Minor project EWI 3615TU

Dutch traffic analysis and traffic jam prediction

José Ignacio de Alvear Cárdenas	4463196
Giulio Dacome	4476867
Thijs-Gerrit Volker	4432126
Ptojr Lengkeek	4433866

TU Delft
Computer science

TU Delft
Computer Science

December 21, 2017

Dutch traffic analysis and traffic jam prediction

by

group 2

Minor project Software Design and Application 2017-2018

José Ignacio de Alvear Cárdenas	4463196
Giulio Dacome	4476867
Thijs-Gerrit Volker	4432126
Ptojr Lengkeek	4433866

Course responsible lecturer: Dr. Georgios Gousios

List of Figures

List of Tables

Contents

1. Initial organization

1.1 Objectives

For this project, three milestones have been defined with the aim of applying Big Data processing and machine learning/artificial intelligence techniques to traffic and weather data.

First of all, the team will perform an analysis on the Dutch traffic data and the weather information collected. Within that frame, three tasks have been defined:

1. **Filtering**: the data will be collected, filtered and structured in order to facilitate its manipulation and reduce the memory storage required.
2. **Statistics**: with the data collected, statistical conclusions will be drawn. Both datasets, the traffic and weather data, will be merge bearing in mind time synchronization. During this stage multiple graphs and tables will be generated.
3. **Results**: conclusions will be drawn and documented in a report with a thorough explanation of the visuals created.

Secondly, the group will apply artificial intelligence techniques (such as an Artificial Neural Network) in order to generate a model that could predict future traffic jams. The ANN will be trained with historical data and the group aims at implementing the final model in a web service that is fed with live data.

Finally, if the group counts with enough time, an optimization model will be created. Such a model would interact with a user, in this case a driver, indicating what would be the optimal speed at which he could ride in order to avoid a traffic jam. Another possibility would be pointing the user to an alternative route such that he/she could avoid an expected traffic jam.

1.2 Data

For this assignment, the team requires historical and live data of the Dutch traffic and the weather conditions on the road.

For the first one, the group retrieves the information from the Nationale Databank Wegverkeergegevens (NDW); the data is collected in an XML format.

With respect to the second information required, the weather data is collected from the Koninklijk Nederlands Meteorologisch Instituut (KNMI) in a text format.

Besides the previous data banks, the schedule of the yearly Dutch holidays will be also included since it will help in the prediction of traffic jams of the AI model.

1.3 Tools

The following software tools will be used in order to carry out the project successfully:

1. **Python**: programming language key for the processing of the data.
2. **Spark**: open-source cluster-computing framework that will be extensively used for the processing of the data. The Spark software also provides us with a set of components, such as the Spark Streaming, the MLLib and the GraphX; which will help with the data analysis and the creation of the artificial intelligent model.
3. **Jupyter notebook**: web application for the code documentation.
4. **GitHub**: development platform to ease the collaboration of the team members to the project's code.
5. **Stack**: online server for the storage of high amounts of data.

2. Traffic Analysis

In the context of the analysis of the traffic data in order to get a prediction of conditions, this chapter is to explain some critical decisions that were taken to detect the presence of a traffic jam.

The first step that was taken was to process historical data and feed them to a machine learning algorithm from the SparkML library. The input to this algorithm is data containing date and time, traffic conditions per sensor and local weather data. The output consists of two values, corresponding to traffic conditions (i.e. traffic flow and average velocity) forecasted for the desired future time interval. That data, though, does not provide any insight on the presence of a congestion, being a congestion defined as *a situation where the local vehicle flow is greater than the local bottleneck flow.* Two models were considered to attempt a successful detection of a traffic jam: one based on a **speed threshold** and another one based on the analysis of the gradient of the data.

3. Traffic Theory overview

The purpose of this chapter is to provide an overview of the traffic measurement systems in the Netherlands. Furthermore, this chapter contains a recap of the main theoretical concepts used to complete the project.

3.1 The Dutch measurement system

Traffic information, together with information regarding the measurement system installed on the network was retrieved from the NDW, the *National Data Warehouse for Traffic Information* [**?**]. The sensor array consists of approximately 60,000 sensors spread throughout the Dutch roads and highways.

3.2 Traffic Theory