Exploiting Context Information for Estimating the Performance of Vehicular Communications

Miguel Sepulcre and Javier Gozalvez
UWICORE, Ubiquitous Wireless Communications Research Laboratory, Miguel Hernandez University of Elche
Elche (Alicante), Spain
msepulcre@umh.es and j.gozalvez@umh.es

Onur Altintas and Haris Kremo
Wireless Communications Group
Toyota InfoTechnology Center Co., LTD.
Tokyo, Japan
onur@jp.toyota-itc.com and hkremo@jp.toyota-itc.com

***Abstract*— The performance of wireless vehicular communications can depend on multiple context factors, such as the propagation conditions, the traffic density, or the location of communication infrastructure units. This paper proposes and evaluates two techniques that are able to identify and quantify such dependencies, and uses them to estimate the vehicular communications performance exploiting context information. The techniques proposed have been evaluated using real-world traces from a Vehicle to Infrastructure (V2I) IEEE 802.11p measurement campaign. The obtained results show that the number of context factors considered in the estimation process can influence its accuracy, and that not all context factors have the same importance in the estimation process.**

***Keywords*— Vehicular communications, Intelligent Transportation Systems, context-awareness, communications performance estimation.**

I. Introduction

Wireless vehicular networks are expected to improve traffic safety and efficiency. To do so, vehicular communications need to be carefully designed and estimated to guarantee maximum reliability, efficiency and scalability. Vehicular communications can be notably influenced by multiple context factors [1][2] such as weather conditions, propagation environment, vehicular speed, road topology, traffic density, presence of obstacles, location of communication units, or communications conditions (e.g. channel load, spectrum available, etc.), etc. For example, NLOS (Non Line-Of-Sight) propagation conditions resulting from the presence of obstacles can drastically reduce the communications range [3], while the local traffic density influences vehicular multi-hop communications [4].

Context information has been exploited in different areas, in particular in pervasive computing where applications are designed to discover and exploit context information. In the wireless and mobile networking domain, previous studies have proposed to exploit context information to improve communications performance and efficiency, for example, through the selection and configuration of protocols. The study reported in [5] presents a framework for the integration of context information into heterogeneous access and handover management. The study reported in [6] proposes exploiting context information provided by the cellular network to improve the forwarding process in multi-hop cellular networks. In addition to communications and networking context information, vehicular systems can exploit advanced context data obtained from positioning devices, digital maps, on-board sensors and even neighboring vehicles. For example, [7] uses information retrieved from digital road maps to improve multi-hop routing protocols. The authors proposed in [8] to use traffic context information to improve the multi-channel management for single-radio transceivers based on the IEEE WAVE/1609 protocol stack. In [9], the authors propose to control the amount of information transmitted wirelessly using context-based prioritization and re-scheduling techniques that take into account vehicular application requirements.

Previous studies have demonstrated the impact of context conditions on vehicular communications. Techniques to identify and quantify such dependency would hence be very useful to design mechanisms capable to estimate the performance of vehicular communications. The availability of such estimates could help configure and optimize vehicular communications and protocols, thereby improving their reliability and efficiency. Based on this idea, this paper proposes and evaluates two techniques for exploiting context information to estimate the performance of vehicular communications. The proposed techniques are able to deal with the uncertainty related to the estimation process, which is especially relevant in the case of wireless communications. An accurate definition and modeling of the context conditions could require a high number of context factors. Extracting such context factors can have a cost in terms of equipment resources needed, computing power for a real-time processing, and possibly communications overhead for their acquisition and sharing with neighboring vehicles. This study also analyzes the impact of the number of considered context factors on the accuracy of the communications performance estimation, and investigates the relevance of different context factors for such estimation. The relevance is here defined as the degree of influence of the context factor on the accuracy of the communications performance estimation. Therefore, this study also contributes to the identification of the most relevant context factors that should be taken into account in future studies.

This paper is structured as follows. Section II reviews techniques traditionally used for reasoning under uncertainty, and that can be useful for our communications performance estimation objective. Based on the conducted review, the two

978-1-4799-2687-9/13/$31.00 ©2013 IEEE

techniques that seem more appropriate for our objective are selected: artificial neural networks and Bayesian networks. Section III presents the specific context factors and communications performance metrics considered in this study. The selection is based on the availability of empirical communication traces. However, it is important noting that the proposed techniques are not restricted to the selected factors and metrics. Section IV and Section V present the design and implementation of the artificial neural network and Bayesian network here proposed. Section VI evaluates the performance of these two techniques using empirical IEEE 802.11p V2I communication traces obtained in an urban measurement campaign. This section also includes the analysis of the impact of the number of context factors on the estimation process, and of the relevance of different context factors.

II. Estimation under uncertainty

To estimate communications performance metrics (e.g. communications range) based on context factors (e.g. traffic density, number of lanes, presence of obstacles), this section reviews techniques traditionally used for reasoning under uncertainty. A technique used to estimate the performance of vehicular communications requires good scalability properties to be able to handle a potentially large set of context factors and performance metrics. Additionally, the technique needs to be able to deal with the uncertainty that can characterize wireless communications and vehicular environments. Some of the existing techniques that can be used for reasoning with uncertain information are fuzzy logic, probabilistic logic, Bayesian networks, artificial neural networks, decision trees or Dempster-Shafer theory. Their main characteristics and suitability for this study are discussed next.

