ISSN: 2414-1895 DOI: 10.6919/ICJE.202011_6(11).0013

Algorithm Model of Traffic Spatio-temporal Sequence Prediction Based on Attention Mechanism

Di Wang, Xueming Zhai

School of North China Electric Power University, Baoding 071000, China.

Abstract

As one of the basic technologies of mobile network intelligent management and scheduling, traffic forecasting has been widely used in network resource scheduling, auxiliary base station site selection and other fields in recent years. However, mobile network traffic sequences often have both time and space dependence. Traditional time series forecasting methods such as autoregressive integrated moving average (ARIMA) model and long-short-term memory (LSTM) model often only pay attention to the time characteristics of the sequence one-sidedly. It ignores the spatial characteristics and is often interfered by irrelevant information in the time dimension. In response to the above problems, this paper proposes an algorithm based on VP-Tree and attention mechanism, through VP-Tree to build a spatial combination, based on the seq2seq structure, introduces the attention mechanism so that the model can identify the key information of the long sequence and assign it to Appropriate weight of each information. Experiments show that the algorithm proposed in this paper takes into account the temporal and spatial characteristics of the traffic sequence, and has higher prediction accuracy than some traditional time-series forecasting algorithms.

Keywords

Traffic prediction, Spatiotemporal data, Vp-tree, Attention mechanism.

1. Introduction

At present, mobile communication technology is evolving to the 5th generation (5G) cellular technology. There is a need to build 5g base stations around the world, but compared with previous base stations, 5g base stations have a sharp increase in power consumption, and a single 5g base station has signal coverage The area is smaller than that of traditional base stations, and the huge construction and operating costs of 5g base stations have imposed a huge burden on telecom operators. Compared with traditional base stations, 5g base stations require more scientific planning and construction and intelligent scheduling management. As one of the basic technologies for intelligent management and scheduling of mobile networks, traffic forecasting has been used by telecom operators in such fields as network resource scheduling, base station optimization and energy saving, and auxiliary base station site selection in recent years. At present, a lot of research is devoted to mining the inherent relationship between time and space of traffic data. Research on the prediction of traffic data based on time and space can help telecom operators grasp the future demand for mobile network resources in a region, so as to expand 5g in China. Construction plays an important role in supporting decision-making. It can be said that accurate forecasting of traffic can help telecom operators improve the stability of mobile networks, allocate network resources rationally, and improve users' network experience, especially to comply with the new needs of future 5g technology and save 5g for operators. The cost of base station construction and operation is of great significance to promote the intelligent transformation of the network management center.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202011_6(11).0013

Aiming at the problem that traditional time series forecasting algorithms often underutilize time and space information, this paper proposes a mobile network traffic forecasting algorithm based on VP-Tree and attention mechanism. The algorithm in this paper mainly consists of two parts. The first part builds a binary tree that divides the space based on VP-Tree to form a spatial combination between communities. The second part forms the input tensor for traffic prediction based on the spatial combination. Introduce a layer of attention mechanism to construct the fusion of external factors and neural network, and the two together determine the output of the neural network (that is, the traffic forecast value).

2. Algorithm model description

2.1 Problem Description

The mobile network load capacity of a single community is provided by a single or multiple base stations covering the location. Assuming there are a total of N communities, the base stations will automatically collect information at regular intervals, including PRB utilization, number of mobile call users, and mobile traffic The attribute characteristics of L dimensions, that is, each community will generate a series of time series data, denoted as X. The time series data of one of the communities is designated as the target of prediction, and the time series data generated by the K communities associated with the community are selected from other communities as the correlation vector. Given all the characteristic data of a specified community within a time window length T, predict the traffic value of the mobile network in the community during the next T+m time period.

