Journal of Computers Vol. 32 No. 4, 2021, pp. 252-262
doi:10.53106/199115992021083204020

# Research on Road Traffic Accident Judgment Method Based on D-S Evidence Theory

Naiyue Chen[1], Zhenjiang Zhang[2*], Jiawei Wang[2], Zihang Yu[2]

[1] School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China
nychen@bjtu.edu.cn

[2] School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China
{zhangzhenjiang, 18120131, 21120163}@bjtu.edu.cn

Received 1 June 2021; Revised 1 July 2021; Accepted 8 August 2021

Abstract. With the rapid development of road traffic and the frequent occurrence of traffic accidents, the accurate judgment and efficient handling of traffic accidents can effectively reduce the impact of accidents to reduce the burden on society. In the road traffic accident judgments, the division of responsibilities made by professionals such as traffic police and judges is usually the basis for the judgment. The language irregularities and inaccurate descriptions in traffic accident determination letters make it necessary for multidimensional professionals to make uniform decisions on traffic accident determination results. Meanwhile, due to the uncertainty of professionals in the judgment process, we propose a data fusion method to express and process the uncertain information in order to improve the accuracy of liability judgment. Considering the vagueness of judgments in road traffic accident judgments, we propose a road traffic accident judgment method based on evidence theory. We design a soft classification reliability model based on information entropy and a hard classification reliability model based on the accuracy of the classifier. The method we proposed can effectively solve the problem of classifier results and improve the accuracy of the liability determination of traffic accident documents.

Keywords: basic probability assignment, data fusion, responsibility for traffic accidents, text categorization

## 1 Introduction

At present, in the era of rapid development of road traffic, traffic accidents occur frequently, have a huge impact on our life and property safety and social order [1]. Because the parties concerned lack a basic understanding of the handling of traffic accidents, it may lead to the expansion of the accident and even cause secondary accidents [2]. Therefore, after a traffic accident occurs, the rapid handling of the traffic accident can prevent the accident from having a greater impact. At the same time, making accurate and fair liability judgments in a timely manner can well protect the rights and interests of both parties in the accident. In addition, the judgment of liability in a traffic accident is a rigorous issue. From the point of view of the judge, as the authority for the determination of liability for accidents, a rigorous and serious attitude is required to quickly make a fair judgment. In addition, in subsequent judgments, due to the complicated circumstances of the accident itself, especially for some accidents with more complicated environments and conditions, there are many influencing factors, and it is difficult to draw accurate conclusions only through manual judgments, or both parties to the accident There is a big disagreement on the accident. At this time, it is necessary to combine previous experience and make a judgment on the accident based on the previous handling results of the related accident. Therefore, the rapid handling and accurate judgment of traffic accidents are of great significance for reducing the impact of the accident and reducing the burden on society [3].

* Corresponding Author

This paper studies the reliability construction method of the classifier in the road traffic accident judgment. In order to reduce the labor cost in the judgment and improve the classification efficiency, we utilize classifiers to determine the party responsible for the accident by traffic accident documents.

The remainder of the paper is organized as follows. In section 2, we review existing approaches to multiple sources of evidence fusion. Section 3 presents the goal of this paper and our proposed method in detail. The experimental performance of this algorithm is discussed in section 4. Thus, the paper concludes in section 5.

## 2 Related Work

In the judgment of road traffic accidents, in order to obtain a more comprehensive judgment result, data fusion technology can be used to make a unified decision on the judgment results of the experts. The DS evidence theory is a data fusion method of multiple sources of evidence [4], which can express uncertain problems through a trust function, which is consistent with the uncertainty in the expert judgment results. Therefore, the uncertainty processing ability of the DS evidence theory can be used to Data fusion is performed on the judgment results of multiple experts [5]. However, in road traffic accident judgments, there are still some shortcomings in the application of evidence theory [6]. Compared with the judgment results of experts, when the classifier is used as the evidence source, the result obtained is a certain category label or membership probability [7]. In order to obtain reliability, uncertain information can be obtained from the classification results, and a reliability construction function can be established to map the classification results to reliability, so that the results of different classifiers can be fused in the form of reliability [8].

