

Research Article

Applying Deep Learning Technologies to Evaluate the Patent Quality with the Collaborative Training

Xindong You ¹, Jiaqi Liu,¹ Zhe Wang,¹ Zhaonan Liu,¹ Xueqiang Lv ¹,
and Jung Yoon Kim ²

¹Beijing Key Laboratory of Internet Culture Digital Dissemination, Beijing Information Science and Technology University, Beijing, China

²Department of Game Media, College of Future Industry, Gachon University, Seongnam-si, 13120, Gyeonggi-do, Republic of Korea

Correspondence should be addressed to Xueqiang Lv; icddtxy@163.com and Jung Yoon Kim; kjyoon@gachon.ac.kr

Received 13 May 2021; Revised 20 August 2021; Accepted 14 September 2021; Published 21 October 2021

Academic Editor: Yu Yinyu

Copyright © 2021 Xindong You et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As the country vigorously promotes the development of science and technology and tries to enhance independent innovation capabilities, more and more attention is paid on the protection of technology ownership. In recent years, China has developed rapidly in many scientific and technological fields, and the number of patent applications increased year by year. However, various patent quality problems including immature patent technology and low patent authorization rate appear. The indicators of patent quantification and quality evaluation are studied in this paper. First, we quantify the patent quality evaluation indicators and combine the content of the patent text to build a patent evaluation model. US patents with patent grade labels are used for training with multitask learning technology. Second, the evaluation model is transferred from the English patents to the Chinese patents, in which the active learning technology and transfer learning technology are used to minimize the work of manual labeling. Finally, a Chinese patent quality evaluation model based on collaborative training was designed and implemented. Methods used in this experiment have notably improved the prediction effect of the model and achieved a better migration effect. A large number of experimental results show that the Chinese patent quality evaluation model has achieved good evaluation results. This research uses deep learning and natural language processing technology to carry out research on patent quality evaluation models from different perspectives, to provide patent decision support for related companies, and to point out research directions for research institutions and patent inventors.

1. Introduction

Since the 18th CPC National Congress, the cause of intellectual property has entered a new era of vigorous development driven by policies. Patent applications are trending year by year. As of 2020, the State Intellectual Property Office has authorized a total of 530,291 disclosed invention patents, an increase of 18% over 2019. According to data published by the State Intellectual Property, in 2020, 683,000 invention patent applications were discovered; a total of 217,000 invention patents were authorized. In order to implement the guiding ideology and effectively promote my country's transformation from a country of intellectual property rights to a country of cre-

ativity, from the pursuit of quantity to the improvement of quality, recently, the National Intellectual Property "Notice on Further Strictly Regulating Judges' Applicants for Understanding" (Guo) Fabaozi 2021 No. 1 during the project verbally promoted the improvement of patent quality, strengthened quality orientation, and strengthened the standardization and supervision of patent transactions. How to quickly, automatically, fairly, press, openly, and scientifically evaluate the quality and value of patents will become a problem that needs to be resolved.

Patent and other intellectual property intangible assets have become an important force in national development, with the development of intellectual property-related technologies at home and abroad and the improvement of their

status in the economic field. In recent years, the patent pledge financing project has gradually taken shape, which has provided a great boost to the growth of small and medium-sized enterprises. In 2020, the total amount of national patent and trademark pledge financing reached 218 billion yuan, an increase of 43.9% year on year. This fully demonstrates that there is a huge demand for patent value evaluation in China and the world, and it is increasingly not negligible in the fields of science and society. While strengthening innovation, my country should also promote the transformation of scientific and technological achievements. Therefore, in recent years, with the country's high requirements on the speed and efficiency of the transformation of the results of authorized national invention patents, how to objectively and automatically evaluate the value of massive Chinese patents has become a hot research topic and a scientific problem that needs to be solved urgently.

Many scholars have been exploring the construction of models based on quantitative indicators and automatic quality assessment. However, there are still the following shortcomings in the evaluation of Chinese patent quality:

- (1) The Chinese patent quality evaluation index system is not sound enough and difficult to quantify: in current research, on the one hand, most of the evaluation indexes of patent quality are still at the theoretical stage, which is far from practical applications; on the other hand, the system is not sound more index dimensions should be considered. Important factors affecting patent quality: the content of the patent text itself, the knowledge information mined in the patent text, and the domain knowledge information mined are not included in the existing patent quality evaluation indicator system
- (2) The difficulty of quantification of the Chinese patent quality evaluation index system and the lack of data make it impossible to implement automatic evaluation: there is no relevant research on the automatic evaluation model of Chinese patent quality in China. On the one hand, due to the aforementioned reasons, most of the Chinese patent quality evaluation indexes are still stuck at the theoretical level; it cannot be quantified, and thus cannot support the automatic evaluation of Chinese patents. On the other hand, due to the lack of data on Chinese patent quality levels, it is impossible to construct a data-driven automatic model of Chinese patent quality

In order to make the most of effective role of patent quality assessment and identify high-quality patents, it is urgent to adopt certain technical means to make this process easier. Faced with the three common problems in the field of patent quality evaluation, this article mainly uses data mining and natural language processing technology to extract indicators from a large number of patent information data of different dimensions and uses deep learning methods to predict patent quality. In addition, multitask learning

method is used to improve the accuracy of patent quality evaluation and prediction. The constructed patent quality evaluation model (PQE-MT) can automatically evaluate the quality of patents with a high accuracy rate (83.9%), saving a lot of valuable manpower and material resources. Regarding the lack of data labeling of Chinese patents, a large number of US patents with patent quality grade are used to construct basic training data. Therefore, a Chinese patent evaluation method based on transfer learning and active learning method is proposed in this paper. Then, the model is divided into two feature spaces, Chinese and English, and a network model is trained on these two features, respectively, to form collaborative training. The results of the evaluation show that the article method achieved a good migratory effect, with micro-F1 reaching 74%.

2. Related Work

2.1. Related Work of Patent Quality Index Quantification. In recent years, scholars have begun to summarize existing indicators from different dimensions and propose a more comprehensive evaluation system. Li et al. (2007) [1] summarized the technology cycle, technical scope, scientific relevance, and claims item number as for the professional indicators. On the other hand, the number of patent applications, patents home race number amount, and patents survival time is summarized as a comprehensive index and litigation. Through the above indicators, comparisons are made in terms of time, difficulty, and cost. Jin et al. (2011) [2] constructed a patent attribute network to optimize the prediction effect of algorithms from the aspects of patent content standardization, innovation, technology relevance, market value, patent inventors, and patentees. Han and Sohn (2015) [3] predicted the remaining effective time of the patent by constructing the similarity feature between the patent abstract and the claims and combining it with 14 other quantitative indicators. The article began to combine the text features of patents with patent attributes to make predictions. Yang et al. (2016) [4] broaden the patent quality impact indicator choice from the patent field correlation and life cycle stages. Leng and Zhai (2017) [5] constructed patent quality evaluation indicators, used the analytic hierarchy process to obtain the importance of the indicators to obtain the corresponding weight results, and used the comprehensive fuzzy evaluation algorithm to evaluate the patents in the medical field. In this article, different indicators are weighted to make them more targeted. Li (2018) [6] quantified patent quality evaluation indicators from the three dimensions of technology, law, and economy. The technical dimension mainly uses dependence, novelty, monopoly, and maturity; the economic dimension uses patent revenue and market evaluation, patent survival time, and other indicators; the legal dimension refers to factors such as patent legal status and litigation conditions. Finally, a comprehensive evaluation of the patent value is carried out in three dimensions. The above article explained and elaborated the patent quality impact indicators at the theoretical level, taking into account indicators of different dimensions and different levels, but some indicators are poor in

availability and difficult to express using algorithms. Meng (2018) [7] conducted a comparative analysis of patent data, combined with traditional patent quality evaluation methods, and quantified patent evaluation indicators in hierarchical levels, set up technical, economic, and legal levels, and further subdivided them into 14 different levels. At two levels, the weights of indicators at different levels are calculated through rough sets, and the final evaluation is performed through cloud models. The article summarizes the advantages and disadvantages of different indicator systems and summarizes the indicators, but the data discretization process and cloud model limit the generalization ability of the method. In general, the abovementioned scholars have put forward numerous patent quality evaluation indicators at the theoretical and application levels, and the topic optimizes and summarizes related indicators and further includes multiple dimensions from time, quantity, technology, law, inventor and agent, and text content. Quantify indicators, compare, and select key indicators as input to the patent quality evaluation model.

2.2. Related Work on Patent Quality Assessing. In terms of patent quality evaluation models, scholars gradually began to apply neural networks to patent quality evaluation. Chen and Chang (2009) [8] extracted indicators from medical patents in the United States, trained deep learning network models, and predicted the market value of patents. Trappey et al. (2011) [9] used the principal component analysis algorithm to obtain the impact factors of different indicators and used them as the input parameters of the backpropagation neural network model to judge the patent quality. As a preliminary screening scheme, the proposed patent quality evaluation method can automatically and effectively evaluate the quality of patents and save the time spent by domain experts on research and development of high-value patents. Wang (2012) [10] and others used indicators such as the actual benefits of patents as input parameters of the algorithm to construct a decision tree for prediction. First, use the analytic hierarchy process to get the value of the patent, then use the decision tree to get the risk of the patent, and combine the two to get the patent status. Zhao et al. (2013) [11] extract numerical indicators such as the number of patent citations and use decision trees, SVM, and deep learning methods to predict patent quality. Articles consider a variety of impact indicators progress in different model experiments; however, due to the use of only numeric attributes, the experimental structure is greatly restricted. Wu et al. (2016) [12] use the self-organizing mapping (SOM) method to cluster the patents to be tested with related patents published in the same field and perform nonlinear space transformation through kernel principal component analysis (KPCA), using SVM established a patent quality classification model, and proposed the somo-kpca-svm model to predict patent quality. The classification method proposed in the article can effectively display the analysis results, but the correctness of the results lacks the reference to the actual categories. Qiu et al. (2017) [13] use CART to extract evaluation indicators, optimize input parameter items, reduce model scale, and improve fitting effect. This method pro-

vides practical methods and theoretical support for the selection of patent indicators. The article uses classification regression trees to solve the patent evaluation problem, but the indicators used in the article are all basic patent attributes, and many influencing factors have not been fully considered. The above three articles did not take into account the textual content of the patent, so that the prediction model lacks vital influencing factors. Lin (2018) [14] applied neural networks to the quality evaluation research of patents and proposed the DLPQE model. The model is composed of two-part vector of the attribute representation structure ANE composed of the patent citation network and the text sequence feature represented by the convolutional neural network connected with the attention mechanism. The article proposes for the first time that the deep learning method is used for patent document analysis, taking into account different indicators and text content, but due to lack of the deep quantification of patent indicators, the convolutional neural network will lose the time series characteristics of the text. The subject combines quantitative indicators with text content and applies multitask learning to the patent quality evaluation model for the first time.

2.3. Related Work on Transfer Learning and Active Learning. Since data labeling is very time-consuming and labour-intensive, high-quality labeling data is scarce. The learning of transfer attracted increasing attention from the academy. Banea et al. (2008) [15] proposed a method of learning to transfer to a foreign language because of the numerous English entry data in this document, and a corpus of English data is used to generate the source domain data set of the target language through automatic translation, which demonstrates the feasibility of using automatic translation for learning to transfer through languages. Pan and Yang (2009) [16] propose transferring domains with sufficient training data to domains with similar data distribution and avoiding costly data annotation tasks via knowledge migration to greatly improve the learning effect and discuss transfer learning and domain adaptation, sample selection, the relationship between deviation, and other related machine learning technologies. Xu and Yang (2011) [17] use transfer learning and multitask learning to extract and transfer useful knowledge from auxiliary domain data. Weiss et al. (2016) [18] introduced the latest developments in the field of biological information and explained the information on learning and transfer solutions and discussed possible future research work. Transfer learning solution has nothing to do with the size of the data. Yosinski et al. (2017) [19] conducted research on the transferability of deep neural networks. Perform finetune experiments layer by layer in different layers to explore network transferability. Note that adding an end-to-end to the deep migration network will improve the effectiveness and overcome better data differences. Migration of the number of layers of the network can speed up learning and network optimization, general characteristics of the low-level network learning, and the high-level network learning field feature.

