15th International Road Weather Conference
Quebec City, Canada, 5-7 February 2010

ROADIDEA **Session**

Following paper/presentation deals with modeling of traffic and road weather and is a part of WP3 within the EU/FP7 project ROADIDEA where a goal is to study the adverse effects of weather on traffic and to develop new and innovative methods and tools to increase traffic fluency and safety. (Full report is available as D3.3 Data fusion - road weather at ROADIDEA homepage: *www.roadidea.eu*)

Title: **Traffic Data and Road Weather**

Jörgen Bogren and Torbjörn Gustavsson *

Klimator AB
GVC GU
Box 460
SE-405 30 Göteborg Sweden

jorgen.bogren@klimator.se

ABSTRACT

To combine weather information with other data source is especially challenging task because the data is very special. Weather data from different sources and with variation in spatial and temporal scale are analyzed together with traffic data. The overall aim is to be able to determine the actual road conditions. The task also involves work concerning quality of data and how different data sources can be combined to increase the information. A new innovation is to use the information about variations in traffic to detect effects by weather elements for example precipitation or slipperiness.

The analysis shows that weather has an obvious impact on traffic and also that it is possible to build a model with the ability to recognize the weather (with weather history), which affects traffic in a negative way. These findings can be used for future development of new information systems. This paper describes a method for modeling weathers impact on traffic, as well as the results obtained when applying that method. The analysis comprises preprocessing, a method for visualizing the effect of weather on traffic parameters (velocity and speed per time of day) and also model building via a decision tree classifier. The visualization is applied to build a dataset with classified samples; "traffic disturbed by weather" or "normal traffic".

A decision tree classifier is used to train models to recognize the combinations of weather parameters that lead to disturbed traffic. The visualization shows a distinct correlation between precipitation and changes in traffic pattern and the decision tree models have a good/useful performance.

*) Contributors to the report D3.3 are as follows: *Olle Wedin (Klimator), Torbjörn Gustavsson (Klimator), Poul Grashoff (Demis), Igor Grabec (Amanova), Ville Könönen (VTT), Matthieu Molinier (VTT), Sami Nousiainen (VTT), Pär Ekström (Semcon).*

1. Introduction

To combine weather information with other data source is especially challenging task because the data is very special. Weather data from different sources and with variation in spatial and temporal scale are analyzed together with traffic data. The present paper deals with modeling of traffic and weather where the analysis is focused on a set of data from outside Gothenburg, Sweden where the amount of traffic is available together with road weather parameters.

The overall aim is to be able to determine the actual road conditions. The task also involves work concerning quality of data and how different data sources can be combined to increase the information. A new innovation is to use the information about variations in traffic to detect effects by weather elements for example precipitation or slipperiness.

The analysis shows that weather has an obvious impact on traffic and also that it is possible to build a model with the ability to recognize the weather (with weather history), which affects traffic in a negative way.

In the study two examples of alternative classifiers, for determent of weathers impact on traffic, are presented. The output of the models produced by the classifiers is presented and can be compared with the decision tree models in the main section.

2. Method

The method in the study to model the weather impact on traffic following steps was applied:

- Get a picture of the weather- and traffic datasets by visualization.
- Select a training dataset containing weather fused with corresponding traffic.
- Find a method to tell the reaction of traffic patterns for known bad weather conditions.
- Find all similar traffic patterns and class them as traffic disturbed by weather.
- From the classified fused weather- traffic data, build a classifier that is able to tell whether or not traffic will be disturbed from a given weather.
- Test the performance of the classifier via test on "never seen" data (cross validation)

For the study data from the city of Gothenburg in Sweden is used.

The given data is taken from 3 positions during 2 years. At each of the positions traffic behaviour and weather parameters are recorded. The weather data is collected at 3 RWIS-stations (Road Weather Information System) The weather data is reported and logged every 30 minutes. The traffic data, used for the analysis, is taken from Bäckebol lane 6 and consists of approx. 400 000 samples. The weather data used is taken from a road weather station close by. The data fusion makes it possible to see during which traffic circumstances the precipitation-samples are more probable.

A classifier is used used to build the models. The task for the classifier is to detect patterns in the set of attributes, in order to be able to build a model that as well as possible classify samples correctly (order them according to class), figure 1.

In the present case the attributes are the weather parameters with history and the classes are divided into disturbed or normal traffic.

