Virtual patrolling

Martie van der Vlista, Luc Wismansa,b,*, Paul van Beekc, Leon Suijsa,c

aDAT.Mobility, P.O. Box 161, 7400 AD, Deventer, The Netherlands
bUniversity of Twente, Enschede, The Netherlands
cGoudappel Coffeng, Snipperlingsdijk 4, 7417 BJ, Deventer, The Netherlands

Abstract

Approximately 25 per cent of all congestion on motorways is caused by incidents. By virtual patrolling, incidents e.g. queues, accidents and car breakdowns on a road network, can be predicted or detected in an early stage. This early detection and prediction of an incident likely to happen, offers the opportunity to act faster and therefore reduce congestion compared to current practice. A module is developed capable of virtual patrolling and based on four key features: data fusion, real-time estimation of the fundamental diagram, fuzzy traffic state estimation and artificial neural networks. In CHARM, the cooperation between the Highways England (UK) and Rijkswaterstaat (NL), the potential of this module has been demonstrated in a simulation environment and in 2015 further developed, deployed and tested in real life.

Extended Kalman filtering and traffic flow theory algorithms are used to fuse data from loop detectors with FVD (floating vehicle data) to provide detailed traffic state data (i.e. flow and speed for 10 seconds for every segment of approximately 200 meters). For the prediction of the future traffic state, information on the actual road capacity is necessary. Road capacity varies for weather, lightness, number of heavy vehicles et cetera. A self-adapting module is used to make a real time estimation of the fundamental diagram and road capacity. The traffic state estimation itself is based on fuzzy logic to determine to what extent the traffic state is free flow or congested. If only speed data is available, alternative simplified fuzzy logic rules are applied. The fuzzy traffic state estimations for separate segments is combined into a combined traffic state estimation. This traffic state estimation is used as input for artificial neural networks (ANN) to predict whether it is likely that a queue will form within the short-term future (within 10 minutes) on a relevant segment. This is defined as a segment that suffers from queues on a frequent basis.

The results of these estimations will be used for virtual patrolling to distinguish between normal and expected congestion and other situations. Information from the modules can also be used for the detection of prediction of shockwaves. This can be of help to act faster when incidents happen and to prevent or solve shockwaves by giving (in-car) advices on speed and lane choice to car users.

Keywords: Charm; virtual patrolling; detection; prediction; traffic state
1. Introduction

The Highways England (UK) and Rijkswaterstaat (the Netherlands) are responsible for the operation and maintenance of the English and Dutch motorway and trunk road networks. Since 2011 they have teamed up in order to develop requirements for a new generation of traffic management (centre) systems that may be jointly procured. Setting up this architecture is being executed under the CHARM programme. The development of virtual patrolling is a solution for one of the challenges formulated within this programme.

1.1. CHARM-PCP objectives

The objective of CHARM is to move towards an open modular architecture for traffic management systems that enables to prevent future vendor lock-in and allows to plug in more easily new modules that can bring breakthrough innovations to traffic management services. CHARM has also the objective to create three new modules for the CHARM architecture that correspond to three sub-challenges that form part of the overarching challenge by the CHARM traffic management authorities to achieve radical improvements in traffic management services that contribute to optimising the performance of the road network, improving road safety and reducing CO₂ emissions. The CHARM programme formulated three challenges:

- Challenge 1: Advanced distributed network management
  To realise a module that provides automated support for management of large (nationwide) traffic networks. The module should be a multi-layered, self-learning engine that is able to manage large networks and balances between different types of goals.

- Challenge 2: Detection & Prediction of Incidents
  To realise a module that provides early identification and prediction of near future events on the network (accidents, queues, etc.), called "virtual patrolling". Detection and prediction should be targeted at the top 3 incidents: accidents, car breakdowns and queues.

- Challenge 3: Support of Cooperative ITS Functions
  To realise a module that supports the implementation of cooperative system services requiring a participation of intelligent infrastructure, in order to optimise the performance of the road network.

