the road network based on spatial geographic attributes assumes that geographically close locations exhibit similar traffic flow variations. In contrast, the method based on the pattern similarity of traffic flow focuses on identifying sequence similarities in traffic flow.

To this end, FedTFP implements road network division based on the pattern similarity strategy (PSS). This is a clustering problem based on time series similarity in essence. It is

divided into two steps: similarity calculation and clustering. The dynamic time warping algorithm (DTW) [25] is used for similarity calculation, which can compute time series similarity with varying lengths and time offsets efficiently. And the K-means algorithm [26] is adopted in clustering, because it is suitable for clustering data with distinct patterns.

The process is shown in Figure 2. Firstly, the similarity of the historical traffic flow sequences S is calculated by DTW to obtain the pattern similarity matrix A_s . $S = (X^1, X^2, \dots, X^N)$ represents the historical feature sequences of all nodes, where $X^i = (x_1^i, x_2^i, x_3^i, \dots, x_T^i)$ represents the values of all features of all moments of node v_i and d_{ij} denotes the shortest distance between X^i and X^j . Secondly, according to A_s , K-means is used to cluster the traffic flow with the similar pattern into one class with a total of K classes. The clustering results are denoted as $C = (c_1, c_2, \dots, c_K)$, where c_k denotes the k -th subnetwork.

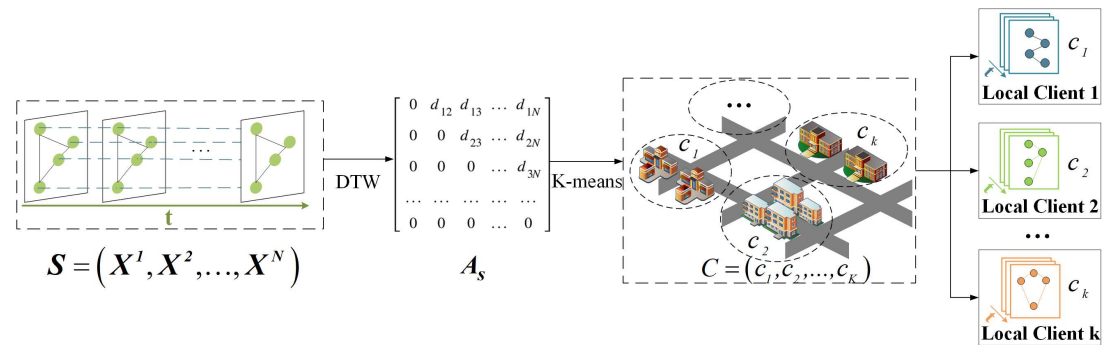


Figure 2. Process of RND.

4.2. FLP

The FLP is subdivided into the local training phase (LTP) and the global aggregation phase (GAP) further.

4.2.1. LTP

This phase completes model training on each subnetwork and addresses model heterogeneity and objective heterogeneity.

In the LTP, existing studies typically assume that all clients have similar data distribution and therefore utilize the same deep learning network structure. However, due to the differences in the distribution of input data, this assumption is not reasonable. But if the structure and training objectives of the model in each client are designed based on the characteristics of local data, it will lead to model heterogeneity and objective heterogeneity, which makes it difficult to fully integrate the characteristics of each local model in global aggregation. Therefore, how to solve the model heterogeneity and objective heterogeneity is difficult.

The key to solving the problem is to address the impact of heterogeneity on global aggregation. To this end, in the LTP, on the one hand, personalized models are trained to conform to the local data distribution and learn the local personalized features. On the other hand, to address model heterogeneity and objective heterogeneity, meme models are designed as intermediate models between personalized models and the global model and to shield the differences in personalized models for the central server. DML is introduced to realize the knowledge distillation between meme models and personalized models.

Specifically, the model structure and scale of the personalized models are determined based on the local data. Meme models are consistent with the global model. Both models use AFSTGCN as the basic unit to capture spatio-temporal dependencies. Finally, through DML, personalized models and meme models learn from and guide each other.