Fuzzy logic measures the degree to which some event occurs or some conditions exist [10] and is well suited for describing subjective context conditions (congested or non-congested, low/medium/high, etc.). In the transportation area, fuzzy logic has been used, for example, to identify traffic congestion conditions since no clear boundaries exist between types of traffic flow [11], or to detect traffic incidents in a highway [12]. Fuzzy logic could be exploited in our study to quantify certain context conditions that are not clearly measurable or do not have clear boundaries. However, fuzzy logic itself does not provide the means for reliably estimating the communications performance using context information.

Probabilistic logic is based on the concept of probability and can hence associate logical assertions with a probability. Probabilistic logic allows writing rules that reason about events' probabilities. As an example, [13] uses probabilistic logic for an autonomous robot to identify its environment using isolated features detected from images. The application of individual probabilistic rules would not allow modeling the interaction and dependencies among different context factors that are necessary in our study. As a result, probabilistic logic is not suitable to estimate the communications performance using different performance metrics and context factors.

Bayesian Networks (BNs) represent a natural extension of basic probabilistic logic, and are probably one of the most popular formal approaches for reasoning under uncertainty. BNs are directed acyclic graphs, where the nodes are random variables and the arcs between nodes represent causal relationships or dependencies. The main property of BNs is that the joint distribution of a set of variables can be written as the product of the local distributions of the corresponding nodes and their parents [14]. The parents of a variable are the variables with an arch directed to it. The flexibility of BNs for presenting probabilistic dependencies, and the efficiency of existing algorithms to perform inference make BNs a powerful tool for solving problems involving uncertainty. In general, modeling BNs requires first designing the network, which includes the selection of significant variables and the identification of the network structure. A learning process is then required to adjust the conditional probabilities of the BN. Learning capabilities are very interesting in our scenario to continuously refine the performance estimation process as new context information is acquired. Once the BN is designed and the conditional probabilities have been adjusted through the learning process, probabilistic inference algorithms are used to compute the probability distribution for any variable given observations of other variables. The variables of a BN that model our problem would be the context factors and communication performance metrics. BNs could be used in our study to calculate the probability distribution of e.g. the transmission range (a performance metric) given existing context conditions (defined by a set of context factors). In addition, inferences can be performed even when not all variables can be observed or data is missing. This can be useful in scenarios where all context factors cannot always be measured, but still a decision needs to be made.

An Artificial Neural Network (ANN) is a structure comprised of densely interconnected adaptive simple processing elements (called artificial neurons or nodes). The output of each artificial neuron is a function (linear, sigmoid, threshold, step, Gaussian, etc.) of its weighted inputs, known as activation function. The attractiveness of ANNs comes from their remarkable information processing characteristics, mainly in terms of nonlinearity, high parallelism, and learning and generalization capabilities [15]. ANNs have been used to solve problems related to pattern classification, clustering, function approximation (estimate the output of an unknown function based on observed inputs), or optimization [16]. One of the most popular and powerful network architectures for function approximation is the so called feed forward network, or multi-layer perceptron, which is formed by an input layer, an arbitrary number of hidden layers, and an output layer. In multilayer perceptron networks, each node of a layer has a direct connection to all nodes of the following layer. With ANN, the goal of the training process is to find the weight value of each connection that will cause the output from the neural network to match the actual target values as closely as possible. Once the network is trained, the information is presented (or fed) to the neurons of the input layer and propagated to the next layers, until the information is converted into the network output at the last layer. ANNs can be exploited in our study to estimate the vehicular communications performance (considering various performance metrics) using observed context conditions. In this setting, the input nodes of the ANN could correspond to the context factors measured/observed, and the output nodes would represent the communications performance metrics.

Decision trees (also known as hierarchical classifiers) are one of the most popular classification algorithms. A decision tree is characterized by an ordered set of nodes, where each node is associated with a decision function of one or more features. Decision trees can be used for regression analysis to estimate the value of an output function based on a set of input values. ANNs are often compared to decision trees because both techniques can model data that has nonlinear relationships between variables, and both can handle interactions between variables. The main advantage of decision trees compared to ANNs is that decision trees better handle binary categorical inputs. However, only discrete/categorical output values are admitted in the case of decision trees, whereas ANNs have the capability of providing output values for input values for which they were not specifically trained (i.e. generalization).

The Dempster-Shafer theory is a mathematical theory widely used due to its ability to differentiate between uncertainty and ignorance, and its ability to merging multiple pieces of information (i.e. sensor fusion) [17]. The Dempster-Shafer theory would be interesting in our study as a technique to evaluate if sufficient context information has been acquired so that a decision can be performed. However, it does not represent a plausible solution for estimating the communications performance since it is not suitable to represent the existing dependencies among context factors and performance metrics.

ANNs and BNs have been selected in this study to estimate the communications performance using context information. Both techniques are able to express the non-linear relationship that can exist between varied context factors and performance metrics. In addition, they both can learn the dependencies between context factors and performance metrics through their respective training processes. ANNs have been selected over decision trees given their limitation to generalize and their discrete output nature. Fuzzy logic, probabilistic logic and Dempster-Shafer theory cannot be used for the purpose of this study since they do not have learning capabilities and cannot reflect multiple dependencies among context factors and performance metrics. ANNs and BNs have different advantages and disadvantages. ANNs present higher scalability in terms of number of nodes. However, ANNs calculate the output values that best match with the training data with which they have been trained. This implies that for a given context scenario the output of an ANN is fixed. In BNs, the output is not a specific value (in our case, the estimation of communications performance metrics), but rather its probability distribution for a given context scenario. Additionally, ANNs require that all the input nodes are used to calculate the output values. This would limit their use in scenarios in which the vehicle has only partial knowledge of the context conditions.

