

Corporate Governance Data Storage Tort Liability for Artificial Intelligence and Emotion Recognition

Zimi Wang

School of Law, University of International Business and Economics, Beijing, China

Email: 275070365@qq.com

How to cite this paper: Wang, Z. M. (2023). Corporate Governance Data Storage Tort Liability for Artificial Intelligence and Emotion Recognition. *Beijing Law Review*, 14, 215-232.

<https://doi.org/10.4236/blr.2023.141012>

Received: November 23, 2022

Accepted: March 13, 2023

Published: March 16, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The rapid development of information technology in the 21st century has realized the transformation of people's production and lifestyle. At the same time, the current research on data infringement is increasing. At present, the legal field has gradually begun to apply informatization, using its advanced data exploration methods and collection technologies. This paper aimed to realize the research on tort liability of corporate governance data storage under the background of artificial intelligence and emotion recognition, so as to promote the ecological development of tort liability and the protection of rights and interests in the era of big data. The principles and laws of neural network technology in artificial intelligence were used for reference to promote the development of data infringement tracking technology. In the experiment of numerical information extraction under the neural network, the recognition rate of the system algorithm only exceeded 98% in time, and the highest recognition rate in location type was only 95.39%. Therefore, it showed that the method proposed in this paper achieved a certain recall rate, and the method was flexible and feasible. Due to the limited training sample size, the accuracy of subjects and attributes was relatively low. However, accuracy topics and attributes had relatively little impact on search results. The subsequent search module was based on a full-text search engine. Therefore, it is of great significance to use neural network technology to study the tort liability of data storage in the context of corporate governance.

Keywords

Data Storage Infringement Liability, Numerical Extraction Technology, Artificial Intelligence, Emotion Recognition

1. Introduction

Since the 21st century, with the continuous development and progress of science and technology, human society has moved from the information age to the data age. Data has become a key factor driving the rapid development of contemporary society and economy. Most scholars generally agree that Big Data is a concept opposed to traditional data. It is distinguished from traditional data, known as “small data” in terms of volume, speed and diversity. It is the large amount of data that cannot be processed or analyzed by traditional software or computing systems (Wang, 2021). Data economy has high innovation, strong permeability and wide coverage. It is not only a new economic growth point, but also a fulcrum to transform and upgrade traditional industries. China’s data economy started late, but China’s Internet infrastructure is gradually improving. The industrial development has taken the lead in the world, and the basic environment has been upgraded in an all-round way, which provides a good soil for China to develop the data economy. In addition, the Chinese government takes the data resource sharing transaction as a development strategy, and actively deploys the data industry. Some professional associations and alliances have also issued a number of related group bids. They are trying to actively promote standardization at different levels, and trying to ensure that the data economy remains ahead in new areas. At present, the number of data service providers such as data demanders, providers, and trading platforms in China’s data trading market has gradually increased. Data products, data types, and data transaction volumes are becoming more and more numerous, which reaches an exponential level of growth. The data economy is about to become a powerful driver of China’s rapid economic growth. At present, there is a relatively lack of relevant research on the risk of data rights infringement in data transactions. Existing research is mainly about data rights and data transactions. The relevant researches on data rights by relevant scholars around the world mostly focus on the content of data rights and the process of proposing a certain data right. The related researches on data transaction focus on the practical status of the transaction center, transaction dilemma, transaction security and so on. However, there are few studies discussing issues related to data rights in the context of data transactions, and even fewer studies discussing the risk of infringement in data transactions. The research on infringement risk in data transaction is conducive to improving the data transaction environment and improving transaction efficiency. At present, China’s research on data infringement and its tracking is still in its infancy. Basic research at all levels is still very poor, and relevant experience in practical application is lacking. Infringing acts such as piracy of data products are difficult to trace, discover and protect rights. Therefore, this paper first defines data rights and summarizes relevant laws and regulations, which makes the concept of data infringement clearer and lays the foundation for subsequent research. Secondly, data productization and entering the market for circulation is a means to maximize the value of data utilization. However, data products are easy to reproduce

and open, which makes most data controllers unwilling to trade data products, resulting in low utilization of data resources. The research on data infringement tracking not only effectively promotes the free flow of data, but also breaks the segmentation of data information chains and improves the efficiency of data information sharing and utilization. It provides strong support for creating a safe, reasonable, efficient and convenient data transaction environment. Effective tracking and monitoring of the infringement of data products helps to overcome the dilemma of data transaction development in the emerging data industry. It can improve the operation efficiency of the data product trading and circulation market, stabilize the order of the data product trading and circulation market, and ensure the stable development of the trading market.

The research on the tort liability of the company's data has always been a relatively vague issue in the era of big data. The following scholars have conducted research in it. Leung P proposed that Orange sued a U.S. affiliate in the U.S. District Court of Delaware, accusing it of infringing five patents related to data storage (Leung, 2018). Cao J's research showed that on the basis of summarizing and analyzing previous research work, the research status and significance of intellectual property expression and educational resource data storage should be expounded (Cao, 2021). Chang C C stated that due to the rapid increase in copyright infringement, it was very important to protect confidential information during transmission over the Internet. More and more researchers had proposed solutions such as stegan (Chang et al., 2017). Menyas P filed a ruling that delayed U.S. government access to server information seized from a Canadian company that leased servers to Megaupload Inc, pending a court-supervised analysis of the servers (Menasz, 2017). Hoy M proposed that the EU allowed persons accused of infringing company names and trademarks to challenge the legal validity of names or trademarks in criminal cases, rather than civil cases (Hoy, 2018). However, the above infringement cases only stop at the level of copyright and patent rights. For the data storage class level, it is only a generalization without in-depth analysis.