AFSTGCN

AFSTGCN embeds the GCN into the GRU to learn the spatio-temporal dependencies in traffic flow comprehensively. Its structure is shown in Figure 3. At first, the spatial

dependency is captured by the attention layer and GCN. And then, taking the time series with spatial dependency as input, the temporal dependency is obtained by the GRU through information transmission between units.

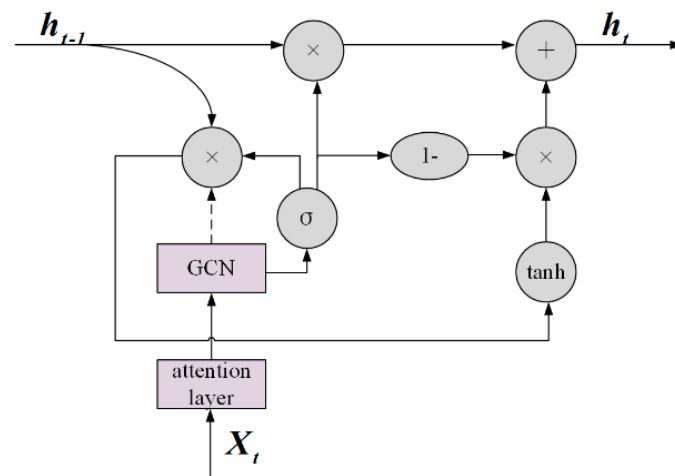


Figure 3. The structure of AFSTGCN.

- Attention layer

The attention layer aims to capture complex spatial dependency better. Firstly, the node features are taken as input to learn the correlation between nodes. Secondly, the attention mechanism is used to conduct the weighted aggregation of neighboring nodes. Finally, the final output features are generated, as shown in Equations (2)–(4).

$$e_{ij} = \mathbf{w}x_t^i \cdot \mathbf{w}x_t^j \tag{2}$$

$$a_{ij} = \text{softmax}(e_{ij}) \odot A^{c_k} \tag{3}$$

$$x_t^i = \sum_{j \in \mathcal{N}_i} a_{ij} \cdot \mathbf{w}x_t^j + b \tag{4}$$

where e_{ij} denotes the attention coefficients, \mathbf{w} is the weight matrix, a_{ij} is the result of normalization for the attention coefficients and \odot denotes the Hadamard product. A^{c_k} is the adjacency matrix of the local client k , \mathcal{N}_i denotes the set of first-order neighbors of node v_i and b is the bias.

- Spatial dependency

A two-layer GCN is used to capture the spatial dependency in the traffic flow, as shown in Equation (5).

$$f(\mathbf{X}_t, A^{c_k}) = \sigma \left(\hat{A}^{c_k} \text{Relu} \left(\hat{A}^{c_k} \mathbf{X}_t \mathbf{W}_0 \right) \mathbf{W}_1 \right) \tag{5}$$

where $\hat{A}^{c_k} = \tilde{D}^{-\frac{1}{2}} \tilde{A}^{c_k} \tilde{D}^{-\frac{1}{2}}$ is the preprocessing of A^{c_k} , $\tilde{A}^{c_k} = A^{c_k} + I$ and I is the unit matrix. \tilde{D} is the degree matrix. \mathbf{W}_0 and \mathbf{W}_1 denote the weight matrices of the first and second layers of the GCN. $\sigma(\cdot)$ denotes the activation function.

- Temporal dependency

The temporal dependency is learned by the GRU, as shown in Equation (6).

$$\mathbf{h}_t = \text{GRU}(f(\mathbf{X}_t, A^{c_k}), \mathbf{h}_{t-1}) \tag{6}$$

where \mathbf{h}_{t-1} denotes the hidden layer state at time $t - 1$.

The above shows the calculation process for one AFSTGCN unit. In the actual calculation, the number of units needs to be determined according to the parameters of the

model, such as the model scale, etc. The calculation process of the meme model is denoted as $AFSTGCN(\cdot)$, and that of the personalized model is denoted as $AFSTGCN_k(\cdot)$.

