ISBN: 978-80-245-2125-1

DOI: 10.18267/pr.2015.pav.2125.10

SYSTEM APPROACH TO INCREASING SAFETY OF ROAD TRAFFIC

DOI: 10.18267/pr.2015.pav.2125.10

Ing. Bc. Marián LAMR doc. , Ing. Jan SKRBEK, Dr

Department of Informatics , Faculty of Economics

Technical University of Liberec

marian.lamr@tul.cz , jan.skrbek@tul.cz

ABSTRACT

This paper deals with a systematic approach to increase traffic safety using modern information technology. Potential data sources are described together with a chosen sample of potential attributes suitable for modeling of dangerous situations. Possible methods for creation of clusters of traffic accidents are outlined and used in a conceptual model of an early warning system about traffic accidents.

KEY WORDS

Road traffic safety, traffic accident, cluster analysis, data mining.

INTRODUCTION

As was already stated in our previous paper (Lamr & Skrbek, 2015) there has been a permanent increase in the amount of traffic accidents since 2011. The police have repeatedly tried to find suitable methods to increase road traffic safety. These methods include, but are not limited to an increased number of police controls, increased penalizations, utilization of a penalty point system and an overall increase of regional and state-wide traffic police operations. Despite all these precautions which aim to eliminate the amount of traffic accidents and injuries, the number of traffic accidents, serious injuries and deaths keeps increasing. The police investigated 16 000 traffic accidents during the holiday in 2015. Except for 150 deaths, more than 500 severe injuries were recorded and the overall damage costs increased. 70 people were killed on roads in august 2014, which is 12 more compared to 2014 (ČTK, 2015). The Ministry of Transport plans to increase sanctions for certain traffic offences even though it was shown that this approach only helps in the short term. (ČTK, 2015). This is a necessary course of action, however, not all road users will obey the new rules and regulations and they do not realize the danger of their behavior. If these people could be warned in time with a high reliability of the warning it is probable that at least a part of them would take the warning into consideration. We assume that the possibility to inform drivers about locations of possible dangers can significantly improve the current situation.

Our recent projects are focused on finding a system solution, which would contribute to reduction of the number of traffic accidents. Unfortunately, this number is constantly growing together with material damage, injuries, and deaths. We try to connect advanced approaches from the field of data mining with progressive methods of vital information transfer and distribution. We aim to identify problems and problematic locations and connect them to other aspects of traffic accidents, such as weather, road conditions, day of the week and others.

ISBN: 978-80-245-2125-1

DOI: 10.18267/pr.2015.pav.2125.10

POTENTIAL DATA SOURCES

The Ministry of Transport provides a database of traffic accidents which took place in the Czech Republic as a part of a project called "Unified Traffic Map" (UTM) since 2006. This project is available online at http://www.jdvm.cz. Because of this project it is possible to find accident data online, to obtain a detailed list of parameters for a given accident and to plot it on a map.

Police of the Czech Republic also regularly create accident statistics which are available online. (Policie ČR, 2015). These statistics are, however, only a simple projection of accident parameters to charts and tables. Usually it is a representation of an examined value in a given period and its comparison to a previous period. To be specific, these can represent numbers of accidents each day, numbers of accidents in accordance with their culprits, numbers of accidents in accordance with accident types in each month, main causes of accidents, numbers of accidents in counties in specific periods and others. This data is available in a PDF format and gives a basic overview of traffic accidents.

Because the data is only available in a PDF format from the Ministry of Transport, it is necessary to download it from the web and convert it to a more suitable format. Another option would be acquiring the data in a suitable format directly from the Police of the Czech Republic. More than 600 000 thousand records can be obtained from the UTM project, each containing 44 attributes. Table 1 shows only selected attributes from such data set. An example is assigned to each attribute for better clarity.

