

## Research Article

# Enterprise Complementarity Based on Patent Information

Lin Sun  and Wenzheng Xu

College of Intelligent Equipment, Shandong University of Science and Technology, Tai'an, Shandong, China

Correspondence should be addressed to Lin Sun; [sunlandbell@163.com](mailto:sunlandbell@163.com)

Received 23 May 2022; Accepted 28 June 2022; Published 26 July 2022

Academic Editor: Chia-Huei Wu

Copyright © 2022 Lin Sun and Wenzheng Xu. 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.

The combination of complementary patents can better promote the advancement of technology, and complementary companies can help companies bring better benefits. This topic has carried out patent complementarity research and proposed a patent complementarity algorithm based on patent technology tree. By calculating the complementarity calculation method of a given patent, the company with the strongest complementarity with known patents is obtained. First, the patent text is represented sentence vectors by LSTM model. Second, the generated sentence vectors are clustered, and the patent information is marked with different categories. Then, the patent word vector is obtained by the CNN text classification model to obtain the confidence that the patent belongs to the corresponding category. Finally, the confidence of the patent and the complementarity coefficient based on the patent classification number are combined to obtain the interpatent complementarity. Calculate the average of the complementarity between all patents and a given patent under the same company, and the company with the greatest complementarity is the complementary company sought. The method combines the patent text and the patent classification number, obtains the similarity of the text semantics through the deep learning model, and quantifies the complementarity of the patents, and the discriminating basis is more accurate and reliable.

## 1. Introduction

As early as 2003, Hoskisson and Yiu [1] proposed complementary technologies to bring less pressure to the employees and management of the company. In 2010, Makri [2] proposed a concept of relevance to the relationship between patents, including two concepts of similarity and complementarity. The concept of patent relevance consists of patent similarity and patent complementarity. Patent similarity is the degree of overlap of two patent information in the same technical field. That is to say, a similar patent refers to a patent that belongs to the same or similar technical field and achieves the same or similar technical effects. Patent complementarity is that two patent information pay attention to different narrow areas in their common fields.

It is very important for enterprises to study the relevance of patents. Marki also pointed out that the complementarity of patented technologies can promote the production of more and better quality products than similarities. The research on patent relevance can support enterprises to find their potential partners and competitive enterprises and play a guiding role in

enterprise M & A. There is a competitive relationship between enterprises with high similarity and low complementarity, so enterprises should strengthen their own enterprise characteristics. For enterprises with certain similarity and great complementarity, they can strengthen cooperation among enterprises to achieve the purpose of mutual benefit and win-win.

Although patent relevance is very meaningful for enterprise development. At present, the research of patents mainly focuses on patent similarity, and there is relatively little research on patent complementarity. Wei [3] proposed that enterprises should pay attention to enterprises that are similar or complementary to themselves when selecting acquisitions. Zhang [4] used the LDA topic model to generate industrial trees in the study of industrial-oriented patents. Experts used complementary scores for each technology branch to calculate the complementarity between patents. Li et al. [5] designed a combination of patent portfolios based on expected failure analysis, including patent portfolio analysis, failure analysis, TRIZ tools to solve problems, and evaluation of umbrella patent portfolios. Wang [6] uses the method of correlation analysis to discover the complementary relationship between patents.

<u>A</u>	<u>63</u>	<u>H</u>	<u>3/</u>	<u>42</u>
Section	Category	Subclass	Large group	Group

FIGURE 1: Patent component presentation.

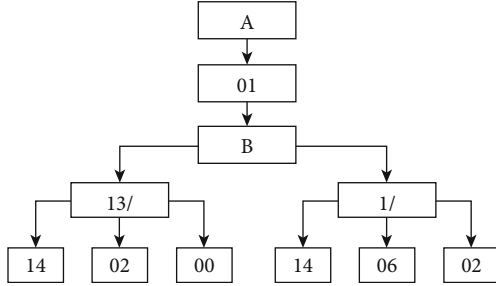


FIGURE 2: Patent technology tree.

