Metaphor Detection with Effective Context Denoising

Shun Wang1*, Yucheng Li2*, Chenghua Lin1†, Loïc Barrault1,3, Frank Guerin2
1 Department of Computer Science, University of Sheffield, UK
2 Department of Computer Science, University of Surrey, UK
3 Meta AI
{swang209, c.lin}@sheffield.ac.uk
{yucheng.li, f.guerin}@surrey.ac.uk
loicbarrault@meta.com

Abstract

We propose a novel RoBERTa-based model, RoPPT, which introduces a target-oriented parse tree structure in metaphor detection. Compared to existing models, RoPPT focuses on semantically relevant information and achieves the state-of-the-art on several main metaphor datasets. We also compare our approach against several popular denoising and pruning methods, demonstrating the effectiveness of our approach in context denoising. Our code and dataset can be found at https://github.com/MajiBear000/RoPPT.

1 Introduction

Metaphor is a pervasive linguistic device, which attracts attention from both the fields of psycholinguistics and computational linguistics due to the key role it plays in the cognitive and communicative functions of language (Wilks, 1978; Lakoff and Johnson, 1980; Lakoff, 1993). Linguistically, metaphor is defined as a figurative expression that uses one or several words to represent another concept given the context, rather than taking the literal meaning of the expression (Fass, 1991). For instance, in the sentence “This project is such a headache”, the contextual meaning of headache is “a thing or person that causes worry or trouble”, different from its literal meaning, “a continuous pain in the head”.

Metaphor detection is challenging, as it requires understanding the nuanced relationships between abstract concepts embodied by the metaphoric expression and its surrounding context. Recent studies on this direction show its potential in benefiting a wide range of NLP applications, including sentiment analysis (Li et al., 2022a), metaphor generation (Li et al., 2022b,c) and mental healthcare (Abd Yusof et al., 2017; Gutiérrez et al., 2017).

When modelling relevant context for metaphor detection, various strategies have been proposed. These range from using highly restricted forms of linguistic context such as subject-verb and verb-direct object word pairs (Gutiérrez et al., 2016), to a wider context accounting for a fixed window surrounding the target word (Do Dinh and Gurevych, 2016; Mao et al., 2018), and modelling the full sentential context (Gao et al., 2018; Choi et al., 2021). While it has been argued that modelling a wider context is beneficial (Cheng et al., 2021), it has also been noted that a wider context is likely to introduce noise into the representations, and hence hinder model’s performance in metaphor detection (Le et al., 2020).

Some recent efforts (Le et al., 2020; Song et al., 2021a) attempt to improve context modelling by explicitly leveraging the syntactic structure (e.g., dependency parse tree) of a sentence in order to capture important context words, where the parse trees are typically encoded with graph convolutional neural networks. MelBERT (Choi et al., 2021) employs a simple chunking method which separates sub-sentences by commas. The sub-sentence that contains a target word is then marked with a special token type, signalling its contextual importance to the target. However, these strategies are either difficult to apply to batch optimisation due to their tree-dependent encoding process, or have limited effectiveness for context denosing. For instance, the simple chunking mechanism misses the syntactic structure, and thus can neither determine the degree of importance of context words, nor connect information across different subsentences.

In this paper, we propose a novel metaphor detection model RoPPT: RoBERTa with Pruning on target-oriented Parse Tree. RoPPT introduces a flat, target-oriented tree structure by reshaping and pruning the ordinary parse trees to extract semantically relevant neighbours of a target word. The resulting tree representation allows the model to

---

* The two authors contributed equally to this work.
† Corresponding author
focus on syntactically relevant information of a target word, and ignore irrelevant parts despite their position. It thus retains more relevant context for metaphor detection.

