Sensors and Materials, Vol. 36, No. 3 (2024) 1243–1249
MYU Tokyo

S & M 3595

Application of Big Data Analysis of Traffic Accidents and Violation Reports for Improving Traffic Safety

Hung-Cheng Yang,* Mu-Quan Chen, and I-Long Lin

Department of Computer Science and Engineering, Tatung University
No. 40, Sec. 3, Zhongshan N. Rd., Taipei City 104, Taiwan (R.O.C.)

(Received January 3, 2024; accepted March 19, 2024)

Keywords: traffic accident file, traffic violation file, traffic information, Tableau software, big data analysis

The causes of traffic accidents are diverse, including weather, road conditions, road design, and psychological factors. With the advancement of information technology, big data on traffic accidents can be collected and analyzed more easily than before. To identify the causes of traffic accidents, we analyzed the Traffic Enforcement Case Database and Traffic Accident Database of the Traffic Division of the National Police Agency in Taiwan. The main causes of traffic accidents from 2013 to 2020 were lane drifting, overspeeding, illegal turning, running red lights, and drunk driving. The number of traffic violations has increased every year in the same period, and the number of casualties and injuries has increased since 2018. It is necessary to customize sensor technologies to monitor such violations to prevent related accidents. Advanced data mining technologies should be used to analyze the data and obtain better information to prevent violations and accidents. The results of this study provide a basis for further study related to developing preventive measures for traffic violations and accidents using advanced sensor technologies.

1. Introduction

Traffic accidents threaten lives and damage properties. They also cause traffic congestion, violate the rights of road users, and burden the public with social costs related to traffic management, medical care, insurance, and others. The causes of traffic accidents are diverse and include physical factors such as the conditions, signs, and design of the road, and weather, as well as human factors such as violation or negligence. If the main causes can be found, related traffic accidents can be prevented effectively: this necessitates the analysis of traffic accidents and their causes. In Taiwan, the total number of traffic violations caught by the police is increasing every year.(1,2) The increasing number of traffic violations on the road may make drivers and pedestrians nervous and involved in similar violations. Therefore, to reduce the occurrence of traffic accidents, the three E's of traffic management are important: education, engineering, and enforcement.

*Corresponding author: e-mail: yangyeh5046@gmail.com
https://doi.org/10.18494/SAM4868

ISSN 0914-4935

Research on the causes of traffic accidents has been conducted mostly to analyze accident patterns and individual behavior in accidents. However, the relationship between traffic violations and accidents has been little studied as the potential impact of violations on traffic accidents has not been paid much attention by the authorities and insurance companies. Therefore, it is necessary to study how to use the data on traffic accidents and violations to determine the main causes of accidents on the road. The results can be used to establish law enforcement and safety measures to reduce the occurrence of traffic accidents. For effective law enforcement and prevention of violations and accidents, it is necessary to develop appropriate sensor technology and deploy developed sensors and cameras in the right places. To enhance safety and prevent accidents, several cameras with sensors, including red-light cameras, speed cameras, and traffic surveillance cameras, are used. Loop sensors, CCD sensors, radar, and light detection and ranging (LiDAR) are used for such cameras.

Using the data of A1 accidents (causing the death of people involved in traffic accidents on site or within 24 h) and A2 accidents (causing injury to the people involved or death after 24 h of accidents) in Taiwan from 2013 to 2020, we analyzed the causes of accidents, time, road conditions, and accident rates and the correlation between traffic accidents and violations. The results can be used to propose effective measures to prevent traffic violations and accidents and enhance road safety. The results can also help improve appropriate sensor technology for effective monitoring and enforcement.[3]

2. Methodology

2.1. Databases

We used the Traffic Enforcement Case Database (TECD) and Traffic Accident Database (TAD), which are managed by the Traffic Division of the National Police Agency in Taiwan. TAD contains the original files of A1 and A2 accidents from 2013 to 2020, and TECD includes the data of the National Traffic Violation File from 2013 to 2020.

