

通信行业语言大模型 技术和应用研究

Research on Large Language Model Technology and Application in Communication Industry

刘贤松, 陈凌, 尹俊 (中国联通智网创新中心, 上海 200050)

Liu Xiansong, Chen Ling, Yin Jun (Intelligent Network & Innovation Center of China Unicom, Shanghai 200050, China)

摘要:

ChatGPT 的出现迅速引爆了 AI 的又一波热潮。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。语言大模型在通信运营商中有着非常广阔的应用前景。阐述了语言大模型开发的基本技术方案及原理并对其在通信行业的应用进行了研究与展望。

关键词:

语言大模型; 人工智能; 通信; 运营商

doi: 10.12045/j.issn.1007-3043.2023.07.003

文章编号: 1007-3043(2023)07-0014-06

中图分类号: TP183

文献标识码: A

开放科学(资源服务)标识码(OSID):



Abstract:

The emergence of ChatGPT quickly ignites another wave of AI boom. In the communication industry, network planning, construction, maintenance, optimization, and operation are very time-consuming, complicated, and labor-intensive tasks. The language big model has a very broad application prospect in communication operators. It expounds the basic technical scheme and principle of large language model development, and researches and looks forward to its application in the communication industry.

Keywords:

Large language model; Artificial intelligence; Communication; Operator

引用格式: 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.

0 前言

ChatGPT 的出现迅速引爆了 AI 的又一波热潮。作为一种人工智能技术驱动的语言大模型, ChatGPT 使用了 Transformer 神经网络架构, 拥有语言理解和文本生成能力。ChatGPT 不单是几乎与人无异的聊天机器人, 还能执行撰写邮件、视频脚本、文案、翻译、代码等任务。ChatGPT 极大便利了 AI 的开发和应用, 给 AI 的发展带来革命性的影响。每个行业、每个领域都值得 AI 化, 人工智能将从服务的规模化、个性化、普惠化

和持续化带动井喷式增长, 促进经济高速发展。AI2.0 时代, AI 将成为数字经济的重心, 有望超越云计算成为第 4 波计算浪潮。

越来越多的企业开始关注如何将大模型应用于生产实践中, 以提高业务效率和降低成本。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。我国通信网络规模庞大, 网络结构复杂, 网络需要支撑的业务要求很高, 网络稳定运行和保证优秀的网络服务质量的压力很大, 这些都给网络日常生产运营工作带来了巨大的挑战和压力。如何把大模型应用到网络生产运营过程中, 助力网络向自智网络演进是我们需要重点考虑

收稿日期: 2023-05-08

的问题。

ChatGPT 通用大模型采用自回归生成式训练、单向注意力机制、多层 Transformer 架构,通过海量参数和海量数据产生了极大的性能提升,辅以有监督调优和人类反馈的强化学习,形成了突破性的 AI 产品,但是通用大模型并不能“包治百病”,通用大模型的直接应用存在数据可信度不高、数据安全不可控、行业知识缺乏等问题,无法完全满足运营商的生产运营需求,企业从头训练一个私有化的通用大模型则需要海量数据和海量算力,还需要一批顶尖的 AI 人才,难度极大。为了解决这些问题,可以通过构建通信行业领域大模型来提高网络相关工作效率,并为一线网络运维工程师提供优质的智能化网络应用产品。

1 语言大模型技术方案研究

1.1 基础模型选择

语言模型是一种利用人工智能技术来预测或生成自然语言表达的方法,它们在自然语言处理(NLP)领域有着广泛的应用。语言模型的发展经历了从基于统计的方法到使用神经网络的方法,再到最近的预训练语言模型(PLM)技术的转变。PLM 技术是一种利用 Transformer 模型在大规模文本数据上进行预训练,然后在特定任务上进行微调的方法,它可以提高语言模型的语言理解和表示能力,提升 NLP 任务的性能。

大型语言模型(LLM)是一种规模非常庞大的 PLM,它们具有超过一定阈值(例如 10 亿或 100 亿)的参数数量,可以显著提高语言模型的性能,并展示出一些小型语言模型所不具备的能力,例如上下文学习、常识推理、知识融合等。LLM 具有强大的语言理解和生成能力,可以应用于多种 NLP 任务,例如机器翻译、问答、摘要、对话等。但是,LLM 也带来了一些技术和伦理方面的挑战和风险,例如计算资源消耗、数据质量控制、模型可解释性、社会责任等。