DML

After basic spatio-temporal dependencies learning through AFSTGCN, DML is used to achieve mutual learning between the meme model and personalized model.

Traditional DML is generally applied to classification tasks [27], but traffic flow prediction is a regression task, so a new loss function suitable for the task is needed. Here, the meme model and the personalized model need not only accept the supervision of the truth value in the training process but also leverage the learning experience of the other party to enhance their generalization ability. So, the loss function is designed to consist of a primary task loss and an auxiliary task loss. The primary task is to optimize the performance of the model itself. The auxiliary task promotes the meme model and personalized model to learn from each other.

$$L_o = MSE(y, AFSTGCN(A^{c_k}, S^{c_k})) \quad (7)$$

$$L_c = MSE(y, AFSTGCN_k(A^{c_k}, S^{c_k})) \quad (8)$$

$$L_a = MSE(AFSTGCN_k(A^{c_k}, S^{c_k}), AFSTGCN(A^{c_k}, S^{c_k})) \quad (9)$$

$$L_b = MSE(AFSTGCN(A^{c_k}, S^{c_k}), AFSTGCN_k(A^{c_k}, S^{c_k})) \quad (10)$$

$$L_m = \beta L_o + (1 - \beta) L_a \quad (11)$$

$$L_p = \alpha L_c + (1 - \alpha) L_b \quad (12)$$

where MSE is the Mean Squared Error. y denotes the true value. L_o and L_c denote primary task loss. L_a and L_b represent auxiliary task loss. L_m and L_p are the total loss for the meme model and the personalized model separately. α and β are fixed to 0.5 so that attention is paid equally to the primary task and the auxiliary task.

The local training process based on DML is shown in Figure 4. After the training, the meme models upload model parameter θ_k to the central server.

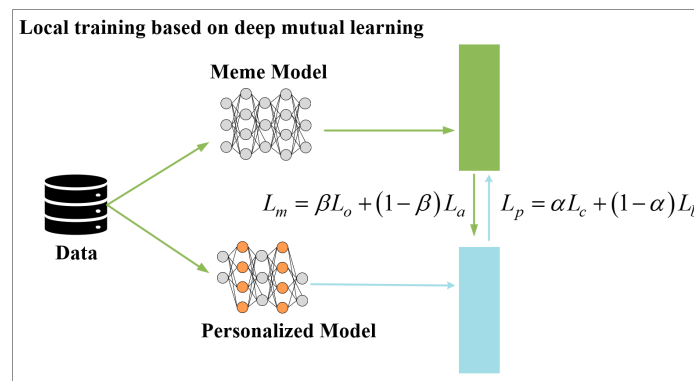


Figure 4. Local training based on DML.

4.2.2. GAP

This phase implements global aggregation and updates the global model.

In FL, the global aggregation strategy refers to aggregating the model weights uploaded by local clients to update the global model. From the perspective of enhancing model performance and reducing the risk of overfitting, it is crucial to choose an appropriate aggregation strategy. Federated Averaging (FedAVG) [28] is one of the common aggregation strategies, which uses the weighted average of the model parameters from the clients to generate a new global model. In FedAVG, the weighted coefficient of each client depends on the size of the local data. However, the measurement method that only considers the importance of data size ignores other possible influencing factors, which may affect the generalization ability of the global model and the fairness of aggregation.

Therefore, to improve the fairness of aggregation, the MFWA strategy is proposed to improve the fairness of aggregation and enhance the robustness of the global model. In MFWA, the weighted coefficient of the client is determined by the model performance and the data size. Firstly, after the training of the local model, the model performance MSE of the k -th client is obtained as p_k . Secondly, the local client uploads current model parameters θ_k and p_k to the central server. Then, the central server calculates the aggregation weighted coefficient of the client based on p_k and $|c_k|$, namely the size of data on the k -th client. The calculation is in Equation (13).

$$M_k = \gamma e^{-p_k} + (1 - \gamma)|c_k| \quad (13)$$

where M_k denotes the weighted coefficient of the k -th client. γ is a hyperparameter that controls the contribution of model performance and data size, and the value of γ should be determined by experiments.

