

# CCL23-Eval 任务3系统报告：基于多任务pipeline策略的汉语框架语义解析

黄舒坦 邵艳秋\* 李炜  
北京语言大学/信息科学学院,  
国家语言资源监测与研究平面媒体中心,  
北京市海淀区学院路15号, 100083

shutan2022@163.com yqshao163@163.com liweitj47@blcu.edu.cn

## 摘要

本论文为2023届CCL汉语框架语义解析评测任务提供了实现方法。针对汉语框架语义解析任务是多任务的特点，考虑到各子任务之间具有较强的时序性和关联性，方法采用了多任务pipeline策略的框架结构，主要由框架分类，论元识别，角色分类三个子模块组成，分别对应框架识别，论元范围识别，论元角色识别三个子任务。本文将框架识别和论元角色识别任务建模为文本分类任务，将论元范围识别任务建模为实体识别任务。考虑到各子任务之间具有较强的时序性和关联性，方法在每个模块均充分考虑了如何利用完成其他子任务时所抽取到的特征和信息。比如在进行角色分类时，利用了框架分类模块识别出的框架类别，以及论元识别模块识别出的论元范围。考虑到目标词及其上下文语境的重要性，本文使用预训练语言模型进行finetune。观察到模型的表现不稳定，训练时使用了对抗训练等策略提升模型性能。最终A榜分数值达到71.91，B榜分数值达到70.60，排名第2，验证了本文方法的有效性。

**关键词：** 汉语框架语义解析；多任务；信息抽取；文本分类；实体识别

## System Report for CCL23-Eval Task 3: Chinese Frame Semantic Parsing Based on Multi task Pipeline Strategy

Shutan Huang Yanqiu Shao\* Wei Li

Information Science School, Beijing Language and Culture University,  
Language Resources Monitoring and Research Center,  
15 Xueyuan Road, HaiDian District, Beijing, 100083

shutan2022@163.com yqshao163@163.com liweitj47@blcu.edu.cn

## Abstract

This paper provides an implementation method for the 2023 CCL Chinese Framework Semantic Parsing Evaluation Task. Because the Chinese Framework Semantic Parsing Task is multitasking, considering the strong temporal and correlation between each subtask, our method adopts the framework structure of the multitask pipeline strategy, mainly consisting of three sub modules: framework classification, argument recognition, and role classification, corresponding to three sub tasks: framework recognition, argument range recognition, and argument role recognition. We model framework recognition and argument role recognition tasks as text classification tasks, and argument range recognition tasks as entity recognition tasks. Considering the strong temporal and correlation between each subtask, the method fully considers how to

\* 通讯作者 Corresponding Author

©2023 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

utilize the features and information extracted from completing other subtasks in each module. For example, when conducting role classification, the framework categories identified by the framework classification module and the argument range identified by the argument recognition module are utilized. Considering the importance of the target word and its contextual context, we use a pre trained language model for finetune. The unstable performance of the model was observed, and strategies such as adversarial training were used to improve the model’s performance during training. The final score of leaderboard A reached 71.91, while the score of leaderboard B reached 70.60, ranking second, verifying the effectiveness of the method proposed in this paper.

**Keywords:** Chinese Frame Semantic Parsing , Multitask , Information Extraction , Text classification , Entity Recognition

## 1 引言

框架语义解析(Gildea and Jurafsky, 2000)是自然语言处理领域中的一项重要任务，其目标是从句中提取框架语义结构，实现对句子中涉及到的事件或情境的深层理解，在文本摘要(Guan et al., 2021a)(Guan et al., 2021b)、关系抽取(Zhao et al., 2020)和阅读理解(Guo et al., 2020a)(Guo et al., 2020b)等下游任务有着重要意义。汉语框架语义解析是基于汉语框架语义资源的语义解析任务，该任务分为以下三个子任务：框架识别、论元范围识别和论元角色识别。框架识别任务的目标是识别句子中给定目标词激活的框架，论元范围识别任务的目标是识别句子中给定目标词所支配论元的边界范围，论元角色识别任务的目标是预测论元范围识别任务所识别论元的语义角色标签。