2.2 Algorithm model description

Community mobile network traffic is not only related to changes in the time dimension, spatial characteristics and the influence of external factors are equally important. However, traditional algorithms such as ARIMA and LSTM only use the historical traffic data information of a single cell and often ignore the spatial relationship between cells. The core idea of this algorithm is to first construct a prediction tensor by mining the spatial relative relationship between the cells and the attribute characteristics of the target community, and combine the attention mechanism to use the deep neural network to analyze and predict the traffic sequence. The overall schematic diagram of the algorithm proposed in this paper is as follows:

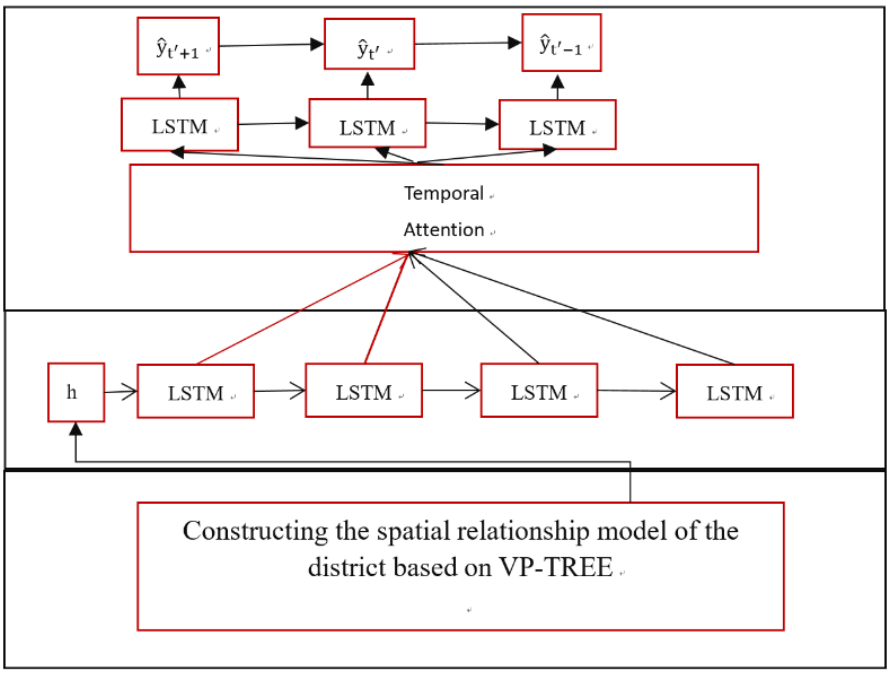

Figure 1. Traffic prediction algorithm model based on vp-tree and attention mechanism

ISSN: 2414-1895 DOI: 10.6919/ICJE.202011_6(11).0013

2.2.1 Establish spatial combination based on VP-TREE

The method proposed in this paper constructs a spatial combination by finding adjacent cells within M meters of the target cell. VP-TREE is a binary tree structure based on distance to measure space proposed by Uhlmann in 1991. It mainly uses the difference between the commanding height and the target point. The distance information between the space is divided. Assuming that the data set containing the latitude and longitude information of all cells is $S = \{S_1, S_2, \ldots., S_n\}$, using the VP-TREE algorithm, randomly select a cell S_h in the data set S as the commanding height of VP-TREE, and the cell to be searched for For S_q, the purpose is to quickly find several cells whose distance from cell S_q is less than r in the formed VP-TREE by constructing VP-TREE:

(1) When |S|=0, the tree is empty, return;

(2) When |S| is not 0, randomly select a cell S_h as the commanding height of VP-TREE, and calculate the distance from other n-1 cells to cell S_h to form a data set as $D = \{D_1, D_2, \ldots, D_{n-1}\}$.

(3) Calculate the median value M in the data set D. Elements smaller than the distance M are assigned to the left subtree, and elements larger than the distance M are assigned to the right subtree.

