Research on Key Information Processing Technology of Environmental Law Enforcement Based on Big Data

Tengfu Liu¹, Lingyao Cheng²
1 Sichuan University, Law School, China, 610207
2 Southwest Jiaotong University, The School of Information Science and Technology, China, 611756

*Corresponding author e-mail: scutfl123@163.com

Abstract. In the enforcement of environmental cases, the acquisition of key information such as plots and evidence has always been a key difficulty. However, the traditional search engine-based key information processing technology for environmental law enforcement suffers from long-term defects and high information loss rate due to the excessive reliance on manpower. In this regard, based on big data technology, this paper proposes new key information processing methods for environmental law enforcement. Specifically, it uses a combination of data warehouse and agent software to initially store the production information of the law enforcement objects; then generates metadata according to the specific requirements of environmental law enforcement; and then classifies the relevant metadata through the K-NN classifier; Finally, the BP neural network algorithm is used to aggregate the classified data to form effective plots and evidence. In order to verify the feasibility of this technology, this paper designs a comparison experiment with its traditional method. The experimental results show that the key information processing technology of environmental law enforcement based on big data has faster information acquisition rate and lower information loss rate, which is a feasible new approach.

Keywords: Environmental Law Enforcement, Assistive Technology For Key Information Acquisition, Big Data, K-NN classifier, BP Neural Network Algorithm

1. Introduction

Enforcement is the lifeblood of law, especially environmental law. Relevant studies show that the current environmental legal system in China has poor implementation effect and is difficult to implement. Among them, the difficulty in obtaining key information such as plot and evidence is quite important. In a general sense, key information such as scenes and evidence at the level of environmental law enforcement is mainly attached to the internal data set of the object of law enforcement, which is not difficult to collect according to the provisions of the Environmental Protection Law and other laws, the object of law enforcement is obligated to provide its own internal data to the law enforcement department. However, due to such factors as complex internal materials and deliberate concealment by law enforcement targets, the data set obtained by law enforcement
departments is rather primitive and crude, and valuable information such as plots and evidence can only be discovered after "finishing". To this end, Chinese law enforcement authorities have developed a key information processing and retrieval technology for environmental law enforcement based on search engines. Through searchers, indexers and user interfaces, relevant information in the data set is filtered and manually screened and processed to determine the available key plot and evidence information. Objectively speaking, although the searcher and indexer operated by robot can achieve the effect of information preprocessing and database establishment, its accuracy and repetition rate are frequently criticized. Therefore, environmental law enforcement agencies in addition to the search engine to add manual review steps; In this way, the cycle of obtaining key information of environmental law enforcement is greatly prolonged. Due to the limitation of human resources, key information may be lost in any link, thus affecting the final output. Thus, the author believes that although the traditional key information processing technology of environmental law enforcement based on search engine can be used, it is obviously not suitable for the needs of contemporary environmental law enforcement, and new methods are in the pipeline.

In the past five years, big data has become a hot topic in China and the world. Especially in the field of law, with the introduction and application of big legal data, legal institutions and legal personnel are at a crossroads and will face drastic changes in the next 20 years, which will exceed the changes in the past two centuries combined. As far as the key information of environmental law enforcement is concerned, it is to some extent related to the principle and form of big data, and high-speed and accurate big data processing technology can be involved in this field to a large extent. Therefore, this paper will study the key information processing technology of environmental law enforcement based on big data in order to overcome the defects of traditional technologies and improve the environmental law enforcement system.

2. Design of key information processing technology for environmental law enforcement based on big data

2.1. Acquisition of original information

As mentioned above, for environmental law enforcement agencies, the initial information resource, namely the internal data set of the target, has a fixed access channel. Therefore, there is no need to use crawlers and other technologies to find all over the sky -- many times a simple data transmission interface can meet the requirements. However, the key point is that the initial data acquisition is simple, but further acquisition is difficult: the internal data set itself has a large volume and high repetition rate, which makes it difficult to store and de-weight in general. In this regard, this paper proposes to establish a special Data Warehouse for environmental law enforcement information to store and update possible environmental law enforcement information and facilitate information acquisition. The Data Warehouse structure is shown in Figure 1:

![Data Warehouse structure](image)

**Figure 1.** Schematic diagram of environmental law enforcement information data warehouse
At the same time, in order to overcome the internal data of the law enforcement set obstacles of high repetition rate is put forward, in this paper with the preliminary processing agent software data collection way, namely by the type of data, information such as keyword as a tag, with agent software has processed tag, so as to eliminate repetition, relevant data resources, for the first time after extraction of the metadata generated.