目前对于框架语义解析的研究主要有以下方法：多种联合学习模型、多任务的pipeline策略和基于框架知识建模方法等。针对汉语框架语义解析任务是多任务的特点，且考虑到各个子任务之间具有较强的时序性和关联性，本文提出了一种基于多任务pipeline策略的汉语框架语义解析方法。方法采用了多任务pipeline策略的框架结构，主要由框架分类，论元识别，角色分类三个子模块组成，分别对应框架识别、论元范围识别和论元角色识别三个子任务。

其中，框架分类模块根据给定句子中的目标词及其上下文语境，为其寻找一个可以激活的框架，实现了框架识别，抽取出了框架特征。论元识别模块在目标词已知的条件下，从句子中自动识别出目标词所搭配的语义角色的边界，实现了论元范围识别，抽取出了论元特征。角色分类模块充分利用了框架分类模块和论元识别模块所抽取到的框架特征和论元特征，确定了句子中每个论元对应的框架元素，实现了论元角色识别。



Figure 1: 整体流程框架图

本文的贡献总结如下，

- 本文为2023届CCL汉语框架语义解析评测任务提供了实现方法，方法采用了多任务pipeline策略的框架结构，将框架识别和论元角色识别任务建模为文本分类任务，将论元范围识别任务建模为实体识别任务，并针对3个子任务分别进行了模型选择实验和消融实验
- 本文在框架识别子任务中，充分利用目标词信息，尝试使用EMA、Warm-up策略和FGM对抗训练提升效果，进行了消融实验，证明了对抗训练可有效提升模型在框架识别任务上的效果
- 本文在论元范围识别子任务中，针对任务特点尝试使用Span标注、MRC标注和Softmax标注，进行了模型选择实验，证明了Softmax标注在本次评测任务上效果最佳
- 本文在论元角色识别子任务中，对于模型输入文本的不同标注、插入和拼接方式，进行了消融实验，证明了在输入文本中插入框架类别信息、目标词信息和论元范围信息可以有效提升论元角色识别的效果

## 2 相关工作

近些年，在汉语框架语义解析任务中，传统机器学习算法中的条件随机场模型(李济洪et al., 2010)和最大熵模型(王蔚林, 2010)获得不错的效果。深度学习算法中的神经网络(王臻et al., 2014)，以及在此基础上充分利用词信息的方法(党帅兵, 2015)也取得了不错的效果。

基于双向长短时记忆网络Bi-LSTM(张苗苗et al., 2018)的算法更加全面的考虑到了句子中的长距离依赖信息，融入了Self-Attention机制的框架(王晓晖, 2022)捕获到了句子中每个词的句法信息，提升了汉语框架语义角色标注模型的性能，得到了更好的语义解析能力。

本文针对汉语框架语义解析任务是多任务的特点，采用了多任务pipeline策略的框架结构，同时引入具有强大语义建模能力的预训练语言模型BERT进行语义解析，使模型充分考虑到各子任务之间较强的时序性和关联性，充分利用完成其他子任务时所抽取到的特征和信息，进而获得了更强的语义解析能力。

## 3 方法

本章节介绍比赛使用的具体方案，包括框架分类，论元识别，角色分类三个部分。分别对应框架识别，论元范围识别，论元角色识别三个子任务。

### 3.1 框架分类

考虑到目标词对框架具有重要的触发和激活作用，为了充分利用先验知识，本文给Bert模型(Liu et al., 2019)的词表添加了<t>和</t>作为额外的Special Token，并用它们将句子中的目标词包裹起来作为输入文本，以此来提升模型在给定目标词的条件下的分类性能。因为目标词的上下文语境对框架分类的准确率非常重要，所以本文将目标词的所有Token进行平均池化后，作为特征向量输入线性分类层进行分类。

通过对数据集进行数据分析发现，存在少数句子同时属于多个框架类别的情况，但是实验发现多标签分类效果较差。进行原因分析，可能因为训练集中绝大多数句子只属于一个框架，且目标词对框架的影响较大，即使一个句子拥有多个框架类别标签，不同框架类别的目标词基本不同。

