

CCL23-Eval 任务3总结报告:汉语框架语义解析评测

李俊材^{1,‡}, 闫智超^{1,‡}, 苏雪峰^{1,3,‡}, 马博翔^{1,‡}, 杨沛渊^{1,‡}, 李茹^{1,2,*†}

¹山西大学 计算机与信息技术学院, 山西 太原 030006

²山西大学 计算智能与中文信息处理教育部重点实验室, 山西 太原 030006

³山西工程科技职业大学现代物流学院, 山西 晋中 030609

[‡]{202122407024, 202022408073, 201912407008, 202222405024}@email.sxu.edu.cn

[‡] 202222407058@email.sxu.edu.cn ; *liru@sxu.edu.cn

摘要

汉语框架语义解析评测任务致力于提升机器模型理解细粒度语义信息的能力。该评测数据集包括20000条标注的框架语义解析例句和近700个框架信息。评测任务分为框架识别、论元范围识别和论元角色识别三个子任务，最终成绩根据这三个任务的得分综合计算。本次评测受到工业界和学术界的广泛关注，共有55支队伍报名参赛，其中12支队伍提交了结果，我们选取5支队伍的模型进行结果复现，最终来自四川的李作恒以71.49的分数排名第一。该任务的更多信息，包括系统提交、评测结果以及数据资源，可从CCL-2023汉语框架语义解析评测任务网址¹查看。

关键词： 汉语框架网；框架识别；汉语框架语义解析；框架角色标注

Overview of CCL23-Eval Task 1: Chinese FrameNet Semantic Parsing

Junca Li^{1,‡}, Zhichao Yan^{1,‡}, Xuefeng Su^{1,3,‡}, Boxiang Ma^{1,‡}, Peiyuan Yang^{1,‡}, Ru Li^{1,2,*†}

¹School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi 030006, China

²Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, Shanxi 030006, China

³School of Modern Logistics, Shanxi Vocational University of Engineering Science and Technology, Jinzhong, Shanxi 030609, China

[‡]{202122407024, 202022408073, 201912407008, 202222405024}@email.sxu.edu.cn

[‡] 202222407058@email.sxu.edu.cn ; *liru@sxu.edu.cn

Abstract

The Chinese Frame Semantic Parsing Evaluation Task aims to enhance the machine models' ability to understand fine-grained semantic information. The evaluation dataset consists of 20,000 annotated examples of frame semantic parsing and nearly 700 frame annotations. The evaluation task is divided into three subtasks: frame identification, argument identification, and role identification. This evaluation has attracted wide attention from both industry and academia, with a total of 55 teams participating, and 12 teams submitting their results. We selected models from 5 teams for result reproduction, and ultimately, Li Zuoheng from Sichuan ranked first with a score of 71.49. More information about this task, can be found on the website of the CCL-2023 Chinese Frame Semantic Parsing Evaluation Task.

Keywords: Chinese FrameNet, Frame Identification, Chinese Frame Semantic Parsing, Frame Role Labeling

¹任务网址: <https://tianchi.aliyun.com/competition/entrance/532083/introduction>

[†] 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

1 背景和动机

语义分析是自然语言处理的基础任务(Kate et al., 2005), 是将自然语言表达解析为结构化语义表示的过程。它能为各种自然语言处理的理解和生成任务提供一定的语义支撑, 对于阅读理解(Guo et al., 2020b; Guo et al., 2020a; 王智强et al., 2016)、文本摘要(Guan et al., 2021a; Guan et al., 2021b)、关系抽取(Zhao et al., 2020)、文本生成(谭et al., 2018)等下游任务有着重要意义。

随着技术的进步, 现有的模型在一些语义分析任务(如词性标注、依存句法分析、语义角色标注)上取得了较好的性能。如今ChatGPT和GPT-4等大模型相继出现, 在这些任务的Zero-Shot和Few-Shot场景下都取得了很好的效果, 但在细粒度语义分析方面, 这些模型的语义场景刻画能力仍然表现不佳。

框架语义解析(Frame Semantic Parsing, FSP)最早由Gildea和Jurafsky基于英文框架网数据集FrameNet提出(Gildea and Jurafsky, 2002), 是一种细粒度语义分析任务, Baker在2007年的SemEval(Baker et al., 2007)中正式提出框架语义解析评测任务。目前, FSP已有大量研究, 主要围绕FrameNet数据集(FN1.5&FN1.7)展开。如基于预训练的方法(Tan and Na, 2019; Jiang and Riloff, 2021)、基于联合学习的方法(Chen et al., 2021; Peng et al., 2018)和基于框架知识建模的方法(Su et al., 2021b; Zheng et al., 2022; Zheng et al., 2023)。

