Accident Analysis and Prevention 111 (2018) 43–55

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Full length article

Investigation on the wrong way driving crash patterns using multiple correspondence analysis

Subasish Das[a,*], Raul Avelar[a], Karen Dixon[a], Xiaoduan Sun[b]

[a] Texas A&M Transportation Institute (TTI), 3135 TAMU, College Station, TX 77843-3135, United States
[b] Department of Civil Engineering, University of Louisiana at Lafayette, Lafayette, LA 70504, United States

ARTICLE INFO

Keywords:
Wrong way driving crashes
Multiple correspondence analysis
Dimensionality reduction
Contributing factors

ABSTRACT

Wrong way driving (WWD) has been a constant traffic safety problem in certain types of roads. Although these crashes are not large in numbers, the outcomes are usually fatalities or severe injuries. Past studies on WWD crashes used either descriptive statistics or logistic regression to determine the impact of key contributing factors. In conventional statistics, failure to control the impact of all contributing variables on the probability of WWD crashes generates bias due to the rareness of these types of crashes. Distribution free methods, such as multiple correspondence analysis (MCA), overcome this issue, as there is no need of prior assumptions. This study used five years (2010–2014) of WWD crashes in Louisiana to determine the key associations between the contribution factors by using MCA. The findings showed that MCA helps in presenting a proximity map of the variable categories in a low dimensional plane. The outcomes of this study are sixteen significant clusters that include variable categories like determined several key factors like different locality types, roadways at dark with no lighting at night, roadways with no physical separations, roadways with higher posted speed, roadways with inadequate signage and markings, and older drivers. This study contains safety recommendations on targeted countermeasures to avoid different associated scenarios in WWD crashes. The findings will be helpful to the authorities to implement appropriate countermeasures.

1. Introduction

Wrong way driving (WWD) crashes on different roadways are considered as constant traffic safety problems. Although wrong way crashes are not large in numbers, the outcomes of these crashes tend to involve disproportionally higher number of fatalities or serious injuries. According to Pour-Rouholamin and Zhou (2016), “WWD crashes happen when a driver, inadvertently or deliberately, drives against the main direction of traffic flow on a controlled-access highway”. A study conducted by Friebele et al. (1971) mentioned that “the wrong-way driver, travelling head-on into an unsuspecting traffic stream, is simply a time bomb ticking off the seconds toward a possible disaster”. Pour-Rouholamin et al. (2014) found 1.34 fatalities per fatal WWD crashes in the U.S. from 2004 to 2013, while for other crashes the fatalities per fatal crash rate is 1.10 during the same time period. According to National Highway Traffic Safety Administration (NHTSA) statistics, around 350 people are killed each year nationwide due to WWD crashes (NHTSA, 2013). In Louisiana, around 300 WWD crashes (0.2% of total crashes) happened every year. Around 0.45% of total crashes in Louisiana are fatal crashes, but for wrong way crashes this percentage is higher (around 1.6% of the total WWD crashes). Thus, it is crucial to identify key risk factors associated with WWD crashes.

The Federal Highway Administration (FHWA) Highway Safety Improvement Program (HSIP) includes a project to monitor WWD crashes and identify hot spots of WWD crashes. It includes a wrong-way study warrant based on total crash and fatal crash rates. The National Transportation Safety Board (NTSB) recommends that the FHWA develop a HSIP policy memorandum for use by state department of transportation agencies to establish wrong-way monitoring programs (NTSB, 2012). The outcomes of the monitoring programs can help in developing improved signage and marking as well as technology like wrong way navigation alerts on vehicles. For an effective monitoring program, determining key association factors in WWD crashes would be particularly helpful.

One of the major tasks in highway safety analysis is the identification of the key contributing factors for different types of crashes. Multiple Correspondence Analysis (MCA) is a dimensionality reduction method useful to describing the significance of co-occurrence of groups of variables or variable categories from a high dimension dataset. This method is also referred to as the pattern recognition method that treats

* Corresponding author.
E-mail addresses: s-das@tti.tamu.edu (S. Das), r-avelar@tti.tamu.edu (R. Avelar), k-dixon@tti.tamu.edu (K. Dixon), xsun@louisiana.edu (X. Sun).

https://doi.org/10.1016/j.aap.2017.11.016
Received 21 November 2016; Received in revised form 3 November 2017; Accepted 13 November 2017

arbitrary data sets as combination of points in a larger dimensional space. It uniquely simplifies complex data into knowledge extraction in a completely different way than parametric estimation does. In MCA analysis, the objective is to investigate associations between multiple variables, as opposed to the more traditional characterization of associations between a set of predictor variables and a single response variable of interest (i.e., number of crashes).

The study team used five years (2010–2014) of Louisiana WWD (for the remainder of this paper Louisiana wrong way crashes, both driving and cycling, will be referred as WWD crashes for consistency) crashes to determine the relationship of the variables and their significance. The objectives of this study are: (1) to identify the relative closeness of the key association factors to determine meaningful co-occurrence, and (2) to recommend countermeasures when appropriate. The findings of this study could help authorities to determine effective and efficient crash countermeasures.

2. Literature review

Although traffic safety research includes an extensive array of research areas, the most prominent are- crash frequency analysis, and crash severity analysis. Lord and Mannering (2010) provided a detailed overview of the properties of crash-frequency data and associated methodological alternatives and limitations for examining such data. Savolainena et al. (2011) provided a similar assessment on crash-severity analysis. Recently, Mannering and Bhat (2014) bridged and extended the previous studies of Lord and Mannering (2010) and Savolainena et al. (2011) by overviewing both count data models and crash severity models. Interested readers can consult these studies as well as a hyperlinked webpage developed by Das (2016) for further information. That webpage lists 592 research papers on statistical and algorithmic methods as well as hyperlinks to all corresponding papers.

