

**研讨课调研报告**

|  |  |
| --- | --- |
| 课程名称 | 自动规划 |
| 调研方向 | 最近一年某相关论文的复现 |

|  |  |
| --- | --- |
| 组 别： | 10 |
| 学号姓名： | 58122231陆文韬 |
| 学号姓名： | 58122227焦子豪 |
| 学号姓名： | 58122225曹映波 |
| 指导教师： | 张志政 |
| 调研日期： | 2024-2025秋季学期 |

1. 调研方向说明

我们选取了一篇名为ChatEV: Predicting electric vehicle charging demand as natural language processing的文章进行了关于自动规划的调研，并进行了对论文的复现，并进行了一些优化。这篇论文主要实现了对某充电站在某一时间段下的充电占用率的预测，为电动汽车寻找空闲充电桩，规划未来充电行为提供了可靠的支持。

1. 国内外研究现状

近年来，大语言模型取得了蓬勃发展，例如 BERT（Kenton 和 Toutanova，2019）、Sentence-T5（Ni 等人，2022）和 ChatGPT，它们在大规模混合语料库上扩展了参数数量至百万级别，并在自然语言理解和推理方面展现了突破性的能力。因此，开创性的研究（Xue 和 Salim，2023；Gruver 等人，2024；Lai 等人，2024；Li 等人，2024b）尝试利用提示新范式探索大语言模型在时间序列预测中的潜力，将时间序列编码为自然语言句子，并生成语义标记作为模型预测。例如，PromptCast（Xue 和 Salim，2023）实现了对天气温度、能源消耗和客户流量的大语言模型赋能预测，无需对大语言模型架构进行任何修改。基于大语言模型的时间序列基础模型的最新进展，如 Time-LLM（Jin 等人，2024a）和 UrbanGPT（Li 等人，2024a），展示了它们令人惊讶的性能，通过重新编程为语言任务。与传统的预测模型相比，这些模型通常需要复杂的参数搜索和从头开始的训练，新的范式（即大语言模型赋能的时间序列预测）提供了一种更简单且更易访问的选择，并且在时间序列预测领域展示了其令人印象深刻的泛化能力。

1. 主要难题

现在高质量电动汽车充电数据稀缺，与加油站相比，充电站的数量仍然相对较少，导致数据点不足，预测模型训练不充分。

充电行为受基础设施分布、定价方案和天气条件等多种因素的影响，而现有的方法难以理解异构数据之间的关系，如时间序列、地理信息和社会经济因素，将它们纳入并进行分析对数据驱动的预测模型构成了挑战。

另外，充电站的使用模式在不同时间和不同地点之间存在差异。因此，需要一个统一且可扩展的预测器，能够有效地处理需求预测中的常见模式和特定上下文，并泛化到未见过的地区，现在还缺少一个能够在不同区域实现稳定需求预测的一般化基础模型。

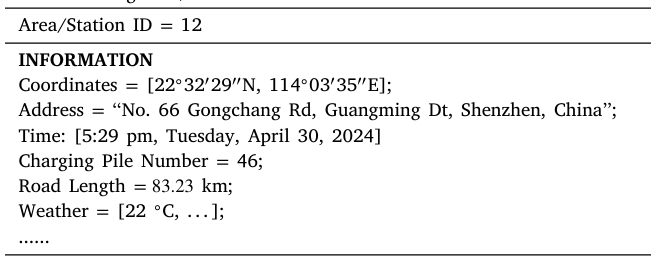
不过将现有的大语言模型用于一般时间序列预测并不能解决问题，需要开发一个专门针对电动汽车充电需求预测的模型，以满足现实世界场景中多样化的需求。而且，尽管电动汽车充电需求和其他形式的人类活动和驾驶在城市交通系统中的需求预测存在相似之处，但它们受到不同的因素影响，如充电价格等，程度也各不相同。这些细微的差异使得电动汽车充电需求预测不同于一般的时间序列预测。而精心设计的提示指令对于充分利用大语言模型在优化电动汽车充电需求预测方面的理解和推理能力至关重要，但这一问题需要进行仔细研究。

1. 常见技术方案

ChatEV的关键步骤有两步：第一、基于提示的任务重述：将预测器置于一个完整的语言环境中，在此环境中，通过使用个性化的提示将电动汽车充电需求预测任务重新表述为文本到文本的格式。通过这种方式，ChatEV 可以利用大语言模型预训练语料库中嵌入的丰富开放世界语义知识，并在相同的语言语义空间内对齐异构数据，从而实现各个领域的准确预测；第二、多领域对齐调优：采用一种高效的模型无偏元学习方法——Reptile对大语言模型主干进行微调，从而结合了与电动汽车充电需求预测相关的领域知识。在这个过程中，使用了语言建模目标，即基于先前的词生成下一个词，而不是使用特定任务的损失函数。从而实现语言模型对电动汽车充电需求预测的有效和高效的适应。

