Proceedings of the 2007 IEEE Symposium on
Computational Intelligence and Data Mining (CIDM 2007)

# Hazard Situation Prediction Using Spatially and Temporally Distributed Vehicle Sensor Information

Thomas Schön
Faculty of Computer Science
and Mathematics
University of Passau, Germany
Email: schoent@fmi.uni-passau.de

Bernhard Sick
Faculty of Computer Science
and Mathematics
University of Passau, Germany
Email: sick@fmi.uni-passau.de

Markus Strassberger
BMW Group
Research and Technology
Munich, Germany
Email: Markus.Strassberger@bmw.de

***Abstract*— Driver Assistance Systems are the key technology to improve traffic safety and lower the number of deadly accidents. Direct communication between cars will further enhance this field of driver safety. In the context of foresighted driving, Bayesian Networks can be used to determine a traffic situation at the current position of a car. Communicating this awareness for the current time and position will help other traffic participants. However, situations change dynamically and cars cannot trust all the information provided by other cars over time. Reasoning with this information is difficult as Bayesian Networks cannot use spatial and temporal data in an appropriate way. This article outlines the spatial and temporal problems in predictive driver assistance and demonstrates how they can be solved by considering spatial and temporal influences by applying weighting techniques. The pre-processed information is utilized by a Bayesian Network for further refinement. Thus, the proposed approach enables the detection and correct prediction of traffic situations. The approach is evaluated by predicting hazardous rain fields in a car by means of information received from other cars.**

## I. INTRODUCTION

Inter-vehicle-communication offers new opportunities to the world of cars. Especially active safety systems will greatly benefit from car-2-car communication. Being aware of a hazard situation, a car may notify other cars in advance [1]. Thus, cars having access to remote information are able to foresee a hazardous situation and warn the driver accordingly in time. In order to accomplish this task, the following three steps are necessary:

- Detection of a local hazardous situation (i.e., at the current position of the car) without interaction of the driver, utilizing only widely deployed on-board sensor systems.
- Exchanging corresponding information with other vehicles by means of car-2-car communication techniques.
- Predicting remote hazardous situations along the route by means of that received information.

Research concentrates so far on wireless information exchange between vehicles [1], and on detecting hazardous situations with on-board sensor systems [2]. For the latter, one of the main challenges is to deal with temporal aspects of sensor information and uncertainty. We showed in our previous work [2] that in many cases Dynamic Bayesian Networks (DBN) are a suitable means to infer both the current driving situation and its causes from standard on-board sensor systems.

However, in order to effectively inform the driver about the situation to come, the future driving situation of a vehicle has to be forecasted. Consolidating a variety of observations from different vehicles at different locations can thereby significantly improve this forecast. This consolidation process raises the following additional research issues:

- By sharing individual knowledge, vehicles get access to sensor information from other cars. However, different and potentially conflicting observations of those vehicles have to be fused to a consistent model of the individual, future driving situation.
- The environmental conditions and, thus, the individual observations are subject to a continuous change. Therefore, temporal aspects of situations and observations have to be taken into account.
- In particular critical weather conditions such as areas of heavy rain or fog have a spatial extension. Consequently, the critical conditions are not limited to the reported locations. Thus, spatial influences and dependencies have to be considered.

A stable prediction of hazards using observations provided by other vehicles requires the consideration of all available evidences. These evidences have properties which distinguish them by their relative position to each other (space) and their timeliness. In a nutshell: It makes a difference for a prediction whether an evidence is old and far away or new and near. Bayesian Networks (BN) cannot handle those spatial interdependencies properly. They can only reason upon given evidence. They cannot distinguish if evidence comes from north or south of a current position without introducing two separate nodes for those positions. They also cannot handle evidences grouped in the west of a position or following a route from west to east without introducing new nodes. However, inserting a large amount of additional nodes worsens the accuracy.

Relative spatial information concerning evidences of one kind is a key issue for a precise recognition of hazard situations. However, information related to different kinds of hazard situations influence each other as well. For example, with a rain evidence the probability of hydroplaning rises, too. BNs are predestined to deal with these causal dependencies.

In this article we present a fast and accurate method taking into account the specific spatial and temporal properties of distributed sensor information. All the fused evidences are fed into a BN again. Consequently, the accuracy of the resulting prediction increases.

After presenting related work in Section II, we will present a new solution for spatio-temporal information fusion, discussing various problems and their solution in detail (Section III). Section IV shows our experimental results and finally we summarize our conclusions in Section V.

## II. Related Work and Prelimiaries

Research in inter-vehicle-communication became very active recently. This is a result of national and international projects [3], [4], [5], [6], focusing on enhancing driver safety and traffic throughput. However, so far research focused on communication issues and not on driver-assistance.