This paper on virtual patrolling concerns challenge 2: the detection and prediction of queues and the detection of non-expected traffic situations caused by accidents and car breakdowns. Challenge 2 specific minimum requirements are:

- The module should at least be able to provide identification of an incident anywhere on the network, or should provide near future probability of (traffic) queues.
- Identification of an incident should be based on the evaluation of actual traffic situation, using the actual and past traffic data, it shall identify where and what type of incident took place.
- Near future probability of (traffic) queues should be based on the evaluation of the actual traffic situation. The facility shall identify where, when (in near future) and with what probability a queue will appear.
- The Road Network Operator must be able to monitor and manage these facilities and hence the use of the simulation and modelling tools.

1.2. This paper

In this paper we will describe the developed framework and the test results within a simulation environment. Furthermore, the innovations and suitability of the results will be discussed. In Chapter 2 of this paper we explain the used theory. Chapter 3 focuses on the research method and Chapter 4 the results are discussed. The paper ends with a discussion of the results.
2. Theory

2.1. Introduction

This section offers a short literature review of state-of-the-art techniques related to both incident detection as prediction. First an introduction is given on Kerner’s three phase theory which is the backbone of our methodology. Subsequently, various incident detection techniques are discussed. Lastly, state-of-the-art prediction techniques are described.

2.2. Incident detection

Fundamental diagram

Since Greenshields’ observations in the 1930’s, a lot of research has been done on the subject of traffic flow dynamics. New and extensive observations of traffic flows resulted in an altered shape of the fundamental diagrams. Currently, the inverse-lambda shape of the flow-density diagram (Figure 1) is considered to be the best approximation. The left “branch” of this flow-density diagram can be seen as the “free flow branch” whereas the right “branch” can be seen as the “congested branch” (Treiber and Kesting 2013).

Kerner’s three phase theory

Based on the difference between the theoretical relation and non-heterogeneous traffic, Kerner developed the concept of “synchronized flow” and its related three phase theory (Kerner 2003). Within the “congested branch” which can be identified in the lambda-shaped flow-density diagram, Kerner distinguishes two phases: synchronized flow and wide moving jam. Additional to these two phases, Kerner identified the free flow phase resulting in three different phases as can be seen in Figure 2:
- Free flow (F)
- Synchronized flow (S)
- Wide moving jam (J)

Fig. 2. Kerner’s three phase diagram including Free flow (F), Synchronized flow (S) and Wide moving jam (J) (Kerner, 2000).

The distinction which Kerner makes between synchronized flow and a wide moving jam is based on spatial-temporal features. The wide moving jam is characterized by the upstream movement of the downstream jam front with a constant speed. The downstream front of synchronized flow is normally fixed at a bottleneck. In Figure 3 both a wide moving jam as a synchronized flow can be identified. It is clear that both phases show a specific spatial-temporal feature. The wide moving jam moves upstream through time, while the synchronized flow is fixed at the bottleneck.

Fig. 3. Space-time diagram including traffic speeds (Kerner, 2000).

Besides this fundamental distinction between synchronized flow and the wide moving jam, Kerner also noted that two well-known effects of congested traffic can occur in both of these phases; (1) Synchronization of the average vehicle speed between different lanes and (2) a wide spreading of empirical data in the flow-density plane. Furthermore, Kerner also states that a wide moving jam has more the tendency to standstills within the traffic flow, whereas a synchronized flow has more the tendency to a synchronization of vehicle speeds across lanes with relatively higher speeds.