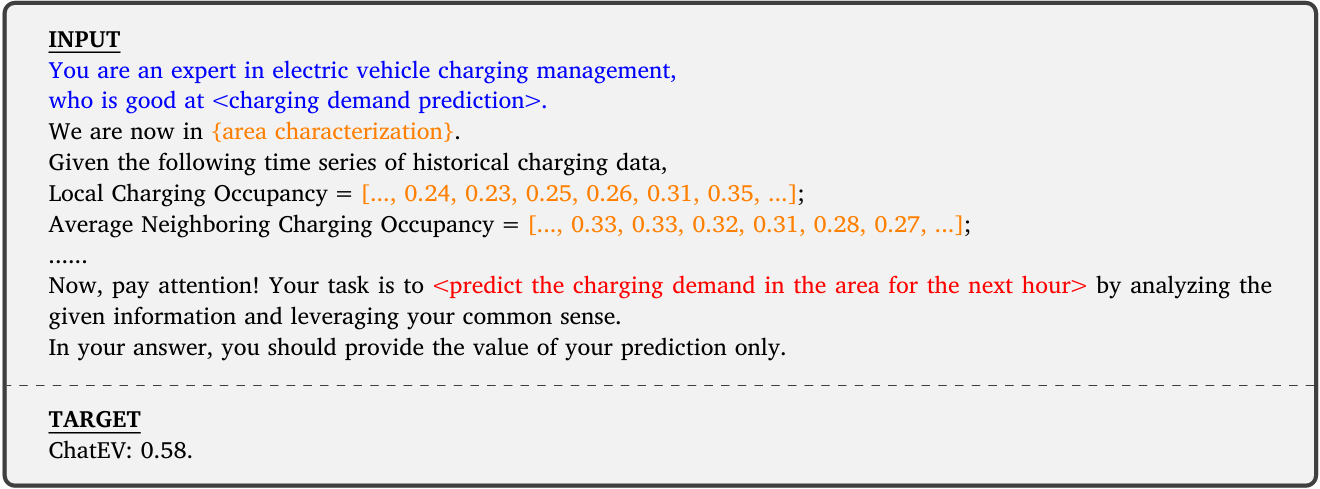
其中，基于提示的任务重述将预测任务转化为自然语言处理，故创建了一个个性化的提示模板，该模板涵盖了基本信息和额外的指令。一般来说，提示包括两个模板，用于模型输入和目标，以及一组相关的元数据。

首先是地理特征，为了向大语言模型主干提供特定于某一充电区域或站点的个性化知识，本文中设计的提示将包括对充电相关特征的详细自然语言描述，比如地理信息（坐标和地址）、社会经济特征（道路密度和 POI 密度）、天气条件（温度和湿度），以及其他更多内容。此外，还需从邻近区域计算出充电相关的变量（如需求和价格）的平均值，并将其整合到区域描述中，以增强空间意识。通过这种方式，ChatEV 整合了邻近节点的特征，将空间数据转换为文本元素，从而消除了在当前由大语言模型驱动的时间序列预测模型中对时间序列片段进行额外训练的需要。总而言之，这种表征有助于大语言模型主干高效地识别研究的充电区域/站点，从而促进个性化适应。

然后是信息检索，为了从不同来源获取最新的元数据以进行区域特征描述，本文采用了检索增强生成（RAG），通过从有用数据的语料库中提取信息来丰富 大语言模型的知识，意味着创建一个与电动汽车充电相关的元数据知识库，根据充电区域/站 ID 可以轻松检索到相应信息并插入到提示模板中。所有检索到的元数据都将被转换成大语言模型兼容的文本标记，如下表所示。

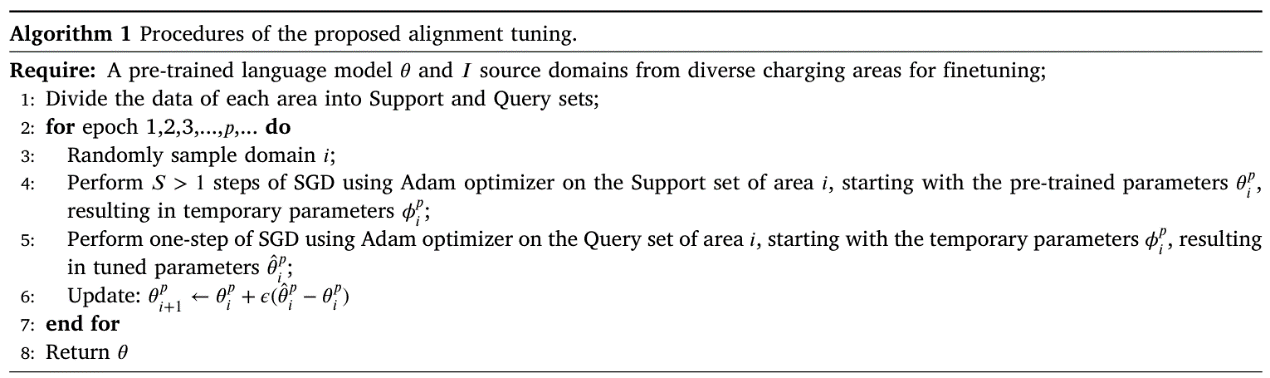
最后是零样本增强指令，为了提升 ChatEV 的推理能力，本文呢在提示设计中应用了一种简单而有效的零样本扩增指令，即角色扮演。角色扮演要求大语言模型假设一个特定的角色，AI 利用这一角色更高效地完成分配的任务。例如，指令可以呈现为“一个擅长 <特定任务> 的 <特定领域> 专家”。借鉴心理学原理，角色扮演可以帮助更有效地编程大语言模型，从而利用“心态转变”的力量，预期能够改进基于大语言模型的预测器的输出。

