Journal of Engineering Sciences
Vol 11, Issue 4 , April/2020
ISSN NO:0377-9254

ROAD TRAFFIC SPEED PREDICTION: A PROBABILISTIC MODEL FUSING MULTISOURCE DATA

Dr. K. UDAY KUMAR REDDY[1], M.MADHAVI[2], P.DIVYA[3], P.NIMAGNA[4], N.PAVAN KALYAN[5], B.GIRISH[6]
[1]Assistant Professor, Dept of CSE, AITS, Rajampet, AP, India.
[2,3,4,5,6]Student, Dept of CSE, AITS, Rajampet, 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 validate the effectiveness and efficiency of our model.

I. INTRODUCTION

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

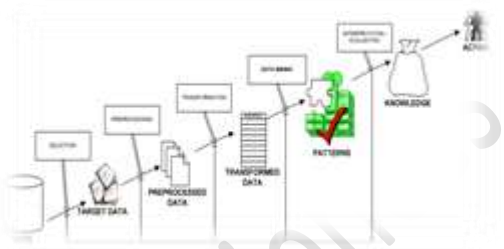


Fig.1: Structure of Data Mining

While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. Generally, any of four types of relationships are sought:

- **Classes**: Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.
- **Clusters**: Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.
- **Associations**: Data can be mined to identify associations. The beer-diaper example is an example of associative mining.
- **Sequential patterns**: Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

II. EXISTING SYSTEM

A trajectory-based community discovery method is proposed in the existing, 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). The prediction problem of rents/returns bike number is tackled using multiple features, e.g., time and meteorology, as measures

of similarity functions in multi-similarity based inference model.

DISADVANTAGES OF EXISTING SYSTEM:

- ❖ Existing methods mainly focus on raw speed sensing data collected from cameras or road sensors, and suffer severe data sparsity 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.

III. PROPOSED SYSTEM

In this paper we introduce new-type traffic related data arising from public services: 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. In this paper, we propose a unified statistical framework, entitled Topic Enhanced Gaussian Process Aggregation Model (TEGPAM) fusing multi-source data, which includes traditional speed sensing data, and new type "sensing" data from social media and map services. The framework combines the location disaggregation model to decompose vague locations into specific links, the traffic topic model to handle the language ambiguity in tweets and the Gaussian Process model to capture the spatial correlation in traffic sensing data. Real-time map and traffic services, e.g., Google Map and Uber, featured by location and navigation functions make travel more convenient. Given an origin-destination (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.

ADVANTAGES OF PROPOSED SYSTEM:

- ❖ Integration of data from multiple cross-domain sources: We implement the idea of improving traffic speed prediction by integrating speed sensing data with new-type traffic-related data, such as tweets and trajectories.
- ❖ Formulation of the unified TEGPAM framework: We propose a unified probabilistic framework TEGPAM that combines the disaggregation model, topic model with Gaussian Process model and is learned by variational methods and a stochastic EM algorithm.
- ❖ Extensive experiments to validate the performance of the proposed method. We validate our approach using real-world data collected from two large American cities. The extensive experiments show the effectiveness of TEGPAM, as well as the model efficiency and reliability.
- ❖ Elaborate analyses of introduced traffic-related data: We explore the impacts of different data sources, by decomposing TEGPAM into sub models and changing the combination ratio of datasets. Comparative experiments demonstrate the effectiveness of each data source.

IV. IMPLEMENTATION

MODULES:

- ❖ System Model
- ❖ Traffic Related Tweets
- ❖ Disaggregation of Tweets
- ❖ Traffic Topic Model

MODULES DESCSRIPTION:

System Model

In this module, we develop a System with a disaggregation model for location uncertainty in tweet and trajectory data, a traffic topic model for tweet language ambiguity and a GP model for capturing the spatial correlation of speed sensing data. In this module, first we develop the system construction entitles required for the proposed model. The system provides the new user for the registration and then login authorization. The authorized users can post their tweets. The users are provided with the option of the posting comments too. The module is designed with the features of Online Social Network modeled base, with the functionalities which are correlated to the proposed model.

Traffic Related Tweets

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. Tweets in the same time period and cities are collected via the Twitter REST search API. Traffic related tweets are preliminarily extracted by matching at least one term of a predefined vocabulary developed by domain experts, which included terms like "traffic", "accident", "stuck", "crash", etc., then further classified and filtered.

Disaggregation of Tweets

To handle the challenge of location uncertainty of new-type data, this section presents a disaggregation strategy to map the low-resolution data, which are tweets and trajectories, into specific road links. Since only 1 percent of tweets have geo-coordinates, most location informations are extracted from tweet text by mapping road names or alias.

Traffic Topic Model

To address the challenge of language ambiguity and capture the traffic description in tweets, a traffic topic model is proposed. With road records containing the geo-coordinates, names and aliases, we geocode tweets to road links by matching their geo-tag and text content to the front end of those links, which corresponds to the driving out direction and is denoted as Head. Different driving directions are denoted as different road links.