Concerning the forecast of a situation, similar work was published by Kennet et al [7], who were able to predict weather conditions in a sea bay using Bayesian techniques. But this scenario was not mobile and the forecast used almost complete data to predict statically for one day at one place, i.e., an influence of an evidence is not moving and can be modeled in a BN as a single node. This node represents a position in the bay. However, in the application domain of vehicles, evidence is randomly distributed in the plain and may be incomplete and sparse.

Kriging is a commonly used technique to estimate spatially dependent aspects. It is capable of using spatially accumulated data as well as sparse data [8]. This is accomplished by clustering and weighting the inputs and interpreting occurring patterns in the data with an established variogram. In fact, the approach presented in this article is similar to kriging. However, the specific characteristics of foresighted vehicle safety impose additional requirements. First, temporal dependencies have to be considered, too, because a prediction of rain heavily depends on timeliness. In addition, probability and trust of different evidences must be regarded. Clustering is not used in our approach because traffic situations may change rapidly and do not behave as the data kriging was originally intended for (mining purposes). Finally, the presented approach accomplishes a fast forecast for a lot of different situations along the route while the car is driving. Comprehensive calculations for all possible traffic situations and for every bit of the route are not possible.

Bearing that in mind, we adapted our prior work from [2], [9] which continued to investigate aspects from [1]. Detection of a hazardous situation is possible by using BNs which only use sensor data or data derived from them. Thereby we do not rely on Dynamic Bayesian Networks (DBN), because of their need for memory and calculation time. In those networks, the time series expansion of sensor data is encoded in the structure itself. Building that structure and rebuilding it in each time slice is costly. For this reason, we introduce an additional type of BN nodes, which does not handle the direct output of sensors but the change underlying the data, i.e., the first derivation of a time series. Figure 1 delineates the structure of the BN, comprising two nodes for speed: one for the current speed and one for the change of speed. As a consequence, the propagation of temporal influences is much faster compared to DBN. On top, exploiting higher derivations of the time-series increases accuracy.

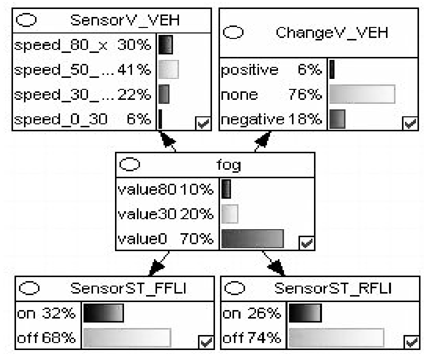


Fig. 1. The intern net for fog detection. Input nodes: current speed, change of speed, front fog light, and rear fog light.

For each kind of hazard which can be detected, there exists a so called *internal* BN, which uses only local data from a car's own on-board sensor systems. This information can be transmitted to other cars. In order to derive a stable picture of the situation to come, it is necessary to process and consolidate all received evidences. However, it is not sufficient to only compute the mean of all evidences. Instead, interdependencies of observations and evidences have to be considered. Further, both the spatial position of evidences relative to each other and the temporal decay of the timeliness of an information are important issues. BNs lack the possibility of modeling spatial dependencies. If there is an evidence for rain in one place, it is more likely to rain in a nearby location, i.e., the probability of rain increases for a particular location, if there are rain evidences close to that location. The same evidence can be far away and in that case the influence must be small. This form of influence is mutual: If one place is adjacent to another it is influencing this place, the same way as it is influenced by it.

In addition, according to the Bayes theory a BN must be acyclic (a directed acyclic graph – DAG). That property contradicts real requirements as mentioned above. Influences in a DAG cannot be mutual, i.e., cyclic (see Figure 2). Thus, it is only possible to model influences from one position to another but not vice versa.

Even with probabilistic paradigms which do not require directionality such as Markov models, such spatial dependencies cannot be modeled appropriately. Directionality makes it difficult to predict whether or not there is rain at a certain location. It makes a difference if rain observations are only reported in the north-west of a vehicle's location, or both in the west and east of it. In the first case, a rain area may be still far away. In the latter case, it is more likely that the rain area has a greater dimension and is overlaying the vehicle's

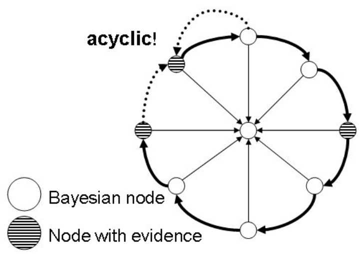


Fig. 2. Imposible spatial interconnection.

route. Thus, those situations have to be distinguished.

