Joint Information Extraction Model Based on Feature Sharing

Zhangchi Gao [1] and Shoubin Li [1]
[1]The Institute of Software, Chinese Academy of Sciences, Beijing, China
shoubin@iscas.ac.cn

***Abstract*—To address the challenge of efficiently and accurately extracting entities, relationships, and events from unstructured text, a joint information extraction model based on feature sharing is proposed. This model utilizes the contextual information of entities, relationships, and events, and integrates entity extraction, relationship extraction, and event extraction tasks through a multi-feature cascade encoder to achieve joint extraction. To validate the effectiveness of the model, comparative analysis was conducted on military news datasets, comparing against two typical information extraction models. Results demonstrated superiority over current state-of-the-art baselines.**

***Index Terms*—Natural Language Processing, Information Extraction, Entity Extraction, Relation Extraction, Event Extraction**

I. Introduction

With the development of automation and intelligence technologies, data presents features like diverse sources, multiple types, and rapid changes. How to achieve fast fusion of multi-source heterogeneous data and construct knowledge graphs by knowledgeizing it has become the key to the successful intelligent transformation across industries [1]. For example, in the field of military confrontation capability construction, as project construction and planning are developing towards multi-domain and pan-domain directions, the informatization level of project resources is also improving. In project construction, there exist complex mapping relationships of capabilities-capabilities, projects-projects, resources-resources, as well as capabilities-projects, projects-resources. These mapping relationships often imply a complex knowledge system. Knowledge graph is the key technology to systematize the important knowledge system in project data. Managing and analyzing investment projects, resource allocation, and capability indicators by knowledge graph technologies is also an important guarantee to understand project construction and investment opportunities, and analyze construction and investment risks [2]. Additionally, the traditional approach of manually constructing large-scale knowledge graphs can no longer meet the needs of the intelligent age. Therefore, using machine learning to automatically extract knowledge from massive textual data has become the core means to construct large-scale knowledge graphs.

As a key process of knowledgeizing massive data, information extraction has become one of the most important tasks in natural language processing. It aims to extract key information like entities [3–5], relationships [6–8], and events [9, 10] from unstructured texts, and present the key information in texts in a structured form. Most existing research works treat entity extraction, relation extraction, and event extraction as independent tasks. Few studies pay attention to the correlations among these tasks. Li et al. [11] proposed a structured prediction model to jointly extract entities and events. Yang et al. [12] presented an intra-document joint model for event and entity extraction using a joint factorial Markov network to jointly learn the internal event structures, cross-event relations, and entity information within a document, improving entity and event extraction performance. The above joint extraction models rely on feature engineering with complex manual features or other auxiliary tasks like syntactic and dependency parsing, which can lead to cascading errors. To address these problems, Wu et al. [13] adopted a hybrid neural network model with bottom-level global sharing to exploit the dependencies between entity extraction and event extraction. Although the global sharing approach can avoid the error propagation brought by manual features or other auxiliary tasks, it fails to well learn the specific information of individual tasks, resulting in decreased model generalization capability.

To address the above issues, this paper proposes a joint information extraction model based on feature sharing, which can accomplish entity extraction, relation extraction and event extraction simultaneously. Firstly, the core is inserting a shared encoder between the private encoder for entity relation extraction and the private encoder for event extraction, to learn the private features for entity relation extraction, private features for event extraction, as well as the shared features between the two tasks, establishing contextual semantic associations among entities, relationships, and events. Secondly, BiLSTM and CRF structures are utilized to simulate the interactions between labels and decode labels corresponding to specific tasks in the sentence. Finally, experiments on sports news datasets demonstrate superiority over current baseline models.

II. Model Description

This section introduces the proposed model, including the entity-relationship joint extraction model, event extraction model, and joint extraction model based on feature sharing.

A. Entity-Relationship Joint Extraction Model

The entity-relationship joint extraction model consists of an encoder and a decoder. The encoder maps the input text into a high-dimensional semantic vector, and the decoder structures the semantic vector into a triplet sequence. As illustrated in Fig. 1, in the joint extraction process, a BERT [14] encoder first maps the input text into a high-dimensional vector with contextual semantic correlations. Then the vector is fed into a BiLSTM structure to capture contextual information between words and enhance contextual understanding. Next, a self-attention mechanism is introduced to weight the BiLSTM encodings and enable the model to better focus on important contextual information. Finally, a conditional random field (CRF) decodes the weighted semantic vectors to predict labels for each word, considering contextual information as well as relationships between labels, generating the optimal sequence labels. Based on the predicted sequence labels, entities and relationships between entities can be extracted, achieving joint extraction of entities and relationships from unstructured texts.