2.2. Generation of key information metadata
Metadata is the data that describes and qualifies the information resource data. The first extracted information is transformed into metadata, which is the necessary step of the subsequent information processing. The key information of environmental law enforcement is normative, unified, descriptive and intuitive. Based on the above characteristics, this paper formulates the corresponding metadata format:

| Table 1. Key information metadata formats |
|------------------------------------------|
| Metadata attribute | Instruction | Metadata attribute | Instruction |
| Metadata name       | Environmental Enforcement Information | Compound type | Yes |
| Short name          | information | Control types | Text control |
| Date type           | String | Length | 8 bytes or more |
| Check way           | Input data is validated | Source | Related data warehouse |
| domain              | More value | Whether it occurs more than once | No |

According to this format, the information extracted for the first time can be generated by the corresponding metadata, so that the information data can be classified and processed later. Of course, there is still room for adjustment of the above format considering the particularity of the presentation of key information such as plot and evidence in some cases.

2.3. Key information metadata classification and processing
After the first extracted environmental law enforcement information is transformed into the metadata of key environmental law enforcement information according to the metadata format, the next step is data classification processing. To this end, this paper proposes a method for processing key information metadata through k-NN classifier, whose formula is as follows:

\[
X_{\text{train}} = (x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(i)})
\]

(1)

In formula (1), \(X_{\text{train}}\) is the sum of the training set. According to the principle of k-NN classifier, the formula of the test set is given below:

\[
X_{\text{test}} = (x^{(i)}, x^{(2)}, x^{(3)}, \ldots, x^{(i)})
\]

(2)

In formula (2), \(X_{\text{test}}\) is the sum of the test sets. When the test set and training set are clear, the training of k-NN classifier is completed. The next step is to classify the metadata using the distance algorithm.

\[
d(x^{(i)}, x^{(j)}) = \sum_{l=1}^{1} |x_l^{(i)} - x_l^{(j)}|
\]

(3)

This paper classifies metadata by calculating Manhattan distance. In formula (3), the distance of all training sets can be calculated by \(d(\text{x}(I), \text{x}^{(j)})\), thus completing the construction of k-NN
classifier. Thus, the trained K-NN classifier will classify the metadata of key information of environmental law enforcement, and provide a prerequisite for the final integration and processing and result output.

2.4. Use BP neural network to achieve the final processing
After metadata generation and K-NN classifier classification, the available key information of environmental law enforcement has begun to take shape. However, from the perspective of process, such information is still scattered and needs to be further integrated and processed before it can be output as effective plots and evidence. In this paper, BP neural network is proposed to realize the final integration.

The BP neural network proposed in this paper consists of an input layer, three hidden layers and an output layer, and the activation function is used to de-linearize the input metadata. In this method, the activation function of BP neural network selects SIGmoid function, whose formula is as follows:

$$g(z) = \frac{1}{1 + e^{-z}}$$ (4)

The Sigmoid function is not as fast as some functions, but it is the easiest to understand and fully applicable to the relatively simple form of critical information metadata for environmental law enforcement.[5] In addition, the BP neural network loss function adopted in this paper adopts the maximum boundary loss function to regulate various parameters and reduce losses. Its formula is as follows:

$$loss = max(0, 1 - y \cdot a), y = \pm 1$$ (5)

After BP neural network is designed, the training formula above is used to train the neural network. In the training process, the relevant parameters are adjusted continuously until the loss function value reaches the minimum, and the iteration is finally stopped. After being processed by BP neural network, the metadata can be output and the corresponding information can be output to relevant systems according to the mapping rules, thus completing the process from the original data set to the key information of available and effective environmental law enforcement. So far, the key information processing technology of environmental law enforcement based on big data has been designed.

3. Performance test of key information processing technology for environmental law enforcement based on big data

3.1. Experimental content
In the previous section, this paper elaborated the principle and content of key information processing technology for environmental law enforcement based on big data. [6] Theoretically speaking, it should be superior to the traditional key information processing technology for environmental law enforcement based on search engine. However, its final performance needs further verification. Therefore, this paper designs a comparative experiment between the key information processing technology of environmental law enforcement based on big data and the key information processing technology of environmental law enforcement based on search engine. The comparison index of the experiment is the integration path of the two technologies and the integration effect of information resources. By comparing the above two indicators, we can measure the effect of the two technologies compared in this experiment on the processing of key information of environmental law enforcement, so as to judge which is better and which is worse.

3.2. Experimental process
The experiment designed in this paper is divided into two parts:
The comparison index of experiment 1 is the processing path of the two technologies. The internal data sets of several law enforcement objects were selected as data sources before the experiment. First of all, this paper will mark different internal data sets of the objects of law enforcement with identification codes of different lengths, and then use two technologies to process the internal data sets of the objects of law enforcement. By tracing the processing path of the two methods to the data set with the identification code, the processing path diagrams of the two technologies are obtained, and further analysis is made to draw the conclusion of experiment 1.

The comparison index of experiment 2 is the ratio of data loss and repetition when the two processing technologies process the same data set. By sorting and analyzing the experimental data, the conclusion of experiment 2 will be drawn in this paper.

