DeepSense: A Novel Learning Mechanism for Traffic Prediction with Taxi GPS Traces

Xiaoguang Niu*, Ying Zhu, Xining Zhang*
School of Computer Science, Wuhan University, Wuhan, China
{xgniu, yingzhu91, zhangxn}@whu.edu.cn
*Co-corresponding author

***Abstract*—The urban road traffic flow condition prediction is a fundamental issue in the intelligent transportation management system. While extracting the high-dimensional, non-linear and random features of the transportation network is a challenge, which is very useful to improve the accuracy of traffic prediction. In this paper, we propose DeepSense, a novel deep temporal-spatial traffic flow feature learning mechanism, with large scale Taxi GPS traces for traffic prediction. DeepSense includes two switchable feature learning approaches. DeepSense exploits a temporal-spatial deep learning approach for traffic flow prediction with the sufficient spatial and temporal taxi GPS traces in dynamic pattern. Meanwhile, DeepSense takes advantage of a supplementary temporal sequence segment matching approach with the temporal transformation of traffic flow state for a given road segment when there are not enough traffic traces. Experimental results show that DeepSense can achieve higher prediction accuracy with nearly 5% improvements compared with existing methods.**

Keywords- *traffic flow condition prediction, deep learning, smart city, temporal-spatial, intelligent transportation system*

I. INTRODUCTION

With the development of intelligent transportation technology, the urban road traffic condition prediction facilitates better utilization of available capacity, which can be widely used for finding best routes[1], urban computing [2][3][4], coordinating traffic flow, and balancing traffic control[5]. It is necessary to explore the traffic dynamics and analyze the evolution pattern of traffic flow.

The urban road traffic condition prediction usually utilizes historical and real-time traffic information effectively for predicting city road conditions in the future time. Most of the existing methods present prediction trend either by using probability and statistics of the time-dependent evolution of current road, or only using spatial relationships among various road segments. Moreover, although available spatial and temporal information are combined to model the traffic network pattern in some previous works, the information does not play out its full potential.

Traffic network possesses complicated relations of time and space. The traffic flow is a high dimensional, non-linear and non-stationary random process. Deep learning that can model high-level abstractions by using architectures composed of multiple non-linear transformations has been used [6]. However, what it uses is still a static network which has not taken the essential temporal information and overall space-time dynamic pattern into consideration. The condition of a road segment is inevitably influenced by the temporal-spatial information in the traffic network. However, the prediction works perform not so well at certain times, which occur especially when there are insufficient taxi GPS traces. Based on this finding, it's proper to consider more supplementary aspects such as traffic flow state transformation rule contained in temporal changes.

In this paper, we propose a temporal-spatial traffic flow feature learning mechanism, namely DeepSense, containing two switchable approaches. It comprehensively utilizes both temporal-spatial dynamic pattern and temporal supplementary information in transportation network. Firstly, in order to analyze sufficient temporal and spatial information of taxi GPS traces, a deep learning approach Restricted Boltzmann Machine (RBM) is trained, which can greatly reduce the dimension of data and fit the non-linearity of data distribution by minimizing the energy function. Then the extracted low-dimension data is put into a support vector machine (SVM) to gain good prediction results. Secondly, when there are not enough traffic information, a temporal sequence segment matching approach, which utilizes the match on historical dimension to predict the future state, is trained to explore the regularity of traffic flow temporal transformation for the forecasted road segment.

The major contributions of this paper consist of the following aspects:

1. Propose DeepSense mechanism, and especially a trained temporal-spatial deep learning approach that can extract features from the high-dimensional, non-linear and random traffic flow to obtain good prediction results.
2. Experiments show that DeepSense mechanism can achieve great performance in prediction accuracy, which proves that deep learning approach is an effective method to extract features for traffic prediction in intelligent transportation system.

This paper is organized as follows: Section II summarizes the related works. Section III presents an overview of our approach. Section IV gives details to concrete processes of the traffic flow condition feature learning. Section V implements experiments and empirical results are given with analysis. Finally, we draw conclusions of our work.