As can be seen in Figure 3, transitions between the three phases can occur through space and time. However, one of the basic principles of the three phase theory is that a transition only takes place between the free flow and the synchronized phase and the synchronized phase and the wide moving jam. This means that no direct transition between free flow and a wide moving jam is possible. Therefore, the free flow traffic first has to move to synchronized flow where after it can make a transition to a wide moving jam. The same principle holds for the transition from a wide moving jam to the free flow phase. Furthermore, research by Kerner and Rehborn (1997) showed that once traffic has fallen from free flow into synchronized flow it does not easily return into free flow again. Traffic can be "trapped" in synchronized flow for up to several hours. This phenomenon is also known as the capacity drop which is discussed next.
Capacity drop

Kerner’s three phase theory can also help to understand the phenomenon of capacity drop. Kerner (2003) describes that traffic can be either stable or metastable in the free flow state. The free flow is metastable if the flow rate is equal or higher than the outflow rate, $q_{out}$. For a metastable free flow, perturbations in the traffic flow can result in a transition from free flow to jam. The higher the flow rate of the free flow is compared to $q_{out}$, the smaller the perturbation has to be to initiate the transition to jam.

As can be observed from Kerner’s three phase diagram the flow rate of the free flow state for densities between 20 to 30 veh/km is generally higher than the flow rate of synchronized flow. Because of perturbations in the metastable free flow, a transition to synchronized flow is initiated. Consequently, for the same densities lower flow rates can be achieved. This phenomenon is called capacity drop. Once the capacity drop emerged, traffic flow cannot easily “jump” back to its free flow state. Therefore, first the inflow to the jammed area has to fall to a much lower value (Treiber and Kesting 2013). The capacity drop phenomenon has been subject in various studies. Leclercq (2011) listed literature with capacity drops observed ranging from 10% up to 30%.

2.3. Congestion prediction

Only a few articles dealt with traffic prediction related research until the 1990’s when research interest increased significantly (see also Van Arem et al., 1997). However, congestion prediction remained a virtually untouched subject. Kaysi et al. (1993) set up a framework for an Advanced Traveller Information System (ATIS). In this framework historical and updated Origin-Destination data with real-time data from a surveillance system were used as input to predict congestion. This may lead to scenarios where traffic control and route guidance are used to change traffic conditions. Since the framework was never implemented it was presented only at a conceptual level. The authors suggested that a Dynamic Traffic Assignment (DTA) model, fed by 3-dimensional O-D matrices - with time as the third dimension -, would be necessary to make good predictions. More recently Wismans et al. (2014) reported on a DTA model approach implemented for the Assen region in the Netherlands, showing it’s feasibility. However, scalability of such approach and quality of predictions especially during non-regular traffic conditions are important points of attention. Kaysi presented two alternatives to the DTA model: Least Squares Estimation and the Kalman Filtering Procedure. Dougherty et al. (1993) described a congestion prediction research study. Motorway data were collected in the Netherlands through dual induction loops and consisted of speed, volume and occupancy at lane level. Data were aggregated to 1 minute time bins and a professional traffic manager decided whether the traffic state was congested or not. The prediction horizon was 10, 20, and 30 minutes. Only one method was used (Multi Layer Feedforward Artificial Neural Networks – MLF-ANN’s) and the authors’ main conclusion was that the results were promising but comparisons should be made with other methods.
Huisken and Coffa (2000) used motorway data that were aggregated to 1 minute time bins to predict congestion 5, 10, and 15 minutes in advance. They used time series analysis and compared the results with ANN-based models. If errors occurred they were categorised as false alarm (falsely predicting congestion) or error (failing to predict congestion) and were measured in percentage of time. The ANN technique outperformed the time series analysis technique. In Huisken (2001) two methods were compared: time series analysis and Adaptive-Network-based Fuzzy Inference Systems (ANFIS) fuzzy logic whereas the input data sets were identical to those in the study of Huisken and Coffa (2000). In this case, the two methods gave similar results. Huisken and Van Maarseveen (2000) again use the same input data sets to compare 6 different methods: Multi Linear Regression, time series analysis, Multi Layer Feedforward neural networks, Radial Basis Function neural networks, Self Organising Map neural networks, and ANFIS fuzzy logic. The performances of the methods were comparable with the exception of the SOM neural networks (results were not given due to high errors) and Multi Linear Regression that was outperformed by the other methods.

