

# Research on Patent Document Classification Based on Deep Learning

Bing Xia, Baoan LI\* and Xueqiang Lv

Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, Beijing Information Science and Technology University, Beijing 100101, China

\*Corresponding author

**Abstract**—Science and technology can be called the first productive force of development, and the patent as an important source of science and technology innovation is getting more and more attention. With the increase in the number of patent applications, followed is patent document classification problem. At present, the patent document classification include artificial method and automatic method, but mostly using the artificial method, which resulted in the classification process is time consuming and not efficient, so the research of patent document classification has the practical significance. In this paper, based on the premise of deep learning, the patent document classification problem is studied.

**Keywords**—deep learning; patent document classification; sparse automatic encoder; deep belief network; softmax

## I. INTRODUCTION

Patent document has attracted the attention of researchers for its creativity, novelty and practicality. According to the European Patent Office statistics, 80% of all technical information in the world can be found in the patent document. The relevant departments of World Intellectual Property Organization (WIPO) have also done statistics, 90% to 95% of the world's inventions are published in the form of patent document, of which about 70% inventions have never been published in other non-patent document. If you can apply the patent documents, scientific research institutions can save 40% of R & D expenditures, saving 60% of the time.

The new patent application needs to be strictly audited before being granted the patent right. Patent examiners in the intellectual property department need to classify the new applications, give the corresponding IPC (International Patent Classification) classification. Artificial classification needs to spend a lot of manpower and material resources, and the classification requires repeated experiments, reading patent abstract and judging the IPC classification and key words of search topics to quickly and accurately retrieve the patent, it requires specialized knowledge; it is difficult for the non-professional. At the same time, because of the different cultural backgrounds and professional knowledge of the patent examiners, the classification of different people will be different. Therefore, the use of automatic or semi-automatic auxiliary classification, can greatly reduce the workload of examiners, and reduce the uncertainty and classification errors of human classification, at the same time, the IPC classification is established to improve the retrieval efficiency.

## II. RESEARCH STATUS OF PATENT DOCUMENT CLASSIFICATION

At present, most of the patent text classification researches still focus on improvement of the existing algorithm and they apply it to patents; IPC as the primary reference for classification system; there are three kinds of data sources: NTCIR data, from the Patent Office or database for data.

Li Chengxiong [1] uses the method of SVM and KNN combination, test under eight large classes in seven parts, and is the classification of the IPC level. Research team of South China University of Technology [2] uses the nuclear vector space and the Bayes mode [3] to classify the patents obtained from Guangdong Intellectual Property Office. Jiang Jianan [4] uses cosine similarity in the IPC big class level in his hierarchical classification method, uses KNN in small class level. Li Shengzhen [5] uses BP neural network to classify the patents downloaded from the State Intellectual Property Office network, the category system reaches the main group level of IPC. Ji Duo [6] uses KNN and its improved algorithm to do classification research according to NTCIR-8 specification, evaluation depth reaches the IPC sub group, using MAP as evaluation indicators, getting better results in a small class. Li [7] completes the F term patent classification sub task of NTCIR-6 based on SVM, and he believes that the information provided by the F term could help to improve the classification results.

Besides the above content based patent text classification research, there is also the classification based on the citation relationship. Lai [8] uses the method of principal component analysis to classify the patent text according to the Co citation relationship among the patents. Li [9] uses the combination of nuclear function and reference network to classify nanotechnology patent.

## III. THE DEVELOPMENT OF DEEP LEARNING

Machine learning is applied to many aspects of our lives, from web search to filter the content of social network, and then to the recommendation system of e-commerce website, and more and more high-tech products is using machine-learning techniques, such as cameras and intelligent mobile phone. Machine learning systems used in the field of image recognition, speech recognition, machine translation, recommendation systems, search optimization, etc.. Deep learning technology is used in these areas.

