CCL25-Eval任务12总结报告：面向中文语音的实体关系三元组抽取

穆文轩[1], 宁金忠[1,*], 潘怡霖[1], 帕尔哈提·吐拉江[2],孙媛媛[2],
李松涛[1], 尹伟鸣[1], 季延旭[1], 张益嘉[1], 林鸿飞[2]
[1]大连海事大学 [2]大连理工大学
ningjinzhong@dlmu.edu.cn

摘要

中文语音实体关系三元组抽取任务（Chinese Speech Entity-Relation Triple Extraction Task，CSRTE）是第二十四届中国计算语言学大会中的一项技术评测，旨在从中文语音数据中自动识别并提取实体及其相互关系，构建结构化的语音关系三元组（头实体、关系、尾实体）。本任务的目标是提升中文语音关系三元组抽取的准确性与效率，增强模型在不同语境和复杂语音场景下的鲁棒性，实现从语音输入到文本三元组输出的全流程自动化处理。通过本次评测，有助于推动中文语音信息抽取技术的发展，促进语音与自然语言处理技术的深度融合，为智能应用提供更加丰富且精准的基础数据支持。此次评测共有257支队伍报名参赛，其中59支队伍提交了A榜成绩。成绩排名前15的队伍晋级B榜，并且表现突出的前7支队伍提交了技术报告。

关键词： 中文语音 ； 实体关系三元组抽取 ； 技术评测

Overview of CCL25-Eval Task 12: Chinese Speech Entity-Relation Triple Extraction

Wenxuan Mu[1], Jinzhong Ning[1,*], Yilin Pan[1], Paltati Tulajiang[2], Yuanyuan Sun[2],
Songtao Li[1], Weiming Yin[1], Yanxu Ji[1], Yijia Zhang[1], Hongfei Lin[2]
[1]Dalian Maritime University [2]Dalian University of Technology
ningjinzhong@dlmu.edu.cn

Abstract

The Chinese Speech Entity-Relation Triple Extraction Task (CSRTE) is a technical evaluation held as part of the 24th China Conference on Computational Linguistics. It aims to automatically identify and extract entities and their relationships from Chinese speech data, constructing structured speech relation triples (head entity, relation, tail entity). The goal of this task is to improve the accuracy and efficiency of CSRTE, enhance the model's robustness under various contexts and complex speech scenarios, and realize a fully automated process from speech input to structured text-based triple output. This evaluation contributes to the advancement of Chinese speech information extraction technologies, promotes the deep integration of speech and natural language processing, and provides richer and more accurate foundational data for intelligent applications. A total of 257 teams registered for the evaluation, with 59 teams submitting results on the A leaderboard. The top 15 teams advanced to the B leaderboard, and the top 7 outstanding teams submitted technical reports.

Keywords: Chinese Speech , Triple Extraction , Technical Evaluation

第二十四届中国计算语言学大会论文集，第475页-第480页，济南，中国，2025年8月11日至14日。

1 引言

近年来，随着人工智能技术的快速发展，语音作为人机交互的核心载体，在智能客服、语音搜索、智能家居等场景中展现出巨大应用价值。传统信息抽取技术主要聚焦于文本模态，通过识别实体及其语义关系，从文本中提取所需信息，例如用于构建知识图谱、支撑结构化知识库。然而，语音数据因其非结构化、时序性、口语化等特征，加之中文独有的复杂性（如同音词歧义、声调变化、方言差异），使得直接从语音中提取结构化关系三元组面临诸多挑战。因此，实现语音到结构化知识的端到端高效转换，不仅在自然语言处理与语音技术交叉领域具有重要研究价值，也亟需相关技术和方法的深入探索与创新。

中文语音实体关系三元组抽取（Chinese Speech Entity-Relation Triple Extraction, CSRTE）任务旨在突破传统基于文本的信息抽取边界，探索语音信号与结构化语义知识之间的直接映射关系。该任务不仅需要解决语音识别（ASR）中的噪声鲁棒性、口语化表达归一化、多轮对话上下文建模等关键问题，还需在此基础上实现实体边界识别、关系分类与三元组对齐的联合优化。