2.2. Big data analysis

Big data has features known as 4V: volume, veracity, variety, and velocity. As data is generated at any time and any place, its amount can be too large to handle. Thus, the first task in dealing with big data is to store the data safely and efficiently. Distributed processing is used for the data storage while backing up the data. To use the stored data effectively, the rules of data processing must be defined. Decision trees, genetic algorithms, and artificial neural networks are widely used nowadays. Then, visualization tools are used to present data in a manner that is sufficiently easy to understand.

In processing the collected big data, we used the Hadoop distributed file system (HDFS) to split the data into smaller files and store them in distributed locations. This technology enables the handling of continuously increasing data to ensure data integrity. Hadoop MapReduce was used to process the data with Spark since Spark is 100 times faster than Hadoop MapReduce. As

Hadoop MapReduce stores the data while computing, the data must be continuously transmitted between the memory and the processor. However, Spark uses in-memory computing to save energy and time in data transmission. Because Spark analyzes data that are stored, it needs to be combined with Hadoop MapReduce. As a visualization tool, we used Tableau. Tableau was used to analyze and visualize big data, SQL databases, spreadsheets, and cloud applications in on-premises or cloud environments. Multiple data sources can be integrated to transform data for further analysis. Charts and graphics can be created to understand the results of data analyses. Various data formats including Excel, txt, and XML are used in Tableau, and its operation is simple and easy. Wordcloud was also used to analyze the frequency of words in the document.[3]

2.3. Statistical methods

We explored the relationship between traffic accidents and violations by using statistical methods to analyze the data. Frequency analysis and cross-analysis were used to understand the characteristics of violations and accidents. Correlation and multivariate analyses were conducted to understand the relationship between major causes, violations, and accidents. We used a logistic regression model to analyze categorical data. Through the analysis, the relationship between the categorical response and explanatory variables was found. The results of logistic regression analysis are more meaningful than those of conventional regression analysis for discrete data.

Hypotheses were tested using the chi-square test, goodness-of-fit test, independence test, and homogeneity test.[4,5] We explored the frequency or proportion distribution of the data. The goodness-of-fit test was conducted to test whether the population distribution of the data followed a specific theoretical distribution. The homogeneity test was employed to test whether the distribution of each category of two or more populations was the same or similar. In the homogeneity test, contingency tables were used to categorize the data. The independence test was used to test whether two variables were independent. The independence test result was presented in the form of a contingency table or cross table. Nonparametric statistical methods were used to estimate or test variables without parameters. Because this statistical method does not require a specific distribution of the population, data sorted by magnitude or occurrence order can be analyzed.

2.4. Analysis procedure

We integrated various data sources to connect the data in cloud environments, including big data, SQL databases, spreadsheets, and cloud applications such as Salesforce using Tableau. Visualization tools were used to draw charts and graphs to a better understand the data. We set data ranges or conditions with Tableau's filters using parameters to increase flexibility in analysis and adjust parameters. Using Tableau's analysis tools, we performed trend analysis and correlation analysis.[6,7] The analysis results were presented in interactive dashboards to support related decision-making.

3. Results and Discussion

3.1 Traffic accidents and casualties

From 2013 to 2023, there were over 400000 national A1 and A2 accidents with an annual average of 1751 casualties and 442601 injuries (Fig. 1). The casualties from traffic accidents within 24 h decreased from 2013 to 2018, and then began to increase from 2019.[2]

3.2 Traffic violations

From 2013 to 2023, an average of 11.92 million traffic violations per year were reported. There were 8.05, 8.17, 9.28, 10.11, 10.53, 11.25, 12.7, 14.64, 13.86, 15.02, and 17.55 million violations from 2013 to 2023 (Fig. 2), which showed a steady increase every year.