Attribute	Record (example)
Accident Location	Prague (capital city)
Accident Date	1.1.2013
Day of the Week	Monday
Accident time	19:00
Accident caused by	Motorized vehicle driver
Alcohol	Yes, blood alcohol content < 0,99 ‰
People Killed	0
Heavily Injured	0
Cause of Accident	Improper turning or reversing
Road Surface condition	Wet surface
Visibility	Nighttime, visibility not reduced by weather conditions
Technical Road Condition	Good, free of defects

Table 1 – Selected illustrative attribute set of a traffic accident record (Authors, 2015)

Another significant and non-negligible source of information for prediction models could be traffic accident databases of insurance companies. It is generally known that a traffic accident record is made after each traffic accident, while this record is identical for all insurance companies. Further information is collected by employers of insurance companies during their communication with clients. Utilization of this information source is hindered by its availability, because it can be considered an intellectual property of a private company and it would be necessary to remove personal information from the data by the company prior to its use.

CLUSTER ANALYSIS AS TOOL FOR FINDING CONNECTIONS BETWEEN TRAFFIC ACCIDENTS

Cluster analysis (CA) is a method that makes it possible to find accidents with similar properties. As its name suggests, this method (which is also used out of data mining) creates clusters in input data. A cluster can be defined as a set of objects with similar properties.

Most algorithms of cluster analysis usually try to represent object properties (attributes) using numbers and it is therefore necessary to transform categorical and dichotomous attributes to numbers in a suitable

ISBN: 978-80-245-2125-1

DOI: 10.18267/pr.2015.pav.2125.10

way. This transformation must be applied to most attributes of data obtained from the Ministry of Transport, because most of the attributes is categorical data.

Attributes entering cluster analysis should also be standardized in a specific way - for example to prevent an attribute with higher values to easily dominate other attributes. Such dominance of some attributes over others would significantly distort similarity evaluation. Approaches preventing these problems are called attribute standardization and one method of achieving it can be described by formula:

$$y_j^i = \frac{x_j^i - \bar{x}_j}{s_j}$$

where y_j^i denotes a standardised j-th attribute of i-th object

x_j^i denotes a j-th attribute of original value, related to i-th object

$\bar{x}_j$ is the mean value of j-th attribute for all objects:

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_j^i$$

s_j denotes the standard deviation of j attribute over all objects

$$s_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_j^i - \bar{x}_j\right)^2}$$

The core issues of cluster analysis are, however, methods of attribute similarity evaluation. There are many methods to express similarity and none of them can be considered universal and applicable to all types of tasks. To give an example we can mention evaluation using object association coefficient or various metrics, such as Euclidean, Manhattan or Chebyshev.

Clustering methods can be divided to hierarchical and non-hierarchical. Hierarchical methods can be divided further to agglomerative (nearest-neighbor method, farthest neighbor method, centroid-based clustering) and divisive (e.g. MacNaughton-Smith).

Agglomerative methods are based on clusters containing only one object. Divisive methods, on the other hand, begin with a single cluster containing all objects. Each cluster divides in two during each decomposition, while decomposing continues until the number of clusters is equal to the number of objects.

Hierarchical structure of partitioning is not applied during non-hierarchical clustering. Methods of nonhierarchical clustering evaluate the quality of object distribution to clusters and this distribution can be modified if necessary. Decomposition quality functional is used for optimal distribution of clusters and their objects. This functional is based on localization of positions where object extrema occur. Non-hierarchical clustering methods can be also divided into two distinct variations. In the first one the number of clusters remains constant, while in the other one the number of clusters is optimized. K-Means algorithm is an example of a method where the number of clusters remains constant. (Řezanková, Húsek, & Snášel, 2009)

One of the first objectives of a research over accident data is data cleansing and standardization, followed by selection of a suitable algorithm for clustering of similar accidents. Clusters created in this process will be further utilized in an accident early warning system.

ISBN: 978-80-245-2125-1

DOI: 10.18267/pr.2015.pav.2125.10

UTILIZING CLUSTER ANALYSIS RESULTS IN AN EARLY WARNING SYSTEM ABOUT A HIGH RISK OF TRAFFIC ACCIDENT

As was already described in (Lamr & Skrbek, 2015) this concept can be simply explained as a complex system which uses suitable models to predict accident risk in real time based on current vehicle location and other attributes. This early warning system about a high risk of traffic accident consists of two parts (control part and user part). A scheme showing the most important parts of the system is shown in Figure 1.