III. CONTEXT FACTORS AND PERFORMANCE METRICS

The selected estimation techniques are tested in this paper using a large set of empirical data obtained in an urban IEEE 802.11p V2I measurement campaign [3]. As a result, this study is restricted, without any loss of generality, to the context factors and communications performance metrics that can be extracted from the available set of data. The measurement campaign was aimed at analyzing the impact of urban characteristics, RSU (Road Side Unit) deployment conditions, and communication settings on the quality of IEEE 802.11p V2I communications. The campaign included 22 different RSU locations carefully selected to study the impact of various operating and propagation conditions on V2I communications. Seventy different RSU deployment configurations (combination of RSU location, transmission power, antenna height and type of mast) were analyzed during the campaign. For each configuration, the vehicle performed multiple test drives to/from the RSU to provide valuable indications on the quality of V2I communications. More than 700 test drives were conducted in total, with around 950km of testing distance traveled during more than 35 hours of wireless measurement tests being recorded. The complete set of empirical data can be obtained by interested readers from http://www.uwicore.umh.es/V2I-measurement-campaign/.

A. *Context factors*

The context factors considered in this study are presented in Table I. These six context factors have been extracted for all the conducted test drives.

Table I CONTEXT FACTORS.

Context factor	**Possible values**	**Description**
Buildings	No, same, opposite, both	Represents the relative position of surrounding buildings to the RSU (No buildings, buildings on the same/opposite side of the street as/to the RSU, or buildings in both sides of the street).
Number of lanes	[Integer]	Total number of lanes of the street considering both driving directions.
Traffic density	Low, medium, high	Approximate traffic density observed in the vehicle driving direction.
Trees	No, same, opposite, both	This variable represents the presence of trees relative to the RSU (No trees, trees on the same/opposite side of the street as/to the RSU, or trees in both sides of the street). Trees in a median are considered in an additional variable.
Median	Yes, no	Indicates the presence of a median with trees obstructing the view between the vehicle and the RSU.
Distance to NLOS conditions	[Float]	Distance between the RSU and the location where the vehicle losses visibility with the RSU due to buildings.

A more detailed analysis of the context factors could be performed (e.g. to reflect the percentage of street occupied with buildings or trees instead of simply reflecting their presence), but their current definition already allows demonstrating the potential of ANN and BN to estimate the communications performance using context information. In a practical deployment, static context factors such as Buildings or Number of lanes could be extracted from digital maps. The local traffic density could be estimated through the periodic exchange of beacons among nearby vehicles.

B. Communications performance metrics

The communications performance metrics measured during the test drives and used in this study are:

- Reliable Connectivity Range (RCR): distance to the RSU up to which the experienced PDR (Packet Delivery Ratio) is above 0.7. The RCR represents then the range over which high quality V2I communications can be established.
- Unreliable Connectivity Range (UCR): distance to the RSU from which the experienced PDR is below 0.1, and only very sporadic and low quality V2I transmissions can take place.
- Packets Per Test drive (PPT): number of packets correctly received during a single test drive to/from the RSU. It represents the volume of information that could be downloaded from the RSU during a single test drive.

Fig. 1 illustrates an example of the PDR measured under two different context scenarios, and how the RCR and UCR metrics are obtained. This figure compares the V2I performance with and without the presence of a median with trees obstructing the visibility between the vehicle and the RSU. The figure was obtained considering a transmission power of P_t=20dBm, a packet transmission frequency of T_f=10Hz and an RSU height of h=6.5m. The specific context conditions of these test drives were: Buildings = both, Number of lanes = 6, Traffic density = medium, Trees = no (trees in the median are not taken into account in this variable), Median = yes/no and distance to NLOS = 395m.

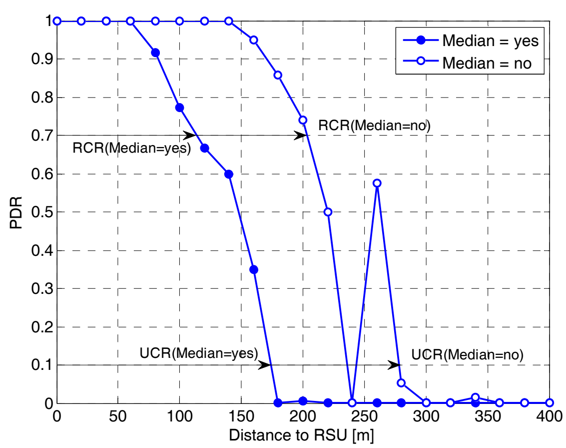


Fig. 1. Effect of a median with trees on the PDR (Packet Delivery Ratio) as a function of the distance of the vehicle to the RSU (h=6.5m, Pt=20dBm).

