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.
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.
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.

General information
Publication status: Published
Organisations: Transport, Machine Learning, Department of Technology, Management and Economics, Transport Demand
Contributors: Rodrigues, F., Markou, I., Pereira, F. C.
Pages: 120-129
Publication date: 2019
Peer-reviewed: Yes

Comparison of Four Types of Artificial Neural Network and a Multinomial Logit Model for Travel Mode Choice Modeling

Discrete choice modeling is a fundamental part of travel demand forecasting. To date, this field has been dominated by parametric approaches (e.g., logit models), but non-parametric approaches such as artificial neural networks (ANNs) possess much potential since choice problems can be assimilated to pattern recognition problems. In particular, ANN models are easily applicable with their higher capability to identify nonlinear relationships between inputs and designated outputs to predict choice behaviors. This article investigates the capability of four types of ANN model and compares their prediction performance with a conventional multinomial logit model (MNL) for mode choice problems. The four ANNs are: backpropagation neural networks (BPNNs), radial basis function networks (RBFNs), probabilistic neural networks (PNNs), and clustered probabilistic neural networks (CPNNs). To compare the modeling techniques, we present the algorithmic differences of each ANN technique, and we assess their prediction accuracy with a 10-fold cross-validation method. Furthermore, we assess the contribution of explanatory variables by conducting sensitivity analyses on significant variables. The results show that ANN models outperform MNL, with prediction accuracies around 80% compared with 70% for MNL. Moreover, PNN performs best out of all ANNs, especially to predict underrepresented modes.

General information
Publication status: Published
Organisations: Transport, Machine Learning, Department of Technology, Management and Economics, University of Illinois
Corresponding author: Lee, D.
Contributors: Lee, D., Derrible, S., Pereira, F. C.
Publication date: 2019
Peer-reviewed: Yes

Research output: Contribution to journal › Journal article – Annual report year: 2019 › Research › peer-review

Publication information
Journal: Transportation Research Record
Volume: 2672
Issue number: 49
ISSN (Print): 0361-1981
Ratings:
BFI (2019): BFI-level 1
Factors driving the adoption of mobility-management travel app: a bayesian structural equation modelling analysis

**General information**
Publication status: Published
Organisations: Transport, Machine Learning, Department of Technology, Management and Economics, Network and Route Choice, Hebrew University of Jerusalem, Technical University of Lisbon
Contributors: Mehdizadeh Dastjerdi, A., Kaplan, S., de Abreu e Silva, J., Nielsen, O. A., Pereira, F. C.
Publication date: 2019

**Host publication information**
Title of host publication: Proceedings of the 98th Annual Meeting of the Transportation Research Board
Electronic versions:
TRB_paper_Revised_Version_03_11_2018.pdf
Source: PublicationPreSubmission
Source-ID: 166480678

**Online Framework for Demand-Responsive Stochastic Route Optimization**
This study develops an online predictive optimization framework for operating a fleet of autonomous vehicles to enhance mobility in an area, where there exists a latent spatio-temporal distribution of demand for commuting between locations. The proposed framework integrates demand prediction and supply optimization in the network design problem. For demand prediction, our framework estimates a marginal demand distribution for each Origin-Destination pair of locations through Quantile Regression, using counts of crowd movements as a proxy for demand. The framework then combines these marginals into a joint demand distribution by constructing a Gaussian copula, which captures the structure of correlation between different Origin-Destination pairs. For supply optimization, we devise a demand-responsive service, based on linear programming, in which route structure and frequency vary according to the predicted demand. We evaluate our framework using a dataset of movement counts aggregated from WiFi records of a university campus in Denmark, and the results show that our framework outperforms conventional methods for route optimization, which do not utilize the full predictive distribution.

