ISSN 2616-5775 Vol. 4, Issue 5: 85-90, DOI: 10.25236/AJCIS.2021.040512

Technical Demand Analysis of Judicial Big Data Application under Risky Driving Behavior of Drivers

Yan Fan[1,a], Yilan Wu[2,3,b]

[1]Institute for Advanced Studies in Humanities and Social Science, Chongqing University, Chongqing, China
[2]Criminal Law School, East China University of Political Science and Law, Shanghai, China
[3] Police Department, Anhui Vocational College of Police Officers, Hefei, China
[a]20172801001@cqu.edu.cn, [b]wuyilan@ahjgxy.edu.cn

Abstract: *According to the rapid development of the global economy and the continuous improvement of people's living standards, more and more people have a car driving license. We must pay attention to standardize the driving behavior of drivers, which is important to ensure personal safety. The survey also found that more than half of traffic accidents are related to dangerous driving behavior. However, in judicial judgment, there are many problems in judicial application which can not be solved because of the abstract provisions of laws in various countries. If the method of big data can be used as a reference for determining risky driving behaviors of drivers, it can provide evidence for identification, improve the accuracy of judicial judgment, and timely avoid some accidents. This paper aims to explore what application technical support is needed by judicial big data in determining risky driving behaviors of drivers.*

Keywords: *judicial big data, risky driving, application technology requirements*

1. Introduction

Dangerous driving refers to the driving behavior that the driver seriously violates the road traffic safety laws and regulations in the process of transportation, resulting in serious danger, including driving without a license, drunken driving, speeding, fatigue driving, driving the overload, taking drugs or taking of sedative drugs after driving, etc. How to prevent and reduce the occurrence of dangerous driving behavior so as to more effectively protect the safety of people's life and property is an important subject concerned by both theoretical and practical departments. By focusing on the unexpected data situation, judicial big data can predict the trend of possible cases, judge the inevitable trend of cases, and provide real-time data support for the court's trial execution, development situation research and judgment and assist scientific decision-making.

Judicial big data mainly refers to the combination of justice and big data, mainly supported by laws and regulations, relevant cases, judgment documents and other data, so a large amount of capital investment is needed in the early stage. It also needs the ability to analyze relevant data, such as win rate, factor weight, legal relationship and regulatory relationship when necessary. If judicatory big data is used well, it can improve a lot of convenience for case determination.

This paper mainly discusses the driver for dangerous driving situations can use judicial big data for evidence collection, determine the case and consider the realization of the four key technologies: (1) the relevant cases, laws and regulations of data collection (2) the data storage (3) the accurate positioning, video capture (4) the posture analysis, intelligent identification.

2. Application technology requirement analysis

2.1 Data collection

Before participating in a case, it is necessary to collect the data of laws and regulations, judgment conditions and previous cases in traffic. Too much data is collected artificially, so data collection technology is needed. Web crawler technology can be used to fetch data in the relevant judgment documents and other websites.

ISSN 2616-5775 Vol. 4, Issue 5: 85-90, DOI: 10.25236/AJCIS.2021.040512

Web crawler is an automatic web page extraction program, it downloads web pages from the world wide web for search engine, it is an important component of search engine. It can automatically read specific information in the network according to pre-modulated conditions. The traditional crawler starts from one or several URL of the initial web page and obtains the urls on the initial web page. In the process of crawling web pages, new URL are continuously extracted from the current page and put into the queue until certain stop conditions of the system are met. The work flow of focused crawler is complicated, and it is necessary to filter the links irrelevant to the topic according to certain webpage analysis algorithm, reserve the useful links and put them into the URL queue to be captured. Then, it will select the next web page URL from the queue according to a certain search strategy, and repeat the above process until reaching a certain condition of the system to stop. In addition, all crawler web pages will be stored by the system for certain analysis, filtering, and index establishment, so as to facilitate the subsequent query and retrieval; For focused crawler technology, the analysis results obtained from this process may also provide feedback and guidance for the subsequent grasping process. Compared with general web crawler, focused crawler only needs to crawl topic-related pages, which greatly saves hardware and network resources. The saved pages are also updated quickly due to the small number of pages, and it can well meet the needs of some specific people for information in specific fields. Focus crawler is needed to collect information about relevant laws and regulations. Compared with ordinary web crawlers, focused crawlers need to solve the following three main problems: First, description or definition of grasping target. Second, analysis and filtering of web pages or data; third, search strategy for URL.

