



Research on the Driving Force of Civil Legal Aid Willingness Under the Discrimination of Big Data—Based on the Analysis of Economic Case Texts

Xiu Ye¹, Huihui Dong¹(✉), Pengxiang Jiang¹, and Yanling Li²

¹ Law School, Guangzhou Xinhua University, Baiyun, Guangzhou, China
392863499@qq.com

² ScEquity Investment Division, Dongguan Private Investment Group Co., Ltd., Nan Chen, Dongguan, China

Abstract. Individuals are subject to three types of big data discrimination: biased selection, false association, and malicious recommendation, and victims should be relieved through civil legal aid. This paper analyzes the text of network cases and constructs a regression model, and finds that there are two types of attitude words in the case text, explicit and implicit, which reflect the true attitude of the willingness to seek civil legal aid and have a significant impact. Through the texts of existing network cases, it has a good inspiration for the construction of a legal aid relief system for China's big data discrimination cases.

Keywords: Big Data Discrimination · Willingness Economy Of Civil Legal Aid · Network Case · Text Analysis

1 Introduction

The Civil Code of the Peoples Republic of China, which will be implemented from January 1, 2021, has made it clear that personal information is a citizen's right and is protected by law. On November 1, 2021, the "Personal Information Protection Law of the People's Republic of China" was implemented to protect the rights and interests of personal information and regulate the processing of personal information by means of special legislation. In the context of the rise of e-commerce, most of personal information is used within the scope of civil laws such as economics and society, but the phenomenon of using big data algorithms to discriminate against individuals has also emerged. When individuals are discriminated against by big data, they maintain their rights through various means, such as complaints, community comments, news media exposure, and legal assistance. However, in modern society, citizens' legal aid is obviously an effective way to protect their legal rights. This paper uses Internet cases and relevant comment texts as research materials, combined with text analysis methods, to analyze the economic benefits of legal aid willingness in China's big data discrimination [1].

2 Concept and Theory Overview

2.1 The Concept of Big Data Discrimination

Prejudice-based selective big data discrimination violates information distributive justice. Big data discrimination can be divided into three main types: biased selection, false association, and malicious recommendation. Based on their own biases, data users set classification standards that do not comply with legal and ethical constraints on the collected individual data, resulting in individuals being at a disadvantage when the system operates under these standards [1]. Through the writing of data processing programs, the data user associates multiple data of the data subject using an algorithm and deduces the characteristics of the data subject on this basis, and thereby classifies the data subject. Discrimination in big data by false associations is the creation of false relationships among multiple pieces of information of data subjects by forcibly and unnecessary association of some data through programs, [2], resulting in the creation of false relationships between data subjects in the later stages of the system in data retrieval operations and service provision. Malicious recommendation of big data discrimination is clearly stipulated in Article 48 of the “Cyber Security Law”, that is, “any electronic information sent by any individual or organization or application software provided shall not be set up with malicious programs, and shall not contain legal, administrative or Regulations prohibit the release or transmission of information”. Malicious recommendation big data discrimination mainly occurs between different information platforms or subsystems under the platform.

2.2 Discrimination Against Big Data in the Civil Legal System

Section 4 of the “Cyber Security Law” stipulates relevant provisions for network information security, but bias-based selective big data discrimination does not violate legal provisions such as information collection and information storage, nor does it violate Article 47, the specific terms of information release and transmission in Article 48. Article 43 of the “Cyber Security Law” stipulates: Individuals have the right to request network operators to delete their personal information. Section 2 of the Personal Information Protection Law, a special chapter stipulates the rules for the processing of personal information. For sensitive information, it is required to meet “specific purposes”, “sufficient necessity”, “individual consent”, etc. Malicious recommendation of big data discrimination is easier to distinguish and detect, and it is also the content that is currently monitored by administrative, and judicial departments. There are also many restrictions in laws and regulations.

2.3 The Concept and Provisions of Civil Legal Aid

The promulgation of the “Regulations on Legal Aid” in 2003 marked the institutionalization of legal aid on the right track. The establishment of the civil legal aid system reflects the justice and superiority of my country’s socialist system on the one hand, and the responsibility of the Chinese government to protect the rights of citizens on the other hand. By sorting out the Civil Code, Cybersecurity Law and other laws, the main

provisions on the rights of civil legal aid subjects are as follows: First, the subject's ownership and right to use information are clarified, although the intangibility of information leads to certain definitions. The rights include data use, [3], data use platforms and institutions, and structured explanations of data being used by algorithms. Finally, the subject's restrictions on data use and the right to withdraw must be stipulated.