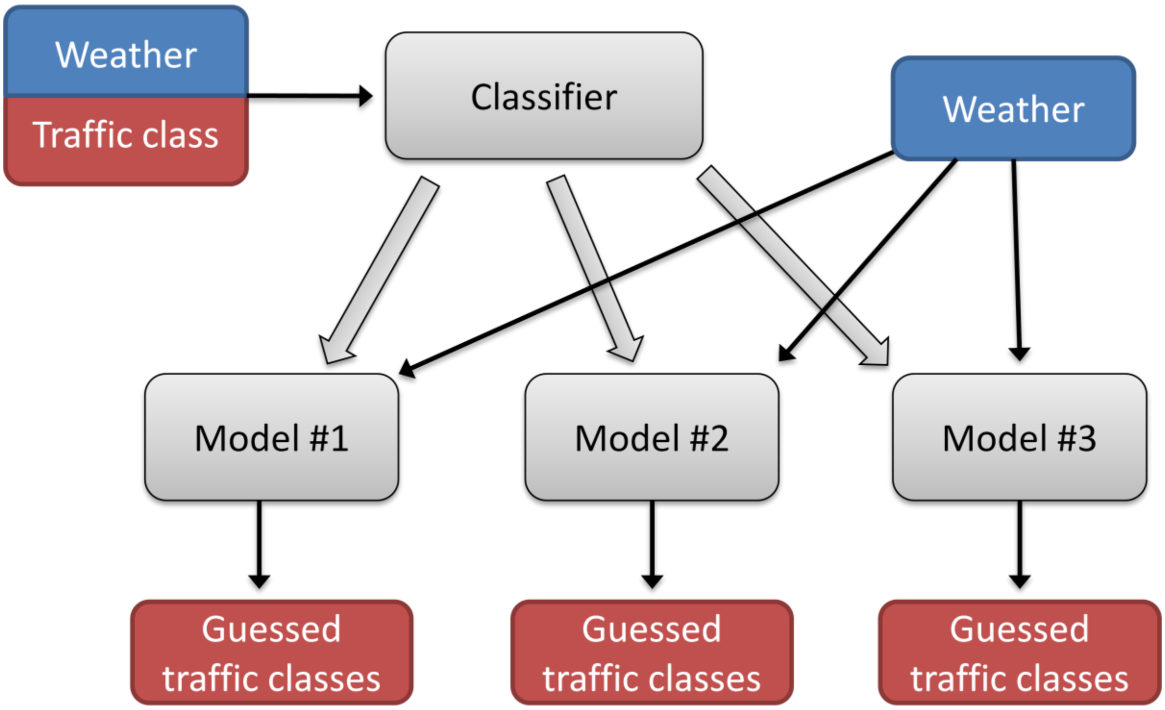

Figure 1. illustrates how samples are processed.

So in other words:

- The classifier takes samples with weather attributes and corresponding classes as input and produces models as output.
- The models take weather data as input and produce guesses of traffic classes as output (disturbed or normal).

Each of the models can be fed with the weather data used by the classifier (training data) or data it has never seen before, either for testing/validation or real use. When used for testing the correct corresponding traffic classes are known, so that the performance is apparent. In this case the data is known as "validation data" and the process is known as "cross validation".
The classifier can be of many kinds such as neural networks, Bayesian classifiers or nearest neighbor classifiers. Here a so called "decision tree classifier", will be used. The reasons for this are:

- It is straight forward to construct, use and explain.
- The models it produces are the easiest to interpret once they are built.

3. Results

The objectives were:

- To see if it can be indicated that weather affects the traffic pattern.
- If so; to be able to tell the combinations of weather parameters, which disturb road traffic in a negative way, at one static point along the road, by the use of a classifier.

To meet objective 1, methods to visualize the impact of rain on traffic patterns were described and implemented. The resulting visualizations indicate that distance between cars will increase and velocity of cars will decrease during periods where precipitation is registered. During the rush hours (6:00-9:00 and 15:00-19:00) no effect of the precipitation on traffic could be indicated.

To meet objective 2, a decision tree classifier was described and used. The model that was produced, performed way above random guessing, but was far from ideal. The significant results here are the methods presented to produce the models, not the models themselves.
One example is that via cross validation of a data set containing approx 350000 samples of which in forehand 2.3% where considered as situations where traffic had been disturbed by weather, a model managed to distinguish 12% of the "disturbed" samples, by classifying just 0.9% of the samples as disturbed in total. Of the total 0.9% that was classified as disturbed 31% actually were disturbed.

4. Discussion

The resulting performance of the model might seem low at first:
Recall=11.95% (fraction of actually disturbed samples that were found by model)
Precision=31% (fraction of classified disturbed samples that were actually disturbed)
A recall of 11.95%, means that many more samples are falsely classified than correctly. Maybe one think that 50% is the least you could ask for, but this way of putting things is deceptive. The proportions of actual classes have to be taken into account: The validation dataset contains just 2.4% disturbed samples. When the model makes guesses so that only 0.9% of the total dataset is considered disturbed, it finds 11.95% of the disturbed samples (called recall). The case with random guessing would have been 0.9%, which is way lower.
The precision with random guessing would have been 2.4% instead of 31% which the classifier reaches.

The result shows not only that it is possible to visualize the impact of precipitation on traffic pattern, but also that it is possible to build a model which is able to classify the traffic situation from weather.

The result of the visualization is probably the more distinct one. If further analysis is done, the suggestion is to keep this first part more or less intact and improve the succeeding model building. Improvements could include running the same analysis on larger amounts of data. The 2 years of data contains many samples, but not so many situations with bad weather. One could also experiment with the weather parameters and their history. Also the classifier could be changed to another kind such as support vector machine or minimum-distance classifier. Another expansion would be to look at many points along a road or a road network in order to discover the impact of weather.

The models can be used as they are, to indicate where traffic will be disturbed or they can be used together with general regression, so that two different general regression models are built; one for normal traffic and one for disturbed, analogous to the selection of weekdays or weekends. For development of future road weather information systems these findings are of great value.

5. References

Bogren Jörgen (Klimator), Wedin Olle (Klimator), Gustavsson Torbjörn (Klimator), Grashoff Poul (Demis), Grabec Igor (Amanova), Könönen Ville (VTT), Molinier Matthieu (VTT), Nousiainen Sami (VTT), Pär Ekström (Semcon): Report D3.3 from WP3 within the EU/FP7 project ROADIDEA (Full report is available as D3.3 Data fusion - road weather at ROADIDEA homepage: *www.roadidea.eu*).