IV. ANN DESIGN AND TRAINING

The first step to build an ANN is the identification of the relevant variables. While the input nodes will be represented by the context factors detailed in Table I, the output nodes will be the communications performance metrics that need to be estimated. Based on the identified nodes, we propose the use of a feed forward network with a single hidden layer, given its potential to approximate any set of functions if the activation functions are continuous [18]. There is no specific technique to select the activation function to be used by the hidden nodes of an ANN, but in general, non-linear activation functions are preferred to enable the representation of non-linear relationships between inputs and outputs of the ANN. Given its potential demonstrated in related studies, the Sigmoid function has been selected in this study to implement the hidden nodes of the ANN. The criteria shown in [16] has been here followed to select the activation functions of the output nodes, and linear functions have been selected. Fig. 2 illustrates the structure of the ANN here proposed to estimate the communications performance based on the the considered context factors.

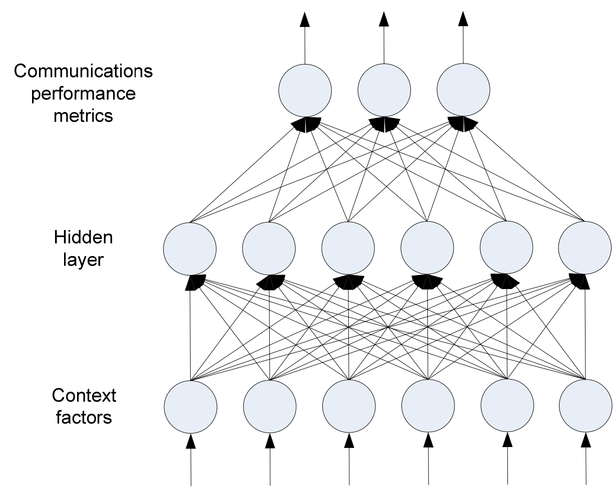


Fig. 2. ANN proposal.

To build, train and evaluate ANNs, this study has used the Matlab library for ANNs. This library has been used to build a feed forward neural network with one hidden layer, 6 input nodes (context factors) and 3 output nodes (communications performance metrics). The number of hidden nodes in the hidden layer has been set to 10. While a lower number of hidden nodes decreased the accuracy of the obtained results, a higher number did not significantly improve the accuracy but increased the computational requirements. While the transfer function of the hidden nodes has been fixed to a Sigmoid (*tanh*), the output nodes present a linear transfer function. TRAINLM is the back-propagation network training function employed for ANN training. It updates weights and bias states according to Levenberg-Marquardt optimization [19]. This algorithm is one of the fastest methods for training moderate-sized feed forward neural networks.

V. BN DESIGN AND TRAINING

To design a BN, the relevant variables need to be identified first. These variables are again the context factors and communications performance metrics. Once the variables have been identified, the structure or topology of the BN needs to be defined. The approach here proposed considers a BN structure in which all variables that can be observed are parents of all variables that need to be estimated. In our study, this corresponds to the BN illustrated in Fig. 3. This structure assumes that all context factors influence all the communications performance metrics. One of the advantages of using this structure is that it has only one undirected path between any two nodes, known as polytree, and therefore can be inferred in polynomial time.

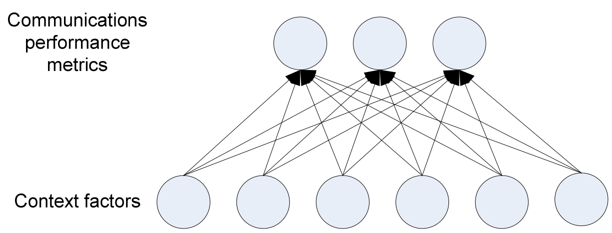


Fig. 3. BN proposal.

To build, train and use BNs, this investigation has used the Bayes Net Toolbox (BNT). BNT is an open-source Matlab package for directed graphical models such as Bayesian Networks. BNT supports different kinds of nodes (probability distributions), exact and approximate inference, parameter and structure learning, and static and dynamic models. BNT is widely used in teaching and research, and is freely available at https://code.google.com/p/bnt/. The BN built has 9 nodes. The first six nodes represent the context factors (Table I) and will be used as observed nodes. The last three nodes represent the communications performance metrics previously defined. The BN built considers Buildings, Number of lanes, Traffic density, Trees and Median as discrete/categorical nodes, and the rest of nodes as continuous nodes with Gaussian distributions (the most popular implemented training and inference algorithms are designed for this type of random distribution).

The algorithm employed for training the implemented BN is the EM (Expected Maximization) algorithm, which finds the maximum likelihood parameters of the different nodes for a fully observed model. In our case, it will find the discrete probability distribution of the discrete nodes, and the median and standard deviation parameters of the continuous nodes. Once the network is trained, we can enter an observation in a given set of variables and calculate the probability distribution of other variables. This is clearly one of the key advantages of BNs with respect to ANNs, since ANNs simply provide the 'best' value of the output variables, but do not provide information about their statistical distribution. In our study, the context factors observed are introduced, and the probability distribution of the performance metrics is calculated.

VI. PERFORMANCE EVALUATION

As previously mentioned, the ANN and BN solutions proposed in this study are here evaluated using the empirical data set obtained from the IEEE 802.11p V2I measurement campaign reported in [3]. Both ANN and BN have been trained with this data set obtained with a transmission power of P_t=20dBm and an RSU height of h=6.5m. Once the networks were built and trained, they have been used to estimate the RCR (Reliable Connectivity Range), UCR (Unreliable Connectivity Range) and PPT (Packets Per Test drive) metrics using as input the context factors reported in Table I and derived from the data set.

