Big Data for Incident management – prediction with neural networks

A project for:

Corporate Informatie Voorziening (CIV)

Contents

1

BACKGROUND

Project background

The large number of road users on the crowded Dutch roads leads to significant congestion and incidents.

On an annual basis there are more than 100.00 incidents, approximately 270 incidents occur every day.

In addition to the human toll, road incidents cause severe disruptions to the circulation on the road network, with economic and social consequences.

Incident Management (IM) is one of the key management practices to address incidents. The primary goals in IM are to ensure the safety of the roads and to restore the vehicles flow after an incident. More specifically, IM aims at:

- Increasing the responsiveness of incident handling, including securing the incident site;
- Improve the control or redirection of the flows;
- Reduce the consequences for incident victims;
- Streamline the cooperation between emergency response actors.

Increasing the visibility on risk sections of the road highway, or on circumstances that may lead to incidents, is of general interest for incident management. Increasing the ability to predict when and where incidents will happen would have a major influence on our ability of mitigating the causes of incidents and, when this fails, respond appropriately.

Given the multitude of factors that lead to incidents this may well be impossible. However, improving the ability of predicting, even marginally, could be of significant value. The increasing availability of data (incidents data, traffic, weather, other data) opens the possibility of exploring methods of big data analysis for incident prediction.

This project explores a set of these methods, namely artificial neural networks, to inform RWS on the opportunity of establishing a more structural effort on incident prediction based on Big Data and Data Science.

Incident causes

The factors contributing to road traffic accidents are commonly grouped into three categories:

- causes attributed to the environment (traffic, weather, visibility road conditions)
- causes attributed to the vehicle (vehicle failure, tire explosion)
- causes attributed to the driver (sleepiness, distractions, errors)

Analysis shows that in about 30% of cases the causes could be attributed to the environment, slippery roads, bad visibility etc.

Only 10% of contributing factors are attributed to technical issues related to the vehicles involved; tire explosions or poor maintenance for example.

In 90% of cases the contributing factor is human error.

A significant proportion of accidents are caused by a combination of the three categories. For example slow driver reactions during adverse weather conditions (speeding when visibility is low).

Predicting incidents implies grasping the implication of each of these factors at a specific time and place in the road network, in near real-time.

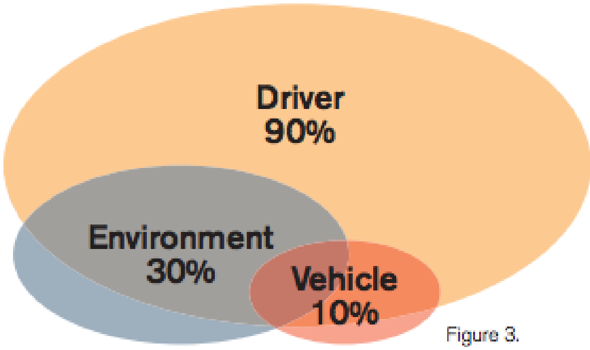


Figure 3.

- Distraction
- Speed
- Risk awareness
- Limited visibility
- Road design
- Weather
- Blind spots
- Tire explosion
- Technical error

Incident prediction and Big Data

In literature, there are several attempts to predict road incidents. Literature focuses on one or more aspects of an incident such as:

- Location (where will it happen?)
- Time (when will it happen?)
- Consequences (will it cause material damage or casualties?)
- Size (will it involve one or two vehicles or more?)
- Duration (will circulation be restored quickly or will it require significant effort ?)

Visibility on the above is at the basis of, for instance:

- Taking pro-active measures for traffic control
- Re-allocating resources to where and when they are most likely needed, from monitoring to response
- Providing warning to travelers

The long history of incident recordings and analysis in most developed countries it at the basis of rules and policies that aims at reducing the chance of incidents, and the continuous decrease in number and consequences is testament to the effectiveness of these measures.

However, the growing availability of data on on any aspect of traffic and of the factors that influence incidents, as well as methods for analyzing massive amounts of data, may underline a growing capability for predictions, something worth exploring.

Data sources for incident prediction

Example causes	Data
Driver: distraction	• Sensors detects driver behavior • Proxies, such as texting or calling while driving • Erratic speed/tracks based on navigation data
Driver: speed	• Speed measures • Transition times between checkpoints/loops • Navigation/telecom data
Driver: Alcohol	• Vehicles sensors installed in vehicle • GPS paths combined with context data to infer alcohol consumption
Environment: visibility	• Local fog detectors • Sunlight orientation and angle
Environment: road design/conditions	• Bind spots • Restrictions or road works • Steepness
Environment: weather	• Rain detections (local) • Icing conditions • Wind • Thunderstorms
Vehicle: tire condition	• Status and condition sensors • Maintenance records
Vehicle: general vehicle health	• Maintenance records • General vehicle sensing data • History of incidents

These variables and data points can be grouped into two main classes:

- Context monitoring factors, that can be measured for all vehicles and roads (traffic volumes, weather). These are generally available.
- Personal and vehicle data factors, which are captured by the car or under the supervision of the driver. They are usually not accessible without permission.

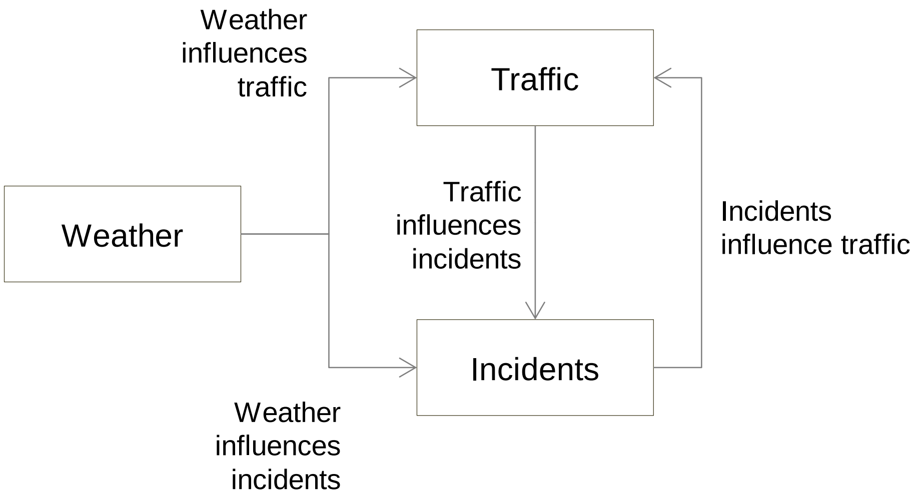
Incident prediction would ideally require access to all these data, but for feasibility reasons is usually restricted to context monitoring factors. This will be the case in this study.

THE TEST SCENARIO: TRAFFIC AND WEATHER DATA

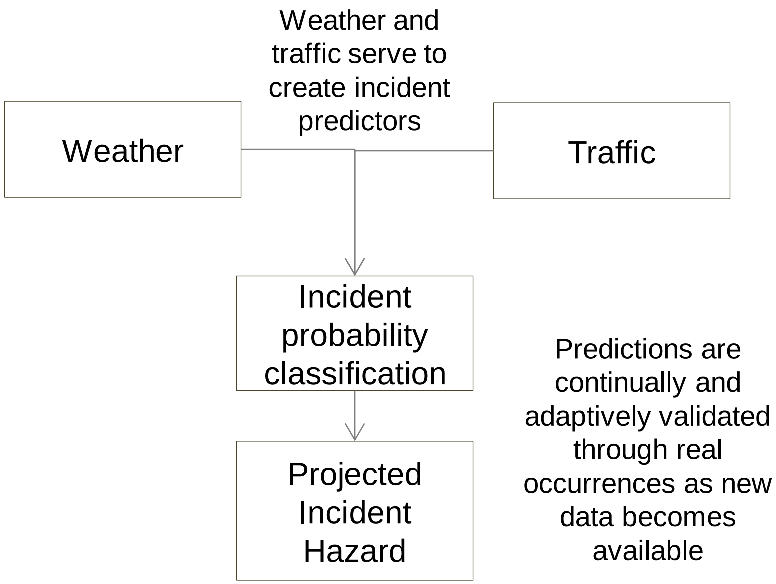
Predictions range from individual incident prediction to aggregated estimates (probability that at least one incident will occur in a certain area during a certain period of time). The model that we use looks at predicting the occurrence/non-occurrence of an incident in a specific road segment on a specified time interval in the future (e.g. next 2 hours).