**General information**
Publication status: Accepted/In press
Organisations: Transport, Machine Learning, Department of Technology, Management and Economics, Network and Route Choice, Nanyang Technological University
Corresponding author: Peled, I.
Contributors: Peled, I., Lee, K., Jiang, Y., Dauwels, J., Pereira, F. C.
Number of pages: 32
Publication date: 2019
Peer-reviewed: Yes

**Publication information**
Journal: ArXiv
Original language: Undefined/Unknown
Keywords: stat.ML, cs.LG, math.OC, stat.AP
Source: FindIt
Source-ID: 2444432998

**Participating in environmental loyalty program with a real-time multimodal travel app: User needs, environmental and privacy motivators**
The increasing complexity and demand of transport services strains transportation systems especially in urban areas with limited possibilities for building new infrastructure. The solution to this challenge requires changes of travel behavior. One of the proposed means to induce such change is multimodal travel apps. However, understanding the motivators underlying individuals’ travel intentions is essential to design and evaluate their effectiveness. This paper pinpoints and analyses the drivers and barriers that influence individual travel decisions when using such apps. The analytical
framework relies on Alderfer's ERG model of human needs that relate the individual's intentions to three domains, namely (1) Existence, (2) Relatedness and (3) Growth needs. Furthermore, environmental attitude, information privacy concerns and perceived difficulties when using the system are incorporated as to better explain user-sided heterogeneity. The case-study focuses on a new travel information system in Copenhagen (Denmark), which is not yet operational, through a technology-use preference survey among 828 travelers. Structural equation models revealed that the motivation for choices are specific to individual users and depend on wide-ranging factors that go beyond traditional economic and socio-demographic methods. The study revealed (1) different intentions among individuals according to the perceived value of the new information system, (2) a relation between different environmental attitude constructs and users' needs, (3) a stronger appeal to use the system for individuals with higher needs of developing social self-concept and eco-travel self-efficacy as well as with lower perceived privacy risk and perceived difficulties, (4) that both functional and psychological factors affect adoption intention.

**Predicting taxi demand hotspots using automated Internet Search Queries**

Disruptions due to special events are a well-known challenge in transport operations, since the transport system is typically designed for habitual demand. Part of the problem relates to the difficulty in collecting comprehensive and reliable information early enough to prepare mitigation measures. A tool that automatically scans the internet for events and predicts their impact would strongly support transport management in many cities in the world. This study addresses the challenges related to retrieving and analyzing web documents about real world events, and using them for demand explanation (if related to a past event) and prediction (if a future one). Transport demand is predicted with a supervised topic modeling algorithm by utilizing information about social events retrieved using various strategies, which made use of search aggregation, natural language processing, and query expansion. It was found that a two-step process produced the highest accuracy for transport demand prediction, where different (but related) queries are used to retrieve an initial set of documents, and then, based on these documents, a final query is constructed that obtains the set of predictive documents. These are then used to model the most discriminating topics related to the transport demand. A framework was proposed that sequentially handles all stages of data gathering, enrichment, and prediction with the intention of generating automated search queries.

Population synthesis is concerned with the generation of synthetic yet realistic representations of populations. It is a fundamental problem in the modeling of transport where the synthetic populations of micro agents represent a key input to most agent-based models. In this paper, a new methodological framework for how to grow pools of micro agents is presented. This is accomplished by adopting a deep generative modeling approach from machine learning based on a Variational Autoencoder (VAE) framework. Compared to the previous population synthesis approaches based on Iterative Proportional Fitting (IPF), Markov Chain Monte Carlo (MCMC) sampling or traditional generative models, the proposed method allows unparalleled scalability with respect to the number and types of attributes. In contrast to the approaches that rely on approximating the joint distribution in the observed data space, VAE learns its compressed latent representation. The advantage of the compressed representation is that it avoids the problem of the generated samples being trapped in local minima when the number of attributes becomes large. The problem is illustrated using the Danish National Travel Survey data, where the Gibbs sampler fails to generate a population with 21 attributes (corresponding to the 121-dimensional joint distribution). At the same time, VAE shows acceptable performance when 47 attributes (corresponding to the 357-dimensional joint distribution) are used. Moreover, VAE allows for growing agents that are virtually different from those in the original data but have similar statistical properties and correlation structure. The presented approach will help modelers to generate better and richer populations with a high level of detail, including smaller zones, personal details and travel preferences.

