

组会汇报

陈钶杰
专业:计算数学

August 15, 2023

目录

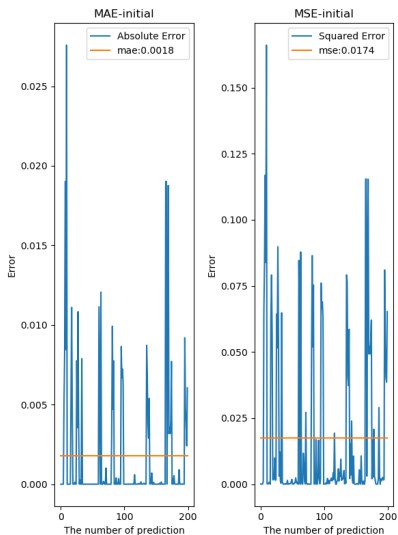
1

代码调试相关

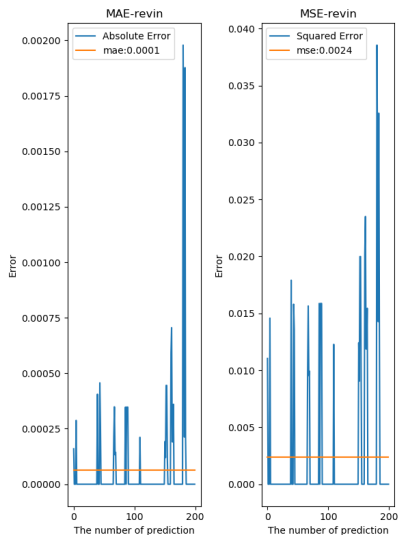
- 对训练数据集进行RevIN处理
- 使用粗粒化的相关实验
- 混合多种类数据集测试
- 下一步的计划

RevIn的具体做法

- 1 将数据集进行规范化
将训练和测试的数据集进行规范化
- 2 使用语言模型对规范化的数据集进行训练
- 3 预测规范化后的数据结果
- 4 将预测的规范化数据进行反规范化，来得到最终的结果(还没做完)



(a) 未用RevIN的测试结果, pass:100%

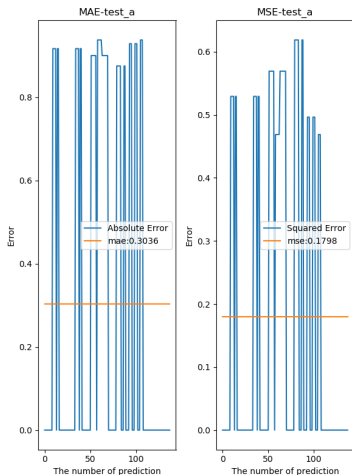


(b) RevIN的测试结果, pass:100%

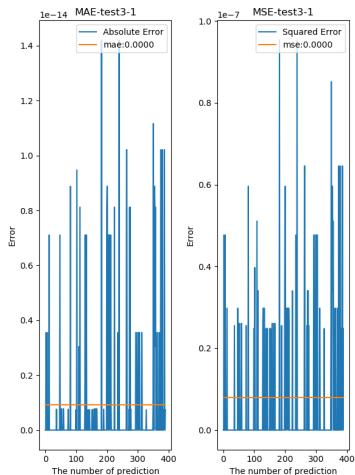
实验结论分析

- 使用RevIN后，每个数会变的很长，举个例子,本来是1,经过规范化以后就是0.2340001了。
- 如果均保留4位有效数字，就会导致对于实际相近的数据，经过归一化以后，都变成同一个值了。
- 对于RevIN的使用方法是否有问题？

关于预测长度的粗粒化测试

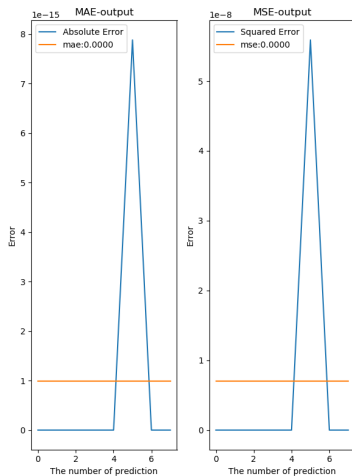


(c) 通过长度/预测 l , pass:36.51%

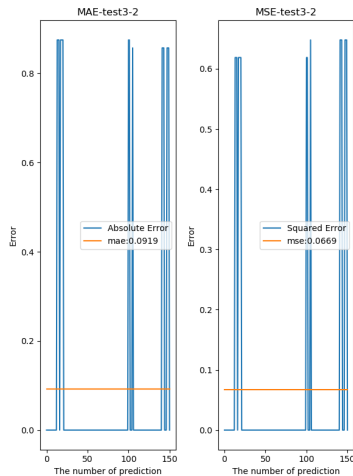


(d) 通过长度/预测 $\frac{l}{2}$, pass:97.5%

泛化性测试



(e) 通过长度/预测 l , pass:4.3%

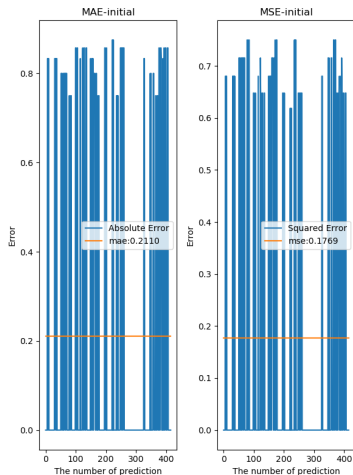


(f) 通过长度/预测 $\frac{l}{2}$, pass:75.5%

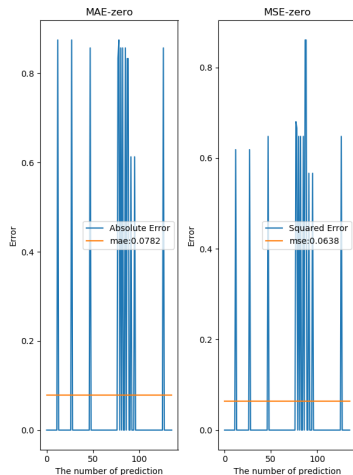
实验结论分析

- 即使使用粗粒化以后，预测长度变长以后对结果的精确度影响以及通过率影响都很大
- 修改的通过率低有一部分原因是语言模型预测长度有限导致的。
- 修改预测长度对通过的影响很大。

将现有的所有训练数据集随机选取部分混合进行测试



(g) 测试集数据, pass: 37.14%



(h) 泛化性测试, pass: 68%

实验结论分析

- 对于见过的任务结果比未见过的任务的通过率都要低，这个可能的原因是对于语言模型而言，对于数值越大的序列，预测的结果可能会于原本的数值差距很大，比如997,998,999，当语言模型可能预测出0.5，导致误差很大。
- 选取的数据集种类多，但是数量相对较少，所以预测结果不太精确。

下一步计划及相关问题

- 将RevIN的预测结果进行解码
- 使用auto-cot，对比当前的结果
- 看相关论文
- 提取模型中的注意力权重，查看模型对于输入信息的处理细节
- 问题：
 - ① 不确定RevIN的做法是否正确？
 - ② 如何解决经过规范化以后数值比较接近的问题？

谢谢老师和同学们的聆听!