

ConnPrompt: Connective-cloze Prompt Learning for Implicit Discourse Relation Recognition

Wei Xiang, Zhenglin Wang, Lu Dai, Bang Wang*

School of Electronic Information and Communications,

Huazhong University of Science and Technology, Wuhan, China

{xiangwei, wangzhenglin, dailu18, wangbang}@hust.edu.cn

Abstract

Implicit Discourse Relation Recognition (IDRR) is to detect and classify relation sense between two text segments without an explicit connective. Vanilla *pre-train and fine-tuning* paradigm builds upon a Pre-trained Language Model (PLM) with a task-specific neural network. However, the task objective functions are often not in accordance with that of the PLM. Furthermore, this paradigm cannot well exploit some *linguistic evidence* embedded in the pre-training process. The recent *pre-train, prompt, and predict* paradigm selects appropriate prompts to reformulate downstream tasks, so as to utilizing the PLM itself for prediction. However, for its success applications, prompts, verbalizer as well as model training should still be carefully designed for different tasks. As the first trial of using this new paradigm for IDRR, this paper develops a *Connective-cloze Prompt* (ConnPrompt) to transform the relation prediction task as a connective-cloze task. Specifically, we design two styles of ConnPrompt template: *Insert-cloze Prompt* (ICP) and *Prefix-cloze Prompt* (PCP) and construct an *answer space* mapping to the relation senses based on the hierarchy sense tags and implicit connectives. Furthermore, we use a multi-prompt ensemble to fuse predictions from different prompting results. Experiments on the PDTB corpus show that our method significantly outperforms the state-of-the-art algorithms, even with fewer training data.

1 Introduction

Implicit Discourse Relation Recognition (IDRR) aims at detecting and classifying some latent relation in between a pair of text segments (called arguments) without an explicit connective word. As illustrated in Fig. 1, an implicit discourse relation of "Contingency" is held between Argument-1 and Argument-2, and the implicit connective 'so'

is inserted by annotators in the PDTB corpus. It is of great importance for many downstream Natural Language Processing (NLP) applications, such as question answering (Liakata et al., 2013), machine translation evaluation (Guzmán et al., 2014), information extraction (Xiang and Wang, 2019), sentiment analysis (Wang and Wang, 2020), and etc. However, due to the absence of an explicit connective word, inferring discourse relations from the contextual semantics of arguments is still a challenging task.

Existing *pre-train and fine-tuning* paradigm (Liu et al., 2021) builds upon a Pre-trained Language Model (PLM) with a well-designed sophisticated neural network to encode the *semantic content* and *interactive evidence* of argument pairs (Liu and Li, 2016; Lei et al., 2017; Bai and Zhao, 2018; Ruan et al., 2020; Li et al., 2020; Liu et al., 2020; Wu et al., 2022). Although the PLMs are adapted to these task-specific neural networks that can effectively learn a kind of contextual semantics of arguments, they introduce some additional parameters that need to be trained by a large amount of labelled data. Moreover, the task objective function is often not in accordance with that of the PLM. As such, the PLM needs to be fine-tuned for solving downstream tasks, resulting in poor utilization of the encyclopedic *linguistic evidence* embedded in the pre-training process.

On the one hand, we notice that the pre-training process of a PLM often uses a kind of cloze task, called Masked Language Model (MLM), to predict a piece of masked text from context. On the other hand, it has been reported that an explicit connective of an argument pair can greatly improve the relation classification performance in the *explicit discourse relation recognition* task (Pitler et al., 2008). Although explicit connectives are not available in the IDRR task, it is of great interests to explore whether we can transform the relation prediction task as a connective-cloze task, such that we

* Corresponding author: Bang Wang

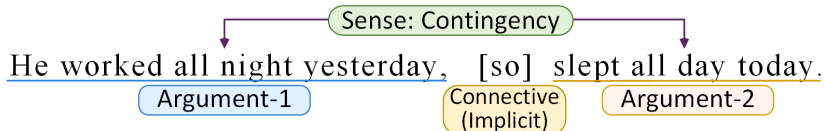


Figure 1: An example of implicit discourse relation annotation with manually inserted connective.

can exploit a pre-trained masked language model to predict a missing (yet possibly latent) connective for implicit relation classification. This is actually in accordance with the philosophy of the *prompt learning* paradigm, that is, predicting a connective as an *answer word* in some predefined template and then mapping the answer word to one relation sense.

The recent *pre-train, prompt, and predict* paradigm models the probability of text directly based on PLMs to perform prediction task. Specifically, it selects appropriate prompts to reformulate downstream task, so as to utilizing the PLM itself to predict the desired output (Liu et al., 2021). Moreover, the prompt paradigm is capable of performing few shot even zero-shot learning, as the PLM is sufficiently pre-trained and no external parameters need to be trained. However, for its successful applications in many downstream NLP tasks (Ding et al., 2021; Wang et al., 2021; Seoh et al., 2021), *prompts engineering, verbalizer* as well as *training strategies* should still be carefully designed for different tasks. In this paper, we explore how to transform the IDRR task against the prompt learning paradigm. To the best of our knowledge, this is the first paper for such explorations.