2.4 Web Text Analysis

Ahmed M, Chen Q, & Li ZH (2020) pointed out that web text analysis can obtain people's opinions, opinions, attitudes and emotions by analyzing the text, audio and images precipitated on the Internet, which are usually divided into documents, sentences and key Word three-level sentiment analysis, document and sentence analysis can determine the sentiment tendency of documents and sentences, and keyword analysis is a more fine-grained sentiment classification for objects or entities in the corpus [4]. There are a large number of big data discrimination cases on the Internet. For case texts, researchers can obtain data in the interactive communication process based on the case texts, and then understand the reviewers' views and evaluations on case events and their willingness to take measures to safeguard their rights. The portrait of the commenter's behavior is used to calculate the commenter's demand for safeguarding rights and means of preference, so as to obtain the real willingness and economy of individual civil legal aid under the discrimination of big data.

3 Design of the Study

3.1 Sample Selection and Data Sources

This article collects the texts of big data discrimination cases from multiple platforms such as China Legal Service Network, Guangdong Legal Aid Center, Legal Online, Zhihu, etc., including the case facts and interactive texts of typical cases of big data discrimination in the three years from 2019 to 2021. The interactive text content includes: analysis texts of lawyers and professionals, comments and messages on the online platform, dialogue texts between legal professionals and individuals on the case platform, and analysis of viewers.

Based on special dictionaries such as the Chinese general dictionary Jieba and the HowNet dictionary, the content of the collected text is segmented, and stop words are removed, so as to convert the unstructured text data into word vectors for storage. [4].

Calculate the word frequency of the word set corresponding to the economic indicator of willingness to seek legal aid in big data discrimination cases. Obtain the results of legal aid for relevant cases. Exclude data from non-legal aid channels in cases. Deal with some missing values.

3.2 Economic Measurement of Willingness to Seek Civil Legal Aid Under Big Data Discrimination

Pennebaker et al. (2003) pointed out that human traits and attitudes are captured by analyzing word types and word frequencies used in the subjects' language [5]. In this paper, through the collection of comment keywords in big data discrimination cases, it represents the economic strength of the willingness to seek civil legal aid. It is also possible to evaluate the willingness and economy of civil legal aid through the means by which parties and legal professionals reach the ultimate protection of rights.

Based on this, this paper uses Word2Vec machine learning to formulate a Chinese word set that can reflect the economy of civil legal aid willingness based on the characteristics of the existing Chinese corpus. The individuals below use the attitudes, viewpoints and other elements of civil legal aid [5].

In this paper, the CBOW model in Word2Vec is used to train the collected text corpus of "big data discrimination cases". Based on the dictionary method, this paper calculates the proportion of the total word frequency of "civil legal aid willingness economy" to the total word frequency. The larger the value of this indicator is, the stronger the willingness to seek civil legal aid is, and vice versa. The calculation formula of the CBOW model is as follows:

The specific index measurement process is:

$$\max \sum_{w \in c} \log(f(w/context(w))) \quad (1)$$

Collect the evaluation data below the case text, such as the attitude of "the parties use legal means to protect their rights", using a 1–5 incremental evaluation, the higher the score, the more willingness to seek civil legal aid the higher the economy. By collecting the bottom of the text, it expresses support for likes by legal means. The calculation formula is: (number of likes/total number of comments) * 100. The number of clicks and forwarding of the case text. Big data discrimination includes three types: biased selection, false association, and malicious recommendation. Considering that different types of definitions involve the analysis of many legal terms, supervised learning is used to label the research cases. Combining existing legal provisions and legal professional analysis, this paper uses the category labeling method for big data discrimination cases.

3.3 Model Design

By setting the economic indicators of the willingness to seek civil legal aid as the dependent variable, and setting the factors that affect the economy of the willingness to seek civil legal aid as independent variables, a measurement model constructed. The variables are described in Table 1.