(4) According to the above method, recursively build the left subtree and the right subtree to form VP-TREE

2.2.2 Attention mechanism

There is an obvious problem when predicting the traffic of the community at a time. The traffic at a time does not simply rely on each unit of the input vector, but mainly refers to the historical data at key time points. If the attention mechanism is not introduced, the encoder layer will be the last As the initial state of the decoder, the state is used to predict the value of the next hour, and so on, the pre-training model is established until the value at the time of prediction, then due to the seq2seq mechanism, the state transmitted to the decoder is constant. That is to say, the weight assigned to each time point is consistent, and the influence of each time on a certain time point in the next moment is constant. This is obviously unreasonable. This phenomenon will cause the predicted curve to become too " Straight", lack of trend change. Therefore, this paper introduces the attention mechanism to make the algorithm focus on information at key time points in the prediction process and reduce the interference of irrelevant information.

The attention mechanism is actually to add an AttentionLayer layer after the encoder to calculate the similarity between each expected input of the decoder and each time point of the encoder, thereby obtaining an attention matrix. During the decoding process of the encoder, the predicted output of each time point All correspond to different context variables, so it will reflect a better trend.

If mathematical notation is used to represent the core idea of the attention mechanism, the encoder and decoder layers process each vector as follows:

Encoder:

$$\mathrm{h_i} = \tan\mathrm{h}\,(\mathrm{W}[\mathrm{h_{i-1}}, \mathrm{x_i}]) \tag{1}$$

$$\mathrm{o_i} = \mathrm{softmax}(\mathrm{Vh_i}) \tag{2}$$

Decoder:

$$c_{t'} = \sum_{i=1}^{T} \alpha_{t'i} h_i \tag{3}$$

$$\alpha_{t'i} = \frac{exp(e_{t'i})}{\sum_{k=1}^{T} exp(e_{t'k})} \tag{4}$$

$$e_{t'i} = v_d^T \tan h\,(W_d[d_{t'-1}; h_i]) \tag{5}$$

Among them, $c_{t'}$ corresponds to the vector at time t'; $e_{t'i}$ corresponds to the hidden layer state of the Encoder at time i in the Encoder process h_i vs. the hidden layer state at time t' in the Decoder $s_{t'}$; the probability of $e_{t'i}$ is normalized to $\alpha_{t'i}$ through the SoftMax function.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202011_6(11).0013

2.2.3 Encoder-decoder

At the beginning of training, given the initial random seed tree of the network, the number of training steps and the learning rate, set the minimized loss function as the training target and apply the Adam optimizer to update the neural network weights. The loss function during the training process uses MSE as the error calculation The formula, expressed in mathematical symbols, can be defined as:

$$\text{loss} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{6}$$

The attribute features in the time series data of the target community and the attribute features contained in the K communities whose distance from the target community is less than d are selected to form a spatial combination, and the input tensor constructed by the two is defined as $I_0 = \{i_1, i_2, \dots, i_t\}$, Standardize all the elements contained in the input tensor sequence I_0. Put the standardized sequence I_0 into the encoder layer, and then summarize the hidden layer states output by each LSTM unit to generate an intermediate semantic vector $\mathrm{H} = \mathrm{q}(h_1, h_2, \dots, h_t)$. By calculating the similarity between the intermediate semantic vector and the expected output vector of the decoder layer, the weight matrix c_t can be obtained. The predicted value $\hat{y}_{t'}^{i}$ of the decoder at t' will be composed of the context vector c_t formed according to the attention weight matrix, and the predicted value $\hat{y}_{t'-1}^{i}$ of the decoder at the previous time.Finally iteratively output the prediction results, the output prediction vector $y = \{y_1, y_2, \dots, y_{t'}\}$, indicating that the prediction of the neural network is based on the traffic data at the previous t The traffic value of the target community at t' time in the future.

