A Dynamic Embedding Method for Passenger Flow Estimation

IIAI AAI 2021

Wei-Yi Chung, Yen-Nan Ho, Yu-Hsuan Wu, Jheng-Long Wu

Presentation: Wei-Yi Chung Date: 2021/07/12

Outlines

Introduction

PART ONE

Introduction

Problem

Public transportation has become an indispensable means of transportation in people's daily lives. Therefore, passenger flow estimation become an important issue, and most of past research used the static traffic embedding in flow prediction, only used corresponding vector space to obtain embeddings.

Motivation

Traffic data will have different embedding feature due to time and flow values. Most of the past research using static embedding method e.g. Word2Vec, this will result in the same features no matter what kind of traffic flow input value.

Solution

This paper proposes to use BERT as a dynamic embedding pre-training model, and transfer it to the GMAN flow estimation model as embedding, and the experimental results show that dynamic embedding does bring good performance on passenger flow estimation.

Related Works

PART TWO

Related Works

- Input data of flow estimation

Name of paper	Year	Input Data type
Traffic flow prediction with big data: a deep learning approach	2014	Temporal
Using LSTM and GRU neural network methods for traffic flow prediction	2016	
A hybrid deep learning based traffic flow prediction method and its understanding	2018	Spatial, Temporal
GMAN: A graph multi-attention network for traffic prediction	2020	
Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting	2020	

- Pre-trained embedding in flow estimation

Name of paper	Year	Embedding type
Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data	2019	Static
Deep and embedded learning approach for traffic flow prediction in urban informatics	2019	
Learning dynamic graph embedding for traffic flow forecasting: A graph self-attentive method	2019	
Predicting dynamic embedding trajectory in temporal interaction networks	2019	
Embedding Traffic Network Characteristics Using Tensor for Improved Traffic Prediction	2020	

Related Works

- Method of flow estimation

Name of paper	Year	Method type
Using LSTM and GRU neural network methods for traffic flow prediction	2016	LSTM, GRU
Short-term traffic speed forecasting based on attention convolutional neural network for arterials	2018	Attention-CNN
An effective spatial-temporal attention based neural network for traffic flow prediction	2019	Attention
Attention based spatial-temporal graph convolutional networks for traffic flow forecasting	2019	Attention-GCN
A cascaded R-CNN with multiscale attention and imbalanced samples for traffic sign detection	2020	Attention-RCNN

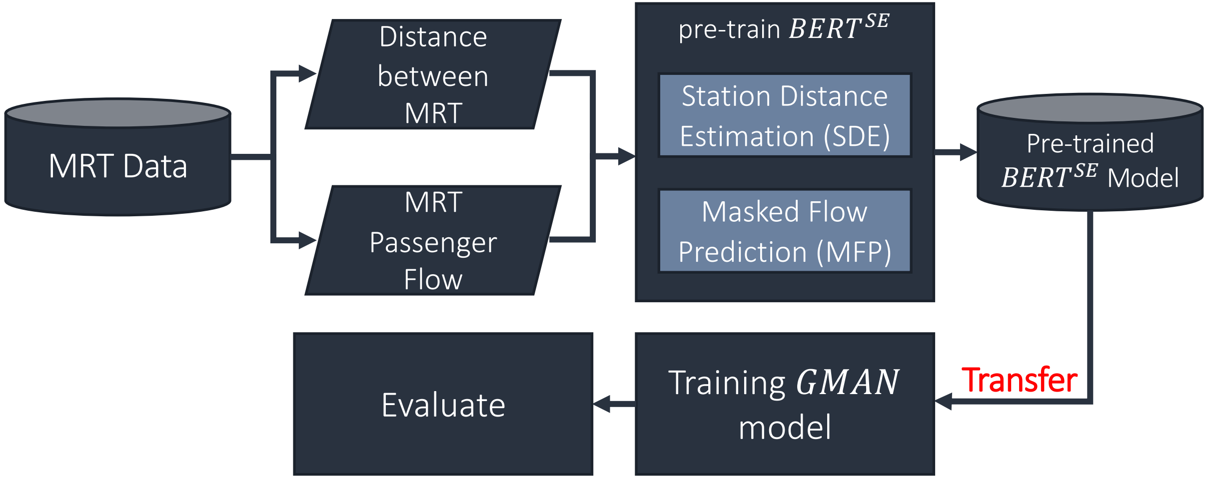
Proposed Method

PART THREE

Proposed Method

Overview of research structure

- modified BERT model as a pre-training dynamic embedding model to representing MRT stations
- transfer to the GMAN model to predict passenger flow



Proposed Method

Pretraining BERT model

- Use the original BERT (Delvin *et. al*, 2018) architecture and replace the target layer for pre-training

- The model input two passenger flow sequence at the same time stamp from two random sites:

$$H^{out} = BERT^{SE}([X^a, X^b])$$

where H^{out} is the hidden state from BERT, X^a, X^b are the flow sequence.

