

# 基于网络大模型的自动路测系统

Research and Practice on Empowering Automatic  
Road Testing Application Assistant Based on  
Network Large Model

## 研究与实践

杨飞虎,刘贤松,石旭荣,易峰,徐静静(中国联通网络运营事业部,上海 200050)

Yang Feihu,Liu Xiansong,Shi Xurong,Yi Feng,Xu Jingjing(China Unicom Network Operations Division,Shanghai 200050,China)

### 摘要:

为了进一步降低运营商网络测试成本,提升无线网络服务质量解决效率,通过在自动路测系统中引入网络大模型,实现以用户目标为中心的交互新范式,形成图形化界面+自然语言交互模式,可提供知识问答、关键指标查询、灵活数据分析、根因定位等功能,成为自然语言驱动的自动路测应用助手,从而提升无线网络优化效率。

### 关键词:

自动路测;网络大模型;交互模式;应用助手  
doi:10.12045/j.issn.1007-3043.2026.02.007  
文章编号:1007-3043(2026)02-0033-06  
中图分类号:TN929.5  
文献标识码:A  
开放科学(资源服务)标识码(OSID): 

### Abstract:

In order to further reduce the cost of network testing for operators and improve the efficiency of solving wireless network service quality, it introduces a network big model into the automatic road testing system to achieve a new paradigm of user centered interaction, forming a graphical interface+natural language interaction mode, which can provide functions such as knowledge Q&A, key indicator query, flexible data analysis, root cause localization, etc., becoming a natural language driven automatic road testing application assistant and improving the efficiency of wireless network optimization.

### Keywords:

Automatic road test; Network large model; Interactive mode; Application assistant

引用格式:杨飞虎,刘贤松,石旭荣,等. 基于网络大模型的自动路测系统研究与实践[J]. 邮电设计技术,2026(2):33-38.

## 1 概述

伴随着人工智能技术(AI)<sup>[1]</sup>的迅猛发展,大模型语言如Codex<sup>[2]</sup>、GPT-1到GPT-4等生成式AI技术的出现迅速引爆了AI的又一波热潮,大模型已被广泛应用于文本生成、语言翻译、自然语言理解等任务,显示出卓越的学习和推理能力,给AI的发展带来革命性的影响。越来越多的研究人员和企业开始探索如何将大模型与特定领域知识相结合,解决行业的难题<sup>[3]</sup>。

在通信行业中,网络规划、建设、维护、优化、运营是非常耗时、复杂且需要大量人力成本的工作<sup>[4]</sup>。为

了降低运营商网络测试成本,改善无线网络服务质量的问题,本文提出一种基于网络大模型的自动路测应用助手实践方案。一方面,该方案利用网络大模型全面赋能自动路测应用,与无线网络“随心测”系统联动,实现主要道路指标查询、问题路段查找、功能导航等功能,形成图形化界面+自然语言交互的多模态交互模式,提高产品使用的灵活性,提升用户体验。另一方面,通过引入大模型,该方案实现以用户目标为中心的交互新范式,调用系统已有的分析模块能力,通过使用提示工程、大小模型协同调用等技术,为路测场景提供不同的解决方案,利用已有应用系统的API接口和AI小模型根因分析、智能决策结果,为一线网优人员在国道省道、高铁、高速等场景提供一种有

收稿日期:2026-01-05

效的服务质量改善方法,有助于提升网络优化工作的效率。

## 2 基于网络大模型的无线网络优化场景实现

网络大模型定位于L1层大模型,在中国联通元景大模型基础上,以“懂网络、懂运营”为目标,是面向网络自智运营形成的通信领域行业大模型(见图1)。它提供知识问答、人机交互、数据分析、方案生成等能力,协同场景化小模型、知识图谱<sup>[5]</sup>和OSS系统API,实现故障处置、网络优化、智能开通、决策辅助等能力,赋能解决网络规划、建设、维护、优化、运营的复杂问题,大幅提升作业效率。

