

# CCL23-Eval 任务2系统报告: WestlakeNLP, 基于生成式大语言模型的中文抽象语义表示解析

高文场<sup>1</sup>, 白雪峰<sup>2</sup>, 张岳<sup>1</sup>

<sup>1</sup>西湖大学, 工学院, 杭州

<sup>2</sup>哈尔滨工业大学, 计算机科学与技术学院, 深圳

{gaowenyang, zhangyue}@westlake.edu.cn

xfbai.hk@gmail.com

## 摘要

本文介绍了我们在第二十二届中国计算语言学大会中文抽象语义表示解析评测任务中提交的参赛系统。中文抽象语义表示(Chinese Abstract Meaning Representation, CAMR)不仅以图的方式表示句子的语义, 还保证了概念对齐和关系对齐。近期, 生成式大规模语言模型在诸多自然语言处理任务上展现了优秀的生成能力和泛化能力。受此启发, 我们选择微调Baichuan-7B模型来以端到端的形式从文本直接生成序列化的CAMR。实验结果表明, 我们的系统能够在不依赖于词性、依存句法信息以及复杂规则的前提下取得了同现有方法可比的性能。

**关键词:** 中文抽象语义表示; 语义解析; 大模型; 微调

## System Report for CCL23-Eval Task 2: WestlakeNLP, Investigating Generative Large Language Models for Chinese AMR Parsing

Wenyang Gao<sup>1</sup>, Xuefeng Bai<sup>2</sup>, Yue Zhang<sup>1</sup>

<sup>1</sup>School of Engineering, Westlake University, Hangzhou

<sup>2</sup>School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen

{gaowenyang, zhangyue}@westlake.edu.cn

xfbai.hk@gmail.com

## Abstract

This paper presents our participating system in the Chinese Abstract Meaning Representation Parsing Evaluation Task at the 22nd China National Conference on Computational Linguistics. Chinese Abstract Meaning Representation (CAMR) not only captures sentence semantics through graphical representation but also ensures the alignment of concepts and relations. Recently, generative large language models have demonstrated exceptional abilities in generation and generalization across various natural language processing tasks. Motivated by these advancements, we fine-tune the Baichuan-7B model to directly generate serialized CAMR from the provided text in an end-to-end manner. Experimental results demonstrate that our system achieves comparable performance to existing methods, eliminating the need for part-of-speech, dependency syntax, and complex rules.

**Keywords:** Chinese Abstract Meaning Representation, Semantic Parsing, Large Language Model, Fine-tuning

## 1 引言

抽象语义表示(Abstract Meaning Representation, AMR)以有根的有向无环图来表示句子的语义 (Banarescu et al., 2013)。AMR图中的节点是句中的词或由词抽象而来的概念; 节点之间的关系通过图中的边来表示, 它们反映了句中概念的语义关系。AMR语义解析旨在从文本中自动获取抽象语义表示 (Flanigan et al., 2014; Konstas et al., 2017; Lyu and Titov, 2018; Cai and Lam, 2020; Bevilacqua et al., 2021; Bai et al., 2022a; Bai et al., 2022b)。得益于AMR语义解析技术的进步, AMR被广泛应用于机器翻译(Nguyen et al., 2021)、问答(Deng et al., 2022)、自然语言推断(Opitz et al., 2023)、关系抽取(Xu et al., 2022)、文本摘要(Liao et al., 2018)、对话系统(Bai et al., 2021)等下游领域。

AMR的数据和标注规范起源于英文, 近年来, Li et al. (2016)将AMR推广到中文, 称为Chinese Abstract Meaning Representation (CAMR)。CAMR在保留了AMR较强语义表示能力的同时, 还增加了“词与概念”的对齐信息和“词与关系”的对齐信息。目前, CAMR解析方法大致可以分为三类: 基于转移的方法(Transition-based)、基于图的方法(Graph-based)和基于序列到序列的方法(Seq2Seq-based)。基于转移的方法先将输入的句子解析为依存关系树, 然后通过一系列“行动”转换为AMR图。Wang et al. (2018)首先在CAMR解析任务中使用了基于转移的框架, Wu et al. (2019)又在此基础上进行了一系列的改进。基于图的方法通过在序列上进行概念预测和关系分类来逐步构建AMR图。在上一届CAMR解析评测任务中拔得头筹的SUDA-HUAWEI (Zhou et al., 2022)和PKU (Chen et al., 2022)都是此类方法。基于序列到序列的方法利用如Transformers、BART等序列到序列模型直接从句子得到序列化的AMR。Huang et al. (2021)首先在CAMR解析任务中使用了序列到序列的方法, 摆脱了人为设计“行动”的过程, 也极大地简化了训练过程。

