An Intelligent System to Enhance Traffic Safety Analysis

Andreas Gregoriades
Dept of Computer Science and Engineering
European University Cyprus, Cyprus
a.gregoriades@euc.ac.cy

Neville Parker
Civil Engineering, CCNY-CUNY
Director CUNY ITS
New York, NY, USA
parker@utrc2.org

Kyriacos C. Mouskos
CTL Cyprus Transport and Logistics Ltd
Nicosia, Cyprus
mouskoskc@gmail.com

Ismini Hadjilambrou
CTL Cyprus Transport and Logistics Ltd
Nicosia, Cyprus
topoismini@gmail.com

Natalia Ruiz-Juri
The University of Texas at Austin
Austin, Texas, USA
natiruizjuri@gmail.com

Aneesh Krishna
Dept of Computing
Daculty of Science & Engineering
Curtin University of Technology, Australia
a.krishna@curtin.edu.au

***Abstract*. Traffic phenomena are characterized by complexity and uncertainty, hence require sophisticated information management to identify patterns relevant to safety. Traffic information systems have emerged with the aim to ease traffic congestion and improve road safety. However, assessment of traffic safety and congestion requires significant amount of data which in most cases is not available. This work illustrates an approach that aims to alleviate this problem through the integration of two mature technologies namely, simulation-based Dynamic Traffic Assignment (DTA) and Bayesian Belief Networks (BBN). The former generates traffic information that is utilised by a Bayesian engine to quantify accident risk. Dynamic compilation of accident risks is used to gives rise to overall traffic safety. Preliminary results from this research have been validated.**

Keywords - Traffic Safety; Dynamic Traffic Assignment; Bayesian Belief Networks.

I. INTRODUCTION

It is well recognized that traffic accidents contribute substantially to urban congestion and traffic safety. Even a minor accident can cause significant traffic congestion in directly impacted areas which can cause safety issues to the overall road network due to secondary crashes as a result of rerouting [3]. Therefore, predicting traffic dynamics has always been an important issue in road safety. However, predicting the behavior of a road network under extreme conditions using historical records is a complex task. This work addresses this issue through the development of a novel Intelligent Traffic Information System that leverages the capabilities of two mature methodologies namely simulation-based Dynamic Traffic Assignment (DTA) [9] and Bayesian Belief Networks (BBN) [1]. The former is widely used in transportation planning and operations to predict drivers' decisions (where and when to travel on the road network), while the latter is powerful uncertainty modelling technique. Both technologies gained significant acceptance with the invention of powerful computational algorithms that enabled their exploitation.

Simulation-based DTA models depart from the traditional static analysis of traffic phenomena that employ analytic approaches to represent traffic conditions. DTA use traffic simulation to replicate the complex traffic flow dynamics especially for signalized systems where the vehicle and signal interactions are difficult to model analytically. This enables dynamic control and management systems to anticipate problems before they occur rather than simply reacting to existing conditions. The main output of DTA is the dynamic user equilibrium paths of each Origin-Destination (OD) pair [3]. These paths define the optimum route on the network that each vehicle will follow given its origin and destination.

BBN are ideal for modelling problems that are characterised by complexity, uncertainty and incomplete information. They have been used extensively in reliability engineering, risk management and decision support. In our case BBN are used to model and assess accident risk using dynamic and static transportation information.

The motivation of this work resides to the fact that current traffic information systems use multi-state safety databases that contain crash, roadway inventory, and traffic volume data. These are analysed to identify safety issues and evaluate the effectiveness of accident countermeasures. The main limitation of these systems is their retrospective approach. Effective safety management requires a prospective viewpoint. The proposed method combines dynamic modelling of traffic conditions with knowledge-based accident prediction to leverage the benefits of

 ISBN: 978-1-61208-132-8

computational intelligence in road safety and in this way provide forecasts of traffic safety. The paper illustrates the integration of BBN accident risk technique with the VISTA (Visual Interactive Systems for Transport Algorithms) simulated DTA framework. The method is applied to study the future behaviour of a road network in Cyprus.