在汉语框架语义解析(Chinese Frame Semantic Parsing, CFSP)方面, 山西大学自2004年率先构建汉语框架网(Chinese FrameNet, CFN)(You and Liu, 2005)并开展相关研究。如李济洪等人基于CRF在CFN数据集上进行语义角色标注(李济洪et al., 2010), 屠寒非等人提出一种基于主动学习的方法并取得一定的效果(屠寒非et al., 2016), 王晓晖提出基于自注意力机制的汉语框架语义角色标注方法以获取句子的长距离信息(Wang et al., 2020)。

本次评测首次推出汉语框架网数据集并提出汉语框架语义解析任务。CFSP作为细粒度语义分析的一种方法, 通过识别汉语语义框架和论元角色, 提供了一种表示句子语义的结构化信息, 这种结构化信息更具表达力, 能够更好地捕捉句子中的语义关系和细粒度语义信息。

2 相关概念及任务描述

2.1 相关概念

框架语义学是认知语言学的一个重要分支, 最早由Fillmore 提出并倡导, 框架语义学把“框架”这一概念的认知结构引入语义学, 为理解词义、句义以及篇章含义提供了认知层面上的解释, 在实现计算机的语言认知理解上具有独特优势。

汉语框架网是以框架语义学为理论基础, 以汉语语料事实为依据构建的汉语框架语义知识库。目前, 汉语框架网中包含1322个框架, 框架元素1000余种。在汉语框架网中, 有以下几个重要概念。

框架: 框架框架是由词语在语言使用者大脑中所激活的图式化认知场景, 是理解和使用语言的背景和动因(郝晓燕et al., 2007)。如表1中的【安排】框架, 该框架表示施动者执行一系列不确定的任务使事件在计划时间和处所发生。

框架名称	安排	
框架定义	施动者执行一系列不确定的任务使事件在计划时间和处所发生。	
框架元素	框架元素名称	框架元素定义
	施动者	为使事件在某一时间和处所发生而做出安排的人。
	事件	施动者作出安排的事件。
	计划时间	施动者安排事件的时间。
	受益人	被安排事件的受益人。
	处所	施动者安排事件发生所处的地点。
	目的	施动者安排事件的动机。
时间	该框架元素描述了做出安排的时间。	

Table 1: 框架【安排】及其所包含的框架元素信息

框架元素: 框架所对应的语义场景中的参与者, 对应于语义角色(本文提到的框架语义角色指框架元素), 例如【安排】框架中的施动者和事件都是该框架的框架元素, 框架元素极大

的丰富了句子目标词所激发场景的语义信息。

词元：词元是指可以激活CFN 框架库中某个框架的词语。每个词元通常可以激活一个或者多个框架，但在具体的某个句子中，每个词元只能归属于某个特定的框架。在本文展示的示例中，框架【安排】包含的词元除了“组织”以外，还包含了举办、举行等词。



Figure 1: 框架语义解析任务示意图

目标词：待标注句子中可以激活框架的词语，通常为词元库中词元。如图1的例句中，“组织”是激活框架的目标词。

2.2 任务描述

汉语框架语义解析任务分为框架识别（Frame Identification, FI）、论元范围识别（Argument Identification, AI）和论元角色识别（Role Identification, RI）三个子任务。

框架识别：框架识别是给定可激活框架的目标词，根据上下文语境，从多个所属框架中选取最符合该目标词语境的语义框架的任务。如图1中框架识别部分所示，目标词“组织”可激活【安排】以及【使结合】框架。但依据上下文语境可最终确定为【安排】框架。

该任务的形式化定义如下：给定包含目标词的句子 S ，记为 $S = (w_1, w_2, \dots, w_n)$ ，其中 w_i 为组成句子的第 i 个词，其中 $1 \leq i \leq n$ 。待识别目标词记为 w_t ， $w_t \in S$ 。要求通过上下文的语义场景从给定的框架库 $F = \{f_1, f_2, \dots, f_n\}$ 中选择出合适的框架 f_t ，记为式：

$$f_t = \operatorname{argmax}_{f_i \in F, w_t \in S} P(f_i | S, w_t) \quad (3.1)$$

