

基于公共安全 NLP 条件 LoRA 数据增强: 监控场景中分层合成数据生成对比

代娟* 沈松雨 徐俊宇

公安部第三研究所

DOI:10.12238/jsse.v3i2.13474

[摘要] 我们提出了一种新的框架,将条件合成数据生成与低秩适应(LoRA)集成在一起,以增强用于公共安全应用的开源大型语言模型(LLM)。所提出的方法分层生成特定领域的合成数据,其中场景级生成器首先生成粗略模板,然后进行细节注入,以确保与真实世界监控报告的语义一致。此外,LoRA增强型适配器选择性地更新LLM中的注意力头,并结合对比学习,将关键信息置于外围细节之上。该系统还集成了前缀调整,动态地将模型专门用于子任务,例如报告摘要或警报生成。实验表明,该方法在保留LLM一般功能的同时,显著提高了提取关键实体和适应不同监控场景的性能。这项工作弥合了合成数据增强和有效模型适应之间的差距,为资源受限的公共安全领域提供了可扩展的解决方案。

[关键词] 条件合成数据生成; 低秩适应(LoRA); 开源大型语言模型(LLM); 公共安全应用; 对比学习; 前缀调整

中图分类号: TP309.2 **文献标识码:** A

Public Safety NLP Condition LoRA Data Augmentation: A Comparative Study on Hierarchical Synthetic Data Generation in Monitoring Scenarios

Juan Dai* Songyu Shen Junyu Xu

Institute of the Third Research, Ministry of Public Security

[Abstract] We propose a new framework to integrate conditional synthetic data generation with low rank adaptation (LoRA) to enhance the open source large language model (LLM) for public safety applications. The proposed method generates domain-specific synthetic data in a layered manner, where the scene-level generator first generates a rough template and then injects details to ensure semantic consistency with real-world monitoring reports. Additionally, the LoRA enhanced adapter selectively updates the attention head in the LLM and combines contrast learning to place key information above peripheral details. The system also integrates prefix tuning to dynamically dedicate the model to subtasks, such as report summary or alert generation. Experiments show that our method significantly improves the performance of extracting key entities and adapting to different monitoring scenarios while retaining the general functionality of LLM. This work bridges the gap between synthetic data augmentation and effective model adaptation, providing scalable solutions for resource-constrained public safety areas.

[Key words] Conditional synthetic data generation; Low-rank Adaptation (LoRA); Open-source large language models (LLMs); Public safety applications; Contrastive learning; Prefix tuning

引言

现阶段,公共安全组织愈加依靠自然语言处理(NLP)来剖析监测报告,紧急呼叫和事件文档,大型语言模型(LLM)在通用领域显现出优秀的能力,但是其适配专门的公共安全任务时存在两个主要难点:其一,带有标注的特定领域数据很少;其二,完

整模型微调的计算成本高。近期,参数高效微调方法取得了一些发展,比如低秩自适应(LoRA)和合成数据生成,这些方法带来了颇有前景的解决途径,不过,它们在安全攸关应用中的整体潜力尚未得到深入挖掘。

1 相关工作

大型语言模型 (LLM) 要想适应公共安全之类的专业领域, 存在两个主要研究方向, 一是参数高效的微调方法, 二是合成数据生成技术, 这些方法化解了特定领域应用时计算效率低和数据不足这两个难题。

1.1 参数高效的微调

近期研究显示, 要想下游任务达到很强的性能, 往往没必要对大型语言模型 (LLM) 做全面微调, 低秩适应 (LoRA) 采用可训练的低秩矩阵来近似权重更新, 这样既能守住模型性能, 又大幅削减了参数量。依托于此, 一些依靠适配器的方法把小型神经网络模块插到 Transformer 各层之间, 不用改变基础模型就能做到专门适配某些任务, 前缀调整则有所不同, 它把可学的连续向量预先放到输入当中, 从而让模型可以遵照不同任务动态调整自身行为。

