

一种大语言模型支持的用电异常智能检测方法

刘同阳¹, 张强¹, 胡新雨², 徐晓轶², 毛艳芳², 吕晓祥², 李正佳³, 孙大军³
(1. 中国电力科学研究院有限公司, 北京 100192; 2. 国网江苏省电力有限公司 南通供电公司, 江苏 南通 226000; 3. 苏州华天国科电力科技有限公司, 江苏 苏州 215000)

摘要:时间序列分析在工业领域极其重要。精准识别时间序列中的异常,能够有效提高生产效率,降低生产成本。针对电力使用,用户用电量时间序列的异常检测对于降低电力公司的运用成本、提升其服务质量具有重要的作用。随着大语言模型的快速发展,大语言模型被用在越来越多的领域,包括新兴的时间序列预测。因此提出一种基于大语言模型的用户用电时间序列异常检测方法,该方法通过提示工程与大语言模型交流,通过对时间序列进行预处理,使大语言模型能够理解输入数据信息,并正确执行异常检测任务。通过在用户用电时间序列数据集上进行实验分析,结果表明,与现有的异常检测模型相比,所提方法能够有效检测出异常数据。

关键词:用户用电量时间序列;用电异常检测;大语言模型;提示工程

An intelligent detection method for electrical anomalies with large language models

LIU Tongyang¹, ZHANG Qiang¹, HU Xinyu², XU Xiaoyi², MAO Yanfang², LYU Xiaoxiang²,
LI Zhengjia³, SUN Dajun³

(1. China Electric Power Research Institute Co., Ltd., Beijing 100192, China; 2. Nantong Power Supply Company, State Grid Jiangsu Electric Power Co., Ltd., Nantong 226000, China; 3. Suzhou Huatian Guoke Electric Power Technology Co., Ltd., Suzhou 215000, China)

Abstract: Time series analysis is considered extremely important in the industrial field, anomalies in time series can be accurately identified to effectively improve production efficiency and reduce costs, and for electricity usage, operating costs of power companies are reduced and service quality is improved through anomaly detection in users' electricity consumption time series. With the rapid development of large language models, they are being applied to more fields, including time series prediction. A method for anomaly detection of users' electricity consumption time series based on large language models is proposed, communication with large language models is achieved via prompt engineering, the time series is preprocessed so that input data can be understood by the models and the anomaly detection task can be correctly executed, and experimental analysis on real user electricity consumption data shows that abnormal data can be effectively detected by the proposed method compared with existing models.

Key words: consumer electricity consumption time series; anomaly detection of power utility; large language modeling; prompt engineering

0 引言

随着国家电网提出加快培育和发展新质生产力,为电网增“智”^[1],电网发展进入智能化时代。电网运行中的配电损失受到越来越多的重视,配电损失可以分为技术性损失和非技术性损失两大类。其中,最严重的非技术性损失为用户的窃电行为,造成了严重的电网财产损失。因此,大规模配用电数据的有效处理与管理变得越来越重要^[2],特别是从大量用户的历史用电数据中快速有效地识别出异常用电行为与设备^[3-4],能够有效减少配电过程

中的损失,对促进电网的安全使用具有重要的意义。

随着电网智能化程度加深,从海量用户用电数据中准确识别异常行为成为关键挑战。语义学习通过解析电力使用模式中的隐含意义,为该问题提供了新思路:它不仅能捕捉用电行为的细微差异和复杂模式,还能超越简单的电量波动分析,深入探究行为成因。例如,在窃电检测中,传统基于规则或阈值的方法难以应对多样化的异常模式,而基于语义学习的大语言模型可结合历史数据与上下文信息,动态调整正常与异常的判断标准,实现更精准的检测。此外,该方法依赖少量示例即可快速适应不同场景,显著降低对大规模标注数据的需求,提升效率与灵活性。这种基于语义理解的智能检测方式,不仅有助于减少配电损失、提升电网安全

收稿日期:2025-09-22;修回日期:2025-10-28

基金项目:国家电网公司总部科技项目(5700-202318598A-3-2-ZN)

