Journal of Computers Vol. 33 No. 2, April 2022, pp. 211-219
doi: 10.53106/199115992022043302019

Design and Implementation of Road Traffic Responsibility Identification System Based on Semantic Understanding and Similar Cases

Hongde Zhou, Zhenjiang Zhang*, Nayue Chen, Yinglong Li, Ruiheng Zhao

Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing100044, China
21120181@bjtu.edu.cn, zhangzhenjiang@bjtu.edu.cn, nychen@bjtu.edu.cn, 19111026@bjtu.edu.cn, 1692359975@qq.com

Received: 1 February 2022; Revised: 7 April 2022; Accepted: 15 April 2022

Abstract. The deep integration of deep learning and the judicial field has led to the continuous development of judicial intelligence. Judicial intelligence can not only assist judicial practitioners to improve their work efficiency, but also better serve the public and promote judicial convenience. Liability determination is the main basic task of judicial intelligence, but with the increase in the number of cases, staff will spend a lot of time dealing with similar cases, reducing work efficiency and consuming energy, so this paper combines natural language processing technology, Recommendation strategy and deep learning, research and implement a road traffic responsibility identification system based on semantic understanding and similar cases. To a certain extent, it can assist sentencing decision-making and standardize judgment standards. This paper firstly studies and analyzes related technologies, then identifies similar cases, uses triples to extract text keywords, and then uses gensim library and text2vec library to calculate text similarity, and then uses D-S evidence theory to compare the above two methods. The fusion of the similarity calculation results, and combined with the road traffic responsibility identification system for application, in which the main function of the evidence theory is to complete the similar case recommendation and similarity calculation through the case description. The establishment of a road traffic responsibility identification system has the important value of improving the predictability of judicial activities and realizing formal and substantive justice.

Keywords: judicial intelligence, similar case retrieval, responsibility determination, deep learning

1. Introduction

1.1 Research Background

An important purpose of today's traffic accident information analysis is to provide an objective basis for "responsibility determination". In the process of handling each case, the determination of responsibility is an indispensable key link. At present, the traditional manual liability determination method used in my country has different liability determination results in the same type of accident occurring in different regions due to the difference in the understanding of law enforcement by law enforcement officers [1], which affects the fairness of law enforcement. In order to change this situation, in recent years, some provinces and cities have organized technical forces to develop a computer responsibility identification system, trying to achieve the purpose of unifying law enforcement standards and standardizing law enforcement behavior through technical means [2]. Therefore, we need to weigh the pros and cons, avoid greed for perfection, and establish a practical and practical system to transform technology into productivity as soon as possible.

Due to the influence of various factors such as system, man-made, geographical and so on, there is a phenomenon of "different judgments for the same case" in judicial practice [3]. This not only affects the solemn status of the judiciary, but also affects the construction of my country's legal system and the rule of law system. Guiding Opinions on Unifying the Application of Laws and Strengthening Searches for Similar Cases" to meet the needs. But it does not actually solve the problem once and for all, nor does it combine with today's technology in a modern context to meet complex and changing situations. Therefore, this paper intends to build a retrieval system that can satisfy diverse needs while improving efficiency. At present, establishing a similar case retrieval system and making full use of it is a key step in unifying the standards for the application of law, and it can be fully inte-

* Corresponding Author

grated with the case guidance system to solve the problem of guiding cases being a mere formality. However, the traditional case retrieval database cannot meet the requirements of "similar cases and similar judgments" [4]. It is necessary to promote the deep integration of judicial theory, judicial needs and science and technology, establish an efficient and intelligent similar case retrieval system, and use technology to solve judicial problems and difficulties, to promote the modernization of my country's judicial system and judicial capacity.

Responsibility determination is an audit mechanism based on identity authentication and authorization management, and plays an important role in network system security. Aiming at the problems that should be solved in responsibility identification, a theoretical prototype and functional model of the responsibility identification platform are proposed [5]. Based on the data warehouse, the new generation of information analysis methods and computational intelligence technologies such as online analysis, data mining, and knowledge discovery are used.

