Influences of Road Works on the Traffic Conditions in the Amsterdam Area

Transport Domain
Primary Topic: DPV, Secondary Topic: SEMI
Course: 2018-1A – Group: 30 – Submission Date: December 11, 2018

Duncan Jansen
University of Twente
d.j.jansen@student.utwente.nl

Jacco van Ekris
University of Twente
c.j.vanekris@student.utwente.nl

ABSTRACT

Visualizations of the traffic conditions in the Amsterdam area were analyzed to asses the impact of road works. The goal was to obtain correlations between roadworks and traffic conditions, which could be used for future road work scheduling.

The data for these visualizations were provided by the Nationale Databank Wegverkeergegevens (NDW). The available data for speed, traffic flow and road works was pre-processed in R for visualizations in Tableau.

Numerous visualizations were made that contributed to the assessment of the impact; location of measurement points, location of road works, geographical heat maps of average traffic flow/speed and geographical heat maps of the significant deviation of traffic flow/speed.

Summarize Discussion
Summarize Conclusion

KEYWORDS

Visualizations, Traffic flow, Traffic speed, Road works, Amsterdam Area.

1 INTRODUCTION

Where and when does one schedule road works? It might seem trivial, but there are some real advantages of efficient road work scheduling. Efficient road work scheduling will limit the nuisance of traffic jams for commuters and costs for transportation companies. But an even more important reason is that it will limit the vehicle emissions. Vehicle emissions cause air pollution which has serious consequences for the climate [**?**] and health risks for the population [**?**].

One method to increase the efficiency of road work scheduling is to learn from past situations. If road works could be correlated to traffic conditions, one could derive knowledge from these correlations and apply this during future road work scheduling. Per example, if it is known how road works impact traffic conditions on a certain road, it can be determined whether the bottle neck of the road will be reached (resulting in a traffic jam). In such a case it might be better to schedule roadworks at a different time or even a different day in the week.

The aim of this project is to derive such correlations by analyzing visualizations of the traffic conditions in the neighborhood of road works. The scope of this project is limited to the Amsterdam area and makes use of the data provided by NDW on speed, flow and road works.

This report continues with the method, which analyses the data, explains the pre-processing- and visualization steps; Results consisting of visualizations and cases; Discussion of the results; Conclusion, reflecting back on the aim of the project.

2 APPROACH

As previously mentioned visualizations are analyzed to derive correlations between roadworks and traffic conditions. The following visualizations are made to help derive such correlations:

- **Visualization of measurement point locations and roadwork locations per hour.** Location of measurement points will help with the following two visualization. The visualization of roadwork locations is more important, because it indicates where and at which time roadworks occur.
- **Visualizations of the average speed and flow at each measurement point per hour.** These visualizations indicate the traffic conditions per hour for the various measurement points.
- **Visualizations of the significant deviation of the speed and flow at each measurement point per hour.** These visualizations indicate whether there is a significant deviation from the "normal" traffic conditions per hour for the various measurement points.

How to derive correlations from these visualizations? The idea is to look at when and where a roadwork occurs or whether it is already active. Then determine whether this roadwork results in a significant deviation for some measurement points. If this is the case, the traffic conditions can be obtained from the second visualization. The chapter method will elaborate on the process of obtaining these visualizations.

3 METHOD

This section describes what data is available, how the data was pre-processed and how the visualizations were made.

3.1 NDW data

The NDW data consists of data sets for flow and speed in CSV format where each part is divided in one meta data file and multiple files containing the measurement data. For each measurement location the meta data file provides the coordinates and other information such as road number or lane number.

Furthermore, the NDW data contains a data set with Status data which provides all kinds of information that may influence the flow and speed data like the opening of bridges, cars that broke down

and road maintenance. For each situation that occurred, a separate XML file is given.

Exact information of the original NDW data set can be found in the NDW Manual [**?**].

3.2 Pre-processing the data

In order to derive correlations, Flow, Speed and Roadworks data is obtained from the NDW data set. To make the data sets suitable for visualization in Tableau, the database structure in Figure **??** on page ?? is used. The construction of the database in Figure ?? on page ?? is explained in the sections below.

3.2.1 Creation of metaFlow and metaSpeed tables.

These data sets contain the unique information for each measurement point. It is limited to measurement points which are anyVehicle only.

- Load CSV file containing the meta data and select the columns measurementSiteReference, specificVehicleCharacteristics, latitude and longitude.
- Remove all rows that are not anyVehicle using the specificVehicleCharacteristics column and remove the specificVehicleCharacteristics column.
- Arrange, group by measurementSiteReference, make distinct and ungroup to obtain all the unique measurement points.
- Give each measurement point an ID number, respectively flowID and speedID.

3.2.2 Creation of metaRoad table.

This data set contains the unique information of roadworks and is limited to the Amsterdam area.

- From each XML file, extract the overallStartTime, overallEndTime, latitude, longitude, ProbabilityOfOccurance, OperatorActionStatus, SourceName and carriageway and save as one CSV file called "road_data.csv".
- Remove all roadworks which are not within the Amsterdam area using maximum and minimum latitudes and longitudes.
- Remove all information except latitude and longitude.
- Give each roadwork an ID number (roadID).

3.2.3 Creation of dataRoad table.

This data set contains the data of the roadworks and is limited to the Amsterdam area.

- Load the "metaRoad.csv" and "road_data.csv" files.
- From the "road_data.csv" file, remove all roadworks which are not within the Amsterdam area using maximum and minimum latitudes and longitudes
- Make sure that each row of the data is unique and combine the data with the meta data ("metaRoad.csv") using a full join by carrigeway, latitude and longitude.
- Arrange the data by overallStartTime and overallEndTime.