Extensive experiments conducted on three public benchmark datasets (i.e., VUA, MOH-X, TroFi) show that RoPPT can significantly improve metaphor detection on all datasets against several popular denoising and pruning methods. Our model also yields better or comparable performance to the state-of-the-art models (Choi et al., 2021; Song et al., 2021a) in Micro F1 measure. To further validate our approach, we conducted an additional investigation to assess the effect of sentence length on the performance of our model. Experimental results demonstrate a positive correlation between the increase in the performance of RoPPT and the length of the input sentence.

In summary, our paper makes three contributions: (1) we propose a flat, target word-oriented tree structure by reshaping and pruning the ordinary parse trees to retain the most relevant context for a target word; (2) we propose RoPPT, a RoBERTa-based model which can effectively encode the target-oriented parse tree for metaphor detection, achieving state-of-the-art results on three benchmark datasets; (3) we compare and evaluate a range of context denoising methods for metaphor detection, demonstrating the effectiveness of our proposed tree structure in context denoising.

2 Method

The overall architecture of RoPPT is shown in Figure 1, which can be divided into two parts: a target-oriented parse tree pruning module and a RoBERTa (Liu et al., 2019) contextual encoder.

2.1 Target-oriented Dependency Parse Tree

Connecting target words with their most relevant context words is crucial for metaphor detection and comprehension. While there have been attempts to employ dependency parse trees in graph convolutional neural networks to improve context modelling (Wang et al., 2020), it raises challenges of how to effectively encode and leverage such syntactic structure information for transformer-based mask language models for metaphor detection.

We tackle this challenge by introducing a target-oriented parse tree generated by three steps: 1) reshape the original parse tree from existing parsers such as spaCy (Honnibal and Montani, 2017) and Biaffine (Dozat and Manning, 2016); 2) root the tree at the target word; 3) prune the tree according to the distance between leaves and root, coined as neighbor range. The rationale behind is that the target word is the focus of the task rather than the original root. So the re-rooting allows us to focus on the connections between target words and their relevant context. The resulting flat, target-oriented tree structure also enables simple encoding process into the model. Figure 1 shows an example of our reshaped tree, which retracts words with neighbor range $con = 1$ to the root ‘bogged’.

2.2 RoBERTa-based Context Encoder

We employ two metaphor identification theories in our model, i.e., Metaphor Identification Procedure (Steen, 2010, MIP) and Selectional Preference Violation (Wilks, 1978, SPV). In MIP, a metaphor is detected when there is a contrast between target word’s contextual and literal meanings, whereas in SPV a metaphorical word is identified by the semantic difference from its surrounding words. Therefore, we model three types of semantic representations for implementing MIP and SPV, i.e., the literal meaning and the contextual meaning of a target word, and the context meaning.

Formally, given a sentence $S = (w_0, ..., w_n)$, we first employ the RoBERTa network to produce representations for each word.

$$H = \text{RoBERTa}_\text{Enc}(\text{emb}_{c\text{ls}}, ..., \text{emb}_{n})$$

Here $c\text{ls}$ is a special token indicating the start of an input, $H = (h_{c\text{ls}}, h_0, ..., h_n)$ the output hidden states, and $\text{emb}_i$ the input embedding for word $w_i$. Specifically, $\text{emb}_i = \text{emb}_w + \text{emb}_\text{pos}$, where $\text{emb}_w$ is the word embedding, and $\text{emb}_\text{pos}$ the position encoding.

Context denoising with the target-oriented parse tree. When modelling sentence representation, existing works directly employed the $c\text{ls}$ embedding as a common practice (Choi et al., 2021; Song et al., 2021b). In contrast, RoPPT employs the target-oriented parse tree to retain the most relevant context for a target word when computing the sentence embedding. Specifically, our sentence embedding is computed as follows.

$$v_S = \frac{1}{n} \sum_{i \in C_n} h_i$$

Here $v_S$ is the sentence representation; $C_n$ represents the $n$ neighbour words within the neighbour
range of the target-oriented parse tree, and \( h_i \) is the hidden state of \( w_i \). In other words, we do average pooling on the most relevant context words as the sentence representation and ignore other words in the sentence. We also design an alternative strategy by directly masking the original input sentence to the encoder according to the pruned parse tree. We denote this intuitive solution as RoPPT with Input Mask (RoPPT IM) and discuss the performance difference between these two variants in §4.