A. ANN and BN comparison

The estimations obtained with the ANN and the BN proposals have been compared with the actual measurements, and the relative error has been calculated as RE=100·(E-A)/A, where E represents the estimated value and A the actual one. Fig. 4 shows the CDF (Cumulative Distribution Function) of the relative error RE obtained using the constructed ANN. The figure shows that 60% of the UCR estimations resulted in relative errors between -12.4% and 21.7%, which correspond to the 20th and 80th percentiles, respectively. This result means that if the estimated UCR value is e.g. 1000m, the actual one is between 876m and 1217m with 60% probability. The interval between the 20th and 80th percentiles of the relative error of the performance metrics will be used as a metric to evaluate the estimation process, and will be referred to as *IRE60*. Fig. 5 compares the performance obtained with ANN and BN using *IRE60* for the three performance metrics. The depicted results show that the higher accuracy with ANN is obtained for the UCR metric, which could indicate that the selected context factors better represent this metric than RCR and PPT. The highest error is observed when estimating the PPT metric. This could be the case because the evaluated context factors do not adequately represent this metric. In fact, the PPT metric notably depends on the presence of traffic lights in the path towards the RSU, because if the vehicle stops in a traffic light close to the RSU, the number of packets correctly received notably increases. As a result, to reduce the error in the estimation of the PPT, additional context factors such as the presence of traffic lights could be considered. In addition, more detailed context information could be acquired (tree densities, building heights, etc.) to reduce in general the estimation error. Fig. 5 also depicts the *IRE60* values obtained with the BN proposal. Since the output of the BN is the probability distribution of each performance metric, the *IRE60* metric was obtained with the estimated value E, equal to the mean of the output distribution; other approaches would also be valid (e.g. the median, a given percentile, etc.). Fig. 5 shows that the relative errors obtained with the BN are lower than the ones obtained with the ANN.

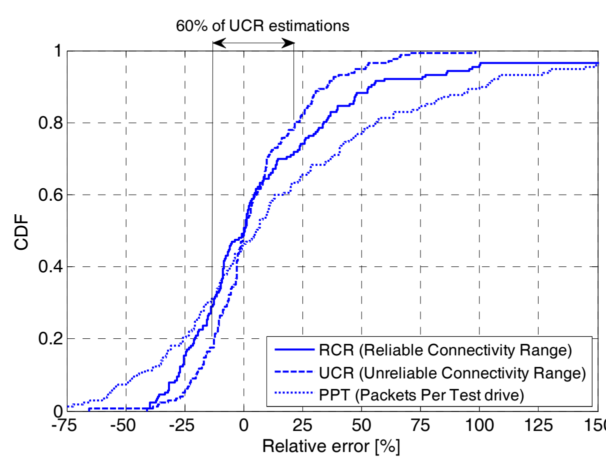


Fig. 4. CDF (Cumulative Distribution Function) of the relative error RE obtained with an ANN.

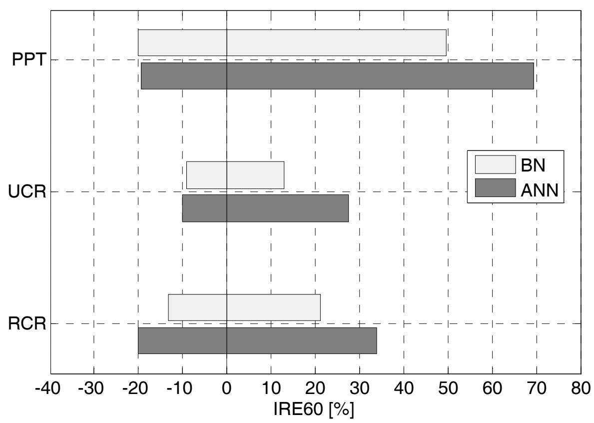


Fig. 5. Comparison of IRE60 values obtained with ANN and BN for the three performance metrics: RCR (Reliable Connectivity Range), UCR (Unreliable Connectivity Range) and PPT (Packets Per Test drive).

B. *Bounds of performance estimation with the BN proposal*

The variability of the radio channel results in that the communications performance metrics measured can differ between consecutive test drives, i.e. different RCR, UCR and PPT values can be obtained under similar context conditions. To illustrate this trend, the bars in Fig. 6 represent the PDF (Probability Density Function) of RCR values measured during the drive tests under similar context conditions. Given this variability, estimating the lower and/or upper performance bounds is of higher interest than estimating e.g. their average value. The proposed BN can be used to estimate such performance bounds since it provides the probability distribution of the performance metrics. Fig. 6 also depicts the PDF of the RCR metric estimated with the BN (a Gaussian curve with certain mean and variance)[1]. The PDF estimated with the BN can be used to estimate the lower and upper bounds of the RCR metric, e.g. the 10th and 90th percentiles, and thereby the interval where to expect the RCR value under certain context conditions. Such interval is highlighted in the figure with a double arrow.

To evaluate the capability of the BN proposal to estimate the metrics' performance bounds, the context factors are used as inputs for each test drive. The estimation process is considered successful if the actual performance metric measured in a test drive is above the estimated lower performance bound for the context conditions under which the test drive was performed. Table II shows the percentage of successful estimations obtained for the different metrics and using the proposed BN. The results in Table II are shown for different percentiles used to obtain the lower performance bound and decide whether an estimation is successful or not. The depicted results show that the percentage of successful estimations increases with lower percentiles.