In order to make the model conform to the data distribution characteristics of the target area, a small number of

materials still need to be marked. Active learning can predict the results and select samples according to the model, and it is most helpful to improve the model. Thompson et al. (1999) [20] use active learning to try to select the most useful example as training data for transfer learning. Experimental results show that skilled learning can significantly reduce the number of samples labeled when the algorithm achieves the same effect. Tong and Koller (2001) [21] use pool-based active learning. The algorithm does not need a randomly selected training group and can request samples for marking. The new algorithm of support vector machine is used for active learning, which provides a theoretical basis for the concept usage algorithm of space version. Experimental results show that the use of the active learning method in the article can significantly reduce the need for labeled samples. Active learning, literature reviews, and less labelled training examples were introduced by Settles (2009) [22]. Discuss the query plan and analyze the experience and theoretical evidence of active learning. Li et al. (2016) [23] propose an active multigrade multilabel learning based on the minimum SVM classification range to take the minimum SVM classification distance as the confidence of the selected example, effectively reduce the number of sample annotations, and improve classification performance. Zhou et al. (2017) [24] propose the impact of active and transfer learning, data expansion, fault selection, continuous information extraction, and other methods on remote data and diagnosis. According to Zhu and Bento (2017) [25], GAN is used for active learning, and an object-oriented learning model GAN and a learning algorithm for motion value samples are given. Konyushkova (2017) [26] tried to transform active learning into a regression problem and dealt with the overfitting of previous query schemes. The experimental results show that the order reduction of data points is feasible in multivariable finite element analysis.

2.4. Related Work on Collaborative Training. After the transfer learning and active learning are used to promote the model to adapt to the feature distribution of the new data set, since the patent data has two views in Chinese and English, the experiment is carried out collaborative training in the remaining unlabeled sample set, making full use of the different learning features of the model on the two views. Features enhance the overall effect of the model. Blum and Mitchell (1998) [27] first tried collaborative training on web content prediction, learning from different views according to the model, using feature differences in different views to predict the labeled samples, and selecting high-confidence data as algorithms in other views training data. Wan (2009) [28], according to the feature difference between Chinese and English bilinguals, combined English data sets with a large number of emotional labels and unlabeled Chinese data, using machine translation technology to translate each other, training models for the two languages separately, and predicting based on the model the probability of each category of the sample is obtained, and the sample with high confidence is obtained as the training data of other models to achieve the purpose of bilingual collaborative training. Guo et al. (2012) [29] first proposed the use of graph-

based confidence estimation semisupervised collaborative training algorithm, which improved the speed of collaborative training in obtaining the data to be labeled to a certain extent. According to the information of the sample itself, the probability of the category of the unlabeled sample is calculated, and the confidence of the unlabeled data is estimated by a multiclassifier, which improves the effect of the collaborative classification algorithm, and proves the effectiveness of the algorithm on the UCI data set. Qiao et al. (2018) [30] use a collaborative training method on the neural network model to divide the sample features into multiple different views and fit the corresponding network parameters in different views. Experiments found that different models learned different characteristics of the data. The characteristics learned by different models are generally complementary. Gong and Lv (2019) [31] jointly use active learning, density peak clustering, and collaborative training to increase the amount of information in the data to be labeled and to a certain extent alleviate the mislabeling of fuzzy samples. Huang and Huang (2019) [32] use collaborative training for machine translation technology to improve the translation effect of the model. The experimental comparison shows that after the use of collaborative training, the results of machine translation are more accurate. When there are fewer labeled samples, there are still better results and translation quality. This topic makes full use of the unique Chinese-English bilingual features of the research and applies collaborative training to the Chinese patent quality evaluation model.

3. Patent Quality Evaluation Model

This paper uses transfer learning and active learning methods based on the PQE-MT model built with English patent data, further proposes a cross-language transfer of patent quality evaluation model based on active learning data expansion, and transfers the original model to Chinese patents. Finally, cooperative training is carried out. We will introduce the model from these three parts.

3.1. Patent Quality Evaluation Model Based on Multitask Learning. We first proposed the PQE-MT_USA model to predict the quality of US patents and selected the best-performing model structure through the comparison of different algorithms to facilitate the migration to Chinese patents in subsequent research. The related technologies involved in this part of the content are as follows, among which Word2vec and Bert are used for vector representation of text content to extract more semantic information; LSTM is used for learning the long-term dependence of the sequence, and Bi-LSTM also considers the information of the article context; CRF can improve the effect of named entity recognition based on the dependence of label results; multitask learning combines named entity recognition tasks with patent quality level prediction tasks to speed up model training and improve model fitting effects; TextCNN is used as comparison experiment of the effect of product neural network and recurrent neural network in this research.

3.1.1. Construction and Quantification of Patent Quality Evaluation Indicators. Due to the lack of systematic and fair evaluation grades and evaluation methods for Chinese patents for a long time, there is no clear quality grade. Grading the quality of patents requires professional knowledge and analysis of a large number of related patents, making it difficult to obtain the labels of Chinese patent data, and there is no reference scale for the evaluation process. Through research, it is found that the US patent has a patent quality grade, and the patent quality grade is recognized by scholars. Therefore, the experiment selected US patents in the field of new energy as the research object.

The experiment analyzes the information contained in the patent and specifically removes redundant information (such as patent application number: a numerical combination of other patent information) and overly complex information (such as patent specification: the text of the detailed description of the patent). Part, the content is tens of thousands of words, most of the content is the detailed introduction of the patent, which contains less effective information, and most of the main information exists in the abstract and claims of the patent and image information (patent drawing: prototype of the patent). As shown in the figure, this experiment mainly conducts research in the direction of natural language processing, focusing on patented text and numerical data. Part of the image information will not be considered for the time being.

The original attribute information obtained after screening is shown in Table 1 based on the attributes of these existing patents, and the experiment will quantify patent quality evaluation indicators in multiple dimensions and directions.

The experiment carried out in-depth processing of the above data items, mainly quantified and combined the following dimensions, including time dimension, quantity dimension, technical dimension, legal dimension, inventor, and agent dimension, and obtained related to patent quality quantitative indicators. And Figure 1 shows the division of some related attributes in different dimensions.

- (1) *Time Dimension.* The experiment will process the various time attributes of the patent and convert it into computable numerical data. In the end, indicators such as the time from patent application to publication, the time from application to approval, the time from publication to approval, the survival time, and the remaining time to termination are obtained
- (2) *Quantity Dimension.* The change in the number of patent applications in a certain field shows the development trend of patents to a large extent. The increase in the number of patent applications means good prospects, and the decline in the number of patent applications indicates that this field is gradually being replaced
- (3) *Technical Dimension.* CPC is the joint patent classification, and IPC is the international patent classification. Both have different partitions for different parts in detail, but the overall structure is similar. Each level is a more detailed classification of the previous level

- (4) *Legal Dimension.* Different legal status indicates the degree of importance the assignee of the patent attaches to the patent, including survival, revocation, withdrawal, reissue, and termination. Since the patent right was granted, the assignee has to pay a certain fee every year to maintain the validity of the patent, and the fee increases year by year. When the patent cannot satisfy the interests of the assignee, the assignee will stop paying the fees, resulting in the cancellation of the patent
- (5) *The Dimensions of Inventor and Assignee.* A good inventor and assignee means a good patent. We get the sum of the number of patents invented by all inventors in the field; the average number of invention patents per inventor in the field, etc. In addition to the same indicators as the inventor, the patent assignee also extracts whether the assignees of previous and future generations have changed, the type of organization, and so on. The assignee involves different organizations, including companies, research institutes, universities, colleges, foundations, and individuals
- (6) *Combined Indicators.* We combine the feature of different dimensions to get combination indexes. Such as the proportion of patent survival, withdrawal and expiration of inventors, the number of patents in different technical fields and different years, and so on. The following is a list of the combination indexes of the assignment. In a specific patent agent and a specific field, Num_Application means the experimental definition: the number of patents applied; Num_Approval means the number of patent approved, Q_2 means the number of IPC and CPC, Q_3 means the size of patent family, and Q_4 means the average patent quotation rate. There are the following quantitative indexes.

$$\text{Research_Focus}_i = \frac{\text{Num_Application}_{\text{filed}}}{\sum_{i=\text{filed}1}^{\text{filed}n} \text{Num_Application}_i}, \quad (1)$$

$$Q_1 = \frac{\text{Num_Approval}}{\text{Num_Application}}, \quad (2)$$

$$\text{Strength} = \text{Num_Apply} * \sum_{i=1}^4 Q_i, \quad (3)$$

$$\text{Technical_Portion}_i = \frac{\text{Strength}_{\text{inventor}}}{\sum_{i=\text{inventor}1}^{\text{inventor}n} \text{Strength}_i}. \quad (4)$$

In addition, the experiment adds long text description type attributes. On the one hand, the experiment takes the publication time of the patent as the node, uses the TF-IDF algorithm to calculate the text similarity of other patents before and after the node, and selects the top 50 with the highest similarity as the result of the similarity before and after the publication of the patent. The smaller

TABLE 1: Filtered patent raw data information.

Attributes	Value
Patent title	Automated electric vehicle charging system and method
Quality grade	6
Legal status	Alive
Country of origin	US
Agent	Interim Design Inc.
Inventor	Haddad; Joseph C.; Elizabethtown; Lysak; Daniel B.; state college
Application date	2016.01.29
Public day	2016.05.26
Approval date	2018.10.23
Expiry date	2032.04.25
IPC	H02J-007/00; B60L-011/18
CPC	B60L11/1835; B60L53/14; B60L53/35; Y04S30/12; Y04S30/14
International patent documentation center	6
Number of families	
Number of ordinary families	5
Cited by other patents	3
Number of cited patents	4
Number of references	0
Cited by other documents	0
Abstract	A system and method for charging an electric vehicle include identifying vehicle information corresponding...
Claims	A system for charging an electric vehicle, comprising: a camera configured to acquire an electronic image of the electric vehicle...

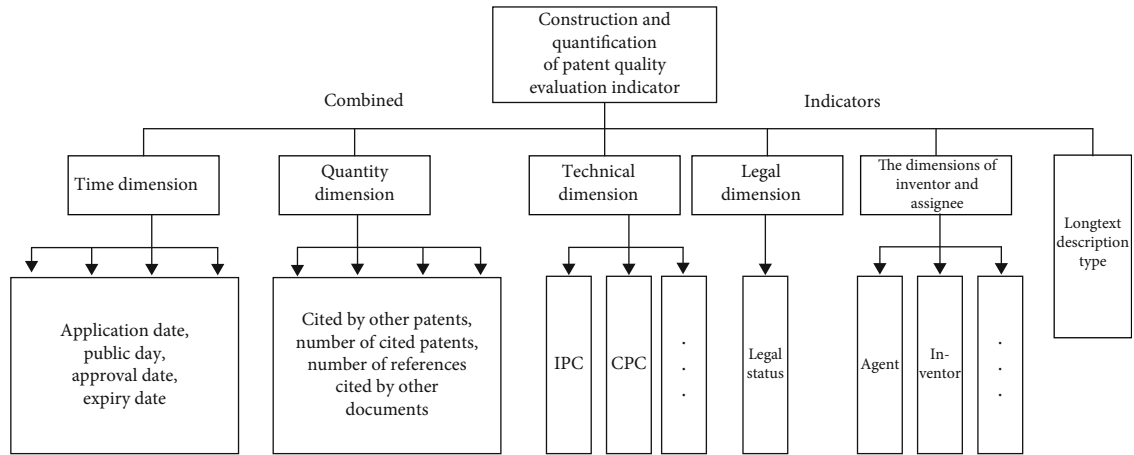


FIGURE 1: Construction and quantification of patent quality evaluation indicator.

the similarity result before the node, the more novel the patent; the greater the similarity result after the publication time node is, it indicates that the patent is groundbreaking and has been used for reference by other related patents, as well as the number of technical elements (field terms) in the main claim and the positive words (can, enable, so as to, etc.). On the other hand, the patent name, patent abstract, and patent claims are input into the sequence model to participate in the final prediction.

The PQE-MT_USA model mainly involves two parts: a quantitative index model and a sequence model. The sequence model uses a multitask learning structure. Next, we will introduce the model from these two parts.

3.1.2. Patent Quality Evaluation Model Based on Quantitative Indicators. The patent quality evaluation model mainly includes input, index quantification, fully connected layer, softmax layer, and output. Its structure is shown in Figure 2.

