Annals of DAAAM for 2011 & Proceedings of the 22nd International DAAAM Symposium, Volume 22, No. 1, ISSN 1726-9679
ISBN 978-3-901509-83-4, Editor B. Katalinic, Published by DAAAM International, Vienna, Austria, EU, 2011
Make Harmony between Technology and Nature, and Your Mind will Fly Free as a Bird
Annals & Proceedings of DAAAM International 2011

CONCEPTUAL MODEL FOR STRUCTURING TRAFFIC ACCIDENT DATA

ZOVAK, G[oran]; SARIC, Z[eljko] & COP, A[ndrej]

Abstract: *The wealth of data available on traffic accidents represents a large database with information on more than 50000 traffic accidents per calendar year. The aim of this paper is to outline a conceptual model of structuring traffic accident data, suggesting the use of a database which would through the data mining method provide valuable information incorporated into further analysis aimed at increasing traffic safety*

Key words: *data mining, traffic accidents, road safety, conceptual mode*

1. INTRODUCTION

Data mining is the process of finding useful information from large sets of data. A major advantage of this method is that it can be applied to almost every business activity. It provides a response to predictable questions referring to either expected or unexpected behaviour of the observed object. The result of data mining can be used in advising companies or their clients on their decision making. Furthermore, the method can also be a means of obtaining useful information from databases on traffic accidents. Such databases contain a massive amount of information covering more than 50 000 traffic accidents per calendar year. A clear indicator of the sheer size and complexity of such data storage is the fact that every form to be filled in case of a traffic accident contains 38 fields. In order to take full advantage of such a comprehensive database, the latter has to be well structured and organized in a way which enables a comparison of all data on several levels, providing the users with the relevant information in the shortest possible time.

2. DATA MINING

Researchers Ying, J, Jing, D, Chang-Tien, L. (Ying, J. et al., 2006.). defined Data mining as the search for patterns in data The data studied can be stored in databases or in various text data; they can appear as unstructured data or data organized into time slots. Mrsic (Mrsic, 2004.) defines two main types of data mining:

- verification of a hypothesis – aimed at verifying whether an idea or assumption of the importance of a relationship between certain data is well-grounded or not
- discovery of new knowledge – the possibility of discovering new information by offsetting certain phenomena

Data mining is particularly relevant in systems containing large amounts of data where it is possible to find the facts which previously did not seem to exist. Iyengar (Iyengar, 2004.) considered that Data mining has significant potential for instance in economics, mechanics, medicine, genetics, the pharmaceutical industry, telematics, etc. In general, data mining can be used in every area where an attempt is made at finding an underlying principle among a vast amount of data. Once the process of data mining reveals a number of useful rules, these should be linked up and formalized so as to enable a successful and appropriate exploitation of the newly discovered knowledge. In his paper Han consider (Han, 2000.) that the advantage of data mining is that it is independent of the subject matter in question since the focus is on the data itself rather than the area of the analysis: this in turn makes it suitable for structuring data on traffic accidents.

3. TRAFFIC ACCIDENTS DATABASE

Traffic accident data are organized in a database. The data are entered after an accident occurs, i.e. when police officers return to the police station.The problem arises when police officers enter the same information about a traffic accident several times. The first data are recorded on the scene of the accident and include personal information on the participants and the description of the accident. Upon arrival to the station, police officers re-enter the data on the accident for the purpose of drawing the investigation report, the criminal complaint, various checks of the vehicles and the people involved in the accident as well as filling in a statistical form whose data is used to plan and carry out preventive measures and the methodological work related with traffic safety. At some police stations with a weak hardware and software support the systems used to enter data on traffic accidents do not enable a simultaneous generation of multiple documents (record of the investigation, the statistical form, reports, etc.) and instead require re-entering of the same information for every document. Considering that in 2009 a total of 50 388 traffic accidents occurred in the Republic of Croatia (Table 1), it is safe to assume the amount of time police officers spent on rewriting the same data: a practice which is unnecessary and unacceptable in today's age of advanced computer systems. Saric (Saric, Z., et al., 2011.) consider that in all the documents that are filled in after the accident, the statistical form is the most important for the analysis of traffic safety, since it provides a basis for obtaining clear information on the accident.