Most researchers choose to use text analysis to process the content of patent texts and then discover the complementary relationship between patents. In the past, researchers have studied the patent texts by using the word bag model to deal with the content of patent texts, which does not take into account the semantic relationship. This topic uses the LSTM model to generate sentence vectors when dealing with patent text content. This way can represent the content of the text and maintain the word order relationship between words. The researcher proposed the method based on expert evaluation [4–6]. The measurement of patent complementarity is costly and feasible in practice. Therefore, this study designed the complementarity calculation based on patent technology number and text category judgment to quantify the complementarity degree calculation method.

## 2. Computation of Complementary Coefficients Based on Patent Technology Tree

The complementarity coefficient is the similarity of patents in classification. The higher the complementarity coefficient, the higher the patent classification similarity. The calculation of complementarity coefficient is realized through the patent IPC classification number. IPC patent classification number adopts the classification principle of combining function and application, with function as the main and application as the auxiliary.

The IPC classification number of a patent consists of five parts, namely, section, category, subclass, large group, and group. The department consists of 1 letter, the large category consists of 2 numbers, the subcategory consists of 1 letter, the large group consists of 1 to 3 numbers, the group consists of 2 to 4 numbers, the large group represented by the fourth part and the group represented by the fifth part are separated by /, and the patent component presentation is shown in Figure 1.

For example, the section of the patent with the main classification number A63H3/42 is A, indicating that the patent is essential for human life. Such patents mainly include the contents of agriculture, light industry, and med-

icine. 63 indicates the category of this patent, and A and 63 together express sports, games, and entertainment activities. Letter H represents a subclass of patents, and A63H represents toys such as gyros, building blocks, and dolls. 3 means a large group under H, and A63H3 means dolls. 42 is the group which represents the doll of the eye. This patent is how to make the moving eyes of the doll. The method of calculating the complementary coefficient of the patented technology will be realized by constructing a patented technology tree. In each of the children's nodes in each of the penultimate layers of the patented technology tree, a complementary coefficient is set for each of the two child nodes, and a patent information is determined by text mining. In the position of the constructed patent technology tree, combined with the set complementary coefficient to calculate the complementarity between the patented technologies, the patent pair with the strongest complementarity in the patent classification number is obtained, that is, to find different patents supporting the same product.

According to the components of the patent main classification number, a patent technology tree is constructed from top to bottom. The root node of the setting tree is the section represented by the patent department, the second layer node is the category of patents, the third layer node is the subclass of patents, the fourth layer node is a large group of patents, and the fifth layer is a patent group. Through this patented technology tree of Figure 2, the location of the final patent information can be got. Given two patents, the five parts from the patent classification number and the same parts from left to right are the complementary coefficients of the two patents as the following equation.

$$A_{i,j} = \frac{x}{5}, \quad (1)$$

where  $x$  represents the number of parts of the patent  $P_i$  and the patent  $P_j$  that are identical from left to right in dividing the patent into five parts. For example, A01B13/00 and A01B13/02, the parts with the same classification number of the two patent numbers are A01B13, the difference is group number which is 00 and 02, and the complement coefficient is 0.8 according to the two identical parts.

## 3. Patent Confidence Calculation Based on Text Analysis

Confidence indicates the probability value of the patent belonging to its corresponding text category, that is, the probability value of the patent text belonging to the category when the patent text is clustered according to the text word vector. The degree of confidence in the patent can be discriminated by the degree of confidence. The higher the confidence, the higher the similarity of the patent text.

As the patent classification involved by enterprises in Tai'an City is limited, if the patent classification is carried out according to the patent classification number, some categories may be dense and some sparse, and the classification effect may be poor. Therefore, patents are expressed as word vectors in this topic. After text clustering, classification