Deep learning has solved many practical problems in recent years. When the data is high dimensional data or structure is very complex, deep learning has been proved very effective in dealing with the data, so it is applied in many fields, such as science, enterprise, government and so on. In addition to the image recognition [10-12] and speech recognition [13-15] to beat other machine learning techniques, in the prediction of potential drug molecules [16], analysis of particle accelerator data and the reconstruction of brain circuits [17], the deep learning method has also defeated other machine learning techniques. Perhaps more surprising is that in some important subjects of natural language understanding, particularly theme classification, sentiment analysis, question-answering system and machine translation, deep learning method also brings encouraging results.

Deep learning will have more success in the near future because it requires very little manual work, so it can easily be used by increase the available computing and data. It will be bound to accelerate the development of this process new learning algorithms and architectures are currently being developed for deep neural networks.

#### IV. PATENT DOCUMENT CLASSIFICATION BASED ON DEEP LEARNING

After detailed research and analysis on the patent document automatic classification and deep learning theory, this paper uses two kinds of deep learning models, which are the sparse automatic encoder and the deep belief network. The hybrid model consists of three parts, the front part has two layers of sparse auto encoder structure, the middle part uses a three-layer deep belief network, finally Softmax regression is used as the classification layer, the deep learning theory is applied to patent document classification.

##### A. Design of Patent Document Classifier

In the completely patent document classifier, it is generally divided into three modules: the patent document-preprocessing module, the feature-learning module and the classification and recognition module.

- In the patent document-preprocessing module, the documents in the document set should be processed in advance. De-noising, word segmentation, remove stop words and other process to the document.
- In the feature-learning module, the data obtained from the pre process of the patent document. Feature learning is performed using sparse automatic encoder layer and depth belief network in feature module.
- In the classification recognition module, using Softmax regression to document classification, the classifier can be very good for the classification of patent document, in the classification module, we can also use other machine learning methods for classification, such as the use of support vector machines or Naive Bayesian.

##### B. Patent Document-Preprocessing Module

In the patent document-preprocessing module, the documents in the document set should be processed in advance. The preprocessing operation includes the following steps:

- First, to deal with useless information in the document, to remove document format, punctuation, numbers, special characters and some other irrelevant information in the documents.
- Word segmentation for documents and divide the sentences in the document into words.
- Remove stop words in the document, there are many useless features in the document, no discrimination and prediction abilities such as articles, pronouns, prepositions, conjunctions and auxiliary.
- Set up the digital mapping table of vocabulary in the text, statistics on the words, using the text space vector to store, so that the text is expressed as a feature vector.

##### C. Feature-Learning Module

In the process of patent document classification, due to the large amount of data, high dimension, it is very easy to cause the feature is not clear. In this paper, we use sparse automatic encoder layer and deep belief network layer for feature learning. In the feature learning stage, the sparse automatic encoder is used to carry out the preliminary study of the feature.

The feature learning process of patent document with sparse automatic encoder, the feature vector of the patent document are used as input to the encoder and get a kind of code. Then the code is added to the decoder, and the reconstruction is obtained. By adjusting the parameters of encoder and decoder, to achieve the minimum error between input and reconstruction, then another representation of the text feature vector is obtained by adjusting the parameters of the code, which is the first feature after the sparse automatic encoder learning. Then training is done at the same way in the second layer. The code obtained from the first layer will be the input of the second layer, minimize the reconstructed error and get the parameters of encoder and decoder in the second layer, the code of second layer obtained by the parameters is the feature obtained by the preliminary extraction of sparse encoder.

In the patent document feature learning process, parameters are crucial to sparse automatic encoder, for solving parameters of sparse automatic encoder first to determine the loss function of sparse automatic encoder by the network structure, this function is based on the loss function in the network without the sparse constraint, add sparse limit quantitative KL divergence. The expression for the loss function of the sparse automatic encoder is:

$$J_{sparse}(W, b) = J(W, b) + \beta \sum_{j=1}^{S_2} KL(p||\hat{p}_j) \quad (1)$$

Among them,  $\beta$  is sparse penalty factor, which can effectively control the sparsity.

$$KL(p||\bar{p}_j) = p \lg \frac{p}{\bar{p}_j} + (1-p) \lg \frac{1-p}{1-\bar{p}_j} \quad (2)$$