In this paper, we develop a *Connective-cloze Prompt* (ConnPrompt) framework to transform the relation prediction task as a connective-cloze task for the IDRR task. Specifically, we design two styles of ConnPrompt template: *Insert-cloze Prompt* (ICP) and *Prefix-cloze Prompt* (PCP), in which the [MASK] token is added for connective answer prediction. The ICP template concatenates two arguments as an entire word sequence, and the [MASK] token is inserted in between two arguments; The PCP template uses a [SEP] token to mark the boundary of two arguments, and the [MASK] token is added at the beginning of argument-1 or argument-2 as a prefix. Besides, we construct an *answer space* mapping an answer word to relation senses according to the hierarchy sense tags and implicit connectives in the training dataset. Furthermore, in order to leverage the com-

plementary advantages of different prompt templates, we use a multi-prompt ensemble to fuse predictions from different prompting results.

We conduct the experiments on the PDTB corpus with four advanced masked language models: BERT, RoBERTa, ERNIE and DeBERTa. Experiment results show that our ConnPrompt significantly outperforms the state-of-the-art algorithms with full training data. Furthermore, our ConnPrompt can also achieve comparable performance even with fewer training data.

2 Method

In this section, we first introduce the overall framework of our ConnPrompt, then explain the details of connective-cloze prompt templates, verbalizer construction, multi-prompt ensembling, and model training strategies.

2.1 Overview

As illustrated in Fig. 2, our ConnPrompt has three main processes, including prompt templating, answer prediction and verbalizer.

Prompt templating: an input argument pair $x = (Arg_1; Arg_2)$ is reformulated into a prompt template $T(x)$ by concatenating two arguments and inserting some PLM-specific tokens such as [MASK], [CLS], [SEP], as the input of a PLM. The [MASK] token is added for PLM to predict an answer word v ; While the [CLS] and [SEP] tokens are used to indicate the beginning and ending of an input word sequence, respectively. Note that some PLMs use other tokens like $\langle mask \rangle$, $\langle s \rangle$, and $\langle /s \rangle$, but they have the same meaning as described above.

Answer prediction: the pre-trained masked language model estimates the probability of each word in its vocabulary V for the [MASK] token as follows:

$$P_v([\text{MASK}] = v \in V | T(x)). \quad (1)$$

We define a discrete answer space $V_a = \{v_1, v_2, \dots, v_n\}$ containing the words manually selected according to the hierarchy sense tags and

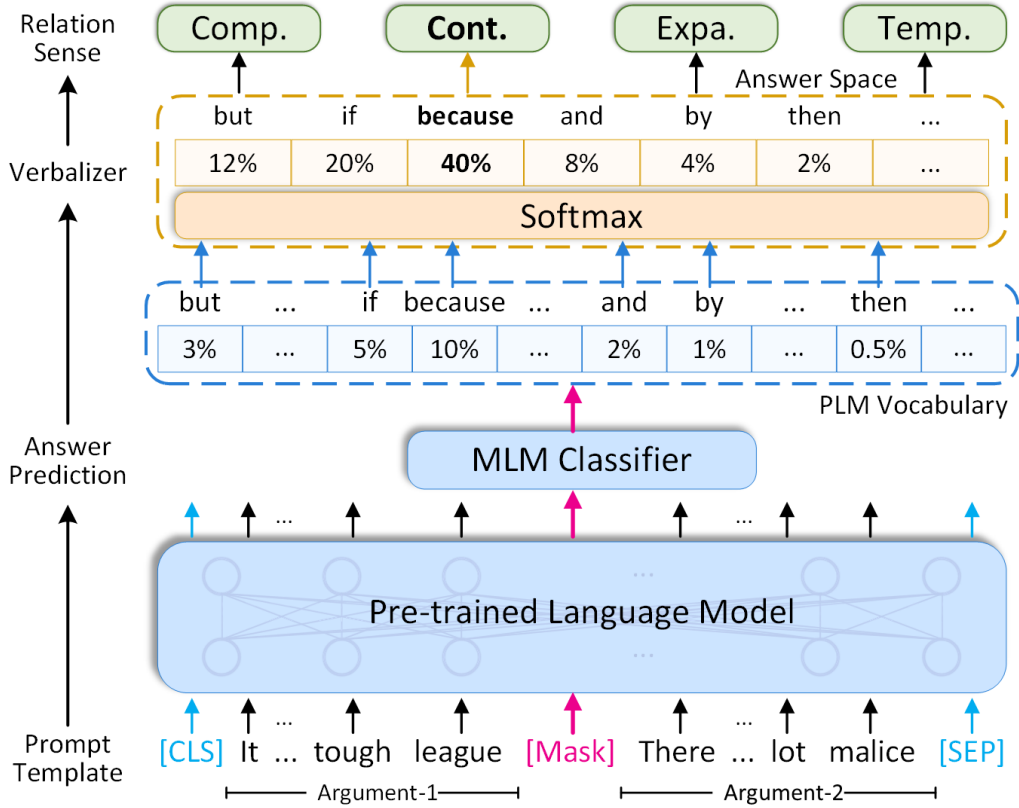


Figure 2: Illustration of our ConnPrompt framework.