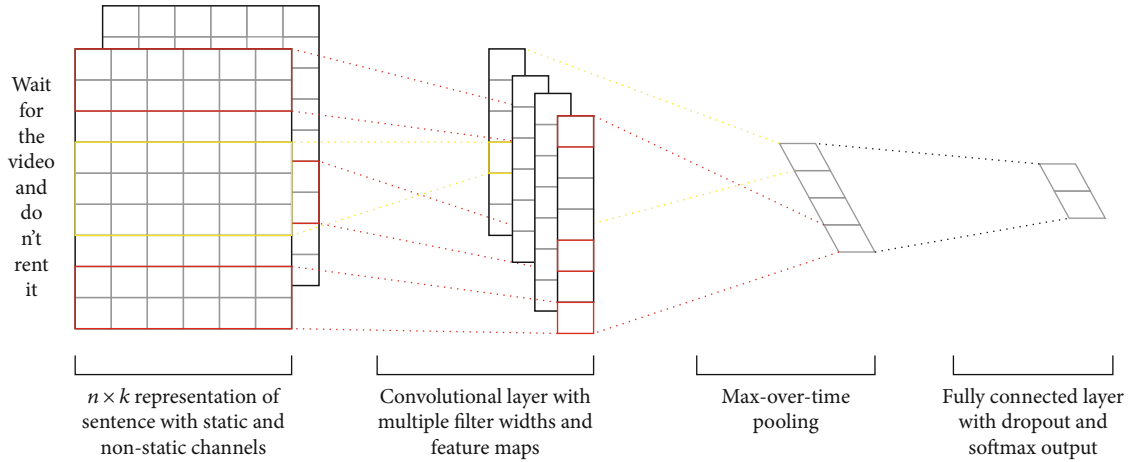


FIGURE 3: Text classification based on TextCNN.

TABLE 1: Results of patent clustering.

Category	0	1	2
Number of enterprises	1956	2426	2962

training is carried out according to the clustering results to determine the category of patents.

**3.1. Patent Text Representation Model Implementation.** When researching patent data, because the computer cannot directly process the text, the researchers proposed VSM, word vector, and other methods [7–10] to represent the text as a computer-processable mode. As deep learning techniques go deeper, more researchers apply it to text analysis. RNN is widely used to deal with natural language issues. Although RNN can process time series more efficiently, it can also cause problems such as gradient disappearance or gradient explosion. The researchers proposed the LSTM model to solve the problem of gradient disappearance, which is suitable for processing and predicting important events with relatively long intervals and delays in time series. LSTM is better able to handle long distance dependencies. Wang et al. [11] proposed a part-of-speech tag based on LSTM. Pappas and Popescu-Belis [12] proposed the HAN model for text categorization, that is, adding the attention layer to operate and select the most important content of the text.

Because the patent text research pays more attention to the keywords of patents, the researcher's research on patent texts uses the word bag model to deal with the content of patent texts, but this method does not consider semantic relations. Therefore, this paper uses the deep learning mode to process the clustering and classification of patent texts not only considering keywords but also semantic relation. In this paper, the text is obtained by the embedding layer, and the LSTM model is used to characterize the content of the patent text in order to obtain a word vector that can represent the text content.

**3.2. Text Clustering.** The clustering algorithm is used to group data objects and divide similar objects into the same group. The labels of the data using the clustering algorithm

are unknown, and the specific criteria of the grouping are also known and are therefore also referred to as unsupervised learning. After clustering the data, the similarity between the groups is as small as possible, and the similarity within the groups is higher.

After the sentence vectors used to represent the text content is obtained by the text representation model, the sentence vectors can be clustered by the Kmeans clustering algorithm which the patent texts are similar. After clustering, each patent texts are tagged with a clustering label, and patents with similar technology are placed in the same category. Density clustering and hierarchical clustering algorithms are also used to complete the text clustering, but the effect is not as good as Kmeans.

**3.3. Text Classification.** The research and implementation of text categorization are a basic research direction in natural language processing. For the text classification method, there are methods such as naive Bayes [13] and support vector machine [14] with machine learning algorithms. With the rapid development of deep learning, researchers have focused on deep learning and began to use deep learning methods for text classification. CNN is better to extract more local important information, so more applications of computer vision in the early days are using CNN. Kalchbrenner et al. [15] and Kim [16] proposed TextCNN model on ACL and EMNLP, respectively, which applied CNN to text categorization. Conneau et al. [17] proposed the VDCNN model in natural language processing that captures key information over longer distances.