There are related researches on intelligent justice at home and abroad. Abroad, Winkels et al [6] invented a legal recommendation system based on Dutch case law. In China, Chen Yanguang et al. [7] published more than 300,000 drug-related cases The knowledge map of drug-related cases was constructed by extracting entities from the Chinese Academy of Sciences; Huang Qianqian and Yang Jianlin take the field of "Internet fraud" as an example, and use a "top-down" approach to build a knowledge graph. The case domain ontology is constructed by combining the document content and expert opinions; then Neo4j is used to generate the case knowledge map. Finally, a smart judicial knowledge service framework based on knowledge graph is proposed [8]; Zhang Hu et al. [9] proposed a method based on multi-model fusion. Ye Jingjing [10] proposed a recommendation method combining text similarity and collaborative filtering technology. Collaborative filtering narrowed the range of alternative laws and solved the problem of excessive number of laws. Since a single recommendation method has its own advantages and disadvantages, combining different recommendation methods and fusing various features for hybrid recommendation can usually produce better recommendation performance. The hybrid recommendation method that integrates multiple heterogeneous auxiliary information can alleviate the problems of data sparse and cold start to a certain extent, but because auxiliary information often has complex characteristics such as multi-modality, data heterogeneity, large-scale, and data sparse, this Methods also face serious challenges. The existing methods for mixing different recommendation methods include weighting, stacking, transformation, feature combination, blending, feature enhancement, etc. Although the hybrid recommender system can alleviate the problem of data sparsity, a big disadvantage of this method is the representation of auxiliary data. Deep learning can learn effective user and item latent representations from auxiliary data information through automatic feature extraction.

Starting from the actual situation, this paper firstly analyzes the domestic and foreign research in related fields, as well as the realization methods and advantages and disadvantages of related research, so as to obtain some generally feasible methods. Then, the system design ideas are expounded. At the same time, combined with natural language processing technology, recommendation strategy and deep learning, a variety of libraries are used to calculate the similarity of texts, and a traffic responsibility identification system based on semantic understanding and similar cases is studied and implemented. Finally, it analyzes the advantages and disadvantages of the system, as well as what advantages it has compared with other similar systems, and how it can be developed in the future.

2. Related Works

2.1 Referral-Based Judicial Research

Recommendation refers to inferring information that users may be interested in through some algorithms, analyzing and processing specific data information through recommendation algorithms, and then recommending the processed results to users with relevant needs. It is a key link in the recommendation system.

The recommendation system can use the distributed link fusion method to build a data storage model, realize the analysis of the feature data in the multi-dimensional feature space, analyze the target feature distribution set based on this, and realize the intelligent data fusion processing according to the distribution characteristics of the association rules of big data [11], the traditional recommendation methods are mainly divided into three categories as shown in Fig. 1. Including: content-based recommendation, collaborative filtering recommendation and hybrid recommendation.

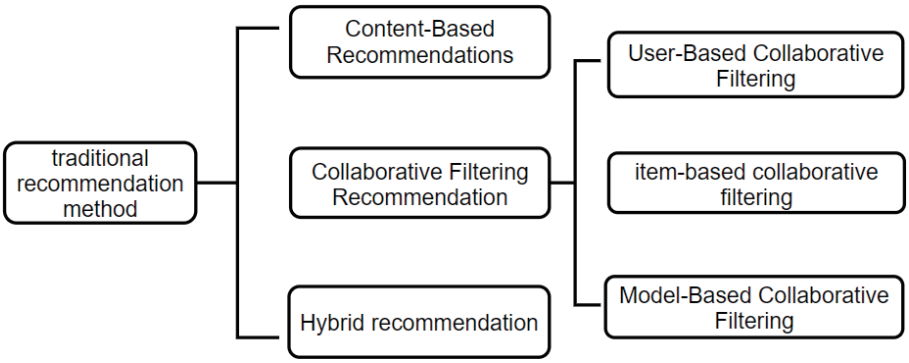

Fig. 1. Classification of traditional recommendation methods

The core of the collaborative filtering recommendation algorithm is to obtain the dependencies between users and items by analyzing the rating matrix (usually the user's rating of the item), and further predict the association between new users and items [12]. Collaborative filtering recommendation algorithm is one of the earliest recommended technologies to be studied and discussed. It effectively promotes the development of personalized recommendation, but it is more complicated to implement. Content-based recommendation methods need to discover user needs from known cases, and then make recommendations based on user interests and item characteristics. The advantage is that there is no cold start and data sparse problems; it can be explained by listing the content characteristics of the recommended results. At the same time, it also has shortcomings. It relies on the marked features of the item, and the feature content must be well structured: it cannot measure the quality of the recommended item.