论元范围识别：在框架语义解析任务中，论元范围识别是指确定论元在句子中的起始位置和结束位置的子任务。即给定一条句子及目标词，在目标词已知的条件下，从句子中自动识别出目标词所支配的语义角色的边界。在图1中，目标词“组织”支配的论元包含“班级的同学”、“明天”和“参加由学院举办的学术研讨会”，而“班级”为错误的论元。

论元范围识别的形式化定义为：对于给定句子 $S = (w_1, w_2, \dots, w_n)$ 及其目标词 $w_t \in S$ ，该任务的目标是为论元 $a_\tau \in \{a_1, a_2, \dots, a_k\}$ 查找其边界范围 i_τ^s 和 i_τ^e 使 $a_\tau = w_{i_\tau^s}, \dots, w_{i_\tau^e}$

论元角色识别：论元角色识别任务是框架语义解析任务中的最后一步。该任务旨在确定句子中每个论元对应的框架元素，即每个论元在所属框架中的语义角色。如图1中“参加由学院举办的学术研讨会”的语义角色为“事件”。

该任务的形式化定义为：在给定句子 $S = (w_1, w_2, \dots, w_n)$ ，句子中目标词 $w_t \in S$ 以及目标词所激活的框架 f 时，对于已知边界范围的论元 $a_\tau = w_{i_\tau^s}, \dots, w_{i_\tau^e}$ ，为其识别出正确的角色类型（框架元素） r_τ ，其中 $a_\tau \in \{a_1, a_2, \dots, a_k\}$ ， $r_\tau \in R_f$ ， R_f 为框架 f 所包含的所有框架元素，任务定义记为式：

$$r_\tau = \operatorname{argmax}_{r_i \in R_f, w_t \in S} P(r_i | S, w_t, f_t, a_\tau) \quad (3.2)$$

3 评测数据

本次公开的CFN2.0数据来源于山西大学中文信息处理团队的汉语框架网(CFN)。CFN数据集自2004年开始不断发展，至今已形成了标注例句数量超过10万条的大规模数据集。CFN2.0数据集由框架信息和标注例句两部分组成，其语料来自涉及多种不同领域的1100多篇新闻稿件。标注内容包括目标词激活的框架以及目标词所支配的语义角色，每条标注例句均经过双盲标注、双重审核以及专家答疑的标注流程，用以保障标注数据的质量。

CFN2.0数据集规模如表2所示。需要注意的是，在统计过程中，对于相同的例句，若其目标词不同，将被视为不同的例句进行计数。

数据集划分	Train	Dev	Test_A	Test_B	ALL
例句数	10000	2000	4000	4000	20000
框架数	671	354	432	504	695
框架元素数	947	649	711	796	987
包含词元数	2359	670	931	572	3132

Table 2: CFN2.0数据集规模

在框架语义解析任务中，不同框架往往包含着不同的语义信息，同时其框架元素的组合也较为复杂多样，这些特点对框架语义分析模型提出了较高的要求。除此之外，在框架与例句的对应关系上，大量框架仅具有少数例句，如图2所示，超过半数的框架仅具有20条以下的例句，与其相对的，例句数最多的框架则具有729条例句，虽然呈现长尾分布现象，但符合人类在进行自然语言描述时的现实规律，这种现象增加了数据的复杂性。

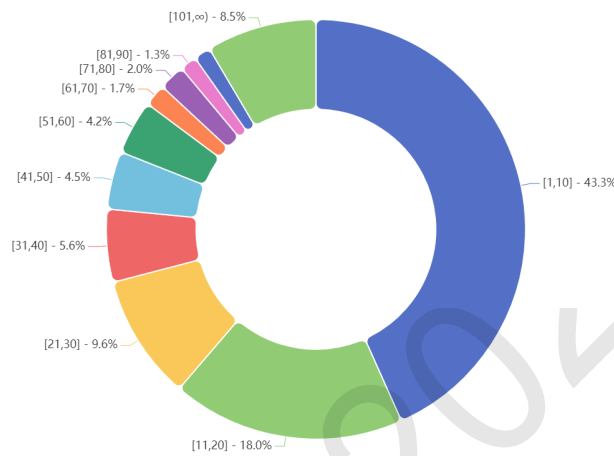


Figure 2: 框架下例句数量区间及其占比

4 评价指标

针对汉语框架语义解析的三个子任务，本次评测的评价指标主要包括框架识别正确率 (Accuracy, Acc)，论元范围识别F1值 (F1-score) 和论元角色识别F1值 (F1-score)，最后将三个子任务的得分加权求和，得到最终的评价分数。