The control part of the system collects processes and distributes traffic accident data. Another task of the control part is a creation and distribution of prediction models. The database of traffic accident itself is a part of the control part. This database is automatically filled with accident data published at www.jdvm.cz. The data is extracted from public sources using suitable scripts and then transformed to be saved into the internal relational database itself. The internal database serves as a data source for traffic accident clusters for accidents with similar attributes.

The user part of the system evaluates actual situation in real time and informs the driver about a high risk of traffic accident in a suitable way. The user part can be realized either by a special device in the vehicle (client) which is connected to the vehicle through a standardized connector, or it can be a part of a cell phone or a navigational unit. The prediction of accident risk is carried out in real time, while the client evaluates the risk based on actual time, location, road condition, vehicle condition, weather and other attributes which reflect the situation. Most of such information is readily available in modern vehicles from various present sensors. The device in the vehicle compares the actual situation with prediction results and if there is a high level of similarity between the situation and prediction, the driver is warned. (Lamr & Skrbek, 2015) The driver is warned if there were multiple accidents under similar conditions on a given location.

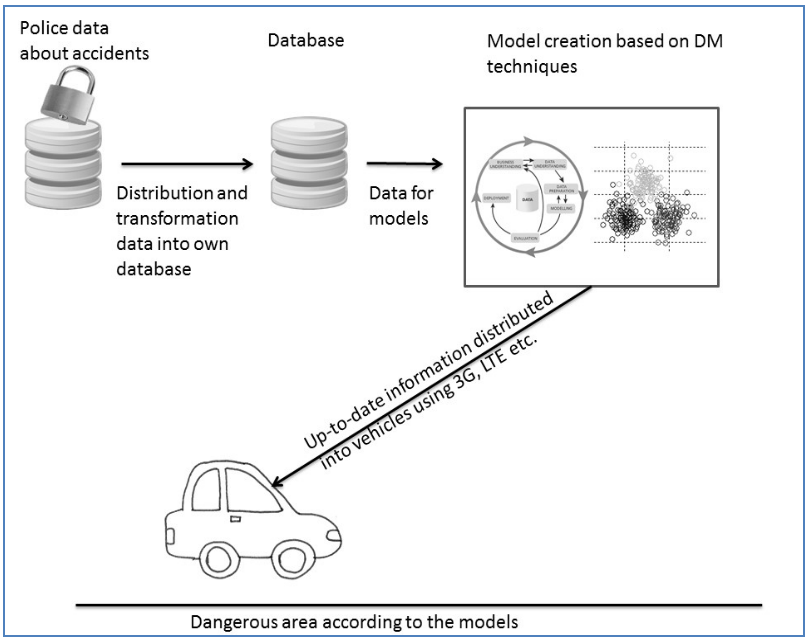

Figure 1. - Warning the driver in real time and place (Lamr & Skrbek, 2015)

It is necessary to determine how the control part should communicate with the client part in order to make the system efficient. Until now we have considered three possible methods of operation.

The first approach (Option A) assumes, that so called "Heat Maps" will be created and built into the base map in the device. These heat maps represent "hot spots" where multiple similar accidents happened. Analysis of such clusters should show, if there are common criteria for such clusters, for example time of the year, i.e. month, day time, temperature or other weather conditions. If an effect of

ISBN: 978-80-245-2125-1

DOI: 10.18267/pr.2015.pav.2125.10

some of the attributes is connected to a given location the cluster will be represented as a single point with specific properties. If a specific characteristic is not found for a given location, the cluster is labeled generally as a "location of frequent traffic accidents". The device in the vehicle monitors the movement of the vehicle in real time and if the vehicle is approaching a cluster representing a location of frequent traffic accidents the driver is warned about the fact. If the vehicle is approaching a location with specific conditions tied to the cluster the device compares these conditions with actual data and warns the driver about a high risk of traffic accident, because multiple accidents occurred at that location under similar conditions. It would be necessary to update the whole base map in the device regularly with generated clusters, but the device itself would not have to be online constantly to operate properly.