2.2 Related Methods

Name Entity Recognition (NER) is the key to get entities in triples. NER is a basic task of natural language processing. The purpose is to identify named entities such as person names, place names, and organization names in the corpus. Information extraction is to extract structured data and specific relationships from unstructured text. Word segmentation is difficult to identify these important key words, and NER can play its power in such scenarios [13]. Traditional named entity recognition tasks use statistical-based machine learning methods to convert the task into sequence labeling or classification problems.

A relationship in a knowledge graph is defined as a certain connection between two or more entities. Entity relationship learning is to automatically detect and identify a semantic relationship between entities from text, also known as relationship extraction. Relation extraction The result is usually a triple (entity 1, relation, entity 2). The schematic diagram is shown in Fig. 2 below.

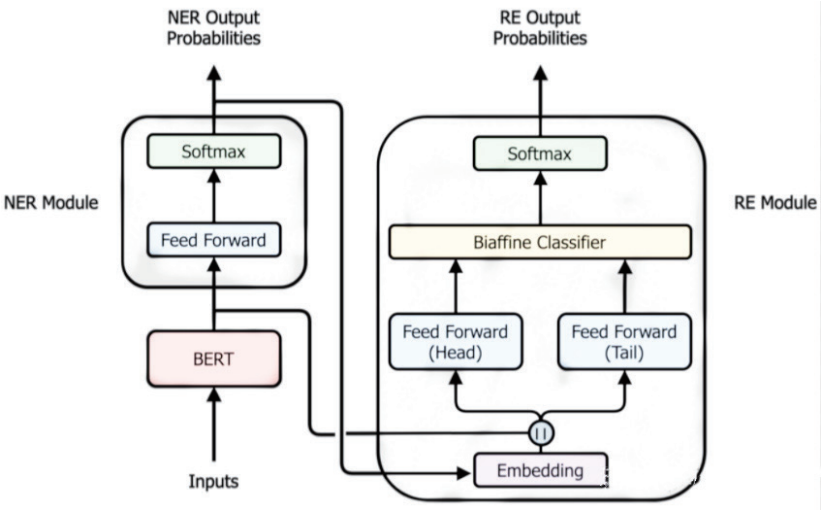

Fig. 2. Triad system

At present, typical entity relation extraction methods include rule-based relation extraction, machine learning-based relation extraction, etc. The rule-based relation extraction method first designs some lexical, syntactic and semantic features by experts who are proficient in linguistic knowledge according to the requirements of the extraction task. of hand-crafted rules (or patterns), and then look for instances that match these patterns in text analysis to deduce semantic relationships between entities. Relation extraction methods based on machine learning are roughly divided into three categories according to the different requirements of machine learning methods for corpora: unsupervised relation extraction, supervised relation extraction and weakly supervised relation extraction.

Responsibility determination is an audit mechanism based on identity authentication and authorization management. It determines the responsibilities of actors by recording, retaining, and auditing security-related events, so as to achieve the goal of verifiable network behavior, accountability for network incidents, and combating cybercrimes. Responsibility identification is to conduct in-depth mining and analysis of data to obtain as much useful information as possible, so as to provide a complete basis for responsibility identification to meet the requirements of a secure network environment that integrates responsibility, rights and interests. How to realize the responsibility identification mechanism has become one of the key issues to be solved in the process of building a network trust system.

Based on the above basic principles, combined with the data content of court judgments, the existing road traffic liability identification systems can be divided into two categories according to different research and development subjects: one is the construction type led by judicial organs, and the other type is the construction type led by civil subjects. Prosecutors can make sentencing recommendations based on search results, and judges can convict and send sentences based on search results. The sentencing auxiliary function can restrain the judge's behavior and overcome the disadvantages of sentencing, but it is necessary to be alert to the mechanical risks, opaque risks and unacceptable risks existing in the sentencing auxiliary system.

3. Design of Road Traffic Responsibility Identification System

3.1 System Design Thinking

First of all, our purpose is to improve work efficiency and reduce the burden of the staff. At the same time, the guiding ideology of the traffic responsibility identification system is to recommend similar cases and give the similarity through the case description input by the user, that is, natural language, so as to achieve the goal of assisting staff. For the purpose of decision-making and improving work efficiency, the road traffic responsibility identification system based on semantic understanding and similar cases can carry out structured processing of unstructured legal judgment documents. For users, this result can be used as an effective reference for case cognition.