At present, with the development of machine learning technology, data fusion has begun to be used more for the fusion of classifiers [9]. Zhang [10] proposed a multi-classifier evidence body fusion rule and the classifier based on reliability. In this article, the evidence theory will be used to fuse the judgment results given by multiple experts. In D-S evidence theory, basic probability allocation is an important direction of its research. Basic probability allocation can also be called reliability construction. Literature [11] performs basic probability distribution according to the distribution of samples in the interval. First, calculate the approximate distribution of the training set, and define the distance according to the approximate distribution of the training set, and finally calculate the similarity according to the distance, and normalize the similarity to obtain the evidence body the basic probability distribution.

## 3 Traffic Accident Judgment Method based on Belief Construction

In order to use the classifier as the source of evidence to make judgments, this paper proposes a road traffic accident judgment method based on belief construction (TAJBC). The method can be divided into five layers, including text preprocessing layer, word embedding layer, classifier layer, reliability building layer, and fusion layer. We will focus on the reliability building layer and the fusion layer which are the critical sections in TAJBC.

The text preprocessing layer contains data cleaning, text segmentation and removing stop words. In data cleaning processing, we will clean meaningless characters such as punctuation marks, carriage returns, and line feeds, in addition to some unrelated letters and numbers that have nothing to do with the result of the judgment, such as the license plate number, time, road number. This method uses a regular matching method for data cleaning. We utilize jieba segmentation to make the word segmentation which can make each word has a relatively complete meaning. We set ‘Chinese Common Stop Word List’ as removing stop word standard which contains 748 stop words including commonly used prepositions, conjunctions, modal particles. On this basis, we add some words such as ‘year, month, day’ and ‘number plate’ are to the stop word list that appear very frequently in road traffic accident judgment documents but have no effect on the judgment result.

In the word embedding layer, we balance the relevance between the word and the classification result. We use the weighted word2vec to generate the word vector which consider the importance of each word. In the judgment documents of road traffic accidents, words such as ‘plaintiff’ and ‘defendant’ appearing in almost all texts appear frequently with little impact on the classification results. Although words such as ‘retrograde’ and ‘drunk’ appear less frequently, they can directly affect the classification results. We use normalized TF-IDF to weight word2vec. The specific process is shown in Table 1.

**Table 1.** Word vector generation algorithm by weighted word2vec

| **Algorithm.** Weighted word2vec word vector generation algorithm |
|---|
| **Input:** Road traffic accident lexicon after text preprocessing *dictionary* |
| **Output:** Weighted word vector matrix $Vector_{t-i}$ |
| 1. **for** *word* **in** *dictionary* |
| 2. word2vec training |
| 3. Get word vector $V$ |
| 4. **for** *word* **in** *dictionary* |
| 5. **if** *word* **in** $V$ |
| 6. $TF-IDF = TF \times IDF$ |
| 7. Weights $w = TF-IDF$ |
| 8. Get the weight matrix $W_{t-i}$ |
| 9. Maximum weight in $W_{\text{TF-IDF}}$ $max = max\{W_{t-i}[i]\}$ |
| 10. Normalized weight matrix $weight_{t-i} = {W_{t-i}}/{max}$ |
| 11. $Vector_{t-i} = V \times weight_{t-i}$ |
| 12. **return** $Vector_{t-i}$ |

In the weight calculation process, we determine whether the word is in the word vector matrix. For words that are not in the word vector matrix, the weight calculation is skipped directly to optimize the calculation process. For the words in the word vector matrix, the TF-IDF value is calculated as the weighted weight. The weight matrix of the word should be generated in the order of the word vector to ensure that the dimension of the weight matrix is the same as the dimension of the word vector. Meanwhile, each word vector also corresponds to the position of its weight value in the matrix. This allows the word vector matrix to be directly multiplied with the weight matrix and get meaningful results. The algorithm will normalize the weight matrix to avoid the phenomenon of too small difference in word vectors caused by too small individual weight values. For example, the weight of the high-frequency word “driving” is only 0.28, and the direct use of this weight will cause the word vector to be close to 0.

