Using Association Rule Mining to Analyze the Accident Characteristics of Intersection with Different Control Types

Yao-Tzu Hsu, Ph.D., P.E.
Associate Professor
Department of Transportation and Logistics
Feng Chia University
Taiwan

Shun-Chi Chang (Corresponding Author)
Ph.D. Candidate
Ph.D. Program for Civil Engineering
Water Resources Engineering, and Infrastructure Planning
Feng Chia University
Taiwan

Abstract

Intersections are the most prone to accidents in the road system. Understanding the characteristics of accidents of intersections under different control types is very important for developing intersection improvement strategies. The main purpose of this research is to mine the accident influence factors and the correlation between each factor at intersections with four control types such as traffic control signal (with pedestrian signal), traffic control signal (without pedestrian signal), flashing light signal and no signal through Association Rule Mining method in data mining technology. The data source is collected from the road traffic accident data of Taichung City Police Department in Taiwan in 2017. The research results show that there are indeed differences in the characteristics of accidents at intersections of various control types. This study uses Association Rule Mining to mine the hidden information of the causes of intersection accidents, which can be used as the references for the future improvement of intersections.

Keywords: At-grade Intersection, Road Traffic Accidents, Association Rule Mining

1. Introduction

In recent years, due to changes in the industrial structure and rapid economic development in Taiwan, the number of motor vehicle holdings has continued to increase. According to statistics from the Ministry of Transport in Taiwan (2019), the total number of all types of motor vehicles has exceeded 20 million in 2018. The area and length of roads have also increased year by year. In 2017, the total length of roads in Taiwan reached 43,206 kilometers and the total area of roads was 532,555 thousand square meters, resulting in a high number of road traffic accidents each year. In the road network system, most road traffic accidents occur at intersections because intersections mix various vehicles and the vehicles movement is very complicated. According to the information on road traffic accidents published by the National Police Agency, Taiwan (2019), more than 50% of road traffic accidents that occur each year occurred at intersections, as shown in Table 1 and Figure 1. In addition, according to the report of the National Highway Traffic Safety Administration, (NHTSA, 2010), 36% of traffic accidents in the United States are related to intersection; in Singapore, traffic accidents at intersections accounted for about 35% of the total number of accidents, and serious injuries and deaths related to traffic accidents accounted for about 32% (Tay and Rifaat, 2011). The foregoing information indicated that traffic accidents at intersections have become serious social problem.

Table 1 Total road accidents and intersection accident numbers (A1+A2)

Year	2011	2012	2013	2014	2015	2016
Intersection accident numbers	140,319	149,070	166,872	185,789	184,073	183,574
Total road traffic accidents	235,776	249,465	278,388	308,742	305,413	305,556

Notes:
1. A1 refers to a traffic accident that caused the death of a person on the spot or within 24 hours.
2. A2 refers to a traffic accident that caused injuries or killed more than 24 hours.

Data source: NPAMI, 2019
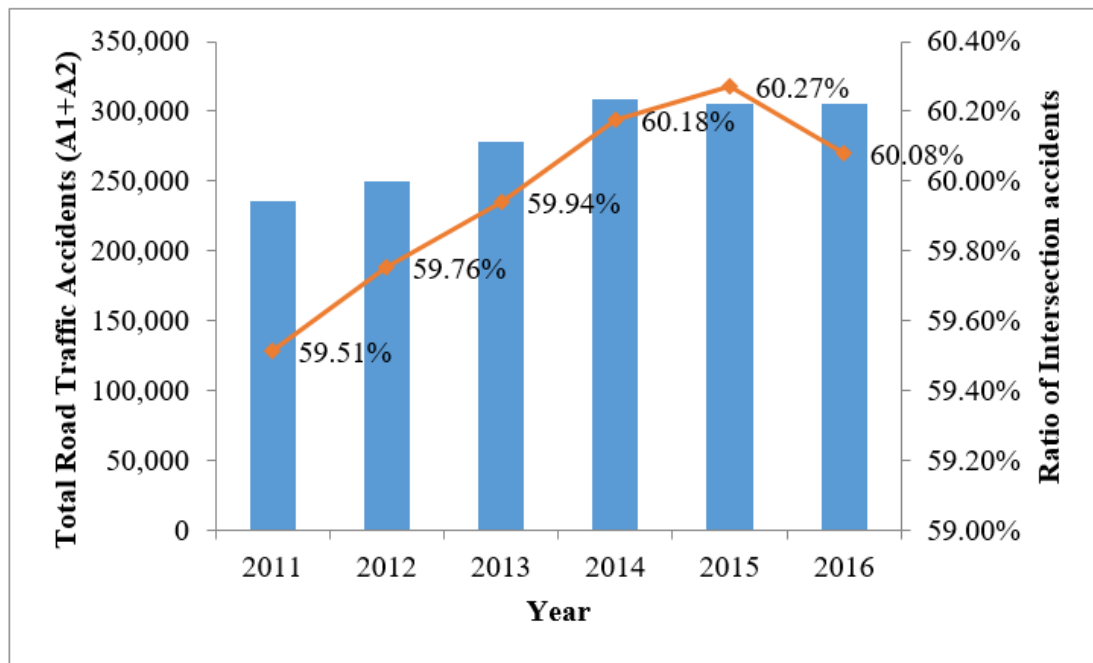
Data source: NPAMI, 2019

