A WPOS-GRU Patent Classification Method Based on Dual-Channel Feature Fusion (Postprint)

Authors: Yu Bengong, Zhang Peixing

Date: 2019-01-03T00:00:00+00:00

Abstract

To improve the efficiency and accuracy of automatic patent text classification, this paper proposes a dual-channel feature fusion WPOS-GRU (word2vec and part of speech gated recurrent unit) method for automatic patent text classification. First, patent abstract texts are acquired and undergo cleaning and preprocessing; next, word vector representation and part-of-speech tagging are performed on the patent texts, mapping them into word2vec word vector sequences and POS part-of-speech sequences respectively; finally, the WPOS-GRU model is trained using these two feature channels, and experimental analysis of the model' s performance is conducted. Compared with traditional patent classification methods and single-channel patent classification methods, the dual-channel feature fusion WPOS-GRU patent classification method enhances classification effectiveness. The proposed method saves substantial labor costs, improves the accuracy of patent text classification, and better satisfies the need for high-efficiency automation in large-scale patent text classification tasks.

Full Text

Preamble

Vol. 37 No. 3
Application Research of Computers
Accepted Paper

WPOS-GRU Patent Classification Method Based on Dual-Channel Feature Fusion

Yu Bengonga,b, Zhang Peihanga
a School of Management; b Key Laboratory of Process Optimization & Intelli-

gent Decision-Making of Ministry of Education, Hefei University of Technology, Hefei 230009, China

Abstract: To improve the efficiency and accuracy of automatic patent text classification, this paper proposes a dual-channel feature fusion WPOS-GRU (word2vec and part-of-speech gated recurrent unit) method for automatic patent text classification. First, patent summary texts are obtained, cleaned, and preprocessed. Then, word vector representations and part-of-speech tagging are performed, mapping patent texts into both word2vec vector sequences and POS tag sequences. Finally, a WPOS-GRU model is trained using the two feature channels, and experimental analysis is conducted on model performance. Compared with traditional patent classification methods and single-channel approaches, the dual-channel feature fusion WPOS-GRU method demonstrates improved classification effectiveness. The proposed method saves substantial labor costs, enhances the accuracy of patent text classification, and better meets the demands for automated, high-efficiency processing of large-scale patent classification tasks.

Keywords: patent classification; part-of-speech tagging; feature fusion; GRU

0 Introduction

In recent years, technological innovation has attracted increasing attention, and patents, as crucial carriers of innovation records, have experienced explosive growth [1]. In the patent application domain alone, China accepted 1.3 million patent applications in 2016, with numbers continuing to rise annually. Manual classification of such massive patent data requires enormous labor costs, and processing efficiency cannot meet practical needs. Consequently, research on automatic patent classification methods has become increasingly important and represents a significant current research hotspot.

Most patent classification research adopts the IPC classification system, a hierarchical structure comprising sections, classes, subclasses, and groups that is widely recognized and used internationally. Researchers currently address automatic patent classification through machine learning, performing text analysis to extract key feature words and combining them with machine learning classifiers to achieve effective results. In recent years, deep learning has demonstrated excellent performance in natural language processing, and its end-to-end workflow better satisfies the requirements for automatic patent classification. Applying deep learning models to achieve automatic patent classification represents a promising solution.

1 Related Work

Patent texts represent an important manifestation of technological innovation, and their analysis and utilization have long been prioritized worldwide. Automatic patent classification is crucial for patent examination and retrieval. Using machine learning for patent classification has been a persistent research focus. Li Shengzhen et al. [2] segmented texts and extracted feature words, mapping patent texts into feature vectors and constructing classifiers using BP neural networks. Ma Fang [3] built classification models using radial basis function neural networks and designed an automatic patent classification system. Compared with ordinary texts, patent texts possess unique characteristics, and purpose-built classifiers better adapt to the requirements of automatic patent classification [4].

Qu Peng et al. [5] argued that patent texts exhibit distinct professional features, and using technical terminology to construct features can improve classification effectiveness, conducting experiments with Naive Bayes, SVM, and other classifiers to compare their performance. Vector space model-based classification methods ignore semantic information between words. Liao Liefa et al. [6] proposed replacing traditional vector space models with topics to incorporate semantic information when building classifiers.