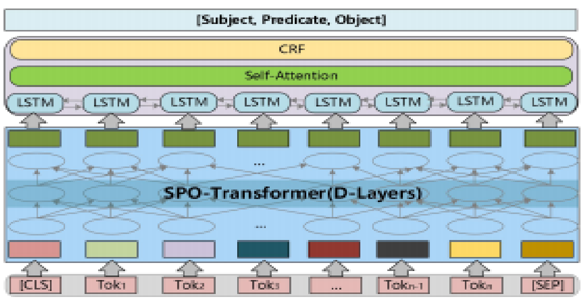

Fig. 1. Entity-relationship joint extraction model

B. Event Extraction Model

The encoder of the event extraction model is the same as the joint extraction model, adopting the BERT encoder. This model takes textual data as input and outputs trigger word identification results, argument identification results, and event categorization results for each trigger word and argument in the form of sequence labeling. As shown in Fig. 2, in the event extraction process, the encoder first maps to high-dimensional semantic vectors. Then the vectors are fed into a BiLSTM structure to learn deep contextual semantics of the text. Next, a self-attention mechanism learns the relationships between event arguments and trigger words scattered in the text. Finally, the self-attention outputs go through two linear layers and softmax functions to predict trigger words and arguments.

C. Joint Extraction Model Based on Feature Sharing

To jointly extract entities, relationships, and events from unstructured texts, a shared encoding layer is introduced, regarding the entity-relationship joint extraction model and event extraction model as private encoders. The private encoders learn task-specific information, while the shared encoder learns common knowledge across tasks, establishing associations between tasks and improving model generalization.

The input data and encoding model for the shared encoder are the same as the information extraction encoder. The specific encodings for entity-relationship extraction and event

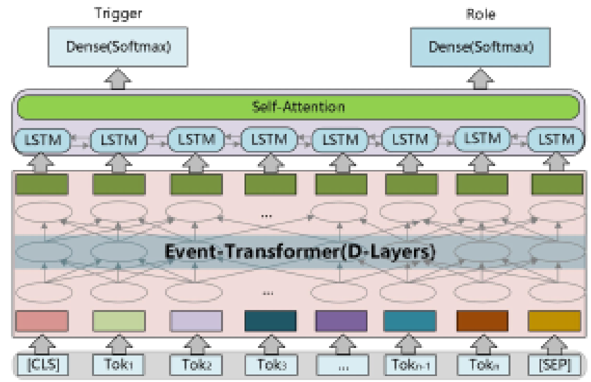

Fig. 2. Event extraction model

extraction, denoted as H_{SPO} and H_{Event}, are first fed into linear layers for linear transformation, obtaining matrices Q_S and Q_E as Formulas (1)-(2), where W_S and W_E are trainable parameters in the linear layers. Similarly, the shared encoding H_{Share} goes through a linear layer to obtain two different matrices K_S and V_S. Then self-attention is computed for each as Formulas (3)-(4), where h indicates the h-th head in multi-head attention, d_k is the dimension per head, Att_S is the result after self-attention between joint extraction features and shared features, and Att_E is the result after self-attention between event features and shared features.

$$Q_S = W_S H_{SPO} \tag{1}$$

$$Q_E = W_E H_{Event} \tag{2}$$

$$Att_S = \text{Softmax}(\frac{Q_S K_S^T}{\sqrt{d_k}})V_S \tag{3}$$

$$Att_E = \text{Softmax}(\frac{Q_E K_S^T}{\sqrt{d_k}})V_S \tag{4}$$

After self-attention computation, results from different heads are concatenated and fed through a feedforward neural network to obtain fused representations H_{Share_S} and H_{Share_E}, where H_{Share_S} denotes the fused features of entity relation and shared features, and H_{Share_E} denotes the fused features of event and shared features.

Moreover, dependencies on shared and private features vary in different training stages for different tasks. To enable automatic learning of ratios between shared and private features during training, trainable weighting factors are set for each subtask, as Formulas (5)-(6), where α_1 and α_2 are parameters learned during training.

$$H_{SPO}^{'} = (1-\alpha_1)H_{SPO} + \alpha_1 H_{Share\_S} \tag{5}$$

$$H_{Event}^{'} = (1-\alpha_2)H_{Event} + \alpha_2 H_{Share\_E} \tag{6}$$

Finally, $H_{SPO}^{'}$ and $H_{Event}^{'}$ are passed to decoders as encodings for entity-relationship extraction and event extraction to generate entity triplets and predict events. The overall architecture is illustrated in Fig. 3.