Dividing the area around the desired point of prediction into sectors and dividing these sectors in different ranges will provoke loss of reality and the resulting BN would be huge. This is due to the fact that for each of those areas there must be a node in the BN. If an evidence is observed for an area, evidence for the corresponding node would be set. This spatial unfolding would be required for each kind of hazardous situations, such as rain, fog, hydroplaning, blockade, or reduced friction. In addition, those nodes need to be densely interconnected. Figure 3 shows the first step of a spatial unfolding.

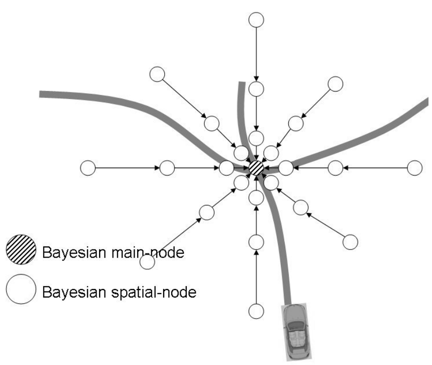


Fig. 3. Spatial unfolding for one kind of information without radial interconnections.

The resulting network cannot be interconnected in all desired directions. It may be convenient to link only dependencies from outer nodes to the center, but again, cycles would not be possible. Therefore, the influences between some evidences is lost, depending on their relative location.

In addition, most of the nodes would remain without evidence or evidences would fall more than once into the same area. Therefore, only one of the evidences could be put in one node. However, the probability and the plausibility for an evidence should increase if there are many evidences supporting each other. On the other hand, if evidences contradict each other, diminishing those properties would be necessary.

The fact that only one evidence per node can be treated leads to the problem that given information is not used. In order to solve this problem for the temporal aspects of evidences, usually DBN are used. These networks clone the whole network in each time slice and then connect these clones from one time slice to the next, realizing a temporal unfolding (see Figure 4).

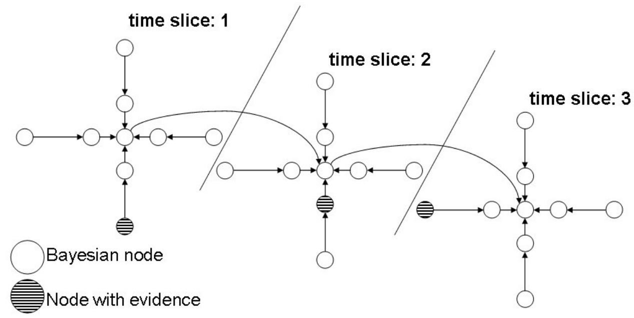


Fig. 4. Temporal unfolding for one node.

Together with the spatial unfolding, the temporal unfolding leads to even bigger networks which are densely connected, but have only few evidences. Thus, propagating the probabilities within the BN depends on only very few evidences. As a consequence, the result mainly depends on the already given a-priori probabilities in most of the nodes.

## III. New Method for Spatio-Temporal Fusion of Vehicle Sensor Information

To avoid the above mentioned problems of BNs in the context of spatial and temporal sensor interpretation, we propose a different concept. Still, this concept relies on the local detection of hazards and critical driving conditions at the car's current location and time. Doing so, space- and time-dependent characteristics are not required. Results are shared with all vehicles within a certain area and saved within the receiving vehicles. The received information is interpolated depending on its spatial and temporal properties. These interpolations are used to estimate the future driving situations of the cars, i.e., whether or not a driver has to be warned of hazards or critical environmental conditions. In order to take the interdependencies of evidences into account, the results of this interpolation are refined using so-called *external* BNs. These external networks typically have the same structure as the internal BNs that are used to derive the current driving situation. The terminology of external BNs should emphasis the fact that the consolidation of observations is done with respect to a remote location.

It is possible to predict moving hazards noticed and announced by moving vehicles with evidences, although we do not know in advance in what quantity those evidences exist.

### A. Local Detection of Hazards

BNs are used to detect hazard areas such as rain at the current position at the present time. This prediction does not require spatial information. Every car has its own sensors whose outputs can be used to feed a small, intern BN. Only direct observations of the current situation are processed. Therefore, the spatial problems of BNs do not affect this calculation.

For each critical driving condition or hazardous event that should be detected there exists one dedicated BN. The structure of those BNs is made up of one input node for each sensor involved in the corresponding recognition, and exactly one output node representing the hypothesis concerning the targeted condition or event. The output node is directly linked with the input nodes. In this way, evidences are not thinned out too much over numerous layers of nodes. Although this structure does not reflect a detailed view of real-world dependencies between causes and effects, it is still a valid model, assuming that casual dependencies can be automatically generated from a large set of recorded real-world trace data.