The rapid development of deep learning in recent years has provided new solutions for many natural language processing problems, demonstrating strong performance particularly in text classification. Some scholars have used autoencoders to process features and extract deep-level information from texts [7], inspiring Ma Shuanggang [8] to apply autoencoders to automatic patent classification with favorable results. Current research on automatic patent classification has focused primarily on feature extraction and processing, while end-to-end deep neural networks can eliminate the constraints of feature engineering and are better suited for automatic classification of large patent datasets [9].

In deep learning models, word2vec word vector training is first performed to map words into low-dimensional vectors, solving the dimensionality problem of traditional bag-of-words models [10] and incorporating contextual information that captures semantic relationships, making it widely adopted in deep learning research. Kim et al. [11] used convolutional neural networks to build text classifiers that extract deep-level features without manual intervention, proving more efficient and effective than traditional feature engineering methods. Hu Jie et al. [12] used CNNs for feature extraction and random forests as classifiers for patent text classification, demonstrating improved performance compared with random forests alone.

Some researchers argue that CNNs' local connectivity characteristics cause them to ignore structural features of texts during feature extraction, whereas recurrent neural network models such as LSTM and GRU are sequence models better suited for text feature extraction [13, 14]. Wang Shuheng et al. [15] used bidirectional LSTM models for text sentiment classification, achieving better accuracy than CNNs. Li Xuelian et al. [16] compared LSTM and GRU model structures and performance, noting that GRU inherits LSTM' s automatic learning capability while having a simpler structure that significantly reduces training time, making it more suitable for large-scale text data research applications. Although deep learning methods have achieved excellent results in public opinion discovery and sentiment analysis, few researchers have applied them to automatic patent classification. This paper analyzes the characteristics of patent texts, extracts patent summary texts, and combines them with LSTM deep learning models to accomplish automatic classification of patent summaries. Furthermore, since each word has different part-of-speech tags that contain important semantic information [17]—particularly in patent texts where nouns generally carry higher importance—while word2vec ignores POS information, this paper combines word2vec vectors with POS information on the basis of GRU to achieve dual-channel modeling of semantics and POS features, proposing a dual-channel feature fusion WPOS-GRU method for automatic patent summary classification.

2 Key Technologies

2.1 Word2vec Training and POS Tagging

The word2vec model, developed by Mikolov et al., is a word vector representation tool comprising CBOW and Skip-gram models. This paper uses the CBOW model for word vector training. The CBOW model structure is shown in Figure 1 [Figure 1: see original paper], consisting of an input layer, projection layer, and output layer. The input layer uses a sliding window to input the n context word vectors of a target word into the model, outputting the vector representation of the current word. Because the word vector representation considers contextual information, the resulting vectors capture semantic relationships, enabling similarity computation through vector distance.

POS tagging is an important task in natural language processing, primarily including rule-based methods, statistical methods, and machine learning approaches. It is commonly used in machine translation, character recognition, and other domains to assign POS tags to each word as a foundation for subsequent semantic analysis. This paper uses the POS tagging standard from the Institute of Computing Technology, Chinese Academy of Sciences, which includes categories such as nouns, temporal words, location words, directional words, verbs, adjectives, etc. This standard provides detailed annotation information and is widely used in related research.

2.2 GRU Recurrent Neural Network Model

The GRU recurrent neural network model incorporates gating mechanisms into traditional RNNs, enabling retention of longer-distance memory. Compared with LSTM, GRU reduces the number of gates and merges the hidden state with the cell state, eliminating redundant information. Consequently, it achieves similar performance to LSTM while significantly improving computational efficiency due to fewer parameters. GRU includes an update gate and a reset gate—mechanisms for selectively passing information—each comprising a sigmoid neural network layer and vector dot product operation. Gate values range within

[0,1], where 1 indicates complete information passage and 0 indicates complete blockage. The GRU structure is shown in Figure 2 [Figure 2: see original paper] and Equation (1).

In the structure, z_t represents the update gate controlling the proportion of current input, R_t is the reset gate controlling which parts of the previous memory affect the current input, W_z, W_r, W_s represent weight matrices, and b_z, b_r, b_s represent bias terms. x_t denotes the input at time step t, s_t represents information to be updated, h_t denotes the hidden state at time step t, and σ represents the nonlinear activation function.

