Multi-Output Gaussian Processes for Crowdsourced Traffic Data Imputation

Traffic speed data imputation is a fundamental challenge for data-driven transport analysis. In recent years, with the ubiquity of GPS-enabled devices and the widespread use of crowdsourcing alternatives for the collection of traffic data, transportation professionals increasingly look to such user-generated data for a good deal of analysis, planning, and decision support applications. However, due to the mechanics of the data collection process, crowdsourced traffic data such as probe-vehicle data is highly prone to missing observations, making accurate imputation crucial for the success of any application that makes use of that type of data. In this paper, we propose the use of multi-output Gaussian processes (GPs) to model the complex spatial and temporal patterns in crowdsourced traffic data. While the Bayesian nonparametric formalism of GPs allows us to model observation uncertainty, the multi-output extension based on convolution processes effectively enables us to capture complex spatial dependencies between nearby road segments. Using six months of crowdsourced traffic speed data or "probe vehicle data" for several locations in Copenhagen, the proposed approach is empirically shown to significantly outperform popular state-of-the-art imputation methods.

General information
Publication status: Published
Organisations: Transport, Machine Learning, Department of Technology, Management and Economics, University of Washington
Contributors: Rodrigues, F., Henrickson, K., Pereira, F. C.
Pages: 594 - 603
Publication date: 28 Apr 2019
Peer-reviewed: Yes

Publication information
Journal: IEEE Transactions on Intelligent Transportation Systems
Multi-output bus travel time prediction with convolutional LSTM neural network

Accurate and reliable travel time predictions in public transport networks are essential for delivering an attractive service that is able to compete with other modes of transport in urban areas. The traditional application of this information, where arrival and departure predictions are displayed on digital boards, is highly visible in the city landscape of most modern metropolises. More recently, the same information has become critical as input for smart-phone trip planners in order to alert passengers about unreachable connections, alternative route choices and prolonged travel times. More sophisticated Intelligent Transport Systems (ITS) include the predictions of connection assurance, i.e. an expert system that will decide to hold services to enable passenger exchange, in case one of the services is delayed up to a certain level. In order to operate such systems, and to ensure the confidence of passengers in the systems, the information provided must be accurate and reliable. Traditional methods have trouble with this as congestion, and thus travel time variability, increases in cities, consequently making travel time predictions in urban areas a non-trivial task. This paper presents a system for bus travel time prediction that leverages the non-static spatio-temporal correlations present in urban bus networks, allowing the discovery of complex patterns not captured by traditional methods. The underlying model is a multi-output, multi-time-step, deep neural network that uses a combination of convolutional and long short-term memory (LSTM) layers. The method is empirically evaluated and compared to other popular approaches for link travel time prediction and currently available services, including the currently deployed model at Movia, the regional public transport authority in Greater Copenhagen. We find that the proposed model significantly outperforms all the other methods we compare with, and is able to detect small irregular peaks in bus travel times very quickly.

General information
Publication status: Published
Organisations: Transport Modelling, Department of Management Engineering, Transport DTU
Corresponding author: Petersen, N. C.
Contributors: Petersen, N. C., Rodrigues, F., Pereira, F. C.
Pages: 426-435
Publication date: 15 Apr 2019
Peer-reviewed: Yes
Combining time-series and textual data for taxi demand prediction in event areas: a deep learning approach

Accurate time-series forecasting is vital for numerous areas of application such as transportation, energy, finance, economics, etc. However, while modern techniques are able to explore large sets of temporal data to build forecasting models, they typically neglect valuable information that is often available under the form of unstructured text. Although this data is in a radically different format, it often contains contextual explanations for many of the patterns that are observed in the temporal data. In this paper, we propose two deep learning architectures that leverage word embeddings, convolutional layers and attention mechanisms for combining text information with time-series data. We apply these approaches for the problem of taxi demand forecasting in event areas. Using publicly available taxi data from New York, we empirically show that by fusing these two complementary cross-modal sources of information, the proposed models are able to significantly reduce the error in the forecasts.

Deep Learning from Crowds

Over the last few years, deep learning has revolutionized the field of machine learning by dramatically improving the state-of-the-art in various domains. However, as the size of supervised artificial neural networks grows, typically so does the need for larger labeled datasets. Recently, crowdsourcing has established itself as an efficient and cost-effective solution for labeling large sets of data in a scalable manner, but it often requires aggregating labels from multiple noisy contributors with different levels of expertise. In this paper, we address the problem of learning deep neural networks from crowds. We begin by describing an EM algorithm for jointly learning the parameters of the network and the reliabilities of the annotators. Then, a novel general-purpose crowd layer is proposed, which allows us to train deep neural networks end-to-end, directly from the noisy labels of multiple annotators, using only backpropagation. We empirically show that the proposed approach is able to internally capture the reliability and biases of different annotators and achieve new state-of-the-art results for various crowdsourced datasets across different settings, namely classification, regression and sequence labeling.
Heteroscedastic Gaussian processes for uncertainty modeling in large-scale crowdsourced traffic data