3.3 Traffic violations causing accidents and casualties

To maintain traffic safety, the police of Taiwan has promoted the Plan to Strengthen the Suppression of Major Traffic Violations, which includes 10 major traffic violations: drunk driving, running red lights (excluding red light when turning right), overspeeding, wrong-way driving, sudden turning, lane drifting, driving on the shoulder, large vehicles and slow-speed vehicles driving outside designated lanes on a highway, driving on prohibited lanes, and illegal turning.[5] The top five violations were lane drifting, overspeeding, illegal turning, running red lights, and drunk driving. In A1 or A2 accidents, the death rate [number

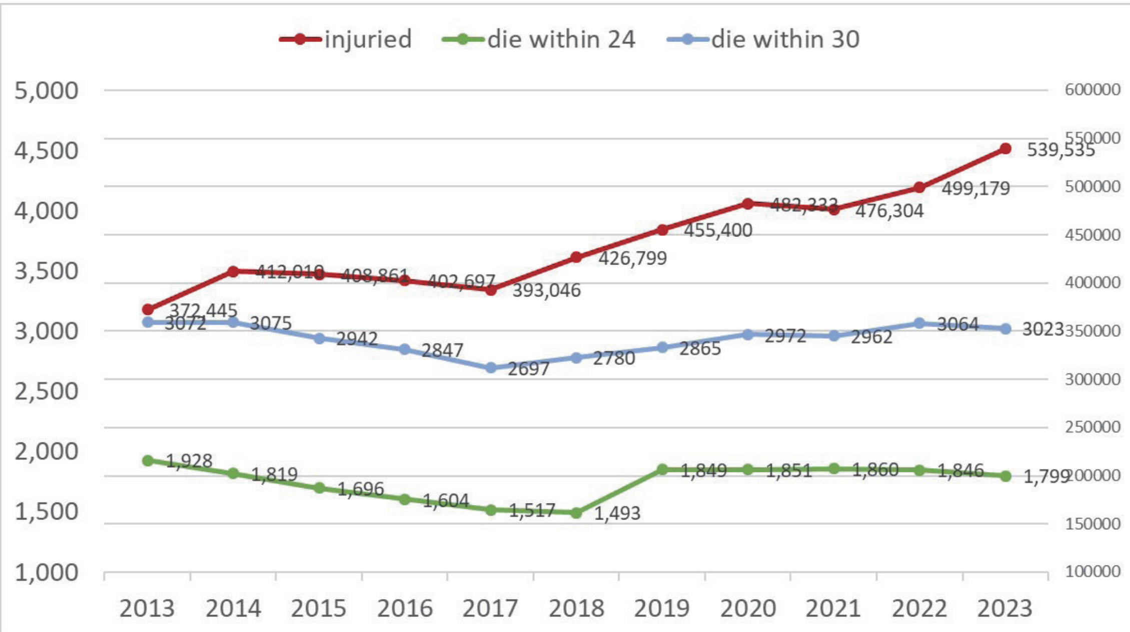

Fig. 1. (Color online) Trend of traffic accidents, casualties, and injury from 2013 to 2023.

