Accident Analysis and Prevention 142 (2020) 105497

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Application of association rules mining algorithm for hazardous materials transportation crashes on expressway

Jungyeol Hong, Reuben Tamakloe, Dongjoo Park*

Department of Transportation Engineering, The University of Seoul, 163 Seoulsiripdae-ro Dongdaemun-gu, Seoul 02504, South Korea

ARTICLE INFO

Keywords:
hazardous material
association rules mining
Apriori algorithm
data mining
expressways

ABSTRACT

Although crashes involving hazardous material (HAZMAT) vehicles on expressways do not occur frequently compared with other types of vehicles, the number of lives lost and social damage is very high when a HAZMAT vehicle-involved crash occurs. Therefore, it is essential to identify the leading causes of crashes involving HAZMAT vehicles and make specific countermeasures to improve the safety of expressways. This study aims to employ the association rules mining (ARM) approach to discover the contributory crash-risk factors of HAZMAT vehicle-involved crashes on expressways. A case study is conducted using crash data obtained from the Korea Expressway Corporation crash database from 2008 to 2017. ARM was conducted using the Apriori algorithm, and a total of 855 interesting rules were generated. With appropriate support, confidence, and lift values, we found hidden patterns in the HAZMAT crash characteristics. The results indicate that HAZMAT vehicle-involved crashes are highly associated with male drivers, single vehicle-involved crashes, clear weather conditions, daytime, and mainline segments. Also, we found that HAZMAT tank-lorry and cargo truck crashes, single vehicle-involved crashes, and crashes on mainline segments of expressways had independent and unique association rules. The finding from this study demonstrates that ARM is a plausible data mining technique that can be employed to draw relationships between HAZMAT vehicle-involved crashes and significant crash-risk factors, and has the potential of providing more easy-to-understand results and relevant insights for the safety improvement of expressways.

1. Introduction

The rapid increase in urbanization and industrialization has led to a significant rise in demand for hazardous materials (HAZMAT). Most of these substances originate from locations other than their destination in large quantities. As such, transportation using trucks becomes inevitable. Currently, research shows that the transportation of HAZMAT on roads form a significant part of all freight transported (Ghaderi and Burdett, 2019; Ghaleh et al., 2019; Holeczek, 2019; Poku-Boansi et al., 2018; Pompone and de Oliveira Neto, 2019), and there is a growing trend in the number of crashes involving HAZMAT vehicles per year (Pompone and de Oliveira Neto, 2019; Yang et al., 2010). Although the frequency of HAZMAT truck-involved crashes is low relative to the crashes of other vehicle types, they often end up in a massive loss of lives and properties. People living around the expressway section have to be evacuated from the HAZMAT vehicle-crash scene, which comes at a considerable cost to the authorities in charge (Frank et al., 2000). Even though HAZMAT provides many benefits to people, its risks to the environment and the lives of humans remains unacceptably high and cannot be overlooked.

South Korea produces approximately 4500 different kinds of HAZMAT, which are transported with trucks on the national expressways. According to Hong et al. (2019d), from 2007 to 2017, there were 315 crashes on expressways involving HAZMAT trucks in South Korea. Considering all the crashes that occurred in the same period, those involving HAZMAT trucks were the most fatal and had longer crash clearance times. Similarly, from a study conducted by Uddin and Huynh (2018), 112 out of 3744 vehicle-involved crashes in the U.S. in 2014 were the most fatal because they were carrying HAZMAT. A ten year (2009–2018) HAZMAT incident report data compiled by the US Department of Transportation showed that there were 166,065 HAZMAT vehicle-related incidents. Out of these, 87.90 % (145,971) of the incidents were highway-related, while 7.96 % (13,211), 3.84 % (6384), and 0.30 % (499) occurred through air, railway and water transport, respectively. The report showed a gradual increase in HAZMAT highway-related incidents (from 12,730 in 2009 to 17,923 in 2018). Even though there is a reduction in fatality and injury severity trends within the period of analysis, transport of HAZMAT by road accounted

* Corresponding author at: The University of Seoul, 21st Century Building 513, Seoulsiripdae-ro 163, Dongdaemun-gu, Seoul 02504, South Korea.
E-mail addresses: jyhong9868@uos.ac.kr (J. Hong), drtamakloe@uos.ac.kr (R. Tamakloe), djpark@uos.ac.kr (D. Park).

https://doi.org/10.1016/j.aap.2020.105497
Received 6 October 2019; Received in revised form 22 January 2020; Accepted 8 March 2020
Available online 19 May 2020
0001-4575/

for 98.10 % of the fatalities and 73.82 % of injuries (US Department of Transportation, 2019). In China, 5203 people lost their lives in 3974 incidents involving the transportation of HAZMAT from 2006 to 2017. This finding shows that more than one person dies in China each day from HAZMAT incidents (Zhao et al., 2018). In south India, 64 crashes occurred due to the transportation of HAZMAT from April 2011 to March 2012 (Palanisamy et al., 2015), and according to the World Health Organization, HAZMAT related traffic deaths per 100,000 people in Thailand and Sweden were 36.2 and 2.8, respectively, in 2017 (Poku-Boansi et al., 2018).

The catastrophic consequences of potential crashes involving HAZMAT vehicles have led to the increase in the number of studies on HAZMAT transportation to reduce the transportation risk (Fang et al., 2017; Haastrup and Brockhoff, 1990; Shen et al., 2014). Some primary reasons for HAZMAT vehicle crashes, such as speed, human-related errors, time of day, vehicle-related defects, and weather, have been identified (Ma et al., 2018; Shen et al., 2014). Most traffic crash analysis has concentrated mainly on using statistical methods to establish the relationship between the dependent and independent variables (Hong et al., 2019a,c; Moudon et al., 2011; Shankar and Mannering, 1996; Shibata and Fukuda, 1994). Nevertheless, it has been reported in the literature that the correlations among crash-risk factors used as independent variables negatively affect the statistical analysis and leads to inaccurate results (Chang and Chen, 2005; Greibe, 2003; Pande and Abdel-Aty, 2009). Since policymakers have shown a high interest in the causalities of HAZMAT vehicle-involved crashes and how to improve its transportation conditions, transportation engineers must use an effective method to investigate the characteristics and contributory factors that can lead to HAZMAT vehicle-involved crashes.

The problems associated with statistical methods mentioned above can be solved by using data mining tools such as classification and regression tree (CART) for predictive modeling problems; however, according to Pande and Abdel-Aty (2009), they are unable to provide quantitative measures for the correlations among independent and dependent variables. In contrast, the association rules mining (ARM) method, which has the capability of discovering the associations among the characteristics and contributory factors of HAZMAT vehicle-involved crashes has been adopted for use as an efficient tool for analyzing crash data (Das et al., 2019b; Das and Sun, 2014; Pande and Abdel-Aty, 2009; Weng et al., 2016; Yu et al., 2019). This data mining methodology has been judged as a more valid technique. It has been identified as having the potential of being used as a decision support tool for road traffic safety engineers as it does not rely on any hypothesis in the discovery of hidden but meaningful connections in datasets (Yu et al., 2019).

Despite this finding, there is limited research that employs data mining tools to explore the hidden associations in HAZMAT vehicle-related crash datasets. To the best of our knowledge, no study has used the more efficient ARM method in this research area. Based on this recognition, the primary objective of this current study is to bridge the knowledge gap by employing the ARM approach to systematically analyze the characteristics and contributory factors of HAZMAT vehicle-involved crashes that occur on expressways in South Korea. HAZMAT vehicle-involved crashes are complex events, and policymakers must consider the numerous contributory factors that occur together when such crashes occur. Therefore, identifying the hidden associations between contributory factors in this present study would help in comprehensively understanding the significant patterns of HAZMAT vehicle-involved crashes to guide policy and decision making that can reduce the frequency and severity of HAZMAT vehicle-related crashes on expressways.

2. Literature reviews

To date, there has been a significant number of studies exploring incidents involving the transportation of HAZMAT (Azimi et al., 2020; Hong et al., 2019a; Liu et al., 2017). Younger drivers HAZMAT truck drivers have been identified as having an increased likelihood of violating road traffic regulations, which is one of the significant causes of crashes on expressways (Hong et al., 2019a). Khattak et al. (2003) analyzed crash-risk factors in large truck rollovers and the injury severity of the occupants of single-vehicle crashes that occurred on highways from 1996 to 1998 using an ordered probability model. They found that the probability of occupants of HAZMAT transporting trucks sustaining injuries in crashes increased by about 16 %, and they were the group that sustained the most severe injuries. Chen and Chen (2011) employed mixed logit models to study injury severity of truck drivers. They noted that HAZMAT truck drivers were either incapacitated or suffered fatal injuries about 22 % and 11 % of the time when involved in a single-vehicle crash or multi-vehicle crash, respectively. Azimi et al. (2020) arrived at similar conclusions after analyzing the injury severity of large truck rollover crashes in the state of Florida using a random parameter logit model. Besides, they identified that crashes tend to be more severe when there are HAZMAT spills.

Regarding risk factors associated with transporting HAZMAT using trucks, a study of 1932 HAZMAT vehicle-involved crashes identified that the release of poisonous substances from most of the trucks caused fires and explosions (Oggero et al., 2006). Similarly, Shen et al. (2014) analyzed that 56 % of 708 crashes involving HAZMAT trucks from 2004 to 2011 occurred on expressways in China, and the majority of them resulted in HAZMAT spills. Their study established that vehicle defects and human errors were the leading causes of HAZMAT crashes.

Ma et al. (2018) conducted a comprehensive study using a Bayesian network model to investigate the factors influencing crashes involving 839 HAZMAT trucks in China from 2015 to 2016. They identified that speed was a significant factor that led to rear-end crashes. They attributed this observation to the fact that it becomes difficult to control the truck under the conditions of speeding. The variables for driver behavior and accident location are associated with rollover crashes, especially when speeding on low-class roads. The drivers' age and the quantity of HAZMAT transported were analyzed as linked with the severity of the crashes. They explained that the quantity of HAZMAT correlates with the inertia exerted on the truck, which makes the truck more prone to crashes. The authors also identified that weather had a significant impact on HAZMAT vehicle-involved crashes. These crashes were identified to be prevalent during summer and had a high chance of occurring during cloudy or rainy weather conditions. Again, their data also showed that fatigued driving led to 62 % of the crashes at night. These findings were in line with the study conducted by Zhao et al. (2012) based on Bayesian network models. They explained that the main factors influencing HAZMAT vehicle-involved crashes were the characteristics of vehicles and road facilities, human factors, and the parking and loading of the HAZMAT.

Existing studies have been devoted to investigating risk factors that affect the occurrence and severity of HAZMAT vehicle-involved crashes on expressways. Some of the methodologies employed include non-parametric techniques such as fault tree analysis and Bayesian networks, while a majority of them used parametric models such as ordered logit or probit models (Chen and Chen, 2011; Khattak et al., 2003; Poku-Boansi et al., 2018; Samuel et al., 2009; Zhao et al., 2012). Even though they have been successful in studying crash risk factors, there are some drawbacks to their application. Parametric methods rely on specific assumptions, which may give rise to the problem of low prediction accuracy. On the other hand, the non-parametric methods used to find the patterns of crashes may suffer from over-fitting as they require a large amount of data (Hong et al., 2020; Weng et al., 2016; Yu et al., 2019).

The current body of literature provides valuable insights into the risk factors associated with HAZMAT vehicle-involved crashes. Nonetheless, instead of finding only the influence of each independent risk factor, practitioners require a good knowledge and understanding of sets of crash contributory factors that often occur together in

HAZMAT vehicle-involved crashes on the expressways. To avoid the shortcomings of the methodologies mentioned above, researchers have shifted their attention to using association rules for mining the significant patterns between crashes and the factors that influence their occurrences. Compared to the parametric methods, this technique does not rely on any assumptions or prior knowledge. They are easy to use and provide readily understandable results.