[1] Under ideal training and building processes, the PDF estimated should match with the measurements. However, in real environments, the communications performance metrics do not necessarily follow a Gaussian probability distribution, and a finite number of measurements can negatively affect the matching.

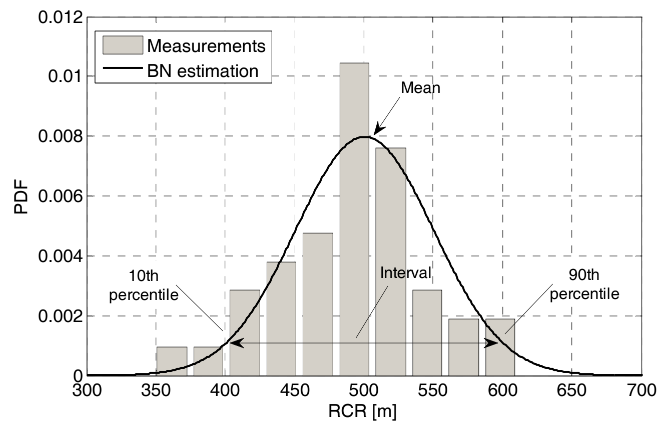


Fig. 6. Example of PDF (Probability Density function) of the RCR (Reliable Connectivity Range) metric for similar context conditions.

To provide an idea of the dispersion of the probability distributions obtained with the BN, Table III shows the average ratio between the interval estimated with the BN and the actual measured metrics. This ratio can be seen as a measure of the variability of the metrics when estimated using the BN. As it can be observed, a more precise estimation of the interval is obtained for the RCR and UCR metrics. However, the higher variability of the PPT metric results in much larger intervals. The accuracy of the obtained estimations could be improved with a higher number of samples/measurements.

Table II PERCENTAGE OF SUCCESSFUL ESTIMATIONS.

	Percentile used for the low performance bound		
Metric	**5th**	**10th**	**20th**
RCR	67.8%	65.5%	62.7%
UCR	68.9%	66.7%	62.7%
PPT	69.5%	63.8%	58.8%

Table III AVERAGE RATIO BETWEEN THE INTERVAL LENGTH AND THE ACTUAL METRICS MEASURED.

	Percentiles used to construct the interval		
Metric	**5th - 95th**	**10th - 90th**	**20th - 80th**
RCR	0.22	0.17	0.11
UCR	0.14	0.11	0.07
PPT	0.98	0.77	0.50

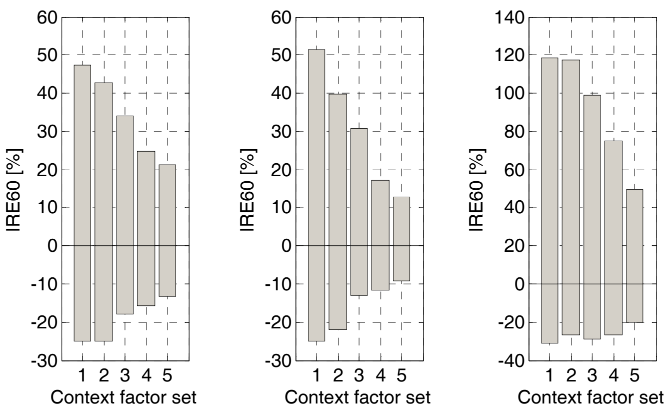
C. *Impact of the number of context factors on the estimation*

This section analyses the impact of the number of context factors on the estimation of the communications performance metrics. The results shown in sections VI.A and VI.B were obtained using all the context factors defined as inputs. In this section, we analyze how reducing the number of context factors influences the accuracy of the obtained estimations. Table IV presents the five context factor sets (CFS) used in this analysis. While CFS1 only includes two context factors, CFS5 includes all the six context factors previously described. Fig. 7

shows the *IRE60* metric obtained when considering the proposed BN and the different CFSs reported in Table IV. The depicted results show that increasing the number of context factors used in the estimation process can considerably reduce the estimation's relative error, although it can also increase acquisition and processing costs. Similar trends have been obtained with the ANN proposal, but are omitted due to space limitations.

Table IV   CONTEXT FACTOR SETS.

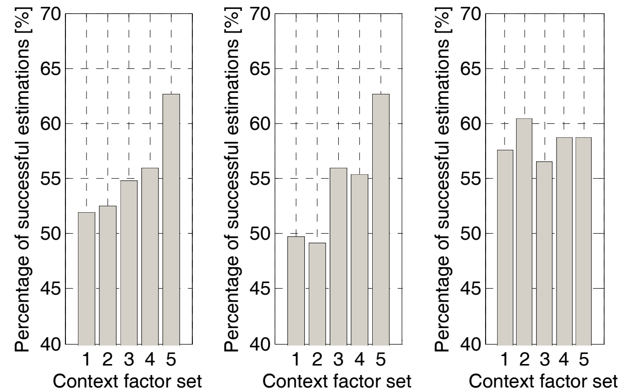
Context factor set	**Context factors included in the communications performance estimation**
CFS1	Buildings, Number of lanes
CFS2	Buildings, Number of lanes, Traffic density
CFS3	Buildings, Number of lanes, Traffic density, Trees
CFS4	Buildings, Number of lanes, Traffic density, Trees, Median
CFS5	Buildings, Number of lanes, Traffic density, Trees, Median, Distance to NLOS conditions