Accurately modeling traffic speeds is a fundamental part of efficient intelligent transportation systems. Nowadays, with the widespread deployment of GPS-enabled devices, it has become possible to crowdsource the collection of speed information to road users (e.g., through mobile applications or dedicated in-vehicle devices). Despite its rather wide spatial coverage, crowdsourced speed data also brings very important challenges, such as the highly variable measurement noise in the data due to a variety of driving behaviors and sample sizes. When not properly accounted for, this noise can severely compromise any application that relies on accurate traffic data. In this article, we propose the use of heteroscedastic Gaussian processes (HGP) to model the time-varying uncertainty in large-scale crowdsourced traffic data. Furthermore, we develop a HGP conditioned on sample size and traffic regime (SSRC-HGP), which makes use of sample size information (probe vehicles per minute) as well as previous observed speeds, in order to more accurately model the uncertainty in observed speeds. Using 6 months of crowdsourced traffic data from Copenhagen, we empirically show that the proposed heteroscedastic models produce significantly better predictive distributions when compared to current state-of-the-art methods for both speed imputation and short-term forecasting tasks.

General information
Publication status: Published
Organisations: Department of Management Engineering, Transport DTU, Transport Modelling
Contributors: Rodrigues, F., Pereira, F. C.
Pages: 636-651
Publication date: 2018
Peer-reviewed: Yes

Publication information
Journal: Transportation Research. Part C: Emerging Technologies
Volume: 95
ISSN (Print): 0968-090X
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
Original language: English
DOIs: 10.1016/j.trc.2018.08.007
Source: FindIt
Source-ID: 2438628180
Research output: Contribution to journal › Journal article – Annual report year: 2018 › Research › peer-review

Real-Time Taxi Demand Prediction using data from the web

In transportation, nature, economy, environment, and many other settings, there are multiple simultaneous phenomena happening that are of interest to model and predict. Over the last few years, the traffic data that we have at our disposal have significantly increased, and we have truly entered the era of big data for transportation. Most existing traffic flow prediction methods mainly focus on capturing recurrent mobility trends that relate to habitual/routine behaviour, and on exploiting short-term correlations with recent observation patterns. However, valuable information that is often available in the form of unstructured data is neglected when attempting to improve forecasting results. In this paper, we explore time-series data and textual information combinations using machine learning techniques in the context of creating a prediction model that is able to capture in real-time future stressful situations of the studied transportation system. Using publicly available taxi data from New York, we empirically show that the proposed models are able to significantly reduce the error in the forecasts. The final mean absolute error (MAE) of our predictions is decreased by 19.5% for a three months testing period and by 57% if we focus only on event periods.

General information
Publication status: Published
Organisations: Department of Management Engineering, Transport DTU, Transport Modelling
Contributors: Markou, I., Rodrigues, F., Pereira, F. C.
Pages: 1664-1671
Publication date: 2018

Host publication information
Title of host publication: 2018 21st International Conference on Intelligent Transportation Systems (ITSC)
Publisher: IEEE (2018 21st International Conference on Intelligent Transportation Systems (itsc)).
Keywords: time series forecasting, taxi demand, special events, textual data, topic modeling, machine learning
DOIs: 10.1109/ITSC.2018.8569015
Source: FindIt
Source-ID: 2442336200
A Bayesian Additive Model for Understanding Public Transport Usage in Special Events

Public special events, like sports games, concerts and festivals are well known to create disruptions in transportation systems, often catching the operators by surprise. Although these are usually planned well in advance, their impact is difficult to predict, even when organisers and transportation operators coordinate. The problem highly increases when several events happen concurrently. To solve these problems, costly processes, heavily reliant on manual search and personal experience, are usual practice in large cities like Singapore, London or Tokyo. This paper presents a Bayesian additive model with Gaussian process components that combines smart card records from public transport with context information about events that is continuously mined from the Web. We develop an efficient approximate inference algorithm using expectation propagation, which allows us to predict the total number of public transportation trips to the special event areas, thereby contributing to a more adaptive transportation system. Furthermore, for multiple concurrent event scenarios, the proposed algorithm is able to disaggregate gross trip counts into their most likely components related to specific events and routine behavior. Using real data from Singapore, we show that the presented model outperforms the best baseline model by up to 26 percent in R-2 and also has explanatory power for its individual components.