The ARM technique is the most popular non-parametric machine learning approach for finding hidden relationships between variables in a database from the concept of data mining (Wu et al., 2019). From the literature, three kinds of basic algorithms can be used for ARM (Yu et al., 2019; Zhao and Bhowmick, 2003). These are the Apriori algorithm, Frequent Pattern Tree (FPT) algorithm, and constraint-based ARM approaches. The Apriori algorithm proposed by Agrawal and Srikant (2005) involves a straight forward approach for mining association rules in datasets. The algorithm generates many candidate itemsets through many passes over the dataset. The robust candidate generation method, together with the new pruning technique employed, makes the Apriori algorithm more efficient compared to other techniques. Another strength of Apriori approaches lies in its ability to avoid the effort wastage of counting infrequent candidate itemsets. Also, the pruning technique dramatically reduces the candidate itemsets, which lessens the computation and memory requirements.

The FPT algorithm introduced by Han and Pei (2000) avoids the candidate generation process, as in the case of the Apriori algorithms. Instead, it employs only two dataset pass overs to generate frequent itemsets, which makes it faster than Apriori algorithms. However, despite its merits, the FPT algorithm is not suitable for incremental mining, and it is very difficult to use it in an interactive mining system where users may make changes to the thresholds used in mining the association rules.

Lastly, the final type of basic approaches used in ARM is the constraint-based ARM. The concept of the constraint-based ARM is to identify only those rules that are interesting to users by making sure that they meet certain user-specified constraints. An example of such a constraint is the item constraint, which places some restrictions on the selection of items or combinations of items that are interesting to users. By including item constraints in the frequent item generation process, both Ng et al. (1998) and Srikant et al. (1997) developed algorithms for faster ARM performance. Another type of constraint which leads to the discovery of rules that follow a specific pattern was developed by Fu and Han (1995). Among these ARM techniques, the most widely used and well-accepted is the Apriori algorithm proposed by Agrawal and Srikant (2005) since it is more basic and involves a precise and iterative method of layer-by-layer search (Wu et al., 2019).

In the literature, the ARM technique has been employed in transportation safety research to investigate rules associated with roadway crashes (Das et al., 2019a; Hong et al., 2020; Montella, 2011; Montella et al., 2012; Weng et al., 2016). Geurts et al. (2005) employed the ARM technique to study crash patterns and characteristics in blackspots. Yu et al. (2019) used association rules to identify factors influencing the patterns of road crashes in Wisconsin. Results from their study showed that the odds for fatal crashes increase when the weather is clear, and the road surface is dry. Das et al. (2019a,b) investigated the patterns of traffic crashes in rainy weather conditions using association rules and demonstrated that young drivers between 15 and 20 years were more prone to being involved in run-off crashes under poor lighting conditions and when the roadways are curved. Xu et al. (2018) identified that inexperienced drivers of overloaded buses or heavy vehicles were more likely to be involved in severe crashes, and improper driving and speeding were commonly related to severe casualty crashes. More recently, Hong et al. (2020) applied ARM to discover hidden patterns and relationships in freight truck-involved crash data. The generated rules demonstrated that over speeding is highly associated with truck crashes during winter, and roadway segment-related crashes were associated with driver's faults and roadway geometry.

Notably, as a machine learning technique, researchers have used the ARM technique to find exciting connections from all types of datasets. Verma et al. (2014) analyzed 843 incidents from a steel manufacturing company using the ARM technique. In the transportation safety field, Montella (2011) and Polders et al. (2014) used the ARM technique to explore crash contributory factors at urban roundabouts using datasets of 274 crashes and 399 crashes, respectively. In identifying the factors influencing the patterns of crashes resulting from wrong-way driving on freeway exit ramps and median crossovers, Das et al. (2018) employed a database that contained 1419 crashes. Weng et al. (2016) also employed the ARM technique to analyze 371 sets of work zone casualty data to identify the patterns and contributory factors and work zone crash risk.

While numerous researchers have analyzed the factors that influence the risk of crashes, most of them focused on frequent crashes involving other vehicle types and not HAZMAT trucks. In reality, HAZMAT vehicle-involved crashes on expressways involve multiple contributory factors that are often interrelated (Zhao et al., 2012). HAZMAT forms a significant part of freight transported on expressways, and the existing transportation mechanism endangers the lives of people dwelling in the vicinity of expressways, the natural environment, other road users, and the drivers of the HAZMAT trucks themselves (Poku-Boansi et al., 2018). Therefore, to manage these unexpected situations, it is imperative to analyze the inherent connection between HAZMAT vehicle-involved crashes on expressways and a variety of crash-risk factors.

3. Methodology

The research approach employed in this study was to determine the levels of association among the contributory factors of HAZMAT vehicle-involved crashes based on the association rule mining (ARM) technique. This mining method has the advantage of being flexible and easy to understand as it lacks a specified function and requires no dependent variables. Owing to its effectiveness in discovering the associations between variables, it has become popular among researchers. The rules generated using the ARM technique are used in decision making or setting countermeasures to reduce crashes caused by some variables that frequently occur in crashes (Weng et al., 2016). Following Cunjin et al. (2015), we define a set of terms and provide primary information about the ARM technique before introducing the methodological concepts.

In the current study, we define an itemset as a set of items that include at least one reported HAZMAT vehicle-involved crash which occurred on the expressway. Therefore, an item is an element that belongs to an itemset. An m-itemset is one that contains m items. For an itemset, $I = \{i_1, i_2, \dots, i_m\}$ of m distinct attributes, consider a database $D = \{t_1, t_2, \dots, t_n\}$ consisting of HAZMAT vehicle-involved crash causality information such that each crash-risk factor in D is a subset of items contained in I. An extracted association rule can be defined for two sets of itemsets X, called the antecedent and Y, called the consequent as an implication of the form $X \Longrightarrow Y$ and satisfies the condition that $X, Y \subseteq I$ and $X \cap Y = \{\}$.

Generally, ARM identifies a group of crash-risk factors that occurs together in the event of a crash (Xu et al., 2018). The most common algorithm for ARM is the Apriori algorithm. Due to its advantages over parametric and other non-parametric methods, we employed the Apriori algorithm in this study. In the Apriori algorithm, the extraction of association rules are based on indicators, namely *support*, *confidence*, and *lift* (Xu et al., 2018). The *support* indicator for the rule $X \Longrightarrow Y$ shows the probability of both X and Y occurring together in the dataset (Lee et al., 2019). It can be mathematically defined, as shown in **Eq. (1)** below:

$$Supp(X \Longrightarrow Y) = P(X \cap Y) = \frac{|X \cup Y|}{|D|} \tag{1}$$

where $|X \cup Y|$ is the number of times both itemsets X and Y occur together, and $|D|$ represents the number of items in the HAZMAT crash database.

Confidence measures the probability that an item Y occurs given that an item X occurs, $P(Y|X)$. It serves as a way of measuring the credibility of the association rule $X \Longrightarrow Y$ (Xu et al., 2018; Yu et al., 2019), defined as

$$Conf(X \Longrightarrow Y) = \frac{Supp(X \cup Y)}{Supp(X)} = \frac{|X \cup Y|}{|X|} \quad (2)$$

where $|X|$ is the number of occurrences of only itemset X, and $|X \cup Y|$ is the number of times both itemsets X and Y occur together.

Given a rule $X \Longrightarrow Y$, with *support* s, *confidence* c, and a total set of HAZMAT vehicle-involved crash-risk factors in a database D, we can say that the rule holds with *support* s if s% of the HAZMAT vehicle-involved crash-risk factors in D are contained in $X \cup Y$. Also, the rule $X \Longrightarrow Y$ holds with *confidence* c if c% of the HAZMAT vehicle-involved crash-risk factors in D that contain X also contain Y (Lai and Cerpa, 2001).

The Apriori algorithm for searching association rules described above can be summarized into three steps (Srikant et al., 1997):

- Step 1: Scan the database to discover frequent itemsets. *Support* for the itemsets are determined, and itemsets that do not meet the minimum *support* thresholds are discarded.
- Step 2: Generate all subsets of frequent itemsets obtained from Step 1 and find the support of all the subsets.
- Step 3: Determine rules from the frequent itemsets and calculate their *confidences*. Select strong rules as those greater than a specified minimum confidence.

In general, the data mining technique described above uses a support-confidence framework to prune infrequent itemsets from the dataset (Wu et al., 2009). However, a significant drawback of the support and confidence indicators is that they are satisfied by many association rules, leading to the generation of a lot of uninteresting rules (Weng et al., 2016). Both indicators have been criticized for not considering the correlation which occurs between $Supp(X \cup Y)$ and $Supp(X)$ in **Eq. (2)**. A more practical indicator called *lift* was proposed to solve this problem. *Lift* considers how much the occurrence probability of Y changes given that X has occurred (Cunjin et al., 2015). Mathematically, it can be calculated using the formula below.

$$Lift(X \Longrightarrow Y) = \frac{Conf(X \Longrightarrow Y)}{Supp(Y)} = \frac{Supp(X \cup Y)}{Supp(X) \bullet Supp(Y)} \quad (3)$$

From **Eq. (3)**, the numerator measures the co-occurrence of itemset X and itemset Y, while the denominator measures the frequency of co-occurrence of itemset X and itemset Y of the rule based on the assumption of conditional independence (Das et al., 2019b). When the ratio is equal to 1, there is no correlation between both itemsets X and Y. When the ratio is less than 1, the occurrence of the itemset X of the rule is exclusive to the occurrence of the itemset Y of the rule. Finally, when the ratio is greater than 1, we recognize the association rule between the itemsets X and Y as valuable. The greater the ratio [$lift(X \Longrightarrow Y) > 1$], the stronger the dependency (Das et al., 2019b; Yu et al., 2019).

According to Han et al. (2012), an association is classified as strong if it satisfies minimum support and confidence thresholds. These thresholds are to be pre-specified by the user. In the literature focused on road safety estimation, the threshold values for support and confidence have usually been set at 1–4 % and 10–20 %, respectively (Xu et al., 2018). Since there is no specific rule for setting the minimum threshold, we gradually change threshold values until we find interesting rules.

4. Case study

4.1. Data description

In South Korea, the KEC oversees the construction and management of the expressways. When there is a crash, officials of the KEC are dispatched to the scene to investigate and collect crash-related information. The crash information is accumulated in a crash database for research purposes (Hong et al., 2019a). For this study, all crashes that occurred on the expressways from 2008 to 2017 were obtained for analysis. From this data, crashes involving HAZMAT vehicles were extracted for analysis. There was a total of 107,173 crashes that occurred on the expressways in the analysis period. Out of these, data for 303 crashes involving HAZMAT vehicles, together with crash information associated with each crash, were used for this study. Since these data are of high quality and were collected by trained personnel of the KEC, the results of the analysis based on it are plausible.

The crash information contained in the data used for the analysis comprises detailed information such as month of year, day of week, time of day, location of crash (route, segment, and region of crash occurrence), weather, roadway geometry (horizontal and vertical alignment characteristics), roadside features (shoulder, median), driver characteristics, vehicle characteristics, the leading cause of crash, and crash severity level.

In South Korea, unlike the KABCO injury severity that has five distinct groupings, the levels of injury severity as designed by the KEC used for reporting injury severities due to crashes are grouped into four (from A through to D). Level A represents fatal crashes (all crashes where the number of deaths > 3, injured persons > 20 or property damage cost > 1 billion Korean Won KRW). Level B shows severe injury (represents all crashes where 1 < number of deaths ≤ 3, 5 < injured persons ≤ 20 or 2.5 million KRW < property damage cost ≤ 1 billion KRW). Level C stands for evident injury (1 < injured persons ≤ 5 or 300 thousand won < property damage cost ≤ 2.5 million KRW), and Level D denotes property damage only (PDO) (damage cost ≤ 300 thousand KRW) (Hong et al., 2019c). Of all the crash observations, the number of crashes with severity level A was the least (0.99 %), and that of severity level D was the highest (53.47 %). The number of crashes tends to increase with a decreasing level of severity.