The literature review reveals a surge of research on WWD crashes since 2014. Table 1 described the research efforts conducted on WWD crashes starting from 1971. Most of the studies used freeway as the main interest group. Few studies focused on all roadways or divided roadways (Ponnaluri, 2016; Kemel, 2015). Many studies performed descriptive statistics to describe the nature of the factors in WWD crashes (Friebele et al., 1971; Copelan, 1989; Cooner et al., 2004; Braam, 2006; Scaramuzza and Cavegn, 2007; SWOV, 2009; Morena and Leix, 2012; Finley et al., 2014; Xing, 2014; Zhou et al., 2015; FDOT, 2015). In many cases, simple descriptive statistics should not suffice to explaining the impact of the contributing factors. It is also important to note that some of these studies focused more on operational considerations than safety (Friebele et al., 1971; Copelan, 1989; Cooner et al., 2004; Braam, 2006; Finley et al., 2014). Several authors explored the analysis of crash outcomes and crash types using a modeling approach. Some studies simply used logistic regression models to differentiate between WWD and non-WWD crashes (Kemel, 2015; Ponnaluri, 2016). As WWD crashes are very small in numbers compared to non-WWD crashes, this small sample size problem is likely to significantly influence the outcomes and statistical power of the models. Pour-Rouholamin et al. (2014) used Firth's penalized-likelihood logistic regression to control the influence of all confounding variables on the probability of WWD crashes while considering the rareness of the WWD event. Pour-Rouholam and Zhou (2016) used generalized ordered logistic regression to perform crash severity analysis using WWD crashes.

The idea of MCA begins in 1970 with French Statistician Jean-Paul Benźecri (Roux and Rouanet, 2010), though there are similarities with Principal Component Analysis (PCA) and Factor Anaysis (FA), two well documented multivariate statistical methods. PCA mainly deals with numerical data, and MCA is a well-accustomed tool for multidimensional categorical data.

MCA has been reinvented many times under different frameworks while keeping the goals similar (De Leeuw, 1973; Hoffman and De Leeuw, 1992). A limited number of studies has been conducted in applying MCA in the transportation safety research. Hoffman and De Leeuw (1992) interpreted MCA as multidimensional scaling method and associated different vehicle models with crash severities. Fontaine (1995) performed MCA on one year of pedestrian crash data to determine the statistical proximity of the significant factors. This study identified few distinctive groups as a basis for more in depth analysis. Factor et al. (2010) applied MCA in determining the association between driver's social characteristics and their involvements in crash severities. This study exposed new facets in the social organization of fatalities. Das and Sun (2015) used eight years (2004–2011) of pedestrian crash data in Louisiana to determine key associations between risk factors. This study determined several significant groups of factors that require deeper exploration in future. Xu et al. (2016) used quasi-indiced exposure method to identify the key factors contributed to pedestrian crashes in Las Vegas from 2004 to 2008. This study later used MCA to determine the interaction between different factors. Das and Sun (2016) applied MCA on eight years (2004–2011) of fatal run-of-road (ROR) crashes in Louisiana to examine the degree of association between risk factors. Das et al. (2017) recently applied MCA on the second Strategic Highway Research Program's (SHRP 2) Washington Roadway Inventory Database (RID) to identify the key association factors for inclement weather crashes. The finding revealed some specific factor groups that require careful attention from the safety professionals.

Table 2 shows variables used in previous studies addressing wrong-way driving crashes using different methods. These past studies will be used to inform the exploratory analysis presented in the following sections.

3. Theory of multiple correspondence analysis

MCA is an unsupervised learning algorithm. In MCA, one does not need to distinguish between explanatory variables and the response variable. It requires the construction of a matrix based on pairwise cross-tabulation of each variable. For example, the dimension of the final dataset of this study is: 1203 × 24. For a table of qualitative or categorical variables with dimension 1203 × 24, MCA can be explained by taking an individual record (in row), *i [i = 1 to 1203],* where 24 categorical variables (represented by 24 columns) have different sizes of categories. MCA can generate the spatial distribution of the points by different dimensions based on these 24 variables.

Let P be the number of variables (i.e., columns) and I is the number of transactions (i.e., rows). This will generate a matrix of *'I multiplied by P'*. If L_p is the number of categories for variable p, the total number of categories for all variables is, $L = \sum_{p=1}^{P} L_p$. It will generate another matrix *'I multiplied by L'*. In this matrix, each of the variables will contain several columns to show all of their possible categorical values.

The cloud of categories is considered as a weighted combination of J points. Category j is represented by a point denoted by C^j with weight of n_j. For each of the variables, the sum of the weights of category points is n. In this way, for the whole set J the sum is nP. The relative weight w_j for point C^j is $w_j = n_j/(nP) = f_j/P$. The sum of the relative weights of category points is 1/P, which makes the sum of the whole set as 1.

$$w_j = \frac{n_j}{nP} = \frac{f_j}{P} \quad \text{with} \quad \sum_{j \in J_q} w_j = \frac{1}{P} \quad \text{and} \quad \sum_{j \in J} w_j = 1$$

Here, $n_{jj'}$ represents the number of individual records which have both categories k and k'. The squared distance between two categories C^j and $C^{j'}$ can be represented by

$$\left(C^j C^{j'}\right)^2 = \frac{n_j + n_{j'} - 2n_{jj'}}{n_j n_{j'}/n} \tag{4}$$

The numerator of Eq. (4) is the number of individual records associating with either j or j' but not both. For two different variables, p and p', the denominator is the familiar "theoretical frequency" for the cell (j, j') of the $J_p \times J_{p'}$ two-way table.