Accident Analysis and Prevention 132 (2019) 105242

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A clustering approach to developing car-to-two-wheeler test scenarios for the assessment of Automated Emergency Braking in China using in-depth Chinese crash data

Bo Sui[a,c,*], Nils Lubbe[b], Jonas Bärgman[c]

[a] Autoliv China, Beihe Highway 1000, 201 807 Shanghai, China
[b] Autoliv Research, Wallentinsvägen 22, 447 83 Vårgårda, Sweden
[c] Chalmers University of Technology, Chalmersplatsen 4, 412 96 Gothenburg, Sweden

ARTICLE INFO

Keywords:
Road safety
AEB
Motorcycle
Bicycle
C-NCAP
ADAS

ABSTRACT

Two-wheeled vehicles (motorized and non-motorized, referred to as TWs) are an important part of the transport system in China. They also represent an important challenge for road safety, with many TW user fatalities and injuries every year. Recently, active safety systems for cars, such as Automated Emergency Braking (AEB), promise to reduce road traffic fatalities and injuries. For these systems to work effectively, it is necessary to understand and define the complex traffic scenarios to be addressed. The aim of this study is to contribute to the development of test procedures for AEB specifically, drawing on the China In-Depth Accident Study (CIDAS) data from July 2011 to February 2016 to describe typical scenarios for crashes between cars and TWs by means of cluster analysis. In total, 672 car-to-TW crashes were extracted. The data was clustered according to five main crash characteristics: time of crash, view obstruction, pre-crash driving behavior of the car driver and the TW driver, and relative moving direction. The analysis resulted in six car-to-TW crash scenarios typical of China. In three scenarios the car and the TW travel perpendicularly to each other before the crash, in two they travel in the same direction, and in one they travel in opposite directions. Further, each scenario can be described with three characteristics (the road speed limit, the TW's first contact point on the car, and the car's first contact point on the TW) that can be included in an AEB test suite.

Some scenarios were similar to those in the Euro New Car Assessment Program (Euro NCAP). For example, in one, a TW moving straight ahead was hit by a car moving perpendicularly, and in the other the car hit a TW traveling in the same direction. Both occurred in daytime, without a visual obstruction. However, in contrast to the Euro NCAP, typical scenarios in China included night-time scenarios, scenarios where the car or the TW was turning, and those in which the TW was hidden from the car by an obstruction. The results contribute to a proposed novel AEB test suite with realistic scenarios specific to China.

1. Introduction

1.1. Background

Two-wheeled vehicles, including traditional pedal bicycles and powered two-wheelers (PTWs: combustion and electric engines motorcycles), have long been popular as a transportation mode in China, especially in congested traffic in big cities, due to their low cost and easy operability (Cherry, 2007). An estimated 400 million traditional pedal bicycles were in use in China at the end of 2011 (Nie and Yang, 2014). By the end of 2017, there were approximately 170 million combustion engine motorcycles in China (National Bureau of Statistics of China, 2017). In addition, the use of the more environmentally friendly electric PTWs is rapidly increasing. Their number in China was estimated to be over 200 million at the end of 2017 (Ministry of Industry and Information Technology of the People's Republic of China, 2018).

Although bicycles and electric PTWs have large environmental advantages, their growing number is a traffic safety concern. In this paper, all types of two-wheeled vehicles will be referred to as TWs. According to national reporting in China, TWs were the transport modality with the largest proportion of fatalities, accounting for 35% of the 58,539 road traffic deaths in China in 2013 (compared with in-vehicle occupants accounting for 28% and pedestrians, 25%; WHO, 2015). Crashes

* Corresponding author at: Autoliv China, Beihe Highway 1000, 201 807 Shanghai, China.
E-mail address: bo.sui@autoliv.com (B. Sui).

https://doi.org/10.1016/j.aap.2019.07.018
Received 25 February 2019; Received in revised form 21 June 2019; Accepted 20 July 2019
Available online 22 August 2019
0001-4575/

and fatalities are largely under-reported in China, as the differences between national reporting and estimates from WHO demonstrate: in 2016, there were 63,093 road traffic fatalities according to national reporting but 256,180 according to WHO's estimate based on hospital data (WHO 2018b). Thus, the nationally reported numbers are best interpreted as indicating increased TW safety concerns, rather than presenting accurate quantification (Bhalla et al., 2014; National Bureau of Statistics of China, 2016; Duan et al., 2017; WHO, 2018a).

Meanwhile, the ownership of motorized vehicles overall has also increased rapidly in China due to the country's fast modernization and economic growth. By the end of 2017, the number of registered motorized vehicles reached 300 million: passenger cars accounted for 180 million, registered motorcycles 82 million, trucks 23 million, and buses 2.3 million (Ministry of Public Security of the People's Republic of China, 2018). The increasing number of passenger cars is a safety threat to TWs. Although detailed statistics on crash characteristics for PTWs are not available for China, it can be noted that in Germany, PTW casualties predominantly occur in crashes with other motorized traffic participants (70%). Most (80%) are with a passenger car (DESTATIS, 2016). For bicyclists, passenger cars are also a large safety threat. In Germany, 80% of bicycle crashes involve other traffic participants; of these, 75% involve a car (DESTATIS, 2016). Some data from China is available for bicyclist fatalities; in fact, Duan et al. (2017) showed that car-bicycle crashes have become a primary contributor to traffic fatalities in China.

The safety threat passenger cars pose to TWs can be addressed by passive and active safety. To protect TWs, passive safety systems in passenger cars are considered to have requirements and solutions similar to those already in use to protect pedestrians. For example, the development of the pedestrian head form test area (the area on the car where head impacts are most likely) has led to countermeasures (e.g., deployable hood and windshield airbags) which also protect TW riders to some extent, although they were initially intended to protect pedestrians (Fredriksson and Rosén, 2012; Zander and Hamacher, 2017). However, the existing test procedures, requirements, and countermeasures might not be adequate for TW riders. For example, the current pedestrian head form test areas may not extend rearward far enough to be effective for TWs; required modifications may include a wrap-around distance (WAD) of 2500 mm, with a of head-impact angle of 70° and an impact speed of 40 km/h (Zander and Hamacher, 2017). In fact, Fredriksson et al. (2015) showed that countermeasures based on cars' hood and windshield areas, designed to mitigate pedestrian injuries, could also protect bicyclists if frame parts higher up on the windshield and A-pillars were considered. These findings could be applied to other types of TWs in China, such as electrical PTWs and motorcycles. A study in China found that head protection for drivers of electrical PTWs needs to cover a WAD of up to 2260 mm to address 80% of real-world impacts (Wu et al., 2018).

Active safety systems for passenger cars have also been regarded as promising measures for improving traffic safety. One such system-Automated Emergency Braking (AEB)-is increasingly available on new vehicles. It applies braking automatically to lower the impact speed, avoiding a crash or mitigating its severity when it is unavoidable.

Computer simulations based on real-world crashes demonstrate AEB's considerable potential for saving lives and mitigating severe injuries in frontal car-to-pedestrian and car-to-bicyclist collisions. Lindman et al. (2010) estimated that AEB can reduce pedestrian fatalities by 24%. Using crash data from Germany, Rosén (2013) showed that AEB reduced the percentage of fatally (severely) injured pedestrians in frontal collisions with cars by 48% (42%). The same study showed that the system is similarly effective at protecting bicyclists, reducing fatalities by 55% and severe injuries by 33% (Rosén, 2013).

Test methods and assessment programs for enhancing vehicle safety have been developed and implemented for AEBs in response to consumer assessment tests. For example, Euro NCAP introduced procedures to test and assess the effectiveness of AEBs in terms of pedestrian safety in 2016, and in terms of cyclists in 2018 (Euro NCAP, 2018a). An assessment of AEBs which considers motorcycles is planned for 2020 (Euro NCAP, 2018b). Similarly, the China New Car Assessment Program (C-NCAP) has included an AEB assessment for pedestrians since 2018 (C-NCAP, 2018), and it is likely that assessment procedures for TW protection will be introduced soon (Sui et al., 2017).