框架识别：框架识别正确率是通过计算模型正确识别的例句数量与总体例句数量之间的比例来打分的，具体的计算公式为：

$$\text{task1_acc} = \text{correct}/\text{total} \quad (4.1)$$

其中，correct为模型预测正确的数量，total为数据总量。

论元范围识别：该任务的评价方式为计算模型识别出的论元范围和数据实际的论元范围之间的F1值，具体计算公式为：

$$\begin{aligned} \text{task2_precision} &= \frac{\text{InterSec}(\text{gold}, \text{pred})}{\text{Len}(\text{pred})} \\ \text{task2_recall} &= \frac{\text{InterSec}(\text{gold}, \text{pred})}{\text{Len}(\text{gold})} \\ \text{task2_f1} &= \frac{2 * \text{task2_precision} * \text{task2_recall}}{\text{task2_precision} + \text{task2_recall}} \end{aligned} \quad (4.2)$$

其中，gold 和pred 分别表示真实结果与预测结果，InterSec(*)表示计算二者共有的token数量，Len(*)表示计算token数量。

论元角色识别：该任务严格判定每一个论元的边界以及角色，同样以F1作为评价指标：

$$\begin{aligned} \text{task3_precision} &= \frac{\text{Count}(\text{gold}, \text{pred})}{\text{Count}(\text{pred})} \\ \text{task3_recall} &= \frac{\text{Count}(\text{gold}, \text{pred})}{\text{Count}(\text{gold})} \\ \text{task3_f1} &= \frac{2 * \text{task3_precision} * \text{task3_recall}}{\text{task3_precision} + \text{task3_recall}} \end{aligned} \quad (4.3)$$

其中，gold 和 pred 分别表示真实结果与预测结果的语义角色集合， $\text{Count}(\ast)$ 表示集合中元素的数量。

最终得分：本次评测的最终得分为三个子任务得分的加权和，具体的计算方式为：

$$\text{final_score} = 0.3 * \text{task1_acc} + 0.3 * \text{task2_f1} + 0.4 * \text{task3_f1} \quad (4.4)$$

5 提交结果

在评测期间，共计55支队伍报名参赛，有12支队伍参与A榜初赛且其中9支参赛队伍进入B榜复赛。最终，我们选取了榜单成绩前5名队伍的模型进行结果复现。

排名	参赛单位/个人	队伍编号	task1	task2			task3			final
			Acc	P	R	F1	P	R	F1	
1	李作恒(个人)	Team.1	74.28	90.79	82.29	86.33	59.59	57.01	58.27	71.49
2	北语(BLCU)	Team.2	71.77	90.35	84.69	87.43	57.06	55.53	56.28	70.27
3	苏大(SUDA)	Team.3	70.59	89.27	82.25	85.62	63.57	49.28	55.52	69.07
4	哈工大(威海)	Team.4	65.14	90.24	85.39	87.75	52.25	52.46	52.36	66.81
5	国关(UIR)	Team.5	65.87	90.47	85.62	87.98	51.12	50.41	50.76	66.46
6	Baseline	—	58.02	89.74	83.89	86.72	49.12	47.87	48.49	62.81

Table 3: 参赛队伍B榜复现成绩

表3中详细列出了这5支参赛队伍以及基线的得分情况（得分以复现结果为准），以最终分数为排名依据。其中任务2和任务3详细列出了各个参赛队伍的正确率、召回率和F1值，下文中我们将以表中队伍编号表示不同的队伍以便于后续表述。

通过对表中数据的分析可以看出，由于基线无法学习到任务1和任务3所需的深层次、细粒度的语义信息，因此，相对于基线，这5支队伍容易取得较大的提升。而基线在任务2上本身表现较好，参赛队伍很难获得显著的提高。

6 方法概述

通过对5支参赛队伍提交的技术报告进行分析以及对其模型进行结果复现，我们整理出参赛队伍所使用的主要方法，以分析不同的任务上各个队伍具有的优势。包括Team.1引入旋转位置编码进行框架识别且效果显著；Team.2使用多种优化策略使模型在各任务的表现都有一定提升；Team.3将CFSP任务转化为图解析任务并取得不错的成绩。此外，我们将评测任务在ChatGPT上进行实验分析，以评估大模型在汉语框架语义解析任务上的能力。