The model uses TextCNN for text categorization as Figure 3. The model includes the convolutional layer, the pooling layer, and the fully connected layer. The word vector processed by embedding enters the convolution layer. After the feature is extracted in the convolution layer, the output result is input into the maximum pooling layer, and the selected feature vector is placed in a fully connected softmax layer to generate a probability which is the sentence category. Word vector matrix' size is  $m \times n$ , that means there are  $m$  words and word vector size is  $n$ . The matrix is fed into input layer. At the convolutional layer, the size of the

convolution kernel is set to  $h * n$ ,  $h$  is the number of words that needs to be convoluted, and  $n$  is the dimension of the word vector. In this layer, the input word vector matrix of each text is convoluted with the convolution kernel to obtain a number of feature information, that is called the feature map; then, the feature map is fed to the next layer. At the pooling level, the maximum pooling method is chosen. The output of the pooling layer is the maximum value in each feature map, which is a one-dimensional vector. The data output from the pooled layer enters a fully connected layer and is processed with softmax. Softmax can generate probability values, each of which corresponds to each category. The category with the highest probability value is the predicted category. Softmax's formula is as the following equation.

$$S_i = \frac{e^i}{\sum_j e^j}. \quad (2)$$

**3.4. Patent Confidence Calculation.** The confidence calculation is performed in 4 steps. First, a vector representation of each patent text is obtained after the LSTM model. Second,  $k$ means algorithm is used for text clustering. Through various clustering index evaluations, the number of clusters is determined to be 3, and the clustering category of each patent text is obtained. Third, the TextCNN model is used to reclassify the text, set the appropriate parameters and optimize the model, and continuously train the optimization model to make the classification model achieve a good effect. The patent's probability value is obtained by the classification model through softmax calculation under each category. Finally, the probability maximum of the patent belonging to a category is selected as its confidence.

Given a patent, the maximum probability which belongs to a text category after TextCNN model is its patent confidence, and the text category to which the patent belongs is its patent confidence class, as follows.

$$C_i = \max(C_{L1}, C_{L2}, C_{L3}), \quad (3)$$

$$LA = \operatorname{argmax}(C_{L1}, C_{L2}, C_{L3}). \quad (4)$$

$C_i$  represents the patent confidence of patent  $i$ .  $C_{L1}$ ,  $C_{L2}$ , and  $C_{L3}$  indicate the probability value of patent  $i$  belongs to categories  $L1$ ,  $L2$ , and  $L3$ .  $LA$  is the patent confidence class.

## 4. Patent Complementarity Analysis

**4.1. Patent Complementarity and Similarity.** Patent complementarity is related to the complementary coefficient and patent confidence. The complementary coefficient defines the similarity in patent's main classification number. Patent confidence defines the similarity in the patent text, so they can be combined to obtain the patent complementarity.

Through analysis, it is found that if patents  $P_i$  and  $P_j$  are under the different patent confidence class  $a$  and  $b$ , i.e.,  $ab$ ; then, the two patents have high complementarity coefficients and high patent confidence, and they are considered to be highly complementary. These two patents are different tech-

nical components of the same product. If the patents  $P_i$  and  $P_j$  are under the same confidence class, i.e.,  $a = b$ ; then, the two patents have high complementarity coefficients and high confidence, and they are considered to be similar. The two patents are similar technical components of the same product. Therefore, patent complementarity  $W_{i,j}$  and similarity  $S_{i,j}$  are defined as the following equation.

$$W_{i,j} = (C_i + C_j) * A_{i,j} \quad \text{if } LA_i \neq LA_j, \quad (5)$$

$$S_{i,j} = 1 - \frac{W_{i,j}}{2}$$

$$S_{i,j} = (C_i + C_j) * A_{i,j} \quad \text{if } LA_i = LA_j. \quad (6)$$

$$W_{i,j} = 1 - \frac{S_{i,j}}{2}$$

Among them,  $C_i$  and  $C_j$  are the patent confidence of patents  $P_i$  and  $P_j$ ,  $LA_i$  and  $LA_j$  are the confidence classes of patents  $P_i$  and  $P_j$ , and  $A_{i,j}$  is the patent complementary coefficient of  $P_i$  and  $P_j$ .