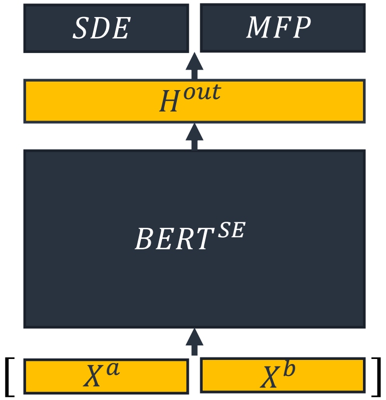
- Two traffic related pre-training target:
  - Station Distance Estimation (SDE):
    - Using the first hidden state h_1^{out} in H^{out} to estimate the distance between two sequence's resource station y_{SDE} :

$$\hat{y}_{SDE} = W_{SDE} h_1^{out} + b_{SDE}$$

    - This target allows the model to learn the spatial features
  - Masked Flow Prediction (MFP):
    - Following MLM concept to predict the masked flow y_{FP} in input sequence:

$$h_i^{FP} = \tau(\varphi(W_{FP1} h_i^{out} + b_{FP1}))$$
$$\hat{y}_{FP} = W_{FP2} h_i^{FP} + b_{FP2}$$

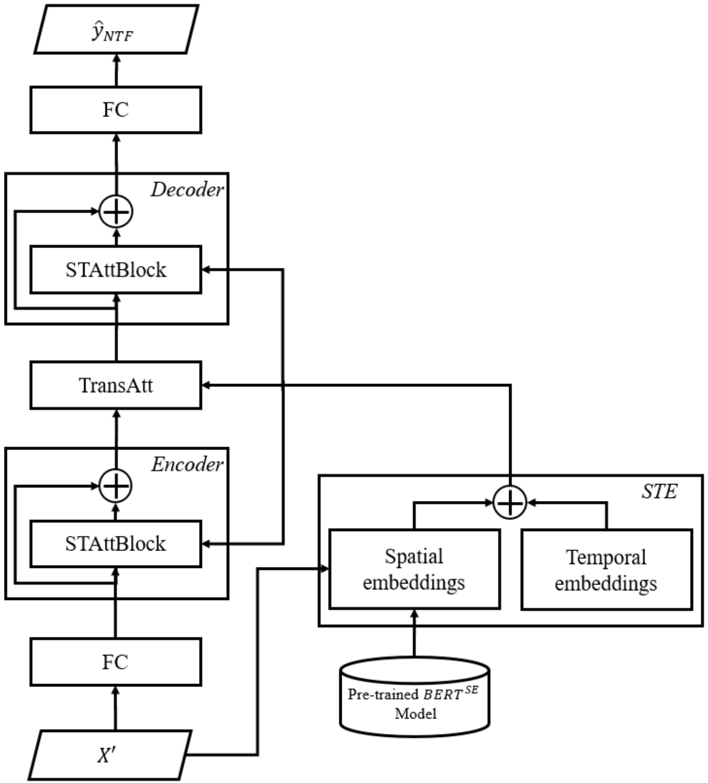
      where φ is the Gaussian error linear unit activation function, τ is layer normalization
    - This target allows the model to learn the flow change pattern



Proposed Method

Structure of GMAN Model

- Encoder-Decoder structure
  - Encoder
    Use historical time embedding to learn historical traffic flow.
  - Decoder
    Use future time embedding to learn how to predict future traffic flow.
- FC
  - Fully-connected which connected every neuron between data and encode-decode layer.
- STAttBlock (STAttention block)
  - Spatial Attention
  - Temporal Attention
  - Gated Fusion
- TransAtt (transform attention layer)
  - Link historical time steps and future time steps
- STE
  - Temporal embedding → Composed of *day of the week* and *time of day.*
  - Spatial embedding → Get through *Node2vec* or *pre-trained BertSE.*



Proposed Method

Training GMAN Model

- Input the original flow sequence X and output of $BERT^{SE}$ as spatial embedding, and predict the next time stamp passenger flow y_{NTF}:

$$\hat{y}_{NTF} = GMAN(X, BERT^{SE}(X))$$

where X denotes the input flow data, $GMAN$ denotes the GMAN model, $\hat{y}_{NTF}$ denotes the predict value of passenger flow.

- Since the passenger flow is the numerical data, we using mean absolute error (MAE) as evaluation metric of $GMAN$ model loss $L(\theta_{GMAN})$:

$$L(\theta_{GMAN}) = |y_{NTF} - \hat{y}_{NTF}|$$

where y_{NTF} denotes the true passenger flow value, θ_{GMAN} denotes all parameter of GMAN model.

Experimental Results

PART FOUR

Experimental Results

Experimental Data Description

- The dataset used in the paper is the passenger flow data of Taipei MRT in Taiwan from January 2017 to March 2019, including 20 stations, 10 of transfer and 10 of non-transfer.
- To reduce the difference in passenger flow between different sites, the paper uses the logarithm method to normalize all data.