Before these test methods and assessment programs can be introduced, a thorough understanding of the complex traffic situation is necessary to obtain the appropriate scenarios to address, to ensure the safety benefits of these systems in the real world (Lenard et al., 2011; Sander and Lubbe, 2018). In several previous studies, the most common scenarios targeted in an assessment were derived from descriptive statistics based on real-world crash data. For example, the European project CATS developed test scenarios for AEB for cyclists, based on descriptive statistics from real-world crashes in Europe (Uittenbogaard et al., 2016a, b). In the APSECSS project, test scenarios for forward-looking integrated systems detecting pedestrian were derived with a similar approach (Wisch et al., 2013). A reliable estimate of real-world AEB performance in China requires a detailed understanding of car-to-TW crash scenarios based on Chinese real-world crash data. Some work has been done on identifying typical car-to-TW scenarios in China. For example, a study from Sui et al. (2017) investigated car-to-TW scenarios in China based on relative moving direction of the car and TW before the crash, and pre-crash driving behavior of the car and TW using descriptive statistics analysis. However, the heterogeneity of real-world traffic data arguably leads to potentially incorrect conclusions when only pure descriptive statistics are used to identify typical crash scenarios, due to the hidden relations between factors (de Ona et al., 2013; Sasidharan et al., 2015; Nilsson et al., 2018). There are other methods, for example clustering, which can alleviate such issues. Cao et al. (2019) found four typical car-to-TW scenarios using hierarchical clustering, but the study focused on a limited set of fatal and severe crashes.

1.2. Clustering

Cluster analysis, or clustering, is an unsupervised data-driven analysis method complementing descriptive statistics in identifying and characterizing typical crash scenarios (Depaire et al., 2008; de Ona et al., 2013; Nitsche et al., 2017). Clustering is the grouping or segmenting of a collection of objects into subsets or "clusters", such that those objects within each cluster are more closely related than objects assigned to different clusters (Hastie et al., 2009).

Generally, two types of clustering methods are available: distance- or similarity-based clustering and model-based clustering. For the former, a matrix containing the distance or dissimilarity between the observations is defined and a clustering algorithm is applied. The choice of distance measures and clustering algorithms depends on types of data and research questions. Typically, Euclidean or Manhattan distance metrics are used for continuous numeric data, Jaccard similarity (coefficient) or Hamming distance for categorical variables, and Gower's generalized coefficient of similarity for mixed-type data (Gower, 1971; Vijay et al., 2012; Sander and Lubbe, 2018).

K-means clustering is the most widely used distance-based clustering algorithm because of its simplicity and efficiency at clustering large data sets (Huang, 1998; Aggarwal and Reddy, 2013). It consists of two main steps: build and iteration. In the build step, K representative points are randomly chosen as initial centroids; the algorithm then assigns each observation to its closest centroid based on the distance matrix built between the observation and the centroid. In the iteration step, the algorithm repeats these steps until the centroid becomes stable or some other alternative criterion is met-for example, the specified maximum number of iterations is reached (Aggarwal and Reddy, 2013).

However, this process cannot be applied to categorical data, since Euclidean distances and means are not meaningful in a traditional sense (Huang, 1998; Sharma and Gaud, 2015). For categorical variable clustering, several extensions of K-means clustering are available which

are based on distance or similarity. These include K-mode (Huang, 1998) and K-medoid methods (Kaufman and Rousseeuw, 1990). Of these two, K-medoid was chosen in this study because K-mode, contrary to K-medoid, produces locally optimal results which depend heavily on the initial mode selection and the order of objects in the dataset, thus increasing the risk of suboptimal results (Aggarwal and Reddy, 2013). Instead of using the centroid as in K-means clustering, K-medoid uses an available data point, medoid, as the center of each cluster (Nitsche et al., 2017). A common algorithm of the K-medoid clustering is Partitioning Around Medoid (PAM) (Aggarwal and Reddy, 2013). During the build step, PAM chooses one representative object for each cluster, referred to as a medoid, and each observation is assigned to the closest medoid based on the corresponding coefficient in a dissimilarity matrix. In the iteration step, each object is tested as a potential medoid by checking if the sum of within-cluster distances gets smaller. If so, the object is defined as a new medoid. The algorithm ends when medoids are no longer being changed (Nitsche et al., 2017). The final medoid is the data point, at the center of a cluster, whose average dissimilarity to all the objects in the cluster is minimal (Sander and Lubbe, 2018).

Seman et al. (2013) compared performance in clustering categorical data between K-mode and K-medoid, and K-medoid showed higher clustering accuracy on nine common categorical datasets. K-medoid is also commonly used in traffic safety analysis (Saunier et al., 2011; Nitsche et al., 2017; Sander and Lubbe, 2018). For an introduction to these distance- or similarity-based clustering methods; see Aggarwal and Reddy (2013).

In addition to the simplicity and efficiency, another major advantage of distance-based clustering is that the clustering quality can continue to improve through the iteration step (Aggarwal and Reddy, 2013). However, if K is not properly defined, the possibility of a suboptimal clustering result increases with successive iterations as well. Therefore, measures to assess the clustering quality are necessary to ensure optimal results.

In model-based clustering, probabilistic models are used to explain the distributions in the data. It is typically assumed that the data are from a mixture of underlying probability distributions (Fraley and Raftery, 1998; Vermunt and Magidson, 2002). In other words, model-based clustering assumes that objects match a model, which is often a statistical distribution. One example of a model-based clustering method that has been applied to traffic safety is Latent Class Clustering (LCC). It was used in a study examining fatal, serious road crashes involving young New Zealand drivers (Weiss et al., 2015), and in another which investigated injury severity in pedestrian crashes in Switzerland (Sasidharan et al., 2015).

Since model-based clustering algorithms cluster data which match a defined distribution, there is a risk of inaccurate clustering results produced from inaccurate user assumptions about the underlying distributions of the data (Aggarwal and Reddy, 2013). Another disadvantage of model-based clustering is low computational efficiency, so processing large datasets is very time-consuming (Aggarwal and Reddy, 2013). To compare clustering on the real data: Tribal Art dataset without prior knowledge of the truth, K-medoid performed better than LLC in the study by Anderlucci and Hennig (2014), resulting in the choice of K-medoid for this study.

1.3. Study aim

The aim of this study is to contribute to the development of AEB by describing typical car-to-TW crash scenarios in China, for subsequent use in AEB test procedures. Cluster analysis of crash characteristics based on the China In-Depth Accident Study (CIDAS) data has been carried out. The results can provide policymakers and consumer assessment programs (e.g., C-NCAP) with realistic test scenarios involving TWs in order to evaluate the real-life performance of AEB in China. This study specifically addresses a knowledge gap about car-to-TW scenarios by taking data heterogeneity and all injury severities into account (through clustering) and provide a minimal, but sufficient set of scenarios for use in AEB assessment, complementing recent work done by Cao et al. (2019).

2. Method

2.1. Data

This study used the CIDAS database, one of the most detailed crash databases available in China. It contains crash data collected in six cities: Changchun, Beijing, Weihai, Ningbo, Chengdu, and Foshan (Chen et al., 2014). Approximately 600 crashes are collected annually; in each crash, at least one person has been injured, one four-wheeled vehicle is involved, and the crash scene is well preserved until the investigation team arrives. The CIDAS database contains 31 data tables and 2800 data items.

There were 1470 TW crashes in the CIDAS database from July 2011 to February 2016, including all levels of injury severity and all types of impacts. When only crashes with M1 vehicles (vehicles used for the carriage of passengers and comprising not more than eight seats in addition to the driver's seat) (CATARC, 1994), were considered, there were 1087 cases. Excluding crashes in which a TW impacted a four-wheeled vehicle from the rear, those lacking complete case photos, and those with TW users less than twelve years old resulted in 830 cases being identified as potential candidates for this study. A further 158 cases were excluded due to missing data, yielding 672 cases in the final analysis.

2.2. Clustering method

Several aspects must be considered when choosing the most appropriate clustering algorithm (Nitsche et al., 2017). In this study, the K-medoid clustering method based on the Hamming distance matrix was considered most suitable, mainly because it can deal with categorical variables (as explained previously) and is robust to outliers (Aggarwal and Reddy, 2013; Nitsche et al., 2017; see also Introduction for method comparisons). We also tried K-mode clustering in this study, but K-medoid showed better clustering quality (see Discussion). The common K-medoid clustering algorithm, PAM, was applied, since it has the advantage of obtaining better clustering accuracy with categorical data than other K-medoid algorithms (Nitsche et al., 2017; Sander and Lubbe, 2018).

To avoid the possibility of suboptimal clustering results produced by the K-medoid algorithm, the Average Silhouette Width (ASW) was calculated to assess the clustering quality. The ASW measures how well the observations are clustered and estimates the average distance between clusters. Generally, the higher the ASW value is, the better the clustering quality is. The silhouette plot, typically used in ASW analysis, shows how close each point in one cluster is to points in the neighboring clusters (Rousseeuw, 1987). The ASW can also indicate the optimal number of clusters for distance- or similarity-based clustering methods, since the optimal number of clusters results in the highest ASW (Sander and Lubbe, 2018).