In Huisken (2006), an elaborate study is carried out on the comparison of several methods used to come to congestion prediction. The methods consist of several supervised ANN methods, an unsupervised ANN model, ARMA time series analysis, Multi Linear Regression, Fuzzy Logic and a so-called ‘naïve’ method (do-nothing method) as a minimal hurdle to pass for the other methods. The prediction horizon is (increasingly) growing: 5, 10, 15, 20, 25 and 30 minutes. Obviously, results deteriorate with increasing prediction horizon. The conclusion of the study reveals that supervised ANN’s outperform the other methods, and this method is used in the current virtual patrolling project.

3. Research

This chapter discusses the methodology which has been proposed for incident detection and prediction. This methodology has been derived from an extensive literature review on state-of-the-art techniques on both subjects. Subsequently, most essential elements within the framework are discussed.

3.1. Framework

In Figure 5 the methodological framework of the proposed methodology is presented. Each step is briefly described.

![Methodological framework](image)
• Initialization: Within this first step, the network, division in segments, handling of nodes and connection with measurements is initialized. Further, the initial FD is estimated for all segments and parameters are set.

• Database: The database contains the network configuration, parameters needed for the various algorithms, the initial FD and all segment data, FD estimates, state estimates, predictions of previous time-steps and the outcome of the incident detection.
  ○ Parameters: The parameters used in the algorithms are stored in a database and are potentially updated based on the evaluation.
  ○ Historical data: The historical database contains the initial FD per segment (potentially updated based on the evaluation). Furthermore it lists historical segment data for every time step (flow, speed and/or occupancy), state estimate and predictions. This can be used for the evaluation step.

• Segment data: The proposed methodology to detect and predict incidents makes use of fused segment data. This segment data is provided by the data layer within the CHARM architecture. For each segment the data layer provides speed and preferable also flow data and is based on fusing loop detector data and floating vehicle data using extended Kalman filtering and traffic flow theory algorithms to provide traffic state data on a more detailed level (i.e. segment lengths of 50-200 m and moving minute averages every 10 seconds) than loop detector data itself typically provides.

• Detection algorithm: The incident detection phase consists of three consecutive steps:
  ○ First a fundamental diagram for each segment is estimated or updated, based on a comparison of the default fundamental diagram and current measurements.
  ○ Subsequently, the fundamental diagram specifics are processed through a fuzzy inference system leading to traffic state estimations for each segment based on (Wismans et al. 2015).
  ○ Finally, both a temporal as spatial evaluation is performed on these traffic state estimations to identify and classify incidents.

• Prediction algorithm: The prediction algorithm uses historical data to derive explanatory variables and parameters using neural networks.

• Evaluation: The evaluation loop is used to update the parameters if necessary

3.2. Testcase

A micro simulation study has been performed in order to develop more knowledge about the proposed methodology. A simulation study provides the possibility to test the methodology in a controlled environment. In this controlled environment we know exactly when and where incidents take place. This provides a reference, the ground truth, which can be compared with the incident detection and prediction by the developed algorithms. This enables quantitative statements about the detection and prediction quality of the proposed algorithms in terms of time-to-detection, false alarm ratio and success ratio. The results of this simulation are described in the next chapter.

The current phase of the project also uses real data for detection and prediction of incidents. Two regions are considered, one in the Netherlands (Eindhoven) and one in the United Kingdom (Birmingham). The algorithms are fed by fused loop detector data and floating car data. The results of this part of the Charm project will be available in February 2016 and may be presented at TRA2016.

4. Results

This chapter presents the results of the simulation study and the algorithm development. First, the results for the incident detection algorithm are visualized. Next, a more quantitative analysis of the quality of the detection algorithm is presented. Finally, the results for the incident prediction algorithm are presented.
4.1. Detection

The incident detection algorithm has been tested on a dataset consisting of 50 random peak periods. A 200 meter aggregation level (segment length is 200 meter) has been used for this study. In time, minutes are used as data aggregation level for speed and flow. This minute data is updated every 10 seconds.