The paper is organised as follows. Next section gives an overview of the literature. This is followed by the methodology. Subsequent sections concentrate on data pre-processing and BBN development. The integration of VISTA with the BBN along with the results that emerge from the amalgamation of the two technologies is described next. The paper finishes with the conclusions section.

II. Related Work

Traffic information systems mainly provide retrospective analysis of traffic safety using historical data [2] due to, coverage, cost and real-time issues of traditional sensor-based schemes to traffic data collection. To escape from this problem simulation based traffic estimation systems emerged. The DTA simulation method employed herein constitute the state of the art in traffic forecasting. The two DTA approaches commonly used to emulate the path choice behavior of drivers are dynamic assignment en-route and dynamic equilibrium assignment. The former models behavioral rules that determine how drivers react to information received en-route while the latter only pre-trip path choices are considered and the goal is to minimize each driver's travel time by finding optimal or sub optimal paths [5]. This work is based on latter approach. Alternative simulation methodologies such as CORSIM, VISSIM, PARAMICS, WATSIM, AIMSUN do not have a true traveler behavior routing component [10]. They instead move traffic by splitting it probabilistically at every intersection. Thus the above-mentioned simulation models cannot be used to accurately predict traffic flow of each road section.

Related accident prediction approaches such as the one employed by Simoncic [6] utilise probabilistic modelling through BBNs. Their work illustrates the application of BBN to model road accidents and accordingly make inferences for accident analysis. The main limitation of this effort is that it concentrates solidly on the development of the BBN model without providing any substantial evidence of its performance. Work by Hu et al. [7] also uses a probabilistic approach to predicting road accidents through intelligent surveillance of vehicle kinematics; however, their method does not address the causal aspects that lead to observed behaviours and hence cannot be easily generalised. State-of-the-art tools in accident prediction, such as SafeNET 2 (Software for Accident Frequency Estimation for Networks), use traffic flows and geometric information to assess accident risk [8]. However, unlike our method, SafeNET 2 does not address the dynamic aspects of road networks using simulation. Hence, their traffic flow estimates are generic which in effect could lead in inaccurate conclusions.

III. Methodology and Rationale

The traffic safety assessment system proposed herein is the amalgamation of probabilistic risk assessment [1] with mesoscopic traffic simulation [10]. The need for this integration boils down to the limitations of traditional traffic information systems that mainly concentrate of data warehousing. The methodology proposed utilises data marts to generate projections of future system behaviour. To that end, intelligent information management techniques have been employed to distil knowledge necessary for the development of models that enable the prospective analysis of system behaviour. The two models that emerged from this process are the accident risk assessment model and the traffic simulation model. The accident risk assessment employed is causality-based and uses a technique popular in the AI domain, namely, BBN. BBNs gained widespread acceptance with the introduction of computational algorithms that enabled their exploitation by Pearl [1]. BBNs are causal networks based on the concept of Bayesian probability, and provide a language and calculus for reasoning under uncertainty. A BBN in essence is a directed graph. It consists of vertices or nodes and directed edges (arrows). Each edge points from parent node to child node. In a belief network each node is used to represent a random variable, and each directed edge represents an immediate dependence or direct influence. Inference is achieved by belief propagation through the models topology.BBN technology is used to model how traffic and infrastructural factors influence accident risk. The second component of the approach is a road traffic simulator based on DTA. DTA evolved rabidly over the past two decades. This advancement has been fueled by the needs of application domains ranging from real-time operations to long term planning. DTA models constitute a natural progression in transportation [3] that evolved from static assignment approaches that assume that traffic flow is static and independent of time. One of the main features of DTA models is the dynamic analysis of road networks using time-varying traffic demands. DTA models effectively the complex interactions between supply and demand in a transport network. As a result, they capture the spatio-temporal trajectories (from origin to destination) of every vehicle and in return mimic in real-time the basic driver behaviors of road users. This constitutes a great advantage over traditional models that do not track the movement of individual vehicles but instead split traffic at intersections [3]. The DTA model is used in VISTA through the Dynamic User Equilibrium (DUE) model [4]. DUE assumes that no user can improve his/her travel time by changing their travel paths or by altering their departure or arrival times. DTA methods are divided into two groups the analytical and the simulation-based models. The former uses mathematical techniques to solve traffic problems while the latter represents problems as a set of interrelated components that dynamically change. The use of DTA model enhances the limitations of existing practices by providing a consistent way of producing estimates of traffic flow conditions of road networks using limited information from traffic flow detectors. Moreover, it produces timely and complete traffic volume estimates for all sections of a road network and hence, can be used to assess accident risk using time varying conditions. The integration of BBN with VISTA in the proposed traffic information system enables the dynamic assessment of accident risk using simulated traffic conditions and prior knowledge embedded in the BBN. A pilot study conducted with the system aimed to assess the safety performance of the Nicosia road network in Cyprus and to investigate how it will behave under different scenarios.