implicit connectives, which is a subset of PLM’s vocabulary, $V_a \subset V$. Then, a softmax layer is applied on the prediction scores of our answer words to normalize them into probabilities:

$$P_a(v_i \in V_a | T(x)) = \frac{e^{p_{v_i}}}{\sum_{j=1}^n e^{p_{v_j}}}, \quad (2)$$

Verbalizer: the predicted answer word is projected to a unique discourse relation sense based on our pre-defined connection regulation.

2.2 Connective-cloze prompts

Motivated by the fact that connective words can effectively indicate the relation sense between two arguments, we design a kind of connective-cloze prompt template to predict a connective-bearing answer word for IDRR. In English syntax and grammar, connective words are usually located at the beginning of a sentence or between two adjacent clauses. Thus we design two styles of prompt templates for connective-cloze prompts: *Prefix Cloze Prompt* (PCP) and *Insertion Cloze Prompt* (ICP), as shown in Fig. 3:

$T_1(x)$ is an ICP template, in which Arg_1 and Arg_2 are concatenated as an entire word sequence,

and the [MASK] token is inserted between two arguments. $T_2(x)$ and $T_3(x)$ are PCP templates, in which the [SEP] token is also used to mark the boundary between Arg_1 and Arg_2 , and the [MASK] token can be either added at the front of Arg_1 ($T_3(x)$) or Arg_2 ($T_2(x)$).

Relation Sense	Answer words
Comparison	<i>similarly, but, however, although</i>
Contingency	<i>for, if, because, so</i>
Expansion	<i>instead, by, thereby, specifically, and</i>
Temporal	<i>simultaneously, previously, then</i>

Table 1: Answer space of our ConnPrompt and their connection to the top-level class discourse relation sense tags in the PDTB.

2.3 Verbalizer Construction

Table 1 presents our verbalizer connection from the answer space to discourse relation sense labels. Note that the answer space is a small subset of the vocabulary in a PLM. We select sixteen answer words from nearly two hundred connectives in the PDTB corpus. In our verbalizer construction, the following four design issues are considered to satisfy the representative of each relation sense.

Insertion Cloze Prompt:

$$T_1(x) = [\text{CLS}] + \text{Argument}_1 + [\text{Mask}] + \text{Argument}_2 + [\text{SEP}].$$

Prefix Cloze Prompt:

$$T_2(x) = [\text{CLS}] + \text{Argument}_1 + [\text{SEP}] + ([\text{Mask}] + \text{Argument}_2) + [\text{SEP}].$$

$$T_3(x) = [\text{CLS}] + ([\text{Mask}] + \text{Argument}_1) + [\text{SEP}] + \text{Argument}_2 + [\text{SEP}].$$

Figure 3: Illustration of our ConnPrompt Templates.

- **Answer Shape:** Only individual word connectives are selected as answer words, as most masked PLMs predicts only a single word.
- **Ambiguity:** Each answer word has one unambiguous connection with one discourse relation sense. Those words that can be used for multiple senses are not selected.
- **Frequency:** High frequency appearance connectives are prior to be selected as the answer words.
- **Semantic:** For those words with similar semantics in the same relation sense, we select a representative one to alleviate the answer confusion issue.

Specifically, we first eliminate those ambiguous connectives each for multiple senses, so that each answer word corresponds to only one discourse relation sense. We next rank the rest connectives according to their appearance frequencies in PDTB corpus to obtain a candidate set of answer words from the top of connective ranking. Finally, we select a representative word for those words with similar semantics in the same relation sense to alleviate the answer confusion issue and construct an answer space with sixteen words V_a as in Table 1. Each of them has a unique connection to one of the four top-level classes of relation sense $Y = \{\text{Comparison, Contingency, Expansion, Temporal}\}$.

2.4 Training Strategies

In model training, we tune the parameters of PLM using the IDRR training dataset based on our created prompt templates and answer space. Note that the final verbalizer layer is a projection and has no parameters to train. For model training, we assign an answer word for each instance of an argument pair as its ground truth label according to its manually annotated implicit connectives and the

hierarchical sense tags in the PDTB. Specifically, if the implicit connective of an argument pair instance is in our answer space, we directly use it as the answer label; Otherwise, we take the most frequent answer word that has the same subtype-level sense tag as its label.