最终形成的提示模板如下。其中蓝色文本是零样本增强指令，橙色文本表示通过区域 ID 检索到的充电相关元数据，红色文本是任务描述。对于传统的统计或深度学习模型，很难对这些异构特征进行对齐并建模它们之间的关系。相反，ChatEV 通过自然语言将它们统一起来，并利用大语言模型的开放世界知识来实现有效的特征理解。总之，最终的提示包括指导大语言模型主干、列出充电相关因素，并以序列到序列的格式呈现某些任务。



微调是一种非常有效的技术，可以将预训练的语言模型专门化以执行下游任务。而在现实场景中，电动汽车的充电模式通常会随时间和空间而变化。直接使用来自不同充电区域的各种样本进行微调可能会导致大语言模型主干网络学习错误的方向，从而在其他数据集上表现不佳。为了改进学习过程，本文应用了元学习方法Reptile进行知识适应。它的工作原理是反复抽样每个任务，并以标准方式对每个任务执行随机梯度下降（SGD）以实现无偏且高效的调整。

假设存在 𝐼 个电动汽车充电站/区域作为源领域，首先将它们的数据分为支持集和查询集。在支持集上进行 𝑆 步随机梯度下降（SGD）。这个过程能够获得一组学习参数。随后，基于支持集中学到的参数，在查询集上进行一步 SGD，以确定一个成熟的优化方向。这一步允许细化优化过程。具体来说，这个调优过程可以用下面这个算法描述。



该方法的创新性可以概括为：(1) 具有知识性。ChatEV 可以结合大语言模型中呈现的开放世界知识，以支持电动汽车充电需求预测。(2) 兼容性。在统一的语言语义空间中，ChatEV 可以轻松融合异构数据，并充分利用来自不同来源的各种信息。(3) 具有泛化能力。ChatEV 利用大语言模型强大的推理能力，通过提示工程和对齐调优，在多种未见过的领域实现准确预测。

1. 常见实验设计和测试指标

（在调研基础上，简要介绍常见测试方法、测试数据、测试指标等方面）

本文在一个名叫ST-EVCDP的数据集上进行了训练，这个数据集包含了深圳市 18,061 个公共电动汽车充电站 247 个交通区域的实时充电占用信息，覆盖了从 2022 年 6 月 19 日到 7 月 18 日的 30 天时间。使用充电占用作为特定区域充电需求的指标。此外，还收集了研究区域的各种重要特征，包括充电定价方案、坐标、区域相邻关系、道路密度、兴趣点以及充电类型（即快速充电或慢速充电）。根据以上信息，只需要提供对应的充电站以及时间段，ChatEV就能提供对应充电站接下来一段时间的占用率。

基线分别使用了多种具有代表性的统计方法、神经网络方法以及最近开发的基于大语言模型的方法。具体来说，统计方法包括 ARIMA，这是一种利用滞后移动平均来预测未来值的统计分析模型；以及 Lasso，这是一种结合了变量选择和正则化技术的回归分析方法，能够在提高预测准确性的同时保持模型的可解释性。比较的神经网络包括：FCNN（全连接神经网络）；LSTM（长短期记忆网络），用于电动汽车充电需求估计；GCN-LSTM，一种时空交通流预测的代表性方法；STGCN，一种基于卷积的典型时空预测模型；HSTGCN，一种最近开发的方法，将图卷积层和门控递归单元结合，分别学习电动汽车充电需求的时空模式；PIAST，一种考虑充电价格潜在影响的电动汽车充电需求预测的先进模型。基于大语言模型的基线是 PromptCast，它使用预训练的语言模型进行零样本设置下的提示驱动时间序列预测；LLMTIME，它对时间序列数据进行分词，并将离散分布转换为高度可适应的数字。值得注意的是，为了验证将基于大语言模型的方法适应下游数据集的重要性，在后续实验中允许 LLMTIME 在训练语料上进行微调，而 PromptCast 则不进行微调。