3 Automatic Classification Model Based on Dual-Channel Feature Fusion WPOS-GRU

This paper first preprocesses text data, including cleaning, segmentation, and stopword removal. Then, word vector representation and POS tagging are performed, with both word vectors and POS information fed into a dual-channel GRU recurrent neural network model for training. Finally, the trained model is used for text classification testing.

3.1 Patent Data Acquisition and Preprocessing

Patent data includes fields such as title, abstract, main body, and main classification number. The abstract text contains the core content of the patent, enabling readers to grasp its category through reading the abstract alone, whereas reading the full text incurs substantial time costs. Therefore, this paper extracts patent summary texts and main classification numbers.

Patent summary texts undergo data cleaning to remove noise from web sources, followed by segmentation and POS tagging using the Institute of Computing Technology standard. The segmented and tagged results are vectorized, mapping words and POS information into low-dimensional vectors separately.

3.2 Patent Text Classification Model

Patent summary texts exhibit more domain-specific and professional language compared with other text data, with extensive technical terminology. Traditional feature-word-based methods cannot adequately cover terminology, requiring feature vector redesign whenever new terms emerge. This paper proposes a recurrent neural network-based approach with stronger generalization capability, as new terms are often phrases composed of existing words whose word2vec semantics can be computed from constituent words. This eliminates manual feature extraction, accommodates the difficulty of feature word extraction from patent summaries, saves substantial labor costs, and better suits automatic classification of large patent datasets.

Considering that different POS tags carry varying importance in patent texts while word2vec ignores POS information, this paper combines word2vec vectors with POS information in a WPOS-GRU framework. POS information contains semantic content that can be input independently or as supplementary information to word2vec vectors. Therefore, this paper designs three model structures for comparative experiments: word2vec single-channel GRU, POS single-channel GRU, and dual-channel fusion WPOS-GRU.

3.2.1 Dual-Channel Feature Fusion WPOS-GRU Model The dual-channel feature fusion WPOS-GRU model comprises word2vec and POS channels, with its structure shown in Figure 3 [Figure 3: see original paper]. The model training process is as follows:

Dual-Channel Feature Fusion WPOS-GRU Algorithm Flow
Input: Training set $D = \{(x_k, p_k, y_k)\}_{k=1}^m$, hyperparameters: Θ
Initialization: Training parameters W

```
repeat for all (x_k, p_k, y_k)   D do
    4. Update parameter values
end for
until: training error converges to a sufficiently small value
Output: Classification network with determined parameters
```

Where D is the training set of m patent summaries, x_k represents the word vector sequence of the k-th patent summary, p_k represents the POS vector sequence of the k-th document, y_k is the category vector of the k-th document, and y'_k denotes the predicted category vector of the k-th text.

3.2.2 Single-Channel Feature GRU Model The single-channel feature GRU model contains only one channel—either word2vec or POS—with its structure shown in Figure 4 [Figure 4: see original paper]. Using the example sentence "保证加强板刚性结构，使其具有良好的承载能力和抗机械冲击能力"(Ensure the rigid structure of the reinforcement plate, giving it good load-bearing capacity and mechanical impact resistance), words like "保证"(ensure) and "加强"(reinforce) are mapped to vectors. In the word2vec single-channel GRU model, words are mapped to word2vec vectors; in the POS single-channel GRU model, words are mapped to POS tag vectors.

The GRU sequence layer feeds the mapped vector sequences into each time step of the first GRU layer according to word order, with each GRU layer' s output serving as input to the next layer. Only the final node' s output is retained from the last GRU layer. The fully connected layer assumes n categories in the patent dataset with n nodes, mapping the GRU sequence layer output to an n-dimensional vector. The softmax classification layer normalizes the previous layer' s output into a new n-dimensional vector where each element represents the probability of belonging to that category.