Initially the road traffic model of Nicosia was specified, implemented, verified and validated in VISTA. Models in VISTA are represented by nodes connected by unidirectional links that represent flow of traffic in one direction. It is possible to have more than one link between two nodes to indicate separate lanes and lane direction. The completed VISTA simulation model was integrated with an accident risk assessor implemented in Java. The simulator provided the risk assessor with the traffic volumes of all road sections of the network for every 15 min interval. Traffic volumes in combination with infrastructural properties of the road network were used by the BBN to assess accident risk. Dynamic input to the BBN is provided by the DTA simulation on a step by step basis. For the development of the BBN topology and the parameterization of its prior knowledge, historical road accident data were complied. The BBNs accuracy was evaluated cross validation technique.

IV. SYSTEM ARCHITECTURE.

The proposed system was developed using the component-based software engineering methodology. With the initial specification of the system requirements we proceeded in the identification of suitable software components that matched the initial system requirements. These components were subsequently integrated to implement parts of the system's functionality. In particular the Bayesian inference engine and the charting components were selected after thorough investigation. The glue code that enabled components interaction was implemented in Java. The architecture of the system is comprised of five main components: the BBN accident assessor, the VISTA simulator, the data preprocessor, the results analyser and the results visualiser, as depicted in Figure 1. The risk assessor quantifies accident risk using a Bayesian inference engine that utilises a probabilistic model. Input to the BBN assessor is categorized into static and dynamic. The former is obtained from the VISTA database and the latter is the output of the VISTA simulation. Input evidence is pre-processed before running the BBN model that propagates the evidence down the network to produce the posterior probability. The integration of the VISTA with the BBN model was realised through asynchronous data interchange. Inherently, results from the VISTA simulator were accumulated in a database along with infrastructural information for each road section. The Risk Assessor component accesses the traffic flow and infrastructure records from the database and accordingly propagates the input evidence down the BBN network to produce the posterior probability for accident risk.

To establish communication between VISTA and the risk assessor it was imperative to pre-process VISTA's output data prior to being utilised by the BBN of the risk assessor. Specifically, VISTA variables are continuous by nature, while the BBN model uses categorical/discrete variables. Hence, it was necessary to discretise the output from VISTA prior to instantiating the BBN model. For the discretisation process it was necessary to refer to domain experts that specified the cut-off values for each variable. Specifically, for traffic volume three states were defined, namely, low, average and high. The first corresponding to less than 100 vehicles per 15 time interval, the second to less than 350 and the last to greater than 350.