(a) RCR (b) UCR (c) PPT

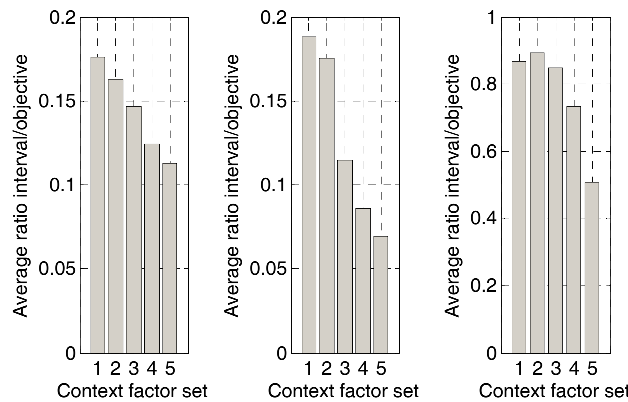
Fig. 7. IRE60 metric as a function of the context factor set for the BN proposal. The estimations are obtained considering the mean of the output distribution of the three performance metrics: RCR (Reliable Connectivity Range), UCR (Unreliable Connectivity Range) and PPT (Packets Per Test drive).

The influence of the number of context factors on the performance estimation process can also be analyzed taking into account the definition of successful estimation previously proposed. Fig. 8 shows the percentage of successful estimations obtained for RCR, UCR and PPT when varying the number of context factors. As it can be observed, this percentage increases as the number of context factors increases for RCR and UCR, which again demonstrates the benefit that increasing the number of context factors could provide if the associated acquisition and processing costs can be tolerated. For the PPT metric, increasing the number of context factors is reflected in a more precise estimation of the interval where to expect the metric, and therefore decreases the ratio between the interval estimated and the actual measurement (see Fig. 9c).



(a) RCR (b) UCR (c) PPT

Fig. 8. Percentage of successful estimations with the BN proposal, and considering the 20th percentile of the output distribution as the lower bound of the interval for the three performance metrics: RCR (Reliable Connectivity Range), UCR (Unreliable Connectivity Range) and PPT (Packets Per Test drive)



(a) RCR (b) UCR (c) PPT

Fig. 9. Average ratio between the interval length estimated and the actual measurements while while varying the number of context factors for the three performance metrics: RCR (Reliable Connectivity Range), UCR (Unreliable Connectivity Range) and PPT (Packets Per Test drive). Percentiles used to construct the interval: 20th – 80th.

D. Quantifying the relevance of context factors

The previous section has shown that increasing the number of context factors can positively influence the estimation process. However, not all context factors might be equally relevant for the estimation process. If this is the case, irrelevant context factors should be excluded from the learning and estimation processes to avoid wasting communication and processing resources. To date, there is not a widely adopted method to calculate the relevance of a context factor. A factor is here considered more relevant than other factors if it has a higher influence on the accuracy of the communications performance estimation.

The previous analysis has been extended to analyze the relevance of each of the six context factors considered in this study. Fig. 10 compares, for the RCR metric, the relative error *IRE60* obtained with CFS5 (vertical black lines and double arrows) to that obtained when removing from CFS5 one of the context factors analyzed (the removed context factor is indicated in the y-axis). The depicted results show that poorer

IRE60 metrics are obtained when removing from CFS5 the Trees or Median context factors. On the other hand, removing the Number of lanes context factor from CFS5 does not significantly increase the *IRE60* metric compared to CFS5. These results thereby demonstrate the varying effect of different context factors on the estimation process, and the importance of correctly identifying the more relevant ones. Fig. 10 also shows that the relevance of a context factor can be different for different metrics. For example, Dist. NLOS is one of the most relevant context factors for estimating the PPT metric, which was not the case for the two other metrics.

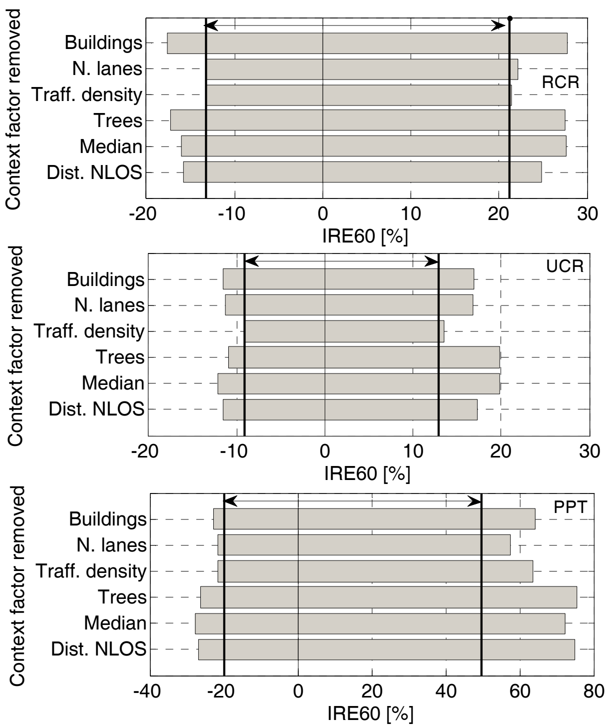


Fig. 10. IRE60 for the RCR (Reliable Connectivity Range), UCR (Unreliable Connectivity Range) and PPT (Packets Per Test drive) metrics when using the BN proposal. The estimations use the mean of the output distribution, and consider only five context factors. The vertical lines and double arrows represent the IRE60 metric obtained with CFS5.