In the test model we focus only on traffic and weather data.

The simplified cause effect chain



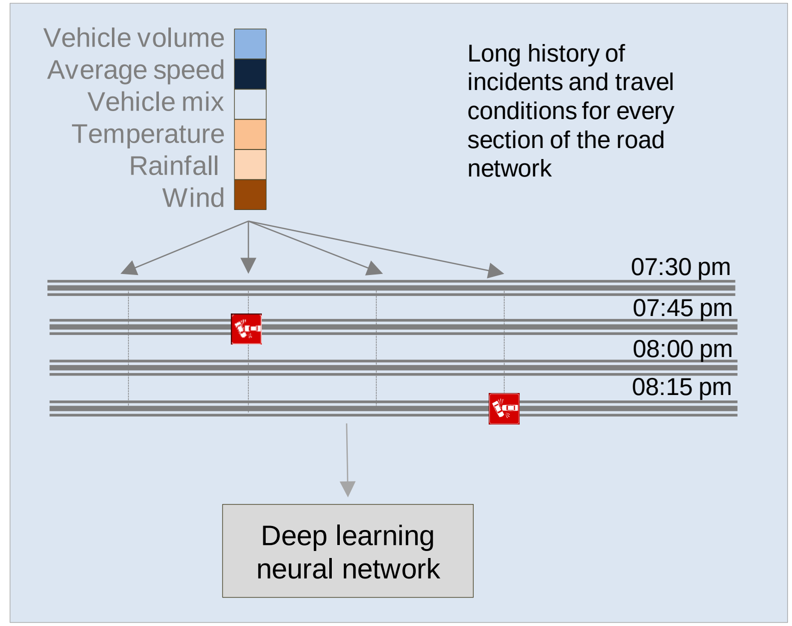
The prediction chain



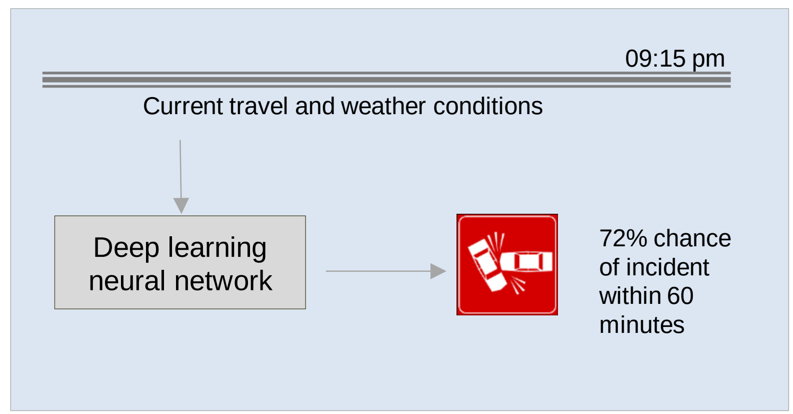
Project snapshot: predicting incidents in space and time

Use case	Predict if in the next two hours there will be an incident in a specific road section.
Rationale	• Road safety management policies are based on many inputs, including evidence of past incidents (e.g. speed limitations around incident hot spots). By providing additional evidence on the chance of incident at specific times and locations, it would be possible to fine tune both traffic management and incident preparedness. • Traffic is heavily monitored through loops or indirectly (floating car data, gps traces, cameras). The vast history of incidents and contributing factors (e.g. weather) can be explored to improve on crude incident prediction measures.
Stakeholders	• Traffic management • Emergency response organization
Methods	• A deep learning neural network is used to assign chances of an incident to combinations of input data (traffic patters, weather patterns)
Data and sources	• Traffic data, loop data • Weather data

TRAINING



USING



2

INCIDENT PREDICTIONS WITH NEURAL NETWORKS

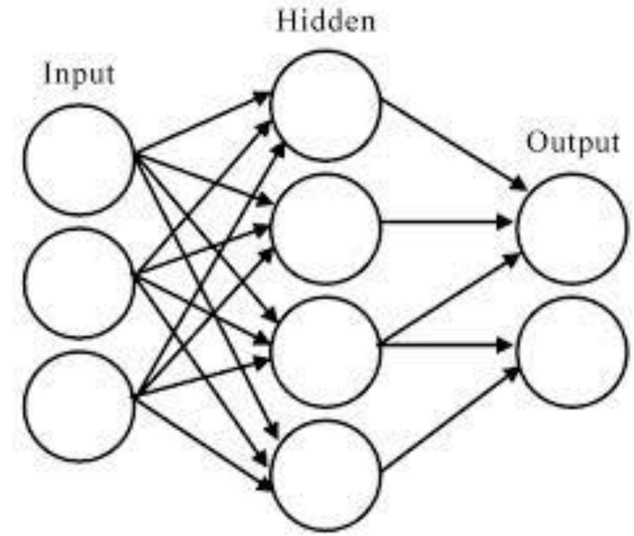
Linking traffic predictions and neural networks

- Availability of loop data offers a wide scope of possibilities for short-term prediction on a dense network of road segments:
- The 1-year traffic dataset encodes a virtually complete record of traffic dynamics in the country.
- The incidents dataset allows mining correlations between traffic conditions at road segments and the likelihood of incidents.
- Combination of the two datasets in a predictive framework can provide short term prediction of incidents.
- Weather data is used complementarily, to test whether it increases the accuracy of prediction.

- To encode the connections between traffic/weather patterns and the occurrence of incidents we use neural networks, adaptive algorithms that learn from training with the examples in the dataset and can then be used to predict incidents from real-time traffic/weather data.
- Strong correlations between traffic/weather conditions at the loops of a road translate to accurate prediction of future incidents by the neural net.
- When the link between traffic/weather and incidents is weak, the neural net will fail to predict accurately.

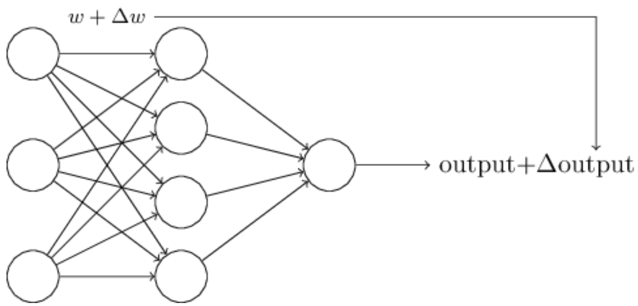
NEURAL NETS BASICS

- Neural Networks are machine learning algorithms used for prediction and classification. Their structure is inspired by basic connectivity features of biological neurons.
- They were first explored widely in the ’80s and ’90s. Crucial breakthroughs in the mid ‘00s led to rapid development, and neural nets are now achieving record accuracies in important classification tasks like image and voice recognition.
- A neural network consists of layers of “neurons” (i.e. variables that take on real values) connected in pairs between one layer and the next by weights. It is usually represented by a graph where neurons are the nodes and weights are the edges connecting them:
- The input layer of nodes takes in the data values. Each arrow represents a weight. When a weight is large, the value of the node at the tip of the arrow is very sensitive to the value of the node at the base of the arrow. A small weight means the two nodes are insensitive to each other.
- The values at the output nodes for a specific input represent the network’s prediction. For example, with two nodes at the output representing two possible future states (e.g. incident/no incident), the prediction of the network is the state corresponding to the node with the higher value.

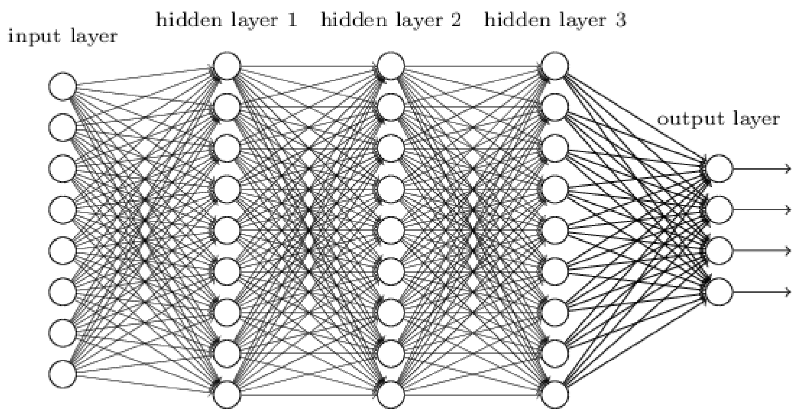


NEURAL NETS BASICS(cont.)

