AN APPROACH FOR TRAFFIC COLLISION AVOIDANCE: MEASURING THE SIMILAR EVIDENCE ON THE CAUSAL FACTORS OF COLLISIONS

Liangguo KANG1,2*, Shuli ZHANG1,3, Chao WU1,2#

1School of Resources and Safety Engineering, Central South University, Changsha, China
2Safety and Security Theory Innovation and Promotion Center, Central South University, Changsha, China
3Taiyuan Municipal Public Utilities Administration, Taiyuan, China

Submitted 6 December 2019; resubmitted 1 April 2020, 7 July 2020; accepted 27 July 2020; first published online

Abstract. The lessons learned from each Traffic Collision (TC) will help safety practitioners to avoid similar occurrences in the future. However, few studies and methods have focused specifically on the similar features among different collisions. Thus, the development of a measurement method for investigating the best evidence on the causal factors of TCs was warranted. In this study, a similarity analysis method based on the Analytic Hierarchy Process (AHP) and Similarity (S) theory, the AHP-S method, was constructed. This method was designed to identify the similar elements and similar units of collision scenes according to the analysis criteria and sub-criteria and further to calculate the degree of similarity between recognized similar pairs among TCs. Six TC cases were randomly selected as examples, and the degrees of similarity between cases 1 to 5 and case 6 were calculated separately. The calculation results showed that out of the five collision cases (cases 1–5), case 1 provided the best evidence for analysing the causal factors of case 6. This study promotes the development of quantitative analysis methods for collision incidents and provides an effective evidence-based method for TC avoidance.

Keywords: traffic collision, causal factors, similarity analysis, similar evidence, collision analysis.

Notations

AHP – analytic hierarchy process;
AHP-S – AHP with similarity theory;
CI – consistency index;
CR – consistency ratio;
RI – random index;
TC – traffic collision.

Introduction

TCs are a global problem and primary concern in the 21st century. Globally, the number of fatalities resulting from TCs is 1.25 million per year (Pérez et al. 2019). In the US, there are approximately 33000 traffic fatalities per year (Wu et al. 2013). In Europe, 26009 people were killed as a result of TCs in 2013 in the 28 EU countries (Chen et al. 2016). In Australia and New Zealand, respectively, 31 and 16% of fatal occupational incidents are TCs (Nævestad et al. 2015). In Great Britain, over the period 2004 to 2013, the average annual number of road fatalities was 2452 (Sarkar et al. 2018). Road safety is an issue that needs to be solved urgently in China, where 57277 people were killed in TCs in 2012 (Chen et al. 2015; Zhao, Deng 2015). Accordingly, safety countermeasures based on information sources related to such incidents should be taken to decrease the number of traffic fatalities and injuries. Causal analysis is one of the basic ways to study TCs and to discover the weaknesses of traffic safety measures. Under the circumstances, understanding the various factors that cause TCs is crucial.

Empirical research has identified a large number of factors related to TCs. The occurrence of TCs is usually related to human, vehicle, road, environment, and management factors (Prenktovskis et al. 2010; Podvezko, Sivilevičius 2013; Chen et al. 2015; Cvitanić, Vukoje 2018; Sze, Song 2019). Among the human factors, speeding or driver negligence are the main causes of TCs (Goel, Sachdeva 2016), and approximately 20% of TCs worldwide are related to driver fatigue (MacLean et al. 2003; Fernandes et al. 2010). Chen et al. (2004) found that 4.94% of collisions are caused by vehicle factors. Regarding road factors, creating reasonable traffic facilities could reduce the TC rate (Hu, Li 2014). With respect to environment factors, the number of TCs is significantly affected by weather variables: for example, the risk of a collision

*Corresponding author. E-mail: kangliangguo@126.com
#Corresponding author. E-mail: wuchao@csu.edu.cn

Copyright © 2021 The Author(s). Published by Vilnius Gediminas Technical University

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
increases if the precipitation is falling as snow (Anderson, Chapman 2011). However, these factors are not fully independent of each other, and each factor is influenced by many other factors and indicators (Chen et al. 2015).

Learning from past incidents is fundamental to incident prevention (Goh, Ube 2017). A good way to gain safety insights from collisions is through similarity analysis, which can create an overall evaluation indicator to calculate the degree of similarity between different collisions. This calculation model is based on a wide literature review, which involves five types of factors: human, vehicle, road, environment, and management. Currently, there are two major types of operations research methods in terms of indicator aggregating, namely, data envelopment analysis and multiple-criteria decision analysis (Zhou, Ang 2009). These two methods are commonly used to evaluate road safety performance for a set of decision-making units (Chen et al. 2015).

The AHP has been used in nearly all applications related to multiple-criteria decision analysis (Hruska et al. 2014; Stević et al. 2019). In the AHP, factors related to a decision-making issue are categorized and then formed into a hierarchy (Ilbahar et al. 2018). The application of the AHP has become common practice in road safety to evaluate and determine the importance weightings of indicators or criteria (Duleba 2019; Farooq et al. 2019). However, due to the lack of a computational procedure to assess degree of similarity, it is not enough to only use the AHP method for analysing the causal factors of collisions. Currently, there are no methods available to investigate similar evidence in relation to the causal factors of collisions. Thus, it is worthwhile exploring and testing new methods to make up for this deficiency. In this study, we developed a similarity analysis method (i.e., AHP-S method) for investigating the best evidence on the causal factors of TCs. The computational procedure of similarity analysis is as follows.