时间序列预测中常用的两个评价指标是均方根误差（RMSE）和平均绝对误差（MAE）。

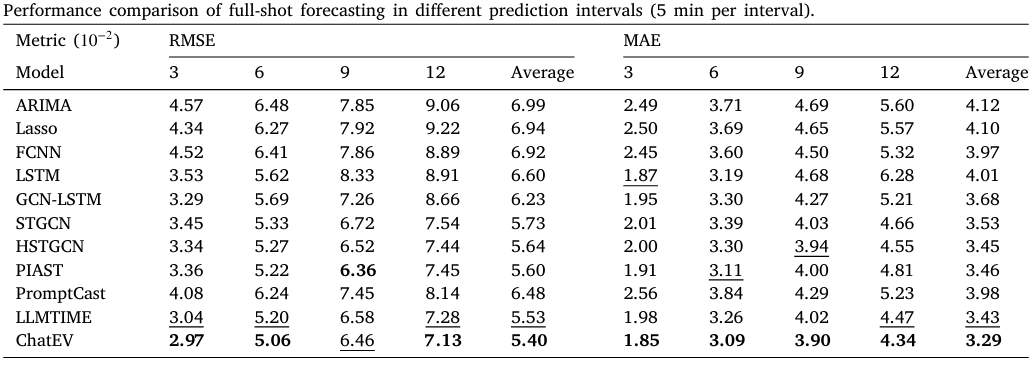
1. 分析：存在问题和感受

虽然大语言模型已经得到了很大的发展，但还没有研究探索大语言模型在电动汽车充电优化中的应用，目前仍处于很大程度上的未开发状态。首先，当考虑到邻近区域的各种异质因素时，设计个性化的提示以重新表述这一复杂任务极具挑战性。其次，尽管大语言模型拥有大量的开放世界知识，但其训练语料库可能缺乏关于电动汽车充电需求预测的信息。最后，尽管已经设计了一些适用于时间序列的通用模型，但专门针对电动汽车充电需求的模型仍然是必不可少的，以探索具有较小模型规模和在特定场景下提供个性化功能的替代方案。

为解决这些不足，本文引入了一种微调的语言模型作为预测器，并针对电动汽车充电需求预测进行了定制提示。迄今为止，ChatEV 代表了从语言视角进行可泛化的电动汽车充电需求预测的开创性努力。

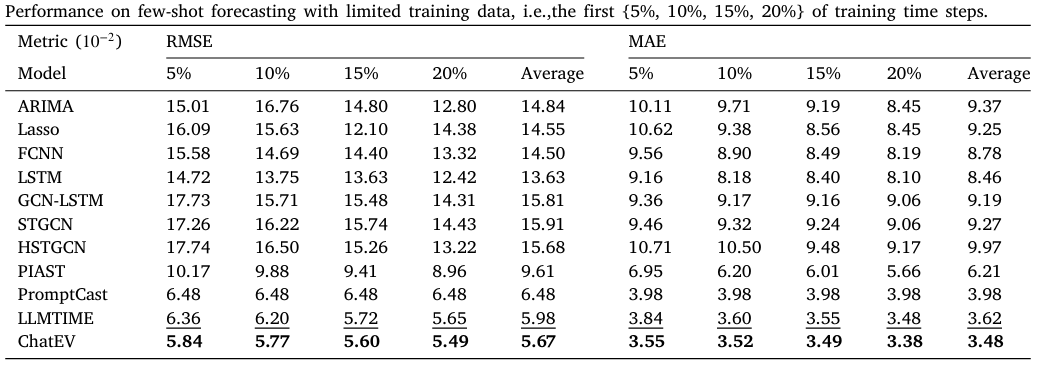
1. 复现情况说明

本文进行比较的方法从四个方面进行讨论：（1）全样本预测，以展示模型在预测电动汽车充电占用率方面的表现；（2）少样本预测，以展示模型快速适应特定区域的能力；（3）零样本预测，以展示模型在处理未见过区域的预测任务方面的能力；（4）消融实验，以验证 ChatEV 中每个关键组件的作用。