Active Learning for Input Space Exploration in Traffic Simulators

Urban environments are systems of overwhelming complexity and dynamism, involving numerous variables and idiosyncrasies which not usually easy to model from a functional perspective. Simulation modeling is a common and well-accepted approach to study such systems, specially those that prove to be too complex to be analyzed by standard analytic methods. However, such urban simulation models can become computationally very expensive to run. To address this drawback, simulation metamodels can be employed to approximate the underlying simulation function. In this paper, we propose a batch-mode active learning strategy based on Gaussian Processes metamodeling that searches for the most informative data points in batches with respect to their corresponding predictive variances. These points are selected in such a way that they originate from different high variance neighborhoods. Eventually, this allows us to analyze the simulation output behavior with fewer simulation requests. Using an illustrative traffic simulation example, the results show that the proposed restricted batch-mode strategy is able to increase the simulation input space exploration efficiency in comparison with standard batch-mode strategies.

Host publication information
Title of host publication: 2018 International Joint Conference on Neural Networks (IJCNN)
Activity Recognition for a Smartphone and Web-based Human Mobility Sensing System

Activity-based models in transport modeling and prediction are built from a large number of observed trips and their purposes. However, data acquired through traditional interview-based travel surveys is often inaccurate and insufficient. Recently, a human mobility sensing system, called Future Mobility Survey (FMS), was developed and used to collect travel data from more than 1,000 participants. FMS combines a smartphone and interactive web interface in order to better infer users activities and patterns. This paper presents a model that infers an activity at a certain location. We propose to generate a set of predictive features based on spatial, temporal, transitional, and environmental contexts with an appropriate quantization. In order to improve the generalization performance of the proposed model, we employ a robust approach with ensemble learning. Empirical results using FMS data demonstrate that the proposed method contributes significantly to providing accurate activity estimates for the user in our travel-sensing application.

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.
Efficient Transport Simulation With Restricted Batch-Mode Active Learning

Simulation modeling is a well-known and recurrent approach to study the performance of urban systems. Taking into account the recent and continuous transformations within increasingly complex and multidimensional cities, the use of simulation tools is, in many cases, the only feasible and reliable approach to analyze such dynamic systems. However, simulation models can become very time consuming when detailed input-space exploration is needed. To tackle this problem, simulation metamodels are often used to approximate the simulators’ results. In this paper, we propose an active learning algorithm based on the Gaussian process (GP) framework that gathers the most informative simulation data points in batches, according to both their predictive variances and to the relative distance between them. This allows us to explore the simulators’ input space with fewer data points and in parallel, and thus in a more efficient way, while avoiding computationally expensive simulation runs in the process. We take advantage of the closeness notion encoded into the GP to select batches of points in such a way that they do not belong to the same high-variance neighborhoods. In addition, we also suggest two simple and practical user-defined stopping criteria so that the iterative learning procedure can be fully automated. We illustrate this methodology using three experimental settings. The results show that the proposed methodology is able to improve the exploration efficiency of the simulation input space in comparison with non-restricted batch-mode active learning procedures.

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 crowdsourcethe 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 6months 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

**Just-In-Time Traffic Model Adaptation to Non-Recurrent Incidents**

**General information**
Publication status: Published
Organisations: Transport DTU, Transport Modelling, Department of Management Engineering
Contributors: Peled, I., Pereira, F. C.
Number of pages: 1
Publication date: 2018
Peer-reviewed: Yes
Electronic versions:
abstract_for_iccs_2018.pdf
Research output: Contribution to conference › Conference abstract for conference – Annual report year: 2018 › Research › peer-review

**Opening Up the Conversation: Topic Modeling for Automated Text Analysis in Travel Surveys**
The difficulties associated with the collection and analysis of open-ended questions in surveys (e.g. “Tell us your opinion about …”), have encouraged the widespread use of closed-ended responses (e.g. “Your opinion according to a scale of 1–5 …”). These, in some circumstances, are very restrictive, curbing the recognition of nuances in the survey population, hence limiting our models. Latent Dirichlet Allocation (LDA), now opens avenues for extraction of open-ended responses. In this paper, we analyze the potential of LDA in open-ended responses, by comparing with closed-ended counterpart options. A questionnaire designed based on Theory of Planned Behavior (TPB), is used to collect information on the intentions to use shared autonomous mobility. Two versions of the questionnaire were used, that alternately allow for open- and closed-ended versions of the same questions. Factor analysis was used to construct factors from the Likert scale questions and LDA was used to extract information from open-ended questions. Ordered Probit models were estimated to predict the intention to use shared autonomous mobility services. Attitudes, socio-demographics and current travel behavior of individuals were used in the model. The questionnaire instrument played an important decision in an individual's decision to use the proposed service. Comparable results were obtained for the two versions of the questionnaire, which emphasizes the need for further research on use of open-ended questions in measuring attitudes.