如果使用多标签分类，难以利用不同目标词的信息，且样本标签过于稀疏，不利于模型的训练。所以最终方案使用单标签多分类而不是多标签分类。

观察到模型的表现不稳定，为了提高模型鲁棒性，提升模型的性能，本文在训练中还使用了以下优化策略：

#### (1) 指数滑动平均Exponential Moving Average(EMA)

指数滑动平均EMA是一种通过给予近期数据更高权重，对模型参数做平均的方法，使得模型参数的更新与一段时间内的历史取值有关，可以提高测试指标并增加模型鲁棒性。

#### (2) 学习率预热Warm-up

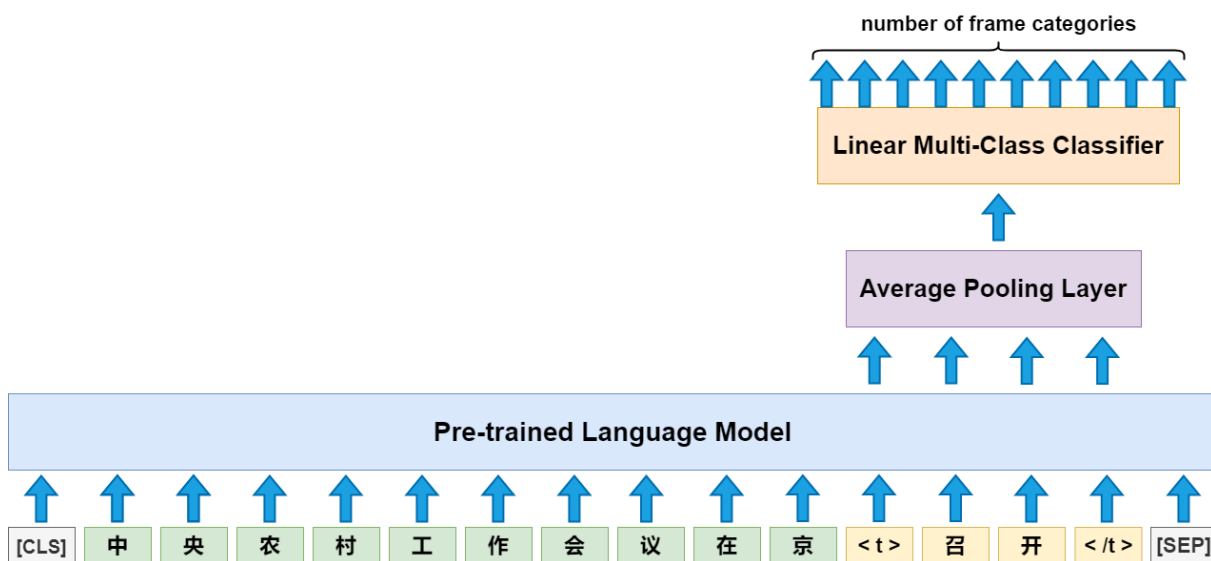


Figure 2: 框架分类模型

训练时使用了学习率Warm-up策略，在训练开始前先使用一个较小的学习率进行一定的迭代次数，以使得模型逐渐适应数据集的特征，使得模型的权重更新更加平稳，减少训练时的震荡和不稳定性，从而提高了模型的训练效果。

### (3) 快速梯度法Fast Gradient Method(FGM)

训练时使用了FGM(Miyato et al., 2017)对抗训练，对embedding层在梯度方向添加扰动，引入噪声，这种训练方式既提高了模型的泛化能力，又提高了模型的鲁棒性。

## 3.2 论元识别

在整个pipeline语义解析框架中，论元识别起着至关重要的作用。由于误差的传递性，是否成功召回论元实体对后续角色分类的准确率影响较大。因此，在训练过程中本文优先保证召回率，保存召回率最优的模型。

具体实现上采用了序列标注模型，对输入文本以Token为单位进行BIO标注，标注出论元范围。其中，该模块的输入文本和框架分类模块的输入相同，是使用<t>和</t>作为额外的Special Token标识出了目标词范围的句子。