This algorithm uses a classifier instead of experts to conduct road traffic is a responsibility judgment, and the output of the classifier is the data source in the evidence fusion in this algorithm. In the judgment of road traffic accidents, the text classifier is suitable for various judgment document modes as the data source. The algorithm classifies the division of responsibility in the accident according to the description of the accident in the road traffic accident judgment document. We select the classifiers suitable for different feature data, which are the LSTM classifier, the Text-CNN classifier and the KNN classifier. LSTM classifier is a kind of soft classifier, which is suitable for the classifier of time series sequence. For text classification, it can better solve the influence of word order. Compared with the LSTM classifier, the Text-CNN classifier is also a soft classifier and ignores word order, which is more suitable for classification based on local features. The KNN classifier is a hard classifier that can extract more features in a sentence and has a higher classification accuracy. Through the traffic accident document, the algorithm will divide the judgment results into five categories according to the size of the responsibility of the defendant, and represent the defendant to assume full responsibility, primary responsibility, equal responsibility, secondary responsibility, and no responsibility.

### 3.1 Reliability Building Layer

The reliability building layer is the core layer of the model, and its main function combines the classifier with the D-S evidence theory. We propose corresponding optimized classification models based on the characteristics of different classifiers to more efficiently perform evidence fusion.

#### 3.1.1 Belief Construction Method Based on Information Entropy

For soft classifiers, we introduce the information entropy to construct reliability. The method will transform each classification result and determine the uncertain part of each classifier. First, we define the $i$ th membership probability of the classifier to the $j$ th focal element as $P_{ij}$. The weighted entropy is shown as $H(P)$

$$H\left(P\right)=-P*\log\left(P\right). \tag{1}$$

We analyze the monotonicity of the $H(P)$. The corresponding function expression is shown in the following formula (2), where the independent variable $x$ is the probability $p$ and its range is $x\in\left[0,1\right]$.

$$f\left(x\right)=-x*\log\left(x\right). \tag{2}$$

The function image is shown in Fig. 1. The abscissa is the independent variable and the ordinate is the function value. It can be seen from the Fig. 1 that the function has two monotonic intervals.

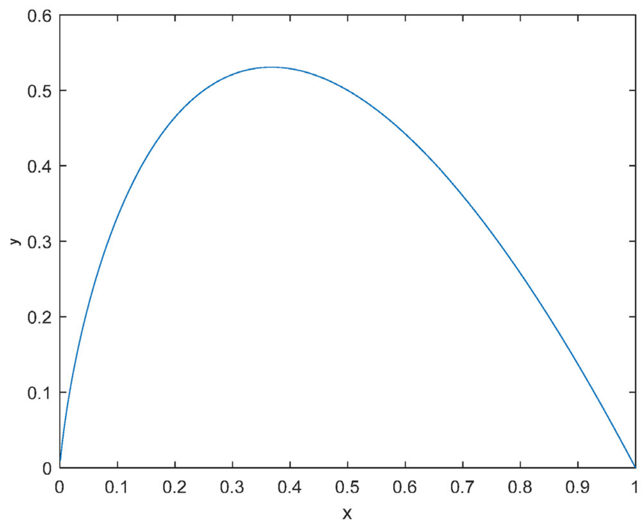


**Fig. 1.** The full life cycle of privacy information

We calculate the derivative of the function in formula (1) and analyze its monotonic interval in detail. The expression for obtaining its derivative is shown in formula (3).

$$f^{'}\left(x\right)=-\log\left(x\right)-1. \tag{3}$$

Combined with the trend of the image, the two monotonic intervals of the function are monotonic increasing intervals $\left[0,{}^{1}\!/_{e}\right]$, monotonic decreasing interval $\left[{}^{1}\!/_{e},1\right]$.