As justified before, we avoid temporal unfolding in such a network introducing special input nodes which do not use the sensor data directly as input. Those nodes will use the change of data input and therefore catch the temporal development of a situation. So, after breaking from 100 km/h to 80 km/h, in such a node we will not set "speed = 80" but "speed = -20", thus realizing the first derivative of the speed function. It is also possible to use the second derivative to get an even more accurate temporal model. As explained before, this is similar to the concept set out in [9].

### B. *Sharing of Local Knowledge*

In order to provide access to remote vehicle sensor information, local observations have to be communicated to other vehicles using wireless communication links. As already mentioned in Section II, there is a variety of research activities in the field of vehicular ad-hoc networks. Note that the prediction of the future driving conditions by means of distributed vehicle sensor data is basically decoupled from the specific communication technology used in the vehicles. The latter is beyond the scope of this paper. However, the resulting accuracy of the prediction depends on the both the specific message content and the number of reported and received evidences about specific observations.

*Message Content:* The impact of hazard events and critical road conditions and, therefore, the importance of related information for other drivers is event-specific and differs to a great extent. As a consequence, the type of the observed hazard must be transmitted. Furthermore, a time stamp and GPS coordinates must be included in order to calculate the time which has passed (= temporal distance) and the distance between the reported observation and the location it influences. In addition to these basic properties, a detailed description of the specific observation is necessary comprising the following parameters: First, a quantification of the intensity of an observed driving condition is necessary. This is in particular important for critical weather conditions. Obviously, the metric quantifying the intensity must be standardized. Rain could, for example, be quantified by liter per hour and square-meter, or, if sufficient, in discrete values such as *none*, *light*, *medium*, or *hard*. Second, the probability of an observation must be transmitted. Obviously it makes a difference whether a hazard event has a probability of 100% or only 40%. In that context we want to stress the difference between probability and intensity. In a Bayesian node, the state for, e.g., *temperature = -20* can be most likely with 80%. Third, it is also important for the inference to measure how the conclusion was reached. It makes a difference whether the car concludes *hard rain*, but is equipped only with an active sensor for wiper speed, or it concludes the same by means of an additional moisture sensor. Therefore, also a value that expresses the trust into a conclusion must be transmitted. For simplicity, we calculate the trust as follows: A BN with 4 input nodes of which only 3 have evidence would have a trust of 75%. Note, that this is just a straightforward estimation of trust. The corresponding trust value could be also derived using Dempster-Shafer theory [11] instead of BNs, for instance. All these parameter values can be derived from a BN. As mentioned, the type of the hazard to be detected corresponds with a specific output node. The probability corresponds with the state of the output node. Finally, the trust into a result is the percentage of input nodes that have evidence. Altogether, a message consists of:

- type: the type of hazard (e.g., rain, temperature, blockade),
- intensity: the amount which has been calculated,
- probability: the probability for this intensity in the Bayesian node,
- trust: the plausibility of a conclusion,
- position: GPS coordinates,
- time: global unique time from GPS signal.

For the following considerations we assume that each message contains all these informations.

*Limited Bandwidth:* The available bandwidth of the wireless communication channels is obviously limited. Therefore, not all raw sensor information can be shared among all vehicles within a certain area. In addition, sensor systems have typical update rates in the order of milliseconds, making it impossible to communicate every update event. Instead, only conclusions about hazards or road conditions are transmitted. As a consequence, also overall computing time is saved, because conclusions from sensor data are only computed once in a car and will not have to be computed in every car receiving the raw sensor data. Furthermore, we cannot assume that our application is the only one which uses the wireless capabilities of the car. Instead, the wireless communication capabilities of vehicles will serve as a basis for a great variety of different applications. A further restriction of the amount of data which is broadcasted is necessary for that reason, too. Assuming an ad-hoc network with a non-optimal and non-collision free medium access strategy such as IEEE802.11 wireless LAN as suggested in [3], [4], [5], [6], the critical parameter is the number of messages rather than the average message size. Therefore we assume that each car continuously calculates the above mentioned parameters (intensity, probability, trust) for all types of hazards, but only send them to other vehicles if the corresponding values change significantly. Hazard areas such as rain move and change their intensity and extension. Thus, a transmission when the own values change is not enough. All cars passing through such an area would indeed encounter different circumstances as expected, but would not send this new information as their own sensor data might not have changed. Therefore, an information is also sent if the forecast for a certain location differs from the actual local observation when the vehicle reaches the reported location. This has the effect that over time, newer information are broadcasted. Due to the dynamic of situations, newer evidences obviously have greater influence on a prediction as old evidences. Therefore, they enable an implicit update strategy leading to a more accurate prediction.

### C. Storage and Retrieval of Messages

In our implementation, the spatially distributed data is stored using a quadtree data structure. This data structure has superior performance properties, as it can identify data by its coordinates. A quadtree divides an area into four quadrants: northwest, northeast, southeast, southwest. These quadrants correspond to nodes of the quadtree. So, a quadtree is a four-ary tree.