- During training, the values of the output neurons are compared to the actual values in the training data. The difference of the output values with expected (real) values is *backpropagated* through the network and weights are adjusted to reduce the error.



- *Backpropagation* is a technical term, referring to the precise manner adjustments are made to the weights when the network makes a wrong prediction. Neural nets first attracted interest as *universal approximators,* algorithms able to approximate any functional relationship between input and output. In practice however neural networks with many hidden layers proved hard to train. It was advances in backpropagation after the mid-'00s that allowed the training of many layers and the achievement of record accuracies, especially in image recognition. Modern neural network methods based on these advances are known as Deep Learning.

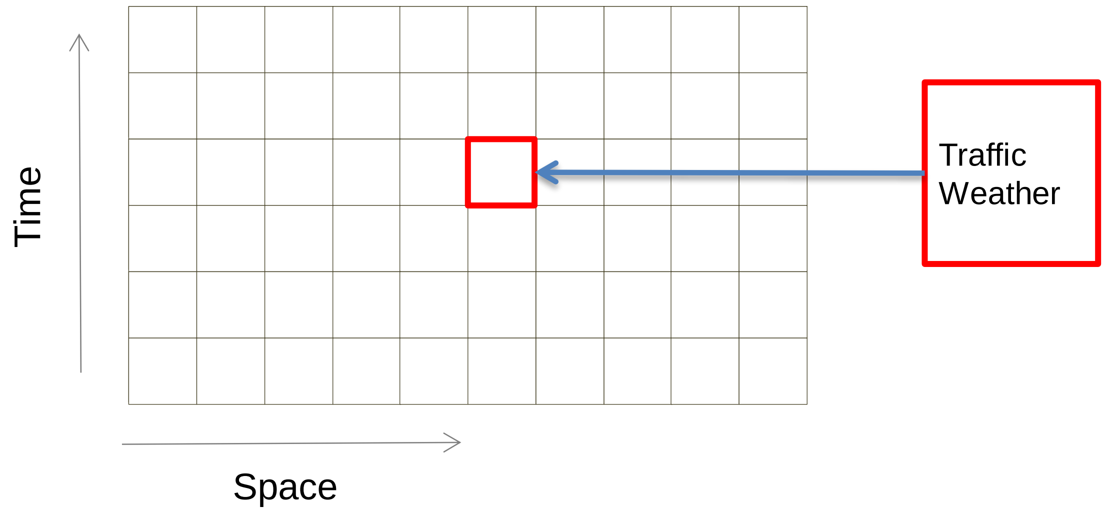


Why neural networks

- Neural nets can perform universal approximation, and so one would expect that any causal links between traffic and weather conditions and incident hazard will be revealed by an appropriately trained network.
- Neural networks are becoming better than humans at image recognition, and steadily gaining in accuracy in many other applied fields of prediction or classification. The concept of a classifiable image can be extended to other patterns, such as the loop traffic/weather conditions in our case.
- Accumulated knowledge of how to train neural nets for real images is not directly relevant to other uses, such as incident prediction. Independent research is needed to extend the tremendous recent successes of neural networks in image and voice recognition to vital applications in other sectors.

IMAGE INTERPRETATION AND INCIDENT PREDICTION

- Neural nets for image recognition have shown impressive results in recent years. International benchmarking competitions, such as that of the MNIST, are routinely won by deep learning neural net setups. Recent entries achieve accuracies over 99% in handwriting recognition and more than 96% in object recognition in photographs.
- Measurements of traffic over a time window can be organized into a spatio-temporal “image”. Each “pixel” of this image contains the values of traffic for a certain timestamp (and possibly side information like weather).
- The traffic images are tagged with a binary class (incident/no incident) depending on whether there was at least one incident on the road segment in some adjacent *future period* (here 2 or 4 hours).
- The tagged images are used to train a neural net to recognize in the traffic image the conditions that raise incident hazard and will lead to incidents in the next 2 to 4 hours, in a similar manner as an image recognition network is trained to recognize certain objects.



3

IMPLEMENTATION, DATA AND DATA PREPARATION

The data set: traffic data

The datasets contain three types of data:

1. Loop measurements (aggregated to 10 min timestamps). We use data for the year 2010.
   - loop id: the unique identifier of the loop in the road network detecting flow and speed
   - Index of lane number: the lane numbering, starting with # for the left-most lane
   - average vehicle speed (km/h)
   - car flow (# of vehicles in the last 10 min)
2. Lane index lookup table
   - indicates the distinction between lanes for loops
3. XML loop metadata
   - loop locations plus other loop metadata

Challenges related to the datasets:

> loops do not measure continuously (many breaks in the data – check slide loop selection)

> locations of some loop IDs can abruptly change, sometimes even by hundreds of kilometers (probably errors)

> loop locations are not associated to hecto-points, which are used to locate incidents in the incident dataset.

> loop direction indication does not match the incidents direction indication (alertcdirectioncoded ≠ RijbaanAanduiding)

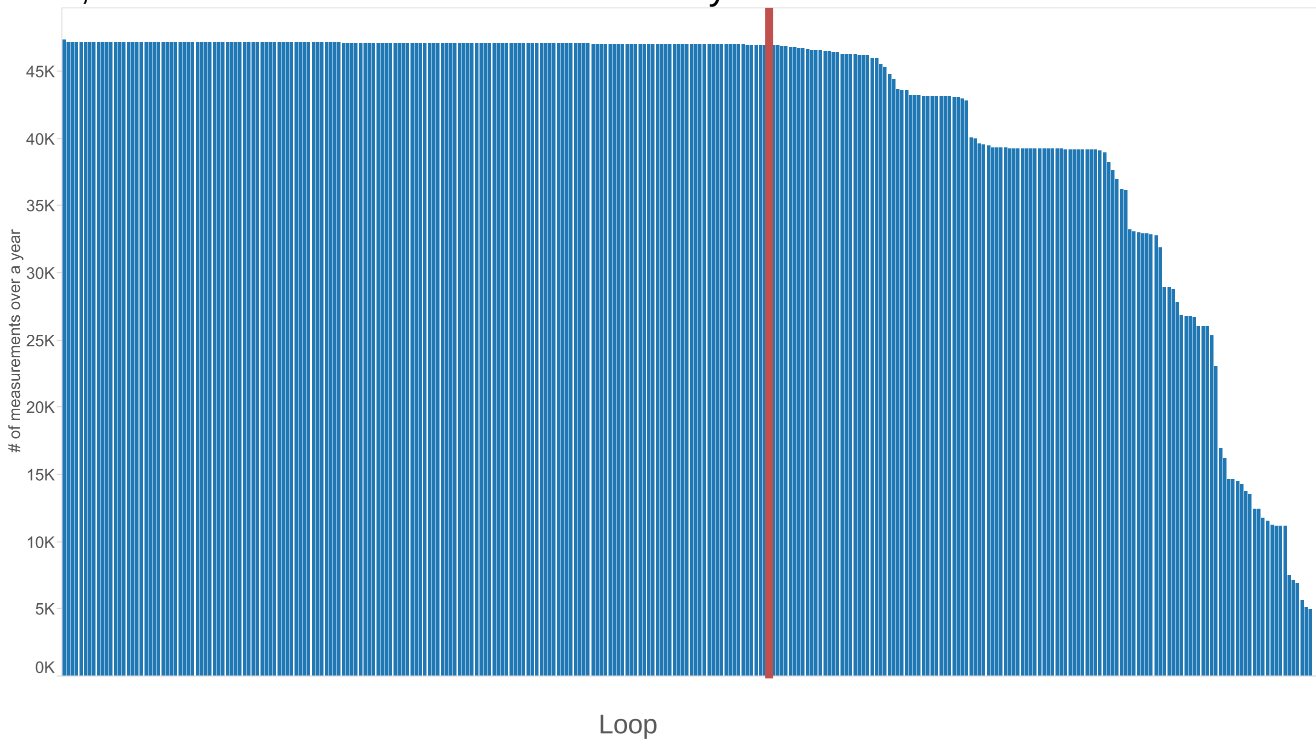
Preprocessing traffic data

- From the loop data we extract average speed and average car flow per 10 minutes
  - We do not distinguish between vehicle types: every vehicle counts as one
  - The values for one loop are the averages across the lanes
- For a road segment with L loops, we have 2L traffic variables (speed and flow) every 10 minutes.
- In the timeframe of one hour, 6 of these 10-minute variable sets can be combined to form an “image”, or signature, of traffic conditions in the road segment during the past hour.
- A new traffic image starts every 10 min, so successive images are partially overlapping.