**General information**
Publication status: Published
Organisations: Transport DTU, Transport Modelling, Department of Management Engineering, Universidade Técnica de Lisboa
Contributors: Baburajan, V., De Abreu e Silva, J., Pereira, F. C.
Pages: 3657-3661
Publication date: 2018

**Host publication information**
Title of host publication: 2018 21st International Conference on Intelligent Transportation Systems (ITSC)
Overview of traffic incident duration analysis and prediction
Non-recurrent congestion caused by traffic incident is difficult to predict but should be dealt with in a timely and effective manner to reduce its influence on road capacity reduction and enormous travel time loss. Influence factor analysis and reasonable prediction of traffic incident duration are important in traffic incident management to predict incident impacts and aid in the implementation of appropriate traffic operation strategies. The objective of this study is to conduct a thorough review and discusses the research evolution, mainly including the different phases of incident duration, data resources, and the various methods that are applied in the traffic incident duration influence factor analysis and duration time prediction. In order to achieve the goal of this study, we presented a systematic review of traffic incident duration time estimation and prediction methods developed based on various data resource, methodologies etc. Based on the previous studies, we analyse (i) Data resources and characteristics: different traffic incident time phases, data set size, incident types, duration time distribution, available data resources, significant influence factors and unobserved heterogeneity and randomness, (ii) traffic incident duration analysis methods, mainly including hazard-based duration model and regression and statistical tests, (iii) traffic incident duration prediction methods and evaluation of prediction accuracy. After a comprehensive review of literature, this study identifies and analyses future challenges and what can be achieved in the future to estimate and predict the traffic incident duration time.

General information
Publication status: Published
Organisations: Department of Management Engineering, Transport DTU, Transport Modelling, Tsinghua University, Massachusetts Institute of Technology
Contributors: Li, R., Pereira, F. C., Ben-Akiva, M. E.
Number of pages: 13
Publication date: 2018
Peer-reviewed: Yes

Publication information
Journal: European Transport Research Review
Volume: 10
Issue number: 2
Article number: 22
ISSN (Print): 1867-0717
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
Original language: English
Keywords: Incident duration analysis, Traffic incident duration prediction, Hazard-based duration model, Data mining, Influence factors
Electronic versions:
filestore_2_.pdf
DOIs:
10.1186/s12544-018-0300-1

Bibliographical note
© The Author(s). 2018 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Population Synthesis Meets Deep Generative Modelling
Agent-based transport models depend to a high degree on the formation of the underlying population. Models and methods for generating such populations, possibly under constraints to reflect future margins, are commonly referred to as “population synthesis” models. Historically, different approaches have been proposed, ranging from deterministic
approaches, such as the Iterative Proportional Fitting (IPF) algorithm (Deming and Stephan, 1940; Rich and Mulalic, 2012), to simulation based approaches of the Markov Chain Monte Carlo (MCMC) type (Farooq et al., 2013). Often, matrix fitting methods such as the IPF has been used in combination with postsimulation based methods in order to translate prototypical individuals into true micro agents. While the existing models are capable of producing acceptable results for agents with relative few socioeconomic and spatial characteristics, these methods do not scale well when the dimensionality of the underlying distribution becomes large. As a result, in many cases, these methods are not able to accommodate the increasing need for more dimensions that result from, e.g. smaller zones, the combination of household-based and individual-based synthesis and more detailed variables in general. In this paper, we propose a different approach to population synthesis based on generative models from the deep learning framework. Contrarily to existing methods, these new methods are scalable and can handle a very large number of both numerical and categorical attributes at the same time.