The description and definition of fetching target is the basis of deciding how to make web page analysis algorithm and URL search strategy. The web page analysis algorithm and candidate URL sorting algorithm are the key to determine the service form provided by search engine and the crawler web page grabbing behavior. The algorithms of these two parts are closely related. It can be implemented by a system consisting of Linux +mysql+ Redis + Django +scrapy+ Webkit, where scrapy+ WebKit serves as the fetching end, Redis serves as the link library, mysql serves as the web page information storage, and Django serves as the crawler management interface.

Noun analysis:

Fetching ring: Fetching ring refers to the process by which a spider retrives a URL from a store, downloads a web page from the Internet, stores the web page in a database, and finally retrives the next URL from the store.

Linkbase: the storage module of a link library, containing general link information; Is the core of the crawl system, using Redis storage.

XPATH: A language for finding information in XML documents. XPATH is used for traversing elements and attributes in XML documents and is a major element of the W3C XSLT standard. Link extraction and information extraction are performed using XPATH and related tools lib.

XPathOnClick: a Chrome plugin that allows you to click on a page element to get an XPATH path for editing configuration templates.

Redis: an open source in-memory database for KV, with good data structure characteristics and high access performance. Used to store LinkBase information

Django: crawler management tool for template configuration and system monitoring feedback. The main purpose of Django is to manage a database, using the Admin function.

Pagebase: page library, mainly store the results of web crawling, as well as the results of page extraction, and dump interaction, using mysql implementation.

Scrapy: An open source opportunity twisted framework for python stand-alone crawler, this crawler actually contains most of the web scraping toolkit for crawlers on the download side as well as the extract side.

2.2 Data Storage

For a large amount of data to crawl down, we need to use the database to store. There are many existing storage frameworks available, such as traditional relational databases: Oracle, MySQL; Emerging NoSQL: HBase, Cassandra, Redis; Full-text retrieval framework: ES, Solr, etc.

MySQL: Relational database, mainly for OLTP, supports transaction, secondary index, SQL,

ISSN 2616-5775 Vol. 4, Issue 5: 85-90, DOI: 10.25236/AJCIS.2021.040512

master/slave, Group Replication schema model

HBase: Based on the Hadoop Distributed File System (HDFS), HBase supports massive data read and write (especially write) and column-oriented NoSql databases with hundreds of millions of rows and columns. Natural distributed, master-slave architecture, no transaction support, no secondary index support, no SQL support.

ElasticSearch: ES is a distributed full-text retrieval framework, and the bottom layer is implemented based on Lucene. Although ES also provides storage and retrieval functions, I never think ES is a database, but with the increasingly powerful functions of ES, the boundary between ES and database is becoming increasingly blurred.

For judicial large database, it is recommended to choose MySQL, which is the most mature, and supports transactions, and secondary indexes, and has the most mature disaster recovery backup scheme.At the same time, you can use Python for data processing, which is much more efficient than directly executing SQL statements in MySQL for data processing.

After establishing a good database, we also need to maintain the database and pay attention to the operation of LAN switches, while sending the log information of LAN switches to the total log server, the administrator needs to analyze and process the log files every day. Python's powerful database processing capability (combined with plug-ins) automates the tedious alarm log processing network operation and maintenance . In addition, the generated result can be named by the date of the day and fault type, greatly improving the efficiency of network operation and maintenance.

Based on the above, the basic data collection and analysis functions are completed.