OpenAI 和 Google 是 LLM 的领导者,他们分别开发了 GPT 系列和 BERT 系列等多个大型模型集合。这些大型模型集合包括了各种不同领域和任务的 PLM,可以应用于多种场景,例如搜索引擎、社交媒体、教育、医疗等^[1]。但是,这些 PLM 可能存在一些偏见、误解、伦理等问题,例如歧视、误导、欺骗等。因此,需要对 LLM 进行更深入的研究和监督,以确保它们能够为人类社会带来积极的影响,而不是负面的后果。同时,也需要加强对 LLM 的公平性、可靠性、安全性等方面

的评估和改进。

1.2 大模型训练调优

1.2.1 大模型的预训练

LLM 是一种利用大量文本数据训练的深度学习模型,能够在多种自然语言处理任务中表现出优异的性能。然而,创建或复制 LLM 并不是一件容易的事情,它需要解决技术和计算资源的难题。我们可以借鉴已有的 LLM 经验和公共资源,如开源模型检查点、API、语料库和支持 LLM 的库等,来降低开发成本和难度。

LLM 的核心是数据预训练,即使用大量无标注或弱标注的数据来训练模型的基础参数,使其能够学习到语言的通用规律和知识。数据预训练需要高质量的数据集,包括通用数据和专业性数据。通用数据是指覆盖广泛领域和主题的文本数据,如网页、对话、书籍、多语言文本等;专业性数据是指针对特定领域或任务的文本数据,如科学文本和代码数据等。不同类型的数据可以提供不同层次的语言信息,有助于提升 LLM 的泛化能力和适应能力。LLM 的预训练数据来源分布如图 1 所示。

除了数据预训练外,LLM 还需要进行数据预处理、网络框架设计和训练方式选择等步骤。数据预处理是指对原始数据进行清洗、分词、编码等操作,以便于模型输入和输出。网络框架设计是指选择合适的神经网络结构,以实现高效的信息传递和表示学习。训练方式选择是指采用合适的自监督学习方式,以利用无标注或弱标注的数据进行模型训练。目前,最常用的网络框架是 Transformer 模型,最常用的自监督学习方式是掩码语言建模(MLM),它们已经在多个 LLM 中得到了广泛应用。

1.2.2 大模型微调调优

为了进一步提升预训练的语言模型的性能和适应性,可以通过指令微调和对齐微调进行优化。

1.2.2.1 指令微调

指令微调是一种基于自然语言指示的微调方法,它旨在提升或唤醒 LLM 的功能,使其能够理解并执行用户给出的自然语言指示,例如生成文本、回答问题、执行计算等。指令微调不仅可以增强 LLM 的通用性和灵活性,还可以改正模型在某些情况下的缺陷,例如生成不合理或不一致的内容、缺乏常识或逻辑推理能力等。