Finally, the central server uses the weighted coefficient M_k of each client for global aggregation. The calculation is described in Equation (14).

$$\theta_g = \sum_{k=1}^K M_k \cdot \theta_k \quad (14)$$

where θ_g is the parameter of the global model. Then, the central server sends θ_g to the meme model of the clients and starts the next round of training.

5. Experiments and Analysis

To verify the effectiveness of FedTFP, three sets of experiments are designed to try to answer the following questions:

- Question 1: Compared with the existing FL-based models for traffic flow prediction, has the prediction performance of FedTFP improved?
- Question 2: What are appropriate values of parameters?
- Question 3: How do different strategies adopted by FedTFP contribute to the prediction performance?

5.1. Dataset

To verify the prediction performance of FedTFP and ensure the reliability of the experimental results, two public traffic flow datasets are used, namely PeMS04 [20] and PeMS08 [20]. They are all derived from the dataset of PeMS (<https://pems.dot.ca.gov/>) (accessed on 1 June 2024) (Caltrans Performance Measurement System), published by the California Department of Transportation. The basic information of the two datasets is as follows.

- PeMS04 is 59-day traffic flow data collected from 307 sensors on 29 roads in San Francisco Bay Area, with a time interval of 5 min, containing a total of 16,992 traffic flows. The dataset spans from January to February 2018.
- PeMS08 is 62-day traffic flow data collected from 170 sensors on 8 roads in San Bernardino, with a time interval of 5 min, containing 17,856 traffic flows. The dataset spans from July to August 2016.

We apply the Z-Score to normalize the dataset and divide it into the training set, validation set and test set in a ratio of 8:1:1. The Adam optimizer is utilized to train the model. The batch size and learning rate are set to 32 and 0.001 separately. The sliding window is set to 12, indicating that the traffic flow of the 60 min is used to predict the traffic flow of the next 5 min. The communication round of FL is 50, and the training epoch of local models is 50. To reduce the overhead of communication and model training, K is set to 6 in the experiment. The appropriate value of K can match the real-world scenario on the one hand and control the training cost on the other hand.

5.2. Evaluation Metrics

Mean Absolute Error (MAE) [29], Mean Absolute Percentage Error (MAPE) [29] and Root Mean Square Error (RMSE) [29] are used to evaluate the prediction performance of FedTFP. The smaller the value of the metrics, the better the predictive performance of the model.

5.3. Baselines

Two competitive baselines based on the model homogeneity assumption are used for performance comparison.

1. The FCGCN combines community detection, FL and the GCN to achieve traffic flow prediction.
2. DST-GCN adopts an improved FL to prevent learning failure and utilizes the GCN and GRU to capture spatio-temporal dependencies.

5.4. Experiment 1: Prediction Performance of FedTFP

Aiming to answer Question 1, FedTFP compares two FL-based traffic flow prediction models, the FCGCN and DST-GCN. The experimental results are shown in Table 1.

Table 1. Baseline comparison.

Dataset	Model	MAE	RMSE	MAPE
PeMS04	FCGCN	27.68	38.21	0.15
	DST-GCN	29.55	42.17	0.17
	FedTFP	17.95	27.98	0.13
PeMS08	FCGCN	24.82	32.64	0.15
	DST-GCN	26.07	37.11	0.15
	FedTFP	14.52	21.95	0.09

The experimental results demonstrate that the prediction performance of FedTFP on the two datasets have significant improvement compared to baseline models.

Firstly, FedTFP considers the pattern similarity of traffic flow in road network division, while both baselines do not. DST-GCN divides the road network evenly based on the proportion of the dataset, and the FCGCN employs the Louvain algorithm to divide the road network based on the connection patterns between road nodes. Both methods focus on geographical spatial attributes and neglect the variations in traffic flow, resulting in poor prediction performance.