The use of artificial intelligence techniques to conduct research on corporate data storage violations is a novel topic. Among others, Kumar R developed a software plagiarism detection tool, which was used to find the content of newly created works to understand which parts of it were plagiarized and to what extent from local repositories/databases and from which sources or from sources available on the internet (Kumar & Tripathi, 2017). Nayak M argued that infringed tech companies such as Microsoft Corp, SAP and Teradata were entering the enterprise data analytics and warehouse software market. These software helped businesses store and analyze operational data (Nayak & Tyler, 2018). Twining G proposed to launch targeted action on behalf of the International Transport Workers Federation against one of Germany's oldest shipping companies. This was due to the company's alleged violation of the rights of others with respect to the data storage project. Menyas P proposed that Canadian courts had confirmed that only pharmaceutical companies holding patents could choose how to

compensate for data storage infringement. Jahner K proposed that US Patent No. 7,840,437 described a system for managing digital data and transactions to ensure that audio and video data rented by consumers could only be viewed for a limited time (Jahner, 2018). However, the above research is only limited to the determination of tort liability of traditional network service providers. Its own ambiguity makes artificial intelligence technology unable to highly integrate with the research on data storage infringement and give full play to its advantages.

The innovation of this paper is that the research on data infringement tracking in the context of data transaction makes up for the blank of related issues in China in the field of data infringement tracking technology. It is beneficial to improve the ability and efficiency of sharing and comprehensive utilization of data information resources in the data trading market. It also promotes the rational and efficient allocation of data and information resources. Therefore, the research in this paper also has important theoretical significance and social practical significance. Based on the description of data rights in global laws, as well as the analysis and investigation of the current situation and problems of data transactions, the connotation of data rights had been studied and discussed in depth. The background of data transaction and data circulation market was explained, which laid a theoretical foundation and background environment for the research of data infringement system.

The innovation of this paper is that the research on data infringement tracking in the context of data transaction makes up for the blank of related issues in China in the field of data infringement tracking technology. It is beneficial to improve the ability and efficiency of sharing and comprehensive utilization of data information resources in the data trading market. It also promotes the rational and efficient allocation of data and information resources. Therefore, the research in this paper also has important theoretical significance and social practical significance. Based on the description of data rights in global laws, as well as the analysis and investigation of the current situation and problems of data transactions, the connotation of data rights had been studied and discussed in depth. The background of data transaction and data circulation market was explained, which laid a theoretical foundation and background environment for the research of data infringement system.

2. Tort Liability for Data Storage Based on Neural Network

2.1. Data Transaction and Data Storage

Data infringement tracking is based on data transaction and circulation. In order to better determine the direction of data infringement tracking service, it is necessary to study the connotation of data transaction and data circulation (Langlois & Mylan, 2018). Data is not a physical commodity, the transaction process is relatively complex, and the transaction forms are various. This paper believes that data transactions can be divided into four categories according to the “subject matter” and “process” of transactions, as shown in **Figure 1**.

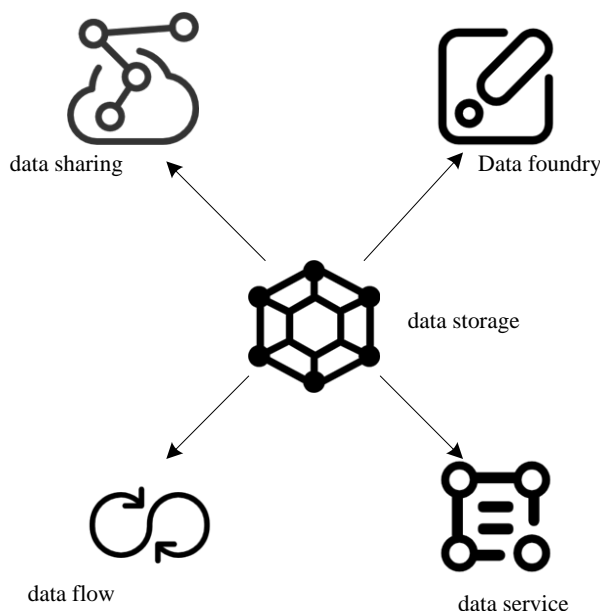


Figure 1. Classification of data-based transaction activity.

As shown in **Figure 1**, common data transactions are divided into four categories: data circulation, data foundry, data sharing, and data services. In the three types of data transaction activities of data foundry, data sharing and data service, the scope of data circulation are limited, and their infringers are easy to be identified and discovered. In data circulation, the transaction process is similar to the transaction of physical commodities, and “data” is passed from the seller to the buyer as a commodity through market trading activities (McDonald et al., 2020). Data is easily leaked in the process of multiple transfers, and it is difficult to determine the leaking party. Therefore, this paper mainly studies the problem of infringement tracking in the data circulation market in data transactions.

2.2. Data Storage Infringement Tracking Mode and Requirements

To accurately analyze the data infringement tracking mode and framework, first of all, it is necessary to determine the infringement scenarios and infringements in the data trading market. By combining the specific scenarios, transaction types and data product types of data circulation, the corresponding data infringement tracking solutions are designed according to different scenarios (Shields, 2017). **Figure 2** shows the transaction types and product types of the data exchange market.

