IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 February 2024, accepted 21 February 2024, date of publication 29 February 2024, date of current version 6 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3371494

RESEARCH ARTICLE

Establishment of a Copyright Regulations Knowledge Base and Development of a Case Recommendation System Utilizing Transformer and Graph Convolutional Network

JIAMING MO[1], JUNYU ZHENG[1], AND DEXU BI[2]

[1]Tourism Business School, Guangzhou Panyu Polytechnic, Guangzhou, Guangdong 510000, China
[2]Department of Elementary Education, Guangxi Police College, Nanning, Guangxi 530028, China

Corresponding author: Dexu Bi (bidexu@gxjcxy.edu.cn)

ABSTRACT To disseminate copyright regulations and address issues related to lapse, inexactitude, and inadequacy in the eminence of case data, we propose an initial methodology for the construction of textual content based on copyright regulations and cases. This methodology involves the processing of regulatory and case information, followed by the exploration of interrelationships. Subsequently, we use the Transformer algorithm for semantic information processing to extract nuances like conceptual terminology, pivotal keywords, and elucidating annotations from cases. This effort facilitates the creation of a concept index for cases, promoting case archiving. Concurrently, we introduce a methodology relying on keywords for the extraction of legal or case-related concepts. Recognising the multifaceted nature of cases with diverse sub-nodes, we propose a feature alignment approach grounded in Graph Convolutional Networks (GCN). This innovation serves as the basis for logically acclaiming copyright regulations within our knowledge framework. Empirical validations accentuate the effectiveness of our case recommendation system, showcasing an accuracy rate of 86.5%. Additionally, our compilation of copyright regulatory knowledge garners outstanding accolades in subjective evaluations.

INDEX TERMS Copyright law knowledge dataset, case recommendation system, integrating network information, machine learning.

I. INTRODUCTION

Copyright regulations are an important part of the modern intellectual property protection system. By stipulating and protecting the rights of copyright owners, they encourage and promote the progress of knowledge innovation, cultural creation, scientific and technological development, and other aspects. Copyright laws and regulations can protect the intellectual property rights of authors, artists, and inventors, ensure that they can obtain corresponding economic benefits and social reputation, promote innovation and development in the fields of culture, science and technology, art, and other fields, ensure that the public can legally obtain and use relevant knowledge, culture, and information resources, and prevent embezzlement, copying, or tampering. And maintain market order and fair competition [1], [2], [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Mohamad Afendee Mohamed.

Copyright regulations encompass diverse professional fields with a broad business scope, presenting challenges in qualitative assessment, legal adjudication, and sentencing due to inconsistent standards. To overcome these obstacles, leveraging network information dissemination to construct a comprehensive knowledge base for copyright laws is essential, breaking down information barriers [4], [5]. This approach facilitates the creation and sharing of online resources dedicated to copyright law, fostering widespread awareness. The knowledge base can take various forms, including websites, social media platforms, online forums, and blogs. Unlike prevalent case recommendation systems in e-commerce, medicine, libraries, and film, there is a notable gap in research for such systems tailored to the realm of copyright regulations. Consequently, there is a pressing need to explore and develop case recommendation systems specifically designed for the intricacies of copyright regulations [6], [7], [8].

At present, there are still many difficulties in the construction of a copyright knowledge base. The construction of a knowledge base requires so many data sources, and the acquisition of data is often a time-consuming and laborious process. Some data sources may require authorization or payment to access, and some data may have copyright or privacy issues that need to be processed and filtered. Data acquired from data sources usually needs to be cleaned and processed, including deduplication, missing values, data format conversion, and so on [9], [10], [11]. The quality of the data and how it is processed will affect the quality and usability of the final knowledge base. The data structure of the knowledge base needs to be properly designed to store and retrieve the data. A well-designed data structure can improve the efficiency and accuracy of knowledge base queries, but the structure of the knowledge base varies greatly in different fields, so it needs to be adjusted and optimised appropriately. The data in the knowledge base needs to be classified and labelled so that users can query and retrieve it [12]. Classification and annotation need to consider many factors, such as domain characteristics, user needs, classification standards, and so on, which require professionals to carry out more complex operations. The content of the knowledge base needs to be constantly updated and maintained to ensure its timeliness and accuracy. Knowledge updating and maintenance should be carried out regularly, and attention should be paid to the reliability and availability of data sources to avoid the degradation of the quality of the knowledge base due to the failure of data sources [13], [14].

In addition, there are also many difficulties in the case recommendation method based on the knowledge base of copyright regulations. The quality of the case data is missing, inaccurate, and incomplete, and it has to be entire cleaned up to increase the credibility and availability of the data [15]. Case recommendation needs to compute the difference among cases to recommend the most relevant cases. The similarity calculation method needs to be customised for different case types and fields, and the comprehensive use of multiple similarity calculation methods should be considered to boost the reliability of recommendations [16], [17]. Case recommendations need to be personalised according to the needs and preferences of users. The user's preference recognition needs to consider the user's historical behaviour, interests, cultural background, and other factors, and it needs to use a variety of data analysis methods and technical means for processing and analysis [18]. A case recommendation needs to adopt an appropriate recommendation algorithm. It needs to consider many factors, such as recommendation scene, data volume, and data quality, and needs to be selected and optimized reasonably. The outcomes should be elucidated for users, providing a comprehensive understanding of the rationale and basis behind the recommendations. The interpretation of recommendation results requires various methods, such as knowledge graph construction and data visualization, so that users can better understand and use the recommendation results [19].

To solve the above problems, this paper constructs a knowledge base of copyright regulations. It proposes a case recommendation system to provide assistance for professionals and reduce the work pressure of legal practitioners based on the concept of network information dissemination.

II. RELATED WORKS

A. RESEARCH STATUS OF KNOWLEDGE BASE CONSTRUCTION

The concept of a knowledge base originates from the field of intelligent control. The knowledge base is a system containing all available knowledge in a certain field, and it is an integral part of an automatic programming system.

