MEDINFO 2010
C. Safran et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-588-4-510

Traffic Accident In Cuiabá-Mt: An Analysis Through The Data Mining Technology

Noemi Dreyer Galvão[a], Heimar de Fátima Marin[b]

[a] *Professor, Federal University of Mato Grosso - UFMT, Technician of the Health State Secretariat of Mato Grosso – MT, Brazil*
[b] *Professor, Federal University of São Paulo - UNIFESP - São Paulo -SP, Brazil*

Abstract

The traffic road accidents (ATT) are non-intentional events with an important magnitude worldwide, mainly in the urban centers. This article aims to analyzes data related to the victims of ATT recorded by the Justice Secretariat and Public Security (SEJUSP)in hospital morbidity and mortality incidence at the city of Cuiabá-MT during 2006, using data mining technology. An observational, retrospective and exploratory study of the secondary data bases was carried out. The three database selected were related using the probabilistic method, through the free software RecLink. One hundred and thirty-nine (139) real pairs of victims of ATT were obtained. In this related database the data mining technology was applied with the software WEKA using the Apriori algorithm. The result generated 10 best rules, six of them were considered according to the parameters established that indicated a useful and comprehensible knowledge to characterize the victims of accidents in Cuiabá. Finally, the findings of the associative rules showed peculiarities of the road traffic accident victims in Cuiabá and highlight the need of prevention measures in the collision accidents for males.

Keywords

Traffic accidents, Data mining, Public health informatics

Introduction

The traffic road accidents are non-intentional events with an important magnitude worldwide, mainly in the countries with medium and low income. The numbers show high number of deaths, disabilities and psychological sequelae, being one of the negative factors of the epidemiological transition that Brazil has experienced over the last decades [1-2].

Brazil is among the countries that have registered the largest frequencies of individuals injured in traffic accidents, with rates showing that in 2002 the number of victims were 219,5 per 100 thousand inhabitants [3] and 18,7 deaths per 100 thousand inhabitants [4].

In 2006, death by traffic road accidents were in a rate of 39,4 per 100 thousand inhabitants. In the same year, the state of Mato Grosso, presented a risk of 56,8 per 100 thousand inhabitants, and the capital Cuiabá recorded a lower rate of 25,2 per 100 thousand inhabitants [4].

The relevance of the traffic accidents in Cuiabá is revealed by the mortality rate per external causes. The data of hospital morbidity in Cuiabá is also higher than the indexes presented by the national record, the Center-West region and by the State of Mato Grosso [5].

This article presents an analysis of relating data in traffic road accidents recorded by the Justice Secretariat and Pubic Security using hospital morbidity and mortality records at the city of Cuiabá in 2006, using data mining technology.

Materials and Methods

Type of study

It is an observational, cross-sectional, retrospective and exploratory study of the secondary databases of security and public health in the municipality of Cuiabá, Mato Grosso State, in 2006.

Source of data and ethnic considerations

The sources of data were those of the State Secretariat of Justice and Public Security (SEJUSP) that made available the database of the victims in traffic road accidents assisted by the Military Fire Brigade and SAMU including the Police Records.

Complementing the database of the public security area, we included data from the national health information systems which contain the register of the accidents and the hospital admissions records at SUS - the Hospital Information System of the Unified Health System (SIH/SUS). In addition, data from the system for recording deaths, the Mortality Information System (SIM) were also included.

The databases and the collecting data files or instruments of data collection, were provided by the Municipal Health Secretariat in Cuiabá-MT and by the State Secretariat of Justice and Public Security (SEJUSP), authorized by their directors and approved by the Ethic Committee at the Federal University of São Paulo under protocol number 1595/07.

Data processing

The databases received from SEJUSP (the Municipal Health Secretariat of Cuiabá-MT), SIH/SUS and SIM were structured in an Excel datasheet, Microsoft Office Excel 2003, comprising 2,547; 38,273 and 3,198 registers, respectively.

A manual checking of the databases was made to investigate the existing variables to verify key attributes for record linkage. The phase of the selection of the registers in the databases took into account all the registers that had the key attributes. In this cleaning phase 1.137 registers were excluded from the SEJUSP databases totalizing 1.410 selected registers.