性和经济性,也为电网管理提供了更智能的解决方案。

值得注意的是,尽管大语言模型最初面向自然语言序列设计,但其核心架构本质上适用于任意有序序列建模,包括时间序列数据。在时间序列任务中,大语言模型通常将数值型时间序列转化为类文本格式作为输入;模型则可输出预测值、异常标签,甚至以自然语言形式解释异常原因。这种“数值序列→文本提示→语义推理→结构化/自然语言输出”的范式,使得大语言模型能够利用其强大的上下文理解和泛化能力,对时间序列中的复杂模式进行建模,而无需针对特定任务重新训练模型。

近年来,大语言模型^[5-6]飞速发展,在自然语言任务中展现出卓越能力,如阅读理解^[7]、对话^[8]和翻译^[9]。此外,其应用已拓展至代码生成^[10-11]以及根据自然语言描述生成图像^[12]和视频等任务,显示出广阔前景。越来越多研究者尝试将大语言模型应用于新任务与数据模态,揭示了其跨领域的潜力。相比传统深度学习的时间序列异常检测方法,预训练的大语言模型依托海量公开数据训练,具备丰富的背景知识,能更好地理解复杂模式与上下文,从而提升异常识别准确性。借助提示工程实现“即插即用”,无需为特定任务设计、训练或调优模型,仅通过少量调整现有模板即可快速适配新需求,显著降低开发与部署成本。

本文将大语言模型引入电力时间序列异常检测,针对用户用电量开展研究,这对大语言模型而言是一项挑战。模型采用固定提示模板引导其识别异常时段;输入前对时间序列进行预处理,以增强特征可读性。实验选用3种大语言模型:“gpt-3.5-turbo”、“llama3-8b-instruct”和“qwen2-7b-instruct”,在真实用户用电数据集上通过随机注入异常片段(模拟用电激增或骤减)构建测试样本。结果表明,本文方法优于传统异常检测算法,虽在精度上略逊于专用深度学习模型,但凭借无需训练的优势,在时间序列异常检测中展现出良好的应用价值。

1 时间序列数据预处理

时间序列分为“一元时间序列”和“多元时间序列”,在本文中,主要研究居民用电时间序列,为一元时间序列。本文将居民用电时间序列定义为 $X=(x_1, x_2, \dots, x_n)$,其中 x_n 为时间步长为 n 处时用电量数据。为了使大语言模型能更好地处理时间序列数据,本文将时间序列 X 转换为标记化的序列,本文将在下面对标记化过程进行详细介绍。

缩放:时间序列数据包括不同大小的数值,可能包括正值和负值,覆盖较大的数值范围。为了加快运行效率,同时减少过大或过小的数值对大语言模型的影响,本文采取缩放的方法,从时间序列中减去最小值 $x_{s_i} = x_i - \min(x_1, x_2, \dots, x_n)$,得到新的时间序列 $X_s=(x_{s_1}, x_{s_2}, \dots, x_{s_n})$ 。缩放后的时间序列消除了负数,不需要大语言模型对复试的单独处理。即引入了一个映射函数 $\varepsilon:R \rightarrow R_{\geq 0}$ 。

其他的缩放方法,如减去最大值的缩放,或者将数据规则化到 $[0, 1]$ 的范围内,都可以实现相同的目标,但过小的规则化范围,可能会导致特征信息的丢失。同时,增加范围将导致更多的数字来标记。使用减去最小值的方法,只需移动信号值的范围,因此需要减少标记的单个数字的数量,并保持序列数据之间的原始差距。此外,通过将值投射到非负范围,我们消除了符号指示符“-/+”的需要,并节省了一个额外的标记。

量化:与用于大语言模型训练的有限词汇集不同,时间序列数据存在小数、负数等不同大小数据,即使经过缩放后的 X_s ,其词汇集依旧是无限的,大语言模型难以直接处理。因此需要对时序数据进行量化处理。本文采用四舍五入的方法,将每个值四舍五入到预定的小数位数,然后缩放为整数格式,以避免在小数点上浪费记号。因此,时间序列变为 $X_q=(x_{q_1}, x_{q_2}, \dots, x_{q_n})$,量化示例如下:

“0.357, 0.003 1, 0.274 2, 0.514” $\rightarrow$ “357, 3, 274, 514”

如上文所述,本文使用了两个映射函数,缩放函数 $\varepsilon:R \rightarrow R_{\geq 0}$ 和量化函数 $Q:R_{\geq 0} \rightarrow Z_{\geq 0}$ 。

滑动窗口:由于大语言模型对输入上下文长度有限制,例如GPT-3.5-turbo的上限为16 k,而时间序列数据长度远大于窗口限制,因此,本文采用滑动窗口的方式,来顺序输入时间序列数据。该方法将每个时间序列分割成具有预定长度和步长特征的滑动窗口;也就是将量化后的时间序列 X_q 分割成集合 $\{\{x_{1 \dots w}^i\}\}_{i=1}^K$ 。

标记化:不同的标记化处理方案在数据处理上有所不同,几种不同的开源大语言模型,如文献[13]和文献[14],使用空格断句对数据进行标记化处理,将数字通过添加空格分割成单个数字。然而,不同的大语言模型对该方式标记的接收程度不同,例如, llama3-8b-instruct无法准确的处理通过空格断句后的数据,因此,在实验过程中,本文对不同的大语言模型采用不同的标记化方案。空格断句对数据进行标记化处理示例:

“357, 3, 274, 514” $\rightarrow$ “3 5 7, 3, 2 7 4, 5 1 4”

2 基于提示工程的时间序列异常检测

基于提示工程,时间序列异常检测将给出一组单变量的时间序列 $X=(x_1,x_2,\cdots,x_N)$,假设其存在一到两组长度未定的异常片段 $A=\{(t_s,t_e)|1\leq t_s\leq t_e\leq N\}_{i=1}^m$,其中 t_s,t_e 分别为区间的起始点和终点。异常检测的目标是找到一组 m 个异常片段,使该组片段尽可能覆盖异常片段。下面首先给出大语言模型提示工程的概念,然后给出基于提示工程的时间序列异常检测方法。

2.1 提示工程

提示工程(也称上下文提示)通过在不更新大语言模型权重的前提下,利用输入文本引导模型解决特定问题。提示是提供给模型的问题描述,可以是一段文字、关键词或其他形式的文本,用于激发模型产生预期响应。有效的提示设计能显著提升模型在不同任务上的表现。

提示可分为硬提示和软提示。硬提示是人工设计的文本模板,直接描述任务或数据,具有良好的可读性和通用性。虽然设计灵活,但要达到最佳效果,通常需反复优化提示内容。相比之下,软提示是在训练过程中自动学习的一组连续向量,无法直接查看或编辑;它通过梯度更新针对特定任务进行优化,本质上是从大模型中提取知识的一种方

式。软提示通常由一个小型可训练模型生成,并作为虚拟令牌插入输入文本前,再送入大语言模型。由于该过程不可见,软提示缺乏可解释性。本文采用硬提示方式进行模板设计,以确保方法的透明性与可解释性。

2.2 提示模板设计

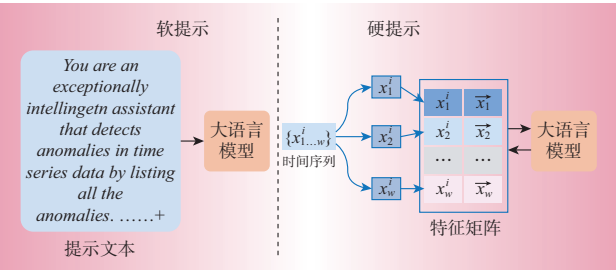


图1 软提示与硬提示对比

Fig. 1 Comparison of soft prompts and hard prompts

如图1所示,基于提示工程的大语言模型时间序列异常检测的输入是提示模板加时间序列。由于输入长度限制,因此,本文采用滑动窗口方式拆分时间序列数据集。输入可表示为 $u_{1\cdots k}^i = \text{prompt} + \{x_{1\cdots w}^i\}$,其中 k 为每轮输入数据的总长度,大语言模型将输出下一个结果文本为 u_{k+1} 。

