Research on Deep Extraction of Financial Entity Relations based on GAT Fusion Model

Jiayu Hong*

School of Economics, Jinan University, Guangzhou, Guangdong, China

*Corresponding author: 3084353519@qq.com

Abstract

In response to the complexity of extracting entity relationships in Chinese financial texts, this study proposes a joint model (SynPOS-GAT) based on Graph Attention Network (GAT) that integrates dependency syntax and part-of-speech information. This model combines BERT for semantic encoding, utilizes BiLSTM to enhance the semantic representation of part-of-speech, and employs GAT to capture syntactic dependency structures, effectively addressing the problems of ambiguous entity boundaries and overlapping relationships in traditional methods. Experiments on the FinRE dataset show that the SynPOS-GAT model outperforms the comparison models in terms of accuracy, recall rate, and F1 value, with an F1 value of 44.78%, an average improvement of 10.08% compared to the baseline. Ablation experiments verified the crucial role of part-of-speech and syntactic information in the model performance. This study provides efficient technical support for the construction of financial knowledge graphs.

Keywords

Financial entities, relation extraction, graph attention, syntactic fusion.

1. Introduction

With the development of globalization and information technology, a large amount of unstructured text data has been generated in financial transactions. These pieces of information are crucial for financial analysis and decision-making, but traditional methods encounter challenges when extracting complex relationships. The core task of Chinese entity relation extraction is to accurately identify entities and their relationships in financial texts and convert them into structured triples (such as "China Eastern Airlines", "merger", "Shanghang Airlines"). However, the professionalism of financial texts and the diversity of relationships make this task more complex. Traditional methods have encountered problems such as ambiguous entity boundaries and poor performance when dealing with overlapping relationships.

In the research of Chinese entity relation extraction, the main techniques include template matching, feature engineering and deep learning. The feature fusion technology significantly improves the model's performance by integrating language features from different levels. Reference [1] integrates word embeddings, vocabulary and grammatical features to enhance the extraction accuracy; Reference [2] improves the extraction effect in multi-object situations through semantic role labeling and dependency syntactic analysis; Reference [3] proposes a bidirectional long short-term memory network model that improves the extraction performance through self-attention mechanism. In addition, the traditional pipeline method is prone to error propagation, while the joint entity relation extraction method solves this problem by sharing parameters, sequence labeling or graph structure methods. Reference [4] combines bidirectional encoder representations and label dependency knowledge to improve the relationship extraction effect; Reference [5] realizes the joint extraction of entities and relations through convolutional neural networks and BiLSTM; Reference [6] generates triples through sequence labeling. And the graph structure-based methods, Reference [7] proposes the graph convolutional network, which performs outstandingly in solving the problem of overlapping relations.

Through the above analysis, the existing methods mainly encounter the following challenges when dealing with Chinese financial texts: Firstly, the traditional pipeline approach treats entity recognition and relationship classification separately, resulting in cumulative errors; Secondly, most joint extraction methods do not fully utilize language features and syntactic structure information, making it difficult to accurately grasp the complex entity boundaries and diverse relationship expressions in Chinese financial texts. To address these issues, this paper proposes a dependency syntax and part-of-speech fusion model based on graph attention networks. By integrating the semantic representation capabilities of BERT, the capture of part-of-speech information by BiLSTM, and the modeling of syntactic dependency structures by GAT, this model achieves the ability to identify complex financial entities and relationships. This model not only can effectively handle complex situations such as ambiguous entity boundaries and overlapping relationships, but also can establish a bridge between entity recognition and relationship extraction through a fine-grained triad classifier, accurately identifying implicit relationships in financial texts, providing more reliable technical support for the construction of knowledge graphs in the financial field.

2. SynPOS-GAT Model

This paper proposes a joint entity relationship extraction model based on GAT,which integrates syntactic and lexical part-of-speech information.The model mainly consists of four modules:the BERT word embedding module,the BiLSTM part-of-speech semantic enhancement module,the GAT syntactic information fusion module,and the joint entity relationship extraction module.Firstly,BERT[8]is used to encode the input text and generate context-dependent vector representations for each word;secondly,the model incorporates the part-of-speech tagging results and integrates the part-of-speech information into the output of BERT through the part-of-speech embedding layer.To effectively integrate these pieces of information,this paper adopts BiLSTM as the core component for feature fusion.By combining the part-of-speech embeddings at each time step of BiLSTM,the model can further refine the syntactic attributes of each word and fuse them with the context information to generate enhanced word vector representations.Additionally,this paper constructs a graph structure based on the results of syntactic dependency analysis.GAT[9]can perform weighting on these dependency relations and identify the most critical syntactic structures for entity and relationship recognition.Based on feature extraction and graph structure analysis,the model employs a fine-grained triplet classifier to jointly extract entities and relationships.This classifier is based on a scoring mechanism,scoring all possible word pairs in the sentence with predefined relationships to determine whether they form a reasonable triplet.Finally,the model uses a decoding algorithm,based on the output of the classifier and index strategy,to recover the boundaries of entities and relationship types from the scoring matrix.