In the SIH/SUS all the registers were selected (38.273). However, in the SIM database, there was an exclusion of 97 registers. Consequently, the total was 3.101 selected death records.

Data processing was divided into two phases: (a) the record linkage of the databases and (b) the data mining technology.

Record Linkage

The database linkage probabilistic method was used to join the data of the Information System of SUS with the data from the State Justice Secretariat and from Public Security (SEJUSP) to obtain information about common individuals that were victims of traffic accidents in the city of Cuiabá-MT.

Three record linkages were carried out separately, SEJUSP *versu.* SIH/SU which resulted 111 pairs; SEJUSP *versus* SIM with 25 pairs; and with the registers of the first record linkage SEJUSP/SIH *versus.* SIM which resulted three pairs, totalizing 139 real pairs. This process formed a single database with 139 registers. Subsequently, the 13 variables were selected as showed in Table 1.

Data mining

The tool used to obtain knowledge from the data bases of accidents in Cuiabá was the WEKA - *Waikato Environment for Knowledge Analysis.* The reasons for selecting it were: it is a tool developed in the JAVA language that has as the main feature the portability (it is easy to be executed in several platforms of the Operational System); has the source code open; easy access by internet and, available at http://www.cs.waikato.ac.nz/ml/weka/.

After processing the formating of the variables and the structure of the database for data input at the selected tool, the Weka software version 3.6.0 was prepared. It reads a text formatted file with .arff extension. In addition, in the pre-processing phase of tool, the preliminary analysis of the data was done. Next, it was perfomed the selection of the association (*Associate)* in the WEKA tool. The selected algorithm was APriori - which consists in identifying and describing associations among the variables of the same item or associations among different items that frequently occur simultaneously in the data base [6-7].

Table 1 – Variables, description and their categories

Variable	Description	Categories
VictimCode	Victim Code	Arabic number
	Gender	M, F
Age group	Age group (from 0	0-9; 10-19; 20-29; 30-39; 40-49; 50-59; 60 years and more
Month of occurrence	Month of occurrence	Jan, Feb, Apr, May, June, July, Aug, Sept, Oct, Nov, Dec
Type of Accident	Type of accident	collision, overrunning, overturn, shock, vehicle crash, fall, none
Victim's transport	Victims's means of transport	On foot, vehicle, motorcycle, automobile, bicycle, none
Type of victim	Type of victim	driver, passenger, pedestrian, ignored
Days of hospitalization	Days of hospitalization	Arabic number
Residential zone	Residential Zone	north, south, east, west, none
Ocurrence zone	Occurrence zone	north, south, east, west, none
Medical Assistance	Medical assistance	yes, no
Place of assistance	Place of Assistance	HCANCERMT, HGU, HUJM, HJDCBA, HMBOMJESUS, HPSMCBA, HSTAHELENA, HSTAROSA, SOTRAUMA, STACASA, None
Evolution	Evolution	Hospital discharge, death

As parameters for the execution of the association the *default values* of the software 0,1 (10%) minimum support (sup.); 0,9 (90%) confidence (conf.); 1,1 minimum Lift and number of rules 10 were maintained.

Results

With the application of the APriori algorithm and its parameters previously established, three sets of data were created. The first comprises five attributes which were combined to create a second set (*itemsets)* that contain six combinations; each one with two attributes. With the combination of the second *itemsets,* a third item/sets was created with two combinations and three variables.

In the three *itemstes* created, it is observed that the supports of each combination are higher than the minimum (10%) suggested by the *default* of the WEKA software. Of the last two combinations in the third *itemsets*, it is verified that 64,0% of the 139 male victims received medical assistance and were discharged from hospital.

In the second combination, of the 139 victims, 41,0% occurred by collision and also had medical assistance, receiving hospital discharge. After the combinations of the *Itemsets*, the 10 best associative rules created by the *APriori* algorithm were ordered by the *lift*. Of the 10 best rules created by the algorithm,

four were not considered since they presented one of the parameters smaller than what was established.

