ISSN: 1813-4890

Research on traffic flow prediction based on spatio-temporal interactive dynamic graph convolutional network

Hongyan Wang [a], Qian Zhang [b], Man Li [c], Li Den [d]

Chongqing College of Mega City Digital Governance, Chongqing College of International Business and Economics, Chongqing, 401520, China.

[a]wanghongyan0905@yeah.net, [b]zhangqian20210701@163.com, [c]LiM012018@163.com, [d]2686739520@qq.com

Abstract

Accurate prediction of traffic flow is fundamental for effective urban traffic guidance and control, playing a vital role in intelligent traffic management. However, achieving precise traffic flow prediction remains a challenging task due to the intricate spatiotemporal dependencies involved. To address this issue and capture the dynamic spatiotemporal characteristics of traffic flow concurrently, this paper introduces a pioneering approach for traffic flow prediction: the Spatio-Temporal Interactive Dynamic Graph Convolutional Network (STIDGCN). Firstly, an interactive learning structure is devised to dynamically aggregate spatiotemporal features of hidden nodes within the traffic network. Secondly, a Gated Temporal Convolutional Network (Gated TCN) is constructed, employing dilated causal convolutional networks at multiple granularity levels to capture the temporal dependencies present in traffic flow. Moreover, the spatial correlation and temporal dependence of the traffic flow at each node and time step are adaptively controlled by a gated fusion mechanism. Experimental results substantiate that the proposed STIDGCN successfully extracts dynamic spatiotemporal features of traffic flow, delivering superior prediction performance compared to prevalent baseline methodologies.

Keywords

Traffic flow prediction, Dynamic spatiotemporal characteristics, Graph attention mechanism, Graph convolutional network.

1. Introduction

With the advancement of society and the economy, the challenges of road congestion and frequent traffic accidents have become increasingly prevalent. Consequently, enhancing the efficiency of public transportation utilization has emerged as a prominent research focal point, and traffic flow prediction stands as a valuable tool in addressing this issue effectively. Traffic flow forecasting endeavors to anticipate forthcoming traffic patterns by leveraging historical traffic data. By harnessing the valuable insights offered by traffic flow forecasts, transportation authorities can proactively intervene in traffic conditions, thereby mitigating congestion, enhancing transportation efficiency, and minimizing traffic accidents.

Traffic flow forecasting methodologies are primarily categorized into statistical methods, machine learning methods, and deep learning methods. Statistic methods typically employ statistical analysis principles to capture the temporal dependencies of traffic flow, often disregarding dynamic spatial characteristics. Consequently, the predictive performance of statistical methods is generally subpar. Common statistical models employed in traffic flow forecasting encompass the historical average model (HA)[1], autoregressive moving average model (ARMA)[2], and vector autoregressive model (VAR)[3]. However, due to the challenges posed by nonlinear traffic flow data and the limited ability to capture dynamic spatial characteristics, statistical forecasting methods often fall short. Consequently, an abundance of deep learning-based forecasting methods has emerged, as they demonstrate the capability to effectively address spatiotemporal traffic data and extract valuable insights. Mathew et al.[4] employed the K-nearest neighbor optimization classifier and the sine K-nearest neighbor optimization classifier to incorporate data-related information among traffic elements into the classification process. In a similar vein, Huang et al.[5] introduced a network architecture consisting of a deep confidence network and a regression model, enabling more precise traffic prediction. Nevertheless, the aforementioned machine learning methods still fail to account for the dynamic spatial characteristics of traffic flow within the traffic flow prediction process.

Considering the intricate influence of the road network structure on the traffic situation, it is noticeable that adjacent nodes tend to exhibit similar traffic states as vehicles traverse the roads. In order to comprehensively capture the temporal and spatial characteristics, it is imperative to incorporate spatial characteristics during the traffic flow prediction process. Such an approach not only enables the extraction of both temporal and spatial attributes but also enhances the accuracy of long-term traffic flow prediction. Consequently, there has been a significant proliferation of deep learning-based traffic flow forecasting methods, as they have proven to be highly effective in addressing these challenges. For instance, Shi et al.[6] proposed a traffic flow forecasting network that integrates Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). By incorporating CNN to learn the spatial structure characteristics of the road network, the model effectively captures the temporal and spatial correlations of traffic flow. Similarly, Ke et al.[7] employed a combination of Convolutional LSTM, standard LSTM, and a convolution layer FCL-Net to predict future traffic flow, taking into account the temporal and spatial dependencies. However, it is important to note that traffic flow data exhibits non-Euclidean characteristics, rendering CNN less effective in extracting features from such data. Consequently, the Graph Convolutional Network (GCN), which excels in handling non-Euclidean data, has gained widespread adoption in traffic flow forecasting. As an example, Li et al.[8] introduced the utilization of Graph GCN to construct a self-learning graph convolutional module, aimed at capturing the spatial characteristics of traffic flow. However, their approach did not delve into the extraction of the underlying spatiotemporal relationships within the traffic flow. On the other hand, T-GCN[9] leveraged both the feature matrix and adjacency matrix to capture temporal and spatial information, respectively. ASTGCN[10] synergistically combines spatiotemporal convolution and a spatiotemporal attention mechanism to effectively learn the intricate characteristics of traffic flow in both temporal and spatial domains. In contrast, STSGCN[11] constructs a local spatiotemporal graph by connecting each node with its neighboring nodes in the preceding and succeeding time steps. However, STSGCN does not explicitly account for the concealed spatiotemporal relationships between nodes, limiting its ability to handle long time series adequately. Similarly, STGCN[12] combines GCN with gated Temporal Convolutional Networks (TCN) to extract both temporal and spatial characteristics of traffic flow. Nonetheless, it fails to fully address the intricate and dynamic nature of the temporal and spatial characteristics within traffic flow data. Existing traffic flow forecasting methods face challenges in modeling, data processing, and model training, struggling to effectively capture spatiotemporal dependencies, dynamic changes, and uncertainties. This limits their ability to simultaneously extract dynamic spatiotemporal features of traffic flow, resulting in suboptimal prediction accuracy and difficulties in long-term forecasting.