II. RELATED WORKS

A. Time-Series-Based Approach

The majority of existed approaches merely utilize time-dependent traffic flow evolution rule for prediction [10][11].

Auto regressive integrated moving average (ARIMA) model [7][8] simply relying on historical traffic flow data of forecasted point takes temporal variation into consideration for prediction. The traffic flow prediction model based on m_{th}-order Markov chain [9] counts the transition probability of traffic flow state with historical data of forecasted road. However, there is just a statistical method to establish model in time sequence and forecast to some extent.

B. Spatial-relation-Based Approach

There are many methods use spatial information in transportation networks to analyze the trend of traffic flow. Castro et al. [13] employs adjacent road to forecast traffic conditions, which does not consider the influence of distant segments and temporal-spatial correlation of predicted road. Markov logic networks [12] can predict the traffic conditions at simultaneous locations in the different future time.

C. Temporal-Spatial-Correlation Based Approach

The methods mentioned above model traffic flow trends merely from one aspect. A Bayesian network approach [14] for traffic flow forecasting proposed by Sun *et al.* applies information of adjacent links and its spatial-temporal information in a transportation network to construct a Bayesian network. Sun *et al*. also develops a selective random subspace predictor (SRSP) model [15] utilizing traffic flows of some most closely correlated links ranked by the measurement of Pearson correlation coefficient in the subspace to forecast the given road link. Although considering spatial-temporal information, these approaches did not extract the characteristics of high dimension, non-linearity and randomness from a viewpoint of the whole network.

III. OVERVIEW

A. Preliminary

Definition 1*: Taxi trajectory*. A trajectory is a time series of GPS points for a trip, where there is a timestamp Δt between consecutive points.

Definition 2*: Road segment/link*. A road link or segment represents a directed edge between crossroads consisting of a direction signal, two terminal points and a length.

Definition 3*: Traffic flow condition.* We select speed of taxis to denote traffic state of road segments. Based on the range of speed, traffic states are classified into five categories: congestion, slow, normal, moderate and unimpeded. Given different speed situations and speed limits on distinct road segments, classified traffic state based on absolute speed is obviously inaccurate. Hence, in our paper, states of each road segment are defined based on K-means algorithm which can cluster speeds in the different time period each day. According to the degree of traffic congestion, we can categorize the traffic conditions with five labels, namely congesting, slow, normal, moderate and unimpeded, represented by $\boldsymbol{S}$={1,2,3,4,5}.

In detail, for the road segment $\boldsymbol{R}$ from 0 o'clock to 24 o'clock, a speed average x_i is collected every 15 minutes, getting the set of speed observations $\boldsymbol{V}=\{v_1,v_2,\ldots,v_{96}\}$ during the period. Setting clustering center as 5 in K-means, $\boldsymbol{V}$ can be divided into 5 subsets and obtain the center of each subsets. According to average values of five subsets, we can acquire different traffic states.

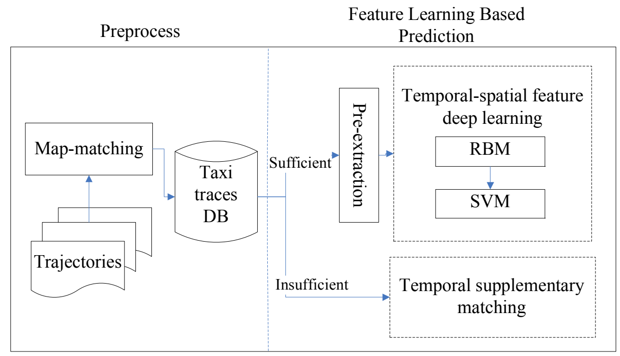


Fig. 1 Framework of DeepSense

B. Framework

In order to comprehensively learn the features of high-dimension, non-linearity and randomness of the road network, we propose a temporal-spatial traffic flow feature learning mechanism, namely DeepSense. As shown in Fig. 1, applying our mechanism for traffic flow condition forecast, there are two parts, including preprocessing and feature learning based prediction.

Preprocessing: Spatial trajectories collected by GPS-equipped vehicles are mapped onto a road network using a map-matching algorithm [16] and then stored into a taxi traces database.