在下面的实验中,选用3种不同的大语言模型(GPT-3.5-turbo、Llama3-8B-Instruct 和 Qwen2-7B-Instruct),采用不同的提示文本模板。表1所示为

表1 GPT-3.5-turbo模型提示模板示例及反馈结果
Table 1 GPT-3.5-turbo model prompt template example and feedback results

实验	提示模板	输出信息的内容
1	$\{x_{1\cdots w}^i\}$, Help me find the anomaly in the above time series.	(1) 提出可以解决该方法的方法,可以使用统计分析、机器学习算法或可视化方法等技术。 (2) 提供的数据量过大,手动分析异常不可行,建议使用Python等编程语言以及Pandas、NumPy 和 scikit-learn 等库来自动化异常检测过程。 (3) 建议提供结构化格式的数据,提供数据异常的模板,能够帮助完成识别数据子集中异常的过程。
2	Find the range of indices that are anomalous in this series, $\{x_{1\cdots w}^i\}$.	(1) 要识别给定时间序列数据中的异常情况,可以计算每个数据点Z-score值。Z-score值衡量的是一个数据点与数据集平均值的标准差数量。通常情况下,Z-score 大于3或小于-3都会被视为异常。 (2) 给出Z-score具体的计算过程。
3	The anomaly indices in <i>timeseries_1</i> = $\{x_{1\cdots w}^1\}$. is: $\{t_{k\cdots w}^1\}$. The anomaly indices in <i>timeseries_2</i> = $\{x_{1\cdots w}^2\}$. is: $\{t_{k\cdots w}^2\}$. The anomaly indices in <i>timeseries_3</i> = $\{x_{1\cdots w}^3\}$. is: Help me find the anomaly.	(1) 生成错误的索引列表。 (2) 没有发现异常。
4	You are an exceptionally intelligent assistant that detects anomalies in time series data by listing all the anomalies. Below is a sequence, please return the anomalies in that sequence. Do not say anything like 'the anomalous indices in the sequence are', just return the numbers. $\{x_{1\cdots w}^i\}$	输出一个包含异常的起始点和终点的索引数组。

GPT-3.5-turbo模型提示模板的优化过程。初始提示模板仅向GPT-3.5-turbo提供了任务描述和所需处理的数据序列。然而,尽管输入了基本的任务需求,该模型未能准确理解任务意图。鉴于初次实验的结果,后续调整了提示模板的设计中引入了具体的异常案例作为示例,通过实例教学帮助模型更好地理解任务要求。这一策略显著提升了模型的输出质量,但模型有时仍会给出错误的结果。在最终模型中,逐步丰富了提示模板中的任务描述信息,明确了模型的角色定位为“智能异常检测专家”,并严格限定了其功能范围为识别数字序列中的离群点。核心指令明确要求助手直接输出异常值对应的索引位置,同时禁止添加任何描述性前缀,确保输出结果可以直接被机器解析。这种设计实现了极简化的输入输出规范:用户只需提供原始序列,系统即返回纯数字形式的异常索引列表。通过这种方式,不仅提高了输出的准确性,也增强了系统的可靠性和实用性。最终形成的3种模型的提示模板示例如下所示:

以GPT-3.5-turbo为例,所采用的输入提示为:
“You are an exceptionally intelligent assistant that detects anomalies in time series data by listing all the anomalies. Below is a sequence, please return the anomalies in that sequence. Do not say anything like ‘the anomalous indices in the sequence are’, just return the numbers.”