指令调整比预训练更有效率,因为它只需用少量

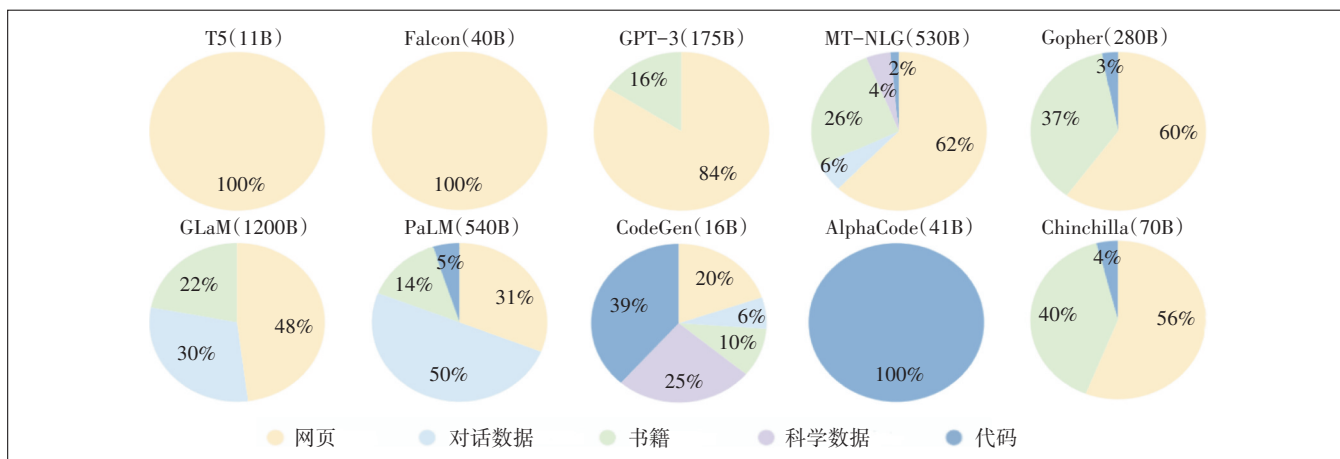


图1 LLM的预训练数据来源分布^[2]

实例进行训练。在进行指令调整时需要平衡数据分布,常规操作是使用等额抽样,并增大高品质数据集的抽样比例以提升性能。另外,指令调整可以和预训练一起使用,同时利用原始文本数据和格式化数据进行多任务学习,从而兼具预训练和指令调整的优势^[3-5]。LLM 指令调整过程如图2所示。

1.2.2.2 对齐微调

对齐微调是一种基于人类反馈的微调方法,它旨在将 LLM 的行为与人类价值取向或偏好相对应,例如遵守道德规范、尊重社会多样性、保护隐私权等。对齐微调需要收集高质量的人类反馈,并考虑一些非常不同的评判标准,例如公平性、可解释性、可信赖性等。对齐微调可以提高 LLM 的安全性和可靠性,防止模型产生有害或不良的影响。

为了实现语言模型(LM)与人类价值观的一致性,

工业界提出了一种从人类反馈中进行强化学习(Reinforcement Learning from Human Feedback, RLHF)的方法。RLHF 使用强化学习算法,根据人类反馈来学习一个奖励模型,从而对 LM 的行为进行调节和优化。RLHF 系统主要由以下3个部分构成。

- a) 一个预训练的LM。可以生成自然语言文本或执行其他任务。
- b) 一个从人类反馈中学习得来的奖励模型。可以评估LM的输出质量和符合度。
- c) 一个用于训练LM的强化学习算法。可以利用奖励模型的指导来更新LM的参数。

如图3所示,RLHF包括3个阶段:监督微调、奖励模型训练和RL微调。在有监督微调阶段,使用标注数据来对LM进行初始训练,以适应特定的任务和领域。在奖励模型训练阶段,使用人类反馈数据来对奖

