Design and Implementation of an Intelligent Legal Affairs Platform Based on Machine Learning

Xu Ding(✉)

University of Belarusian State University, Minsk, Belarus
guang429710356@qq.com

Abstract. The legal affairs platform is the most important mechanism for reflecting fairness and justice in our market economic system. It often plays a role in supervising the implementation of laws, the business legal norms of enterprises, the judicial effectiveness of official documents, and the specific process of legal affairs. In the past, there were problems of high risk and low efficiency of human resources in the process of handling legal affairs. Based on the analysis and research of the existing legal affairs management methods and business processes in enterprises, this article adopts the Word2vec recommendation model and MVC three-level structure in natural language processing to conduct research on the design of an intelligent legal affairs platform based on machine learning. It mainly includes: CBOW algorithm optimization, personalized recommendation authorization, dispute management, etc. Finally, based on the optimization method proposed above, an intelligent legal affairs platform based on machine learning was designed and implemented. After experimental verification, the optimized system improved the problems existing in previous legal affairs processing.

Keywords: machine learning · legal affairs · natural language processing

1 Introduction

Currently, with the rapid development of economic globalization, the concepts of the rule of law economy and contract economy are deeply embedded in the development process of the market economy. As an important component of the market economy, enterprises need to build a good legal environment for business operations and maintain the order of the market competition environment. As an important constraint tool for enterprise activities, contracts are the basic conditions for economic activities between enterprises, and also an important embodiment of enterprises' compliance with the laws of market economy [1].

Currently, there are many problems in the legal management platform of private companies, such as: parallel use of multiple systems, lack of unified standards such as templates and permission control, poor user experience, and single functionality. The technological advantages and natural expansion of artificial intelligence have led to the

P. Qi and Z. Chen (Eds.): ICBDIE 2023, AISR 178, pp. 536–541, 2024.
https://doi.org/10.2991/978-94-6463-238-5_74

profound reshaping of human civilization by this technology, with the continuous development of machine learning algorithms having a profound and significant impact on the traditional legal industry. In order to improve the competitiveness of the enterprise market and regulate the corresponding legal transaction processing behavior, it is necessary to introduce an intelligent legal transaction processing platform based on machine learning for management [2].

2 Word2vec Algorithm Architecture

In the context of the study of neural network language models, Word2vec is a machine learning model used to train word vectors. Word2vec is a training and optimization process based on a neural network language model. The main process includes: (1) Statistical frequency of all words in the corpus is a prerequisite for training data. Because adding words with very small word frequencies to the dictionary not only requires a large amount of memory space, but also reduces the speed of training [3]. At the same time, these words with very small word frequencies will reduce the accuracy of the trained word vector due to insufficient training. Therefore, Word2vec will remove these noises, that is, ignore words with very small word frequencies in the corpus [4]. So that is word2vec.

(2) The addition of word vectors as input to the input layer is different from the original way of arranging the word vectors of input words in order, which will reduce the dimension of the input vector.

$$p(w_{t-n+1}, w_{t-n+2}, \ldots\ldots w_{t-1}) = \frac{e^{yw_t}}{\sum_i e^y} \tag{1}$$

(3) Remove the hidden layer, that is, after summing the vectors of the input layer, connect it directly to the output layer. This does not affect the training effect, but reduces the amount of calculation during training. As shown in Fig. 1:

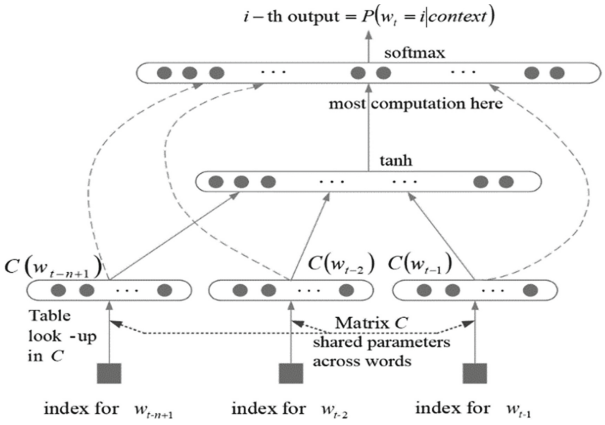

Fig. 1. Word2vec algorithm flow

(4) The output layer abandons the Huffman tree and adopts Softmax based on the Huffman tree, which uses Hoffman encoding to encode and represent each word in the dictionary. (5) In order to improve the training speed and reduce the computational complexity of model training, a negative sampling algorithm is introduced.