Secondly, both the FCGCN and DST-GCN are based on the model homogeneity assumption, which cannot meet the personalized needs of local clients. FedTFP breaks the assumption of model homogeneity and can meet the personalized requirements of local models better, obtaining good performance improvement.

Finally, both the FCGCN and DST-GCN use the FedAVG algorithm for global aggregation, which has a data imbalance problem, so as to impact the generalization ability of the global model. FedTFP addresses the data imbalance problem using the MFWA strategy. MFWA fully considers the impact of data size and model performance on the global model, thereby improving the robustness of the global model.

5.5. Experiment 2: Parameter Determination

To address Question 2, different values of γ are tested to observe their impact on model performance.

When selecting the value of γ , it is essential to balance the influence of client model performance and data size. Typically, γ is within the range of 0 to 1, with adjustments made in steps of 0.1. However, choosing too low γ values (in the range of 0 to 0.2) or too high γ values (in the range of 0.8 to 1) will make the model over-biased to a certain factor, resulting in increased sensitivity to outliers and reduced model performance. Therefore, the selection

range of γ is restricted to between 0.3 and 0.7, which helps improve the stability of the model. As shown in Figures 5–7, MAE and RMSE values decrease first and then increase with the increase in γ , and the MAPE does not change first and then increases. When γ is 0.4, the prediction performance is the best. So γ is set to 0.4.

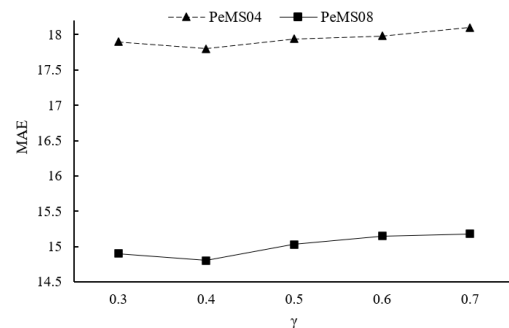


Figure 5. MSE under different γ .

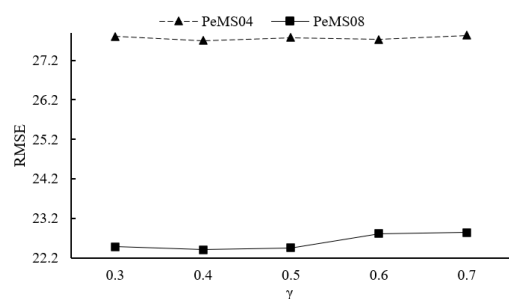


Figure 6. RMSE under different γ .

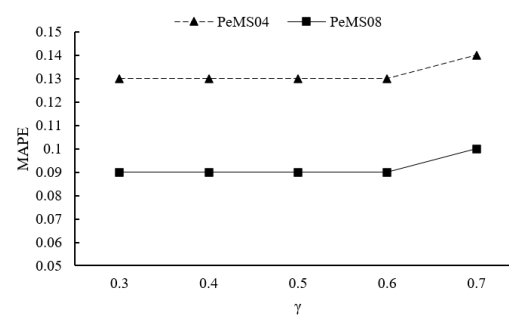


Figure 7. MAPE under different γ .

5.6. Experiment 3: Ablation Experiments

The PSS, DML and MFWA are strategies designed or adopted by FedTFP. To verify their effectiveness, three sets of ablation experiments are conducted below.

5.6.1. Ablation Experiment for PSS

To validate the effectiveness of the PSS, strategies for road network division adopted by DST-GCN and the FCGCN are used to form two contrasting models separately, including FedTFP-R, dividing the road network randomly according to a proportion, and FedTFP-L, dividing the road network by the Louvain algorithm. The experimental results are shown in Table 2.

Table 2. Ablation experiment for PSS.