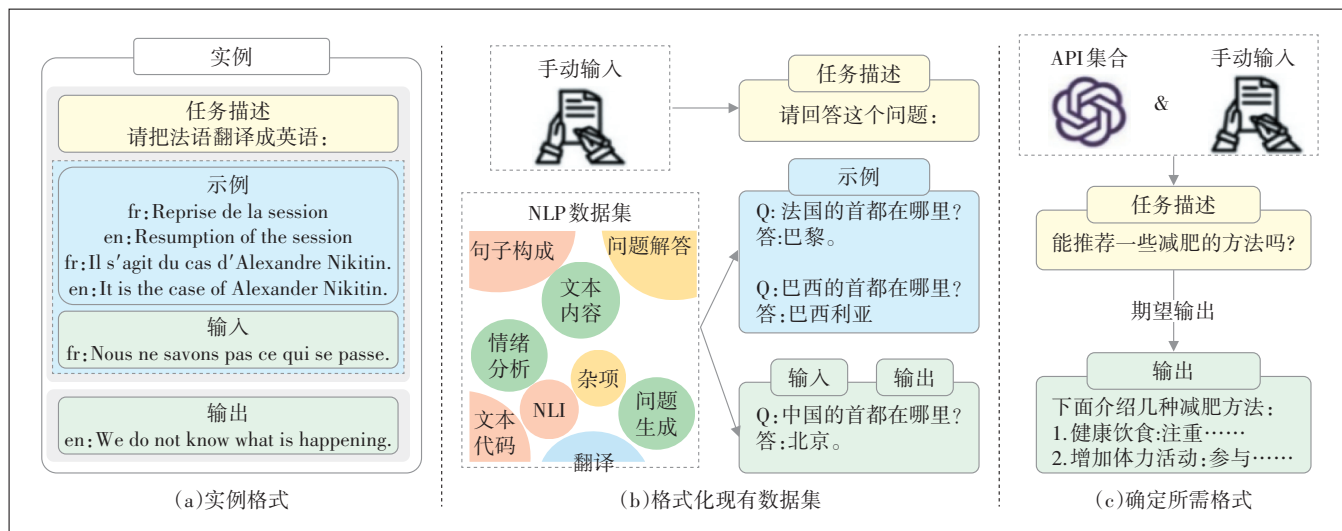


图2 LLM指令调整过程^[2]

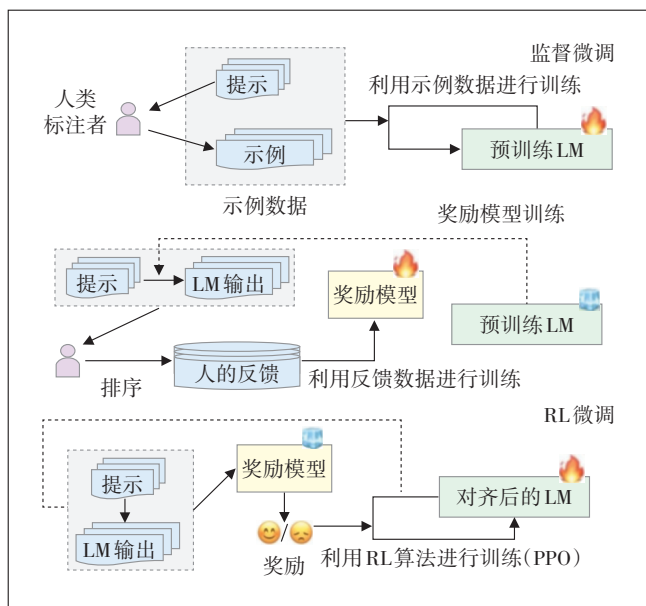


图3 基于人类反馈的强化学习流程^[2]

励模型进行训练,以捕捉人类对LM输出的偏好和评价。在RL微调阶段,使用强化学习算法来对LM进行进一步训练,以最大化奖励模型的期望值。这个过程可以重复多次,以更充分地对齐LM与人类价值观。

1.3 大模型与知识图谱/知识库结合

知识图谱和大模型是2种不同的知识存储和使用方式,各有优势和劣势。知识图谱是一种将知识以实体、关系、属性等概念进行显式、结构化、可控的表示和存储的方法,可以方便地进行知识查询、理解、更新和归因,同时也有利于进行逻辑推理和提供可解释性。大模型是一种利用神经网络参数学习和存储知识的方法,可以通过生成式的方式表达知识,表现出色的交互任务理解和生成能力,例如自然语言处理、计算机视觉等领域。然而,大模型也存在一些问题,如事实准确性、来源追溯、专业知识获取等方面的挑战。

如何更好地融合知识图谱和大模型,发挥其优点,克服其弱点,是当前研究的重点。一方面,可以利用知识图谱为大模型提供结构化、可信、可解释的知识来源,增强大模型的事实准确性、来源追溯、专业知识获取等能力。另一方面,可以利用大模型为知识图谱提供生成式、灵活、多样的知识表达方式,增强知识图谱的交互任务理解和生成能力。这样,可以实现知识图谱和大模型之间的互补和协同,提高人工智能系统的整体性能和效率。