Police administration	Traffic accidents								
	Total			With dead people			With injured people		
	2008.	2009.	+ - %	2008.	2009.	+ - %	2008.	2009.	+ - %
zagrebačka	16.319	**15.157**	-7,1	99	**82**	-17,2	3.359	**3.491**	+3,9
splitsko-dalmatinska	5.067	**5.037**	-0,6	55	**42**	-23,6	1.819	**1.745**	-4,1
primorsko-goranska	4.478	**4.135**	-7,7	35	**37**	+5,7	1.186	**1.204**	+1,5
osječko-baranjska	2.686	**2.530**	-5,8	36	**37**	+2,8	1.045	**1.018**	-2,6
istarska	3.530	**3.307**	-6,3	31	**33**	+6,5	1.070	**1.010**	-5,6
dubrovačko-neretv.	1.251	**1.190**	-4,9	23	**18**	-21,7	561	**482**	-14,1
karlovačka	2.372	**2.113**	-10,9	23	**19**	-17,4	517	**481**	-7,0
sisačko-moslavačka	1.940	**1.892**	-2,5	30	**23**	-23,3	756	**666**	-11,9
šibensko-kninska	1.273	**1.210**	-4,9	20	**17**	-15,0	497	**419**	-15,7
vukovarsko-srijem.	1.596	**1.494**	-6,4	30	**24**	-20,0	631	**594**	-5,9
zadarska	2.771	**2.723**	-1,7	24	**20**	-16,7	788	**753**	-4,4
bjelovarsko-bilogor.	1.153	**1.063**	-7,8	18	**17**	-5,6	346	**394**	+13,9
brodsko-posavska	1.650	**1.511**	-8,4	41	**20**	-51,2	648	**636**	-1,9
koprivničko-križev.	1.108	**959**	-13,4	13	**16**	+23,1	442	**384**	-13,1
krapinsko-zagorska	1.008	**1.018**	+1,0	19	**19**	0,0	397	**373**	-6,0
ličko-senjska	1.140	**1.116**	-2,1	16	**14**	-12,5	267	**286**	+7,1
međimurska	1.065	**879**	-17,5	11	**15**	+36,4	355	**290**	-18,3
požeško-slavonska	746	**751**	+0,7	12	**7**	-41,7	265	**263**	-0,8
varaždinska	1.586	**1.528**	-3,7	26	**15**	-42,3	458	**451**	-1,5
virovitičko-podrav.	757	**775**	+2,4	23	**18**	-21,7	291	**297**	+2,1
TOTAL	**53.496**	**50.388**	**-5,8**	**585**	**493**	**-15,7**	**15.698**	**15.237**	**-2,9**

Tab. 1. Number of traffic accidents in the Republic of Croatia

The data entered in the statistical form are permanently stored in the central information system which automatically processes the collected data. Subsequently an analysis is made of the collected and organised statistical data. in order to determine trends in accidents on Croatian roads, their temporal and spatial parameters as well as the causes and consequences for such accidents. On the basis of such analyses the police take the necessary measures and actions aimed at reducing the number of traffic accidents and increasing road safety. The form to be filled in for each traffic accident contains a total of 38 fields which police officers must fill in with details regarding the accident. Such data represent a quality knowledge base on traffic accidents. The advantage of such data is that the vast majority of them are correct. Although some irregularities may occur due to errors made while entering data, they can be ignored considering the simple method of filling in the form.

Other types of errors that can occur include subjective statements made by those involved in the accident or an investigation of the traffic accident scene carried out improperly. Based on the data referring to the number of traffic accidents in 2009, the traffic accident forms provided information on 50 388 accidents, which constitutes a large database that should be exploited to its fullest potential. With 38 fields within each of the 50 388 forms, there clearly is a need for a system that would enable a quality in-depth analysis of all that information in order to utilize all the advantages of such a database. Seen from this perspective, the data mining system lends itself as the ideal solution for data analysis.