1.1. Mathematical method of similarity analysis

The core work of this section introduces the method we developed for computing the degree of similarity between TCs. This method combines the advantages of the AHP and Similarity (S) theory, and so we named it the AHP-S method. The AHP has been widely applied to solve complex multiple-criteria problems by structuring them into a hierarchy (Ho, Ma 2018). The overall decision-making goal is placed at the top of the hierarchy, with lower levels of criteria and alternatives placed below. Each criterion or sub-criterion can be further divided into appropriate levels of detail, and decision alternatives are laid down at the last level of the hierarchy (Calabrese et al. 2016). Similarity theory is widely used to provide evidence-based practice support for safety practitioners, but it can only qualitatively analyse the similar features of TCs. Currently, similarity theory lacks a suitable quantitative description for analysing the degree of similarity between incidents, and thus there is a need to construct computational procedures. In this context, we developed a similarity analysis method (i.e., AHP-S method) for investigating the best evidence on the causal factors of TCs. The computational procedure of similarity analysis is as follows.

1.1.1. The relative weight \( w_{ij} \)

To obtain the degree of importance of similar units, a structure is established which takes the criteria as the target level, the sub-criteria as the middle level, and the similar units as the bottom level (Figure 1).

The relative importance of two similar units is established on the basis of Saaty’s Fundamental Scale (Herva, Roca 2013), which consists of nine possible numeric values. Table 1 provides descriptions of each degree of the scale.

The comparison of two similar units \((u_i, u_j)\) can be mathematically presented as:

\[
S_{ij} = \frac{u_i}{u_j} \quad (i, j = 1, 2, 3, \ldots, k),
\]

where: \( S_{ij} \) denotes the weight exchange value of the pairwise comparison of \(u_i\) and \(u_j\) (Table 2).

![Figure 1. Structure for analysing the importance of similar units](image)

<table>
<thead>
<tr>
<th>Intensity of importance</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>equal importance</td>
</tr>
<tr>
<td>3</td>
<td>moderate importance</td>
</tr>
<tr>
<td>5</td>
<td>essential or strong importance</td>
</tr>
<tr>
<td>7</td>
<td>very strong importance</td>
</tr>
<tr>
<td>9</td>
<td>extreme importance</td>
</tr>
<tr>
<td>2, 4, 6, 8</td>
<td>intermediate values between the two adjacent judgments</td>
</tr>
</tbody>
</table>
Table 2. Pair-wise comparison matrix for similar units