The partial derivative of the function is obtained according to the loss function of the sparse automatic encoder, and the error of the neuron node in the output layer is:

$$\delta_i^{(3)} = \frac{\partial}{\partial z_i^{(3)}} \frac{1}{2} \|y - h_{W,b}(x)\|^2 = -(y_i - a_i^{(3)}) f'(z_i^{(3)}) \quad (3)$$

The error of the neurons node in the input layer is:

$$\delta_i^{(2)} = \left[ \left( \sum_{j=1}^{S_2} W_{ji}^{(2)} \delta_j^{(3)} \right) + \beta \left( -\frac{p}{\bar{p}_j} + \frac{1-p}{1-\bar{p}_j} \right) \right] f'(z_i^2) \quad (4)$$

The final partial derivative is:

$$\frac{\partial}{\partial w_i^{(l)}} J(W, b; x, y) = a_j^{(l)} \delta_i^{(l+1)} \quad (5)$$

$$\frac{\partial}{\partial b_i^{(l)}} J(W, b; x, y) = \delta_i^{(l+1)} \quad (6)$$

Sparse automatic encoder through constant training, when the partial derivative of the loss function achieves the minimum value, the parameter of the sparse automatic encoder is obtained. According to this parameter, the preliminary characteristics of the learning of sparse automatic encoder are obtained.

Then, it can be using the deep belief network for further feature extraction, which makes it easier to classify the features. The training of deep belief network consists of two steps: pre training and fine-tuning.

During the first step of pre training, it is no supervise mode. By training each layer of RBM, RBM hidden layer vector is obtained by mapping the vector of visual layer, the hidden layer vector as the visual layer vector of the next RBM. And so on, according to the difference between the hidden layer and the visual layer to update weight of each layer in the whole network structure.

In the training of RBM, this paper uses a fast learning algorithm, contrastive divergence algorithm. Firstly, take the state of the neurons in the visual layer as a training sample  $V$ , calculate the state in the hidden neuron node layer, according to the calculated results, and then calculate the state of the neurons in the visual layer, thus we get a new visual layer. When the original visual layer and the reconstructed visual layer are simultaneously, the other expression of the visual layer can be replaced by the hidden layer. The existing error between original visual layer and reconstructed visual layer can be used to adjust the system parameters  $\theta = (w, a, b)$ .

In the second step fine-tuning process, it is supervise mode, use of labeled data sets. Using the back propagation algorithm to optimize the weight of the whole network, get better network weights. In the training process of the deep belief network, the problem of local optimum caused by random initial weights is avoided, and the whole weight value of the

network is continuously updated through the fine-tuning process, which makes the classification performance better.

#### D. Classification Recognition Module

In the classification recognition module, the Softmax regression is used to classify, and the features learned from feature-learning module are used as the input of Softmax classifier.

In the Softmax classifier, there are  $m$  texts in the training set,  $m$  texts from  $k$  classes. The training set is expressed as  $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ ,  $x^{(i)} \in R^{n+1}$  represents section  $i$  training text,  $y$  represents a class that is capable of taking a different value of  $k$   $y^{(i)} \in \{1, 2, \dots, k\}$ . Using the assumed function to calculate the probability value of  $x$  which belongs to each category of  $j$   $P(y = j|x)$ , using  $K$  dimensional vector to represent these probability values. The assumed functions are as follows:

$$h_{\theta}(x^{(i)}) = \begin{bmatrix} P(y^{(i)} = 1|x^{(i)}; \theta) \\ P(y^{(i)} = 2|x^{(i)}; \theta) \\ \vdots \\ P(y^{(i)} = k|x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \quad (7)$$

Among them, each sub vector of the vector  $h_{\theta}(x^{(i)})$  is the probability that the document belong to different classes, in order to make the probability of the entire sub vectors equal to one, the probability is normalized,  $\theta_1, \theta_2, \dots, \theta_k \in R^{n+1}$  represents the parameters used in the model.

According to the hypothesis function  $h_{\theta}(x)$  vector to solve the probabilities of test text  $x$  belongs to all categories, among which probability is the largest, the text will belong to which class.