3. Experimental results and analysis verification

3.1 Datasets

The data in this experiment is selected from the urban cell traffic statistics table of the telecom operator and the corresponding base station data information. By arranging the data in ascending order of time and deduplicating the data, the communities in the area are obtained from July 2, 2019 to July 22, 2019. A total of 10,302 detailed traffic records with hour as the smallest statistical unit per day. The statistics of this data set are shown in the following table:

Table 1. Data set information statistics table

Datasets	Cell traffic data and related information statistics
Target	Mobile network traffic value
Time	July 2, 2019 to July 22, 2019
Time unit	The smallest unit is hour
Feature	PRB utilization, number of community users, community latitude and longitude information

3.2 Experiment platform

The computer configuration used in this experiment is as follows: the processor model is Intel Core i7-7820HQ, the cpu reference frequency is 2.90GHz, the memory is 32GB, the graphics card is NVIDIA Quadro M1200, and the operating system used is Ubuntu 16.04. This experiment runs on the Ubuntu system through the integrated development environment for Pycharm Community Edition 2020.1.4. The design language used in the program is Python 3.7.8 (64-bit). During this period, Tensorflow is called to build the neural network, and Pandas performs data statistics. With analysis, Numpy processes the data, and matplotlib processes the graphics.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202011_6(11).0013

3.3 Hyperparameters

According to the hardware performance of the experimental platform, the batch size of the neural network is set to 64, the learning rate is 0.002, the dropout is set to 0.15, the loss function uses MSE, the neural network weight optimizer uses adam algorithm, and the specified filter is within 1000 meters from the target cell. Community. The flow sequence is based on hours. The input sequence contains one week of data. The characteristics of the input sequence select the flow value, PRB utilization rate, the number of community users and the flow value of other cells within 700 meters of the target cell. The output sequence contains each 24 hours of the day. The hourly flow value.

3.4 Baselines

(1) ARIMA: Differential integrated moving average autoregressive model. The model not only takes into account the dependence of main factors on the time series, but also takes into account the interference of random fluctuations. It is a very common and effective time series forecasting analysis method.

(2) GBRT: Progressive gradient regression tree, as an algorithm with strong generalization ability, the core is that each tree is learned from the residuals of all previous trees, and it has a good predictive time series real value application.

(3) SVM: Support vector machine, which occupies a very important position in traditional machine learning, and has been verified to be used in time series forecasting.

(4) LSTM: Long and short-term memory model, which solves the problems of gradient disappearance and gradient explosion in traditional RNN networks through a unique "gate structure", and has a good predictive effect on time series.

(5) Seq2seq: Seq2seq is an Encoder-Decoder structure network, which was first applied to NLP-related fields such as machine translation, speech recognition, chatbots, etc., as a method that can generate another sequence based on a given sequence and a specific method The method of a series can also be applied to the forecast of time series.

3.5 Evaluation Metrics

This article mainly adopts two indicators to compare the accuracy of models between different algorithms, namely RMSE and MAE. RMSE is often used as a standard for measuring the prediction results of machine learning models and the true value. It refers to the square root of the ratio of the square of the difference between the observed value and the true value to the number of observations. MAE can better reflect the actual situation of the predicted value error, it is the average of the absolute error. The definitions of RMSE and MAE in mathematical notation are as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_t - y_t)^2} \tag{7}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|(x_t - y_t| \tag{8}$$

Among them, N represents the overall length of the time sequence, x_t represents the true value of the time sequence at time t, and $y_t$represents the predicted value of the time sequence at time t.

3.6 Analysis of results

A piece of flow data is randomly selected in the data test set for verification. The comparison between the flow value predicted by the algorithm proposed in this article and the actual value, and the comparison between the predicted value and the actual value of other time series forecasting models are as follows:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202011_6(11).0013