<table>
<thead>
<tr>
<th></th>
<th>u₁</th>
<th>u₂</th>
<th>u₃</th>
<th>u₄</th>
<th>u₅</th>
<th>u₆</th>
<th>u₇</th>
<th>u₈</th>
<th>u₉</th>
<th>u₁₀</th>
<th>u₁₁</th>
<th>u₁₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>u₁</td>
<td>s₁₁</td>
<td>s₁₂</td>
<td>s₁₃</td>
<td>s₁₄</td>
<td>s₁₅</td>
<td>s₁₆</td>
<td>s₁₇</td>
<td>s₁₈</td>
<td>s₁₉</td>
<td>s₁₁₀</td>
<td>s₁₁₁</td>
<td>s₁₁₂</td>
</tr>
<tr>
<td>u₂</td>
<td>s₁₂</td>
<td>s₂₂</td>
<td>s₂₃</td>
<td>s₂₄</td>
<td>s₂₅</td>
<td>s₂₆</td>
<td>s₂₇</td>
<td>s₂₈</td>
<td>s₂₉</td>
<td>s₂₁₀</td>
<td>s₂₁₁</td>
<td>s₂₁₂</td>
</tr>
<tr>
<td>u₃</td>
<td>s₁₃</td>
<td>s₂₃</td>
<td>s₃₃</td>
<td>s₃₄</td>
<td>s₃₅</td>
<td>s₃₆</td>
<td>s₃₇</td>
<td>s₃₈</td>
<td>s₃₉</td>
<td>s₃₁₀</td>
<td>s₃₁₁</td>
<td>s₃₁₂</td>
</tr>
<tr>
<td>u₄</td>
<td>s₁₄</td>
<td>s₂₄</td>
<td>s₃₄</td>
<td>s₄₄</td>
<td>s₄₅</td>
<td>s₄₆</td>
<td>s₄₇</td>
<td>s₄₈</td>
<td>s₄₉</td>
<td>s₄₁₀</td>
<td>s₄₁₁</td>
<td>s₄₁₂</td>
</tr>
<tr>
<td>u₅</td>
<td>s₁₅</td>
<td>s₂₅</td>
<td>s₃₅</td>
<td>s₄₅</td>
<td>s₅₅</td>
<td>s₅₆</td>
<td>s₅₇</td>
<td>s₅₈</td>
<td>s₅₉</td>
<td>s₅₁₀</td>
<td>s₅₁₁</td>
<td>s₅₁₂</td>
</tr>
<tr>
<td>u₆</td>
<td>s₁₆</td>
<td>s₂₆</td>
<td>s₃₆</td>
<td>s₄₆</td>
<td>s₅₆</td>
<td>s₆₆</td>
<td>s₆₇</td>
<td>s₆₈</td>
<td>s₆₉</td>
<td>s₆₁₀</td>
<td>s₆₁₁</td>
<td>s₆₁₂</td>
</tr>
<tr>
<td>u₇</td>
<td>s₁₇</td>
<td>s₂₇</td>
<td>s₃₇</td>
<td>s₄₇</td>
<td>s₅₇</td>
<td>s₆₇</td>
<td>s₇₇</td>
<td>s₇₈</td>
<td>s₇₉</td>
<td>s₇₁₀</td>
<td>s₇₁₁</td>
<td>s₇₁₂</td>
</tr>
<tr>
<td>u₈</td>
<td>s₁₈</td>
<td>s₂₈</td>
<td>s₃₈</td>
<td>s₄₈</td>
<td>s₅₈</td>
<td>s₆₈</td>
<td>s₇₈</td>
<td>s₈₈</td>
<td>s₈₉</td>
<td>s₈₁₀</td>
<td>s₈₁₁</td>
<td>s₈₁₂</td>
</tr>
<tr>
<td>u₉</td>
<td>s₁₉</td>
<td>s₂₉</td>
<td>s₃₉</td>
<td>s₄₉</td>
<td>s₅₉</td>
<td>s₆₉</td>
<td>s₇₉</td>
<td>s₈₉</td>
<td>s₉₉</td>
<td>s₉₁₀</td>
<td>s₉₁₁</td>
<td>s₉₁₂</td>
</tr>
<tr>
<td>u₁₀</td>
<td>s₁₁₀</td>
<td>s₂₁₀</td>
<td>s₃₁₀</td>
<td>s₄₁₀</td>
<td>s₅₁₀</td>
<td>s₆₁₀</td>
<td>s₇₁₀</td>
<td>s₈₁₀</td>
<td>s₉₁₀</td>
<td>s₁₁₀</td>
<td>s₁₁₁</td>
<td>s₁₁₂</td>
</tr>
<tr>
<td>u₁₁</td>
<td>s₁₁₁</td>
<td>s₂₁₁</td>
<td>s₃₁₁</td>
<td>s₄₁₁</td>
<td>s₅₁₁</td>
<td>s₆₁₁</td>
<td>s₇₁₁</td>
<td>s₈₁₁</td>
<td>s₉₁₁</td>
<td>s₁₁₁</td>
<td>s₁₁₂</td>
<td>s₁₁₃</td>
</tr>
<tr>
<td>u₁₂</td>
<td>s₁₁₂</td>
<td>s₂₁₂</td>
<td>s₃₁₂</td>
<td>s₄₁₂</td>
<td>s₅₁₂</td>
<td>s₆₁₂</td>
<td>s₇₁₂</td>
<td>s₈₁₂</td>
<td>s₉₁₂</td>
<td>s₁₁₂</td>
<td>s₁₁₃</td>
<td>s₁₁₄</td>
</tr>
</tbody>
</table>

$S_{ij} = 1$ means that $u_i$ is equally as important as $u_j$, and $S_{ij} = 9$ means that $u_i$ is extremely more important than $u_j$. Odd degrees are usually used for judgments, and even degrees (2, 4, 6, and 8) are used if the decision maker is unable to make an explicit assessment between two judgments. The determination of similar units is based on the axiom of reciprocity. If the decision maker considers that $u_i$ is moderately more important than $u_j$ ($u_i = 3 \cdot u_j$, i.e., $S_{ij} = 3$), then $u_i$ will be three times weaker than $u_j$ ($u_j = 1/3 \cdot u_i$, i.e., $S_{ij} = 1/3$). The matrix $S$ can be mathematically presented as:

$$S = \begin{bmatrix}
    s_{11} & ... & s_{1j} & ... & s_{1k} \\
    ... & ... & ... & ... & ... \\
    s_{ij} & ... & s_{jj} & ... & s_{jk} \\
    ... & ... & ... & ... & ... \\
    s_{ik} & ... & s_{jk} & ... & s_{kk}
\end{bmatrix}. \quad (2)$$

By normalizing the solution, the relative weight $w_i$ can be mathematically presented as:

$$w_i = \frac{1}{k} \sum_{j=1}^{k} s_{ij} \cdot \frac{1}{\sum_{j=1}^{k} s_{ij}}. \quad (3)$$

The maximum eigenvalue $\lambda_{max}$ equals the number of orders. It can be presented as follows:

$$\lambda_{max} = \frac{1}{k} \sum_{i=1}^{k} \left( S \cdot w_i \right). \quad (4)$$

If $S$ is a consistent matrix, then the maximum eigenvalue of $S$ is equal to its number of orders. However, the pair-wise comparison matrix cannot achieve complete consistency in practice. The difference in value between $\lambda_{max}$ and $k$ can be used to judge the degree of consistency. The CI can be calculated as follows:

$$CI = \frac{\lambda_{max} - k}{k - 1}. \quad (5)$$

For each comparison matrix, a corresponding RI is used. RI is an index of a randomly generated reciprocal matrix. Its values for computation purposes are presented in Table 3.