As can be seen from **Figure 2**, there are mainly three types of data transactions, namely entrusted transactions, agency transactions and direct transactions. Regardless of the type of transaction, infringement may occur during the transaction. Because of the particularity of data products, in the data source service, since the data provider provides online data services in real time for many times, it is difficult for the data buyer to technically retain the data. Even if the local retention of data can be achieved, it is difficult to reuse these data due to

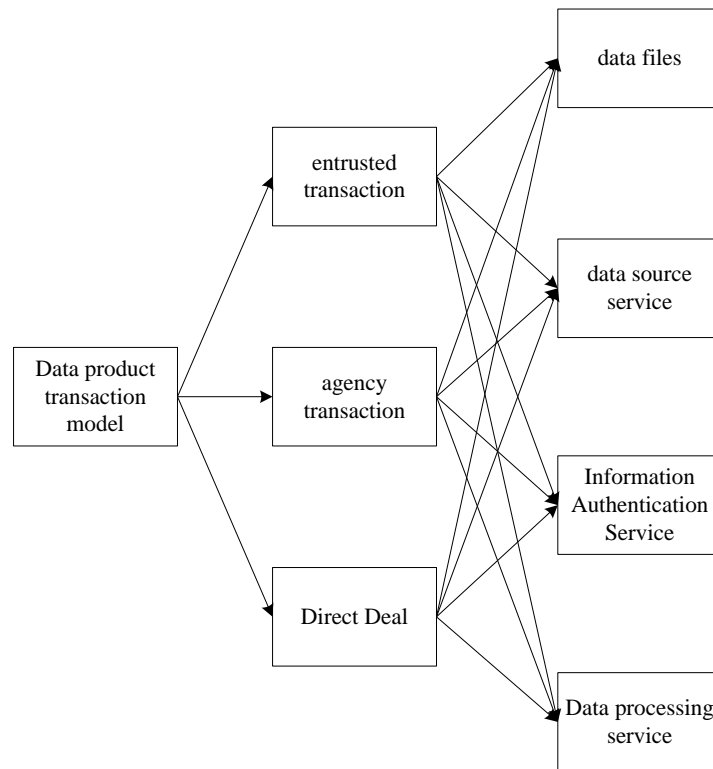


Figure 2. Data type transaction pattern diagram.

the lack of corresponding scene assistance. In the information authentication service and data processing service, the data buyer cannot access the original data of the product, but can only access the data service provided by the data seller, which makes it more difficult to infringe on the data. Therefore, in general, data infringement mostly occurs in the transaction mode of one-time delivery of data product files. Therefore, the infringement scenario discussed in this paper mainly occurs in the data commodity transaction in which the data product file is delivered one-off in the data product exchange market.

2.3. Construction of Data Storage Infringement Tracking System

The architecture design of the data infringement tracking system based on the numerical information extraction technology is mainly divided into three layers, namely the display layer, the logic layer and the storage layer. The architecture of the data infringement tracking system is shown in **Figure 3**.

As can be seen from **Figure 3**, the interface display layer is mainly used for the interaction between the data infringement tracking system and the user. Users submit their data products to the data infringement tracking system, and the data infringement tracking system provides infringement tracking services. When the infringement tracking system finds an infringement, the warning message and relevant clue evidence would be returned and displayed on the interface layer, which prompts the user to take reasonable rights protection measures. The business logic layer mainly realizes the entry of data product information and

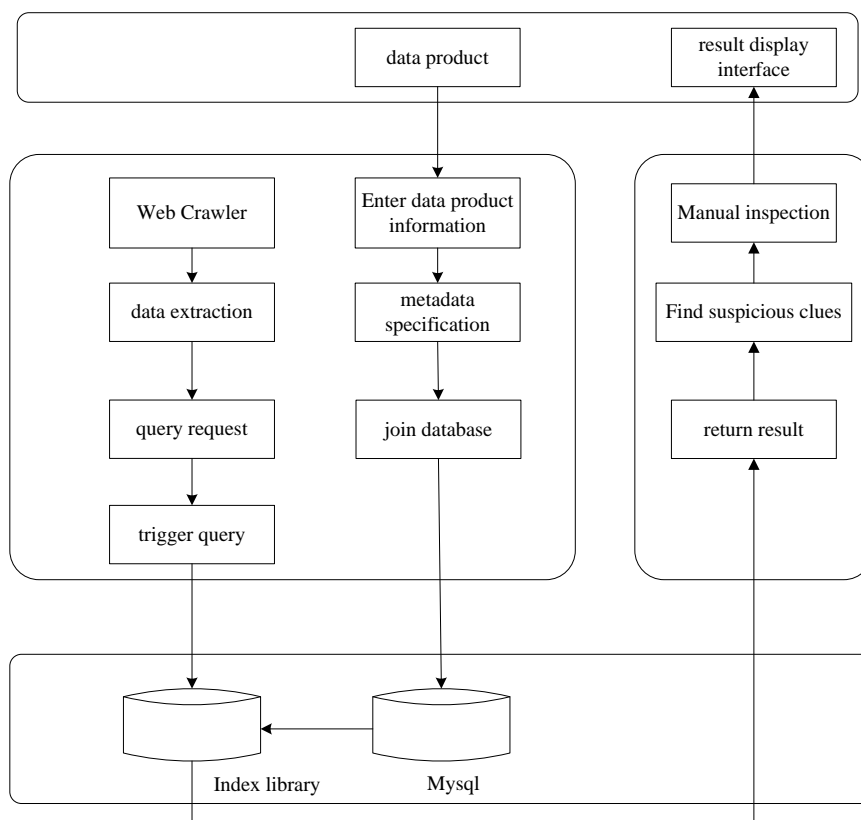


Figure 3. System architecture diagram.

the normalization of metadata, the extraction of web crawler and numerical information, and the query and feedback of infringement information. The infringement tracking system uses crawling technology to monitor key websites and extract web pages from the Internet. Artificial intelligence NLP (Natural Language Processing) techniques are used to extract numerical metrics from web pages. The indicators are compared with the data of the “data product library” and its derivatives, and suspicious clues are found, which prompts manual inspection. Infringements are discovered and evidences are kept in a timely manner. The storage layer is mainly used to store data. In the data infringement tracking system, there are mainly two databases. One is the MySQL database and the other is the ElasticSearch index library. MySQL database mainly stores some structured information. The product data of the data provider is written into the ElasticSearch index library. The data information of the data extraction module is also put into the ElasticSearch index library for comparison and analysis. In the data infringement tracking system, data is the core existence. **Figure 4** shows the data flow of the system.