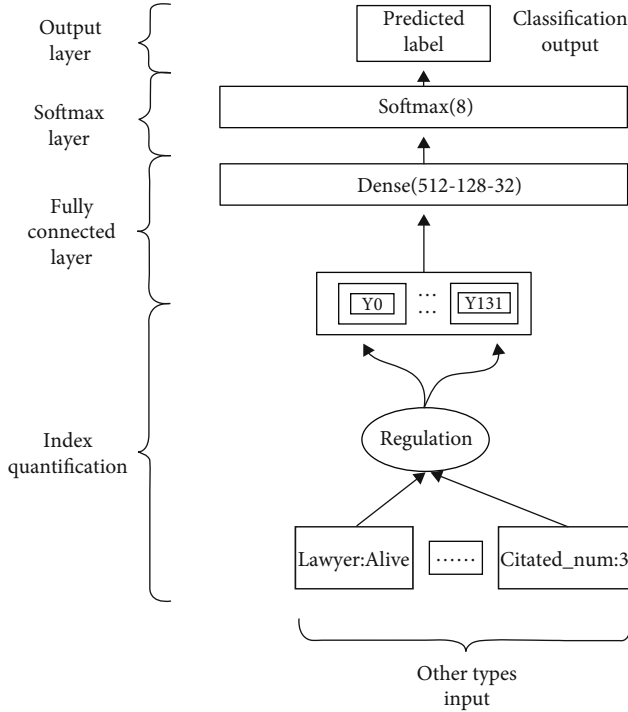


FIGURE 2: Quantitative index model structure.

First, we enter the initial patent attributes, including legal status and number of citations. According to relevant rules, 15 initial indicators and 117 quantitative indicators in different dimensions are constructed, which together constitute 132 indicators. We combine the patent text with 132 indicators into a fully connected layer composed of 512-128-32 nodes. Finally, multiple classifications are performed through the softmax layer to predict the patent quality level. Patent quality levels are divided into eight categories. Take the “Dynamic power supply management system for electric vehicle” patent as an example, its legal status is “Alive,” the number of citations is “4,” and other patent attributes and 132 evaluation indicators are combined to finally get its patent quality grade prediction result. The result is 8.

3.1.3. Multitask Learning Model

(1) *Task 1: Text Classification Model.* This task preprocesses the title, abstract, and claims of the patent and predicts the quality of the patent through the deep learning model described below. The detailed structure of the text classification model is as follows.

The model takes the three parts of patent title, abstract, and claims as input and first performs text processing on it, including splicing the input items, removing stop words, removing special symbols, converting to lowercase, and converting to word roots. The experiment uses Word2vec and Bert as the initial vector representation of words. The processed data is fine-tuned in Bert to obtain the word vector of a specific field as a word embedding. Next, through the Bi-LSTM layer, fusing the attention mechanism, LSTM can

learn long-term dependencies, and after bidirectional splicing, it can effectively use the context information of the text to dig out more hidden features.

The Bi-LSTM in this model is equivalent to inputting the sequence into a forward LSTM and a back-end LSTM, respectively, and the output result is used as the result. The structure of LSTM is shown in Figure 3. Taking the title part of the patent text “Test system for battery management system” as an example, the basic unit of LSTM is used to encode the text. First, input the word vector “Test” and pass through the forget gate, memory gate, and output gate. The calculation generates a new hidden state h_t (test) corresponding to it. The calculation process of the three gates is as follows:

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f), \quad (5)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i), \quad (6)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o), \quad (7)$$

$$c_t = f_t \circ c_{t-1} + i_t \sigma_c(W_c x_t + U_c h_{t-1} + b_c), \quad (8)$$

$$h_t = o_t \circ \sigma_h(c_t). \quad (9)$$

Among them, x_t is the input of LSTM, f_t is the forget gate, it is the input gate, o_t is the output gate, h_t is the hidden gate, c_t is the cell state, W , U , and b are the matrices in training and the network learning calculation element value, and σ is the sigmoid function, which ensures that the output result is between 0-1, which is used to control the proportion of information passing.

The second word vector “system” and the hidden state h_{t-1} (test) output at the previous moment are used as new inputs. After three gate calculations again, the updated hidden state h_t (system), h_t (system) is obtained. Contains information about the characters entered in the preceding text. Repeat the above coding process until the end of the patent text.

When the input sequence is very long, it is difficult for the model to learn a reasonable vector representation. The attention mechanism [33] retains the intermediate results of the LSTM encoder on the input sequence, selectively learns the input and associates it with the output sequence, so that the model focuses on the words that are considered more important in the input sequence. Then, through a fully connected layer network composed of 512, 128, and 32 nodes, the softmax layer is used for multiclassification to predict and output the patent quality level. The specific structure is shown in Figure 4.

(2) *Task 2 (NER Task): Sequence Labeling Model.* This task is used to identify the terms in the patent text. The first few layers of the model are the same as task 1. Here, we focus on the CRF layer and the output layer.

In the NER task, the decoding layer usually uses the conditional random field (CRF) model to perform label

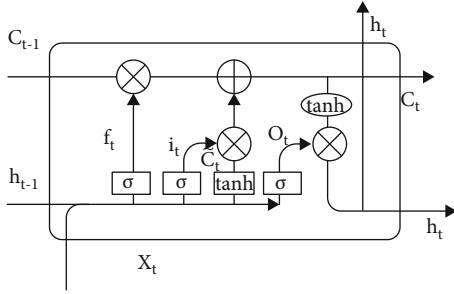


FIGURE 3: LSTM structure diagram.

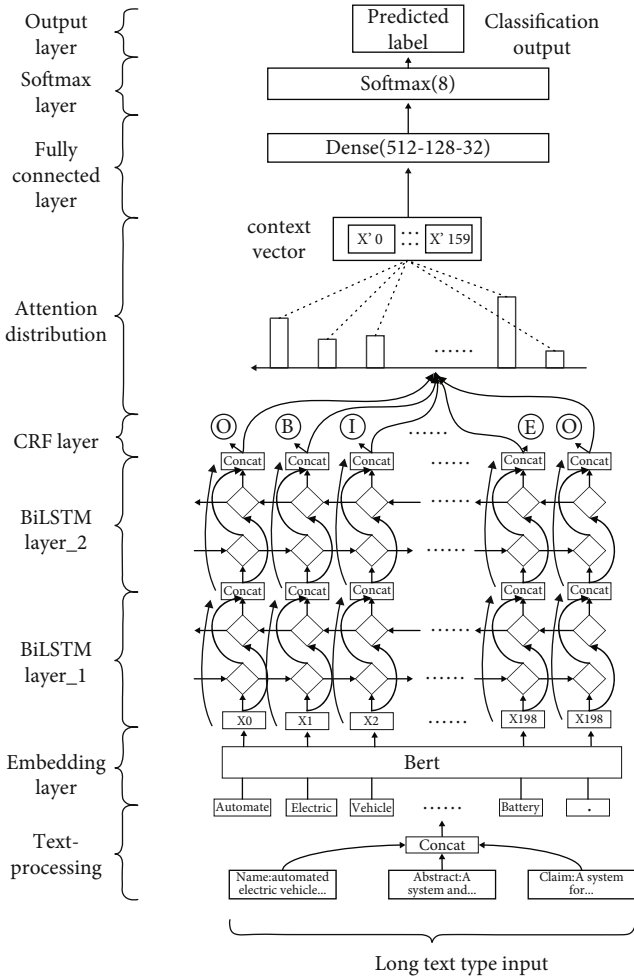


FIGURE 4: Text classification model.

inference on the entire character sequence in turn. The decoding process is as follows.

For an input sequence, the input vector is obtained after embedding to the LSTM, and the score on the label corresponding to each word is obtained after the linear layer is applied. According to the label transition matrix T , we can get the score of the label at the previous moment as y_i and the label at the next moment as y_{i+1} , namely, $T[y_i, y_{i+1}]$. For a sequence x , if the length of the sequence x is n and there are m possible labels, then there are a total of m^n possible labeling results. We can use the LSTM + CRF model to

calculate the score (y) of each possible annotation result, and then use softmax to normalize to find the probability of a certain annotation result, and choose the one with the largest probability as the annotation result. The specific calculation process is as follows:

$$\phi_t(y', y | x) = \exp(w_{y', y}^T h_t + b_{y', y}), \quad (10)$$

$$p(y | x; \theta) = \frac{\prod_{t=1}^n \phi_t(y_{t-1}, y_t | x)}{\sum_{y' \in Y_{(x)}} \prod_{t=1}^n \phi_t(y'_{t-1}, y'_t | x)}. \quad (11)$$

Among them, $w_{(y', y)}$ and $b_{(y', y)}$ are the training parameters of the label pair (y', y) , h_t represents the output of the coding layer at time t , θ represents the model parameters, $Y_{(x)}$ represents all possible tag sequences corresponding to the character sequence x . In the tag reasoning process, CRF needs to find the tag sequence y^* that maximizes the conditional probability given the input sequence x :

$$y^* = \operatorname{argmax}_{y \in Y_{(x)}} p(y | x; \theta). \quad (12)$$

The search problem of the tag sequence y^* can be solved efficiently using the Viterbi algorithm.

The experiment defines the label space as $\{B, I, E, O\}$. “B” indicates that the element is the beginning of a domain term, “I” indicates that the element is the middle part of a domain term, “E” indicates that the element is the end of a domain term, and “O” indicates that this element does not belong to the domain term part. The CRF layer can determine the domain terms in the sequence data according to the label sequence output by the model. In the final output, each element is marked as a label in $\{B, I, E, O\}$. Take the patent text of “Test system for battery management system” as an example, the sequence length is 6, there are four possible tags for BIEO, and a total of 6^4 possible tagging results, including $y = (OOBIE)$, $y = (BEOOBI)$, \dots , after the calculation (OOBIE) marked the highest score, that is to identify “battery management system” as a patent term. The specific structure of the model is shown in Figure 5.

The experiment considers that the two tasks overlap greatly, and the domain terms in the text sequence are the part that has a greater impact on the patent level. The sequence labeling task backpropagates to update the parameters, which is conducive to the model to obtain a better sequence representation and pay more attention to the learning of domain terms. The two tasks together form a multitask learning structure, which promotes each other and simplifies the overall task complexity and prediction time. Take the patent text of “Automated electric vehicle” as an example, its specific structure is shown in Figure 6.

This figure combines the previous two models. The left side is the sequence model part of multitask learning, and the right side is the quantitative index model part. The PQE_MT_USA model is obtained after the two parts of vectors are spliced to predict the final result and used for subsequent migration training of Chinese patents.

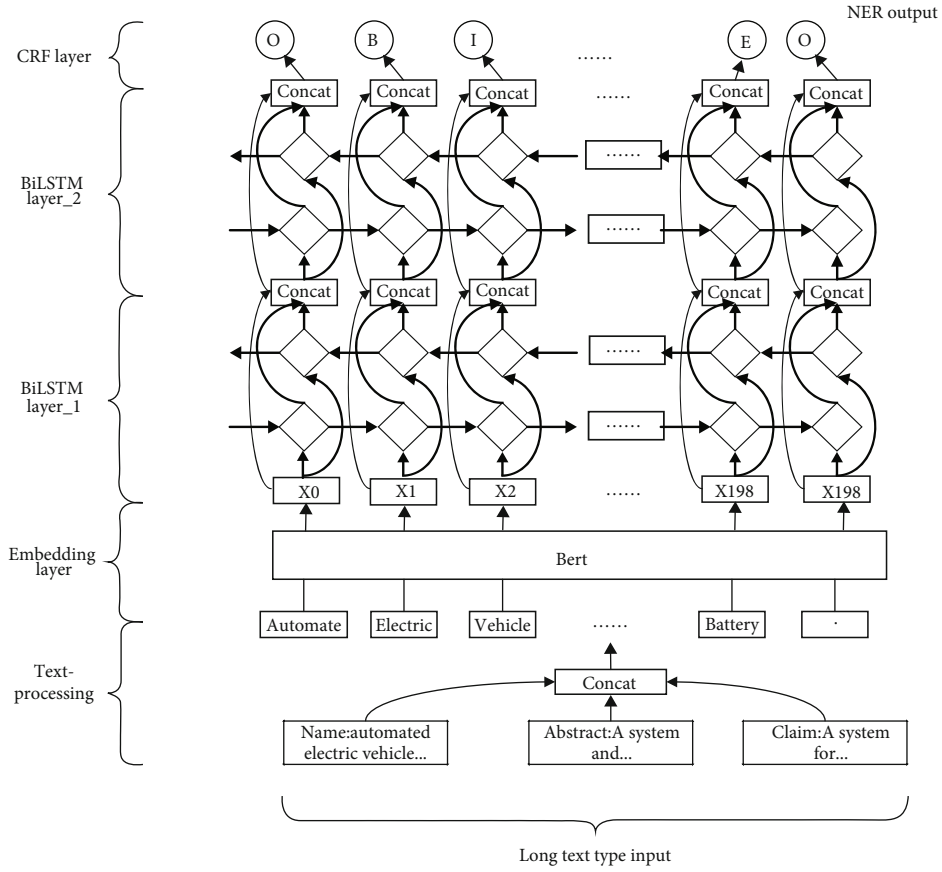


FIGURE 5: Sequence labeling model.