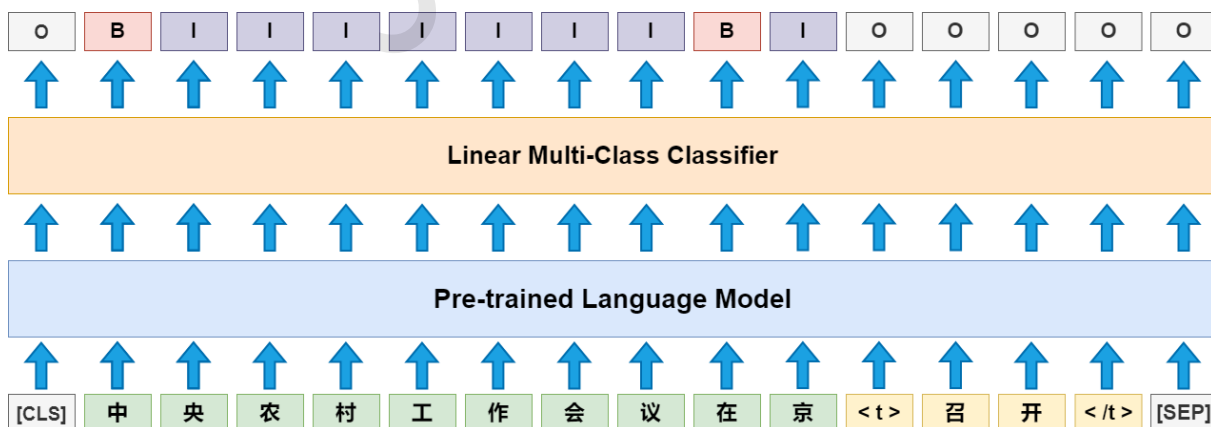


Figure 3: 论元识别模型

### 3.3 角色分类

角色分类模块作为pipeline框架的最后一部分，为了能充分利用框架分类模块和论元识别模块的预测结果，本文方法修改了模型输入的文本。

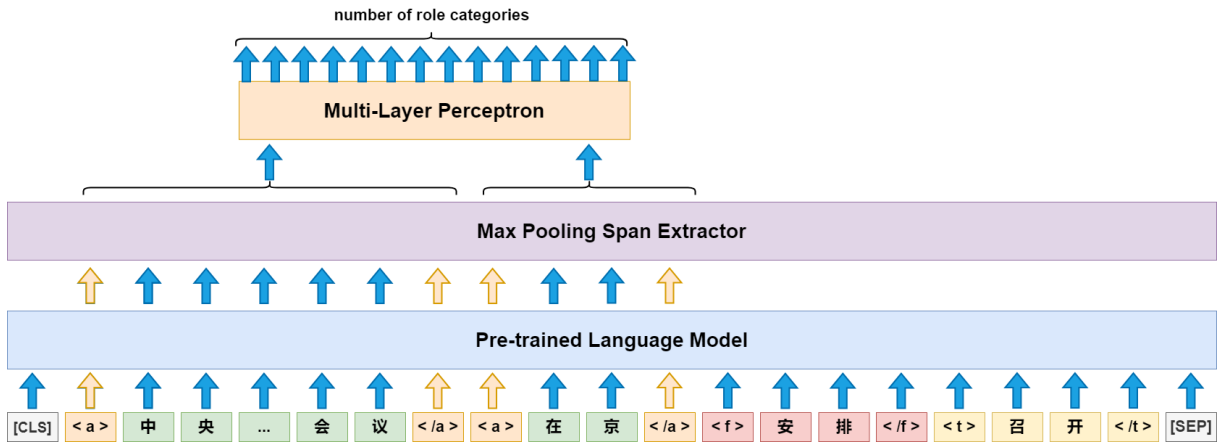


Figure 4: 角色分类模型

首先，给基座模型Bert的词表添加<t>和</t>用于标识目标词，添加<a>和</a>用于标识论元范围，添加<f>和</f>用于标识框架信息。然后，将框架分类模块的预测结果作为框架特征、将论元识别模块的预测结果作为论元特征加入到输入文本中。具体实现是将框架类别插入到目标词前，然后使用<t>和</t>包裹目标词，使用<a>和</a>包裹论元，使用<f>和</f>包裹框架类别。最后，对论元以Span为单位进行Max Pooling，将最终得到特征向量传递给激活函数为ReLU的神经网络层进行分类。

比如，对于输入文本“中央农村工作会议在京召开”，本文将该句子的框架类别“安排”插入到目标词“召开”前，使输入文本变为“中央农村工作会议在京安排召开”，并且使用Special Token分别包裹目标词、论元和框架类别。