2.3. Clustering variables selection

Test scenarios must accurately represent real-world crashes as well as being practical, in terms of the simulation and hardware used in the assessment procedures (Zander and Hamacher, 2017). The car-to-TW scenarios in this study were defined by the characteristics of the environment and the road users involved. The scenarios were characterized by five variables describing the environmental conditions and kinematics of the car and TW before the crash. Four were identified directly from the CIDAS database: TIME, time of crash (TZEIT in CIDAS), OBSTRUCT, visual obstruction (SICHTBV in CIDAS), INTENT_Car, pre-crash driving behavior of the car driver (ABSICHT of car

Table 1
Clustering variable specifications.

Variable name	Description	Reason for clustering
TIME	Time of crash	Night-time scenarios require the car's cameras and sensors to have better detection performance compared to those in daytime. Further, Euro NCAP has included this variable in its assessment of AEB performance for pedestrians.
OBSTRUCT	Existence of vision obstruction, hiding TW from the car driver.	Sensors have difficulty detecting TWs when they are hidden behind an obstruction; cooperative communication-based technology (V2X) is not (yet) common in AEB.
INTENT_Car	Pre-crash driving behavior of the car	It is an important factor that describes the general kinematics of the car before the crash.
INTENT_TW	Pre-crash driving behavior of the TW	It is an important factor that describes the general kinematics of the TW before the crash.
RELMOT	Relative moving direction between the car and the TW before the crash	While car and TW's intents describe the individual road user's kinematics, information about their relative motion is needed to describe test scenarios.

in CIDAS), and INTENT_TW, pre-crash driving behavior of TW driver (ABSICHT of TW in CIDAS). The fifth variable, RELMOT, the relative moving direction between the car and the TW before the crash, was derived through a case-by-case study. Following the precedent of Sui et al. (2017), the variable was based on: the variable UTYP in CIDAS database, the crash description, driver interviews, and CAD sketches. All variable names and descriptions, together with the reason for choosing each one as a clustering variable, are shown in Table 1. The clustering variables were chosen based on the intended use of the results - selecting the smallest number of variables that would adequately describe the NCAP test scenarios (Nitsche et al., 2017; Sander and Lubbe, 2018). Medoid scenarios, created from clustering algorithm, were combinations of these five crash characteristics. They are subsequently recommended as NCAP testing scenarios.

2.4. Correlation analysis of clustering variables

Correlation between clustering variables would introduce noise into the data, which could influence the results of clustering (Ben-Hur and Guyon, 2003). Cramer's V statistic was used to measure the strength of association between pairs of two categorical variables. Values closer to 1 showed strong association while values closer to 0 show weak or no association.

2.5. Crash characteristics analysis

Because of their relevance to the test scenario settings in the practical implementation of C-NCAP, three variables were further analyzed: the road speed limit, the TW's first contact point on the car, and the car's first contact point on the TW (see Table 2). These variables, important for a complete description of a test setup, were not included in the main clustering as they are reasonably easy to manipulate in a full-scale (C-NCAP) test, within the respective clustering scenarios. Furthermore, these three variables substantially affect the severity of the crash and the AEB algorithm timing design. Consequently, instead of including them in the clustering, they were included in a subsequent crash characteristics analysis.

Table 3
Cramer's V results.

Cramer's V	TIME	OBSTRUCT	INTENT_Car	INTENT_TW	RELMOT
TIME	1	0.0553	0.1251	0.0776	0.1942
OBSTRUCT		1	0.0496	0.0941	0.1805
INTENT_Car			1	0.1526	0.2668
INTENT_TW				1	0.1259
RELMOT					1

3. Results

3.1. Correlation analysis results

The results of Cramer's V correlation statistics are shown in Table 3. Weak associations were found across all variables, indicating that the associations probably did not affect clustering results substantially.

3.2. Clustering

Fig. 1 shows the cluster validity metric ASW, which was calculated across a range of three to fifteen clusters (red line). In addition, the minimum number of samples in each cluster for different numbers of clusters is also shown in the figure (blue line).

The minimum number of samples in each cluster decreased drastically to less than 40 for seven or more clusters. The ASW value for six clusters was as high as it is for seven and eight clusters, increasing only again for nine and more clusters. Choosing six clusters seemed to be the best compromise between high ASWs with small sample sizes in the clusters and low ASWs with relatively large sample sizes. Furthermore, silhouette coefficient plots (see Figs. A1–A3 in Appendix A; Rousseeuw, 1987; Nitsche et al., 2017) showed which observations have negative silhouette coefficients; fewer observations were placed in the wrong cluster for six clusters than for five or seven.

3.3. Clustering results

Frequency distributions of the clustering variables are listed in Table 4. Each medoid scenario, the combination of typical crash characteristics generated from each clustering, was identified by the gray cells in Table 4.

Table 2
Crash characteristics variable specifications.

Variable name	Interpretation	Categories
Vlimit	Speed limit of the road the car is traveling on	≤40 km/h, 50-60 km/h, ≥70 km/h
OVERLAP	TW's first contact point on the car	Front: 0-25%, Front: 25%-50%, Front: 50%-75%, Front: 75%-100%, left side_front, left side_middle, left side_rear, right. side_front, right. side_middle, right side_rear.[a]
FirstCont2W	Car's first contact point on the TW	Left or right side from front: 0-25%, left or right side from front: 25%-50%, left or right side from front: 50%-75%, left or right side from front: 75%-100%, left or right side from front: 0-50%, left or right side from front: 50%-100%.

[a] In addition to the categories, the contact point was also provided as a percentage of the car's width or length (depending on region of impact), relative to the front left corner of the car, and the front of the car, respectively.

Fig. 1. ASW values and minimum number of samples for different number of clusters.

For example, the medoid scenario from Cluster 1 was characterized as a daytime crash without a visual obstruction and a car moving straight ahead impacted a TW approaching perpendicularly to the car.

For the medoid scenarios of Clusters 1, 4, and 6, the typical crash characteristics were a car moving straight ahead impacting a TW moving straight ahead perpendicularly (see Table 4 and Table 6). Differences between these three medoid scenario were found in the crash times and the visual obstruction. Cluster 1's medoid scenario occurred in daytime without a visual obstruction, Cluster 4's occurred at night-time without a visual obstruction; Cluster 6's medoid scenario, in the daytime with a visual obstruction, was the only one with a visual obstruction.

For the medoid scenarios of Clusters 2 and 3, typical crash characteristics were the car and the TW driving in the same direction longitudinally in daytime, without a visual obstruction. The separation between medoid scenarios of Clusters 2 and 3 was that the former was characterized by the car turning right before the crash. In the medoid scenario of Cluster 3, more crashes occurred when the TW turned left.

In the medoid scenario of Cluster 5, the car was turning left with the TW going straight in the opposite direction, at night-time, without a visual obstruction.

3.4. Comparison across the medoid scenarios and the top ten most frequent scenarios

Table 5 shows a comparison between medoid scenarios and the top ten most frequent scenarios (where No. 1 is the most frequent) created from all possible combinations of the five selected variables. These ten represented 68% of the whole sample. Of all the frequent scenarios, only 4, 6, 8, and 9, highlighted in gray, were not also medoid scenarios (using six clusters). These four scenarios were merged into other clusters. For example, 4 and 6, with the car turning left, were mostly merged into the scenarios with the car driving straight (Cluster 2 and Cluster 1, respectively); the only difference was the pre-crash driving behavior of the car driver.

3.5. Crash characteristics analysis

The following figures show the distributions of crash characteristics variables (road speed limit, car and TW impact points) for each cluster. Note that this analysis only included crashes with the medoid scenario characteristics of the cluster; cases in a cluster with at least one variable not equal to the medoid scenario were omitted.

3.5.1. Road speed limit

Fig. 2 shows that a speed limit between 50–60 km/h is common for most medoid scenarios, except in medoid scenario of Cluster 6 with a visual obstruction, in which speed limits at or under 40 km/h were more common.

3.5.2. TW's first contact point on the car

To describe the contact area with precision, the front part of the car was divided into four areas, starting from the driver's side: 0-25%, 25%-50%, 50%-75%, and 75%-100% (see Fig. 3). Each side of the vehicle was divided into three areas: left side-front, left side-middle, and left side-rear on the left, and right side-front, right side-middle, and right side-rear on the right. Further, we presented the contact point also as a percentage of the car's width or length (depending on region of impact), relative to the front left corner of the car, and the front of the car, respectively.