	09:30 AM	09:40 AM	09:50 AM	10:00 AM	10:10 AM	10:20 AM	10:30 AM	10:40 AM	10:50 AM	11:00 AM	11:10 AM	11:20 AM
Speed	65	64	66	70	73	78	80	85	90	87	82	80
Flow	460	460	450	450	420	420	410	400	360	300	320	340
Image 1	65	64	66	70	73	78						
	460	460	450	450	420	420						
Image 2		64	66	70	73	78	80					
		460	450	450	420	420	410					
Image 3			66	70	73	78	80	85				
			450	450	420	420	410	400				

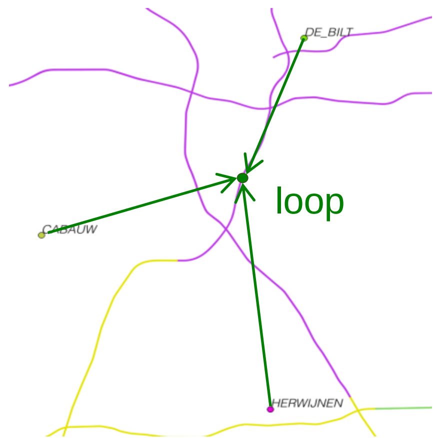
Loop selection

- Neural nets work better with complete data. In our case this requires that all “pixels”, i.e. loops, in our traffic “image” have values, without missing measurements.
- None of the loops has a complete record of measurements (has reads on every of the 52,560 10-minute intervals in the test year).
- Including every loop in the traffic image would result in most images containing missing values, which reduces the network accuracy.
- To construct full snapshots of the traffic that will minimize the chance of confusing the network, we pick only loops without too many gaps in the records. We selected loop locations with more than 47,000 10-minute measurements over a year.



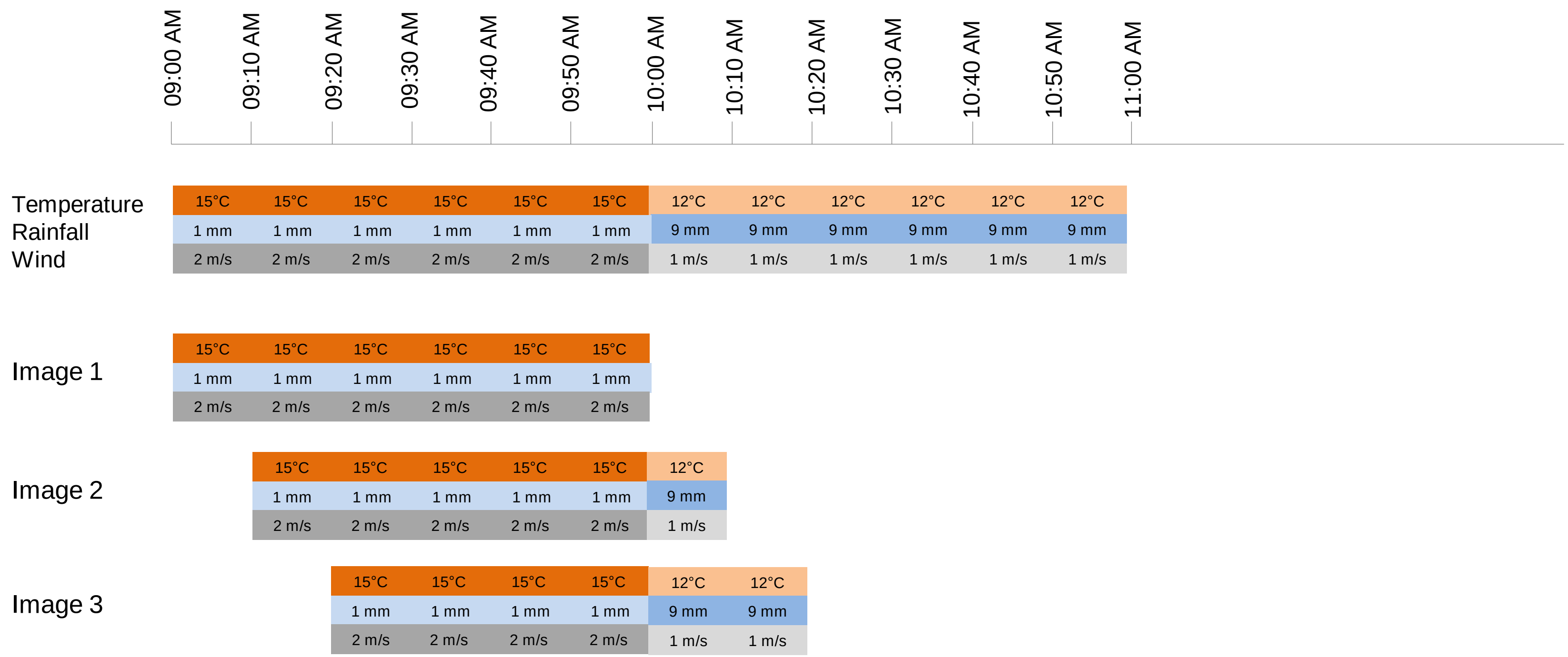
The data set: weather data

- We use public weather data collected by KMNI: http://www.knmi.nl/klimatologie/uurgegevens/selectie.cgi
- We use data from 35 weather stations covering the whole country.
- The station measurements are interpolated to provide estimates of several weather variables (temperature, visibility, wind speed, and others) at every loop location per one hour.
- For each loop location, we interpolated the weather variables from the 3 nearest weather stations. Weather variables are interpolated using inverse square distance averaging.



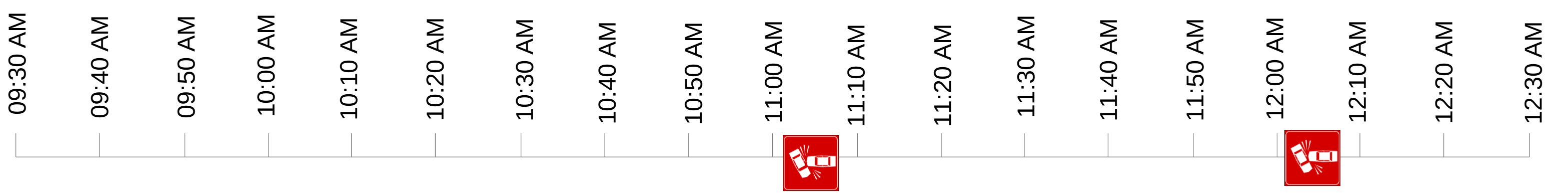
Preprocessing weather data

- Assuming we use W weather variables, the weather variables are structured as 1-hour images each one containing W variables
- The data can be further aggregated to averages across multiple loops