6.1 基于旋转位置编码的框架识别方法

针对在框架识别中会丢失目标词与整体句子之间的位置信息关系的问题，参赛队伍Team.1使用了一种基于旋转式位置编码(Su et al., 2021a)的方法来计算实体之间注意力信息然后进行分类和抽取，该方法使用旋转矩阵对绝对位置进行编码，同时将显式的相对位置依赖性纳入自注意公式中，模型结构如图3所示。结果显示，在引入旋转式位置编码以后，模型对于框架识别等任务有较为显著的提升。

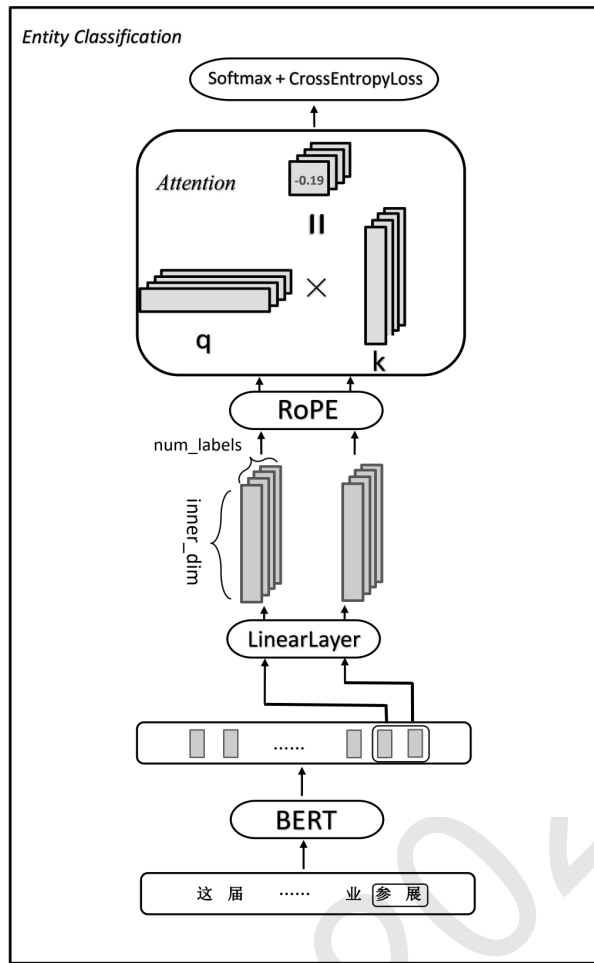


Figure 3: 基于RoPE的实体分类框架图

6.2 基于多种优化策略的方法

为解决模型不稳定的问题，提高模型鲁棒性，*Team.2*在训练过程中使用了多种优化策略，通过设置滑动平均、在梯度方向添加扰动等方式，减少训练时的震荡，提高模型的泛化能力。

指数滑动平均：指数滑动平均(Exponential Moving Average,EMA)是一种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，可以提高模型在测试集上的鲁棒性。

Warm-up策略：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。

基于快速梯度上升的对抗训练：训练时使用了FGM(Miyato et al., 2021)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。

6.3 基于词的图解析方法

*Team.3*为了提升模型在多个子任务上的性能，并确保句子语义的结构化表达，采用了BES的图表示结构(Zhou et al., 2021)。他们将基于片段的语义框架解析转换成了一个基于词的图解析任务，并在端到端框架中将句子中目标词对应的框架和论元识别放在一起。

图4-a和图4-b描述了该方法如何表示框架语义解析的标注结果。首先，在句子开始位置添加了“root”标签，然后将其与目标词相连，接下来将目标词与论元相连。对于目标词与论元之间的连接边的构造，当论元为单个词时，使用“S-r”进行连接；否则，使用“B-r”和“E-r”分别连接论元中开始位置和结束位置的词。