全样本预测的性能对比结果如下。

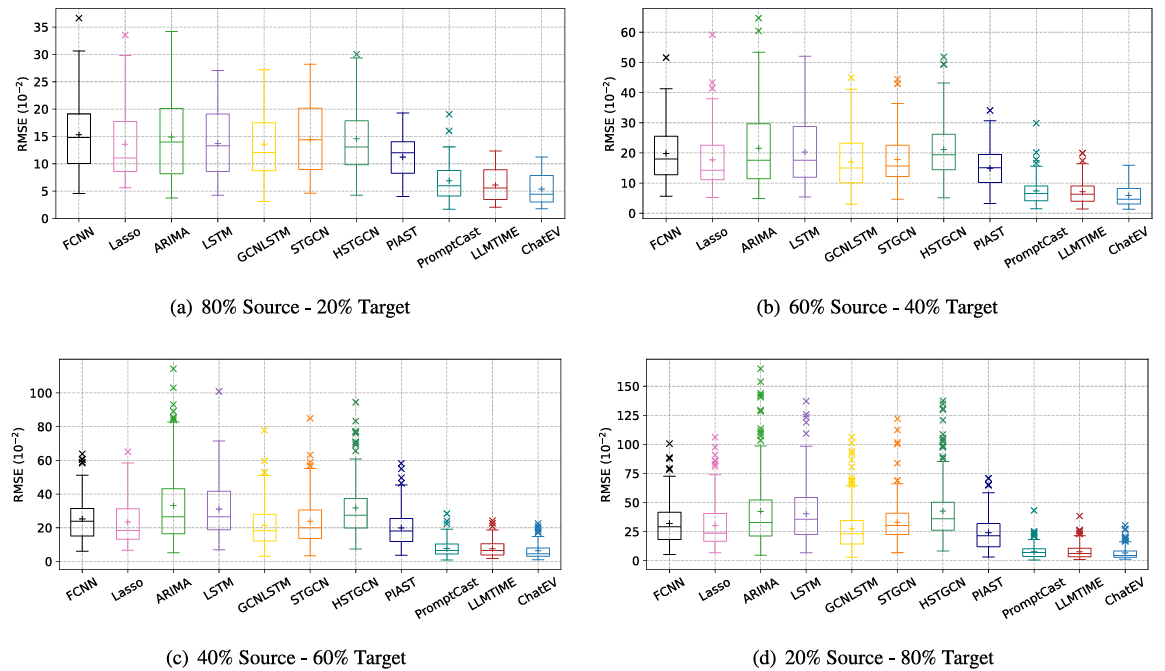
其中最佳和次佳结果分别用粗体和下划线标出。可以得出以下观察结果。首先，ChatEV 在所有指标上都取得了最先进的性能，并且在所有测试区间内始终优于所有基线方法。平均而言，ChatEV 在均方根误差（RMSE）和平均绝对误差（MAE）上分别优于最强基线 3.7%和 4.6%。这种改进表明ChatEV的有效性以及微调大型语言模型在电动汽车充电需求预测方面巨大的潜力。在全样本场景中，这可以归因于基于 Transformer 的大参数量语言模型所获得的出色学习能力，使ChatEV能够有效地捕捉电动汽车数据中的潜在模式。ChatEV 还利用了所提出的对齐调优和设计提示的好处，收集了从时间到空间方面全面的信息。其次，作为一种早期的 RNN 方法，LSTM 用于捕捉时间序列中的隐藏模式，其性能优于 ARIMA、Lasso 和 FCNN。这一观察表明，将非线性时间知识整合到时间序列预测中的潜力。然而，LSTM 在空间信息捕捉方面不如时空预测方法（即 GCN-LSTM、STGCN、HSTGCN 和 PIAST），这表明其捕捉空间信息的能力不足。第三，时空预测方法在传统方法（ARIMA 和 Lasso）和代表性的递归方法（即 LSTM）中表现出相对更强的性能。这突显了通过学习邻接关系来整合空间特征的重要性。 然而，这些先进的方法仍然难以整合社会经济特征等异构数据，以预测电动汽车充电需求。对于基于大语言模型的方法，PromptCast 是一个有竞争力的基线，因为它在所有指标上都实现了令人印象深刻的预测准确性，无需额外的适应或微调。它通过利用大语言模型的自然语言理解能力来整合多种多样因素作为文本上下文进行预测。此外，LLMTIME 优于 PromptCast，这归功于其分词方法，该方法在数字之间插入空格，并采用典型的微调过程，从而从电动汽车数据中提取特定知识。这两种基于大语言模型的方法突显了自然语言处理在作为下一代电动汽车充电需求预测有前途的替代方法方面的潜力。然而，本文提出的方法之所以优于 PromptCast 和 LLMTIME，是因为引入了对齐微调，这增强了模型有效对齐和捕捉相关信息的能力。