3.2 Model Settings

Based on the above research background, the guiding ideology of our modeling is: put the case information analysis into the larger environment involved, and integrate multiple traffic management such as traffic accident handling, motor vehicle management, driver management, traffic violations, and traffic monitoring. For the data of the business system, a data warehouse is established with different analysis requirements as the theme; on this basis, the relevant theories and technologies of online analysis, data mining and expert systems are used to construct a system model of traffic responsibility identification based on semantic understanding and similar cases. The system structure designed according to the guiding ideology is shown in Fig. 3.

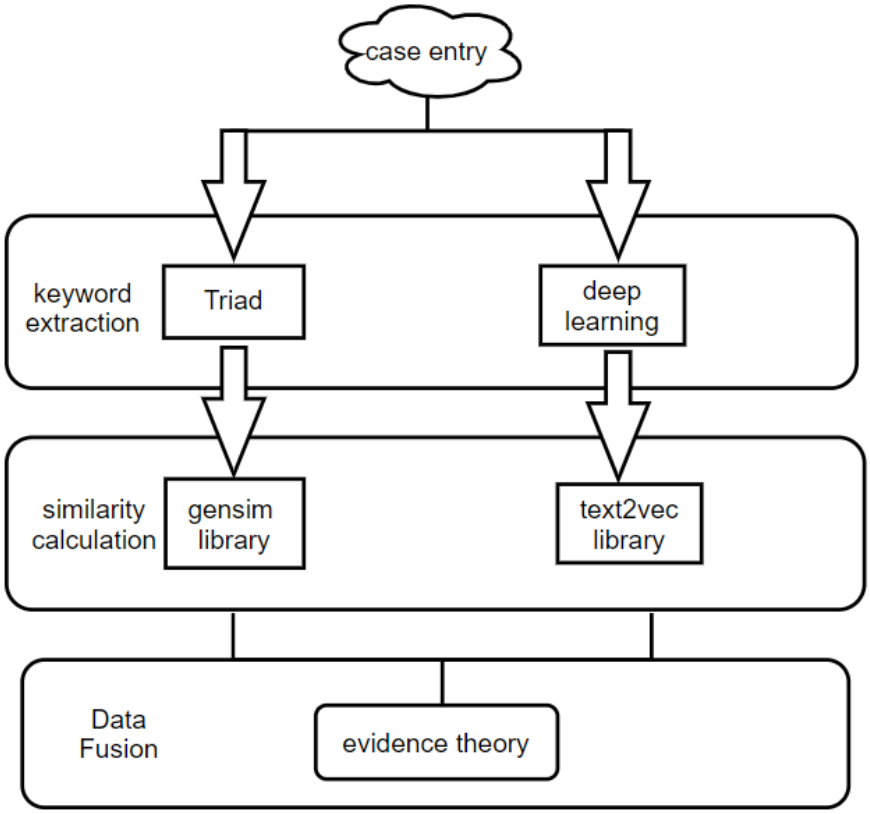

Fig. 3. System structure

The responsibility identification system in this paper is an audit mechanism based on identity authentication and authorization management. It determines the responsibility of the behavior subject by recording, retaining, and auditing security-related events, so as to achieve the goal of accountability for events, verifiable behavior, and crime fighting. Responsibility determination is to conduct in-depth mining and analysis of data to obtain as much useful information as possible, so as to provide a complete basis for determination of responsibility, and at the same time provide a reference for case judgment, so as to meet the requirements of a secure network environment that integrates responsibility, rights and interests.

The responsibility identification system needs to perform security processing and analysis on the application system logs collected in real time to check the behavior of system users. Therefore, the system should satisfy

1)Specify a unified log format and carrier: Due to the diversity of system event log forms and usually poor readability, this brings difficulties to log analysis. Therefore, specifying a unified log format and carrier can provide complete, clean, accurate, and more targeted log data for log analysis, and improve the efficiency and accuracy of log analysis.

2)Store past cases: The system should store cases in the database for comparison during case analysis and identification of similar cases, so as to calculate the similarity and provide reference for the staff to determine responsibility.

3)Guarantee the reliability of the database: Security logs and cases are the original basis for responsibility determination, and their integrity and reliability are the basic guarantee for correct and effective responsibility determination results. Timestamps are mainly used for tamper-proofing and post-repudiation of electronic documents to determine the exact time when electronic documents are generated. Stamping the time stamp when the security log is stored can effectively improve the reliability.