$$L = \frac{1}{T}\sum_{t} logf\left(w_{t-n+1}, w_{t-n+1}, ..., w_{t-1}; \theta\right) + R(\theta) \tag{2}$$

A negative sample is a method of constructing words that do not exist in the corpus by replacing the central word of a word sequence with another word. During the weight update process, randomly select some negative samples from all negative samples to update. Using this method to train, the smaller the probability of negative samples appearing, the better the training effect [5].

3 Legal Transaction Data and Experiments Based on Word2vec Algorithm

3.1 Experimental Parameters and Steps

First, create corresponding data tables in the database. The structured attributes contained in several corporate legal affairs related data obtained from major websites, including legal person names, business scope, employment qualifications, credit records, business hours, ownership change records, and other information, are used to construct a basic information table for legal affairs processing, Table-a. Based on the structured attributes in the questions raised by consumers and enterprises, including unit price, purchase price, contract signing, fund use records Basic information table such as repayment records Table-b. The corresponding information table is shown in Table 1.

Table 1. Information Table

Table-a	Legal person name, business scope, employment qualification, credit record, business hours, ownership change record
Table-b	Unit price, purchase price, contract signing, fund usage records, repayment records

Table 2. Rule Table

group	id	password	level
Table-1	1	123	master
Table-2	2	234	master
Table-3	3	345	master
Table-4	4	456	master

Table 3. Precision Comparison of Legal Affairs Recommendation Models

network structure	FPS	Size	Flops
FPN+GAM	89.2	83.2 MB	54
FPN+GAM+CAM	79.48	23.1 MB	103
MobileNet-A0+GAM+CAM	78.12	10.1 MB	121

3.2 Prediction Process and Verification

To verify the effectiveness of recommendations for legal affairs platforms based on machine learning algorithms, three sets of ablation experiments were designed on the ICDAR2015 dataset for different sequential structures. The experimental comparison results are shown in Table 3. Firstly, keeping the model backbone network unchanged, based on using the FPN network as the backbone network, a depth feature extraction module (GAM) is added to verify the effectiveness of the FPN+GAM structure. On the basis of the FPN+GAM structure, a channel feature fusion module (CAM) is added, and experiments are added to verify the effectiveness of the FPN+GAM+CAM structure. Finally, the effectiveness of the MobileNet-A0+GAM+CAM structure is verified [6] (Table 2).

3.3 TextRank Algorithm

The TextRank algorithm is proposed on the basis of the PageRank algorithm, and is a extractive unsupervised summary method that transforms the analysis of text into a network graph pattern (Fig. 2).

This allows you to determine the importance of nodes by analyzing the weight of each node in the network graph. Think of each sentence in a text as a node. If there is similarity between two sentences, there is an undirected weighted edge between the two

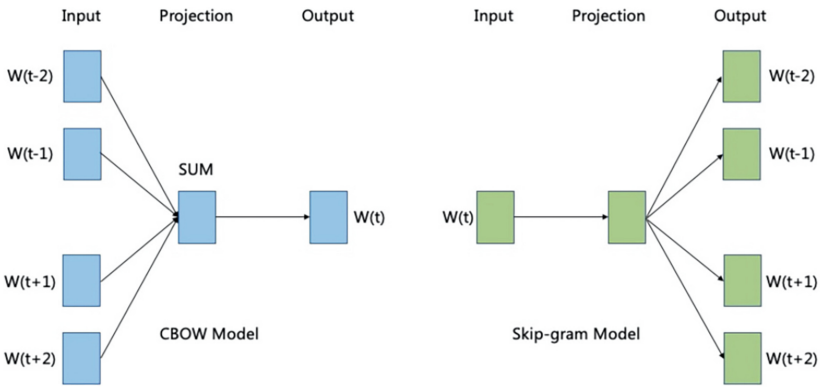

Fig. 2. Model Wording Method

sentences. The calculation formula for aggregation similarity is:

$$Simil(s_i, s_j) = \frac{\left|\{w_k | w_k \in s_i \text{ and } w_k \in s_j\}\right|}{log(|s_i|) + log(|s_j|)} \tag{3}$$