3.2. Patent Quality Evaluation Model Based on Transfer Learning. If the model PQE_MT_USA trained by the US patent is directly applied to the Chinese patent, the predicted patent rating result will have a large error. The reason is that if you use traditional machine learning methods, you need the data to obey the same distribution and sufficient labeled data. On the issue of labeling data, the process of labeling data consumes a lot of manpower and time costs; on the issue of data distribution, on the one hand, there are cultural differences in the writing process of Chinese and English patents. On the other hand, there are differences in the distribution of various attributes and quantitative indicators of Chinese and US patents. By comparing Chinese and American patent data, it is found that the inventors, attorneys, IPC classification numbers, and CPC classification numbers of Chinese and American patents have the same expression. Although the inventors and agents of the Chinese and American patents belong to two different sets, the experiment quantifies the indicators based on the overall situation in the field, and the training model will not be affected by the specific values. The differences in patents between the two countries are mainly reflected in the following four aspects: (1) the legal status of patents is different; (2) there are differences in the distribution of patents between the two countries in the following 6 attributes: number of cited patents, number of citations by other patents, number of common families, number of citations by other documents, number of references, and international patent docu-

mentation center; (3) the patent origination level is different; and (4) the content of the text type is different, including the title, abstract, and claims of the patent. The writing language of the US patent is English, and the Chinese is Chinese. In terms of related indicators, US patents are generally higher than Chinese patents.

Because the US patent has the patent strength label required for experimentation, the Chinese patent has the contradiction between a large amount of data and a small amount of labeling. At the same time, the impact of various attributes and indicators on patent strength has the same trend, with only the difference in feature distribution. Transfer learning breaks the same distribution assumption of traditional machine learning and can adapt to the requirements of experiments well.

Transfer learning mainly involves domains and tasks. Given a labeled source domain, that is, an English patent text with a large number of labeled data sets: $D_s = \{x_i, y_i\}_{i=1}^n$, and an unmarked target domain, that is, the Chinese patent text without data annotation: $D_t = \{x_j\}_{j=n+1}^{n+m}$. The data distribution of these two fields $P(x_s)$ and $P(x_t)$ are different, $P(x_s) \neq P(x_t)$ [34]. The purpose of transfer learning is to use the knowledge of D_s to learn the knowledge (label) of the target domain D_t . That is to say, the process of transfer learning is to transfer knowledge from the source domain to the target domain, complete the model update, and predict the target domain label. Transfer learning can be divided into different

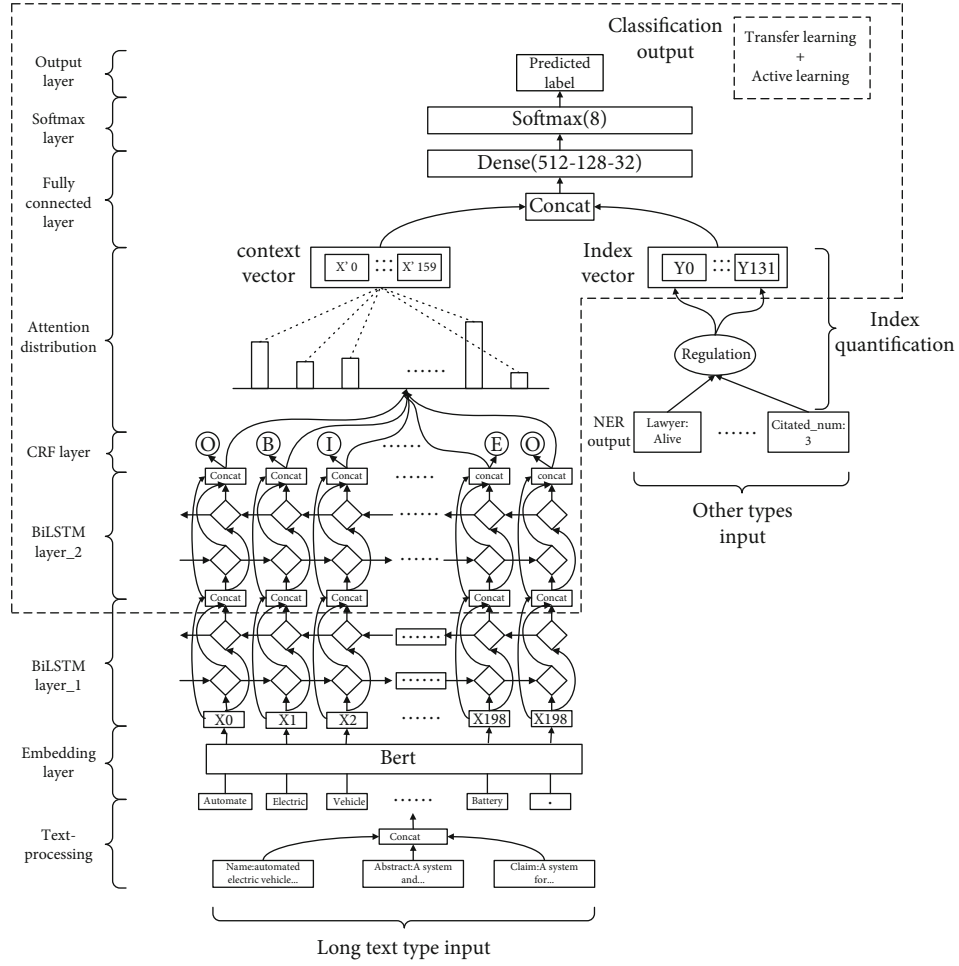


FIGURE 6: PQE_MT_USA model structure.

types, and there are many classification methods. This article mainly uses the model-based transfer learning method, that is, based on the self-adaptation of model parameters to find new parameters θ , and the transfer of parameters makes the model better at the target work on the domain to minimize its loss. The calculation formula is as follows:

$$\min \frac{1}{n} \sum_{i=1}^n L(x_i^s, y_i^s, \theta). \quad (13)$$

The experiment uses US patents as a domain of origin and Chinese patents as a reference domain. It is impossible to transfer directly between the different languages. This concerns the problem of the classification of texts in the cross language. The experiment uses automatic translation to translate the text of Chinese and English patents and marks the patent levels of some Chinese patents. The model adapts to the distribution of indices and to the linguistic characteristics of Chinese patents. Figure 7 shows the transfer learning process.

To put it simply, taking the patent text in the field of new energy vehicles used in the experiment as an example, the process of transfer learning is to first translate the English

patent text in the United States (source domain data) into Chinese patent text (target domain data) and use it as the input of the PQE_MT_USA model mentioned in the previous section is vectorized using Bert, and the model is retrained through the Bi-LSTM encoding layer and CRF decoding layer. The initial parameters are the final parameters obtained by PQE_MT_USA training English text. After the training is completed, new parameters suitable for the Chinese patent grade prediction model are obtained, and then other Chinese patent texts (target domain data) are tested and adjusted to obtain the transferred parameters.

Through comparative experiments, try to freeze the parameters of different layers to obtain the best migration effect. The specific experimental process and results will be described in detail in the fourth part of the experimental results and analysis. The final experiment selects the part shown by the dotted line in Figure 5 to transfer the model parameters.

Although migration learning requires less data, too little migration data for larger networks will still cause overfitting problems in the model. The model needs more migration samples to make the network model have better generalization capabilities. The experiment contains more than 20,000 Chinese patent data. Labeling a large number of data

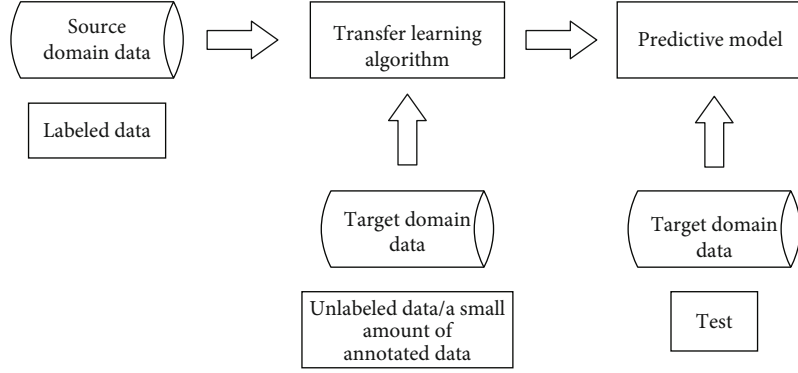


FIGURE 7: Transfer learning flowchart.

samples requires a lot of manpower and material resources. In the process of transfer learning, which patents should be selected for labeling has a greater impact on the results of the transfer. With the same amount of target domain data, selecting samples that are difficult for the model to choose has a better effect than selecting samples that can be clearly classified by the model. The process of letting the model actively propose which data needs to be labeled is active learning. Transfer learning reduces the amount of sample labeling, and active learning improves the quality of labeling samples.

The proposal of active learning is mainly due to the different amount of information provided by each sample in the training set for model training, that is, different samples have different contributions to the improvement of the model effect. Active learning mainly includes classifiers, labeled training data sets, unlabeled data sets, query functions used to extract samples with large amounts of information in unlabeled data sets, and supervisors who label the extracted samples.

The experiment selects more efficient and mature methods based on uncertainty reduction. The main query functions include methods based on classification uncertainty $U(x)$, classification margin $M(x)$, and classification entropy $H(x)$. The specific calculation method is shown in the following equations.

$$U(x) = 1 - P(\hat{x} | x), \quad (14)$$

$$M(x) = P(\hat{x}_1 | x) - P(\hat{x}_2 | x), \quad (15)$$

$$H(x) = -\sum_k p_k \log(p_k). \quad (16)$$

The experiment uses the most effective method based on classification margin for multiclassification problems, namely, the best versus second best (BVSB) strategy [35] to calculate the uncertainty of the sample. This method calculates the difference of the category with the highest probability among the sample prediction results as a measure of sample selection. The smaller the difference, the greater is the uncertainty. Figure 8 lists the two real samples in the experiment. In contrast, the sample on the right has a smaller probability difference and greater uncertainty, and

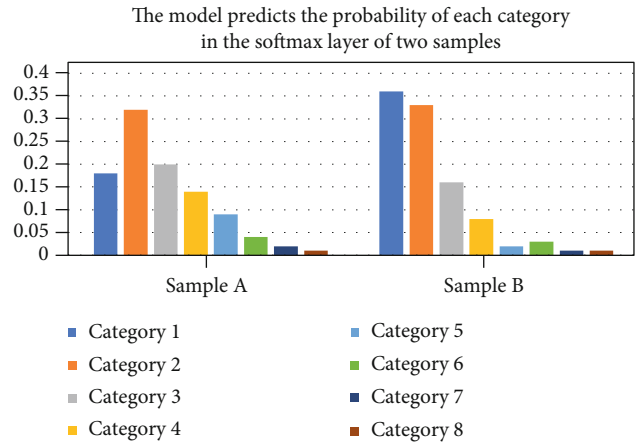


FIGURE 8: The model predicts the probability of each category in the softmax layer of two samples.

the model is more inclined to select it as the sample to be labeled.

By analyzing the output value of each category of the two samples, it is observed that the highest probability of sample A is category 2, its probability is 0.32, the category with the highest probability of sample B is category 1, and its probability is 0.36. Although the final predicted result of sample B is more probable, it can be observed from the overall probability of each category that the model is more confusing for sample B because the difference between the highest category probability and the second highest category probability is smaller, based on classification, the active learning algorithm of margin is good at extracting samples with higher uncertainty for multiclassification problems.

The following is a detailed introduction to the process of transfer learning and active learning algorithms based on the Chinese patent.

First, the English text Patent_USA_En in the source domain data set is translated into Chinese text Patent_USA_Ch, the Chinese patent quality evaluation model PQE-MT_Ch based on the US patent is obtained by Patent_USA_Ch training, and the frozen model PQE-MT_Ch does not require a training layer. Next, we use the unlabeled target domain Chinese patent data set Patent_CHN_Ch as the training set to continue training the model PQE-MT_

Ch. Then, we use this model to classify and predict the sample set Patent_CHN_Ch, calculate the probabilities C1 and C2 of the two categories with the highest predicted probability for each sample, and get the set S of 100 samples with the largest uncertainty, that is, the smallest C1-C2. Then remove the set S from the sample set Patent_CHN_Ch, manually mark the samples in the set S, add the marked set S to the training set, repeat the above training process, and finally get the migration model PQE-MT-CHN_Ch. This model comprehensively utilizes transfer learning and active learning algorithms and is suitable for the English patent quality assessment of Chinese patents. Figure 9 is a diagram of the overall process of transfer learning and active learning.