Feature learning based prediction: Feature learning based prediction includes two switchable approaches. When there are enough taxi GPS traces, i.e. more than a constant traffic flow value Φ, we apply deep learning approach to learn temporal-spatial features with pre-extracted temporal-spatial traffic flow information from taxi traces database. Considering the traffic flow of adjacent road segments converge on the current road segment, the traffic flow of most correlated road segments are utilized to forecast the traffic condition of current road segment. Moreover, in order to overcome the data sparsity with insufficient traffic information, a temporal sequence segment matching approach on historical dimension is used to estimate the state of traffic flow and explore the temporal sequence transformation rule. Details are presented in Section IV.

IV. FEATURE LEARNING BASED PREDICTION

This section details the approaches applied by the temporal-spatial traffic flow feature learning mechanism to predict traffic flow condition of current road segment, which includes temporal-spatial feature deep learning and temporal feature supplementary learning.

A. Temporal-spatial feature deep learning

By employing temporal-spatial feature deep learning approach, we can reduce the dimension of the input variables

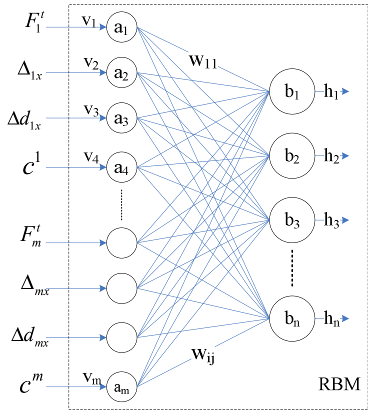


Fig. 2 Temporal-spatial feature deep learning model

and get rid of the redundancy. In addition, as the input data is too complex to be described by manually selected models, deep network is used to automatically fit the distribution of data and learn good features. This approach includes three main steps, feature pre-extraction, feature learning with RBM and training a SVM model.

Feature pre-extraction: We adopt correlation coefficient, the index of measuring the correlation between other roads and the predicted point to select the most correlated roads as input variables.

Considering m samples $\{x_i(k), y_j(t_{ct}+t_l)\}(i,j=1,2,\cdots,m)$ in a spatial accessible space, $x_i(k)(k=1,2,\cdots,n)$ represents the traffic flow of road segment i at time stamp k. And $y_j(t_{ct}+t_l)$ stands for the traffic flow of road segment j at predicted time $(t_{ct}+t_l)$, where t_{ct} is the current time and t_l means the duration of the prediction. The Pearson correlation coefficient between $x_i(k)$ and $y_j(t_{ct}+t_l)$ is defined as follows:

$$r(k)=\frac{\sum_{i=1}^{m}(x_i(k)-\bar{x})(y_j(t_{ct}+t_l)-\bar{y})}{\sqrt{\sum_{i=1}^{m}(x_i(k)-\bar{x})^2\sum_{i=1}^{m}(y_j(t_{ct}+t_l)-\bar{y})^2}}$$
$$=\frac{m\sum_{i=1}^{m}x_i(k)y_j(t_{ct}+t_l)-\sum_{i=1}^{m}x_i(k)\sum_{i=1}^{m}y_j(t_{ct}+t_l)}{\sqrt{m\sum_{i=1}^{m}x_i(k)^2-(\sum_{i=1}^{m}x_i(k))^2}\sqrt{m\sum_{i=1}^{m}y_j^2(t_{ct}+t_l)-(\sum_{i=1}^{m}y_j(t_{ct}+t_l))^2}} \quad (1)$$

It means higher correlation if the absolute value of spatial and temporal correlation coefficient is closer to 1. Then the information of the most correlated m links will be chosen as extracted features.

We design four parameters to describe the above extracted features, where F_m, c_m denotes the average speed and the certain state of the m_{th} road segment at the correlated time stamp. Δ_{mx} is defined as the interval of the time stamp between predicted road segment and its m_{th} most correlated road segment. For example, the Δ_{mx} between $x_m(t\text{-}t_0)$ and $y_x(t_{ct}+t_l)$ is t_l+t_0. d_m, another metric of the features, is measured as the geo-distance between predicted road segment and its m_{th} most correlated road segment.