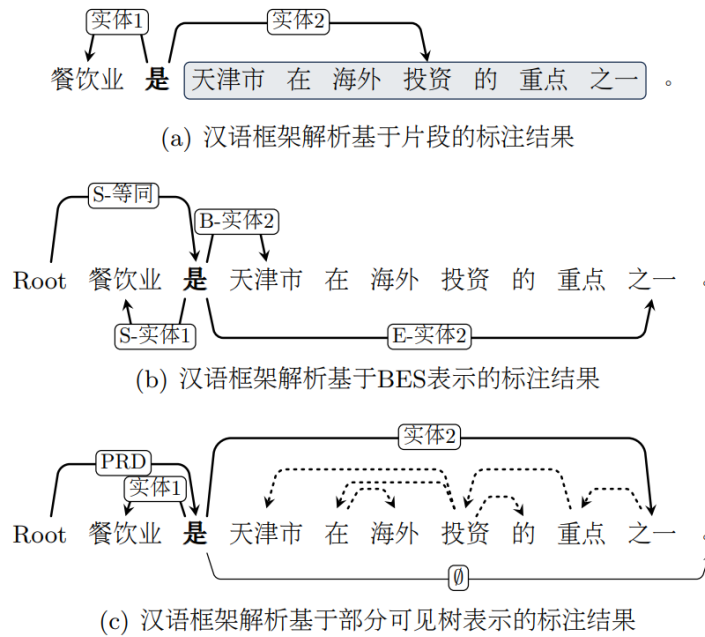


Figure 4: 汉语框架语义解析转化为树结构解析任务示例

同时, *Team.3*提出, 针对论元识别的任务2和任务3, 显式的建模论元内部结构可能对框架解析有帮助。因此借鉴一种角色标注方法(Zhang et al., 2021), 将基于span的框架语义解析转换为树结构解析任务, 将序列形式的论元建模为树结构(图4-c), 对于多个词的论元, 将所有潜在的论元子树作为目标词的后代, 并为其使用Eisner算法(Eisner, 2000)找出最高得分的子树, 将该论元片段的语义角色标签分配到目标词和子树的词头“之一”之间的边上, 最后利用抽取子树的后代来识别论元的范围, 并将目标词和词头之间的标签作为论元的标签。

6.4 基于多任务预训练的方法

来自哈尔滨工业大学的*Team.4*, 鉴于任务2、任务3与传统的NER任务相似, 因此对预训练模型进行进一步修改。该队伍采用多任务预训练策略来提高模型的性能, 具体的, 该队伍使用LERT(Cui et al., 2022)作为预训练模型, 除了原始的Bert预训练任务外, LERT还进行了词性标注 (Part-of-Speech tagging, POS), 命名实体识别 (Named Entity Recognition, NER) 和依存分析 (Dependency Parsing, DEP)。LERT在不同的汉语自然语言理解任务中表现出显著的提高, 在本次评测中的任务2和任务3上也分别取得了88.53和57.93的F1值。

6.5 基于数据增强的方法

在CFN2.0数据集中不同框架所包含例句的数量呈现出典型的长尾分布, 如常见的【陈述】框架拥有超过300条例句, 然而超过半数的框架在训练集中出现的次数不超过十次, 个别框架如【上下级关系】更是只出现了一次。因此, 就框架识别而言, 大部分框架需要在低资源场景下进行。为解决此问题, *Team.5*采用了针对包含例句少的框架的数据扩充方法, 该方法将出现频率在10到20之间的数据重复3次, 将出现频率小于10的数据重复10次。在数据扩展之后共包含大约26,000个标注实例。如图5所示, 增强后的数据集更加平滑, 长尾效应比原始数据集明显减弱。这在一定程度上降低了分类任务的难度。