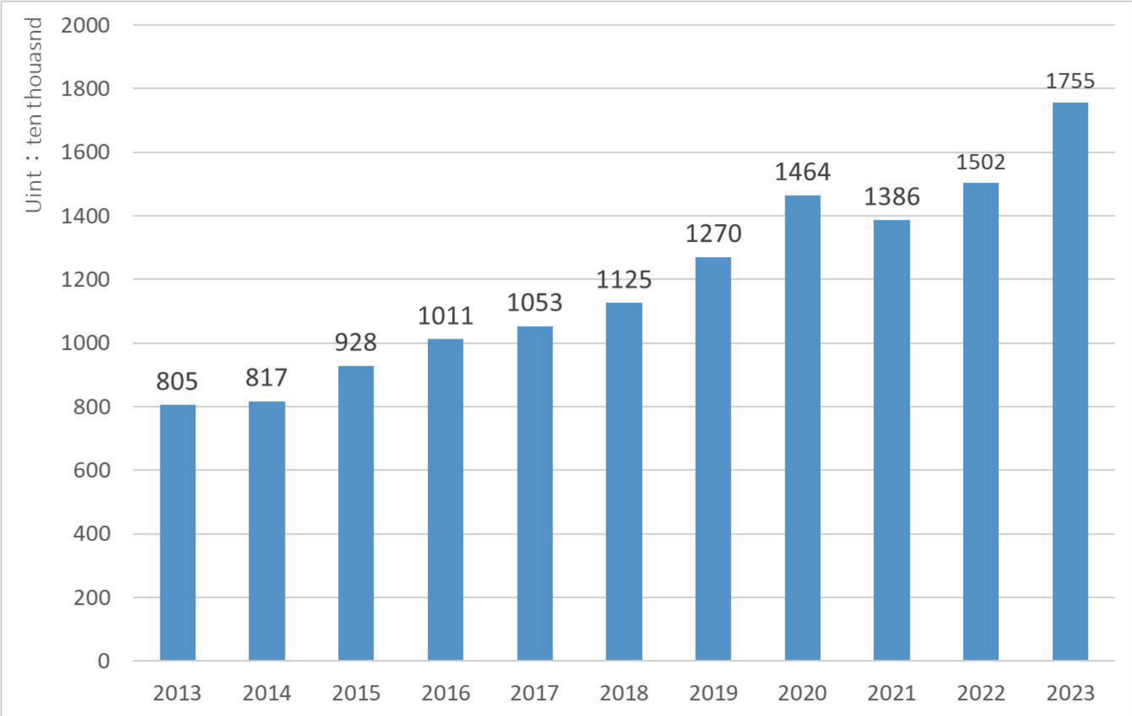

Fig. 2. (Color online) Traffic violations from 2010 to 2023

of deaths/(number of deaths + number of injuries) × 100] was calculated. The top five main violations for casualties were drunk driving, overspeeding, illegal turning, running red lights, and driving on prohibited lanes.[1,2]

3.4. Discussion

The police have promoted the Plan for Preventing Major Traffic Violations and strengthened law enforcement for traffic violations. For the effective prevention of violations and enhanced traffic safety, it is necessary to use technology and analyze related data. Road design must also be improved to secure safety, and appropriate education is necessary for drivers and pedestrians to obey the law and become more cautious on the road.[8]

The presence of police officers on the road prevents overspeeding and other traffic violations significantly. Surveillance cameras and speed sensors were proven to be far more effective in preventing traffic violations and accidents than the law enforcement by police officers on the road as the use of cameras reduced violations by 40−50% and casualties by 25−30%. Therefore, it is important to use cameras to monitor traffic violations on the road. There are several types of cameras used for monitoring and preventing violations, including red-light cameras, speed cameras, and traffic surveillance cameras.[9]

Using the TECD and TAD, we determined that lane drifting, overspeeding, illegal turning, running red lights, and drunk driving are the main causes of casualties on the road. Thus, it is necessary to develop and install appropriate sensors and cameras to detect such violations on the road and in vehicles to prevent fatal accidents. Nowadays, inductive loop sensors, CCD sensors, radar, and LiDAR are used for the monitoring and enforcement of traffic violations and accidents. The results of this study reveals a need to develop or improve the performance of sensor technology to monitor and prevent violations more effectively.(10)

4. Conclusion

Using the TECD and TAD of the Traffic Division of the National Police Agency in Taiwan, we explored how traffic violations caused accidents and affected the number of casualties from the accidents from 2013 to 2020. The main violations causing many casualties were lane drifting, overspeeding, illegal turning, running red lights, and drunk driving. The number of traffic violations caught by the police has increased every year. The number of casualties and injuries caused by traffic accidents had been increasing since 2018. To prevent violations and accidents in traffic, it is important to implement enforcement, education, and engineering effectively.(11) For enforcement and engineering, appropriate technology and its deployment on the road are significant in enhancing road safety. It is necessary to develop and customize various sensor technologies, including inductive loop sensors, CCD sensors, radar, and LiDAR, and to keep analyzing the data of violations and accidents. To obtain better information on violations and accidents, advanced data mining technologies such as decision trees, neural networks, and genetic algorithms should be used.(12) Along with such advanced technology, socioeconomic factors must be investigated to explore the relationship between violations and accidents and establish preventive measures using appropriate technology.(13)