To address the aforementioned challenges, this study introduces a novel approach called Spatio-Temporal Interactive Dynamic Graph Convolutional Network (STIDGCN) for accurate traffic flow prediction. The interactive learning module serves as a pivotal component within our proposed framework, predominantly consisting of Interactive Dynamic Graph Convolution

Network (IDGCN) implemented with a strategic divide and conquer approach. The temporal dependence of traffic flow is captured using Gated Temporal Convolutional Network (Gated TCN) with extended causal convolution networks of varying granularity levels.

The primary contributions are summarized as follows:

1. We propose STIDGCN, a traffic flow prediction model that integrates interactive learning, gated temporal convolution, and gated fusion mechanism. This model offers a comprehensive exploration of the dynamic spatiotemporal features within traffic flow data, while simultaneously reducing computational complexity through parallel computation.

2. We developed an interactive learning module utilizing dynamic graph convolutions to capture spatiotemporal dependencies effectively. The module incorporates a divide-and-conquer approach, with one dynamic graph convolution generating dynamic graphs by fusing adaptive and learnable adjacency matrices. Additionally, we introduced a Gated TCN that captures the temporal dependence of traffic flow using dilated causal convolutional networks with varying granularities.

3. Extensive experiments on two traffic datasets demonstrate that STIDGCN outperforms existing popular baseline methods, achieving superior prediction performance.

2. Related Works

In the domain of modeling spatial features in traffic flow, the primary methods encompass traditional convolutional networks, recurrent neural networks, graph convolutions, and more. Traditional convolutional networks excel at extracting local features from standard grid data, while their applicability is limited. In contrast, GCN display remarkable proficiency in learning from nonlinear graph-structured data, allowing them to effectively capture the comprehensive features of the data.

Currently, two prominent methods for GCN prevail: the spatial domain-based method and the spectral domain-based method. The spatial domain-based method leverages the aggregation of feature information from neighboring nodes to extract node features. While this approach offers computational efficiency, the selection of an appropriate node neighborhood poses a formidable challenge. Niepert et al.[13] put forward a linear approach for selecting the neighborhood of the central node and demonstrated that accurate node neighborhood selection can significantly enhance prediction accuracy. In a related context, Li et al.[14] incorporated graph convolution into human action recognition, employing the division of each node's neighborhood into distinct subsets. In the context of traffic flow analysis, the utilization of neighborhood information for nodes is regarded as prior knowledge and remains constant throughout the training process. Spectral domain-based methodologies employ spectral analysis to aggregate the neighborhood information of each node. However, a notable limitation of such approaches arises from the fact that GCN operates on the entire graph, necessitating the processing of the complete graph in each iteration. This computational requirement poses a significant challenge due to its substantial computational complexity. Hence, to enhance the computational efficiency of the model, Bruna et al.[15] introduced a generalized graph convolution framework based on Laplace operators. Building upon this framework, Yu et al.[12] proposed a gated graph convolution network for traffic flow prediction, aiming to capture the dynamic characteristics of traffic flow. However, this model did not account for the dynamic spatiotemporal dependencies inherent in traffic flow.

Attention mechanisms have experienced rapid development and are widely applied in various fields, including speech recognition, natural language processing, and image processing. For instance, Guo et al.[16] combined spatiotemporal graph convolution with self-attention mechanisms to capture the dynamic spatiotemporal features of traffic flow. Shi et al.[17] developed a spatiotemporal neural network based on self-attention mechanisms to capture dynamic spatiotemporal correlations. Zheng et al.[18] proposed a traffic flow prediction method based on a spatiotemporal attention mechanism, effectively extracting dynamic spatial features and complex temporal dependencies. Furthermore, Zheng et al.[19] embedded self-attention mechanisms into the graph convolutional network to construct a traffic flow prediction network capable of capturing the dynamic spatiotemporal dependencies of traffic flow. Although these methods use self-attention mechanisms to capture spatiotemporal dependencies, most approaches often overlook the stacked implicit relationships and hidden spatiotemporal features within the channel dimensions, thereby weakening the model's ability to capture spatiotemporal features.

STIDGCN draws inspiration from the above studies and addresses both the temporal dependency and spatial correlation problems present in traffic flow. This is achieved by integrating GCN, TCN and multi-head graph attention mechanisms to effectively model traffic flow data. The interactive learning strategy is able to comprehensively explore the dynamic spatio-temporal features in the traffic flow, while TCN can effectively capture the intricate temporal dependencies and hidden dynamic spatial features. In addition, a gated fusion mechanism is adopted to intelligently combine the dynamic spatio-temporal features derived from TCN-a and TCN-b to minimize the error propagation in the prediction process and improve the prediction accuracy.

3. Methodology

3.1. Problem Definition

The traffic road network is represented as a graph $G=(V,E,A)$, where $|V|=N$ denotes the set of nodes representing observation sensors within the road network. The set of edges E connects the nodes, with edge weights corresponding to the distances between the nodes. The initial adjacency matrix $A\in\mathbb{R}^{N\times N}$ is generated from graph G. A_{ij} is assigned a value of 1 when $v_i,v_j\in V$ and $(v_i,v_j)\in E$, and 0 otherwise. By utilizing the initial adjacency matrix A, derived from the original traffic network, as prior knowledge, the future traffic flow $\hat{X}^{T+1:T+H}=\left\{x_G^{t+1},x_G^{t+2},\cdots,x_G^{T+H}\right\}$ is predicted based on the historical time series $X^{1:T}=\left\{x_G^1,x_G^2,\cdots,x_G^t,\cdots,x_G^T\right\}$. In this context, $x_G^t\in\mathbb{R}^{N\times C}$ represents the observation of graph G at time t, C denotes the number of feature channels, T denotes the length of the given historical time series, and H denotes the length of the predicted future traffic sequence. The mapping relationship for the traffic flow prediction problem can be expressed as follows:

$$\hat{X}^{T+1:T+H}=\Gamma\left(X^{1:T};\Theta\right) \tag{1}$$

where $X^{1:T}\in\mathbb{R}^{T\times N\times F}$, $\hat{X}^{T+1:T+H}\in\mathbb{R}^{H\times N\times F}$, F denotes the output feature dimension of each node in the STIDGCN model, while Θ encompasses all the learnable parameters within the model. Additionally, Γ symbolizes the prediction function.