2.1. Data Preprocessing

This paper first uses a tokenizer to divide the continuous text into independent lexical units. Secondly, it employs the LTP tool developed by the Language Technology Platform of Harbin Institute of Technology [10] for part-of-speech tagging and syntactic dependency analysis. Among them, part-of-speech tagging classifies each lexical unit through a tagger and assigns corresponding grammatical attributes. Additionally, to integrate the boundary information of words, this paper adopts the BIO marking method to mark the part-of-speech results; for dependency syntactic analysis, the dependency tree identifies the relationships between words and constructs the dependency structure of the sentence, with nodes representing words and edges representing dependency relations.

2.2. Word Embedding Processing based on BERT

For a sentence sequence T, it is input into BERT, and its output is H, where T = [CLS , T1 , ... , Ti , ... , Tn] and H = [h0 , h1 , ... , hi , ... , hn]. H represents the word vector representation of the sequence T, and the calculation formula is:

$$H = \text{MultiHead}(T) \tag{1}$$

In the formula, MultiHead refers to the multi-head attention function of BERT.

In the input encoding stage of BERT, pre-trained word embeddings are first utilized, followed by position encoding. The calculation formula is as follows.

$$PE(pos, 2i) = \sin(pos/10\,000^{2i/d}) \tag{2}$$

$$PE(pos, 2i+1) = \cos(pos/10\,000^{2i/d}) \tag{3}$$

In the formula: pos represents the position of the word in the sequence; i represents the dimension range of the word vector; d represents the embedding dimension.

In the attention module of BERT, the attention mechanism generates the output by calculating the weighted sum of the query and a set of key-value pairs. Specifically,

$$\text{Attention}\,(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{\boldsymbol{d}_k}}\right)\boldsymbol{V} \tag{4}$$

In the formula, Q, K, and V represent query, key, and value respectively.

Furthermore, each encoding layer also includes a fully connected feedforward network, consisting of two linear layers and a ReLU activation function. The calculation formula is as follows:

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{5}$$

In the formula: x is the input vector; FFN(x) is the feedforward neural network function; W1 and W2 are the weight matrices of two linear transformations, respectively used for dimension expansion and reduction; b1 and b2 are the corresponding bias vectors.

2.3. Word Part-of-Speech Semantic Enhancement based on BiLSTM

During the data preprocessing stage, first, the part-of-speech of each word is annotated. Then, these discrete labels are transformed into continuous vectors through the part-of-speech embedding layer. Finally, these part-of-speech vectors are concatenated with the semantic vectors output by BERT in the concatenation layer to form a comprehensive feature representation, which is further processed by the BiLSTM layer.

Specifically, first, H = [h0, h1, ..., hi, ..., h_n] represents the sequence of word vectors output by BERT, and P = [p0, p1, ..., pi, ..., p_n] represents the sequence of part-of-speech labels generated by the sentence. Firstly, the part-of-speech labels are transformed into part-of-speech vectors V = [vp0, vp1, ..., vpt, ..., vpn] through an embedding layer. Then, they are concatenated with the word vectors output by BERT as [H;V] = [h0 || vp0, h1 || vp1, ..., ht || vpt, ..., h_n || vpn], which serves as the input for BiLSTM. The output result S is the new feature representation obtained after processing by BiLSTM, and the calculation formula is:

$$S = \mathrm{BiLSTM}[H;V] \tag{6}$$

Since BiLSTM is composed of two layers of LSTM neural networks in the forward and backward directions,it passes information in opposite directions to the two LSTMs with the same input information.The two LSTMs operate in chronological order and in reverse chronological order respectively,and at time t,the hidden states are defined as pt and qt,where pt is called the forward LSTM and qt is called the backward LSTM.The calculation formula is:

$$p_t = f(U^{(1)} h^{1}_{t-1} + W^{(1)} x_t + b^{(1)}) \tag{7}$$

$$q_t = f(U^{(2)} h^{2}_{t+1} + W^{(2)} x_t + b^{(2)}) \tag{8}$$

$$h_t = p_t \oplus q_t \tag{9}$$

In the formula: U represents the hidden state transition matrix; W represents the input transition matrix; b represents the bias vector; ⊕ represents the vector concatenation operation; h_t represents the bidirectional LSTM, and here h_t is the output result S described above, which is St.