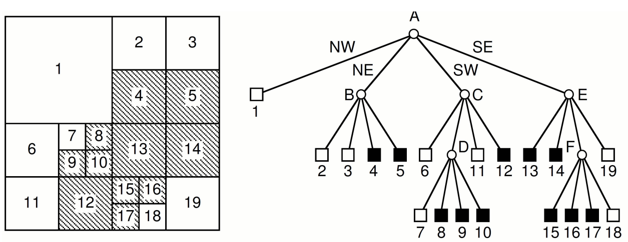


Fig. 5. A quadtree of an area with some data [12].

As Figure 5 shows, a node has only children if it is not empty and not completely filled. But if it contains data in a subarea, the node containing those data is divided into quadrants again. As a consequence, only those children have successors that have more than one item to store.

### D. Situation Prediction 1: Weighting Evidence

The availability of remote evidences from other cars finally enables the effective individual in-vehicle prediction of hazards and critical road conditions on the vehicles' routes. Therefore, the effects of spatial and temporal distance (Figure 6) have to be considered.

Evidences obviously have more influence if spatial or temporal distance is low and vice versa, and less influence if the evidence is old or far away. In addition, there might be other parameters with similar influences on the prediction, such as for example trust or plausibility. In general, the desired behavior can be achieved by weighting an evidence depending on multiple, related properties. Basically, the value of the evidence is multiplied with different weighting factors, summed up, and finally normalized by dividing this sum with the sum of all weighting factors. Therefore, the weighted *average* value is dominated by those evidences with higher influence:

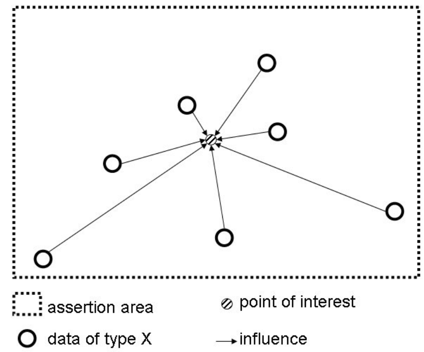


Fig. 6. Spatial positions create different influences.

$$weightedVal. = \frac{\sum_{i=0}^{\#evid.} \sum_{j=0}^{\#factors} weightFactor_{i,j} \cdot evid.Val._i}{\sum_{i=0}^{\#evid.} \sum_{j=0}^{\#factors} weightFactor_{i,j}}.$$

There exists a weighting factor for all properties of the evidence (Section III-B) which need to be weighted. Therefore, there must be a weighting factor for time, distance, probability, and trust. We want to emphasize that there is no special difference in spatial or temporal distance. Both are represented as influence factors such as probability or trust. Even though, in our simulation those two factors differ in that the space factor is calculated linearly and the time factor non-linearly (see below).

Choosing the weighting factors and their calculation can be done easily and individually as needed. The factors for probability and trust are the values themselves. The specific function for the other factors is dependent on the characteristic dynamic of a certain hazard type. Two of them have proven to be very useful for the reason of their simplicity and effectiveness. First, linear interpolation: This interpolation decreases the factor of influence slowly but linearly from the current point of interest (the position we want to know the value for) to the horizon of the maximum influence range:

$$linearFactor = 1 - \frac{distance}{max.distance}.$$

On the other hand, a non-linear behavior might be interesting as well, depending on the influence to be modeled. That is, evidences from nearby locations have disproportionately high influence compared to evidences rather far away. This is accomplished with the reciprocal value of the distance:

$$asymptoticFactor = \frac{1}{distance}.$$

The curves of these two weighting factors are shown in Figure 7.

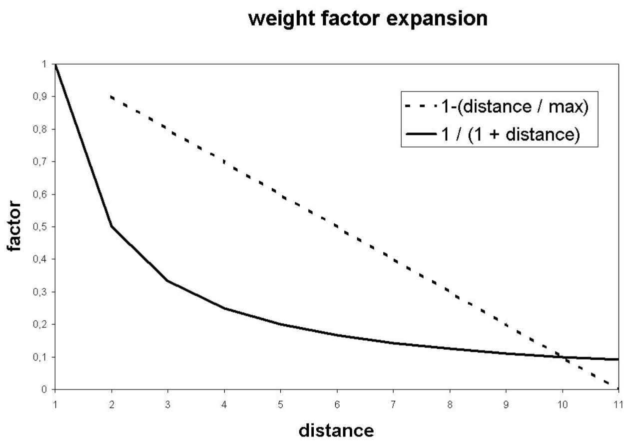


Fig. 7. Curves of the weighing factors.