## 4 实验

### 4.1 数据集

CFN1.0(Chinese FrameNet, CFN)数据集是由山西大学以汉语真实语料为依据构建的框架语义资源，数据由框架知识及标注例句组成，包含了近700个语义框架及20000条标注例句。其中，训练集包含标注数据10000条，验证集包含标注数据2000条。其中，训练集中有418个框架类别是样本数小于10的few-shot框架标签，有66个框架类别是zero-shot框架标签。存在标签分布不均的情况，但训练集和验证集框架类别的标签分布基本一致。A榜测试集和B榜测试集各包含4000条样本。

### 4.2 框架识别效果

本文使用Roberta模型，尝试使用CLS向量和目标词Avg Pooling向量进行分类。其中，使用目标词Avg Pooling向量分类效果提升明显，Acc提升16.3%。训练时使用EMA、Warm-up策略和FGM对抗训练，Acc提升7.7%。消融实验结果如下：

Table 1: A榜测试集

Model	Acc
Roberta+CLS	48.51%
Roberta+Avg Pooling	64.87%
+EMA & Warm-up	68.44%
+EMA & Warm-up & FGM	72.50%

Table 2: B榜测试集

Model	Acc
Roberta+Avg Pooling	65.31%
+EMA & Warm-up & FGM	72.33%

Table 3: 各队伍结果对比

队伍名	Acc
四川大学	74.27%
苏州大学	70.58%
哈尔滨工业大学	65.13%
国际关系学院	65.86%
本文方法	71.77%

### 4.3 论元范围识别效果

对于论元范围识别，本文尝试了Span标注、MRC标注(Li et al., 2019)(Li et al., 2020)和Softmax序列标注等方式，其中Softmax序列标注效果最佳。针对数据集的标签分布不均，尝试换用Focal Loss(Lin et al., 2017)进行训练，但模型效果未见明显提升。

其中，Span标注的解码部分使用2个线性层作为Language Model Head分别预测Span的起始位置和终止位置。MRC标注则尝试在输入文本前拼接上“请找出句子中论元的范围：”作为Query，以句子级的原始文本作为Document，以阅读理解的方式来处理序列标注任务。解码部分和Span标注一样，使用2个线性层分别预测Span的起始位置和终止位置。考虑到数据集中重叠实体较少，采用严格解码形式，对于重叠实体选取Logits最大的一个，保证精确率。实验表明，在本次评测数据集上，MRC标注的精确率较高，但召回率和F1值不如Softmax序列标注，且训练和推理效率较低。

消融实验结果如下：

Table 4: A榜测试集

Model	P	R	F1
Roberta+MRC	<b>91.37%</b>	84.56%	87.83%
Roberta+Span	89.21%	85.88%	87.51%
Roberta+Softmax	90.08%	<b>86.60%</b>	<b>88.31%</b>
Roberta+Softmax+Focal Loss	89.98%	86.50%	88.20%

Table 5: B榜测试集

Model	P	R	F1
Roberta+Softmax	90.37%	84.27%	87.21%

Table 6: 各队伍结果对比

队伍名	P	R	F1
四川大学	90.78%	82.29%	86.33%
苏州大学	89.27%	82.24%	85.61%
哈尔滨工业大学	90.23%	85.39%	87.74%
国际关系学院	90.47%	85.61%	87.97%
本文方法	90.35%	84.68%	87.42%

#### 4.4 论元角色识别效果

针对模型的输入文本是否插入框架类别以及是否使用Special Token进行包裹, 本文进行了消融实验。首先使用Special Token中的<t >和</t >将目标词包裹, 使用<a >和</a >将论元角色包裹。将处理后的文本输入模型, 得到的评测指标相比于未使用Special Token进行输入文本标识的指标提升明显, f1值提升约4.5。

其次, 使用Special Token中的<f >和</f >包裹框架类别后插入到输入文本中, 为输入文本引入框架类别信息。论元角色识别效果提升明显, f1值提升约2.2。

消融实验的结果如下:

Table 7: A榜和B榜测试集

Input	P	R	F1
text	52.12%	52.32%	52.21%
+target word & argument span	56.58%	56.74%	56.65%
+target word & argument span & frame(A榜测试集)	58.76%	58.98%	58.87%
+target word & argument span & frame(B榜测试集)	57.90%	55.82%	56.84%