Figure 1 The trend of intersection accident rate in Taiwan

The control types at intersection can be divided into three types: traffic control signals, flashing light signals, and no signals. Among them, traffic control signals can be divided into with and without pedestrian signals. The characteristics of road traffic accidents at intersections with different control types may be different, which has a significant impact on the development of intersection improvement strategies. In the past researches on road traffic accidents, most of them focused on intersections, but rarely explored the correlation between the causes of accidents and factors based on traffic control types. The traditional statistical methods are applied to analysis for most of traffic accident related researches, but the data collection must meet certain assumptions. Therefore, there are more and more studies using data mining technology related methods in traffic accident analysis.

The main purpose of this research is to use data mining technology to explore the factors affecting the intersection accidents at various traffic control types, such as traffic control signals (with and without pedestrian signals), flashing light signals, and no signals, and the relationship between the factors, and then analyze the major differences in the characteristics of intersection accidents of various types of control methods. The research results can provide a reference for traffic and transportation authority to improve the lack and prevent the accident at intersections under different traffic control types. The remnant of the paper is then organized as follows: section 2 is literature review that includes accident analysis of intersection and traffic accident analysis using data mining technology, and describes the collecting of data, the concept of Association Rule Mining and Apriori algorithm in section 3; the results of association rule analysis are presented in section 4. Finally, section 5 presents the suggestions and conclusions of the study.

2. Literature Review

Because the traffic flow is more complicated, the accident rate at the intersection is more prominent than other places on the road. Therefore, in the analysis of road traffic accidents, issues related to intersections have been hot research topics. Many experts and scholars hope to find out the key factors that affect the occurrence of accidents at intersections.

Wong et al. (2007) analyzed the accident data at signalized intersection in Hong Kong, and used Poisson regression and negative binomial regression to quantify possible impacts of variables that include accident rate, traffic flow, geometric design, road environment and traffic control information on severe fatal injuries and slight injuries, and controlled the possible interference of traffic flow at the same time. The research results showed that the road environment, the degree of road curvature and the presence or absence of tram stations have a significant impact on severe fatal injuries and slight injuries. Tay and Rifaat (2011) pointed out that intersections are considered to be the most dangerous location for accidents on the road, but there was no study in Singapore to investigate the factors that influence the severity of

intersection accidents. This study took traffic accident data from 1992 to 2002 as the analysis object, and analyzed the impact of various accident-related factors on the severity of intersection accidents by the Ordered Probit Model.

The research results showed that vehicle type, road type, collision type, driver's characteristics and time of day are important factors that affect the accident severity at intersection. Chen et al. (2012) wanted to study the risk factors that influence the accident severity of road user at intersections. Based on the analysis of Logistic regression, the research results showed that a total of 7 risk factors are significantly related to the severity of traffic accidents, including driver age and gender, speed zone, traffic control type, time of day, collision type and seat belt usage. Ahmed et al. (2014) mentioned that effective accident analysis must classify attributes of count data, such as geometry and traffic control parameters. This study discussed the impact of road width, land use, lane markings, and traffic control on the safety of unsignalized intersections. The results showed that minor road intersections with single lane markings and no signal control in non-urban areas is the most prone to accidents. Wang et al. (2014) explored the collision types of vehicle crashes at intersections. The study analyzed the causes of various types of collisions and constructed a safety performance function using the Bayesian model to calculate the probability of various types of collisions. Parameters in the study include road geometric design, signal control type, traffic flow and traffic accident type.

In the past, in the analysis of road traffic accidents, traditional statistical methods were mostly used, but their use was limited. They must meet some basic conditions, such as normality and independence of residuals (Kumar and Toshniwal, 2015). Data mining is an emerging technology in the field of artificial intelligence and database systems. It can extract hidden, unknown, credible, and effective knowledge from large data. The data mining is based on parameters of user-defined settings to find the information which users are interested in from a group of unprocessed resources. After some special processing, it can be used as a basis for policy makers to develop strategies. With the advancement of information technology in recent years, the technology of data mining become increasingly mature, and it has the characteristics of finding useful information from resources without being limited by the statistical assumptions. Association Rule Mining technology is an important data mining technology. It is one of the commonly used methods. This technology can mine the correlation of consumer purchasing behavior from the transaction database (Han and Kamber, 2006; Weng, 2011). More and more studies have applied it to the analysis of road traffic accidents, and also proved its applicability.