在本次第二十四届中国计算语言学大会（CCL 2025）中，首届CSRTE任务受到了广泛关注，共有257支队伍报名参赛，最终A榜有59支队伍提交成绩，前15支优秀队伍晋级B榜，并由其中表现突出的7支队伍撰写技术报告。这一评测任务专注于从中文语音数据中端到端地抽取实体及其关系三元组，涵盖语音识别、命名实体识别、关系抽取等多个关键环节，促进语音与自然语言处理技术的深度融合趋势。

目前，关于中文语音信息抽取的研究尚处于起步阶段。尽管已有大量面向文本的实体关系抽取技术取得显著进展，如实体识别（Named Entity Recognition, NER）和关系分类（Relation Classification, RC）任务已在多个标准数据集上开展深入探索，但这些研究多以文本为输入，未能充分考虑语音数据在信息抽取中的潜在价值。相比之下，语音模态因其连续性、非结构性、含噪性及口语化表达特征，在实体识别及关系建模上面临更大挑战，特别是在中文语境下，诸如同音异义、声调变异、口音差异等问题尤为突出。因此，CSRTE任务的提出，旨在突破传统信息抽取任务在模态上的限制，探索从语音信号直接构建结构化知识的可能路径。

作为首次面向中文语音的三元组抽取评测，CSRTE 任务具有以下几个显著特点。第一，在数据集方面，本次评测任务基于高质量中文语音数据构建，并同步提供标注的实体关系三元组标签，既保证了语料的真实性，又增强了模型训练的监督信号。第二，在任务设置方面，评测主张采用端到端的建模范式，要求系统能够从原始语音中直接输出结构化的“头实体–关系–尾实体”三元组。第三，在评测机制方面，任务分为A榜与B榜，采用多阶段评审流程，使得比赛更加客观公正，促进研究者对该领域的兴趣与投入，以及CSRTE任务的发展。

2 相关工作

近年来，随着对个人信息保护和提高语言理解系统需求的增加，基于语音的命名实体识别（NER）领域受到了越来越多的关注。Cohn等人介绍了音频去识别这一新任务，即识别并遮蔽音频中包含个人信息的部分。此任务在保护医疗记录中的隐私方面尤其重要，与文本去识别任务中的实体识别有相似之处。该研究建立了用于评估的基准数据集，强调了音频去识别在隐私保护中的潜在应用(Cohn et al., 2019)。Yadav等人探讨了从英文语音中识别命名实体的挑战。传统上，NER采用两步流程：自动语音识别（ASR）和随后进行的NER标注。然而，这种串行方法可能导致步骤间的错误累积。最新研究表明，集成的端到端方法在性能上优于基于音素的ASR系统。论文还探讨了使用法语数据集进行端到端NER的方法，包括使用特殊符号进行NER标注的新尝试(Yadav et al., 2020)。

并且，Shon等人讨论了用于语音语言理解（SLU）任务的新基准套件，重点关注NER、情感分析和ASR。与ASR不同，SLU尚未得到足够的资源和关注。为了解决这一问题，研究提供了新的微调和评估集注释，提供了当前系统的基线评估。未来方向包括增加更多的SLU任务以及改进测试集注释和人类性能衡量方法(Shon et al., 2022)。此外，Chen等人强调了中文语音中NER的挑战，特别是同音字和多音字对识别结果的影响。由于缺乏公开数据集，中文

*Corresponding author

语音的NER研究较为有限。为此，提出了基于AISHELL-1语料库的新数据集AISHELL-NER，涵盖了多个领域，并标注了个人（PER）、地点（LOC）和组织（ORG）三种类型的命名实体(Chen et al., 2022)。Zhou等人进一步探索了中文口语中的NER，使用了富含自然对话现象的数据集，如间歇词和口吃。通过标注实体较多的对话并采用Roberta-CRF模型，该研究在涵盖多种主题的多个对话中进行了基准测试。研究还与多个大型语言模型（LLM）进行对比，使用一致的ASR生成文本，并通过精确度、召回率、F1分数和命名实体准确度（NEA）等指标评估NER表现(Zhou et al., 2024)。

