Analysis and Application of Judicial Data Based on Machine Learning

Juan Yang
(Sichuan Judicial and Police Officers Professional College, Deyang Sichuan, 618000)
2014326215@dlvtc.edu.cn

Abstract—At this stage, judicial trial work has received increasing attention from the people and society, and its rationality is regarded as one of the important evaluation criteria for the fairness and effectiveness of the national judicial organs. This article analyzes the methods and types of machine learning. The author studies the relevant content of the machine learning judicial data process. The purpose of this article is to improve the reliability of the analysis and forecast results and provide a theoretical basis for the further improvement of the national legal system.

1. INTRODUCTION
Traditional Chinese judicial circles have problems such as cumbersome documents, numerous regulations, complicated case information, heavy burdens on judges, and low efficiency in judgments. In the context of artificial intelligence technology, people can use the advantages of machine learning such as fast speed, high accuracy, and strong continuity to learn from judicial data and build corresponding predictive management models. This also provides a reliable data reference for the trial of judicial cases and improves the credibility of the trial results.

2. METHODS AND TYPES OF MACHINE LEARNING

2.1. Main Learning Method

2.1.1. Principal Component Analysis
Principal component analysis is a method based on statistics. Its core is to transform a group of variables that may have linear correlations into another group with the same information through orthogonal transformation. However, for variables with uncorrelated lines, the group of variables obtained after conversion is called principal components. The principal component analysis method can clean up the overlapping part of the characteristic variables during the use process, and establish a smaller number of new variable groups. Moreover, this can also make these variables uncorrelated with each other without discarding information. In specific applications, this method combines multiple features in a high-dimensional space into several irrelevant principal components with lower dimensionality through orthogonalization, and uses variance to measure the change in information. One of the vectors in the high-dimensional space is selected as the first principal component, and the sum of the distances from all points to the determined regression line is required to be the smallest. According to the same method, the second principal component and the third principal component can be selected...
in sequence. At this time, the linear combination formed is \( Z_1 = \Phi_{11}X_1 + \Phi_{21}X_2 + \ldots + \Phi_{p1}X_p \). Subsequently, we can use the regression model for processing to obtain the required application data.

2.1.2. Partial Least Squares
Partial least squares method is often used to find the basic relationship between two different matrices, using the hidden variable method of modeling the covariance in these two spaces. Denote \( X = T^P + E \), \( Y = U^Q + F \). Among them, \( X \) represents an \( m \times n \) dimension eigenvariable matrix, \( Y \) represents another \( m \times p \) dimension response matrix, \( T \) and \( U \) are \( m \)-dimensional vectors, which are projections of \( X \) and \( Y \), respectively. \( P \) and \( Q \) are \( n \)-dimensional and \( p \)-dimensional orthogonal load vectors, respectively, \( E \) and \( F \) are error term matrices, and conform to the normal distribution of independent and identical distribution. Subsequently, when we process the regression equation, we can get the corresponding regression coefficient vector, and thus get the optimal solution needed.

2.1.3. K Proximity Algorithm
The K-nearest algorithm is the simplest classification algorithm. Assuming that the characteristic variable \( X \) of each sample data is an \( n \)-dimensional vector, the following equation relationship exists for \( X_i \): \( X_i = (x_{i1}, x_{i2}, \ldots, x_{in}), i = 1, 2, 3, \ldots n \). In the established \( n \)-dimensional space, the system can find a certain coordinate according to the algorithm to guide other coordinate content. We can place all the known sample data in the training set in the same \( n \)-dimensional space. Subsequently, the researcher will also enter other acquired sample data into the space. At this time, the staff needs to calculate the Euclidean distance for a sample in the space to ensure the accuracy of the data information in the training set. Moreover, in the data calculation process, the staff can also find the number of coordinates closest to \( X_i \), denoted as \( k \), and the highest label value obtained from the \( k \) values is the predicted value in its application stage.