Similar to Choi et al. (2021), we use the hidden state of target word \( w_t \) as the contextual target word embedding (i.e. \( v_{S,t} = h_t \)), and the literal target word embedding \( v_t \) is obtained by feeding the single target word \( w_t \) to the RoBERTa network.

\[
v_t = \text{RoBERTa} \_\text{Enc}(\text{emb}_t)
\]

We then model SPV (\( h_{SPV} \)) by concatenating the sentence embedding \( v_S \) and contextual target embedding \( v_{S,t} \), and MIP (\( h_{MIP} \)) by concatenating the contextual and literal target embeddings \( v_t \), followed by a MLP layer (i.e. \( f_1(\cdot) \) and \( f_2(\cdot) \)).

\[
\begin{align*}
  h_{SPV} &= f_1([v_S, v_{S,t}]) \\
  h_{MIP} &= f_2([v_{S,t}, v_t])
\end{align*}
\]

Finally, we combine two hidden vectors \( h_{MIP} \) and \( h_{SPV} \) to compute a prediction score \( \hat{y} \), and use binary cross entropy loss to train the overall framework for metaphor prediction.

\[
\hat{y} = \sigma(W^T [h_{MIP}; h_{SPV}] + b)
\]

\[
\mathcal{L} = -\sum_{i=1}^{N} [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)]
\]

3 Experimental Setup

Dataset. We conduct experiments on four public benchmark datasets. VUA-18 (Leong et al., 2018) and VUA-20 (Leong et al., 2020) are the largest available datasets, released in the metaphor detection shared tasks in 2018 and 2020. VUA-20 extends VUA-18 with about 12K sentences for training set and 3.6K sentences for test and validation sets. The MOH-X dataset is constructed by sampling sentences from WordNet (Miller, 1998). Only a single target verb in each sentence is annotated. The average sentence length is 8 tokens, the shortest of our three datasets. TroFi (Birke and Sarkar, 2006) consists of sentences from the 1987-89 Wall Street Journal Corpus (Charniak et al., 2000), with an average length of 28.3 tokens per sentence.

Baselines. RoBERTa_SEQ (Leong et al., 2020) is a fine-tuned RoBERTa sequence labeling model for metaphor detection. MelBERT (Choi et al., 2021)
null
its subject, which is correctly labeled by RoPPT but incorrectly by baseline models. For the instance with metaphorical target word *bogged*, "a routine exercise in extending the government’s borrowing power to $3.1 thousand billion became bogged down.", the target word *bogged* is separated from its subject by a long phrase, which causes baselines (including MelBERT) to fail to detect the metaphor. Thanks to the parse tree, RoPPT links *exercise* directly to the target and produces the right label.

5 Conclusion

In this paper, we proposed, RoPPT, an effective approach to extract contextual information for target words for metaphor detection based on a target-oriented parse tree structure. Extensive experiments show that our model can yield better performance compared to the state-of-the-art. In addition, our method is particularly effective in denoising long sentences, despite its simplicity.

6 Limitations

Empirical experiments show that our method is more effective in denoising long sentences with the proposed target-oriented parse tree. While this is somewhat expected as shorter sentences tend to have cleaner context, it raises a question or limitation of how can we improve the proposed method to better deal with short sentences and improve its performance in these cases. One possibility is to exploit external knowledge (e.g. ConceptNet) to support the detection of the most important contextual words.

References


E. Darío Gutiérrez, Ekaterina Shutova, Tyler Marghetis, and Benjamin Bergen. 2016. Literal and metaphorical senses in compositional distributional semantic