Table 8: 各队伍结果对比

队伍名	P	R	F1
四川大学	59.58%	57.01%	58.27%
苏州大学	63.57%	49.27%	55.51%
哈尔滨工业大学	52.25%	52.45%	52.35%
国际关系学院	51.12%	50.41%	50.76%
本文方法	57.06%	55.52%	56.28%

## 5 结论

针对汉语框架语义解析任务, 本文引入具有强大语义建模能力的预训练语言模型BERT来更好地进行语义解析, 在此基础上, 针对汉语框架语义解析任务是多任务的特点, 本文采用了多任务pipeline策略的框架结构, 使模型充分利用了完成其他子任务时所抽取到的特征和信息, 并验证了方法的有效性。

但是, 本文方法还存在着不足, 虽然预训练语言模型有着强大的性能, 在语义解析能力上还是存在优化的空间。针对数据集中few-shot和zero-shot的框架标签, 容易出现分类错误的情况。因此在未来的研究中会更加关注如何更好的提升模型的语义解析能力, 除此之外, 也会对如何提升模型对few-shot和zero-shot框架标签的语义解析能力进行进一步的研究。

## 参考文献

Daniel Gildea and Daniel Jurafsky. 2000. Automatic labeling of semantic roles. In *38th Annual Meeting of the Association for Computational Linguistics, Hong Kong, China, October 1-8, 2000*, pages 512-520. ACL.

- Yong Guan, Shaoru Guo, Ru Li, Xiaoli Li, and Hongye Tan. 2021a. Frame semantic-enhanced sentence modeling for sentence-level extractive text summarization. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4045–4052, Online and Punta Cana, Dominican Republic, November. Association for Computational Linguistics.
- Yong Guan, Shaoru Guo, Ru Li, Xiaoli Li, and Hu Zhang. 2021b. Integrating semantic scenario and word relations for abstractive sentence summarization. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2522–2529, Online and Punta Cana, Dominican Republic, November. Association for Computational Linguistics.
- Shaoru Guo, Yong Guan, Ru Li, Xiaoli Li, and Hongye Tan. 2020a. Incorporating syntax and frame semantics in neural network for machine reading comprehension. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2635–2641, Barcelona, Spain (Online), December. International Committee on Computational Linguistics.
- Shaoru Guo, Ru Li, Hongye Tan, Xiaoli Li, Yong Guan, Hongyan Zhao, and Yueping Zhang. 2020b. A frame-based sentence representation for machine reading comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 891–896, Online, July. Association for Computational Linguistics.
- Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and Jiwei Li. 2019. Entity-relation extraction as multi-turn question answering. In Anna Korhonen, David R. Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 1340–1350. Association for Computational Linguistics.
- Xiaoya Li, Jingrong Feng, Yuxian Meng, Qinghong Han, Fei Wu, and Jiwei Li. 2020. A unified MRC framework for named entity recognition. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 5849–5859. Association for Computational Linguistics.
- Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 2999–3007. IEEE Computer Society.
- 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. *CoRR*, abs/1907.11692.
- Takeru Miyato, Andrew M. Dai, and Ian J. Goodfellow. 2017. Adversarial training methods for semi-supervised text classification. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.
- Hongyan Zhao, Ru Li, Xiaoli Li, and Hongye Tan. 2020. Cfsre: Context-aware based on frame-semantics for distantly supervised relation extraction. *Knowledge-Based Systems*, 210:106480.
- 党帅兵. 2015. 基于词分布表征的汉语框架语义角色识别研究. 硕士论文, 山西大学.
- 张苗苗, 张玉洁, 刘明童, 徐金安, and 陈钰枫. 2018. 基于gate 机制与bi-lstm-crf 的汉语语义角色标注. *计算机与现代化*, (4):1–6.
- 李济洪, 王瑞波, 王蔚林, and 李国臣. 2010. 汉语框架语义角色的自动标注. *软件学报*, 21(4):597–611.
- 王晓晖. 2022. 基于self-attention的句法感知汉语框架语义角色标注. *中文信息学报*, 36(10):38–44.
- 王臻, 常宝宝, and 穗志方. 2014. 基于分层输出神经网络的汉语语义角色标注. *中文信息学报*, 28(6):56–61.
- 王蔚林. 2010. 基于最大熵模型的汉语框架语义角色自动标注. Ph.D. thesis, 太原: 山西大学硕士学位论文.