Feature learning with RBM: The main idea of deep learning algorithm RBM is that a plenty of unlabeled data are used to train an unsupervised neural network, and then the learned features can be used to train an efficient classifier such as SVM.

RBM is a generative stochastic neural network that can learn a probability distribution (e.g. $p(v,h)$) over the input set, which is good at reasoning about and predicting irregular and stochastic behavior in the traffic flow. As shown in Fig.2, RBM is in the shape of a bipartite graph with no intra-layer connections. The hidden unit activations h (low dimensional data) are mutually independent given the visible unit activations v (our four parameters, namely F_m , c_m , Δ_{mx} , d_m) and conversely, the visible unit activations are mutually independent given the hidden unit activations. When constrains are given on v, all hidden units are conditionally independent, i.e. $p(h|v) = p(h_1|v)\cdots p(h_n|v)$, and vice versa. Given visible layer v, hidden layer h can be obtained through $p(h|v)$, meanwhile we can also gain the value of units by getting the hidden units $p(v|h)$. By adjusting the parameters, the visible layer v_l obtained from hidden layer can approximately equal the original input layer v. From this perspective, the outputs of hidden units are another representation of the visible units, namely the original high-dimensional temporal-related information have been transformed into low-dimensional meaningful data.

The standard type of RBM has binary-valued hidden and visible units, and consists of a matrix of weights in which w_{ij} denotes the connection weight between hidden unit h_j and visible unit v_i, a_i denotes the bias of v_i and b_j denotes the bias of h_j. Given parameters $\theta = (w_{ij},a_i,b_j)$, the energy of a configuration (v,h) is defined as:

$$E(v,h;\theta)=-\sum_{i=1}^{m}a_iv_i-\sum_{j=1}^{n}b_jh_j-\sum_{i=1}^{m}\sum_{j=1}^{n}h_jw_{ij}v_i \quad (2)$$

This energy function is analogous to that of a Hopfield network. As in general Boltzmann machines, probability distributions over hidden and visible vectors are defined in terms of the energy function:

$$P(v,h)=\frac{1}{Z}e^{-E(v,h)} \quad (3)$$

where Z is a partition function defined as the sum of $e^{-E(v,h)}$ over all possible configurations. Similarly, the marginal probability of a visible vector of Booleans is the sum over all possible hidden layer configurations:

$$P(v)=\frac{1}{Z}\sum_{h}e^{-E(v,h)} \quad (4)$$

Then the probability under the condition $(h_j=1)$ and the probability under the condition $(v_i=1)$ are given like this:

$$P(h_j=1\,|\,v;\theta)=\sigma(b_j+\sum_{i=1}^{m}w_{ij}v_i) \quad (5)$$

$$P(v_i=1\,|\,h;\theta)=\sigma(a_i+\sum_{j=1}^{n}w_{ij}h_j) \quad (6)$$

where $\sigma(x)=\frac{1}{1+e^{-x}}$ denotes the logistic sigmoid function, the parameters $\theta = (w_{ij},a_i,b_j)$ can be learned by gradient-based contrastive divergence algorithm.

As we discussed so far, RBM can be used to process temporal-spatial traffic features into more informative low dimensional features.

Training a SVM model: After input data being compressed by RBM, more representative features are extracted as the inputs of SVM, and better classification performance will be gained.

Through training a SVM classifier, we can get 5 classifications of traffic flow state, and their probabilities construct a vector:

$$P(t) = [p(state{=}1), p(state{=}2), p(state{=}3), p(state{=}4), p(state{=}5)] \quad (7)$$

B. Temporal feature supplementary learning

We evaluate the similarity between temporal sequence segments in two days based on the speed distribution pattern. Specially, if the speed distribution follow a similar pattern of change in the same period of two days, the similarity value between the two will be high. Hence, the speed distribution we want to obtain at the predicted time can be estimated according to these referable days.