最后，Ning等人也在探索中文口语中的NER，重点关注体现自然对话特征的对话。采用与前述研究类似的Roberta-CRF和ASR模型，该工作评估了多个对话和话题的性能，为自发性语音中NER技术的有效性提供了宝贵的见解。通过多种评估指标的使用，突出强调了在推进这一领域中的综合性能评估的重要性(Ning et al., 2024)。

3 评测细则

3.1 评测赛道设置

本次评测设立双技术路径赛道，面向语音信息抽取任务的两类主流技术方案：端到端方案（End-to-End, E2E）与流水线分阶段方案（Pipeline）。参赛团队需根据所采用的技术路径选择相应赛道，并严格遵循路径定义及相关参数限制。

端到端赛道要求参赛方案基于单一模型，从语音信号中直接完成语义理解、实体识别及关系抽取等多任务联合建模，过程中不依赖任何独立的自动语音识别（ASR）模块。为鼓励技术创新与模型紧凑性，端到端方案所使用模型的参数总量不得超过11B。典型符合要求的模型架构包括Qwen-Audio、GLM-4-Voice等公开发布的端到端多模态模型。

流水线赛道采用“语音转文本-关系抽取”两阶段处理流程，第一阶段需通过自动语音识别模块将语音转写为文本，第二阶段再对转写文本执行关系抽取。为保障技术开放性与资源公平性，该赛道规定语音识别模块必须为开源且具备非商业授权，且其参数量不超过800M。

奖项设置与双赛道结构紧密结合。其中，一等奖专属授予端到端赛道表现最优团队，以鼓励具备先进技术路线的方案。二等奖与三等奖则按1:1比例分别分配至两赛道。若某一赛道有效参赛队伍数量不足，相应奖项名额将自动划拨至另一赛道。评委会将对路径归属争议作最终裁定，确保评测公平性。

为确保合规性与可复现性，所有参赛团队必须在限定数据范围内完成模型训练与优化，不得使用任何外部语音或文本数据。伪数据生成仅限基于主办方提供的数据集或所选基座模型进行。端到端方案若采用第三方预训练模型进行微调，须在赛前完成架构核验；流水线方案则需提交语音识别模块的开源协议证明。进入复赛的前六支队伍必须提交完整的私有部署版本，包括源代码、模型权重及技术说明文档（不超过六页），否则将取消获奖资格。

评测严格依据功能性标准对技术路径进行划分。端到端方案的核心特征为模型能够在无转写中间结果的情况下，直接从语音输入中生成结构化三元组；而流水线方案则要求显式区分语音识别与文本信息抽取两个阶段。对于路径判定存疑的参赛队伍，主办方提供提前申请界定通道，以保障参赛流程顺利进行。

3.2 评测数据

本次评测所采用的数据集基于开源语音资源构建，主要来源包括Common Voice 17的中文子集与AISHELL语音识别数据集(Chen et al., 2022)。在此基础上，由专业标注人员对语音内容中的实体及其语义关系进行细致标注，形成面向中文语音三元组抽取任务的高质量评测数据集。

该数据集总计包含约40小时的中文语音数据，覆盖约20,000条语音样本。每条样本均提供原始语音文件及其对应的结构化标注信息，包括实体文本及其之间的关系类型，但不包含直接转录的文本内容，以契合端到端处理需求。实体类别涵盖人物、组织、地点、产品、著作等10大类，关系类型预定义超过50种，涵盖丰富的语义关系场景，适用于评估模型的多类关系识别能力。数据内容来源广泛，涵盖新闻播报、日常对话、正式演讲等多种真实语境，以增强模型在复杂多变语音环境下的鲁棒性与泛化能力。通过数据的多样性设计，评测任务能够更全面地反映系统在实际应用中的表现。

