International Journal of Research
Available at https://pen2print.org/index.php/ijr
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 06 Issue 01
January 2019

Role of Data Fusion Prediction of Road Traffic Speed

V.Prasanthi & T.KishoreBabu
1PG Scholar, Dept of CSE, Malineni Lakshmaiah Engineering College, Singarayakonda, Prakasam (Dt), AP, India.
2Assistant Professor, Dept of CSE, Malineni Lakshmaiah Engineering College, Singarayakonda, Prakasam (Dt), AP, India

Abstract - Road traffic speed prediction is a challenging problem in intelligent transportation system (ITS) and has gained increasing attentions. Existing works are mainly based on raw speed sensing data obtained from infrastructure sensors or probe vehicles, which, however, are limited by expensive cost of sensor deployment and maintenance. With sparse speed observations, traditional methods based only on speed sensing data are insufficient, especially when emergencies like traffic accidents occur. To address the issue, this paper aims to improve the road traffic speed prediction by fusing traditional speed sensing data with new-type "sensing" data from cross domain sources, such as tweet sensors from social media and trajectory sensors from map and traffic service platforms. Jointly modeling information from different datasets brings many challenges, including location uncertainty of low-resolution data, language ambiguity of traffic description in texts and heterogeneity of cross-domain data. In response to these challenges, we present a unified probabilistic framework, called Topic-Enhanced Gaussian Process Aggregation Model (TEGPAM), consisting of three components, i.e. location disaggregation model, traffic topic model and traffic speed Gaussian Process model, which integrate new-type data with traditional data. Experiments on real world data from two large cities in America validate the effectiveness and efficiency of our model

1. Introduction

Traffic flow prediction is a key part and core content of intelligent transportation system as well as the important basis for transportation information service, traffic control and guidance [1-2]. Forecasting timely and accurately is premise of the intelligent transportation system realizing dynamic traffic management. Crossroads are the key component of transportation network. The size of traffic volume in intersections decides directly the passage capacity of road network, which becomes the bottleneck of road transportation network and plays a significant role in the entire road transportation network [3-4]. To solve the problem of predicting the short-time traffic flow in crossings, it proposes a mining algorithm, which experimentally shows good performance in the real transportation data set [5-6].

According to the time duration of prediction, traffic volume predication can be long-time and short-time predication; as far as the predicted object is concerned, there is crossroad traffic prediction and high-way traffic prediction. Road transportation system is a huge and complicated nonlinear system which has human be involved and is time-varying. The system has higher uncertainties, which may derive from environmental factors like road condition, climate changes *etc.* or emergency situations like traffic accidents, mass gathering *etc.* Those factors bring about certain difficulties to the anticipation of road traffic flow, especially for the short-time prediction. For that, researchers have presented lots of models such as ARIMA model and nonparametric regressive model, which are specifically designed to predict highway and road segment traffic volumes [7-8]. Crossroads are important to the road transportation network. Its transportation is very complex. The traffic flow in each direction in the crossing roads is relevant to not only its own traffic flow and also flow in other direction and the timing plan of traffic signal lights. The traffic flow in crossroads is more volatile than flow series in road sections, particularly the short-time flow series [9]. The volatility occurs not merely because of more accidents taking place in the intersections but also affected by traffic signal lights, specifically variable timing scheme of the signal lights. At present, the acquisition coils for the traffic characteristics of urban intelligent transportation

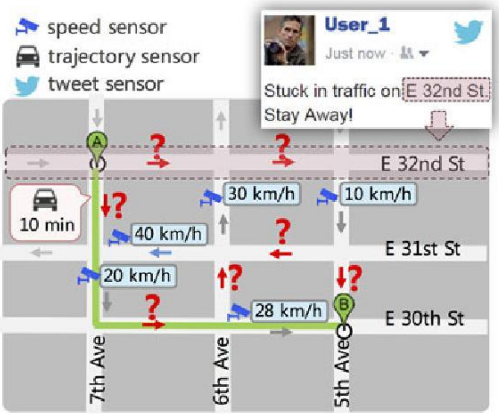

Fig. 1: Problem setting. Our goal is to predict the traffic speed of specific road links, as shown with the red question marks, given: 1) some speed observations collected by speed sensors, as shown in blue; 2) trajectory and travel time of OD pairs. Note that speeds of passed road links are either observed or to be predicted; 3) tweets describing traffic conditions. Note that the location mentioned by a tweet may be a street covering multiple road links. as “Slow traffic on I-95 SB from Girard Ave to Vine St.” posted by local transportation bureau account. Such text messages describing traffic conditions and some of them tagged with location information are accessible by public and could be a complementary information source of raw speed sensing data.