3.2. Framework of STIDGCN

Fig.1 illustrates the overall framework of the STIDGCN model, comprising an input layer, an interactive learning module, stacked STAHGCN modules, and an output layer. Notably, the interactive learning module incorporates an interactive strategy with an IDGCN, which employs Dynamic Graph Convolution Network (DGCN). The STAHGCN module encompasses the Gated TCN and the Spatial Gate Fusion, and each STAHGCN layer is connected residually. The Gated

TCN employs an expansive causal convolutional network to effectively capture temporal dependencies.

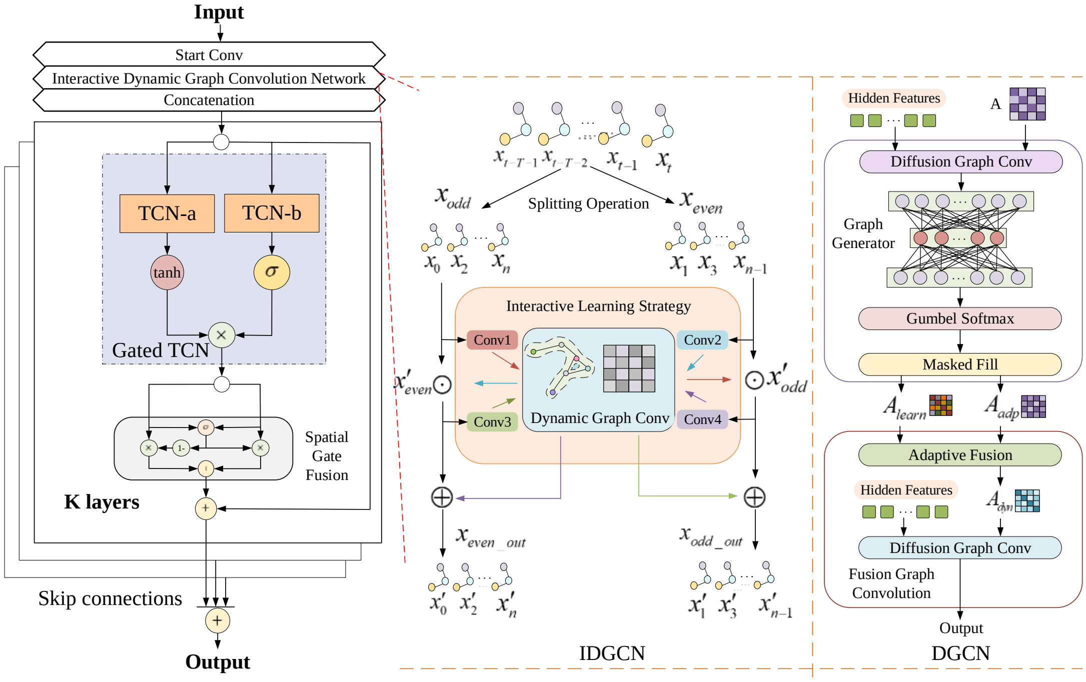

Fig.1 Overall framework of STIDGCN model

3.2.1. Interactive Learning

In contrast to methods relying on CNN and TCN, GCN exhibits notable efficiency in handling non-Euclidean data, superior ability to learn the temporal-spatial dependence of traffic flow, and more comprehensive capture of complex time dependencies and dynamic spatial characteristics. Consequently, the proposed STIDGCN model leverages both CNN and GCN to construct an interactive learning module. By employing an interactive learning framework that partitions the input data into intervals, the model effectively acquires a comprehensive understanding of the dynamic spatiotemporal characteristics of traffic flow.

(1) Interactive Dynamic Graph Convolution Network (IDGCN)

To enable interactive learning, dynamic graph convolutional networks are seamlessly integrated into IDGCN, facilitating the exchange of their learned dynamic spatiotemporal features. Within the IDGCN framework, the input sequence is partitioned into two subsequences, enabling the shared learning of their respective dynamic spatiotemporal features and thus facilitating the simultaneous extraction of these features. Additionally, each subsequence shares parameter weights and incorporates a 1D convolutional preprocessing step to expand the receptive field and enhance the overall representation of the features.

The input for IDGCN is denoted as $X \in \mathbb{R}^{T\times N\times F}$, which undergoes interleaving and sampling to acquire two subsequences, referred to as parity sequences, namely $X_{odd} \in \mathbb{R}^{T/2\times N\times F}$ and $X_{even} \in \mathbb{R}^{T/2\times N\times F}$. The output of the initial interactive learning phase in IDGCN is represented by $X'_{odd} \in \mathbb{R}^{T/2\times N\times F}$ and $X'_{even} \in \mathbb{R}^{T/2\times N\times F}$. Furthermore, X'_{odd} and X'_{even} are subject to subsequent interactive learning, resulting in the final output sequences denoted as $X'_{odd\_out} \in \mathbb{R}^{T/2\times N\times F}$ and $X_{even\_out} \in \mathbb{R}^{T/2\times N\times F}$. The specific operations within IDGCN are indicated as follows:

$$X_{even}, X_{odd} = Split(X) \tag{2}$$

$$X'_{odd} = tanh\left(DGCN\left(Conv1\left(X_{even}\right)\right)\right) \odot X_{odd} \tag{3}$$

$$X'_{even} = tanh\left(DGCN\left(Conv2\left(X_{odd}\right)\right)\right) \odot X_{even} \tag{4}$$

$$X_{odd\_out} = X'_{odd} + tanh\left(DGCN\left(Conv3\left(X'_{even}\right)\right)\right) \quad (5)$$

$$X_{even\_out} = X'_{even} + tanh\left(DGCN\left(Conv4\left(X'_{odd}\right)\right)\right) \quad (6)$$

Where the symbols $Conv1$, $Conv2$, $Conv3$, and $Conv4$ represent 1D convolution operations, $tanh$ denotes the Hadamard product and activation function, additionally, $DGCN$ represents the dynamic graph convolution utilized within IDGCN.