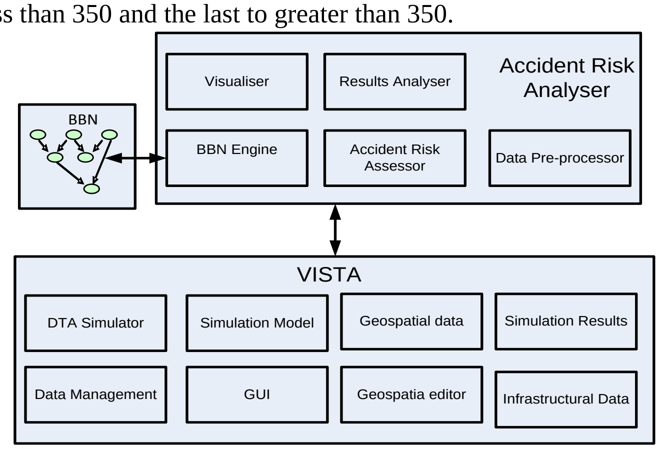

Figure 1. The System architecture

V. DATA COLLECTION AND PRE-PROCESSING

The development of the BBN required the specification of the topology and the parameterisation of its prior knowledge in conditional probability tables. To that end historical accident records were obtained from the Cyprus Police department specialising on traffic safety. Preliminary compilation of the data was performed with the SPSS statistical package. The accident dataset covered all accidents occurred in the Nicosia area from 2002 until 2008 and comprised over 9000 records. Each record consisted of 43 (six continuous and 37 categorical) input parameters covering global, local, temporal, accident, driver and car characteristics collected at the site of the accident by the police officers, eye witnesses and the involved parties. Each record was associated with a single categorical output parameter pertaining to accident severity, namely light,

ISBN: 978-1-61208-132-8

severe and fatal, as evaluated by the police officer at the site of the accident.

However, for the development of the BBN's topology it was imperative to specify also the influence of infrastructural properties to the accident risk. However, due to limited information regarding infrastructural properties in the accident reports, it was necessary to map each accident on a geospatial GIS platform and subsequently import these on VISTA to obtain more information regarding the infrastructure at the accident location. Once this mapping was achieved additional information regarding the dynamic aspects of the road network at the accident scene was obtained from VISTA. This helped to define the causal relationships of the BBN variables that described the infrastructure and the traffic dynamics. Figure 2 shows the accidents geospatial layer of the dataset on ARCGIS using the baseline map. Each lollypop on the map corresponds to one or more accidents that occurred at the specified location.

Figure 2. Accident data on ARCGIS

The same accident dataset was used to identify locations on the network with high accident frequency the so called black spots. These points were used to validate the system after it was implemented. Specifically, a subset of the original dataset was used to validate the system. Black-spots (Figure 3) that were associated with that subset were used to test its performance on unknown conditions.

VI. DEVELOPMENT OF THE BBN

To develop the BBN model it was imperative to firstly identify the variables that adequately describe the problem, subsequently define the possible states that each variable could take and finally define the dependencies among them. A preliminary analysis of the accident data provided a generic indication of the influence of each variable to road accident. Database pre-processing involved two steps (a) replacement of missing and erroneous (e.g. falling outside the acceptable range) parameter values by the mean value of the parameter values of the other (assumed correct) records, and (b) grouping neighboring or related values of multi-valued (i.e. containing more than 12 values) categorical parameters so as to have a manageable number of intelligible as well as regular categories per parameter.