To check the correctness of comparisons, a CR is designated. It is calculated to determine the inconsistency in the evaluation. It is determined in accordance with the following equation:

$$CR = \frac{CI}{RI}. \quad (6)$$

Table 3. Average consistency values of random matrices

<table>
<thead>
<tr>
<th>k</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>RI</td>
<td>0</td>
<td>0</td>
<td>0.52</td>
<td>0.89</td>
<td>1.12</td>
<td>1.24</td>
<td>1.36</td>
<td>1.41</td>
<td>1.46</td>
<td>1.49</td>
<td>1.52</td>
<td>1.54</td>
</tr>
</tbody>
</table>

Table 4. Pair-wise similarity scale

<table>
<thead>
<tr>
<th>Degree of similarity</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>slight similarity</td>
</tr>
<tr>
<td>0.3</td>
<td>moderate similarity</td>
</tr>
<tr>
<td>0.5</td>
<td>essential or strong similarity</td>
</tr>
<tr>
<td>0.7</td>
<td>very strong similarity</td>
</tr>
<tr>
<td>0.9</td>
<td>extreme similarity</td>
</tr>
<tr>
<td>0.2, 0.4, 0.6, 0.8</td>
<td>intermediate values between the two adjacent judgments</td>
</tr>
</tbody>
</table>

Figure 2. Structure for analysing the values of similar units

If $CR \leq 0.1$, it means that the evaluation within the matrix is acceptable. Otherwise, the judgments are untrustworthy.

1.1.2. The values of similar units $q_i$

A structure is constructed to obtain the values of similar units by means of similarity theory, which takes sub-criteria as the target level, similar units as the middle level, and similar elements as the bottom level (Figure 2).

It assumes that case I involves $m$ elements (denoted by $a$) and case II involves $n$ elements (denoted by $b$). The number of similar units between case I and II is $k$, which may be mathematically presented as:

$$k = m \cap n. \quad (7)$$

where: $a_i$, $b_j$ denote similar elements; $u_i$ denotes a similar unit formed by $a_i$ and $b_j$; $q_i$ denotes the degree of similarity between $a_i$ and $b_j$, i.e., the values of $u_i$ ($i = 1, 2, 3, \ldots, k$). Also, $q_i$ within 0 to 1 are selected. $q_i = 0$ means that $a_i$ is totally different from $b_j$, and $q_i = 1$ means that $a_i$ is the same as $b_j$. The degrees 0.1, 0.3, 0.5, 0.7, and 0.9 are usually used, and the degrees 0.2, 0.4, 0.6, and 0.8 are used if it is not possible for the decision maker to make an explicit assessment (Table 4). $q_i = 0.7$ means that there is very strong similarity between $a_i$ and $b_j$, which indicates that the decision maker assessed the degree of similarity between two elements to be very strong.

1.1.3. The degree of similarity $Q(I, II)$

According to the relative weight $w_i$ and the values of similar units $q_i$, the degree of similarity between case I and
case II, i.e., \( Q(I, II) \), is calculated as follows:

\[
Q(I, II) = \frac{k}{m + n - k} \cdot \sum_{i=1}^{k} w_i \cdot q_i .
\]  

(8)

Case I will provide the best evidence available for analysis in case II when \( Q(I, II) \) is close to 1. Otherwise, it cannot provide evidence for analysis in case II.

1.2. Research process of similarity analysis

Similarity analysis of the causal factors of TCs can be realized by using the principle of “similarity safety systematics” (Wu, Jia 2016; Jia et al. 2016), which is intended to analyse the degree of similarity between collisions. As the large number of traffic cases provides a practical basis for analysis, similarity analysis of collisions is one of the important tools that can be used to avoid similar occurrences by using the data and detailed information on recorded collisions. Similarity analysis has the following advantages:

» it makes up for the shortcomings of the traditional frequency analysis of collisions;

» it provides the decision maker with effective evidence by calculating the degree of similarity between the causal factors of TCs;

» it enables safety countermeasures based on current incident information sources to be taken to improve traffic safety practices.

Similarity analysis investigates the similar features of study objects, thus providing evidence-based information for traffic practices (Wang et al. 2017). Taking TCs as an example, the research process of similarity analysis is as follows (Figure 3):

» TC cases are selected according to the research objectives;

» the decision maker sets specific criteria and sub-criteria of TCs: e.g., criteria are set as driver factors, and sub-criteria are set as driver's behaviour, decision-making ability, and reaction speed;

» the similar elements of TCs are identified: e.g., the similar elements of a causal factor are identified by the criteria and sub-criteria, and then the degree of similarity between similar elements is analysed in a mathematical way;

» the similar units of TCs are analysed: e.g., similar units are identified by means of the elements of different cases; then, the decision maker mathematically compares the relative importance of the similar units;

» the degree of similarity between case I and case II is calculated under the condition of \( 0 \leq Q(I, II) \leq 1 \);

» the decision maker applies the results to improve traffic safety practices.

2. Application of proposed method

2.1. TC case selection

There were approximately 960000 TCs in China over the period 2012 to 2016; among these, 68 were major accident level or above. The term major accident level or above refers to an accident that results in over 10 deaths or over 50 serious injuries or in direct economic losses of over 50 million CNY (SC PRC 2007). Information on collisions was obtained from the accident investigation reports of the State Administration of Work Safety (SAWS 2016) of China as well as related studies by Li and Xiao (2016a, 2016b, 2016c, 2016d, 2016e, 2017). The 68 cases were classified as follows: 16 head-on crashes, 6 rear-end crashes, 38 overturns, 3 fires, and 5 others. Six cases were randomly selected from among the 16 head-on crashes to analyse the degree of similarity between the causal factors of TCs (Table 5).

