Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Legal Regulation Strategies for Civil Abuse of Litigation Rights Based on Algorithmic Research

Bihua Ou[1], Baomin Wang[1], Xiaoying Zhao[2,] ✉

[1] *Law School, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China*
[2] *School of Foreign Studies, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China*

ABSTRACT
The research selects the documents related to the legal regulation of civil abuse of rights of action as the research object, crawls the central and local legal regulation database through Python, and uses the social network analysis method to quantitatively analyze the dimensions of the subject of legal regulation from the composition of the subject of legal regulation, the density of the network, the centrality, and the cohesive subgroups, etc. The data preprocessing is carried out on the valid data obtained. Secondly, we pre-processed the acquired valid data, extracted high-frequency words using the improved TF-IDF algorithm, and obtained the probability distribution of the subject strength of "document-subject" and "subject-phrase-item" by calculating the degree of perplexity and utilizing the LDA subject model, and obtained the probability distribution of the subject strength at different stages of civil abuse litigation. In order to obtain the themes and evolution characteristics of the legal regulation of civil abuse of rights of action at different stages, the research results are combined with the results of the study from multiple dimensions. Finally, the research results are combined to design the strategy of legal regulation of civil abuse of rights of action from multiple dimensions.

Keywords: TF-IDF algorithm, social network analysis, lda topic model, civil abuse of rights of action

1. Introduction

The right of civil litigation is the right of the parties to request the state to give protection in civil litigation, that is, the parties in their civil rights and interests are infringed upon or disputes arise in civil relations, enjoy the right to request the state court to exercise judicial power to protect civil rights and interests or to resolve civil disputes [1-3]. Of course, freedom is freedom with limits, and any rights enjoyed by citizens must be exercised within the limits of freedom and without jeopardizing the interests of the state, the public interests of society and the legitimate rights and interests of others. If any person exercises a civil right of action outside the freedom permitted by law, this constitutes an

✉ Corresponding author.
E-mail address: xiaoying_cherry@163.com (X. Zhao).
Received 03 October 2024; Revised 20 December 2024; Accepted 15 March 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-054

abuse of the civil right of action. Therefore, the abuse of the civil right of action, the infringement of the legal interests of not only the legitimate rights and interests of citizens, but also the order of society and the judicial order of the State, with great harm.

The state has the obligation to guarantee the parties to exercise the right of action, through the law to ease or resolve disputes, in order to protect the legitimate rights and interests of the parties as well as to stabilize the judicial order, to achieve judicial fairness and justice, and to ensure the stable development of society. But the abuse of civil litigation, but precisely with the establishment of civil procedure law task deviation, the behavior occurs in order to create disputes, so that the original does not need judicial resources to intervene to achieve a stable state of the legal order, but intentionally create a conflict, and then require judicial intervention to restore the stability of the state again [4-5]. Abuse of civil litigation is against the principle of honesty and credit of civil litigation law, the legitimate rights and interests of others. The principle of good faith, each litigant is required to be honest and trustworthy, in order not to infringe on the state, society or the legitimate interests of others as a prerequisite for the implementation of litigation behavior, the exercise of litigation rights [6]. However, the abuse of civil litigation behavior caused by the implementation of the harmful results is to disrupt the judicial order, infringement of the legitimate rights and interests of others, including the property rights of others, the right to reputation and so on, obviously violates the principle of honesty and good faith.

The laws in force in each country are not so perfect and have some institutional shortcomings. And litigants, especially those who do not want to accept their own defeat, and self-defense litigants, use these shortcomings to conduct malicious litigation, abuse of personal right of action, causing waste of resources to government agencies and other situations, as well as infringing on the legitimate rights and interests of the other party [7]. Among them, the abuse of procedural rights is a kind of abuse of the right to sue, but civil procedural fairness and good faith litigation is difficult to define a status quo, the civil abuse of the right to sue the development of the system has a certain degree of difficulty [8]. Literature [9] analyzes the legal behavior of the parties based on legal rights in civil litigation procedures, but is an abuse of procedural rights, and further analyzes its root causes, with the intention of optimizing the management of procedures, and controlling the abuse of procedural rights at a socially acceptable level. In addition, there are some poor procedural litigation strategies in civil litigation with the intention of using repeated arbitration to weaken the rights of the other party [10]. In this regard, literature [11] mentions the need for a comprehensive understanding of the motives of the parties involved in procedural abuses, as well as the actions of the relevant stakeholders, from legal, social, economic, and individual psychological aspects, in order to solve the problem of duplicative litigation, which may be an abuse of the right to litigation, at its source. The literature [12], on the other hand, makes legislative and non-legislative recommendations for abusive litigation situations and emphasizes that any change in abusive litigation situations must be balanced with the existing judicial proceedings and must not be biased. More desirably, Ontario Civil Procedure Rule 2.1 provides for the court's ability to dismiss all types of abuse of rights of action [13]. It not only saves litigation costs and resources, but also protects the rights of the other party. Therefore, there is still a need for some definite legal regulation to provide for civil abuse of the right of action in this case, the judge's practice and the abuse of the right of action party whether to dismiss the claim and not allowed to re-litigation and a series of issues. And in recent years, the algorithmic research let day by day, the application of a wide range of applications, in the regulation of the state's discretionary power also has aided decision-making applications [14]. This can be used as a reference, algorithm-assisted decision-making and prediction tools to help civil abuse of rights of action problem of legal regulation strategy development.

In this paper, we use Python to crawl the relevant legal policies related to the regulation of civil abuse of the right to sue, and use social network analysis to conduct text mining from the composition of the subject of legal regulation, network density, centrality, and cohesive subgroups to determine the distribution of subgroups of the subject of legal issuance. On this basis, the improved TF-IDF is applied

to analyze the word frequency of text keywords to derive the core vocabulary. The LDA subject model is used to model the text data, calculate the subject strength of each subject, and identify the subject according to the subject-word item distribution, and determine the number of subjects in the model with the help of visualization technology.