(2) Dynamic Graph Convolution Network (DGCN)

The DGCN consists of two key components: the graph generation network and the fusion graph convolution. The graph generation network is responsible for generating a learnable adjacency matrix that emulates the time-varying dynamic associations between nodes. The fusion graph convolution combines two adjacent matrices and performs graph convolution. Moreover, DGCN leverages diffusion graph convolution and the graph generation network to deeply learn the dynamic spatial characteristics, thereby enhancing the spatial heterogeneity capturing capability of STIDGCN and ultimately improving its overall performance. The DGCN takes the hidden feature $\ell \in \mathbb{R}^{T\times N\times F}$ and the predefined initial adjacency matrix $A \in \mathbb{R}^{N\times N}$ as inputs, which are then passed through the diffusion graph convolutional network. The output of this network is subsequently fed into the Graph Generator, resulting in the generation of the discrete matrix $A' \in \mathbb{R}^{N\times N}$, which contains spatiotemporal information. The representation of A' is as follows:

$$X_{even\_out} = X'_{even} + tanh\left(DGCN\left(Conv4\left(X'_{odd}\right)\right)\right) \quad (7)$$

Where GCN represents the operations of diffusion convolution and Graph Generator, MLP denotes the multilayer perceptron.

To ensure the feasibility of the sampling process during training, STIDGCN employs Gumbel reparameterization:

$$X_{even\_out} = X'_{even} + tanh\left(DGCN\left(Conv4\left(X'_{odd}\right)\right)\right) \quad (8)$$

Where $g \sim Gumbel(0,1)$ represents a random variable, τ signifies the temperature value of 0.5, and A_{learn} denotes a neighbor matrix generated by the graph generator that effectively captures the dynamic dependencies among nodes. Furthermore, an adaptive adjacency matrix $A_{apt} \in \mathbb{R}^{N\times N}$, which does not rely on any prior knowledge, is constructed and can be represented as follows:

$$X_{even\_out} = X'_{even} + tanh\left(DGCN\left(Conv4\left(X'_{odd}\right)\right)\right) \quad (9)$$

Where $E_1 \in \mathbb{R}^{N\times c}$ and $E_2^T \in \mathbb{R}^{N\times c}$ represent the learnable parameters, and the initial value of A_{apt} is a predefined adjacency matrix $A \in \mathbb{R}^{N\times N}$ derived from the original graph data.

The DGCN utilizes an adaptive fusion module to combine A_{learn} and A_{apt}, resulting in the dynamic adjacency matrix $A_{dyn} \in \mathbb{R}^{N\times N}$. This matrix is then fed into the diffusion graph convolutional network to capture the hidden dynamic spatiotemporal correlations within the traffic road. The fusion module operates as follows:

$$A_{dyn} = \alpha A_{apt} + (1-\alpha) A_{learn} \quad (10)$$

where α denotes the learnable adaptive parameter factor.

STIDGCN employs diffusion map convolution within the graph generator network, along with fusion graph convolution and concatenation fusion modules. The input for diffusion graph convolution is uniformly defined as $X_{in} \in \mathbb{R}^{T\times N\times F}$. Specifically, in the graph generator network, the diffusion map convolution is defined as follows:

$$GCN\left(X_{in}, A_{apt}\right) = \sum_{k=0}^{K} A_{apt}^{k} X_{in} W \tag{11}$$

Where k is the diffusion step size, K is the maximum number of diffusion steps, and W denotes the parameter matrix.

In the fusion graph convolution module, A_{dyn} represents the adjacency matrix of the input for fusion graph convolution. The diffusion graph convolution can be expressed as follows:

$$GCN\left(X_{in}, A_{dyn}\right) = \sum_{k=0}^{K} A_{dyn}^{k} X_{in} W \tag{12}$$

Within the Concatenation module, the dynamic spatiotemporal features derived from the interactive learning structure are merged in a time-indexed order. These features are then passed to the diffusion graph convolution layer to effectively capture and rectify the entire time series features. Notably, in contrast to previous studies, STIDGCN utilizes both the predefined initial neighbor matrix $A \in \mathbb{R}^{N \times N}$ and the dynamic neighbor matrix $A_{dyn} \in \mathbb{R}^{N \times N}$ generated by the interactive learning structure in the Diffusion Graph Convolution. Furthermore, $P_f = A / rowsum(A)$ and $P_b = A^T / rowsum\left(A^T\right)$ represent the forward and backward transfer matrices of the initial neighbor matrix A, respectively. In this context, the representation of the diffusion graph convolution in the concatenation fusion module is as follows:

$$GCN\left(X_{in}, A, A_{dyn}\right) = \sum_{k=0}^{K}\left(A_f^k X_{in} W_1 + A_b^k X_{in} W_2 + A_{dyn}^k X_{in} W_3\right) \tag{13}$$

The DGCN module serves to extract profound latent spatial features by uncovering the hidden dependencies among nodes in the traffic network. By incorporating the DGCN module into IDGCN, STIDGCN can effectively harness its capability to capture dynamic spatiotemporal features, thereby enhancing its overall performance in this regard.

3.2.2. Gated TCN

The Gated TCN component within the STIDGCN model is comprised of two parallel time convolution modules, namely TCN-a and TCN-b. TCN-a is primarily responsible for processing the time series data of each node, enabling the capture of time dependence and trend changes. Following TCN-a, TCN-b is introduced to operate at a higher level of temporal abstraction. By processing the output of TCN-a through additional convolutional layers and incorporating residual connections, TCN-b effectively extracts higher-level temporal features and patterns. Moreover, each time-space map volume layer is equipped with a residual connection to preserve essential information throughout the model. In this study, the time convolution layer utilizes an extended causal convolution network to effectively capture the time dependence of nodes. By employing a stacked configuration of convolution layers, the extended causal convolution network achieves an expanded receptive field, enabling a more comprehensive understanding of temporal relationships. Additionally, the extended causal convolution adopts a specific step size during the sliding operation, facilitating efficient processing of long time series data. Moreover, non-recursive parallel computing techniques are employed to enhance learning speed and alleviate the issue of gradient vanishing, resulting in improved training performance.

Utilizing extended causal convolution with a kernel size of 2 and an expansion factor of k, the input is subsampled every k steps, and standard 1D convolution is applied to the selected input. Considering a one-dimensional sequence input $x \in \mathbb{R}^{H}$ and a filter $\varpi \in \mathbb{R}^{K}$, the expanded causal convolution operation of x and ϖ at step t is illustrated in Equation 14.

$$x * \varpi(t) = \sum_{s=0}^{K-1} \varpi(s) x(t - d \times s) \tag{14}$$

Where d is an expansion factor that controls the jump step size.