As shown in **Figure 4**, the data flow mainly goes through three major processes: acquisition, indexing, and searching. First, user-supplied data products are used to build ES’s inverted index library. Then, the relevant text information is taken from the key monitoring website for data cleaning, and some irrelevant information, null values, and duplicate values are excluded. Next, the structured

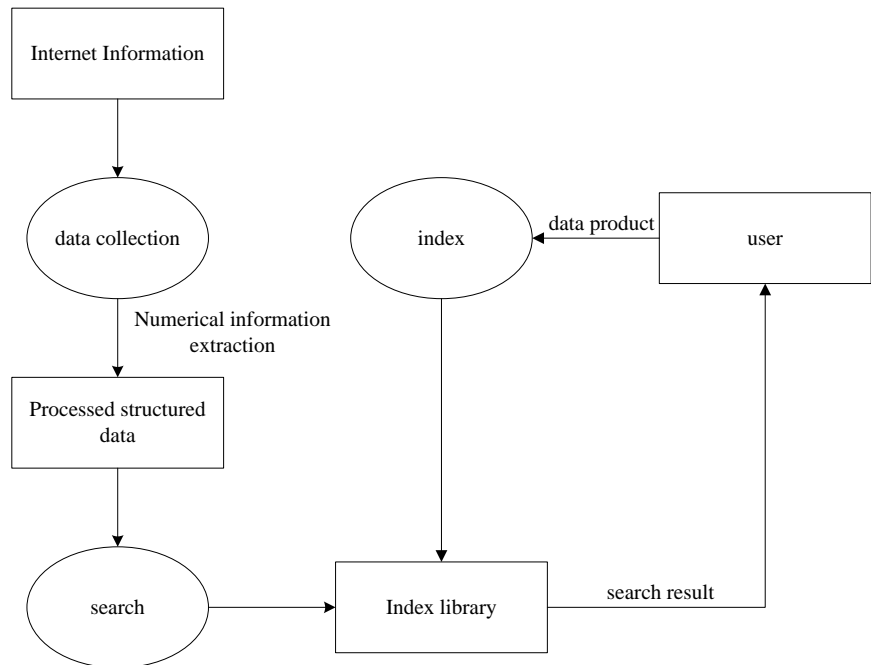


Figure 4. Data flow chart of data infringement tracking system.

data is obtained by extracting numerical information, the keywords are matched in the index database, and the search results are returned to the user (Cai et al., 2017). At this point, the flow of data in the infringement tracking system and the processing process are completed.

2.4. Functional Module Design of Data Storage Infringement Tracking System

This data infringement tracking system is mainly divided into three modules, a data acquisition module, a numerical information extraction module, and a data infringement tracking and search module. The functional block diagram of the data infringement tracking system is shown in Figure 5.

As can be seen from Figure 5, the functions of each module are as follows.

Data acquisition module: Mainly according to the analysis of business needs, the data of the relevant website is taken from the network and synchronized to the database, and the real-time update of the database is realized. The sub-modules included in the data collection module mainly include: web crawler, establishment of collection database.

Numerical information extraction module: It mainly uses algorithms based on rules, statistics and deep learning to extract the text in the database to complete the structuring of the data. The sub-modules included in the numerical information extraction module mainly include: data preprocessing, rules, statistics and deep learning extraction.

The developer designs the corresponding database schema and database table on the existing database management system according to the user's needs. The database tables are shown in Table 1.

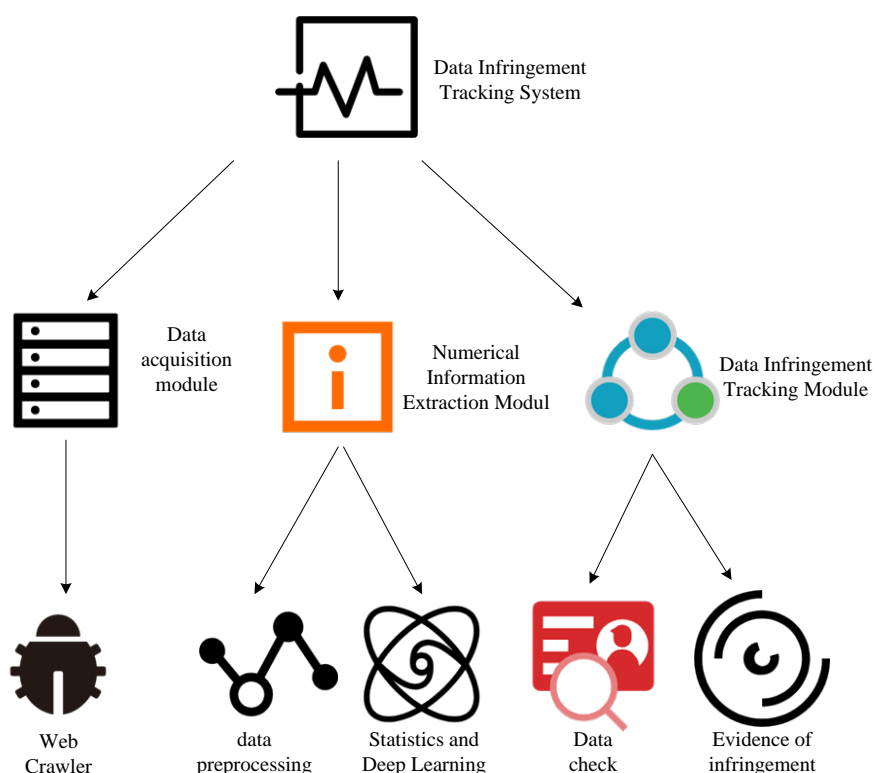


Figure 5. Functional block diagram of data infringement tracking system.

