Cross-language Transfering the Patent Quality Evaluation Model Based on Active Learning Data Extension

Jiaqi Liu, Xindong You, Zhe Wang, and Xueqiang Lv(✉)

Beijing Key Laboratory of Internet Culture Digital Dissemination, Beijing Information Science and Technology University, Beijing, China
lxq@bistu.edu.cn

Abstract. At present, China has become a major patent production country, and the number of patent applications has been ranked first in the world for many years. As the number of patents has increased, the quality of patents has begun to draw people's attention. At present, there is no clear evaluation method for Chinese patents. Manual evaluation of patents requires a large number of relevant experts to research and compare patents in different fields, which is time-consuming and labor-intensive. In the previous study, the author constructed an English patent quality evaluation model PQE-MT using U.S. Patents that represent patent strength. This paper introduces this model into Chinese patents through transfer learning and active learning, thereby reducing the workload of manual labeling. The evaluation results show that the method in the experiment has achieved a good migration effect, with Micro-F1 reaching 74%.

Keywords: Patent quality assessment · Transfer learning · Active learning · Multi-task learning · Cross-language text classification

1 Introduction

In recent years, Sino-US trade disputes and US sanctions against Huawei have all reflected the importance of technological innovation and intellectual property rights. Traditional manufacturing industries need rapid transformation, which also reflects the importance of invention patents. China's R&D spending rose from 0.9% of GDP in 2000 to 2.22% in 2018. The total investment is nearly 2 trillion yuan [1]. It can be seen that the country has made great efforts to develop technology and actively innovate. In recent years, the number of entity patent applications in China has increased rapidly. In 2018, the number of patents applied by Chinese applicants was 1.542 million, up 11.6% year on year [2]. Under the phenomenon that the number of patent applications is leading the world, it does not mean that China's technological innovation is already at the top of the world. There are still some quality problems in China's patents, such as low application rate and lack of patents with important strategic significance and substantial improvement. It can be seen that China still lags behind other developed countries in terms of

Y. Weng et al. (Eds.): TridentCom 2020, LNICST 380, pp. 120–133, 2021.
https://doi.org/10.1007/978-3-030-77428-8_10

innovation, and the increase in the number of patent applications masks the fact that the quality of patents is low.

Based on these problems, putting high-quality patents in a strategic position has become an urgent task. Improving the quality of Chinese patents is of great significance to China's development. For enterprises, it is helpful to help enterprises understand the development trend of industry technology and choose the development direction in a targeted way [3]. From the perspective of the government, it is conducive to the government agencies to efficiently analyze the trend of science and technology, optimize the investment in science and technology, and formulate targeted policies for the development of science and technology. At the same time, it is helpful for scientific research institutions to analyze scientific and technological trends, grasp scientific and technological trends, track scientific and technological hot spots, and clarify technological directions [4]. Besides, the improvement of China's patent quality helps investors and patent inventors to analyze the direction of emerging technologies, quickly find better investment and development goals, and reduce potential legal risks.

Using data mining and natural language processing technologies, effective indicators are extracted from a large number of patent information data of different dimensions, and the US patent quality evaluation model PQE-MT is constructed. On this basis, using transfer learning and active learning methods, the index distribution and writing differences of Chinese patents have been improved.

2 Related Work

Due to the time-consuming and labor-intensive work of data labeling and the scarcity of high-quality labeling data, transfer learning has attracted more and more attention from the academic community. Banea [5] proposed a method of cross-language transfer learning due to the abundant data of English tagging. In this paper, the tagged corpus of English data is used to generate the source domain data set of the target language through machine translation, which proves the feasibility of using machine translation for cross-language transfer learning. Pan [6] proposed to transfer the domain with sufficient training data to the domain with similar data distribution, to greatly improve the learning effect by avoiding expensive data labeling work through knowledge transfer, and discussed the relationship between transfer learning and related machine learning technologies such as domain adaptation and sample selection deviation. Xu [7] uses migration learning and multitask learning to extract and transfer useful knowledge from the data in the auxiliary domain, thus helping to solve the problem of insufficient data in the target domain. The latest progress in the field of biological information is introduced. Weiss [8] explains transfer learning and information about solutions and discusses possible future research work. Where the migration learning solution is independent of data size. Yosinski [9] studied the mobility of deep neural networks. Fine-tune experiments are carried out layer by layer on different layers to explore the mobility of the network. It is pointed out that adding fine-tune to the deep migration network will greatly improve the effectiveness and better overcome the differences between data. The migration of network layers can accelerate the learning and optimization of the network. The bottom network learns the common characteristics and the top network learns the domain characteristics.