近期, 随着预训练语言模型的参数量和数据量不断增大, 大规模语言模型(大模型)在自然语言处理领域的多个任务上取得了显著的进展 (Zeng et al., 2022; Zhao et al., 2023; Wang et al., 2023a; Wang et al., 2023b; Ren et al., 2023; Touvron et al., 2023)。相较于传统预训练语言模型, 更多的参数使得大模型能够记忆更广泛的世界知识, 并展现了小模型所不具备的涌现能力。通过使用指令微调和代码预训练等技术, 大模型具备了响应人类指令、泛化到新任务、代码理解与生成、利用思维链推理以及处理长距离依赖等多项能力。

受此启发, 我们采用百川智能发布的中文大模型Baichuan-7B进行CAMR解析。为了赋予大模型处理CAMR图的能力, 我们采用基于深度优先遍历的序列化算法来将CAMR图转化为一个线性序列。在此基础上, 我们通过对大模型进行微调来以序列到序列的方式从文本生成对应的CAMR图。实验结果表明, 我们所提出的方法在不依赖于词性、依存句法信息以及复杂规则的前提下, 仅利用原句和分词信息就取得了同现有工作可比的结果。

## 2 方法

我们首先对CAMR进行序列化, 在此基础上通过微调Baichuan-7B模型来将文本转换为序列化的CAMR图。模型的输入数据包括提示词、原句以及经过分词并带有词编号的句子。输出是序列化的CAMR。为了确保生成的CAMR序列是合法的, 我们还采用了一些基于规则和匹配的方法进行后处理。

### 2.1 CAMR序列化

由于大模型通常在文本上进行训练, 因此不能直接用于生成CAMR图。为了解决这个问题, 我们对CAMR图进行序列化, 通过将CAMR表示为一个线性序列, 我们能够通过大模型以自回归的方式生成线性化的CAMR图。具体而言, 我们采用了基于深度优先遍历的序列化方式, 原因是深度优先遍历获得的线性化CAMR图与原文的顺序最接近。图 1展示了一个CAMR图及其序列化的结果, 在序列化的CAMR图中, 我们通过左括号来记录当前的深度, 利用右括号来匹配左括号进行图结构还原。通过这种方式, 我们能够通过线性化的CAMR来还原原始CAMR图。

```

# ::snt 新华社 北京 十二月 二日 电
# ::wid x1_新华社 x2_北京 x3_十二月 x4_二日 x5_电 x6_

(x5 / 电-02
  :arg0() (x9 / organization
    :name() (x1 / name :op1 x1/新华社)
    :location() (x11 / city
      :name() (x2 / name :op1 x2/北京 )))
  :time() (x13 / date-entity
    :month() (x3 / 12)
    :day() (x4 / 2)))
(x5 / 电-02 :arg0() (x9 / organization :name() (x1 /
  name :op1 x1/新华社) :location() (x11 /
  city :name() (x2 / name :op1 x2/北京 ))) :time()
(x13 / date-entity :month() (x3 / 12) :day() (x4 / 2)))

```

Figure 1: CAMR序列化实例，左侧为CAMR图，右侧为序列化结果(实际为单行)。

## 2.2 Baichuan-7B

Baichuan-7B<sup>1</sup>是一个基于Transformer结构的中英双语大模型。该模型是在大约1.2万亿个标记(tokens)上进行训练的，拥有约70亿个参数，并且上下文窗口长度为4096。在标准的中文和英文权威基准测试(C-EVAL/MMLU)上，Baichuan-7B取得了与同样规模模型相比最佳的效果。