Dataset	Model	MAE	RMSE	MAPE
PeMS04	FedTFP-R	18.67	28.53	0.15
	FedTFP-L	18.21	28.31	0.13
	FedTFP	17.84	27.75	0.13
PeMS08	FedTFP-R	15.08	22.54	0.10
	FedTFP-L	14.85	22.49	0.09
	FedTFP	14.71	22.31	0.09

The experimental results indicate that FedTFP outperforms FedTFP-L and FedTFP-R across all three performance evaluation metrics, validating the effectiveness of the road network division method based on the pattern similarity of traffic flow. FedTFP-R randomly divides the road network, overlooking the underlying structure of the traffic road network, resulting in significant data discrepancies within one subnetwork. FedTFP-L considers node connectivity patterns but neglects spatial information and the dynamics of traffic flow, resulting in suboptimal performance.

5.6.2. Ablation Experiment for DML Strategy

To validate the effectiveness of the DML strategy, the comparison models FedTFP-A (without DML) and FedTFP (with DML) are formed. The performance of both models is shown in Table 3.

Table 3. Ablation experiment for DML.

Dataset	Model	MAE	RMSE	MAPE
PeMS04	FedTFP-A	18.52	28.88	0.13
	FedTFP	17.86	27.78	0.13
PeMS08	FedTFP-A	15.02	22.65	0.09
	FedTFP	14.53	21.97	0.09

From Table 3, it can be seen that FedTFP has better performance than that of FedTFP-A, demonstrating that DML can enhance the prediction performance overall.

So, is DML also effective for each local model? To answer this question, further experiments are conducted, and the results are shown in Table 4.

Table 4. Performance of each local model with DML.

Dataset	Client	MAE	RMSE	MAPE
PeMS04	Client 0	18.91	29.24	0.13
	Client 1	15.64	23.17	0.08
	Client 2	17.62	26.93	0.14
	Client 3	18.84	28.54	0.15
	Client 4	17.88	28.09	0.10
	Client 5	18.49	27.81	0.13
	Central server	17.84	27.76	0.13
PeMS08	Client 0	16.56	24.62	0.11
	Client 1	17.87	25.31	0.12
	Client 2	12.76	20.55	0.07
	Client 3	14.17	21.46	0.07
	Client 4	13.35	19.40	0.09
	Client 5	5.69	8.24	0.19
	Central server	14.71	22.31	0.09

From Table 4, we can see that although the performance of local models with personalized structures is lower than the global model, they also perform well, compared to

baselines listed in Table 1. This demonstrates that the DML strategy can help to address model heterogeneity and objective heterogeneity.

In conclusion, the experimental results confirm the effectiveness of DML.

5.6.3. Ablation Experiment for MFWA Strategy

To verify the effectiveness of the MFWA strategy, the comparison models FedTFP-B, adopting the FedAVG strategy, and FedTFP, adopting the MFWA strategy, are formed. The performance comparison of the two models is shown in Table 5.

Table 5. Ablation experiment for MFWA.

Dataset	Model	MAE	RMSE	MAPE
PEMS04	FedTFP-B	18.70	28.48	0.16
	FedTFP	17.95	27.98	0.13
PEMS08	FedTFP-B	15.21	22.59	0.11
	FedTFP	14.78	22.42	0.09

Table 5 shows that the prediction performance of the FedTFP-B model on the two datasets is lower than that of the FedTFP model, and it proves the effectiveness of the MFWA strategy. Compared to FedAVG, on the one hand, MFWA gives larger weights to local models with excellent performance and larger data size, to learn the overall data distributions and features better; on the other hand, for local models with poor performance or small data size, MFWA assigns weights reasonably to ensure relative fairness, as well as to prevent biases caused by over-reliance on local models with larger data size.

In summary, the experimental results show that the prediction performance of FedTFP is higher than that of the baseline methods, and the PSS, DML and MFWA strategies are effective. However, FedTFP does not consider environmental constraints [30], which may limit its application in the real world, and it is vulnerable to attacks when uploading model weights or sending global weights in the FLP.

6. Conclusions

FedTFP is proposed for traffic flow prediction based on FL and AFSTGCN. It divides into two stages: RND and the FLP. In RND, FedTFP is able to divide the road network more reasonably by the PSS. In the FLP, the personalized model and meme model are designed to meet the needs of the personalized learning of local clients and the global aggregation of central server separately, and the MFWA strategy is used to improve the generalization ability of the global model. Besides traffic flow prediction, FedTFP can also be applied to traffic accident prediction, which helps reduce traffic congestion, lower the environmental impact of transportation and promote urban sustainability.

