改进LIME并应用于负载预测的研究

1. 研究背景：

   负载预测是预测未来一定时间区间内电网的电力负载消耗情况。精确的负载预测对电力部门至关重要,可以帮助制定电力生产计划,进行经济调度,确保电网稳定运行。随着分布式能源接入,电力负载呈现出空间分布式、时间动态变化的新特征。传统预测模型难以适应这种复杂情况。当前黑盒模型如LSTM虽在预测精度上有优势,但可解释性差。如果预测错误,不知道原因,也难以提升模型。负载预测涉及重要决策,如果模型不可靠,会导致严重后果。为确保预测结果可信,需要能对预测进行解释,发现模型错误的原因。同时,解释也有助于operators更好地理解和应用预测,提高决策质量。

2. 相关综述：

   a. 数据驱动的负载预测模型

模型	论文	优势	劣势
LSTM	[1,2,3]	利用时间相关性	内部状态不可解释
CNN-LSTM	[4,5]	CNN提取局部特征,LSTM建模时间相关性	无法解释两者融合
ResNet	[6]	残差学习提高性能	模型不可解释
XGBoost	[7,8]	集成决策树,泛化性能好	决策过程不可视化
SVR	[9]	参数优化提高健壮性	线性假设简单化

3. 方法选择

   a. 解释技术

解释技术	方法	优点	缺点	来源文献
基于注意力机制	-	生成特征重要性	需要修改网络结构	[10]
梯度反向传播	Grad-CAM等	生成显著性图	结果可能有噪声	[11]
相关性传播	LRP	追踪特征贡献	计算复杂,不够灵活	[12,13]

解释技术	方法	优点	缺点	来源文献
Shapley值	-	近似全局解释,理论基础好	计算复杂	[14,15,16]
局部模拟模型	LIME等	解释任何黑盒模型	解释可能不稳定	[17,18]

b. 选择LIME的理由

   i. 模型无关性

      LIME可以解释任意黑盒模型,无需知晓模型细节,这样即使更换模型也无需修改代码。

   ii. 局部解释

      LIME专注单个样本附近,可以定位预测错误的具体原因,而不是全局解释。

   iii. 采样策略

      LIME采样生成样本邻域的思路可行且易实现,为获得本地解释数据提供了思路。

   iv. 简单可解释

      LIME定义了“可解释性”, Outputs对人更可解释。

   v. 直观表达

      LIME用特征重要性条形图等直观展示结果,便于用户理解。

   vi. 算法公开

      LIME算法公开,可基于它进行改进来解决时间序列的解释问题。

   vii. 已有应用

      LIME已在类图像领域应用,可证明其解释黑盒模型的有效性。

c. 在时间序列方面的不足

   i. 采样策略问题

      原始LIME是针对图像等数据,其采样策略是局部的随机扰动。这可能会产生不代表性的邻域样本。对时间序列而言,随机扰动会破坏时间顺序结构。

   ii. 距离度量问题

      LIME中的样本权重依赖距离度量。原始LIME直接使用欧式距离,而对时间序列需要设计合理的距离度量,以反映时间相关性。

   iii. 稳定性问题

LIME每次运行会采样不同的样本邻域,产生的解释可能不稳定。对时间序列而言,需要采取措施提高解释的稳定性。

iv. 线性假设问题

LIME使用简单的线性模型拟合复杂模型。对具有时序动态的模型,简单的线性关系可能不够描述。

v. 计算效率问题

LIME需针对每个样本进行采样和拟合,计算效率较低。在线解释时间序列还需要提升效率。

vi. 评价指标问题

LIME当前评价指标可能不完全适合时间序列,需要设计针对稳定性、准确性等的评价指标。

综上所述,当前LIME在时间序列数据上存在采样策略、距离度量、稳定性、线性假设等问题,需要进行改进才能很好地应用到时间序列可解释性任务上。

4. 改进方案

a. 采样策略

可以考虑先用时间序列聚类算法(如K-means)对训练数据进行聚类,获得多个具有相似时间模式的聚类。然后在为每个样本生成解释时,仅从其所属的聚类中采样,而不是从整个训练数据中完全随机采样。这可以获得与当前样本在时间模式上更加一致的邻域样本。

b. 距离度量

可以设计融入时间相关性的距离度量,如加权编辑距离。编辑距离计算两个序列变换到完全相同需要的编辑操作次数,通过设计操作的时间相关性权重,可以反映时间顺序信息。这可以获得比直接欧式距离更可解释的样本权重。

c. 提高稳定性

可以设置LIME的随机数生成器种子,使每次运行采样结果一致,避免解释波动。也可以设计确定性采样,如划分聚类后固定抽取每个聚类的k个样本。

d. 使用非线性模型

考虑将LIME中的线性模型改为非线性回归,如神经网络、树回归等,以处理时间序列的动态非线性。但是模型复杂度不能太高,需要权衡可解释性。

e. 提升计算效率

可以只对部分相关变量进行采样扰动,降低不必要的开销。也可以优化代码,利用并行计算等手段加速距离计算和模型拟合过程。

f. 评价指标

可以设计时间序列可解释性的定量指标,如同一样本多次解释的一致性来评估稳定性,用真实数据进行归一化来评估准确性等。

5. 后续计划

a. 实现流程方面:

- [x] ~~理解LIME方法的细节,分析其在时间序列数据上的不足~~
- [x] ~~改进LIME采样策略,使邻域样本更有代表性~~
- [x] ~~设计时间序列距离度量,用于LIME样本权重计算~~
- [x] ~~收集数据,进行负载预测模型训练~~
- [x] ~~应用改进LIME进行预测结果解释,可视化关键变量~~
- [ ] 设计算法评价指标,和原LIME比较以显示提升
- [ ] 撰写论文阐明方法与结果

[1]Kong, Weicong, et al. "Short-term residential load forecasting based on LSTM recurrent neural network." IEEE transactions on smart grid 10.1 (2017): 841-851.

[2]Kwon, Bo-Sung, Rae-Jun Park, and Kyung-Bin Song. "Short-term load forecasting based on deep neural networks using LSTM layer." Journal of Electrical Engineering & Technology 15 (2020): 1501-1509.

[3]Muzaffar, Shahzad, and Afshin Afshari. "Short-term load forecasts using LSTM networks." Energy Procedia 158 (2019): 2922-2927.

[4]Alhussein, Musaed, Khursheed Aurangzeb, and Syed Irtaza Haider. "Hybrid CNN-LSTM model for short-term individual household load forecasting." Ieee Access 8 (2020): 180544-180557.

[5]Rafi, Shafiul Hasan, Shohana Rahman Deeba, and Eklas Hossain. "A short-term load forecasting method using integrated CNN and LSTM network." IEEE Access 9 (2021): 32436-32448.

[6]Chen, Kunjin, et al. "Short-term load forecasting with deep residual networks." IEEE Transactions on Smart Grid 10.4 (2018): 3943-3952.

[7]Abbasi, Raza Abid, et al. "Short term load forecasting using XGBoost." *Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 33rd International Conference on Advanced Information Networking and Applications (WAINA-2019) 33*. Springer International Publishing, 2019.

[8]Wang, Yuanyuan, et al. "Short-term load forecasting of industrial customers based on SVMD and XGBoost." *International Journal of Electrical Power & Energy Systems* 129 (2021): 106830.

[9]Tan, Zhenqi, et al. "Short-term load forecasting based on integration of SVR and stacking." *IEEE Access* 8 (2020): 227719-227728.

[10]Li, Ao, et al. "Attention-based interpretable neural network for building cooling load prediction." *Applied Energy* 299 (2021): 117238.

[11]Ardito, Carmelo, et al. "ISCADA: Towards a Framework for Interpretable Fault Prediction in Smart Electrical Grids." *IFIP Conference on Human-Computer Interaction*. Cham: Springer International Publishing, 2021.

[12]Kim, Seung Geun, et al. "Enhancing the Explainability of AI Models in Nuclear Power Plants with Layer-wise Relevance Propagation." *Proceedings of the Transactions of the Korean Nuclear Society Virtual Autumn Meeting, Jeju, Korea*. 2021.

[13]Houidi, Sarra, Dominique Fourer, and François Auger. "On the use of concentrated time–frequency representations as input to a deep convolutional neural network: Application to non intrusive load monitoring." *Entropy* 22.9 (2020): 911.

[14]Wang, Wenyu, et al. "Diversity factor prediction for distribution feeders with interpretable machine learning algorithms." *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2020.

[15]Wali, Syed, and Irfan Khan. "Explainable Signature-based Machine Learning Approach for Identification of Faults in Grid-Connected Photovoltaic Systems." *2022 IEEE Texas Power and Energy Conference (TPEC)*. IEEE, 2022.

[16]Zhang, Di, et al. "A bi-level machine learning method for fault diagnosis of oil-immersed transformers with feature explainability." *International Journal of Electrical Power & Energy Systems* 134 (2022): 107356.

[17]Fan, Cheng, et al. "A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning." *Applied Energy* 235 (2019): 1551-1560.

[18]Machlev, Ram, et al. "Measuring explainability and trustworthiness of power quality disturbances classifiers using XAI—Explainable artificial intelligence." *IEEE Transactions on Industrial Informatics* 18.8 (2021): 5127-5137.