Baichuan-7B 的优异性能主要来自于三部分，高质量数据、分词和模型架构。在数据方面，其训练语料兼顾质量和多样性，相较于其他同参数规模的开源中文预训练模型，数据量提高了超过50%；在分词方面，它对SentencePiece中的字节对编码(Byte Pair Encoding, BPE)进行了优化，提升了对于中文的压缩率，还能做到未知字词的全覆盖；在模型架构方面，整体模型基于标准的Transformer结构，采用与LLaMA相同的模型设计，使用旋转式位置编码(Rotary Position Embedding)。同时，尺寸为4096的上下文窗口使得模型的应用场景更加广泛。

## 2.3 大模型微调

在微调的过程中，我们构造的输入数据包括任务指令、原句以及带有词编号的分词后的句子(上下文)，标签则是序列化的CAMR。以“这几天关于中俄战略合作伙伴关系成了大热点。”为例，以下为构造的输入数据：

给定如下输入句子以及分词结果，输出句子所对应的中文抽象语义表示图：这几天关于中俄战略合作伙伴关系成了大热点。 x1\_这 x2\_几 x3\_天 x4\_关于 x5\_中 x6\_俄 x7\_战略 x8\_合作 x9\_伙伴 x10\_关系 x11\_成 x12\_了 x13\_大 x14\_热点 x15\_。

除此之外，在微调的过程中，我们取消了对于任务指令、原句和上下文的损失计算，使得模型专注于生成的内容本身。

形式化而言，假设 $\mathcal{D}$ 代表整个训练数据集， $I$ 表示任务指令， $X$ 表示原句， $C$ 表示上下文， $Y$ 表示序列化的CAMR，本方法通过优化如下损失函数进行训练：

$$\ell = - \sum_{\{I, X, C, Y\} \in \mathcal{D}} \log P(Y | I \oplus X \oplus C) \quad (1)$$

其中 $I \oplus X \oplus C$ 代表三个序列的拼接。为了方便表示，我们将微调后的模型称为LLMCP。

## 2.4 后处理

针对微调后的模型生成的CAMR序列，我们进行了后处理，使其符合CAMR的规范。该部分主要分为括号补全和节点信息补全。

- **括号补全：**由于Baichuan-7B模型没有经过代码训练，所以对于生成的CAMR需要进行补全和调整。为了解决这个问题，我们采用了相对简单的规则进行处理。根据左右括号的相对数量，我们在生成的序列末尾添加或删除右括号。考虑到CAMR的特点，起始的左括号应当在序列末尾处被右括号闭合，因此，我们将提前闭合的右括号移动至序列末尾。

<sup>1</sup><https://github.com/Baichuan-inc/Baichuan-7B>

- **节点信息补全**: 模型在推理过程中, 会出现生成的节点信息不全的问题, 即丢失节点的编号或词语。例如, 模型对于图1中的例子, 可能会生成如“(电-02 :arg0() ...)”或“(x5 :arg0() ...)”的结果。我们通过规则匹配, 为节点补全丢失的编号或词语, 使其与输入中的编号和分词保持一致。

### 3 实验

#### 3.1 实验设置

本次评测分为closed和open两个赛道, 我们选择了参加open赛道。尽管训练集、开发集和测试集TestA提供了词性和依存句法信息, 但我们在模型微调过程中仅利用了原句和分词信息。本次测评新增的测试集TestC是2000句问句语料, 旨在增强模型对汉语问句的焦点和一句多问的分析能力。

#### 3.2 实验结果

	P	R	F1
Dev	79.24	77.20	78.20
TestA	74.40	70.24	72.26
TestB	70.42	68.63	69.52
TestC	82.73	82.49	82.61