El Tayeb et al. (2015) used two algorithms, Apriori and Predictive Apriori, to analyze the relationship between the attributes of road traffic accidents and the severity of accidents recorded in the database. The analysis results of the association rules generated by the two algorithms were summarized to obtain the most meaningful and interesting rules. The empirical results showed that the Apriori algorithm is more effective than the Predictive Apriori algorithm and generated more meaningful rules between the attributes of road traffic accidents and the severity of the accident. Ait-Mlouk et al. (2016) used Association Rule Mining and multi-criteria analysis methods in an attempt to mine hidden new information from historical data on traffic accidents on one of the busiest roads in Morocco that can assist traffic authorities and law enforcement agencies to develop relevant countermeasures to improve road safety. Gupta et al. (2016) used Indian traffic accident data as the analysis object, and proposed a research framework based on the analysis of association rules. The association rules were used to explore the important factors affecting the severity of different accidents. The research results showed that some hidden factors can be discovered through this analysis framework, which is very useful for understanding the factors behind road traffic accidents in the area. Kumar et al. (2016) pointed out that the incidence of traffic accidents in certain locations is more obvious, and analyzing for these locations will help understand the characteristics of traffic accidents. This research used data mining technology to identify locations with a high incidence of traffic accidents and find out the various impact factors of accidents at different locations of accident frequency. The study took traffic accident data in Dehradun County, India as the research object, and applied K-means algorithm to classify accident data into three groups: high incidence rate, moderate incidence rate, and low incidence rate according to the accident site incidence rate. And then used Association Rule Mining to find out the characteristics and impact factors of traffic accident of each group. Janani and Ramya Devi (2017) pointed out that using data mining technology to analyze road traffic accidents can obtain rich and meaningful results.

The analysis of accident sites helps to determine certain road accident characteristics that make road accidents frequent in these locations. In this study, the K-means algorithm was used to divide the accident sites into three clusters according to the frequency of accidents, and then used Association Rule Mining to mine the accident characteristics of these locations. Li et al. (2017) analyzed the 2007 road traffic fatal accident data recorded by the National Highway Traffic Safety Administration, and explored the relationship between mortality and traffic accident attributes. The attributes include collision type, weather condition, road surface condition, lighting conditions, and whether drunk.

The Apriori algorithm was used to explore its association rules, and a classification model was established with Naïve Bayes classifier, and K-means algorithm was used for clustering. Specific recommendations for safe driving were proposed based on the research results.

Looking at the above analysis, in the past research, much attention has been focused on the analysis of accidents at intersections, such as the cause analysis of accidents at signalized intersections, the cause analysis of accidents at unsignalized intersections, and the analysis of the severity of intersection accidents. Few studies explored the relationship between intersection control types and road traffic accidents; In addition, Agrawal and Srikant (1994) pointed out that in the field of data mining, the Association Rule Mining is the most commonly used method. Its main concept is to find the correlation between attributes in the database. The Association Rule has become more common in related research on traffic safety, but it has not been applied to the analysis of intersection accidents.

3. Methodology

The main purpose of this research was using Association Rule Mining method in data mining technology to explore the accident characteristics and the relationship between accident influence factors at intersection under different control types such as traffic control signals (with pedestrian signals), traffic control signals (without pedestrian signals), flashing light signals, and no signals. The research framework was provided in Figure 2. This section will explain the association rule analysis theory and the concept of Apriori algorithm that the most commonly is used in association rule analysis. It will also state the collection and selection of road traffic accident data and the selection of attribute variables.
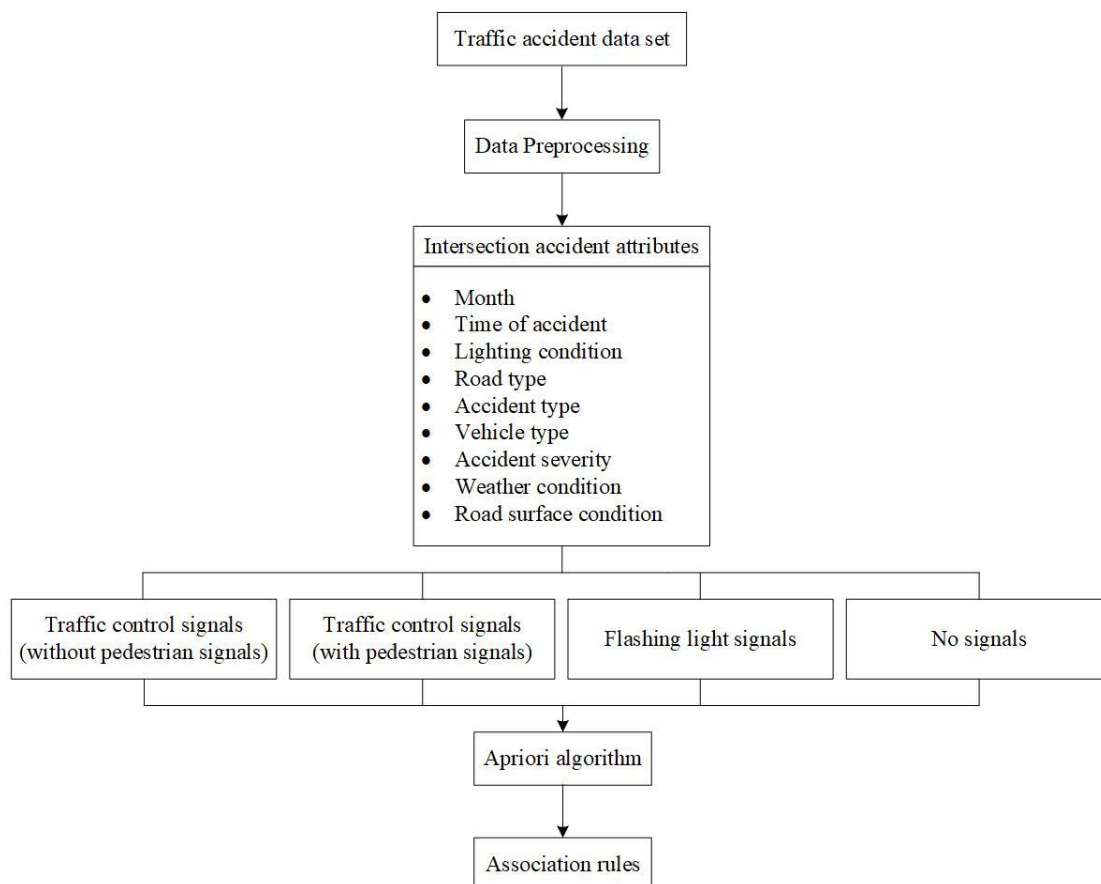