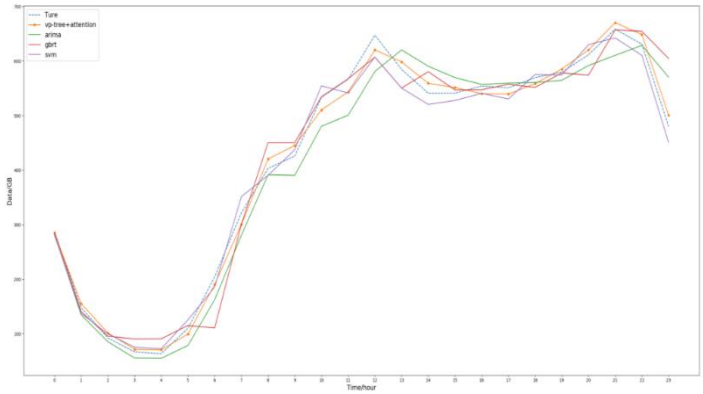

Figure 2. Comparison between model predicted value and actual value

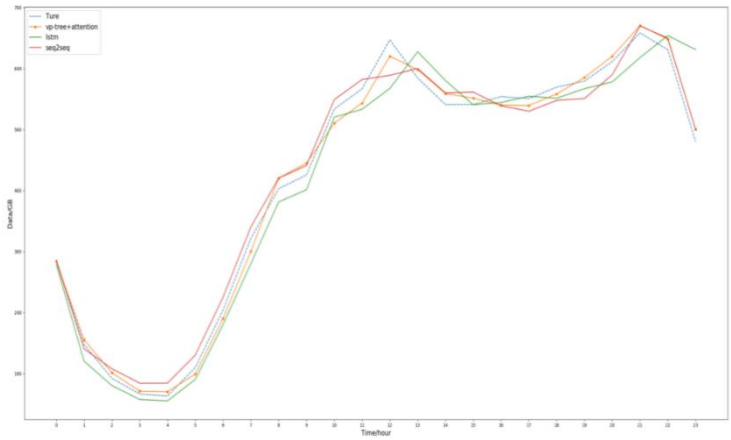

Figure 3. Comparison chart of model and traditional neural network prediction model

As shown in the figure, the predicted value generated by the traffic prediction model based on vp-tree and attention mechanism proposed in this paper has a small error with the actual value, and it reflects the trend of traffic changes within 24 hours. Ordinary ARIMA and LSTM are very serious for the accumulation of errors in multi-step forecasting, and there is obvious lag. The reason can be imagined. Because the previous sliding window is all true values in single-step forecasting, the error accumulation is However, once the rolling forecast is adopted, the value in the sliding window may be based on the forecast value at the previous moment, and basically every forecast value has a deviation, so the accumulation of errors will be more serious, so in the end, although A 24-hour traffic forecast can be obtained, but the deviation will be greater. The GBRT algorithm itself is suitable for small data samples, but the data set used in this article is too large. It can be seen from the results that the prediction effect of GBRT is not good. Compared with SVM, the biggest difficulty of SVM is the choice of kernel function and parameter adjustment. The choice of kernel function and the level of parameter adjustment will greatly affect the accuracy of prediction. The algorithm proposed in this paper is based on the attention mechanism. It is a black box model with better versatility, and from the experimental results, it also guarantees the prediction accuracy. Finally, the problem of Seq2seq in the flow prediction process is that the key information in the sequence cannot be effectively paid attention to, which makes the error too large, and the response trend is not particularly consistent with the actual change trend. This is also the reason why the algorithm in this article introduces the attention mechanism. The algorithm proposed in this article actually adopts the basic structure of Seq2Seq and introduces the vp-tree algorithm. By constructing a cell space combination, the algorithm proposed in this article reduces the error, and between the encoder and the decoder, The

ISSN: 2414-1895 DOI: 10.6919/ICJE.202011_6(11).0013

attention mechanism is further introduced to help this algorithm better reflect the trend of traffic changes.

Table 2. Model prediction accuracy statistics table

Method	MAE	RMSE
ARIMA	30.84	39.74
GBRT	25.48	33.65
SVM	19.45	28.79
LSTM	29.25	37.48
Seq2seq	20.78	29.96
VP-Tree+Attention	14.72	25.56

As shown in the table, the MAE of this method is reduced by 52.27%, 42.23%, 24.32%, 49.68%, and 29.16% compared with ARIMA, GBRT, SVM, LSTM, and Seq2seq. The RMSE is compared with ARIMA, GBRT, SVM, LSTM, and Seq2seq. It decreased by 36.44%, 29.94%, 12.23%, 32.61%, 15.69% in turn. Through experimental observation, it can be seen that the algorithm proposed in this paper has obvious advantages in forecast accuracy compared with traditional time series forecasting methods.