Transfer learning still needs to label a small amount of corpus for the model to adapt to the distribution of data features in the target domain. Active learning can predict the results through the model and select the samples that contribute the most to the improvement of the model. Tohompson [10] uses the method of active learning to try to select the example with the largest amount of information as the training data for transfer learning. The experimental results show that active learning can significantly reduce the number of labeled samples when the algorithm achieves the same effect. Tong [11] uses pool-based active learning. The algorithm does not need a randomly selected training set and can mark the requested samples. A new algorithm for active learning using a support vector machine provides a theoretical basis for the algorithm by using the concept of version space. The experimental results show that the active learning method in this paper can significantly reduce the demand for labeled samples. Settles [12] introduces active learning and reviews relevant literatures. Machine learning algorithms can achieve higher accuracy through fewer labeling training examples. The query scheme is discussed, and the experience and theoretical evidence of active learning are analyzed. Li Jielong [13] takes the minimum classification distance of SVM as the confidence level of selecting examples and proposes multi-example multi-label active learning based on the minimum classification interval of SVM, which effectively reduces the amount of sample labeling and improves the classification performance. Zhou [14] uses active learning and transfer learning, data expansion, majority selection, continuous fine-tuning, and other methods to verify the data set in the field of medical images, and points out that the introduction of active learning can reduce the amount of data annotation by at least half. Zhu [15] combines GAN with active learning for the first time, obtains the generator model by training GAN, and obtains the most valuable samples through active learning for experts to the label. Konyushkova [16] solves the problem of the insufficient generalization ability of traditional selection strategies. By transforming active learning into regression problems for learning, the experimental results have achieved good results on data sets in many different fields.

3 Cross-language Transfering Patent Quality Evaluation Model Based on Active Learning Data Extension

3.1 Prediction Model of Chinese Patent Quality Grade

Based on the PQE-MT model, this paper uses transfer learning and active learning methods to further propose a cross-language transfer patent quality evaluation model based on active learning data expansion, and transfer the original model to Chinese patents. The following article will introduce the model from these two parts respectively.

3.2 PQE-MT Model Structure and Transfer Learning

The PQE-MT model consists of two parts: a quantitative index model and a multi-task learning model. Multi-task learning includes text classification tasks and named entity recognition tasks. The following describes the specific content and functions of each part of the network model. The overall model structure is shown in Fig. 1.