4. CONCEPTUAL MODEL FOR STRUCTURING TRAFFIC ACCIDENT DATA

Considering the content complexity of traffic accident database, due to the quantity of data, there is a need to set up a model that would provide a clear and user-friendly view of reliable information. In order to achieve the desired quality it is necessary to enable connecting the relevant data together with the aim of ensuring reliable and accurate information. First of all, the system should be organised in a way that it dynamically links the necessary data while filtering out those that are irrelevant and displaying them only when requested by the user. Currently the part of the data available through the annual statistical report is not sufficiently interconnected and can only display up to two to three attributes compared with one another. It is for example possible to find out the number of accidents occurring on roads in poor condition but not also who were in most cases the people responsible for causing accidents on roads in poor condition nor what were the consequences. It is therefore necessary to maximize the correlation of data in order to outline the interrelationship between the data connected in this way. To this aim, and by taking into account the vast amount of data, a hierarchical structure should be formed containing on the first level the following elements:

- type of accident
- the vehicle involved in the accident
- the people involved in the accident
- the location of the accident
- the time and weather conditions
- the cause of the accident
- the consequences of the accident

All seven sets are interconnected so that for a single accident, different sets can be associated with one another.
Figure 1 shows the relational model of data connectivity on the *first level*, i.e. for the main volume. Together these carriers represent data sets which in turn account for individual entries (tables); however, in order to make the system function properly, they have to be, as already stated, linked together.

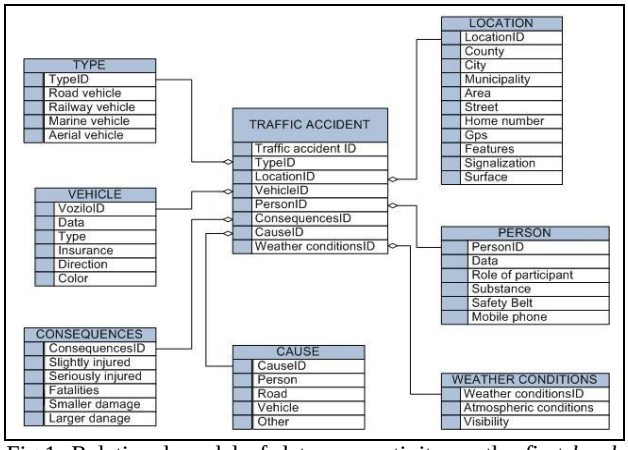

Fig.1. Relational model of data connectivity on the *first level*, i.e. for the main volume

For this reason, every accident should be awarded an identification number (ID) which would then be used to identify a certain accident during the search through data.

Such an ID would enable the system to determine to which car accident certain data belong. Furthermore, an identification number should also be attributed within every set of data carriers (*PERSON, LOCATION, CAR, CAUSE* ...) so as to enable a correlation of data from different data sets for the same accident.

5. CONCLUSION

Statistical data on traffic accidents are the most important element in determining and designing actions to increase traffic safety. The structuring of data on traffic accidents is a demanding and complex problem.
Based on the analysis of the current situation, it is evident that while a comprehensive body of data already exists, there is a lack of sufficient correlation between them.
The statistical report on road safety provides individual information on the number of traffic accidents, the people who most often cause them, the circumstances in which they occur etc. It is however not possible to find information on several levels, with correlated data on traffic accidents, e.g. which types of accidents most frequently occur during bad weather on the roads outside the city and what is the most likely age of the drivers who cause them.
Such data would on the one hand provide a detailed insight into the issue of what causes traffic accidents and on the other prove a useful tool in defining concrete steps towards accident prevention.

6. REFERENCES

Han J., Kamber M. *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2000

Iyengar, V. *On detecting space-time clusters*, Proceedings of ACM Conf. on Knowledge Discovery in Data Mining, pp. 587-592, 2004

Mrsic, L. *Applying data mining methods in the trade of textiles and related products*, Faculty of economics and business Zagreb, 2004

Saric, Z, Zovak, G., Koronc, N., „*Comparison of methods for determining crash hotspots in the road traffic*"**,** Proceedings of the Scientific-technical union of mechanical engineering, 19th International Conference trans&MOTAUTO'11, 1313-5031,Varna, Bulgaria 2011

Ying,J, Jing,D, Chang-Tien, L.: *Spatial-Temporal Data Mining in Traffic Incident Detection*, SIAM DM 2006 Workshop on Spatial Data Mining, Maryland, 2006