Table 1. Acquisition database tables.

name	type	length	Field explanation	Remark
id	int	32	id	non-null, primary key
url	varchar	255	url	non empty
title	varchar	255	Article title.	non empty
content	longtext	265	Text content	primary key
create date	datetime	32	Article creation date	non empty

As can be seen from **Table 1**, in order to better support the operation of other programs of the data infringement tracking system, it is necessary to reasonably plan the attributes and relationships of the data objects in the database. Reasonable allocation type and length are also required to reduce memory. Database tables are designed as rigorously as possible. If the subsequent application process is inconsistent with its function, the database would be overturned, and the data needs to be redesigned and imported.

After sorting out and summarizing and referring to other literatures, this paper proposes a numerical information storage format suitable for data infringement tracking system. The original corpus and its corresponding numerical information storage format are shown in **Table 2**.

At present, the vast majority of research still mainly adopts the method based on pattern matching. This paper aims at the problems of single information

Table 2. Numerical information storage format table.

main body	Attributes	attribute value	time
real estate development	invest	15144.3 billion yuan	January to December 2020
real estate development	invest	up 4.0%	January to December 2020
Residential	invest	11444.6 billion yuan	January to December 2020
Residential	invest	up 8.6%	January to December 2020

processing, poor extraction effect and poor portability of pattern matching methods. Combined with the characteristics of numerical information, the neural network method based on deep learning is applied to information extraction. A numerical information extraction framework combining rules, statistics and deep learning is proposed. This method effectively reduces manual intervention, improves portability, and solves the problem of mode conflict.

2.5. Neural Networks and Natural Language Processing Algorithms

1) Long and short-term memory network

Long short-term memory network (LSTM) is designed on the basis of recurrent neural network. Unlike standard feedforward neural networks, LSTMs have feedback connections (Woodman et al., 2017). LSTM utilizes three gates to control the inflow and outflow of long-term states. These three gates exist in the core part of LSTM, the LSTM unit, which is the data flow calculation process of the three gates. The following is a detailed introduction.

Forget Gate: It determines how much of the unit state C_{t-1} at the previous moment is retained to the current moment C_t . It is determined by the Sigmoid function. By looking at the previous state h_{t-1} and the input content x_t , each number in the Cell state C_{t-1} is output a number between 0 and 1, the Formula is: $i_t = \sigma(W_i * [h_{t-1}, X_t] + b_i)$

Input Gate: It determines how much of the network's input X at the current moment is saved to the cell state C_t . The tanh function weights the passed value to determine its importance level (-1, 1). Its calculation Formula (2) is:

$$i_t = \sigma(W_i * [h_{t-1}, X_t] + b_i) \quad (2)$$

The current input unit state calculation Formula (3) is:

$$\tilde{C}_t = \tanh(W_c * [h_{t-1}, X_t] + b_c) \quad (3)$$

Output Gate: It is used to determine the final output. The tanh function weights the passed value to decide its importance level (-1, 1) and multiplies the output of the sigmoid. The calculation Formula (4) is:

$$O_t = \sigma(W_o * [h_{t-1}, X_t] + b_o) \quad (4)$$

The calculation Formula (5) of the unit state C_t at the current moment is:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (5)$$

The final output calculation Formula (6) of the LSTM at the current moment is:

$$h_t = O_t * \tanh(C_t) \tag{6}$$

2) Bi-LSTM-CRF model

On the basis of the LSTM model, relevant scholars first proposed the Bi-LSTM-CRF model and used it to solve the sequence labeling problem in natural language processing. The BI-LSTM-CRF model is a combination of bidirectional long short-term memory network and conditional random field. The model is shown in **Figure 6**.

It can be seen from **Figure 6** that for sentence sequence $X = (x_1, x_2, \dots, x_n)$, it is first entered into the bidirectional long-term memory network. Then it is introduced into the state transition matrix $[A]$, and the matrix P is set as the output of the bidirectional long short-term memory network. $[A]_{i,j}$ represents the probability of transitioning from the i th state to the j th state at each time step. $[P]_{i,j}$ represents the probability that the i -th word in the input observation sequence is the j -th label. Then the prediction output Formula (7) of the label sequence $y = (y_1, y_2, \dots, y_n)$ corresponding to the observation sequence $[X]$ is:

$$s([X], y) = \sum_{i=1}^n ([a]_{y_{i-1}, y_i} + [P]_{i, y_i}) \tag{7}$$

Among them, each element A in the transition matrix A represents the score from the i th label to the j th label. y_o and y_n represent the beginning and end of a sentence, respectively. Therefore, A is a matrix of $R(k+1)*(k+2)$, and finally the softmax logistic regression model is used to normalize the label score to the conditional probability. Its Formula (8) is as follows:

$$P(X | Y) = \frac{e^{s(X,Y)}}{\sum_{y \in Y_x} e^{s(X,Y)}} \tag{8}$$

The choice of loss function is particularly critical, affecting the quality of the entire training effect. Here, the maximum likelihood function is used as the loss function, and its Formula (9) is as follows:

$$\log(P(Y | X)) = s(X, Y) - \log\left(\sum_{y \in Y_x} e^{s(X,Y)}\right) \tag{9}$$

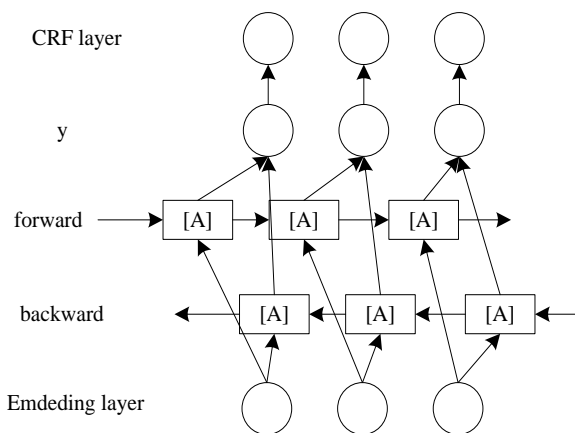


Figure 6. BI-LSTM-CRF model structure.