**4.2. Enterprise Complementarity Analysis.** The enterprise complementarity analysis distinguishes the enterprise with the strongest complementarity patent of a known patent, that is, the enterprise where the highest patent complementarity of a patent set is located.

Calculating the patent complementarity  $W_1, W_2 \dots W_n$  of a given patent  $p$  with other patents, if some patents belong to the same company, it will be averaged the patent complementarity of the company. The company with the highest average value of patent complementarity is the most complementary company it is looking for. The company complementarity  $Y$  is shown in the following equation.

$$Y = \bar{W} = \frac{\sum_{i=1}^n W_i}{n}. \quad (7)$$

## 5. Experimental Results and Analysis

**5.1. Data Source.** Since it is great significance for the enterprise complementarity of companies in the same region, this paper is based on the patent of Tai'an of Shandong Province that is the author's city. Use the reptile to obtain 7344 patents from baiten website between 1983 and 2016 in Tai'an. The patent data includes the patent name, application number, applicant author, application date, main classification number, and abstract, all of this is stored in the csv file.

### 5.2. Model Building

**5.2.1. Text Clustering.** This topic uses LSTM model to achieve text representation. After the preprocessing of the patent's summary, the word vector is obtained through the embedding layer, and the dimension of each word vector is 64 dimensions.

The word vector representation of each word is input into the LSTM model. The number of cells in the hidden layer of the LSTM model is 128. Through the LSTM model,

TABLE 2: Patent complementarity coefficient display.

Patent	$P_1$	$P_2$	$P_3$
Patent number	CN200920028726.0	CN201020173582.0	CN201110318971.7
PCN	A01D13/00	A01D13/00	A01C7/18
$P_1$		1	0.4
$P_2$	1		0.4
$P_3$	0.4	0.4	

TABLE 3: Text classification results.

Category	0	1	2
Classification probability	0.05	0.92	0.03

a sentence vector for characterizing the text content can be obtained which size is 128.

The text sentence vector uses  $k$ means for text clustering, the Calinski-Harabasz score and silhouette score are used to evaluate the clustering model, it is found that when the cluster number is 3, and the effect of clustering is best, so the patents are divided into three categories as shown in Table 1.

**5.2.2. Text Classification.** This paper uses TextCNN model to achieve text classification. This model includes the convolution layer, the pooling layer, and the fully connected layer and performs softmax processing.

First, after the preprocessing of the patent's summary, the word vector is obtained through the embedding layer, and the size of each word vector is 64. For a patent text which has  $m$  words, the matrix input to the convolutional layer is  $m * 64$ . The length of the convolutional layer is 64, and the width is the number of words processed by the convolution kernel at a time. Kernel sizes = (2, 3, 4) is set here, and the convolution kernel can obtain different fields of view. The convolution layer is mainly for extracting local feature information. The column vector is obtained after the convolutional layer. In the pooling layer, the largest pooling is selected to extract a maximum value from each feature map of the upper layer, kernel num is 4, and each of the four feature maps are joined together to form a new column vector. It will be input to the fully connected layer via the relu activation function. The input of the fully connected layer is the product of the length of kernel num and kernel sizes, and the output is the number of categories obtained by clustering. The output result is processed by softmax, the probability value corresponding to each category is obtained, and the maximum value is taken out as the patent confidence that the patent text belongs to its corresponding category.

In order to ensure the validity of patent complementary computing, 50% patent texts are selected as training set and 50% patent data as test set. The parameters of TextCNN model are adjusted. Finally, the classification accuracy of the model reaches 87%, and the model is saved. All patents are classified under the model and got the classification probability.