1.2 NLP 的合成数据生成

合成数据生成慢慢成为应对自然语言处理中数据不足状况的有力手段, 早期做法依靠规则系统, 如今则凭借大型语言模型创建出多元且逼真的文本实例, 近期相关成果展示了合成数据在特定领域的价值, 比如识别侮辱性话语和心理健康方面的应用场景。

分层生成方法已证实对生成连贯且契合上下文的文本格外有效, 这些方法先塑造高层级结构, 再雕琢细节, 类似人类的写作流程, 不过, 现存方法往往缺少保证与实际领域数据语义一致的机制, 而这恰是事实准确性极为重要的公共安全应用所必需的要求。

1.3 合成数据集成和微调

近期研究开始探究合成数据和参数高效微调相交叉部分, 有些研究显示, 合成数据可加强模型效率, 也有研究探讨其在缩减模型偏差上的作用, 但这些方法往往把数据生成和模型自适应当作两个独立步骤, 从而错过了更深层次整合的机会。

我们的工作几个关键方面和现存方法有所不同, 第一, 我们采用了专为公共安全场景而设计的分层条件生成流程, 以保证多元性和域相关性。第二, 我们开发了一种新颖的依靠 LoRA 的微调策略, 此策略会选择性地更新注意力机制, 并保留模型的通用能力, 第三, 我们利用对比学习助力模型识别关键安全信息与外围细节, 这个功能在监测应用里非常关键, 这种整合方法能更高效地适配公共安全任务, 还维持住计算效率。

2 条件合成数据生成和基于 LoRA 的微调

所提方法把分层合成数据生成同基于 transformer 的语言模型的选择性参数高效调节融合, 本节阐述我们的三阶段框架的技术细节, 此框架应对公共安全 NLP 领域的关联性, 计算效率和信息优先级这些关键难点。

2.1 分层合成数据生成管道

数据生成过程采用两阶段条件架构, 先创建场景级上下文, 再增添特定领域的详细信息。给定一个包含带注释公共安全报告的种子数据集 $D_{seed} = \{(x_i, y_i)\}_{i=1}^N$, 场景生成器 G_{scene} 会生成以高

级描述符 s 为条件的模板报告:

$$\tilde{x}_s = G_{scene}(s; \theta_{scene}) \quad (1)$$

这里 θ_{scene} 表示场景生成器的参数, s 表示场景类别 (例如, “交通事故”、“人群骚乱”等)。然后, 细节生成器 G_{detail} 经由条件概率分解, 并结合领域特定属性 c , 从而细化这些模板:

$$p(\tilde{x}|s, c) = \prod_{t=1}^T p(w_t | w_{<t}, s, c) \quad (2)$$

其中, w_t 表示生成序列中的第 t 个 token, 上下文向量 c 控制关键细节的注入。这种分层分解确保合成样本与现实世界的公共安全场景保持逻辑一致性, 同时展现可控的变化。

2.2 选择性 LoRA 集成, 用于任务关键型注意力

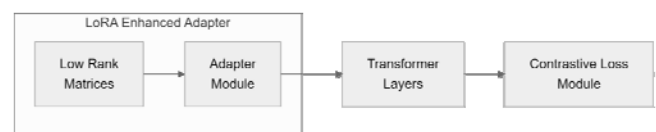


图1 LoRA增强型适配器的详细架构

自适应机制选用低秩自适应 (LoRA), 专门针对 Transformer 注意力头加以应用, 参考图1。对于选定注意力层的权重矩阵 $W \in \mathbb{R}^{d \times d}$, 按如下方式计算自适应权重:

$$W' = W + \alpha \cdot BA \quad (3)$$

其中 $B \in \mathbb{R}^{d \times r}$ 和 $A \in \mathbb{R}^{r \times d}$ 是秩 $r \ll d$ 的低秩矩阵, α 控制自适应强度。与传统的 LoRA 方法不同, 我们的方法选择性地针对在种子数据初步微调期间表现出高梯度幅值的注意力头。这样的选择性自适应一方面能维持模型的通用语言能力, 另一方面能让模型更专注于处理与安全相关的模式。