### 2.1 基于LoRA<sup>[6]</sup>微调大模型的网络指标提取

以中国联通元景大模型为基础模型,针对无线网络优化领域进行训练微调(Fine-tuning),可以采用Q-LoRA轻量级微调大模型的方法,该方法思路主要包含以下3个方面。

a) 在原始PLM(Pre-trained Language Model)旁边增加一个旁路,做一个降维再升维的操作,来模拟所谓的intrinsic rank。

b) 训练的时候固定PLM的参数,只训练降维矩阵A与升维矩阵B。而模型的输入输出维度不变,输出时将BA与PLM的参数叠加。

c) 用随机高斯分布初始化A,用0矩阵初始化B,保证训练的开始,此旁路矩阵依然是0矩阵。

LoRA这个思路类似于残差连接,同时使用这个旁路的更新来模拟Full Fine-Tuning<sup>[7]</sup>的过程,Full Fine-Tuning可以被看做是LoRA的特例(当 $r$ 等于 $k$ 时)。在推理过程中,LoRA也几乎未引入额外的Infer-

ence Latency,只需要计算 $W=W_0 + \Delta W$ 即可。

因此通过收集整理上千条基于网络指标查询问题的基础语料,并使用大模型提示工程进行语料重写、扩写获取上万条训练集,基于LoRA方法对大模型进行微调,通过自适应优化,提高网络指标的准确性和稳定性,适用于各种网络指标提取任务。该方法具有高效性和自适应性的优势,能够在大规模语料数据上进行快速且准确的指标提取,并利用大模型的语义理解和文本生成能力,为用户提供一个友好、自然和高质量的交互对话体验。

### 2.2 基于流形嵌入的q-q匹配相似问题高精度召回

为提升大模型召回准确率,通常的做法是“先检索再整合”,思路如下。

a) 首先准备好文档,并整理为纯文本的格式。把每个文档切成若干个小的chunks。

b) 调用文本转向量的接口,将每个chunk转为一个向量,并存入向量数据库。

c) 当用户发来一个问题的时候,将问题同样转为向量,并检索向量数据库,得到相关性最高的一个chunk。

d) 将问题和chunk合并重写为一个新的请求发给chatglm的api。

这里主要是将用户请求的query和document做匹配,也就是所谓的q-d匹配。q-d匹配的问题在于query和document在表达方式存在较大差异,通常query是以疑问句为主,而document则以陈述说明为主,这种差异可能会影响最终匹配的效果。一种改进的方法是不直接做q-d匹配,而是先通过document生成一批候选的question,当用户发来请求的时候,首先

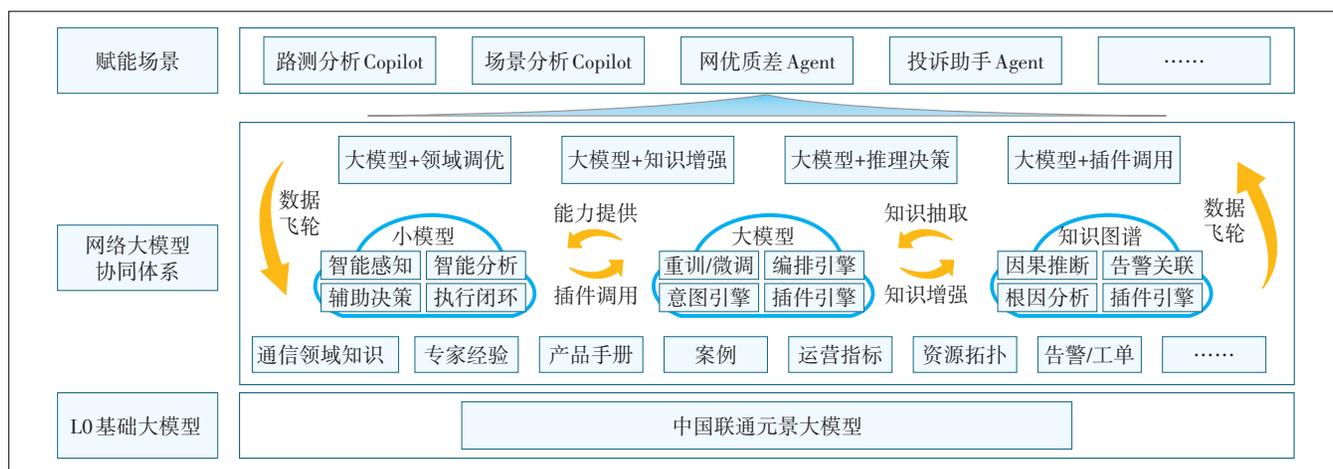


图1 网络大模型赋能网络生产作业场景架构

把 query 和候选的 question 做匹配,进而找到相关的 document 片段,这种方法就是“q-q 匹配”。

通过将所有问题进行 tokenization 和 embedding<sup>[8]</sup>,将问题进行向量化表示,映射成为高维空间中的一个点,再通过相似语句扩写,根据业务层对问题的分类,将原始预训练模型权重下可能相距较远,但业务上实际相似的问题映射成高维流形中的 2 个相近的点,从而完成了高精度的 q-q 相似问题召回。