3.3. Patent Quality Evaluation Model Based on Collaborative Training. In the experiment of the above transfer model, it is not difficult to find that different languages make the model have different effects. The following research will start to use the difference in feature space between Chinese and English languages, and the different features and knowledge learned by the models in different languages, resulting in different prediction effects for different samples, and use bilingual collaborative training to improve the prediction results of the models.

Collaborative training is the mainstream semisupervised machine learning algorithm. The theoretical basis is by training classifiers under multiple views; under the existing small amount of annotated corpus, different features of samples can be learned according to different classifiers, and a large amount of category information of data to be annotated can be further obtained, thus, achieving the purpose of learning by using unlabeled data. The cooperative training process is similar to clustering hypothesis; on the whole, it can be explained that the classification model trained by a small amount of labeled data can only roughly describe the distribution of data, and it is difficult to accurately divide most samples. By using a large amount of unlabeled data, the most accurate data can be found based on the classification of the trained model, so that the model can further find a more accurate classification surface.

In the process of cross-language transfer learning, there is a language barrier between Chinese and English, and the feature spaces of the two languages are different, so it is impossible to transfer the model directly. The experiment used machine translation to unify the languages of the patents of the two countries, so the patents of both countries have two language versions. The experiment hopes to make full use of both Chinese and English feature spaces to further improve the prediction effect of the model.

The experiment trains Chinese patent quality evaluation model PQE_MT_CHN_Ch based on Chinese and Chinese patent quality evaluation model PQE_MT_CHN_En based on English. Through mutual supervision of the two models, a Chinese patent quality evaluation model PQE_MT_CHN_Co based on collaborative training is obtained.

In the experiment, the collaborative training algorithm used machine translation to translate the Chinese text of Chinese annotation data into English text. Train the Chinese-based Chinese patent quality evaluation model

PQE_MT_CHN_Ch on the Chinese text of the labeled Chinese patent data set. In the same way, the English-based Chinese patent quality evaluation model PQE_MT_CHN_En is trained on the English text of the labeled Chinese patent data set. Use these two models to label the unlabeled patent texts in the corresponding data sets and select high-confidence data from them to add to the labeled data set collection. Iterate the above training process. Next, the prediction results of the two evaluation models PQE_MT_CHN_Ch and PQE_MT_CHN_En are combined according to weights, and finally, the Chinese patent quality evaluation model PQE_MT_CHN_Co based on collaborative training is obtained. The overall flow of the algorithm is shown in Figure 10. The algorithm ensures that the sample data distribution is balanced with the real situation by setting the number m_i of the data obtained under each category.

4. Experimental Results and Analysis

4.1. Experimental Corpus. 59147 patents are used in the experiment, and the number of patents with different grades is shown in Figure 11. The experiment randomly shuffled the data set to ensure the uniform distribution of different label samples. 90% samples (53475) are for training and 10% samples (5942) are for testing. The verification method chooses k -fold cross-validation for model optimization. The $k - 1/k$ of the data set is used as the train set, and the rest is used as the validation set. The result is the average of k times. The value of k is set to 10.

There are 21611 Chinese patents, of which 308 are in the United States, which means that the two countries have applied for 308 patents; at the same time, this led to the formation of a monopoly family belonging to the same patent family, in order to meet quality standards, corresponding to Chinese patents. In this paper, 100 Chinese patents with patent grade are taken as the samples of the final model test, 208-as the original data for migration, and the remaining 21303 Chinese unmarked patents are used as the active learning pool for data screening, Mark.

All experimental data are random, and the extraction method is random. In order to ensure the uniform distribution of different label samples and the accuracy and fairness of experimental results, this paper involves Google's machine translation, and this is the most accurate translation system at present.

4.2. Experimental Settings. The accuracy, recall, $f1$ -score, accuracy, microaveraging, and macroaveraging used in the text classification evaluation are used for evaluation. The calculation process is as equations (17)–(26). For category C, the classification results can be divided into the following situations:

- (1) The original type C is classified as type C, and the quantity is recorded as a
- (2) The original non-C category is classified as C category, and the quantity is recorded as b

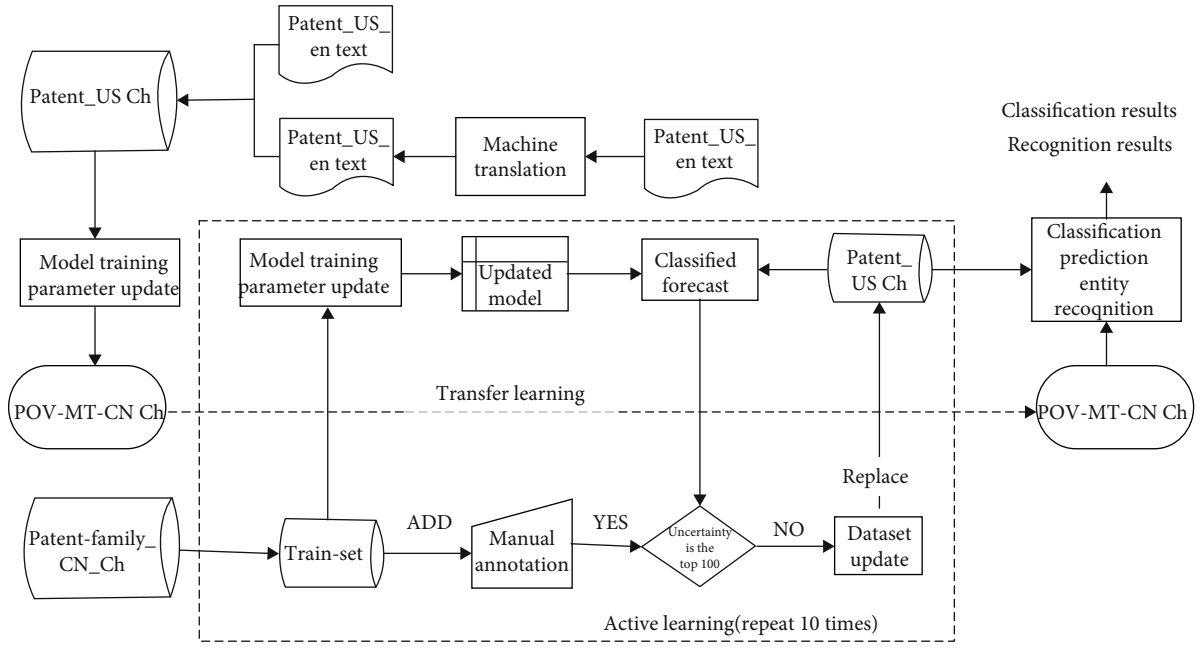


FIGURE 9: Overall process diagram of transfer learning and active learning.

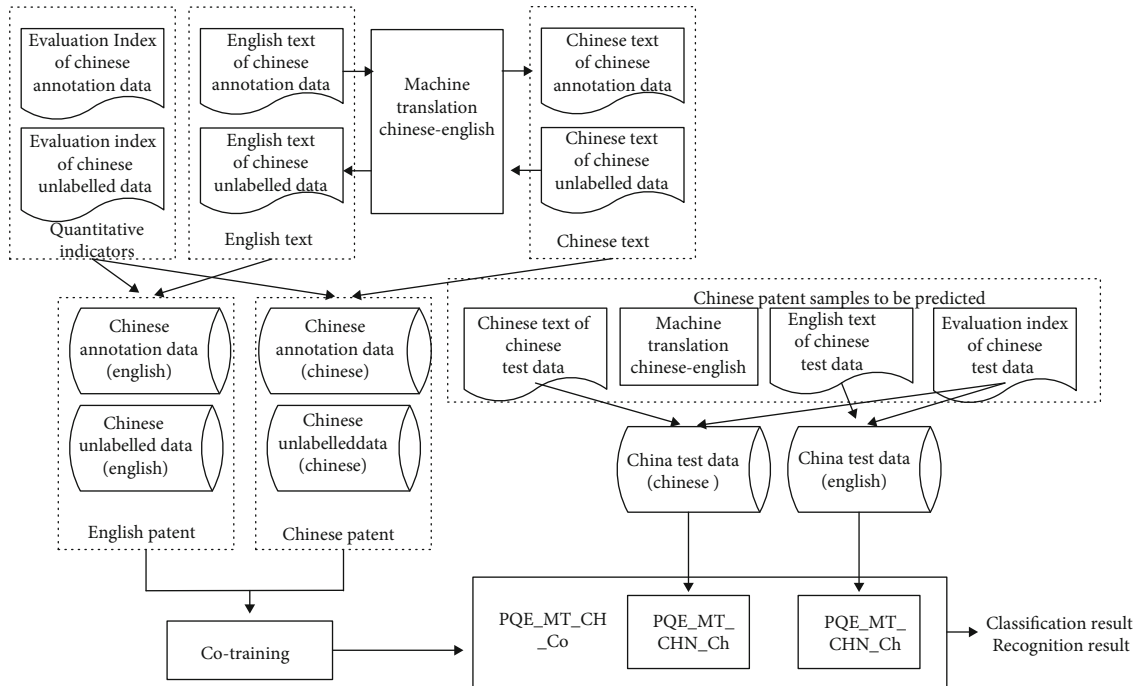


FIGURE 10: Bilingual collaborative training flowchart.

(3) Originally C is classified as non-C, and the quantity is recorded as c

$$R = \frac{a}{a+c} \times 100\%, \quad (18)$$

(4) The original non-C category is classified as non-C category, and the quantity is recorded as d

$$F = \frac{2 \times P \times R}{P + R} \times 100\%, \quad (19)$$

$$P = \frac{a}{a+b} \times 100\%, \quad (17)$$

$$A = \frac{a+d}{a+b+c+d} \times 100\%, \quad (20)$$

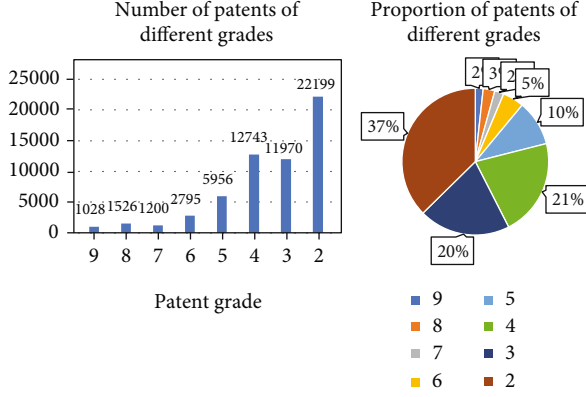


FIGURE 11: Number and proportion of patents with different ratings.

$$\text{Micro}_P = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n a_i + \sum_{i=1}^n b_i}, \quad (21)$$

$$\text{Micro}_R = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n a_i + \sum_{i=1}^n c_i}, \quad (22)$$

$$\text{Micro}_{F1} = \frac{2 * \text{Micro}_P * \text{Micro}_R}{\text{Micro}_P + \text{Micro}_R}, \quad (23)$$

$$\text{Macro}_P = \frac{1}{n} \sum_{i=1}^n P_i, \quad (24)$$

$$\text{Macro}_R = \frac{1}{n} \sum_{i=1}^n R_i, \quad (25)$$

$$\text{Macro}_{F1} = \frac{2 * \text{Macro}_P * \text{Macro}_R}{\text{Macro}_P + \text{Macro}_R}. \quad (26)$$

4.3. Experimental Results and Analysis of Patent Quality Evaluation Model

4.3.1. *Experimental Results and Analysis of Patent Quality Evaluation Model Based on Multitask Learning.* The experiment tested 15 initial patent indexes, 117 quantified indexes, and 132 indexes combined with the two, respectively. The results are shown in Table 2. The results show that both the initial attributes of patents and the quantitative index parameters in the experiment have influence on the evaluation of patent quality, but the evaluation effect is still poor and cannot meet the final demand. It will be combined with sequence model tasks to improve the effect.

The experimental sequence model and multitask learning model (PQE-MT) training results are shown in Table 3. For readability, the model is defined as follows:

- (i) *Model_1.* word2vec as the embedded layer of the model connects with Bi-LSTM
- (ii) *Model_2.* On the basis of Model_1, attention mechanism is added to form patent quality evaluation model

TABLE 2: Accuracy of initial patent attributes and quantitative index models.