有3种主要的结合知识图谱和大模型的策略,可

以提高语言模型的知识表示和推理能力。

a) 将知识图谱转化成文本,作为文本语料的一部分去做训练。这种策略可以将知识图谱中的信息以自然语言的形式融入到大模型中,从而增强其语义理解和生成能力。例如,K-Adapter^[6]是一种基于Transformer的语言模型,它使用了一个适配器层来融合来自知识图谱的文本信息,从而提高了在多个下游任务上的性能。

b) 保留知识图谱本身的结构化信息,在训练过程中作为特殊内容或结构融入大模型中。这种策略可以利用知识图谱中的图结构和关系信息,为大模型提供更丰富和准确的知识表示。例如,KT-NET^[7]是一种基于BERT的语言模型,它使用了一个图注意力网络来编码来自知识图谱的结构化信息,并将其与文本信息相结合,从而提高了在问答和阅读理解等任务上的性能。

c) 使用外挂知识增强。将知识图谱创建为一个外挂知识库,当需要陈述事实、进行推理时,从知识图谱中获取事实部分,然后让语言模型自行处理推理部分。这种策略可以在不改变大模型本身结构和参数的情况下,动态地为其提供相关的知识支持。例如,GPT-4^[8]是一种基于GPT-3的语言模型,它使用了一个外挂知识库来存储来自知识图谱的事实信息,并在生成过程中根据上下文和查询条件从中检索相关信息,从而提高了在对话、摘要和故事生成等任务上的性能。基于向量检索知识库的文档问答任务流程如图4所示。

2 通信行业语言大模型应用场景

通信行业语言大模型在通用语言大模型的基础上引入通信行业的数据和知识。通过重训练、调优、知识增强等方式形成具有强大知识理解和生成能力的行业领域大模型。它可以自动解答常见网络专业问题,自动识别并修复网络故障,自动优化网络性能,并且可以与人类进行自然语言交互。主要有以下一些场景。

2.1 通信知识智能问答

大模型在通信领域通用知识问答领域有着广泛的应用前景,主要面向通信领域的运维工程师、网络架构师、技术支持人员等角色,解决他们在通信领域知识获取、沉淀、管理、应用难的问题,提供更便捷、高效的解决方案。

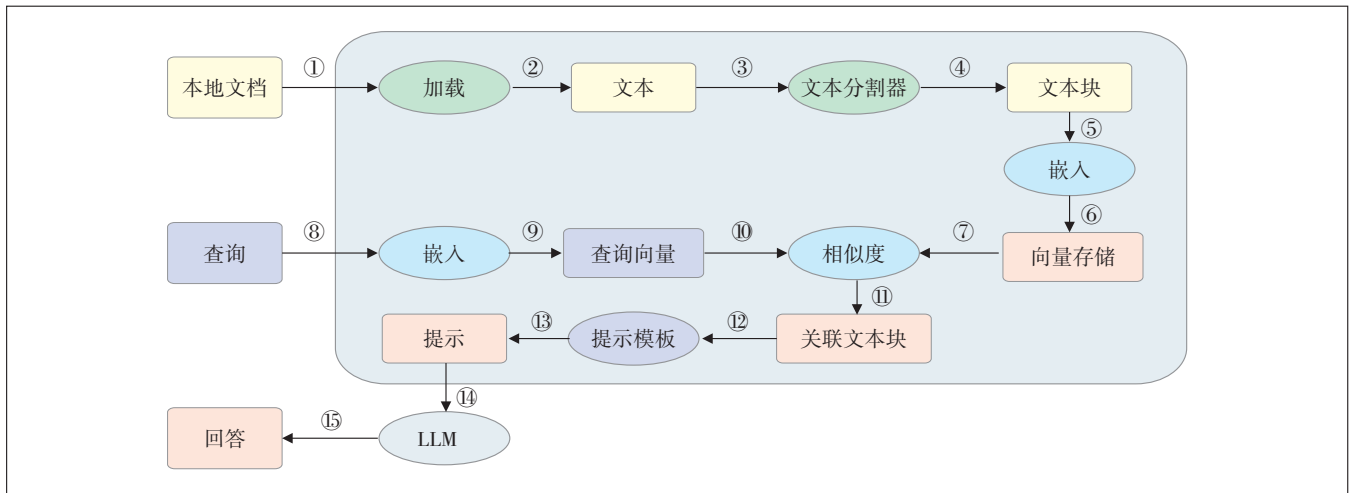


图4 基于向量检索知识库的文档问答任务流程^[4]