Factors such as horizontal and vertical alignment, and median and shoulder barrier types comprise roadway geometry characteristics. In reference to these factors, the majority of the crashes were identified to have occurred on straight (38.61 %) and slope less (32.34 %) roads. Next, drivers were also seen to be susceptible to crashes at sharp curves (500 m) towards the left (25.08 %), and gentle downward slopes (25.08 %). Median and shoulder barriers are installed at the median or on the shoulders of roadways to lower the overall severity potential of crash outcomes. From our data, roadway segments with fixed wall median barriers (34.98 % for fixed walls of height 127 cm and 23.10 % for fixed walls of height 81 cm) and guardrail shoulder barriers (46.20 %) are major crash zones. Rockfall barriers, which are installed to shield drivers from falling rocks, were in the minority in terms of observations. In all, only four (1.32 %) HAZMAT vehicle-involved crashes were associated with rockfall barriers. The reason for the few counts is because the expressways in these locations were constructed in rural and mountainous areas. It is understandable to have low crash observations in these segments since the annual average daily traffic in such zones are meager. Of the 303 HAZMAT vehicle-involved crashes studied in this research, 198 (65.35 %) occurred at mainlines, while 68, representing 22.44 % of all crashes, occurred at ramp sections. These findings show that most of the crashes occurred on straight and leveled mainline sections of the roadway and sharply curved ramp sections.

Driver's attributes are classified into two factors, namely, gender and age. Concerning the drivers' attributes, males had the most crashes (96.04 %) compared to females (3.96 %). Of the accident observations considered for the study, drivers within their 30s to 50s were mostly

Table 1
Descriptive statistics.

Factors	Items	Count	%	Factors	Item	Count	%
Region	Ulsan	20	6.60	Weather	Windy	1	0.33
	Yangsan	17	5.61		Cloudy	51	16.83
	Changwon	14	4.62		Fine	223	73.60
	Daegu	14	4.62		Foggy	2	0.66
	Gumi	14	4.62		Rainy	24	7.92
	Hwasung	14	4.62		Snowy	2	0.66
	(Other)	210	69.31	[1]Horizontal alignment	Left curve < 500 m	76	25.08
Month of year	April	28	9.24		Left curve > 1000 m	51	16.83
	May	34	11.22		500 m ≤ Right curve ≤ 1000 m	1	0.33
	June	38	12.54		Right curve > 1000 m	57	18.81
	July	31	10.23		Right curve < 500 m	1	0.33
	September	30	9.9		Straight	117	38.61
	October	27	8.91	Vertical alignment	0 % < Downslope < 1 %	76	25.08
	(Other)	115	37.95		1 % ≤ Downslope ≤ 3 %	37	12.21
Day of week	Monday	49	16.17		Downslope > 3 %	18	5.94
	Tuesday	42	13.86		Slope = 0	98	32.34
	Wednesday	59	19.47		1 % ≤ Upslope ≤ 3 %	41	13.53
	Thursday	59	19.47		Upslope > 3 %	13	4.29
	Friday	52	17.16		0 % < Upslope < 1 %	20	6.60
	Saturday	28	9.24	Median barrier types	Fixed wall (height = 127 cm)	106	34.98
	Sunday	14	4.62		Fixed wall (height = 81 cm)	70	23.10
Day/Night	Daytime	210	69.31		No median barrier	48	15.84
	Nighttime	93	30.69		Guardrail	30	9.90
Route	Gyeongbu-line	70	23.10		Fence	25	8.25
	Namhae-line	31	10.23		Green median	10	3.30
	Jungbu Naeryuk-line	24	7.92		(Other)	14	4.62
	Jungbu-line	22	7.26	Shoulder barrier types	Unknown	3	0.99
	Seohaean-line	22	7.26		Concrete wall	24	7.92
	Youngdong-line	16	5.28		Fence	1	0.33
	(Other)	118	38.94		Guardrail	140	46.20
Segment	Mainline	198	65.35		No shoulder barrier	57	18.81
	Ramp	68	22.44		Rockfall barrier	4	1.32
	Rest area	7	2.31		(Other)	74	24.42
	TG (hi-pass)	4	1.32	Driver's age group	Unknown	16	5.28
	TG (TCS)	13	4.29		Age 20s	11	3.63
	Tunnel	13	4.29		Age 30s	61	20.13
Severity of crash	A	3	0.99		Age 40s	89	29.37
	B	32	10.56		Age 50s	80	26.40
	C	106	34.98		Age 60s	26	8.58
	D	162	53.47		Under 20 years old	20	6.60
Cause of crash	Over-speeding	52	17.16	Vehicle types	Container	4	1.32
	Drowsy driving	51	16.83		Tank-lorry	143	47.19
	Negligence	50	16.50		Cargo truck	156	51.49
	Driver fault others	40	13.20	Vehicle sizes	Large size	196	64.69
	Improper load	33	10.89		Mid-size	52	17.16
	Flat tire	29	9.57		Small size	55	18.15
	(Other)	48	15.84	Gender	Male	291	96.04
Type of crash	Multiple vehicle involved	78	25.74		Female	12	3.96
	Single vehicle involved	225	74.26				

[1] The number of HAZMAT vehicle-related crashes at the expressway segment with horizontal alignment 500 m ≤ left curve ≤ 1000 m is 0.

involved in HAZMAT vehicle-related crashes as compared to older and drivers. Most of the crashes can be attributed to drivers' faults (over-speeding, drowsy driving, and negligence). A look at factors such as weather conditions and type of crash shows that the majority of crashes occurred during fine weather (73.60 %), and 74.26 % of the crash observations were single-vehicle crashes. Concerning factors for vehicle characteristics (vehicle type and vehicle size), 51.49 % of the HAZMAT vehicles involved in the crashes were cargo trucks, whereas the remaining, namely tank-lorry and container carrying vehicle-involved crashes, accounted for 47.19 % and 1.32 % of the crashes, respectively. Finally, concerning the factors for temporal variables (day/night and day of week), most crashes were observed to have occurred during the weekday, and during the daytime.

Table 1 shows the descriptive statistics of HAZMAT vehicle-involved crash data employed in the study. From the table, crash factors were diversified into more detailed information such as human factors, vehicle defects, region of crash, route of crash, and roadway factors. For example, some human factors leading to crashes include drivers' traffic violations such as over-speeding and drivers' conditions such as drowsy driving. This research considered a total of 18 explanatory items with 174 sub-items for the ARM analysis.

4.2. Generation of rules using the association rules mining technique

It is essential to calibrate the thresholds for support and confidence to arrive at significant findings. Defining appropriate minimum support and confidence thresholds would lead to the discovery of interesting rules. Lower threshold values result in a large number of rules, which would be very difficult to interpret due to noise and overlaps. Also, using higher thresholds would result in the generation of less and uninteresting rules (Das et al., 2018). The existing literature on ARM technique has noted that there are no criteria for setting the minimum support and confidence thresholds; however, the threshold values for which interesting rules are generated is maintained (Das et al., 2018; Hong et al., 2020; Weng et al., 2016; Xu et al., 2018; Yu et al., 2019).

A trial and error method was used to provide a reasonable set of thresholds for the study. Even though the selection of these thresholds is subjective, it was not random. As per the previous studies, we chose two sets of minimum support and confidence threshold values as 0.01 and 0.1, and 0.7 and 0.2, respectively. Upon using the first set of threshold values (0.01 support, and 0.1 confidence), the Apriori algorithm resulted in 1,283,874 rules, with the lift values distributed from 0.3–60.8. The second set showed only seven rules with lift values ranging from 0.99 to 1.02. These results did not provide significant association rules, and it was difficult to extract any patterns of contributory factors that lead to HAZMAT vehicle-involved crashes for our study.

It is desirable to have a high level of support, a large confidence factor, and a lift factor greater than 1 to attain interesting rules. Previous studies set their minimum lift threshold as one (Das et al., 2018, 2019b; Das and Sun, 2014). However, since the strength and interest of the association rule increase with increasing lift value, we set our lift value as 1.1, which is lesser than those reported by other studies (Montella et al., 2020, 2012). Given the nature of the data used for the study (HAZMAT vehicle-crashes, which are infrequent events), increasing the lift threshold further would result in a drastic reduction in the number of rules. Specifically, to obtain relevant association rules, we continuously iterated the combination of support and confidence probabilities and defined the minimum support, confidence, and lift thresholds as 0.2, 0.4, and 1.1, respectively. Defining a 20 % minimum support implies that no item or sets of items would be considered frequent if it does not appear in at least 60 HAZMAT vehicle-involved crashes (20 % of the total 303 crashes). It is worth noting that ARM is an unsupervised machine learning approach that is used to find meaningful patterns between sets of elements in every distinct transaction (crash) and is not intended to replace parametric methods (Pande and Abdel-Aty, 2009).

5. Results of analysis

Generation of the association rules was conducted in R (R Core Team, 2017) using the package "arules" (Hahsler et al., 2020), and the visualization of the results was facilitated using the package "arulesViz" (Hahsler, 2019). A total of 855 association rules were produced, and the top 30 frequent items for the HAZMAT vehicle-involved crashes on the expressways are presented in Fig. 1. From Fig. 1, readers can gain a full understanding of the dataset used in the generation of association rules through the frequency of items. The five most frequent items were found to be {Gender: male}, {Number of vehicles: single}, {Weather: fine}, {Time: daytime}, and {Segment: mainline}. The results from the basic analysis show that HAZMAT vehicle-involved crash occurrence is highly associated with male drivers, and most of the observations in the dataset were single-vehicle crashes. Also, such crashes frequently occurred during the daytime, when the weather is clear, and on the mainline segment of expressways.

Fig. 2 provides an easy way of visualizing the association rules generated using the Apriori algorithm. Essentially, it describes the relationship between the confidence, support, and lift values for the rules generated. In the graph, each dot represents a rule, and the two measures of interest, namely support and confidence are on the x and y axes, respectively. The intensity of the colors signifies the confidence. High-frequency rules are those with lower support and lift values. The support values were distributed from 0.2 to 0.96, whereas the confidence values were found to range from 0.4 to 1.0. The lift values of the rules ($X \Longrightarrow Y$), representing how much the occurrence probability of itemset Y changes given that itemset X has occurred, were observed to be in the ranges of 0.79 and 2.33. The lift value of a rule was higher when the confidence value of the rule was also large, and when the support value becomes smaller. From the scatter plot, a total of 630 rules were found to have a lift value greater than 1. As discussed in the methodology, this signifies a high association between Y and X. A value of lift less than 1 showed that having Y factors in the HAZMAT vehicle-involved crash decreased the chances of the occurrence of X in the HAZMAT vehicle-involved crash. In all, 215 rules showed associations with lift values less than 1. Ten rules were found to have lift values equal to 1, representing that there was no relationship between Y and X factors in the HAZMAT vehicle-involved crash.