In the process of training, the use of dynamic programming algorithm can be more convenient than other methods. Therefore, it is used to perform fast computation of the state transition matrix, which enables optimization of the label sequence. Its Formula (10) is as follows:

$$y = \arg \max_{y \in Y_x} s(X, \tilde{Y}) \tag{10}$$

Without manual annotation, the model achieves the best results in named entity recognition, semantic block and part-of-speech tagging.

3) Centralized storage algorithm based on dynamic clustering

It is assumed that the number of times a node collects data per unit time in the wireless sensor network is r_{cb} and the size of the data collected each time is s_d . Node i is the sensor node that is not elected as a cluster head. Node h is the node elected as the cluster head, and o is the sink node. The storage cost of the data collected by node i per unit time is:

$$C_{store}(i) = r_d(i) s_d C(i, h) + a r_d(i) s_d C(h, o) \tag{11}$$

The energy of wireless sensor network nodes is consumed in acquisition, communication, computing and other operations. However, the energy consumed in the communication process far exceeds the consumption of other aspects. Therefore, the energy consumption of the node sending lbit data is calculated by Formula (12):

$$E_{tx}(l, d) = E_{elec} * l + \epsilon_{fs} * d^2 * l, d < d_0 \tag{12}$$

$$E_{tx}(l, d) = E_{elec} * l + \epsilon_{mp} * d^4 * l, d \geq d_0 \tag{13}$$

d_0 is calculated by Formula (14):

$$d_0 = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}} \tag{14}$$

The energy consumed to receive lbit data is:

$$E_{rx}(l) = E_{elec} * l \tag{15}$$

Therefore, the threshold formula for selecting cluster heads is:

$$T(i) = \frac{p}{1 - p \left(r \bmod \frac{1}{p} \right)} * \frac{E_{i,res}}{E_{init}}, i \in G \tag{16}$$

Among them, the member nodes in the cluster are alternately and selectively removed from the cluster or added to the cluster. This keeps clusters from being too large or too small in the network, balancing power consumption between clusters. The threshold for the number of cluster members is preferentially defined as:

$$V_1 = \left[\left(\frac{N}{k} - 1 \right) * (1 + \lambda) \right] \tag{17}$$

$$V_2 = \left[\left(\frac{N}{k} - 1 \right) * (1 - \eta) \right] \tag{18}$$

The exiting node i can join cluster h_z , and the conditions are met:

$$h_{z,num} \leq V_1 \quad (19)$$

$$C_{store,h_z}(i) \leq C_{store,h_j}(i) \quad (20)$$

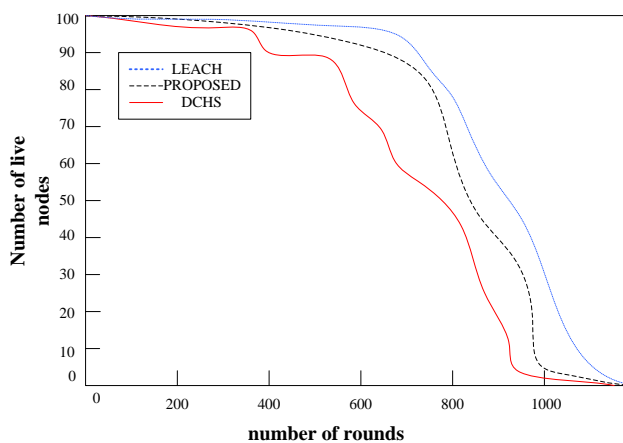
$$t_1 \leq h_{j,num} - V_1 \quad (21)$$

$$j_1 \leq V_1 - h_{z,num} \quad (22)$$

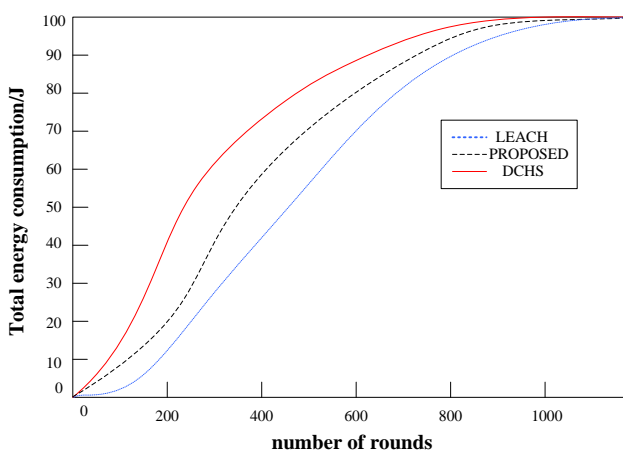
3. Data Storage Infringement Tracking System Experiment

3.1. Number of Surviving Nodes in the System

In order to verify the rationality and efficiency of the system algorithm, a comparative test is carried out under the centralized data storage and network experimental parameters under the LEACH and DCHS protocols. Its inventory node comparison chart is shown in **Figure 7**.



(a) The number of surviving nodes varies with the number of rounds



(b) The total energy consumption of the network as a function of the number of rounds

Figure 7. Comparison of the number of surviving nodes and the total energy consumption of the network under the change of the number of rounds.