Table 1: LLMCP在开发集和测试集的Align-Smatch得分

本次评测所采用的Align-smatch指标在Smatch的基础上增添了概念对齐信息和关系对齐信息。表 1列出了我们的方法的得分。我们提出的LLMCP在开发集上获得了78.2的F1分数, 在三个测试集上分别获得了72.26、69.52以及82.61的分数。TestC的结果相比于其他两个测试集更高, 原因是TestC的数据由问句组成, 其长度相对较短、句式相对简单, 因此解析难度相对较低。相比于Huang et al. (2021)的序列到序列方法(该方法在TestA上获得了70.29的F1分数), LLMCP取得了显著的性能提升, 这证明了大模型在CAMR解析任务上的有效性。相比于往年的系统, LLMCP在不使用外部句法知识和复杂规则的前提下获得了可比的性能, 这表明CAMR解析能够以一种端到端的方式进行有效建模。

#### 3.3 结果分析

在使用Baichuan-7B之前, 我们也尝试在相同的数据上对较为轻量的mT5模型进行微调。然而, mT5在开发集上的表现(52.3)远低于Baichuan-7B模型的表现(78.2)。这一结果表明基座模型对于解析效果有着显著影响, 而大模型在预训练阶段获得的世界知识和涌现能力对于CAMR解析任务具有重要帮助。

我们微调的模型也存在大模型普遍面临的一个问题, 即幻觉(hallucination)现象。当模型生成的文本不遵循原文或者不符合事实(Ji et al., 2022), 我们就认为模型出现了幻觉。尽管我们在输入中提供了每个词的编号, 但有时生成的CAMR序列的编号会发生整体错位的情况。这一现象主要集中在较长的句子中, 可能是因为模型在推理过程中解码器错误地参考了编码器编码的信息所导致的。

此外, 我们还观察到模型在处理较长输入时存在CAMR不完整的问题。例如, 有时模型只解析了句子前四分之三的内容, 而对于后四分之一的概念和关系则完全未生成。这可能是由于模型的生成窗口有限导致的。

### 4 结语

在本次CAMR解析评测任务中, 我们探索了通过微调大模型从文本生成序列化的CAMR。实验结果表明, 我们的方法在不依赖词性、依存句法信息以及复杂规则的前提下取得了同现有方法可比的性能。未来, 针对长句的CAMR解析问题, 我们将尝试使用分句推理后融合的方法, 通过将长句进行拆分、推理和融合, 有效地缓解生成窗口受限问题, 从而确保生成的CAMR结构完整且准确。

## 致谢

感谢匿名审稿人对本文提出的宝贵意见，我们在此基础上进行了完善。本研究受国家重点研发计划项目（2022YFE0204900）资助，张岳是本文的通讯作者。

## 参考文献

- Xuefeng Bai, Yulong Chen, Linfeng Song, and Yue Zhang. 2021. Semantic representation for dialogue modeling. 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 4430–4445, Online, August. Association for Computational Linguistics.
- Xuefeng Bai, Yulong Chen, and Yue Zhang. 2022a. Graph pre-training for AMR parsing and generation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6001–6015, Dublin, Ireland, May. Association for Computational Linguistics.
- Xuefeng Bai, Sen Yang, Leyang Cui, Linfeng Song, and Yue Zhang. 2022b. Cross-domain generalization for AMR parsing. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10907–10921, Abu Dhabi, United Arab Emirates, December. Association for Computational Linguistics.
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, pages 178–186.
- Michele Bevilacqua, Rexhina Blloshmi, and Roberto Navigli. 2021. One spring to rule them both: Symmetric AMR semantic parsing and generation without a complex pipeline. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14):12564–12573.
- Deng Cai and Wai Lam. 2020. AMR parsing via graph-sequence iterative inference. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1290–1301, Online. Association for Computational Linguistics.
- Liang Chen, Bofei Gao, and Baobao Chang. 2022. A two-stage method for Chinese amr parsing. *ArXiv*, abs/2209.14512.
- Zhenyun Deng, Yonghua Zhu, Yang Chen, Michael Witbrock, and Patricia Riddle. 2022. Interpretable AMR-based question decomposition for multi-hop question answering. *arXiv preprint arXiv:2206.08486*.
- Jeffrey Flanigan, Sam Thomson, Jaime Carbonell, Chris Dyer, and Noah A. Smith. 2014. A discriminative graph-based parser for the Abstract Meaning Representation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1426–1436, Baltimore, Maryland. Association for Computational Linguistics.
- Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Wenliang Dai, Andrea Madotto, and Pascale Fung. 2022. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55:1 – 38.
- Ioannis Konstas, Srinivasan Iyer, Mark Yatskar, Yejin Choi, and Luke Zettlemoyer. 2017. Neural AMR: Sequence-to-sequence models for parsing and generation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 146–157, Vancouver, Canada. Association for Computational Linguistics.
- Bin Li, Yuan Wen, Weiguang Qu, Lijun Bu, and Nianwen Xue. 2016. Annotating the little prince with Chinese AMRs. In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*, pages 7–15, Berlin, Germany. Association for Computational Linguistics.
- Kexin Liao, Logan Lebanoff, and Fei Liu. 2018. Abstract Meaning Representation for multi-document summarization. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1178–1190, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Chunchuan Lyu and Ivan Titov. 2018. AMR parsing as graph prediction with latent alignment. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 397–407, Melbourne, Australia. Association for Computational Linguistics.