Model input (size)	Train	Dev	Test
Initial index (15)	0.4809	0.4811	0.4813
Quantitative indicators (117)	0.6292	0.6248	0.6156
Combine (132)	0.6604	0.6609	0.6555

- (iii) *Model_3.* On the basis of Model_1, using the Bert after fine-tuning by all patent text data instead of word2vec
- (iv) *Model_4.* On the basis of Model_2, CRF is added to identify patent terms, forming a patent quality evaluation model based on multitask learning
- (v) *Model_5.* On the basis of Model_3, CRF is added to identify patent terms, forming a patent quality evaluation model based on multitask learning
- (vi) *PQE_MT_USA.* Model_5 combined with quantitative index model

The detailed results of our proposed PQE_MT_USA model on different indicators in each category are shown in Table 4.

Through the analysis of the experimental results, it can be concluded that

- (i) Model_2, which uses word2vec as embedding layer and Bi-LSTM as sequence layer, can achieve 58.25% accuracy on train set, 56.52% accuracy on validation set, and 55.22% accuracy on test set
- (ii) After adding the attention mechanism, the accuracy of the model is improved by nearly 2 percentage points, which proves the effectiveness of attention mechanism
- (iii) After replacing the embedding layer with BERT, the accuracy of the model has been greatly improved, but the generalization ability has slightly decreased
- (iv) After using the sequence model based on multitask learning, the effect of model is further optimized and the evaluation effect is further improved. Although multitask learning increases the scale of the model, but the training time of the model does not significantly increase because two tasks share learning parameters, only more 4128 parameters are trained. Each epoch in word2vec and BERT models only increases 54 s and 62 s, indicating that the model has better usability
- (v) The final model combines features of two parts and achieves good results in classification tasks. The accuracy of train set, validation set, and test set is 87.29%, 82.29%, and 83.90%, respectively. It shows that two parts of the model in this paper have an important impact on the evaluation of patent quality

TABLE 3: Results of models.

Model	Text classification			Train	NER		Time s/epoch
	Train	Val	Test		Val	Test	
Model_1	0.5825	0.5652	0.5522	N/A	N/A	N/A	1420
Model_2	0.6026	0.5805	0.5734	N/A	N/A	N/A	1526
Model_3	0.6373	0.5910	0.5962	N/A	N/A	N/A	1922
Model_4	0.6326	0.6054	0.6024	0.8461	0.8457	0.8490	1580
Model_5	0.6890	0.6253	0.6220	0.8541	0.8584	0.8566	1984
PQE_MT_USA	0.8729	0.8292	0.8390	0.8689	0.8795	0.8788	2310

TABLE 4: Results of PQE_MT.

Category	Precision	Train set			Precision	Test set		
		Recall	F1	Support		Recall	F1	Support
1	0.98	0.99	0.98	19967	0.96	0.98	0.97	2232
2	0.93	0.92	0.92	10743	0.89	0.87	0.88	1227
3	0.93	0.87	0.90	11534	0.89	0.81	0.85	1209
4	0.77	0.82	0.80	5330	0.65	0.72	0.68	626
5	0.58	0.80	0.67	2502	0.53	0.65	0.58	293
6	0.54	0.50	0.52	1087	0.48	0.36	0.41	113
7	0.53	0.67	0.60	1382	0.44	0.59	0.50	144
8	0.69	0.74	0.72	930	0.57	0.53	0.55	98
Micro	0.89	0.90	0.89	53475	0.84	0.85	0.84	5942
Macro	0.74	0.79	0.76	53475	0.67	0.69	0.68	5942

- (vi) The model has a good effect on a large number of categories, but a little worse effect on a small number of categories, and the result is positively correlated with the size of the categories as a whole

4.3.2. Experimental Results and Analysis of Patent Quality Evaluation Model Based on Multitask Learning. Different layers of the neural network model usually learn the characteristics of different levels of the sample. “Visualizing and Understanding Convolutional Networks” [36, 37] visualizes the layers of CNN and shows the above theory more clearly. The bottom layer 1 and 2 of the model network can usually learn the basic color and edge features of the object; the third layer of the model network generally learns the texture features of the object; the fourth layer that continues upward can learn local features, such as wheels; the top level learns more discernible overall characteristics.

For the image training model, the results of each network layer are easy to observe. In contrast, the knowledge learned by the text training sequence model is difficult to clearly display, but the above rules are also met on the whole. The underlying structure of the network learns the text the part-of-speech and word-sense characteristics of the word; high-level results learn semantic and syntactic characteristics. In response to this phenomenon, experiments try to perform transfer learning in different layers of the model and compare the optimal transfer part.

In the process of transfer learning, the bottom part of the model is usually kept intact to reduce the risk of model over-

fitting. At the same time, for the model in this article, the corpus is first converted into the same language through machine translation. In this process, there is usually no major change in part of speech and word meaning. Due to the differences in culture and writing habits, the migration part is closer to semantics, context, and syntax. At the same time, it meets the requirements of coping with the high-level migration of the network model. The detailed information of each layer of the model PQE_MT_USA is shown in Table 5.

According to the structural characteristics of the network model, the experiment gradually freezes at the bottom of the network, compares the results of different parts of the migration, and obtains the final-migration part. Since only 308 Chinese patents have American patents, they have quality ratings. Take the English data migration results as an example to show and use a total of 308 Chinese patent samples to translate into English text and then perform model migration. Use 100 of them as a test set to compare the effects of different models. Actually, the Chinese patent data used for migration are only 208 articles. The experimental results, respectively, enumerate the microaverage and macroaverages of the accuracy, recall, and F1 value on the training set and the test set. The results are shown in Table 6.

It can be seen from the experimental results that as the low-level parameters of the network gradually freeze, the indicators of the training set show a downward trend, and the indicators of the test set, especially the macroaverage, have a process of first increasing and then decreasing. The main reason for this phenomenon is that the new data set

TABLE 5: PQE-MT layer information.

Number of layers	Type	Output shape	Number of parameters
0	Text input layer	(None, 200)	0
1	Embedding layer	(None, 200, 768)	17999616
2	Bidirectional LSTM_1	(None, 200, 160)	543360
3	Bidirectional LSTM_2	(None, 200, 160)	154240
4	Attention	(None, 160)	32400
5	Number input layer	(None, 132)	0
6	Concatenate layer	(None, 292)	0
7/9/11	Dense layer_1/2/3	(None, 512/128/32)	150016
8/10/12	Dropout layer_1/2/3	(None, 512/128/32)	65664
13	CRF layer	(None, 200, 4)	4128
14	Softmax layer	(None, 8)	668

TABLE 6: Final results of different migration parts.

Migration layer	Training set (microavg/macroavg)						Test set (microavg/macroavg)					
	Precision		Recall		F1-score		Precision		Recall		F1-score	
0-14	0.85	0.70	0.85	0.68	0.85	0.69	0.61	0.38	0.60	0.44	0.60	0.41
2-14	0.83	0.70	0.83	0.63	0.83	0.66	0.66	0.43	0.65	0.46	0.65	0.44
3-14	0.82	0.66	0.82	0.65	0.82	0.65	0.66	0.61	0.66	0.54	0.66	0.57
7-14	0.78	0.66	0.77	0.60	0.77	0.63	0.60	0.44	0.60	0.38	0.60	0.41
9-14	0.60	0.43	0.60	0.39	0.60	0.41	0.58	0.44	0.58	0.36	0.58	0.40
11-14	0.53	0.37	0.53	0.34	0.53	0.35	0.52	0.42	0.52	0.35	0.52	0.38

is small. The larger the migration part of the network during the migration process, the better the fitting effect in the training set, but the problem of overfitting occurs. In the end, the experiment freezes the training of the parameters of the 0-2 layers and selects the 3-14 layers of the network for migration. According to the parameter items in Table 6, freezing the underlying model training can reduce a large number of parameter updates, which improves the model effect and reduces the migration time.

In addition, only the English text of the Chinese patent has experimented with training data, and the patent of model migration was not used. Results of the two experiments are shown in Tables 7 and 8.

From the comparison of Tables 7 and 8, it can be verified that the transfer learning has a certain degree of improvement in each indicator of the model. Due to the lack of training samples in normal supervised learning, the entire TEM model has two obvious problems: first, model TEM poor forecast results in categories with a small number of samples. The results of the accuracy, collection, and f value of the main measurement indicators in category 7 and category 8 were all 0; second, the capacity to generalize the model is very poor, and the problem of overadaptation is obvious. The test set is about 35% lower than the training set in both micro and macroaverages. Comparing results after forwarding learning results in both problems improved. For a few categories of samples, the model has better adaptation, the gap between the microaverage indicators is reduced to 17% on the training set and the test set, and the three indicators of the macroaverage, respectively, reduce to 10%, 16%, and

15%. These two problems have been solved to some extent thanks to the model having learned how the relevant characteristics affect the results of a large number of US patents. Although the distribution of characteristics is different, the overall trend is approximately the same. The model adapts even more to changes in several Chinese patent indicators based on the trend initially learned. Although the model has improved, the effect still cannot meet the final requirements for patent assessment. Hope to improve the model effect by increasing migration data.

We randomly selected 200 unlabeled Chinese patents and manually labeled them as data_random and used the active learning algorithm to obtain the top 200 Chinese patents with the largest classification margin and labeled them as data_active_learning. The experiment starts from 0 and increments each time by 20 on these two data sets, until it reaches 200. Compare the effects of randomly selected data and active learning to select data on model migration. The comparison of the prediction accuracy of the model in the two cases is shown in Figure 12.

By comparing the two trends, it has been found that using the active learning method to select data is better than randomly selected data. The effect of the model that used active learning to select about 140 data to be labeled is equivalent to the effect of randomly selecting 200 data labels. Then, the experiment uses active learning to further expand the labeled data, and active learning can get the most informative data. And each round obtained the first 200 data with the largest classification margin for manual labeling. A total of 2000 samples labeled as new migration data were

TABLE 7: Results of training the network using only Chinese patent data.

Category	Training set				Test set			
	Precision	Recall	F1-score	Support	Precision	Recall	F1-score	Support
1	0.92	1.00	0.96	61	0.55	0.66	0.60	35
2	0.79	0.70	0.74	33	0.41	0.39	0.40	18
3	0.80	0.82	0.81	49	0.18	0.19	0.19	21
4	0.61	0.69	0.65	29	0.30	0.30	0.30	10
5	0.42	0.65	0.51	17	0.33	0.38	0.35	8
6	1.00	0.80	0.89	5	0.00	0.00	0.00	3
7	0.00	0.00	0.00	9	0.00	0.00	0.00	3
8	0.00	0.00	0.00	5	0.00	0.00	0.00	2
Microavg	0.76	0.76	0.76	208	0.40	0.40	0.40	100
Macroavg	0.57	0.58	0.57	208	0.22	0.24	0.23	100

TABLE 8: Data transfer learning results of the initial 208 Chinese patent English.

Category	Training set				Test set			
	Precision	Recall	F1-score	Support	Precision	Recall	F1-score	Support
1	1.00	0.92	0.96	61	1.00	0.77	0.87	35
2	0.83	0.91	0.87	33	0.54	0.78	0.64	18
3	0.82	0.92	0.87	49	0.78	0.67	0.72	21
4	0.81	0.72	0.76	29	0.20	0.20	0.20	10
5	0.50	0.76	0.60	17	0.35	0.75	0.48	8
6	0.40	0.40	0.40	5	0.33	0.33	0.33	3
7	0.60	0.33	0.43	9	0.67	0.33	0.44	3
8	0.75	0.60	0.67	5	1.00	0.50	0.67	2
Microavg	0.83	0.83	0.83	208	0.66	0.66	0.66	100
Macroavg	0.71	0.70	0.70	208	0.61	0.54	0.57	100

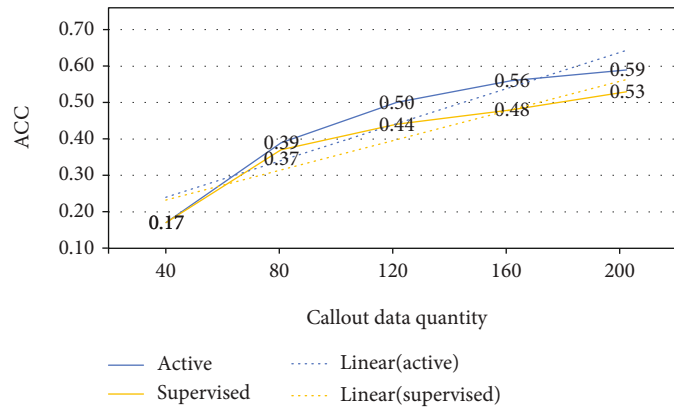


FIGURE 12: Comparison of the accuracy of transfer learning between active learning and randomly selected labeled data.

obtained, the model uses 2-14 levels with the best migration effect obtained from the experimental comparison as a migration part, and the result of the migration of 2208 Chinese patent data obtained after active learning is shown in Table 9. Furthermore, the experiment was conducted in the same way using the patented English language, and the results are given in Table 10.