2. Text mining and analysis of legal regulation of civil abuse of rights of action

2.1. Legal Regulation Text Mining Model and Data Collection

2.1.1. Legal regulation text mining process. The flowchart for the construction of the legal rule text mining model in this study is shown in Figure 1. Specifically, the construction of legal regulation text mining can be divided into 4 steps:

1) Collection of legal regulation text on civil abuse of litigation rights, applying web crawler technology to collect raw legal regulation text from relevant legal databases in China and the United States respectively, and constructing the raw corpus, which is saved in machine-readable form for subsequent processing.

2) Data preprocessing, the original legal rules corpus is cleaned to eliminate the text noise and get the model training corpus.

3) LDA model training, the original corpus through the training to get the civil abuse of the right to sue the legal regulation of the "theme - word" distribution, and then calculate the theme heat to get the civil abuse of the right to sue the legal regulation of the hot topic distribution.

4) Word2Vec training, combining the words in the training corpus with the "topic-word" distribution of the LDA model to represent the topic vectors, and obtaining the topic evolution trend of legal rules according to the policy topic results of each legal rule stage.

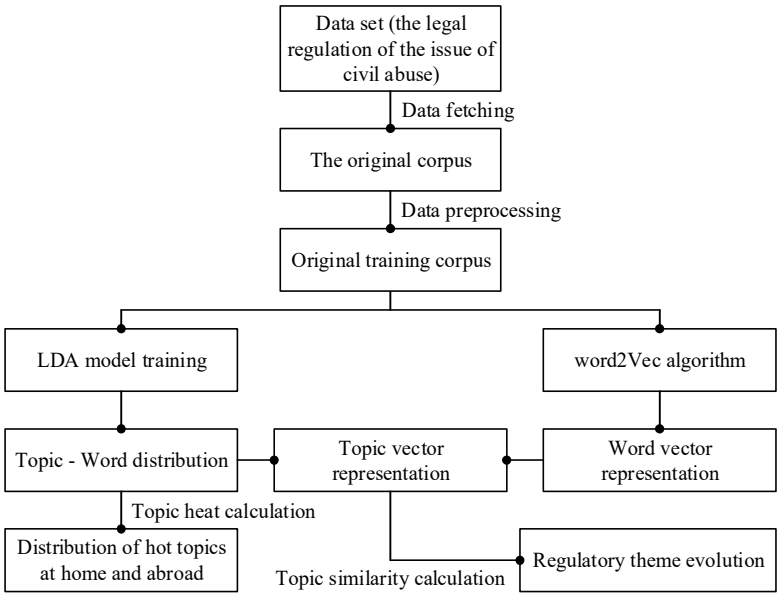


Fig. 1. The legal rules of text mining model construction process

2.1.2. Data sources. This study focuses on analyzing the current status and characteristics of the implementation of legal regulation of civil abuse of litigation rights in China, and retrieving and crawling data from legal databases through Python. In this paper, the scope of legal texts targeted for collection is defined as laws, regulations, rules, etc. issued in the form of documents by power or administrative organs at all levels, while expert consultation comments, meeting minutes, draft laws, etc. are not included in the scope of data collection for this study. The textual data come from the database of laws and regulations of the Ministry of Justice of China, the China Court Network-Legal

Library, the National Standardization Administration, the Procuratorate Laws and Regulations Library, the National Laws and Regulations Database and the legal regulation related to the issue of civil abuses of the right to litigation, with some chapters explicitly mentioning the legal regulation of the issue of civil abuses of the right to litigation.

2.1.3. Text pre-processing. After crawling and collecting data at all levels, it is necessary to perform data cleaning preprocessing operations. For Chinese text, first of all, the python open source tool of jieba Chinese word splitting is utilized to slice the statements of each statute, so as to make them into machine-readable forms of text content. And set the deactivation word list to eliminate the semantic dummy words, punctuation marks or adjectives and nouns that have no effect on the semantics of the subject, deactivation words can improve the efficiency of information processing and save computer processing resources, which is an important step in text preprocessing. In order to ensure the accuracy of the model theme mining, professional nouns related to personal privacy protection are manually input as the key lexicon to ensure that professional proper nouns will not be cut off.

2.2. Thematic model of the text on legal regulation of civil abuse of rights of action

2.2.1. TF-IDF model with improvements:

1) TF-IDF model. Among all the keyword extraction algorithms, the principle of TF-IDF is simpler, easier to understand and more effective.TF-IDF (term frequency-inverse document frequency) is a commonly used weighting technique, which is often used in information retrieval, text mining, etc.[15]. TF-IDF is a keyword extraction algorithm based on statistical features, which determines the degree of representation of a word or phrase for a document in the corpus by means of statistical features such as word frequency and inverse document word frequency. Eliminated by the inverse document frequency in the whole corpus, it can weaken the importance of the commonly used high-frequency words in most of the articles in the corpus, let TF be the word frequency in which a word occurs in a certain document, and IDF is the logarithm of the inverse of the number of documents containing the word to the total number of documents in the corpus, the TF-IDF formula is:

$$tf-idf_{ij}=tf_{ij}\times idf_{ij}=\frac{n_{ij}}{\sum_k n_{kj}}\times\log\left(\frac{|D|}{\left|\left\{j:t_i\in d_j\right\}\right|}\right) \tag{1}$$

where n_{ij} is the frequency of occurrence of word t_i in document d_j and D is the total number of corpus documents.

Since there is a phenomenon that the high word frequency of individual feature words suppresses other lower frequency words, in order to eliminate this effect, normalization can be used as follows:

$$tf-idf_{ij}=\frac{tf-idf_{ij}}{\sqrt{\sum(tf-idf_{ij})^2}} \tag{2}$$

Wanting to apply this method to legal rule text analysis requires certain improvements.

2) Improved TF-IDF model. In this paper, based on the understanding of the original theory and the existing improvements, the characteristics of legal rule texts are considered: rigor. The rigor of the law involves the masses, obtaining opinions or references from a wide range of people, and constantly modifying and revising to finally get a perfect legal document. The improved formula is expressed as:

$$tf-idf_{ij}^{;}=tf_{ij}\times idf_{ij}^{;}=\frac{n_{ij}}{\Sigma_k k_{kj}}\times\log\left(\frac{\sum E}{\Sigma_{\left\{j:t_i ed_j\right\}}E_j}\right) \tag{3}$$

2.2.2. LDA Subject Modeling. The basic idea of LDA topic model is that there are multiple potential topics in each text, link the text to the text according to the potential topics, and then use the distribution vector of the text topics to represent the text features [16].The LDA topic model is shown in Fig. 2.