In order to ensure that a larger degree of membership corresponds to a larger information entropy, the reliability construction function should be monotonically increasing within its range. Therefore, it is necessary to ensure that the range of the independent variable $p_{ij}$ is in the monotonically increasing interval of the function.

it is necessary to compress the range of the independent variable caused by the overflowed of the function argument range. According to the above monotonic interval, the turning point of the increase or decrease interval is ${}^{1}\!/_{e}$. It's about 0.37, which is between one-third and one-fourth of the interval length. In order to ensure that the entire probability range is in the monotonically increasing interval, the entire interval length should be compressed to at least a quarter of the original length. The compressed probability $P_{ij}$ is shown in formula (4). In the following calculations, the probabilities used are all compressed probabilities.

$$P_{ij} = \frac{p_{ij}}{4}. \quad \textbf{(4)}$$

For each classifier's compression probability for each focal element, we calculate its weighted information entropy by formula (5).

$$H\left(X_{ij}\right) = -P_{ij} * \log\left(P_{ij}\right). \quad \textbf{(5)}$$

According to the probability given by each information source, the redundant information or uncertain information is calculated, and this part of information is removed from the weighted entropy as the uncertain part. In the calculation of joint probability, for the same source of evidence, if each classification result is regarded as an event, each event is independent of each other, and the joint probability of each source of evidence is the form of multiplication of each probability. Here it is still expressed in the form of weighted entropy, and the weighted joint entropy is shown in formula (6).

$$H\left(\prod_{j=1}^{m} X_{ij}\right) = -P_i * \log\left(P_i\right) = -\left(\prod_{j=1}^{m} P_{ij}\right) * \log\left(\prod_{j=1}^{m} P_{ij}\right). \quad \textbf{(6)}$$

According to the weighted information entropy of each focal element and the weighted joint entropy of each classifier, the non-redundant information entropy of each focal element is calculated, as shown in formula (7).

$$H^{'}\left(X_{ij}\right) = H\left(X_{ij}\right) - H\left(\prod_{j=1}^{m} X_{ij}\right). \quad \textbf{(7)}$$

It is solved on the first monotonically increasing interval, and the result obtained is the reliability of the construction by using the above analysis of the properties of the weighted entropy function. Since the function situation is complex and cannot be solved directly, we adopt an approximate solution method and use its monotonicity to continuously approximate the final result.

The final solution $m_{ij}$ is the reliability of the $i$-th classifier for the $j$-th focal element in equation (8). The function $f$ is the inverse function of the weighted entropy function, which is the solution function of the reliability.

$$m_i\left(A_j\right) = f\left(\left(\prod_{j=1}^{m} P_{ij}\right) * \log\left(\prod_{j=1}^{m} P_{ij}\right) - P_{ij} * \log\left(P_{ij}\right)\right). \quad \textbf{(8)}$$

The solving process of the function is shown in Table 2.

**Table 2.** Method of reliability solution

**Algorithm.** Reliability Solving Process

**enter:** Weighted entropy $H_w^{'}\left(X_{ij}\right)$

**Output:** Reliability $m_{ij}$

1. **for** $x = 0:0.01:0.25$
2. $y = -x\log_2\left(x\right)$
3. **if** $y \geq H_w^{'}\left(X_{ij}\right)$
4. $dis1 = y - H_w^{'}\left(X_{ij}\right)$
5. $dis2 = \left(x-0.01\right)\log_2(x-0.01) - H_w^{'}\left(X_{ij}\right)$
6. **if** $dis1 \geq dis2$
7. $m = x - 0.01$
8. **else**
9. $m = x$
10. **break**
11. $m_{ij} = m \times 4$
12. **return** $m_{ij}$

### 3.1.2 Belief Construction Based on Accuracy

The output of the KNN classifier is that the label value of the classification result belongs to the hard classifier.

For hard classifiers, the uncertainty in classification is mainly caused by the accuracy of the classifier. The accuracy rate $p_{acc}$ of each classifier indicates the probability of the classification result being accurate, and it can also be considered as the probability of the appearance of the true judgment result. In hard decision-making, the reliability is usually artificially assigned based on the parameters of the evidence source itself. This method is highly subjective and inaccurate. In the road traffic accident judgment, the accuracy of the classifier is obtained by extensive training, which can accurately reflect its classification performance and has a strong representativeness. Therefore, the classification accuracy can be used as the reliability of the certain part.