Demand pattern analysis of taxi trip data for anomalies detection and explanation

Due to environmental and economic stress, strong investment exists now towards adaptive transport systems that can efficiently utilize capacity, minimizing costs and environmental impacts. The common vision is a system that dynamically changes itself (the supply) to anticipate traveler needs (the demand). In some occasions, unexpected and unwanted demand patterns are noticed in the traffic network that lead to system failures and cost implications. Significantly low speeds or excessively low flows at an unforeseeable time are only some of the phenomena that are often noticed and need to be explained for transport system’s better future response. The objective of this research is the formulation of a proper methodology that identifies anomalies on traffic networks and correlates them with special events using internet data. Our main subject of interest is the investigation of why traffic congestion is happening as well as why there are demand fluctuations in days were there are no apparent reasons for the occurrence of such phenomena. We evaluated our system using Google’s NYC taxi trips public dataset. We defined initially the “normality” baseline and thereafter we studied individual days’ demand patterns for outliers’ detection. Our approach enabled us to detect demand fluctuations, analyze and correlate them with disruptive events scenarios like extreme weather conditions, public holidays, religious festivities and parades. Using kernel density analysis, the affected areas as well as the significance of the observed differences compared to the average day are depicted.
Event characteristics that disrupt transport system's balance

The life of the city is often reflected in traffic patterns: popular sporting events draw crowds, holidays create disruptions, protests may result in road closures, etc. Decades of research on travel demand and network modelling already provide satisfying predictive tools. However, the main research focus has been on regular behaviour, such as peak/off-peak cycles, regular functioning of the infrastructure, and normal weather conditions. Consequently, non-recurrent events severely challenge such models. Under non-recurrent circumstances, the typically expected correlation structures (e.g. between demand flows in neighbor areas; between current and recent values of traffic speeds or travel times) are drastically affected, severely affecting predictions. It is therefore necessary to take into consideration data from different sources. The objective of this research is the development of a methodology that correlates high taxi demand observations with popular events retrieved from Social Media platforms. Using NYC taxi trips public dataset, the average demand of the day was determined using kernel density analysis. Days that showed significant outliers compared to the average day were further studied using a dataset of around 116000 events. The second dataset was retrieved from the Web for the same 6 months period through the direct use of APIs. The correlation step includes the comparison of spatial and temporal kernel density depiction of taxi pick-up locations and events retrieved details. Through the correlation evaluation of traffic data and semantic information, conclusions were made on how the demand of taxi pick-ups changes based on certain event characteristics.

Learning Supervised Topic Models for Classification and Regression from Crowds

The growing need to analyze large collections of documents has led to great developments in topic modeling. Since documents are frequently associated with other related variables, such as labels or ratings, much interest has been placed on supervised topic models. However, the nature of most annotation tasks, prone to ambiguity and noise, often with high volumes of documents, deem learning under a single-annotator assumption unrealistic or unpractical for most real-world applications. In this article, we propose two supervised topic models, one for classification and another for regression problems, which account for the heterogeneity and biases among different annotators that are encountered in practice when learning from crowds. We develop an efficient stochastic variational inference algorithm that is able to scale to very large datasets, and we empirically demonstrate the advantages of the proposed model over state-of-the-art approaches.
Use of Taxi-Trip Data in Analysis of Demand Patterns for Detection and Explanation of Anomalies

Because of environmental and economic stress, current strong investment in adaptive transport systems can efficiently use capacity, minimizing costs and environmental impacts. The common vision is of a system that dynamically changes itself (the supply) to anticipate the needs of travelers (the demand). In some occasions, unexpected and unwanted demand patterns are noticed in the traffic network; these patterns lead to system failures and cost implications. Significantly, low speeds or excessively low flows at an unforeseeable time are only some of the phenomena that are often noticed and need to be explained for a transport system to develop a better future response. The objective of this research was the formulation of a methodology that could identify anomalies on traffic networks and correlate them with special events by using Internet data. The main subject of interest in this study was the investigation of why traffic congestion was occurring as well as why demand fluctuated on days when there were no apparent reasons for such phenomena. The system was evaluated by using Google's public data set for taxi trips in New York City. A "normality" baseline was defined at the outset and then used in the subsequent study of the demand patterns of individual days to detect outliers. With the use of this approach it was possible to detect fluctuations in demand and to analyze and correlate them with disruptive event scenarios such as extreme weather conditions, public holidays, religious festivities, and parades. Kernel density analysis was used so that the affected areas, as well as the significance of the observed differences compared with the average day, could be depicted.