We adopt the cross entropy loss as the cost function:

$$J(\theta) = -\frac{1}{K} \sum_{k=1}^K \mathbf{y}^{(k)} \log(\hat{\mathbf{y}}^{(k)}) + \lambda \|\theta\|^2, \quad (3)$$

where $\mathbf{y}^{(k)}$ and $\hat{\mathbf{y}}^{(k)}$ are the gold label and predicted label of the k -th training instance respectively. λ and θ are the regularization hyper-parameters. We use the AdamW optimizer (Loshchilov and Hutter, 2019) with $L2$ regularization for model training.

2.5 Multi-prompt Ensembling

Multi-prompt learning uses multiple unanswered prompts for an input at inference time to make prediction (Lester et al., 2021). In accordance with English convention, we have designed three prompt templates. As each of them can output prediction probabilities for answer words, we make a decision fusion of majority voting as multi-prompt ensembling for final relation sense prediction. After each prompt predicting a specific relation sense, if two or more prompts have the same prediction, then their predicted relation sense is used as the final output. In a case that each prompt predicts differently, we choose the prediction from the prompt template with the highest F1 in the validation dataset.

3 Experiment Settings

In this section, we present our experimental settings, including dataset, PLMs, parameter settings and competitor models.

3.1 The PDTB Dataset

We conduct our experiments on the Penn Discourse TreeBank (PDTB) 3.0 corpus (Webber et al., 2019),

which contains more than one million words of English texts from Wall Street Journal. Following the conventional data splitting, we use sections 2-20 as the full training set, sections 21-22 as the testing set and 0-1 as the development set (Ji and Eisenstein, 2015). Our experiments are conducted on the four top-level classes of relation sense, including Comparison, Contingency, Expansion, Temporal.

For few shot learning, we randomly down-sample the full training set to construct some subsets containing {Full, 50%, 30%, 20%, 10%} instances of the full training set. Table 2 summarizes the statistics of training instances in the training set and subsets.

Relation	Expa.	Comp.	Cont.	Temp.	Total
Train-Full	8645	1937	5916	1447	17945
Train-50%	2794	1937	2794	1447	8972
Train-30%	1346	1346	1346	1346	5384
Train-20%	898	898	898	898	3592
Train-10%	449	449	449	449	1796
Dev.	748	190	579	136	1653
Test	643	154	529	148	1474

Table 2: Statistics of the PDTB training set and down-sampling sets with four top-level relation senses.

3.2 Pre-trained Language Models

We use four masked pre-trained language models (PLM) for comparison:

- **BERT** (Devlin et al., 2019): The most representative PLM proposed by Google ¹, which is pre-trained using a *cloze task* and a *next sentence prediction* task.
- **RoBERTa** (Liu et al., 2019): A BERT-enhanced PLM proposed by Facebook ², which removes the next sentence prediction objective and is pre-trained on a much larger dataset with some modified key hyper-parameters.
- **ERNIE** (Sun et al., 2019): A knowledge-enhanced PLM proposed by Baidu ³, which uses some knowledgeable masking strategies in pre-training.

¹<https://github.com/google-research/bert>

²<https://github.com/pytorch/fairseq/>

³<https://github.com/PaddlePaddle/ERNIE>

- **DeBERTa** (He et al., 2021): The latest masked PLM proposed by Microsoft ⁴, which improves BERT and RoBERTa models using a disentangled attention mechanism and an enhanced mask decoder.

3.3 Parameter Setting

Table 3 presents the configuration of each English masked pre-trained language model. All these PLM models are implemented in PyTorch ⁵ framework by HuggingFace transformers ⁶ (Wolf et al., 2020), and run with CUDA on NVIDIA GTX 1080 Ti GPUs. From our statistics, 99.46% of arguments do not exceed 50 words in PDTB. So we set the maximum length of each prompt template to 100 tokens, in which the maximum length of argument-1 is 50 tokens, and the rest 50 tokens are for argument-2 and [MASK], [CLS] and [SEP] tokens. We train all the four masked PLMs with the same mini-batch of 16 and learning rates of 5e-6, 1e-5, 2e-5 and 5e-5. We release the code at: <https://github.com/HustMinsLab/ConnPrompt>.

PLM	Model	Vocab. size	Layer	Dim.
BERT	bert-base-uncased	30522	12	768
RoBERTa	roberta-base	50265	12	768
ERNIE	ernie-2.0-en	30522	12	768
DeBERTa	deberta-base	50265	12	768

Table 3: Configuration of four pre-trained masked language models.

3.4 Competitors

We compare our ConnPrompt with the following advanced models:

- **DAGR**N (Chen et al., 2016) encodes word-pair interactions by a neural tensor network.
- **NNMA** (Liu and Li, 2016) combines two arguments' representations for stacked interactive attentions.
- **IPAL** (Ruan et al., 2020) propagates self-attention into interactive attention by a cross-coupled network.
- **PLR** (Li et al., 2020) uses a penalty-based loss re-estimation method to regulate the attention learning.

⁴<https://github.com/microsoft/DeBERTa>

⁵pytorch.org

⁶<https://github.com/huggingface/transformers>

- **BMGF** (Liu et al., 2020) combines bilateral multi-perspective matching and global information fusion to learn a deep contextualized representation.