In addition, of the 10 rules created by the algorithm of association of the traffic accident victims it was verified that, in all the rules created, the attributes referring to hospital medical assistance (Yes) and evolution (discharge) were present, as antecedent or consequent. Two other frequent variables in the rules created were the type of accident by collision and the male gender. The frequency of these variables in the rules indicates possible standards found in the related data.

In relation to the rules, it can be observed that the first presented confidence of 92,0% (0,92), which is higher than the minimum established, (TP_ACCID=COLLISION - MED. ASSIST=SIM 62 ==> EVOL=H. DISCHARGE 57). Also, it can be observed that 92,0% of the victims according to the type of accident, more specifically, the collision, received medical assistance and were discharged from the hospital, which means that 62 collision accident victims received hospital medical assistance and 57 were discharged. Five victims died. Related to the support of 41,0% (also higher than the minimum established) of the 139 victims of traffic Road accidents that occurred in the city, 41,0% were due to collision, received medical assistance and were discharged from the hospital. This rule was classified as the Best rule by the *lift* parameter. The *lift* of the rule (1,16) was higher than the minimum *lift* (1,1) established in the *default* of the *software*.

The second rule, (MED-ASSIST=SIM 121 ==> EVOL=h. discharge 110) shows that 121 victims of accidents received hospital medical assistance and 110 presented evolution for hospital discharge with confidence level of 91,0%. From the total number of the Road traffic accident victims, 79,0% received medical assistance and were also discharged.

The third rule shows that 100,0% of the victims were discharged from hospital because they received medical assistance.

In the fourth rule, the variables male gender and hospital discharge appear as antecedents, indicating that 100,0% of the male victims were discharged because they received hospital assistance. It also can be observed that 64,0% of the total number of accident victims were males and also were discharged because received hospital assistance.

The fifth rule shows that 100,0% of the traffic accident victims by collision were discharged also due to medical assistance. But of the 139, only 57 (41,0%) collision accident victims received hospital discharge due to medical assistance.

The last rule considered by the parameters higher than those established, shows that 90,0% of the victims were males and received medical assistance, and because of this factor they were discharged.

Discussion

In this study the use of the data mining technology was presented through a data base related by the probabilistic method, since the Brazilian health databases do not present a common single identifier. Leles [8] highlighted the importance of the probabilistic method to integrate databases, mainly those database in the healthcare area.

Through the Apriori algorithm, the six rules selected, have parameters higher than those established. The knowledge obtained from the rules was comprehensible in consonance with the results and previous studies performed in Cuiabá-MT [9], and Cuiabá surroundings [10-11]. However, it is worth mentioning that these studies did not use the data mining technology to analyze the data.

The two variables that constantly appear in the subsets of the created rules indicated that there is an association between medical assistance and evolution to hospital discharge. It is also known that these can be called valid standards [12], since they were frequent presented in all the subsets. Two other frequent variables were the male gender and the type of accident by collision. The result (standards) is very similar to that found in other studies carried out by other methods and technologys of analysis in Cuiabá-MT [9], in the surroundings of Cuiabá [10-11], Alta Floresta-MT [13] and in the Distrito Federal [14].

It is worth emphasizing that the rules created, though trivial, show that the type of accident by collision makes victims and fatal victims; illustrating that all the types of vehicles are not invulnerable armors and that the drivers need to be prepared to drive.

In the rules presented, it is also emphasized that males were the main accident victims, even if a large number received medical assistance, and were discharged. It is known however, that the situation interferes in the family economy, since the most affected age group was that of the individuals economically active. There is also the immeasurable social loss that is caused by the trauma. It is difficult to evaluate quantitatively the social impact of the sequelae that the trauma causes for the population; on the other hand, there is a concern about the financial cost of their attendance (medical assistance), follow-up rehabilitation.

Finally, despite a few rules created and considered, there are enough elements to propose preventive measures of these accidents by collision involving male drivers. It is highlighted that these rules were based on real data that effectively show the scenery and can support programs dedicated to prevent violence in the traffic.