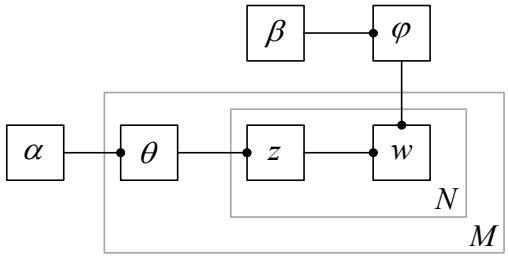


Fig. 2. LDA topic model

In Fig. 2, θ denotes the topic distribution, φ denotes the word distribution, α denotes the Dirichlet prior parameter for θ, β denotes the Dirichlet prior parameter for φ, z denotes the topic extracted from θ, w denotes the final generated word, M denotes the number of documents, and N denotes the number of words.

The LDA topic model adheres to a certain distribution by sampling words w through topic z and prior parameter β, and topic z is extracted from Dirichlet prior parameter α, which $P(w_i \mid f_i)$ is known in any text f, and by training on text-word $P(w_i \mid f_i)$, text-topic $P(z_k \mid f_i)$ and topic-word $P(w_j \mid z_k)$ are obtained, whose joint probability distributions $P(w_i \mid f_i)$ have the formulas shown below:

$$P(w_i \mid f_i) = \sum_{k=1}^{K} P(w_j \mid z_k) P(z_k \mid f_i) \tag{4}$$

The generation probability $P(f_i, w_i)$ for each word in the text is shown below:

$$P(f_i, w_i) = P(f_i) P(w_i \mid f_i) = P(f_i) \sum_{k=1}^{K} P(w_j \mid z_k) P(z_k \mid f_i) \tag{5}$$

The estimated parameter θ is:

$$\theta = (P(w_j \mid z_k), P(z_k \mid f_i) \tag{6}$$

In order to completely realize the “probabilistic”, LDA topic model as a random variable parameters. LDA topic model has three layers of text, topic, word structure, the training of the text-topic $P(z_k \mid f_i)$ and topic-word $P(w_j \mid z_k)$, so as to be able to identify the topic information in a large-scale text set.

The LDA topic model training requires only the text set and the number of topics, given that the text dataset of legal rules related to civil abuse of rights of action used in this thesis is large in number and is not annotated, so the LDA topic model is selected to model the dataset.

2.2.3. Selection of algorithm parameters:

1) Setting a priori parameters. The core of LDA topic modeling is the Dirichlet function, and the Dirichlet distribution needs to determine the values of two a priori parameters α and β, which represent the a priori distributions of document-topic and topic-word, respectively, and the setting of these two a priori parameters directly affects the reliability of the analysis results of the LDA topic model. According to the experimental studies conducted by existing scholars on the setting of the relevant parameters: for a priori parameter α, Jonathan compares the human-learned topic patterns

with the LDA topic generation model by setting different α values, and the α values of the topic patterns that agree with the human judgment fluctuate above and below $1/k$, so the study sets α to $1/k$; for a priori parameter β, Thomas et al. believe that a larger a priori value will Thomas et al. argued that a larger a priori value would cause the model to fail to discover more topics, and that a smaller a priori value could discover more domain-specific topics, and set the value of β to 0.01 in reference to the value of β used in the related study, so as to mine more interpretive topics in the corpus. Summarizing the basis of previous related studies, the setting of a priori parameters in the LDA topic model framework is completed.

2) Set the number of topics. Generally speaking, the number of topics k is the adjustment of the granularity of the topic model, and the larger the number of topics k will make the obtained topics more specific and nuanced, and there will be the problem that the topics will cover each other or be unable to differentiate the content of the corresponding topics; on the contrary, the smaller the number of topics k will lead to the failure of the text to carry out a complete representation of the subject, and will increase the complexity of the subject. In order to obtain effective LDA topic modeling results, it is necessary to set an appropriate number of topics k for the model.

In order to make the model analysis results have enough reliability and interpretability, the UMass algorithm is proposed for calculating the semantic consistency of the topic model, which measures the score of a topic by measuring the semantic similarity between the words with higher scores in the topic, and automates the judgment of the identification of the inferred topic superiority. The metric also helps one to identify the concepts implied by the themes. The formula for this metric is as follows:

$$coherence(V) = \sum_{v_i, v_j \in V} score(v_i, v_j, \int) \tag{7}$$

$score(v_i, v_j)$ is calculated as follows:

$$score(v_i, v_j) = \log \frac{D(v_i, v_j) + 7}{D(v_j)} \tag{8}$$

where V is a set of words belonging to a topic, ò is the smoothing coefficient, $D(v_i, v_j)$ calculates the number of documents containing word v_i, v_j, and $D(v_j)$ calculates the number of documents containing word v_j.

In this study, the Coherence metric is used to determine the optimal number of topics by comparing the change in model quality under different numbers of topics, using different numbers of topics as variables.

2.3. Social network analysis of the text of the civil abuse of rights of action statute

A social network is a collection of nodes and their connecting relationships between them, in layman's terms, a social network is a social network structure constructed from different nodes linked to each other by intermediate nodes. A node is a social enabler, which can be any social unit or social entity that represents a transmitter or receiver of information, such as an individual, company, college, village, organization, city, country, etc. Social network analysis can be used to judge the subjective enabler nature of nodes in the social network structure, including socioeconomic status, social value destination, and network roles. Connectivity refers to the relationship between nodes, which can be a real existing social relationship or a nominal relationship based on a certain principle. The arrow of the connectivity indicates the direction of information transfer, and the thickness or assignment of the connectivity indicates the frequency of information transfer or the strength of the connectivity relationship. Among the real relationships are concerns, friends and acquaintances.

2.3.1. Centrality measure of network nodes. Based on topology theory social network can be abstractly represented as $G=(V,E)$, set V denotes all the nodes in the social network and set E denotes the connectivity between the nodes based on a particular relationship. $g=|V|$ denotes the total number of all nodes.