第二十四届中国计算语言学大会论文集，第475页-第480页，济南，中国，2025年8月11日至14日。

实体1	关系类型	实体2
沙玛什	持有称号/职业	正义之神
沙玛什	隶属于	苏美尔乌图神
沙玛什	隶属于NORP组织	苏美尔乌图神
苏美尔乌图神	设立职位/称号/奖项	正义之神

Table 1: 语音样本`sample` 的三元组注释示例

Table 3.2中是一个案例中的实体和三元组。原始文件是一个后缀名为".wav"的音频文件，需要从音频文件中识别出表格中的多种实体及其关系。

本数据集由大连海事大学智能技术实验室整理并拥有所有权，已通过人工筛查与百度智能云模型的联合审核，剔除潜在的违规及敏感内容。评测所提供的数据资源仅供学术研究用途，禁止任何形式的商业使用。数据集中所包含的个别有害示例仅作为系统安全能力验证用途，不代表主办方任何立场。

3.3 评价指标

在本次任务中，采用了F1分数（F1 scores）作为性能评估的主要指标。这些评价指标的计算公式如下所述:

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{3}$$

其中，涉及四个变量TP，FP，FN，FP，主要含义如表3.3:

	预测为正类别（预测值）	预测为负类别（预测值）
实际正类别	True Positive (TP)	False Negative (FN)
实际负类别	False Positive (FP)	True Negative (TN)

Table 2: 评价指标变量

其中，三元组的匹配模式按照严格匹配，并且顺序需要严格按照test集合中的顺序，否则F1值为0。

4 参赛情况

本次比赛吸引了257支队伍报名参赛，包括清华大学、大连理工大学、电子科技大学、北京理工大学、北京航空航天大学、西安电子科技大学、北京交通大学、东北大学、中国科学院、华东师范大学、香港科技大学等高校参加。此外，还吸引了中电海康集团有限公司、南宁铁路局科研所、连云港日报社、赣西肿瘤医院等单位参赛。

本次比赛分A榜和B榜两个阶段。A榜一共有59支队伍提交实验结果，其中A榜最高成绩F1值达到了64.41，第15名成绩为46.27。本次评测取A榜前15名队伍晋级B榜，进行决赛，然后取B榜前7名作为最终的获奖队伍（优先端到端赛道）。前7名获奖队伍如表4所示:

其中，队伍排名分别为第二、第三和第四的三支队伍提交了评测论文，其余队伍提交了技术报告。

5 队伍方法

来自个人参赛者（广东珠海）的队伍在初赛和复赛中分别采用了Pipeline与End-to-End两种方案进行中文语音实体关系三元组抽取。在初赛阶段，该队伍基于FireRedASR模型完成语音

最终排名	参赛队伍	单位	得分	赛道
1	一二三四1234	个人	0.5302	E2E
2	out of memory	中国石油大学	0.5937	Pipeline
3	伊托洛斯基	赣西肿瘤医院	0.5292	E2E
4	橘子猫	香港科技大学	0.5308	Pipeline
5	我不讲组会	西安电子科技大学	0.5181	E2E
6	金风细雨楼	连云港日报社	0.5152	E2E
7	结束乐队	北京航空航天大学	0.4846	Pipeline

Table 3: 参赛队伍成绩表

转文本任务，并使用LLaMA Factory框架对InternLM3-8B-Instruct模型进行LoRA微调以提取三元组。在复赛阶段，转向端到端方案，最终选用Qwen2.5-Omni-7B多模态大语言模型作为底座，结合指令微调设计，将语音输入与格式化prompt一同送入模型进行结构化三元组抽取。该方法通过统一格式转化、嵌套结构输出和LoRA高效微调策略，在B榜取得了0.5302的F1得分。实验表明，Qwen2.5-Omni在处理复杂语音场景下的语义抽取任务中具备强大能力，且端到端设计有效避免了传统ASR-NLP串联模式下的误差累积问题。