(OD) pair on a map, such services can recommend optimal route from the origin to the destination with least time, and trajectories can be collected once drivers use the service to navigate. Here a trajectory is a sequence of links for a given OD pair, and a link is a road segment between neighboring intersections. Correspondently, a trajectory travel time is an integration of link travel times, which are related to the real-time road traffic speeds. Longer trajectory travel time indicates that some involving road links may be congested with lower traffic speed. Trajectory data is useful for a wide range of transportation analyses and applications [49] [9].

Based on the above observations, where traditional traffic sensing data are limited while new-type data from social media and map service begin to spring up, our goal is to predict the road-level traffic speed by incorporating new-type data with traditional speed sensing data. To motivate this scenario, consider a road traffic prediction example depicted in Fig.1. Those links in red question marks are not covered by traditional speed sensors, but may be passed by trajectories attached with travel time information, or mentioned in tweets describing traffic conditions, so their speeds can be inferred fusing multiple cross-domain data.

2. RELATED WORKS

Traffic prediction problem can be broadly classified into short-term and long-term prediction [1], considering three main basic traffic measurements: traffic flow, an equivalent flow rate in vehicles; speed, mean of the observed vehicle speeds; lane occupancy, the percentage of time that the sensor is detecting vehicle presence. This paper focuses on the short-term traffic speed prediction combining multi-source heterogeneous data, which, as far as we know, has not been well explored before. This part gives a summary on short-term traffic speed prediction and the exploration on fusing multiple information sources.

Short-term Traffic Speed Prediction: The presented methods can be classified into two categories:

1) **parametric methods**, assume that traffic speed follows a probability distribution based on a fixed set of param-eters. Time series analysis technique is applied in traffic speed prediction based on the periodicity of traffic speed during a day or a week. Auto-Regressive Moving Average (ARMA) models are adopted in [46] and [38], where Mul-tivariate Spatial-Temporal Auto-Regressive (MSTAR) model is adopted to include dependency among observations from neighboring locations. A review about Auto-Regressive In-tegrated Moving Average (ARIMA) time series methods can be found in [55]. ARIMA and Winters exponential smooth-ing techniques are used to forecast urban

freeway flow in [54]. [53] separate ARIMA models for a set of loop detectors that incorporate information from upstream measurement locations. A single Space-Time Auto-Regressive Integrated Moving Average (STARIMA) model is proposed to describe the spatiotemporal evolution of traffic flow in an urban network in [26], which is essentially a constrained Vector Autoregressive Moving Average (VARIMA) model [13] with constraints that reflect the topology of a spatial network and result in a drastic reduction in the number of parameters. A Generalized Space-Time ARIMA (GSTARIMA) method is proposed in [57], which extends ARIMA in spatial and tem-poral dimension and is more flexible because parameters are designed to vary per spacial location. Kalman filter-based approaches are used in [11] and [14], and show advantages for on-line estimation of traffic flows. Markov logic network is used to simultaneously predict the congestion state in [30]. A structured time series model is proposed in multi-variate form for short-term traffic prediction in [12].

2) **non-parametric methods**, make no distribution as-sumptions and the number of parameters scales with the number of training data. K-nearest neighbor (KNN) non-parametric regression methods, e.g. [9], [21], [58], find the k-nearest neighbors using Euclidean distance and calculate the weight. Neutral Networks (NNs), e.g. [50], [27], are biologically-inspired systems and can be trained to ap-proximate virtually any nonlinear function given adequate data and a proper network architecture. NNs have many derivatives for short-term prediction, such as back prop-agation neutral network with genetic algorithms [1] and wavelet networks [22]. Travel speed of each road segment is computed using the GPS trajectories by a context-aware matrix factorization approach in [45]. To adaptively route a fleet of cooperative vehicles under the uncertain and dynamic road congestion conditions in [33] and [34], a GP probabilistic model is proposed to capture the spatial and temporal relationships of travel speeds over road segments and temporal contexts, especially with estimating the mean and covariance of the GP prior from the historical data. Geostatistical interpolation techniques named Kriging are proposed to capture spatial and temporal evolutions of traffic flows in [48].