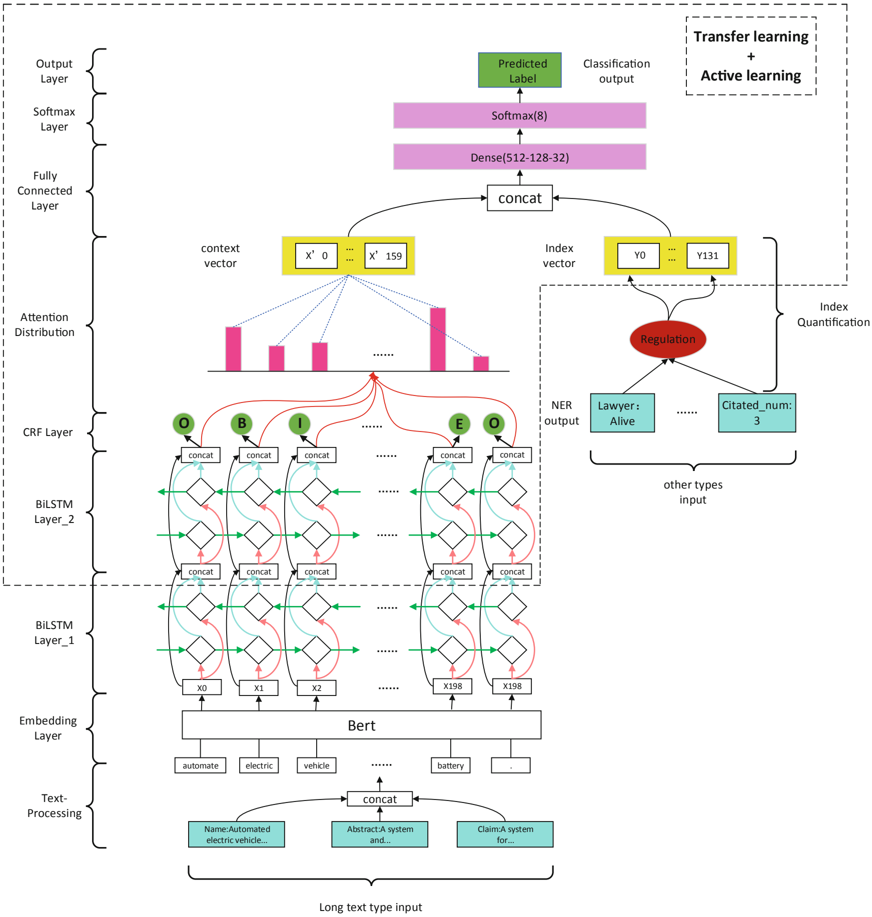

Fig. 1. PQE-MT model structure and migration

- Input: Initial Patent Attribute + Patent Title, Abstract, Claims.
- Index quantification: 132 indexes, composed of 15 initial indexes and 117 multidimensional quantified indexes.
- Text processing: The patent title, abstract, and sovereign claim are spliced together, and stop words and special symbols are removed. Convert all characters to lowercase letters and roots for English text; Word segmentation is carried out for Chinese texts.
- Word embedding: BERT [17] is used as the embedding layer of the model in the experiment. BERT is a pre-training model, which trains the language model on 3.3 billion text corpus and makes fine adjustments on different downstream tasks. The model has achieved the best results so far on different text tasks. BERT is a combination of Transformer. Transformer has made model innovations in multi-head attention mechanism, self-attention mechanism, and position coding. BERT enhances the semantic expression of sentences by learning contextual relationships in the large-scale corpus.

In the experiment, the word vector fine-tuned by BERT on the corpus in the field of new energy vehicles is used as the embedding of words.

- Bidirectional LSTM: After the embedding layer, the experiment uses multitask learning as part of the sequence model. Multitask learning includes two tasks: bi-directional LSTM, text classification task composed of attention mechanism, CRF, and named entity recognition task. The long-term memory network [18] can learn the long-term dependency of texts. The structure uses storage units to record historical information, thus ensuring the integrity of the information. The concept of the control gate is introduced, and the information flow of the model is controlled by the update gate, forgetting gate, and output gate. Based on LSTM, bidirectional LSTM takes into account the past and future timing features of the sequence, and effectively uses context information to mine more hidden features.
- Attention mechanism: When the input sequence is very long, LSTM is difficult to obtain a reasonable vector. Note that the machine retains the intermediate vectors of the LSTM encoder, and then trains a structure to selectively learn these inputs and associate the output sequence with them so that the model can focus on some words considered important in the input sequence.
- CRF: In the Named Entity Recognition task, the results of the bidirectional LSTM are entered into the CRF. CRF improves the accuracy of identification through the dependency relationship between tags. The words in the sentence correspond to each time node in the sequence. The output of the model is a series of tags, and the tag space is defined as $\{B, I, E, O\}$. "B" denotes the beginning of the domain term, "I" denotes the middle part of the domain term, "E" denotes the end of the domain term, and "O" denotes that the element does not belong to the domain term. According to the tag sequence output by the model, the domain words in the sequence can be determined.
- Fully connected layer: input a fully connected neural network composed of 512, 128, and 32 nodes.
- SoftMax Layer: The structure of the full connection layer is finally input into SoftMax Layer for multi-classification.
- Output: Eight types of patent quality grades.