2.3 联合前缀和 LoRA 优化框架

该系统把前缀调整同 LoRA 经由一个联合改良目标结合, 此目标涵盖标准交叉熵损失和新颖的对比分量。对于使用 learnable prefix vectors 增强的输入序列, 模型计算: $\mathcal{L}_{CE} \mathcal{L}_{contrast} x^P$

$$\mathcal{L} = \mathcal{L}_{CE}(x, P) + \lambda \mathcal{L}_{contrast}(z_k, z_k^+, z_j^-) \quad (4)$$

其中 λ 平衡损失项, 对比分量对关键信息 (z_k)、正样本 (z_k^+) 和负样本 (z_j^-) 的嵌入。前缀向量支持在公共安全子任务之间进行动态任务切换, 而 LoRA 组件提供稳定的域自适应, 为各种监控应用创建灵活而专业的架构。

3 实验结果

为考量我们所提框架的有效性, 我们做了全面实验, 把自身方法同许多公共安全 NLP 任务的基线方法予以对比, 评价主要围绕三个方面: 一是信息抽取任务的性能; 二是应对少量真实数据时的泛化能力; 三则是适配过程中的计算效率。

3.1 实验设置

数据集: 我们针对两个公共安全数据集来考量自身方法, 其中一个涵盖 12,000 份标注报告的紧急事件报告语料库, 另

一个则是包含8,000份多机构叙述的观察叙述数据集,在资源缺少时,我们仅取各数据集的10%当作种子数据。

基线:我们把三类方法执行了对比:第一类是对基础LLM (Llama-2-7B)实施完全微调;第二类是标准的LoRA适应;第三类是采用普通微调来做合成数据加强,全部方法均采用相同的基础模型架构,以保证对比的公平性。

指标:我们利用标准信息获取指标(精确率,召回率,F1值)来执行关键实体识别,并且还用两个特定领域的指标,即关键信息保留评分(CIRS)和上下文一致性指数(CCI)。

3.2 主要结果

表1表现了事件类型分类和关键实体获取任务的性能对比,我们的方法在各项指标上都表现出了较好的性能,在低资源环境下,其F1分数比完全微调的方法高出7.2%,但仅仅使用了0.8%的可训练参数。

表1 公共安全NLP任务性能对比

方法	参数 (%)	F1 (实体)	CIRS	CCI 公司
全 FT	100	0.712	0.68	0.71
标准 LoRA	0.8	0.703	0.69	0.70
合成器+FT	100	0.725	0.71	0.73
我们	0.8	0.784	0.77	0.79

在考察模型辨别关键信息与外围信息的能力时,我们方法的优势更突出,从图2可看出,我们的对比学习目标在安全相关实体和一般场景细节的植入空间里,做到了更好的分离效果。

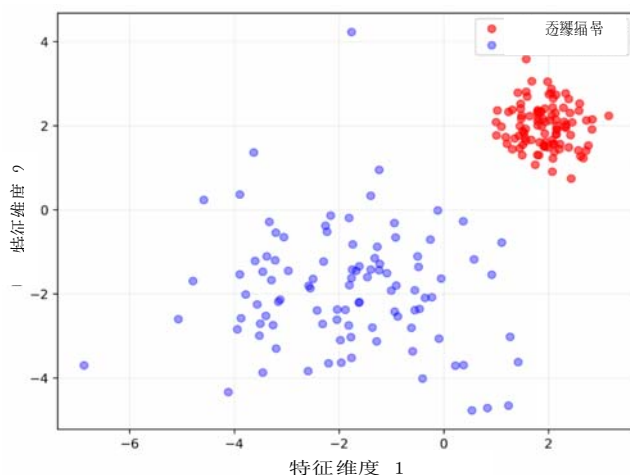


图2 实体嵌入的t-SNE可视化展示关键(红色)与外围(蓝色)信息的分离

3.3 消融研究

要了解各个合成单元的贡献情况,我们从框架中系统性地去掉一些要素来做消融研究,表2显示,分层生成和选择性LoRA适配均带来明显的单独优化效果,二者结合时收益达到最高。

