Paper number EU-TP0068

Relating different types of (traffic) data in the NDW Traffic Observatory at DiTTlab, Delft University of Technology

Edoardo Felici[1*], Hans van Lint[2]

1. National Data Warehouse for Traffic Information (NDW), The Netherlands.

edoardo.felici@ndw.nu

2. Delft University of Technology

Abstract

The National Data Warehouse in the Netherlands has teamed up with the DiTTlab research group at the Delft University of Technology to relate different types of (traffic) data together. Combining these data types (real-time traffic-, incident-, roadworks-, event- and weather data), each containing different characteristics, geographical referencing and timeframes, will lead the team to novel insights and analysis possibilities of the traffic state, further explored in a multi-year project proposal that will allow the multi-scale analysis of traffic with all relevant data, from road section level to the national scale. This will lead to innovative applications to predict the traffic state that can be used in all future ITS developments. The first steps have produced promising results: an application has been realized, making it possible to relate traffic data with related road works, incident and the weather conditions. This gives novel insights and analysis possibilities for traffic engineers and policy makers.

Keywords: Big data, traffic data analysis, multi-scale

Introduction

The National Data Warehouse for Traffic Information

The Dutch National Data Warehouse for Traffic Information (NDW) is a public partnership that brings together traffic data from various sources, in one database, and delivers these data to road authorities and service providers. Founded in 2009 by 15 road authorities, this amount of has grown to 19 partners including the national road authority Rijkswaterstaat, all provinces, the four biggest municipalities and a number of city regions. NDW has taken the role of a shared service organization for the procurement of traffic data on behalf of these partners.

The initial procurement of traffic data by NDW in 2009 provided speed measurements, travel times and vehicle counts with vehicle categorization measured on approximately 24.000 locations throughout the country on the main road network and a number of provincial roads. The network of measurements has expanded over the years to include higher coverage on provincial roads and main urban roads, delivering data on 7.100 kilometers of road as of November 2015. Besides speed measurements, travel times and vehicle counts, information on the status of infrastructure is also delivered through NDW. Data on the opening of bridges, roadworks, rush-hour lanes and incidents is gathered within NDW and delivered to stakeholders on occurrence. (1)

DiTTlab: from struggling through data towards searching in information

NDW produces a large amount of data: since starting the collection of data in 2009, NDW has produced about 6 terabytes of compressed XML data. This data is stored in an historical database that allows the export of specific data and locations and offers a number of statistical tools to analyze the behavior and development of traffic. Yet this offers too little flexibility to detect patterns and relationships between different data sources. Can we relate the weather conditions on a particular day to the traffic situation that was observed? What influence do road works have on travel times on the surrounding road networks? What can we learn from the occurrence of similar traffic patterns on different road networks and conditions? These questions can only be answered by creating relations between different data types and sources and organizing this in a way that the data produces useful information.

In 2015 NDW has teamed up with prof. dr. ir. Hans van Lint of the Delft University of Technology to develop this within the Delft integrated Travel and Transport Laboratory (DiTTlab). Within DiTTlab, a university research team collaborates with companies and governments to develop an open source, multi-scale and integrated environment for the assimilation of traffic data from all possible sources. For NDW, this has resulted in the creation of the NDW Traffic Observatory, wherein the big data foundations for the full exploitation of NDW data can be laid down (2).

Relating real-time traffic data to events and weather conditions

Research questions

The first challenge is to relate different types of data that are collected by NDW. Can we systematically and technically relate these different data sources and make the results easily accessible? Once this is done, which products and services can be developed to increase the functional utility of the data and can a database be made ‘intelligent’ to search through the data?

Relating different data types

During 2015 the team at DiTTlab worked on the first step: technically relating the data types to each other. This has resulted in the development of a webtool wherein a geographical route selection can be made, and the resulting traffic data including the related events (related road works, events and incidents) is generated. A time-space diagram is then available to analyse the traffic situation on the chosen route and day, including information about weather conditions. An export tool allows the user to download all the relevant data.