The TextRank algorithm is equivalent to a sorting algorithm that can divide patent text into several units, construct a connection graph through sentence nodes, and calculate the TextRank value of a sentence through cyclic iteration using similarity [7]. The calculation matrix is:

$$S_{n\times n} = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{pmatrix} \tag{4}$$

Find a vector representation for each segmented sentence, calculate the similarity between the sentences, and store it in a matrix. Then, calculate and sort it based on the similarity matrix and network diagram. Finally, the highest ranked n sentences are used as the final summary result.

4 Conclusions

Starting from maintaining the interests of the Group's legal responsibility, standardizing the Group's contract management business process, and improving the efficiency and standardization of the Group's contract formulation, the Group's legal affairs management system has improved and enhanced the functions of the original contract management system [8], expanded its scope of use, clarified its management functions, standardized its work processes and approval processes, in order to prevent contract risks, standardize economic behavior Reducing contract execution disputes provides strong support. The system is compatible with both PC and mobile terminals, fully improving the efficiency of management and service work [9].

The legal affairs management and legal risk assessment system designed in this article is mainly aimed at dealing with affairs personnel, legal business processes, and the handling of cases involved. After the system design is completed, all required processes will be used using electronic office management.

Having a legal affairs management and legal risk assessment system can effectively improve the work efficiency of lawyers, and can also improve the management level of lawyers. Based on the development background of W, I have developed a legal affairs management and legal risk assessment system. The purpose of designing this system is to better mobilize the work enthusiasm of kings through running this system, more effectively make legal affairs better, and free themselves from the complicated physical labor [10].

References

1. Tang Linan. Design and Implementation of Legal Affairs Management and Legal Risk Assessment System [D]. Dalian: Dalian Maritime University, 2016.

2. Hao Junfang. Problems and Optimization Paths in Enterprise Legal Affairs Management [J]. China Collective Economy, 2019 (12): 126-127.
3. Liu G, Yin YM, Liu W D, et al. Visualizing the Intellectual Structure and Evolution of Innovation Systems research: a Bibliometric Analysis[J]. Scientometrics, 2015(1):135-158.
4. Zhao Hongrui. Total Theory of World Civilization: The Rise of China's Civilization and the Principles of National Security and Rule of Law. Beijing: China Legal Publishing House, 2015.
5. Zhang Xudong. Blue Book of Artificial Intelligence: A Report on the Development of Artificial Intelligence in Medical Sciences in China. Beijing: Social Science Literature Press, 2020.
6. Office of The Ministry of Health. National Health and Family Planning Commission China Health Statistics Year-book 2018 [D]. Beijing: Beijing Union Medical University Press, 2018.
7. Jia Z J, Hong B, Chen D M, et al. China's Growing Contribution to Global Intracranial Aneurysm Research (1991–2012): A Bibliometric Study[J]. PLoSOne, 2014(3):91594.
8. Qin Qian, Ma Zhiguo. An empirical study on the protection and upgrading of regional intellectual property rights in China against the backdrop of economic and trade frictions between China and the United States [J]. China Science and Technology Forum, 2020 (2): 99-109.
9. Yu Fenglei, Zhang Ge. Research on the Protection Path of Big Data Intellectual Property Law from the Perspective of Trade Secrets. Guangxi Social Science, 2020(1): 99-104.
10. Wang Chao, Jin Yan. CPTPP Agreement and Intellectual Property Protection in China's Legal System [J]. Law Application, 2022 (2): 48-55.