International Journal of Transportation Science and Technology 7 (2018) 114–123

Contents lists available at ScienceDirect

International Journal of Transportation Science and Technology

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

Factors influencing the patterns of wrong-way driving crashes on freeway exit ramps and median crossovers: Exploration using 'Eclat' association rules to promote safety

Subasish Das [a,*], Anandi Dutta [b], Mohammad Jalayer [c], Apoorba Bibeka [a], Lingtao Wu [a]

[a] Texas A&M Transportation Institute, 3135 Texas A&M University, College Station, TX 77843, United States
[b] Computer Science and Engineering Department, Texas A&M University, College Station, TX 77843, United States
[c] Center for Advanced Infrastructure and Transportation, Rutgers, The State University of New Jersey, 100 Brett Road, Piscataway, NJ 08854-8058, United States

A R T I C L E   I N F O

Article history:
Received 16 December 2017
Received in revised form 27 February 2018
Accepted 27 February 2018
Available online 9 March 2018

Keywords:
Wrong-way driving crashes
Traffic safety
Association rules mining
Eclat algorithm

A B S T R A C T

Wrong-way driving (WWD) has been a constant traffic safety problem in certain types of roads. These crashes are mostly associated with fatal or severe injuries. This study aims to determine associations between various factors in the WWD crashes. Past studies on WWD crashes used either descriptive statistics or logistic regression to identify the impact of key contributing factors on frequency and/or severity of crashes. Machine learning and data mining approaches are resourceful in determining interesting and non-trivial patterns from complex datasets. This study employed association rules 'Eclat' algorithm to determine the interactions between different factors that result in WWD crashes. This study analyzed five years (2010–2014) of Louisiana WWD crash data to perform the analysis. A broad definition of WWD crashes (both freeway exit ramp WWD crashes and median crossover WWD crashes on low speed roadways) was used in this study. The results of this study confirmed that WWD fatalities are more likely to be associated with head-on collisions. Additionally, fatal WWD crashes tend to be involved with male drivers and off-peak hours. Driver impairment was found as a critical factor among the top twenty rules. Despite several other studies identifying only the WWD contributing factors, this study determined several influencing patterns in WWD crashes. The findings can provide an excellent opportunity for state departments of transportation (DOTs) and local agencies to develop safety strategies and engineering solutions to tackle the issues associated with WWD crashes.

Introduction

Wrong-way driving (WWD) crashes occur when a driver, intentionally or unintentionally, drives in the opposite direction of traffic flow. These crashes have a higher probability of fatal consequences (1.34 fatalities per fatal crash) compared to other types of crashes (1.10 fatalities per fatal crash) since, being likely head-on or opposite-direction sideswipe collisions.

Peer review under responsibility of Tongji University and Tongji University Press.

* Corresponding author.

E-mail addresses: s-das@tti.tamu.edu (S. Das), anandixd@tamu.edu (A. Dutta), mohammad.jalayer@rutgers.edu (M. Jalayer), a-bibeka@tti.tamu.edu (A. Bibeka), l-wu@tti.tamu.edu (L. Wu).

https://doi.org/10.1016/j.ijtst.2018.02.001
2046-0430/

According to the Federal Highway Administration (FHWA), approximately 300–400 people die each year due to WWD crashes in the U.S. (Federal Highway Administration, 2017). These values highlight the need for novel preventive approaches to mitigate the number of WWD crashes, which we address in this paper.

The National Transportation Safety Board (NTSB) defines that 'WWD is vehicular movement along a travel lane in a direction opposing the legal flow of traffic on high-speed divided highways or access ramps' (NTSB, 2012). This definition restricted WWD crashes only on controlled-access highways. This report has not included WWD crashes that result from median crossover encroachments. While the majority of the studies address WWD crashes on freeways, the crash analysis on median crossover WWD crashes has not been conducted in depth. This study aims to overcome this gap.

WWD crashes have been a subject of intense scrutiny over the past decades. Researchers have studied WWD crashes in different states in the U.S. including, Texas, Illinois, North Carolina, Michigan, and New Mexico (Braam, 2006; Finley et al., 2014; Lathrop et al., 2010; Morena and Leix, 2012; Zhou et al., 2015). Most of the previous studies used the descriptive statistics to explore the role of different factors associated with WWD crashes on high speed roadways and freeways. Pour-Rouholamin, Kemel, and Ponnaluri have employed other techniques such as Firth's penalized likelihood logistic regression, logistic regression and generalized order logit model to analyze the WWD crash data (Kemel, 2015; Ponnaluri, 2016; Pour-Rouholamin et al., 2016). Based on several previous studies (Braam, 2006; Cooner et al., 2004; Copelan, 1989; Lathrop et al., 2010; Morena and Leix, 2012; Scaramuzza and Cavegn, 2007; Zhou et al., 2015), deadly WWD crashes appear to involve intoxicated drivers. Moreover, these studies identified other significant confounding factors associated with WWD crashes such as driver age, driver gender, and time of day. Several previous studies (Morena and Leix, 2012, Zhou et al., 2015) also found the significant role of dark roadway conditions on the likelihood of WWD crashes. Studies using parametric techniques such as logistic regression and Firth's Penalized Likelihood Logistic Regression also explored that intoxicated drivers and darkness cause WWD crashes as significant factors (Kemel, 2015; Ponnaluri, 2016; Pour-Rouholamin et al., 2016). These studies are pivotal in coming up with countermeasures to reduce WWD crashes. However, these studies either used descriptive statistics alone or employed parametric methods to model the underlying relationships in the data. It should be noted that descriptive statistics might be insufficient to untangling complex relationships between different variables, whereas parametric models make assumptions about the distribution of the independent and dependent variables that might not always be true. In the recent years, machine learning and data mining methods have been widely used in transportation safety research to overcome the assumption issues associated with statistical modeling (Chen and Xie, 2015; Sun et al., 2014; Dong et al., 2015; Das et al., 2018b,c; Khan et al., 2015; Iranitalab and Khattak, 2017; Das and Sun, 2016, 2015). Machine learning models can detect interactions between different features and mask them if necessary. Data mining methods do not depend on any assumptions because the generated rules and patterns would show either interestingness or redundancy.