Based on the conclusions of the above two experiments, this paper will comprehensively evaluate the performance of the two technologies in this comparative experiment.

3.3. Experimental results
The results of experiment 1 are shown in the figure below:

![Figure 2](image)

Figure 2. The two techniques are compared in the processing paths of the internal data sets of the law enforcement objects

The figure 2 shows that in the use of verification in this paper, based on large data of key information processing technology, environmental law enforcement of the law enforcement internal data collection only after data transit point (transfer site) to enter the save point (storage site), then the output, the path of a single, relatively short, it will be relatively short processing cycle, is not easy to cause confusion and loss of information. Using the traditional search engine based key information processing and retrieval technology for environmental law enforcement, the internal data set of the object shall first pass through multiple routing sites and then through multiple transfer sites before entering the storage site and finally output the results. In this process, the path is longer and more complex, and the nodes also intersect with each other. In this way, the cycle is longer and the information is relatively easy to be confused and lost, thus affecting the output results.

The results of Experiment 2 further verified the conjecture of Experiment 1:
Table 2. Comparison of parameters of processing results of the two technologies

| group | Authentication technology | Compare the technology |
|-------|--------------------------|------------------------|
|       | Presented According % | Repeta Rate % | Processing Cycle min | Presented According % | Repeta Rate % | Processing Cycle min |
| 1    | 1.54                     | 1.21                  | 1.2                  | 7.69                    | 21.33                  | 12.7               |
| 2    | 1.33                     | 1.19                  | 1.1                  | 8.16                    | 20.96                  | 10.8               |
| 3    | 2.02                     | 1.42                  | 1.2                  | 7.97                    | 22.07                  | 13.2               |
| 4    | 1.76                     | 1.36                  | 1.2                  | 9.23                    | 19.75                  | 11.5               |
| 5    | 1.49                     | 1.38                  | 1.3                  | 8.08                    | 23.41                  | 12.3               |
| 6    | 1.98                     | 1.57                  | 1.1                  | 8.71                    | 20.17                  | 10.6               |
| 7    | 2.06                     | 1.15                  | 1.2                  | 7.43                    | 18.69                  | 11.2               |
| 8    | 2.11                     | 1.03                  | 1.1                  | 9.05                    | 22.56                  | 13.4               |
| 9    | 1.42                     | 1.59                  | 1.3                  | 8.54                    | 19.82                  | 12.6               |
| 10   | 1.68                     | 1.45                  | 1.1                  | 8.39                    | 21.31                  | 10.9               |

It can be seen from Table 2 that when the key information processing technology of environmental law enforcement based on big data verified in this paper is used, the average information loss rate is 1.74%, the average repetition rate is 1.34%, and the average processing cycle is 1.2min. When the traditional search engine-based key information processing technology of environmental law enforcement is used as a comparison, its average information loss rate, average repetition rate and average processing cycle are 8.33%, 21.01% and 11.9min. It can be seen that the information loss rate, repetition rate and processing cycle of verification technology are far less than those of comparison technology. Compared with the traditional key information processing technology of environmental law enforcement based on search engine, the key information processing technology of environmental law enforcement based on big data has shorter cycle, lower information loss rate, and features of efficiency and precision. It is a feasible new approach.

4. Conclusion

In view of the defects of the traditional key information processing technology of environmental law enforcement based on search engine, this paper proposes the key information processing technology of environmental law enforcement based on big data. Through experimental verification, this technology has the characteristics of high efficiency and precision, overcomes the defects of traditional technology, and is worthy of extensive promotion. In the future, the technology can be further improved to promote technological progress in the field of environmental law enforcement.

References

[1] Chang Jiwen: New & LT; Environmental Protection Law & GT; The strictest in history but the hardest to implement, Environmental Protection, 10th issue, 2014, page 24
[2] Zhu Jie, Lu Xinghua: Multimedia Integrated Learning Resource Information Integration System in Cloud Storage Environment, Information Technology 2019, 6th issue, page 125.
[3] Richard Susskind. Tomorrow's Lawyers: An Introduction to Your Future, Oxford: Oxford University Press, 2013, pp.2.
[4] Liu Feng, Zhang Xiaolin: Review of Scientific Data Metadata Standard and Research on Its Generalization Design, Modern Library and Information Technology, 2015, 12th issue, page 3.
[5] Huang Yi, Duan Xiuisheng, Sun Shiyu, Lang Wei: Research on Deep Neural Network Training Algorithm Based on Improved SIGmoID Activation Function, Computer Measurement and Control, 2nd issue, 2017, page 126.
[6] Xiong Zhiqiang, Zhang Xianjing. Research on Computer Information Processing Technology Based on Big Data[J]. Journal of Beijing Institute of Graphic Communication, 2020.