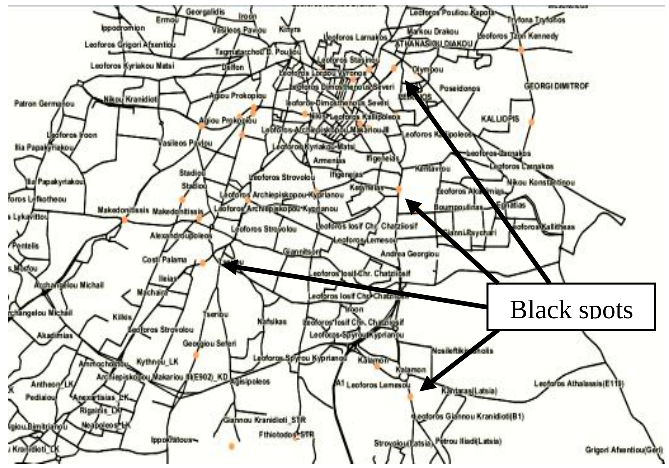

Figure 3. Network's Black Spots as overlaid dots

Statistical analysis relating the 43 input parameters (independent variables) to accident type (dependent variable) reveals that the Spearman correlation coefficient values between the inputs and the output are low (Figure 4), while the Spearman p-values are relatively high. Owing to the sufficient size of the database however, it is still possible for some of the correlations to be statistically significant. In support of that, accident type prediction was found far from satisfactory when only the statistically significant/correlated parameters were employed, thus demonstrating that statistically derived feature selection cannot be performed on a statistical basis for extracting the input parameters that affect the output and discarding those that do not provide accident type-related information.

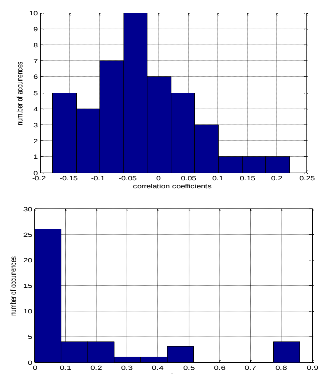

Figure 4. Statistically derived correlation coefficients (left) and p-values (right) between the 43 input parameters and accident type.

To that end, the processed accident data was subsequently used to identify the core variables of the BBN model along with their dependencies. In order to reduce the

 ISBN: 978-1-61208-132-8

complexity of the process and the model itself, the dimensionality of the initial data set was reduced using Principal Component Analysis (PCA). This helped to identify the core variables of the model. In principle, PCA projects the original data into a new set of orthogonal axes in such a way that the original multidimensional dataset with possibly correlated parameters is linearly transformed into a novel dataset of identical dimensions but with totally uncorrelated parameters. Owing to the fact that each new axis is selected so as to maximally expose the (remaining) variability of the dataset, it is not unusual for the first few axes of the PCA mapping to account for most of its variability. Hence, small PCA axes are generally sufficient in representing the original data with minimal loss of information. Results from this process yielded 12 artificial variables.

In addition, a subset of the accident data acquired by the police was mapped into ARCGIS geospatial application and subsequently on VISTA as a GIS layer. Afterward, VISTA was utilized to associate the accident parameters with the infrastructural properties of the road network at each accident point. Finally the traffic volumes along with traffic speed of vehicles were used to calculate traffic density of each road section of the network. These were subsequently associated with each accident record. Results from the dimensionality reduction using PCA together with the data merge that associated accidents locations with traffic density metric and infrastructure, yielded 19 candidate variables for the development of the BBN topology. The topology depicted in Figure 5, was initially learned from these processed data using the Expectation Maximisation algorithm. A refinement of the initial topology was performed using domain knowledge from the literature and subject matter experts.

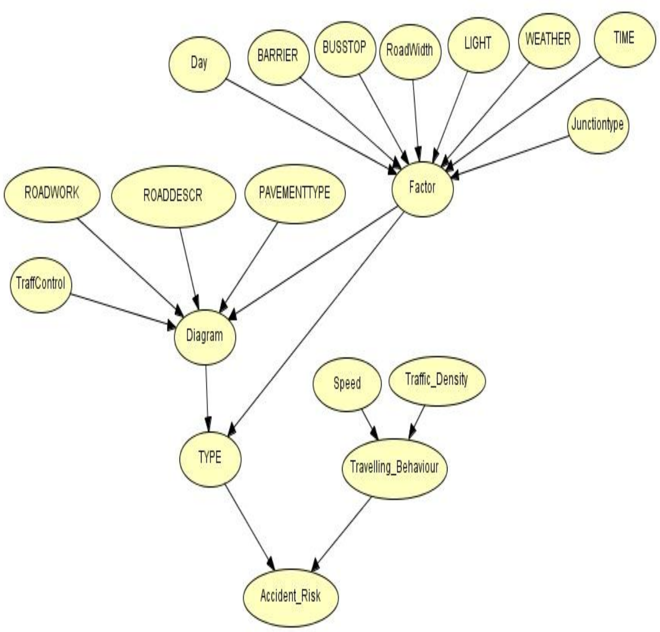

Figure 5. Learned BBN topology

To validate the effect of each variable to the target variable, namely, accident risk, a preliminary validation study was conducted using attribute relevance analysis (Fig. 6). Envisioner, data mining tool was used to compute the relevance between each identified causal factor to accident risk. The relevance of each variable to accident risk was compared against the learned conditional probability that emerged from EM algorithm and the posterior probabilities computed by the BBN with the independent instantiation of each of the leaf node variables.