Two recent studies used multiple correspondence analysis (MCA) method (Jalayer et al., 2017; Das et al., 2018a) to explore WWD crashes. Compared to the parametric methods, MCA has a lower bias as this method does not make any prior assumptions about the variables. One of the limitations of this method is that it does not enable researchers to conduct significance test on the clusters. Moreover, this technique only informs about the correlation between variables but not about causation. To overcome the limitation of MCA at the same time retaining the low bias advantage of MCA, we analyzed five years (2010–2014) of Louisiana WWD crashes (for the remainder of this paper, Louisiana wrong way crashes will be referred as WWD crashes for consistency) using association rules. Frequent pattern mining (FPM) and association rules are powerful machine learning tools to identify the relationship between different factors. FPM recognizes the set of items that tend to occur together and association rules help understand the causal relationship between a set of antecedent items with the following item. The researchers are interested in finding long patterns as a multitude of factors can cause WWD crashes. Vertical data mining technique called "Eclat" performs better when long patterns are required thus the researchers used "Eclat" to analyze the data. This study aims to identify interdependence between various factors leading to WWD crashes.

Theory

According to Aggarwal and Han, frequent pattern mining (FPM) involves finding relationships among the items in a database (Aggarwal and Han, 2014). Several recent studies have employed association rules mining to determine significant associations for different safety problems (Das et al., 2018d; Das et al., 2017a,b; Das and Dutta, 2014; Das and Sun, 2014; Geurts et al., 2005; Mirabadi and Sharifian, 2010; Montella, 2011; Pande and Abdel-Aty, 2009).

FPM can be utilized to obtain association rules. A frequent pattern is obtained by comparing the support for a pattern with a certain minimum threshold. A pattern consists of various items in the dataset. Association rules mining uses three parameters to perform the analysis. These parameters are: support, confidence, and lift. The support of a pattern is determined by how frequently that pattern occurs in the dataset. The confidence of an association rule can be obtained by identifying the ratio of the number of times antecedent and consequent occur together by the number of time the antecedent occurs in the dataset. A third parameter called "lift" helps us determine the type of association between the antecedent and the consequent. Lift value higher than 1 indicates positive interdependence between the antecedent and the consequent, whereas a value less than 1 indicates a negative interdependence. These parameters can be expressed as:

$$S(X) = \frac{N_X}{N} \tag{1}$$

$$C(X \rightarrow Y) = \frac{S(X \rightarrow Y)}{S(X)} \tag{2}$$

$$L(X \rightarrow Y) = \frac{S(X \rightarrow Y)}{S(X).S(Y)} \tag{3}$$

where:
X = Antecedent
Y = Consequent
N_X = Number of incidents with pattern X
N = Total number of incidents
$S(X)$ = Support of pattern X
$S(X \rightarrow Y)$ = Support of frequent pattern formed by X and Y
$C(X \rightarrow Y)$ = Confidence of the association rule $(X \rightarrow Y)$
$L(X \rightarrow Y)$ = Lift of the association rule $(X \rightarrow Y)$

It can be computationally expensive or often impossible to explicitly enumerate all the patterns and then find if they are frequent patterns. Various algorithms have been developed to identify frequent patterns and determine the association rules efficiently. These algorithms are designed to work on the specific data format. Data can be either in horizontal layout or vertical layout (Maimon and Rokach, 2010). In the horizontal layout, each row contains the list of items that occur together whereas in the vertical layout the data comprises an item and a set of unique identifiers for the rows in which that item was present. 'Aprori' and 'Eclat' are the standard algorithms for dealing with horizontal and vertical layout data, respectively. Both of these algorithms exploit the monotonicity property of the support sets to eliminate patterns (Maimon and Rokach, 2010). Monotonicity property means that the support of an item set would always be less than the support of its subsets. Taking advantage of this feature, only those item sets, whose subsets are determined as a frequent pattern in the previous iteration, are evaluated. The main difference between the two algorithms occurs in the method used for counting support (Bernus et al., 1998). To identify the support for an item set in horizontal data layout, algorithm scans through all the rows whereas it can be directly determined for a vertical data by finding the intersection of the set of unique identifiers for the items in the item set. According to Bernus et al. (1998), vertical mining algorithms (Eclat) perform better when researchers are interested in finding long patterns in the data. The authors have thus used Eclat algorithm in this study. To make the Eclat algorithm more efficient, all the frequent patterns with two item sets are determined by using Apriori algorithm, and the rest are found by using Eclat algorithm.