来自中国石油大学（华东）的参赛队伍提出了一种基于语音识别与大语言模型协同的中文语音三元组抽取方法。该方法首先利用语音识别模型将语音转换为文本，再通过热词检测、拼音相似度匹配和阿拉伯数字转中文等手段对识别文本进行纠错与优化，随后采用微调后的大语言模型执行实体关系三元组抽取任务。在模型训练方面，该队伍构建了高质量的指令微调数据集，统一采用嵌套结构格式标注三元组信息，并基于参数高效微调策略优化大模型表现。实验结果显示：所提方法在提升专有名词识别准确率、减少同音字错误率方面成效显著，在测试集上取得了优于基线的整体性能，同时消融实验验证了热词库构建、拼音匹配算法与两阶段语音识别策略对任务效果的关键贡献。

来自赣西肿瘤医院的参赛队伍采用端到端的中文语音三元组抽取方法，基于LLamafactory框架，选用qwen2-audio和qwen2.5-omin两个7B量级大语言模型进行LoRA参数高效微调，并通过模型对比与多轮参数组合试验最终确定qwen2.5-omin为最优模型。该队伍设定多组LoRA参数（如rank=64，alpha=128），在step=3000时取得最佳效果。此外，实验过程中还评估了复杂与简化Prompt、RAG信息引入等策略对模型性能的影响，结果表明其对最终性能提升有限。最终，该队伍在B榜End-to-End赛道中以0.5292分获得第二名。

来自香港科技大学的参赛队伍提出了一种轻量高效的pipeline方法，用于中文语音中的实体与关系抽取任务。该系统由工业级语音识别模块FireRedASR和基于span的联合实体关系抽取模型组成，前者采用Conformer-Transformer结构，后者借鉴SpERT架构并适配中文语境（使用Chinese-BERT-wwm编码器）。相较于近年来流行的端到端LLM方案，该方法具备结构清晰、可解释性强、部署便捷等优势。在CCL2025-Eval Task 12评测中，该pipeline在仅1.2B参数量下，以A榜F1值0.56显著超越了如Qwen2-Audio（7B，F1=0.46）等大模型，验证了传统模块化方法在中文语音信息抽取中的强大竞争力。

来自西安电子科技大学的参赛队伍针对中文语音三元组抽取任务，设计了一种基于Qwen2.5-Omni-7B的大模型端到端微调方案。该方案利用Qwen2.5-Omni原生支持语音输入的多模态能力，结合Schema约束与结构化Prompt模板，引导模型直接从语音中生成符合规范的结构化三元组。为提升鲁棒性，团队构建了高质量微调数据，并引入语音数据增强策略，包括TTS合成语音辅助训练与无效音频样本过滤。训练过程基于ModelScope Swift框架，采用LoRA微调，最终在CSRTE评测中分别在A榜与B榜获得F1得分0.5171和0.5181，排名第六与第五。该研究充分展现了多模态大模型在复杂语音场景下的强大结构化信息抽取能力。

来自连云港日报社的参赛队伍采用端到端（E2E）建模方式，结合多模态大模型进行语音实体关系三元组抽取。方法上，团队以大模型生成伪三元组作为数据增强手段，构建语音-文本-结构三元组的训练数据，在此基础上对预训练多模态模型进行微调以适应任务需求。训练过程包括两个阶段：第一阶段使用大模型自动生成伪标注数据进行预训练，第二阶段在官方提供的Common Voice 17中文子集与AISHELL语音数据上进行微调。在此基础上，团队还尝试引入