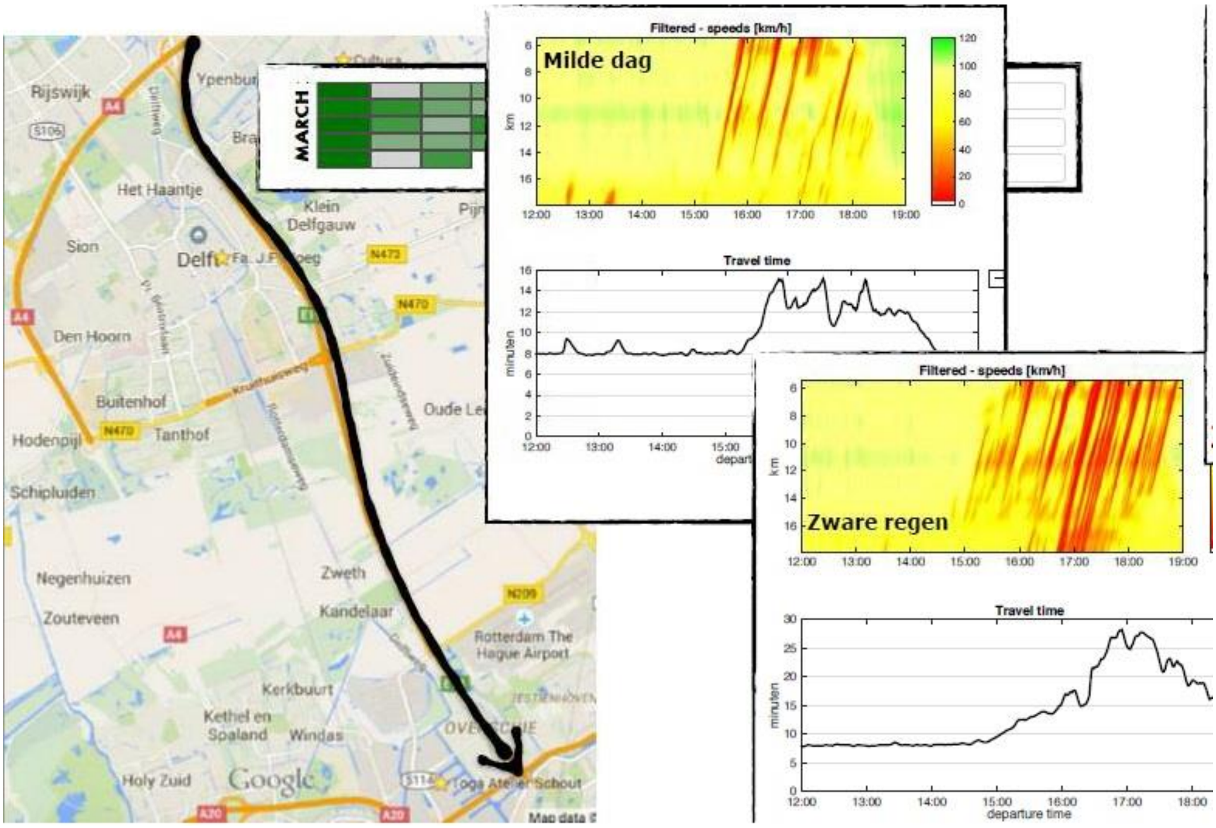

Figure 1 – Choosing a route to create time-space diagrams with all relevant traffic and event data

Figure 1 shows the traffic conditions in a time-space diagram on a day with 'mild' weather (top), compared to a day with heavy rainfall (bottom). The time-space diagram shows the time of day on the horizontal axis and the distance along a road on the vertical axis. This analysis can now be done on-the-fly through the DiTTlab application. This is achieved through the creation of a cross-section of the selected route. All relevant events along this cross-section, including the weather conditions on the chosen date and time period, are then extracted from the NDW-data and visualized as input for the analysis.

This has resulted in a database with 12.447 cross sections, created from 66.589 measuring points (which amount to 5.1M if changes on the network over time are considered). On these cross-sections, 2.382.489.886 lane and vehicle class records with measurement data can be dynamically related and visualized in the DiTTlab application.

Automatically recognizing congestion patterns

Time-space diagrams show a variety of possible congestion patterns that can relate to specific traffic situations. Incidents, shockwave traffic jams or roadworks each result in a specific and recognizable pattern in a time-space diagram.