In future research, multi-source data such as road conditions, weather, holidays, etc., will be introduced to improve the prediction performance of traffic flow. The challenge may be feature fusion. Additionally, this work will be promoted for traffic flow prediction for multi-city, to help address the data sparsity of some cities. The challenge is heterogeneous transportation modes in different cities.

Author Contributions: Conceptualization, J.F.; methodology, C.D.; validation, C.D.; formal analysis, J.F.; data curation, J.F. and C.D.; writing—original draft, J.F. and C.D.; writing—review and editing, J.F. and C.D.; supervision, Q.M.; funding acquisition, Q.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Program of China (No: 2022YFB3304401).

Data Availability Statement: Data can be provided upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, J.; Zhang, Y.; Hu, Y.; Yin, B. Survey on Graph Convolutional Neural Network-Based Traffic Prediction. *Beijing Gongye Daxue Xuebao/J. Beijing Univ. Technol.* **2021**, *47*, 954–970.
2. Afandizadeh Zargari, S.; Amoei Khorshidi, N.; Mirzahosseini, H.; Heidari, H. Analyzing the Effects of Congestion on Planning Time Index—Grey Models vs. Random For Regression. *Int. J. Transp. Sci. Technol.* **2023**, *12*, 578–593. [[CrossRef](#)]
3. Yin, X.; Wu, G.; Wei, J.; Shen, Y.; Qi, H.; Yin, B. Deep Learning on Traffic Prediction: Methods, Analysis, and Future Directions. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 3054840. [[CrossRef](#)]
4. Li, F.; Liu, K.; Chen, J. Traffic Status Prediction Based on Multidimensional Feature Matching and 2nd-Order Hidden Markov Model (HMM). *Sustainability* **2023**, *15*, 14671. [[CrossRef](#)]
5. Zhang, C.; Xie, Y.; Bai, H.; Yu, B.; Li, W.; Gao, Y. A Survey on Federated Learning. *Knowl. Based Syst.* **2021**, *216*, 106775. [[CrossRef](#)]
6. Zhang, R.; Wang, H.; Li, B.; Cheng, X.; Yang, L. A Survey on Federated Learning in Intelligent Transportation Systems. *arXiv* **2024**, arXiv:2403.07444.
7. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated Machine Learning: Concept and Applications. *ACM Trans. Intell. Syst. Technol.* **2019**, *10*, 1–19. [[CrossRef](#)]
8. Qiao, W.; Haghani, A.; Hamedi, M. A Nonparametric Model for Short-Term Travel Time Prediction Using Bluetooth Data. *J. Intell. Transp. Syst. Technol. Plan. Oper.* **2013**, *17*, 165–175. [[CrossRef](#)]
9. Kumar, S.V.; Vanajakshi, L. Short-Term Traffic Flow Prediction Using Seasonal ARIMA Model with Limited Input Data. *Eur. Transp. Res. Rev.* **2015**, *7*, 21. [[CrossRef](#)]
10. Agarap, A.F.M. A Neural Network Architecture Combining Gated Recurrent Unit (GRU) and Support Vector Machine (SVM) for Intrusion Detection in Network Traffic Data. In Proceedings of the ACM International Conference Proceeding Series, Macau, China, 26–28 February 2018.
11. Wang, J.; Deng, W.; Zhao, J.B. Short-Term Freeway Traffic Flow Prediction Based on Multiple Methods with Bayesian Network. *Jiaotong Yunshu Xitong Gongcheng Yu Xinxiji/J. Transp. Syst. Eng. Inf. Technol.* **2011**, *11*, 147.
12. Bharti; Redhu, P.; Kumar, K. Short-Term Traffic Flow Prediction Based on Optimized Deep Learning Neural Network: PSO-Bi-LSTM. *Phys. A Stat. Mech. Its Appl.* **2023**, *625*, 129001. [[CrossRef](#)]
13. Wang, X.X.; Xu, L.H. Short-Term Traffic Flow Prediction Based on Deep Learning. *Jiaotong Yunshu Xitong Gongcheng Yu Xinxiji/J. Transp. Syst. Eng. Inf. Technol.* **2018**, *18*, 81–88. [[CrossRef](#)]