The experiment uses US patents as the source domain and Chinese patents as the target domain. Due to the problem of language barrier between different languages, it is impossible to directly transfer between different languages. This involves the problem of cross-language text classification. At present, the mainstream solutions include methods based on language knowledge base [19], methods based on multilingual models [20] and methods based on machine translation technology [21]. The main research contents of this paper are as follows: 1. It is difficult to build a complete and accurate language knowledge base, which needs to be updated for different fields; 2. The effect of the existing machine translation has been greatly improved, and the experiment itself does not need very accurate translation, only the translation of domain words needs to be as accurate as possible; 3. The difference of feature space between different languages can help the model learn different features. In the experiment, machine translation was used to translate the text parts of Chinese and English patents, and the patent grades of some Chinese patents were marked. Through transfer learning, the model was adapted to the index distribution and language features of Chinese patents [22].

Through comparative experiments, the parameters of different layers are frozen to obtain the best offset effect. The specific experimental process and results will be explained in detail in the experimental results and analysis in the fourth part. Finally, the part shown by the dotted line in Fig. 1 is selected for the migration of model parameters.

3.3 Efficient Transfer Using Active Learning

Figure 2 is the flow chart of active learning.

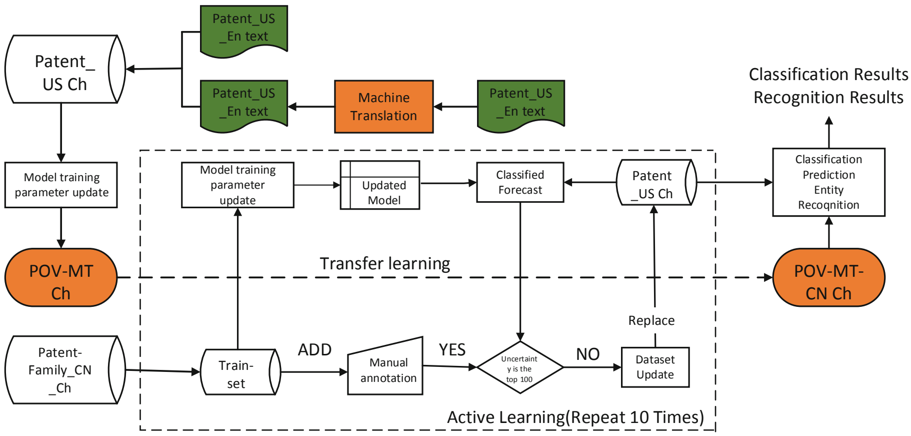

Fig. 2. Overall process diagram of transfer learning and active learning

4 Experimental Results and Analysis

4.1 Experimental Data

The experiment used 59,147 U.S. Patents, of which the number of patents with different ratings is shown in Fig. 3. 90% of the samples (53475) are selected as training data and 10% of the samples (5942) are selected as test sets for testing. The verification method selects the K-fold cross-validation for model optimization. K-1/K of the training data is selected as the training set and the rest as the verification set. As a result, the average value of K times is obtained, and the value of K is set to 10.

There are 21,611 Chinese patents, 308 of which exist in the U.S. Patent data set, i.e. 308 patents are applied in two countries at the same time, forming a patent family and belonging to the same family. In the experiment, its score in the US patent corresponds to the Chinese patent, so as to obtain the quality level corresponding to the Chinese patent. In this paper, 100 Chinese patents with the patent rating are obtained as the test set of the final model, another 208 are used as the initial migration data, and the remaining 21303 unlabeled Chinese patents are used as the data pool for active learning to select the data to be labeled.