4)Case analysis: After obtaining the case data stored in the database, semantic understanding of the current case, similarity calculation and correlation analysis should be performed, and the premise set and result set should be supplemented for the case, so as to prepare knowledge for responsibility determination.

5)Responsibility determination for security incidents: The responsibility determination personnel can deter-

mine the responsibility for a specific security incident, that is, find the person responsible for the occurrence of the case, or the cause of the case.

In a word, the core idea of recommendation technology based on content filtering is: take the user's historical selection records or preference records as reference recommendations, and mine other unknown records that are highly relevant to the reference recommendation as the content recommended by the system. Obtain the user's interaction records within a certain period of time through the user's explicit feedback and implicit feedback (such as browsing time, number of clicks, number of searches, dwell time, etc.), and then learn the user's preferences in these records and mark them as features, Then, the similarity in content between the user's preference and the recommended object to be tested can be calculated, and finally the similarity of the recommended object to be tested and the user's preference can be sorted, so that the recommended object that conforms to the user's interest and preference can be selected for the user. Calculation of similarity is a key part, which directly affects the recommended strategy. There are many ways to calculate similarity. The following formula is commonly used to calculate similarity:

$$u(i, s) = score(user, item) \tag{1}$$

'u' represents the user, 'i' is the set of objects that the user can recommend, and the similarity score score has a variety of calculation methods, usually calculated by the distance of the cosine of the vector angle. in line with the userss preferences.

In practical applications, the most common application is to find other similar items as the recommendation result based on the item that the user is interested in. Among them, similarity can be defined according to feature attributes. Therefore, we can also use a similar content-based recommendation. The recommendation process of the entire system is shown in Fig. 4 below.

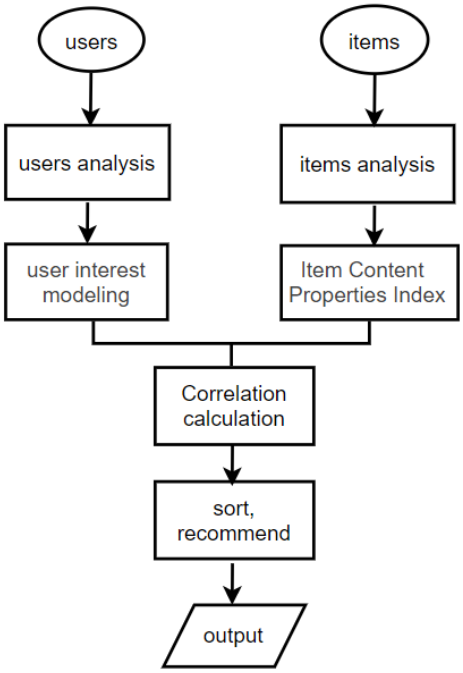

Fig. 4. System recommendation flow chart

Because content-based recommendation is mainly about information acquisition and information filtering. At the same time, the research on text information acquisition and filtering is relatively mature, so this paper decided to combine natural language processing technology to extract text keywords, use gensim library and text2vec library for similarity calculation, and then score and show it to users as a reference, and finally merge the results.

4. System Function Realization and Analysis

4.1 Development Environment

The road traffic responsibility identification system in this paper is implemented by the currently popular B/S architecture. The B/S (Browser/Server) architecture is mainly divided into front-end browsers and back-end servers. The server side is responsible for receiving HTTP from the front-end browser. Request and return the corresponding data processing result data, and the front-end browser is responsible for rendering the returned data and displaying it on the page. The main advantage of the B/S architecture is that it can be operated on any computer system without the need to install specific client software, just install a browser to run, so compared with other architectures such as C/S (Client/Server), its portability and compatibility are better. The overall operation process of the B/S system architecture is shown in Fig. 5.