The expanded causal convolution network employed in STIDGCN significantly increases the receptive field of the time convolution network layer by stacking expanded causal convolution layers with expansion factors in ascending order. This approach enables the model to effectively capture longer sequences using fewer layers, resulting in computational resource savings without compromising prediction accuracy. By expanding the receptive field, STIDGCN enhances its capability to capture and incorporate long-term dependencies in the input data, thereby improving the accuracy of traffic flow prediction. By providing the input $X \in \mathbb{R}^{T\times N\times F}$, the Gated TCN can be expressed as follows:

$$h = g(\zeta_1 * X + b) \odot \sigma(\zeta_2 * X + c) \tag{15}$$

Where ζ_1, ζ_2, b and c are the model parameters, $\odot$ is the product of elements, $g(\cdot)$ is the activation function of the output, and $\sigma(\cdot)$ is the Sigmoid function which determines the rate of information transmission to the subsequent layer.

3.2.3. Spatial Gate Fusion

STIDGCN incorporates the Spatial Gate Fusion module. The computation of the spatiotemporal gated fusion mechanism can be expressed as follows:

$$H^{(L)} = z \cdot H_S^{(L)} + (1-z) \cdot H_T^{(L)} \tag{16}$$

$$z = sigmoid\left(H_S^{(L)} W_{z,1} + H_T^{(L)} W_{z,2} + b_z\right) \tag{17}$$

Among them, $W_{z,1} \in R^{D\times D}$, $W_{z,2} \in R^{D\times D}$, $b_z \in R^D$ are learnable parameters, and z is the gate. The gating fusion mechanism adaptively controls the spatial correlation and time dependence of the traffic flow in each node and time step.

4. Experiment

4.1. Data Description

The robustness and effectiveness of the STIDGCN model's predictive capabilities have been rigorously evaluated using two prominent public transport datasets: METR-LA and PEMS-BAY. The METR-LA dataset encompasses a comprehensive collection of traffic speed statistics obtained from 207 sensors installed along the highways in Los Angeles County over a span of four months. Similarly, the PEMS-BAY dataset comprises a rich compilation of traffic speed information recorded by 325 sensors strategically placed across the traffic roads within the San Francisco Bay Area, spanning a duration of six months.

Table 1 Description of experimental dataset

Data	METR-LA	PEMS-BAY
Type	sequentially	sequentially
Attribute	speed	speed
Location	highways of Los Angeles	the Bay Area
Edges	1515	2369
Time Steps	34272	52116
Nodes	207	325

Both METR-LA and PEMS-BAY datasets meticulously recorded essential information such as detection location, date of detection, and data type. In our experimental setup, we adhered to a chronological split of the dataset into training, testing, and validation sets, maintaining a proportion of 7:2:1 respectively. This division ensured a robust evaluation framework for predicting traffic flow at various time intervals, namely 15 minutes, 30 minutes, and 60 minutes.

To provide a comprehensive overview of the experimental dataset, we present the detailed specifics in Table 1.

4.2. Parameter Setting

In the conducted experiment, the historical traffic data spanning one hour is employed to accurately forecast the traffic flow for the subsequent 60 minutes, encompassing 12 distinct horizons. The entirety of the data sets is meticulously partitioned, following a chronological order, into three subsets: the training set, the verification set, and the test set, maintaining a proportionate division ratio of 6:2:2 respectively. The experiment was conducted utilizing the computational capabilities of a powerful system comprising 22 vCPU AMD EPYC 7T83 64-Core Processor with a RTX 4090 GPU Card. For the purpose of modeling, an 8-layer network architecture was employed, incorporating a sequential spreading factor sequence of 1, 2, 1, 2, 1, 2, 1, 2. Additionally, the model training process employed the Adam optimizer, initialized with a learning rate of 0, and integrated a discard rate parameter is $p = 0.3$. To facilitate a comprehensive analysis of the experimental outcomes and assess the predictive capabilities of the model, the mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) metrics are employed for evaluating the disparities between the actual traffic flow speed and the corresponding prediction results. These metrics enable a thorough evaluation of the model's prediction performance. These metrics are calculated as follows:

$$MAE = \frac{1}{\kappa}\sum_{i=1}^{\kappa}\left|\hat{\partial}_i - \partial_i\right| \tag{18}$$

$$MAPE = \frac{1}{\kappa}\sum_{i=1}^{\kappa}\left|\frac{\hat{\partial}_i - \partial_i}{\hat{\partial}_i}\right| \times 100\% \tag{19}$$

$$RMSE = \sqrt{\frac{1}{\kappa}\sum_{i=1}^{\kappa}\left(\hat{\partial}_i - \partial_i\right)^2} \tag{20}$$

Where κ is the number of samples, $\hat{\partial}_i$ and ∂_i denote the ground truth and the predicted value of the i-th sample, respectively.

4.3. Baselines

Comparative analysis is conducted between STIDGCN and the following models:

(1) HA[1]: Historical average model that relies on historical average traffic flow information for forecasting.

(2) VAR[3]: Vector Auto-Regression model.

(3) SVR[21]: Support Vector Regression utilizing a linear vector machine to train the model and predict traffic flow based on input-output relationships.

(4) ARIMA[22]: Auto-Regressive Integrated Moving Average model with Kalman filter.

(5) FC-LSTM[22]: Recurrent neural network with fully connected LSTM hidden units.

(6) WaveNet[23]: Convolutional Neural Network designed for sequence data prediction.

(7) Graph WaveNet[24]: Combination of graph convolution and extended accidental convolution in a WaveNet model.

(8) STGCN[12]: Spatial-temporal graph convolution network that combines graph convolution and 1D convolution.

(9) STSGCN[11]: Spatial-Temporal Synchronous Graph Convolutional Network that captures spatial-temporal characteristics by stacking multiple local GCN layers in the time direction.

(10) MRes-RGNN[25]: Utilizes graph neural networks to jointly capture graph-based spatial dependencies and temporal dynamics.