1) Point degree centrality. Point centrality refers to the total number of connections of a node, i.e., the more direct connections it has, the higher its participation in that network. The specific concept can be interpreted in both absolute and relative terms:

In an undirected graph, the

$$X_{ij}=X_{ji}=\begin{cases}1 & \text{Node } i \text{ is related to node } j\\ 0 & \text{Node } i \text{ is unrelated to node } j\end{cases} \tag{9}$$

The absolute centrality is calculated as:

$$C_D(I)=d(i)=\sum_j X_{ij}=\sum_j X_{ji} \tag{10}$$

The relative centrality is calculated as:

$$C_{D'}(i)=\frac{d(i)}{g-1} \tag{11}$$

where, $g-1$ is the maximum possible absolute centrality degree in that network and hence the relative centrality degree is in the range of [0, 1], where 0 denotes the case of an isolated point, i.e., the point is not connected to any other node. 1 is to denote that the point is directly connected to any other node in the network.

In directed graphs, the point degree centrality is the point out degree and the point in degree, respectively.

$$X_{ij}=\begin{cases}1 & \text{Node } i \text{ Select node } j\\ 0 & \text{other}\end{cases} \tag{12}$$

Similarly $X_{ji}=1$ when indicates that node j selects node i.

The point out degree is calculated as:

$$C_{DO}(i)=d_O(i)=\sum_j X_{ij} \tag{13}$$

The point out degree is relative to the center degree:

$$C'_{DO}(i)=\frac{d_O(i)}{g-1} \tag{14}$$

$$C_{DI}(i)=d_I(i)=\sum_j X_{ji} \tag{15}$$

The point incidence is calculated as the relative centrality of the point incidence:

$$C_{DI}(i)=\frac{d_I(i)}{g-1} \tag{16}$$

2) Proximity centrality. The proximity centrality index describes the ability of a node to approach other nodes without being controlled by other nodes, if the sum of the distances between a node and other nodes in the network is smaller, the higher the proximity centrality of the node is, indicating that the node is more closely connected to other nodes, and the node is able to make the information spread further in the network, the proximity centrality is calculated by the formula:

$$C_C(i) = \left[\sum_{j \neq i}^{g} d(i,j)\right]^{-1} \tag{17}$$

where $\sum_j x1^g d(i,j)$ denotes the sum of distances between node i and all other nodes except node i.

The standardized proximity centrality degree is:

$$C_{C'}(i) = \frac{g-1}{\sum_{j=1}^{g} d(i,j)} \tag{18}$$

It can be regarded as the inverse of the average distance between node i and all other nodes, so $C_c^{'}(i) \in [0,1]$. When a node in the network is adjacent to all other nodes, it means that the node has the maximum proximity to any other node, and at this time, the standardized proximity centrality degree of the node is 1.

3) Mediator centroid degree. Intermediary center point is the point on the shortest path between two nodes, and its intermediary center degree describes the size of the intermediary role of the node to contact other nodes that do not have a direct connection between each other, if the number of shortest paths through the point the more, indicating that the intermediary role of the point the more, and the information in the network can flow faster.

Its normalization formula in the undirected graph is:

$$C_B(i) = \frac{2\sum_{j<k} g_{jk}(i) / g_{jk}}{(g-1)(g-2)} \tag{19}$$

The normalization formula in directed graphs is:

$$C_B(i) = \frac{\sum_{j<k} g_{jk}(i) / g_{jk}}{(g-1)(g-2)} \tag{20}$$

where g_{jk} denotes the number of shortest paths through node j and node k, $g_{jk}(i)$ denotes the number of shortest paths through node i in the shortest paths of node j and node k, and $\sum_j g_{jk}(i) / g_{jk}$ is the absolute mediator centrality of node i.

2.3.2. Structural Hole Theory

1) Effective size. This is the size of the individual network minus the redundancy, which is the average degree of the individual network members of the node, calculated as follows:

$$ES_i = \sum_j (1 - \sum_q P_{iq} P_{jq}) = n - \frac{1}{n} \sum_j \sum_q P_{jq} \tag{21}$$

In the above equation, $_n$ represents the degree of node i, j denotes the node that is directly connected to node i, while q denotes the node that is jointly connected to node i and node j, and P_{iq} denotes the proportion of weight that node q has among all the connected nodes of node i, and similarly for P_{jq}.

2) Efficiency. It refers to the ratio of the effective size of the actors to the actual size, in general, the efficiency of the nodes that are in the position of the structural hole is relatively large, then the efficiency of node i is calculated as follows:

$$EF_i = \frac{ES_i}{n} \tag{22}$$

3) Limit system. It refers to the node in the network by all the other points of the sum of the limitation system, the larger the value, it means that the node itself in the network, the greater the dependence, the more unlikely to occupy the structure of the hole position. Node i is subjected to the limiting system of a particular node in the network measured by the following formula:

$$C_{ij} = g(P_{ij} + \sum_{q,qzi,qz\,j} P_{iq} P_{qj} g)^2 \tag{23}$$

Then the restriction system of node i is:

$$C_i = \sum_j C_{ij} \tag{24}$$

2.3.3. Community structure detection. Community structure detection algorithm based on edge intermediacy. The basic idea is that edges connecting separate modules have high edge centrality because all shortest paths from one module to another must cross them. The algorithmic step is then to remove the edge's with the highest intermediary centrality by repeated iterations until no edge exists in the network.

2.3.4. Factions. A faction is, by the relevant definition of graph theory, a maximal complete subgraph consisting of at least three points, where all nodes within the graph are neighboring each other and the faction is not contained by any other faction. In an undirected graph, all points of a faction are in direct reciprocal relationships, and the relationships between members inside the faction are closer relative to the outside.