In the above algorithm, frequent patterns with k items is used to determine frequent patterns with $k + 1$ items. It happens due to the monotonicity property. As stated above, superset of a pattern, will have less support than the pattern thus if a pattern is not a frequent pattern then definitely its superset is not a frequent pattern. Next, the algorithm finds the unique identifier for a pattern with $k + 1$ items by taking the intersection of two sets in S_k that are used to form this pattern that is they have k common items. Sets are arranged using lexicographic order thus can be compared with each other. After finding the set of frequent patterns, confidence is used to find the associative rules. This study has used two R packages 'arules', and 'aruleViz' to perform the analysis (Hahsler et al., 2005, Hahsler, 2017).

Data

To perform this study we used state maintained traffic crash database compiled from 2010 through 2014 in Louisiana. The primary dataset was prepared by merging information from three different databases (crash, roadway inventory, and vehicle). The final database contains 1419 WWD crashes. These crashes involved 2651 individuals. In this study, we analyzed the factors associated with WWD crashes. The variable importance was determined by using random forest algorithm. For the final analysis, we have used 16 variables with a set of 99 categories. Table 1 illustrates the descriptive statistics of different factors. Based on this table, the majority (54%) of the WWD crashes are head-on crashes. The vehicle damage data (front and front left contribute 64% of crashes) is also representative of the higher likelihood of head-on crashes. Other studies found similar finding like this (Pour-Rouholamin et al., 2016; FDOT, 2015). Around 50% of the crashes are fatal, disabling or evident or possible injury. A majority of WWD crashes occur in a business locality, which is representative of urban areas. Nighttime is a key contributing factor in WWD crashes. Other studies also showed similar finding (Copelan, 1989; Lathrop et al., 2010; Morena and Leix, 2012). As this study used both exit ramp WWD crashes and median crossover WWD crashes, two way roads show higher density in crash distribution. Around 18% of the crashes happened due to the median crossover (crashes involved with roadway locations with yellow no passing lane, and yellow dash line) encroachments. Louisiana WWD crashes do not show anomaly patterns in weather and alignments. Around 27% of Louisiana WWD crashes were exit ramp WWD crashes. In our study, a higher percentage of crashes is seen in lower speed roadways due to the inclusion of median crossover WWD crashes in the data analysis. Around 89% of WWD crashes occur on straight roads as compared to under 11% crashes on more treacherous alignments such as curves, grades, hillcrests, and dip.

The majority of drivers involved in these crashes are young adults. Males are more involved in WWD crashes (65%) as compared to females. Younger and older drivers are usually more prone towards WWD crashes. Louisiana crash data showed

Table 1
Descriptive statistics.

Category	Percentage	Category	Percentage
Violation		**Damage**	
Careless	24.5	Front	53.8
Dr. Cond.	11.0	Front Left	12.9
Disregard Traffic Control	10.0	Rear	10.3
Driving Left of Centre	5.5	None	4.5
Fail to Yield	3.7	Left	4.4
Improper Backing	2.4	Right	4.1
Improper Pass.	2.0	Total	1.7
Improper Turn	1.9	Under Carriage	1.1
No Violations	1.9	Top Carriage	0.0
Too Close	1.0	Other	4.4
Turned from Wrong Lane	0.8	Unknown	2.8
Speeding	0.4	**Alignment**	
Improper Parking	0.1	Straight	88.8
Other	32.6	Curve	6.6
Unknown	2.2	On Grade	2.6
Speed		Hillcrest	0.9
<20 mph	11.5	Dip	0.3
31–40 mph	26.0	Other	0.5
21–30 mph	24.7	Unknown	0.3
41–50 mph	18.2	**Collision**	
51–60 mph	12.6	Head-On	22.7
61–70 mph	7.0	Right Angle	20.1
Light		Sideswipe	16.2
Daylight	55.1	Single vehicle	11.8
Dark – Continuous Street Light (SL)	24.9	Left Turn	5.5
Dark – No SL	11.7	Rear End	4.8
Dark – Int. SL	4.8	Right Turn	3.9
Dusk/Dawn	2.5	Other	15.0
Other	1.0	**Severity**	
Driver Age (Dr_Age)		O	50.9
15–24	24.1	C	29.5
25–34	23.5	B	14.6
35–44	17.2	A	2.8
45–54	14.3	K	2.2
55–64	11.2	**Driver Condition (Dr_Cond)**	
65–74	6.0	Distracted	40.8
>75	3.7	Normal	22.1
Driver Gender (Dr_Gender)		Impaired	19.2
Male	64.6	Fatigued/Illness	3.2
Female	31.2	Other	3.0
Unknown	4.3	Unknown	11.7
Traffic Condition		**Traffic Control**	
Continuous Functioning (Cont. Func.)	79.8	White Dash Line	27.8
No Controls	16.7	Yellow No Passing Lane (PL)	11.2
Not Continuous Functioning	0.7	Sign	6.8
Other	1.4	Signal	6.7
Unknown	1.3	Yellow Dash Line	6.3
		Other	41.2
Road		**Locality**	
Two-Way-No Separation	37.8	Business	64.1
Two-Way-Separation	29.8	Residential	23.2
One-way	26.8	Other	12.7
Two-Way-Barrier	4.6	**Weather**	
Other	1.0	Clear	77.2
Hour (Hr)		Cloudy	14.0
Off-peak	66.4	Inclement	8.3
Peak	33.6	Other	0.5

similar trends like other studies showed for younger drivers (Finley et al., 2014; Xing, 2014; Zhou et al., 2015; FDOT, 2015; Ponnaluri, 2016) and older drivers (Scaramuzza and Cavegn, 2007; Pour-Rouholamin et al., 2016; Ponnaluri, 2016;). Moreover, the majority of crashes are caused by violation such as carelessness, driver condition (distracted, fatigued, normal etc.), disregarding traffic control and driving left of center. Distracted drivers showed higher trends in WWD crashes in Louisiana. Pour-Rouholamin et al. (2016) also associated inattention in older ages with higher number of WWD crashes. Table 1 also shows that male drivers were nearly twice in numbers when compared with female drivers. Similar findings are found in other studies (Lathrop et al., 2010; Morena and Leix, 2012; Finley et al., 2014; Ponnaluri, 2016). Driver error (driver condition and driver violation) contributed significantly to WWD crashes. Similar results were found in Caltrans (2015) and Scaramuzza and Cavegn (2007).