本文评估了有限微调数据（即≤训练时间步的前 20%）的场景，得到的少样本预测的性能对比结果如下。

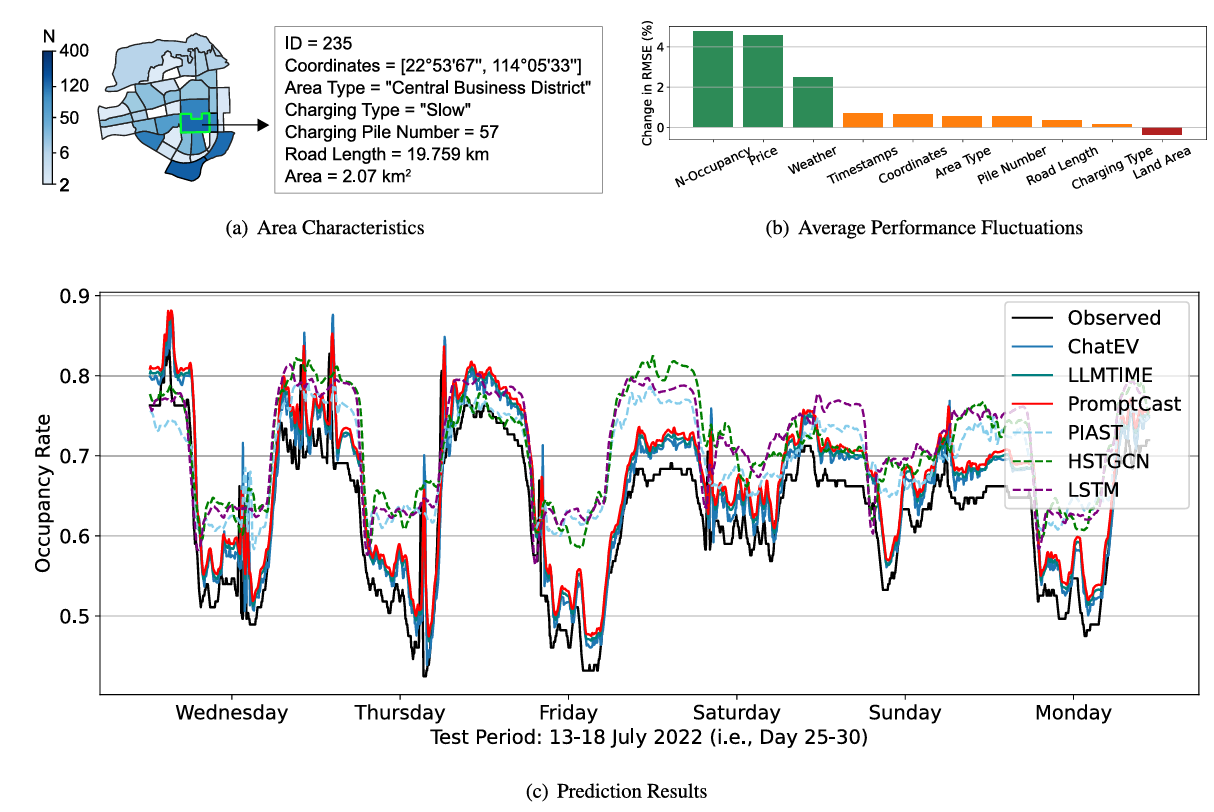


结果显示，现有的数据驱动预测方法，包括统计方法和深度学习方法，在面对仅包含训练数据的 5%、10%、15%和 20%的少样本场景时表现不佳。尽管 PIAST 引入了额外的先验知识（即约束），以帮助其在少样本场景中优于其他学习模型，但其性能仍然不尽如人意。相比之下，新颖的基于提示的预测方法 PromptCast 在这些数据驱动方法中表现显著优于其他方法，这可能可以归因于其预训练语料库中存在的开放世界知识。值得注意的是，由于 PromptCast 是一个基于提示的模型，不进行任何微调过程，因此其预测结果在所有少样本设置中保持一致。 此外，LLMTIME 在有限的训练数据中融入了本地知识，并从大语言模型中提取先验知识，从而在 PromptCast 上取得了更好的性能。而本文提出的方法超越了所有基线，突显了语言模型作为高效的电动汽车充电需求预测器的潜力。研究结果还表明，ChatEV 只需少量微调数据就能快速适应特定区域的具体下游预测任务，展示了该模型的快速学习能力。

本文进行了零样本预测的性能对比，即当模型在其他领域优化时（即在这些领域中模型未见过领域的任何数据样本），预测模型在未见过的领域中的表现如何。本文从研究的 247 个交通区域中随机抽取 20%、40%、60%和 80%作为源领域，而剩余的区域作为测试的目标领域，并使用 30 分钟的预测协议进行这次零样本评估，预测指标的 RMSE 结果以箱线图的形式展示在下图中，其中预测指标的均值和中位数分别用“+”和横线表示，“×”表示异常值。

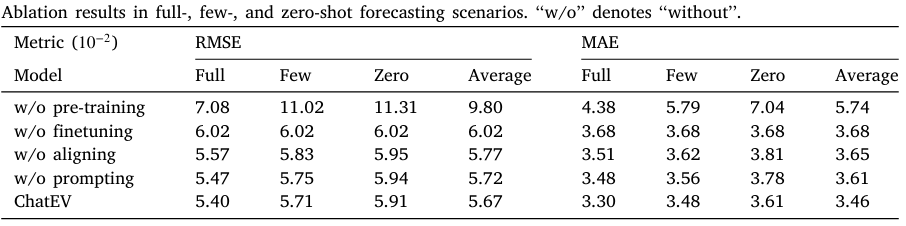


可以看到，ChatEV 在跨区域零样本适应中表现出显著的泛化能力，即异常值较少，并且预测误差低于所有基线模型。这表明所提出的模型可以根据其开放世界知识做出合理的预测，即使在特定区域没有见过相关数据。这种泛化能力突显了 ChatEV 作为可靠的在线电动汽车充电需求预测器的巨大潜力，在现实场景中高质量数据通常稀缺。同样，PromptCast 和 LLMTIME 利用预训练语言模型进行基于提示的预测，快速适应能力仅次于本文中的模型。这是因为 ChatEV 利用了针对电动汽车充电需求预测定制的提示，使其能够超越 PromptCast 和 LLMTIME 等基于大语言模型的方法，后者遵循的是针对一般时间序列预测设计的提示。相比之下，在零样本预测任务中，统计和学习方法显得力不从心：随着源区域的减少，它们在目标区域的表现越来越差。这些数据驱动的模型很难从如此少量的数据中提取出适用于所有区域的充电模式，这再次突显了引入开放世界先验知识以支持泛化预测任务的重要性。