### 2.3 基于大小模型协同<sup>[9]</sup>的解决方案

通过插件和知识库的方式进行能力调用实现大小模型协同技术,完成基于结构化数据挖掘推理的网优方案生成,基本流程如下(见图 2)。

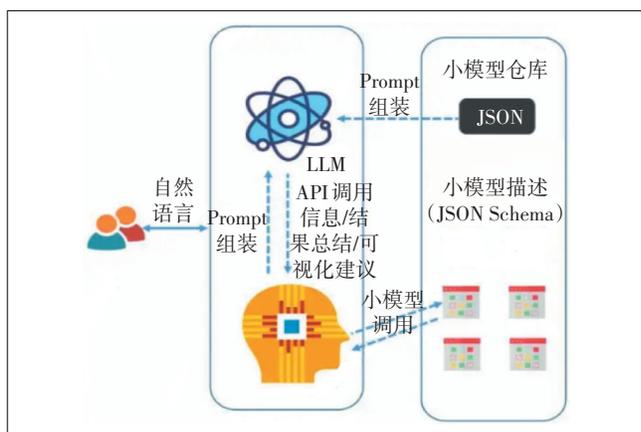


图 2 大小模型协同调用流程

a) 首先需要定义小模型的 API 接口(如现有部署在 AI 平台上的开放 API),这个需要根据各自的业务情况进行充分设计与实现,形成小模型 API 的使用“说明书”(JSON Schema 描述,也就是小模型部署仓库里面的具体描述)。

b) 使用者输入自然语言,系统借助 LLM 将使用者的输入问题转化为对小模型 API 工具的调用,包括小模型 API 的名称与提取的参数。

c) 根据 LLM 的响应,调用指定的小模型 API,取得返回的数据。根据情况需要,在一些场景下可能还需要将返回的数据再附加到用户输入,再次交给 LLM,由 LLM 来输出最终响应给客户的分析结果。

充分利用大模型的语义理解和文本生成能力,结合小模型的专业和高效能力,提供一个完整、智能和可靠的解决方案。

## 3 基于网络大模型的自动路测系统

### 3.1 整体思路

通过网络大模型赋能,自动路测系统打破用户必须按现有应用产品设计规则点击操作的模式,将图形界面(GUI)封装的能力 API 化,用户可利用网络大模型生成 SQL、可视化图表,通过 API 编排呈现原系统没有开发构建过的指标,甚至将来可实现跨系统的功能编排调用。

### 3.2 系统模块

系统共分为 3 个技术模块:自动路测随心测系统、能力编排引擎、大模型系统。

a) 自动路测随心测系统。可提供 4G/5G 道路免测试免分析能力,充分利用了网络大数据和 AI 能力,通过采集并关联多维 4G/5G 网络的 MDT/XDR/PM/CM 数据,实现全网道路用户 7×24 h 的众筹测试,支持覆盖、质量、语音感知、数据感知等四大类 10 多个关键指标的地理化呈现及分析。以颠覆路测生产模式,提升工具化能力,让产品从好用到全面好为目标,成为网优人员的必备工具。

b) 能力编排引擎。作为整个方案的中枢,编排引擎是一个集成平台,它通过定义和执行接口调用流程,实现不同系统间的信息交互和数据整合。它不仅管理应用系统与大模型之间的交互,还负责对话历史记录、任务调度和业务流程的自动化,确保整个系统的高效协调运作。

c) 大模型系统。它负责进行复杂的数据分析和生成智能响应,利用先进的算法和大量的数据,大模型系统能够理解用户的自然语言与业务数据含义,提供深度的洞察力和预测能力,支持决策的制定和问题的解决。