Figure 2 Proposed framework for analysis

3.1 Association Rule Mining

The association rule is a discussion of probability relationships. Based on past historical records, it tries to explore the relevance among data in the database. The Association Rule Mining is an important technology in the process of data mining. The theory of Association Rule Mining is conditional probability. For example, if factor A appears, X% of the probability will also have factor B at the same time. We assume that the relevant information data set in the road traffic accident database is $L=\{l_1,l_2,\ldots l_n\}$, and accident attributes A and B are a subset of L and are independent of each other. The expressions of association rule are: A→B, A⊂L, B⊂L, and A∩B = ∅.

Agrawal et al. (1993) pointed out that all association rules must meet two thresholds, namely minimum support and minimum confidence. With these two indicators, we can find meaningful rules from many rules.

1. Support (S_p): It indicates that the probability of A and B appearing at the same time.

$$S_P = \frac{p(A \cap B)}{N} \qquad (1)$$

2. Confidence (C_f): It indicates that the probability of B appearing on the premise of A appearing.

$$C_f = \frac{P(A \cap B)}{P(A)} \qquad (2)$$

In addition to the above two required indicators, there are three other indicators that can be used to measure the effect of association rules: Lift, Leverage and Conviction.

1. Lift (L_t): The probability of B appearing alone is compared with the Confidence. The value must be greater than 1 to show that the rule is valid. A larger Lift value means better results.

$$L_t = \frac{P(A \cap B)}{P(A) \times P(B)} \qquad (3)$$

2. Leverage (L_v): It compares the increase of the probability of simultaneous occurrence of A and B to the individual occurrence probability of the two factors. Its value will be between -0.25 and +0.25. If it is greater than 0, A and B has possibility of simultaneous occurrence.

$$L_v = P(A \cap B) - P(A) \times P(B) \qquad (4)$$

3. Conviction (C_v): It is used to indicate the correlation between A and B. The Conviction will be between 0.5 and infinity. If the value is 1, it indicates that the two factors are not related.

$$C_v = \frac{P(A) \times P(B)}{P(A \cap \bar{B})} \qquad (5)$$

3.2 Apriori Algorithm

In the analysis of association rule, the Apriori algorithm proposed by Agrawal and Srikant (1994) is the most widely used. The core concept of the algorithm is to repeatedly generate Candidate Item Set to find all high-frequency item set. After sifting through the minimum support and minimum trust, all relevant rules are derived. The analysis procedure of Apriori algorithm is as follows (Liao and Wen, 2009):

1. Set minimum support and confidence.

2. According to the concept of candidate item sets, if the support of the candidate item set is greater than or equal to the minimum support, the candidate item set is a high-frequency item set.

3. First, read all data from the database to get the support of the first candidate item set, and then find the first high-frequency item set. A combination of these high frequency single item sets is used to generate a second candidate item set.

4. After scanning the database to obtain the support of the second candidate item set, find the second high-frequency item set and use the combination of these second high-frequency item set to generate a third candidate item set.

5. Scan the entire database repeatedly, and compare it with the lowest support level to generate a high-frequency item set, and then combine to generate the next level of candidate item set, until it is no longer combined to generate a new set of candidate item set.

3.3 Data collection and variable selection

The road traffic accident data used in this research were all the accidents that occur in Taichung City in 2017 and the data was collected from Taichung City Government Open Data Platform (2018). Through the setting of conditions, road traffic accidents occurring at intersections were selected. After the accident data was collected, the data must be preprocessing (Kumar and Toshniwal, 2015). It is one of the important tasks in data mining. It mainly deals with missing values, deletes irrelevant attributes, and classifies the data attributes so that the data conforms to the type required for subsequent analysis. After preprocessing the data, a total of 38,660 samples will be analyzed. There were 8 attributes used in the study that include month, time of accident, lighting condition, road type, accident type, vehicle

type, accident severity, weather condition, and road surface condition. The classification of each attribute is shown in Table 2.

As Association Rule Mining is used to extract different attribute values that frequently occur in certain accidents in order to understand the different factors related to road traffic accidents, it cannot be used for classification purposes, and it cannot specify what kind of traffic control type certain factors belong to.