General information
Publication status: Published
Organisations: Department of Management Engineering, Transport DTU, Transport Modelling
Contributors: Markou, I., Rodrigues, F., Pereira, F. C.
Pages: 129-138
Publication date: 2017
Peer-reviewed: Yes

Publication information
Journal: Transportation Research Record
Volume: 2643
ISSN (Print): 0361-1981
Ratings:
BFI (2017): BFI-level 1
Scopus rating (2017): CiteScore 0.83 SJR 0.589 SNIP 0.708
Web of Science (2017): Impact factor 0.695
Web of Science (2017): Indexed yes
Original language: English
Electronic versions:
. Embargo ended: 01/01/2018
DOIs:
10.3141/2643-15
Source: FindIt
Source-ID: 2389539187
Research output: Contribution to journal › Journal article – Annual report year: 2017 › Research › peer-review
Using internet search queries to predict human mobility in social events

While our transport systems are generally designed for habitual behavior, the dynamics of large and mega cities systematically push it to its limits. Particularly, transport planning and operations in large events are well known to be a challenge. Not only they imply stress to the system on an irregular basis, their associated mobility behavior is also difficult to predict. Previous studies have shown a strong correlation between number of public transport arrivals with the semi-structured data mined from online announcement websites. However, these models tend to be complex in form and demand substantial information retrieval, extraction and data cleaning work, and so they are difficult to generalize from city to city. In contrast, this paper focuses on enriching previously mined information about special events using automated web search queries. Since this context data comes in unstructured natural language form, we employ supervised topic model to correlate it with real measurements of transport usage. In this way, the proposed approach is more generic and a transit agency can start planning ahead as early as the event is announced on the web. The results show that using information mined from the web search not only shows high prediction accuracy of public transport demand, but also potentially provides interesting insights about popular event categories based on extracted topics.

General information
Publication status: Published
Organisations: Department of Management Engineering, Transport Modelling, Transport DTU, KTH - Royal Institute of Technology, MIT-Singapore Program, Los Alamos National Laboratory
Contributors: Borysov, S., Lourenco, M., Rodrigues, F., Balatsky, A., Pereira, F. C.
Number of pages: 6
Pages: 1342-1347
Publication date: 2016

Host publication information
Title of host publication: Proceedings of 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)
Publisher: IEEE
Keywords: Data models, Predictive models, Internet, Urban areas, Web search, Planning
DOI:
10.1109/ITSC.2016.7795731
Source: FindIt
Source-ID: 2350291944
Research output: Chapter in Book/Report/Conference proceeding › Article in proceedings – Annual report year: 2016 › Research › peer-review

Activities:

6th AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2018) (Event)
Period: 5 Jul 2018 → 8 Jul 2018
Filipe Rodrigues (Member)
Department of Management Engineering
Transport DTU
Transport Modelling

Description
Program committee member
Degree of recognition: International

Related event
6th AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2018)
05/07/2018 → 08/07/2018
Activity: Membership › Membership of commitees, commissions, boards, councils, associations, organisations, or similar

Deep Learning from Crowds
Period: 2 Feb 2018 → 7 Feb 2018
Filipe Rodrigues (Speaker)
Department of Management Engineering
Transport DTU
Transport Modelling
Degree of recognition: International

Related event

**The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI), 2018**
02/03/2018 → 07/03/2018
Activity: Talks and presentations › Conference presentations

*Predicting for the adaptive transport system and other necessary ingredients for resilient urban mobility*
Period: 15 Dec 2017
Filipe Rodrigues (Invited speaker)
Department of Management Engineering
Transport DTU
Transport Modelling
Degree of recognition: International

Related event

**Leveraging Big Data for Future Mobility Workshop**
15/12/2017 → 15/12/2017
Munich, Germany
Activity: Talks and presentations › Guest lectures, external teaching and course activities at other universities

31st Conference on Neural Information Processing Systems
Period: 4 Dec 2017 → 9 Dec 2017
Filipe Rodrigues (Participant)
Department of Management Engineering
Transport DTU
Transport Modelling
Degree of recognition: International

Related event

31st Conference on Neural Information Processing Systems
04/12/2017 → 09/12/2017
Long Beach, United States
Activity: Attending an event › Participating in or organising a conference

30th Annual Conference on Neural Information Processing Systems
Period: 5 Dec 2016 → 10 Dec 2016
Filipe Rodrigues (Participant)
Department of Management Engineering
Transport DTU
Transport Modelling
Degree of recognition: International

Related event

30th Annual Conference on Neural Information Processing Systems
05/12/2016 → 10/12/2016
Barcelona, Spain
Activity: Attending an event › Participating in or organising a conference
**Machine Learning for Transportation**
Period: 1 Jun 2016 → 3 Jun 2016
Filipe Rodrigues (Guest lecturer)
Department of Management Engineering
Transport DTU
Transport Modelling
Degree of recognition: International

**Related event**

**2016 Summer School on Big Data, Mobility Patterns and Transport Analytics**
01/06/2016 → 03/06/2016
Activity: Talks and presentations › Guest lectures, external teaching and course activities at other universities