### 3.3 业务流程

基于网络大模型的微调和提示工程,利用语音/自然语言对话的方式来执行系统功能、处理数据返回结果,并生成专家建议,主要包括如下 6 个步骤(见图 3)。

步骤 1:用户在网络大模型对话页面输入问题(如请问上海市长宁路 7 月 1 号的整体指标情况怎么样?),网络大模型对话页将该问题传递给“能力编排引擎”。

步骤 2:“能力编排引擎”收到用户的 query 后,开始调用网络大模型的服务接口,将该问题 query 传递给网络大模型。

步骤 3:大模型服务收到 query 后,开始做意图识别和参数抽取,一部分是识别用户问题意图对应哪个意图类别(如指标查询类、质差分析类等),并将该问

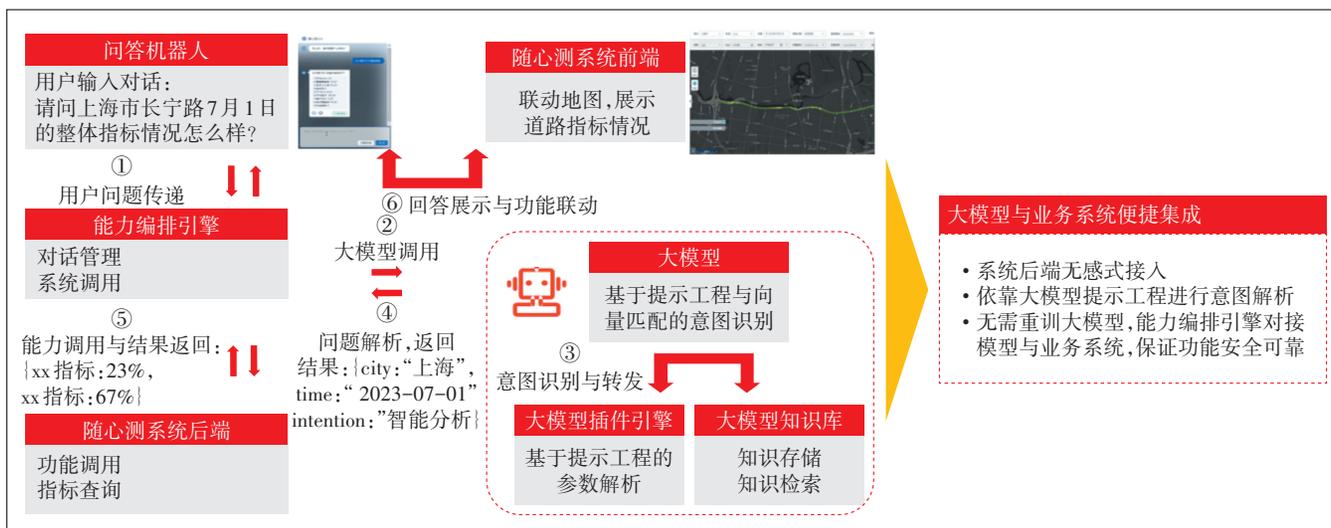


图3 网络大模型赋能自动路测随心测系统技术实现方案

题中关键的参数进行抽取,包括时间、地点、指标类型。通过插件方法实现意图识别和参数抽取,主要通过填写唯一标识(operationId)、概述(summary)、详细描述(description)定义工具功能,填写parameters下每一个参数的标识(name)、详细描述(namedescription)、是否必须(required)、默认值(default)定义工具的使用方式。比如图3中随心测指标查询分为道路指标和小区指标2个细粒度的意图,它们分别提取道路+指标和小区名称+指标2个参数。

步骤4:将大模型参数抽取的结果进行解析,对于可枚举的参数,可通过配置字典库中的同义词词库进行精确抽取;此外,还可通过配置正则表达式、语境抽参等规则完成参数提取,在参数pattern字段配置。常用的参数类型可通过内置的多种ner模型完成较精准的抽取,典型的参数如时间/时间段、地址、人名、组织名、手机、邮箱等,在参数description有实体触发词,则触发对应的实体抽取。根据插件中配置的参数描述,通过组prompt由大模型完成参数抽取,如图3中的步骤④中抽取的“城市”字段是Shanghai,转换为“上海”,并组装成可调用服务的请求json格式。

步骤5:将该json调用“随心测系统后端”,调用相关的后端api接口,得到相应的结果返回,如从随心测系统查询在xxx时间点xx地点xx指标具体值,并将该结果返回“能力编排引擎”。

步骤6:收到相关信息后,将该信息一方面回吐网络大模型对话页,在对话框展示对话结果,同时自动路测“随心测系统前端界面”同步完成系统的链接和跳转。

### 3.4 关键技术

#### 3.4.1 意图识别

使用RAG+提示工程技术,配合系统功能描述,完成系统功能调用的意图识别,使用Qwen2-72B模型,对用户问题进行描述扩充和问题仿写,并生成用户意图相关的推理结果,经专业网络工程师进行数据标注,微调Qwen2-7B模型,让意图识别模型准确率、推理效率、部署代价得到非常大的提升。