The data set: incident data

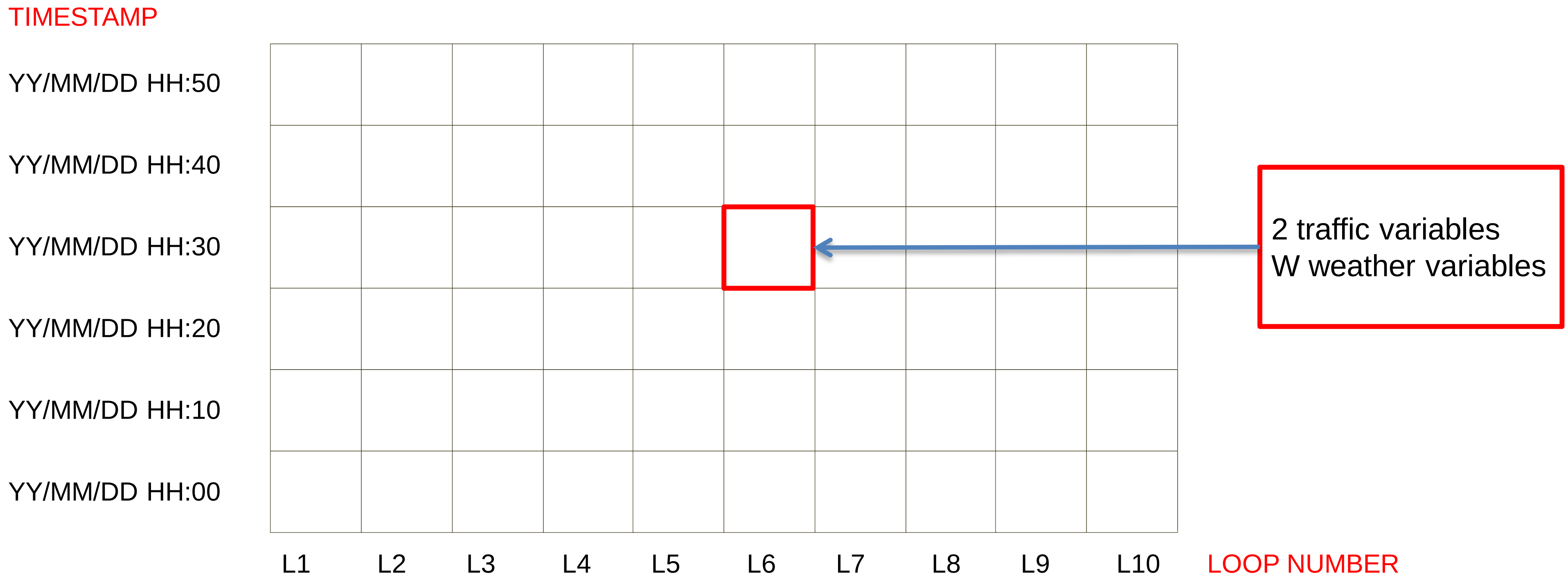
- The incident dataset contains incidents assigned the X/Y coordinates of the nearest hecto-point.
- In addition the incident dataset contains a host of attributes for each incident like severity of injuries, first response time and others. These attributes are not used in this project: we predict occurrence of incidents independently of type.
- Incidents are used to assign incident class (binary, incident/no incident) to the “image” of traffic and weather condition for a given road segment.
- Every 1-hour traffic image is assigned a 0/1 tag denoting whether *at least one* incident happened on the road segment during a definite future period. In the example below a traffic image is tagged with the incidents that occur in the next 2 hours.



Speed	65	64	66	70	73	78
Flow	460	460	450	450	420	420
Temperature	15°C	15°C	15°C	12°C	12°C	12°C
Rainfall	1 mm	1 mm	1 mm	9 mm	9 mm	9 mm
Wind	2 m/s	2 m/s	2 m/s	1 m/s	1 m/s	1 m/s

The “image” is tagged with 2 incident occurrences

Structuring traffic/weather like an image to train the network



- The grid of square “pixels” represents the traffic & weather conditions “image” for a hypothetical road with 10 loops L1 to L10.
- Each pixel of the image contains (2+W) values, the average traffic & weather measurements for the corresponding loop at the specific 10-minute interval.
- Each row of pixels corresponds to the whole road segment for a 10-minute interval.
- Each column is a single loop’s measurements every 10 min for an hour.
- The timeframe here is chosen here to be 1 hour, but other durations are also possible. We test image time durations of 1 or 2 hours.

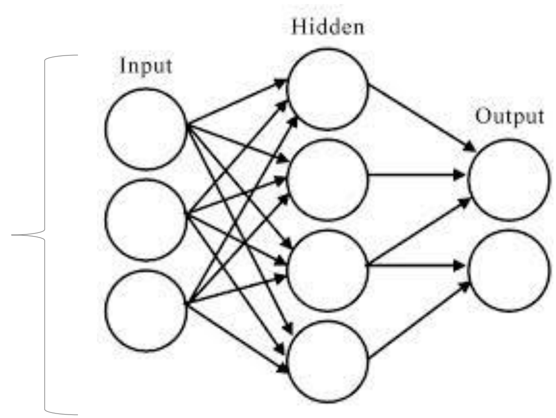
IMPLEMENTATION OF THE NETWORK

For our tests we implement a deep learning architecture for training the net:

- The input layer consists of T(2+W)L nodes, one for each of the the (2) traffic and (W) weather variables. The factor T in the front can be 6 or 12, depending on the duration of traffic data that we use to construct the traffic image (1 or 2 hours).
- The output layer consists of just 2 nodes, indicating **incident** or **no incident** in an adjacent future period (2 or 4 hours).
- The number of hidden layers together with a number of network parameters is set by testing to achieve the best possible accuracy.
- We explored configurations and parameterizations of the networks to determine the best combination. Deep learning setups with between 3-5 hidden layers provided the best accuracy.
- The hidden layers can contain varying numbers of nodes, as long as they are more than the number of loops in the image. Beyond this threshold accuracy proved insensitive to this parameter in the tests.

(2-5 hidden layers)

6(2+W)xL
traffic & weather
variables



Incident

No incident

TECHNICAL IMPLEMENTATION OF THE NETWORK

- Technical implementation of the neural network was done in Python using the specialized library Theano (http://deeplearning.net/software/theano/)
- Theano allows training of neural networks on the computer's GPU (Graphics Processing Unit) using the CUDA architecture developed by NVIDIA. Training on the GPU is up to 50x times faster than on the CPU taking advantage of parallel computation
- We trained the network using:
  - A local server with GeForce graphic card
  - An Amazon GPU virtual instance - https://aws.amazon.com/ec2/instance-types/

  http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using_cluster_computing.html
- We used many different networks to test configurations, parameters and depth (number of hidden layers); we tested them on a local computer as well as on Amazon.
- The training time for the network highly depends on the number of nodes and number of hidden layers. The networks used to perform incident predictions are relatively large, reaching 10 000 nodes per layer and up to 5 hidden layers. Typical training time is about 1 hour.

4

TEST AREA AND RESULTS

Test area and training sets

- We train the neural net on data from a section of 48km of Highway A12 south of Utrecht, extending from Utrecht to Den Haag and from Utrecht to Arnhem.
- We use traffic, weather and incident datasets from 2010.
- The traffic/weather data is used to construct traffic “images”, which are tagged with a binary incident/no incident class depending on the occurrence of an incident in an adjacent future time window. The duration of these images is 1 or 2 hours, depending on which choice offers the best accuracy in tests.
- 80% of the images are used for training and 20% for testing the accuracy of the network.
- To decide the duration of the prediction window, we perform error-rate analysis for a range of time windows, from 10 min to 12 hrs. We test the resulting uncertainty by using incidents-only strategies as benchmark. These are prediction algorithms that can be built from the knowledge of average incident probabilities only, agnostic to traffic patterns. We refer to these as *incidents-only strategies*.
- We optimize the neural net configuration and measure the prediction accuracy on the whole road and 1/2 , 1/4 and 1/8 segments.
- Weather data did not add significantly to the prediction accuracy. At optimal network configurations, when weather is added to traffic at the input the improvement is always within statistical error, coming at significant computational cost because of a multiplication of input nodes. Hence best performance tests were conducted with traffic data only.

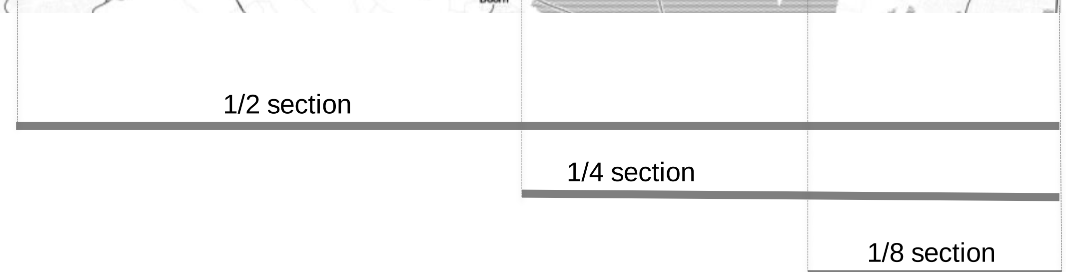
Test area: highway A12/E35 next to Utrecht

The test area is a section of 48km on the A12/E35 south of Utrecht. We use 153 loops in the test area (shown in the figure below).

The image below shows the density of incidents on the test road in 2010 (3229 incidents in total).