2.2. Criteria and sub-criteria of TCs

TCs are among the leading causes of death and injuries of various levels in China (Chen et al. 2015). The key focus of similarity analysis is to investigate the cause of collisions on the basis of current collision information sources. Llopis-Castelló et al. (2018) proposed that the most important factors in the occurrence of road crashes are related to infrastructure, vehicle, and human factors. Lum and Reagan (1995) studied crash reports in the US, which showed that collisions are due to the roadway conditions, drivers, and vehicles and the interaction among several factors. Chen et al. (2004) analysed the causes of TCs in China across 2000; their analysis indicated that collisions are caused by human, vehicle, road, and environment factors. The findings indicate that road safety is a complex system issue comprising five types of factors (human, vehicle, road, environment, and management) and that TCs can be caused by a single factor or a combination of these five factors (Chen et al. 2015). The relationship among the five factors is illustrated in Figure 4.

The causal factors of TCs can be divided into internal factors (management) and external factors (human, vehicle, road, and environment), and these factors contain similar elements:

» the causal factors of TCs mutually affect one another and are connected: for instance, poor weather (e.g., rainy weather) has a great influence on driver’s sight, which increases the probability of a collision occurring;

» the causes of TCs mainly consist of one or several factors: for instance, human factors (e.g., drink-driving) may be the main cause of a TC even if environment, vehicle, and road conditions are normal;
the condition of an external factor can reflect a flaw of the internal factor: for instance, a defect in a safety facility (e.g., no traffic lights) indicates a road safety management deficiency.

Through the above analysis, the TC criteria in this study were set as five factors: human, vehicle, road, environment, and management. The sub-criteria of TCs were determined in terms of the respective criteria, as shown in Table 6.

Figure 5 illustrates the scheme of the AHP-S hierarchy model used in this study. In this model, the goal, criteria, sub-criteria, similar elements, similar units, and degree of similarity between TCs are defined and divided into specified groups. All of them are put into the cor-

<table>
<thead>
<tr>
<th>Case</th>
<th>A brief introduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>On 23 April 2012, Cao was driving a heavy truck. He crossed the double yellow lines in the center of the highway, rushing into the opposite lane. Then, at 132 km + 500 m on provincial highway 220 in China, the truck collided with a medium-sized bus driven by Gao. The incident resulted in 14 deaths, 11 injuries, and a direct economic loss of over 7.2 million CNY.</td>
</tr>
<tr>
<td>2</td>
<td>On 20 August 2012, Xu was driving a minibus. He turned right at 22 km + 600 m on provincial highway 110 in China. He crossed the center line and collided with a heavy truck, driven by Diao, travelling in the opposite direction. The incident resulted in 12 deaths, 1 injury, and a direct economic loss of over 6.0 million CNY.</td>
</tr>
<tr>
<td>3</td>
<td>On 12 August 2013, Li was driving a heavy truck. He crossed the center line and his truck collided with a large bus travelling in the opposite direction at 768 km + 260 m on national highway 312 in China. The incident resulted in 11 deaths, 12 injuries, and a direct economic loss of over 12.6 million CNY.</td>
</tr>
<tr>
<td>4</td>
<td>On 26 August 2013, Shao was driving a heavy truck. At 305 km + 140 m on national highway 310 in China, Shao took the wrong measures and turned left into the opposite lane when the vehicle in the front suddenly braked. Then, the truck had a frontal collision with a medium-sized bus driven by Ge. The incident resulted in 10 deaths, 5 injuries, and a direct economic loss of over 8.1 million CNY.</td>
</tr>
<tr>
<td>5</td>
<td>On 2 May 2015, Yuan was driving a minibus when he suddenly crashed into the opposite lane. Then, the minibus collided with the front part of a heavy truck at Jinqi road in China. The incident resulted in 10 deaths, 3 injuries, and a direct economic loss of over 3.7 million CNY.</td>
</tr>
<tr>
<td>6</td>
<td>On 13 October 13 2016, Zhu was driving a heavy semitrailer, while Yu was driving a three-wheel vehicle in the opposite direction. Both of them slightly crossed the center line and consequently collided with each other. The incident resulted in 11 deaths and a direct economic loss of over 3.1 million CNY.</td>
</tr>
</tbody>
</table>

Note: 1 CNY is approximately 0.1571 USD; direct economic losses only involve property damage.
rect hierarchical order, and then the connection between them is proposed (Fabjanowicz et al. 2018). Specifically, the goal is connected to each of the criteria; each criterion is linked with sub-criteria; each sub-criterion is connected to a similar element; each similar element is linked with a similar unit, and a similar unit is connected to the degree of similarity between two TC cases.