As it is likely that most AEBs would perform equally well for TWs coming from the right or the left (see, for example, Scenarios 1, 4, and 6 in Table 6), symmetry was also assumed when describing the test scenarios. For example, medoid scenario of Cluster 1, with the TW coming from the right side and the car impacting the TW with a 75%–100% overlap, was symmetrical to the TW coming from the left side and the car impacting the TW with a 0–25% overlap. This symmetry was considered when characterizing the overlap in the final scenarios. The grouping schema is shown in Appendix A Table A1. Further, the mode of the distribution of contact points within the most frequent sector (considering symmetry in Fig. 4) became the recommended impact point in the C-NCAP tests.

3.6. Car's first contact point on the TW

In CIDAS, the possible contact areas on the TW were divided into several sectors as shown in

Fig. 5. Unfortunately, this information was only coded for 269 out of the 672 cases. Results were therefore only indicative for scenario construction (see Table 6).

3.6.1. Summary of results

Detailed typical car-to-TW scenarios, combining the clustering and

Table 4
Frequency of clustering variables in each cluster: the five gray cells in each cluster (row) are the typical crash characteristics that form the medoid for that cluster.

Variable	Category	Cluster 1 (n = 160, 24%)	Cluster 2 (n = 142, 21%)	Cluster 3 (n = 113, 17%)	Cluster 4 (n = 105, 16%)	Cluster 5 (n = 81, 12%)	Cluster 6 (n = 72, 11%)
TIME	Day	160	129	98	0	30	59
	Night	0	13	15	105	51	13
OBSTRUCT	Yes	0	6	0	0	4	72
	No	160	136	113	105	77	0
RELMOT	Perpendicular	147	28	19	73	0	59
	Longitudinal same	0	108	71	16	15	3
	Longitudinal opposite	13	6	23	16	66	10
INTENT_Car	Go straight	129	0	102	86	0	43
	Turn left	31	37	7	11	77	16
	Turn right	0	102	4	8	2	13
	others	0	3	0	0	2	0
INTENT_TW	Go straight	156	140	25	98	79	59
	Turn left	0	0	85	4	1	8
	Turn right	4	2	2	3	1	4
	Others	0	0	1	0	0	1

Table 5
Comparison between top ten scenarios and medoid scenario. The rows in grey mean that this frequent scenario was not a medoid scenario (using six clusters), instead those were merged into other clusters.

Top 10 scenario	TIME	OBSTRUCT	RELMOT	INTENT_Car	INTENT_TW	Frequency	Cluster
1	Day	No	Perpendicular	Go straight	Go straight	111	Cluster 1
2	Night	No	Perpendicular	Go straight	Go straight	49	Cluster 4
3	Day	No	Longi. same	Turn right	Go straight	49	Cluster 2
4	Day	No	Longi. same	Turn left	Go straight	37	Cluster 2
5	Day	No	Longi. same	Go straight	turn left	32	Cluster 3
6	Day	No	Perpendicular	Turn left	Go straight	31	Cluster 1
7	Night	No	Longi. opposite	Turn left	Go straight	30	Cluster 5
8	Day	No	Longi. opposite	Turn left	Go straight	29	Cluster 5
9	Day	No	Perpendicular	Turn right	Go straight	28	Cluster 2
10	Day	Yes	Perpendicular	Go straight	Go straight	27	Cluster 6

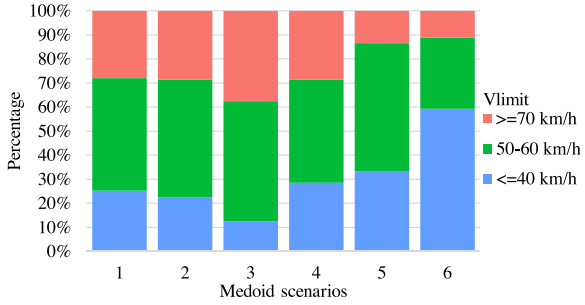


Fig. 2. Distribution of road speed limit in each medoid scenario group.

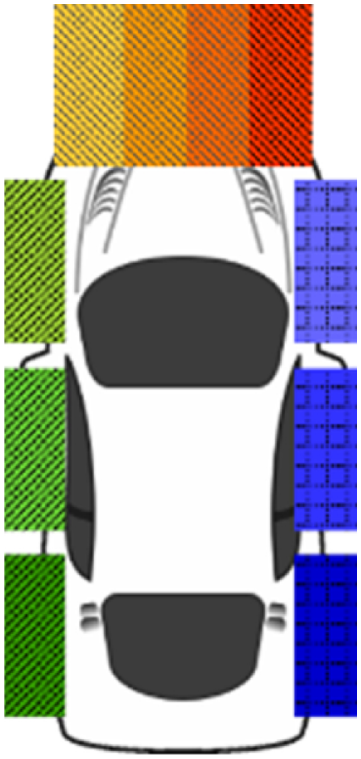


Fig. 3. Contact areas on the car.

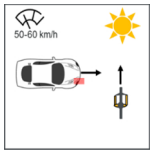
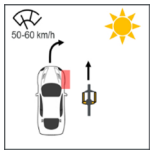
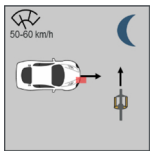
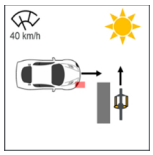
crash characteristics analyses, are presented in Table 6. Both pre-crash kinematics and environmental factors were included.

4. Discussions

This study identified and described typical car-to-TW crash scenarios in China by applying an unsupervised data-driven clustering method to data from an in-depth crash database. Six crash scenarios were identified and described. The choice of variables to include in the clustering was guided by the aim to provide test scenarios for the assessment of AEB, targeting car-to-TW crashes in China.

There are three common types of two-wheeled vehicles on the roads in China: bicycles, electric PTWs and motorcycles, and it is not obvious whether they should be studied separately or together (Sui et al., 2017). However, previous work has shown no statistically significant

Table 6
Typical scenario descriptions.

Scenario	Description	Illustrative pictograms
1	A car moving straight ahead impacting a TW moving straight ahead from a perpendicular direction in daytime without a visual obstruction. The road speed limit is 50-60 km/h. The contact point is at 88% of the car's front, measured from the driver's side. The first contact point on the TW is the first quarter of its front.	
2	A car turning right impacting a TW moving straight ahead in the same direction in daytime without a visual obstruction. The road speed limit is 50-60 km/h. The contact point is at 2% of car's length on the right side, measured from the front. The first contact point on the TW is the first quarter of its front.	
3	A car moving straight ahead impacting a left-turning TW previously moving in the same direction in daytime without a visual obstruction. The road speed limit is 50-60 km/h. The contact point is at 85% of the car's front, measured from the driver's side. The first contact point on the TW is on its rear half.	
4	A car moving straight ahead impacting a TW moving straight ahead from a perpendicular direction at night without a visual obstruction. The road speed limit is 50-60 km/h. The contact point is at 81% of the car's front, measured from the driver's side. The first contact point on the TW is on its front half.	
5	A left-turning car impacting a TW moving straight ahead from the opposite direction at night without a visual obstruction. The road speed limit is 50-60 km/h. The contact point is at 84% of the car's front, measured from the driver's side. The first contact point on the TW is on its front half.	
6	A car moving straight ahead impacting a TW moving straight ahead from a perpendicular direction at night with a visual obstruction. The road speed limit is 40 km/h. The contact point is the at 89% of the car front, measured from driver's side. The first contact point on the TW is on its front half.	

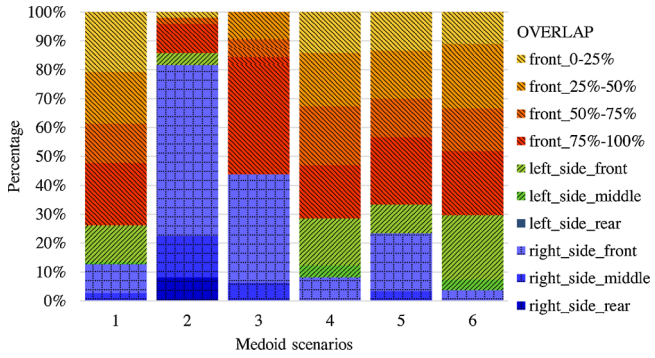


Fig. 4. Distribution of TW's first contact area on the car.

Distributions of the point on the car where the TW first made contact are shown in Fig. 4.