随着网络结构日趋复杂,网络运维、优化的难度将成倍增加,需要运维工程师一专多能,具备较宽的知识面。行业领域大模型可以代替目前的知识百科、专家经验库。通过大模型的训练和应用,提高通信领域知识问答的准确率和效率。用户只需简单的语言描述问题,即可得到精准、自然流畅的回答。

同时,大模型技术还可以实现知识沉淀、知识管理、知识提取和知识生产的能力,为员工提供自动存储、检索和生成网络专业知识的服务,大幅提升网络业务条线复杂度较高的智力密集型工作质效,形成人机协同的工作模式。与传统的知识问答系统相比,大模型可以结合上下文信息和场景分析来提高回答的合理性,具备更加自然、流畅、准确地回答能力,从而提高用户满意度和信任度。

2.2 配置自动生成

随着网络接入终端数量的不断增加,业务越发的复杂,传统人工配置和管理网络设备的方式已无法快速响应不同业务需求,网络运营成本居高不下。基于意图驱动网络与软件定义网络技术,将网络管理与业务需求相结合,使网络管理者或应用通过意图语言的方式与网络进行交互^[9],基于AI算法完成网络的动态配置,是进一步提高网络管理运维效率,支持业务的连续性、敏捷性和动态性以及实现高度自动化和人性化网络管理的必然趋势^[10]。

网络运维管理人员可通过大模型知识引擎以人机交互的方式,无须关注底层设备的具体配置和操作,只需要指定网络的目标和目的,便可自动生成适配不同设备厂商的南向网络配置报文,轻松实现网络

业务动态配置开通或一键生成网元配置模板,并通过与网管系统的集成实现自动配置激活。此外,还可以通过多轮对话高效查询网络资源位置、状态等信息,降低操作门槛。

2.3 辅助编码

辅助编码是大模型的典型应用场景,大模型的智能化和语义理解能力以及上下文对话能力,可以帮助代码软件研发人员更好地完成编程工作,提高编码效率。

研发人员在编写代码时可能会遇到各种问题,如编译错误、语法错误等,大模型可以辅助研发人员快速解决这些问题,它可以充当一个在线教程,提供代码示例、错误原因定位、解决方案等,从而缩短研发人员使用搜索引擎寻求解决方案的时间。通过大模型生成代码片段、SQL语句和命令脚本,进行代码优化、代码评审与代码重构,可以大幅度降低编写代码的时间成本,从而显著提高研发人员的生产力。在自动化测试方面,大模型可以快速生成全面的测试用例和测试代码,从而提高测试的完整度和准确性。

另外利用大模型辅助编码,还可以降低编码初学者的学习难度和时间,例如通过简单的prompt,大模型就可以快速地生成一些工具类代码,初学者可即取即用,从而实现快速上手编程。

2.4 运维日志自动分析

传统的运维日志分析需要人工逐条查看,来分析故障可能产生的原因,耗费大量时间和人力。大模型通过解析网络故障告警和关键日志的文本信息,自动识别出异常事件、故障原因等信息,为运维人员提供