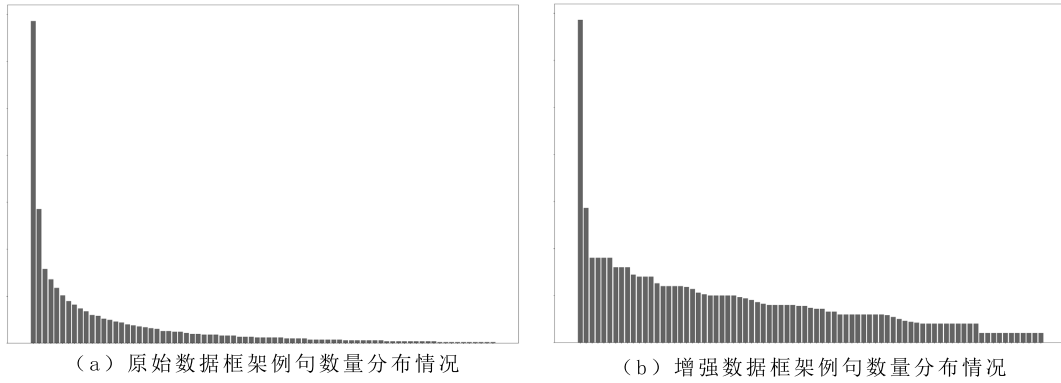


Figure 5: 数据增强前后框架例句分布图

Team.3则通过从CoNLL09(Hajic et al., 2009)数据集中抽取一定量的数据，然后用训练好的模型进行预测，将预测好的数据扩充到训练数据集，增加了样本的多样性和数量，以增强模型的鲁棒性和泛化能力，使得模型性能有所提升。

6.6 基于知识融合的方法

为了提高模型在框架分类任务中有效区分不同目标词的能力，并充分利用这些目标词中嵌入的信息，Team.4引入了称为“Gloss data augmentation”的数据增强方式，如表4所示，该方法从字典中读取目标词的释义作为额外输入。

Target Word	Example of Gloss
使用	为达到某种目的，让人员、资金等为其服务。
寻求	(动) 寻找追求：~真理—~人生的真谛。
取得	(动) 得到：~成绩—~联系。

Table 4: Gloss data augmentation 示例

图6展示了该方法如何将释义信息与原本的数据信息结合，Team.4使用了一个线性层来结合注解和原始输出，产生了一个新的输出向量，其中包含了与目标单词相关的更精确的分类信息。通过这种方法，可以使该模型有效地利用释义提供的语义线索和语境线索，最终增强其区分目标词所属框架的能力。

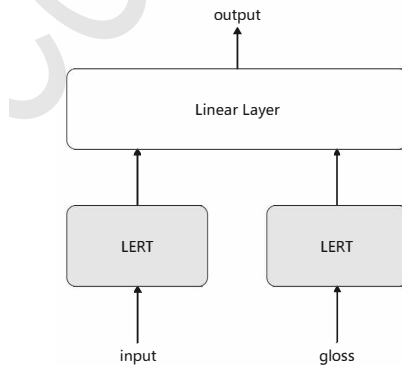


Figure 6: Gloss Enhanced LERT 模型图

6.7 大语言模型分析

对于本次的评测任务，我们同样也设计实验测试了大模型在不同子任务上的能力。我们从测试集中抽取了部分样例，通过构建不同的提示信息，使用ChatGPT (gpt-3.5-turbo-16k) 完成相应的子任务。具体来说，对于框架识别任务，我们测试了ChatGPT在Zero-Shot和Few-

Shot场景下的结果（表5）。对于论元范围识别和论元角色识别任务，我们通过设计思维链的提示方法，与ChatGPT进行多轮对话，引导其生成出更加可靠的结果（表6）。

样例数量	可选提示	正确率
Zero-Shot	无框架定义	40%
	有框架定义	37%
Few-Shot	无框架定义	54%
	有框架定义	53%

Table 5: ChatGPT在FI任务的实验结果

task	Precision	Recall	F1
论元范围识别	60.98	22.52	32.90
论元角色识别	6.38	7.59	6.93

Table 6: ChatGPT在AI任务和RI任务的实验结果

实验结果表明，ChatGPT在汉语框架语义解析的三个子任务上的性能表现均不理想，在思维链的引导下，ChatGPT依然无法很好的从框架、论元及框架元素的角度理解输入的文本，难以使其适应汉语框架语义解析的任务需要。

7 总结

本次评测对于细粒度语义分析具有重要意义，同时也吸引了大量来自学术界或工业界的队伍报名参赛。由于评测任务难度较高、语义粒度细，小模型面对大量的框架时语义理解能力不足，在角色标注时无法应对大量的角色类型；大模型则缺乏框架语义知识，难以正确识别句子中的论元角色，这反映出该任务仍有巨大的发展前景。总体而言，本次评测针对现有模型在细粒度语义分析方面不足的问题，以汉语框架语义解析任务来对模型的场景刻画能力进行评估。未来的评测可以考虑扩大数据覆盖的领域，涵盖更多的语义场景，更全面的评估模型对细粒度语义场景的理解能力，进一步推进汉语框架网的发展。

致谢

感谢国家自然科学基金重点项目（61936012）和科技创新2030-“新一代人工智能”重大项目（2020AAA0106100）的支持。感谢CCL评测组委会的支持。