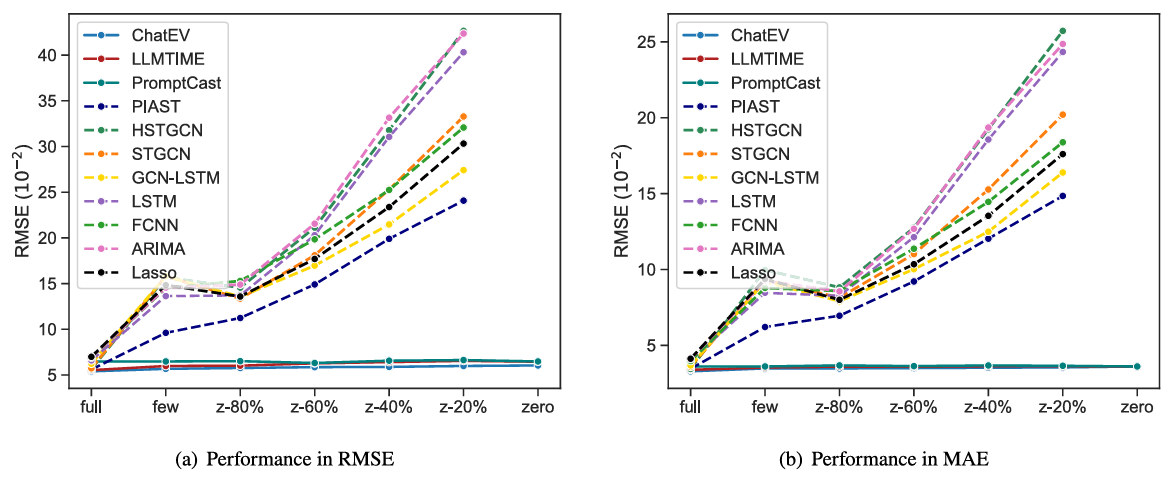


上图展示了零样本性能的示例，直接体现了大语言模型在预测充电需求（即未来占用率）方面的高效性和有效性。具体而言，子图（a）展示了具有多种特征的区域特性示例；子图（b）展示了这些特征在预测中的重要性，通过逐一移除每个特征来表示性能变化（正值表示移除特征后模型性能下降，反之亦然）；子图（c）展示了基线预测曲线与实际占用率曲线的对比。根据图中的观察，可以得出以下几点。首先，子图（a）显示了来自多种数据资源的多样化特征，如坐标、充电价格和天气条件，可以在大语言模型兼容的自然语言空间中高效融合。在子图（b）中，可以看到静态特征，如坐标、区域类型、道路长度、桩数和充电类型会导致性能下降。 这突显了在基于大语言模型的电动汽车充电需求预测中，将静态特征以文本描述形式整合的重要性，以便更有效地确定目标区域。另一方面，在实验场景中，包含“土地面积”信息可能会引入预测过程中的噪声。这是因为该特征在不同交通区域可能会不规则地变化；例如，许多大面积区域可能缺乏充电设施。与静态特征相比，动态变量，即邻近占用率（N-Occupancy）、充电价格（Price）和天气（Whether），对 ChatEV 的预测准确性影响更大。从模型中排除这些动态特征会导致预测性能显著下降，表现为 RMSE 的增加。最后，根据子图（c），两种基于大语言模型的预测器，即 PromptCast、LLMTIME 和 ChatEV，在零样本设置中能够与真实占用率保持同步。这种出色的表现归功于大语言模型提供的开放世界知识，这使它们能够快速学习并有效泛化。 相比之下，三种深度学习基线模型容易出现欠拟合的问题。最后，ChatEV 由于从提出的提示和对齐过程中学习到的个性化知识，表现出更优的性能，并实现了最小的预测误差。

本文还进行了消融实验研究各环节对结果的影响。包括（1）不使用提示（w/o prompting）：移除设计好的提示，包括区域特征化和任务指令，仅保留时间序列作为模型输入；（2）不进行对齐（w/o aligning）：禁用提出的对齐调优过程，并在我们的大语言模型主干上进行常规微调；（3）不进行预训练（w/o pre-training）：使用空白的 T5 模型而不是预训练模型，以消除大语言模型中的开放世界知识；（4）不进行微调（w/o finetuning）：仅通过提示语言模型进行预测。