Important for the weighting with these factors is to multiply them with an additional factor to match the range of the evidence value. As our evidences are within a range from 0 to 100 the weighting factors must be in that range as well. So they are multiplied by 100. Without this measurement, the weighting by values from 0 to 1 is not effective.

The following short example demonstrates the behavior of the weighting process by using only one weighting factor assuming two evidences: One with the intensity of 40 and probability of 80% and the second evidence with the value 80, but only 40% probability. With linear weighting we obtain $53,\overline{33}$. As intended, the value drifts towards 53, which is nearer to 40 because 40 had the higher probability.

### E. Situation Prediction 2: Inter-Hazard-Influence

Spatial and temporal characteristics of separated hazard types or critical road conditions are considered by calculating the weighted value for any type of hazard or condition at a certain location. However, interdependencies of the various types of hazard observations (such as rain, snow, hydroplaning, temperature) are not considered so far. For example, rain and hydroplaning can be detected using their evidences, but if a car has only rain evidences, it has to presume a certain probability for hydroplaning as well.

As the purified values resulting from the interpolation do not contain direct spatio-temporal characteristics, we can feed those weighted values again into a BN. This enables a selective cross-attribute refinement of the results already obtained. The weighting process already accounts for the spatial and temporal influences. As a consequence, the refinement can be treated in a way very similar to the local hazard detection (see Section III-A). Note that not all types of hazards have the same magnitude of influence. For example, spatial influence of rain evidences is typically larger than evidences of a street blockade. Therefore, the specific weighting factors depend on the specific type of hazard. Figure 8 delineates a spike through the layers of evidences taking the corresponding value of each layer for the related hazard type.

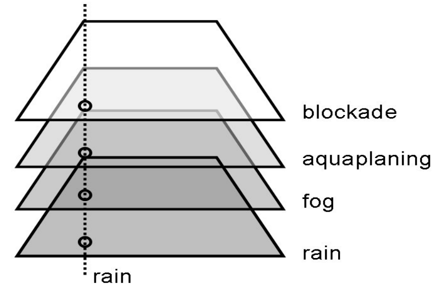


Fig. 8. Spike through the layers of hazards.

The values extracted from the layers are put again into an *external* BN which uses only external information as input (remember that the internal networks use only sensor data from within the car). Figure 9 shows an exemplary external rain BN. Note that this time the BN comprises two *rain* nodes. On the one hand, the lower node is an input, where the interpolated value for rain is applied. On the other hand, the upper node is the output node, determining the refined result.

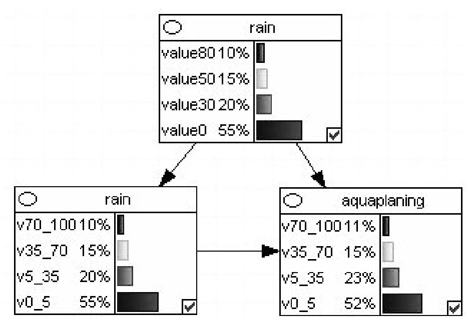


Fig. 9. Extern BN for rain prediction (two rain nodes for input and output).

The proposed three-step approach of interpolation and refinement enables both the integration of spatial and temporal effects and the interdependencies of individual observation of different hazard types and road conditions.

## IV. Experimental Results

In order to evaluate the effectiveness of the proposed approach, we simulated an area of approximately 3 km $\times$ 4 km in the center of Munich. In this area, 200 cars are equipped with communication and sensing capabilities. These cars are moving along the streets following the Krauss mobility model [13]. Additionally, road conditions and hazards such as rain, fog, or a street blockade (as result of an accident) can be introduced into the simulated scenario. Those conditions affect the on-board sensors of cars in a specified manner. The hazards can be configured to move randomly, in a certain pattern, or not at all.

The cars equipped with the proposed system can predict a hazard situation in the course of the drive. The distance between a car and a potential hazard has no influence on the prediction, as the prediction relies only on received remote evidence and not on own sensor data.

Thereby the accuracy of the achieved prediction decreases with decreasing availability of remote observations. This is because the hazard area may move meanwhile the predicting car approaches. In a worst case scenario a car may predict the future road condition using only a few and rather old evidences (of course, then with little trust and probability, which is accomplished by the interpolation process). It should be noted that the cooperative system as described can never predict a critical situation that has not been encountered by any means before. However, sharing new information with other cars significantly improves the safety for following cars.

We evaluated the proposed three-step prediction by simulating a variety of different scenarios. For example, we used a moving rain field with sharp edges, which means that a sensor would jump from a value of 0 to a value of 50 at the moment the car enters that rain field. We also simulated scenarios with seamless transitions where the sensors report values increasing towards the center of the rain field, but with a slight and random discrepancy (simulating white noise and imperfect sensors). We also equipped cars with defective sensors that deliver wrong information. Nevertheless, the prediction works well, depending on the number of locally available evidences in the predicting cars. This in turn depends on the number of vehicles that pass and encounter a critical condition per time, the communication range, and the available channel capacity.