Word2vec Single-Channel Feature GRU Algorithm Flow
Input: Training set $D = \{(x_k, y_k)\}_{k=1}^m$, hyperparameters: Θ
Initialization: Training parameters W

```
repeat for all (x_k, y_k)   D do
    4. Update parameter values
end for
until: training error converges to a sufficiently small value
Output: Classification network with determined parameters
```

Where D is the training set of m patent summaries, x_k represents the word vector sequence of the k-th patent summary, y_k is the category vector of the k-th document, and y'_k denotes the predicted category vector of the k-th text.

POS Single-Channel Feature GRU Algorithm Flow
Input: Training set $D = \{(p_k, y_k)\}_{k=1}^{m}$, hyperparameters: Θ
Initialization: Training parameters W

```
repeat for all (p_k, y_k)   D do
    4. Update parameter values
end for
until: training error converges to a sufficiently small value
Output: Classification network with determined parameters
```

Where D is the training set of m patent summaries, x_k represents the POS vector sequence of the k-th patent summary, y_k is the category vector of the k-th document, and y'_k denotes the predicted category vector of the k-th text.

3.3 Model Evaluation

To design comparative experiments and evaluate the feasibility of the proposed method, this paper first splits all summary data into training and validation sets. After training models using the above algorithms, performance is evaluated using accuracy, precision, recall, and F1-score.

Accuracy refers to the ratio of correctly classified patent texts T to total patent texts N, as shown in Equation (2):

$$\text{Accuracy} = \frac{T}{N}$$

Precision refers to the proportion of patent texts predicted as belonging to a category that actually belong to that category, as shown in Equation (3), where TP represents documents correctly predicted as belonging to the category and FP represents documents incorrectly predicted as belonging to the category:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall refers to the proportion of patent documents actually belonging to a category that are correctly predicted, as shown in Equation (4), where TP represents documents correctly predicted as belonging to the category and FN

represents documents belonging to the category but incorrectly predicted:

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score balances precision and recall, as shown in Equation (5):

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

4 Experiments

4.1 Patent Data Acquisition and Preprocessing

For experiments, this paper selected the computer domain, which the authors are familiar with, following the patent selection method from reference [8]. First, the SooPAT patent search engine was used to search Chinese patents under the "computer" theme, limited to invention patents, identifying five main classification numbers: G06F1/16, G06F1/18, G06F1/20, G06F3/02, and G06F3/14 at the subgroup level. These five categories were then searched on the Shanghai Intellectual Property Platform. To ensure timeliness, patent data from the past four years was selected, with 2,000 patents downloaded per category. After screening and deduplication, patent summary texts were retained and stored by category.

Summary texts underwent segmentation and stopword removal using the jieba segmentation tool, which demonstrates good performance and is widely used in related research. Since word2vec training requires large corpora, Wikipedia texts and patent texts were combined, with the Wikipedia corpus being publicly available online at approximately 1.3 GB and the patent dataset containing 10,000 entries. This combination satisfies large corpus requirements while incorporating domain information to ensure effective word vector training. Finally, the word2vec model mapped words in the summary texts into 100-dimensional vectors.

POS tagging used the jieba toolkit with the Institute of Computing Technology standard, which is comprehensive and widely used in related research to support natural language processing tasks. POS tags were one-hot encoded into 50-dimensional vectors; for example, adjective "a" was encoded as [0,0,1,0,0,⋯] with "a" corresponding to the third position in the vector.

4.2 Model Training

The experimental environment configuration is shown in Table 1 . The GRU recurrent neural network sequence length matches sentence length. GRU node counts of 50, 100, and 150 achieved high F1-scores, but since patent text classification requires high efficiency, this paper set the node count to 50.

The 10,000 patent texts across five categories were split into training and validation sets, with 8,000 training samples and 2,000 validation samples. Cross-entropy was used as the loss function and Adam as the optimizer, with model convergence achieved after three training epochs.

To verify the proposed method' s effectiveness, three models were trained and compared with commonly used patent classification methods: NB (Naive Bayes), SVM (Support Vector Machine), NN (Neural Network), and RF (Random Forest). Experimental results are shown in Table 2 and Figure 5 [Figure 5: see original paper]. Additionally, to verify the varying importance of different POS tags, word2vec single-channel GRU models were trained using only nouns, only adjectives, and only verbs for evaluation.