The second approach (Option B) uses the control part as a powerful database of traffic accident clusters and more operations are transferred to the client in user part. Unnecessary attributes for predictions would be removed from the database after thorough analysis, which would reduce the amount of accident data transferred to the client. Clusters in the control part would be created based solely on the amounts of traffic accidents and their geographical coordinates. The device in the vehicle would monitor the movement of the vehicle similarly to Option A but the client would download accident data for locations in the vicinity of the vehicle only. Each downloaded cluster would be tested according to the time of the year i.e. month, date, time, weather conditions and accident cause (for example: not keeping a safe distance behind other vehicles). This approach would not require regular massive base map updates for the client, because this data would not be stored in the device – only short-term data would be downloaded. The disadvantage of this approach is the necessity of constant internet connection of the client, which might pose a problem in locations with no signal. Hardware requirements on the client would also be higher, because the specificity tests (for example accidents that happened during a heavy rain at a nearby location) would take place in the device in real time.

The third approach (Option C) could be a combination of the first two. Similarly to option A, the control part would contain a database of general clusters formed according to the number of accidents at a given location as well as specific clusters tied to specific attributes. Data downloading would work similarly to option B, which means that only data necessary for actual prediction in the vicinity of the vehicle would be downloaded.

Only an in-depth analysis of traffic accident data obtained from the unified traffic map and further testing will suggest which of these three operational modes is the most suitable for real application.

CONCLUSION

The principle of an early warning system which was outlined in this paper should contribute to a system solution of prevention of frequent traffic accidents, which pose a significant life and property loss in the Czech Republic every year. This conceptual suggestion is in its early stages of development and at this moment we focus mainly on data preparation. The advantage of this system compared to other approaches is that it does not aim to solve consequences of crisis situations, such as eCall and RDS-TMC, but tries to actively prevent dangerous situations from happening.

REFERENCES

ČTK. (2015, August 17). Ministerstvo dopravy chce zvýšit pokuty za rychlost. Až trojnásobně. Retrieved November 13, 2015, from http://www.denik.cz/automoto-denik/ministerstvo-dopravy-chce-zvysit-pokuty-za-rychlost-az-trojnasobne-20150817.html

ČTK. (2015, September 1) V srpnu zemřelo na silnicích 70 lidí, o 12 víc než loni. Retrieved November 13, 2015 from http://zpravy.aktualne.cz/domaci/v-srpnu-zemrelo-na-silnicich-70-lidi-o-12-vice-nezloni/r~9d183314509e11e5adcb0025900fea04/

Kubát, D., Lamr, M., & Skrbek, J. (2015). New Approaches to Smart Solutions for eliminating Car Accidents. Proceedings of the 12th International Conference Liberec Economic Forum 2015 (pp. 392-401). Liberec: Vysokoškolský podnik, spol. s.r.o.

ISBN: 978-80-245-2125-1

DOI: 10.18267/pr.2015.pav.2125.10

Lamr, M., & Skrbek, J. (2015). The options for actively increasing road safety. In: Doucek, P., Chroust, G., & Oškrdal, V. *IDIMT2015 Information Technology and Society Interaction and Interdependence*. (pp. 487-494). Linz: Trauner Verlag.

Ministerstvo dopravy chce zvýšit pokuty za rychlost. Až trojnásobně. (2015). *DENÍK.CZ*. Retrieved from: http://www.denik.cz/automoto-denik/ministerstvo-dopravy-chce-zvysit-pokuty-za-rychlost-az-trojnasobne-20150817.html

Policie ČR. (2015, November 1). Statistika nehodovosti - Policie České republiky. Retrieved November 13, 2015, from http://www.policie.cz/clanek/statistika-nehodovosti-900835.aspx

Řezanková, H., Húsek, D., & Snášel, V. (2009). Shluková analýza dat. (2., rozš. vyd., p. 218). Praha: Professional Publishing.