Figure 10 shows the prediction ahead of a car. The smooth line is the actual value for rain at the given position ahead. While the car passes through the rain area it predicts the rain values ahead on the road. The stepped line indicates the prediction compared to the real value at the same position. In this case we predicted the situation 20 seconds ahead of the car. This forecast horizon is considered sufficient to warn the driver appropriately in time. The shown values represent the actual and the predicted values at a location 20 seconds ahead of the car along its most probable route. While the car is moving, the location of forecast also moves ahead.

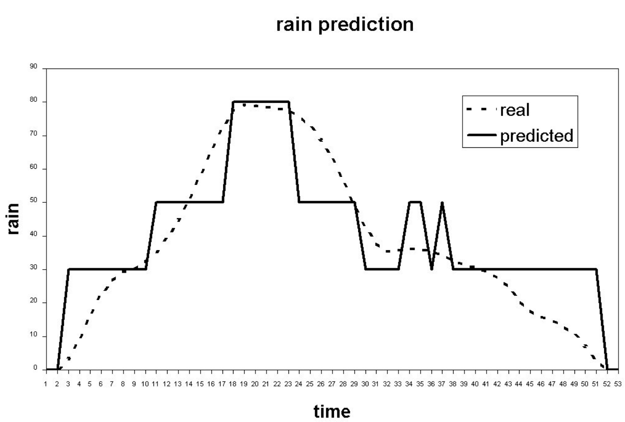


Fig. 10. Prediction of the rain values far ahead of a car.

We see that the predictions are very close the real values. The steps are due to a discretization of the rain values, as the rain prediction only foresees discrete values such as *none*, *light*, *medium*, or *hard*, which we consider sufficient for the task of driver notification. These discrete values are encoded in the simulation as scalar values *0*, *30*, *50* and *80*. These numbers represent sections of all possible outcomes. That is, 0 stands for any value between 0 and 5. The other states represent the intervals 5 – 35, 35 – 70 and 70 – 100, respectively. We encoded rain as a value between 0 and 100, but a different encoding is possible. To minimize the error produced by this discretization we can introduce a bigger number of states with smaller intervals. It is not considered an error, if the real value at a position is 65, for instance, and the prediction equals 50, as this is only a matter of discretization in that case.

In the following, we focus on real (not discretization) errors. Due to the dynamic movement of rain fields, the prediction may still rely on comparable, old evidences. Therefore, it differs from the real value by at least one state. But as the rain field moves slowly and there are enough cars to update the position frequently, these changes are noted. In that way the cars are feeding the weighted interpolation for the external BN with newer evidences. Those evidences will have more influence in the prediction as the old evidences.

In a different set of scenarios the number of sensors installed in a car was reduced. Thus, the installed internal BNs could not use most of their input nodes, which led to less trust in the observation. In such scenarios the accuracy of the prediction did not decrease significantly. The corresponding graph is almost equal to that of Figure 10. This can be explained by the fact that the trust value is taken into account as weighting factor for the weighted interpolation. Vehicles which do no have all sensors installed put a low trust in their own findings. Therefore, evidences from vehicles with many sensors and therefore higher trust in their observations are given more weight in the interpolation. Although if there are only few evidences in the internal network, the network would detect hazards with an intensity based on the sensor data available. Only accuracy will degenerate depending on the missing sensors. As long as there exists at least one evidence, the prediction will work. This can be explained by the direct feedback of evidences. In the proposed network, the input nodes are directly linked to the targeted output node. Thus, already one evidence influences the outcome.

In another set of scenarios we removed all internal networks for hydroplaning detection. With this step, no hydroplaning evidences where shared among the vehicles. Nevertheless, the cars predict an hydroplaning hazard based on the other kind of hazards for which evidences are given, showing the benefit of external BN. Even if there are no direct evidences for a situation, rain evidences, for example, are influencing the hydroplaning probability and intensity.

As mentioned before we could deteriorate the prediction by moving the hazards faster and providing less or false evidences. But with additional parameters this could be compensated. The moving direction of a hazard could be modeled as parameter and take influence in a prediction as well. We were able to adjust parameters such as the maximum validity of an evidence or its range of influence to enhance the prediction in a given set of scenarios. It seems possible that these parameters could be adjusted automatically and dynamically during a simulation (and in practical use).