4. Conclusion

In this paper, the traditional ARIMA and LSTM tend to only pay attention to the time characteristics of the sequence one-sidedly, ignore the spatial characteristics, and are often interfered by irrelevant information in the time dimension, and propose a neural network based on VP-Tree and Attention. The network model is used to solve the problem of mobile network traffic data prediction. From a spatial perspective, this algorithm uses the Haversine formula to calculate the distance of multiple community data according to latitude and longitude, and uses VP-Tree to construct a binary tree to divide the community. From the perspective of time, select the eigenvalues that affect the change of community traffic on the time scale, and construct the input tensor together with community combination. The overall Seq2Seq architecture is adopted, and Temporal Attention is introduced between the encoder and the decoder, so that the prediction model can obtain the ability to intercept key information in the long-term sequence, so that the prediction model can better cope with the periodic factors of traffic changes. Through the above methods, the algorithm model proposed in this article takes into account the time angle and the space angle, and can effectively predict the changing trend of the traffic sequence.

Experimental results show that the algorithm route proposed in this paper has better versatility than traditional mathematical prediction methods, and has achieved higher prediction accuracy than traditional neural network algorithms. However, there is room for further optimization in this model. First, the process of constructing the network is more complicated, the data needs to be preprocessed, and there are many hyperparameters that need to be set. If the data is improperly processed or the parameters are not set properly, it is difficult to have good results. Second, the amount of data required is large, the training process consumes hardware resources seriously, and the training time is long. How to simplify the neural network model and reduce the training time without reducing the prediction accuracy will become the focus of future research work.

Acknowledgements

This thesis was actually completed with the help of the tutor and classmates. The tutor’s profound professional knowledge, rigorous academic attitude, excellence in work style, tireless noble teacher ethics, strict self-discipline, lenient and noble demeanor, simple and approachable personality charm It has a profound impact on me. I can't help but set up ambitious academic goals and master the basic research methods, but I still understand many principles of dealing with others and dealing with others. From the selection of the topic to the completion of this thesis, every step is completed under the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202011_6(11).0013

guidance of the supervisor, and a lot of hard work has been devoted to the supervisor. Here, I would like to express my high respect and heartfelt gratitude to the tutor!

References

[1] HUANG C M, CHIANG M S, DAO D T, et al. Y2V data offloading for cellular network based on the Software Defined Network (SDN) inside Mobile Edge Computing (MEC) architecture[J]. IEEE Access, 2018, 6(99): 17741 -17755.

[2] SAXENA N, SAHU B J R, HAN Y S. Traffic-aware energy optimization in green LTE cellular systems [J]. IEEE Communications Letters, 2014, 18(1):38 -41.

[3] Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015-2020[EB/OL]. White Paper, Feb. 2016.

[4] Barbour A D, Holst L, Janson S. Poisson approximation[M]. Oxford: Clarendon Press, 1992, 51(51):299-316.

[5] Leland W E, Taqqu M S, Willinger W, et al. On the self-similar nature of Ethernet traffic (extended version) [J]. Networking, IEEE/ACM Transactions on, 1994, 2(1): 1-15.

[6] Zhu L, Wang Y, Fan Q. MODWT-ARMA model for time series prediction[J]. Applied Mathematical Modelling, 2014, 38(5): 1859-1865.

[7] Haviluddin H, Alfred R. A genetic-based backpropagation neural network for forecasting in time-series data[C] //International Conference on Science in Information Technology. 2015.

[8] ZHOU X, ZHAO Z, LJ R, et al. The predictability of cellular networks tra ffic [C]// Proceedings of the 2012 International Symposium on Communications and Information Technologies. Piscataway, NJ: IEEE, 2012:973 -978.

[9] Xingjian Shi. Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model[C].// Neural Information Processing Systems 30 (NIPS 2017).

[10] Di Chai, Leye Wang . Bike Flow Prediction with Multi-Graph Convolutional Networks[C]// 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.