Faction Common Algorithm - Fast Discovery Algorithm

The specific steps of this algorithm are as follows:

Step 1: Initialize the factions, $_n$ node in the network is initialized to $_n$ factions. Where elements e_{ij} and a_i of matrix $_e$ measure the formula:

$$e_{ij} = \begin{cases} \frac{1}{2m} & \text{There is an edge between vertices } i \text{ and } j \\ 0 & \text{Or else} \end{cases} \quad a_i = \frac{k_i}{2m} \tag{25}$$

Step 2: After merging all the faction pairs in the network, calculate their modularity growth ΔQ and then merge ΔQ the largest faction pair while updating matrix $_e$ and modularity Q .

$$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j \tag{26}$$

Step 3: Repeat the previous two steps until there is only one faction left.

3. Analysis of the legal regulation of civil abuse of rights of action

3.1. Quantitative indicators of the subjective dimension of legal regulation

UCINET is one of the social network analysis software widely used in academic research, which shows the connection between data through visualization. In this paper, we use the analysis indexes of the three dimensions of cohesive subgroup analysis, point degree centrality analysis and network density analysis in UCINET to get the evolution trend of the dimension of the legal regulation subject of the civil abuse of right of action problem, firstly, we construct the N×N order adjacency matrix between the legal regulation subjects of the civil abuse of right of action problem, and if there is a joint issuance between the two subjects, we assign a value of 1, otherwise, it is 0. The collation of the data can be find

that there are 10 policy subjects involved in joint issuance in this research. Secondly, the matrix is imported into UCINET to draw the cooperation network mapping between policy subjects, and on the basis of analyzing the structural characteristics of the cooperation network, the data relationship is displayed through the intuitive network mapping, which provides scientific theoretical basis for the study in the paper.

3.1.1. Network density analysis. After binarizing the data, the network nodes composed of 10 joint issuing subjects, calculated by the Network tool in UCINET as shown in Table 1, the network density is 0.5944, and the density between the network nodes is at a moderately high level, which indicates that the connection between the nodes presents a relatively close state, and that there is a certain degree of cooperation between the subjects of the legal regulation of the issue of civil abuse of the right to sue. At the same time, the standard deviation between network nodes is very large, which means that there is a large difference in the degree of connection between different classes of issuing subjects. From this, it is inferred that there is a network phenomenon of small groups among the subjects of legal regulation of civil abuse of rights of appeal issues, and later on, in order to verify this hypothesis, this paper will also analyze the degree of centrality and cohesive subgroups indicators among the subjects of the policy.

Table 1. Network density

Parameter	Density
Density(matrix average)	0.5944
Standard deviation	3.2566

3.1.2. Centrality analysis. **Table** 2 shows the point centrality of the 10 subjects of joint issuance measured by UCINET software, and the point centrality of the National People's Congress (NPC) deputies and the Standing Committee of the National People's Congress (NPCSC) is the highest at 35.000, which indicates that the NPC deputies and the NPCSC have the closest connection with other nodes, and have the most prominent roles and status in this network. They are followed by local representatives and the State Council with a point centrality of 8.000, local governments with a point centrality of 7.000, ministries and commissions of the State Council with a point centrality of 6.000, and the people's government with a point centrality of 2.000, which accounts for the lowest proportion.

The data shows a large difference in the point centrality of the NPC deputies and the NPC Standing Committee compared to the other policy-issuing subjects, indicating that the NPC deputies and the NPC Standing Committee occupy a central position in the overall network, and that the rest of the policy-issuing subjects play a supporting role in the legal regulation of civil abuses of rights of action.

Table 2. Point centrality of the publication subject

Issuing unit	Point center	Proportion
NPC representative	35.000	0.303
The standing committee of the National People's Congress	35.000	0.303
Local representative	8.000	0.088
The state council	8.000	0.079
Local government	7.000	0.071
The ministries of the state council	6.000	0.062
People's government	2.000	0.026

3.1.3. Analysis of cohesive subgroups. The policy subjects of the legal regulation of civil abuse of rights of appeal are divided into three factions, as shown in **Table** 3. The presence of NPC deputies in the three factions indicates that NPC deputies are highly involved in the issuance of legal regulation of

civil abuses of rights of action, and are the core members of the entire network of policy subjects by jointly issuing relevant policies with different policy subjects. In addition the State Council exists in two factions and plays an active role in the joint policy subjects. The rest of the policy subjects are present in only one faction, with average interaction in joint issuance.

Table 3. Factional classification of policy subjects

Factions	Policy subject
1	NPC deputies, the NPC Standing Committee, local deputies, The State Council, ministries and commissions of The State Council, local governments, and people's governments
2	The NPC deputies, the State Council, the Supreme People's Court, and the Supreme People's Procuratorate
3	NPC deputies, The State Council, and administrative organs

Using UCINET's block model to analyze the cohesive subgroups of policy subjects, **Figure** 3 shows the distribution of subgroups of policy issuing subjects, and **Table** 4 shows the density matrix of subgroups of issuing subjects, which can be found to be divided into six subgroups, and from the point of view of the internal connection of the subgroups, the density of subgroup 1, which is composed of the deputies to the National People's Congress (NPC) and the Standing Committee of the National People's Congress (NPCSC) is the largest, at 3.000, which indicates that the deputies to the NPC and the Standing Committee of the National People's Congress (NPCSC) are closer in their joint issuance. From the point of view of inter-subgroup connection, subgroup 1 is most closely connected with subgroup 5, i.e., the Standing Committee of the National People's Congress (NPC), the Standing Committee of the NPC, and the local governments form a close cohesive subgroup phenomenon, and, secondly, subgroup 2 and subgroup 5 also maintain a closer connection, i.e., the State Council and the cohesion of the local governments are more obvious as well.