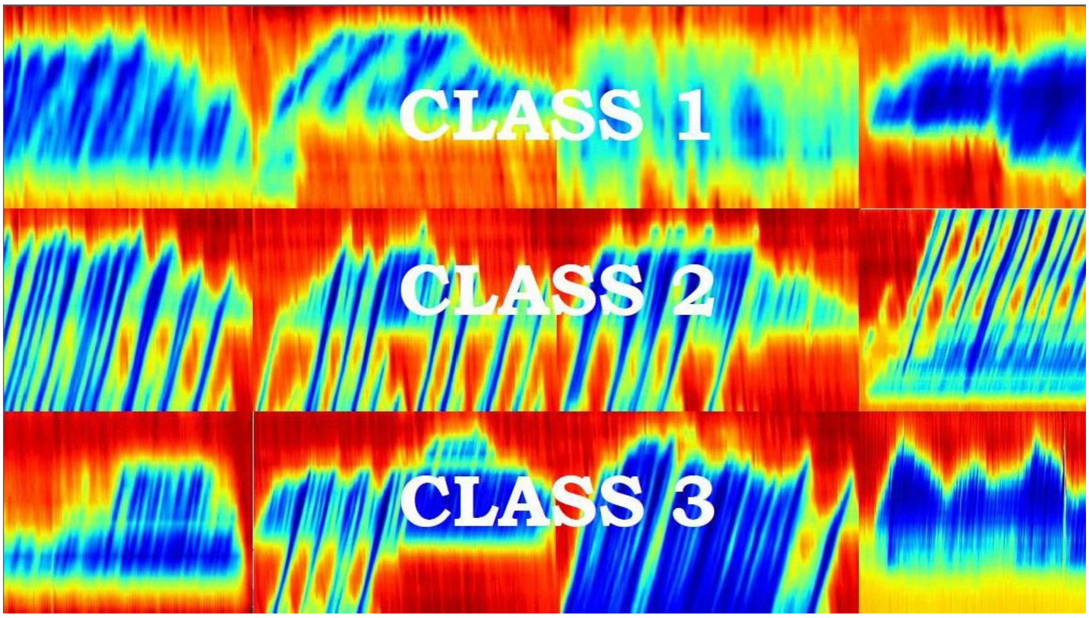

Figure 2 – Classifications of time-space diagrams

The next step in the NDW-project at DiTTlab is to create an application that is able to recognize these patterns on the basis of a colormap. Color keypoints are automatically extracted from the time-space diagrams to create a description of specific features. This allows to summarize traffic patterns into a histogram of vectorised features using the SURF and K-means algorithms.