针对性处理方法和建议,通过人机交互方式将结果反馈给运维人员,辅助运维人员快速定位和解决网络故障问题。

利用大模型进行运维日志自动分析主要的价值有以下几点。

a) 提高效率。利用大模型进行自动分析,可以大大提高效率,快速定位问题。

b) 减少错误。人工分析运维日志容易出现疏漏和错误,而大模型可以通过深度学习算法自动识别和分析日志,减少错误率。

c) 实时监控。利用大模型进行运维日志自动分析可以实现实时监控,及时发现问题并进行处理,避免出现严重故障。

d) 提高可靠性。运维日志是系统运行的重要记录,利用大模型进行自动分析可以提高可靠性,保证系统稳定运行。

e) 降低成本。传统的运维日志分析需要大量人力和时间,而利用大模型进行自动分析可以降低成本,提高效率^[11]。

2.5 智慧运营系统助手

传统系统运营需要耗费大量人力和时间成本,定期输出运营周报、月报、场景评估报告和对应的优化方案,复杂场景甚至要对多个系统的多个模块进行数据汇聚和集中监控。基于自主训练的通信行业领域大模型,建立智慧运营系统助手服务于网络运营智能化水平提升。利用生成式AI大模型多模态、跨模态的内容生成能力,对接运营商集约化系统,通过插件将大模型的多维能力与外部工具、资源、知识等优势融合。以交互式助手的形式赋能系统智能监控和运营,为一线运营人员提供时效性更高、交互更便捷、内容更丰富、边际成本接近于零的智慧运营支撑能力。如指标趋势分析,运营图表生成、运营报告生成、运营分析和方案推荐等。在未来,预估40%的工作时间都可得到大语言模型的助力^[12],智慧运营系统助手将深入企业运营,赋能产业变革。

3 结束语

没有最好的模型,只有面向应用场景最合适的模型。在工业界,有时候不仅需要算法的突破,更需要对于一类问题的思路和观念的突破。这些突破可以引导我们更好地去解决一类问题。GPT系列不完全是算法的创新,但一定是产品级的创新。通过组合各类

技术,面向实际的业务应用的切实需求,综合考虑先进性、效率、成本、可控性等等,构建解决方案才是最佳之道。

参考文献:

- [1] ANSELL S M. Harnessing the power of the immune system in non-Hodgkin lymphoma; immunomodulators, checkpoint inhibitors, and beyond[J]. Hematology. American Society of Hematology. Education Program, 2017, 2017(1): 618-621.
- [2] ZHAO W X, ZHOU K, LI J Y, et al. A survey of large language models[DB/OL]. [2023-04-02]. <https://arxiv.org/abs/2303.18223>.
- [3] WEI J, BOSMA M, ZHAO V Y, et al. Finetuned language models are zero-shot learners [DB/OL]. [2023-04-02]. <https://arxiv.org/abs/2109.01652>.
- [4] OUYANG L, WU J, JIANG X, et al. Training language models to follow instructions with human feedback [DB/OL]. [2023-04-02]. <https://arxiv.org/abs/2203.02155>.
- [5] SANH V, WebSON A, RAFFEL C, et al. Multitask prompted training enables zero-shot task generalization [DB/OL]. [2023-04-02]. <https://arxiv.org/abs/2110.08207>.
- [6] WANG R Z, TANG D Y, DUAN N, et al. K-Adapter: infusing knowledge into pre-trained models with adapters[C]//Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Online: Association for Computational Linguistics, 2021: 1405-1418.
- [7] YAMAMICHI K, HAN X H. MCGKT-Net: multi-level context gating knowledge transfer network for single image deraining [C]//15th Asian Conference on Computer Vision. Kyoto, Japan: Springer-Verlag, 2020: 68-83.
- [8] OpenAI. GPT-4 [EB/OL]. [2023-05-30]. <https://openai.com/research/gpt-4>.
- [9] 周洋程, 闫实, 彭木根. 意图驱动的6G无线接入网络[J]. 物联网学报, 2020, 4(1): 72-79.
- [10] 杨巽, 童贞理, 朱周华, 等. 基于意图网络的智能运维优化研究[J]. 消费电子, 2021(9): 73-74.
- [11] 张召路. 一种基于智能分析的通信运营商网络故障管理方法及系统: CN202211299727.5[P]. 2023-03-07.
- [12] 巴比特. 埃森哲技术负责人: 各行业40%的工作时间将受到ChatGPT等大型语言模型的影响[EB/OL]. [2023-05-30]. <https://new.qq.com/train/a/20230411A04XF900?no-redirect=1>.

作者简介:

刘贤松, 毕业于武汉大学, 硕士, 中国联通智网创新中心网络AI产品部总监, 主要负责网络AI平台和网络AI能力研发、赋能工作; 陈凌, 毕业于复旦大学, 硕士, 中国联通智网创新中心网络AI产品部首席算法专家, 主要负责算法技术研究工作; 尹俊, 毕业于中国传媒大学, 硕士, 中国联通智网创新中心网络AI产品部算法工程师, 主要从事面向运营商网络的人工智能算法模型研发工作。