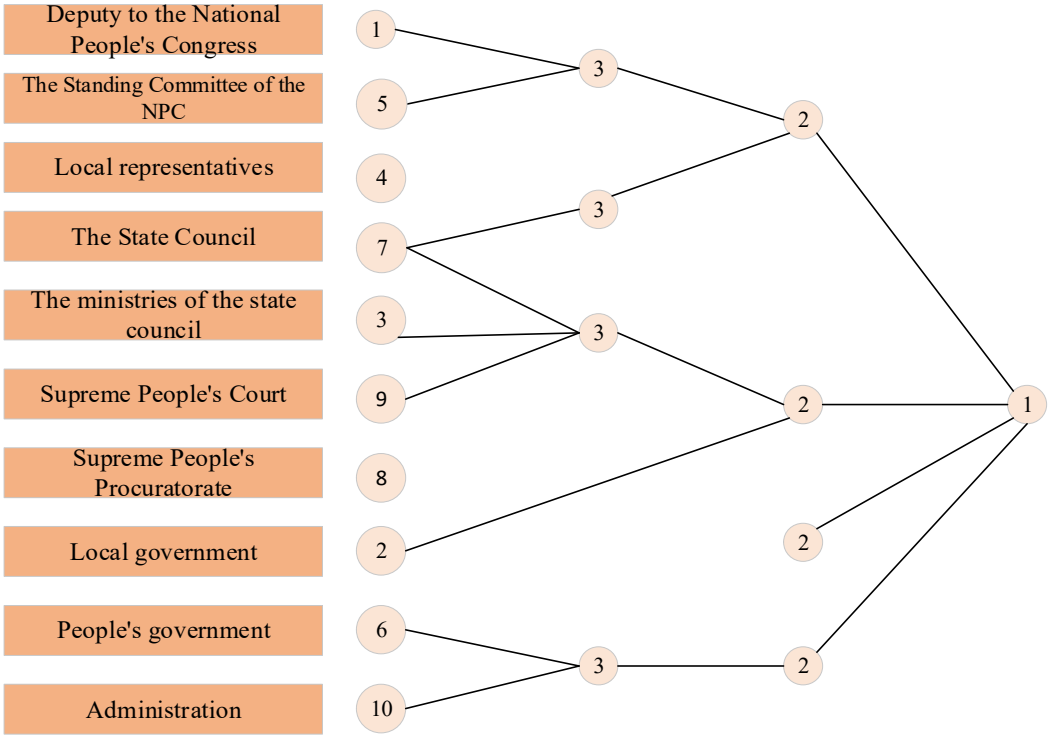


Fig. 3. The density matrix of the published subject subgroups

Table 4. Subject subgroup density matrix

Subgroup	1	2	3	4	5	6
	Density	Density	Density	Density	Density	Density
1	3.000	0.400	2.000	0.400	11.000	0.400
2	0.400		2.000	2.000	5.000	0.000
3	2.000	2.000	2.000	2.000	2.111	0.000
4	0.400	2.000	2.000		2.000	0.000
5	11.000	5.000	2.111	2.000		0.000
6	0.400	0.000	0.000	0.000	0.000	2.000
R^2=0.732						

3.2. Analysis of Hot Topics in Legal Regulation of Civil Abuse of Litigation Rights

3.2.1. Improving TF-IDF keyword analysis. The improved TF-IDF algorithm was used to extract words with key meaning from the text data that had been processed by word splitting and word deactivation. The results of the pre-processed logistic policy text word-splitting were calculated with the help of the TfidfTransformer algorithm in the Python programming language. This step involves not only the statistics of the frequency of occurrence of individual words in the text, but also considers the distribution of these words in the entire set of legal texts, thus ensuring that the extracted keywords are both representative and accurately reflect the theme and focus of the text. In order to demonstrate the application effect of the improved TF-IDF algorithm more intuitively, this paper selects the 38 keywords with the highest TF-IDF values of the text on civil abuse of litigation rights from the calculation results, and arranges them in the order of their TF-IDF values from high to low. The statistical results of some keywords are shown in **Table** 5.

As you can see from the above table, the word "idea" stands out with a high frequency of 0.124 and becomes the keyword with the most occurrences. This shows that the concept of law occupies a dominant position in the issue of civil abuse of litigation rights, and also reflects that the main reason for the occurrence of civil abuse of litigation rights is the lack of awareness of the law by the parties. This was followed by keywords such as "system", "fear", "understanding", "awareness" and "rights and interests", which together caused the problem of civil abuse of litigation rights.

Further analysis of these keywords reveals that they are centered around the themes of malicious prosecution, false prosecution, and repetitive prosecution. This indicates that these three types of lawsuits are the main ones that appear in the issue of civil abuse of rights of action, reflecting their influence on the issue of civil abuse of rights of action.

Table 5. The keyword statistical results

Sequence	Key words	Word frequency	Sequence	Key words	Word frequency
1	Idea	0.124	20	Reputation	0.031
2	System	0.117	21	Collusion	0.028
3	Fear	0.083	22	Take advantage of	0.025
4	Understand	0.052	23	Conspiring	0.023
5	Consciousness	0.079	24	Impartiality	0.021
6	Equity	0.076	25	Consume	0.019
7	Law	0.073	26	Integrity	0.017
8	Convenience	0.068	27	Subjectivity	0.015
9	Fiction	0.065	28	Risk	0.013
10	Overstate	0.064	29	Rights	0.011
11	Benefit	0.057	30	Hide	0.009
12	Fraud	0.055	31	Forgery	0.005

13	Legitimacy	0.054	32	Competition	0.003
14	Awe	0.048	33	Performance	0.0022
15	Endorse	0.047	34	Principle	0.0019
16	Reduplication	0.045	35	Qualification	0.0017
17	Flippancy	0.039	36	Time	0.0015
18	Falsehood	0.035	37	Factor	0.0013
19	Time	0.033	38	Behavior	0.0011

3.2.2. Optimal number of themes. By constantly adjusting the number of iterations of the model, it was found that the model started to stabilize at the number of iterations niters = 200. Under this setting, the change in perplexity is shown in **Figure** 4. When the number of topics is 21, the perplexity degree is relatively small and the decline gradually slows down. Meanwhile, in order to verify the reasonableness of the selection of the number of topics, the pyLDAvis package of Python is called to carry out the visual analysis of the distance between topics, as shown in **Figure** 5. When the number of themes, K, is 21, the themes show a clear distinction between them. In summary, the optimal number of themes is confirmed to be 21, based on which the subsequent research is carried out.