2.3. Similar elements and similar units of the causal factors of TCs

According to the criteria and sub-criteria (Table 6), the six cases were investigated to examine the degree of similarity between the causal factors of TCs. The causal factors of these cases are shown in Table 7. This paper took case 6 as

![Figure 5. Hierarchical structure of the AHP-S model used for the similarity analysis of the causal factors of TCs](image)

<table>
<thead>
<tr>
<th>Goal</th>
<th>Similarity analysis of causal factors of TCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria</td>
<td>Human</td>
</tr>
<tr>
<td>Human sub-criteria</td>
<td>Psychological state</td>
</tr>
<tr>
<td>Vehicle sub-criteria</td>
<td>Load</td>
</tr>
<tr>
<td>Road sub-criteria</td>
<td>Road condition</td>
</tr>
<tr>
<td>Environment sub-criteria</td>
<td>Weather</td>
</tr>
<tr>
<td>Management sub-criteria</td>
<td>Safety management</td>
</tr>
</tbody>
</table>

| Degree of similarity | Cases 1 and 6 | Cases 2 and 6 | Cases 3 and 6 | Cases 4 and 6 | Cases 5 and 6 |

Table 7. Causal factors of six TC cases

<table>
<thead>
<tr>
<th>Case</th>
<th>Road sub-criteria</th>
<th>Environment sub-criteria</th>
<th>Management sub-criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>road condition</td>
<td>road line</td>
<td>road facilities</td>
</tr>
<tr>
<td>1</td>
<td>flat road</td>
<td>complete</td>
<td>complete</td>
</tr>
<tr>
<td>2</td>
<td>gentle slope</td>
<td>complete</td>
<td>complete</td>
</tr>
<tr>
<td>3</td>
<td>flat road</td>
<td>complete</td>
<td>complete</td>
</tr>
<tr>
<td>4</td>
<td>flat road</td>
<td>complete</td>
<td>complete</td>
</tr>
<tr>
<td>5</td>
<td>gentle slope</td>
<td>fuzzy</td>
<td>incomplete</td>
</tr>
<tr>
<td>6</td>
<td>long slope</td>
<td>complete</td>
<td>complete</td>
</tr>
</tbody>
</table>

Note: The items in italic style represents the incident-causing factors of cases 1 to 6.
the central study subject for the similarity analysis of the causal factors of TCs; this case was compared with each of the other cases separately. Taking case 1 and case 6 as an example, they were compared from the five perspectives: human, vehicle, road, environment, and management (Table 7). It was assumed that the similar elements of case 1 is recorded as \( a_k \) and \( b_k \) of case 6. A similar unit \( u_1 \) was formed when \( a_k \) is similar to \( b_k \). As shown in Table 7, cases 1 and 6 include similar elements, such as violation and illegal operation; thus, violation and illegal operation can form a similar unit (i.e., \( u_4 \)), recorded as “\( u_4 \) = violation and illegal operation”. Following the same method, we recorded “\( u_2 \) = speeding”, “\( u_3 \) = overload”, “\( u_4 \) = imperfect safety management system”, and “\( u_5 \) = insufficient traffic supervision”. The values of a similar unit \( q_i \) between cases 1 and 6 were computed by the similarity method and recorded as \( q_i = (q_{11}, q_{22}, q_{33}, q_{44}, q_{55}) \). Following the above steps, the similar units between case 6 and cases 2 to 5, respectively, were analysed separately. The similar units between case 6 and cases 1 to 5 are summarized in Table 8.

2.4. Degree of similarity between the causal factors of TCs

Cases 1 and 6 were taken as examples to calculate the degree of similarity between the causal factors of TCs using the AHP-S method. The main steps of the calculation were as follows.

2.4.1. The relative weight \( w_i \) of the causal factors of TCs

According to the statistical results of crash reports, most collisions are caused by human factors rather than by vehicle defects. For example, Lum and Reagan (1995) found that 93% of collisions are related to human factors, and Chen et al. (2004) found that 89.95% of collisions are caused by human factors. From the perspective of collision avoidance, human factors have a higher weighted value than vehicle factors. Excessive speed is a leading risk factor on the road, and most drivers determine their speed by observing their surroundings (Matínez et al. 2013). Speeding is also considered as a human factor. As \( u_1 \) involves two risk factors, we reasoned that compared to \( u_2 \), the intensity of \( u_1 \) falls in between equal importance and moderate importance, namely \( S_{12} = 2 \) and \( S_{21} = 1/2 \). In addition, compared to \( u_3 \), the intensity of \( u_2 \) falls in between equal importance and moderate importance, meaning \( S_{32} = 2 \), \( S_{23} = 1/2 \), \( S_{13} = 3 \), and \( S_{31} = 1/3 \).

Moreover, the human, vehicle, road, and environment risk prevention factors are tightly related to the management system factor. In terms of collision avoidance, management factors have a higher weighted value than human factors. As \( u_4 \) and \( u_5 \) are considered as management factors, we reasoned that \( u_4 \) is equally as important as \( u_5 \), namely \( S_{54} = 1 \) and \( S_{54} = 1 \). In addition, we reasoned that \( u_4 \) is moderately more important than \( u_2 \), meaning \( S_{24} = 1/3 \), \( S_{42} = 3 \), \( S_{25} = 1/3 \), \( S_{52} = 3 \), \( S_{14} = 1/2 \), \( S_{41} = 2 \), \( S_{15} = 1/2 \), and \( S_{51} = 2 \).