Given a predicted time t ($t{=}t_{ct}{+}t_l$), we first calculate the similarities between the period of T time slots (per time unit is Δt) in current day d and the corresponding period in H historical days (let $h \in H$ denotes a day in the historical day set H), and then choose referable days to produce a prediction of the speed distribution by taking a weighted combination. More specially, according to the above classification standard of traffic flow condition, there are *5* speed interval and $P_d(t_{ct}) = [p_{d,t_{ct},1}, p_{d,t_{ct},2}, \cdots, p_{d,t_{ct},5}]$, where $p_{d,t_{ct}j}$ $(1 \le j \le 5)$ denotes the proportion belonging to j_{th} speed classification at current time t_{ct} in current day d. For example, assuming the range of speed is 0~60 km/s and per 10 km/s is the average value of five speed subsets, hence $p_{d,t_{ct}j,1}$, $p_{d,t_{ct}j,2}$,…, $p_{d,t_{ct}j,5}$ respectively represents the proportion of vehicle speed 0~15km/s, 15~25km/s, 25~35km/s, 35~45km/s, 45~60km/s. Similarly, $P_h(t_{ct}) = [p_{h,t_{ct},1}, p_{h,t_{ct},2}, \cdots, p_{h,t_{ct},5}]$ denotes the speed distribution vector at time t_{ct} of the day h in history.

Therefore, we utilize the cosine similarity measure to calculate the similarity of two speed distribution vectors at the same time t_{ct} between the day d and h as follows:

$$w_{d,h}(t_{ct}) = \frac{\sum_{j=1}^{5} p_{d,t_{ct},j} \cdot p_{h,t_{ct},j}}{\sqrt{\sum_{j=1}^{5} p_{d,t_{ct},j}^2}\sqrt{\sum_{j=1}^{5} p_{h,t_{ct},j}^2}} \quad (8)$$

Then the total speed distribution similarity in T time slots between the day d and h can be calculated:

$$w_{d,h} = \sum_{i=0}^{T} w_{d,h}(t_{ct} - i \cdot \Delta t) \quad (9)$$

A constant threshold ε is defined to filter adequate traffic flow speed distribution sequence segments in history according to the similarity which are more than the chosen threshold, i.e. $w_{d,h} \ge \varepsilon$. Assuming there are n days in the historical sets matching the requirement, we calculate the speed pattern distribution at predicted time utilizing the total speed pattern similarity between the day d and matched historical day as its corresponding weight. Finally, we can get the weighted vector at the predicted time t ($t{=}t_{ct}{+}t_l$) as the final state result:

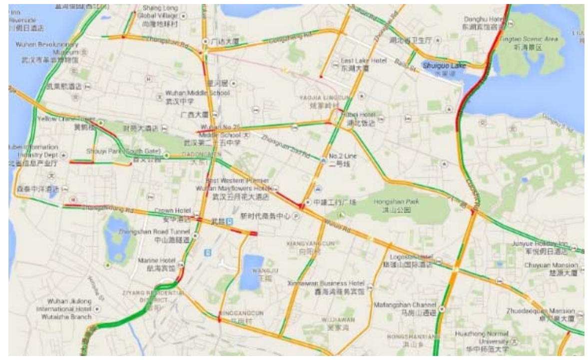


Fig. 4 The road network of Wuchang district in Wuhan

$$P_d(t) = \sum_{h=1}^{n} \frac{w_{d,h}}{\sum_{h=1}^{n} w_{d,h}} \cdot P_h(t) \quad (10)$$

V. Evaluations

A. Datasets

Datasets of Wuhan are used for our traffic flow prediction. We select a representative region to verify the validity of our mechanism. The following three available datasets are used:

1) Taxi Trajectories: Our evaluations are performed based on the trajectory data sets generated by 30,000 taxis over a period of three months from January to March in 2013.Because of the different traffic patterns on weekdays and weekends, we extract 13 weeks working data from three months data for prediction. Total distance of the data set is about 450 million kilometers, and whole number of GPS points reaches 890 million. The interval sampling each time is 4.5 minutes with the distance of 500 meters between two consecutive points.
2) Road Network: The road network of Wuhan is adopted to perform the experiments. In Fig. 4, a snapshot of the Wuchang district road network in Wuhan is displayed in the rush hour (5pm).