As mentioned before in this report, accurate flow data is only available for loop detectors. Floating vehicle data (or many alternative speed data sources) do not typically produce flow data. This means that flow data needs to be estimated from floating vehicle data. As this estimation process is likely to result in less accurate flow data, the algorithm has been tested on its sensitivity for errors in flow data. Therefore, the algorithm has been used on the same dataset as mentioned before, however now with a certain error on the flow data. Also, a variant of the algorithm with linear membership functions for speed has been tested, which represents a situation in which no flow data at all is available. In table 1 the results of these variants of the algorithm are presented. It is clear that an error of up to 20% does not have a significant impact on the quality of the algorithm in terms of any of the three indicators. However, the variant with linear membership function does result in a much higher false alarm rate. At the same time the time to detection is reduced compared to the use of curved membership functions. From these results can be concluded that flow data is desired to limit the amount of false alarms. However, this can either be “exact” loop data or estimated flow data from floating vehicle data.

|                  | Curved FD | 10% error | 20% error | Linear (no flow) |
|------------------|-----------|------------|-----------|-----------------|
| Time to detection| 25,3 s    | 25 s       | 25,5 s    | 19,2 s          |
| Success ratio    | 98%       | 98%        | 98%       | 98%             |
| False alarm ratio| 13%       | 13%        | 12%       | 17%             |

4.2. Prediction

Since cross-validation is used to determine the results, the ANN’s were trained with seven out of the eight available peak period data subsets, while the remaining peak period subset was used to validate the predictive value of the trained ANN. Each time one peak period data subset was not used to train the ANN and could consequently act as validation data subset. After eight cycles the total validation of the ANN with the particular architecture and prediction horizon was concluded.

As stated before Huisken (2006) already showed that a prediction of queues, using ANNs, with real traffic data is possible 5 minutes in advance with a false alarm rate less than 5-8%. Based on this PHD study we already know that the challenge for prediction is feasible.

A closer look at the results shows a reduction in speed below the threshold of the definition of a queue, but not for long enough to meet our criteria; in fact, sometimes the speed picks up again. We conclude that the ANN is able to identify onsets of queues, even minor ones, but that – for now – the algorithm is too sensitive. We expect that when we train the ANN on real data as done within the current project, this sensitivity might become less of an issue, since real data incorporates more noise. ANNs are well-known to give good results under noisy circumstances.

5. Discussion

5.1. Innovations

The approach combined various state of the art scientific methods to detect and predict incidents. As far as we know, these methods in combination have never been used in practice or at this scale and for this purpose. This also means that there were various challenges which we addressed as follows:
Various decisions were required for which there are not necessarily default values available. This varied between the level of aggregation of data, e.g. segment length and time step length and required thresholds e.g. within the fuzzy state estimation and incident detection. Therefore, the estimation of default parameters was an important part of the research.

The data and data availability was a challenge. To be able to train the algorithms in the initialization step, sufficient training data is required. In addition, the algorithms should also cope with missing or erroneous data. Partly this can be solved by using predicted, historical data or data of upstream or downstream segments, but this certainly influences the performance or even the ability to detect or predict incidents. Next, is the required penetration of floating vehicle data to be able to estimate accurate segment data and the fact that we only know speeds based on floating vehicle data for segments without loop detection.

There were issues which are not comprehensively addressed in scientific research but important for application, such as dealing with intersections, diverge and merges. These were especially relevant for the prediction of the future segment data and future traffic state.

The required target detection and prediction intervals determine the available time for the controls. Although time may be saved by parallel/distributed computing, testing the calculation times resulted in adapting the approach. This was achieved by using very simple decision rules for the algorithms.

The evaluation and therefore adjustment of parameters was a challenge, because there are no standard rules how to update these parameters (i.e. which decision rules to use). The determination of these decision rules is part of the research, which has started but will be extended in the current phase of the project.