General information
Publication status: Published
Organisations: Transport DTU, Transport Modelling, Department of Management Engineering
Contributors: Borysov, S., Rich, J., Pereira, F. C.
Number of pages: 3
Publication date: 2018
Peer-reviewed: Yes
Research output: Contribution to conference → Paper – 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
Research output: Chapter in Book/Report/Conference proceeding → Article in proceedings – Annual report year: 2018 → Research → peer-review

Social network analysis in future transportation systems: Contributions on observability, behaviour and structure

General information
Publication status: Published
Organisations: Department of Management Engineering, Transport DTU, Transport Modelling, Technical University of Munich, University of Cyprus
Contributors: Pereira, F. C., Antoniou, C., Dimitriou, L.
Number of pages: 2
Pages: 369-370
Publication date: 2018
Towards Dynamic Bayesian Networks: State Augmentation for Online Calibration of DTA Systems

A key component of Dynamic Traffic Assignment (DTA) systems is the online calibration of simulation parameters, which is crucial in generating accurate predictions of network states. A widely used approach for online calibration is the Kalman filter which allows for the incorporation of demand and supply parameters and any type of measurement data. This paper presents a Dynamic Bayesian Network extension for traditional Kalman filters with a technique called state augmentation. Although it has been discussed in the calibration literature, the usage and applicability were not fully investigated. The state augmentation technique is particularly useful for delayed systems, for example in large networks with high travel times. In this paper, we discuss state augmentation for Kalman filtering and illustrate its modeling advantages via a Dynamic Bayesian Network (DBN) representation. These advantages are demonstrated by a case study using the Singapore expressway network. The results indicate that employing state augmentation yields better estimation and prediction accuracy of traffic states, around 10% less error than the standard extended Kalman filter.
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 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.
Enabling Bus Transit Service Quality Co-Monitoring Through Smartphone-Based Platform

The growing ubiquity of smartphones offers public transit agencies an opportunity to transform ways to measure, monitor, and manage service performance. The potential of a new tool is demonstrated for engaging customers in measuring satisfaction and co-monitoring [Editor's note: This is the authors' word, meaning "agencies using public feedback to supplement official monitoring and regulation."] bus service quality. The pilot project adapted a smartphone-based travel survey system, Future Mobility Sensing, to collect real-time customer feedback and objective operational measurements on specific bus trips. The system used a combination of GPS, Wi-Fi, Bluetooth, and accelerometer data to track transit trips while soliciting users' feedback on trip experience. Though not necessarily intended to replace traditional monitoring channels and processes, these data can complement official performance monitoring through a more real-time, customer-centric perspective. The pilot project operated publicly for 3 months on the Silver Line bus rapid transit in Boston, Massachusetts. Seventy-six participants completed the entrance survey; half of them actively participated and completed more than 500 questionnaires while on board either at the end of a trip, at the end of a day, or both. Participation was biased toward frequent Silver Line users, the majority of whom were white and of higher income. Indicative models of user-reported satisfaction reveal some interesting relationships, but the models can be improved by fusing the app-collected data with actual performance characteristics. Broader and more sustained user engagement remains a critical future challenge.

Enhancing Resilience to Disasters using Social Media

During the last decade, Social Media (SM) have emerged as a prominent trend in social communication, with online platforms such as Facebook and Twitter to conquer the internet space with millions of visitors per day. SM usage generates an astonishing amount of information, which could be used for scarcely experienced situations, such as mass convergence and emergency events. This study presents a preliminary exploratory analysis on examining the capacity of Social Media to extract information on individuals choices during evacuation. We collect tweets from the evacuation in Oroville, California USA due to danger of flood and the evacuation. The data is used for the creation of a user sample which allows the collection of historical data. The historical data is compared with the data collected during and after the evacuation.
evacuation. The goal of this comparison is the extraction of potential information related to the evacuation.

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.
Probabilistic Modeling and Visualization for Bankruptcy Prediction

In accounting and finance domains, bankruptcy prediction is of great utility for all of the economic stakeholders. The challenge of accurate assessment of business failure prediction, specially under scenarios of financial crisis, is known to be complicated. Although there have been many successful studies on bankruptcy detection, seldom probabilistic approaches were carried out. In this paper we assume a probabilistic point-of-view by applying Gaussian Processes (GP) in the context of bankruptcy prediction, comparing it against the Support Vector Machines (SVM) and the Logistic Regression (LR). Using real-world bankruptcy data, an in-depth analysis is conducted showing that, in addition to a probabilistic interpretation, the GP can effectively improve the bankruptcy prediction performance with high accuracy when compared to the other approaches. We additionally generate a complete graphical visualization to improve our understanding of the different attained performances, effectively compiling all the conducted experiments in a meaningful way. We complete our study with an entropy-based analysis that highlights the uncertainty handling properties provided by the GP, crucial for prediction tasks under extremely competitive and volatile business environments.