We test the neural network on the full test areas and on subsections:

- 1/2 section (88 loops, 1619 incidents)
- 1/4 section (37 loops, 942 incidents)
- 1/8 section (16 loops, 727 locations)



Benchmarking predictions

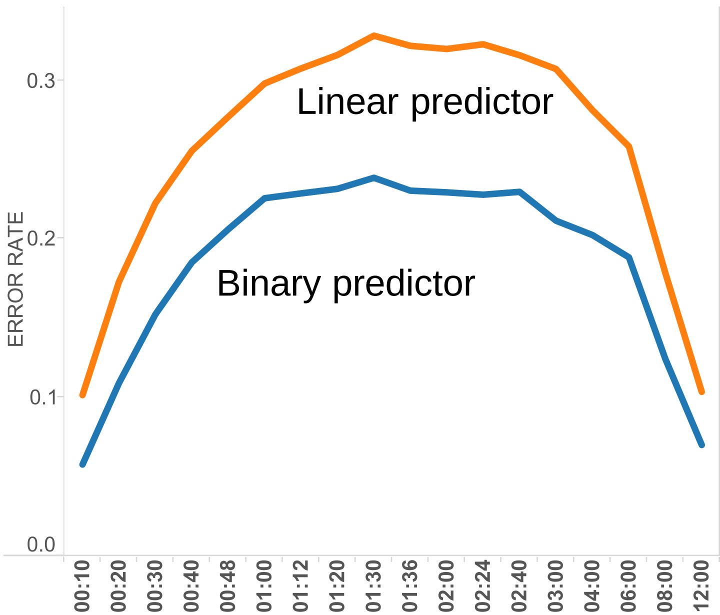
Every prediction project faces the need to benchmark the quality of the outputs. For this project we benchmark against what can be achieved based solely on past incident data.

We consider the following question: if the only data known is the occurrence of incidents on the selected road section in the past, how well can we predict if an incident is going to occur in the future (e.g. in the next 1 hour or 2 hours)?

There are multiple ways to derive the likelihood of an incident. We select two ways that correspond to two algorithms:

- **The binary predictor**. This algorithm looks at the past incident data and for every hour of the day and of the week and predicts incident (no-incident) if in that hour the chance of past incidents is more than 50% (less than 50%). This is an aggressive algorithm that does not distinguish between high or low probability of incidents.
- **The linear predictor**. This algorithm looks at the past incident data and for each hour of the day and of the week computers the chance of incidents (e.g. 67% probability of one or more incidents on a selected stretch of the road on Thursdays between 2pm and 3pm). For the future it randomly generates an incident at a certain time and place that will have the same probability of past incidents.

Benchmarking predictions, cont.



The image to the left illustrates the error rate for the binary and linear predictors for variable time intervals for the test road section on the A12.

When the time interval is very short (e.g. 10 minutes) any algorithm is precise in that the chance of incidents is always low and it easy to predict. The same holds, for opposite reasons, when the interval is very large: on a 12 hours interval the chance of incident is high and predictable.

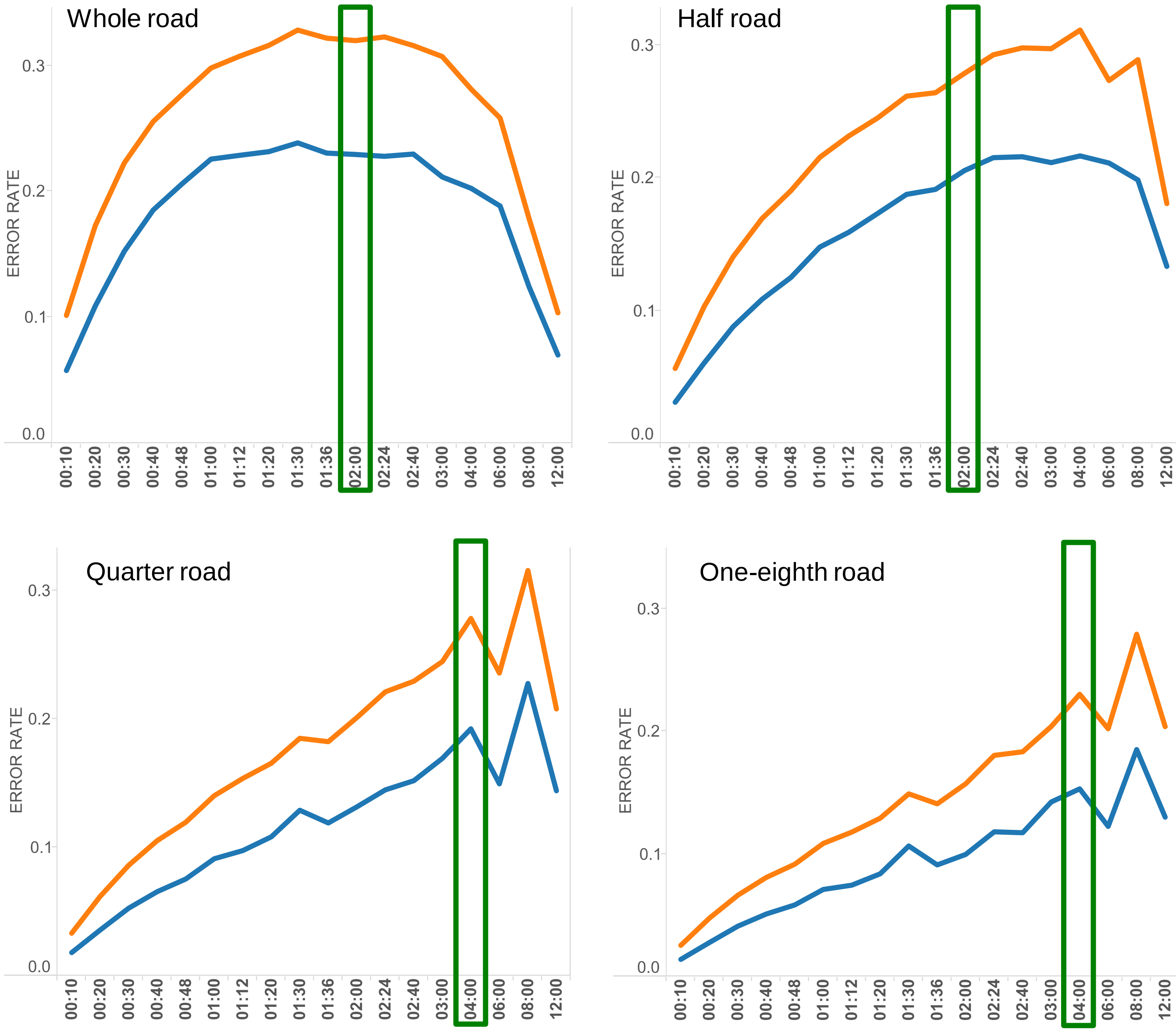
The errors in the range between 1 and 4 hours, which are relatively stable, are a better characterization of the predictive capability of tools based only on past incident data.

The binary predictor performs systematically better than the linear predictor with error rates around 20%, compared to 30% of the linear predictor.

This means that only by looking at past incidents it is possible to predict future incidents with relatively high accuracy.

These estimates provide a benchmark for the neural network method.

Time window for predictions: window of maximum uncertainty



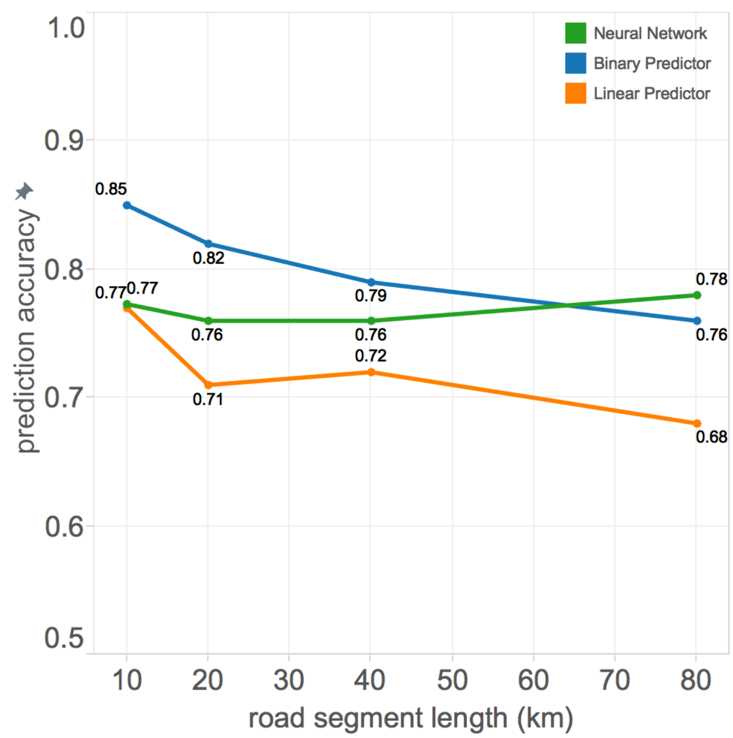
To test the neural networks we look for the duration of the prediction window that would incur the maximum average error rate on the binary and linear predictors.