There is little published experience in training a neural network for predicting incidents. The outcome of this research will contribute to the general knowledge.

5.2. Suitability of the results

We have developed the academic ideas further into an environment in which we have managed to achieve the majority of the goals for this phase of the project with respect to the detection and prediction of queues and incidents. Where we have not been able to achieve the objectives (the prediction of emergency situations such as accidents and breakdowns), this is due to the random and rare nature of these events, rather than methodological shortcomings.

We have started to develop practical parameter values that should enable the practical implementation of the techniques using standard data sources that are generally available on Dutch and English Motorways: loop detectors providing flow and speed data. We have also shown that the methods are suitable where speed data (i.e. floating vehicle data) is available, but not flows. This gives us confidence that the module can be and should be developed into a prototype and tested in the next phase of the project in a practical context, linking the module to real-life data and testing the module’s ability to predict queues and incidents with acceptable confidence and within acceptable computation times.

References

Beek, P., Suijs, L., Wismans, L., Huisken, G., Van Vuren, T., Cornwell, I. (2014). Identification and prediction of incidents by learning algorithms, Deventer, The Netherlands.

Dougherty, M.S., Kirby, H.R. and Boyle, R.D. (1993). The use of neural networks to recognize and predict traffic congestion, Traffic Engineering and Control, 34 (6), pp. 311 – 314. Greenshields (1935). “A study of Traffic Capacity”.

Huisken, G. (2001). Short-term forecasting of traffic flow on freeways, in: Proceedings of the 9th World Conference on Transportation Research, Seoul, South Korea.

Huisken, G. (2003). Short-term forecasting of traffic flow on freeways, System Analysis Modelling Simulation, 43 (2), pp. 165 – 173.

Huisken, G., Inter-Urban Short-Term Traffic Congestion Prediction, T2006/8, December 2006, TRAIL Thesis Series, The Netherlands.

Huisken, G. and Coffa, A. (2000). Short-Term Congestion Prediction: comparing time series with neural networks, IEE Conference Publication 472, pp. 60 – 63.

Huisken, G. and Van Maarseveen, M.F.A.M. (2000). Congestion prediction on motorways: a comparative analysis, in: Proceedings of the 7th World Congress on Intelligent Transportation Systems, Turin, Italy.

Kaysi, I. Ben-Akiva, M.E., and Koutroupos, H.N. (1993) Integrated approach to vehicle routing and congestion prediction for real-time driver guidance, Transportation Research Record, 1408.

Kerner, B. S. (2003). "Three-phase traffic theory and highway capacity." Physica A 333: 379-440.
Kerner & Rehborn (1997). “Experimental Properties of Phase transition in Traffic Flow.” Physical review Letters Volume 79, number 20.

Kerner, B. S., H. Rehborn, et al. (2004). “Recognition and tracking of spatial-temporal congested traffic patterns on freeways.” Transportation Research Part C 12(5): 369-400.

Leclercq, L., J. Laval, et al. (2011). “Capacity Drops at Merges: an endogeneous model.” Transportation Research Part B 45(9): 1302-1313.

Treiber, M. and A. Kesting (2013). Traffic Flow Dynamics, Springer.

Van Arem B., Kirby, H.R., Van Der Vlist, M.J.M. and Whittaker, J.C. (1997). Recent advantages and applications in the field of short-term traffic forecasting. International Journal of Forecasting, 13 (1), pp. 73 – 85.

Wismans, L.J.J., E. de Romph, K. Friso, & J. Zantema (2014). Real-time traffic models, decisions support for traffic management. Procedia Environmental Sciences, 22, pp. 220-235, doi:10.1016/j.proenv.2014.11.022

Wismans, L.J.J., Suijs, L.C.W., Krol, L., Berkum, E.C. van (2015). In-car advice to reduce negative effects of phantom jams. In proceedings of the 94th annual meeting of the Transportation Research Board, Washington D.C.