2.2. Learning Type

As shown in Figure 1, the process of mechanical learning mainly involves known data analysis, learning and training, and establishing application models. If a sufficient amount of basic data is not provided in the learning process, the machine learning algorithm can find many application hypotheses from the set. If the content is assumed to have the corresponding accuracy, the staff can use the set algorithm to effectively deal with this contradiction, so as to make it more close to the real application objective function. The common types in specific learning are as follows.

2.2.1. Artificial Neural Networks
As shown in Figure 2, the artificial neural network is generally composed of an input layer, \( n \) hidden layers, and an output layer in the application process. Among them, the input layer is used for the import and basic sorting of the data to be learned. The hidden layer is used to capture the nonlinear relationship between variables, which is also the core part of the algorithm. The output layer is the application output for learning data to obtain reliable data analysis results.
2.2.2. Deep Neural Network Algorithm
In the process of machine learning, deep neural network algorithms are also commonly used algorithms. In the application process, the most complex structure in the data is screened and processed by abstract processing from the lower layer layer by layer. At the same time, it can also obtain effective application features from many sample data, which is convenient for subsequent analysis and processing work to proceed smoothly. When learning a deep neural network algorithm, it can be understood as a system structure containing a large number of neural networks, and the structure contains a large number of neuron structures. Meanwhile, neurons will maintain a high correlation with other neurons. The use of the corresponding weights of different neurons to modify the application functions in the network is conducive to completing the study of corresponding data information.

2.2.3. Swarm Intelligence Optimization Algorithm
The swarm intelligence optimization algorithm refers to obtaining the required solution by means of the swarm mode that simulates the activities of organisms, and the individuals in the solution have certain wisdom and experience. The algorithm can be combined into powerful group intelligence by means of interaction, and use this to complete the smooth solution of complex problems. Judging from the current usage, this method also has sub-categories such as particle swarm optimization, ant colony optimization, and longhorn beard search algorithm. It can be used in conjunction with other algorithms to improve the accuracy of the algorithm results. It should be noted that when using this algorithm to complete related data information processing, it is necessary to understand that the algorithm has relatively high requirements for the total number of parameters and the quality of the parameters. Otherwise, the staff also need to prepare the parameter information in the actual processing link to improve the reliability of the parameter calculation results.

3. MACHINE LEARNING JUDICIAL DATA PROCESS ANALYSIS

3.1. System Requirements Analysis

3.1.1. Functional Requirements
Judging from the current implementation of judicial and legal affairs, the total amount of judicial documents to be compiled by judicial workers every day is huge. They not only need to find similar cases from massive data, but also need to conduct in-depth research in conjunction with the country's newly revised legal provisions to meet people's judicial needs. At this stage, lawyer consultation fees are relatively expensive. In many cases, people will choose to look up existing cases as a reference to determine further work behavior. Based on this, how to provide high-quality case content has become a basic condition for better serving the public. In the learning process, the machine will not only complete case screening and sorting in the early stage, but also mark keywords for the collected cases. This is convenient for people to provide more convenient conditions for user case search.
3.1.2. Non-functional Requirements
The non-functional requirements of machine learning are mainly reflected in the following two parts. Firstly, time characteristics. It can control the time interval within 0.5 milliseconds during the page jump process when the user makes a request. This can ensure the validity of the data jump process. Moreover, users can use the model established by machine learning to control the response time within a reasonable range when making a request. In this way, foot-and-mouth disease ensures the timeliness and reliability of the information content response process. Secondly, it needs to have a certain load capacity in practical applications. That is, it needs to satisfy several users for remote access and provide users with better service content.