2.3 Accurate positioning and camera capture

In order to find out whether a driver has dangerous driving behavior in time, the precise positioning camera capturing and monitoring system for a specific target must be equipped first. The specific target here is the driver, and precise positioning of the driver saves time. Video surveillance system can be seen everywhere in various fields. It can vividly convey the image information of the object we need to observe, so as to obtain the information we are looking for. In order to expand the monitoring range, the pylon can be used to carry the camera rotation, so that the camera can rotate different angles to obtain images in all directions. However, in the scenario of dangerous driving, we need to obtain the real-time position information of the driver to know the real time state. Therefore, accurate and fast acquisition of specific target location information has become an important requirement.

At present, the technologies that can be used for regional positioning mainly include CSS(Chirp Spread Spectrum, linear frequency modulation Spread Spectrum technology) positioning, GPS satellite positioning, Bluetooth positioning, WIFI network positioning, GPRS/CDMA mobile communication technology and so on. Among them, CSS technology has great advantages in wireless communication positioning. CSS technology is the use of Chirp signal transceiver detection to achieve ranging positioning, it through the Chirp signal pulse compression, so that its surge sharp time domain, energy becomes very concentrated, is very easy to be detected. At the same time, the Chirp signal itself in anti-multipath, anti-Doppler frequency shift and other aspects have excellent performance, so CSS technology in addition to the advantages of the traditional spread spectrum technology itself, it is also very good in anti-multi-frequency, anti-frequency bias ability, and low reflected power. It is precisely because of these unique advantages that CSS technology in recent years in the field of ranging and positioning has received more and more attention and application.

Chirp spread spectrum signal was first applied in the radar field, because it has the characteristics of long time, wide band and wide product, and is used to solve the problems of radar measuring range and range resolution. The ultra bandwidth signal based on Chirp spread spectrum technology has the advantages of transmitting power transmission distance longer than the traditional ultra bandwidth signal. In 2007, IEEE802.15.4A working group adopted CSS technology as one of the new standards to the physical layer of IEEE802.15.4A. At the same time, The German Nanotron company provided RFID solutions with CSS technology under this standard for communication positioning. Moreover, a large number of related products such as nanoPAN5375 are introduced, which makes the widespread application of CSS technology in communication positioning become a reality.

The CSS positioning system consists of a base station (BS), a mobile tag, and a location processor. Where, BS is fixed and its coordinates in the whole positioning network are known as reference points. Tag refers to a mobile Tag. During positioning, the base station and mobile Tag communicate through

ISSN 2616-5775 Vol. 4, Issue 5: 85-90, DOI: 10.25236/AJCIS.2021.040512

chirP signal, and the distance between mobile Tag and each base station is obtained according to the length of transmission time. Then these data are transmitted to the processor through the network node for calculation, and the coordinates of nodes in the positioning network are obtained according to the positioning algorithm to achieve positioning. Localization algorithm with TDOA arrive (lag) algorithm, it is to the TOA (doa) algorithm is improved, arrival time is different from the direct use of signal measuring the distance between target, but with multiple base stations receive the same signal to determine the difference of mobile station location, so the algorithm need not require clock synchronization between base station and measured label, Only clock synchronization between base stations is required, which has much lower technical requirements than TOA algorithm, reducing the difficulty and error of system implementation, and has been widely used in positioning systems. TDOA determines the position of the mobile tag through the hyperbola determined by the time difference between the two base stations receiving signals. Multiple sets of TDOA can provide the intersection of multiple sets of hyperbola, so as to obtain the target position. In view of the above realistic requirements and technical realizability, the idea can use the positioning camera monitoring system based on CSS technology to achieve real-time image information acquisition for drivers.