3种模型的提示模板均首先明确了模型的角色定位,而后描述任务及输出格式要求,以便后续处理,区别在于指令风格和约束强度:GPT-3.5-turbo擅长语义理解故需较少格式限制,因此采用对话式自然语言指令,侧重任务描述而允许较自由的输出格式;Llama3-8B-Instruct需要显式顺序约束以弥补序列处理弱点,明确要求禁止注释并保留原始顺序;Qwen2-7B-Instruct的数学倾向使其最适合规则化提示,因此采用函数式规则列表,通过编号指令强制大模型理解任务,并按照严格要求的输出格式进行输出。

经过滑动窗口的方式提示,大语言模型将输出数个异常段,对于每个大语言模型认为的异常段,将其转换为时间序列对应的索引,并将所有索引合并到一起,作为最终的大语言模型生成的异常段。

3 实验结果与分析

3.1 数据集及评价指标

本文采用真实的用户用电数据集ECL,该数据

集包含从2012年到2014年收集的321个客户每小时电力消耗数据。采样频率为h,其中,321个客户的平均用电量范围在10 kWh到20 053 kWh,最大用电量范围在85 kWh到76 400 kWh,从中可以看出,不同用户之间的用电范围跨度较大,覆盖的用电情况较为广泛,基本覆盖所有用电情况的用户。为了模拟用电异常,本文对数据集进行预处理,随机构建一到两段连续的异常用电数据,每段异常长度大于1 000个步长小于5 000个步长,每段异常的构建过程为:对用电数据乘以1.5~2.0的倍数系数,模拟用户用电异常增长的情况,或者乘以0.1~0.5的系数,模拟用户的用电量异常降低情况,其中系数为随机生成。为了对模型有效性进行评价,本文选择了3个评价指标对模型进行有效性判断,分别为:精确率、召回率和F1。

$$P_{\text{recision}} = TP / (TP + FP) \quad (1)$$

$$R_{\text{ecall}} = TP / (TP + FN) \quad (2)$$

$$F1 = 2(P_{\text{recision}} R_{\text{ecall}}) / (P_{\text{recision}} + R_{\text{ecall}}) \quad (3)$$

式中:TP为异常检测正确的数值;FP为异常检测错误的数值,FN为没有正确检测到的异常数值。可以看出, P_{recision} 为检测的准确率; R_{ecall} 为异常检测结果正确的数值占全部检测数值的比例;而F1为查全率,即正确预测为正的占全部实际为正的的比例。

3.2 对比模型

在对比实验中,本文采用3种类型不同的模型进行实验对比,分别为Z-Score、LSTM和Anomaly Transformer,下面将对3种模型进行介绍。

Z-Score:该方法属于基于统计的异常检测方法,利用滑动窗口机制,通过计算平均值和方差,与之后的节点进行对比,如果后一节点值不在这个范围内,则判断该节点为异常节点,即计算离群点来识别异常。

LSTM:LSTM是一种循环神经网络的变体,通过引入门控循环单元来捕获序列数据中的长期依赖关系。使用LSTM进行异常检测时,通过预测下一节点数值,若下一节点值与预测值偏离较大,则判断该节点为异常节点。在本实验中,采用两层堆叠LSTM结构,每层包含128个隐藏单元,后接全连接输出层。模型使用Adam优化器,训练周期为100轮,早停策略防止过拟合。训练集包含前80%的时序数据,剩余20%作为测试集。

Anomaly Transformer:该模型将标准Transformer中的自注意力机制改进为异常注意力机制,通过关联差异量化时序点异常程度。具体实现中,采用4层编码器,搭配提出的极大极小策略来放大正常与

异常点的特征差异。同样采用训练集包含前80%的时序数据,剩余20%作为测试集。

3.3 实验结果及分析

所有的实验均在一台 i7-9700 CPU, 32G RAM 以及 RTX2080Ti GPU 的个人电脑上进行,值得注意的是,本文模型使用 API 调用大语言模型,能够做到即插即用,开发和使用成本较低。例如 GPT-3.5-turbo 的 API 调用费用为 0.11 元/(10⁴·Tokens)。