	Non-transfer station	Transfer station
Max. flow	10,566	46,225
Min. flow	0	0
Avg. flow	1,167	4,532
St. D.	1,356	5,329
Number of example	172,710	172,710

Experimental Results

Evaluation Metric

- Since the target of passenger flow prediction is numerical, in the evaluation of the results for these models, we use two common indicators of regression tasks

A. Mean Absolute Error (MAE)

$$MAE = \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

B. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{N}}$$

Experimental Results

Comparison on Spatial-Embedding

- This section compares the performance of the static embedding $node2vec^{SE}$ and dynamic embedding $BERT^{SE}$ proposed by the paper.

Spatial-embedding	Transfer		Non-transfer	
	MAE	*RMSE*	*MAE*	*RMSE*
$node2vec^{SE}$	1.47	0.89	1.23	0.30
$BERT^{SE}$	0.93	0.72	1.24	0.33

- The MAE and RMSE on $BERT^{SE}$ at the transfer station have reached 0.93 and 0.72, respectively, which is much better than node2vec model $node2vec^{SE}$.

- The result shows that the greater the amount of traffic change, the better the stability of the dynamic embedding estimation performance.

Experimental Results

Comparison on Time Length of History Passenger Flow

- This section compares compare the performance of using different time length of historical time point input data into the $GMAN$ model. The best performance can be achieved in $BERT^{SE}$ when a single historical time point is used.

Time length of historical passenger flow	$BERT^{SE}$		$node2vec^{SE}$	
	MAE	*RMSE*	*MAE*	*RMSE*
1	0.85	1.51	2.05	2.94
3	2.56	3.64	1.69	2.17
5	1.97	2.86	1.57	1.99

- The $BERT^{SE}$ have reached 0.85 and 1.51 on MAE and RMSE, respectively. The high impact information on traffic flow prediction is from the previous time point.

- The results of $node2vec^{SE}$, the longer the input length, the better the prediction performance. This is due to the characteristics of the structure of the $GMAN$ model. When the SE is fixed, the more historical time steps data as input into the $GMAN$ model, the better it can be predicted.

Experimental Results

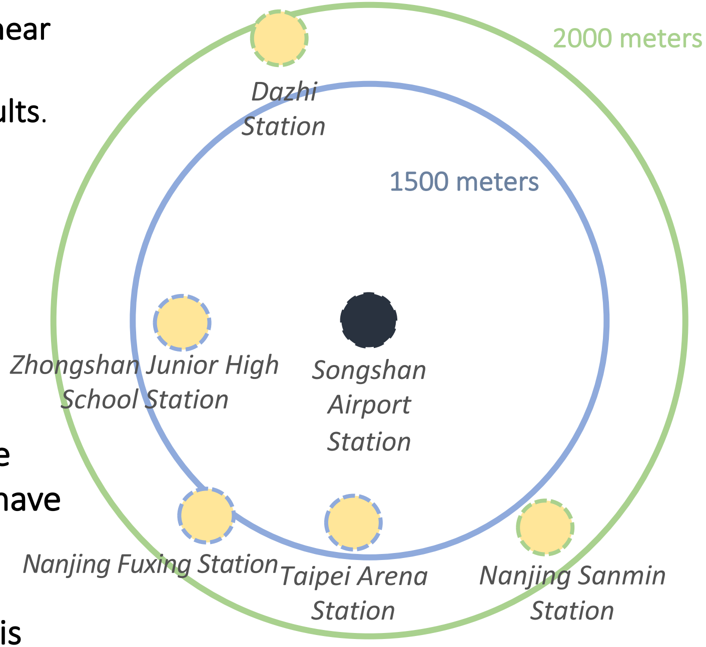
Comparison on Filter Stations by Distance

- This section compares the number of filtered stations by distance measurement. The purpose of filter stations is to only take the information of MRT stations near the target station to predict the passenger flow of that station and use the passenger flow of stations that are closer to each other to achieve better results.

Filter Stations	$BERT^{SE}$		$node2vec^{SE}$	
	MAE	*RMSE*	*MAE*	*RMSE*
None	1.46	2.04	1.21	1.8
1500 m	0.83	1.5	1.18	1.59
2000 m	0.83	1.5	1.19	1.61

- MAE and RMSE in the range of 1500 and 2000 meters after filtered have reached 0.83 and 1.5 in the $BERT^{SE}$ model; 1.18 MAE and 1.59 RMSE have reached 1500 meters in $node2vec^{SE}$.

- The prediction performance using the small number of filtered stations is much better than no filtered station, which also means that the impact of the passenger flow at the MRT station is only in the nearby stations.



Conclusion & Future Works

PART FIVE

Conclusion & Future Works

- Conclusion
  - Proposed a dynamic embedding method for passenger flow estimation.
  - The experimental results shows that dynamic embedding has better performance than static embedding (node2vec) in passenger flow estimation.
  - The experimental results also comparison the filter stations by distance, shows that reducing the number of nearby stations of the target station can effectively reduce the difficulty of model learning, thereby improving the performance of flow estimation.
- Future Works
  - different training target such as the next station prediction task to train the model $BERT^{SE}$ can be added, to learn more perspective of flow data.

Thanks for your listening