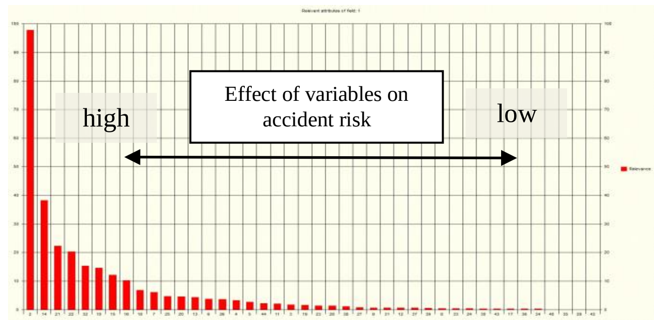

Figure 6. Relevance Analysis results.

VII. Model Validation

To estimate the accuracy of the developed BBN model, five-fold cross validation has been employed. To that end the accident database that was enhanced with traffic data was randomly partitioned into five folds of equal number of records. Subsequently, and for each fold, four of the sets were employed for training the model while the remaining set was reserved for testing. Prediction accuracy was calculated by the weighted average of the test results of the five folds. Overall, the results of the validation process demonstrate that the model can accurately predict accident risks. However, the fact that the traffic volume used for validation is based on simulated results biases the outcome. To that end an additional validation study needs to be performed to verify that the model performs well in realistic settings.

VIII. Results

Results from the accident risk assessor were used to calculate the accident risk index of each road section on the network. A road segment was labeled as accident prone if the predicted BBN accident risk probability was above a pre-specified threshold value. BBN estimates that fall below the cutoff value were ignored. This enables the safety engineer to alter the granularity of the analysis by altering the threshold value. To produce the accident risk index is was imperative to normalize the number of accidents that were predicted by the BBN with the traffic volume per time interval, for each road section. This aimed to escape the Simpsons paradox that defines phenomena that falsely prove the reverse of the truth. Inherently the Sympson's Paradox implies false causation, a logical fallacy by which two events that occur together are claimed to be cause and

 ISBN: 978-1-61208-132-8

effect. For example: statistically more accidents occur while the weather is good. Therefore, one erroneously could infer that good weather causes road accidents. The above argument commits to this fallacy, because in fact the explanation is that in good weather more cars are in the road and this causes more accidents. To find the actual effect of weather on accidents it is hence important to normalize the accidents that occurred in relation to the cars that are in the road. To that end, the proposed system uses a systematic approach that utilizes the traffic volume estimates from the VISTA simulation and the accidents predicted using the BBN risk assessor. Traffic volume acts as a normalizing factor for the number of accidents predicted using the BBN risk assessor. This gives rise to the accident risk index for each road section of the network that inherently highlight network's black spots. An illustration of the results produced by the traffic safety system is depicted in Figure 7. This figure illustrates a subset of the results and indicate that sections with IDs, 3, 21 and 47 have the highest accident risk index. The system enables the safety engineer to provide appropriate countermeasures to alleviate the problem. These are reported in the system and subsequently implemented in the simulation model. Each countermeasure then undergoes an evaluation procedure in the system to verify that the problem is eliminated prior to being implemented.