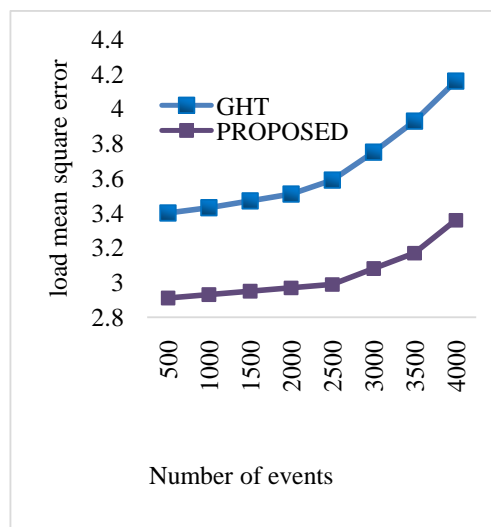
As can be seen from **Figure 7(a)**, the first node death value of the centralized data storage under the LEACH protocol was 401, and its total energy consumption percentage was 61.10%. The first node death value of DCHS was 492, and its total energy consumption percentage was 49.95%. The first node death value of the proposed algorithm was 646, and its total energy consumption percentage was 81.25%. The results showed that the total energy consumption percentage of the proposed algorithm was 20.15% and 31.30% higher than that of LEACH and DCHS, indicating that the lifetime could be extended in addition to cluster head optimization (Esposito, 2018). The proposed algorithm optimized the cluster size by dynamically adjusting the leaving or joining of cluster members, balancing the energy load among the clusters so that most nodes die at a later stage. It can be seen from **Figure 7(b)** that DCHS consumed less energy than LEACH. Because the residual energy factor of the node was added in the cluster head selection process, the load was distributed to each node in the network. The proposed method consumed less energy than the total energy of LEACH and DCHS centralized data storage. Because in addition to the remaining energy factor of nodes in the cluster head selection stage, the storage cost and proximity to the cluster head allow nodes to send data to nodes through the cluster head more efficiently during the cluster formation stage. Whereas the dynamic cluster phase allowed for a more even cluster distribution and more balanced energy consumption among the clusters, thereby reducing the overall energy consumption of the entire network. Therefore, as shown in **Figure 7**, compared with the centralized data storage under the LEACH protocol, the improved algorithm provides better network energy efficiency and load balancing, which in turn prolongs the life cycle.

3.2. System Load Balance

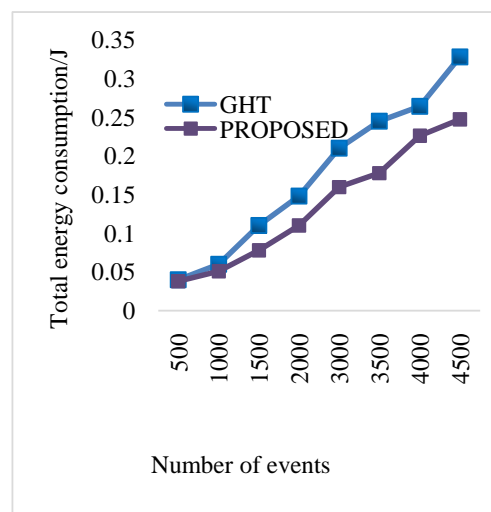
To evaluate whether the load in the network is balanced, it is mainly to check whether the resource consumption of the nodes is balanced. This paper defines the current load of node i as $(LS)_i$, which is a comprehensive evaluation function of the node's energy and storage capacity resources. The line comparison chart of the mean square error and total energy consumption with the number of events is shown in **Figure 8**.

It can be seen from **Figure 8(a)** that the proposed algorithm has a smaller load mean square error than the GHT algorithm, which indicates that the proposed algorithm has better load balance than the GHT algorithm as a whole. The GHT algorithm uses a standard hash function to evenly map the data in the network. The area where the nodes are sparsely distributed allocates more data, resulting in an excessively high load in the area. The mapping mechanism proposed in this paper is to map storage nodes according to data priority and node distribution. Storage nodes are more likely to be allocated in areas where nodes are densely distributed. The data priority is matched with the regional node density, which makes the load in the network more balanced. **Figure 8(b)** shows

that the energy consumption of GHT and the algorithm proposed in this paper increases as the number of events increases. However, the total energy consumption of the algorithm proposed in this paper is smaller than that of the GHT algorithm for the following reasons. The GHT algorithm selects storage nodes through a hash map without considering the user's query requirements. The algorithm proposed in this paper combines external storage and data-centric storage algorithms to select storage nodes that can make queries faster and consume less energy. Part of the data is directly stored in the sink node itself, which saves a lot of energy consumed by the query. Therefore, **Figure 8** shows a plot of load mean squared error versus total network energy consumption as a function of the number of events.



(a) Load mean square error with the number of events



(b) Total network energy consumption with the number of events

Figure 8. Comparison of mean square error and total energy consumption of different algorithms when the number of events changes.

3.3. System Performance Evaluation

In Chinese texts, the expression of time is very fixed. Therefore, it is suitable to use the rule-based method to extract the time accordingly. Location is a proper noun, such as country, province, city and region, which is relatively limited and fixed. Therefore, the location can be extracted by using a dictionary, which is more convenient and simple. However, it is necessary to pay attention to the actual situation such as the abbreviation of some place names. In practical applications, the extraction technology of time and place is very mature, and the existing extraction tools can be used to extract them. After analyzing the algorithm, this paper also evaluates and analyzes the performance of the data infringement tracking system. Among them, 10 interpretation articles published from January 2020 to March 2020 were taken from the latest release and interpretation column on the homepage of the National Bureau of Statistics of China. It covers six categories including cultural industries, industry, energy production, real estate, consumer goods retailing and manufacturing. At the same time, it includes statistical data and descriptive data, with comprehensive content and various forms. In all experimental corpora, there are 124 sentences with numerical information, of which there are 430 attribute values. All the original data are pre-processed and manually annotated to become experimental expectations, which are randomly divided into training set and test set according to the ratio of 4:1. In the Windows10 environment, the attribute value algorithm is tested by running the program and through the collected experimental text. The results are shown in **Table 3**.

It can be seen from **Table 3** that the recognition experiment of attribute value has obtained quite good results. The accuracy rate, recall rate and F1 value all reached more than 98%. It showed that the rule-based numerical information extraction method was very suitable for the extraction of attribute values. This is because the attribute value is mainly identified by numbers as a special identifier, and the expression is relatively fixed, so it is easy to identify. The main reasons for errors were incomplete rule matching templates, inaccurate word segmentation, and ambiguity of word meanings. The identification results of time and place are shown in **Table 4**.