We regard the wrong judgment as uncertain information, the reliability formula $m_i\left(A_j\right)$ can represent multiple different classification results, where $s$ is the actual result and $j$ is the source of evidence.

$$m_i\left(A_j\right)=\begin{cases} p_{acc}, j=s \\ 0, j\neq s \end{cases}. \tag{9}$$

The reliability of the uncertain part is $m_i\left(\Omega\right)$.

$$m_i\left(\Omega\right)=1-p_{acc}. \tag{10}$$

In the process of evidence fusion, the TAJCES method will automatically allocate uncertain parts of information. At the same time, for the reliability of 0, the algorithm will use the conflict processing ability of the fusion formula to correct the reliability.

The reliability construction process is shown in Table 3. After reliability construction, the five classification results of the classifier are converted into five focal elements under the recognition framework, from $A_1$ to $A_5$ to represent the defendant to assume full responsibility, primary responsibility, equal responsibility, secondary responsibility, and non-accountability.

**Table 3.** Belief constructing method of classifier

**Algorithm.** Belief constructing method of classifier

**Input:** $p_{ij}$

**Output:** Reliability $M$

1. **if** soft classifiers
2. Calculate probability compression for $p_{ij}$
3. Calculate weighted entropy $H\left(X_{ij}\right)$
4. Calculate the weighted joint entropy as the uncertainty $H\left(X_i\right)$
5. Calculate non-redundant information entropy $H^{'}\left(X_{ij}\right)$
6. Calculate reliability $m\left(A_{ij}\right)$ by Table 1
7. **else if** soft classifiers
8. Calculate the accuracy of the classifier $P_{acc}$
9. Calculate reliability $m\left(A_{ij}\right)$
10. Calculate unknown part reliability $m\left(\Omega\right)$
11. **end**
12. **return** Reliability $M$

### 3.2 Fusion Layer

The TAJBC method uses the fusion layer to perform data fusion and decision-making on the reliability of the classification results. Since the TAJBC method is an open framework with strong compatibility, the classifier can be selected arbitrarily according to the data type and the required classification results. $m_i\left(A_j\right)$ is the reliability of the *i*-th classifier to the *j*-th focal element. The result is shown in forluma (11), where $c_i$ is the class of the classifier, 0 represents the soft classifier, and 1 represents the hard classifier.

$$m_i\left(A_j\right)=\begin{cases} f\left(\left(\prod_{j=1}^{m}\left(p_{ij}/4\right)\right)*\log\left(\prod_{j=1}^{m}\left(p_{ij}/4\right)\right)-\left(p_{ij}/4\right)*\log\left(p_{ij}/4\right)\right), c_i=0 \\ p_{acc}, c_i=1 \end{cases}. \tag{11}$$

The algorithm uses the Dempster combination formula to iteratively fuse, each time two evidence sources are selected for data fusion, and the combination formula is shown in formula (12).

$$\left(m_1 \oplus m_2\right)\left(A\right)=\frac{1}{K}\sum_{A_1 \cap A_2} m_1\left(A_1\right)m_2\left(A_2\right). \tag{12}$$

After multiple iterations, the mass function of each focal element obtained in the last iteration is used as the data fusion result. The algorithm is mainly based on the data fusion result to make a decision to obtain the final decision result. The decision rule is shown in formula (13), which can effectively reduce the risk of decision-making.

$$\begin{cases} m_i\left(A\right)-m_j\left(A\right)>\varepsilon_1 \\ m\left(\Omega\right)<\varepsilon_2 \\ m_i\left(A\right)>m\left(\Omega\right) \end{cases}. \tag{13}$$

where $m_i\left(A\right)$ is the final judgment result, and $m_j(A)$ is the mass function of the *j*-th focal element. The threshold $\varepsilon_1$ is the minimum difference to be satisfied between the mass functions and $\varepsilon_2$ means the maximum value of the uncertain part. The value of the threshold parameters are both 0.1 to ensure that there is sufficient discrimination between the reliability.