**5.3. Patent Complementarity Analysis.** This paper proposes a patent complementary coefficient calculation method based on the patent technology tree. In the patent technology tree, two patents are calculated to count the number of nodes they pass together from top to bottom, and Equation (1) is used to get the patent complementary coefficient. Taking the patents of CN200920028726.0, CN201020173582.0, and C- N201110318971.7 as examples, which PCN (Patent Classification Number) are A01D13/00, A01D13/00 and A01C7/18, the patent complementarity coefficient is as follows Table 2. All patents are transformed into text vectors and clustered into three categories. Using the TextCNN for text classification, write down the classification probability of each patent with softmax. Taking patent CN201310333695.0 as an example, this is a patent for drum type pressurized multiring composite sealing ring which PCN is F16J15/16. The results of the patent classification model are as follows Table 3. Based on the above results, the patent confidence class is category 1, with a patent confidence of 0.92. According to Equation (3), three patents with the highest similarity with CN201310333695.0 are shown in Table 4 below. These patents are also classified as category 1. It can be seen from the similarity analysis that CN201310333698.4 has the strongest similarity with the target patent. From the patent text, the patent CN201310333695.0 is a drum type pressurized multiring composite sealing ring, CN201310333698.4 is a drum type pressurized multiring composite sealing plunger, CN200920026108.2 is a pharmaceutical shut-off valve, and CN200420052295.9 is a multifunction exhaust valve. Although several patents have a sealing function, CN201310333698.4 is more similar to the target patent, so the data analysis result is correct.

Patent CN200920240814.7 is a bare electric protective cover for the high voltage part of dry type transformer, and the main classification number is H01F27/29. The patent confidence class is category 0. According to Equations (5) and (6), the three patents with the highest patent complementarity are found in the following Table 5. From the results of the patent complementarity analysis, CN201220235178.0 has the strongest patent complementarity with the target patent. From the patent text, CN201220235178.0 is the internal oil separator for the transformer body, CN200920019747.6 is the transformer small radial coil electrostatic ring, and CN201320108624.6 is a new crimping current transformer connection structure. CN201220235178.0 and the target patents are different parts of the transformer. They are also in the same section from the main classification number. Therefore, the two patents

TABLE 4: Similarity analysis.

Patent number	CN201310333698.4	CN200920026108.2	CN200420052295.9
PCN	F16J9/00	F16K1/00	F16K1/00
Confidence class	1	1	1
Patent similarity	1.76	1.24	1.12
Patent complementarity	0.12	0.38	0.44

TABLE 5: Complementarity analysis.

Patent number	CN201220235178.0	CN200920019747.6	CN201320108624.6
PCN	H01F27/30	H01F27/34	H01F38/30
Confidence class	1	1	2
Patent complementarity	1.72	1.34	1.15
Patent similarity	0.14	0.33	0.425

are different technical details of the same technology product, so the patent complementarity is stronger. The above analysis shows that the results of patent complementarity analysis are correct.

Patent CN200920240814.7 is a patented product of Shandong Luneng Mount Tai Electric Equipment Co., Ltd. For this patent, the average value of all enterprise patents and their patent complementarity is calculated. The company with the strongest patent complementarity is Shandong Taikai Prefabricated Substation Change Co., Ltd. Both are local enterprises in Tai'an, and they are engaged in different technical studies of transformers in terms of power equipment.

## 6. Conclusion

This topic uses a text processing model based on deep learning and turns the patent text into quantitative patent complementarity and similarity calculation. It proposes a new patent complementarity measurement model, which has technical support for the confirmation of complementary enterprises. Better promote the technical cooperation and exchange of enterprises and produce more technological achievements.

There are three important promotion in the patent complementarity analysis of enterprises based on patent text. First is text representation model. Patent texts preprocessed are processed into word vectors through embedding layer. Sentence vectors are input into LSTM model to represent patent text replacing word bag representation which ignores the semantics of sentences. Second is text categorization model. Because the crawled patents are not classified, the patent complementarity calculation method proposed in this paper needs the confidence that the patents belong to their corresponding categories. Therefore, the generated sentence vectors are labeled with different categories for patent information by Kmeans clustering algorithm. The patent word vectors are input into the text classification model of TextCNN to obtain the confidence that patents belong to

their corresponding categories. This method improves the previous LDA topic clustering, only the probability of keywords in classification can be obtained, and the classification probability of the whole patent in different categories can be obtained. Third is the calculation of patent complementarity. The patent confidence and patent complementarity coefficient are calculated to obtain the patent complementarity. The patent complementarity of all patents in one company means the patent complementarity between the company and the given company. The company with the largest patent complementarity is the complementary company sought. This method connects and differentiates the text category and patent main classification number and then proposes the probability of patent similarity and patent complementarity.