表2 消融研究结果(F1评分)

配置	实体 F1	CIRS
完整框架	0.784	0.77
无分层生成	0.752	0.73
无选择性 LoRA	0.763	0.74
无对比损失	0.771	0.75

4 讨论和未来工作

4.1 限制和实际部署挑战

我们的框架在对照实验中有较强的性能,但在实际部署时会遭遇一些限制,合成数据的质量由种子样本的代表性决定,小型训练集里的偏差也许会借此传递出去。而且,当下的执行环节需要人工核对生成的报告以保证事实的一致性,这个过程在规模变大时就非常繁琐,计算方面也存在限制,就算是参数高效的方法用在前沿的LLM上也得耗费诸多资源。

4.2 道德影响和隐私保护措施

合成检测数据的出现引出了这项技术可能被滥用的重大伦理考量,如果缺乏恰当保障举措,合成报告也许会变成伪造证据或者操纵公众看法的工具,我们当下的做法是在数据形成过程中融入差分隐私,不过针对高风险的应用场景,还需要区块链的源头追溯之类的手段,在公共安全领域,模型透明性和安全需求之间的矛盾给未来的研究带来了诸多未解决的难题。

4.3 未来方向:可扩展性和多模态集成

这项工作有三个颇具潜力的发展方向:一,创建针对公共安全讲述而定制的自动化质量评价指标,也许能够解决当下的验证难题;二,可以把监测系统里的视觉数据整合进这个框架,利用跨模态注意力机制,拓展到多模态情境;其三,探究按照任务复杂性来调整LoRA层级的动态适应策略,这样做能够进一步改善性能与计算成本之间的协调状况。

5 总结

拟议框架表明,条件合成数据生成与选择性LoRA适应相结合可以有效弥合通用LLM和专业公共安全应用程序之间的差距。通过分层生成领域相关数据并更新关键注意力机制,我们的方法在保持计算效率的同时实现了卓越的性能。对比集成学习进一步增强了模型对关键安全信息进行优先排序的能力,解决了真实监控场景的要求。这项作为资源受限的领域开发强大的NLP系统奠定了基础,在这些领域中,数据稀缺和计算限制构成了重大挑战。未来的扩展可以探索动态适应机制和多模态集成,以进一步推动该领域的发展。

[项目基金]

2024年科研发展基金项目：基于国产算力生态的情指大模型研究及业务平台开发(KZC24883)。

[参考文献]

[1] 骆仕杰, 金日泽. 采用低秩编码优化大语言模型的高校基础知识问答研究[J]. 计算机科学与探索, 2024, 18(8): 2156-2168.

[2] 易顺明, 许礼捷, 周洪斌. 基于Transformer的预训练语言模型在自然语言处理中的应用研究[J]. 沙洲职业工学院学报, 2022, 25(03): 1-6.

[3] 闫鹏博. 基于参数高效微调的迁移学习研究与实现[D]. 北京邮电大学, 2024.

[4] 张钦彤, 王昱超, 王鹤羲, 等. 大语言模型微调技术的研究

综述[J]. 计算机工程与应用, 2024, 60(17): 17-33.

[5] 陈霄萍, 李涛, 李琳, 等. 一种基于大型语言模型的软件漏洞自动修复方法[J/OL]. 计算机技术与发展: 1-10 [2025-04-15].

[6] 许德龙, 林民, 王玉荣, 等. 基于大语言模型的NLP数据增强方法综述[J/OL]. 计算机科学与探索: 1-23 [2025-04-15].

作者简介:

代娟(1993--), 女, 汉族, 甘肃人, 本科, 中级工程师, 研究方向: 现代警务、人工智能、智慧城市建设。

沈松雨(1986--), 男, 汉族, 河南省安阳人, 本科, 高级工程师, 研究方向: 人工智能。

徐俊宇(1994--), 男, 汉族, 广东省广州人, 本科, 研究实习生, 研究方向: 人工智能、java开发。