The theoretical research on knowledge bases includes the construction principle, architecture design, construction mode, knowledge system, application framework, and so on. These results lay the foundation for the construction of a knowledge base. Literature [20] summarised the construction mode of the knowledge base through investigation, put forward the current ontology mode and alliance mode of knowledge base construction, and discussed the dynamic mechanism, coordination and incentive mechanism, control mechanism, and sharing mechanism of the knowledge base. Literature [21] sorted out the fundamental differences between knowledge bases and databases in the process of constructing the ontology-based FMEA knowledge base framework. Literature [22] focused on knowledge extraction, knowledge fusion, and application of knowledge graph updates in the study of knowledge base construction and system integration methods for multi-source heterogeneous data in big data environments. Literature [23] constructed the hierarchical architecture of the knowledge base, including the data resource layer, knowledge processing layer, knowledge storage layer, and application service layer. Guided by user needs, it constructed the system's functions and hierarchy of knowledge based on demand. Literature [24] discussed the important role of applying the knowledge base to services and sorted out the knowledge flow and the steps that can be implemented in the knowledge base system.

The knowledge scope of the knowledge base of copyright regulations mainly includes copyright-related legal cases and related human knowledge and documents. Copyright mainly includes a natural person, organisation, work, right, event, time, region, etc. Knowledge resources mainly provide knowledge content for domain knowledge bases, including databases and text compilations. The collection of knowledge resources should be as comprehensive as possible, including all types of knowledge. For example, copyright knowledge needs to be collected comprehensively from copyright laws and regulations and copyright case knowledge. In addition, we should also collect the knowledge resources of the author, work-related knowledge, and case-based knowledge of copyright.

B. RESEARCH STATUS OF CASE RECOMMENDATION SYSTEM

Employing big data and AI technology to achieve accurate similar case recommendations can make similar cases the reference and yardstick for new cases, which is conducive to improving the case-handling efficiency of legal workers, reducing their work pressure, and better realising the fairness and justice of the rule of law society.

Many researchers focus on the method of legal text processing, which makes good progress and is of great help to judicial workers. Literature [25] proposed a method to extract semantic representations that utilises common legal document formats to identify chunks of structural and semantic information and models them according to popular legal meta-patterns. Literature [26] ultilised Natural Language Processing (NLP) methods to generate annotated text. In addition, ontologies are widely used in the field of law. Literature [27] used the semantic web and ontologies for knowledge engineering of specific subdomains of Indian law, and each case was built into a file in eXtensible Markup Language (XML) format and saved into the system. Literature [28] also used ontology technology to build a search system in the legal domain. Since the terms include the features of words and handle various features such as multiple words and synonyms, the model is more friendly to non-legal people who do not know legal terms. Literature [29] combined the citation network with machine learning to obtain similarities among legal texts. They considered each paragraph of the legal text as a separate node and then measured the similarity score of all pairs of paragraphs. If the similarity between any two paragraphs was above a threshold, a strip was added between the documents containing the two paragraphs. Finally, document coupling was used to calculate the similarity of the texts. Literature [30] used the nearest neighbour by case feature to determine the similarities among cases. Literature [31] used the Bayesian statistic to achieve the similarity of cases by word frequency.

III. CONSTRUCTION OF COPYRIGHT LAWS AND REGULATIONS KNOWLEDGE BASED ON NETWORK INFORMATION DISSEMINATION

Copyright knowledge is divided into two parts: copyright regulations and copyright cases. The amount of knowledge about copyright regulations is small, but the quality of the laws and regulations is high. There are many copyright case knowledge texts, and the case texts are semi-structured data, so it is easy to extract knowledge from them. For legal cases, the general structure of judgement documents is clear, which is suitable for automatic construction. The planning and design of the copyright knowledge base is mainly preparatory work before construction. This step determines the knowledge, the technology adopted, the characteristics, the services provided, and the description language of the knowledge base.

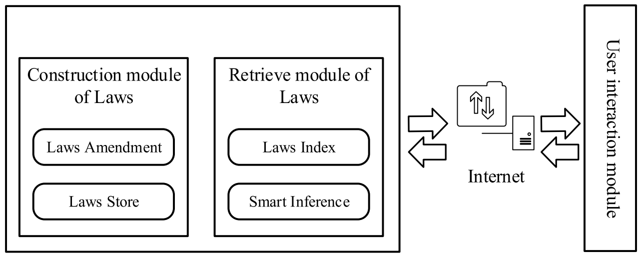


FIGURE 1. Construction of a copyright law knowledge base integrating network information dissemination.

According to the analysis of the characteristics of the copyright law knowledge base integrating network information dissemination, we focus on the modification method, storage mode, regulation index, and intelligent reasoning when designing the architecture of the knowledge base. In addition, we integrate the way of network information dissemination to connect our constructed copyright knowledge base to the Internet to facilitate the public's inquiries at any time. The knowledge base of copyright laws and regulations designed by us is shown in Figure 1. We construct the copyright law knowledge base through the Internet, and we can apply the communication online to achieve mass sharing and access to databases, which means network information dissemination.

First of all, we structured the copyright instances and saved them into the database, which can improve the efficiency of reading and writing. It should be noted that we store these laws and regulations and the corresponding case knowledge in the form of text and establish the corresponding index file to improve reading efficiency. To further optimise the storage space, we use WebSQL technology to support the deletion, modification, reading, and saving of file instances in the knowledge base, and the object is the text index of copyright regulations and copyright cases.

Then, we construct the text of copyright regulations and copyright cases in the knowledge base. In addition to the function of building ontology, which means copyright cases, we also need to provide the functions of ontology storage and ontology adding, deleting, modifying, and searching. Through the external framework of the knowledge base constructed manually, we add the basic functions of adding, deleting, modifying, and querying to the knowledge base. In addition, the update of copyright regulations and copyright cases needs to take into account the handling of their relationship. For example, when a new regulation is added, the concept related to the regulation should first be searched in the copyright regulations and copyright cases. If it already exists, the relationship should be directly established; if not, a new regulation should be created. The construction of regulations in the knowledge base of copyright regulations is the core part of the whole construction work because the whole construction work of the knowledge base of copyright regulations is based on regulations. The construction of regulations in the knowledge base of copyright regulations is a process of building relevant information layer by layer from top to bottom according the top ontology of copyright regulations. The concept structure is obtained by analysing the knowledge of copyright regulations, and then the concept is integrated from top to bottom according to the hierarchical structure of the concept. Finally, the structure, attribute and relationship of the concept are extracted.