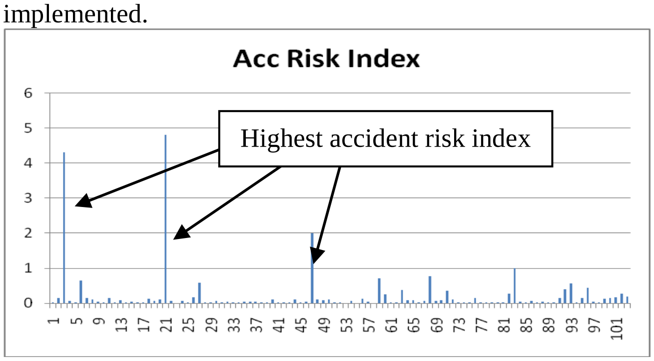

Figure 7. Links on the road network with highest accident risk index

IX. Conclusions and Future Work

The traffic information system described herein illustrates a novel approach to quantifying road safety using probabilistic inference with DTA simulation. Integration of VISTA with BBN, as presented, enables the combination of known and uncertain evidence for accident risk quantification. The system combines state of the art technologies in traffic simulation and accident risk assessment. Integration of these provides the safety engineer with the necessary mean to perform a holistic and intelligent analysis of road safety. The method escapes from the problem of traffic data shortage that most traditional approach are suffering, through the use of DTA simulation. VISTA provides complete traffic volume data estimates for all road sections of the network on a 24 hour basis. This constitutes advancement over existing methods that base their analysis on limited data obtained from a scarce number of traffic sensors on the network. Therefore, the proposed method and the supporting system enable the identification of safety hazards in road networks using dynamic data and thus improved safety analysis that escapes the Sympson's paradox.

It should be noted that, once a black spot is identified, it requires a microscopic safety analysis to examine the behavioral aspects of the vehicle kinematics that leaded to an accident. This requires the development of the corresponding microscopic simulation. This requires detailed specification of roadway geometry, comprehensive traffic control data, lighting, environmental and traffic conditions in a microscopic simulation model. A future enhancement of our method will be developed to incorporate the above attributes into a model that will inherit properties from the mesoscopic analysis described in this study.

Acknowledgment

This research was co-funded by the Cyprus Research Promotion Foundation under the program: Targeted International Collaboration (ΔΙΕΘΝΗ/ΣΤΟΧΟΣ/0308).

References

[1] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Information. San Francisco: Morgan Kaufmann, 1988.
[2] C. Kapp, "Who acts on road safety to reverse accident trends", The Lancet, vol. 362, 2003, pp. 1125-1125.
[3] V. Sisiopiku and X. Li, "Overview of Dynamic Traffic Assignment Options" Proceedings, 2006 Spring Simulation Multiconference, Society for Modeling and Simulation International, 2006, Huntsville, AL, pp. 110-116.
[4] A. Ziliaskopoulos and S. Lee, "A Cell Transmission Based Assignment-simulation Model for Integrated Freeway/Surface Street Systems", Proc., 75th Transportation Research Board Annual Meeting, 1996, Washington, DC, pp. 427-444.
[5] M. Florian et al., "Application of a simulation-based dynamic traffic assignment model", European Journal of Operational Research, vol. 189, 2008, pp. 1381–1392
[6] M. Simoncic, "A Bayesian Network Model of Two-Car Accidents" Journal of Transportation and Statistics, vol. 7, no. 2,3, 2004, pp13-27.
[7] W. Hu, X. Xiao, D. Xie et al., "Traffic accident prediction using 3-D model-based vehicle tracking", IEEE Transactions on Vehicular Technology, vol. 53, no. 3, 2004, pp. 677-694.
[8] TLR, "SafeNET2", http://www.trlsoftware.co.uk/news/detail.asp?pid=51&iid=29, 2007.
[9] H. Lo, "A dynamic traffic assignment formulation that encapsulates the cell transmission model", In A. Ceder (ed.) Transportation and Traffic Theory, Pergamon, Oxford,1999, pp. 327-350.
[10] FHWA, "Surrogate Safety Measures From Traffic Simulation Models", FHWA-rd-03-050 U.S. Department of Transportation Federal Highway Administration Research, Development, and Technology, 2010