Traffic Modeling with Multi-Source Heterogeneous Data:

Some researchers attempted to combine traffic sensing data with other data sources, to handle external factors such as traffic accidents (e.g. [36], [42]), mobile sensors (e.g. [39], [40]) and weather (e.g. [37], [2]). [37] reviews the literature on the impact of weather on traffic demand, traffic

International Journal of Research
Available at https://pen2print.org/index.php/ijr
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 06 Issue 01
January 2019

safety, and traffic flow relationships. A trajectory-based community discovery method is proposed in [32], where the trajectory similarity is modeled by several types of kernels for different information markers (e.g. semantic properties of the locations and the movement velocity). [29] tackles the rents/returns bike number prediction prob-lem using multiple features, e.g. time and meteorology, as measures of similarity functions in multi-similarity-based inference model. While [32] and [29] introduce different information sources as new features for computing the similarity, our work assumes the latent relations between these informations, and constructs a Bayesian generative process. As crowdsourcing data from a crowd of online social platform become more available, researchers begin utilizing social content to estimate traffic conditions. Twitter data are matched to detect traffic incidents in [36]. In [39], traffic anomaly detection uses crowd sensing with two forms of data, human mobility and social media, and the detected anomalies are described by mining representative terms from the social media that people posted when the anomaly happened. Few methods incorporate social me-dia text data (e.g. Twitter data) to improve traffic speed prediction. [31] extends spatiotemporal GP in [34] to three dimensional topic-aware GP, where topics on road links are probabilistic modeled based on the user, space and time of tweets. [15] do not tackle the location uncertainty problem of tweets, because the inference of traffic status based on words of tweets only focuses on the average regional traffic flow, which is insufficient for predicting road speed.

3.GAUSSIAN PROCESS PRELIMINARIES

Gaussian Processes (GPs) have been widely studied in many fields, such as spatio-temporal modeling [24], [38]. Given a set of road segments S under a specified time stamp, we spatially model the traffic speed of road segments via a function f : S ! R$_{þ}$, which outputs the traffic speed for a given road link s. Assume that f is sampled from a Gaussian process prior: fðsÞ GPðmðsÞ; kðs; s^{0}ÞÞ, which is fully specified by the mean function and the covariance, or kernel, function An important property of GP is that if two sets of varia-bles are jointly Gaussian, the conditional distribution of one set conditioned on the other is Gaussian, that is the basis to compute the posterior analytically [39].

Suppose that there are currently observed links **S** S with speed observations **V** ¼ fv$_{s}$; s 2 **S**g, where the traffic speed v$_{s}$ for each link s 2 S follows v$_{s}$ N ðfðsÞ; s^{2}Þ, where s^{2} is i.i.d. Gaussian noise. Then we can calculate the posterior distribution given the prior distribution with mean and kernel function, and the current observations **V** , which is still a GP distribution.

4.MODEL DESIGN

This section begins by formalizing the speed prediction problem in Section 4.1. Then we introduce three models from Sections 4.2, 4.3, and 4.4 to tackle the challenges afore-mentioned in the introduction, i.e., a disaggregation model for location uncertainty in tweet and trajectory data, a traffic topic model for tweet language ambiguity

International Journal of Research
Available at https://pen2print.org/index.php/ijr

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 06 Issue 01
January 2019

and a GP model for capturing the spatial correlation of speed sensing data. Section 4.5 integrates three models dealing with different information source into a novel probabilistic model, named TEGPAM, under the Bayesian framework.

4.1 Problem Formulation

Consider a road network denoted by S ¼ f1; . . . ; Sg contain-ing S road links, and a time duration denoted by T ¼ f1; . . . ; T g containing T time stamps. Our goal is to predict traffic speed of some links at a certain time stamp using the past and current observations from multiple data sources, including traffic sensing data, Tweets and trajectories. The terms and formal definitions used throughout this paper are listed as follows.

Traffic Condition. Road traffic condition is described by two variables: continuous traffic speed and binary traffic status. The speed at time t 2 T and link s 2 S is denoted by $v_{t;s}$ 2 R, and the status is denoted by $x_{t;s}$ 2 f0; 1g, where 1 refers to congested traffic and 0 refers to normal traffic. Denote $\mathbf{S}_t$ S as speed-observed links at time t, and

V $_t$ ¼ f$v_{t;s}$; s 2 $\mathbf{S}_t$g correspondingly as observations.

Traffic Related Tweet. A tweet d is depicted as a tuple ðt; $\mathbf{S}_d$; $\mathbf{w}_d$Þ, where t 2 T denotes the time that the tweet is posted, $\mathbf{S}_d$ S is the set of possible links implied by the tweet text, and $\mathbf{w}_d$ ¼ fw_1; . . . ; w_{Nd} g denotes the sequence of traffic related words in the tweet text. Note that $\mathbf{S}_d$ will con-tain multiple links if the location mentioned in tweet d is not specific, such as a street name containing multiple road seg-ments without finer information. Denote D_t ¼ fd_1; . . . ; d_{Dt} g as the traffic related tweet set at time t.