From what has been discussed above, we reasoned that compared to \( u_3 \), the intensity of \( u_4 \) falls in between moderate importance and essential importance, namely \( S_{45} = 1/4 \), \( S_{54} = 4 \), \( S_{35} = 1/4 \), and \( S_{53} = 4 \). Finally, we established a pair-wise comparison matrix (Table 9).

There exist 7 and 5 similar elements in cases 1 and 6, accordingly, and 5 similar units of these two cases were formed (Table 8) i.e., \( m = 7 \), \( n = 5 \), \( k = 5 \), \( RI = 1.12 \). The data (i.e., Table 9, \( k \), and RI) were put into Equations (3)–(6) to calculate the relative weight \( w_i \), as follows:

\[
\begin{align*}
    w_i &= (0.1841, 0.11, 0.0067, 0.3186, 0.3186)^T.
\end{align*}
\]

The CR was calculated as follows:

\[
\begin{align*}
    CR &= 0.0068 < 0.1.
\end{align*}
\]

Consequently, the evaluation within the matrix was acceptable.

2.4.2. The values of the similar unit \( q_i \) of the causal factors of TCs

The values of the similar unit \( q_i \) were calculated as follows using the similarity method:

\[
\begin{align*}
    q_i &= (0.4, 0.7, 0.5, 0.8, 0.8).
\end{align*}
\]

Table 8. The similar units of the five groups

<table>
<thead>
<tr>
<th>Grouping</th>
<th>Similar units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases 1 and 6</td>
<td>( u_1 ) = violation and illegal operation; ( u_2 ) = speeding; ( u_3 ) = overload; ( u_4 ) = imperfect safety management system; ( u_5 ) = insufficient traffic supervision</td>
</tr>
<tr>
<td>Cases 2 and 6</td>
<td>( u_1 ) = violation; ( u_2 ) = speeding; ( u_3 ) = overload; ( u_4 ) = imperfect safety management system; ( u_5 ) = insufficient traffic supervision</td>
</tr>
<tr>
<td>Cases 3 and 6</td>
<td>( u_1 ) = violation and illegal operation; ( u_2 ) = speeding; ( u_3 ) = overload; ( u_4 ) = imperfect safety management system; ( u_5 ) = insufficient traffic supervision</td>
</tr>
<tr>
<td>Cases 4 and 6</td>
<td>( u_1 ) = violation and illegal operation; ( u_2 ) = speeding; ( u_3 ) = overload; ( u_4 ) = imperfect safety management system; ( u_5 ) = insufficient traffic supervision</td>
</tr>
<tr>
<td>Cases 5 and 6</td>
<td>( u_1 ) = violation and illegal operation; ( u_2 ) = speeding; ( u_3 ) = overload; ( u_4 ) = imperfect safety management system; ( u_5 ) = insufficient traffic supervision</td>
</tr>
</tbody>
</table>

Table 9. Pair-wise comparison matrix of similar units of the causal factors of TCs

<table>
<thead>
<tr>
<th>( S )</th>
<th>( u_1 )</th>
<th>( u_2 )</th>
<th>( u_3 )</th>
<th>( u_4 )</th>
<th>( u_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u_1 )</td>
<td>1</td>
<td>2/3</td>
<td>1/2</td>
<td>1/2</td>
<td></td>
</tr>
<tr>
<td>( u_2 )</td>
<td>1/2</td>
<td>1</td>
<td>2/3</td>
<td>1/3</td>
<td>1/3</td>
</tr>
<tr>
<td>( u_3 )</td>
<td>1/3</td>
<td>1/2</td>
<td>1</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>( u_4 )</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>( u_5 )</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

From (9)
2.4.3. The degree of similarity between the causal factors of TCs

$m$, $n$, $k$, $q_i$, and $w_i$ were put into Equation (8), and the degree of similarity between cases 1 and 6 was calculated as follows:

$$Q(1,6) = \frac{5}{7 + 5 - 5} \cdot \left(0.1841 \cdot 0.4 + 0.11 \cdot 0.7 + 0.0687 \cdot 0.5 + 0.3186 \cdot 0.8 + 0.3186 \cdot 0.8\right) = 49.63\%.$$  (12)

In addition, the degrees of similarity between cases 2 to 5 and case 6 were calculated respectively following the above steps; the results were as follows:

$$Q(2,6) = 41.68\%;$$  (13)

$$Q(3,6) = 40.76\%;$$  (14)

$$Q(4,6) = 46.11\%;$$  (15)

$$Q(5,6) = 33.98\%.$$  (16)

3. Discussion of results

Our analysis of the casual factors of six TC cases clearly shows that every case included causes such as driver violations, speeding, and overload. According to the investigation reports, driver factors, such as crossing the center line, illegal overtaking, speeding, and driver fatigue, are the direct causes of collisions. In addition, vehicle factors, such as overload and brake problems, increase the severity of collisions (Vorko-Jović et al. 2006).

Due to knowledge gaps or biases, decision makers may ignore crucial information in the investigation and analysis of incidents. Identifying the similar units of the causal factors of TCs will help decision makers to find evidence-based information for preventing the occurrence of similar incidents. For instance, case 1 happened in 2012, which was earlier than case 6 happening in 2016. Case 6 could have been avoided if decision makers had learnt lessons from case 1 in terms of the human, vehicle, road, environmental, and management related causal factors and taken appropriate actions to control similar conditions.