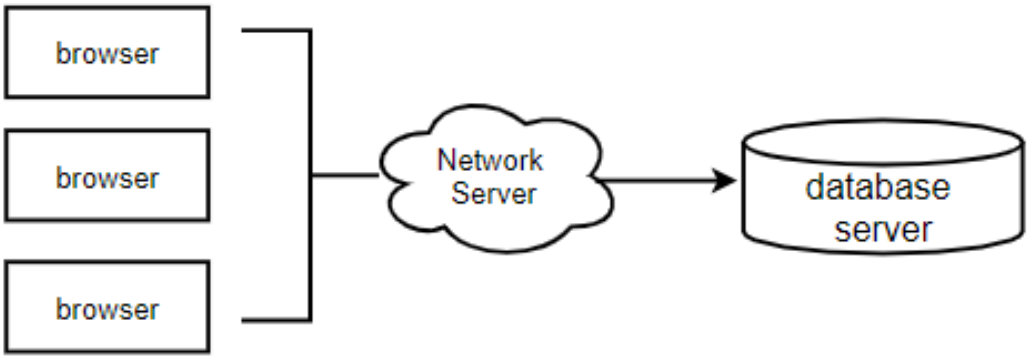

Fig. 5. B/S architecture flow chart

SQL Srever2000 adopts the user/server architecture, which can realize the cooperative operation between different database servers, and realize the creation, configuration and maintenance of the database.

4.2 Main functions and Implementation Methods of the System

The design of the system needs to consider data storage, background program processing, and system interface. Therefore, this system includes three architectural layers: data layer, logic layer, and presentation layer. The function of this system model is mainly divided into two modules: responsibility extraction module and responsibility identification module.

In the responsibility extraction module, users can enter case information, and the system will give judgments for reference.

In the responsibility identification module, similar cases can be identified. First, two methods are used to calculate the text similarity. One is to use triples to extract text keywords, and then use the gensim library to calculate the similarity. The other is to use the text2vec library to calculate the text. Similarity calculation, and then use evidence theory to fuse the similarity calculation results of the above two methods. Among them, D-S evidence theory can deal with the uncertainty of data, and it is widely used in the fields of information fusion, decision analysis and artificial intelligence. Finally, the content of the case, the similarity, the responsibility of the defendant and the relevant laws and regulations are displayed. Thereby assisting staff to determine the responsibility. Its function is shown in Fig. 6 below.

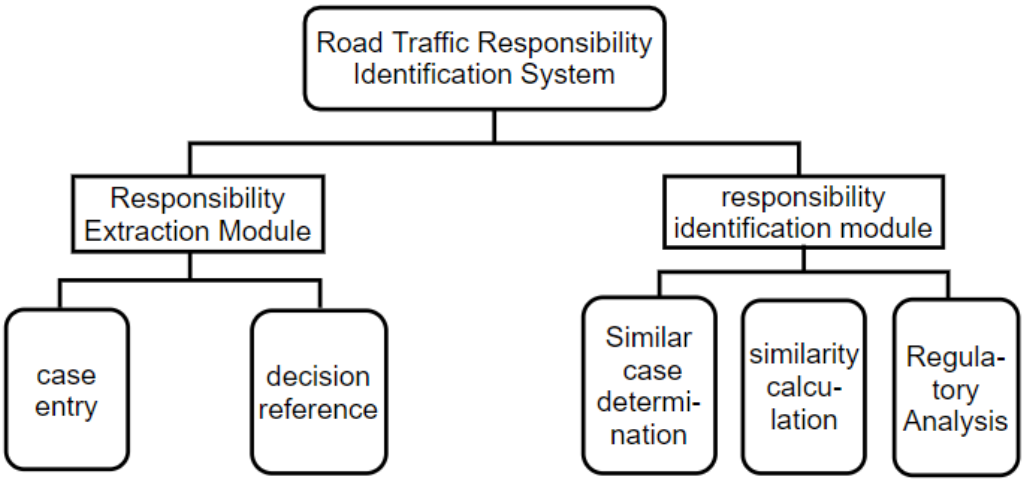

Fig. 6. System functions

The road traffic responsibility identification system based on semantic understanding and similar cases designed in this paper uses feature extraction based on triples, and then combines gensim library and text2vec library to obtain text similarity and finally performs information fusion. Compared with other responsibility identification systems, the recommended speed is faster. After many experiments, there is only a delay of about 1s, which can quickly return the results and improve the work efficiency of personnel. At the same time, the operation interface of the system is simple and easy to operate, and the similarity in the form of percentage can be visually checked, which is easy to operate. Moreover, compared with other time-stamp-based liability identification systems, the specific details of the case can be seen, and the liability identification can be carried out more effectively.

Recommending similar cases is one of the basic tasks of judicial intelligence. It can help people quickly get the relevant information about the cases they want, while saving time and improving work efficiency, as well as promoting judicial convenience for the public.