- Long H. B. Nguyen, Viet H. Pham, and Dien Dinh. 2021. Improving neural machine translation with AMR semantic graphs. *Mathematical Problems in Engineering*.
- Juri Opitz, Shira Wein, Julius Steen, Anette Frank, and Nathan Schneider. 2023. AMR4NLI: Interpretable and robust NLI measures from semantic graphs. *ArXiv preprint*, abs/2306.00936.
- Xiaozhe Ren, Pingyi Zhou, Xinfan Meng, Xinjing Huang, Yadao Wang, Weichao Wang, Pengfei Li, Xiaoda Zhang, A. V. Podolskiy, Grigory Arshinov, A. Bout, Irina Piontkovskaya, Jiansheng Wei, Xin Jiang, Teng Su, Qun Liu, and Jun Yao. 2023. Pangu- $\sigma$ : Towards trillion parameter language model with sparse heterogeneous computing. *ArXiv preprint*, abs/2303.10845.
- Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models. *ArXiv preprint*, abs/2302.13971.
- Chuan Wang, Bin Li, and Nianwen Xue. 2018. Transition-based Chinese AMR parsing. In *North American Chapter of the Association for Computational Linguistics*.
- Xiao Wang, Wei Zhou, Can Zu, Han Xia, Tianze Chen, Yuan Zhang, Rui Zheng, Junjie Ye, Qi Zhang, Tao Gui, Jihua Kang, J. Yang, Siyuan Li, and Chunsai Du. 2023a. Instructuie: Multi-task instruction tuning for unified information extraction. *ArXiv preprint*, abs/2304.08085.
- Yidong Wang, Zhuohao Yu, Zhengran Zeng, Linyi Yang, Cunxiang Wang, Hao Chen, Chaoya Jiang, Rui Xie, Jindong Wang, Xingxu Xie, Wei Ye, Shi-Bo Zhang, and Yue Zhang. 2023b. Pandalm: An automatic evaluation benchmark for llm instruction tuning optimization. *ArXiv preprint*, abs/2306.05087.
- Runxin Xu, Peiyi Wang, Tianyu Liu, Shuang Zeng, Baobao Chang, and Zhifang Sui. 2022. A two-stream AMR-enhanced model for document-level event argument extraction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5025–5036, Seattle, United States. Association for Computational Linguistics.
- Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, et al. 2022. Glm-130b: An open bilingual pre-trained model. *ArXiv preprint*, abs/2210.02414.
- Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Z. Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jianyun Nie, and Ji rong Wen. 2023. A survey of large language models. *ArXiv*, abs/2303.18223.
- Shilin Zhou, Qingrong Xia, Yang Li, Zhefeng Wang, and Zhenghua Li. 2022. Suda-huawei camr2022 比赛技术评测报告.
- Huang Ziyi, Li Junhui, and Gong Zhengxian. 2021. 基于序列到序列的中文AMR 解析(Chinese AMR parsing based on sequence-to-sequence modeling). In *China National Conference on Chinese Computational Linguistics*.
- 吴泰中, 顾敏, 周俊生, 曲维光, 李斌, and 顾彦慧. 2019. 基于转移神经网络的中文AMR 解析. *中文信息学报*, 33(4):1–11.