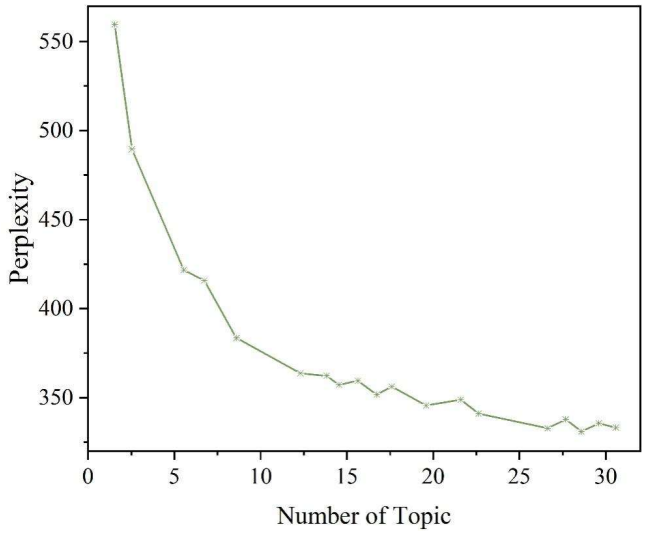
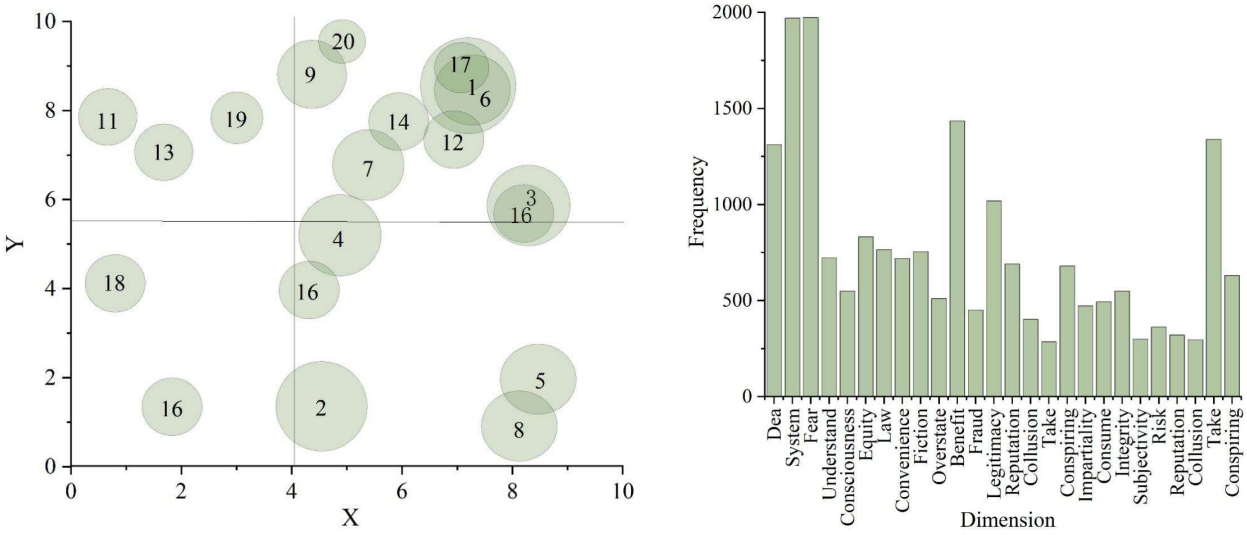


Fig. 4. Topic confusion



(a)Intertopic Distance Map (b) Topic frequency

Fig. 5. LDAvis visual interactive image

3.2.3. Thematic modeling:

1) Document-topic probability distribution. Through topic modeling, a "document-topic probability distribution table" containing 21 topics was obtained. The table presents the 56 policy documents processed by paragraphs in the form of a matrix, as shown in **Table** 6. Each row represents a different policy document, and each column corresponds to a different topic. The value in the matrix is the probability that this document belongs to the corresponding topic. The larger the probability value corresponding to a topic for a given document, the greater the strength of that topic in the document.

Table 6. Document topic probability distribution table

Documentation	Topic0	Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7	Topic8
1	0.0026	0.0024	0	0.0023	0.0026	0	0.256	0.0047	0.0542
2	0	0.645	0.0022	0	0.0037	0.0025	0	0	0.0047
3	0.0167	0	0.0014	0	0.0064	0	0.0047	0	0.0014
4	0	0	0	0.4675	0.0127	0	0.2453	0	0.0357
5	0	0.4554	0	0	0.0949	0	0.0094	0	0.0073
6	0	0.0063	0.7841	0	0	0.0951	0	0.0093	0
7	0.0034	0.0228	0	0.0467	0.7203	0	0	0	0.0092
8	0.05757	0	0.0562	0.0373	0	0	0.1784	0.0069	0.6423
9	0	0.0564	0	0.4743	0.0216	0.6512	0	0	0.0164
10	0	0.0574	0.0765	0	0	0.0156	0.0083	0.6496	0
11	0.0054	0	0.0432	0.0443	0.0678	0	0.0943	0.0343	0.1621
……	0	0.0357	0.4752	0	0	0.0043	0.0025	0	0.2452
21	0.0251	0	0.0676	0.1632	0.0854	0	0.0051	0.1321	0.0342

2) Topic-item distribution. In addition to the "document-topic probability distribution table", the "topic-word item distribution table" for 21 topics (1 to 20) is also obtained. The table is also in the form of a matrix, where the rows represent different topics and the columns represent the N words with the highest probability extracted from each topic. Based on the probability distribution of topic-phrase items, the distribution of words in each topic and the probability of each topic corresponding to each word are obtained, and finally the topic-phrase item distribution is obtained, and the words in the table are arranged in ascending order of probability. In this study, the top 9 words with the highest probability in each topic are selected as the representatives of the meaning of that topic, and the topics are judged and recognized manually. The data of the topic-word-item distribution is shown in **Table** 7.