Experimental results in Tables 8 and 9 are compared to observe the effect of enabled learning on the effect of the data expansion model. Using the initial data migration model, micro-F1 values in the training set and in the test set were 83% and 66%, and macro-F1 values were 70 and 57%. After active learning for data expansion, the values of micro-F1 were 79% and 74% and the macro-F1 values were 68% and

TABLE 9: Data transfer learning results of 2208 Chinese patent English text data after active learning expansion.

Category	Training set				Test set			
	Precision	Recall	F1-score	Support	Precision	Recall	F1-score	Support
1	0.93	0.96	0.95	676	0.93	0.83	0.88	35
2	0.85	0.71	0.77	404	0.88	0.78	0.82	18
3	0.79	0.85	0.81	506	0.74	0.81	0.77	21
4	0.59	0.77	0.67	258	0.39	0.70	0.50	10
5	0.61	0.64	0.62	154	0.50	0.38	0.43	8
6	0.41	0.54	0.47	69	0.33	0.33	0.33	3
7	0.63	0.54	0.58	87	0.33	0.67	0.45	3
8	0.78	0.37	0.50	54	1.00	0.50	0.67	2
Microavg	0.79	0.80	0.79	2208	0.74	0.74	0.74	100
Macroavg	0.70	0.67	0.68	2208	0.64	0.63	0.63	100

TABLE 10: Data transfer learning results of 2208 Chinese patent Chinese text data after active learning expansion.

Category	Training set				Test set			
	Precision	Recall	F1-score	Support	Precision	Recall	F1-score	Support
1	0.95	0.89	0.92	676	0.87	0.79	0.83	35
2	0.68	0.68	0.68	404	0.51	0.54	0.53	18
3	0.65	0.79	0.71	506	0.52	0.64	0.57	21
4	0.55	0.52	0.53	258	0.45	0.42	0.43	10
5	0.45	0.28	0.34	154	0.36	0.22	0.27	8
6	0.43	0.36	0.39	69	0.33	0.67	0.45	3
7	0.56	0.71	0.48	87	0.33	0.33	0.33	3
8	0.70	0.65	0.67	54	1.00	0.50	0.67	2
Microavg	0.73	0.71	0.72	2208	0.62	0.61	0.61	100
Macroavg	0.61	0.61	0.61	2208	0.55	0.51	0.53	100

63%. Due to the increase in the amount of migrated data, the ability to generalize the model became stronger, and the effect of the FOI model improved to a certain extent by providing the test set data.

The results of Table 9 were compared with Table 10, and the effects of Chinese and English were observed on the model effect. The model transition effect for English texts is generally better than Chinese text. The authors believe that the use of machine translated Chinese data leads to degradation of the initial model effect in the case of using source models trained by U.S patents due to the limitations of machine translation techniques, resulting in the final effect of the transfer model is reduced; the number of Chinese patent translations used for the target model has a lower overall influence factor; in addition, the differences in the Chinese and English writing specifications and feature spaces also cause difference in model learning effects.

4.3.3. Experimental Results and Analysis of Patent Quality Evaluation Model Based on Multitask Learning. The article uses SVM as a classification model and Chinese patent text as an example to compare and demonstrate the use of TF-IDF feature vectors and the bag-of-words method for feature word extraction using mutual information, information gain, and chi-square, as well as the three

TABLE 11: Forecast accuracy of different algorithms.

Method	Threshold	Accuracy (train)	Accuracy (test)
TF-IDF	N/A	0.68	0.60
MI	0.1	0.59	0.53
IG	0.03	0.66	0.61
CHI	3	0.67	0.64
Union	N/A	0.71	0.62

methods to extract feature words take the result of the union. The SVM classification algorithm based on bag-of-words is very critical in the selection of feature words. The experiment uses feature extraction formulas for different types of patents to calculate, by manually setting a threshold as a reference value, and extracting words higher than this value as the input features of the bag of words [38]. Table 11 shows the accuracy of algorithm prediction in different situations.

The experimental results show that the SVM classification algorithm based on TF-IDF achieves 68% accuracy on the training set, but in contrast, the prediction result on the test set is reduced by nearly 10%, showing a trend of overfitting. Although the bag-of-words method has a lower accuracy in the training set compared with the TF-IDF method, the accuracy in the test set has been greatly improved. Due to the

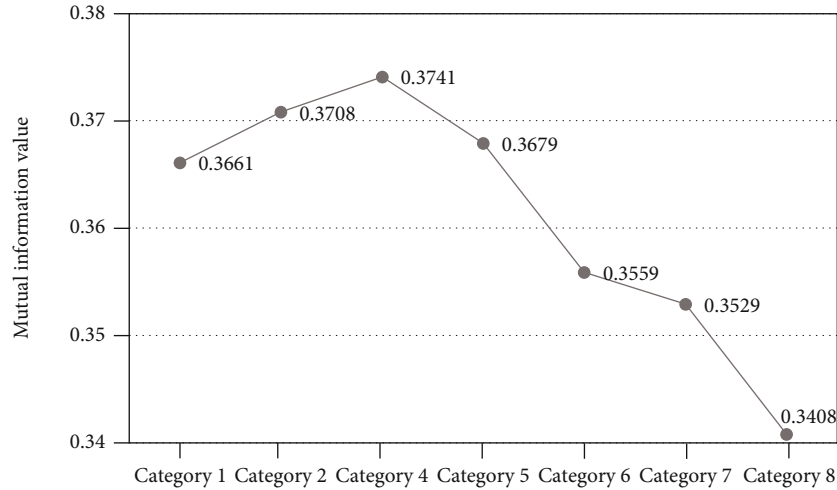


FIGURE 13: Mutual information value between category 3 and other categories.

higher dimension of the union set, it is easy to cause overfitting, which reduces the accuracy of the test set. The experiment finally uses the chi-square algorithm for feature extraction. In addition, the experiment uses the mutual information value that can calculate the similarity between the categories to test different categories. The mutual information value of category 3 and other categories is selected as the display. The test results are shown in Figure 13.

It can be observed that the closer the distance between the samples, the greater the mutual information value. This result fully demonstrates the effectiveness of the training set data annotation. Experiments with different numbers of feature words, the number of feature words with the best model effect is obtained, and the result is shown in Figure 14. The figure shows the dimensions of feature words, and the input of the model is a vector of word bag features and quantitative indicators.

A comparative analysis of the experimental results shows that the accuracy of the model in the training set increases as the dimension of the feature word increases. When the dimension of the feature word is too high, overfitting problems will occur. The effect of the model on the test set gradually decreases. Use the 400-dimensional bag-of-words vector with the best comprehensive results to further test different classification algorithms. The results are shown in Table 12.

It can be seen that the classification accuracy of Bayesian and kNN algorithms is poor, and the decision tree algorithm is not suitable for patent quality evaluation classification due to the overfitting problem caused by the complexity of the sample. Finally, SVM is selected as the classification method for classification.

In the experiment, CHI with better results is selected as the feature word extraction method. Extract 400-dimensional bag-of-words vectors, perform SVM classification on the text, perform collaborative training with the migrated deep learning model PQE_MT_CHN at the same time, and finally compare the prediction effects of the two methods.

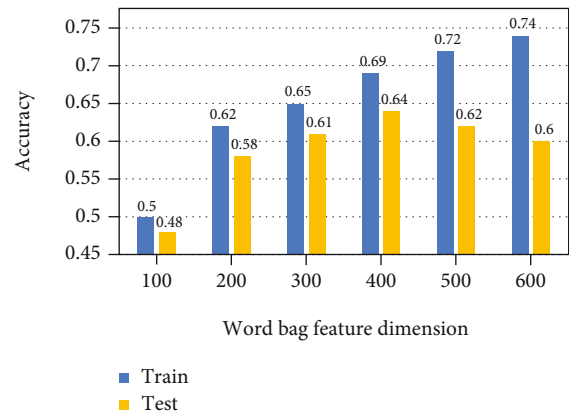


FIGURE 14: SVM classification results of different dimensional bag-of-words vectors.

TABLE 12: CHI extracts the classification results of 400-dimensional bag-of-words vectors combined with quantitative indicators in different algorithms.

Model	Train			Test		
	Precision	Recall	F1	Precision	Recall	F1
Naive Bayes	0.66	0.65	0.65	0.62	0.60	0.61
kNN	0.62	0.58	0.60	0.56	0.54	0.55
SVM	0.70	0.69	0.69	0.64	0.63	0.63
Decision tree	0.74	0.71	0.72	0.59	0.57	0.58

In the experiment, the proposed collaborative training method was compared with the following benchmark methods:

- (i) *SVM_Ch*. Use the combination of Chinese features and quantitative indicators as input, and use SVM for classification. There is no need for text translation in the process, only the marked data is used
- (ii) *SVM_En*. Use the combination of English features and quantitative indicators as input and use SVM for classification. The process requires a Chinese-

TABLE 13: Classification results of PQE_MT_CHN_Co and other benchmark models.

Model	Train			Test		
	Precision	Recall	F1	Precision	Recall	F1
SVM_Ch	0.70	0.69	0.69	0.64	0.63	0.63
SVM_En	0.66	0.61	0.63	0.60	0.57	0.58
SVM_Ch + En	0.75	0.67	0.71	0.62	0.60	0.61
SVM_Co	0.75	0.70	0.72	0.68	0.67	0.67
PQE_MT_CHN_Ch	0.73	0.71	0.72	0.62	0.61	0.61
PQE_MT_CHN_En	0.79	0.80	0.79	0.74	0.74	0.74
PQE_MT_CHN_Ch + En	0.81	0.79	0.80	0.67	0.66	0.66
PQE_MT_CHN_Co	0.84	0.82	0.83	0.78	0.77	0.77

English translation, and only the marked data is used

- (iii) *SVM_Ch + En*. Use the combination of Chinese and English bilingual features and quantitative indicators as input and use SVM for classification. The process requires a Chinese-English translation, and only the marked data is used
- (iv) *SVM_Co*. Use SVM_Ch and SVM_En for collaborative training, and sample prediction is obtained by weighting the results of the two models. The process requires translation from Chinese to English, using unlabeled data for collaborative training
- (v) *PQE_MT_CHN_Ch*. Use the combination of Chinese features and quantitative indicators as input, and use Chinese text transfer learning to obtain the model PQE_MT_CHN for classification. There is no need for text translation in the process, only the marked data is used
- (vi) *PQE_MT_CHN_En*. Use the combination of English features and quantitative indicators as input and use English text transfer learning to obtain a model for classification. The process requires a Chinese-English translation, and only the marked data is used
- (vii) *PQE_MT_CHN_Ch + En*. Use the combination of Chinese and English bilingual features and quantitative indicators as input and use the Chinese and English bilingual text transfer learning model for classification. The process requires a Chinese-English translation, and only the marked data is used
- (viii) *PQE_MT_CHN_Co*. Use PQE_MT_CHN_Ch and PQE_MT_CHN_En for collaborative training, and the sample prediction is obtained by weighting the results of the two models. The process requires Chinese-English translation, using unlabeled data for collaborative training

The results of the above model are shown in Table 13. In addition, the experiment compares the number of samples

of each category and optimizes the experimental parameters for the iteration rounds of collaborative training and each iteration process. The specific results of the model are shown in Figures 15 and 16. The following text will analyze and summarize the results of this part.