The respective variables together constitute each dimension of the dependent variable "economics of willingness to seek civil legal aid". The regression model is constructed through the collected data, and the degree of influence of each influencing factor on the dependent variable is analyzed.

Table 1. Indicators for each variable.

Independent variable X	Dominant word set indicator measure. Prosecute (x1); seek arbitration (x2); report to the police (x3); lawyer assistance (x4); file a case and record (x5)
	Invisible word set indicator measure. Number of likes for legal rights protection and support (x6); text forwarding (x7); link clicks (x8)
Dependent variable Y: Willingness to seek civil legal aid economics.	

4 MODEL RESULTS AND ANALYSIS

4.1 Descriptive Statistical Analysis

As shown in Table 2, through the processing of case texts, a total of 219 big data discriminatory cases were observed. In the text segmentation, explicit and implicit words are used, and the attitudes of the parties and reviewers are obtained respectively according to the corresponding relationship of the independent variables. In addition to supporting the use of relevant laws to safeguard rights, there were also those who opposed or failed to express their attitudes. Number of people. Through the keyword method of text analysis, the number of people who support, oppose, and have unclear attitudes is obtained as the measurement data of each indicator.

Secondly, according to the marked results, 219 big data discriminatory cases are classified, and the specific results are shown in Table 3. A lower number of single discrimination was presented in 219 cases. 1 case of biased selection discrimination, 1 case of false association discrimination, and 3 cases of malicious recommendation discrimination. All three types of discrimination occurred in the same case in 149 cases. In the same case, the total number of two types of discrimination was 65, 21 cases of prejudice selection and malicious recommendation discrimination, and 30 cases of false association and malicious recommendation discrimination. In Table 4, Sue means x1, Seek arbitration x2, Call the police deal with x3, Lawyer assist x4, File a case filing x5, Rights protection likes x6, Text forward x7, Link click x8.

4.2 Correlation Analysis

As shown in Table 4, the explicit and latent indexes of “willingness to seek civil legal aid” are analyzed by correlation coefficients, and there is a correlation between the variables, which are significant at the levels of 0.01 and 0.05, respectively. *Analysis of Model Results.*

By constructing the model: Willingness to seek civil legal aid economy

$$(\Psi) = a_1 * x_1 + a_2 * x_2 + a_3 * x_3 + a_4 * x_4 + a_5 * x_5 + a_6 * x_6 + a_7 * x_7 + a_8 * x_8 + b \tag{2}$$

The independent variables constitute various dimensions of the dependent variable. Through the estimation of the parameters of the regression model, the economic impact

Table 2. Quantity distribution after segmentation.

Classification name		Category tag	Number of categories
Prejudicial selection discrimination		1	1
False association discrimination		2	1
Hybrid discrimination		3	3
Prejudice selection	False association discrimination	4	14
	Malicious recommendation discrimination	5	21
	False association, malicious recommendation discrimination	6	30
	Total factor discrimination (three types of discrimination exist at the same time)	7	149

Table 3. Types and numbers of big data discriminatory cases.

Variable name	Umber of observations	Support attitude	Opposition	Attitude is not clear
Take prosecution(x1)	219	144	21	54
Seek Arbitration (x2)	219	173	35	11
Alarm handling (x3)	219	188	16	25
Legal assistance (x4)	219	127	80	12
Filing and filing (x5)	219	56	47	116
Legally defending rights and supporting acts Likes (x6)	219	201	8	10
Text Forwarding(x7)	219	109	24	86
Link Click (x8)	219	117	41	61

Table 4. Correlation coefficient matr.

Matters	X1	X2	X3	X4	X5	X6	X7	X8
X1	1	.316**	.270**	.266**	.407**	.331**	.129*	.347*
X2		1	.430**	.319**	.236*	.301**	.237**	.385**
X3			1	.226*	.168**	.411**	.339**	.239**
X4				1	.322*	.387**	.209**	.530*
X5					1	.237**	.371**	.288**
X6						1	.526*	.522*
X7							1	.427*
X8								1

** .Indicates a significance level of 0.01

*.Indicates a significance level of 0.05

Table 5. Regression Model Results.