B. Performance

Evaluation of temporal-spatial deep learning approach. Several parameters need to be determined for utilizing a temporal-spatial deep learning approach. Since different network structures influence our forecast result obviously, effect of parameters need to be tested. One of the most important problems to design a neural network is to determine the size of network. 128 nodes are chosen to be the nodes number of input layer. As the change of input layer node number, different weighted mean accuracy (WMA) can be obtained. As shown in Fig. 5, the best choice at the coordinate point of 64 nodes based on our experiments. If there are more nodes, it increases the redundancy and computational expense for training model. However, few nodes

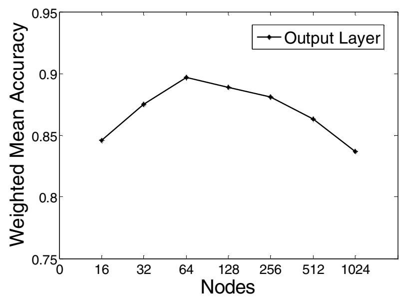


Fig. 5 Effect of nodes in output layer

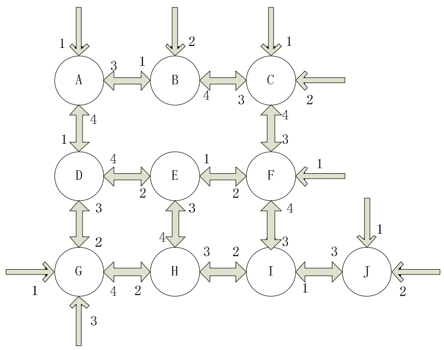


Fig. 6 Illustration of road network segment

cannot make full use of neural network to depict the characteristics of high-dimension, non-linearity and randomness in the transportation network.

As shown in Fig. 6, road segment E_1 chosen as the forecasted road segment denotes the traffic flow from upstream link F to downstream road E. In order to predict the object traffic flow of $E_1(t+t_l)$ at time $t+t_l$, M (M = 32 can be calculated through number of input nodes in RBM) most closely correlated traffic flow with different time indexes are selected from all the following available traffic flows: $\{A_1(t), A_1(t-1\cdot\Delta t), \ldots, A_1(t-l\cdot\Delta t)\}$, $\{A_3(t), A_3(t-1\cdot\Delta t), \ldots, A_3(t-l\cdot\Delta t)\}$, … , $\{E_1(t), E_1(t-1\cdot\Delta t), \ldots, E_1(t-l\cdot\Delta t)\}$, … , $\{J_3(t), J_3(t-1\cdot\Delta t), \ldots, J_3(t-l\cdot\Delta t)\}$, etc. In this paper, we choose 15 min as the time unit Δt and l = 20.

A road segment state is valued as the state number n^r_t of the road segment r at the time stamp t. When a road segment traffic flow state is predicted as the state n^t_r with the probability p generated by two different approaches, the predicted state value of the road segment is defined as n^t_r-1+p, e.g., when a predicted state is 2 and the corresponding probability is 0.9 generated by deep learning approach, then the deep learning-based predicted state value is 1.9.

Fig. 7 shows the traffic state prediction of temporal-spatial deep learning approach with the prediction time φ=15min. There are two time slots (about 8am and 6pm) where the prediction values are more close to the actual values. It can be inferred that in the rush hours, we have enough taxi trajectories which can provide us with sufficient data to learn the states of traffic flow and produce a more accurate prediction. Temporal-spatial deep learning approach can effectively extract the temporal-spatial features