2.4 Posture analysis, intelligent identification

After accurately positioning and capturing the driver's image and location information, it is necessary to analyze the driver's posture to identify whether the driver has dangerous driving behavior. The analysis and understanding of human behavior has become one of the hotspots of research in recent years. In the development process of human behavior analysis and understanding, researchers have overcome many technical difficulties and formed some classic algorithms, but there are still many unsolved problems. From the perspective of the development trend of research, the research of human behavior analysis is developing from the use of single feature and single sensor to the use of multiple features and multiple sensors. Human pose estimation, as an important feature of human behavior recognition, is to conduct human behavior analysis. The foundation is one of the research directions that have attracted much attention in the field of human behavior analysis. Human body pose estimation refers to detecting the position of each part of the human body from the image and calculating its direction and scale information. Human behavior analysis is based on the analysis and understanding of the context of multiple frames of images, while human posture recognition is for processing a single frame of static images. The correct recognition of the posture information of multiple frames of continuous static images provides the possibility for the realization of correct behavior analysis and understanding. Therefore, the accuracy and real-time of human body posture estimation directly affect the accuracy and real-time of human behavior analysis, and ensuring real-time and accurate posture recognition is the basis for the next step of behavior analysis. Automatic radar recognition technology can be used for the analysis of the driver's behavior.

Radar automatic target Recognition (RATR) is an important field in radar research and application. High resolution range profile (HRRP), namely one-dimensional range image, is an important radar feature for real-time target recognition. HRRP represents the vector sum projected by the echo of the target scattering center in the direction of radar line of sight. It contains a wealth of target structure characteristic information, such as target size, scattering point distribution, etc., and is easy to obtain and process. However, there are still many difficulties in target recognition using HRRP, among which attitude sensitivity is the main factor affecting the performance of HRRP-RATR. When the high resolution radar transmits a signal with enough bandwidth, the target echo occupies multiple range units, and then the radar target can be simplified as a scattering point model. Since the peaks in HRRP are the vector synthesis of all scattering points in each range element, if the scattering points do not change over the range element, the relative position of each peak remains unchanged. The backscattering property of each scattering point in the target varies with the Angle of view between the target and radar, so the peak amplitude changes rapidly with the attitude. The above two reasons cause HRRP to be very sensitive to the change of the target attitude Angle, that is, the attitude sensitivity of HRRP. As a result, the performance of the key step feature extraction of HRRP-RATR is difficult to be guaranteed, and the driver's attitude cannot be accurately analyzed.

Since traditional methods cannot solve the HRRP pose sensitivity problem well, the deep learning technology, which has been successful in image recognition, has been widely applied to HRRP-RATR tasks, especially convolutional neural network (CNN). Deep network model differs from traditional shallow machine learning model in its deep network structure, large-scale network model parameters and big data conditions required for training model. The convolutional layer of deep CNN acts like a feature extractor, which can learn rich internal structural features from low level to high level.

ISSN 2616-5775 Vol. 4, Issue 5: 85-90, DOI: 10.25236/AJCIS.2021.040512

CNN is a feed-forward neural network. Its artificial neurons can respond to a part of the surrounding units in the coverage area and have excellent performance for large-scale image processing. The convolutional neural network is mainly composed of an input layer, a convolutional layer, a ReLU layer, a pooling layer, and a fully connected layer. In practical applications, the convolutional layer and the ReLU layer are often collectively referred to as the convolutional layer. CNN stacks multiple convolutional layers, which not only has the advantages of traditional neural networks, such as good error tolerance, adaptability and self-learning ability, but also has the ability of automatic feature extraction and weight sharing. CNN consists of 10 convolutional layers, each followed by a nonlinear activation layer, four maximum pooling layers, two full connection layers of 1024 units, and one softmax output layer of full connection. The input is pre-processed HRRP, and the dimension of each HRRP is 1 x 1024. The output is an approximate posterior probability distribution of the target type. One dimensional convolution kernel is adopted which is different from ordinary image recognition. A large number of successful cases in the field of image recognition show CNN's powerful ability in feature extraction and representation learning. Therefore, it is assumed that CNN can be used to analyze driver behavior, and the database can be improved, updated and enriched through systematic deep learning.