The copyright content involved in this paper is laws and regulations and copyright cases. The content of laws and regulations is small, but the accuracy requirement is high. There are a large number of cases, but the case text is semi-structured data, and the information structure of the concept involved is obvious and easy to extract. Therefore, we adopt different ways to construct knowledge and cases of copyright regulation. As shown in Figure 2, we extracted keywords from the laws and regulations of copyright, constructed the concept, and saved the examples at the same time, where experts in the respective domain or annotators typically carry out artificial labeling. They utilise specialised tools or software interfaces to observe information such as data, text, or images and add relevant labels or classification information to each data point. Creating a database index typically involves selecting one or more columns in a database table and establishing data structures on these columns. Instance generation encompasses the utilisation of models and algorithms to produce new data points. However, considering the number of cases, we use the semantic information intelligent processing algorithm (Transformer), which is involved in the intelligent extracting module, to extract the information such as concept words, keywords, and annotations in the cases and realise the generation of a case concept index and save case instances. Transformer is a deep learning model architecture that has advantages in processing sequential data and has become a standard model architecture for many NLP tasks. The core of Transformer is Self-Attention, whose formula is as follows:

$$Q = f_{all}\ W^{Q} \tag{1}$$

$$K = f_{all}\ W^{K} \tag{2}$$

$$V = f_{all}\ W^{V} \tag{3}$$

$$Z = \text{Softmax}\left(\frac{QK^{T}}{\sqrt{d_k}}\right)V \tag{4}$$

where Z is the enhanced feature, d_k is the dimension, W refers to the linear transformation, and f is the input feature. In the Transformer, Q (Query), K (Key), and V (Value) are the core components of the attention mechanism. These symbols represent three different representations of the input sequence after linear transformation, used to calculate attention weights. Query represents obtaining a Query matrix by performing a linear transformation on the input sequence. In the self-attention mechanism, it represents information

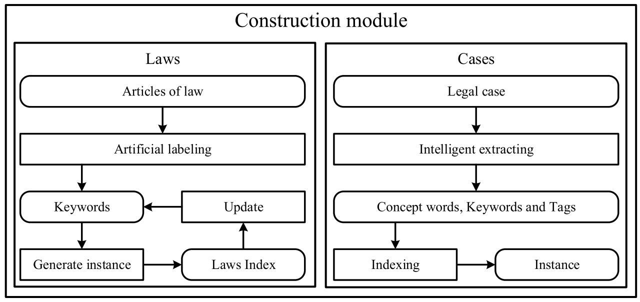


FIGURE 2. **Copyright las and cases building module.**

about the current position and is used for comparison with other positions. The key represents: Similarly, by performing a linear transformation on the input sequence, the Key matrix is obtained. Key is used to measure the similarity between Query and other positions, helping the model determine the attention weight assigned to each position. Value (V) represents the value matrix obtained by performing a linear transformation on the input sequence. Value is the information used in calculating the attention-weighted sum, which determines the impact of other positions on the current position.

Finally, considering that the main purpose of building a copyright knowledge base is to facilitate users retrieval and learning of copyright knowledge, providing users with copyright knowledge services is the ultimate purpose of building a copyright knowledge base. The copyright knowledge service is mainly the visual display function of knowledge and the knowledge retrieval function. The case recommendation system will be introduced in the next section.

IV. COPYRIGHT CASE RECOMMENDATION SYSTEM

With the progress of artificial intelligence and the openness of copyright judgement documents, case recommendation has become a hot issue. In order to support the construction of the knowledge base of copyright regulations and further expand the dissemination scope of copyright regulations in the network, we designed a copyright case recommendation system, as shown in Figure 3, to facilitate the matching and dissemination of copyright cases.

Firstly, we extract keywords from the query text and use Transformer to quantify its features. Then, it is compared with the features of keywords, concept words, and tags of cases in the copyright regulation knowledge base. Considering that each case contains many different child nodes, we propose a GCN-based method to match the features. GCN excels in effectively capturing the structural information inherent in graph data. It achieves this by learning relationships between nodes and generating node embeddings, allowing for meaningful feature matching within the graph. This goes beyond relying solely on the local neighbourhood information of nodes. Through hierarchical information transmission within the graph, GCN facilitates the transmission and integration of feature information across nodes. This mechanism empowers nodes to leverage the holistic context of the entire graph for

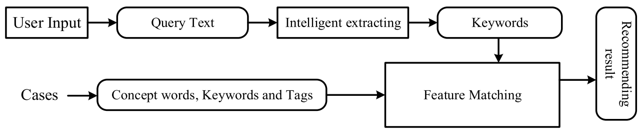


FIGURE 3. **Copyright las and cases building module.**

refining their representations, contributing more effectively to feature-matching tasks.

Furthermore, GCN adeptly learns low-dimensional embedding representations for each node, effectively serving as distinctive features. In the context of feature matching, these node embeddings serve as a metric for gauging similarity between nodes, thereby facilitating feature matching and alignment. The semantic features and their relative position relationships with the keywords of copyright cases and query texts are described by relation features. The set of vertices and edges is determined to establish a directed topological graph G=(V, E, s), where V denotes the nodes in graph G, and E refers to the relative positions of nodes relative to other nodes, including the semantic relative positions of copyright keywords, and s denotes the direction of edges. Then, the feature matrix H=(V, E) is shown by the following formula (5), as shown at the bottom of the next page:
where n refers to the number of the nodes and s=$\{-1, 0, 1\}$, -1 and 1 can refer to the different examples of effectivity and ineffectiveness, respectively, and 0 represents no connection between two nodes. The direction information is represented by the adjacency matrix A of the graph, in which there are positive, negative, and 0 values. These three cases represent positive connection, negative connection, and no connection between nodes, respectively, and the adjacency matrix A is expressed as:

$$A = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ -1 & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & 1 \\ 0 & \cdots & -1 & 0 \end{pmatrix} \tag{6}$$