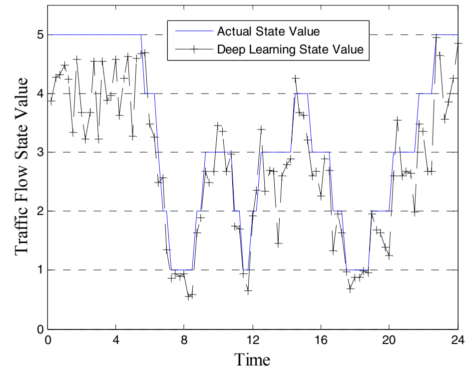


Fig. 7 Traffic state prediction of temporal-spatial deep learning approach

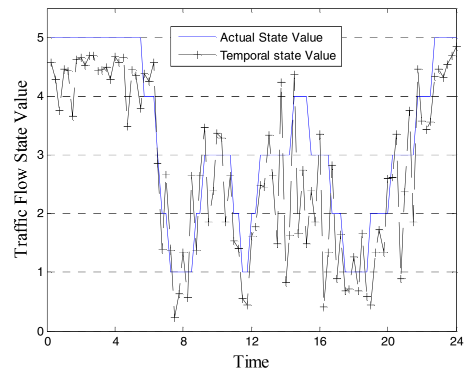
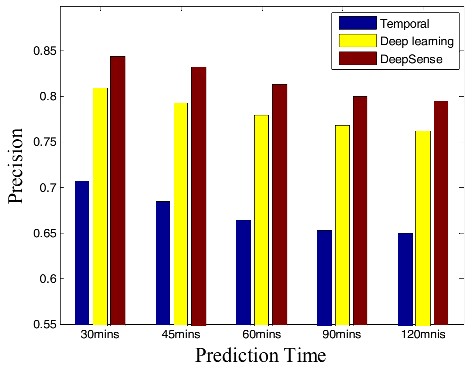


Fig. 8 Traffic state prediction of temporal supplementary approach

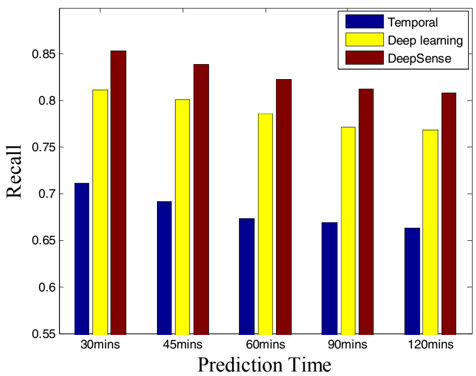
of high-dimensional, non-linear and random traffic flow. However, due to the lack of sufficient Taxi GPS traces, the traffic prediction accuracy in the early morning hours (from 1:00am to 5:00am) is significantly lower than that in the day time. With the increase of the vehicles at 6am, the prediction accuracy improves gradually. It can be seen that deep learning approach cannot train the effective network to extract features because of the sparsity problem of data.

Evaluation of temporal supplementary approach. Fig. 8 demonstrates the prediction result of temporal supplementary approach with the prediction time φ=15min. According to the calculation, the prediction precision of temporal supplementary approach is nearly 0.70. Comparing the results of two approaches to predict separately in Fig. 7 and Fig. 8, it can be inferred that the effect of deep learning approach is much better than supplementary approach in most cases. However, temporal supplementary approach have better prediction effect on making up for the problem of sparse in deep learning approach because of taking full advantage of the supplementary temporal change rule of traffic flow, which can provides effective information to improve the prediction accuracy when in the midnight time.

Evaluation of DeepSense mechanism. DeepSense comprehensively uses temporal-spatial information and temporal supplementary rule in the traffic network. Fig. 9 shows the precision and the recall of three approaches: under the same prediction time interval, DeepSense can obtain 5% higher precision and recall than other approaches. DeepSense converges after about 120mins. The effect of the



A) Precision of three different approaches



B) Recall of three different approaches

Fig. 9 Study on three different approaches

deep learning approach is close to DeepSense. For combining temporal supplementary matching approach to overcome the problem of sparse in deep learning, the prediction result of DeepSense can be further improved. When less than Φ (200 vehicles per hour), deep learning approach is utilized from 6am to 11pm, with the rest of time for another approach. It can be apparently inferred that in the early morning, especially between 1 to 5 a.m., due to the sparse data in the traffic network, deep learning approach may have a low degree of recognition. However, supplementary matching approach can obtain higher prediction precision during that period of time by exploring temporal change rule of traffic flow. Hence, a more accurate result can be gained by DeepSense which utilizes two approaches.