14. Wang, H.; Wang, Z.; Li, L.; Kong, H.; Wang, Q.; Xu, J. Spatial-Temporal Prediction Model of Urban Short-Term Traffic Flow Based on Grid Division. *J. Comput. Appl.* **2022**, *42*, 2274–2280.
15. Zhang, J.; Zheng, Y.; Qi, D. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. In Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI 2017, San Francisco, CA, USA, 4–9 February 2017.
16. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 2978386. [[CrossRef](#)] [[PubMed](#)]
17. Yu, B.; Yin, H.; Zhu, Z. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In Proceedings of the IJCAI International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018.
18. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; Li, H. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *212*, 935152. [[CrossRef](#)]
19. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Li, P.; Bengio, Y.; Veličković, P.; Casanova, A.; Liò, P.; Cucurull, G.; et al. Graph Attention Networks. In Proceedings of the 6th International Conference on Learning Representations, ICLR 2018—Conference Track Proceedings 2018, Vancouver, BC, Canada, 30 April–3 May 2018.
20. Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wan, H. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. *Proc. AAAI Conf. Artif. Intell.* **2019**, *33*, 922–929. [[CrossRef](#)]
21. Liu, Y.; Yu, J.J.Q.; Kang, J.; Niyato, D.; Zhang, S. Privacy-Preserving Traffic Flow Prediction: A Federated Learning Approach. *IEEE Internet Things J.* **2020**, *7*, 7751–7763. [[CrossRef](#)]
22. Xia, M.; Jin, D.; Chen, J. Short-Term Traffic Flow Prediction Based on Graph Convolutional Networks and Federated Learning. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 1191–1203. [[CrossRef](#)]
23. Wang, H.; Zhang, R.; Cheng, X.; Yang, L. Federated Spatio-Temporal Traffic Flow Prediction Based on Graph Convolutional Network. In Proceedings of the 2022 IEEE 14th International Conference on Wireless Communications and Signal Processing, WCSP 2022, Nanjing, China, 1–3 November 2022.
24. Zhang, C.; Zhang, S.; Yu, J.J.Q.; Yu, S. FastGNN: A Topological Information Protected Federated Learning Approach for Traffic Speed Forecasting. *IEEE Trans. Industr. Inform.* **2021**, *17*, 8464–8474. [[CrossRef](#)]
25. Li, M.; Zhu, Z. Spatial-Temporal Fusion Graph Neural Networks for Traffic Flow Forecasting. In Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI 2021, Virtually, 2–9 February 2021; Volume 5A.
26. Aghabozorgi, S.; Seyed Shirkhorshidi, A.; Ying Wah, T. Time-Series Clustering—A Decade Review. *Inf. Syst.* **2015**, *53*, 16–38. [[CrossRef](#)]
27. Zhang, Y.; Xiang, T.; Hospedales, T.M.; Lu, H. Deep Mutual Learning. In Proceedings of the Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.

28. Brendan McMahan, H.; Moore, E.; Ramage, D.; Hampson, S.; Agüera y Arcas, B. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Proceedings of the Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, Ft. Lauderdale, FL, USA, 20–22 April 2017.
29. Han, X.; Gong, S. LST-GCN: Long Short-Term Memory Embedded Graph Convolution Network for Traffic Flow Forecasting. *Electronics* **2022**, *11*, 2230. [[CrossRef](#)]
30. Mirzahosseini, H.; Safari, F.; Hassannayebi, E. Estimation of Highway Capacity under Environmental Constraints vs. Conventional Traffic Flow Criteria: A Case Study of Tehran. *J. Traffic Transp. Eng.* **2021**, *8*, 751–761. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.