Model	PLM	Acc	F1
DAGRN (ACL, 2016)	Word2vec	57.33%	45.11%
NNMA (EMNLP, 2016)	Glove	57.67%	46.13%
IPAL (COLING, 2020)	BERT	57.33%	51.69%
PLR (COLING, 2020)	BERT	63.84%	55.74%
BMGF (IJCAI, 2020)	RoBERTa	69.95%	62.31%
Our ConnPrompt	BERT	69.67%	64.00%
	RoBERTa	75.17%	70.88%
	ERNIE	72.93%	68.37%
	DeBERTa	74.63%	70.19%

Table 4: Overall results of comparison models on the PDTB corpus.

4 Result and Analysis

4.1 Overall Result

We implement a four-way classification on the top-level relation sense of the PDTB, in which macro $F1$ score and accuracy (Acc) are used for evaluation. Table 4 compares the overall performance between our ConnPrompt and the state-of-the-art models with pre-train and fine-tuning paradigm. We note that the first two competitors both use a kind of distributed representation based static word embeddings: Word2vec and Glove, provided by Google⁷ and Stanford NLP Group⁸. While the others use Transformers based pre-trained masked language model BERT and RoBERTa, which are dynamic and contextual. We also compare the performance of our ConnPrompt with different PLMs.

The first observation is that the IPAL, PLR and BMGF model can obviously outperform the first two competitors, viz., the DAGRN and NNMA model. This might be attributed to the use of more advanced dynamic PLMs which are pre-trained with deeper neural networks and larger scale of parameters based on Transformers. Indeed, transformer-based PLMs have been proven to be more effective for many downstream NLP tasks (Devlin et al., 2019; Liu et al., 2019). Although these competitors’ well-designed task-specific neural networks also have a certain impact on the performance, the gaps between dynamic

PLMs and static word embeddings are still apparent in the IDRR task.

The second observation is that in our ConnPrompt employing different PLMs, the ConnPrompt-BERT performs the worst. We note that although they all employ Transformer based model in pre-training, the RoBERTa, ERNIE and DeBERTa have applied some adjusted and optimized pre-training processes. Specifically, the RoBERTa removes the next sentence prediction task and uses a much larger dataset for training; While the ERNIE uses some knowledgeable masking strategies; and the DeBERTa applies a disentangled attention mechanism to encode context and position information separately. This suggests that the improvements and optimization in the pre-training process can effectively improve the performance of prompt learning.

Finally, our ConnPrompt with all four PLMs have achieved better performance than all conventional pre-train and fine-tuning paradigm models in macro $F1$ score, even some of the competitors have used advanced PLMs like RoBERTa and BERT, to train an elaborate downstream task model. Besides, the ConnPrompt-RoBERTa and ConnPrompt-DeBERTa model have achieved significant improvements over all competitors and PLMs in terms of much higher macro $F1$ score and Acc. We attribute its outstanding performance to our task transformation of connective-cloze prediction into the training of PLMs, other than using task-specific model built-upon a PLM, by which our ConnPrompt can better enjoy the encyclopedic linguistic evidence embedded in a PLM during the model training process.

4.2 Prompt Template Effections

In the prompt paradigm, using different templates may impact on the task performance. Table 5 compares the results of our designed single-prompt templates and multi-prompt ensembling of ConnPrompt.

It can be first observed that using different prompt templates do result in some performance disparity, even though the gaps are not obvious. For single-prompt learning, the BERT and RoBERTa have achieved the best performance in Prompt-1, while the ERNIE and DeBERTa have achieved the best performance in Prompt-2. This suggests that the semantic encoding might play the central role in BERT and RoBERTa, as the Prompt-1 does

⁷code.google.com/archive/p/word2vec

⁸<https://nlp.stanford.edu/projects/glove/>

PLM	BERT		RoBERTa		ERNIE		DeBERTa	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Prompt-1	69.74%	<u>63.95%</u>	<u>74.36%</u>	<u>69.91%</u>	72.05%	67.25%	72.32%	67.74%
Prompt-2	69.34%	63.69%	73.61%	69.63%	<u>72.25%</u>	<u>67.59%</u>	<u>72.66%</u>	<u>67.98%</u>
Prompt-3	67.64%	62.65%	73.54%	69.00%	70.15%	66.21%	<u>72.66%</u>	67.92%
Multi-Prompt	<u>69.67%</u>	64.00%	75.17%	70.88%	72.93%	68.37%	74.63%	70.19%

Table 5: Results of single-prompt templates and multi-prompt ensembling with different PLMs on the PDTB corpus.

not mark the boundary of two arguments, and the [MASK] token is inserted between them to form an entire word sequence.

By contrast, ERNIE and DeBERTa might have more consideration on position encoding, as the Prompt-2 uses a [SEP] token to distinguish two input arguments and the [MASK] token is added at the front of argument-2. They have shown more powerful ability with these position information. We also observe that Prompt-3 cannot outperform the other two prompts. This may be attributed to its infrequent grammar structure that places the connective at the beginning of the first argument.