Conclusion

Finally, the findings in the associative rules revels the peculiarities of the road accident victims in the city of Cuiabá, Mato Grosso State, in 2006. It emphasize the need for preventive measures in the accidents by collision and also educational program dedicated to the male gender for the reduction of these events. The public policies on transportation, the guiding plan of the city, and the statute of the city must be reevaluated

Finally, the mining data technology can be considered a powerful tool in the analysis of secondary data, even with related data, and also to help the decision-making process, provinding

useful knowledge of databases from the health information systems and public security.

References

[1] Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise de Situação em Saúde. Mortalidade por acidentes de transporte terrestre no Brasil / Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Análise de Situação em Saúde. – Brasília : Ministério da Saúde, 2007. 80 p

[2] Duarte EC, Duarte E, Sousa MC, Tauil PL, Monteiro RA. Mortalidade por acidente de transporte terrestre e homicídios em homens jovens das capitais das Regiões Norte e Centro-Oeste do Brasil, 1980-2005. Epidemiol. Serv. Saúde. 2008; 17(1):7-20.

[3] OMS Organización Mundial de la Salud. Informe mundial sobre prevención de los traumatismos causados por el tránsito: resumen. Washington DC: Organización Mundial de la Salud; 2004.

[4] [MS] Ministério da Saúde. Departamento de Informática do Sistema Único de Saúde (DATASUS) [homepage na Internet] Banco de dados dos Sistemas de Informações sobre Mortalidade (SIM). Brasil; 2009. acesso em 29 jul 2009. disponível em http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/obtmt.def

[5] [MS] Ministério da Saúde. Departamento de Informática do Sistema Único de Saúde (DATASUS) [homepage na Internet] Banco de dados dos Sistemas de Informação de Morbidade Hospitalar.Brasil; 2008. [acesso em 10 abr 2008]. Disponível em: http://w3.datasus.gov.br/datasus.

[6] Bayam E; Liebowitz J; Agresti W. Older drivers and accidents: A meta analysis and data mining application on traffic accident data. Expert Systems with Applications. 2005; 29: 598–629.

[7] Abbott PA; Lee SM. Data mining and Knowledge Discovery In: Saba VK, Mccormick KA. Essentials of Nursing Informatics. Four edition, Nova Iorque: Mcgraw Hill – Medical Publisning Division; 2006.

[8] Leles AD. Técnicas de *record linkage* e inteligência artificial para a construção de um *data warehouse* aplicado à área de Saúde. Revista Científica do IMAPES. 2004;2(2):41-46

[9] Oliveira LR, Mello Jorge MHP. Análise epidemiológica das causas externas em unidades de urgência e emergência em Cuiabá/Mato Grosso. Rev Bras Epidemiol. 2008; 11(3): 420-30.

[10] Soares BAC, Scatena JHG, Galvão, ND. Evolução e Características Da Morbidade por Acidentes e Violências na Grande Cuiabá - Mato Grosso. Revista Espaço para a Saúde. 2008; 9(2):26-38.

[11] Soares BAC, Scatena JHG, Galvão, ND. Acidentes e Violência na Grande Cuiabá: o que retrata a demanda dos serviços de emergência. Epidemiol. Serv. Saúde. 2009; 18(3):265-276.

[12] Fayyad UM, Shapiro GP, Smyth P, Uthurusamy R. "Advances in Knowledge Discovery and Data Mining". Califórnia: American Association for Artificial Intelligence; 1996; 611p.

[13] Marchese VS, Scatena JHG, Ignoti E. Caracterização das vítimas de acidentes e violências atendidas em serviço de emergência: município de Alta Floresta, MT (Brasil). Revista Brasileira de Epidemiologia 2008;11(4):648-659.

[14] Modelli MES, Pratesi R, Tauil PL. Alcoolemia em vítimas fatais de acidentes de trânsito no Distrito Federal, Brasil. Rev. Saúde Pública 2008 42(2): 350-352

Address for correspondence

Noemi Dreyer Galvão
Rua Veneza, 105, Bairro Jardim Itália, Cuiabá – MT, BRASIL
CEP 78060-810. e-mail: noemidgalvao.mt@terra.com.br-

The Presenting
Rua Napoleão de Barros 754 Vila Clementino, São Paulo - SP
BRASIL, CEP04024-002 –
e-mail : hfmarin@unifesp.br