In the judgment of road traffic accident responsibility, the final judgment must meet the following three constraints. First, the finally selected $m_i\left(A\right)$ is the largest among all mass functions, and the mass difference with other focal elements is greater than the threshold $\varepsilon_1$. Secondly, the mass function of the unknown part is smaller than the threshold $\varepsilon_2$. Finally, the final decision result $m_i\left(A\right)$ is greater than the unknown part $m\left(\Omega\right)$. After decision-making, the final judgment result is the first focal element. The value of from 1 to 5 corresponds to the defendant's full responsibility, primary responsibility, equal responsibility, secondary responsibility, and no responsibility, respectively. The final judgment result is the *i*-th focal element, and its value of 1 to 5 corresponds to different responsibilities, which are that the defendant bears full responsibility, primary responsibility, equal responsibility, secondary responsibility, and no responsibility.

## 4 Experiments

### 4.1 Comparison of Word Vector Generation Algorithms

This section uses the TAJBC method to compare the word vector generated performance with word2vec algorithm and the weighted word2vec algorithm. The experiment selects Text-CNN classifier training the road traffic accident judgment document data. The data set includes 1,000 pieces of evidence, and there are five types of labels, which bear full responsibility, primary responsibility, equal responsibility, secondary responsibility, and no responsibility for the defendant. Each type of label has 200 pieces of data.

In the judgment process, the weighted word vector and the ordinary word vector are used to generate the word vector. The curve of the classification accuracy with the iteration round is shown in Fig. 2. In the Fig. 2 we can find with the number of training rounds increasing, the accuracy of the weighted word vector has been higher than that the ordinary word vector. The model uses weighted word vector classification until the tenth round of training approaches convergence.

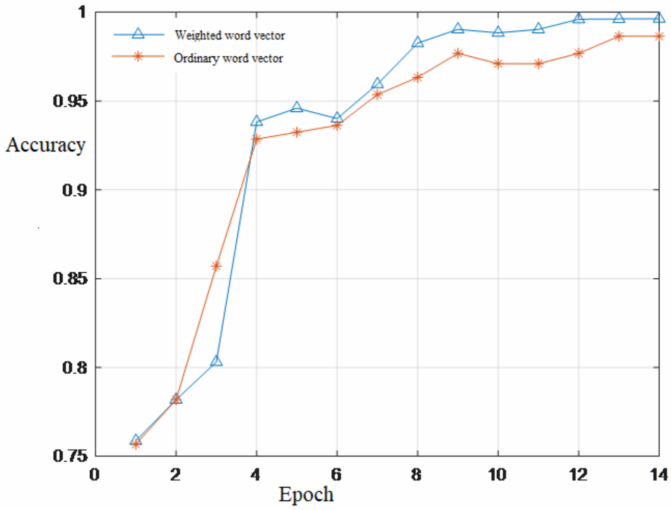


**Fig. 2.** Change of model accuracy

## 4.2 Reliability Experiments Based on Information Entropy

This experiment compares the reliability allocation algorithm based on information entropy with the method of directly fusing classification results. $A$, $B$, and $C$ are the three possible outcomes of the accident judgment. $A$ represents the defendant to assume primary responsibility, $B$ represents the defendant to assume equal responsibility, and $C$ represents the defendant to assume secondary responsibility. There are three classifiers as evidence sources, and the specific membership of each classification result is shown in the following formula (14).

$$\begin{aligned} &p_1(A)=0.6, p_2(A)=0.4, p_3(A)=0.5. \\ &p_1(B)=0.3, p_2(B)=0.3, p_3(B)=0.3.. \\ &p_1(C)=0.1, p_2(C)=0.3, p_3(C)=0.2. \end{aligned} \tag{14}$$

According to the evidence, the reliability obtained by the reliability construction algorithm based on information entropy is shown in formula (15).

$$\begin{aligned} &m_1(A)=0.48, m_2(A)=0.28, m_3(A)=0.36. \\ &m_1(B)=0.24, m_2(B)=0.2, m_3(B)=0.2. \quad . \\ &m_1(C)=0.04, m_2(C)=0.2, m_3(C)=0.14. \end{aligned} \tag{15}$$

Compared with the direct probability of the classifier, the reliability has different degrees of discount, and it is not a simple linear weighting relationship. For ordinary linear weighting, after the reliability is redistributed, the result will be exactly the same as the reliability without construction. Therefore, this non-linearity guarantees the availability of the construction results. And for each classifier, the sum of the reliability is less than 1, which is equivalent to introducing uncertain factors into the result of the classifier.