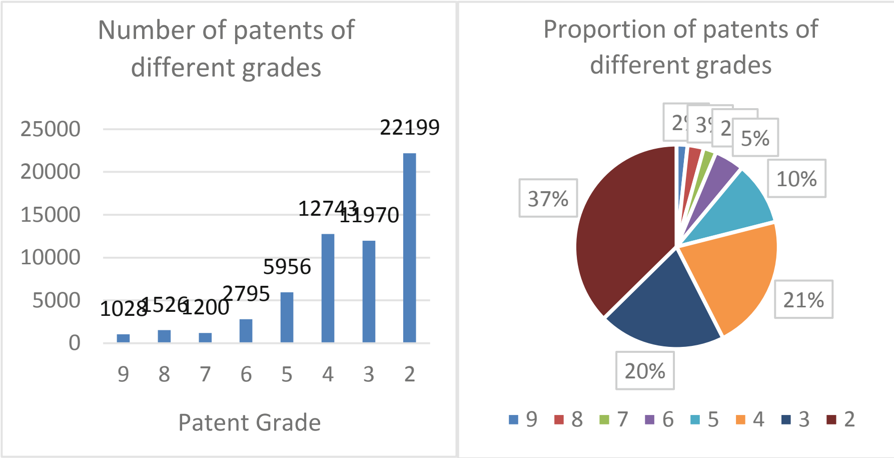

Fig. 3. Number and proportion of patents with different ratings

All the data in the experiment are randomly scrambled and the extraction method is random, to ensure the uniform distribution of different label samples and the accuracy and fairness of the experimental results. The machine translation part involved in this article uses Google Translation, which is currently the most accurate translation system.

4.2 Evaluation Indicators

The accuracy, recall, F-score, accuracy, Micro-averaging, and Macro-averaging used in text classification evaluation are used for evaluation. The calculation process is shown in formula (1)–(6). For Category C, the classification results can be divided into the following situations:

1) Originally Category C was divided into Category C, and the quantity was recorded as a;
2) Originally non-C was classified as C, and the quantity was recorded as B;
3) Originally Class C was classified as Non-C, and the quantity was recorded as C;
4) Originally non-C was classified as non-C, and the quantity was recorded as D;

$$\text{Micro\_P} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i} \tag{1}$$

$$\text{Micro\_R} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} a_i + \sum_{i=1}^{n} c_i} \tag{2}$$

$$\text{Micro\_F1} = \frac{2 * \text{Micro\_P} * \text{Micro\_R}}{\text{Micro\_P} + \text{Micro\_R}} \tag{3}$$

$$\text{Macro\_P} = \frac{1}{n}\sum_{i=1}^{n} P_i \tag{4}$$

$$\text{Macro\_R} = \frac{1}{n}\sum_{i=1}^{n} R_i \tag{5}$$

$$\text{Macro\_F1} = \frac{2 * \text{Macro\_P} * \text{Macro\_R}}{\text{Macro\_P} + \text{Macro\_R}} \tag{6}$$

4.3 Parameter Setting

The experiment uses a 768-dimensional BERT vector with 12-layer transformer and 12 multi-head attention mechanisms as the embedding layer of the model. When the accuracy rate of the verification set does not increase after 10 epoch, the training is stopped, and the best model parameters are obtained by the checkpoint. Besides, the experiment hopes to obtain a more accurate classification model in the multi-task learning of the experiment. The final loss value of the model is the sum of the loss value of the classification result and the loss value of the entity labeling result according to the weight of 3:1. Due to the imbalance of categories, the experiment gives corresponding class weight to different categories, so that the number of categories and the weight of each category are multiplied to the same value, ensuring that better results can be achieved in small categories. Based on this, the above models are compared and tested to obtain the best sequence model, which is combined with the quantitative index model. The active learning process is set up in 10 rounds. Each round obtains the top 100 data with the highest score for labeling, and finally obtains 1000 labeled samples for further migration training of the model.

4.4 Model Training Results

Selection of Transfer Layer in Transfer Learning.