3.1.3. System Use Case Design

As shown in Figure 3, in the system use case design process, the following points are mainly included. Firstly, perform remote text input. It mainly refers to the fact that the user completes the case description through the interface for subsequent case interaction operations. Secondly, remote file upload. This is also the content of the information interaction between the system and the user. Returning the data with incorrect format can ensure the unity of the information transmitted. Thirdly, analysis and storage of judgment documents. It not only analyzes documents uploaded to the system, but also organizes analysis suggestions. Subsequently, the system will store these analysis suggestions together in the system to facilitate subsequent data access. Fourth, obtain screening cases. After simple processing and screening, similar cases can be obtained. Subsequently, the system can choose to download and view or analyze directly, providing users with more choices.

3.2. Case Screening System Design

3.2.1. System Logic Design
In the process of establishing the case screening system, the logical design of the system is a basic work content. The upload module and input module mentioned in the above article are examples. The two types of modules are file acquisition means in use, and there is a strong complementarity between the two. If the user's content is relatively simple when searching, he can select the text input method to obtain information. If you need a lot of narrative content, you can choose the file upload method to edit the content. The two modules can be logically selected according to the actual situation of the text during use, which can improve the use value of the analysis results.
3.2.2. System Development and Design

In the process of system development and design, the Python language is mainly used to complete the content design, and the content mentioned in the above modules is used to establish the development framework. This can improve the reliability of the analysis results. Many categories such as Http Response, File Upload, Case Analyse, etc. will be used in the specific development process. Take the Http Response class as an example. Its main function is to smoothly use the basic functions in the web communication system to ensure the reliability of the data analysis results.

3.3. Text Input and Transmission Module Design

As shown in Figure 4, the above process needs to be followed to complete the comprehensive processing of the corresponding content during the application of the text input module. The corresponding functional features will also be completed with the help of custom views in the Django framework system. In the specific application process, the user will return to the corresponding interactive form after making a request. The system will judge its submission status, and after judging that it has been submitted, the system will store the form content. Moreover, the system will directly transfer the relevant content to the corresponding text function for unified arrangement, thus completing an interactive application process.

3.4. Analysis Module Design

3.4.1. Judgment Document Analysis Module

In the process of machine learning judicial data, a very important work content is to organize the judgment documents to facilitate the follow-up to provide users with more objective suggestions. In the design process of this module, it is necessary to sort out the uploaded analysis data, and filter out the key information with value. Under normal circumstances, the length of the judgment document is relatively long, and the time and cost required to read it directly is higher. In this regard, when designing the application of the module, the staff will use the advantages of machine learning to extract key words from them, instead of traditional manual search for related words. This can play a role in reducing time costs [1].

3.4.2. Key Information Extraction Module

When extracting key information, the content of the information that needs to be extracted includes the specific type of the document, the court of the judgment of the document, the legal clauses involved, the summary of the case, and the sentenced crimes. The system needs to store the collected key information in a limited set for value. Regarding this, among the judgment documents to be processed, the system needs to treat it as a collection of words or phrases after word segmentation. This also has a strong correlation with the interactivity between information, thus forming a data set that is easy to
filter. For some cases with more complicated plots, the system can establish multiple separate subsets to be reflected in a union. This helps to improve the reliability of the analysis results [2].

3.5 Case Screening Module Design

3.5.1 Data Preprocessing
A high-quality judicial database needs to contain complete judicial factual data, and these data and information also need to be marked in advance. In the meantime, when valid data is obtained, the system can form reliable training data to help users solve some application problems. In the process of data preprocessing, the first task is to ensure the integrity and adequacy of data information. Simultaneously, the system also needs to do a good job in the comprehensive processing of the "accusation" and "imprisonment" fields. Take the "accusation" field as an example, which records the crime and the course of the case. It will not only use different hidden fields to complete data processing, but also improve the integrity and reliability of the processed field content [3].