Trajectory and Travel Time. A trajectory or path p is denoted as a tuple ðt; $\mathbf{S}_p$; c_pÞ, where t 2 T is the time when the trajectory is generated given an OD pair, $\mathbf{S}_p$ S repre-sents consecutively connected links in the trajectory and c_p 2 R is the time cost traveling through the trajectory. Denote P_t ¼ fp_1; . . . ; p_{Pt} g as the trajectory set at time t, then C_t ¼ fc_1; . . . ; c_{Pt} g is the corresponding travel time cost set. The road length is represented as **L** 2 R^S with each compo-nent l_s equal to the road length of link s 2 S.

Problem Formulation (Road Traffic Speed Prediction Fusing Multi-source Data). Consider a set of road links S in the time duration of T , given a traffic related tweet corpus D ¼ f$D_t$jt 2 T g, a set of travel times C ¼ f$C_t$jt 2 T g with known road length **L**, and speed observations **V** ¼ f**V** $_t$jt 2 T g, our problem is to predict these unobserved traffic speed variables f$v_{s;t}$jt 2 T ; s 2 S $\mathbf{S}_t$g.

5.CONCLUSION

This paper proposes a novel probabilistic framework to pre-dict road traffic speed with multiple cross-domain data. Existing works are mainly based on speed sensing data, which suffers data sparsity and low coverage. In our work, we handle the challenges arising from fusing multi-source data, including location uncertainty, language ambiguity and data heterogeneity, using Location Disaggregation Model, Traffic Topic model and Traffic Speed Gaussian Process Model. Experiments on real data demonstrate the effectiveness and efficiency of our model. For Future work, we plan to implement kernel-based and distributive GP, so the traffic prediction framework can be applied into a real-time large traffic network.

References

[1] X. Yu and P. D. Prevedouros, "Performance and challenges in uti-lizing non-intrusive sensors for traffic data collection," Advances Remote Sens., vol. 2, pp. 45–50, 2013.

[2] S. Clark, "Traffic prediction using multivariate nonparametric regression," J. Transp. Eng., vol. 129, pp. 161–168, 2003.

[3] B. Williams, P. Durvasula, and D. Brown, "Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models," Transp. Res. Rec., vol. 1644, pp. 132–141, 1998.

[4] M. Kamarianakis and P. Prastacos, "Forecasting traffic flow condi-tions in an Urban network: Comparison of multivariate and uni-variate approaches," Transp. Res. Rec., vol. 1857, pp. 74–84, 2004.

[5] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," Transp. Res., vol. 19, pp. 606–616, 2011.

[6] S. M. Turner, W. L. Eisele, R. J. Benz, and D. J. Holdener, Travel Time Data Collection Handbook. Office Highway Inf. Manage., Federal Highway Administration, US Dept. Transportation, Washington, DC, USA, 1998.

[7] B. Abdulhai, H. Porwal, and W. Recker, "Short-term traffic flow prediction using neuro-genetic algorithms," J.-Intell. Transp. Syst., vol. 7, no. 1, pp. 3–41, 2002.

[8] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow fore-casting," Transp. Res., vol. 10, pp. 303–321, 2002.

[9] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicu-lar traffic flow as a seasonal stochastic time series process," No. LTVA/29242/CE99/103, 1999.

[10] Y. Kamarianakis and P. Prastacos, "Space-time modeling of traffic flow," Computers & Geosciences, vol. 31, no. 2, pp. 119–133, 2005.

[11] R. Giacomini and C. W. Granger, "Aggregation of space-time processes," J. Econometrics, vol. 118, pp. 7–26, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/ S0304407603001325

[12] M. Xinyu and H. Jianming, "Urban traffic network modeling and short-term traffic flow forecasting based on GSTARIMA model," in Proc. Int. IEEE Conf. Intell. Transp. Syst., 2010, pp. 19–22.

[13] B. Ghosh, B. Basu, and M. O'Mahony, "Multivariate short-term traffic flow forecasting using time-series analysis," IEEE Trans. Intell. Transp. Syst., vol. 10, no. 2, pp. 246–254, Jun. 2009.

[14] J. Guo and B. M. Williams, "Real-time short-term traffic speed level forecasting and uncertainty

quantification using layered kal-man filters," Transp. Res. Rec., vol. 2175, pp. 28–37, 2010.

Author's Profile:

V.Prasanthi Studying M.Tech in Malineni Lakshmaiah Engineering College, Singarayakonda, Prakasam(Dt), AP,India.

T.KishoreBabu
is completed his B.TechinMalineniLakshmaiah Engineering College in Singarayakonda ,prakasam dt and Completed his M.Tech in VidyaVikas Institute of Technology(Autonomous) Chevella, Rangareddy dist. He is interested in the subjects of Networks and Database. He was guided 20 batches of B.Tech students and 8 batches of M.Tech students. He has a total of 9 years experience in teaching .He is working as an associate professor in CSE department in Malineni Lakshmaiah Engineering College, Singarayakonda, Prakasam(Dt), AP, India.