VII. CONCLUSIONS

This paper has proposed and evaluated the use of artificial neural networks and Bayesian networks to estimate the performance of vehicular communications using context information. Both techniques have learning capabilities and provide sufficient flexibility for considering multiple communications performance metrics and context factors. The obtained results show that the BN proposal outperforms the AN one in terms of the estimates' accuracy. In addition, BNs provide relevant statistical information about the communications performance estimates. The probability distribution of the estimated metrics can be used to derive the estimation performance bounds under certain context conditions. The study here reported has also demonstrated the impact of the number of used context factors on the accuracy of the performance estimates, and the varying relevance of context factors for different performance metrics. This relevance analysis could be useful to identify which type of data needs to be stored and which factors should be evaluated in future studies.

REFERENCES

[1] T. Moazzeni, "A wireless propagation channel model with meteorological quantities using neural networks", *Proc. Int. Conf. IEEE Consumer Communications and Networking Conference (CCNC)*, Las Vegas (USA), pp. 1307-1309, 8-10 Jan. 2006.

[2] H. Menouar, M. Lenardi, F. Filali, "A Movement Prediction based Routing Protocol for Vehicle-to-Vehicle Communications", *Proc. Vehicle-to-Vehicle Communications Workshop (V2VCOM)*, San Diego (USA), July 2005.

[3] J. Gozalvez, M. Sepulcre and R. Bauza, "IEEE 802.11p Vehicle to Infrastructure Communications in Urban Environments", *IEEE Communications Magazine*, vol. 50, no. 5, pp. 176-183, May 2012.

[4] J. Gozalvez, M. Sepulcre and R. Bauza, "Impact of the Radio Channel Modelling on the Performance of VANET Communication Protocols", *Springer Telecommunication Systems*, vol. 50 (3), pp. 149-167, July 2012.

[5] C. Mannweiler, A. Klein, J. Schneider, H.D. Schotten., "Exploiting User and Network Context for Intelligent Radio Network Access", *Proc. Int. Conference on Ultra Modern Telecommunications*, St. Petersburg (Russia), Oct. 2009.

[6] B. Coll-Perales, J. Gozalvez, V. Friderikos, "Context-Based Opportunistic Forwarding in Multi-hop Cellular Networks using Mobile Relays", *Proc. ACM Workshop on High Performance Mobile Opportunistic Systems (HP-MOSys),* Barcelona (Spain), Nov. 2013.

[7] J. Tian, et al., "Spatially aware packet routing for mobile ad hoc inter-vehicle radio networks", *Proc. IEEE Int. conference on Intelligent Transportation Systems*, Shangai (China), October 2003, pp.1546-1551.

[8] Zhe Wang and M. Hassan, "Context-Aware Channel Coordination for DSRC", *Proc. IEEE Automotive Networking and Applications (AutoNet) Workshop*, New Orleans (USA), pp.1-6, Dec. 2008.

[9] T. Kosch, C.J. Adler, S. Eichler, C. Schroth, and M. Strassberger, "The scalability problem of vehicular ad hoc networks and how to solve it", *IEEE Wireless Communications*, vol. 13, no.5, p. 22-28, Oct. 2006.

[10] B. Kosko, S. Isaka, "Fuzzy Logic", Scientific American, vol. 269, no. 1, pp. 62–67, July 1993.

[11] R. Bauza and J. Gozalvez, "Traffic congestion detection in large-scale scenarios using vehicle-to-vehicle communications", *Journal of Network and Computer Applications*, In Press, March 2012.

[12] X. Binglei et al, "Fuzzy-logic-based Traffic Incident Detection Algorithm for Freeway", *Proc. Int. Conf. on Machine Learning and Cybernetics (ICMLC)*, Kunming (China), 12-15 July 2008.

[13] R. Belizia Polastro et al., "Probabilistic Description Logics for Context Awareness in Mobile Robotics: A Case Study in Imprecisely Specified Probabilities", *Proc. Int. Symposium on Imprecise Probability: Theories and Applications (ISIPTA),* Durham (UK), 14-18 July 2009.

[14] R. Daly, et al., "Learning Bayesian networks: approaches and issues", *The Knowledge Engineering Review - Cambridge University Press*, vol. 26, pp 99-157, 2011.

[15] M. Correa, et al., "Comparison of Bayesian networks and artificial neural networks for quality detection in a machining process", *Elsevier Expert Systems with Applications*, vol. 36, no. 3, part 2, pp. 7270-7279, April 2009.

[16] Anil K. Jain, et al., "Artificial Neural Networks: A Tutorial", *IEEE Computer*, vol. 29, pp.31-44, 1996.

[17] D. Koks and S. Challa, "An introduction to Bayesian and Dempster-Shafer data fusion", *DSTO Systems Sciences Laboratory DSTO–Technical Report–1436*, Edinburgh (Australia), 2003.

[18] G. Cybenko, "Approximation by Superpositions of a Sigmoidal Function", *Math. Control Signals Systems*, vol. 2, pp. 303-314, 1989.

[19] M.T. Hagan and M. Menhaj, "Training feed-forward networks with the Marquardt algorithm", *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989-993, 1994.