RailNorrköping 2019 Short Course

Big data in railway operations

Offered at the 8th International Conference on Railway Operations Modelling and Analysis, RailNorrköping2019

Date: Thursday June 20th, 2019, 9.00—16.00
Place: Building *Kåkenhus*, 2nd floor.

PROGRAM

9.00–10.00 **Lecture 1: Introduction to big data in traffic planning**
Clas Rydergren, Linköping University, Sweden

10.00–10.20 *Refreshments*

10.20–11.50 **Lecture 2: Mining, analysis and modelling of passenger flow data, time and route choice in railway networks**
Otto Anker Nielsen, Technical University of Denmark, Denmark

11.50–13.10 *Lunch at Louis de Geer*

13.10–14.40 **Lecture 3: Analysis of train traffic and track capacity use to improve the quality of operations**
Giorgio Medeossi, TRENOlab, Italy

14.40–15.00 *Refreshments*

15.00–16.00 **Lecture 4: Machine Learning: theory and application**
Pavle Kecman, Allianz Data Office, The Netherlands

Lecture 1: Introduction to big data in traffic planning

Large amounts of data on individual traveller's choices from diverse sources provide great opportunities for traffic demand modelling. However, working with the data requires knowledge of how to filter and massage the data and how to gain insights into the quality of the data. This lecture will give an overview of data sources that are available today for use in travel demand modelling and traffic planning and give some examples of possible use.

We will discuss a range of data sources, from sources that have been installed mainly for collecting travel related data, like radars sensors for measuring speeds and flows of vehicles on highways, to data that may provide insights of travel patterns but comes from sensors that is mainly installed for another purposes, like data generated in tele-communications networks for enabling mobile phone services. Examples from structuring, processing and analysis of radar measurements, smart card data, GPS data and mobile phone operator data are given, and insights from analyses where the data is combined with digital map data and time table information are shown. The examples cover both applications where insights are generated directly from the data and applications where the data is used as input to traffic operations and traffic planning models.

Associate Professor Clas Rydergren, Linköping University, Sweden

Clas Rydergren has a background in Mathematics from Linköping University and received his PhD in Optimization from the same university in 2001. He is the head of the Transport Systems group at the division of Transport and Communications Systems and Linköping University and is the coordinator of the master program Communications, Transport and Infrastructure (KTS).

Clas is involved in several national and international research projects in the area of traffic demand modelling and traffic network assignment. In many of the projects, analysis of GPS data, ticket data and data from mobile phone operators, are used for travel demand analysis and traffic demand modelling.

Lecture 2: Mining, analysis and modelling of passenger flow data, time and route choice in railway networks

When looking at railway planning, a discrepancy exists between planners who focus on the train operations and publish fixed railway schedules, and passengers who look not only at the schedules but also at the entirety of their trip, from access to waiting to on-board travel and egress. Traditional surveys on passenger's travel patterns from door to door – e.g. postcard analysis, stop interviews and other types of interview surveys – have been very expensive. Knowledge on passenger flows, route choices and preferences have therefore limited to some extent.

The lecture gives an overview of recent research in large-scale data sources, including smart card data, smartphone-based surveys, internet-based surveys, automatic counting systems and GPS-based tracing of vehicles and travelers, and how they can be merged and combined to obtain a better knowledge on passengers choices and preferences. Different applications of this is presented related to automatic OD-estimation, estimation of route choice models and analyses of arrival time distributions to stations. At the end of the lecture, examples are presented on how this knowledge can be used for a passenger oriented planning and optimization of public transport time-tables.

Professor Otto Anker Nielsen, Technical University of Denmark, Denmark

Otto Anker Nielsen is leading the IPTOP project on Integrated Public Transport Planning and Optimisation. He is professor in transport modelling at the Technical University of Denmark (DTU), Dept. of Management, and head of the Transport Division. He has 27 years of research experience and applied work experience within the field of transport modelling and transport behaviour research. He has been leading several large-scale transports modelling projects in Denmark and at EU level over the period from 1994 until now. He has been/is supervisor for 27 PhD-students, 10 visiting PhD-students and more than 100 MSc thesis projects and about 70 BSc-thesis.

Lecture 3: Analysis of train traffic and track capacity use to improve the quality of operations

Over two centuries after their invention railways are still the backbone of mobility in the largest and most congested urban areas. However, facing the increasing transport demand appears more and more challenging given the rigidity of the system: maximising the utilisation and quality of service on the existing networks is crucial for all operators and transport authorities.

Understanding the current operations in detail is a key prerequisite in this process, since it allows right-dimensioning the timetable and the investments.

The most useful data source for a network-wide analysis are surely the train describer/ track circuit data, since they are available for the entire network and long time periods, and at the same time are easily manageable.

The presentation will show methods for verifying the quality of datasets, identifying the critical elements on the network and suggesting improvements, as well as using the current process times as an input for testing the performances of investments and future timetables.

Dr. Giorgio Medeossi, TRENOlab, Italy

Giorgio Medeossi studied Transportations Systems Engineering and received a PhD at the University of Trieste. He was involved in several projects in the field of rail operations analysis, modelling and simulation. With the paper "A method for using stochastic blocking times to improve timetable planning" he was awarded the "Young Researcher Award" of the International Association of Railway Operations Research (IAROR) in 2011. In 2015 he founded TRENOlab, a software development and consulting company that supports the railway undertakings in improving operations. In four years TRENOlab has been involved in railway operations studies including some of the busiest lines in France, UK, Norway, Italy and other Countries in Europe and America.

Lecture 4: Machine Learning: theory and application

A recent article in the Forbes magazine states: "There will be no data science job listings in about 10 years, and here is why. There are no MBA jobs in 2019, just like there are no computer science jobs". Data Science (DS), and Machine learning (ML) as its essence, is therefore about to become a part of standard toolkit expected from any professional in their domain. Railway operations generate a lot of data that can be used for understanding complex phenomena and improving the current processes by means of ML. This lecture has the purpose to introduce and enable academics and professionals to be able to effectively use ML in their work. In the first part we will introduce the main theoretical concepts and decompose a typical machine learning project into a sequence of steps and procedures from problem understanding to model deployment. A life cycle of an ML project will then be demonstrated on an example from railway operations. After the course, interested participants will have an opportunity to work on a ML use case using their preferred analytic tool.

Dr. Pavle Kecman, Allianz Data Office, The Netherlands

Pavle developed interest in Railway operations and engineering as a student at the University of Belgrade. His Ph.D. and post-doctoral research, conducted at Delft University of Technology and Linköping University, respectively, was focused on developing predictive and prescriptive machine learning models for railway traffic prediction and optimisation. In 2016 he moved from academia to industry, working first as a cross-industry data science consultant at IBM and currently as a Lead Data Scientist at Allianz focused on working with telematics data and natural language processing.