Our multi-prompt model achieves performance improvements over the single-prompt models for almost all PLMs. This indicates that our multi-prompt ensembling is effective for fusing multiple single-prompts for discourse relation classification.

4.3 Few Shot Learning

Some researchers have reported that the prompt paradigm is of some robustness to using fewer training data in other NLP tasks, like text classification (Wang et al., 2021) and entity typing (Ding et al., 2021). We would also like to examine the performance of ConnPrompt and competitors with few shot learning. We adopt down-sampling to construct smaller training datasets; While the development set and test set remain unchanged.

Fig. 4 summarizes the few shot learning results. It is not unexpected that both our ConnPrompt and competitors suffer from the reduction of training data. The left column presents the ConnPrompt results when using different PLMs. It is again observed that the RoBERTa is still the best PLM choice for the ConnPrompt with fewer training data. The center column compares the ConnPrompt built upon BERT with two competitors, viz., IPAL and PLR, also employing BERT as their PLM. We select the single-prompt with the best performance in the validation dataset for comparison. It is ob-

served that the performance improvements of ConnPrompt are quite significant in few shot learning. In particular, when using 10% training data, the ConnPrompt (F1 48.32% and Acc 50.41%) outperforms the IPAL and PLR using 50% training data, (IPAL: F1 45.53% and Acc 50.00%; PLR: F1 47.13% and Acc 50.07%). Similar results can also be observed for ConnPrompt built upon RoBERTa (right column). These results validate the effectiveness of ConnPrompt even with fewer training data.

5 Related Work

5.1 Implicit Discourse Relation Recognition

The pre-train and fine-tuning paradigm for the IDRR task is usually approached as a classification problem, and the key is to construct a downstream task model built-upon some PLM for the argument representation learning.

Deep learning models have prevailed for their capabilities of automatic learning argument representation upon PLM (Zhang et al., 2015; Rutherford et al., 2017). For example, the SCNN model (Zhang et al., 2015) obtains each argument representation via a single convolution layer, and the concatenation of two arguments’ representations is used for relation classification. Rutherford et al. (2017) employ a LSTM network to capture word contextual semantics for argument representation. Some hybrid models have attempted to combine CNN, LSTM, graph convolutional networks and etc., for more sophisticated argument representation (Zhang et al., 2021; Shi and Demberg, 2019; Jiang et al., 2021). These approaches, however, have ignored the fact that different words may contribute differently in argument representation learning.

Attention mechanisms can guide a neural model to unequally encode each word according to its contextual importance for argument representa-

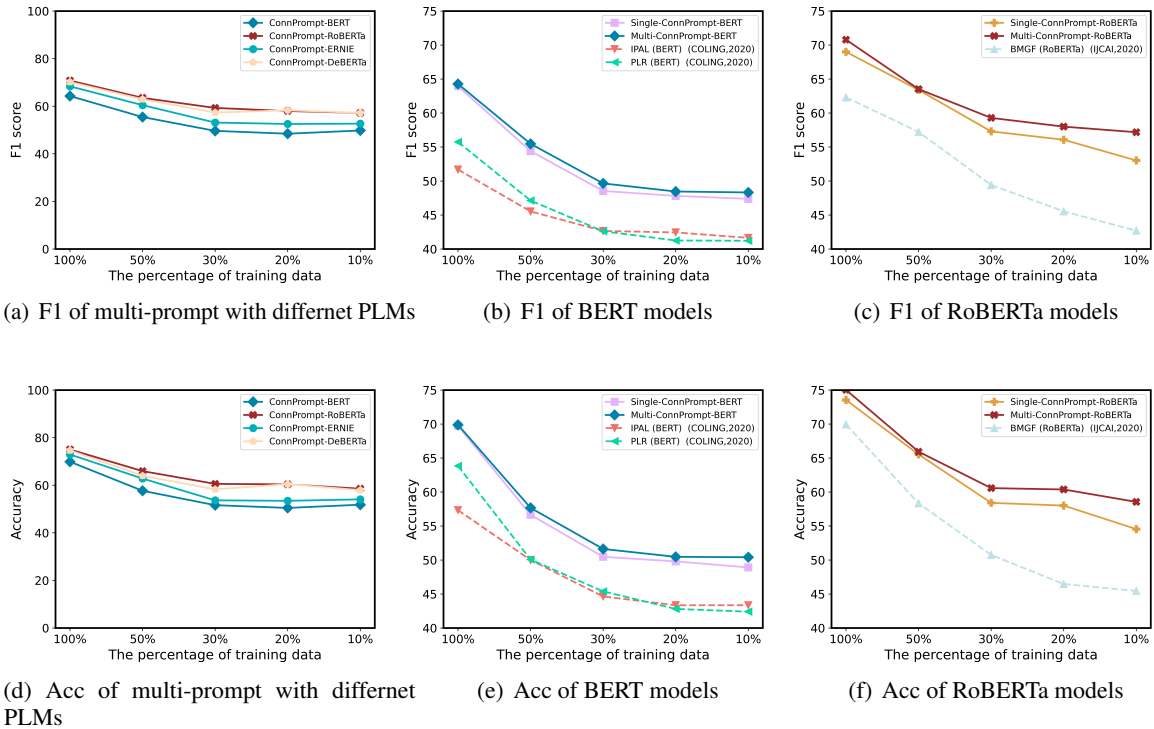


Figure 4: Performance comparison of few shot learning on the PDTB corpus.