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

Mapping Social Media for Transportation Studies

General information
Publication status: Published
Organisations: Department of Management Engineering, Transport DTU, Transport Modelling, Technical University of Munich
Contributors: Chaniotakis, E., Antoniou, C., Pereira, F. C.
Number of pages: 7
Pages: 64-70
Publication date: 2016
Peer-reviewed: Yes

Publication information
Journal: IEEE Intelligent Systems
Volume: 31
Issue number: 6
ISSN (Print): 1541-1672
Ratings:
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 3.7 SJR 0.826 SNIP 2.104
Web of Science (2016): Impact factor 2.374
Web of Science (2016): Indexed yes
Original language: English
Uncertainty in Bus Arrival Time Predictions: Treating Heteroscedasticity With a Metamodel Approach

Arrival time predictions for the next available bus or train are a key component of modern traveler information systems (TISs). A great deal of research has been conducted within the intelligent transportation system community in developing an assortment of different algorithms that seek to increase the accuracy of these predictions. However, the inherent stochastic and nonlinear nature of these systems, particularly in the case of bus transport, means that these predictions suffer from variable sources of error, stemming from variations in weather conditions, bus bunching, and numerous other sources. In this paper, we tackle the issue of uncertainty in bus arrival time predictions using an alternative approach.

Rather than endeavor to develop a superior method for prediction, we take existing predictions from a TIS and treat the algorithm generating them as a black box. The presence of heteroscedasticity in the predictions is demonstrated and then a metamodel approach is deployed, which augments existing predictive systems using quantile regression to place bounds on the associated error. As a case study, this approach is applied to data from a real-world TIS in Boston. This method allows bounds on the predicted arrival time to be estimated, which give a measure of the uncertainty associated with the individual predictions. This represents to the best of our knowledge the first application of methods to handle the uncertainty in bus arrival times that explicitly takes into account the inherent heteroscedasticity. The metamodel approach is agnostic to the process generating the predictions, which ensures the methodology is implementable in any system.

General information
Publication status: Published
Organisations: Department of Management Engineering, Transport DTU, Transport Modelling, Singapore-MIT Alliance, Massachusetts Institute of Technology, Northeastern University
Contributors: O'Sullivan, A., Pereira, F. C., Zhao, J., Koutsopoulos, H. N.
Number of pages: 11
Pages: 3286-3296
Publication date: 2016
Peer-reviewed: Yes

Publication information
Journal: IEEE Transactions on Intelligent Transportation Systems
Volume: 17
Issue number: 11
ISSN (Print): 1524-9050
Ratings:
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 4.44
Web of Science (2016): Impact factor 3.724
Web of Science (2016): Indexed yes
Original language: English
Electronic versions:
QRbus.pdf
DOIs:
10.1109/TITS.2016.2547184
Source: FindIt
Source-ID: 2304191671
Research output: Contribution to journal › Journal article – Annual report year: 2016 › 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.
DynaMIT2.0: architecture design and preliminary results on real-time data fusion for traffic prediction and crisis management

The ability to monitor and predict in real-time the state of the transportation network is a valuable tool for both transportation administrators and travellers. While many solutions exist for this task, they are generally much more successful in recurrent scenarios than in non-recurrent ones. Paradoxically, it is in the latter case that such tools can make the difference. Therefore, the dynamic traffic assignment and simulation based prediction system such as DynaMIT (1) demonstrates high effectiveness in the context of sudden network disturbance or demand pattern changes. This paper presents the design, development and implementation of new components and modules of DynaMIT 2.0 which is an extension of its predecessor with recent enhancements on online calibration, context mining, scenario analyser and strategy simulation capability. Also, some preliminary results are presented using Singapore expressway to show the actual benefit of the system.