The adjacency matrix of a graph is a two-dimensional matrix utilised to portray the connections between nodes within the graph. In the case of a directed graph, the adjacency matrix takes the form of a square matrix, where elements A[i][j] signify connections from node i to node j. Positive values indicate positive connections between nodes, often denoting a directed edge from node i to node j, where i points to j, signifying a positive or dependency relationship. Conversely, negative values represent negative connections between nodes. This may indicate a directed edge from node i to node j, where i points to j, symbolising a reverse or inhibitory relationship. A zero value signifies the absence of a direct connection between nodes. In the adjacency matrix, a zero value denotes the lack of a directed edge between node i and node j, indicating the direct nonexistence of a connection. This representation provides a clear depiction of directional relationships within the graph, encompassing forward, backward, and unconnected scenarios. Such clarity proves valuable for analyzing graph structures, comprehending node interactions, and conducting network analysis.

Then, G = (V, E, s), obtained by the above method, is used to describe the child nodes and their attributes of the case. Relationship reasoning is carried out through GCN, and the role awareness node 'x' is defined to quantify the child nodes. Then, the relationship between the child nodes and the attributes and between the child nodes and the root node can be expressed as follows:

$$x_{i+1} = \sigma\left(W_0 x_i + \sum_{r\in R}\sum_{j\in N}\frac{1}{N}W_r W_j\right) \tag{7}$$

Among them, W represents a linear transformation, which is the learnable parameter in deep learning; R refers to the number of the root nodes; and the relationship between child nodes is reasoned by using the method of GCN, as shown in Figure 4. At the outset, it is imperative to initialise each node in the graph and assign an initial feature representation to it. Generate an adjacency matrix from the graph data, delineating the connection relationships among nodes. Typically symmetric, the adjacency matrix's elements denote the presence or weights of edges between nodes. Within GCN, the learning parameters encompass weight matrices responsible for propagating information between nodes, with these matrices being subject to learning during training. Utilise the adjacency matrix and weight matrix for feature propagation across each node. This process entails weighting and summing the features of each node with those of its neighbouring nodes, thereby updating the nodes' representations.

Finally, the cases in the knowledge base of copyright regulations were compared with the corresponding nodes in the query text, and similar cases in the candidate recognition results were screened out. In addition, the reasoning results of GCN are used to compare the features of different cases, and the closest case is selected to replace the features on the original node. The replaced cases strengthen the information in this part, and more effective semantic features can be extracted to improve recognition accuracy.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATASET AND PERFORMANCE METRICS

We use copyright datasets (https://zenodo.org/record/6395957#. ZGOJ73ZBztU, doi: 10.5281 / zenodo. 6395957) [32] to test our knowledge base and the effectiveness of the proposed recommendation system. The dataset comprises 6.5 million distinct licence files, providing a valuable resource for empirical studies on open-source licensing. It can be utilised for training automated licence classifiers, conducting natural language processing (NLP) analyses of legal texts, and exploring historical and phylogenetic aspects of Free and Open Source Software (FOSS) licensing. Our experimental setup is shown in Table 1. In addition, we set the weight decay term to 0.0001 and adopted Adam as the optimizer of the model. The selection of the epoch size in machine learning training is commonly guided by a synthesis of empirical experimentation, computational constraints, and the inherent characteristics of the dataset.

To begin with, opting for a larger epoch size might necessitate increased computational resources and prolonged training durations. Conversely, smaller epoch sizes may restrict the model from thoroughly exploring the entire dataset, potentially leading to underfitting. Therefore, the choice of an epoch size of 50 aims to strike a harmonious balance between computational efficiency and affording the model ample opportunity to glean insights from the data. Moreover, through an extensive series of experiments, it was observed that the model exhibited signs of both inadequacy and overfitting when trained with 50 epochs. This observation underscores the nuanced interplay between epoch size and model performance, prompting a critical reevaluation of the optimal training configuration for the dataset at hand.

To evaluate the model's performance, we used the mean average precision (mAP) as the evaluation criterion, which can be calculated as follows:

$$\mathrm{V_P} = \frac{\mathrm{V_{TP}}}{\mathrm{V_{TP}} + \mathrm{V_{FP}}} \tag{8}$$

$$\mathrm{V_R} = \frac{\mathrm{V_{TP}}}{\mathrm{V_{TP}} + \mathrm{V_{FN}}} \tag{9}$$

$$\mathrm{mAP} = \int_0^1 \mathrm{V_P}\left(\mathrm{V_R}\right) \mathrm{d}\left(\mathrm{V_R}\right) \tag{10}$$

where TP denotes positive samples correctly classified as positive, FP signifies negative samples erroneously classified as positive, FN indicates positive samples incorrectly classified as negative, and V means the number of the corresponding samples.

mAP is an indicator used to evaluate detection or recognition performance. In the fields of computer vision and machine learning, mAP is widely used to measure the accuracy of models in processing data. In particular, the mean Average Precision (mAP) is acquired through the computation of the average accuracy for each category, followed by the calculation of the mean value across all categories. mAP for models is typically calculated based on ground truth annotations. Ground truth annotations refer to the manually labelled or annotated information in the data.