a) RAG和提示工程技术。首先,利用RAG技术从知识库中检索与用户问题相关的信息,然后结合提示工程技术,引导Qwen2-72B模型生成用户问题的扩充描述和仿写,以便更好地理解用户意图。

b) 描述扩充和问题仿写。使用Qwen2-72B模型对用户问题进行描述扩充和仿写,生成更详细的描述和不同形式的同义问题,从而丰富训练数据并增强模型的泛化能力。

c) 数据标注。专业网络工程师对生成的描述和仿写问题进行数据标注,确保数据的准确性和质量,这些标注数据将用于后续的模式微调。使用标注后的高质量数据对Qwen2-7B模型进行微调,通过监督学习方法提升模型在意图识别任务上的表现,确保其能够准确理解和处理用户的输入。

d) 模型部署。将微调后的Qwen2-7B模型部署为在线推理服务,并集成到现有系统中,实现意图识别和功能调用,确保系统能够高效地响应用户请求。

通过上述方案,利用RAG和提示工程技术结合Qwen2-72B模型进行描述扩充和仿写,并对Qwen2-7B模型进行微调,可以显著提升意图识别的准确率和

推理效率,同时降低部署代价。专业网络工程师的数据标注确保了模型训练数据的高质量,从而进一步增强了模型的性能。

### 3.4.2 地址、场景名称检索

将 Elasticsearch(ES)作为检索工具,对系统中全国场景名称、地址数据进行索引,实现高效的检索。

a) 词库索引。构建词库索引,根据系统所有地址与场景名称词库进行统计,基于系统中所有地址和场景名称构建词库,统计关键词的出现频率,并使用 BM25 算法(一种相关性评估指标,能消除文档长度的影响,使评估结果更为平衡)计算相关性分值。

b) 场景名称与地址检索词库。构建了一个包含全国 42 万特定场景与道路的词库,并将其导入到 ES 数据库中。

c) 处理流程。以“上海市中山公园”为例,处理用户问题的流程如下。

(a) 省、市名称分级索引。通过二级词库进行句子分词和省市分流。

(b) 普通词库。统计词频,分词数量超过 2 000 个,如“上海”“中山”“公园”“中山公园”。

(c) 省市分流。锁定“上海”场景名称库。

(d) 分词索引检索。使用 BM25 分值进行排序,例如“公园”的分值为 5.34,“中山公园”的分值为 14.75,“公园广场”的分值为 13,最终确定为“中山公园”。

(e) 补充词库。统计字频,分词数量达到 7.8 万以上,以增强检索的准确性和覆盖面。

总的来说,场景地址提取技术通过构建词库索引,使用 BM25 算法评估关键词的相关性,以及利用 ES 数据库进行高效的检索,从而实现对用户问题中场景名称与道路地址的快速准确提取。

## 4 基于网络大模型的自动路测系统实践效果

### 4.1 更好的软件交互体验,降低系统使用门槛

自动路测应用系统是基于生产作业流程设计的应用功能,使用操作过程比较复杂,主要存在如下问题:操作繁琐,使用门槛高;面向功能实现,菜单式,系统层级复杂;缺乏个性化定制和自适应能力。一线人员需要经过系统培训后才能掌握所需的功能位置,完全掌握应用系统的核心功能需要一定的学习门槛。

基于网络大模型的自动路测系统通过网络大模型与应用系统联动,实现了图形界面和语言界面混合的交互模式,使用自然语言即可直接调用应用系统的

核心能力,改善自动路测分析流程,由原来 7 步操作简化为 1 次问答,提升了网优工作效率和用户体验,同时可提升 85% 的道路问题分析效率。

### 4.2 简化网络优化问题处理过程,提升分析效率

传统路测处理过程效率较低,网络优化人员需要通过多系统进行指标提取、问题分析、日志分析等,跨平台效率低,主要依赖于网络优化工程师经验,无法实时快速响应网络突发问题。

基于网络大模型的自动路测系统通过网络大模型赋能后,可直接面向全国不同省份地(市)的网络优化人员,突破时空限制,利用网络大模型 copilot 能力可快速获取任意道路网络关键指标及道路问题,通过大小模型协同技术完成能力调用,输出生成式的根因分析结果和优化解决方案,辅助并指导网优人员执行优化方案。