For the analysis we pick a time window for prediction that is close to the region of maximum error rate for the two incident-only strategies, i.e. maximum uncertainty when only incidents are known.

This is the 2 hours range and 4 hours range: that is, predicting incidents in the next 2 hours and in the next 4 hours.

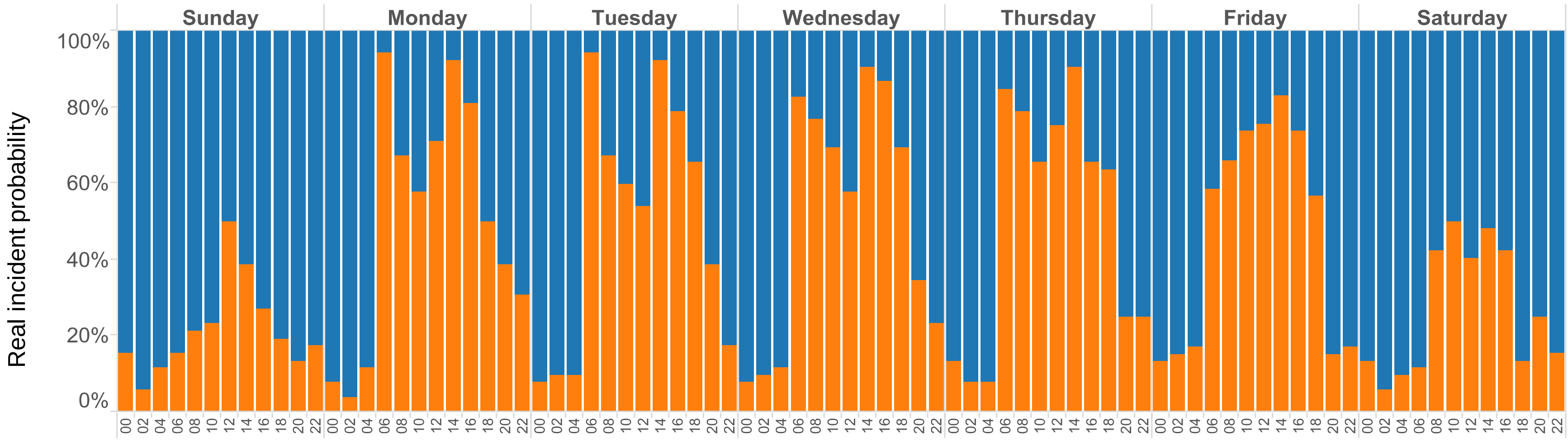
Benchmark: average results of neural net vs. incidents-only strategy

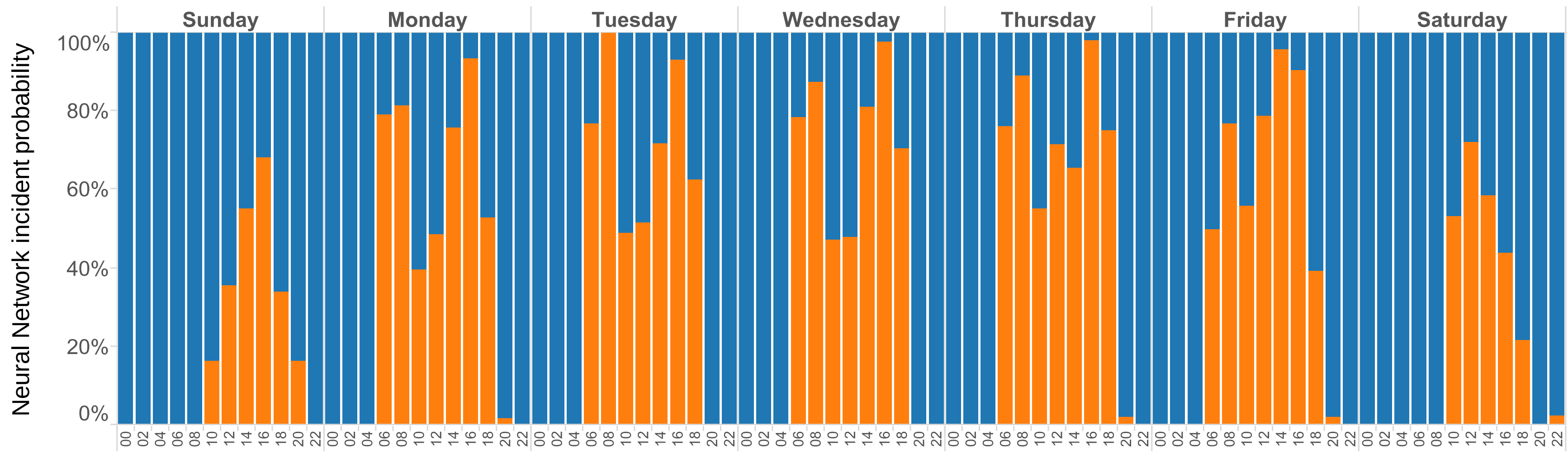
Test road Segment on A12/E35	Image time window	Prediction time window	Neural Network (NN)	Binary predictor (BP)	Linear predictor (LP)
1 (full length)	2h	2h	78%	76%	68%
1/2	1h	2h	76%	79%	72%
1/4	1h	4h	76%	82%	71%
1/8	1h	4h	77%	85%	77%



- The neural net can diagnose incident hazard better than knowledge of incident rates only on long stretches of roads. For shorter lengths the binary predictor is more accurate while the neural network exceed the linear predictor.
- With one year of training data, the prediction accuracy degrades when smaller road segments are examined. This is caused by the reduction of training data available.
- The binary predictor and the linear predictors are minimally impacted by dataset duration: their prediction ability will not change with longer datasets and history data.