2.4. Semantic Structure Information Fusion based on GAT

Using GAT, the dependency relationships between words in the sentence are encoded. GAT is then used to weight the nodes (words) in the graph, highlighting the words that are more relevant to the entity and relation extraction tasks. Through the attention mechanism, the model can adaptively learn the relative importance of different words, enhancing the understanding of the sentence structure.

Specifically, first, S = [s0, s1, ..., st, ..., sn], where si ∈ RF, represents the sentence representation after being processed by BiLSTM. Each si is the feature vector of the i-th word. The input layer of GAT is a set of node feature vectors, which is S, where n is the number of nodes, F is the number of node features, and R represents a certain node feature. The output layer is a new set of node feature vectors S = [s0, s1, ..., st, ..., sn], where si ∈ RF', where F' represents the new feature vector dimension.

GAT predicts the features of new nodes based on the features of input nodes for n nodes. To obtain the corresponding input and output features, at least one linear transformation needs to be performed based on the input features. Therefore, a weight matrix W needs to be trained for all nodes, which represents the relationship between the F input features and the F' output features. For each node i, we define an attention mechanism aij to measure the importance of node j to node i. The attention coefficient is calculated as follows:

$$a_{ij} = \frac{\exp(\mathrm{LeakyRelu}(\boldsymbol{W} \cdot [\boldsymbol{s}_i \parallel \boldsymbol{s}_j]))}{\sum_{k \in n_i} \exp(\mathrm{LeakyRelu}(\boldsymbol{W} \cdot [\boldsymbol{s}_i \parallel \boldsymbol{s}_k]))} \tag{10}$$

In the formula: si || sj represents vector concatenation; LeakyRelu is the activation function. The new feature s of each node is obtained by weighting and summing the features of neighboring nodes and the corresponding attention coefficients.

$$s_i' = \sigma\left(\sum_{j \in n(i)} a_{ij} \cdot s_j \right) \tag{11}$$

In the formula, σ represents the activation function. To enhance the model's expressive power, this study employs the multi-head attention mechanism. For each head k, the independent attention coefficient aik) and the updated feature s(k) are calculated, and the final feature s is the concatenation of all the head outputs.

$$\overline{s_i} = \|_{k=1}^{K} s_i^{(k)} \tag{12}$$

2.5. Joint Entity Relationship Extraction

After obtaining the semantic representation that integrates multiple pieces of information, the model in this paper adopts the joint extraction method proposed by OneRel to identify entities and relations. The framework adopts a relation-specific angle label strategy, which can accurately mark the entity boundaries. For each pair of words wi and wj in the sentence, as well as each predefined relation, the score of the triplet (wi, rk, wj) is calculated. Based on the scoring matrix and label strategy, the probability matrix is traversed, and through rule decoding, all entities and relations are extracted to construct the final triplet list.

The relation-specific angle label strategy identifies the boundaries of entity pairs through specific labels, thereby accurately extracting entities and relations during the decoding stage. The label design takes into account complex situations such as overlap and nesting of entity pairs, and mainly includes the following contents. HB-TB (head begin-tail begin): Indicates that the beginning positions of the head entity and the tail entity occur simultaneously. HB-TE (head begin-tail end): Indicates that the beginning of the head entity and the end of the tail entity occur simultaneously. HE-TE (head end-tail end): Indicates that the end positions of the head entity and the end position of the tail entity occur simultaneously. The specific relation-specific angle label strategy corresponding to the examples is shown in Figure 1.

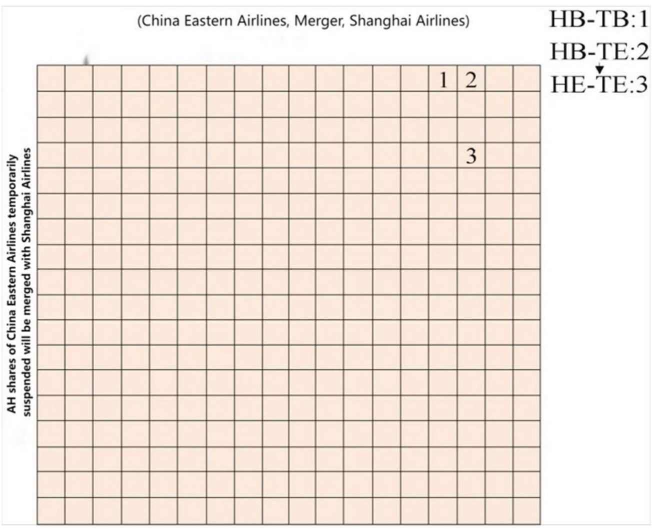

Figure 1. Time-frequency feature images after feature enhancement.