In general, faulty sensors or faulty information introduced by attackers decreased the accuracy of the forecast most drastically. Basically, there exist two classes of false detection. First, the presence of critical road condition is not observed. Second, a critical road condition is detected although no such situation exists. A vehicle would, for example, forecast a rain field based on the remote evidences of other cars. But upon arriving this vehicle would not detect any rain, because of a faulty rain sensor. Thus, the car interprets this circumstance as if the rain had stopped, and transmits this (in this situation) false information. If now the car puts a high trust in this, the prediction of the other traffic participants would be influenced negatively. The same occurs with attacks, where an attacker consciously transmits false information. However, this kind of misinformation can also be compensated due to the cooperative characteristic of the overall system. Other cars will again detect the difference in prediction and received evidence and, therefore, correct the information. Note that we assume that the majority of participating vehicles have neither faulty sensors nor are they an attacker to the system. Also, the influence of misinformation can be reduced if a certificate-based reputation system is used. Doing so could influence the system's internal trust (plausibility) in an evidence received from other cars by adjusting the transmitted value according to the reputation of the originating node.

## V. Conclusion

The cooperative sharing of individual probe vehicle sensor information (probe data) among vehicles enables access to remote sensor information and, therefore, enables effective foresighted driving. Thereby, one of the main challenges is the consolidation of this spatially and temporally distributed vehicle sensor information. Although BNs perform poorly using spatial information and cannot handle sparse temporal information in a densely connected network, they can be used by converting the spatial and temporal characteristics into the evidences itself. In particular, spatial and temporal effects can be taken into account using a multi-dimensional weighted interpolation. We therefore proposed a three-step approach. First, the current situation at the current vehicle location is detected using BNs to handle the inherent uncertainty. Critical observations are transmitted to other cars in the affected area. Second, a forecast concerning the road conditions on the vehicles' routes is accomplished by consolidating the available remote sensor information from other vehicles. In order to gain knowledge from locations on the route where no observations are available, the spatial and temporal influence of certain observations, as well as its trust and probability are interpolated. Third, the calculated values are fed into a BN again to refine the conclusions, taking into account the interdependencies between different hazard events and observations.

We showed in a variety of simulated set of scenarios that the resulting prediction mirrors the simulated conditions very well. Thereby, the accuracy of the prediction is dependent on the number of available observations, their accuracy, and trust. However, even with few equipped vehicles and available sensor systems the approach performed well in most of the cases. Using additional information, for example from third party remote services such as weather forecasts, may further increase the accuracy.

## References

[1] T. Kosch, "Local danger warning based on vehicle ad-hoc networks: Prototype and simulation," in *Proceedings of the 1st International Workshop on Intelligent Transportation (WIT 2004)*, Hamburg, Germany, 2004, pp. 43–47.

[2] M. Reiß, B. Sick, and M. Strassberger, "Collaborative situation-awareness in vehicles by means of spatio-temporal information fusion with probabilistic networks," in *Proceedings of the 2006 IEEE mountain workshop on adaptive and learning systems*, Y. Motai and B. Sick, Eds., Logan, Utah, 2006, pp. 189–194.

[3] NOW, "NOW project homepage." [Online]. Available: http://www.network-on-wheels.de/

[4] INVENT. [Online]. Available: http://www.invent-online.de/

[5] WILLWARN, "WILLWARN project homepage." [Online]. Available: http://www.prevent-ip.org/en/prevent_subprojects/safe_speed_and_safe_following/willwarn/

[6] Car-to-Car Communication Consortium (C2C-CC), "C2C-CC project homepage." [Online]. Available: http://www.car-2-car.org

[7] R. J. Kennet, K. B. Korb, and A. E. Nicholson, "Seabreeze prediction using Bayesian networks," in *Advances in Knowledge Discovery and Data Mining*, ser. Lecture Notes in Computer Science, D. Cheung, G. J. Williams, and Q. Li, Eds., 2001, vol. 2035, pp. 148–153, (Proceedings of the 5th Pacific-Asia Conference, PAKDD 2001 Hong Kong, China).

[8] V. B. Wim and K. Jack, "Kriging interpolation in simulation : A survey," in *Winter Simulation Conference*, 2004.

[9] M. Reiß, "Fusion of Spatio-Temporal Information and Knowledge in Vehicles using Probabilistic Networks," Diploma thesis, Fakultät für Mathematik und Informatik der Universität Passau, 2006.

[10] D. S. Laboratory, "Genie & smile," University of Pittsburgh, 2006. [Online]. Available: http://genie.sis.pitt.edu/

[11] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.

[12] H. Samet, "Data structures and algorithms for multidimensional and metric databases," 2005. [Online]. Available: http://www.cs.umd.edu/class/spring2005/cmsc828s/slides/quad.pdf

[13] S. Krauss, "Microscopic modelling of traffic flow: Investigation of collision free vehicle dynamics," Ph.D. dissertation, University of Cologne, Apr. 1998.