V. INPUT DESIGN AND OUTPUT DESIGN

INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

OBJECTIVES:

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

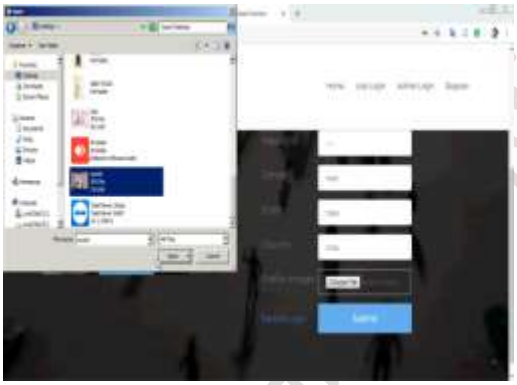
- Convey information about past activities, current status or projections of the Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

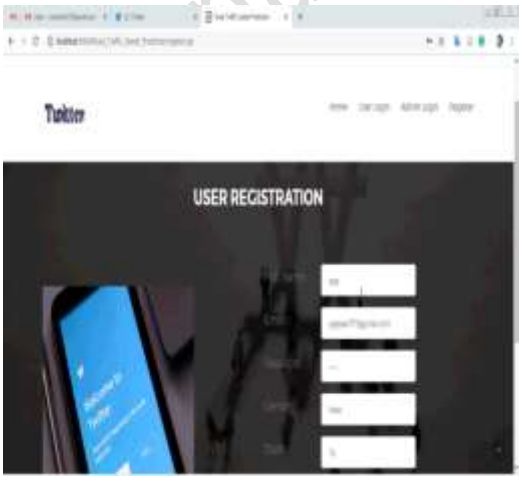
OUTPUTS


USER LOGIN

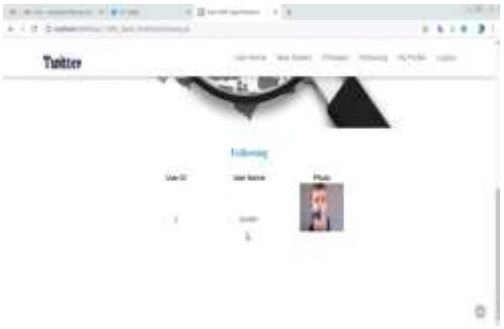

USER REGISTRATION

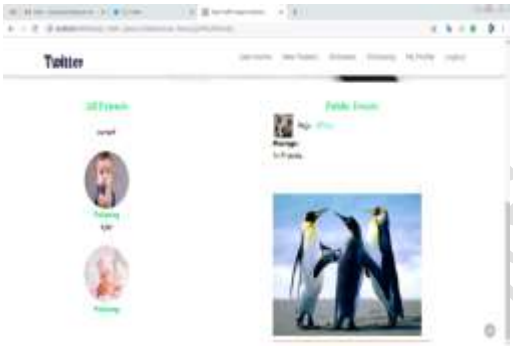

SOCIAL
MEDIA

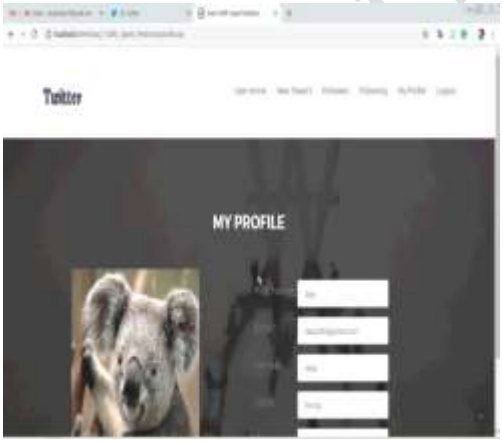



USER REGISTRATION

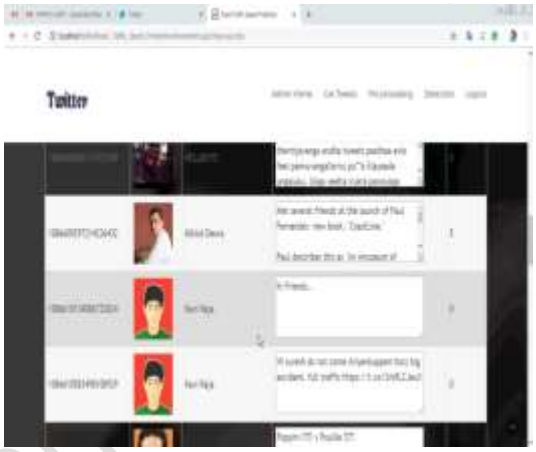
JES
Journal of Engineering Sciences
Vol 11, Issue 4 , April/2020
ISSN NO:0377-9254

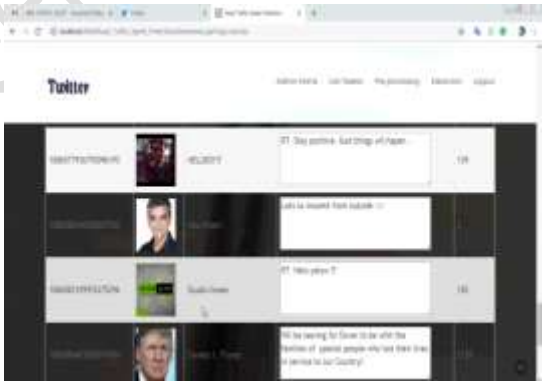


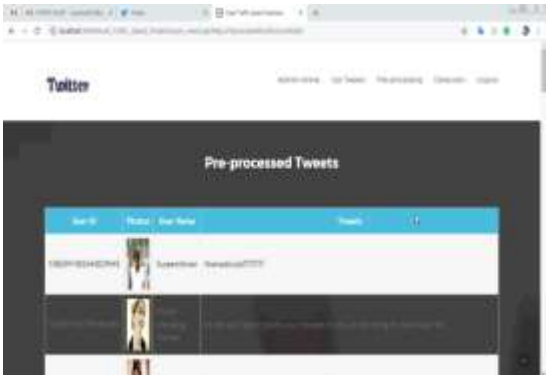




Journal of Engineering Sciences
Vol 11, Issue 4 , April/2020
ISSN NO:0377-9254







CONCLUSION

This paper 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 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 utilizing on-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 low 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 conditions in an Urban network: Comparison of multivariate and univariate 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 forecasting," Transp. Res., vol. 10, pp. 303–321, 2002.
[9] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehiculartraffic flow as a seasonal stochastic time series process," No.LTVA/29242/CE99/103, 1999.