3. Conclusion

This paper proposes four kinds of existing technologies, and conceivably constructs a system based on judicial big data to detect and discriminate whether a driver is driving dangerously with the support of these four technologies, including the use of web crawler technology to collect, and build a database to store huge amounts of data(including the use of python to process and analyze the database and log files), through the positioning of the camera based on CSS technology monitoring system to monitor the driver's posture and through the method of HRRP - RATR based on CNN is used to analyze and judge drivers' behaviors and improve the database through deep learning and then reduce the cost of manual intervention.

This is just the realization technology of some basic functions that I have envisaged and proposed, but I believe that under continuous research, the system will become more and more powerful, the driver’s dangerous driving behavior will become less and less, and the judicial big data tool will also Better application in various fields of justice.

References

[1] Jiayun, W. (2020). Judicial Big Data and the Realization of Judicial Justice. Journal of Jishou University (Social Sciences Edition), 41(2), 137.
[2] Devins, C., Felin, T., Kauffman, S., & Koppl, R. (2017). The law and big data. Cornell JL & Public Policy, 27, 357.
[3] Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. BCL Rev., 55, 93.
[4] Lyon, D. (2014). Surveillance, Snowden, and big data: Capacities, consequences, critique. Big data & society, 1(2), 2053951714541861.
[5] Medvedeva, M., Vols, M., & Wieling, M. (2018). Judicial decisions of the European Court of Human Rights: Looking into the crystal ball. In Proceedings of the Conference on Empirical Legal Studies.
[6] Botsman, R. (2017). Big data meets Big Brother as China moves to rate its citizens. Wired UK, 21, 1-11.
[7] Oviedo-Trespalacios, O., Nandavar, S., Watson, B., Lewis, I., Haque, S., White, K., ... & O'Donnell, K. (2019, September). Benchmarking distracted driving against other key risky driving behaviours. In 2019 Australasian Road Safety Conference.
[8] Zimbardo, P. G., Keough, K. A., & Boyd, J. N. (1997). Present time perspective as a predictor of risky driving. Personality and individual differences, 23(6), 1007-1023.
[9] Schwebel, D. C., Ball, K. K., Severson, J., Barton, B. K., Rizzo, M., & Viamonte, S. M. (2007). Individual difference factors in risky driving among older adults. Journal of safety research, 38(5), 501-509.
[10] Lamba, H. S., & Dubey, S. K. (2015, September). Analysis of requirements for big data adoption to maximize IT business value. In 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions) (pp. 1-6). IEEE.

ISSN 2616-5775 Vol. 4, Issue 5: 85-90, DOI: 10.25236/AJCIS.2021.040512

[11] Burmester, G., Ma, H., Steinmetz, D., & Hartmannn, S. (2018). Big data and data analytics in aviation. In Advances in Aeronautical Informatics (pp. 55-65). Springer, Cham.

[12] Liu, Y. (2018, January). Big data technology and its analysis of application in urban intelligent transportation system. In 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS) (pp. 17-19). IEEE.

[13] Hutchins, M. G., McGrane, S. J., Miller, J. D., Hagen-Zanker, A., Kjeldsen, T. R., Dadson, S. J., & Rowland, C. S. (2017). Integrated modeling in urban hydrology: reviewing the role of monitoring technology in overcoming the issue of 'big data'requirements. Wiley Interdisciplinary Reviews: Water, 4(1), e1177.

[14] Kohlwey, E., Sussman, A., Trost, J., & Maurer, A. (2011, July). Leveraging the cloud for big data biometrics: Meeting the performance requirements of the next generation biometric systems. In 2011 IEEE World Congress on Services (pp. 597-601). IEEE.

[15] Boehm, B. (2000). Requirements that handle IKIWISI, COTS, and rapid change. Computer, 33(7), 99-102.

[16] Cui, B., Mei, H., & Ooi, B. C. (2014). Big data: the driver for innovation in databases. National Science Review, 1(1), 27-30.

[17] Terzi, R., Sagiroglu, S., & Demirezen, M. U. (2018). Big data perspective for driver/driving behavior. IEEE Intelligent Transportation Systems Magazine, 12(2), 20-35.