3.5.2 Model Analysis and Processing
Relying on the machine learning situation to build an application model helps to assist the smooth development of other tasks. In the specific analysis process, we should pay attention to the following points. Firstly, we need to sort out the content of the training sub-modules. Judging from the current application situation, the training model belongs to the further arrangement of pre-processed data. It divides the existing data content into applications and obtains the corresponding processing results [4]. Secondly, the application of the prediction sub-module. This module organizes the collected text data with the help of big data. At the same time, it will also optimize and organize related content to obtain corresponding organized data to improve the efficiency of data analysis results.

3.5.3 Result Display Module
In the design process of the result display module, the following points should also be noted. Firstly, the screening results are displayed. In order to improve the convenience of displaying content, the system can only display the title of the case and match some keywords during the specific display process. This facilitates the user's application selection. Secondly, optimize the layout of the display interface. For example, the system can enter current text information on the left side of the interface, while the right side is used for case lists. This also facilitates the smooth acquisition of other data information, thereby obtaining more accurate data screening results, and enhancing the use value of data screening content [5].

4 Application Case Analysis of Machine Learning Judicial Data

4.1 Coping with the Test Environment
When testing the established machine learning system, we need to establish a corresponding test environment. The functions to be tested include the following aspects. Firstly, test the information interaction function of the system to see if it can successfully complete the file content transmission. Secondly, analyze the judgment documents to determine whether the system can accurately analyze the corresponding analysis results. Thirdly, test the case screening function. After submitting the application, check whether a similar case can be found in a short time. Fourth, test the download function of the system to check the integrity of the downloaded content [6].

4.2 Test Process and Results
Complete the test work in sequence according to the above content, from which the corresponding test results can be obtained. The specific content is as follows. Firstly, after uploading, storage space is added in the system. Moreover, the system also displays the corresponding folder, indicating that the upload function is reasonable. Secondly, when analyzing judgment documents, the system will enter
multiple existing cases. The system will make statistics on its judgment, and the accuracy rate is 85.3%, which needs to be further improved. Thirdly, similar cases can be obtained after entering the request. Its fault tolerance rate is below 10%. Moreover, the data integrity meets the requirements after downloading, and the system has promotion value [7].

5. **CONCLUSION**

To sum up, using the advantages of machine learning to apply it to judicial data analysis has a positive effect on speeding up data analysis and improving the accuracy of data analysis results.

**REFERENCES**

[1] Zhao Dawei, Li Wenhua. Research on the application of artificial intelligence technology in the bond industry [J/OL]. Finance and Economics, 2020(12): 86-90[2021-01-07].

[2] Song Hongqing, Du Shuyi, Zhou Yuanchun, Wang Yuhe, Wang Jiulong. Big data intelligent platform and application analysis for oil and gas resource development [J/OL]. Journal of Engineering Science: 1-14 [2021-01-07].

[3] Xu Kewen, Xu Bo, Wu Ying, Xu Haoran. Overview of the Application of Machine Learning in Ultrasound Images [J/OL]. Computer Engineering and Applications: 1-9 [2021-01-07].

[4] Wang Zhendong, Zhang Lin, Li Dahai. Overview of Internet of Things Intrusion Detection System Based on Machine Learning [J/OL]. Computer Engineering and Applications: 1-12 [2021-01-07].

[5] Toya, Li Shiying, Zhang Ji, Deng Kaifei, Luo Yiwen, Dong Hewen, Huang Ping. Fourier transform micro-infrared spectroscopy combined with machine learning algorithms to identify electric shock injuries[J]. Forensic Medicine, 2020, 36(01): 35-40.

[6] Peng Liqin, Wan Lei, Wang Maowen, Li Zhuo, Zhao Hu, Wang Yahui. Research progress and prospects of machine learning in bone age assessment[J]. Forensic Medicine, 2020, 36(01): 91-98.

[7] Dai Feifan, Si Weiyun, Ni Jinping. Research on Judicial Data Analysis and Modeling Based on Machine Learning——Taking "Intentional Injury Crime" as an Example [J]. Intelligent Computers and Applications, 2020, 10(01): 89 -94.