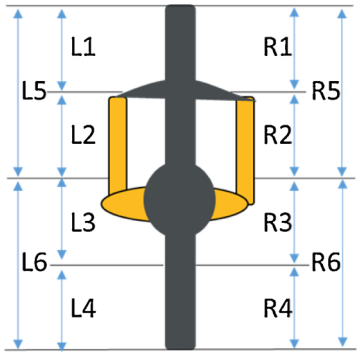


Fig. 5. TW's first contact area sectors.

differences among them in terms of injury severity or head-impact locations on the car (Wu et al., 2018). In this study, the different types were grouped together under the umbrella term TW. However, we recommend that a consumer assessment program like C-NCAP use a range of TW speeds in its test scenarios, to cover the range in speeds for all TWs. Information about the difference (e.g., in speed) between TW types studied can be found in Appendix A, Figs. A4–A6.

Applying clustering techniques to categorical data requires making some preliminary important choices. First, an appropriate clustering algorithm must be selected among several, such as K-modes and PAM (Sharma and Gaud, 2015; Nitsche et al., 2017). In our study, PAM showed better, more stable model performance with higher ASW values than did K-modes (see Table 7).

In addition, the variables to be clustered need to be chosen: they need to be relevant to the aim of the study. In this study, in addition to three basic variables indicating pre-crash kinematics, two more were added for the clustering: night or day (time of crash), and with or without visual obstruction. These two were added for two reasons. The first was that the AEB-pedestrian test scenarios in Euro NCAP already included night-time and visual obstruction testing scenarios. The second was that, in previous research on the safety of vulnerable road users (VRUs), higher-severity injuries have been identified in crashes occurring in darkness (Rosén, 2013; Chen et al., 2014; MacAlister, and Zuby, 2015; Uittenbogaard et al., 2016a). Collecting data on these two variables is an important aspect of AEB evaluation, both in China (e.g., C-NCAP) and elsewhere, because both darkness and obstruction of the car driver's vision represent great challenges to the in-car sensors and as a result, may influence AEB performance substantially.

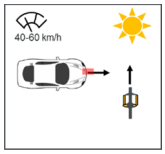
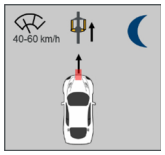
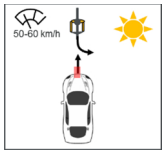
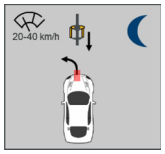
Table 7
ASW values comparison between K-modes and PAM.

Number of clusters (K)	ASW of K-modes	ASW of PAM
4	0.39	0.34
5	0.36	0.39
6	0.28	0.41
7	0.37	0.41
8	0.37	0.41
9	0.37	0.46
10	0.36	0.5
11	0.37	0.53
12	0.36	0.55

The five clustering variables alone were not enough to fully define an AEB test scenario and ensured a valid test scenario assessment. Thus, information about the car speed (here, the posted road speed limit is used, as its actual speed is not available) and the TW's impact point on the car were added to describe the distributions of the variables within each cluster. The impact speed and impact point can then be varied for AEB testing in a test suite. Both these variables are highly relevant in real-world crash scenarios, as well as being practical for testing.

In the study from Cao et al. (2019), four typical car-to-TW scenarios were obtained from clustering based on another crash database in China, NAIS (National Automobile Accident In-depth Investigation System). Table 8 summarized the configurations of the four Cao et al. (2019) scenarios. Comparing Cao et al. (2019) results with this study, two scenarios were common: a car moving straight ahead impacting a cyclist moving straight ahead from a perpendicular direction in daytime without a visual obstruction, and a left-turning car impacting a TW moving straight ahead from the opposite direction at night without a visual obstruction. However, important differences can be observed: an opposite direction scenario was included in Cao et al. (2019), while same direction scenarios were included the current study. This might be explained by different data samples used: Cao et al. (2019) focused on fatal and severe crashes, where head-on scenarios are more prominent. In the current study, all injury severities were considered to obtain typical car-to-TW scenarios. Cao et al. (2019) did not consider visual obstructions a clustering variable. Hence, there are no visual obstructions described in the Cao et al. (2019) test scenarios, even though relevant for AEB performance. In summary, this study complements the Cao et al. (2019) study by a) including all injury severities, and b) basing clustering on another dataset and with a larger number of crashes. Actually, it corroborates the Cao et al. (2019) findings in part as both studies highlight that car turning scenarios and night scenarios are being common in car-to-TW crashes in China and should be considered in AEB assessment.

Table 8
Four typical car-to-TW scenarios from Cao et al. (2019).

Scenario	Description	Exemplary pictograms
1	A car moving straight ahead impacting a TW moving straight ahead from a perpendicular direction in daytime without a visual obstruction. The car speed varies from 40-60 km/h. The contact point is at 50% of the car's front, measured from the driver's side.	
2	A car moving straight ahead impacting a TW moving in the same direction at night without a visual obstruction. The car speed varies from 40-60 km/h. The contact point is at 50% of the car's front, measured from the driver's side.	
3	A car moving straight ahead impacting a right-turning TW from the opposite direction in daytime without a visual obstruction. The car speed varies from 50-60 km/h. The contact point is at 50% of the car's front, measured from the driver's side.	
4	A left-turning car impacting a TW moving in the opposite direction at night without a visual obstruction. The car speed varies from 20-40 km/h. The contact point is at 50% of the car's front, measured from the driver's side.	

Descriptive frequency analysis has been applied in a previous study of most common car-to-TW crash scenarios in China (Sui et al., 2017). In that study, a more limited set of variables (pre-crash driving behavior of the car, pre-crash driving behavior of the TW and their relative moving direction) was analyzed compared to those in the present clustering study. While the limited variable set allows for cross-tabulation and sensible frequency analysis, it is not enough to completely describe test scenarios. For example, the most common scenario from Sui et al. (2017), a car moving straight ahead impacting a TW moving straight ahead from a perpendicular direction, was segmented into three scenarios for clustering: one was a daytime scenario without a visual obstruction, one was a daytime scenario with a visual obstruction, and the final one was a night-time scenario without a visual obstruction. Similarly, the second most common scenario they identified, a left-turning vehicle impacting a TW moving straight ahead from the opposite direction, was also augmented with more detailed environmental information: it was night-time and there was no visual obstruction.

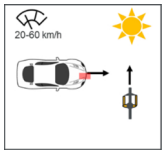
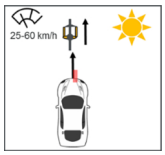
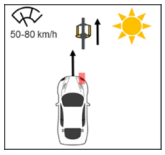
When only frequencies are used, it is possible to draw incorrect conclusions, due to data heterogeneity (Nilsson et al., 2018). Cluster analysis avoids this problem by dividing the data into several relatively homogenous classes or clusters. Moreover, as the number of variables and permutations increases, it becomes more important to capture the interdependencies and correlations between them, which clustering can do; simple frequency analysis cannot.

As seen in Table 5, the medoid scenarios (from cluster analysis) and frequent scenarios (from descriptive statistics) were not the same; some of the latter were not regarded as medoid scenarios. Instead, they were merged into the nearby medoid scenarios. One such example was the frequent scenario involving a left-turning car impacting a TW moving in a straight line from the opposite direction, in the daytime without any visual obstruction. The clustering algorithm merged this scenario with the same scenario at night, indicating that the scenario at night can provide more information when combined with other scenarios than when its descriptive statistic results were considered. When more information is used to derive representative crash scenarios, clustering analysis can minimize bias due to data heterogeneity.

A comparison of car-to-TW crash scenarios in China with those in the Euro NCAP is necessary to check the feasibility of introducing test scenarios in Euro NCAP to C-NCAP directly. As one of the globe's most advanced NCAPs which promotes the enhancement of vehicle safety, Euro NCAP has introduced testing methods for simulating AEB/pedestrian and AEB/bicyclist interactions (Euro NCAP, 2018a). Similarities can be found between applying cluster analysis to CIDAS and Euro NCAP testing scenarios as shown in Table 9: a car moving straight ahead impacting a TW approaching perpendicularly, in the daytime with no visual obstruction. Also, longitudinal, same-direction scenarios can be found both in the clustering results in this paper and in Euro NCAP testing scenarios (Euro NCAP, 2018a). However, our clustering scenarios include scenarios in which the car or the TW was turning, while in Euro NCAP scenarios only include cars or the bicyclist moving straight ahead (Euro NCAP, 2018a). This study also identified two additional variables important enough to merit inclusion in C-NCAP test scenarios — night-time and visual obstruction. Again, this is different from Euro NCAP, where there are no test scenarios that include night-time or visual obstruction (so far, at least). Finally, the scenario in this paper with the TW and the car coming from opposite directions is not currently included in the Euro NCAP 2018, either. The main reason for these differences is probably that traffic situations in China and European countries (in general) are different. Thus, the test scenarios for evaluating AEB performance in C-NCAP must be based on Chinese traffic situations, as the ones in this study are.