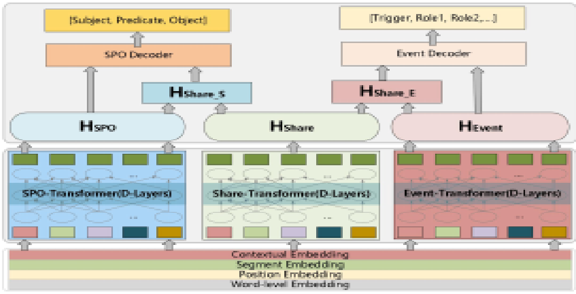

Fig. 3. Joint extraction model based on feature sharing

III. Experiments

To validate the superiority of the proposed model in information extraction, comparative experiments are conducted against two models:

- **BBC**: BERT+BiLSTM+CRF, a commonly used single-task model for entity extraction, relation extraction, or event extraction, effective in capturing contextual information and modeling sequence relationships.
- **HNN-EE** [13]: A hybrid neural network model jointly extracting entities and events via global parameter sharing.
- **Our Method**: The proposed model.

Performance is evaluated on entity-relationship extraction, event extraction, and the overall F1 scores on test sets.

Table I: Results of joint entity-relationship extraction and event extraction (%)

	Entity-Relationship			Event Extraction		
	P	R	F1	P	R	F1
BBC	79.3	77.3	78.3	70.2	68.2	69.2
HNN-EE [15]	85.3	76.5	80.7	74.4	67.3	70.6
Our Method	84.0	82.5	83.2	79.9	72.9	72.5

As shown in Table I, the single-task BBC model achieves the lowest F1 scores on both tasks. The other two joint extraction models improve upon the single-task baselines, demonstrating the feasibility of unified modeling for information extraction tasks and the usefulness of inter-task correlations.

Compared to BBC, our model improves F1 by 4.9% and 3.3% for entity-relationship extraction and event extraction respectively, indicating associations between the two tasks that facilitate mutual enhancement.

Compared to HNN-EE, our model achieves 2.5% and 1.9% F1 gains on the two tasks. Both models exploit joint extraction, but HNN-EE adopts a global encoder for cross-task knowledge sharing while our model utilizes multiple encoders to capture both shared and task-specific knowledge, demonstrating the benefits of modeling task-specific differences.

IV. Conclusion

This paper proposes a joint information extraction model based on feature sharing to address the challenge of efficiently and accurately extracting information from unstructured texts. It builds contextual semantic associations between entities, relationships, and events via multi-feature cascade encoders. Experiments demonstrate advantages over current state-of-the-art baselines on multiple information extraction tasks. In future work, domain-specific datasets will be constructed to further validate the effectiveness, and model optimization will be performed based on data characteristics to improve overall performance.

References

[1] Junyi Wang et al. “Method and device for portfolio analysis based on knowledge graph.” Patent CN202211181071.7, 2023.
[2] Eddie Chen and Xi Cai. “Construction of knowledge map for large-scale construction project management.” Journal of Changsha University of Technology (Natural Science Edition), 2016.
[3] Xiaoyan Zhang, Ting Wang, and Huowang Chen. “A study of named entity recognition.” Computer Science 32.4 (2005): 5.
[4] Hongkui Yu et al. “Chinese named entity recognition based on cascading hidden Markov models.” Journal of Communications 27.2 (2006): 8.
[5] Yuntian Feng, Hongjun Zhang, Wenning Hao. “Named entity recognition for military text.” Computer Science, 2015, 42(7): 15-18.
[6] Jing Dong et al. “A study of feature selection in Chinese entity-relationship extraction.” Journal of Chinese Information Processing 21.4 (2007): 80-91.
[7] Fengshuai Gao and Huabin Yang. “A military named entity relationship extraction method combining word rules and SVM modeling.” Information Communications 11 (2017): 2.
[8] Wenyu Duan et al. “Research on entity and relationship extraction methods for weapon and equipment domain.” Journal of the Chinese Academy of Electronics and Information Science 17.12 (2022): 1165-1172.
[9] Yanyan Zhao et al. “Research on Chinese event extraction techniques.” Journal of Chinese Information Processing 22.1 (2008): 3-8.
[10] Yuanfang Yu et al. “Multi-round event argument extraction guided by role information.” Journal of Peking University (Natural Science) 59.1 (2023): 83-91.
[11] Qi Li et al. “Constructing information networks using one single model.” Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014.
[12] Bishan Yang and Tom Mitchell. “Joint extraction of events and entities within a document context.” 2016.
[13] Wentao Wu, Peifeng Li, and Qiaoming Zhu. “A joint entity and event extraction method based on hybrid neural networks.” Journal of Chinese Information Processing 33.8 (2019): 7.
[14] Jacob Devlin et al. “BERT: pre-training of deep bidirectional transformers for language understanding.” 2018.