伪增强与多样化语境数据以增强模型泛化能力。最终，取得B榜第六的成绩，展现了E2E模型结合数据增强的强大潜力。

来自北京航空航天大学与北方工业大学的参赛队伍采用Pipeline模型结构，分为两个阶段完成语音实体关系三元组抽取任务。第一阶段使用轻量级语音识别模型SenseVoiceSmall将.wav语音文件转写为文本，该模型采用非自回归端到端架构，具备推理速度快、中文适配强的优势；第二阶段则利用LoRA微调后的ChatGLM4语言模型从转写文本中提取结构化三元组。训练过程中，该队伍采用两轮训练策略，先在标注数据上训练，再利用模型生成的伪标签进行增强训练。在此基础上，团队还引入实体边界精确匹配与预定义关系类型约束，增强抽取结果的规范性与准确性。最终在中文语音关系抽取任务中，A榜F1得分为0.4940，B榜得分为0.4846，验证了该两阶段处理框架的有效性与稳定性。

6 总结

本次中文语音实体关系三元组抽取评测任务（CSRTE）依托于第二十四届中国计算语言学大会（CCL 2025）举办，由大连海事大学联合多所高校和科研机构共同组织。评测聚焦语音模态下的实体识别与关系抽取这一前沿课题，首次面向中文语音场景构建结构化知识抽取基准框架，涵盖端到端建模与传统流水线方法两种技术路径，并严格限定模型规模、训练数据及评测标准，确保公平性与技术先进性。

本次评测共吸引了257支队伍报名参赛，最终59支队伍完成A榜成绩提交，前15名晋级B榜，最终有7支表现优异的队伍提交了详细技术报告。从提交方法来看，参赛方案覆盖了Qwen2.5-Omni、GLM、ChatGLM、InternLM等多个大模型平台，涵盖了语音识别优化、LoRA参数高效微调、伪数据增强、多模态prompt设计等多种创新策略，充分展现了当前中文语音信息抽取领域的研究深度与实践活力。值得注意的是，端到端方法在避免ASR误差传播、增强系统一致性方面表现出明显优势，显示出未来语音信息抽取系统的关键发展方向。

评测不仅推动了中文语音信息抽取从“文本转写+结构化处理”向“语音直接生成结构化知识”的模式转变，也为后续多模态语言理解任务提供了统一任务定义、评估体系与数据基准。通过本次评测平台的构建与开放，主办方期望进一步激发语音与自然语言处理交叉领域的研究热情，推动构建更具通用性与实用价值的中文语音理解系统，为智能语音技术的深入发展提供坚实基础。

参考文献

Ido Cohn, Itay Laish, Genady Beryozkin, Gang Li, Izhak Shafran, Idan Szpektor, Tzvika Hartman, Avinatan Hassidim, and Yossi Matias. 2019. *Audio De-identification—A New Entity Recognition Task.* In North American Chapter of the Association for Computational Linguistics (NAACL), pages 197–204.

Hemant Yadav, Sreyan Ghosh, Yi Yu, and Rajiv Ratn Shah. 2020. *End-to-End Named Entity Recognition from English Speech.* In International Speech Communication Association (INTERSPEECH), pages 4268–4272.

Suwon Shon, Ankita Pasad, Felix Wu, Pablo Brusco, Yoav Artzi, Karen Livescu, and Kyu J Han. 2022. *SLUE: New Benchmark Tasks for Spoken Language Understanding Evaluation on Natural Speech.* In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 7927–7931.

Boli Chen, Guangwei Xu, Xiaobin Wang, Pengjun Xie, Meishan Zhang, and Fei Huang. 2022. *AIShell-NER: Named Entity Recognition from Chinese Speech.* In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 8352–8356.

Shilin Zhou, Zhenghua Li, Chen Gong, Lei Zhang, Yu Hong, and Min Zhang. 2024. *Chinese Spoken Named Entity Recognition in Real-world Scenarios: Dataset and Approaches.* In Findings of the Association for Computational Linguistics (ACL), pages 1872–1884.

Jinzhong Ning, Yuanyuan Sun, Bo Xu, Zhihao Yang, Ling Luo, and Hongfei Lin. 2024. *Breaking the Boundaries: A Unified Framework for Chinese Named Entity Recognition Across Text and Speech.* In Findings of the Association for Computational Linguistics: EMNLP 2024, pages 1250–1260.