The relevance of these results to countries other than China is not obvious, as this result from India showed: the scenario of a car traveling straight ahead impacting an oncoming PTW was frequent in India (Painter et al., 2018), but not in China, according to our results. This difference was found in spite of the fact that both China and India have many PTWs and many related traffic deaths. Hence, the test scenarios proposed for China should not be applied to other countries without careful consideration of the local crash situation. In general, several aspects may contribute to the differences in number and type of car-to-TW crashes between China and other countries such as different lifestyles and different driving styles, as well as the different anthropometry of car drivers (which may affect the visibility of TWs).

Table 9
Euro NCAP test scenarios for AEB cyclist (Euro NCAP, 2018a).

Scenario	Description	Exemplary pictograms
CBNA-50	A car moving straight ahead impacting a cyclist moving straight ahead from a perpendicular direction in daytime without a visual obstruction. The car speed varies from 25-60 km/h. The contact point is at 50% of the car's front, measured from the driver's side.	
CBLA-50	A car moving straight ahead impacting a cyclist moving in the same direction in daytime without a visual obstruction. The car speed varies from 25-60 km/h. The contact point is at 50% of the car's front, measured from the driver's side.	
CBLA-25	A car moving straight ahead impacting a cyclist moving in the same direction in daytime without a visual obstruction. The car speed varies from 50-80 km/h. The contact point is at 75% of the car's front, measured from the driver's side.	

Due to the lack of properly validated weighting factors in CIDAS, it is not possible to scale up the results of this work to represent all of China; this lack of scalability is a major limitation of the study. Further, none of the CIDAS crashes used in the analysis have been reconstructed, so more detailed information about the crash characteristics (such as the pre-crash driving speeds of car and the TW) was not available in this study. Another limitation is the ASW value (0.41) for six clusters (see Appendix, Fig. A2). Normally, a higher ASW value would be preferred for more reliable clustering results (Sander and Lubbe, 2018). However, more clusters meant that some had unreasonably small sample sizes, so a relatively low ASW value was accepted, instead of choosing a higher number of clusters with a higher ASW value.

The typical car-to-TW crash scenarios in this work, obtained from clustering results, described key pre-crash scenario parameters from real-world traffic situations involving TWs. This valuable, in-depth knowledge can improve the safety of TW users in China. Assessment scenarios that are firmly anchored in real-world crash statistics from Chinese data are more relevant to Chinese traffic situations than scenarios derived from other data. Thus, they have a better chance of steering the development of vehicle technologies, such as AEB, towards saving more lives on Chinese roads. The scenarios can also be used in consumer assessment programs to improve the acceptability of new technological advances-and perhaps in infrastructure design to improve conditions outside the vehicle, as well.

To further improve in-crash countermeasures in car-to-TW crashes, future research should study the characteristics of the two-wheeler crashes that continue to occur when active safety systems, such as automatic emergency braking, are present in a larger proportion of the vehicle fleet in China. The six scenarios presented in this study may be a good starting point for such a study. Additionally, a car-to-TW pre-crash matrix database which includes trajectories of the car and the TW should be constructed, in order to enable more detailed studies of pre-crash behavior, as well as safety systems assessment through simulations.

5. Conclusions

Six typical car-to-TW scenarios were defined when clustering was applied to in-depth Chinese crash data. The scenarios can be briefly described as follows (see also pictograms in Table 6): (1) a car moving straight ahead, impacting a TW moving straight ahead from a

perpendicular direction, in daytime without visual obstruction; (2) a car turning right, impacting a TW moving straight ahead in the same direction, in daytime without visual obstruction; (3) a car moving straight ahead, impacting a left-turning TW previously moving in the same direction, in daytime without visual obstruction; (4) a car moving straight ahead, impacting a TW moving straight ahead from a perpendicular direction, at night without visual obstruction; (5) a left-turning car, impacting a TW moving straight ahead from the opposite direction, at night without visual obstruction; and finally, (6) a car moving straight ahead, impacting a TW moving straight ahead from a perpendicular direction, at night with a visual obstruction.

These scenarios provide in-depth insights about real-world car-to-TW crashes in China. These insights can be used when assessing the safety impact of current and future active and passive safety systems addressing the large number of car-to-TW crashes in China. In addition, these scenarios can be used as a basis for crash tests-in consumer assessment programs, for example-as they capture the essence of car-to-TW crashes in China.

Acknowledgements

The authors and would like to thank Jiguang Chen for the patient answers about CIDAS coding questions, and Ulrich Sander Daniel Nilsson for valuable discussions on the clustering methods, as well as Tina Mayberry for a language review of most of the manuscript.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The study was financed fully by Autoliv.

Appendix A

Figs. A1–A3 show the cluser silhouette plots. Figs. A4–A6 show variable distribution split for type of TW. Table A1 details grouping of TW travelling direction. Table A2–A3 show the distribution of TIME and OBSTRUCT variables in the dataset.

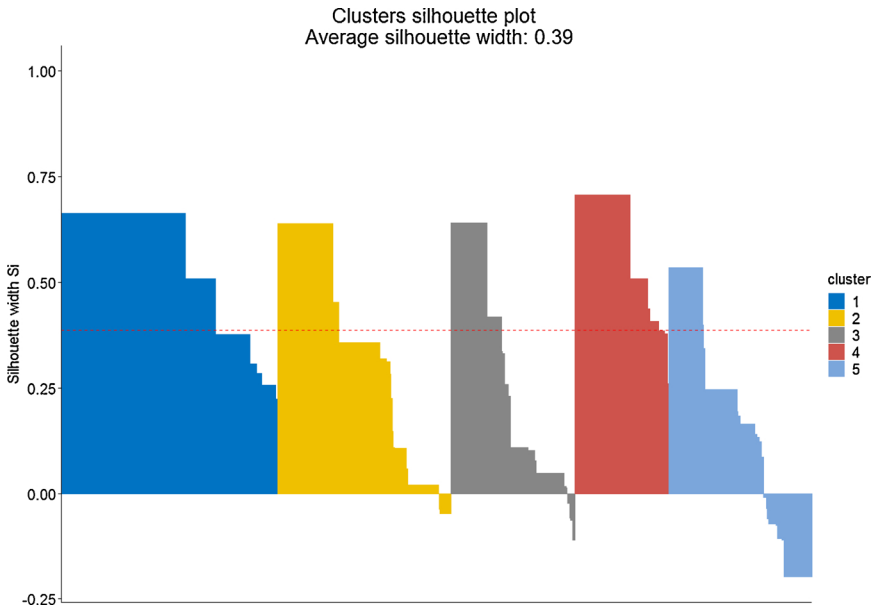


Fig. A1. Cluster silhouette plot: ASW value for the number of clusters equals to five. See Rousseeuw (1987) for a description of silhouette plots.

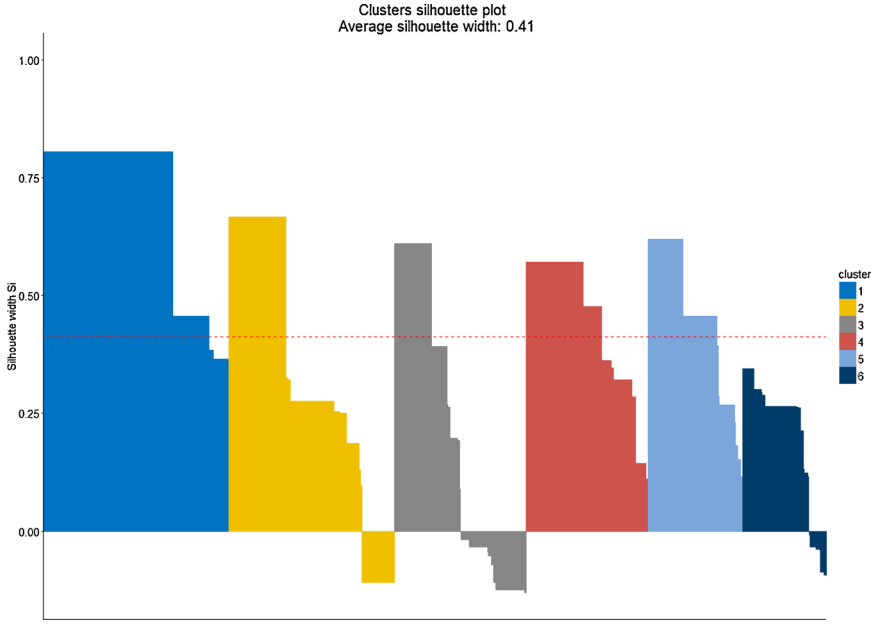


Fig. A2. Cluster silhouette plot: ASW value for the number of clusters equals to six.

