Learning to Ask for Data-Efficient Event Argument Extraction (Student Abstract)

Hongbin Ye1,2, Ningyu Zhang1,2,†, Zhen Bi1,2, Shumin Deng1,2, Chuanqi Tan4, Hui Chen3, Fei Huang3, Huajun Chen1,2†

1 Zhejiang University & AZFT Joint Lab for Knowledge Engine, China
2 Hangzhou Innovation Center, Zhejiang University, China
3 Alibaba Group, China
{yehongbin,zhangningyu,bizhen,zju.231sm,huajunsir}@zju.edu.cn,
{chuanqi.tcq,weidu.ch,f.huang}@alibaba-inc.com

Abstract

Event argument extraction (EAE) is an important task for information extraction to discover specific argument roles. In this study, we cast EAE as a question-based cloze task and empirically analyze fixed discrete token template performance. As generating human-annotated question templates is often time-consuming and labor-intensive, we further propose a novel approach called “Learning to Ask,” which can learn optimized question templates for EAE without human annotations. Experiments using the ACE-2005 dataset demonstrate that our method based on optimized questions achieves state-of-the-art performance in both the few-shot and supervised settings.

Introduction

Event argument extraction (EAE) is an important and challenging task in information extraction, which aims to discover specific role types for each argument in an event. For example, given that the word “declared bankruptcy” triggers a Declare-Bankruptcy event in the sentence “My uncle declared bankruptcy in 2003 and his case closed in June 2004,” EAE aims to identify that “My uncle” is an event argument in this sentence, and its argument role is “Org.”

Figure 1: Event argument extraction as a question-based cloze task, conducted in a masked language modeling manner. In the second example, we use the pseudo question token to search for the most probable argument role type.

Our Model: L2A Framework

Our L2A framework for EAE relies on question templates $X_Q$ that map an input sentence to a QA-formatted input sequence for a standard pre-trained bidirectional transformer: [CLS]question[SEP]sentence[SEP]. For the construction of the template for the question, we introduce two different strategies: (1) L2A (base): the manual questions template for the input text, which replace the tokens of the argument roles with the [mask] and add necessary prompt word information, such as event type and argument span tokens, into the question template. (2) L2A (pseudo): because manual prompts are labor-intensive and may result in sub-optimal EAE performance, we further introduce the automatic construction of the question template. Specifically, we use several unused tokens $[u1] - [u8]$ (e.g., unused or special token in the vocabulary) to form a pseudo question template and fix the other weights of the language model to learn the op-
timized questions.

Training Objectives

Because the argument role label contains semantic information, we can simplify the label mapping in EAE as an injective function. For example, we can project the "Transaction.Transfer-Money" event role as:

$$\mathcal{Y}(\mathcal{R}) = \{ \text{giver, recipient, beneficiary, place} \}$$ (1)

We normalize the vocabulary distribution of a single token of the event role and define the prediction probability as:

$$p(y \mid X_Q) = \frac{\exp{(w_{y(r)} \cdot h_{\text{MASK}})}}{\sum_{r' \in \mathcal{R}} \exp{(w_{y(r')} \cdot h_{\text{MASK}})}}$$ (2)

where $h_{\text{MASK}}$ is the hidden vector corresponding to the [MASK] position. Therefore, we use cross-entropy loss to define the event role prediction as:

$$\mathcal{L}_{EAE} = CE(p(y \mid X_Q))$$ (3)

where $\mathcal{L}_{EAE}$ denotes the EAE loss and $CE$ is the cross entropy loss function. To make the input text resemble the natural language more closely, we leverage an auxiliary optimization object. We randomly mask out other tokens in the sentence and conduct a masked language model prediction as:

$$q(x^m \mid x', u) = \frac{\exp{(\|L(x', u)\| x^m)}}{\sum_{x' \in \mathcal{Y}} \exp{(\|L(x', u)\| x')}}$$ (4)

$$\mathcal{L}_{MLM} = \sum_{m \in M} BCE(q(x^m \mid x', u))$$ (5)

where $u$ denotes the question format input sequence, $x^m$ is the original token which has been randomly masked, $x'$ represents the input sentence after mask processing, and $BCE$ is the binary cross entropy loss function. Finally, we have the following optimization objective:

$$\mathcal{L}_{total} = \mathcal{L}_{EAE} + \mathcal{L}_{MLM}$$ (6)

Experiments

We evaluate our L2A model with the ACE 2005 dataset. For the few-shot scenario, we follow the few-shot settings of (Gao, Fisch, and Chen 2021), which is different from the N-way K-shot setting. From Table 1, we observe that L2A (base) performs better than baselines, which indicates that fine-tuning in a question-based cloze task can bring substantial benefits. To illustrate the effectiveness of optimizing tokens, we conduct a nearest neighbor vocabulary embedding search to project the best optimized pseudo question token to a readable natural language. We note the model with the projected question as L2A (projected). It is worth noting that the performance of L2A (projected) is only lower than the best-optimized result by 0.8%. From Figure 2, we can observe that L2A (pseudo) demonstrates an absolute improvement of up to 15% over previous state-of-the-art models in the few-shot scenario ($k = 4$). Note that the question can include task-specific and argument-relevant information that can boost the performance. Moreover, our approach is consistent with the pre-training paradigm and thus is more convenient for leveraging the knowledge available in the parameter space when learning with sparse data.

Conclusion

In this paper, we present L2A, a simple but effective method for optimizing question templates for EAE tasks, and demonstrate that it performs competitively with state-of-the-art models. Furthermore L2A can search suitable questions for EAE without any manually constructed templates.

Acknowledgments

This work is funded by NSFC91846204/NSFCU19B2027.

References

