2023年5月14日 23:36

Evaluation of interpretability methods for multivariate time series forecasting

主要贡献：1. 介绍了两种新的 对比局部可解释性的 评估标准

2. 对比了3种局部可解释性的方法

方法：

1. 4个数据集：耗电量，Rossmann销售额， 沃尔玛销售额， 俄亥俄的门诊人数
2. 模型：TDNN， LSTM， GBR
3. 可解释性方法：
    1. Random baseline (随机rank)
    2. Omission

       $$\phi_{j\ell,t} = f(\mathbf{X}_t) - f(\mathbf{X}_{t \setminus x_{j\ell,t}})$$

    3. SHAP
4. 评估标准
    1. Area over the perturbation curve for regression (AOPCR)
        1. AOPCR measures the effect of removing the top K Features
        2. focuses on a small percentage of the most important features
        3. Higher values show higher local fidelity
    2. Ablation percentage threshold (APT)
        1. APT measures the percentage of features that need to be removed to pass a certain threshold
        2. usually requires the removal of a higher percentage of features.
        3. Lower percentage shows higher local fidelity.
5. 结论：
    1. SHAP method has the highest fidelity