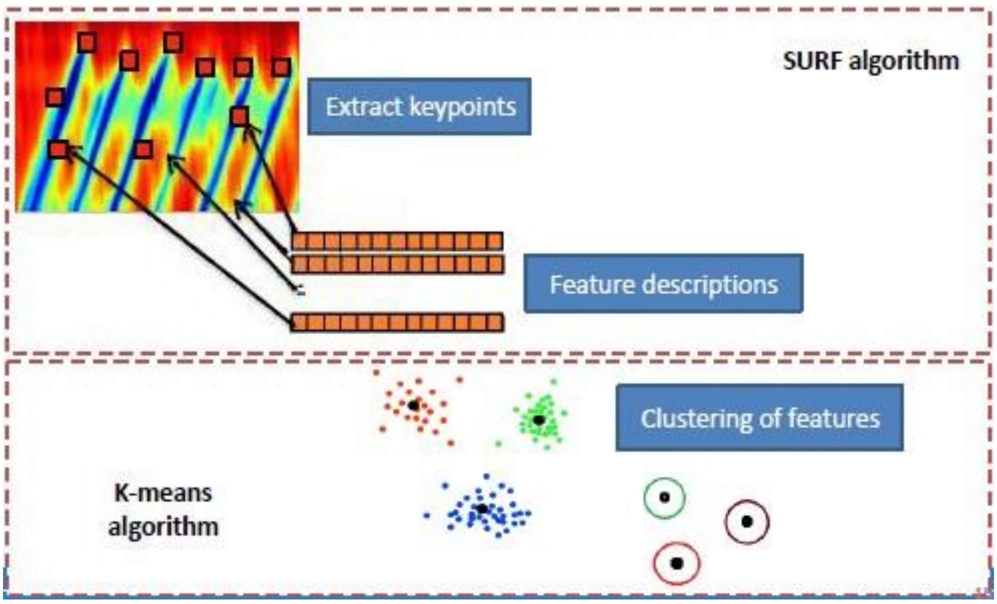

Figure 3 – Extracting features from time-space diagram

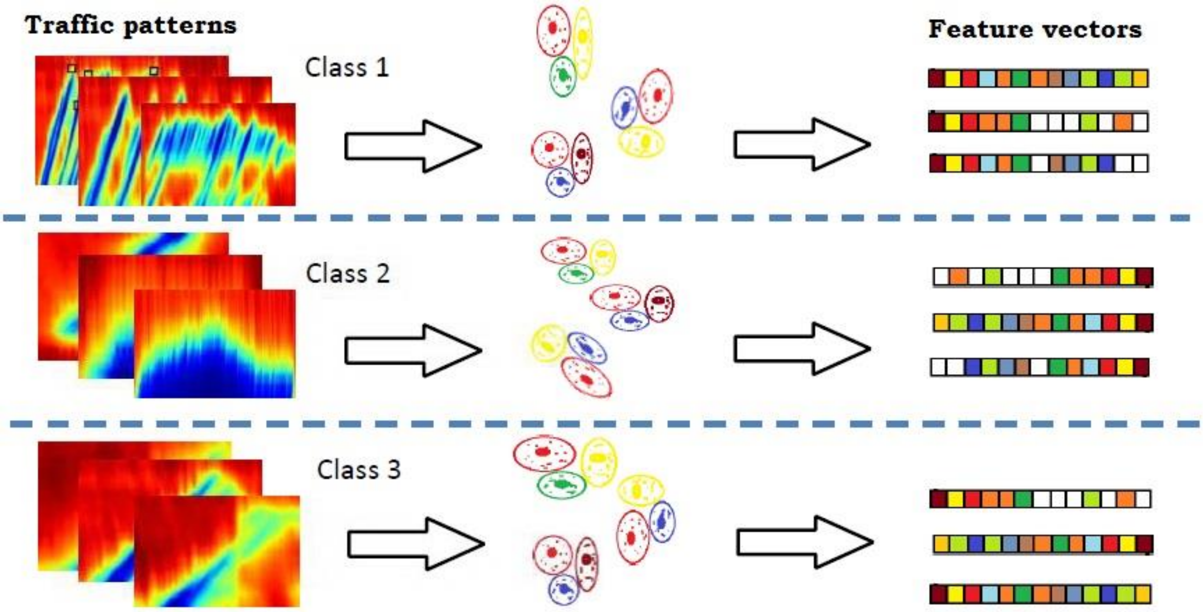

Figure 4 – Creating classifications of feature vectors from traffic patterns

The feature vectors created from the different classes of traffic patterns (see figure 4) can then be processed in an support vector machine-classifier (SVM) to train a model in automatically recognizing these patterns, thus creating a self-learning application.

Creating a congestion search engine

This process allows a user to use search queries based on congestion patterns, and find all relevant similar situations where a chosen traffic pattern occurred. This creates the possibility to 'reverse-search' the database, instead of looking at specific congestion patterns on a selected time, date and route. This potential increases the usability of NDW-data and makes innovative analyses possible. This CoZi (**Co**ngestie-**Z**oek-Eng**i**ne) will be developed in the first half of 2016, making it possible to present results this late spring (3).

This completes the first research project undertaken by DiTTlab for NDW, which sets the standard for new data storage, relation and analysis possibilities for the users of NDW data.