There are still many shortcomings in the research process of this topic, which need to be further improved and perfected. Due to the limited experimental data, the trained model effect may not be applicable to more patents. Therefore, the next step will further increase the number of patents to achieve better training effect. In the representation of sentence vectors, LSTM alone cannot achieve good results, so attention mechanism can be considered to improve the model. Considering text classification, CNN can extract important local information, but it has no good effect in dealing with long-distance dependence problem. Although RNN is suitable for dealing with long-distance text, it often fails to capture local important text effectively. Therefore, to solve this problem, the new model with fusion mechanism using CNN and RNN can get important local text and effectively deal with the long-distance text. And the classification accuracy of TextCNN model for text classification can be further improved to enhance the validity of patent confidence calculation and enhance the accuracy of patent complementary calculation.

## Data Availability

[https://github.com/Eronwin/Patent\\_analysis](https://github.com/Eronwin/Patent_analysis)

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was funded by top talent training plan for young teachers in undergraduate teaching of Shandong University of Science and Technology (no. bjrc20190509).

## References

- [1] R. E. Hoskisson and D. Yiu, "The dynamics of knowledge regimes: technology, culture and competitiveness in the USA and Japan," *Asia Pacific Journal of Management*, vol. 20, no. 2, pp. 283–286, 2003.
- [2] M. Makri, M. A. Hitt, and P. J. Lane, "Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions," *Strategic Management Journal*, vol. 31, no. 6, pp. 602–628, 2010.
- [3] X. Wei, *Overseas M & A Integration Risk of Chinese Enterprises*. PhD thesis, Zhejiang University, 2015.
- [4] D. Zhang, "Research on the measurement of the relatedness of patent technologies for patent integration," *Journal of Intelligence*, vol. 33, no. 11, pp. 54–61, 2014.
- [5] R. T. H. Li and B. Xu, "Patent design around method based on AFD for umbrella-type patent strategy," *Journal of Mechanical Engineering*, vol. 52, no. 5, pp. 1–11, 2016.
- [6] M. Y. Wang, "Exploring potential R&D collaborators with complementary technologies: the case of biosensors," *Technological Forecasting and Social Change*, vol. 79, no. 5, pp. 862–874, 2012.
- [7] G. M. Salton, A. Wong, and C. S. A. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1975.
- [8] G. E. Hinton, *Learning Distributed Representations of Concepts*, Eighth Conference of the Cognitive Science Society, 1989.
- [9] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *Computer Science*, 2013.
- [10] T. Mikolov, I. Sutskever, C. Kai, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in Neural Information Processing Systems*, vol. 26, pp. 3111–3119, 2013.
- [11] P. Wang, Q. Yao, F. K. Soong, H. Lei, and Z. Hai, "Part-of-speech tagging with bidirectional long short-term memory recurrent neural network," *Computer Science*, 2015.
- [12] N. Pappas and A. Popescu-Belis, "Multilingual hierarchical attention networks for document classification," <https://arxiv.org/abs/1707.00896>, 2017.
- [13] D. D. Lewis, "Naive (Bayes) at forty: the independence assumption in information retrieval," in *European Conference on Machine Learning*, Springer, Berlin, Heidelberg, 1998.
- [14] Y. Zheng and Y. Liu, "Support vector machine solves multi-classification problem," *Computer Engineering and Applications*, vol. 41, no. 23, pp. 190–192, 2005.
- [15] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," <https://arxiv.org/abs/1404.2188>, 2014.
- [16] Y. Kim, "Convolutional neural networks for sentence classification," CoRR, abs/1408.5882, 2014.
- [17] A. Conneau, H. Schwenk, L. Barrault, and Y. Lecun, "Very deep convolutional networks for text classification," <https://arxiv.org/abs/1606.01781>, 2016.