Results

Determining the optimum threshold of support and confidence is the key issue while using association rules mining algorithm. Lower values of support and confidence will generate a large number of variables, which would be very difficult to interpret due to the generated noise and overlapping. On the other hand, higher values could create a low number of rules with the less interesting pattern. To overcome this issue, this study used a simple convex optimization method to obtain optimum support and confidence values. The rules are generated for 2-itemset (minimum support = 0.01, minimum confidence = 0.80), 3-itemset (minimum support = 0.05, minimum confidence = 0.70), and 4-itemset (minimum support = 0.05, minimum confidence = 0.70). The rules with lift values greater than 1.0 indicate positive interdependence effects between antecedent and consequent. The number of rules (lift > 1) generated for different itemsets are as follows:

- 2-itemset: 214 rules
- 3-itemset: 204 rules
- 4-itemset: 2041 rules

Rules with high lift value indicate strong associations between the factors. Mining high-lift rules (in this case, top 20 rules in each itemset) can obtain highly associated characteristics in WWD crashes. Top twenty rules for three itemsets are listed in Tables 2–4. Table 2 presents the top twenty 2-itemset association rules based on the lift value. Some of the generated rules lack interestingness. For example, rule 2 (*Traffic Condition = No Controls → Traffic Control = Other*), and rule 16 (*Light = Dark – Cont. SL → Hr=Off-peak*) indicate redundant information. Following insights can be drawn from the significant rules:

- The significant factors for 2-itemset rules are two-way roadway with no physical separation, fatal crashes, male drivers, business locality, and nighttime crashes.
- The rule with highest lift is *Severity = K → Collision = Head-on* (support = 0.02, confidence = 0.88, lift = 3.89). It indicates that there is a strong interdependence between fatal crashes and head-on collision. The support value indicates that 2% of WWD crashes are fatal crashes due to a head-on collision. The confidence value indicates that out of all WWD fatal crashes, 88% were due to head-on collisions. The proportion of fatal head-on WWD crashes was 3.89 times the proportion of fatal WWD crashes in the complete dataset. Zhou et al. found that among multiple vehicles WWD, 45.6% were head-on crashes (Zhou et al., 2015).
- *Violation = Improper Passing → Light=Daylight* because drivers might tend to be more complacent during daytime thus undertaking risky maneuvers.

Table 2
Top 20 two item association rule.

Rules	Antecedent	Consequent	Support (%)	Confidence (%)	Lift
1	Severity = K	Collision = Head-On	2	88.1	3.888
2	Violation = Improper Passing	Road = Two-Way-No Separation	1.8	90.7	2.401
3	Traffic Control = Yellow Dash Line	Road = Two-Way-No Separation.	5.7	90.4	2.391
4	Traffic Condition = No Controls	Traffic Control = Other	15.3	91.4	2.22
5	Traffic Control = Yellow No PL	Road = Two-Way-No Separation	9.4	83.2	2.202
6	Violation = Improper Backing	Severity = O	2.2	93.7	1.839
7	Violation = Improper Passing	Light = Daylight	1.8	88.9	1.612
8	Severity = K	Damage = Front	1.8	81.4	1.514
9	Dr_Age=> 75	Light = Daylight	3	81.6	1.48
10	Hr = Peak	Light = Daylight	27.1	80.6	1.462
11	Traffic Control = Signal	Locality = Business	6.1	91	1.419
12	Collision = Right Turn	Locality = Business	3.5	89.4	1.395
13	Light = Dark - No SL	Hr = Off-peak	10.5	90.3	1.361
14	Severity = K	Hr = Off-peak	2	89.8	1.354
15	Speed = 31–40 mph	Locality = Business	22.3	85.8	1.338
16	Light = Dark - Cont. SL	Hr = Off-peak	22.1	88.8	1.338
17	Light = Dark - Int. SL	Hr = Off-peak	4.2	87.5	1.319
18	Violation = Improper Turn.	Locality = Business	1.6	84.3	1.316
19	Collision = Right Turn	Dr_Gender = Male	3.3	84.6	1.31
20	Severity = K	Dr_Gender = Male	1.8	83.1	1.286

Table 3
Top 20 three item association rule.