The ranking of the degrees of similarity between cases 1 to 5 and case 6 can be described as follows: case 1 > case 4 > case 2 > case 3 > case 5. Compared to the other four cases, case 1 provides the best evidence available for analysing the causal factors of case 6. Due to the degree of similarity between case 1 and case 6 being scored 0.4963, it can be used with larger samples of TCs to find other high-similarity cases for analysing case 6. In the future, this method can be developed into computer software to analyse the degree of similarity between different collisions, which can save time and increase efficiency for safety practitioners in terms of incident analysis. If combined with big data technology, this method can automatically analyse TC information sources and provide the best evidence for road safety policy making.

The management factor of traffic safety defects was a common problem among the six TC cases. On the basis of the similarity analysis between cases 1 to 5 and case 6, traffic safety improvements could be carried out using the “6E” measures – engineering, education, enactment, enforcement, emergency, and evaluation – Gao et al. (2015), (Table 10). The aim is to improve road traffic safety practices in order to avoid the occurrence of similar collisions and to save lives.

It is worth noting that more specific and multifaceted indicators related to traffic safety should be further investigated in order to provide more comprehensive evidence for safety practices. In practical application, the analysis can be performed by changing indicator weights according to the actual situation. In order to enable more effective use of the similarity analysis method, more detailed information related to collision incidents should be collected. Due to the limits of manual computation, this study only selected six cases to compute the degree of similarity between the causal factors of TCs, which may mean that the measurement results are not accurate enough.

Table 10. “6E” measures for traffic safety improvement

<table>
<thead>
<tr>
<th>Perspectives</th>
<th>Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engineering</td>
<td><strong>establish an institution for the training and examination of drivers;</strong>&lt;br&gt;<strong>design automatic alarm devices for speeding or overload;</strong>&lt;br&gt;<strong>develop devices for monitoring drivers’ violation behaviour</strong></td>
</tr>
<tr>
<td>Education</td>
<td><strong>educate drivers to learn the laws and regulations related to traffic safety;</strong>&lt;br&gt;<strong>driving simulations to correct unsafe behaviour;</strong>&lt;br&gt;<strong>build up periodic training system.</strong></td>
</tr>
<tr>
<td>Enactment</td>
<td><strong>revise and perfect laws related to traffic safety;</strong>&lt;br&gt;<strong>establish qualification permit systems for freight drivers;</strong>&lt;br&gt;<strong>standardize the design of traffic safety facilities</strong></td>
</tr>
<tr>
<td>Enforcement</td>
<td><strong>strengthen monitoring in collision-prone sections;</strong>&lt;br&gt;<strong>increase penalties for traffic violations;</strong>&lt;br&gt;<strong>strengthen early warnings of poor environment</strong></td>
</tr>
<tr>
<td>Emergency</td>
<td><strong>formulate emergency schemes;</strong>&lt;br&gt;<strong>develop emergency supplies equipment;</strong>&lt;br&gt;<strong>establish “green passages” for emergencies.</strong></td>
</tr>
<tr>
<td>Evaluation</td>
<td><strong>evaluate degree of safety of road facilities;</strong>&lt;br&gt;<strong>construct electronic evaluation system for driving behaviour and safety education</strong></td>
</tr>
</tbody>
</table>
Conclusions

In order to use existing information resources on TCs, this study developed a similarity analysis method, the AHP-S method, for quantitatively analysing the degree of similarity between the causal factors of collisions that combines the advantages of the AHP and similarity theory. The core of the method is to identify the similar elements and similar units of the causal factors of TCs according to the criteria and sub-criteria. It provides evidence-based practice support for collision investigation and analysis.

Using six TC cases in China as examples, the degrees of similarity between the causal factors of TCs were calculated using the AHP-S method. This study took case 6 as the central study subject, and the calculation results showed that the degrees of similarity between cases 1, 2, 3, 4, and 5 and case 6 were 49.63, 41.68, 40.76, 46.11, and 33.98%, respectively. The ranking of the similarity of the five cases can be presented as follows: case 1 > case 4 > case 2 > case 3 > case 5; that is, compared to the other four cases, case 1 provides the best evidence available for the analysis of the causal factors of case 6. With a larger collision sample, the AHP-S method could find other high-similarity cases for analysing case 6. In doing so, it would supplement the available safety insights gained from the literature or reports.

Moreover, the results of the similarity analysis of collisions were further discussed and the “6E” (engineering, education, enactment, enforcement, emergency, and evaluation) traffic safety improvement measures were put forward. Overall, this study can help decision makers discover targeted measures for traffic risk prevention, and it has reference value for investigating similar evidence in other fields, such as mine safety, construction safety, and chemical safety, etc.

Acknowledgements

The authors would like to thank National Natural Science Foundation of China (Grant No 51534008) for supporting this research.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No 51534008).

Author contributions

Liangguo Kang was responsible for manuscript writing and editing, data analysis, content planning.

Shuli Zhang was responsible for literature search, data collection and analysis, content planning.

Chiao Wu was responsible for manuscript editing, content planning and research guidance.

Disclosure statement

Authors do not have any competing financial, professional, or personal interests from other parties.

References