tion (Zhou et al., 2016; Liu and Li, 2016; Lan et al., 2017; Guo et al., 2020; Ruan et al., 2020; Li et al., 2020). For example, Zhou et al. (2016) apply self-attention to weight a word according to its similarity to its belonging argument. Guo et al. (2020) and Liu et al. (2020) adopt an interactive attention to differentiate words in one argument, where a word is weighted according to the similarity between its encoding and another argument representation. Liu and Li (2016) design a multi-level attention to repeatedly compute word importance in a hierarchical way. Ruan et al. (2020) propose a pipeline workflow to apply interactive attention after self-attention. Li et al. (2020) use a penalty-based loss re-estimation method to regulate the attention learning.

5.2 Prompt Learning for NLPs

After the emergence of large-scale PLMs like BERT (Devlin et al., 2019), RoBERTa (Liu et al., 2019), ERNIE (Sun et al., 2019) and etc., the prompt learning has become a new paradigm for some NLP tasks, which use the probability of text in PLMs to perform a prediction task (Seoh et al., 2021; Wang et al., 2021; Ding et al., 2021). For example, Seoh et al. (2021) propose a cloze question prompt and a natural language inference prompt for

aspect-based sentiment analysis. Wang et al. (2021) propose a transferable prompting framework to capture cross-task knowledge for few-shot text classification. Ding et al. (2021) apply a cloze-style prompt learning on fine-grained entity typing in fully supervised, few-shot and zero-shot scenarios. Up to now, prompt learning has achieved promising results on some NLP tasks, but has not been reported for the IDRR task to the best of our knowledge.

The proposed ConnPrompt transforms the relation prediction task as a connective-cloze task against the prompt learning paradigm.

6 Conclusion

This is the first paper on examining the pre-trained, prompt, predict paradigm for the IDRR task. We have developed a Connective-cloze Prompt (ConnPrompt) to transform the IDRR task as a connective-cloze prediction task based on a pre-trained language model (PLM). Two styles of manually designed prompt template: Insertion Connective Prompt and Prefix Connective Prompt, have been designed to convert input argument pairs into the prompt formulation, and a discrete answer space is constructed with sixteen answer words for verbalizer. Experiments on the PDTB corpus

have validated that our ConnPrompt can significantly outperform the state-of-the-art algorithms, even with fewer training data.

This paper has applied the basic techniques of prompt learning for the IDRR task. In the last year, the prompt paradigm has achieved some new interesting advances, covering the techniques for choosing pre-trained models, designing continuous prompt templates, constructing answer space as well as training and tuning strategies. We note that some of these new techniques shall also be examined and improved for the IDRR task. Besides, the excellent performance of the prompt learning in this paper also motivates us to further investigate its applications in other NLP tasks.

Acknowledgements

This work is supported in part by National Natural Science Foundation of China (Grant No: 62172167). We also want to use our ConnPrompt model on MindSpore⁹, which is a new deep learning computing framework. These problems are left for future work.

References

- Hongxiao Bai and Hai Zhao. 2018. [Deep enhanced representation for implicit discourse relation recognition](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 571–583.
- Jifan Chen, Qi Zhang, Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. 2016. [Implicit discourse relation detection via a deep architecture with gated relevance network](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 1726–1735.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.
- Ning Ding, Yulin Chen, Xu Han, Guangwei Xu, Pengjun Xie, Hai-Tao Zheng, Zhiyuan Liu, Juanzi Li, and Hong-Gee Kim. 2021. [Prompt-learning for fine-grained entity typing](#). *arXiv preprint*, arXiv:2108.10604:1–12.
- Fengyu Guo, Ruifang He, Jianwu Dang, and Jian Wang. 2020. [Working memory-driven neural networks with a novel knowledge enhancement paradigm for implicit discourse relation recognition](#). In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, pages 7822–7829.
- Francisco Guzmán, Shafiq Joty, Lluís Màrquez, and Preslav Nakov. 2014. [Using discourse structure improves machine translation evaluation](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 687–698.
- Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. [Deberta: decoding-enhanced bert with disentangled attention](#). In *9th International Conference on Learning Representations*, pages 1–23.
- Yangfeng Ji and Jacob Eisenstein. 2015. [One vector is not enough: Entity-augmented distributed semantics for discourse relations](#). *Transactions of the Association for Computational Linguistics*, 3:329–344.
- Feng Jiang, Yaxin Fan, Xiaomin Chu, Peifeng Li, and Qiaoming Zhu. 2021. [Not just classification: Recognizing implicit discourse relation on joint modeling of classification and generation](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, page 2418–2431.
- Man Lan, Jianxiang Wang, Yuanbin Wu, Zheng-Yu Niu, and Haifeng Wang. 2017. [Multi-task attention-based neural networks for implicit discourse relationship representation and identification](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1299–1308.
- Wenqiang Lei, Xuancong Wang, Meichun Liu, Ilija Ilicvski, Xiangnan He, and Min-Yen Kan. 2017. [Swim: a simple word interaction model for implicit discourse relation recognition](#). In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pages 4026–4032.
- Brian Lester, Rami Al-Rfou, and Noah Constant. 2021. [The power of scale for parameter-efficient prompt tuning](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3045–3059.
- Xiao Li, Yu Hong, Huibin Ruan, and Zhen Huang. 2020. [Using a penalty-based loss re-estimation method to improve implicit discourse relation classification](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1513–1518.
- Maria Liakata, Simon Dobnik, Shyamasree Saha, Colin Batchelor, and Dietrich Rebholz Schuhmann. 2013. [A discourse-driven content model for summarising scientific articles evaluated in a complex question answering task](#). In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 747–757.