Rules	Antecedent	Consequent	Support (%)	Confidence (%)	Lift
1	Dr_Gender=Male Violation = Dr. Cond.	Dr_Cond. = Impaired	6.9	82.8	4.321
2	Damage = Front Violation = Dr. Cond.	Dr_Cond. = Impaired	5.9	81.3	4.24
3	Traffic Condition = Cont. Func. Violation = Dr. Cond.	Dr_Cond. = Impaired	7.2	80.5	4.201
4	Hr = Off-peak Violation = Dr. Cond.	Dr_Cond. = Impaired	7.1	80.4	4.197
5	Weather = Clear Violation = Dr. Cond.	Dr_Cond. = Impaired	7.0	77.5	4.044
6	Align = Straight Violation = Dr. Cond.	Dr_Cond. = Impaired	7.0	76.4	3.989
7	Access = No Control Violation = Dr. Cond.	Dr_Cond. = Impaired	5.5	71.6	3.735
8	Traffic Condition = Cont. Func. Speed = 61–70 mph	Traffic Control = White Dash Line	5.0	74.7	2.688
9	Traffic Condition = Cont. Func. Speed = 61–70 mph	Road = Two-Way-Separation	5.1	76.4	2.564
10	Road = Two-Way-Separation Dr_Cond. = Impaired	Traffic Control = White Dash Line	5.2	68.8	2.476
11	Access = No Control Traffic Control = Yellow Dash Line	Road = Two-Way-No Separation	5.5	90.2	2.386
12	Align = Straight Traffic Control = Yellow Dash Line	Road = Two-Way-No Separation	5.4	90.0	2.381
13	Traffic Control = Yellow Dash Line Traffic Condition = Cont. Func.	Road = Two-Way-No Separation	5.2	89.5	2.369
14	Collision = Head-On Road = Two-Way-Separation	Traffic Control = White Dash Line	5.5	64.5	2.319
15	Traffic Condition = No Controls Violation = Other	Traffic Control = Other	5.4	95.4	2.315
16	Light = Daylight Traffic Control = Yellow No PL	Road = Two-Way-No Separation	5.8	87.1	2.304
17	Hr = Off-peak Traffic Condition = No Controls	Traffic Control = Other	9.9	93.9	2.280
18	Weather = Clear Traffic Condition = No Controls	Traffic Control = Other	12.6	93.6	2.272
19	Traffic Condition = No Controls Speed = 21–30 mph	Traffic Control = Other	5.5	93.5	2.271
20	Access = No Control Traffic Control = Yellow No PL	Road = Two-Way-No Separation	8.8	85.7	2.266

- WWD crashes are mostly seen in business areas (rule 11, 12, 15, and 18). These crashes are mostly on signalized intersections. The crashes happened due to turn related maneuvers.
- Two-way roadways with no physical separation were found as the dominating consequent in the top five rules. The rule *Violation = Improper Passing → Roadway = Two-lane Roadways with no Physical Separation* has a lift value of 2.40. It indicates that the proportion of improper passing related WWD crashes on two-lane roadways with no physical separation was 2.40 times the proportion of all improper passing related WWD crashes in the complete dataset. These rules also indicate that pavement marking like yellow lines are not sufficient enough in reducing median crossover encroachment related WWD crashes. Physical separation might be considered as a possible countermeasure which requires long-term planning and justifications for the policymakers to proceed due to higher expense.
- Male drivers are more likely to be associated with fatalities and right turn collisions (rule 19, and rule 20). Zhou et al., 2015 also found a higher number of males to be associated with WWD crashes.
- Off-peak hours are obviously associated with nighttime crashes with different lighting condition. The lift of nighttime with no lighting condition is higher than nighttime crashes with lighting (Morena and Leix, 2012; Scaramuzza and Cavegn, 2007; Xing, 2014).

Table 3 presents the three-itemset association rule. Following are the critical observations from this table:

- There are seven rules with *Driver Condition = Impaired* as the consequent. All the rules with *Driver Condition = Impaired* consequent contain *Violation = Driver Condition* antecedent. This is because impaired driving is the subset of driving condition. These rules have lift value greater than 3, which indicates that there are strong interdependences between the antecedents and the consequent.

Table 4
Top 20 four item associaiton rule.

Rules	Antecedent	Consequent	Support (%)	Confidence (%)	Lift
1	Damage = Front Traffic Condition = Cont. Func. Violation = Dr. Cond.	Dr_Cond. = Impaired	5	86.9	4.536
2	Hr = Off-peak Dr_Gender=Male Violation = Dr. Cond.	Dr_Cond. = Impaired	5.8	86.5	4.515
3	Weather = Clear Dr_Gender=Male Violation = Dr. Cond.	Dr_Cond. = Impaired	5.8	85.5	4.461
4	Dr_Gender=Male Traffic Condition = Cont. Func. Violation = Dr. Cond.	Dr_Cond. = Impaired	5.9	84.9	4.429
5	Hr = Off-peak Traffic Condition = Cont. Func. Violation = Dr. Cond.	Dr_Cond. = Impaired	5.8	84.7	4.42
6	Align = Straight Dr_Gender=Male Violation = Dr. Cond.	Dr_Cond. = Impaired	5.9	83.9	4.377
7	Hr = Off-peak Weather = Clear Violation = Dr. Cond.	Dr_Cond. = Impaired	6	81.5	4.255
8	Hr = Off-peak Align = Straight Violation = Dr. Cond.	Dr_Cond. = Impaired	5.8	81.5	4.252
9	Weather = Clear Traffic Condition = Cont. Func. Violation = Dr. Cond.	Dr_Cond. = Impaired	6	80.8	4.217
10	Align = Straight Traffic Condition = Cont. Func. Violation = Dr. Cond.	Dr_Cond. = Impaired	5.9	79.7	4.159
11	Align = Straight Weather = Clear Violation = Dr. Cond.	Dr_Cond. = Impaired	5.8	75.7	3.953
12	Hr = Off-peak Locality = Business Dr_Cond. = Impaired	Light = Dark - Cont. SL	5.8	70.2	2.819
13	Road = Two-Way-Separation. Dr_Cond. = Impaired Traffic Condition = Cont. Func.	Traffic Control = White Dash Line	5.1	75	2.698
14	Collision = Head-On Road = Two-Way-Separation Traffic Condition = Cont. Func.	Traffic Control = White Dash Line	5.5	70	2.518
15	Access = No Control Align = Straight Traffic Control = Yellow Dash Line	Road = Two-Way-No Separation	5.3	89.8	2.376
16	Access = No Control Traffic Control = Yellow Dash Line Traffic Condition = Cont. Func.	Road = Two-Way-No Separation.	5.1	89.3	2.363
17	Light = Daylight Traffic Control = Yellow No PL Traffic Condition = Cont. Func.	Road = Two-Way-No Separation.	5.5	88.4	2.339
18	Access = No Control Light = Daylight Traffic Control = Yellow No PL	Road = Two-Way-No Separation	5.7	88.3	2.336
19	Severity = O Weather = Clear Traffic Condition = No Controls	Traffic Control = Other	7.6	95.7	2.324
20	Hr = Off-peak Weather = Clear Traffic Condition = No Controls	Traffic Control = Other	8.3	95.7	2.322