实验结果如表 2 所示,可以看到在所有的模型中, Anomaly Transformer 在 3 个指标上的结果都取得了最优解, 基于深度学习的模型在经过足够的训练数据进行预训练后, 能够取得较为优异的异常检测结果, 其中, 以 $P_{precision}$ 为对比指标时, 结果优于基于提示工程的 GPT-3.5-turbo 模型 22.8%, 同时, 以 R_{recall} 对比指标时, 优于基于提示工程的 llama3-8b 模型 30.1%, 以 $F1$ 为对比指标时, 同样优于 GPT-3.5-turbo 模型 29.1%。然而深度学习预训练的时间成本和训练数据的标注、模型的调整开销都是不能忽视的, 虽然本文的基于提示工程的模型虽然在实验效果上弱于 Anomaly Transformer 模型, 但提示工程使用预训练的大模型进行异常检测, 无需训练。

表 2 基于提示工程的用电异常检测模型与其他模型实验结果对比

Table 2 Experimental results of prompt engineering-based power usage anomaly detection model with other models

模型	准确率	召回率	$F1$
Z-Score	0.153	0.097	0.119
LSTM	0.128	0.150	0.138
Anomaly Transformer	0.264	0.307	0.284
Prompt-GPT-3.5-turbo	0.215	0.227	0.220
Prompt-llama3-8b	0.142	0.275	0.187
Prompt-qwen2-7b	0.134	0.236	0.171

在本文提出的 3 个基于提示工程的异常检测模型中, 基于 GPT-3.5-turbo 的模型取得了较好的结果, 在 $P_{precision}$ 和 $F1$ 中取得了最优的效果, 而 llama3-8b 在 R_{recall} 上取得了最好的结果。出现这种情况的原因是, llama3-8b 在输出异常的结果时更为保守, 相较于另外两种模型输出的异常段更短, 同时能够使输出的结果中正确判断的异常点占有所有输出异常点的比例更高, 因此取得了最好的 R_{recall} 值。另外两种模型中, GPT-3.5-turbo 的性能最为优异, 能够比较准确地查找出异常, 取得了相较于另外 5 种模型中第二好的 $P_{precision}$ 和 $F1$ 值。

LSTM 在所有模型中取得了最低的 $P_{precision}$ 值, 而 R_{recall} 和 $F1$ 结果优于 Z-Score, 这是因为 Z-Score 是基于前一个滑动窗口地值判断下一节点是否为异常值, 每次只判断一个值, 因此取得了较好的 $P_{precision}$ 值, 然而其受限于对连续异常点的判断较弱, 难以判断第一个异常点之后的异常点, 因此其 R_{recall} 和 $F1$ 结果较差。

由实验结果可知, 虽然基于提示工程的时间序列异常检测效果稍弱于特殊设计的深度学习模型, 但其效果好于基于统计方法和经典机器学习方法, 具有较好的可靠性。基于提示工程的大语言模型时间序列异常检测, 能够避免传统深度学习算法复杂的训练过程, 且同时可减少训练成本和系统开发的复杂性。

4 结束语

针对用户用电异常检测, 本文提出一种基于大语言模型的提示工程时间序列异常检测方法, 并在真实用户用电数据集上进行了实验验证。与 3 种不同模型的对比结果表明, 尽管本方法在性能上略逊于专门设计的深度学习模型, 但仍取得显著效果, 优于统计方法和经典机器学习方法。因此, 将大语言模型引入用户用电时间序列异常检测具有广阔前景, 可为电网安全运行提供更智能、高效的解决方案, 有助于及时发现并处理潜在安全隐患, 提升电力系统的稳定性与可靠性。

需要指出的是, 本文采用固定提示模板。未来将探索更多样化的提示工程策略, 如逐步提示、多轮对话提示等, 以更充分挖掘大语言模型在复杂用电异常检测中的理解与处理能力, 并进一步在多样化实际场景中优化模型应用。D

参考文献:

[1] 周蕴文. 智能电网环境下电力调度安全措施研究[J]. 电气技术与经济, 2024, (8): 248-250.
ZHOU Yunwen. Research on safety measures for power dispatching in smart grid environment [J]. Electrical Equipment and Economy, 2024, (8): 248-250.
[2] 任禹同, 曹晓冬, 李世洁, 等. 基于电力数据的重点人群在室用电行为分析方法研究及应用[J]. 电力需求侧管理, 2022, 24(6): 112-118.