Therefore, in the study, according to the type of signal at intersection the data were divided into 4 clusters, such as traffic control signals (with pedestrian signals), traffic control signals (without pedestrian signals), flash signals, and no signals, as shown in Table 3, then performed association rule mining and extracted association rules for each cluster, so as to understand the accident characteristics of various traffic control types (each cluster) and the correlation between factors, and further explored the differences of accidents characteristics between clusters.

Table 2 Intersection accident attributes for analysis

Attribute name	Type	Values
Month (M)	Nominal	1=spring(3~5), 2=summer(6~8), 3=fall(9~11), 4=winter(12~2)
Time of accident (T)	Nominal	1=0~4, 2=4~8, 3=8~12, 4=12~16, 5=16~20, 6=20~24
Lighting condition (LC)	Nominal	1=day light, 2=road light, 3=no light
Road type (RT)	Nominal	1=three-way, 2=four-way, 3=multi-way
Accident type (AT)	Nominal	1=vehicle-pedestrian, 2=vehicle-vehicle, 3=single vehicle
Vehicle type (VT)	Nominal	1=big-sized vehicle, 2=small-sized vehicle, 3=motorcyclist or cyclist, 4=others
Accident severity (AS)	Nominal	1=fatality, 2=injury, 3= no injury
Weather condition (WC)	Nominal	1=sunny day, 2=cloudy day, 3=bad
Road surface condition (RS)	Binary	1=dry, 2=bad

Table 3 Cluster types and accident numbers

Cluster id	Cluster name	Total accidents
C1	Traffic control signals (with pedestrian signals)	4277
C2	Traffic control signals (without pedestrian signals)	16470
C3	Flashing light signals	2782
C4	No signals	15131

4. Results and Discussion

Initially, according to the traffic control types, the accident collection was divided into four clusters: traffic control signals (with pedestrian signals), traffic control signals (without pedestrian signals), flashing light signals and no signals, as shown in Table 3. The Apriori algorithm [2] in Weka 3.8.3 was used to implement the association rule mining. According to the recommendations of related researches in the past, the minimum support is set to 5% (Geurts, K. et al., 2002; Kumar, S. and Toshniwal, D., 2016; Janani, G. and Ramya Devi, N. 2017); The minimum reliability is set to 30% (Geurts, K. et al., 2002). The results of association rule analysis can understand the correlation between different attribute values that occur together when an accident occurs. Each cluster generates multiple rules, but only based on the Lift value, select some meaningful rules to explain. The rules for each cluster are discussed below:

4.1 Association rules for traffic control signals (with pedestrian signals)

As described in Section 3.1, the Lift value must be greater than 1 to indicate that the rule is meaningful. The higher the value is the better the rule effect between attributes. In the results of the association rule analysis of traffic control signals (with pedestrian signals), the highest Lift value is 1.65 shown in Table 4, and the association rules are sorted according to the Lift value from high to low. The rule with the highest Lift value shows that there are mostly small-sized vehicle accidents in this cluster, and when accidents occur, it is very likely that road users will be injured. Most accidents occur under the conditions of sunny days and dry road surface. The most common type of accident are vehicle-to-vehicle collisions, and they are located at the intersection of four-way. In addition, rules 3 and 5 in the table include the accident severity as uninjured, indicating that among the accidents in the cluster, there are still a certain number of accidents that did not cause road users to be injured.

Table 4 Association rules for traffic control signals (with pedestrian signals)

Rule no.	Rule body	Conf.	Lift	Lev.	Conv.
1	{VT=2} → {AS=2}	0.95	1.65	0.19	8.51
2	{WC=1, AS=2} → {RS=1, AT=2}	0.93	1.16	0.06	2.85
3	{RS=1, AS=3} → {WC=1, AT=2}	0.89	1.15	0.06	2.07
4	{WC=1, LC=1} → {RS=1, AT=2}	0.91	1.13	0.06	2.19
5	{WC=1, AS=3} → {RS=1}	0.99	1.12	0.05	16.47
6	{WC=1, RT=2, AT=2} → {RS=1}	0.99	1.12	0.06	15.32
7	{LC=1, RT=2, RS=1} → {WC=1}	0.95	1.12	0.05	3.23
8	{LC=1, RS=1, AT=2} → {WC=1}	0.95	1.12	0.05	3.23
9	{LC=1, RS=1} → {WC=1}	0.95	1.12	0.05	3.13
10	{RS=1, AT=2 } → {WC=1 }	0.95	1.12	0.08	3.08

4.2 Association rules for traffic control signals (without pedestrian signals)

Table 5 shows the analysis results of the association rule mining of traffic control signals (without pedestrian signals). The maximum Lift value is 1.13. This rule shows that the cluster accidents mainly occur in sunny day, natural light, and dry road surface. There are many collisions between vehicles. According to the Lift value, check other rules in order, and found that in the case of sunny weather and dry roads, most accidents will not cause road users to be injured, and there are more accidents at the four-way intersection type. In dry road surface condition, collisions between vehicles are more likely to result in uninjured consequences. Compared results of the association rules analysis of this cluster with the traffic control signal (with pedestrian signal) cluster, the most important difference is that the attributes of the vehicle type are not included in the association rules of the cluster, and the severity of the accident is mainly uninjured in this cluster.