Real incidents probabilities vs. neural network incident probabilities

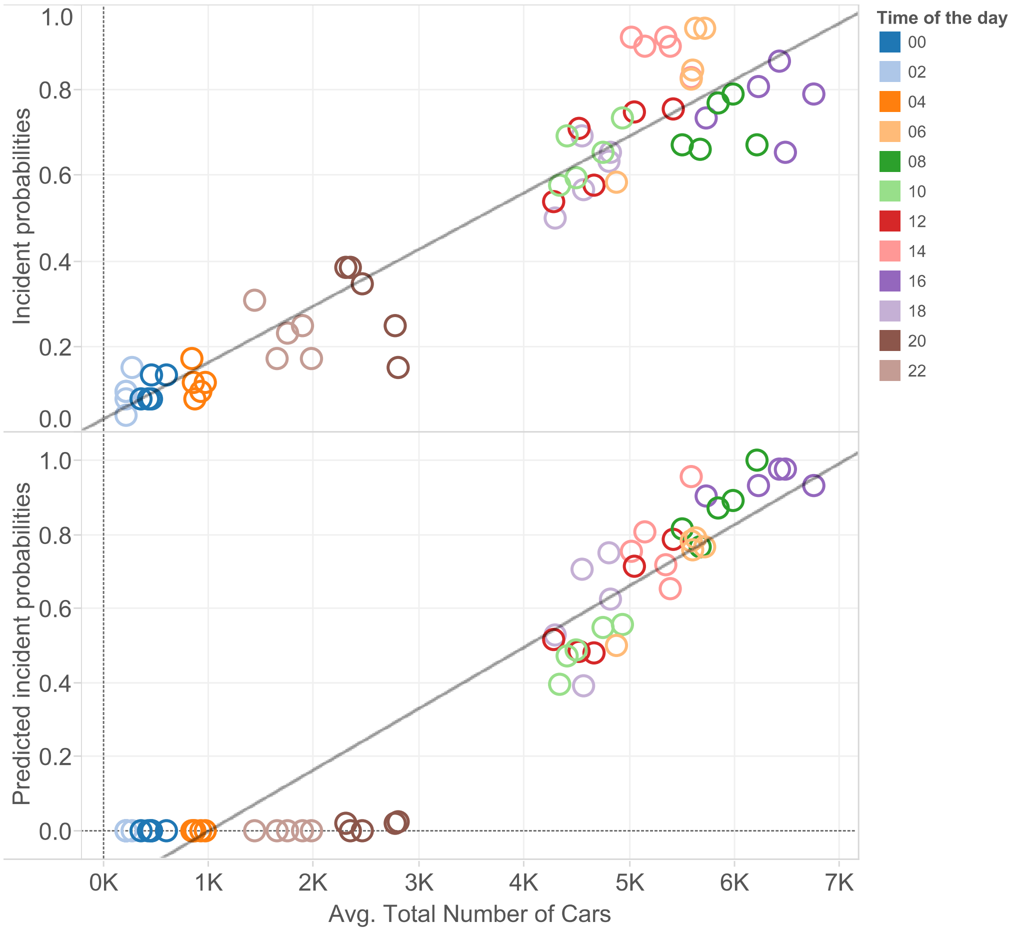




Real incidents probabilities vs. neural network incident probabilities

- During 10:00pm – 06:00am the neural network adopts the optimal rational strategy: it predict no incident to incur a minimum number of errors.
- Incidents during this time have practically zero correlation with large scale traffic patterns. They are due to other factors, not included in the input data.
- On the contrary, during the day the network supersedes the optimal rational strategy, attaining a 3% advantage during day/evening hours.
- There is non-trivial correlation between large-scale traffic patterns on the road and incident hazard during these hours, encoded in “signatures” which the network is able to recognize.

Correlation between number of cars and incident probability



1. The diagram provides another view of predicted probabilities, vs. the number of cars on E35.
2. As expected and well known, there is a strong correlation between number of cars and probability of having an incident (top plot).
3. There are also significant deviations. It is the patterns that cause these deviations which the network is trained to recognize.
4. The day/late-night traffic load dichotomy is clear. Traffic is too sparse during late night hours to correlate with incidents, so the network is unable to predict any of these incidents.
5. During daytime and evening however, the network discovers strong correlations between traffic and incidents, and is also able to reproduce the correlation with the number of cars and part of the deviations.

5

CONCLUSIONS AND RECCOMENDATIONS

Conclusions

- This study explores the use of neural networks to predict incidents on a highway network.
- The rationale of this approach is the increasing availability of data that can be correlated to incidents and the development of new methods to manage and analyze massive amounts of data. The study is meant to assess the usability and usefulness of neural networks for this purpose, but also the limitations of the methods and the which conditions under which it can be developed further for practical implementations.
- The study leverages recent advances in neural networks applied to image recognition and speech recognition and applies the same methods to incident identification and prediction.
- While there are many causes of incidents that can be identified, the project focuses on the relationship between incidents and context data (traffic and weather), data that can be assumed to be generally available.
- The term “prediction” is interpreted in a strict way: the chance of an incident or no-incident occurring during a specific time window and on a specific road segment.
- Ideally the predictions should be implemented to the entire road network: for practical reasons and to validate the concepts first we explore the method for a section of 48km of the A12/E35 near Utrecht.
- We also choose to predict incidents within a 2 hours interval: the choice is dictated by two factors:
  - The interval corresponds to the interval of maximum uncertainty in historical data
  - The data set used is for 1 year of incidents, which includes a relatively limited number of incidents even in 2 hour intervals. Shorter intervals would be insufficient to train the network.

Conclusions, cont.

- From the results we see that large scale traffic patterns (the traffic detected on the A12 section used for the study) correlate with incident hazard during day and early night hours.
- During late night/early morning hours incidents seem to be based on wholly other factors and no correlations are detected.
- Neural networks achieve a better accuracy than the optimal rational strategy that takes only the average incident probabilities into account, indicating the existence of non-trivial traffic patterns that raise a high probability of incident.
- Weather data does not contribute significantly to the accuracy. To some extent this may be an artifact of the 1-hour resolution of weather data, and weather may prove more valuable for prediction if the resolution is increased to 10 min. But at the same time, relevant atmospheric phenomena are already indirectly encoded in the traffic patterns (for instance by drivers, which adapt to weather conditions or by dynamic speed regulation on the network)
- We tested the network prediction capability for the entire section of the test area (48 km) and for two half sections and 4 quarter sections.
- Scaling down the size of the test area hurts relative accuracy, indicating that a different approach is needed for the micro-level (up to a few loops).

Observations and limitations of the study

The results of the study are impacted by several limitations:

- The duration of the data sample appeared as too limited. Data sets of 5 or 10 years would provide a much broader set of patterns against which to train the network. The expected result is an increased prediction precision but also a lower spatial granularity of predictions.
- We aggregated traffic data (for instance averaged speed and flow across lanes) to reduce the number of input parameters to the network. This is linked to the fact that the data sample does not contain sufficient data to train the network with even more degrees of freedom. These limitations are reduced if a longer, data set is used.
- The weather data is coarse and interpolated to one hour intervals. This reduces the weather time and space variability and may explain why weather does not emerge as a useful predictor. Better weather data will change this conclusion.
- The prediction accuracy is measured against the actual occurrences or otherwise of incidents. The neural network is trained to give the same weight to errors, be them false positives and false negatives. A different training pattern (for instance, train the network to recognize positive incident occurrences with maximum precision, accepting for instance false positives) could lead to different conclusions.
- Incident attributes, such as type of car involved, number of passengers, severity of the incident etc. are not considered. This again is linked to the data duration which is too short to train the network on finer attributes of the incidents.

Observations and recommendations for future work

1. The analysis of A12/E35 demonstrated that neural networks can discover meaningful correlations between traffic patterns and incidents. However, scaling down to the level of very small road segments, where predictions are most valuable, requires a different algorithmic approach.
2. For localized prediction, large scale spatial traffic patterns encoded in measurements from many loops are not directly relevant. They relate to incident hazard on long stretches of road, but the occurrence of an incident at a specific location depends more on local fluctuations (e.g. local speed differences, or abrupt changes in vehicle density).
3. To produce accurate localized predictions for incidents, data for one or two loops near selected incident hotspots would be most relevant. Going to the available 1min accuracy (instead of 10 min for the large scale analysis) is the best option to take full advantage of temporal patterns of the local traffic.
4. Previous attempts at local prediction have not made use of temporal patterns, instead using small “snapshots” of local traffic of 5-10 min duration, with the prediction time window being correspondingly short (10 min). The prediction accuracies from these studies approach 70%.
5. With appropriate implementation of last-generation sequential prediction methods, the extent of the duration window and the prediction accuracy may be pushed to levels that signal applicability in real-time incident warning systems (30min-1hr, >85%).
6. Two classes of algorithms based on mining temporal patterns hold promise for local prediction, and have not been tested on loop/incident data:
   - Recurrent Neural Networks: A class of predictive neural network with internal memory that allows it to be trained with sequences of data points and classify them in real time.
   - Bootstrapped Exponential Weights: A sequential learning family of algorithms that combine the predictions of a large number of data-derived predictors into a single decision of higher accuracy.

We propose to test implementations of these algorithms with 5 years of traffic and incident data at 1min resolution, with the goal of raising the accuracy to levels that open the prospect of application to real-time streams of loop data.

Selected references

M. A. Nielsen, "Neural Networks and Deep Learning." Determination Press, 2014.

Y. Bengio, I. Goodfellow, A. Courville. “Deep Learning”. MIT Press, 2015.

M. G. Karlaftis, E I Vlahogianni. “Statistical methods versus neural networks in transportation research: Differences, similarities and some insights”. Transportation Research Part C: Emerging Technologies, 19, issue 3, 2011.

S. Roshandel. “Impact of real-time traffic characteristics on freeway crash occurrence : systematic review and meta-analysis.” Masters by Research thesis, Queensland University of Technology, 2014.

F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. Goodfellow, A. Bergeron, N. Bouchard, D. Warde-Farley and Y. Bengio. “Theano: new features and speed improvements”. NIPS 2012 deep learning workshop.

J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley and Y. Bengio. “Theano: A CPU and GPU Math Expression Compiler”. *Proceedings of the Python for Scientific Computing Conference (SciPy) 2010. June 30 - July 3, Austin, TX.*