4.4. Experimental Results

Table 2 presents a comprehensive comparison between STIDGCN and 12 commonly used baseline models in terms of 15-minute, 30-minute, and 60-minute predictions. Remarkably, STIDGCN demonstrates significant superiority over the baseline models across all evaluation indicators on the two datasets. Specifically, in the 15-minute, 30-minute, and 60-minute predictions on the METR-LA dataset, STIDGCN outperforms the most advanced method, MRes-RGNN, by 0.5%, 3.7%, 4.8% in MAE, 2.6%, 4.5%, 1.3% in RMSE, and 3.3%, 4.3% in MAPE, respectively. On the PEMS-BAY dataset, the improvements are even more pronounced, with STIDGCN surpassing MRes-RGNN by 4.9%, 5.1%, 6.1% in MAE, 5.6%, 0.5%, 0.3% in RMSE, and 5.3%, 2.5%, 0.8% in MAPE, respectively.

Statistical methods such as HA, VAR, and ARIMA, along with traditional machine learning techniques SVR and FC-LSTM, exhibit limited prediction accuracy due to their failure to consider spatial correlation. In contrast, spatiotemporal GCN models, exemplified by STGCN and STSGCN, demonstrate superior prediction performance by effectively handling non-Euclidean traffic data. On the other hand, MRes-RGNN employs a gating mechanism and comprehensively considers the periodic patterns between long- and short-term traffic flows. Furthermore, it mitigates the gradient vanishing problem in model training through the utilization of a residual network.

In contrast to the baseline model, STIDGCN leverages interactive dynamic graph convolution to construct an interactive learning strategy. It combines TCN with adaptive and dynamic graphs, as well as GCN and a multi-head graph attention mechanism, to fully extract the dynamic spatiotemporal features of traffic flow. The interactive learning strategy effectively exploits the dynamic attributes of spatiotemporal traffic data. STIDGCN achieves this by constructing an adaptive adjacency matrix in the static adaptive graph learning module and utilizing a graph attention network in the dynamic graph learning network, enabling effective capture of the dynamic associations among hidden nodes in the road network over time. Additionally, the stacked expansive causal convolutional network and multi-head graph attention mechanism in STIDGCN drive effective long-term prediction. Compared to the baseline model, STIDGCN demonstrates the best prediction performance. As depicted in Fig.7, it is evident that with increasing training time, STIDGCN exhibits improved training performance, higher prediction accuracy, and excellent long-term prediction capabilities.

Table 2 The performance comparison of different models on METR-LA and PEMS-BAY datasets

Data	Models	15min			30min			60min		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
METR-LA	HA	4.16	7.80	13.00%	4.16	7.80	13.00%	4.16	7.80	13.00%
	VAR	4.42	7.89	10.20%	5.41	9.13	12.7%	6.52	10.11	15.80%
	SVR	3.99	8.45	9.30%	5.05	10.87	12.10%	6.72	13.76	16.70%
	ARIMA	3.99	8.21	9.60%	5.15	10.45	12.70%	6.90	13.23	17.40%
	FC-LSTM	3.44	6.30	9.60%	3.77	7.23	10.90%	4.37	8.69	13.20%
	WaveNet	2.99	5.89	8.04%	3.59	7.28	10.25%	4.45	8.93	13.62%
	GWN	2.98	5.90	7.92%	3.59	7.29	10.26%	4.43	8.97	13.64%
	STGCN	2.88	5.74	7.62%	3.47	7.24	9.57%	4.59	9.40	12.70%
	STSGCN	3.31	7.62	8.06%	4.13	9.77	10.29%	5.06	11.66	12.91%
	MRes-RGNN	2.71	5.38	7.36%	3.14	6.53	8.63%	3.61	7.62	10.55%

	STIDGCN	**2.70**	**5.18**	**7.01%**	**3.06**	**6.24**	**8.52%**	**3.49**	**7.29**	**10.07%**
PEMS-BAY	HA	2.88	5.59	6.80%	2.88	5.59	6.80%	2.88	5.59	6.80%
	VAR	1.74	3.16	3.60%	2.32	4.25	5.00%	2.93	5.44	6.50%
	SVR	1.85	3.59	3.80%	2.48	5.18	5.50%	3.28	7.08	8.00%
	ARIMA	1.62	3.30	3.50%	2.33	4.76	5.40%	3.38	6.50	8.30%
	FC-LSTM	2.05	4.19	4.80%	2.20	4.55	5.20%	2.37	4.96	5.70%
	WaveNet	1.39	3.01	2.91%	1.83	4.21	4.16%	2.35	5.43	5.87%
	GWN	1.39	3.01	2.89%	1.83	4.21	4.11%	2.35	5.43	5.78%
	STGCN	1.36	2.96	2.90%	1.81	4.27	4.17%	2.49	5.69	5.79%
	STSGCN	1.44	3.01	3.04%	1.83	4.18	4.17%	2.26	5.21	5.40%
	MRes-RGNN	1.41	2.96	2.94%	1.79	3.86	3.91%	2.09	4.76	4.95%
	STIDGCN	**1.34**	**2.81**	**2.76%**	**1.69**	**3.84**	**3.90%**	**1.98**	**4.64**	**4.91%**

In order to provide a comprehensive understanding of STIDGCN, we present the experimental results of STIDGCN alongside FC-LSTM, Graph WaveNet, and STGCN models on the PEMS-BAY dataset, as depicted in Fig.2, Fig.3 and Fig.5. Analyzing the three figures, it is evident that the prediction performance of STIDGCN surpasses that of FC-LSTM, Graph WaveNet, and STGCN models significantly. This observation indicates that STIDGCN adeptly captures the dynamic spatiotemporal characteristics of traffic flow. Furthermore, the prediction errors exhibit slower growth as the prediction duration increases. Notably, for prediction durations longer than 15 minutes, the prediction errors of STIDGCN are consistently and significantly lower than those of the other comparative models. These findings affirm the superior prediction performance of STIDGCN in long-term prediction scenarios.

Additionally, in Fig.7, we present the ground true and predicted values of STIDGCN for the 30th minute (Horizon 3) and 60th minute (Horizon 12) at nodes 700 to 1200 within the PEMS-BAY dataset. It is evident that all of our models exhibit commendable prediction accuracy, successfully capturing the fluctuations in the data. As the prediction time step increases, the prediction accuracy for Horizon 12 slightly declines compared to Horizon 3. This phenomenon can be attributed to the inherent challenge of accurately capturing the complex dynamic spatiotemporal characteristics of traffic flow for long-term prediction. Nevertheless, as depicted in Figure 6 and Figure 7, STIDGCN demonstrates exceptional performance by effectively predicting the trend of the data, even for long-term predictions.