4.3 Experimental Results Analysis

Comparative experiments reveal that among traditional machine learning models, neural networks perform best with classification accuracy of 0.92. Among the three methods proposed in this paper, word2vec single-channel GRU achieves classification accuracy of 0.95, representing significant improvement over traditional methods and demonstrating the important value of applying deep learning to patent classification. Training single-channel GRU models using only nouns achieves accuracy of 0.91, using only verbs achieves 0.81, while using only adjectives achieves only 0.53, indicating that different POS tags contribute differently to classification, with nouns containing the most information. Furthermore, POS single-channel GRU achieves classification accuracy of 0.46, substantially higher than the random baseline of 0.20 for five-class classification, demonstrating that POS features contain substantial semantic information. The dual-channel fusion model achieves classification accuracy of 0.974, successfully accomplishing automatic patent text classification tasks. The dual-channel feature fusion WPOS-GRU also demonstrates superior F1-score performance, showing substantial improvement over both traditional and single-channel methods.

5 Conclusion

As an important manifestation of technological innovation, patent text classification plays a crucial role in patent analysis and utilization. This paper addresses the problem of automatic and efficient classification of large-scale patent texts by proposing a dual-channel feature fusion WPOS-GRU automatic classification model. By incorporating POS semantic information, the model improves classification accuracy and makes automatic patent classification more reliable and practical. However, limitations remain, such as suboptimal performance for emerging categories with insufficient patent text data.

References

[1] WIPO. World intellectual property report 2017 [EB/OL]. (2017-12-06). http://www.wipo.int/edocs/pubdocs/en/wipo_pub_944_2017.pdf.

[2] Li Shengzhen, Wang Jianxin, Qi Jiandong, et al. Automated categorization of patent based on back-propagation network [J]. Computer Engineering and Design, 2010, 31(23): 5075-5078.

[3] Ma Fang. Research of patent automatic classification based on RBFNN [J]. New Technology of Library and Information Service, 2011(12): 58-63.

[4] Liu Hongguang, Ma Shuanggang, Liu Guifeng. A review of research on patent document classification algorithms based on machine learning [J]. Library and Information Studies, 2016(3): 79-86.

[5] Qu Peng, Wang Huilin. Fundamental research questions in patent text categorization [J]. New Technology of Library and Information Service, 2013, 29(3): 38-44.

[6] Liao Liefa, Le Fugang, Zhu Yalan. The application of LDA model in patent text classification [J]. Journal of Modern Information, 2017, 37(3): 35-39.

[7] Liu Kan, Yuan Yunying. Short Texts Feature Extraction and Clustering Based on Auto-Encoder [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2015, 51(2): 282-288.

[8] Ma Shuanggang. Study of automatic Chinese patent classification based on deep learning theory and method [D]. Zhenjiang: Jiangsu University, 2016.

[9] Tajbakhsh N, Suzuki K. Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs [J]. Pattern Recognition, 2017, 63(3): 476-486.

[10] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space [J]. Computer Science, 2013.

[11] Kim Y. Convolutional neural networks for sentence classification [C]// Proc of Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2014: 1746-1751.

[12] Hu Jie, Li Shaobo, Yu Liya, et al. A patent classification model based on convolutional neural networks and random forest algorithm [J]. Science Technology and Engineering, 2018, 18(6): 268-272.

[13] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures [J]. Journal of International Neural Network Society, 2005, 18(5-6): 602.

[14] Chung J, Gulcehre C, Cho K, et al. Gated Feedback recurrent neural networks [J]. Computer Science, 2015: 2067-2075.

[15] Wang Shuheng, Turgun Ibrahim, Kahaerjiang Abiderexiti, et al. Sentiment classification of Uyghur text based on BLSTM [J]. Computer Engineering and Design, 2017, 38(10): 2879-2886.

[16] Li Xuelian, Duan Hong, Xu Mu. A Gated Recurrent Unit Neural Network for Chinese Word Segmentation [J]. Journal of Xiamen University, 2017, 56(2): 237-243.

[17] Tang Xiaoyan, Cao Jing. Automatic genre classification via n-grams of part-of-speech tags [J]. Procedia-Social and Behavioral Sciences, 2015, 198(7): 474-478.

Note: Figure translations are in progress. See original paper for figures.