3.2.4 Creation of dataFlow and dataSpeed tables.

These data sets contain the flow/speed data. It is limited to measurement points which are anyVehicle only.

- Load the "metaFlow.csv" or "metaSpeed.csv" and the original meta data file of flow or speed.
- Get all indexes where specificVehicleCharacteristics equals anyVehicle.
- Create an empty list and do for each measurement data file the following:
  - Load the data file and select the columns measurementSiteReference, periodStart, periodEnd, numberOfInputValuesused, numberOfIncompleteInputs, dataError and avgVehicleFlow/avgVehicleSpeed.
  - Remove all rows that are not anyVehicle using the saved indexes.
  - Remove all the rows that contain an error based on the columns numberOfInputValuesused, numberOfIncompleteInputs and dataError.
  - Remove the columns numberOfIncompleteInputs and dataError and arrange the data by periodStart.
  - Combine the data with the meta data ("metaFlow.csv" / "metaSpeed.csv") using a full join by measurementSiteReference.
  - Replace the columns periodStart and periodEnd by one date with the corresponding hour (this can be done because the time between periodStart and periodEnd is always 1 minute).
  - Calculate for each measurement point, per hour, the mean of the flow and the corresponding standard deviation for all the lanes combined OR calculate for each measurement point, per hour, the weighted harmonic mean of the speed and the corresponding standard deviation for all the lanes combined.
  - Add the resulting table to the list.
- Combine all the items in the list to one large table and remove overlapping data that may occur in the large table.
- Add day of the week column based on date column.
- Add a difference column that is based on the average for each measurement point (flowID/speedID), day of the week and hour of the day minus the current avg_speed/avg_flow.
- Add a significant difference column based on the difference column plus the standard deviation.
- Save the resulting table to a CSV file called "dataFlow.csv" or "dataSpeed.csv".

3.2.5 Creation of Date table and enhancements to dataFlow, dataSpeed and dataRoad.

The Date table contains all the available dates which have information of the flow, speed and roadworks. The dataFlow, dataSpeed and dataRoad tables have to be modified to replace the dates by a dataID.

- Load the "dataRoad.csv", "dataFlow.csv" and "dataSpeed.csv" files.
- Get the dates + hour column of dataFlow and dataSpeed and create a Date table with the columns date, hour and dataType (flow/speed).
- Remove all rows in dataRoad that are not within the range of the Date table.
- Replace the periodStart and periodEnd columns in dataRoad with a date and a hour column by generating all the dates and hours between the date range periodStart and periodEnd.
- Add the dataRoad dates and hours to the Date table, arrange by date and hour and add a dateID column based on the row number.

- Do an inner join with dataFlow, dataSpeed and dataRoad and the Date table such that the dates are replaced by an dateID.
- Save the resulting dataFlow, dataSpeed and dataRoad tables by overwriting the original "dataFlow.csv", "dataSpeed.csv" and "dataRoad.csv" files.
- Save the Date table to a CSV file called "Date.csv".

3.3 Visualization NDW data

First of all, the data was imported into tableau and the connections between data sets were created. The data sets are linked based on the ID's and make use of an outer join (see Figure **??** on page ??). In the following sections creation of the multiple visualizations is explained.

3.3.1 Visualization of measurement point locations and road-work locations per hour.

- Column: startLocatieForDisplayLong.
- Row: startLocatieForDisplayLat.
- Pages: hour.

3.3.2 Visualizations of the average speed and flow of traffic at each measurement point per hour.

- Column: startLocatieForDisplayLong.
- Row: startLocatieForDisplayLat.
- Pages: hour.
- Mark: avg_speed/avg_flow.

3.3.3 Visualizations of the significant deviation of the speedand flow at each measurement point per hour.

- Column: startLocatieForDisplayLong.
- Row: startLocatieForDisplayLat.
- Pages: hour.
- Mark: sig_dif_speedID_dayWeek_hourDay/sig_dif_flowID_dayWeek_hourDay.

4 RESULTS

Unfortunately, it is not possible to show all of the visualizations in this report. Therefore, it was chosen to present some examples of the visualizations and a case will be presented with more in depth analysis.

4.1 Example available results

In Figure 1 on the next page the visualization of measurement point locations and road-work locations for a random hour is shown. In sub-figure 1a the location of roadworks for a single day are shown. The roadworks can change during time and in tableau the visualization is interactive. It is possible to scroll through time and see roadworks appear and disappear. This also applies to the visualizations in sub-figure 1b and 1c.

In Appendix A on page 5 visualizations of the average speed and flow at each measurement point at a certain hour are shown. In sub-figure 2a the average speed for every measurement location is colour indicated; ranging from 50 km per hour (red) to 120 km per hour (light grey). Similar for the average flow in sub-figure 2b; with the flow ranging from 0 (light grey) to 2000 (dark orange). In tableau it is possible to scroll ,for both applications, through the time interactively; the average speed/flow will change accordingly.

In Appendix **??** on page ?? visualizations of the significant deviation of the speed and flow at each measurement point at a certain hour is shown. In sub-figure **??**a the significant deviation of speed for each measurement point is shown. The deviation of speed ranges from -30 (red) to 0 (light grey). Similar for the deviation of flow, which ranges from -946 (dark orange) to 1202 (dark blue) in sub-figure **??**b. In tableau it is possible to scroll, for both applications, through the time interactively; the significant deviation of speed/flow will change accordingly.

4.2 Results case

Appendix **??** on page ??

5 DISCUSSION

6 CONCLUSIONS

Figure 1: Traffic conditions in the Amsterdam area at 17:00 - 10/06/16 using averages.

Appendices

A TRAFFIC CONDITIONS USING AVERAGES

Figure 2: Traffic conditions in the Amsterdam area at 17:00 - 10/06/16 using averages.