It can be seen from **Table 4** that in terms of time, only the recall rate reached more than 99%, and only the highest recall rate in the location type reached 95.54%. Therefore, this showed that the method proposed in this paper achieved a certain recall rate, and the method was flexible and executable. Due to the limited number of training samples, the accuracy of subjects and attributes was relatively low. Since the subsequent search module was based on a full-text search engine, the accuracy of entities and attributes had relatively little impact on search results. Therefore, using artificial intelligence and emotion recognition technology, data storage infringement liability under corporate governance was analyzed, which could help data infringement tracking technology to be optimized.

Table 3. Attribute value identification results.

Quantity to be drawn	430
Number of draws	430
correct amount	422
Accuracy/%	98.14%
Recall rate/%	98.66%
F1 value/%	98.14%

Table 4. Time and place element identification results.

element type	time	Place
Accuracy/%	96.11%	92.18%
Recall rate/%	99.31%	95.54%
F1 value/%	97.44%	94.33%

4. Conclusion

This study starts from data rights, and after relevant demonstrations, data, as an object of rights, has property attributes, which should protect the legitimate rights and interests of individuals and enterprises over data. Therefore, on this basis, this study conducts research on infringement tracking of data storage to ensure the orderly, safe and efficient storage of big data. Then, a number of existing data infringement tracking technologies are compared and studied, which lays a theoretical foundation for subsequent research. After solving the theoretical foundation problems, this paper studies the application of a number of technologies in data infringement tracking. In particular, the in-depth application of the excellent characteristics of numerical information extraction technology in the data infringement tracking system is discussed, which lays the foundation for the construction of the entire data infringement tracking system. A data infringement tracking based on numerical information extraction technology is proposed. As controllers of personal data, corporate data storage providers have significant control over personal data. Therefore, corporate data storage providers have a greater duty of care than other individuals and organizations. It is also different from traditional tort liability in tort liability. At the same time, equal causation and contributory fault also apply it. The biggest difference between corporate data storage service provider infringement and traditional infringement is the way of damages. Infringement damages for corporate data storage service providers are limited to moral damages and punitive damages to compensate the information owner.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Cai, K., Immink, K. S., Zhang, M., & Zhao, R. (2017). On the Design of Spectrum Shaping Codes for High-Density Data Storage. *IEEE Transactions on Consumer Electronics*, 63, 477-482. <https://doi.org/10.1109/TCE.2017.015067>
- Cao, J. W. (2021). Mode Optimization and Rule Management of Intellectual Property Rights Protection of Educational Resource Data Based on Machine Learning Algorithm. *Complexity*, 2021, Article ID: 1909518. <https://doi.org/10.1155/2021/1909518>
- Chang, C. C., Chen, T. S., Wang, Y. K., & Liu, Y. (2017). A Reversible Data Hiding Scheme Based on Absolute Moment Block Truncation Coding Compression Using Exclusive OR Operator. *Multimedia Tools and Applications*, 77, 9039-9053. <https://doi.org/10.1007/s11042-017-4800-0>
- Esposito, C. (2018). Interoperable, Dynamic and Privacy-Preserving Access Control for Cloud Data Storage When Integrating Heterogeneous Organizations. *Journal of Network and Computer Applications*, 108, 124-136. <https://doi.org/10.1016/j.jnca.2018.01.017>
- Hoy, M. (2018). *Sweden Details New Proposal on Trademarks, Company Names* (p. 23). World Intellectual Property Report, No. 6.
- Jahner, K. (2018). *Dish Network Sinks Digital Rental, Targeted Ad Patents* (p. 11). World Intellectual Property Report, No. 8.
- Kumar, R., & Tripathi, R. C. (2017). An Analysis of the Impact of Introducing the Plagiarism Detection System in an Institute of Higher Education. *Journal of Information & Knowledge Management*, 16, Article ID: 1750011. <https://doi.org/10.1142/S0219649217500113>
- Langlois, G., & Mylan (2018). *Teva Lose Actavis Infringement Defense in Valeant Suit* (p. 35). World Intellectual Property Report, No. 6.
- Leung, P. (2018). *HTC's Bid to Move Patent Lawsuit Rejected by Appeals Court* (p. 23). World Intellectual Property Report, No. 6.
- Mcdonald, H., Berecki-Gisolf, J., Stephan, K., & Newstead, S. (2020). Preventing Road Crashes: Do Infringements for Traffic Offences Have a Deterrent Effect amongst Drivers Aged 40+? An Examination of Administrative Data from Victoria, Australia. *Transportation Research Part F Traffic Psychology and Behaviour*, 69, 91-100. <https://doi.org/10.1016/j.trf.2020.01.004>
- Menyasz, P. (2017). *Ontario Court Delays Providing Megaupload Data to U.S. Officials* (p. 6). World Intellectual Property Report, No. 5.
- Nayak, M., & Tyler, E. (2018). *Teradata Sues SAP for Trade Secrets Theft, Copyright Violations* (p. 9). World Intellectual Property Report, No. 7.
- Shields, T. (2017). *Sony Patent Complaint against Arris Cable Boxes Probed by U.S. Agency* (p. 27). World Intellectual Property Report, No. 5.
- Wang, Z. Y. (2021). Big Data and Predictive Research in Social Sciences: An Analysis of Application Scenarios Based on Conflict Prediction and Election Prediction. *Study and Exploration*, No. 6, 60.
- Woodman, S., Hiden, H., & Watson, P. (2017). Applications of Provenance in Performance Prediction and Data Storage Optimisation. *Future Generation Computer Systems*, 75, 299-309. <https://doi.org/10.1016/j.future.2017.01.003>