Matters	Standardization coefficient	T	Salienc
constant	-1.388	4.291	0.003
Take Prosecution (x1)	0.395	0.804	0.023
Seek Arbitration (x2)	1.47	0.034	0.043
Alarm handling (x3)	0.905	0.197	0.041
Legal assistance (x4)	2.711	0.369	0.014
Filing and filing (x5)	0.047	0.239	0.012
Number of likes for legal rights protection and support (x6)	3.073	0.197	0.007
Text Forwarding (x7)	2.008	0.369	0.032
Link Click (x8)	1.094	0.239	0.04

on the willingness to seek civil legal aid is analyzed. The results of the regression model analysis with the help of SPSS 20.0 are shown in Table 5.

Through the regression model, it can be found that “the number of likes for legal rights protection and support behavior” has the greatest impact on “the economy of willingness to seek civil legal aid”, with a coefficient of 3.073; followed by “lawyer assistance”, with a coefficient of 2.711; third is “text forwarding”, with a coefficient of 2.008; the least influential is “taking prosecution measures” with a coefficient of 0.395. Moreover, seeking the support of others can strengthen the willingness to seek civil legal aid in the economy, for example, through the forwarding of case texts, community support plays a significant role.

5 Conclusions and Future Prospects

Prejudice selection and big data discrimination based on false associations have great concealment. Therefore, in the process of discrimination against big data, civil aid laws should not only aim at the objectivity and acquisition channels of data, but also include the supervision of data algorithms. In order to strengthen the willingness to seek civil legal aid, the economy should be regulated from the following aspects.

5.1 Strengthening the Provisions on the Rights of Subjects in Civil Legal Aid for Individuals Who Suffer from Big Data Discrimination

The rights of subjects can be regulated by combing the Civil Code, the Personal Information Protection Law, the Cyber security Law, and many laws in the Economic Law and Civil Law. The subject uses the information part. Attributable to the subject, government regulatory agencies can use the definable attributable to the regulatory agency and the subject, and third-party use of information belongs to the subject; finally, the subject's data collection, storage, and reuse rights should be attached to the subject's voluntary approval, and other institutions can regulate it. Used within the range.

5.2 The Scope of the Object is Clear the Scope of Data in Civil Legal Aid Must be Clear

First, the content involved in the data must not contravene the scope of the current laws; secondly, the form of data collection and storage must be clear; thirdly, the data processing procedures must be explained, and the purpose of the algorithm program to generate data processing and use must be specified, use, and assessment of consequences; finally, the data includes not only the initial information, but also the content of the reproduced data after the data is processed by the algorithm program.

5.3 Establishment of a Resident Supervision Mechanism for the Data Platform by Regulatory Agencies

Prejudice selection, false association, and malicious recommendation are three big data discrimination behaviors, which are implemented by the data platform in a dynamic process and change with the development of business environment, market demand and information technology. Therefore, if the legal aid legislation provides regulators with the right to resident supervision of data platforms, and designs a resident supervision mechanism in the process of practice, then big data discrimination will be stopped before and during the event, reducing the social impact after the event. Operating costs.

Acknowledgements. This paper is sponsored by Guangdong Provincial Department of Education 2017 Youth Innovative Talents Project (Humanities and Social Sciences) "Research on the Legal Mechanism of Personal Information Use and Protection in the Age of Big Data" (2017WQNCX189).

References

1. Ahmed. M and Chen. Q, Li. Z.H. (2020), Constructing Domain-Dependent Sentiment Dictionary for Sentiment Analysis. *Neural Computing and Applications*”, Vol.32(18), 14719–14732.
2. Liu. Y. (2020), Government Responsibility in Civil Legal Aid: A Comparative Perspective between China and the United States, *Academic Exchange*, Vol.12, pp79–90.
3. Mike. A. (2015), Toward an Ethicx of Algorithms: Convening, Observation, Pribability, and Timeliness, *Science Technology & Human Values*, Vol.41, NO.1, pp93–117.
4. Pennebaker. J.W., Mehl.M.R, and Niederhoffer. K.G.(2003), “Psychological Aspects of Natural Language Use: Our Words, Our Selves”, *Annual Review of Psychology*, Vol.54, pp547–577.
5. See.H.M. (2018), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Berkeley J.Emp & Lab.I, pp285–289.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