In the output layer of the model, for each possible pair of entities and each predefined relationship, the model outputs a score matrix. The model uses a score-based classifier to evaluate whether the triplet (wi, rk, wj) formed by any two words wi, wj and the predefined relationship rk is valid.

$$\mathrm{score}(w_i, r_k, w_j) = \boldsymbol{v}^{\mathrm{T}} \cdot \boldsymbol{\phi}(\boldsymbol{W}^{\mathrm{score}} \cdot [\overline{\boldsymbol{s}_i} \| \overline{\boldsymbol{s}_j} \| \boldsymbol{E}_{\mathrm{REL},\, k}]) + \boldsymbol{b}^{\mathrm{score}} \tag{13}$$

In the formula: $\boldsymbol{W}^{\mathrm{score}}$ and $\boldsymbol{b}^{\mathrm{score}}$ represent the learnable parameters; $[\overline{\boldsymbol{s}_i} \| \overline{\boldsymbol{s}_j} \| \boldsymbol{E}_{\mathrm{REL},\, k}]$ represents the concatenation of $\overline{\boldsymbol{s}_i}$ and $\overline{\boldsymbol{s}_j}$ and their relationship embedded in $\boldsymbol{E}_{\mathrm{REL},\, k}$; $\boldsymbol{\phi}$ represents the nonlinear activation function; $\boldsymbol{v}$ represents the weight vector used for scoring. Subsequently, the cross-entropy loss function is used, and the scores are converted into probabilities through softmax to represent the probability that each triplet belongs to each relationship category. The calculation formula is:

$$\boldsymbol{P}(y \mid w_i, r_k, w_j) = \frac{\exp(\mathrm{score}(w_i, r_k, w_j))}{\sum_{y'} \exp(\mathrm{score}(w_i, r_k, w_j) y')} \tag{14}$$

In the formula, y' represents all possible labels, including valid relationship labels and the invalid label "-".

The decoding process first determines the entity boundaries.It then traverses the scoring matrix based on the specific index strategy for identifying specific patterns within the scoring matrix.For each relation rk,it finds the pair of words wi and wj with the highest score,representing the possible entity boundaries.Secondly,based on these determined positions and entity boundary labels,entities and relations are extracted.Finally,the extracted results are processed.Specifically,triplets with different positions but the same entities and relations are merged,problems of entity overlap or nesting are resolved,and low-confidence extraction results are filtered to optimize the output.

3. Experiments and Result Analysis

3.1. Data Set

This paper selects the financial news relation extraction dataset FinRE from CFLEB.The dataset contains 7,454 training data,1,489 validation data and 3,727 test data.The task is to classify the head-entity-tail-entity pairs into 44 relation categories,and the evaluation metric is F1-Score.

3.2. Parameter Settings and Evaluation Indicators

To verify the accuracy of the proposed model in extracting Chinese entity relationships in the financial field, this paper uses precision, recall rate and F1 value as evaluation indicators.

3.3. Experimental Results

To verify the performance of the proposed model, this paper selected several representative models for comparative experiments. The comparison models include OneRel, GlobalPointer and CASREL, which respectively employ different methods for extracting entity relationships. All models were evaluated under the same dataset and with the same hyperparameters. The experimental results are shown in Table 1.

Table 1. Financial relation extraction accuracy, recall, and F1 value.

Model	Accuracy rate	Recall rate	F1-score
CasRel	15.78	47.96	23.75
GlobalPointer	45.73	34.92	39.60
OneRel	40.63	40.90	40.76
SyPOS-GAT	45.32	44.25	44.78

3.4. Result Analysis

The SynPOS-GAT model performs the best in terms of accuracy,recall rate and F1 value.Among them,the F1 value reaches 44.78%,and it is on average 10.08%ahead of other models.The model proposed in this paper maintains a high accuracy rate while also having a good recall rate,and achieves a relatively high F1 score overall.This performance improvement stems from the combination of part-of-speech tagging and syntactic dependency analysis(part-of-speech tagging helps distinguish the contextual meanings of polysemous words,and syntactic analysis reveals the dependencies between entities).Moreover,jointly extracting entities and relations effectively avoids the cascading errors in traditional methods and improves the recognition accuracy.