#### V. CONCLUSION

Nowadays, with the rapid development of modern science and technology, patents are becoming more and more important in people's minds, this has led to a surge in the amount of patent analysis and utilization of the existing patents has received extensive attention from the related fields. Patent document automatic classification as the basis of patent analysis and utilization, the classification techniques have also been proposed for higher requirements. Patent document classification based on deep learning will have a very good application prospects in the development of the future. With the continuous improvement of deep learning theory, further research on this article will add more new research results of deep learning; this will enhance the performance of the patent document classifier which is based on deep learning.

#### ACKNOWLEDGEMENT

This project was supported by the Funding Project for Natural Science Foundation of China (Grant No. 61671070) and the Opening Project of Beijing Key Laboratory of Internet Culture and Digital Dissemination Research (Grant No. ICDD201608).

## REFERENCES

- [1] Li Chengxiong, Ding Yuehua, Wen Guihua. Application of SVM – KNN Combination Improvement Algorithm on Patent Text Classification[J]. Computer Engineering and Applications, 2006, 42(20),193-195.
- [2] Ding Yuehua, Wen Guihua, Guo Weiqiang. Patent Categorization Based on Kernel Vector Space Model[J]. Journal of South China University of Technology: Natural Science Edition, 2005,33(8),58-61.
- [3] Guo Weiqiang, Wen Jun, Wen Guihua. Patent Categorization Based on Bayes Model[J]. Computer Engineering and Design, 2005,26(8),1986-1987,1996.
- [4] Jiang Jian'an, Lu Jieping, Ni Weiwei. Automatic Text Categorization for Patent Data[J]. Journal of Computer Applications,2008,28(1),159-161.
- [5] Li Shengzhen, Wang Jianxin, Qi Jiandong. Automated Categorization of Patent Based on Back – propagation Network[J]. Computer Engineering and Design,2010,31(23),5075-5078.
- [6] Ji Duo, Cai Yunlei, Cai Dongfeng. Patent Automatic Classification Research Based on Shared Nearest Neighbor[J]. Journal of Shenyang Institute of Aeronautical Engineering,2010,27(4),41-46.
- [7] Li Y Y,Bontcheva K, Cunningham H. SVM Based Learning System for F – term Patent Classification[C]. In: Proceedings of the 6<sup>th</sup> NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and CrossLingual Information Access.2007.
- [8] Lai K K, Wu S J. Using the Patent Co – citation Approach to Establish a New Patent Classification System[J]. Information Processing and Management, 2005, 41(2),313-330.
- [9] Li X, Chen H, Zhang Z. Automatic Patent Classification Using Citation Network Information: An Experimental Study in Nanotechnology[C]. In: Proceedings of the 7<sup>th</sup> ACM/IEEE – CS Joint Conference on Digital Libraries. New York: ACM,2007,419-427.
- [10] Krizhevsky, A., Sutskever, I., Hinton, G. ImageNet classification with deep convolutional neural networks. In Proc. Advances in Neural Information Processing Systems 25,2012,1090–1098.
- [11] Farabet, C., Couprie, C., Najman, L., LeCun, Y. Learning hierarchical features for scene labeling. IEEE Trans. Pattern Anal. Mach. Intell. 35,2013, 1915–1929.
- [12] Tompson, J., Jain, A., LeCun, Y., Bregler, C. Joint training of a convolutional network and a graphical model for human pose estimation. In Proc. Advances in Neural Information Processing Systems 27,2014 1799–1807.
- [13] Mikolov, T., Deoras, A., Povey, D., Burget, L., Cernocky, J. Strategies for training large scale neural network language models. In Proc. Automatic Speech Recognition and Understanding,2011,196–201.
- [14] Hinton, G. Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine 29,2012,82–97.
- [15] Sainath, T., Mohamed, A.-R., Kingsbury, B., Ramabhadran, B. Deep convolutional neural networks for LVCSR. In Proc. Acoustics, Speech and Signal Processing,2013,8614–8618.
- [16] Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E., Svetnik, V. Deep neural nets as a method for quantitative structure-activity relationships. J. Chem. Inf. Model. 55,2015,263–274.
- [17] Helmstaedter, M. Connectomic reconstruction of the inner plexiform layer in the mouse retina. Nature 500,2013,168–174.