References

1 National Police Agency: https://www.npa.gov.tw/en/app/data/list?module=wg055&id=8026 (accessed March 2024).
2 New Taipei City Police White Paper: https://www.police.ntpc.gov.tw/dl-1409-3cc7aefe-cd0d-4e71-afb3-dd19d31f2646.html (accessed March 2024).
3 Bigdata Analytics: https://www.techtarget.com/searchbusinessanalytics/definition/big-data-analytics (accessed March 2024).
4 A. Almeida, S. Brás, S. Sargento, and I. Oliviera: J. Urban Aff. **4** (2023) 100065. https://doi.org/10.1016/j.urbmob.2023.100065
5 L. Hu, X. Bao, H. Wu, and W. Wu: J. Adv. Transp. **2020** (2020) 9084245. https://doi.org/10.1155/2020/9084245
6 W. Frawley, G. Piatetsky-Shapiro, and C. Matheus: AI Mag. **13** (1992) 57. https://doi.org/10.1609/aimag.v13i3.1011
7 R. Stahlbock, S. Lessman, and S. F. Crone: Data Mining: Speical Issue in Annals of Information System (Springer, Berlin, 2009) 1–15. http://doi.org/10.1007/978-1-4419-1280-0_1
8 A. S. Al-Ghamdi: Accid. Anal. Prev. **35** (2003) 717. https://doi.org/10.1016/S0001-4575(02)00050-7
9 A. Fang, C. Qiu, L. Zhao, and Y. Jin: Proc. 2018 IEEE Int. Conf. Intelligent Transportation Engineering (ICITE, 2018) 291. http://doi.org/10.1109/ICITE.2018.8492665
10 Transporation Dissertation: https://www.iot.gov.tw/cp-144-72364-fee4e-2.html (accessed March 2024).
11 Transporation Dissertation: https://www.iot.gov.tw/cp-144-210557-3c8ee-2.html (accessed March 2024).
12 E. Amoros, J. L. Martin, and B. Laumon: Accid. Anal. Prev. **35** (2003) 537. https://doi.org/10.1016/S0001-4575(02)00031-3
13 A. H. Lee, M. R. Stevenson, K. Wang, and K. K. W. Yau: Accid. Anal. Prev. **34** (2002) 515. http://doi.org/10.1016/S0001-4575(01)00049-5

About the Authors

Hung-Cheng Yang received his bachelor's degree in traffic management in Taiwan in 1984 and his master's degree in information management in 2003 from the Central Police University in Taiwan. He has been studying for his Ph.D. degree in computer science and engineering at Tatung University since 2021. Since 1984, he has worked for the police. His research interests include digital evidence, forensics, and traffic accident prevention using big data. (yangyeh5046@gmail.com).

Mu-Quan Chen graduated from the National Open University, Taiwan in 2012 and received a master's degree in innovation management from Yuanpei University of Medical Technology in 2016. Since 2021, he has been studying computer science and engineering at Tatung University, Taiwan. He has worked for the police since 1989. His research interests are in the prevention and identification of online fraud and misinformation.
(muchuan1968@gmail.com).

I-Long Lin received his B.S. degree from Central Police University, Taiwan, in 1983, and his M.S. and Ph.D. degrees from Tamkang University and National Taiwan University of Science and Technology, Taiwan, in 1989 and 1998, respectively. From 1983 to 2011, he was a professor at Central Police University, Taiwan. From 2012 to 2021, he was a professor at Yuanpei University of Medical Technology, Taiwan. Since 2021, he has been a professor at Tatung University. His research interests include digital evidence, forensics, and cybersecurity. (cyberpaul@gm.ttu.edu.tw)