Multi-scale traffic state estimation

Potential

Starting in 2016, DiTTlab will be elaborating the next phase of the NDW-project. Using the NDW-data and combining this with many other potential data sources, the goal is to allow:

- Estimation of unknown traffic parameters, like traffic volumes, vehicle-loss hours, inflows, turn-rates, capacities, critical speeds and fundamental network diagrams;
- Estimation of origin-destination matrices on different spatial levels, from road lane to the

macroscopic regional level;
- Simulation and prediction of traffic on the basis of all available data for the different spatial levels.

This will create the potential for new applications for traffic simulation, prediction and a large increase in data relations leading to novel insights and analyses.

Multi-scale spatial levels

Each spatial level requires a specific and relevant input for an optimal analysis. Clearly, the required input differs for each spatial level. The table below illustrates the various spatial levels and their corresponding data input.

Table 1 – Spatial levels and corresponding data input

State estimation on road section (dynamics between 1 and 10 seconds)	
Relevant phenomena	*Typical data and models for estimation, and ITS applications in which these are used*
Dynamics of individual driving and lane-specific traffic streams (densities, speeds), dynamics of lane distributions and queuing per lane including at on/off ramps	• Dynamics: in the order of 1-10 seconds. • Data: lane-specific individual (and/or aggregated) speeds, headways (distributions), sampled vehicle trajectories , inflows, (destination-specific?) turnfractions • Models: lane- and userclass-specific traffic flow models (discrete or continuum). • Typical ITS applications: in-car control & information, intersection control & ramp metering, managed lanes
State estimation on road corridor level (dynamics between 10 seconds and few minutes)	
Relevant phenomena	*Typical data and models for estimation, and ITS applications in which these are used*
Dynamics of carriageway traffic streams (i.e. densities, speeds), dynamics of inflows, turnfractions	• Dynamics: 10 seconds to several minutes. • Data: carriageway flows, speeds, sampled trajectories, (destination-specific?) inflows, turnfractions , GPS/GSM traces • Models: continuum traffic flow models, data driven tools • Typical ITS applications: in-car information, coordinated corridor control & information

	provision
State estimation on regional network level (dynamics between 5 minutes and 30 minutes)	
Relevant phenomena	*Typical data and models for estimation, and ITS applications in which these are used*
Dynamics of traffic streams, spillback dynamics (queues, densities, speeds), route choice patterns and origin-destination (OD) flow dynamics	• Dynamics are in the order of 5-30 minutes. • Data: carriageway flows, speeds, (destination-specific?) inflows, turnfractions, GPS/GSM traces • Models: continuum traffic flow models, route & departure time choice models, data-driven tools • Typical ITS applications: network management, route guidance, information provision
State estimation on national network level (dynamics between 15 minutes and 60 minutes)	
Relevant phenomena	*Typical data and models for estimation, and ITS applications in which these are used*
Aggregated OD flow & route choice dynamics, dynamics of travel times and costs along corridors, and of aggregate production and accumulation (level of service) within major metropolitan areas	• Dynamics: in the order of 30 minutes to hours. • Data: spatiotemporally aggregated flows, speeds, travel times, GSM/GPS trips ends, survey data. • Models: reservoir models, continuum models, route, departure time, mode choice models • Typical ITS applications: multi-modal network management, route guidance, traffic information

Figure 5 – from traffic lane to higher network level

The multi-scale approach will enable an accurate and reliable estimations of the prevailing traffic state that will help efficiency and effectivity of all applications, from personal traffic apps to traffic control centers. This project will result in a framework that will be based on a fusion engine to incorporate all available data sources (based on the aforementioned initial NDW-project), and turn these into one coherent view of the traffic state. DiTTlab plans to work from an open source perspective, to create a framework available for all to build upon further. This approach makes the project unique in the world, allowing the results to be used in many existing and new applications within the ITS community and transportation in general (4).

References

1. The National Data Warehouse for Traffic Information, the Netherlands (2009). See www.ndw.nu for more information.
2. Lint, J.W.C. van (2015). *Onderzoeksvoorstel NDW Traffic Observatory, Diagnosticeren, Schatten en Voorspellen van Verkeer*.
3. Lint, J.W.C. van (2015). *Presentation of preliminary results at NDW-DiTTlab symposium, 17th November 2015*
4. Lint, J.W.C. van (2015). MiRRORS p*roject proposal for Open Technology Programme*