The layers of the final PQE-MT are shown in Table 1:

For neural networks, different layers of the network model usually learn the characteristics of different levels of samples. "Visualizing and Understanding Convolutional Networks" [23] visualizes each layer of CNN, which shows the above theory more clearly. Layer1 and Layer2 learn basic color and edge features; Layer3 learns texture features; Layer4 learn local features such as wheel; Layer5 learns more discernible overall features. Compared with images, it is easier to observe, and the knowledge learned by each model of the text sequence is difficult to display, but in general it also conforms to the above-mentioned rules, and the bottom layer learns the features of part of speech and meaning of words. Learn semantic and syntactic features at a high level.

The experimental results list the micro-average and macro-average of accuracy, recall rate, and F value on the training set and the test set respectively. The results are shown in Table 2.

Table 1. PQE-MT layer information

Number of layers	Type	Output shape	Parameter quantity
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 2. Final results of different migration parts

Migration layer	Training set (micro avg/macro avg)						Test suite (micro avg/macro avg)					
	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

We conducted a model migration experiment using only English texts of Chinese patents as training data and not using U.S. Patents, and compared with the 3–14 layer model with the best transfer effect to observe the improvement of model effect by transfer learning. The two results are shown in Tables 3 and 4.

Through the comparison of Tables 3 and 4, it can be seen that transfer learning has a certain degree of improvement on each index of the model. Due to the lack of training samples in normal supervised learning, the whole model has two obvious problems: first, the prediction results of the model are very poor in categories with a small number of samples, and the main measurement indexes, accuracy rate, recall rate and F value results in categories 7 and 8 are all 0; Second, the generalization ability of the model is very poor, and the over-fitting problem is obvious. Whether it is micro-average or macro-average, the test set is about 35% lower than the training set. Comparing the results after transfer learning, we can see that the results have improved on both issues.

Table 3. Result of training network use 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
Micro avg	0.76	0.76	0.76	208	0.40	0.40	0.40	100
Macro avg	0.57	0.58	0.57	208	0.22	0.24	0.23	100

Table 4. Results of data transfer learning for initial 208 Chinese patent English text

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
Micro avg	0.83	0.83	0.83	208	0.66	0.66	0.66	100
Macro avg	0.71	0.70	0.70	208	0.61	0.54	0.57	100

For categories with a small number of samples, the model has a better adaptation. For the prediction results, on the training set and the test set, the gap between the micro-average values is reduced to 17%, and the macro-average values are reduced to 10%, 16%, and 15% respectively. These two problems have been solved to a certain extent, thanks to the fact that the model has learned how relevant features affect the results in a large number of U.S. Patents. Although the distribution of features is somewhat different, the overall trend is roughly the same. In the process of transfer learning, the model further adapts to the changes of Chinese patents in various index items on the originally learned trend.

Although the model has been improved, the effect still cannot meet the final patent evaluation requirements. I hope to improve the effectiveness of the model by increasing the migration data.

In the experiment, 200 unlabeled Chinese patents were randomly selected for manual labeling as data_random, and the top 200 Chinese patents with the largest classification margin were labeled as data_active_learning by using an active learning algorithm. In the experiment, the two data sets are respectively incremented by 20 from 0 until reaching 200. The effects of random data selection and active learning data selection on model migration were compared. The comparison of the prediction accuracy of the model in the two cases is shown in Fig. 4.