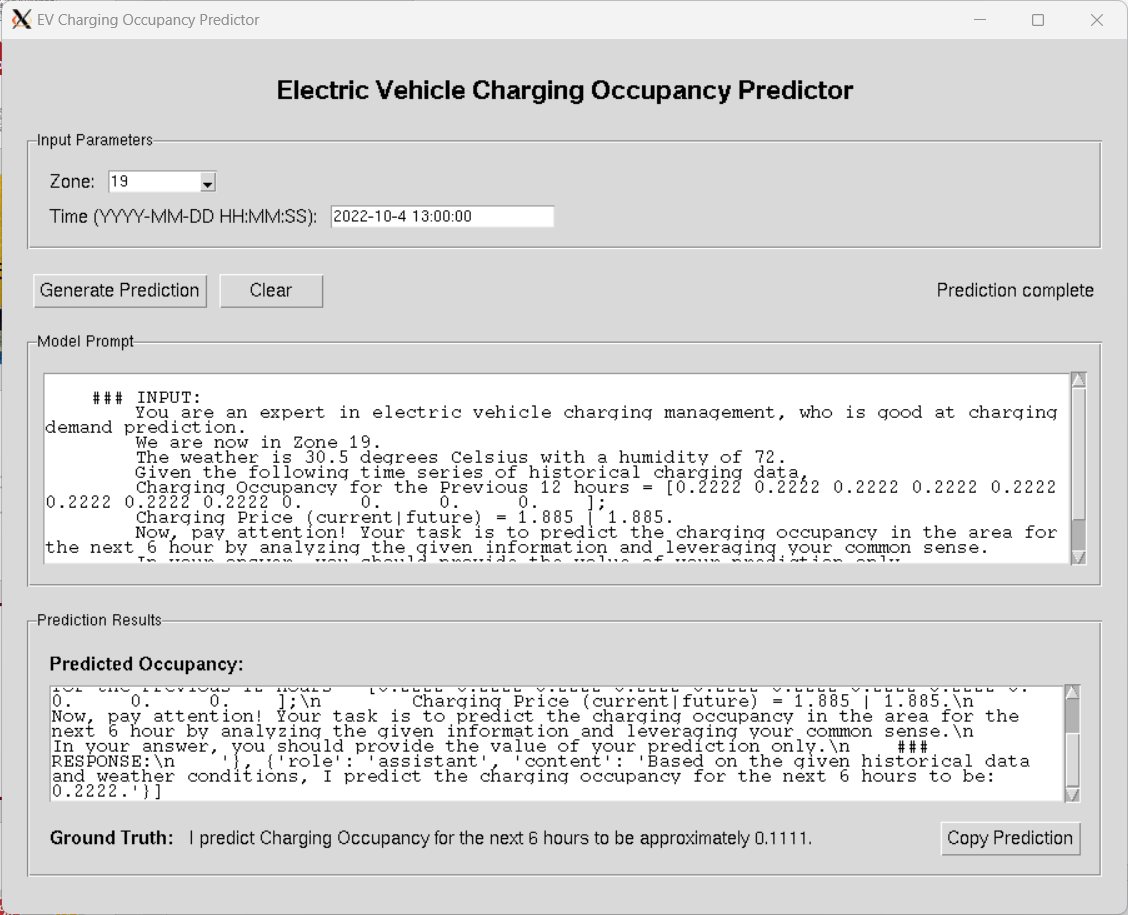
在上表中展示了全样本、少量样本和零样本预测场景的消融结果。其中，所有少量样本指标在四种训练数据大小：{5%，10%，15%，20%}下取平均值，而所有零样本结果在四种未见区域采样比例：{20%，40%，60%，80%}下取平均值。从消融结果中，可以得出以下观察。首先，本文提出的方法中的每个组件都对整体性能有所贡献，因为移除任何一个都会导致性能下降。其次，在移除大语言模型中呈现的开放世界知识（即未进行预训练的情况下），尤其是在少量样本和零样本场景中，性能出现了显著下降。这再次强调了使用大语言模型作为电动汽车充电需求预测器的潜力。此外，在没有任何微调的情况下，ChatEV 的性能有所下降，但仍能实现与 PromptCast 和 LLMTIME 相当的性能，这表明所提出的提示词和预训练知识的有效性。



简而言之，上图概述了在不同场景（全样本、少样本和零样本）下模型评估的情况。其中，训练和微调可用源区域的比例{80%、60%、40%、20%}表示来自源区域的训练数据量，而“zero”场景则表示没有训练数据。可以看出，使用大语言模型进行电动汽车充电占用预测是有效的。此外，在全样本设置下，ChatEV 表现出更优的性能，将其定位为最先进的解决方案。这种改进可以归因于语言空间内多种信息的整合。最后，所提出的模型在与其他基线模型的比较中表现更优，在少样本和零样本场景中展示了其显著的泛化能力，能够在有限数据的情况下适应新场景。

1. 改进情况说明

由于原文直接在控制台中输出了结果，我们制作了一个GUI，在这个界面中只需要选择对应的“zone”即充电站的ID和时间就可以将对应的数据（比如地区、天气、前面一段时间的占用率等）插入到设计好的提示模板中，然后进行预测，生成接下六个小时的充电站占用率。



1. 参考文献

Haohao Qu, Han Li, Linlin You, Rui Zhu, Jinyue Yan, Paolo Santi, Carlo Ratti, Chau Yuen, ChatEV: Predicting electric vehicle charging demand as natural language processing, Transportation Research Part D: Transport and Environment