Table 2 shows the rules for the top 23 lift values ordered according to the decreasing lift value. Higher lift values indicate stronger associations between the consequent of the rule or left-hand-side (LHS or X) and the antecedent of the rule or right-hand-side (RHS or Y). All 23

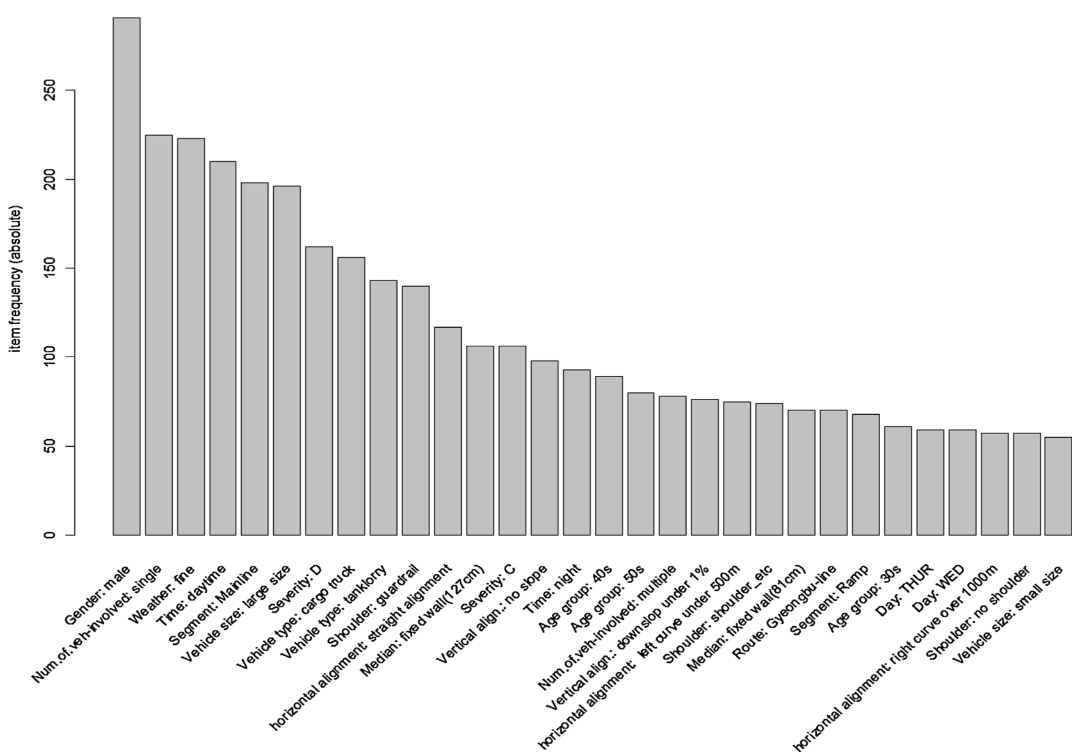


Fig. 1. Top 30 frequent items associated with HAZMAT vehicle-involved crashes.

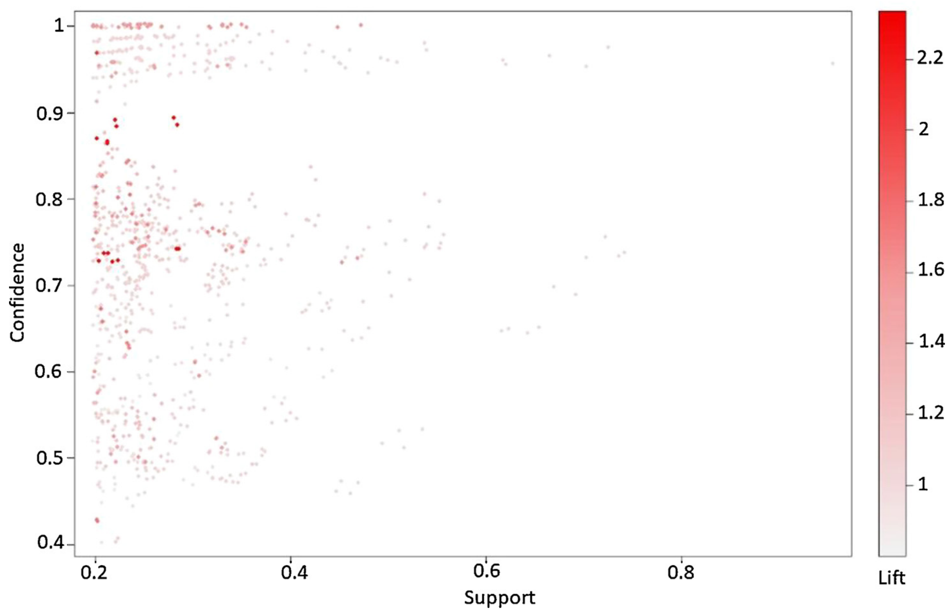


Fig. 2. Plot visualization for confidence and support values for 855 rules.

rules in Table 2 have lift values higher than 1.68, which indicates that the itemsets X and Y are highly associated with each other. The rules mostly relate to roadway alignment and drivers' characteristics. The variable category for straight horizontal alignment (no curve) is the 'consequent' for the top three rules with the highest lift values. The highest lift value is found to be 2.328 with rule {Gender: male, Vertical alignment: Slope = 0}⟹{Horizontal alignment: straight}. The rule signifies that a HAZMAT vehicle-involved crash involving a male driver on a roadway section without a vertical curve is likely to have occurred on a roadway section with no horizontal curve. The support value is 28.3 %, and the confidence value is 89.6 %.

From these values, we interpret the first rule as 28.3 % of HAZMAT vehicle-involved crashes involved males driving on flat roadway segments with no horizontal or vertical curves. It also shows that 89.6 % of the HAZMAT vehicle-involved crashes, which occurred on a road surface with no horizontal curve, occurred when a male driver had an accident on the roadway segment with no vertical curve. From the lift value 2.328, we discovered that HAZMAT vehicle-involved crashes relating to men on flat roadway segments did not happen by coincidence. Notably, the results showed that there is a strong association between segments with no horizontal and vertical curves. When these types of alignments coexist on a roadway segment, there could be a high risk of HAZMAT vehicle-involved crash occurrence.

Another interesting finding is that single HAZMAT vehicle crashes were found to be highly associated with horizontal and vertical alignments, or a tank-lorry vehicle type. Rules 7 and 8 in Table 2 show that

Table 2
Rules sorted by top 23 lift values.

No	LHS	RHS	Support	Confidence	Lift
1	{Gender: male, Vertical align.: Slope = 0}	Horizontal alignment: straight	0.283	0.896	2.328
2	{Gender: male, Vehicle type: cargo truck, Vertical alignment: Slope = 0}	Horizontal alignment: straight	0.220	0.893	2.321
3	{Vertical alignment: Slope = 0}	Horizontal alignment: straight	0.286	0.888	2.307
4	{Gender: male, Horizontal alignment: straight alignment}	Vertical alignment: Slope = 0	0.283	0.741	2.300
5	{Vehicle type: cargo truck, Vertical alignment: Slope = 0}	Horizontal alignment: straight	0.224	0.883	2.295
6	{Horizontal alignment: straight alignment, Number of vehicles-involved: single}	Vertical alignment: Slope = 0	0.211	0.736	2.282
7	{Gender: male, Horizontal alignment: straight alignment, Number of vehicles-involved: single}	Vertical alignment: Slope = 0	0.211	0.736	2.282
8	{Horizontal alignment: straight alignment, Vehicle type: cargo truck}	Vertical alignment: Slope = 0	0.224	0.731	2.268
9	{Vertical alignment: Slope = 0, Weather: fine}	Horizontal alignment: straight	0.201	0.871	2.264
10	{Gender: male, Horizontal alignment: straight alignment, Vehicle type: cargo truck}	Vertical alignment: Slope = 0	0.220	0.728	2.259
11	{Horizontal alignment: straight alignment, Weather: fine}	Vertical alignment: Slope = 0	0.201	0.726	2.253
12	{Number of vehicles-involved: single, Vertical align.: Slope = 0}	Horizontal alignment: straight	0.211	0.865	2.247
13	{Gender: male, Number of vehicles-involved: single, Vertical alignment: Slope = 0}	Horizontal alignment: straight	0.211	0.865	2.247
14	{Horizontal alignment: Left curve < 500 m, Vehicle size: large size}	Vehicle type: tank lorry	0.201	0.968	2.058
15	{Gender: male, Segment: Mainline, Vehicle type: cargo truck}	Horizontal alignment: straight	0.204	0.674	1.751
16	{Horizontal alignment: Left curve < 500 m}	Vehicle type: tank lorry	0.201	0.813	1.729
17	{Vehicle type: tank lorry}	Horizontal alignment: left curve under 500 m	0.201	0.427	1.729
18	{Vehicle size: large size, Vehicle type: tank lorry}	Horizontal alignment: left curve under 500 m	0.201	0.427	1.729
19	{Number of vehicles-involved: single, Segment: Mainline, Vehicle size: large size}	Vehicle type: tank lorry	0.234	0.807	1.715
20	{Segment: Mainline, Vehicle type: cargo truck}	Horizontal alignment: straight	0.207	0.656	1.705
21	{Gender: male, Number of vehicles-involved: single, Segment: Mainline, Vehicle size: large size}	Vehicle type: tank lorry	0.224	0.800	1.701
22	{Segment: Mainline, Time: daytime, Vehicle size: large size}	Vehicle type: tank lorry	0.204	0.795	1.690
23	{Gender: male, Vehicle type: cargo truck, Weather: fine}	Horizontal alignment: straight alignment	0.230	0.648	1.684

both antecedents {Horizontal alignment: straight alignment, Number of vehicles-involved: single} and {Gender: male, Horizontal alignment: straight alignment, Number of vehicles-involved: single} had the same consequent {Vertical alignment: Slope = 0}. The lift value is 2.282, support is 0.211, and confidence is 0.736. Also, rules 13 and 14 show that LHS {Number of vehicle-involved: single, Vertical alignment: Slope = 0} and LHS {Gender: male, Number of vehicle-involved: single, Vertical alignment: Slope = 0} are linked with HAZMAT crashes on roadways with no horizontal curve. The lift is 2.247, support is 0.211, and confidence is 0.865. Similar to the previously discussed rule, these rules demonstrate that HAZMAT vehicle crashes are likely to occur on flat-surfaced roadways with no vertical and horizontal curves. The reason for this observation is that HAZMAT vehicle drivers, especially males, feel more comfortable when driving on such flat roadways, hence making them more likely to become drowsy or drive carelessly. This result is consistent with literature which points out that male drivers are more likely to be at fault (Das et al., 2018).

The consequents from rules 20 and 22 verify that tank-lorry vehicles carrying HAZMAT are mostly involved in with single-vehicle crashes. The antecedent of rule 20 is {Number of vehicle-involved: single, Segment: Mainline, Vehicle size: large size} and that of rule 22 is {Gender: male, Number of vehicle-involved: single, Segment: Mainline, Vehicle size: large size}. They have the same consequent {Vehicle type: tank lorry}. Rule 20 has 0.234 as support and 0.807 as confidence. By these values, 23.4 % of HAZMAT vehicle-involved crashes that were single-vehicle crashes of large trucks and occurred on the mainline, were mostly tank-lorries HAZMAT vehicles. Also, the results show that 80.7 % of the HAZMAT crashes that involved large single-vehicle trucks on the mainline segments also involved tank-lorry HAZMAT vehicles. Rule 22 has 0.224 for support and 0.800 for confidence. The values showed that 22.4 % of HAZMAT vehicle-involved crashes involved a male driving a large-sized tank-lorry truck on a mainline, and such crashes result in single-vehicle crashes. Further, it shows that 80.0 % of the single-vehicle large-sized HAZMAT vehicle-involved crashes on that occurred on mainlines of expressways also involve tank-lorries.