Table 5 Association rules for traffic control signals (without pedestrian signals)

Rule no.	Rule body	Conf.	Lift	Lev.	Conv.
1	{LC=1, RS=1} → {WC=1, AT=2}	0.89	1.13	0.07	1.94
2	{LC=1, RS=1} → {WC=1}	0.96	1.12	0.07	3.45
3	{WC=1, AS=3} → {RS=1}	0.99	1.11	0.04	12.7
4	{WC=1, RT=2} → {RS=1}	0.99	1.11	0.06	11.88
5	{WC=1} → {RS=1}	0.99	1.11	0.08	11.81
6	{WC=1} → {RS=1, AT=2}	0.91	1.11	0.08	2.04
7	{WC=1, AT=2} → {RT=2, RS=1}	0.71	1.1	0.05	1.23
8	{RS=1, AS=3} → {AT=2}	0.96	1.06	0.03	2.54
9	{LC=1, AT=2} → {WC=1 }	0.89	1.04	0.02	1.29
10	{LC=1, RT=2} → {WC=1 }	0.88	1.03	0.01	1.2

4.3 Association rules for flashing light signals

As shown in Table 6, the analysis results of the accident association rule mining at the intersection of flashing light signal show a maximum Lift of 1.66. Most of the accidents at the intersection of flashing light signal are accidents of motorcyclist or cyclist, and if an accident occurs, it will most likely cause injury to the parties. Accidents mainly occur in sunny day, natural light, and dry road surface. The types of accidents are still mostly vehicle-to-vehicle collisions, and they often occur at four-way intersections. The analysis results of the association rule mining of this cluster is compared with the traffic control signals (with pedestrian signals) and the traffic control signals (without pedestrian signals). The main difference is that the vehicle type of motorcyclist or cyclist are more common in the cluster association rules, and that the severity of the accident in the cluster is mainly injury.

Table 6 Association rules for flashing light signals

Rule no.	Rule body	Conf.	Lift	Lev.	Conv.
1	{AS=2} → {VT=3}	0.91	1.66	0.2	4.99
2	{WC=1, LC=1} → {RS=1, AT=2}	0.93	1.14	0.06	2.58
3	{WC=1, RT=2} → {RS=1, AT=2}	0.93	1.14	0.06	2.5
4	{WC=1, AT=2} → {RS=1}	0.99	1.12	0.08	11.2
5	{WC=1, LC=1} → {RS=1}	0.99	1.12	0.06	9.94
6	{WC=1, AS=2} → {RS=1}	0.99	1.12	0.05	9.5
7	{WC=1, VT=3} → {RS=1}	0.99	1.12	0.05	9.06
8	{WC=1} → {RS=1}	0.99	1.12	0.09	9.28
9	{WC=1, RT=2} → {RS=1}	0.99	1.12	0.05	8.99
10	{RT=2, RS=1} → {WC=1}	0.95	1.12	0.06	2.87

4.4 Association rules for no signals

Table 7 shows the analysis results of the association rule mining for accidents at unsignalized intersections. The maximum Lift value is 1.66, and the results are very similar to those of flashing light signals group. There are more accidents of motorcyclist or cyclist than other types of vehicles at the unsignalized intersection, and when an accident occurs, the severity of the accident is mostly "AS = 2". The environment which accident occurs is roughly the same as the intersection of the three control types mentioned above. Natural light, clear weather during the day and the dry road surface are the most prone to accidents. Most of the accident types are collisions between vehicles. When the weather is fine and the road surface is dry in the daytime, because there is no signal at the intersection, drivers are prone to have aberrant driving behaviors, such as not slowing down to pass through intersection or non-yielding. In the event of an emergency, it is likely that there will be no time to respond, resulting in an accident and causing the parties to be injured.

Table 7 Association rules for no signals

Rule no.	Rule body	Conf.	Lift	Lev.	Conv.
1	{AS=2} → { VT=3}	0.96	1.66	0.2	9.98
2	{RS=1, AS=2} → { VT=3}	0.96	1.66	0.18	10.33
3	{LC=1, RS=1} → {WC=1, AT=2}	0.9	1.12	0.07	1.97
4	{RS=1, AT=2, VT=3} → {WC=1}	0.96	1.11	0.04	3.04
5	{LC=1, RS=1} → {WC=1}	0.96	1.11	0.07	3.2
6	{WC=1, AS=2} → {RS=1}	0.99	1.1	0.04	15.2
7	{WC=1, VT=3} → {RS=1}	0.99	1.1	0.05	14.32
8	{WC=1, AT=2} → {RS=1}	0.99	1.1	0.07	12.67
9	{AT=2, VT=3} → {RS=1}	0.92	1.03	0.01	1.32
10	{AT=2, VT=3} → {WC=1}	0.89	1.03	0.01	1.23

4.5 Trend analysis of the number of accidents

Figure 3 shows the monthly accident distribution trends of intersection under four control types as traffic control signals (with pedestrian signals), traffic control signs (without pedestrian signals), flashing light signals and no signals. It can be seen from the figure that the distribution trends of the four control types are very similar. All of them have a high number of accident in March and November. However, the number of accidents at unsignalized intersection increased slightly from April to May, which is different from other control types. It can be seen from this trend that since May to June is the rainy season in Taiwan, people will avoid unnecessary activities, so the number of accidents is relatively low. Due to the colder weather in December, January, and February, people will also reduce the opportunity to go out, especially the motorcyclists or cyclists, so the number of accidents is also low.