3.5. Ablation Experiment

To further verify the impact of each component on the model's performance,this paper conducted ablation experiments.In the experiments,the part-of-speech information,syntactic information,and their combination were removed from the model respectively.The results showed that the OneRel+VocabPos model outperformed the baseline model OneRel by 2.44%,1.07%,and 1.75%in terms of accuracy,recall rate,and F1 value,respectively.This indicates that part-of-speech information helps the model more accurately identify the boundaries of entities.The OneRel+Syntactic information model outperformed the baseline model OneRel by 1.86%,4.44%,and 3.11%in terms of accuracy,recall rate,and F1 value,respectively.This shows that syntactic information enhances the model's understanding of sentence structure.The SynPOS-GAT model outperformed the baseline model OneRel by 4.69%,3.35%,and 4.02%in terms of accuracy,recall rate,and F1 value,respectively.This indicates that after removing any one module,the precision,recall rate,and F1 value of the model decreased. The specific results are shown in Table 2.

Table 2. The results of the ablation experiment comparison test.

Model	Accuracy rate	Recall rate	F1-score
OneRel	40.63	40.90	40.76
OneRel+VocabPos	43.07	41.97	42.51
OneRel+Syntatic information	42.49	45.34	43.87
SynPOS-GAT	45.32	44.25	44.78

4. Conclusion

This study focuses on the task of extracting entity relationships from Chinese financial texts and proposes a joint model (SynPOS-GAT) based on Graph Attention Network (GAT) that integrates dependency syntax and part-of-speech information. This model effectively addresses the problems of ambiguous entity boundaries and overlapping relationships in traditional methods by integrating the semantic representation capabilities of BERT, the capture of part-of-speech information by BiLSTM, and the modeling of syntactic dependency structures by GAT. Experimental results show that the F1 value of SynPOS-GAT on the FinRE dataset reaches 44.78%, significantly outperforming the comparison models, with an average improvement of over 10%. Ablation experiments further verified the contribution of part-of-speech information and syntactic information to the model performance, and the model performed best after their fusion. This study provides more reliable technical support for the construction of financial knowledge graphs, and future research can explore the utilization of more language features and structural information to further improve the model performance.

References

[1] Shiraishi Y ,Kaneiwa K .A self-matching training method with annotation embedding models for ontology subsumption prediction[J].International Journal of Data Science and Analytics,2026,22(1):86-86.

[2] Nguyen T D H ,Le B .Faithful explanation of semantic role labelling with dependency and constituency feature importance[J].Neurocomputing,2026,676132996-132996.

[3] Suthakar G R A ,Sandesh J B .Enhanced image captioning with positional and dual attention using deep convolutional long short-term memory and emotional feedback mechanism[J].Multimedia Tools and Applications,2026,85(3):228-228.

[4] Baek W J ,Chung K .Scene graph optimization using hyperedge-based relationship extraction[J].Applied Soft Computing,2026,194114907-114907.

[5] Zhao Y ,Li Q ,Liu Y , et al.A multi-layer deep learning framework for joint extraction of geological disaster entities and relations[J].Geomatics, Natural Hazards and Risk,2025,16(1).

[6] Chen J K ,Chen Y ,Wang Z , et al. GMN-Zoomer: Learning graph similarity via hierarchical parsing, pooling and matching[J].Neural Networks,2026,200108778-108778.

[7] Xie Y ,Zheng X ,Xie S , et al. Fault prediction for aluminum electrolysis based on spatial-temporal attention dynamic graph convolutional networks[J].Control Engineering Practice,2026,172106878-106878.

[8] Zhou L ,Bai L .Unsupervised temporal knowledge graph entity alignment via BERT and structure adaptation[J].Expert Systems With Applications,2026,315131860-131860.

[9] Tong K ,Gan H Z ,Wang J S , et al.Prelimbic cortex to zona incerta pathway mediates recognition memory recovery delay through down-regulating astrocytic GAT-3[J].iScience,2026,29(3):115078-115078.

[10] Assis B V J ,Machado G H ,Rocha A Y , et al.A validated LLE–LTP–LC–MS/MS workflow for pesticide biomonitoring in human plasma: Application to occupational exposure assessment in rural workers[J].Environmental Chemistry and Ecotoxicology,2026,81342-1355.