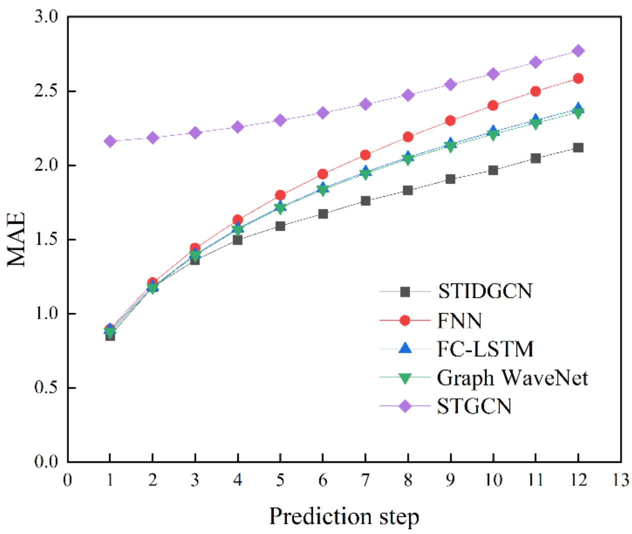

Fig.2 Comparison of MAE of different models on PEMS-BAY dataset

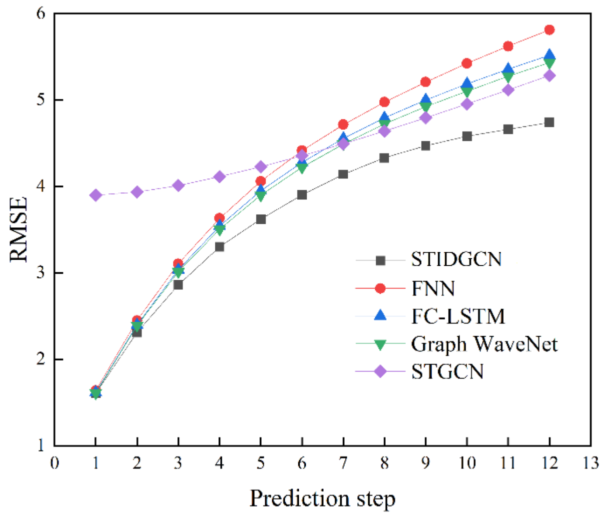

Fig.3 Comparison of RMSE of different models on PEMS-BAY dataset

Fig.4

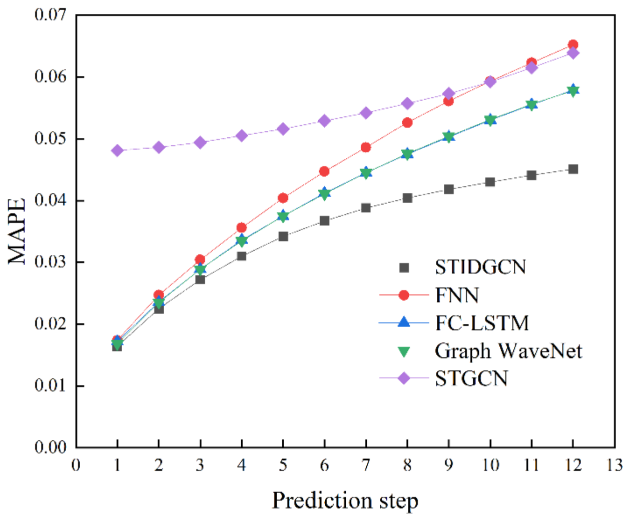

Fig.5 Comparison of MAPE of different models on PEMS-BAY dataset

Fig.6

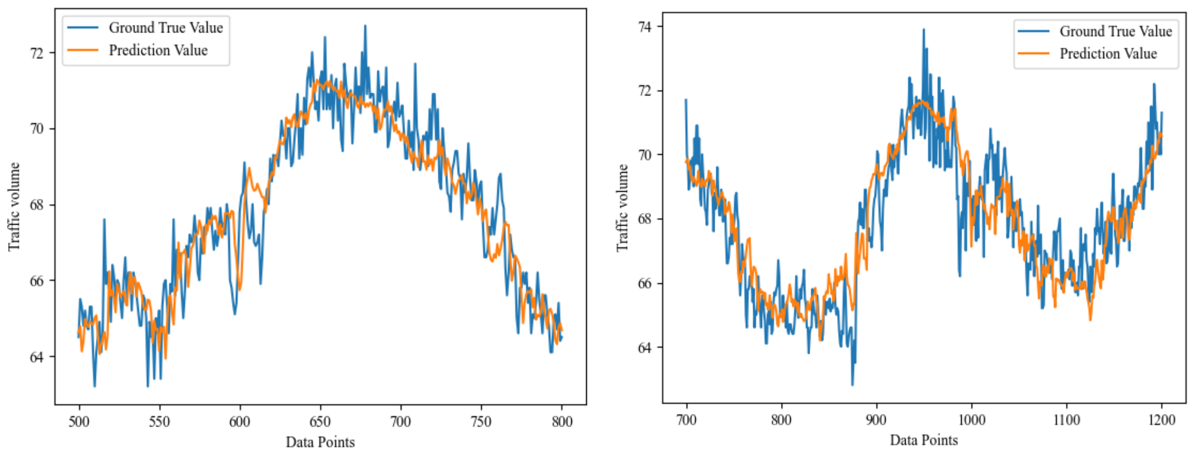

Fig.7 Visualization of traffic prediction on PEMS-BAY dataset (left: Horizon 3, right: Horizon 12)

5. Conclusion

This paper introduces a pioneering traffic flow prediction model, namely STIDGCN, designed to forecast traffic conditions at various time intervals within a road network. STIDGCN represents an innovative spatio-temporal interactive dynamic graph convolutional network capable of effectively capturing the dynamic spatiotemporal characteristics of the intricate road network. It achieves this by autonomously learning the underlying dynamic relationships among road nodes over time, without relying on any prior knowledge. Consequently, STIDGCN excels at capturing the dynamic spatiotemporal features of traffic flow, enabling accurate predictions. STIDGCN combines the strengths of GCN, TCN, and the multi-head graph attention mechanism, allowing it to inherit their respective advantages. The interactive learning strategy effectively extracts dynamic features from spatiotemporal traffic data, while the inclusion of Gated TCN enables the model to precisely capture complex temporal patterns. Experiments on two real-world datasets demonstrate that STIDGCN consistently outperforms baseline methods, showcasing its superior predictive capabilities.