Figure 4 shows the accident distribution trends of the time period of the intersection under four control types as traffic control signals (with pedestrian signals), traffic control signs (without pedestrian signals), flashing light signals and no signals. It can be seen from the figure that the distribution trends of the four control types all show an upward trend from 0: 00 ~ 4: 00 to 4: 00 ~ 8: 00 and 4: 00 ~ 8: 00 to 8: 00 ~ 12: 00. From 8:00~12:00 to 12:00~16:00, the number of accidents decreases, and from 12:00 ~16:00 to 16:00~20:00, the number of incidents shows an upward trend. It drops from 20:00 to 24:00. There are two peaks in this trend distribution, 8: 00 ~ 12: 00 and 16: 00 ~ 20: 00 respectively. These two periods cover the peak time of commuting to and from work, so there are more accidents.
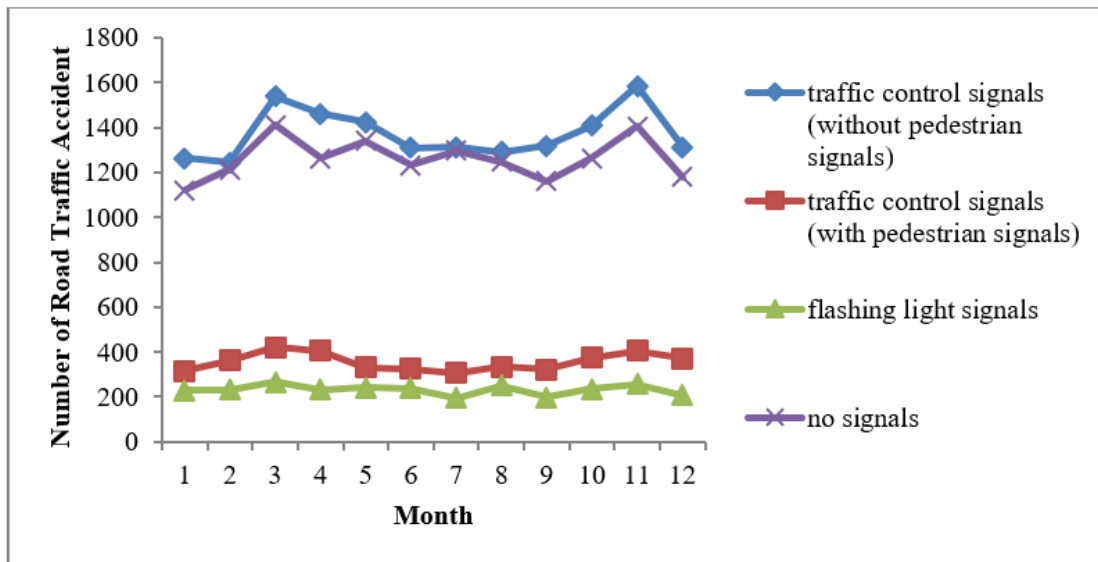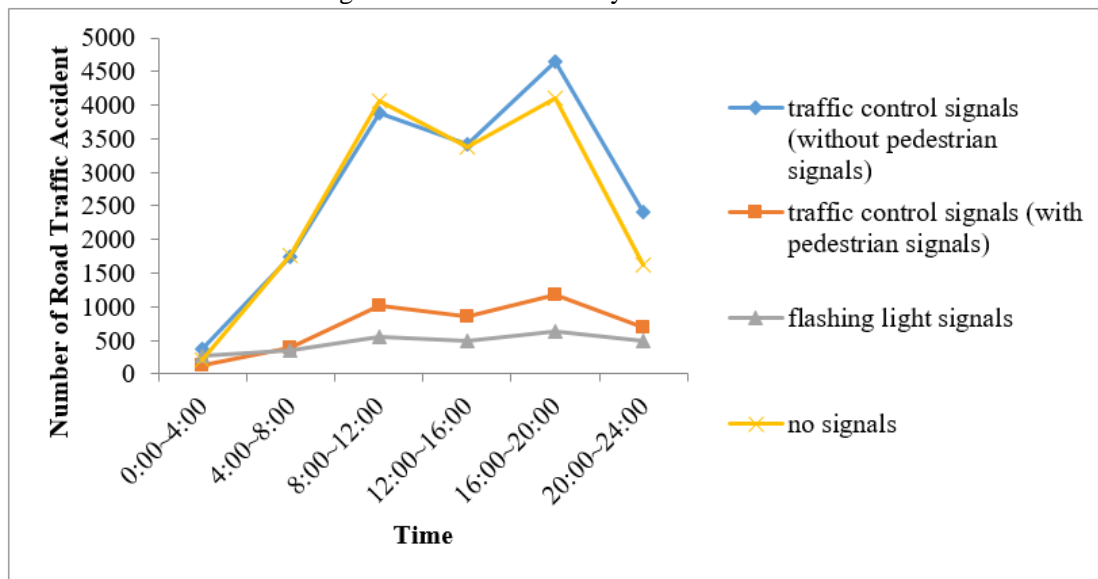
Figure 3 Month trend analysis of accidents

Figure 4 Time Trend Analysis of Accidents

5. Conclusion

In the research of road traffic accident analysis, more and more literatures have adopted data mining technology as the analysis method, and have obtained rich results, showing the suitability of data mining technology for road traffic accident analysis. This study analyzed the traffic accident data of Taichung City in 2017 published by the Taichung City Police Department, and explored the differences in road traffic accident characteristics at intersections under different control types.