⁹<http://www.mindspore.cn/>

- Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021. [Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing](#). *arXiv preprint*, arXiv:2107.13586:1–46.
- Xin Liu, Jiefu Ou, Yangqiu Song, and Xin Jiang. 2020. [On the importance of word and sentence representation learning in implicit discourse relation classification](#). In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, pages 3830–3836.
- Yang Liu and Sujian Li. 2016. [Recognizing implicit discourse relations via repeated reading: Neural networks with multi-level attention](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1224–1233.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. [Roberta: A robustly optimized BERT pretraining approach](#). *arXiv preprint*, arXiv:1907.11692:1–13.
- Ilya Loshchilov and Frank Hutter. 2019. [Decoupled weight decay regularization](#). In *7th International Conference on Learning Representations*, pages 1–18.
- Emily Pitler, Mridhula Raghupathy, Hena Mehta, Ani Nenkova, Alan Lee, and Aravind K. Joshi. 2008. [Easily identifiable discourse relations](#). In *Proceedings of the 22nd International Conference on Computational Linguistics*, pages 87–90.
- Huibin Ruan, Yu Hong, Yang Xu, Zhen Huang, Guodong Zhou, and Min Zhang. 2020. [Interactively-propagative attention learning for implicit discourse relation recognition](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3168–3178.
- Attapol Rutherford, Vera Demberg, and Nianwen Xue. 2017. [A systematic study of neural discourse models for implicit discourse relation](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 281–291.
- Ronald Seoh, Ian Birle, Mrinal Tak, Haw-Shiuan Chang, Brian Pinette, and Alfred Hough. 2021. [Open aspect target sentiment classification with natural language prompts](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6311–6322.
- Wei Shi and Vera Demberg. 2019. [Learning to explicit connectives with seq2seq network for implicit discourse relation classification](#). In *Proceedings of the 13th International Conference on Computational Semantics - Long Papers*, page 188–199.
- Yu Sun, Shuohuan Wang, Yu-Kun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. 2019. [ERNIE: enhanced representation through knowledge integration](#). *arXiv preprint*, arXiv:1904.09223:1–8.
- Chang Wang and Bang Wang. 2020. [An end-to-end topic-enhanced self-attention network for social emotion classification](#). In *Proceedings of The Web Conference 2020*, pages 2210–2219.
- Chengyu Wang, Jianing Wang, Minghui Qiu, Jun Huang, and Ming Gao. 2021. [Transprompt: Towards an automatic transferable prompting framework for few-shot text classification](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2792–2802.
- Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. 2019. [The penn discourse treebank 3.0 annotation manual](#). Philadelphia, University of Pennsylvania, pages 1–81.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. [Transformers: State-of-the-art natural language processing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45.
- Changxing Wu, Liuwen Cao, Yubin Ge, Yang Liu, Min Zhang, and Jinsong Su. 2022. [A label dependence-aware sequence generation model for multi-level implicit discourse relation recognition](#). In *Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence*, pages 11486–11494.
- Wei Xiang and Bang Wang. 2019. [A survey of event extraction from text](#). *IEEE Access*, 7:173111–173137.
- Biao Zhang, Jinsong Su, Deyi Xiong, Yaojie Lu, Hong Duan, and Junfeng Yao. 2015. [Shallow convolutional neural network for implicit discourse relation recognition](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2230–2235.
- Yingxue Zhang, Fandong Meng, Li Peng, Jian Ping, and Jie Zhou. 2021. [Context tracking network: Graph-based context modeling for implicit discourse relation recognition](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association Computational Linguistics*, pages 1592–1599.
- Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. 2016. [Attention-based bidirectional long short-term memory networks for relation classification](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, page 207–212.