Intelligent Transport Systems in the Smart City

General information
Increased urbanisation puts pressure on city infrastructures. In traffic this infers increased congestion detriment to health and to the more general liveability of the city. As a City, Copenhagen has a strong will to find new ways to management the increased urbanisation in order to provide citizens with a high quality of life. Smart city solutions hold potentials to restructure the management of the city with digital solutions rather than cumbersome and expensive physical investments. However, many of these solutions have no or little track record, and needs testing – by the civil services operating the new solutions and by the citizens living in them. Our motivation to apply to this call is a genuine interest in building new solutions to our citizens that provide them higher quality of life. The ability to try and test solutions with agility and citizens’ feedback holds a promise to shortcut and to accelerate the transition of the city to the future.

The aim of this project is therefore twofold; 1) to build capacity to swiftly conduct experiments in the settings where solutions have to be implemented, 2) to generate new knowledge about the relation of traffic management to air quality. Combined the project aims to establish proofs of value of smart city solutions in traffic to health and related public expenses and to the ability to implement new city workflows of prototyping and testing. Previous Organicity participants have already focused on the topics of air quality and traffic: AirPublic through high-resolution, real-time and low cost spatial maps of air pollution in London through small sensors in an effort to produce data that can enable informed decision making related to the problem. While the use of real-time data in both instances is intriguing we are of the opinion that such data is only needed for a limited time and not forever. We believe that the gathering of real-time data through a reasonable time period can provide enough information to still make changes possible if that information is integrated into the planning and management aspect of a city. Thus, we propose a project for improving the air pollution of a city through changes in traffic management through a limited time period of real-time data gathering. The project design monitor two stretches of roads of similar design and with consistent traffic flows. Sensors placed at intersections will monitor air quality (NOx, CO, CO2, PM10 and PM2.5). Traffic lights would then be optimized for better health possibilities by altering the traffic flow and registering differences.
Autonomous Bus Demand Modeling from Big Data
Peled, I., PhD Student, Department of Management Engineering
Pereira, F. C., Main Supervisor
Dauwels, J., Supervisor
Samfinansieret - Andet
01/01/2018 → 31/12/2020
Award relations: Autonomous Bus Demand Modeling from Big Data
Project: PhD

Multi-model bus Arrival Prediction with Intelligent Handling of Uncertainties
Petersen, N. C., PhD Student, Department of Management Engineering
Pereira, F. C., Main Supervisor
Heckscher, A., Supervisor
Ansat eksternt
01/02/2017 → 30/01/2021
Award relations: Multi-model bus Arrival Prediction with Intelligent Handling of Uncertainties
Project: PhD

Machine Learning and Mobility
Markou, I., PhD Student, Department of Management Engineering
Pereira, F. C., Main Supervisor
Larsen, A., Supervisor
Rich, J., Examiner
Jenelius, E. A., Examiner
Viti, F., Examiner
Technical University of Denmark
15/12/2015 → 14/12/2018
Award relations: Machine Learning and Mobility
Project: PhD

ATIS and modal shift: the role and the effectiveness of information provision and perception
Mehdizadeh Dastjerdi, A., PhD Student, Department of Management Engineering
Pereira, F. C., Main Supervisor
Kaplan, S., Supervisor
Mabit, S. E., Examiner
Cherchi, E., Examiner
Shiftan, Y., Examiner
Samfinansieret - Andet
15/12/2015 → 30/06/2019
Award relations: ATIS and modal shift: the role and the effectiveness of information provision and perception
Project: PhD

Intelligent Composition of Buffer Times in Railway Scheduling
Cerreto, F., PhD Student, Department of Management Engineering
Nielsen, O. A., Main Supervisor
Harrod, S., Supervisor
Pereira, F. C., Examiner
Krasemann, J. T., Examiner
Preston, J., Examiner
Technical University of Denmark
15/12/2014 → 06/09/2018
Award relations: Intelligent Composition of Buffer Times in Railway Scheduling
Project: PhD
Activities:

Projects for Mobility in the Smart City
Period: 1 Jun 2018 → 30 Jun 2018
Inon Peled (Organizer)
Francisco Camara Pereira (Organizer)
Department of Management Engineering
Transport DTU
Transport Modelling
Degree of recognition: Local

Related event

Projects for Mobility in the Smart City
01/06/2018 → 30/06/2018
Keywords: course, smart city, projects
Activity: Attending an event › Participating in or organising workshops, courses, seminars etc.