Road Traffic Speed Prediction: A Probabilistic Model Fusing Multi-Source Data

S. SHOBANASHRI [1],Mr. S.SENTHILKUMAR [2], Mr. M.ANNAMALAI [3]

[1]PG Scholar/Department of Computer Science & Engineering/VMKV Engineering College Salem/TamilNadu/India shobanashri.santhanakrishnan@gmail.com

[2]Professor/ Department of Computer Science & Engineering/ VMKV Engineering College Salem/TamilNadu/Indian vmkvec.itsenthilkumar@gmail.com

[3]Professor/ Department of Computer Science & Engineering/ VMKV Engineering College Salem/TamilNadu/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

Background and Motivation

Road traffic monitoring is of great importance for urban transportation system. Traffic control agencies and drivers could benefit from timely and accurate road traffic prediction and make prompt, or even advance decisions possible for detecting and avoiding road congestions. Existing methods mainly focus on raw speed sensing data collected from cameras or road sensors, and suffer severe data scarcity issue because the installation and maintenance of sensors are very expensive . At the same time, most existing techniques based only on past and current traffic conditions do not fit well when real-world factors such as traffic accidents play a part. To address the above issues, in this paper we introduce new-type traffic related data arising from public services:

1) **Social media data**, which is posted on social networking websites, e.g. Twitter and Facebook. With the popularization of mobile devices, people are more likely to exchange news and trifles in their life through social media services, where messages about traffic conditions, such as “Stuck in traffic on E 32nd St. Stay away!”, are posted by drivers, passengers and pedestrians who can be viewed as sensors observing the ongoing traffic conditions near their physical locations. Meanwhile, traffic authorities register public accounts and post tweets to inform the public of the traffic status, such 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) 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.


speed sensor
trajectory sensor
tweet sensor
User_1
Just now
Stuck in traffic on E 32nd St.
Stay Away!
E 32nd St
10 min
30 km/h
10 km/h
40 km/h
E 31st St
20 km/h
28 km/h
E 30th St
7th Ave
6th Ave
5th Ave

2.SYSTEM MODEL, WEATHER NODES, EVENT AND SPECIAL DAY NODES, TIME RELATED NODES, TRAFFIC NODES

System model

A module is a part of a program. Programs are composed of one or more independently developed modules that are not combined until the program is linked. A single module can contain one or several routines.

Weather Nodes:

Five kinds of meteorological indexes, visibility, precipitation, snow depth, temperature and wind speed, are adopted.

Event and Special day Nodes:

Some nodes describe other local things, like local eventsand local special days. Events usually affect the traffic for a period of time around them.

Time Related Nodes:

Time-bin is the basic time unit in the model. Each timebin represents a short period of consecutive time. It is connected with the estimation of traffic conditions discussedlater. In simple terms, traffic conditions can only be estimated from data within certain ranges

Traffic Nodes:

Traffic node is that kind of node holding the traffic condition on each road segment on each time-bin. In this model,there are 2 future time-bins and 4 segments named by readerID pairs for both directions of Route 116 and Route 9. Sothere are 12 traffic nodes including current conditions.

3. METHODOLOGY

Road-Social Network

A road-social network is a pair of graphs (Gr,Gs), where Gr is a road network and Gs is a social network. Each vertex vs ∈Gs is associated with a set of vertices {vr} of Gr, indicating that user vs has been to these locations in {vr}. For example, Figure 1 shows a road-social network.

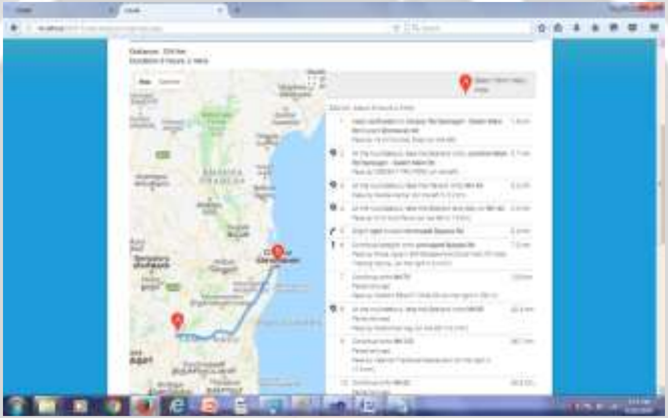
Social Influence in Road-Social Networks

As stated in the IC model, when a vertex u activates its inactiveout-neighbors such as v, u also has a probability to activate v's out-neighbors such as w, even though u has no direct edge to w. This prompted us to introduce the concept of influence and influenced-by for two remote vertices u and v in Gs. Let P = _u = w1, w2, ..., wm= v_ denote a path from u to v. Using the IC model, the probability that v is influenced by u (or u influences v) through this path equals to the product of propagation probabilities on edges along this path, In IGs, if there is a path P from u to v, we say v obtains a single influence in IGs, denoted by u → v. To compute SI(or), in IGs, we should consider multiple single influences to qs, since many users in the user set H have been to the same location.

4. Units

- A network of roads represented by poly lines
- At each intersection of two roads, a point/vertex is placed
- Between any two vertices on the road network, that segment has properties used in calculations (length of segment, time for traveling the segment, etc)
- **5.Results**

Traffic prediction problem can be broadly classified in to 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:



CONCLUSION

This project proposes a novel probabilistic framework to predict road traffic speed with multiple cross-domain data. Existing works are mainly based on speed sensing data, which suffers data spar sity and low coverage. In our work, wehandle the challenges arising from fusing multi-source data,including location uncertainty, language ambiguity and data heterogeneity, using Location Disaggregation Model, TrafficTopic model and Traffic Speed Gaussian Process Model. Experiments on real data demonstrate the effectivenessand efficiency of our model. For Future work, we plan toimplement kernel-based and distributive GP, so the trafficprediction framework can be applied into a real-time largetraffic network.

REFERENCES

[1]B. Abdulhai, H. Porwal, and W. Recker. Short-term traffic flowprediction using neuro-genetic algorithms.ITS Journal-IntelligentTransportation Systems Journal, 7(1):3–41, 2002.

[2]R. Alfelor, H. S. Mahmassani, and J. Dong. Incorporating weatherimpacts in traffic estimation and prediction systems.Technicalreport, US Department of Transportation, 2009.

[3]M. T. Asif, N. Mitrovic, L. Garg, and J. Dauwels.Low-dimensionalmodels for missing data imputation in road networks. 32(3):3527–3531, 2013.

[4]C. M. Bishop.Pattern Recognition and Machine Learning (InformationScience and Statistics). Springer-Verlag New York, Inc., 2006.

[5]D. M. Blei and J. D. Lafferty.Correlated topic models.InNIPS,pages 147–154, 2005.

[6]D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichletallocation.Journal of Machine Learning Research3:993–1022, 2003.

[7]J. Chen, K. H. Low, Y. Yao, and P. Jaillet.Gaussianprocessdecentralized data fusion and active sensing for spatiotemporaltraffic modeling and prediction in mobility-on-demand systems.IEEE Transactions on Automation Science and Engineering, 12(3):1–21, 2015.

[8]P.-T. Chen, F. Chen, and Z. Qian.Road traffic congestion monitoring in social media with hinge-loss markov random fields.InICDM, pages 80–89. IEEE, 2014.

[9]S. Clark.Traffic prediction using multivariate nonparametricregression.Journal of transportation engineering, pages 161–168,2003.

[10] Y. Gal, M. V. D. Wilk, and C. E. Rasmussen. Distributed variationalinference in sparse gaussian process regression and latent variablemodels.Advances in Neural Information Processing Systems, 4:3257–3265, 2014