Experimental results show

- (1) From the comparison of the various data in Table 13, it can be concluded that the Chinese patent quality evaluation model constructed by the experiment using transfer learning is better than the best-performing SVM in traditional machine learning in the overall transfer effect on the small training set. It proves the feasibility of using transfer learning in this field
- (2) After combining the Chinese and English features of the two models, the prediction effect of the training set has been improved, but due to the lack of training corpus, certain overfitting problems have occurred, resulting in the model's effect on the test set decline. It further proves the necessity of using collaborative training instead of splicing Chinese and English directly
- (3) The proportion of each type of data selected in each iteration of collaborative training has a greater impact on the improvement of the training effect. It can be observed from Figure 15 that when the data proportions of each category are the same, the sample space of model learning will change, leading to the reduction of model prediction effect, and the model is the most stable when the data with the same proportion of each category in the initial data set is selected
- (4) After using the collaborative training algorithm, the semisupervised training method further optimizes the fitting effect of the model on the small training set, and all the evaluation indicators of the model are further improved. The PQE_MT_CHN_Co proposed in the article reaches the highest level in all indicators. Figure 14 can analyze that collaborative training can improve the effect of both Chinese and English models at the same time. Initially, the effect

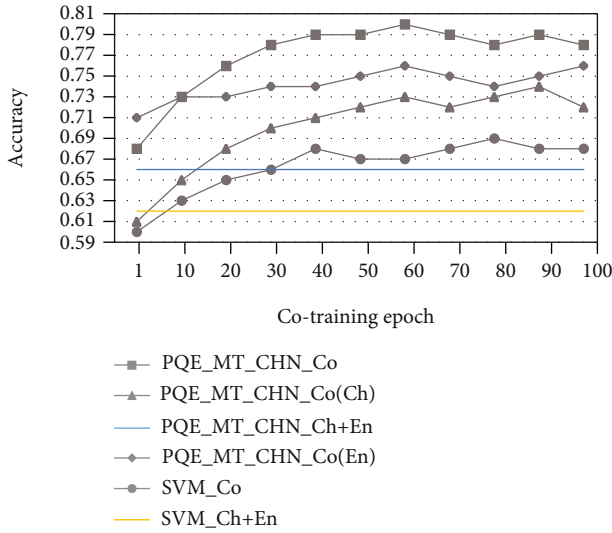


FIGURE 15: Cotraining iterative process model effect changes.

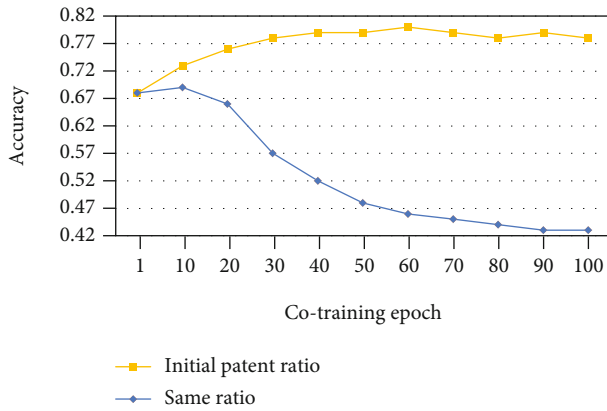


FIGURE 16: The impact of the proportion of data obtained in each category on cotraining.

of the two models is quite different, resulting in a lower effect of the model's joint prediction. However, after multiple rounds of iterations, the prediction effects of the two models gradually approached, and the joint prediction effect of the two models is better than that of one of the models. The micro- F_1 value of the final model reached 83% on the training set and 77% on the test set. It shows that the method proposed in the article has improved the effect of quality evaluation to a certain extent

5. Conclusion and Future Work

Patent quality evaluation model-PQE_MT is proposed in this paper by quantifying the patent indexes from different dimensions and taking the long text part of the patent as one kind of index. The proposed model combines quantitative index model and sequence model based on multitask learning to predict the quality rating of patent. The advantage of this model is that it combines the numerical attribute

part with the text part of the patent as the evaluation indexes. Therefore, the feature extracted from the patent is more comprehensive. In addition, multitask learning is added to train two tasks together, so that the process of updating parameters by backpropagation can promote each other and improve the effectiveness of the model. Compared with the baseline model, the PQE_MT model improves the accuracy of English patent quality assessment.

The migration process mainly includes three parts: the selection of migrant parts, the cross-sectoral transition, and the use of active learning to enlarge data. The advantage of this model is to select the data labeled from the maximum information quantity using the transition learning technology and the positive learning and to reduce the time cost of manual annotation. In the experimental process, the prediction effect of the model gradually improved, and finally, the Chinese patent quality evaluation model achieved good accuracy. At the end of the experiment, we compared the effects of transmission learning between Chinese and English and found that the results were different and the model learned different features. In the follow-up task, we consider two different character models to have different characteristics that may affect each other to improve different results.

In the future, more comprehensive and more meaningful indexes should be quantified. Patent information such as patent drawings, patent research and development cycle, and other internal information of the company are still unused in the experiment, which may improve the predicted results and be more convincing. Moreover, we will try different data expansion methods or semisupervised algorithms to improve the effect of the migration process.

Data Availability

The data used to support the finding of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

Regarding the content of the experiment in the article, we previously published the research results of the active learning part in the EAI conference [39]. Based on the research results, we have carried out more innovations, expansions, and more details on the patent quality evaluation research. The authors also acknowledge the National Natural Science Foundation of China under grant no. 62171043, Natural Science Foundation of Beijing under grant no. 4212020, Defense-related Science and Technology Key Lab Fund project under grant no. 6412006200404, Qin Xin Talents Cultivation Program of Beijing Information Science and Technology University under grant no. QXTCP B201908, and Research Planning of Beijing Municipal Commission of Education under grant no. KM202111232001.

References

- [1] L. Qinghai, L. Yang, S. Wu, and X. Xu, "Overview and AHP of patent value evaluation indexes," *Studies in Science of Science*, vol. 2, pp. 281–286, 2007.
- [2] X. Jin, S. Spangler, Y. Chen et al., "Patent maintenance recommendation with patent information network model," in *2011 IEEE 11th International Conference on Data Mining*, Vancouver, BC, Canada, 2011.
- [3] E. J. Han and S. Y. Sohn, "Patent valuation based on text mining and survival analysis," *Journal of Technology Transfer*, vol. 40, no. 5, pp. 821–839, 2015.
- [4] S. Yang, Y. Wang, C. Hang, Z. Lijuan, L. Baizhi, and D. Lei, "Research on the evaluation of patent legal value," *High Technology News*, vol. 26, no. Z1, pp. 815–823, 2016.
- [5] L. Xionghui and Z. Fuyuan, "Research on the evaluation of invention patent value based on fuzzy comprehensive evaluation method," *Journal of Gannan Normal University*, vol. 38, no. 3, pp. 26–30, 2017.
- [6] L. Dan, "The determination of market dominance in the field of patents—from the perspective of patent value evaluation," *Electronic Intellectual Property*, vol. 5, pp. 21–29, 2018.
- [7] M. Yuying, *Comprehensive Evaluation of Patent Value Based on Rough Set and Cloud Model*, Capital University of Economics and Business, 2018.
- [8] Y. S. Chen and K. C. Chang, "Using neural network to analyze the influence of the patent performance upon the market value of the US pharmaceutical companies," *Scientometrics*, vol. 80, no. 3, pp. 637–655, 2009.
- [9] A. J. C. Trappey, C. V. Trappey, C. Y. Wu, and C. W. Lin, "A patent quality analysis for innovative technology and product development," *Advanced Engineering Informatics*, vol. 26, no. 1, pp. 26–34, 2012.
- [10] J. Wang, Y. Wu, and S. Dashuai, "Research on non-commercial patent value evaluation method based on decision tree model," *Economic Forum*, vol. 10, pp. 131–136, 2012.
- [11] Z. Yunhua, Z. Jing, L. Yan, and Y. Xucheng, "Research on patent value evaluation method based on machine learning," *Information Science*, vol. 31, no. 12, pp. 15–18, 2013.
- [12] J. L. Wu, P. C. Chang, C. Tsao, and C. Y. Fan, "A patent quality analysis and classification system using self-organizing maps with support vector machine," *Applied Soft Computing*, vol. 41, no. C, pp. 305–316, 2016.
- [13] Q. Yihui, Z. Chiyu, and C. Shuixuan, "Research on patent value evaluation index system based on classification regression tree algorithm," *Journal of Xiamen University (Natural Science Edition)*, vol. 56, no. 2, pp. 244–251, 2017.
- [14] H. Lin, *Research on Patent Value Evaluation Method Based on Deep Learning*, University of Science and Technology of China, 2018.
- [15] C. Banea, R. Mihalcea, J. Wiebe, and S. Hassan, "Multilingual subjectivity analysis using machine translation," in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 127–135, Honolulu, Hawaii, 2008.
- [16] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [17] Q. Xu and Q. Yang, "A survey of transfer and multitask learning in bioinformatics," *Journal of Computing Science and Engineering*, vol. 5, no. 3, pp. 257–268, 2011.
- [18] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, p. 9, 2016.
- [19] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," 2014, <https://arxiv.org/abs/1411.1792>.
- [20] C. Tohompson, M. E. Califf, and R. Mooney, "Active learning for natural language parsing and information extraction," in *Proceedings of the Sixteenth International Conference on Machine Learning*, pp. 406–414, San Francisco: Morgan Kaufmann, 1999.
- [21] S. Tong and D. Koller, "Support vect or machine active learning with applications to text classification," *Journal of Machine Learning Research*, vol. 2, pp. 45–66, 2001.
- [22] B. Settles, *Active Learning Literature Survey*, University of Wisconsin-Madison Department of Computer Sciences, 2009.
- [23] L. Jielong, X. Yanshan, H. Zhifeng, R. Yibang, and Z. Liyang, "Multi-instance multi-label active learning based on SVM," *Computer Engineering and Design*, vol. 37, no. 1, pp. 254–258, 2016.
- [24] Z. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway, and J. Liang, "Fine-tuning convolutional neural networks for biomedical image analysis: actively and incrementally," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7340–7351, Honolulu, HI, USA, 2017.
- [25] J. J. Zhu and J. Bento, "Generative adversarial active learning," 2017, <https://arxiv.org/abs/1702.07956>.
- [26] K. Konyushkova, R. Sznitman, and P. Fua, "Learning active learning from data," 2017, <https://arxiv.org/abs/1703.03365>.
- [27] A. Blum and T. M. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the eleventh annual conference on Computational learning theory*, pp. 92–100, the University of Wisconsin, Madison, Wisconsin, 1998.
- [28] X. Wan, "Co-training for cross-lingual sentiment classification," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pp. 235–243, Beijing, China, 2009.
- [29] G. U. O. Tao, L. I. Guiyang, and L. A. N. Xia, "Graph-based semi-supervised cooperative training algorithm," *Computer Engineering and Design*, vol. 33, no. 9, pp. 3584–3587 +3621, 2012.
- [30] S. Qiao, W. Shen, Z. Zhang, B. Wang, and A. Yuille, "Deep co-training for semi-supervised image recognition," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 135–152, Munich, 2018.
- [31] Y.-l. Gong and J. Lv, "Cooperative training algorithm combining active learning and density peak clustering," *Computer Applications*, vol. 39, no. 8, pp. 2297–2301, 2019.
- [32] J.-y. Huang and D.-y. Huang, "Improving the translation performance of neural machine translation system by cooperative training," *Journal of Xiamen University (Natural Science Edition)*, vol. 58, no. 2, pp. 176–183, 2019.
- [33] H. Gao, K. Xu, M. Cao, J. Xiao, Q. Xu, and Y. Yin, "The deep features and attention mechanism based method to dish healthcare under social IoT systems: an empirical study with a hand-deep local-global net," *IEEE Transactions on Computational Social Systems (TCSS)*, vol. 2021, no. 8, pp. 1–12, 2021.
- [34] H. Gao, X. Qin, R. J. D. Barroso et al., "Collaborative learning-based industrial IoT API recommendation for software-defined devices: the implicit knowledge discovery perspective,"

- IEEE Transactions on Emerging Topics in Computational Intelligence (TETCI)*, vol. 2020, no. 9, pp. 1–11, 2020.
- [35] Y. Huang, H. Xu, H. Gao, X. Ma, and W. Hussain, “SSUR: an approach to optimizing virtual machine allocation strategy based on user requirements for cloud data center,” *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 2, pp. 670–681, 2021.
- [36] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *European Conference on Computer Vision*, Springer, Cham, 2014.
- [37] J. Xiao, H. Xu, H. Gao, M. Bian, and Y. Li, “A weakly supervised semantic segmentation network by aggregating seed cues: the multi-object proposal generation perspective,” *ACM Transactions on Multimedia Computing Communications and Applications*, vol. 17, no. 1s, pp. 1–19, 2021.
- [38] Z. Wang, “A study on the subject classification of the NBA match reports,” in *Proceedings of 2018 2nd International Conference on Electronic Information Technology and Computer Engineering (EITCE 2018)*, pp. 104–121, Shanghai University of Engineering Science, China, 2018.
- [39] J. Liu, X. You, Z. Wang, and X. Lv, “Cross-language transferring the patent quality evaluation model based on active learning data extension,” in *International Conference on Testbeds and Research Infrastructures: Tools for Design, Implementation and Verification of Emerging Information Technologies*, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer, Cham, 2020.