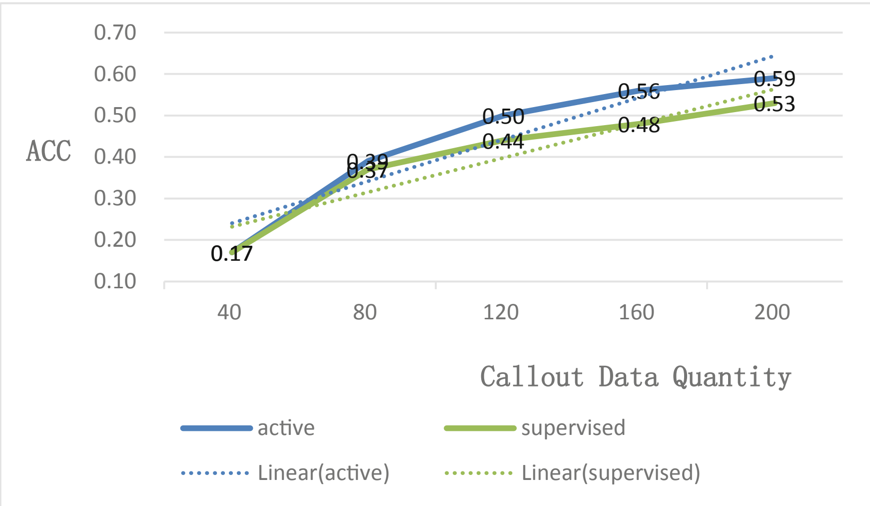

Fig. 4. Comparison of accuracy rate of active learning and randomly selected annotation data migration learning

Through the comparison of the two trend lines, it is found that the active learning selection data is better than the randomly selected data migration learning, and the model effect is equivalent to the random selection of 200 data annotation when the active learning selects about 140 data annotation. Therefore, the experiment uses active learning to further expand market data. Active learning can obtain data with the largest amount of information. In the experiment, 10 rounds of active learning were carried out, and the top 200 data with the largest classification margin were obtained for manual labeling in each round. A total of 2,000 labeled samples were obtained as new migration data, increasing the amount of data by only 10 times. The model uses the 2–14 layers with the best migration effect obtained by experimental comparison as the migration part, and the migration results of 2208 Chinese patent data obtained by active learning expansion are shown in Table 5. Besides, the experiment was carried out, in the same way, using the patented English language, and the results are shown in Table 6.

Table 5. Results of data transfer learning for 2208 Chinese patent english text 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
Micro avg	0.79	0.80	0.79	2208	0.74	0.74	0.74	100
Macro avg	0.70	0.67	0.68	2208	0.64	0.63	0.63	100

Table 6. Results of data migration learning for 2208 Chinese patents 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
Micro avg	0.73	0.71	0.72	2208	0.62	0.61	0.61	100
Macro avg	0.61	0.61	0.61	2208	0.55	0.51	0.53	100

The experimental results in Table 4 are compared with those in Table 5 to observe the influence of active learning and data expansion on the model effect. Using the model of initial data migration, the Micro-F1 values on the training set and the test set are 83% and 66%, and the Macro-F1 values are 70% and 57%, respectively. The overall model tends to over-fitting. After active learning for data expansion, Micro-F1 values are 79% and 74%, Macro-F1 values are 68% and 63%, respectively. Due to the increase in the amount of migrated data, the generalization ability of the model becomes stronger, and the effect of the model is improved to a certain extent when predicting the test set data.

The experimental results in Tables 5 and 6 are compared to observe the influence of Chinese and English languages on the model effect. It can be observed that the transfer effect of the model on English text is generally better than that on Chinese text. The author thinks that due to the limitation of machine translation technology when using the source model trained by US patent, the use of Chinese data after machine translation leads to the deterioration of the initial model effect and the reduction of the final transfer model effect. However, the overall influencing factors of English texts translated by Chinese patents, which are used in quantity and as target models, are relatively low. Besides, the differences in writing specifications and feature spaces between Chinese and English also make the learning effects of the models different.

5 Conclusions

By comparing and analyzing the patent data between China and the United States, the experiment summarizes the similarities and differences between the two countries' patents, aligns the quantitative indicators, and migrates to Chinese patents through the multi-task learning network PQE-MT model trained by the United States patent with patent quality rating labels. The transfer process mainly involves the selection of transfer parts, cross-language transfer, and the use of active learning to expand data. The advantage of this model lies in the use of migration learning technology and active learning to select the data to be labeled with the largest amount of information, thus minimizing the time-consuming manual labeling process. In the whole process, through experiments, the prediction effect of the model was gradually improved, and finally, the Chinese patent quality evaluation model achieved good accuracy. At the end of the experiment, the effect of transfer learning between Chinese and English was compared, and it was found that the result indicators were different, and the model learned different features. In the following work, the author will combine the two language models and consider the different features learned so that they can influence each other and improve the final results.