完成优化调整后,可通过大模型直接询问优化对比效果,无需再安排人员去现场复测,极大提升了问题闭环管控过程效率。如在对话框中输入:“请查询上海市杨高南路 2024 年 5 月 16 日前闭环的问题簇有哪些?”大模型会快速将出现的杨高南路道路问题与近期经优化处理闭环的具体结果进行对比,如点击第某个问题簇 ID,通过大模型调用小模型的系统诊断根因结果为:RF 无主覆盖,优化解决方案为 8430860(锦代\_4)方位角向左扳 10°。经过网优工程师在 2024 年 4 月 1 日优化处理后,道路问题已闭环,网络质量显著提升(见表 1)。

表 1 杨高南路道路问题优化前后关键指标对比结果

| 指标名称              | 闭环前   | 闭环后    |
|-------------------|-------|--------|
| 平均电平/dBm          | -81.6 | -71.42 |
| 下行 SINR/dB        | 9.05  | 15.27  |
| 弱覆盖率/%            | 0.04  | 0.05   |
| 差 SINR(<0 dB)比例/% | 26.5  | 0.42   |
| 重叠覆盖率/%           | 29.78 | 1.02   |
| Mod 3 干扰率/%       | 26.61 | 0.97   |

### 4.3 改变传统菜单式软件开发模式,提升迭代开发效率

现有的自动路测系统是基于传统菜单式软件开发模式,开发周期较长(一般以月为级别),研发团队基于产品 PRD 的功能设计,按照多个迭代周期需要多岗协同,涉及架构、UI、前端、后端、测试等。

基于网络大模型的自动路测系统,是将自然语言中的关键内容进行提取,转化为业务系统可识别的参数,业务系统处理后形成界面所需要的结构化数据,

如图4所示,界面中已经完全没有菜单式选项,只有Chat交互对话框和地图。自动路测系统基于网络大模型提供的插件配置能力进行AI能力的集成,每个插件可以配置多种问题模式,当大模型匹配到对应的问

题模式后,可确定问题的对应的业务“意图”和这个“意图”所需要的参数集。这种开发方式开发周期短(以天为级别)且开发成本低(只需插件配置及前后端联调),可提升60%以上的开发效率。



图4 基于网络大模型软件开发页面呈现效果

## 5 结束语

道路测试优化是无线网络优化过程中的一个重要环节,传统的路测作业需要投入大量的人力物力,包括线路规划、测试资源的协调,并按照计划在在规定时间内完成负责区域的所有线路的数据采集、道路问题分析、优化方案输出及方案落地后效果评估。通过引入网络大模型,重塑自动路测应用系统,为网络优化工程师量身打造全天候远程随心测、随时查的“网络驾驶舱”,实现网络问题的全程闭环管控,提升了运营商数字化执行效率。

### 参考文献:

- [1] 刘予峰,蒋颀阳. 基于人工智能技术(AI)的智能网络优化方法分析[J]. 电子世界,2012(10):14-15.
- [2] CHEN M, TWOREK J, JUN H, et al. Evaluating large language models trained on code [EB/OL]. [2025-04-02]. <https://arxiv.org/abs/2107.03374>.
- [3] RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners [EB/OL]. [2025-02-09]. [https://cdn.openai.com/better-language-models/language\\_models\\_are\\_unsupervised\\_multitask\\_learners.pdf](https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf).

- [4] 刘贤松,陈凌,尹俊. 通信行业语言大模型技术和应用研究[J]. 邮电设计技术,2023(7):14-19.
- [5] JI S X, PAN S R, CAMBRIA E, et al. A survey on knowledge graphs: representation, acquisition and applications [J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(2):494-514.
- [6] LI H, SMITH J, ZHANG L. Efficient fine-tuning of large language models with low-rank adaptation [J]. Journal of Artificial Intelligence, 2023, 45(2):150-160.
- [7] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [EB/OL]. [2025-02-09]. <https://arxiv.org/abs/1810.04805>.
- [8] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [EB/OL]. [2025-02-11]. <https://arxiv.org/abs/1301.3781>.
- [9] WANG Y W, SHEN T, ZHANG S Y, et al. Advances in edge-cloud collaboration and evolution for large-small models [J]. Journal of Image and Graphics, 2024, 29(6):1510-1534.

### 作者简介:

杨飞虎,毕业于华东理工大学,高级工程师,硕士,主要从事无线网络AI产品规划及设计工作;刘贤松,毕业于武汉大学,高级工程师,硕士,主要从事网络AI研究工作;石旭荣,毕业于山东大学,助理工程师,学士,主要从事大模型及AI算法研究工作;易峰,毕业于悉尼大学,硕士,主要从事大数据和AI能力研发工作;徐静静,毕业于长安大学,工程师,学士,主要从事前端研发工作。