Table 7. Topic and term distribution table

Serial number	Subject marking	Term term
1	Interest in the right of action	System, Understanding, Consciousness, Reputation, Law, Legality, Competition, Awe
2	Buy fake products on purpose	Fictitious, False, Exaggerated, Profit,Fraud, Good faith
3	Protection of litigation rights	Subjective, Risk, Justice, Rights and interests, Legality, Performance, Principle,Good faith
4	Misuse of legal procedure	Forging, Collusion, Concealment, Collusion, And energy
5	Civil suit	Time, Behavior, Consumption, Convenience, Fear, Reputation, And justice
6	Abuse of the right of action	Law ,Understanding ,Fiction, Factors, Indiscretion Time
7	Damage to rights and interests	Interests, System, Utilization, Rights, Qualification, Recognition, Duplication
8	Profit competition	System, Understanding, Fiction, Interests, Collusion, Collusion, Use
9	Fraud outfit	Conspiring, False creation, Concealment, Behavior, Fraud, Performance

3.3. Strategies for legal regulation of civil abuse of rights of action

In order to more effectively regulate civil litigation abuses, this section provides strategies for civil

litigation abuse regulation based on the previous thematic exploration of civil litigation abuse issues.

3.3.1. Establishment of a system of liability for tort damages for abusive claims. A subject of civil litigation rights who abuses the rights he or she enjoys may cause wrongful infringement on others, which is a tort in nature and should be sanctioned by law. For the abuse of rights, the subject of the abuse of sanctions, and relief for the victims, this should be a natural process. However, in China's current legislation, the prevention and control of abusive behavior of the norms are not perfect, the judiciary for the party for the other side of the abusive lawsuit filed by the litigation can only be affirmed through the judgment in its favor, but before the party due to maliciously being sued, suffered property and non-property damages, and often can not get adequate relief. Although the Guidelines provide for damages as a form of relief for parties whose interests have been impaired, they do not regulate the mechanism for their realization. Therefore, it is extremely necessary to recognize abusive litigation as one of the types of torts and to establish a corresponding damages system.

3.3.2. Improvement of administrative penalty mechanisms for abusive practices. In judicial practice, most of the abusive acts meet the requirements of tort, so it is necessary for the injured party to claim damages by way of a tort claim. And in some cases, the abusive behavior did not cause property or non-property damage to the other party, but the law should be in a position to adhere to the negative evaluation of abusive behavior, for such cases of the perpetrator can not be tolerated indulgence, but should be punished in the form of administrative penalties. The Code of Civil Procedure currently provides for the imposition of fines and judicial detention, and administrative penalties are imposed on parties who violate their obligations under the Code of Civil Procedure.

4. Conclusion

In this paper, Python is used to crawl the laws related to civil abuse of rights of action for text data collection, and text mining methods are used to pre-process the data, respectively, social network analysis is used, and words with key significance are extracted by using the improved TF-IDF method. Then the LDA model was applied for topic modeling, and the optimal number of topics was determined by the perplexity index. According to the results of LDA model training, the “topic-word” distribution identifies the topics, and 21 topics are finally obtained.

Funding

This research was supported by the Hebei Provincial High People's Court entrusted the project: "The Issue of Civil Abuse of Litigation Right in accordance with Regulations".

References

[1] Karmaza, O. O., Sarana, S. V., Fedorenko, T. V., Gurzhii, T. O., & Nefedova, A. V. (2018). The protection of civil rights and interests in the court. *Journal of Advanced Research in Law and Economics*, 9(8), 2622-2630.

[2] Cox, W. S. (2024). Common law practice in civil actions. *BoD–Books on Demand*.

[3] Warden, D. (2018). The Ninth Cause: Using the Ninth Amendment as a Cause of Action to Cure Incongruence in Current Civil Rights Litigation Law. *Wayne L. Rev.*, 64, 403.

[4] Nikulin, A. S., & Gabaraev, A. S. (2023). Genesis and Legal Essence of the Institution of Abuse of Law. *Economic Problems and Legal Practice*, 19(2), 55-59.

[5] Stifler, L. (2017). Debt in the Courts: The Scourge of Abusive Debt Collection Litigation and Possible Policy Solutions. *Harvard Law & Policy Review*, 11(1).

[6] Safaeva, N. R. (2021). Possibility of Applying the Category" Good Faith and Abuse of Procedural Rights" to Participants in Disputes Arising from Corporate Legal Relations under the Legislation of the CIS and Neighboring Countries. *LEX*, 74(5), 144.

[7] Morissette, Y. M. (2019). Querulous and Vexatious Litigants as a Disorder of a Modern Legal System. *Canadian Criminal Law Review*, 24(3), 265-312.

[8] Gajda–Roszczynialska, K. (2019). Abuse of Procedural Rights in Polish and European Civil Procedure Law and the Notion of Private and Public Interest. *Access to Justice in Eastern Europe*, (2), 53-85.

[9] Dauti, R., & Dauti, A. (2021). ABUSE OF PROCEDURAL RIGHTS IN CIVIL LITIGATION. *KNOWLEDGE-International Journal*, 49(5), 1039-1044.

[10] Gaillard, E. (2017). Abuse of Process in International Arbitration. *ICSID Review*, 32(1), 17-37.

[11] Yablon, C. M. (2018). THE VIRTUES OF COMPLEXITY: JUDGE MARRERO'S SYSTEMIC ACCOUNT OF LITIGATION ABUSE. *Cardozo Law Review*, 40(1).

[12] Fitch, E., & EASTEAL, P. (2017). Vexatious litigation in family law and coercive control: Ways to improve legal remedies and better protect the victims. *Family Law Review*, 7, 103-115.

[13] Kennedy, G. (2018). Rule 2.1 of Ontario's Rules of Civil Procedure: Responding to Vexatious Litigation While Advancing Access to Justice?. *Windsor Yearbook of Access to Justice*.

[14] Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. *Philosophical transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 376(2128), 20170359-20170359.

[15] Arar Al Tawil,Laiali Almazaydeh,Doaa Qawasmeh,Baraah Qawasmeh,Mohammad Alshinwan & Khaled Elleithy. (2024). Comparative Analysis of Machine Learning Algorithms for Email Phishing Detection Using TF-IDF, Word2Vec, and BERT. Computers, Materials & Continua(2),3395-3412.

[16] Li Youzhu,Zhang Miao,Liu Jinsi,Su Bingbing,Lin Xinzhu,Liang Yuxuan... & Zhang Junjie. (2022). Research on the Disturbance Sources of Vegetable Price Fluctuation Based on Grounded Theory and LDA Topic Model. Agriculture(5),648-648.