Future research will focus on developing interactive dynamic graph convolution and dynamic adaptive graph generation networks, incorporating external factors such as weather, traffic incidents, and holidays to further enhance the accuracy of traffic flow prediction.

Acknowledgements

This work is supported by Scientific Research Program of Chongqing College of International Business and Economics (Natural Science Programs). (KYZK2024019).

References

[1] J. Liu and W. Guang: A summary of traffic flow forecasting methods, J. Highway Transp. Res. Develop, Vol. 21 (2004) No. 3, p. 82-85.

[2] B.M. Williams and L.A. Hoel: Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results[J].Journal of Transportation Engineering, Vol. 129 (2003) No. 6, p. 664-672.

[3] E. Zivot and J. Wang: Vector Autoregressive Models for Multivariate Time Series, New York, NY, USA: Springer. (2003).

[4] A. Mathew and F. Rawther: Hadoop based short-term traffic flow prediction on D2its using correlation model and KNN HSsine, Proc.IEEE Int. Conf. Power Control Signals Instrum. Eng. (ICPCSI). (2017) p. 1123-1129.

[5] W. Huang, G. Song, H. Hong and K. Xie: Deep Architecture for Traffic Flow Prediction: Deep Belief Networks with Multitask Learning, IEEE Transactions on Intelligent Transportation Systems, Vol. 15 (2014) No. 5, p. 2191-2201.

[6] X. Shi, Z. Chen, H. Wang, D.Y. Yeung, W.K. Wong, W.C. Woo: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, Proceedings of the 31st International Conference on Neural Information Processing Systems, Vol. 1 (2015), p. 802-810.

[7] J. Ke, H. Zheng, H. Yang and X. Chen: Short-Term Forecasting of Passenger Demand Under On-Demand Ride Services: A Spatio-Temporal Deep Learning Approach, Transportation Research Part C: Emerging Technologies, Vol. 85 (2017), p. 591-608.

[8] Y. Li, R. Yu, C. Shahabi, Y. Liu: Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting, International Conference on Learning Representations (2018).

[9] L. Zhao, Y. Song, C. Zhang, Y. Liu, H. Li: T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction, IEEE Transactions on Intelligent Transportation Systems, Vol. 21 (2020) No. 9, p. 3848-3858.

[10] S. Guo, Y. Lin, N. Feng, C. Song and H. Wan: Attention Based Spatiotemporal Graph Convolutional Networks for Traffic Flow Forecasting, Proceedings of the 33rd AAAI Conference on Artificial Intelligence, Vol. 33 (2019) No. 1, p. 922-929.

[11] C. Song, Y. Lin, S. Guo and H. Wan: Spatial-Temporal Synchronous Graph Convolutional Networks: A New Framework for Spatiotemporal Network Data Forecasting, Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34 (2020) No. 1, p. 914-921.
[12] B. Yu, H. Yin and Z. Zhu: Spatial-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting, Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32 (2018), p. 3634-3640.
[13] M. Niepert, M. Ahmed and K. Kutzkov: Learning Convolutional Neural Networks for Graphs, International Conference on Machine Learning, Vol. 1 (2016), p. 2014-2023.
[14] C. Li, Z. Cui, W. Zheng, C. Xu and J. Yang: Spatio-Temporal Graph Convolution for Skeleton Based Action Recognition, Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32 (2018), p. 3482-3489.
[15] J. Bruna, W. Zaremba, A. Szlam and Y. Lecun: Spectral Networks and Locally Connected Networks on Graphs, International Conference on Learning Representations (2014).
[16] S. Guo, Y. Lin, N. Feng, C. Song and H. Wan: Attention Based Spatiotemporal Graph Convolutional Networks for Traffic Flow Forecasting, Proceedings of the 33rd AAAI Conference on Artificial Intelligence, Vol. 33 (2019) No. 1, p. 922-929.
[17] X. Shi, H. Qi, Y. Shen, G. Wu and B. Yin: A Spatiotemporal Attention Approach for Traffic Prediction, IEEE Transactions on Intelligent Transportation Systems, Vol. 22 (2020) No. 8, p. 4909-4918.
[18] C. Zheng, X. Fan, C. Wang and J. Qi: GMAN: A Graph Multi-Attention Network for Traffic Prediction, Proceedings of the 34th AAAI Conference on Artificial Intelligence, Vol. 34 (2020), p. 1234-1241.
[19] G. Zheng, W. Chai and V. Katos: A Dynamic Spatiotemporal Deep Learning Framework for Traffic Speed Prediction on Large-Scale Road Networks, Expert Systems with Applications, Vol. 195 (2022), p. 116585.
[20] Y. Wang, N. Geroliminis and L. Leclercq: Recent Advances in ITS, Traffic Flow Theory, and Network Operations, Transportation Research Part C: Emerging Technologies, Vol. 68 (2016), p. 507-508.
[21] F. Zhou, Q. Yang, K. Zhang, G. Trajcevski, T. Zhong and A. Khokhar: Reinforced Spatiotemporal Attentive Graph Neural Networks for Traffic Forecasting, IEEE Internet of Things Journal, Vol. 7 (2020) No. 7, p. 6414-6428.
[22] J. Zhao, et al.: Attention-Based Dynamic Spatiotemporal Graph Convolutional Networks for Traffic Speed Forecasting, Expert Systems with Applications, Vol. 204 (2022), p. 117511.
[23] Y. Wang, et al.: An Improved WaveNet Network for Multi-Step-Ahead Wind Energy Forecasting, Energy Conversion and Management, Vol. 278 (2022), p. 116709.
[24] Q. Zhang, C. Li, F. Su and Y. Li: Spatiotemporal Residual Graph Attention Network for Traffic Flow Forecasting, IEEE Internet of Things Journal, Vol. 10 (2023) No. 13, p. 11518-11532.
[25] C. Chen, K. Li, S.G. Teo, X. Zou, Z. Zeng: Gated Residual Recurrent Graph Neural Networks for Traffic Prediction, Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33 (2019), p. 485-492.