For finding the detailed characteristic patterns of HAZMAT vehicle-involved crashes, we extracted rules for specific itemsets from the consequents (RHS), namely {Vehicle type: tank lorry}, {Vehicle type: Cargo truck}, {Number of vehicle-involved: single-vehicle crash}, and {Segment: Mainline}. These consequents of the rules frequently occurred in a Word Cloud, as shown in Fig. 3, and are of most interest to authors. Each pattern of the rules demonstrated by network graphs in Fig. 4 shows the relationship between each antecedent and consequent. In Fig. 4, a large circle size indicates a high support value; and the darker the color, the higher the lift value. The subsections below present the rules generated based on the frequently occurring consequents in the Word Cloud.

5.1. Association rules for HAZMAT tank-lorry (vehicle type)

As shown in Fig. 4(a), the crashes involving tank-lorry had a strong association with daytime, clear weather, male driver, mainline segment, horizontal alignment with a Left curve < 500 m, large size vehicle, single vehicle-involved crash, and guardrail shoulder barrier type. This was confirmed by the high lift values associated with their rules. A total of 50 rules were generated using HAZMAT tank-lorry involved crashes as consequent (RHS). The top ten rules ranked by the lift values were selected and presented in Table 3.

The highest lift value is 2.058 for the LHS {Horizontal alignment: Left curve < 500 m, Vehicle size: large size}, which is interpreted as the probability of large-sized HAZMAT vehicle-involved crashes on roadways with a left horizontal curve under 500 m was 2.058 times the probability of the HAZMAT tank-lorry crashes in our dataset. It means that HAZMAT tank-lorry crashes are highly related to vehicle size and low horizontal curve. As far as severity is concerned, crashes of severity level D was significantly associated with tank lorries. Surprisingly, there is enough evidence to show that most of the crashes involving tank lorries occurred in fine weather, and on the mainlines, with a high proportion of them being single-vehicle crashes.

5.2. Association rules for HAZMAT cargo truck (vehicle type)

HAZMAT transporting cargo trucks showed a higher propensity of being linked to daytime, clear weather, male driver, mainline segment, curve-less horizontal alignment, vertical alignment with 0 % slope, and single-vehicle involved crashes, as shown in Fig. 4(b). We noted that among the HAZMAT vehicle-involved crashes, unlike tank-lorries that are affected by horizontal curves, cargo trucks were not associated with horizontal and vertical alignments. This result can be inferred to be related to the characteristics of the vehicle type and HAZMAT type loaded in the vehicle. The height of tank-lorries is higher than regular cargo trucks, making it more unstable when driving on the curved sections of roadways. Also, since the features of both vehicle types are different, the driving style, the quantity, and type of HAZMAT contained in the vehicle may be essential factors to consider. Cargo truck drivers are more likely to drive faster and with less care compared to their other counterparts. As they drive long distances, they may also be drowsy when passing through straight roadway sections with no vertical or horizontal curves.

A total of 25 rules were generated for HAZMAT cargo truck-involved crashes. Among them, the top 10 rules sorted by the lift values were selected and are shown in Table 3. The highest lift value is 1.647 for the LHS {Horizontal alignment: straight, Weather: fine}. The results reveal that the proportion of HAZMAT vehicle-involved crashes on the expressways with straight horizontal alignment segments under clear weather was 1.647 times the proportion of the HAZMAT cargo truck crashes.

5.3. Association rules for single vehicle-involved crashes (crash type)

We identified strong associations between single HAZMAT vehicle-involved crashes and related items such as vertical alignment with slope = 0, level grade horizontal alignment, daytime, age groups (the 40s and 50s), severity level C and D, vehicle type (tank-lorry), and guardrail shoulder barrier type as shown in Fig. 4(c). Through the Apriori algorithm, a total of 83 rules were produced, and all rules were found to have lift values greater than 1. The top 10 association rules in Table 3 have lift values ranging between 1.123 and 1.182, confirming that the RHS and LHS are highly correlated. The rule with the highest lift value of 1.182 is {Gender: male, Severity: C, Time: daytime}$\Longrightarrow$ {Number of vehicles-involved: single}. The rule has 0.207 support value implying 20.7 % of the HAZMAT vehicle-involved crashes included male drivers involved in crashes with injury severity level C during the daytime produced a single HAZMAT vehicle crash. Also, from the confidence value, 87.5 % of HAZMAT vehicle-involved crashes under the condition {Gender: male, Severity: C, Time: daytime} results in a single HAZMAT vehicle crash.

The results manifest that single HAZMAT vehicle crashes are likely to be associated with good weather and are highly probable to occur during the daytime. Moreover, HAZMAT cargo truck crashes driven by males on straight horizontally aligned roads were found to be related to single HAZMAT vehicle-involved crashes. Unlike the other rules, it was found that HAZMAT vehicle crashes associated with guardrails on a roadside were related to single HAZMAT vehicle crashes. This result indicates that a single HAZMAT vehicle-involved crash is likely to be an accidental collision with a guardrail.

Due to the limitation of the number of observations in this study, we could not find specific results to show that drivers' drowsiness, carelessness, or illegal driving were associated with single HAZMAT vehicle-involved crashes. However, we can find clues from the relationships between the LHS items such as daytime, level grade, and straight roadway geometry, good weather condition, and single HAZMAT

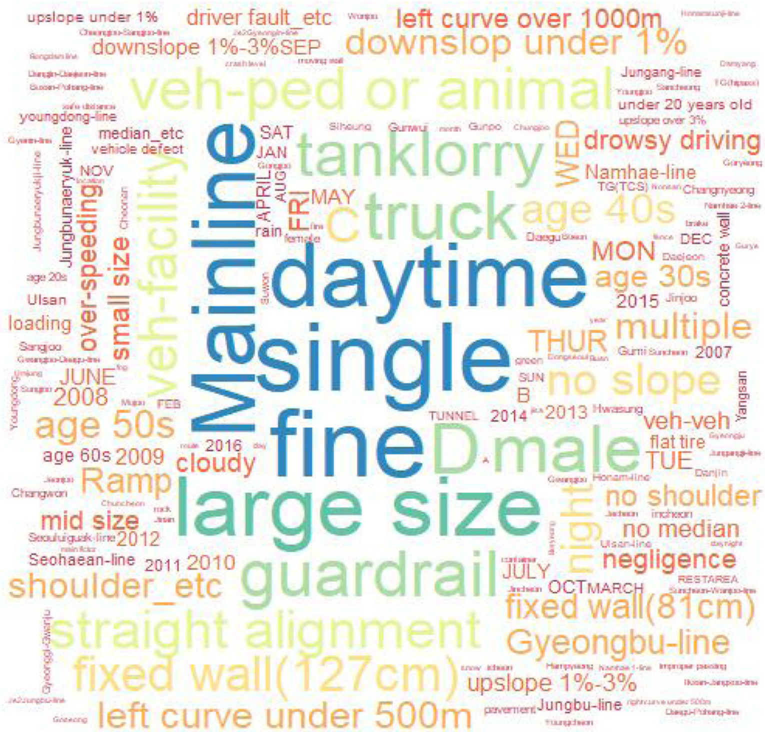
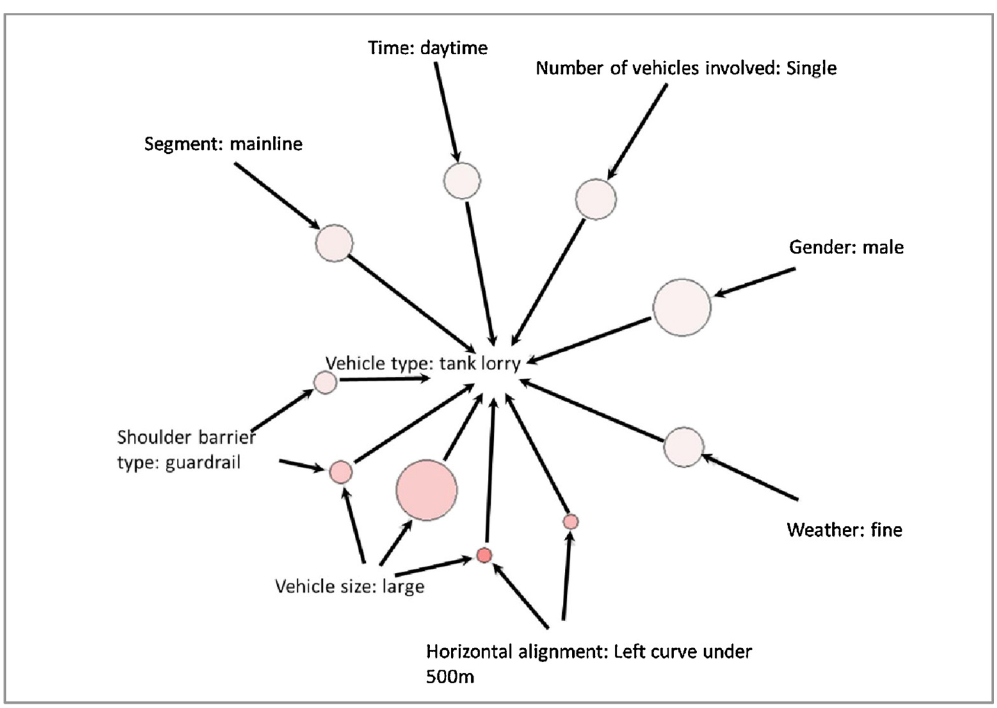


Fig. 3. Word Cloud for HAZMAT vehicle-involved crash data.

vehicle-involved crashes. As such, it is necessary to figure out as to what extent a single HAZMAT vehicle-involved crash is related to the HAZMAT vehicle driver's behavior.

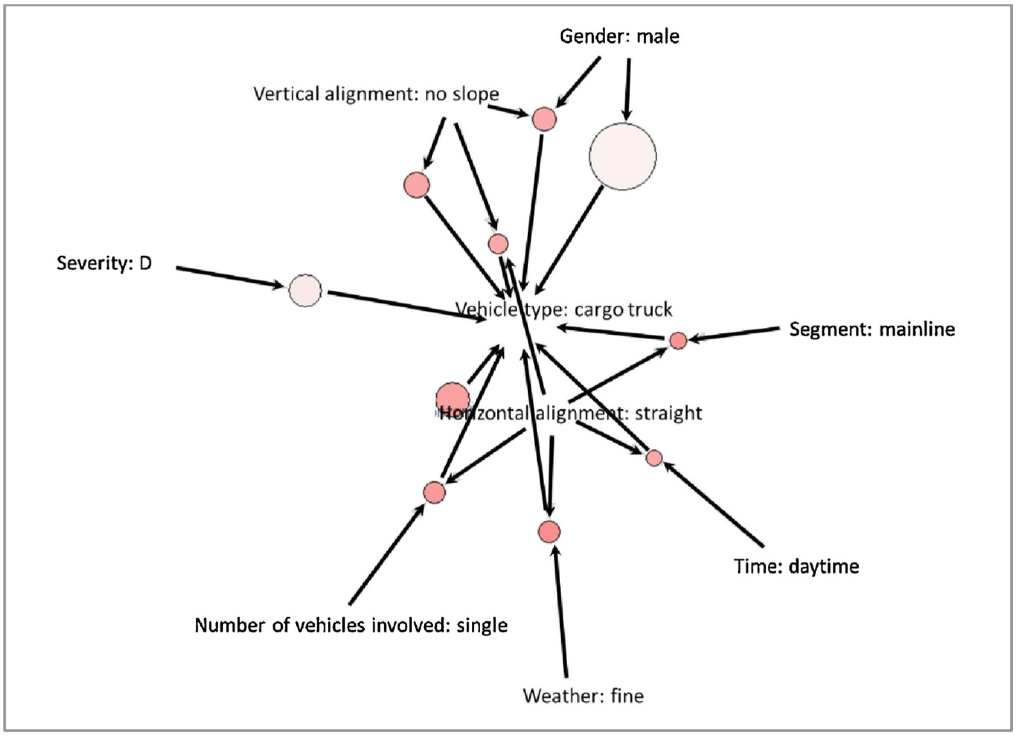
5.4. Association rules with HAZMAT vehicle crashes on mainline segments (location)

HAZMAT vehicle-involved crashes on mainlines are associated with clear weather, nighttime, large size vehicle, tank-lorry, multi-vehicle involved, guardrail shoulder barrier type, a fixed wall of height 81 cm at the median, a fixed wall of height 127 cm at the median, and left



(a) HAZMAT vehicle type: tank lorry

Fig. 4. Graph-based visualization of items and top 10 rules.

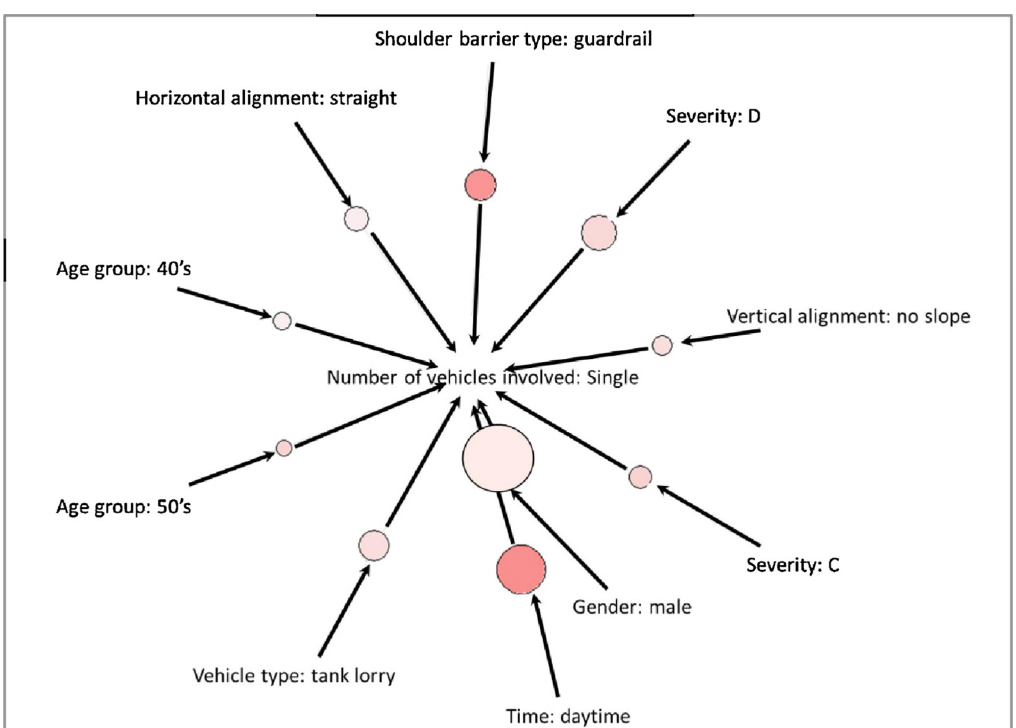


(b) HAZMAT vehicle type: cargo truck

Fig. 4. (*continued*)

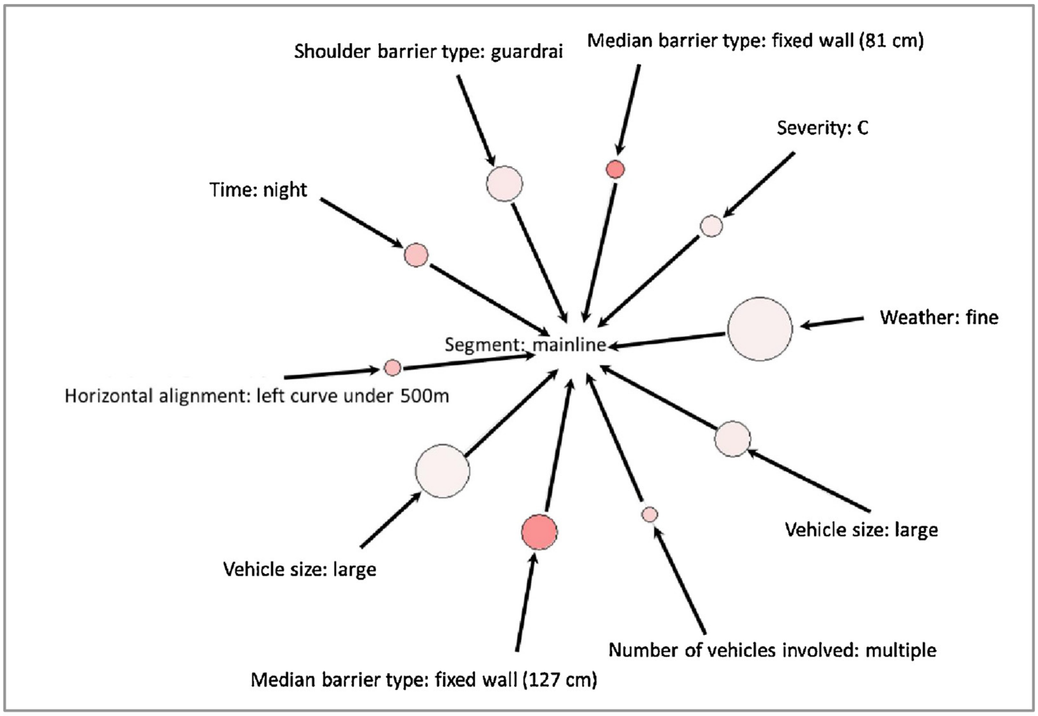
horizontal curve under 500 m. Fig. 4(d) shows their relationship visualized by the network graph. From the Apriori algorithm, a total of 47 rules were generated. Out of these, 46 rules except one showed lift values greater than 1. One rule was found to be equal to 1, which demonstrated that the RHS and LHS in the rule were not associated. As such, the rule was dropped.

The highest lift value was found to be 1.47, and two rules had the same value. One rule's LHS was {Median barrier type: fixed wall (height = 81 cm)} and another's LHS was {Median barrier type: fixed wall (height = 127 cm), Number of vehicle-involved: single}. The



(c) Single vehicle crash

Fig. 4. (*continued*)



(d) Segment: Mainline

Fig. 4. (*continued*)

support value of the rules with the highest lift was 0.22, indicating that 22 % of the HAZMAT vehicle crashes occurred on roadways with a fixed wall median barrier type of height of 81 cm or 127 cm also involved HAZMAT crashes on the mainline of roadways. The rules show that HAZMAT vehicle-involved crashes on mainlines are associated with the existence of the median. Among the top 10 rules in Table 3, eight rules showed a strong relationship with HAZMAT crashes that occur on roadways with the fixed wall median barrier type of height 127 cm.

6. Discussion

In this study, association rule mining (ARM) was applied for analyzing the characteristics of HAZMAT vehicle-involved crashes on expressways. ARM has been identified to be more advantageous compared to the parametric methods, which depend on specific assumptions, and other non-parametric approaches which generally suffer from overfitting (Weng et al., 2016; Yu et al., 2019). After obtaining appropriate thresholds using a trial and error approach, we obtained a total of 855 rules for HAZMAT vehicle-involved crashes. Analysis results showed that HAZMAT vehicles are sensitive to mainline locations, fine weather, leveled roadways, size, and type of truck. The key findings of the study are summarized as follows.

In general, the five most common factors associated with HAZMAT vehicle-involved crashes were identified. From the results, these crashes are likely to occur when the driver is a male, when the weather is clear, during the day time, and on mainline segments. Also, HAZMAT vehicle-involved crashes are likely to be single-vehicle crashes. The rule with the greatest interdependence (strength or interestingness) had a lift value of 2.328. According to the rule, HAZMAT vehicle-involved crashes that occur on roadway segments with no horizontal curves (straight) are likely to be driven by male drivers on roads with 0 % slope. This result is intuitive as drivers are less likely to drive carefully on straight and flat (no vertical slope) roadway segments (Hong et al., 2020; Yu et al., 2019).

With regards to HAZMAT lorry type, the study results highlighted that, male drivers driving large tank-lorry HAZMAT trucks during the daytime, in clear weather, on mainline segments with sharp left curves (< 500 m) are likely to crash into shoulder guardrails. It was found that the probability of large-sized HAZMAT vehicle-involved crashes on the roadway with a left horizontal curve under 500 m was 2.058 times the probability of the HAZMAT tank-lorry crashes in our dataset. This finding has various potential explanations. Large HAZMAT truck drivers are mostly unaware of upcoming horizontal curves or may find it difficult when turning at sharp-curved sections due to the weight of their vehicle. As such, most of the crash involving such vehicles are likely to end up in single-vehicle crashes.

Concerning HAZMAT cargo trucks, the study results provided enough evidence to show that they are highly linked to daytime, clear weather, male drivers, mainline segments with no horizontal alignment and vertical alignment. A crash involving this vehicle type is also likely to result in single-vehicle crashes. An interesting finding is that cargo trucks were related to crashes occurring on level grade and straight roadway segments, while tank-lorries crashes were associated with crashes occurring on horizontal curve sections of the expressway.

Finally, single HAZMAT vehicle-involved crashes were found to have rules with the items such as no grade vertical alignment, flat horizontal alignment, daytime, driver's in their 40s and 50s, severity levels C and D, tank-lorry, and guardrail shoulder barrier type. Primarily, we discovered that HAZMAT vehicles that crash into guardrails on the roadside were related to single HAZMAT vehicle crashes. Moreover, HAZMAT vehicle-involved crashes on mainlines have rules linked with clear weather, nighttime, large-sized vehicle, tank-lorry, multi-vehicle involved, guardrail (shoulder barrier type), both fixed wall (median barrier type) of height 81 cm and 127 cm, and left horizontal curve under 500 m. From the ARM results, HAZMAT vehicle-involved crashes on mainlines are highly linked with the existence of the fixed median barrier types.

Some of the results of this study were consistent with the literature on vehicular crash analysis using ARM. Das et al. (2018) mentioned that male drivers were found to be mostly at fault in crashes. Also, as seen in this study, crashes on level surface roads was identified to be linked with male drivers, and mostly results in single-vehicle crashes during

Table 3
Top 10 rules ranked by lift values for each specific items.

RHS	Top 10 Rules	Support	Confidence	Lift
Vehicle type: tank-lorry	{Horizontal alignment: Left curve < 500 m, Vehicle size: large size}	0.201	0.968	2.058
	{Horizontal alignment: Left curve < 500 m}	0.201	0.813	1.729
	{Number of vehicles-involved: single, Segment: Mainline, Vehicle size: large size}	0.234	0.807	1.715
	{Gender: male, Number of vehicles-involved: single, Segment: Mainline, Vehicle size: large size}	0.224	0.800	1.701
	{Segment: Mainline, Time: daytime, Vehicle size: large size}	0.204	0.795	1.690
	{Gender: male, Severity: D, Vehicle size: large size}	0.230	0.787	1.672
	{Severity: D, Vehicle size: large size}	0.237	0.783	1.664
	{Gender: male, Time: daytime, Vehicle size: large size, Weather: fine}	0.243	0.771	1.639
	{Time: daytime, Vehicle size: large size, Weather: fine}	0.253	0.770	1.637
	{Number of vehicles-involved: single, Vehicle size: large size, Weather: fine}	0.270	0.766	1.629
Vehicle type: cargo truck	{Horizontal alignment: straight alignment, Weather: fine}	0.234	0.845	1.647
	{Gender: male, Horizontal alignment: straight alignment, Weather: fine}	0.230	0.843	1.643
	{Horizontal alignment: straight alignment, Segment: Mainline}	0.207	0.829	1.615
	{Gender: male, Horizontal alignment: straight alignment, Segment: Mainline}	0.204	0.827	1.611
	{Horizontal alignment: straight alignment, Number of vehicles-involved: single}	0.234	0.816	1.590
	{Gender: male, Horizontal alignment: straight alignment, number of vehicles-involved: single}	0.234	0.816	1.590
	{Horizontal alignment: straight alignment}	0.306	0.795	1.549
	{Gender: male, Horizontal alignment: straight alignment}	0.303	0.793	1.546
	{Vertical align.: Slope = 0}	0.253	0.786	1.531
	{Horizontal alignment: straight alignment, Time: daytime}	0.201	0.782	1.524
Number of vehicles-involved: single	{Gender: male, Severity: C, Time: daytime}	0.207	0.875	1.182
	{Gender: male, Time: daytime, Vehicle type: cargo truck, Weather: fine}	0.214	0.867	1.171
	{Severity: C, Time: daytime}	0.207	0.863	1.166
	{Horizontal alignment: straight alignment, Time: daytime}	0.220	0.859	1.161
	{Gender: male, Horizontal alignment: straight alignment, Time: daytime}	0.220	0.859	1.161
	{Time: daytime, Vehicle type: cargo truck, Weather: fine}	0.217	0.846	1.143
	{Gender: male, Shoulder: guardrail, Time: daytime, Weather: fine}	0.207	0.840	1.135
	{Gender: male, Time: daytime, Weather: fine}	0.418	0.836	1.129
	{Gender: male, Time: daytime, Vehicle type: cargo truck}	0.280	0.833	1.126
	{Shoulder: guardrail, Time: daytime, Weather: fine}	0.211	0.831	1.123
Segment: mainline	{Median barrier type: fixed wall(height = 81 cm)}	0.220	0.957	1.470
	{Median barrier type: fixed wall(height = 127 cm), Number of vehicles-involved: single}	0.220	0.957	1.470
	{Gender: male, Median barrier type: fixed wall(height = 127 cm), number of vehicles-involved: single}	0.217	0.957	1.469
	{Gender: male, Median barrier type: fixed wall(height = 81 cm)}	0.204	0.954	1.464
	{Median barrier type: fixed wall(height = 127 cm)}	0.332	0.953	1.463
	{Median barrier type: fixed wall(height = 127 cm), Weather: fine}	0.263	0.952	1.462
	{Gender: male, Median barrier type: fixed wall(height = 127 cm)}	0.326	0.952	1.462
	{Gender: male, Median barrier type: fixed wall(height = 127 cm), Weather: fine}	0.260	0.952	1.461
	{Median barrier type: fixed wall(height = 127 cm), Vehicle size: large size}	0.224	0.944	1.450
	{Gender: male, Median barrier type: fixed wall(height = 127 cm), Vehicle size: large size}	0.217	0.943	1.448

the day. Using parametric models, Hong et al. (2019a) identified that the presence of guardrails increased the probability of having freight truck crashes on the mainlines of expressways in Korea. The authors also identified that all types of fixed median barriers were associated with crashes on expressways. These results support our claim that a HAZMAT truck crash is likely to occur on mainlines when there are shoulder guardrails and fixed median barriers present. Although weather variables were found to be significant in their model, a previous study by Abdel-Aty (2003) found the weather variable insignificant in their model. However, using the ARM technique in this current study, the patterns concerning weather were uncovered more clearly. Hong et al. (2020) also analyzed truck crashes using the Apriori algorithm, and their results also showed strong associations between weather and truck crashes. Specifically, the authors demonstrated that medium-weight trucks were highly linked with crashes occurring during the rainy weather, whereas drowsy driving during the evening had a high propensity of causing crashes during fine weather. These references highlight the potential of ARM to discover hidden patterns in data compared to traditional crash analysis models.

7. Conclusion and recommendation

Assessment of HAZMAT vehicle-related incidents has become an important topic in the field of transportation logistics (Hong et al., 2019b). However, research geared at understanding the contributory factors of such roadway incidents are in the minority. Extracting the relationship between HAZMAT crash attributes and impact factors using ARM provides an opportunity to make informed decisions and enact the right policies for reducing the risk of crashes involving HAZMAT vehicles. Pande and Abdel-Aty (2009) argue that ARM allows researchers to study data without limiting the amount of information the data contains. For researchers, engineers, and policymakers in the HAZMAT transportation field, ARM is a plausible technique that provides a systematic and straightforward way of exhausting all possible crash patterns, and to interpret the relationship between antecedents and consequents of association rules for HAZMAT vehicle-involved crashes.

The investigations carried out in this research are necessary as it provides a meaningful visualization of the common factors that lead to HAZMAT vehicle-involved crashes. A couple of significant factor groups have been identified based on the extracted rules generated in the current study. From the results, more specific countermeasures and policies can be established. Concerning driver-specific factors, it was identified that male drivers and older drivers (aged 40–50) are mostly associated with HAZMAT vehicle involved crashes. It is necessary to find ways of monitoring the attitude of drivers they drive on mainlines during the day. This can be done by installing special cameras on expressways to detect and control risky driving behavior. Images from the camera could be employed in establishing risky driving patterns such as mobile phone use while driving (Berri et al., 2014), and drivers found culpable should be penalized severely to serve as a deterrent to others. Policymakers have to ensure that HAZMAT vehicle drivers receive periodic education and training on how to drive and manage incidents involving HAZMAT vehicles. Also, concerning environmental conditions, HAZMAT vehicle crashes occurring during the nighttime could be addressed if roadway designers provide enough lighting and reflectors on segments that have poor visibility at night. Governments should make and enforce policies to check the working schedule of HAZMAT vehicle drivers to reduce the incidents resulting from the tiredness of old drivers.

With regards to roadway attributes, expressways with long mainline segments that have no vertical or horizontal curves tend to encourage risky driving as drivers become more relaxed as they drive through such segments. As such, most drivers end up in single-vehicle crashes (hitting guardrails and other roadway facilities). The installation of safety devices such as rumble strips and roadway reflectors at the shoulders, at the median and near tollgates of expressways, could help keep the driver alert. Transportation engineers should review their median crash history to identify crash-hotspots for the installation of the rumble strips at the median or shoulders of the roads to alert drivers who may be driving towards such directions. Also, to reduce the vulnerability of HAZMAT vehicle drivers at sharp curves, roadway designers have to provide enough information through the installation of road signs to alert drivers approaching curved segments. The speed limit at those curved segments could be reduced to lessen the number of HAZMAT vehicle-related crashes.

The contribution of this research to the literature is that it seeks to identify hidden relationships in HAZMAT vehicle-involved crash-risk factors and enhances knowledge on the driver, vehicle, environment, and roadway-related factors that are likely to be the cause of HAZMAT vehicle-related crashes on expressways. To our knowledge, the present study is one of the first attempts that employs a more efficient ARM technique to explore HAZMAT truck-involved crashes. From a policy viewpoint, this kind of analysis provides useful insights to decision-makers for the development of appropriate safety measures that can possibly reduce the number of HAZMAT vehicle-related crashes and fatalities. Furthermore, the results of this study provide a basis for setting practical countermeasures for reducing HAZMAT vehicle-related crashes on expressways.

This study is not devoid of limitations. In verifying that each additional item in the rules leads to an increase in the lift value, some researchers employed the lift increase criterion (LIC) condition to select strong association rules (López et al., 2014; Montella et al., 2020, 2012, 2011). Using the LIC criterion, together with the standard criteria for ARM, can be very instrumental in extracting the most interesting conclusions. However, as our article compares and matches with similar studies in other fields that have shown that extracting rules using appropriate lift, support and confidence thresholds can also provide interesting association rules (Das et al., 2018, 2019b; Das and Sun, 2014; Hong et al., 2020; Pande and Abdel-Aty, 2009; Verma et al., 2014; Weng et al., 2016; Yu et al., 2019), we did not consider using the LIC criterion. As a promising avenue for future research, we would consider incorporating the LIC criterion in the future. Another limitation of this study is that it was challenging to select the support and confidence threshold values as there are no specific criteria to set minimum thresholds. We overcame this challenge by iterating the algorithm over a hundred times until we found appropriate thresholds that provided interesting rules. Also, due to the interest in analyzing HAZMAT vehicle-involved crashes, which are infrequent events (fewer crash counts), we could not find any interesting rules having severity levels as consequents. In the future, different thresholds for support and confidence would be applied to a larger dataset to confirm the consistency of the results reported in this study.

CRediT authorship contribution statement

Jungyeol Hong: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - review & editing. **Reuben Tamakloe:** Formal analysis, Writing - original draft, Visualization, Validation. **Dongjoo Park:** Supervision, Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors wish to thank the Korea Expressway Corporation (KEC) for providing the crash data used in this paper. This work was supported by the National Research Foundation of Korea (NRF)grant funded by the Korean government (Ministry of Science and ICT) (no. 2017R1C1B2010175).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.aap.2020.105497.

References

Abdel-Aty, M., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. J. Safety Res. 34 (5), 597–603. https://doi.org/10.1016/J.JSR.2003.05.009.

Agrawal, R., Srikant, R., 2005. Fast algorithms for mining association rules. Proc. 20th Int. Conf. Very Large Data Bases, VLDB 487–499.

Azimi, G., Rahimi, A., Asgari, H., Jin, X., 2020. Severity analysis for large truck rollover crashes using a random parameter ordered logit model. Accid. Anal. Prev. 135, 105355. https://doi.org/10.1016/j.aap.2019.105355.

Berri, R.A., Sinva, A.G., Parpinelli, R.S., Girardi, E., Arthur, R., 2014. A pattern recognition system for detecting use of mobile phones while driving. 2014 International Conference on Computer Vision Theory and Applications (VISAPP) 411–418.

Chang, L.Y., Chen, W.C., 2005. Data mining of tree-based models to analyze freeway accident frequency. J. Safety Res. https://doi.org/10.1016/j.jsr.2005.06.013.

Chen, F., Chen, S., 2011. Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways. Accid. Anal. Prev. 43 (5), 1677–1688. https://doi.org/10.1016/j.aap.2011.03.026.

Cunjin, X., Wanjiao, S., Lijuan, Q., Qing, D., Xiaoyang, W., 2015. A mutual-information-based mining method for marine abnormal association rules. Comput. Geosci. 76, 121–129. https://doi.org/10.1016/j.cageo.2014.12.001.

Das, S., Sun, X., 2014. Investigating the pattern of traffic crashes under rainy weather by association rules in data mining. Transportation Research Board 93rd Annual Meeting. Transportation Research Board, Washington, DC.

Das, S., Dutta, A., Jalayer, M., Bibeka, A., Wu, L., 2018. Factors influencing the patterns of wrong-way driving crashes on freeway exit ramps and median crossovers: exploration using ‘Eclat’ association rules to promote safety. Int. J. Transp. Sci. Technol. 7 (2), 114–123. https://doi.org/10.1016/J.IJTST.2018.02.001.

Das, A., Ahmed, M.M., Ghasemzadeh, A., 2019a. Using trajectory-level SHRP2 naturalistic driving data for investigating driver lane-keeping ability in fog: an association rules mining approach. Accid. Anal. Prev. 129, 250–262. https://doi.org/10.1016/j.aap.2019.05.024.

Das, S., Dutta, A., Avelar, R., Dixon, K., Sun, X., Jalayer, M., 2019b. Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures. Int. J. Urban Sci. 23 (1), 30–48. https://doi.org/10.1080/12265934.2018.1431146.

Fang, K., Ke, G.Y., Verma, M., 2017. A routing and scheduling approach to rail transportation of hazardous materials with demand due dates. Eur. J. Oper. Res. https://doi.org/10.1016/j.ejor.2017.01.045.

Frank, W.C., Thill, J.C., Batta, R., 2000. Spatial decision support system for hazardous material truck routing. Transp. Res. Part C Emerg. Technol. https://doi.org/10.1016/S0968-090X(00)00007-3.

Fu, Y., Han, J., 1995. Meta-rule-guided mining of association rules in relational databases. KDOOD/TDOOD 39–46.

Geurts, K., Thomas, I., Wets, G., 2005. Understanding spatial concentrations of road accidents using frequent item sets. Accid. Anal. Prev. 37 (4), 787–799. https://doi.org/10.1016/j.aap.2005.03.023.

Ghaderi, A., Burdett, R.L., 2019. An integrated location and routing approach for transporting hazardous materials in a bi-modal transportation network. Transp. Res. Part E Logist. Transp. Rev. https://doi.org/10.1016/j.tre.2019.04.011.

Ghaleh, S., Omidvari, M., Nassiri, P., Momeni, M., Mohammadreza Miri Lavasani, S.,

2019. Pattern of safety risk assessment in road fleet transportation of hazardous materials (oil materials). Saf. Sci. https://doi.org/10.1016/j.ssci.2019.02.039.
Greibe, P., 2003. Accident prediction models for urban roads. Accid. Anal. Prev. 35, 273–285.
Haastrup, P., Brockhoff, L., 1990. Severity of accidents with hazardous materials. A comparison between transportation and fixed installations. J. Loss Prev. Process Ind. https://doi.org/10.1016/0950-4230(90)80010-8.
Hahsler, M., 2019. arulesViz: Visualizing Association Rules and Frequent Itemsets, R package version 1.3-3. (accessed. 5.4.20). https://CRAN.R-project.org/package=arulesViz.
Hahsler, M., Buchta, C., Gruen, B., Hornik, K., 2020. arules: Mining Association Rules and Frequent Itemsets [WWW Document]. URL https://cran.r-project.org/package=arules (accessed 5.4.20).
Han, J., Pei, J., 2000. Mining frequent patterns by pattern-growth: methodology and implications. SIGKDD Explor. 2 (2), 14–20.
Han, J., Kamber, M., Pei, J., 2012. Data Mining: Concepts and Techniques. Elsevier Inc., Maryland.
Holeczek, N., 2019. Hazardous materials truck transportation problems: a classification and state of the art literature review. Transp. Res. Part D Transp. Environ. 69, 305–328. https://doi.org/10.1016/j.trd.2019.02.010.
Hong, J., Park, J., Lee, G., Park, D., 2019a. Endogenous commercial driver's traffic violations and freight truck-involved crashes on mainlines of expressway. Accid. Anal. Prev. 131, 327–335. https://doi.org/10.1016/j.aap.2019.07.026.
Hong, J., Tamakloe, R., Lee, G., Park, D., 2019b. Insight from scientific study in logistics using text mining. Transp. Res. Rec. J. Transp. Res. Board 2673 (4), 97–107. https://doi.org/10.1177/0361198119834905.
Hong, J., Tamakloe, R., Park, D., 2019c. A comprehensive analysis of multi-vehicle crashes on expressways: a double hurdle approach. Sustainability 11, 10. https://doi.org/10.3390/su11102782.
Hong, J., Tamakloe, R., Park, D., Choi, Y., 2019d. Estimating incident duration considering the unobserved heterogeneity of risk factors for trucks transporting HAZMAT on expressways. Transp. Res. Rec. https://doi.org/10.1177/0361198119827925.
Hong, J., Tamakloe, R., Park, D., 2020. Discovering insightful rules among truck crash characteristics using Apriori algorithm. J. Adv. Transp. 2020, 1–16. https://doi.org/10.1155/2020/4323816.
Khattak, A.A.J., Schneider, R.J.R., Targa, F., 2003. Risk factors in large truck rollovers and injury severity: analysis of single-vehicle collisions. Transp. Res. Rec. 22.
Lai, K., Cerpa, N., 2001. In: Proceedings of the OPTIMA Conference. Support vs confidence in association rule algorithms. Curicó, Chile.
Lee, S., Cha, Y., Han, S., Hyun, C., 2019. Application of association rule mining and social network analysis for understanding causality of construction defects. Sustainability 11, 3. https://doi.org/10.3390/su11030618.
Liu, X., Liu, C., Hong, Y., 2017. Analysis of multiple tank car releases in train accidents. Accid. Anal. Prev. 107, 164–172. https://doi.org/10.1016/j.aap.2017.07.007.
López, G., Abellán, J., Montella, A., de Oña, J., 2014. Patterns of single-vehicle crashes on two-lane rural highways in Granada Province. Transp. Res. Rec. J. Transp. Res. Board 2432 (1), 133–141. https://doi.org/10.3141/2432-16.
Ma, X., Xing, Y., Lu, J., 2018. Causation analysis of hazardous material road transportation accidents by Bayesian network using genie. J. Adv. Transp. 2018. https://doi.org/10.1155/2018/6248105.
Montella, A., 2011. Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. Accid. Anal. Prev. 43 (4), 1451–1463. https://doi.org/10.1016/J.AAP.2011.02.023.
Montella, A., Aria, M., D'Ambrosio, A., Mauriello, F., 2011. Data-mining techniques for exploratory analysis of pedestrian crashes. Transp. Res. Rec. J. Transp. Res. Board 2237 (1), 107–116. https://doi.org/10.3141/2237-12.
Montella, A., Aria, M., D'Ambrosio, A., Mauriello, F., 2012. Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. Accid. Anal. Prev. 49, 58–72. https://doi.org/10.1016/j.aap.2011.04.025.
Montella, A., de Oña, R., Mauriello, F., Rella Riccardi, M., Silvestro, G., 2020. A data mining approach to investigate patterns of powered two-wheeler crashes in Spain. Accid. Anal. Prev. 134, 105251. https://doi.org/10.1016/j.aap.2019.07.027.
Moudon, A.V., Lin, L., Jiao, J., Hurvitz, P., Reeves, P., 2011. The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County. Accid. Anal. Prev. 43 (1), 11–24. https://doi.org/10.1016/J.AAP.2009.12.008.
Ng, R.T., Lakshmanan, L.V.S., Han, J., Pang, A., 1998. Exploratory mining and pruning optimizations of constrained associations rules. ACM SIGMOD Rec. 27 (2), 13–24. https://doi.org/10.1145/276305.276307.
Oggero, A., Darbra, R.M., Muñoz, M., Planas, E., Casal, J., 2006. A survey of accidents occurring during the transport of hazardous substances by road and rail. J. Hazard. Mater. 133 (1–3), 1–7. https://doi.org/10.1016/j.jhazmat.2005.05.053.
Palanisamy, S., Sebastian, J., Venkatesan, S., 2015. Safety analysis on hazardous chemicals transportation by Indian roads. Sci. Res. Essays 10 (2), 53–57. https://doi.org/10.5897/SRE2014.6142.
Pande, A., Abdel-Aty, M., 2009. Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. Saf. Sci. https://doi.org/10.1016/j.ssci.2007.12.001.
Poku-Boansi, M., Tornyeviadzi, P., Adarkwa, K.K., 2018. Next to suffer: population exposure risk to hazardous material transportation in Ghana. J. Transp. Health 10, 203–212. https://doi.org/10.1016/j.jth.2018.06.009.
Polders, E., Daniels, S., Casters, W., Brijs, T., 2014. Identifying crash patterns on roundabouts. Traffic Inj. Prev. 16 (2), 202–205. https://doi.org/10.1080/15389588.2014.927576.
Pompone, E.C., de Oliveira Neto, G.C., 2019. A survey on accidents in the road transportation of hazardous materials in Sao Paulo, Brazil, from 1983 to 2015. Transp. Res. Rec. 2673 (2), 285–893. https://doi.org/10.1177/0361198119827915.
R Core Team, 2017. A Language and Environment for Statistical Computing [WWW Document]. URL https://www.r-project.org/ (accessed 01.20.20).
Samuel, C., Keren, N., Shelley, M.C., Freeman, S.A., 2009. Frequency analysis of hazardous material transportation incidents as a function of distance from origin to incident location. J. Loss Prev. Process Ind. 22 (6), 783–790. https://doi.org/10.1016/j.jlp.2009.08.013.
Shankar, V., Mannering, F., 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. J. Safety Res. 27 (3), 183–194. https://doi.org/10.1016/0022-4375(96)00010-2.
Shen, X., Yan, Y., Li, X., Xie, C., Wang, L., 2014. Analysis on tank truck accidents involved in road hazardous materials transportation in China. Traffic Inj. Prev. 15 (7), 762–768. https://doi.org/10.1080/15389588.2013.871711.
Shibata, A., Fukuda, K., 1994. Risk factors of fatality in motor vehicle traffic accidents. Accid. Anal. Prev. 26 (3), 391–397. https://doi.org/10.1016/0001-4575(94)90013-2.
Srikant, R., Vu, Q., Agrawal, R., 1997. Mining association rules with item constraints. KDD-97 Proceedings.
Uddin, M., Huynh, N., 2018. Factors influencing injury severity of crashes involving HAZMAT trucks. Int. J. Transp. Sci. Technol. 7 (1), 1–9. https://doi.org/10.1016/j.ijtst.2017.06.004.
US Department of Transportation, 2019. Incident statistics. [WWW Document]. Natl. Transp. Stat (accessed 12.09.19). https://portal.phmsa.dot.gov/analytics/saw.dll?PortalPages.
Verma, A., Das Khan, S., Maiti, J., Krishna, O.B., 2014. Identifying patterns of safety related incidents in a steel plant using association rule mining of incident investigation reports. Saf. Sci. 70, 89–98. https://doi.org/10.1016/J.SSCI.2014.05.007.
Weng, J., Zhu, J.Z., Yan, X., Liu, Z., 2016. Investigation of work zone crash casualty patterns using association rules. Accid. Anal. Prev. 92, 43–52. https://doi.org/10.1016/j.aap.2016.03.017.
Wu, H., Lu, Z., Pan, L., Xu, R., 2009. An improved Apriori-based algorithm for association rules mining. Sixth International Conference on Fuzzy Systems and Knowledge Discovery 51–55.
Wu, B., Zhang, J.H., Yan, X.P., Yip, T.L., 2019. Use of association rules for cause-effects relationships analysis of collision accidents in the Yangtze River. In: Weintrit, A., Neumann, T. (Eds.), Advances in Marine Navigation and Safety of Sea Transportation. CRC Press/Balkema, Leiden, pp. 65.
Xu, C., Bao, J., Wang, C., Liu, P., 2018. Association rule analysis of factors contributing to extraordinarily severe traffic crashes in China. J. Safety Res. 67, 65–75. https://doi.org/10.1016/j.jsr.2018.09.013.
Yang, J., Li, F., Zhou, J., Zhang, L., Huang, L., Bi, J., 2010. A survey on hazardous materials accidents during road transport in China from 2000 to 2008. J. Hazard. Mater. 184 (1–3), 647–653. https://doi.org/10.1016/J.JHAZMAT.2010.08.085.
Yu, S., Jia, Y., Sun, D., 2019. Identifying factors that influence the patterns of road crashes using association rules: a case study from Wisconsin, United States. Sustainability 11 (7), 1–14. https://doi.org/10.3390/su11071925.
Zhao, Q., Bhowmick, S., 2003. Association Rule Mining: A Survey. Nanyang Technological University, Singapore.
Zhao, L., Wang, X., Qian, Y., 2012. Analysis of factors that influence hazardous material transportation accidents based on Bayesian networks: a case study in China. Saf. Sci. 50 (4), 1049–1055. https://doi.org/10.1016/j.ssci.2011.12.003.
Zhao, L., Qian, Y., Hu, Q.-M., Jiang, R., Li, M., Wang, X., 2018. An analysis of hazardous chemical accidents in China between 2006 and 2017. Sustainability 10 (8), 2935. https://doi.org/10.3390/su10082935.