VI. Conclusions

This paper proposes a novel deep temporal-spatial traffic feature learning mechanism, namely DeepSense, for traffic flow condition prediction with large scale taxi GPS traces. Specifically, DeepSense makes full use of the traffic information. In DeepSense, a temporal-spatial deep learning approach with preprocessed traffic-related features can effectively extract characteristics of non-linearity, randomness and high-dimension from the traffic flow change. As a supplementary, by utilizing environment features, CRF classifier can further reflects the dynamic pattern in the transportation network. Finally, a weighted classifier fusion approach is used to obtain a better prediction result. We adopted trajectory data sets generated by 30,000 taxis over a period of three months in Wuhan to evaluate our approach. The experiment results showed that our approach can contribute to improve the prediction accuracy effectively.

Acknowledgements

This work was partially supported by National Key Basic Research Program of China "973 Project" (Grant No. 2011CB707106), Development Program of China "863 Project" (Grant No. 2013AA122301), National Natural Science Foundation of China "NSFC" (Grant No. 61303212) and the Program for Changjiang Scholars and Innovative Research Team in University (Grant No. IRT1278).

References

[1] Yuan N J, Zheng Y, Zhang L, et al, "T-finder: A recommender system for finding passengers and vacant taxis", *IEEE Trans. On Knowledge and Data Engineering*, 25(10), pp. 2390-2403, 2013.

[2] Zheng Y, Liu Y, Yuan J, et al, "Urban computing with taxicabs," *Proc. of ACM Ubicomp*, pp. 89-98, 2011.

[3] Liu S, Liu Y, Ni L M, et al, "Towards mobility-based clustering," *Proc. of ACM SIGKDD*, pp.919-928, 2010.

[4] Liu W, Zheng Y, Chawla S, et al, "Discovering spatio-temporal causal interactions in traffic data streams," *Proc. of ACM SIGKDD*, pp.1010-1018, 2011.

[5] Zhang F, Wilkie D, et al, "Sensing the pulse of urban refueling behavior," *Proc. of ACM UbiComp*, pp. 13-22, 2013.

[6] Huang W, Hong H, Li M, et al, "Deep Architecture for Traffic Flow Prediction," *Advanced Data Mining and Applications*. Springer Berlin Heidelberg, pp. 165-176, 2013.

[7] Van Der Voort M, Dougherty M, Watson S, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transportation Research Part C: Emerging Technologies*, 4(5), pp. 307-318, 1996.

[8] Moorthy CK , Ratcliffe BG, "Short term traffic forecasting using time series methods," *Transportation planning and technology*, 12(1), pp. 45-56.1988.

[9] Yuan J, Zheng Y, et al, "Driving with knowledge from the physical world," *Proc. of ACM SIGKDD*, pp. 316-324, 2011.

[10] Yuan J, Zheng Y, Zhang C, et al, "T-drive: driving directions based on taxi trajectories," *Proc. of ACM SIGSPATIAL*, pp. 99-108, 2010.

[11] Yuan J, Zheng Y, Xie X, et al, "T-Drive Enhancing Driving Directions with Taxi Drivers' Intelligence," *IEEE Trans. on Knowledge and Data Engineering*, 25(1), pp. 220-232, 2013.

[12] Lippi M, Bertini M, Frasconi P. "Collective traffic forecasting," *Machine Learning and Knowledge Discovery in Databases*. Springer Berlin Heidelberg, pp. 259-273, 2010.

[13] Castro P S, Zhang D, Li S, "Urban traffic modelling and prediction using large scale taxi GPS traces," *Pervasive Computing*. Springer Berlin Heidelberg, pp. 57-72, 2012.

[14] Sun S, Zhang C, Yu G, "A Bayesian network approach to traffic flow forecasting," *IEEE Trans. on Intelligent Transportation Systems*, 7(1), pp. 124-132, 2007.

[15] Sun S, Zhang C, "The selective random subspace predictor for traffic flow forecasting," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8(2), pp. 367-373, 2007.

[16] Yuan J, Zheng Y, Zhang C, et al, "An interactive-voting based map matching algorithm," *Proc. of IEEE MDM*, pp. 43-52, 2010.