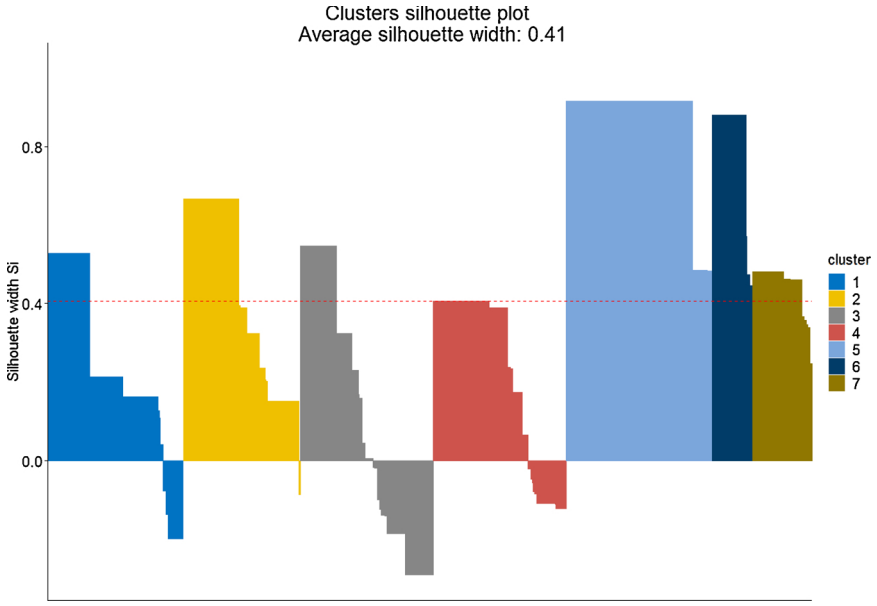


Fig. A3. Cluster silhouette plot: ASW value for the number of clusters equal to seven.

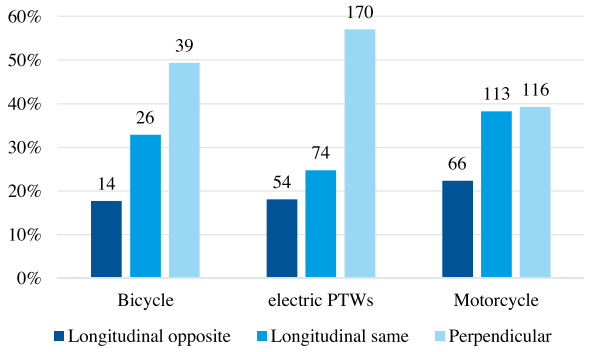


Fig. A4. Distribution of RELMOT in each TWs group.

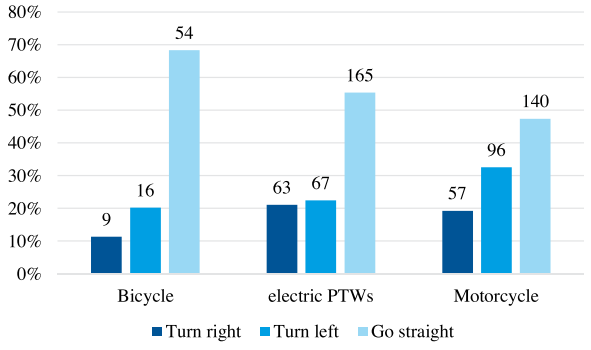


Fig. A5. Distribution of INTENT_Car in each TWs group.

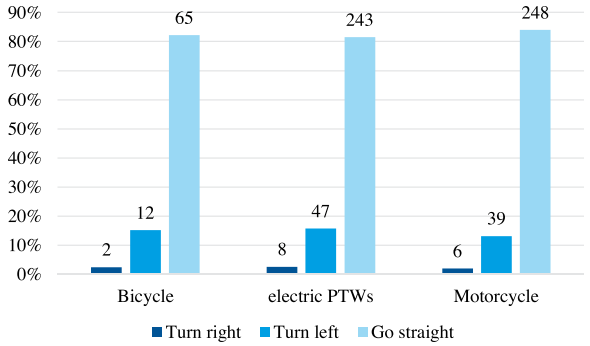


Fig. A6. Distribution of INTENT_TW in each TW group.

Table A1
Grouping table of TW travelling direction and contact point on the car.

Grouping	TW travelling direction	Contact point on the car
Group 1	Left side	Front_0–25%
	Right side	Front_75%–100%
Group 2	Left side	Front_25%–50%
	Right side	Front_50%–75%
Group 3	Left side	Front_50%–75%
	Right side	Front_25%–50%
Group 4	Left side	Front_75%–100%
	Right side	Front_25%–50%
Group 5	Left side	Left side front
	Right side	Right side front
Group 6	Left side	Left side middle
	Right side	Right side middle
Group 7	Left side	Left side rear
	Right side	Right side rear

Table A2
Distribution of TIME.

TIME	Frequency	Percent
Day	475	71%
Night	197	29%

Table A3
Distribution of OBSTRUCT.

OBSTRUCT	Frequency	Percent
No	590	88%
Yes	82	12%

References

Aggarwal, C.C., Reddy, C.K., 2013. Data Clustering: Algorithms and Applications. Chapman & Hall, London, UK.

Anderlucci, L., Hennig, C., 2014. The clustering of categorical data: a comparison of a model-based and a distance-based approach. Commun. Stat. Theory Methods 43 (4), 704–721. https://doi.org/10.1080/03610926.2013.806665.

Bhalla, K., Shotten, M., Cohen, A., Brauer, M., Shahraz, S., Burnett, R., Leach-Kemon, K., Freedman, G., Murray, C.J.L., Dingenen, R., Van Dentener, F., Vos, T., Naghavi, M., Abraham, J., Bartels, D., Weh, P.-H., 2014. Transport for Health: The Global Burden of Disease from Motorized Road Transport.

Ben-Hur, A., Guyon, I., 2003. Detecting stable clusters using principal component analysis. Funct. Genomics Methods Protoc. 1, 159–182. https://doi.org/10.1385/1-59259-364-X:159.

Cao, Y., Xiao, Y., Dong, H., Wang, Y., Wu, X., Li, P., Qiu, Y., 2019. Typical pre-crash scenarios reconstruction for two-wheelers and passenger vehicles and its application in parameter optimization of AEB system based on NAIS database. Proceedings of the 26th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Netherland.

CATARC, China Automotive Technology and Research Center, 1994. GB/T 15089–2001. Classification of Power-driven Vehicles and Trailer. Retrieved from:. http://www.camra.org.cn/static/uploadfile/content/standard/GB-T%2015089-2001.pdf.

Chen, Q., Chen, Y., Bostrom, O., Ma, Y., Liu, E., 2014. A comparison study of car-to-pedestrian and car-to-E-bike accidents: data source: the China in-depth Accident Study (CIDAS). SAE Tech. Paper. https://doi.org/10.4271/2014-01-0519. 2014-01-0519, 2014.

Cherry, C.R., 2007. Electric Two-wheelers in China: Analysis of Environmental, Safety, and Mobility Impacts. Unpublished PhD thesis, University of California, Berkeley.

C-NCAP, 2018. C-NCAP (China New Car Assessment Program) Management Regulation (2018 edition). Retrieved from:. http://www.c-ncap.org/cms/files/cncap-regulation-2018.pdf.

de Ona, J., López, G., Mujalli, R., Calvo, F.J., 2013. Analysis of traffic accidents on rural highways using latent class clustering and Bayesian networks. Accid. Anal. Prev. 51, 1–10. https://doi.org/10.1016/j.aap.2012.10.016.

Depaire, B., Wets, G., Vanhoof, K., 2008. Traffic accident segmentation by means of latent class clustering. Accid. Anal. Prev. 40, 1257–1266. https://doi.org/10.1016/j.aap.2008.01.007.