- The rule with highest lift is *Driver Gender = Male, Violation = Driver Condition → Driver Condition = Impaired* (support = 0.07, confidence = 0.82, lift = 4.32). The support value indicates that 7% of all WWD crashes involved with impaired male drivers. The confidence value indicates that out of all crashes due to the conditions of the male drivers, 82% were due to the impaired driving condition. The proportion of impaired male driver involved crash was 4.32 times the proportion of all crashes due to the driving condition of the male drivers.

- There are eight rules that include pavement marking related countermeasures. The pavement markings are white dashed line, yellow dashed line, and yellow no passing lane marking. These crashes are median crossover encroachment related WWD crashes. It indicated that pavements marking related countermeasures are not sufficient for reducing WWD crashes. Finley et al. (2014) noted that effectiveness of pavement marking related countermeasures for lowering WWD crashes is not apparent when impaired drivers are involved.
- The rules generated for 3-itemset also indicates that impaired driving is a critical contributor to WWD crashes. Many of the previous studies found impaired drivers or intoxicated to be important factors leading to WWD crashes (Statewide Wrong Way Crash Study, 2015; Cooner et al., 2004; Copelan, 1989; Friebele et al., 1971; Ponnaluri, 2016; Scaramuzza and Cavegn, 2007).
- Roadways with higher posted speed limits are likely to be involved in WWD crashes. Possible countermeasures include improved signage and traffic control devices at locations with higher posted speed limits. Finley et al. (2014) showed that WWD detection systems show promise in providing alerts to drivers.

Fig. 1 illustrates a balloon-plot to describe the association between the representative of grouped antecedents and consequents of all the 2041 rules generated for 4-itemset. The reddish-gray indicates the group of association rules with high lift values while the light gray of balloon implies that the group of association rules has small lift values. Similarly, the large-sized balloon represents the group of association rules with high support values while the small-sized balloon represents the group of association rules with low support values. As it is difficult to get a clear picture of 2041 rules, this graphic gives a quick snapshot of the overall trend. The domination factors are impaired driving, roadways with no physical separation, male drivers, and inadequate countermeasures like pavement markings.

Table 4 presents the 4-itemset association rules. Following are the important observations from this table:

- The factor *Dr_Cond. = Impaired* is found as the consequent in 11 of the rules. All of these rules show a strong interdependence between the antecedent and consequent.
- Male drivers are more likely to be associated with impairment.
- The two-way roadway with no physical separation was found as the dominating factor in the 4-itemset rules.
- The 4-itemset rules also indicate that the traffic control countermeasures like white dashed line, yellow dashed line, and yellow no passing lane are not adequate in reducing WWD crashes. Innovative signing and pavement marking (low mounted wrong way sign, reflective wrong way arrows, pavement arrows, and flashing wrong way indicators) are the potential solutions to reduce these crashes.