Acknowledgments. This work is supported by National Natural Science Foundation of China under Grants No. 61671070. Project of Developing University Intension for Improving the Level of Scientific Research–No. 2019KYNH226, Qin Xin Talents Cultivation Program, Beijing Information Science & Technology University No. QXTCP B201908.

References

1. In 2018. China spent nearly 2 trillion yuan on research and development, and major scientific and technological innovation indicators steadily improved. East China Sci. Technol. **2019**(04), 13 (2018)
2. China News Network. State Intellectual Property Office: In 2018, China filed 1.542 million invention patents [EB/OL]. (2019-1-10) [2019-6-21]. http://money.163.com/19/0110/16/E563C6PE00258105.html

3. Akdemir, A.: Research on task discovery for transfer learning in deep neural networks. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pp. 33–41 (2020)
4. Wu, J.L.: Patent quality classification system using the feature extractor of deep recurrent neural network. In: 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 1–8. IEEE (2019)
5. Banea, C., Mihalcea, R., Wiebe, J., et al.: Multilingual subjectivity analysis using machine translation. In: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, pp. 127–135 (2008)
6. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2009)
7. Xu, Q., Yang, Q.: A survey of transfer and multitask learning in bioinformatics. J. Comput. Sci. Eng. **5**(3), 257–268 (2011)
8. Weiss, K., Khoshgoftaar, T.M., Wang, D.D.: A survey of transfer learning. J. Big Data **3**(1), 9 (2016)
9. Yosinski, J., Clune, J., Bengio, Y., et al.: How transferable are features in deep neural networks? Advances in Neural Information Processing Systems, pp. 3320–3328 (2014)
10. Tohompson, C., Califf, M.E., Mooney, R.: Active learning for natural language parsing and information extraction. In: Proceedings of the 16th International Conference on Machine Learning, pp. 406–414. Morgan Kaufmann, San Francisco (1999)
11. Tong, S., Koller, D.: Support vect or machine active learning with applications to text classification. J. Mach. Learn. Res. **2**, 45–66 (2001)
12. Settles, B.: Active Learning Literature Survey. University of Wisconsin-Madison Department of Computer Sciences (2009)
13. Li, J., Xiao, Y., Hao, Z., Ruan, Y., Zhang, L.: Multi-example and multi-tag active learning based on SVM. Comput. Eng. Des. **37**(01), 254–258 (2016)
14. Zhou, Z., Shin, J., Zhang, L., et al.: 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 (2017)
15. Zhu, J.J., Bento, J.: Generative adversarial active learning (2017). arXiv:1702.07956
16. Konyushkova, K., Sznitman, R., Fua, P.: Learning active learning from data. In: Advances in Neural Information Processing Systems, pp. 4225–4235 (2017)
17. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2018)
18. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
19. Amine, B.M., Mimoun, M.: Wordnet based cross-language text categorization. In: 2007 IEEE/ACS International Conference on Computer Systems and Applications, pp. 848–855. IEEE (2007)
20. Gliozzo, A., Strapparava, C.: Cross language text categorization by acquiring multilingual domain models from comparable corpora. In: Proceedings of the ACL Workshop on Building and Using Parallel Texts, pp. 9–16. Association for Computational Linguistics (2005)
21. Bel, N., Koster, C.H.A., Villegas, M.: Cross-lingual text categorization. In: International Conference on Theory and Practice of Digital Libraries, pp. 126–139. Springer, Berlin (2003)
22. Beltz, H., Rutledge, T., Wadhwa, R.R., et al.: Ranking algorithms: application for patent citation network. Information Quality in Information Fusion and Decision Making, pp. 519–538. Springer, Cham (2019)
23. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: European conference on computer vision, pp. 818–833. Springer, Cham (2014)