参考文献

- Collin F Baker, Michael Ellsworth, and Katrin Erk. 2007. Semeval-2007 task 19: Frame semantic structure extraction. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 99–104.
- Xudong Chen, Ce Zheng, and Baobao Chang. 2021. Joint multi-decoder framework with hierarchical pointer network for frame semantic parsing. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2570–2578.
- Yiming Cui, Wanxiang Che, Shijin Wang, and Ting Liu. 2022. Lert: A linguistically-motivated pre-trained language model. *arXiv preprint arXiv:2211.05344*.
- Jason Eisner. 2000. Bilexical grammars and their cubic-time parsing algorithms. *Advances in probabilistic and other parsing technologies*, pages 29–61.
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288.
- 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.

- 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.
- 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.
- 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.
- Jan Hajic, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, M Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, et al. 2009. The conll-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*, pages 1–18.
- Tianyu Jiang and Ellen Riloff. 2021. Exploiting definitions for frame identification. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2429–2434.
- Rohit J. Kate, Yuk Wah Wong, and Raymond J. Mooney. 2005. Learning to transform natural to formal languages. In *AAAI Conference on Artificial Intelligence*.
- Takeru Miyato, Andrew M. Dai, and Ian Goodfellow. 2021. Adversarial training methods for semi-supervised text classification.
- Hao Peng, Sam Thomson, Swabha Swayamdipta, and Noah A Smith. 2018. Learning joint semantic parsers from disjoint data. *arXiv preprint arXiv:1804.05990*.
- Jianlin Su, Yu Lu, Shengfeng Pan, Ahmed Murtadha, Bo Wen, and Yunfeng Liu. 2021a. Roformer: Enhanced transformer with rotary position embedding. *arXiv preprint arXiv:2104.09864*.
- Xuefeng Su, Ru Li, Xiaoli Li, Jeff Z Pan, Hu Zhang, Qinghua Chai, and Xiaoqi Han. 2021b. A knowledge-guided framework for frame identification. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5230–5240.
- Sang-Sang Tan and Jin-Cheon Na. 2019. Positional attention-based frame identification with bert: A deep learning approach to target disambiguation and semantic frame selection. *arXiv preprint arXiv:1910.14549*.
- Xiaohui Wang, Ru Li, Zhiqiang Wang, Qinghua Chai, and Xiaoqi Han. 2020. 基于self-attention的句法感知汉语框架语义角色标注(syntax-aware Chinese frame semantic role labeling based on self-attention). In *Proceedings of the 19th Chinese National Conference on Computational Linguistics*, pages 616–623, Haikou, China, October. Chinese Information Processing Society of China.
- Liping You and Kaiying Liu. 2005. Building chinese framenet database. In *Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE '05. Proceedings of 2005 IEEE International Conference on*.
- Yu Zhang, Qingrong Xia, Shilin Zhou, Yong Jiang, Guohong Fu, and Min Zhang. 2021. Semantic role labeling as dependency parsing: Exploring latent tree structures inside arguments. *arXiv preprint arXiv:2110.06865*.
- 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.
- Ce Zheng, Xudong Chen, Runxin Xu, and Baobao Chang. 2022. A double-graph based framework for frame semantic parsing. *arXiv preprint arXiv:2206.09158*.
- Ce Zheng, Yiming Wang, and Baobao Chang. 2023. Query your model with definitions in framenet: An effective method for frame semantic role labeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 14029–14037.

- Shilin Zhou, Qingrong Xia, Zhenghua Li, Yu Zhang, Yu Hong, and Min Zhang. 2021. Fast and accurate end-to-end span-based semantic role labeling as word-based graph parsing. *arXiv preprint arXiv:2112.02970*.
- 屠寒非, 李茹, 王智强, and 周铁峰. 2016. 一种基于主动学习的框架元素标注. 中文信息学报, 30(4):44–55.
- 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.
- 王智强, 李茹, 梁吉业, 张旭华, 武娟, and 苏娜. 2016. 基于汉语篇章框架语义分析的阅读理解问答研究. 计算机学报, 39(4):13.
- 红叶谭, 真闫, 茹李, and 毅民敬. 2018. 迈向创造性语言生成: 汉语幽默自动生成的探索. 中国科学F辑, 048(011):1497–1509.
- 郝晓燕, 刘伟, 李茹, and 刘开璞. 2007. 汉语框架语义知识库及软件描述体系. Ph.D. thesis.