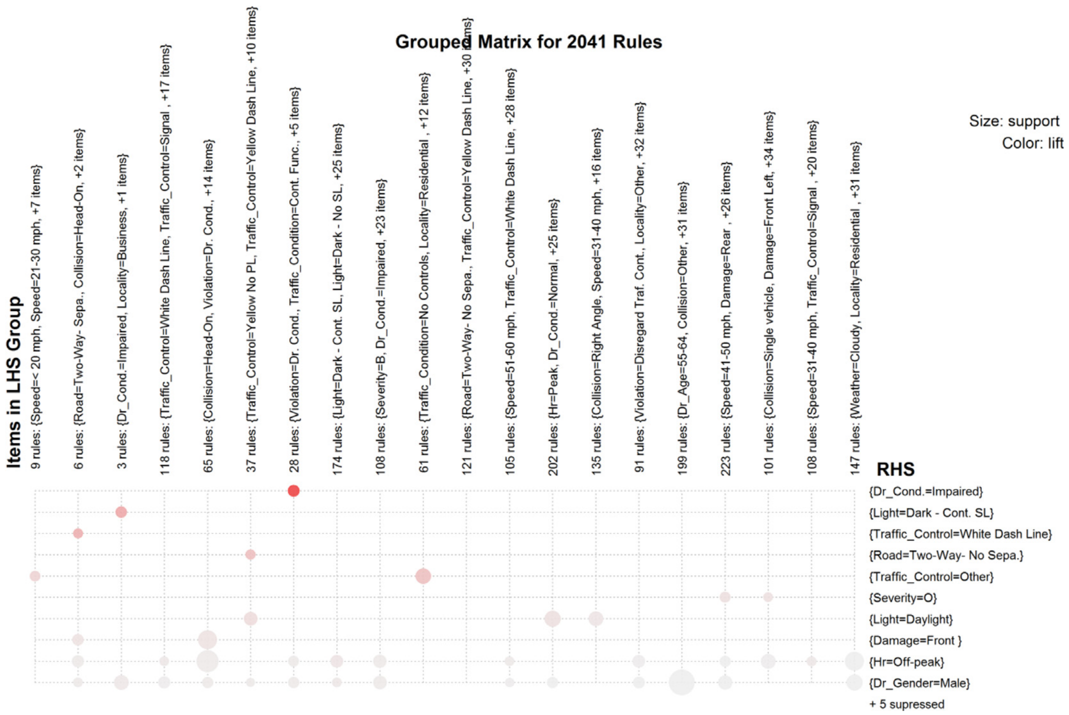

Fig. 1. Balloon plot on 2041 rules generated for 4-itemset rules.

Discussions

In recent years, many studies attempted to identify the key factors that influence the frequency and severity of WWD crashes for the development of effective safety-related countermeasures. However, the number of WWD crashes is still high. It shows that, in addition to statistical analysis, research needs to be conducted with new techniques and in further directions. In parametric models, there are usually inherent assumptions that explanatory variables should be independent whereas many factors have the high correlation in traffic safety analysis in reality. Data mining can overcome this issue by generating the interestingness and uniqueness of the rules. The obvious and redundant rules are thus not required to be examined due to insignificance.

Hence, the focus of policy implication from the parametric models and association rule method could be entirely different. In this paper, a data mining approach (association rules Eclat algorithm) was proposed to investigate WWD crash characteristics and contributory factors. Mining association rule between sets of crash characteristics can obtain essential characteristics and contributory factors reasonably in WWD crashes. By using different values of support and confidence, it is possible to get enough information of the combination of crash characteristics to analyze the potential causes of wrong-way crashes. The rules generated in this study determined several contributing factor groups: two-lane undivided roadways, exit ramps, head-on crashes, male drivers, impaired driving, improper and inadequate pavement markings, inadequate signs and alert systems, and nighttime crashes. Majority of the findings are in line with prior studies. The key difference of this study is the inclusion of median crossover WWD crashes. Majority of these crashes happened on rural two lane undivided roadways. Proper signings and improvement in pavement marking are essential in reducing such crashes. One benefit of the current study is that it provides likelihood values (in the form of lift parameter) for different scenarios, which would be beneficial to safety engineers and policymakers in performing decision making on the countermeasure selection.

The study has a few limitations. The current effort has not utilized any sophisticated optimization techniques to determine the optimized values of the parameters (support and confidence). Future studies can incorporate optimization technique like ant-colony optimization or genetic algorithm to determine the streamlined values of the parameters. In addition, this study provided analysis for top twenty rules for each itemset (2-item, 3-item, and 4-item). A more rigorous discussion on an additional number of rules might provide more intuitive insights, which is currently out of the scope of this study.

Acknowledgements

The authors are grateful to two reviewers for providing important comments and suggestions that enhanced the original version of the paper.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ijtst.2018.02.001.

References

Aggarwal, C.C., Han, J., 2014. Frequent Pattern Mining. Springer.
Bernus, P., Schmidt, G., Shaw, M., 1998. International Handbooks on Information Systems. Springer.
Braam, A., 2006. Wrong Way Crashes: Statewide Study of Wrong Way Crashes on Freeways in North Carolina. North Carolina Department of Transportation, Division of Highways.
Caltrans Division of Research, 2015. Innovation and System Information. Wrong-Way Driving Prevention Methods.
Chen, C., Xie, Y., 2015. Machine learning for recognizing driving patterns of drivers of large commercial trucks. Transp. Res. Rec.: J. Transp. Res. Board 2517, 18–27.
Cooner, S.A., Cothron, A.S., Ranft, S.E., 2004. Countermeasures for Wrong-Way Movement on Freeways: Overview of Project Activities and Findings. Texas Transportation Institute, Texas A & M University System. FHWA/TX-04/4128-1.
Copelan, J., 1989. Prevention of Wrong Way Accidents on Freeways. California Department of Transportation, California, USA. FHWA/CA-TE-89-2.
Das, S., Avelar, R., Dixon, K., Sun, X., 2018a. Investigation on the wrong way driving crash patterns using multiple correspondence analysis. Anal. Accid. Anal. Prev. 111, 43–55.
Das, S., Dutta, A., Dixon, K., Minjares-Kyle, L., Gillette, G., 2018b. Using deep learning in severity analysis of at-fault motorcycle rider crashes. Transp. Res. Rec.: J. Transp. Res. Board.
Das, S., Mudgal, A., Dutta, A., Geedipally, S., 2018c. Vehicle consumer complaint reports involving severe incidents: mining large contingency tables. Transp. Res. Rec.: J. Transp. Res. Board.
Das, S., Dutta, A., Avelar, R., Dixon, K., Sun, X., Jalayer, M.A., Dutta, R., 2018d. Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures. Int. J. Urban Sci., 1–19
Das, S., Minjares-Kyle, L., Avelar, R.E., Dixon, K., Bommanayakanahalli, B., 2017a. Improper passing related crashes on rural roadways: using association rules negative binomial miner. Paper presented at: Transportation Research Board 96th Annual Meeting, Washington D.C.
Das, S., Minjares-Kyle, L., Avelar, R.E., Dixon, K., Bommanayakanahalli, B., 2017b. Improper passing related crashes on rural roadways: using association rules negative binomial miner. Paper presented at: Transportation Research Board 96th Annual Meeting, Washington D.C.
Das, S., Sun, X., 2016. Association knowledge for fatal run-off-road crashes by multiple correspondence analysis. IATSS Res. 39 (2), 146–155.
Das, S., Sun, X., 2015. Factor association with multiple correspondence analysis in vehicle-pedestrian crashes. Transp. Res. Rec.: J. Transp. Res. Board 2519, 95–103.
Das, S., Dutta, A., 2014. Association rules mining on the ridership data of a bike sharing program. Paper presented at: Proceedings in 55th Annual Transportation Research Forum, San Jose, CA.