The TAJBC algorithm uses two different reliabilities for data fusion, and the fusion results obtained are shown in the following Table 4.

**Table 4.** Fusion result of reliability construction methods

| Fusion result | Information entropy reliability | Classification result |
|---|---|---|
| Focal Element *A* | 0.5167 | 0.5 |
| Focal Element *B* | 0.2823 | 0.3 |
| Focal Element *C* | 0.2010 | 0.2 |

It can be seen from the fusion results in Table 4 that when the classification results are directly used as the reliability for fusion, the classification of the classification results is not accurate enough caused by no analysis of the uncertain factors in the case. The reliability construction method based on information entropy can express the degree of membership of the classification result in the form of a trust interval, and the expression of the judgment result is more rigorous.

### 4.3 Road Traffic Accident Judgment Method based on Belief Construction Experiments

The experiment will compare the decision accuracy of the TAGBC method with other multi-classifiers. In the process of judging responsibility, the road traffic accident judgment document data set is divided into two parts, each time 80% of the data is randomly selected for training, and the remaining 20% is used for testing. We conduct each group of experiments 5 times and take the average accuracy rate as the final result, in order to reduce the impact of data randomness on the classification results.

We will compare the TAJBC method with other multi-classifier decision-making methods. The selected comparison methods are direct voting and Bayesian voting. Among them, the direct voting method uses the principle of the minority to obey the majority to make decisions on the classification label value, and the Bayesian voting method weights the label value with the accuracy of the classifier to make the decision. The comparison results are shown in Table 5.

**Table 5.** The influence of decision mode on classification effect

| Decision-making method | Judgment accuracy |
|---|---|
| Direct voting | 80.79% |
| Bayesian voting | 82.93% |
| TAJBC method | 89.62% |

In the Table 5, in the above three decision-making methods, although the direct voting method can make decisions on multiple classifiers, it has a limited improvement in classification accuracy compared with sub-classifiers. This is due to When the degree of support of each classification result is not much different, a small change may cause a large deviation in the classification result. When Bayesian voting is used for decision-making, the classification effect has a certain improvement compared with the direct voting method, but because the uncertain information in the data cannot be distinguished, the improvement effect is also relatively limited. When the TAJBC method is used to make a decision, the classification accuracy is obviously much higher than the other two decision-making methods. This is because the uncertainty and confidence interval are defined by the reliability, so that the two classification results that are very close can show better distinguishing ability under the action of multiple classifiers, and therefore have a better effect.

In Table 5 we can find although the direct voting method can make decisions on multiple classifiers, it has a limited improvement in classification accuracy compared with sub-classifiers. The reason is when the degree of support of a certain classifier for each classification result is not much different, a small change may cause a large deviation in the classification result. When Bayesian voting is used for decision-making, the classification effect is improved to a certain extent compared with the direct voting method, but the improvement effect is also relatively limited because it is impossible to distinguish the uncertain information in the data When using the TAJBC method to make a decision, the classification accuracy is obviously much higher than the other two decision-making methods. This is because the uncertainty and confidence interval are defined by the reliability, so that the two classification results that are very close can show better distinguishing ability under the action of multiple classifiers.

In order to verify the data fusion capability of the TAJBC method, the experimen is compared with other evidence combination methods. We use a variety of evidence fusion methods to make a fusion decision on the reliability construction result of a single classifier. The compared combination formulas include Dempster combination formula, Yager combination formula, Murphy combination formula, Sun combination formula and TAJBC method. The judgment accuracy rates of various methods are shown in Table 6.