- REN Yutong, CAO Xiaodong, LI Shihao, et al. Analysis and application of indoor electricity behavior of key population based on electricity data [J]. *Power Demand Side Management*, 2022, 24(6):112–118.
- [3] 陈慧, 陈适, 郭银婷, 等. 基于正则自编码器及 Optuna 寻优的异常用电数据清洗研究[J]. *电力需求侧管理*, 2023, 25(5):53–58.
- CHEN Hui, CHEN Shi, GUO Yinting, et al. Abnormal power consumption data cleaning based on regular self-encoding and Optuna optimization [J]. *Power Demand Side Management*, 2023, 25(5):53–58.
- [4] 郑世英, 牛清林, 刘伟, 等. 电力用户异常用电的深度神经网络检测方法[J]. *电力需求侧管理*, 2023, 25(6):82–87.
- ZHENG Shiyong, NIU Qinglin, LIU Wei, et al. Deep neural network detection method for abnormal electricity consumption by power users [J]. *Power Demand Side Management*, 2023, 25(6):82–87.
- [5] HUANG H, ZHENG O, WANG D, et al. ChatGPT for shaping the future of dentistry: the potential of multimodal large language model [J]. *International Journal of Oral Science*, 2023, 15(1):29–49.
- [6] 徐浩然, 向阳, 丁玲. 基于大语言模型知识增强和多特征融合的中文命名实体识别方法[J/OL]. *北京航空航天大学学报*, 1–12 [2025–11–27]. <https://doi.org/10.13700/j.bh.1001-5965.2024.0421>
- XU Haoran, XIANG Yang, DING Ling. Chinese named entity recognition method based on knowledge enhancement from large language models and multi-feature fusion[J/OL]. *Journal of Beijing University of Aeronautics and Astronautics*, 1–12 [2025–11–27]. <https://doi.org/10.13700/j.bh.1001-5965.2024.0421>
- [7] CUI C, MA Y, CAO X, et al. A survey on multimodal large language models for autonomous driving[C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2024:958–979.
- [8] HOSSEINI-ASL E, MCCANN B, WU C S, et al. A simple language model for task-oriented dialogue[C]//*Advances in Neural Information Processing Systems*, 2020, 33:20 179–20 191.
- [9] 杨滨瑕, 罗旭东, 孙凯丽. 基于预训练语言模型的机器翻译最新进展[J]. *计算机科学*, 2024, 51(S1):50–57.
- YANG Binxia, LUO Xudong, SUN Kaili. Recent progress on machine translation based on pre-trained language models[J]. *Computer Science*, 2019, 51(S1):50–57.
- [10] XU F F, ALON U, NEUBIG G, et al. A systematic evaluation of large language models of code[C]//*Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming*. 2022:1–10.
- [11] JAIN N, VAIDYANATH S, IYER A, et al. Jigsaw: Large language models meet program synthesis [C]//*Proceedings of the 44th International Conference on Software Engineering*. 2022:1 219–1 231.
- [12] KOH J Y, FRIED D, SALAKHUTDINOV R R. Generating images with multimodal language models [C]// *Advances in Neural Information Processing Systems*, 2024: 36–56.
- [13] SANDMANN S, RIEPENHAUSEN S, PLAGWITZ L, et al. Systematic analysis of ChatGPT, Google search and Llama 2 for clinical decision support tasks [J]. *Nature Communications*, 2024, 15(1):2 050–2 058.
- [14] SHINO M. Transmistrat at semeval-2024 task 10: Using mistral 7b for emotion discovery and reasoning its flip in conversation [C]//*Proceedings of the 18th International Workshop on Semantic Evaluation*. 2024:298–304.
-
- 作者简介：**
- 刘同阳(1995), 男, 助理工程师, 硕士, 研究方向为电力认知智能、知识图谱、生成式大语言模型;
- 张强(1983), 男, 高级工程师, 硕士, 从事电力领域自然语言处理、知识图谱及认知计算工作;
- 胡新雨(1986), 男, 高级工程师, 研究方向为人工智能、输变配电运检、高压试验、带电检测等。
- (责任编辑 安然)