5. Summary and Outlook

Natural language processing has come a long way in the past two decades with the development of computer science. However, in the legal field where language wording is relatively standardized, natural language processing has not achieved many results [14]. At the same time, the existing traditional government system cannot meet the requirements of the times, and its data is complicated and cannot be effectively used [15].

In order to effectively prevent and reduce traffic accidents, and unify the scale of law enforcement to achieve “fairness, justice and openness”, this paper proposes a system model of road traffic liability identification based on semantic understanding and similar cases, and establishes a corresponding application system. The model adopts information intelligent analysis methods and decision support theory, and integrates the research results of knowledge engineering, artificial intelligence, data mining technology and other aspects to intelligently analyze traffic accident information, thereby providing an objective and scientific basis for dynamic traffic management [16].

This paper designs and implements the road traffic responsibility identification system. The system is based on semantic understanding and similar case recommendation. First, two methods are used to calculate text similarity. One is to use triples for text keyword extraction and then use gensim library for similarity calculation, and the other is to use text2vec library for text. Similarity calculation, and then use evidence theory to fuse the similarity calculation results of the above two methods. Through this platform, users can have an intuitive understanding and a deep understanding of the similarity of road crossing cases, and this technology can be used in many fields.

Acknowledgements

This work has been supported by the National Key Research and Development Program of China (Grant number 2018YFC0831900).

References

[1] J. Cao, Research on Judicial Cognizance of Civil Liability of "Wrongful Birth", [dissertation], Bengbu, Anhui Province: Anhui University of Finance and Economics, 2021.
[2] X. Wang, Discussion about Intellectual Property Protection and Antimonopoly Law Enforcement in Unified Market Regulation, Research on China Market Regulation (7)(2018) 14-18+2.
[3] Y. Chen, W. Rong, Ideal or Reality: The Intrinsic Cause and Optimization Path of Different Judgments in the Same Case, Journal of Qiqihar University (Philosophy and Social Science Edition) (2)(2022) 90-93.
[4] H. Xu, The applicable dilemma of adjudication in similar cases and its optimization path, Journal of Taiyuan Urban Vocational College (3)(2022) 181-183.
[5] M. Huang, A new interpretation of network service provider's tort liability determination——from the perspective of directly obtaining economic benefits, Journal of Chongqing Radio and Television University 34(1)(2022) 35-40.
[6] R. Winkels, A. Boer, B. Vredebregt, A.V. Someren, Towards a legal recommender system, in: Proc. International Conference on Legal Knowledge and Information Systems: JURIX 2014, 2014.
[7] Y.G. Chen, H.S. Liu, C.N. Li, J. Liu, Y.Y. Sun, Knowledge graph construction techniques based on criminal cases, Journal of Zhengzhou University (Natural Science Edition) 51(3)(2019) 85-90.
[8] R. Pandey, P.J. Brantingham, C.D. Uchida, G. Mohler, Building knowledge graphs of homicide investigation chronologies, in: Proc. 2020 International Conference on Data Mining Workshops (ICDMW), 2020.
[9] H. Zhang, X. Wang, C. Wang, H. Cheng, H. Tan, R. Li, Law Article Prediction Method for Legal Judgment Documents, Computer Science 46(9)(2019) 211-215.
[10]J.J. Ye, Research on intelligent statutes recommendation technology based on collaborative filtering and text relevance, [Master Dissertation], Nanjing: Nanjing University, 2019.
[11]J. Liu, K. Shi, Big Data Decomposition-Fusion and Its Intelligent Acquisition, Computer Science 47(6)(2020) 66-73.
[12]H. Zhang, D. Ge, S. Zhang, Hybrid recommendation system based on semantic interest community and trusted neighbors, Multimedia Tools and Applications 77(4)(2018) 4187-4202.
[13]Z. Chen, X. Liu, M. Zhang, D. Zhang, Named Entity Recognition Technology for Brief Cas, Computer System Application 31(1)(2022) 47-54.
[14]Z. Liu, Target feature data mining method for information intelligent recommendation system, Information Technology (3)(2022) 162-165+171.
[15]F. Jiang, Problem of "Many Cases but Few Judges" and the Changes of the Way of Governance, Tribune of Political Science and Law 33(2)(2015) 25-37.
[16]J. Long, J. Guo, Review and Prospect of Dynamic Traffic Assignment Research, Journal of Transportation Systems Engineering and Information 21(5)(2021) 125-138.