B. TESTING OF COPYRIGHT CASE RECOMMENDATION SYSTEM

First, we conduct performance experiments on the sub-module of our method, specifically the sopyright case recommendation system, utilising copyright datasets. Simultaneously, we select models renowned for extracting exceptional features, including BiLSTM [33], Dlinear [34], Transformer [35], Transformer-XL [36], Deformer [37], and Informer [38], and assess their performance. The outcomes are presented in Table 2. Our method achieves mAP@0.95, mAP@0.75, and mAP@0.50 values of 0.782, 0.829, and 0.865, respectively. Notably, "mAP@N" denotes the threshold used for calculating the average accuracy (mAP) in the task, with N being the threshold. Comparing our approach with the feature matching method employed by the Transformer framework, we observe improvements in mAP@0.95, mAP@0.75, and mAP@0.50 metrics by 0.022, 0.046, and 0.054, respectively, in comparison to the basic Transformer. Furthermore, in benchmarking against Deformer and Informer, our method consistently outperforms them across all evaluation indicators. Additionally, a comprehensive comparison with basic linear models reveals that our method surpasses BiLSTM and outperforms Dlinear across all metrics. Notably, our method achieves substantial increases of over 1% in mAP@0.75 and mAP@0.50. Within the expansive knowledge base of copyright regulations, the pivotal determinant for achieving superior performance in mAP, derived from precision and recall rates, resides in the proposed Graph Convolutional Network (GCN) inference of case example nodes. To assess the stability of the model, we conducted a reevaluation of the experiments above, introducing changes to the model's training method. In this reassessment, we adjusted the initial learning rate to 0.002 and the batch size to 64. The retrained model yielded mAP@0.95, mAP@0.75, and mAP@0.50 values of 0.792, 0.824, and 0.862, respectively. Comparing these results with other methods, our model continues to outperform the alternatives, showcasing its consistent superiority. This outcome reinforces the stability and effectiveness of our model, even when subjected to variations in the training approach.

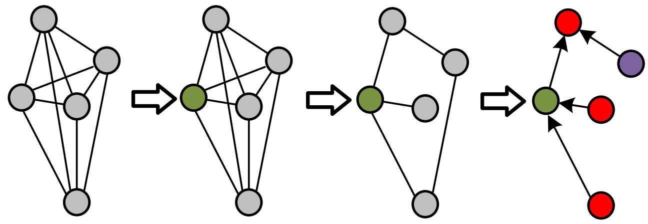


FIGURE 4. GCN feature quantization.

Subsequently, we demonstrate the model convergence of the feature-matching model within the case recommendation system in our approach, alongside semantic information processing in the construction of the copyright regulation knowledge base, as illustrated in Figure 5. It can be found that our proposed copyright regulations knowledge base and case

$$H = \begin{pmatrix} V_1 & -E_1(1,2) & \dots & 0 \\ -E_1(2,1) & \dots & \dots & \dots \\ \dots & \dots & \dots & -E_{n-1}(n,n-1) \\ 0 & \dots & -E_{n-1}(n,n-1) & V_n \end{pmatrix} \tag{5}$$

TABLE 1. Experiment details.

Types	Parameters
CPU	R7-7900
GPU	Rtx 3080Ti
Deep learning framework	Pytorch
Epochs	50
Batch-size	24
Initial learning rate	0.004
Momentum	0.90

recommendation system with integrated network information dissemination can achieve faster convergence in the training process and reach an ideal training state. From the curve, we can see that the training loss of our model decreases very quickly. By the 10th training round, the two models have reached a level of stability, and in the subsequent training, the gradients of our two models will further decrease until they plateau.

Finally, we conduct a comparative analysis of our system against other recommendation systems, including those outlined in Paper [39], Paper [40], Paper [41], Paper [42], and Paper [43]. The results are presented in Figure 6. We can conclude that our system can achieve the best performance with 86.5% mAP@0.50, which can outperform the other recommendation systems and prove the effectiveness of our method.

C. TESTING OF COPYRIGHT LAW KNOWLEDGE BASE

To verify the proposed knowledge base of copyright law integrating network information dissemination, we randomly select 200 copyright laws and regulations and use the method of this paper to build their knowledge. To have a clear comparison effect, we reproduce five additional knowledge base construction methods to compare with our method. Furthermore, we conduct a control test comparing the knowledge construction derived from our method with the actual implementation of copyright laws in real-world scenarios, denoting the conventional usage of this database in practice. The experimental input consists of 200 copyright laws, and we assess the similarity between the knowledge construction achieved through the method proposed in this paper and the knowledge construction observed in real-world scenarios of copyright laws and regulations. Firstly, we represent the data through graph embedding, representing the knowledge in each knowledge construction model as mathematical vectors. Then, choose the cosine similarity measurement method to measure the similarity between the representations of two knowledge construction models. In addition, before conducting similarity measurements, representations are usually standardised to ensure that scales of different dimensions or features do not bias similarity measurements. The results are shown in Figure 7. There are 110 pieces with similarity above 95% and 56 pieces with similarity above 75%, accounting for more than 80% of all samples.

We perform cross-tests on the knowledge base of copyright law constructed using various methodologies. Upon comparing the five reproduced knowledge construction methods, it becomes evident that our proposed knowledge construction method exhibits the highest comprehensive similarity [39] with real knowledge construction results, reaching an impressive 88.2%. Notably, our method maintains a swift computing speed, as illustrated in Table 3. Despite having a parameter count of 359.4M, our knowledge construction method experiences no adverse impact on calculation time, thanks to the integration of a network information transmission method in its design. The substantial number of parameters in our approach contributes significantly to its performance. Harnessing the parallel computing capability inherent in deep learning frameworks, along with the effective management of an extensive parameter set, our approach capitalizes on the accelerated processing facilitated by Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs) within network information transmission hardware. This optimisation results in an overall reduction in time for knowledge construction.

To intuitively demonstrate the performance of our knowledge construction method, we directly conduct knowledge construction of copyright law and demonstrate its process, as shown in Figure 8. First, our approach extracts the objects on which the regulation operates, namely films and television series. Then, our method will find the copyright or authorship contained therein based on this object, such as producers, directors, and so on. From the visualisation results, we can see that our knowledge construction method can cover the objects, rights claims, and functional objectives of the regulations. Our model can further determine the relationship between the action objectives so as to cover all the content of the regulation.

D. DISCUSS

The widespread dissemination of copyright regulations and the improvement of case data quality contribute to various aspects of the legal domain. Firstly, clear regulations and accurate case data help legal professionals better understand and comply with regulations, thereby enhancing legal compliance. Secondly, addressing issues of omission, inaccuracy, and insufficiency in case data allows for the establishment of a more comprehensive and accurate case database, providing judges, lawyers, and researchers with a more reliable legal reference.

The implementation of the Transformer and Graph Convolutional Networks (GCN) technologies in building a copyright regulation knowledge base and case recommendation system has substantial impacts. The semantic information processing capability of the Transformer, coupled with the feature alignment method of the Graph Convolutional Network (GCN), empowers the system to delve into the essential

TABLE 2. Performance of our method and other methods.

Methods	mAP@0.95	mAP@0.75	mAP@0.50
BiLSTM	0.742	0.762	0.778
Dlinear	0.779	0.818	0.849
Transformer	0.760	0.783	0.811
Transformer-XL	0.765	0.798	0.824
Deformer	0.776	0.797	0.823
Informer	0.779	0.817	0.851
Ours	**0.782**	**0.829**	**0.865**

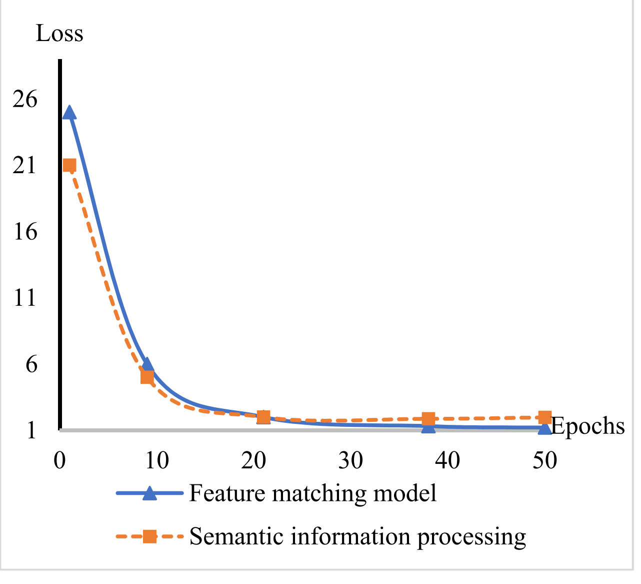


FIGURE 5. The training of our method.

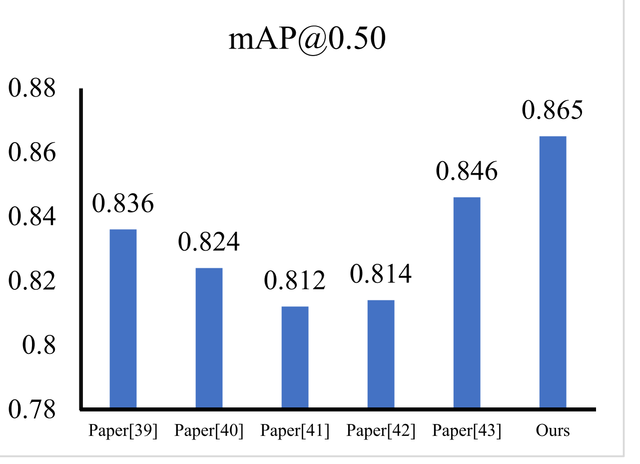


FIGURE 6. Performance of our system and other system.

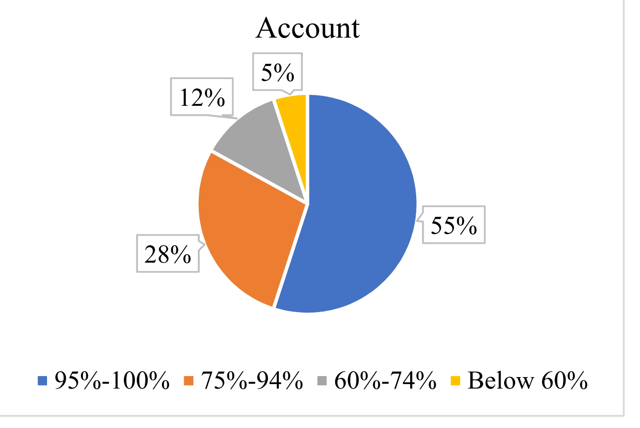


FIGURE 7. The similarity between the copyright knowledge constructed by this method and the artificially constructed knowledge.

concepts and terminologies of regulations and cases, facilitating an understanding of the relationships between them. This contributes to a more comprehensive and accurate representation of the regulatory framework and case network, offering legal professionals a more intelligent and efficient legal query and recommendation service.

Specifically, the application of these technologies aids in establishing a more robust concept index, facilitating case archiving and retrieval. Efficient knowledge repository development is achieved through keyword extraction of concepts related to regulations and cases. The feature alignment approach based on GCN provides a solid foundation for an intelligent recommendation system, making copyright regulation recommendations more intelligent and personalised.

In practice, the application of these technologies not only improves the efficiency of handling regulations and cases but also enables legal professionals to access key information more quickly. Empirical validations demonstrate an impressive accuracy rate of 86.5% for the case recommendation system, confirming the practical value of this technology in the legal domain. Furthermore, subjective evaluations of the copyright regulation knowledge base also receive outstanding acclaim, further affirming the positive impact of these technologies in the legal field.

The copyright regulations knowledge base and case recommendation system based on Transformer and GCN, while possessing numerous advantages, also exhibit certain limitations.

To begin with, the utilisation of transformer and GCN models in this approach necessitates substantial computational resources. Particularly in the legal domain, where data pertaining to copyright regulations and cases is relatively scarce, this could pose challenges in acquiring sufficient training data, consequently impacting the model's performance. Moreover, this high demand for computational resources may present obstacles for some legal professionals who may find it challenging to meet the hardware and software requirements essential for running these models. Hence, a crucial avenue for future enhancement lies in diminishing the computational resource prerequisites of the model to a more pragmatic level and enhancing its operability. Secondly, the

TABLE 3. **Test of Copyright law knowledge base.**

Methods	Cost time	Parameters (M)	Comprehensive similarity
Paper [15]	**1.23s**	314.6	86.4%
Paper [16]	2.35s	286.4	79.4%
Paper [17]	1.56s	339.7	85.2%
Paper [18]	1.56s	274.9	83.9%
Paper [19]	2.39s	**212.8**	80.2%
Ours	1.67s	359.4	**88.2%**

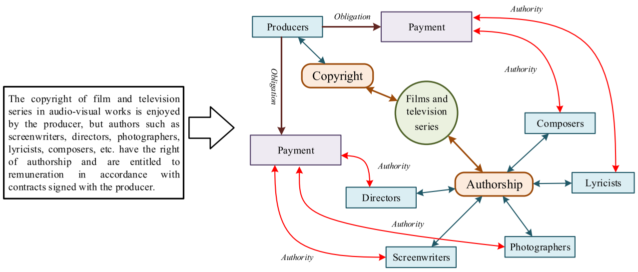


FIGURE 8. **The knowledge construction process of copyright law.**

inherent black-box nature of deep learning models may impede their acceptability in the legal sphere. Legal practitioners typically seek to comprehend the rationale and logic underlying decision-making processes, and models that are overly intricate or opaque may fall short of meeting this requirement. Thus, incorporating explanatory and transparent features in the model design, enabling legal professionals to grasp the decision-making process, will enhance the model's acceptability in practical applications.

Furthermore, the model's potential overreliance on specific types of regulatory and case data could result in suboptimal performance, especially in emerging or specialised fields. In such cases, the model's recommendations may lack accuracy or comprehensiveness. Addressing this issue may involve integrating more diverse and abundant data sources, ensuring the model encompasses a broader spectrum of legal knowledge. Lastly, the dynamic nature of regulations and evolving cases may pose challenges for models to promptly adapt to new legal frameworks or judicial interpretations. Consequently, regular updates to the model, reflecting the latest developments in the legal landscape, have become imperative to ensure its sustained efficacy in the long run.

In the future, our emphasis will be on mitigating data bias concerns and addressing the long-tail effect within databases and recommendation systems. Our objective is to enhance the efficiency of the copyright regulations knowledge base and case recommendation system, all while tackling the scarcity of copyright regulations. Through rigorous research and innovative methodologies, we aim to establish information resources that are more comprehensive and reliable. This, in turn, will contribute to the widespread dissemination of knowledge in the realm of copyright regulations. Simultaneously, we strive to ensure that recommendation systems are equipped to offer more precise and fitting suggestions when confronted with intricate regulatory landscapes.

VI. CONCLUSION

To promote the understanding of copyright, this paper introduces a methodology for constructing a knowledge base of copyright laws and regulations, along with a copyright case recommendation system based on the dissemination of network information. Through an analysis of the attributes of copyright regulations and cases, this paper measures characteristic information inherent in regulations and cases using text as the primary carrier. On the basis of fully constructing the corresponding conceptual model, the copyright regulations knowledge base is constructed. In addition, in order to assist in the dissemination of copyright laws and regulations knowledge base in the network, this paper designs and completes a case recommendation system to realise the intelligent retrieval of copyright law, regulations and cases. The experimental results show that our knowledge base construction method can pass the subjective test of people, and the case recommendation system can also achieve a mAP of 86.5%. In the future, our attention will be directed towards mitigating data bias and addressing the long-tail effects of databases and recommendation systems. We aim to strike a balance in handling the scarcity of copyright regulations while improving the effectiveness of copyright regulation knowledge bases and case recommendation systems.

REFERENCES

[1] P. Ç. Aksoy and Z. Ö. Üner, "NFTs and copyright: Challenges and opportunities," *J. Intellectual Property Law Pract.*, vol. 16, no. 10, pp. 1115–1126, Dec. 2021.

[2] J. Liu, "An empirical study of transformative use in copyright law," *Stanford Technol. Law Rev.*, vol. 22, p. 163, Aug. 2019.

[3] K. J. Boudreau, L. B. Jeppesen, and M. Miric, "Profiting from digital innovation: Patents, copyright and performance," *Res. Policy*, vol. 51, no. 5, Jun. 2022, Art. no. 104477.

[4] C. Joyce, T. T. Ochoa, and M. W. Carroll, *Copyright Law*. New York, NY, USA: Academic, 2020.

[5] B. L. T. Sturm, M. Iglesias, O. Ben-Tal, M. Miron, and E. Gómez, "Artificial intelligence and music: Open questions of copyright law and engineering praxis," *Arts*, vol. 8, no. 3, p. 115, Sep. 2019.

[6] D. Bridge, M. H. Goker, and L. McGinty, "Case-based recommender systems," *Knowl. Eng. Rev.*, vol. 8, no. 3, p. 115, 2005.

[7] S. Milano, M. Taddeo, and L. Floridi, "Recommender systems and their ethical challenges," *AI Soc.*, vol. 35, no. 4, pp. 957–967, Dec. 2020.

[8] A.-Z. Yen, C.-C. Chang, H.-H. Huang, and H.-H. Chen, "Personal knowledge base construction from multimodal data," in *Proc. Int. Conf. Multimedia Retr.*, Aug. 2021, pp. 496–500.

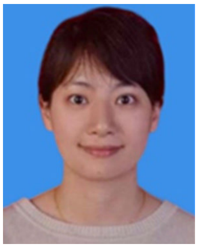
[9] A.-Z. Yen, H.-H. Huang, and H.-H. Chen, "Multimodal joint learning for personal knowledge base construction from Twitter-based lifelogs," *Inf. Process. Manage.*, vol. 57, no. 6, Nov. 2020, Art. no. 102148.

[10] D. W. K. Khong and W. J. Yeh, "The problem of copyright protection for machine learning databases: A comparative study," *NTUT J. Intellectual Property Law Manag.*, vol. 10, no. 2, pp. 86–101, 2021.

[11] B. Smyth, "Case-based recommendation," in *The Adaptive Web: Methods and Strategies of Web Personalization*. Berlin, Germany: Springer, 2007, pp. 342–376.

[12] B. Sarwar, G. Karypis, J. Konstan, and J. T. Riedl, "Application of dimensionality reduction in recommender system—A case study," Univ. Minnesota Digit. Conservancy, 2000.
[13] L. Sharma and A. Gera, "A survey of recommendation system: Research challenges," *Int. J. Eng. Trends Technol.*, vol. 4, no. 5, pp. 1989–1992, 2013.
[14] M. Wang, P. Ren, L. Mei, Z. Chen, J. Ma, and M. de Rijke, "A collaborative session-based recommendation approach with parallel memory modules," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2019, pp. 345–354.
[15] A. Hürriyetoğlu, E. Yörük, O. Mutlu, F. Duruğan, Ç. Yoltar, D. Yüret, and B. Gürel, "Cross-context news corpus for protest event-related knowledge base construction," *Data Intell.*, vol. 3, no. 2, pp. 308–335, Jun. 2021.
[16] H. Zhang, "Neural network-based tree translation for knowledge base construction," *IEEE Access*, vol. 9, pp. 38706–38717, 2021.
[17] B. Pereira, C. Robin, T. Daudert, J. P. McCrae, P. Mohanty, and P. Buitelaar, "Taxonomy extraction for customer service knowledge base construction," in *Proc. 15th Int. Conf. Semantic Syst. Power AI Knowl. Graphs*, vol. 2019, Karlsruhe, Germany. Cham, Switzerland: Springer, 2019, pp. 175–190.
[18] I. Sabek and M. F. Mokbel, "Sya: Enabling spatial awareness inside probabilistic knowledge base construction," in *Proc. IEEE 36th Int. Conf. Data Eng. (ICDE)*, Apr. 2020, pp. 1177–1188.
[19] X. Fang, A. Kalinowski, H. Zhao, Z. You, Y. Zhang, and Y. An, "Prompt design and answer processing for knowledge base construction from pre-trained language models (KBC-LM)," Tech. Rep., 2022.
[20] M. Koniaris, G. Papastefanatos, and Y. Vassiliou, "Towards automatic structuring and semantic indexing of legal documents," in *Proc. 20th Pan-Hellenic Conf. Informat.*, Nov. 2016, p. 4.
[21] A. Wyner and W. Peters, "Towards annotating and extracting textual legal case factors," in *Proc. 3rd Workshop Semantic Process. Legal Texts*, 2010, pp. 26–45.
[22] M. Saravanan, B. Ravindran, and S. Raman, "Improving legal information retrieval using an ontological framework," *Artif. Intell. Law*, vol. 17, no. 2, pp. 101–124, Jun. 2009.
[23] P. Casanovas, M. Palmirani, S. Peroni, T. van Engers, and F. Vitali, "Semantic web for the legal domain: The next step," *Semantic Web*, vol. 7, no. 3, pp. 213–227, Mar. 2016.
[24] A. Mandal, K. Ghosh, S. Ghosh, and S. Mandal, "Unsupervised approaches for measuring textual similarity between legal court case reports," *Artif. Intell. Law*, vol. 2021, pp. 1–35, Jan. 2021.
[25] K. D. Ashley and V. R. Walker, *From Information Retrieval (IR) to Argument Retrieval (AR) for Legal Cases: Report on a Baseline Study*. JURIX, 2013, pp. 29–38.
[26] K. Sugathadasa, B. Ayesha, N. de Silva, A. S. Perera, V. Jayawardana, D. Lakmal, and M. Perera, "Legal document retrieval using document vector embeddings and deep learning," in *Proc. Comput. Conf. Intell. Comput.*, vol. 2, Springer, 2019, pp. 160–175.
[27] S. Zacchiroli, "A large-scale dataset of (open source) license text variants," in *Proc. IEEE/ACM 19th Int. Conf. Mining Softw. Repositories (MSR)*, May 2022, pp. 757–761.
[28] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The performance of LSTM and BiLSTM in forecasting time series," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 3285–3292.
[29] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 9, 2023, pp. 11121–11128.
[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
[31] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. Le, and R. Salakhutdinov, "Transformer-XL: Attentive language models beyond a fixed-length context," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 1–20.
[32] Q. Cao, H. Trivedi, A. Balasubramanian, and N. Balasubramanian, "DeFormer: Decomposing pre-trained transformers for faster question answering," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 4487–4497.
[33] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 12, 2021, pp. 11106–11115.
[34] C. Gao, X. Wang, X. He, and Y. Li, "Graph neural networks for recommender system," in *Proc. 15th ACM Int. Conf. Web Search Data Mining*, Feb. 2022, pp. 1623–1625.
[35] M. M. Afsar, T. Crump, and B. Far, "Reinforcement learning based recommender systems: A survey," *ACM Comput. Surv.*, vol. 55, no. 7, pp. 1–38, Jul. 2023.
[36] H.-W. An and N. Moon, "Design of recommendation system for tourist spot using sentiment analysis based on CNN-LSTM," *J. Ambient Intell. Humanized Comput.*, vol. 13, no. 3, pp. 1653–1663, Mar. 2022.
[37] S. Geng, S. Liu, Z. Fu, Y. Ge, and Y. Zhang, "Recommendation as language processing (RLP): A unified pretrain, personalized prompt & predict paradigm (P5)," in *Proc. 16th ACM Conf. Recommender Syst.*, Sep. 2022, pp. 299–315.
[38] Y. Yang, C. Huang, L. Xia, and C. Li, "Knowledge graph contrastive learning for recommendation," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2022, pp. 1434–1443.
[39] C. Zhou, C. Liu, Q. Zeng, Z. Lin, and H. Duan, "A comprehensive process similarity measure based on models and logs," *IEEE Access*, vol. 7, pp. 69257–69273, 2019.

JIAMING MO mainly engaged in the research and teaching of intellectual property law, civil law, and tourism law. Currently, he has published 13 academic articles in ISSN/CN. He has also participated in two national projects, two provincial and ministerial projects, two municipal projects, and two school level projects in China. He has achieved certain research and teaching results in legal research and teaching.



JUNYU ZHENG is currently with Guangzhou Panyu Polytechnic. She is engaged in the digital education direction of exhibition planning and management. She has ten years of experience in project management in internet companies. She has a solid theoretical foundation and practical experience in the field of information technology. She has PMP Certification, Information System Project Manager and System Planning and Management Certification.

DEXU BI received the master's degree in pedagogy from the College of Education, Southwest University. He is currently pursuing the Ph.D. degree in ethnic education with the College of Education Science, Guangxi University for Nationalities. He is also an Associate Professor with the Department of Elementary Education, Guangxi Police College. His research interests include criminology and educational technology.