Das, S., Sun, X., 2014. Investigating the pattern of traffic crashes under rainy weather by association rules in data mining. Paper presented at: Transportation Research Board 93rd Annual Meeting, Washington D.C.
Dong, N., Huang, H., Zheng, L., 2015. Support vector machine in crash prediction at the level of traffic analysis zones: Assessing the spatial proximity effects. Accid. Anal. Prev. 82, 192–198.
Florida Department of Transportation, 2015. Statewide Wrong Way Crash Study. FDOT Financial Project ID: 190258-1-32-37.
Federal Highway Administration. Wrong Way Driving. Available at: https://safety.fhwa.dot.gov/intersection/other_topics/wwd/. Accessed February 1.
Finley, M.D., Venglar, S.P., Iragavarapu, V., Miles, J.D., Park, E.S., Cooner, S.A., Ranft, S.E., 2014. Assessment of the Effectiveness of Wrong Way Driving Countermeasures and Mitigation Methods. Texas A & M Transportation Institute. FHWA/TX-15/0-6769-1.
Friebele, J.D., Messer, C.J., Dudek, C.L., 1971. State-of-the-Art of Wrong Way Driving on Freeways and Expressways. TxDOT. 139-7.
Hahsler, M., Grun, B., Hornik, K., 2005. arules – a computational environment for mining association rules and frequent item sets. J. Statist. Softw. 14 (15).
Hahsler, M., 2017. arulesViz: interactive visualization of association rules with R. The R Journal 9 (2).
Geurts, K., Thomas, I., Wets, G., 2005. Understanding spatial concentrations of road accidents using frequent item sets. Accid. Anal. Prev. 37 (4), 787–799.
Jalayer, M., Pour-Rouholamin, M., Zhou, H., 2017. Wrong-way driving crashes: a multiple correspondence approach to identify contributing factors. Traffic Injury Prevent. (just-accepted)
Iranitalab, A., Khattak, A., 2017. Comparison of four statistical and machine learning methods for crash severity prediction. Accid. Anal. Prev. 108, 27–36.
Khan, G., Bill, A., Noyce, N., 2015. Exploring the feasibility of classification trees versus ordinal discrete choice models for analyzing crash severity. Transp. Res. Part C: Emerg. Technol. 50, 86–96.
Kemel, E., 2015. Wrong-way driving crashes on french divided roads. Accid. Anal. Prev. 75, 69–76.
Lathrop, S.L., Dick, T.B., Nolte, K.B., 2010. Fatal wrong-way collisions on New Mexico's interstate highways, 1990–2004. J. Forensic Sci. 55 (2), 432–437.
Maimon, O., Rokach, L., 2010. Data Mining and Knowledge Discovery Handbook. Springer, US.
Mirabadi, A., Sharifian, S., 2010. Application of Association Rules in Iranian Railways (RAI) Accident Data Analysis. Saf. Sci. 48 (10), 1427–1435.
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.
Morena, D.A., Leix, T.J., 2012. Where These Drivers Went Wrong. Federal Highway Administration. FHWA-HRT-12-004.
National Transportation Safety Board (NTSB), 2012. Wrong Way Driving. https://www.ntsb.gov/safety/safety-studies/Documents/SIR1201.pdf. Accessed February, 2017.
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. 47 (1), 145–154.
Ponnaluri, R.V., 2016. The odds of wrong-way crashes and resulting fatalities: a comprehensive analysis. Accid. Anal. Prev. 88, 105–116.
Pour-Rouholamin, M., Zhou, H., Zhang, B., Turochy, R.E., 2016. Comprehensive analysis of wrong-way driving crashes on Alabama Interstates. Transp. Res. Rec.: J. Transp. Res. Board 2601, 50–58.
Scaramuzza G and Cavegn M. Wrong-Way Drivers: Extent-Interventions. Paper presented at: The European Transport Conference; Leeuwenhorst, Netherlands.
Sun, J., Sun, J., Chen, P., 2014. Use of support vector machine models for real-time prediction of crash risk on urban expressways. Transp. Res. Rec.: J. Transp. Res. Board 2432, 91–98.
Statewide Wrong Way Crash Study, 2015. Florida Department of Transportation. 190258-1-32-37.
Xing, J., 2014. Characteristics of wrong-way driving on motorways in Japan. IET Intel. Transport Syst. 9 (1), 3–11.
Zhou, H., Zhao, J., Pour-Rouholamin, M., Tobias, P.A., 2015. Statistical characteristics of wrong-way driving crashes on Illinois freeways. Traffic Inj. Prev. 16 (8), 760–767.