DESTATIS, 2016. Kraftrad- und Fahrradunfälle in Straßenverkehr 2015. Table 1.9. Retrieved from:. https://www.destatis.de/DE/Publikationen/Thematisch/TransportVerkehr/Verkehrsunfaelle/UnfaelleZweirad5462408157004.pdf?_blob=publicationFile.

Duan, J., Li, R., Hou, L., Wang, W., Li, G., Li, S.E., Cheng, B., Gao, H., 2017. Driver braking behavior analysis to improve autonomous emergency braking systems in typical Chinese vehicle-bicycle conflicts. Accid. Anal. Prev. 108, 74–82. https://doi.org/10.1016/j.aap.2017.08.022.

Euro NCAP, 2018a. European New Car Assessment Programme (Euro NCAP) — Test Protocol AEB VRU Systems. Retrieved from:. https://cdn.euroncap.com/media/41772/euro-ncap-aeb-vru-test-protocol-v203.201811091314249741.pdf.
Euro NCAP, 2018b. Strategy Working Group. Euro NCAP 20/25 Roadmap. Retrieved from:. https://cdn.euroncap.com/media/30700/euroncap-roadmap-2025-v4.pdf.
Fraley, C., Raftery, A.E., 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. Comput. J. 41 (8), 578–588. https://doi.org/10.1093/comjnl/41.8.578.
Fredriksson, R., Rosén, E., 2012. Integrated pedestrian countermeasures — potential of head injury reduction combining passive and active countermeasures. Saf. Sci. 50 (3), 400–407. https://doi.org/10.1016/j.ssci.2011.09.019.
Fredriksson, R., Ranjbar, A., Rosén, E., 2015. Integrated bicyclist protection systems — potential of head injury reduction combining passive and active protection systems. Proceedings of the 24th International Technical Conference on the Enhanced Safety of Vehicles (ESV.
Gower, J.C., 1971. A general coefficient of similarity and some of its properties. Biometrics 27 (4), 857. https://doi.org/10.1017/CBO9781107415324.004.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning. Elements 1, 337–387. https://doi.org/10.1007/b94608.
Huang, Z., 1998. Extensions to the k-means algorithms for clustering large data sets with categorical values. Data Min. Knowl. Discovery 2 (1998), 283. https://doi.org/10.1023/A:1009769707641.
Kaufman, L., Rousseeuw, P.J. (Eds.), 1990. Finding Groups in Data. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA.
Lenard, J., Danton, R., Avery, M., Weekes, A., Zuby, D., Kühn, M., 2011. Typical pedestrian accident scenarios for the testing of autonomous emergency braking systems. Proceedings of the 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV).
Lindman, M., Ödblom, A., Bergvall, E., Eidehall, A., Svanberg, B., Lukaszewicz, T., 2010. Benefit estimation model for pedestrian auto brake functionality. Proceedings of 4th International Conference on Expert Symposium on Accident Research (ESAR).
MacAlister, A., Zuby, S.D., 2015. Cyclist crash scenarios and factors relevant to the design of cyclist detection systems. Proceedings of IRCOBI Conference 2015 373–384.
Ministry of Industry and Information Technology of the People's Republic of China, 2018. Xin biaozhun yinling woguo diandong zixingche hangye gaozhiliang fazhan [HYPHEN] < diandong zixingche anquan jishu guifan > guojia biaozhun pigao jiedu [The new standard leads the development of electric bicycles to high quality [HYPHEN] < electric bicycle safety standard > interpretation of draft of national standard. Retrieved from:. http://www.miit.gov.cn/n1146295/n1652858/n1653018/c6014865/content.html.
National Bureau of Statistics of China, 2016. China Statistical Year Book 2016. Retrieved from:. http://data.stats.gov.cn/easyquery.htm?cn=C01&zb=A0S0D02&sj=2016.
National Bureau of Statistics of China, 2017. China Statistical Year Book 2017. Retrieved from:. http://www.stats.gov.cn/tjsj/ndsj/2017/indexeh.htm.
Nie, J., Yang, J., 2014. A study of bicyclist kinematics and injuries based on reconstruction of passenger car-bicycle accident in China. Accid. Anal. Prev. 71, 50–59. https://doi.org/10.1016/j.aap.2014.04.021.
Nilsson, D., Lindman, M., Victor, T., Dozza, M., 2018. Definition of run-off-road crash clusters—for safety benefit estimation and driver assistance development. Accid. Anal. Prev. 113 (January), 97–105. https://doi.org/10.1016/j.aap.2018.01.011.
Nitsche, P., Thomas, P., Stuetz, R., Welsh, R., 2017. Pre-crash scenarios at road junctions: a clustering method for car crash data. Accid. Anal. Prev. 107, 137–151. https://doi.org/10.1016/j.aap.2017.07.011.
Painter, M.S., Sonawane, S.B., Patel, M., 2018. Injuries sustained by riders and pillions of motorised two-wheelers in India. Proceedings of IRCOBI Asia Conference 2018.
Rosén, E., 2013. Autonomous emergency braking for vulnerable road users. Proceedings of IRCOBI Conference 2013.
Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.
Sander, U., Lubbe, N., 2018. The potential of clustering methods to define intersection test scenarios: assessing real-life performance of AEB. Accid. Anal. Prev. 113, 1–11. https://doi.org/10.1016/j.aap.2018.01.010.
Sasidharan, L., Wu, K.-F., Menendez, M., 2015. Exploring the application of latent class cluster analysis for investigating pedestrian crash injury severities in Switzerland. Accid. Anal. Prev. 85, 219–228. https://doi.org/10.1016/j.aap.2015.09.020.
Saunier, N., Mourji, N., Agard, B., 2011. Using data mining techniques to understand collision processes. Canadian Multidisciplinary Road Safety Conference.
Sharma, N., Gaud, N., 2015. K-modes clustering algorithm for categorical data. Int. J. Comput. Appl. 127 (17), 975–8887.
Seman, A., Bakar, Z.A., Sapawi, A.M., Othman, I.R., 2013. Amedoid-based method for clustering categorical data. J. Artif. Intell. 6, 257–265. https://scialert.net/abstract/?doi=jai.2013.257.265.
Sui, B., Zhou, S., Zhao, X., Lubbe, N., 2017. An overview of car-to-two-wheeler accidents in China: Guidance for AEB assessment. Proceedings of the 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV).
Uittenbogaard, J., Op den Camp, O., van Montfort, S., 2016a. CATS Deliverable 2.2: CATS car-to-cyclist accident parameters and test scenarios (No. TNO 2014 R11705).
Uittenbogaard, J., Rodarius, C., Op den Camp, O., 2016b. CATS Deliverable 1.2: CATS car-to-cyclist accident scenarios (No. TNO 2014 R11594).CATS Deliverable 1.2: CATS car-to-cyclist accident scenarios (No. TNO 2014 R11594).
Vermunt, J., Magidson, J., 2002. Latent class cluster analysis. Appl. Latent Cl. Anal. 1. https://doi.org/10.1186/1472-6947-9-47.
Vijay, R., Mahajan, P., Kandwal, R., 2012. Hamming distance based clustering algorithm. Int. J. Inf. Retrieval Res. (IJIRR) 2 (1), 11–20. https://doi.org/10.4018/ijirr.2012010102.
Weiss, H.B., Kaplan, S., Prato, C.G., 2015. Fatal and serious road crashes involving young New Zealand drivers: a latent class clustering approach. Int. J. Inj. Contr. Saf. Promot. 7300 (July), 1–17. https://doi.org/10.1080/17457300.2015.1056807.
WHO, World Health Organization, 2015. Global Status Report on Road Safety 2015. Geneva, Switzerland. .
WHO, World Health Organization, 2018a. Global Status Report on Road Safety 2018. Geneva, Switzerland. .
WHO, World Health Organization, 2018b. WHO Methods and Data Resources for Country-level Causes of Deaths 2000–2016. Geneva, Switzerland. .
Wisch, M., Seiniger, P., Pastor, C., Edwards, M., Visvikis, C., Reeves, C., 2013. Scenarios and weighting factors for pre-crash assessment of integrated pedestrian safety systems (No. SST.2011.RTD-1 GA No. 285106).
Wu, H., Sui, B., Lubbe, N., 2018. An overview of car-to-E-PTW accidents in China. Paper presented at the 21st China Conference of Automotive Safety Technology.
Zander, O., Hamacher, M., 2017. Revision of passive pedestrian test and assessment procedures to implement head protection of cyclists. Proceedings of the 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV).