Convolution Neural Network Based Patent Infringement Detection Method

Weidong Liu
College of Computer Science, Inner Mongolia University, China.Institute of Scientific and Technical Information of China.
Hohhot, China
cslwd@imu.edu.cn

Senjun Pei
College of Computer Science, Inner Mongolia University, China
Hohhot, China
peisenjun@gmail.com

***Abstract*—With the development of intellectual property rights in recent years, the number of patent applications has been increasing. At the same time, the number of patent infringement cases has also increased. When there is infringement between patents, the traditional method is for patent examiners to manually search for infringing features to determine whether there is infringement between patents according to the patent law. Since a patent is a complex semi-structured text and involves a wide range of fields, most of the current infringement detection methods cannot determine the infringement features well, and most of the methods only study one-to-one patent infringement and do not solve the problem of one-to-many patent infringement well. In order to solve the above problems, a patent infringement detection method based on convolutional neural network is proposed. The method extracts and represents infringement features from patents, patent claims and independent patent claims respectively, represents patents by different patent text vectorization methods, combines and filters features based on convolutional neural networks so as to obtain semantic information of different abstraction layers of patents, and finally tests the evaluation model on a one-to-many patent infringement data set. The results show that the model has greatly improved the infringement detection accuracy.**

Keywords-CNN; Patent infringement; Infringement detection; Feature representation

I. INTRODUCTION

With the advent of the era of knowledge competition, the number of patent applications as a carrier of intellectual property rights has increased dramatically, leading to a high incidence of various patent infringement cases at the same time. Since patents are complex semi-structured texts, they cover a wide range of fields. When patent infringement occurs, traditional manual detection requires patent examiners to have a high professional background and need to spend time to understand and learn the relevant patent field knowledge, which not only increases the workload of manual detection, but also cannot guarantee the accuracy and timeliness of examination results, resulting in the accumulation of patent applications and unnecessary patent litigation, bringing economic losses to the patentee. Therefore, from the patent infringement cases, we can see the necessity and importance of designing a method for automatic detection of patent infringement.

In the current patent infringement determination process, the related work of infringement detection methods can be divided into supervised methods and unsupervised methods.

1) The unsupervised methods are broadly divided into clustering-based methods [1],which mostly use the k-means algorithm for clustering categories that are clear, and the SOM (Self-Organizing Maps)-based Chinese patent infringement detection algorithm [2] for clustering with fuzzy categories; game-theory-based methods [3], which are based on the claims of both sides of the patent and The method is based on game theory, which establishes a game tree based on the dynamic simulation of the claims and technical features of both parties to the patent, so as to calculate the risk of infringement between patents; based on the content of the patent text for keyword extraction [4], text content analysis to extract the SAO structure [5] to calculate its similarity, and determine whether the patent is infringed according to the size of the similarity; during the litigation processing of patent infringement cases, the patentee establishes a two-segment game based on the cost and profit of the patent to achieve Dynamic negotiation [6] to determine whether the patent is infringed and the post-infringement treatment.

2) Supervised methods are broadly classified into vector space model-based detection methods [7-8], which determine the negation relationship between patents by mapping all texts, paragraphs or words of patents into vectors of fixed size and calculating the similarity between vectors; hierarchical keyword vector construction based on patent claims [9], which calculates the patent-claim-technical feature hierarchical vectors by constructing the similarity between different patents and determine whether they infringe based on the magnitude of the similarity; based on the similarity between patents and considering both semantic and technical similarity, the method uses BERT to measure the semantic similarity based on the patent text [10]. Deep learning-based method [11-12], which extracts parts of patents and vectorizes them based on SOM neural network, and then clusters them using FCM algorithm; based on Doc2vec vectorization to judge the similarity between patent texts [13], which preprocesses patent texts and constructs a corpus, then vectorizes patent texts

DOI:10.18293/SEKE2023-030

based on Doc2vec model Then, the cosine similarity between vectors is calculated, and the size of similarity is proportional to the risk of infringement [14].

However, current infringement detection methods have the following limitations.

1) For unsupervised methods, the limitation of clustering-based methods is that the accuracy of clustering is low when solving large scale patent data sets, and the number of initial clusters cannot be determined, resulting in the k-means method cannot be used and the clustering effect is poor. The method based on game theory can only calculate the infringement risk value among patents to a certain extent, and the feature extraction single does not make good use of the patent text content to extract the features that represent patents. Similarity-based methods, because of the variability of patent text content and structure, increase the cost of manual labeling makes patent feature extraction more limited, and a large number of traditional similarity calculation only extracts part of the patent text information, and does not obtain the deep semantic information of the patent text, thus making the calculation of similarity more complex. The dynamic negotiation based on cost and profit is affected by the uncertainty of the negotiation process, and the result is contingent and cannot be efficiently and accurately determined whether the infringement has occurred.

2) For the supervised method, the vector space model-based method, the complexity of the patent text content makes the corresponding patent vector dimension and the semantic granularity of the patent vector features increased uncertainty, and most of the vector dimension is high and sparse, which is not conducive to the calculation of similarity. The deep learning method Yoon uses self-organizing neural network method to construct feature maps. Deep learning models require a large number of data sets, and the pre-model data processing and labeling work requires a lot of manual participation. The method of using Doc2vec to determine patent infringement is implemented based on text similarity. The method firstly represents the patent text content by embedded vectors with trained neural network models, and then calculates the similarity between the vectors, so as to determine whether the patent constitutes infringement between them based on the similarity magnitude. The limitation of this method is that it has a single feature extraction, does not integrate the patent text content well, cannot represent the patent text content efficiently and accurately, and has a low accuracy rate.

3) Most of the current infringement detection methods deal with one-to-one infringement relationships, and there is no fast and efficient solution for one-to-many patent infringement. Moreover, most of the traditional infringement detection methods are based on the textual content of patents, the cross-referencing relationship of patents and the non-textual content of patents (patent numbers, patent classification codes, application dates, etc.), and do not focus on the overall content of patents. Moreover, the existing detection methods do not have unified high-quality data sets, and the experimental results and models cannot be judged by their merits.

In order to determine whether multiple patents infringe each other,we propose a convolution neural network-based algorithm that uses the neural network to self- learn in order to discover patterns and obtain an efficient detection model

1) A high-quality one-to-many patent infringement data set was constructed to better assess the strengths and weaknesses of the model and its generalization capabilities.

2) Extract features by fusing different parts of the patents.Although traditional patent infringement determination is based on the claims, other elements of the patent can also be used to help determine whether a patent is infringed.

3) Extraction and representation of infringement features using convolutional neural networks.

II. BACKGROUND AND RELATED WORK

A. DOC 2 Vec

One of the main steps of text classification is the word vector representation of text. A good word vector can better express the semantics between words. Word2vec is widely used in the field of word embedding. Although word2vec provides high- quality word vectors, for sentences, documents or paragraphs, these data cannot be well projected into the vector space, nor can it express their rich semantic information. Doc2vec method is an unsupervised algorithm, which can learn from variable length text and obtain fixed length feature representation. Doc2vec model is an extension of word2vec model, and it also has two training methods, PV-DM(Paragraph Vector-Distributed Memory) and PV-DBOW.(Paragraph Vector-Distributed Bag of Words).

B. Convolution Neural Network

Convolution neural network is a variant of feed forward neural network. Convolution layer is the core of convolution neural network.Through convolutionoperation, the two purposes of dimension reduction and feature extraction can be achieved.Convolution operation can select more representative local features, so as to express the important features of data more efficiently. At first, convolution neural network made a breakthrough in the field of computer vision [15],because it shows a high degree of displacement invariance in feature extraction.The two main features of local perception and weight sharing make the neural network effectively reduce the order of magnitude of parameter learning. When this idea is applied to text data mining, it greatly improves the performance of the task.

C. Patent Representation

A patent is a complex structure and a very broad field of public documents. Patent documents contain the title, specification abstract, claims and additional description of the patent, and as the core of the patent, the claims are essentially a collection of technical features of the patent. According to the description of the patent content, the claims are divided into independent claims and dependent claims. The independent claims contain the necessary technical features of the patent, and the dependent claims are attached to the independent claims, which are more detailed definitions and descriptions of the technical features of the independent claims. For patent infringement, the focus is not only on the identification of infringement relationship and infringement relevance mining, but also on the consideration of the complexity of the patent content structure is particularly important for patent text mining [16-17].

Definition 1:Patent Representation P_i

$$P_i = \begin{cases} A \cup C_{all}^i \cup D \\ C_{all}^i \\ C_{ij}^* \end{cases} \tag{1}$$

Where P_i denotes the i^{th} patent, A denotes the abstract of the specification of the patent, C_{all}^i denotes all claims of the i^{th} patent, D denotes an additional description of the patent, $|C_{all}^i|$ denotes the number of all claims of the i^{th} patent, C_{ij} denotes the j^{th} claim of the i^{th} patent, C_{ij}^* ($C_{ij}^* \in C_{ij}$)denotes the j^{th} claim of the i^{th} patent, and the claim is an independent claim.($0 < j < |C_{all}^i|, \sum_{j=1}^{|C_{all}^i|} C_{ij} = C_{all}^i$)

Definition 2: Infringing patent data set IPDS

$$\text{IPDS}= \{ \text{IPA}^k \} \tag{2}$$

Infringing patent association IPA^k

$$\text{IPA}^k=\{P_0^K P_1^K P_2^K \ldots\ldots P_j^K\} \tag{3}$$

IPA^k represents a association of infringing patents analyzed through a set of real infringement cases, consisting of one infringing patent P_0^K and many infringed patents $P_1^K P_2^K \ldots\ldots P_j^K$. k represents the k^{th} infringing patent association with infringing relationship in IPDS, an IPDS consists of n IPA.

TABLE I. SYMBOLS AND THEIR MEANINGS

Symbols	Description
P_i	The i^{th} patent
A	Abstract of patent specification
C_{all}^i	All claims of the i^{th} patent
$\lvert C_{all}^i \rvert$	The claim number in the P_i
D	Additional description of the patent
C_{ij}	The i^{th} Claim in the j^{th} patent
C_{ij}^*	The i^{th} Claim in the j^{th} patent, which is an independent claim
IPA^k	Infringing patent association
$IPDS$	Patent infringement data set
P_0^K	The k^{th} one -to- many infringing patent
P_j^K	The k^{th} one -to- many infringed patent

D. Problem Statement

Patent infringement is the use of patented technology in production without the permission of the patentee or legal protection during the validity of the patent. The essence of one-to-many patent infringement is that when a set of technical solutions wants to be registered as a patent, it is necessary to determine whether the registered patented technical solution has been copied based on the technical features of the technical solution (evidence of infringement).

The paper is to address the patent infringement associations IPA^K, P_0^K whether the patent is infringed $P_1^K P_2^K \ldots\ldots P_j^K$, To solve this problem, we parse large-scale patent infringement cases and patent texts, extract different parts of the patent texts by parsing them, and denotes them as patent word vectors by advance training patent text vectorization methods, and input the word vectors into a convolution neural network model through certain calculations, the convolution neural network includes multiple convolution layers, maximum pooling layers and fully connected layers. decay regularization. An excellent patent infringement detection model is obtained after training. The model can make a determination of whether a single patent infringes other patents.

III. PROPOSED MODEL

When one-to-many patent infringement occurs, infringing patent data set,(IPDS) and infringing patent association (IPA^k) can be obtained by parsing the patent infringement cases.First the data set is divided into training set: validation set: test set as 6:2:2,and then parsed the patent text,obtain three representations of the patent:①all patent contents $A \cup C_{all}^i \cup D$.② all claims of the patent C_{all}^i. ③ independent claim C_{ij}^*.The adjudication process is shown in Fig.1.

We train Doc2vec with the content of ① to get our vectorization method Patent2vec, and train Doc2vec with the content of ② to get our vectorization method Claims2vec.The patented text content needs to be converted into a vector representation by word embedding, different vectorization methods have a significant impact on the accuracy of text classification. Therefore, we train two different vectorization methods by analyzing the three patent contents obtained from the above patents. When the training samples of the input word or sentence vector Paragraph Id of the input word are from representation①, this method is Patent2vec, and when the training samples of the input word or sentence vector Paragraph Id of the input word are from representation ②, this method is Claims2vec.

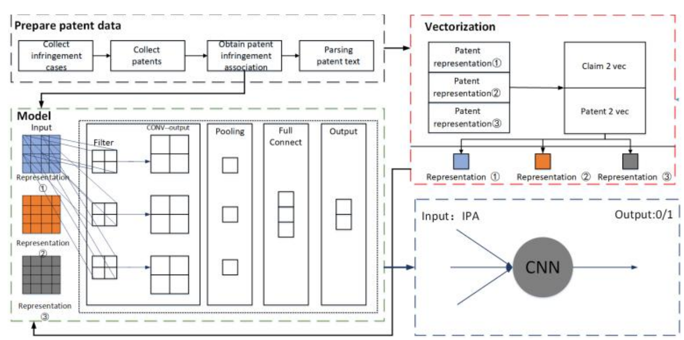

Fig.1.Patent infringement judgment model

As shown in Equation 1, the first patent representation method uses the entire text content of the patent for representation, and the content of the patent text is vectorized with Patent 2 vec, and then trained with the constructed convolutional neural network, whose algorithm is shown in Algorithm.1.

Algorithm 1: Patent representation for patent infringement detection ①

Input: IPDS $=\{ IPA^k\}$, $IPA^k=\{P_0^K P_1^K P_2^K......P_j^K\}$

Output: Infringement result 0/1

1: Express all text contents of the patent as patent features, $P_i=（A \cup C_{all}^i \ \cup D）$

2: Vectorized representation of patent acquisition,

V_{P_i}= Patent2vec（P_i）

3: **for** k =1…K do

4: **for** j=1…J do

5: Cosine_Similarity: $W_i = \langle P_0^K | P_j^K \rangle$

6: Normalization W_i

7: $V=V_{P_0} + \sum_{i=1}^{j} W_i * V_{P_I}$

8: End

9: V_{input}+=V

10: End

11. V_{input} Input to neural network prediction, return results

The patent text is analyzed to obtain a second representation of the patent, using the patent claims(C_{all}^i) denotes the patent, and the content of the patent text is vectorized using Claims2vec and then trained using the constructed convolution neural network， it is shown as Algorithm.2.

Algorithm 2: Patent representation for patent infringement detection ②

Input: IPDS $=\{ IPA^k\}$, $IPA^k=\{P_0^K P_1^K P_2^K......P_j^K\}$

Output: Infringement result 0/1

1: Express all text contents of the patent as patent features, $P_i= C_{all}^i$

2: Vectorized representation of patent acquisition,

V_{P_i}= Patent2vec（P_i）

3: **for** k =1…K do

4: **for** j=1…J do

5: Cosine_Similarity: $W_i = \langle P_0^K | P_j^K \rangle$

6: Normalization W_i

7: $V=V_{P_0} + \sum_{i=1}^{j} W_i * V_{P_I}$

8: End

9: V_{input}+=V

10: End

11. V_{input} Input to neural network prediction, return results

The patent text is analyzed to obtain a third representation of the patent,using the independent claims of the patent(C_{ij}^*) denotes the patent, and the content of the patent text is vectorized using Claims2vec and then trained using the constructed convolution neural network ,it is shown as Algorithm.3.

Algorithm 3: Patent representation for patent infringement detection ③

Input: IPDS $=\{ IPA^k\}$, $IPA^k=\{P_0^K P_1^K P_2^K......P_j^K\}$

Output: Infringement result 0/1

1: Express all text contents of the patent as patent features, $P_i=C_{ij}^*$

2: Vectorized representation of patent acquisition,

V_{P_i}= Claims 2vec（P_i）

3: **for** k =1…K do

4: **for** j=1…J do

5: Regular and Analysis C_{all}^i

6: Obtaining C_{ij}^*

7: End

8: $V_{input}=C_{0j}^* + \sum_{i=1}^{j} C_{ij}^*$

9: End

10. V_{input} Input to neural network prediction, return results

IV. EXPERIMENT

To validate the efficiency of our method, we design experiments as follows.

A. *Experimental Data*

In order to prove the validity of our model in rating prediction, we used the data of USPTO and Google patents as experimental data sets. Patent infringement cases are extracted from the patent trial and appeal board (ptab) inter departmental review (IPR) documents on the website of the U.S. patent and Trademark Office through automatic crawlers. These data sets include 100 patent infringement cases, 251 patent infringement combinations, and 2000 unrelated or similar patents. The patents used in our experiment are shown in Table II.

TABLE II. Experimental data

Source	**USPTO**	**USPTO**	**Google patents**
Data set	Patent infringement cases	IPA	Irrelevant or similar patent
Number	100	251	2000
Does infringe	Yes	Yes	No

B. *Baseline Models*

We compare our approach with state-of-the-art methods:a comparative study of Doc 2 vec based methods for detecting similarity in patent documents [11], which includes the following steps.

1) The patent is decomposed into a patent number document B and a patent abstract document C.The dictionary document set D is generated from document C, and D is used to generate the trained document set H using Doc2vec.

2) The similarity between the B and H patent number files is calculated and a manual secondary analysis is performed to obtain the final infringement detection results.

The essence of the method is to vectorize the patent text, and determine whether there is an infringement relationship between patents by calculating the similarity of the patent text. The size of the similarity is positively related to the probability of infringement.

C. *Evaluation Measurements*

Infringement risk is the evaluation measure in this paper. We will compare the experimental results of our proposed approach with the baseline approach in terms of violation risk.In this study, the patent associations violation problem is essentially a classification task.For the three sets of comparison experiments, we use score in Eq.4 as the evaluation metric on the test set to judge the three sets of experiments, we use the most basic evaluation metric, precision in Eq.5, to validate the predictions, we use a loss function to check the robustness of our constructed model, and we use F1scores to balance precision and recall in Eq.6 and in Eq.7.

$$\text{Precison}_{\text{score}}=\frac{\text{correct predictions}}{\text{all predictions}} \tag{4}$$

$$\text{Accuracy}=\frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{5}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP+FN}} \tag{6}$$

$$\text{F1}=\frac{2\times\text{Recall}\times\text{Precison}}{\text{Precison+Recall}} \tag{7}$$

D. *Experimental Setups*

For the infringing patent associations, we conduct the following experiments.

1) Firstly, we obtain the judgment documents of patent infringement from USPTO, intercept and analyze the content of the judgment documents, obtain the infringing patent pairs and obtain the text content of corresponding patents from Google Patents and similar patents without patents to build the experimental data set.Parsing the patent text, extracting all the contents of the patent text, claims and independent right contents, i.e., Equation 1. generating training corpus to train Doc2vec to get vectorized methods Patent2vec and Claims2vec.

2) Based on the different representations of patents, the infringement detection model based on convolutional neural network is trained by Algorithms 1, 2 and 3, and the infringement features that work best in the infringement determination process are obtained by experimentally comparing the different representations of patent texts. When Patent2vec and Claims2vec are used to integrate features and statements, an excellent infringement detection model is finally obtained through experiments.

3) The Baseline Models is used for infringement detection through a comparative study of Doc2Vec based patent document similarity detection methods.Baseline Models Use IPA^k to verification.According to the vectorization method Doc2Vec in the Baseline Models , vector the different patent, and the infringement probability value between patents, namely the infringement risk, is calculated by using the similarity.

4) We conducted experiments based on the same data set IPA^k ,then,compare the results obtained by our method with those of the reference method in the evaluation measures.

E. *Experimental Results*

For the infringing patent data set,the data set is divided into training set: validation set: test set as 6:2:2 ,after data set partitioning is complete, our experimental flow is shown in Figure.1, where we represent the data by different patent representations and different patent text vectorization methods.The model was trained using the training and validation data, and once the training model was generated, the model was tested using the test data. Based on the evaluation parameters, our model was validated on the test data, as shown in Table III.

TABLE III. Model evaluation parameters

IPDS	**Evaluation measurement**		
Algorithm	**Loss(%)**	**Accuracy(%)**	**F1(%)**
Algorithm1	0.6858	0.6000	0.74
Algorithm2	0.5638	0.7667	0.64
Algorithm3	0.5390	0.6667	0.73

In the Baseline Models, the patent number A to be tested and the abstract text of patent B to be tested are first generated from patent A in the post-training document H based on the trained Doc2vec model, and the list of similar patents is obtained and manually analyzed twice based on the abstract text of patent B. The basic experiment was performed by entering three patents for similarity determination,and the results are shown in Table IV.

TABLE IV. Baseline method results

Patents similar to H	7874783	8995191	7999571
Similarity	0.5317	0.4655	0.4626

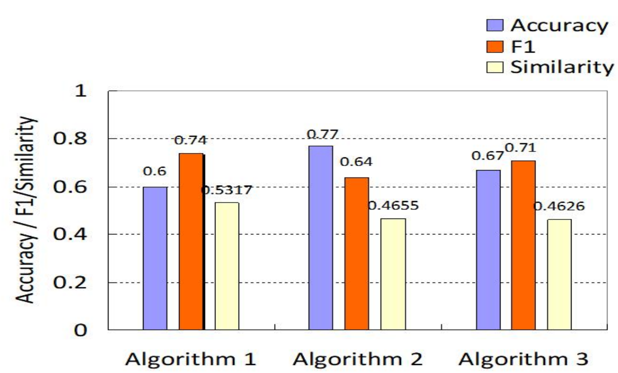

Fig.2. Patent infringement judgment model

Compared with the benchmark method, the method proposed in this paper experiments with different manifestations of the patent and concludes through the experimental results that the best evidence of infringement is the claim part of the patent, because the claim part expresses the content of the patent clearly and in detail, and describes the scope of protection of the patent in detail.Our method can quickly resolve the claims of patents when inputting one-to-many infringement cases and represent the claims by claim2vec, and finally get the result of whether the patent is infringed or not with an accuracy rate of 76.67%,while in the baseline model, when inputting infringing patents and infringed patents, the trained Doc2vec model is used to generate the patents into training After the document, as the greater the similarity between infringing patents, the greater the risk of infringement between patents, so, although the baseline model solves the complex problem in the patent field to a certain extent, its accuracy rate is still low.

V. CONCLUSIONS AND FUTURE WORK

The essence of patent infringement detection is text classification. The key is to accurately extract the central idea of patent documents. The method of extracting the central idea is to extract the keywords of documents or sentences as features, and train classifiers and classify them based on these features. Because the convolution and pooling process of convolution neural network is a feature extraction process, when we can accurately extract the features of keywords, we can accurately extract the central idea of documents or sentences.The baseline model method only calculates the similarity based on the representation of patent documents, does not refine the content of patent documents, but uses mechanical methods to represent patents. It is impossible to accurately judge whether patents are infringed, but it has reference value to some extent.A patent infringement detection method based on convolutional neural network is proposed in this paper. The method obtains the depth information of the patent by selecting and representing the textual content of the patent, and merges different elements of the patent for comparison, so as to obtain an effective determination method. The contribution of our method to previous research is summarized as follows:

1) The lack of standardized and consistent datasets for previous infringement detection methods has led to the inability to evaluate the model, which we support by analyzing decided infringement cases from uspto.

2) Using an innovative and unified convolutional neural network automatic infringement detection framework, the model is trained with different patent text representation methods and text vectorization methods, and the best patent text representation method for infringement detection is obtained through comparative experiments, which improves the detection efficiency to some extent.

3) Most of the previous infringement detection focuses on the infringement detection of a single patent, and this paper solves the infringement determination of multiple patents.

We compare our approach with state-of-the-art methods and the results show that our approach outperforms the benchmark methods in patent infringement detection. In the case of one-to-many patent infringement, in addition to building a high quality dataset for patent infringement detection, better models can also be used to extract semantic information about patents. These issues will be addressed in future work.

ACKNOWLEDGEMENT

The research work in this paper was supported by the National Science Foundation of China (grant no. 62161036 and grant no. 61801251), Postdoctoral Science Foundation of China (2020M680643), Program for Young Talents of Science and Technology in Universities of Inner Mongolia Autonomous Region(NJYT23058)

REFERENCES

[1] Bhatt P C, Lu T C. Identifying firm significance and positions in the patent innovation based on centrality measures' clustering approach[J]. IEEE Access, 2023.4.29

[2] WU Y, MA Y, Zhai D. Research on SOM-based Chinese patent infringement detection[J]. Journal of Intelligence, 2014, 33(2): 33-39.

[3] Liu X. Research and System Implementation of Patent Infringement Detection Method Based on Game Theory [D]. Inner Mongolia University,2021.

[4] Dong L, Pan H. Impact of patent infringement compensation rules on patent quality problems[J]. Complexity, 2021, 2021: 1-12.

[5] Kim S, Yoon B. Patent infringement analysis using a text mining technique based on SAO structure[J]. Computers in Industry, 2021, 125: 103379.

[6] Crampes C, Langinier C. Litigation and settlement in patent infringement cases[J]. RAND Journal of Economics, 2002: 258-274.

[7] Shalaby M, Stutzki J, Schubert M, et al. An lstm approach to patent classification based on fixed hierarchy vectors[C]//Proceedings of the 2018 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2018: 495-503.

[8] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.

[9] Lee C, Song B, Park Y. How to assess patent infringement risks: a semantic patent claim analysis using dependency relationships[J]. Technology analysis & strategic management, 2013, 25(1): 23-38.

[10] Yoo Y, Jeong C, Gim S, et al. A Novel Patent Similarity Measurement Methodology: Semantic Distance and Technological Distance[J]. arXiv preprint arXiv:2303.16767, 2023.

[11] Abdelgawad L, Kluegl P, Genc E, et al. Optimizing neural networks for patent classification[C]//Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part III. Cham: Springer International Publishing, 2020: 688-703.

[12] Lu Y, Xiong X, Zhang W, et al. Research on classification and similarity of patent citation based on deep learning[J]. Scientometrics, 2020, 123: 813-839.

[13] Yun S, Cho W, Kim C, et al. nTechnological trend mining: identifying new technology opportunities using patent semantic analysis[J]. Information Processing & Management, 2022, 59(4): 102993.

[14] CaoQi, Zhao Wei, Zhang Yingjie, et al. Comparative Study of PatentDocuments Similarity Detection on Deep Learning of Doc2VecBased Methods[J]. Library and Information Service, 2018, 62(13): 74-81.

[15] Hou S, Wang Z. Weighted channel dropout for regularization of deep convolutional neural network[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 8425-8432.

[16] Wang X, Kou L, Sugumaran V, et al. Emotion correlation mining through deep learning models on natural language text[J]. IEEE transactions on cybernetics, 2020, 51(9): 4400-4413.

[17] Wang X, Zhang H, Xu Z. Public sentiments analysis based on fuzzy logic for text[J]. International Journal of Software Engineering and Knowledge Engineering, 2016, 26(09n10): 1341-13