**Table 6.** The influence of fusion method on classification effect

| Decision-making method | Accuracy |
|---|---|
| Dempster | 65.51% |
| Yager | 82.59% |
| Murphy | 78.27% |
| Sun | 83.44% |
| TAJBC | 89.62% |

In Table 6, we can find that the classification accuracy with Dempster is very low. Due to the different classification angles of the classifiers, the classification results are different. However, Dempster is weak in dealing with conflicting evidence. Therefore, the result of this conflict-free processing fusion method is relatively poor, even weaker than the decision result of using a single classifier. Although there is a certain improvement in conflict handling with Yager, Murphy and Sun, whether it is the Murphy combination formula that distributes the reliability evenly, or the Yager combination formula and the Sun combination formula that allocate conflicts to the unknown part, the effect of improving the single classifier is limited. Compared with other fusion decision methods, the TAJBC method has a higher classification accuracy. The TAJBC method conflicts the output of each classifier, and the defined discount factor can well represent the reliability of each classifier, and has the highest accuracy rate for accident judgment.

## 5 Conclusion

This paper focus on liability judgments based on road traffic accident documents. We studied the basic probability distribution of classifiers in road traffic accident judgments. We proposed road traffic accident judgment method based on belief construction utilized the text classifiers instead of experts to make decisions, which reduces the labor cost of judgments and ensures the objectivity of judgments. We built a reliability model to solve the problem that the results of the classifiers are difficult to merge directly. We define uncertain information according to the output of the classifier and establish the corresponding basic probability distribution function. We introduced mutual information entropy to define uncertain information, and used weighted information entropy to construct the reliability of the soft classifier. For a hard classifier, the accuracy of the classifier is taken as the definite information, and other results are taken as the uncertain information. Comparison experiments showed that this method has a better fusion effect than single-classifier decision and other multi-classifier decision methods.

## Acknowledgements

This work has been supported by National Key R&D Program of China (2018YFC0831900).

## References

[1] Y. He, Z. Liu, X. Zhou, B. Zhong, Analysis of Urban Traffic Accidents Features and Correlation with Traffic Congestion in Large-Scale Construction District, in: Proc. 2017 International Conference on Smart Grid and Electrical Automation (ICSGEA), 2017.

[2] W.-H. Zhang, H.-X. Deng, X.-B. Wang, Safety factor analysis for traffic accident scene based on computer simulation, in: Proc. 2010 International Conference on Intelligent Computation Technology and Automation, 2010.

[3] A. Fang, C. Qiu, L. Zhao, Y. Jin, Driver Risk Assessment Using Traffic Violation and Accident Data by Machine Learning Approaches, in: Proc. 2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE), 2018.

[4] W. Zhang, X.L. Ji, Y. Yang, J.W. Chen, Z.P. Gao, X.S. Qiu, Data Fusion Method Based on Improved D-S Evidence Theory, in: Proc. 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), 2018.

[5] C.P. Cheng, F.Y. Xiao, A New Distance Measure of Belief Function in Evidence Theory, IEEE Access (7)(2019) 68607-68617.

[6] Z. Deng, J.Y. Wang, Conflicting Evidence Combination Method Based on Evidence Distance and Belief Entropy, in: Proc. 2020 IEEE International Conference on Networking, Sensing and Control (ICNSC), 2020.

[7] Z.G. Liu, Y. Liu, J. Dezert, F. Cuzzolin, Evidence Combination Based on Credal Belief Redistribution for Pattern Classification, IEEE Transactions on Fuzzy Systems 28(4)(2020) 618-631.

[8] Q.F. Yan, Y. Zhou, L. Zou, Y.L. Li, Evidence Fusion Method Based on Evidence Trust and Exponential Weighting, in: Proc. 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2020.

[9] W.M. Lian, T.T. Xue, Y.Y. Lu, M.Y. Wang, W. Deng, Research on Hierarchical Data Fusion of Intelligent Medical Monitoring, IEEE Access 8(2020) 38355-38367.

[10] W. Zhang, X.L. Ji, Y. Yang, J.W. Chen, Z.P. Gao, X.S. Qiu, Data Fusion Method Based on Improved D-S Evidence Theory, in: Proc. 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), 2018.

[11] H.F. Chen, X. Wang, Determination of Basic Probability Assignment Based on Probability Distribution, in: Proc. 2020 39th Chinese Control Conference (CCC), 2020.