First, the accident data is divided into 4 clusters according to the traffic control types: traffic control signals (with pedestrian signals), traffic control signals (without pedestrian signals), flashing light signals and no signals, and then Association Rule Mining is used to analyze the data of each cluster. Association rules revealed different factors related to road accidents in each cluster, and the hidden and meaningful information was mined, which is very helpful for developing safety improvement strategies at intersections of different control types. In the analysis of association rules, in addition to the two necessary indicators of support and reliability, conditions such as Lift, Leverage, and Conviction were also added to the study as a threshold limit to measure the effect of association rules and ensure the accuracy of the rules. The main difference between this study and previous studies was to explore and compare the characteristics of traffic accidents with different control types at intersections. There have been no studies in the past to discuss related issues.

At the same time, this study also validates the appropriateness of data mining techniques in intersection accident analysis. Although there are a large number of historical resources in road traffic accident database, how to select appropriate accident samples for exploration from these resources is worthy of further discussion. Grouping the data according to different accident attributes can discover more meaningful rules and provide an important reference for researchers and transportation authorities.

References

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. Proceedings of ACM SIGMOD, 207-216.

Agrawal, R., &Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In: Proceedings of the 20th international conference on very large data bases, 487-499.

Ahmed, A., Sadullah, A. F. M., &Yahya, A. S. (2014). Accident analysis using count data for unsignalized intersections in Malaysia. Procedia Engineering, 77, 45-52.

Ait-Mlouk, A., Agouti, T., &Gharnati, F. (2016). An approach based on association rules mining to improve road safety in Morocco. International Conference on Information Technology for Organizations Development.

Chen, H., Logan, D. B., & Cao, L. (2012). Analysis of risk factors affecting the severity of intersection crashes by logistic regression. Traffic Injury Prevention, 13, 300-307.

El Tayeb, A. A., Pareek, V., &Araar, A. (2015). Applying association rules mining algorithms for traffic accidents in Dubai. International Journal of Soft Computing and Engineering, 5.

National Highway Traffic Safety Administration, NHTSA (2010). Crash factors in intersection-related crashes: An on-scene perspective, Washington, DC.

Geurts, K., Wets, G., Brijs, T., &Vanhoof, K. (2003). Profiling high frequency accident locations using association rules. In Proceedings of the 82nd Annual Transportation Research Board, Washington DC. (USA).

Gupta, M., Kumar-Solanki, V., & Kumar-Singh, V. (2016). A novel framework to use association rule mining for classification of traffic accident severity. IngenieríaSolidaria, 13, 37-44.

Han, J. W., &Kamber, M. (2006). Data Mining: Concepts and Techniques. San Francisco: Morgan Kaufmann.

Janani, G., &Ramya Devi, N. (2017). Road traffic accidents analysis using data mining techniques. Journal of Information Technology and Applications, 2, 84-91.

Kumar, S., &Toshniwal, D. (2015). A data mining framework to analyse road accident data. Journal of Big Data, 2(1), 1-18.

Kumar, S., &Toshniwal, D. (2016). A data mining approach to characterize road accident locations. Journal of Modern Transportation, 24, 62-72.

Li, L., Shrestha, S., & Hu, G. (2017). Analysis of road traffic fatal accidents using data mining techniques. IEEE 15th Int. Conf. on Software Engineering Research, Management and Applications (SERA), 363-370.

Liao S. H., & Wen, C. H. (2009). Data mining and business intelligence. (1st ed). Taipei City: YehYeh Book Gallery, Taiwan.

Statistical data of Ministry of Transportation and Communications (MOTC), Taiwan (2019). https://www.motc.gov.tw/, (November, 2019).

Statistical data of National Police Agency, Ministry of the Interior (NPAMI), Taiwan (2019). https://www.npa.gov.tw/NPAGip/wSite/np?ctNode=12552&mp=1, (November, 2019).

Taichung City Government Open Data Platform, Taiwan (2018). https://opendata.taichung.gov.tw/, (December, 2018).

Tay, R., &Rifaat, S. M. (2011). Factors contributing to the severity of intersection crashes. Journal of Advanced Transportation, 41, 245-265.

Wang, X. S., Xie, K., Abdelaty, M., Chen, X. H., & Tremont, P. J. (2014). Systematic approach to hazardous-intersection identification and countermeasure development. Journal of Transportation Engineering, 140, 1-22.

Weng, C. H. (2011). Mining association rules with high correlation from the purchasing intension data. Journal of Information Management, 18, 119-138.

Wong, S. C., Sze, N. N., & Li, Y. C. (2007). Contributory factors to traffic crashes at signalized intersections in Hong Kong. Accident Analysis and Prevention, 39, 1107-1113.