

# 组会汇报

陈钶杰  
专业:计算数学

July 18, 2023

# 目录

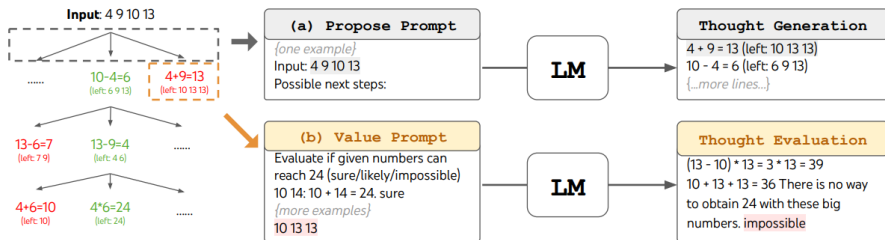
- 1 论文阅读:
  - 思维之树

- 2 代码调试
  - 结果展示
  - 实验结果分析
  - 下一步的计划

# 问题背景

- 大模型的推理过程是一个token级别的，从左到右的一个决策过程，对比人类的思考方式来说，有着非常大的局限（例如人类写文章，写着写着发现写错了，我们可以回过头来，重新修改前面的内容，然后再继续往后写，而大模型LM不能这样），这使得大模型在需要探索，全局分析（前瞻探索或者回溯），初步决策发挥关键作用的任务中效果不太好。
- 即现有的大模型有如下的缺点：
  - ① 局部性：现有LM模型不会在思维过程中探索不同的延续，即不会去探索当前节点的其他分支解决方案。
  - ② 缺乏全局性：大模型LM没有纳入任何类型的规划、展望或回溯来帮助评估不同的选项。
- 人类推理特点：可以重复使用可用的信息来进行启发式的探索，促使挖掘出更多真正有用的信息，找到最终的解决方案。

## Tree of Thoughts方法



- 用一个算24点的例子,来说明具体使用思维树.
  - ① 思维分解:问题的提问方式可能一样(都是一句话),但是问题类型可以进行细分,比如算24点就是运算一行方程.
  - ② 思维生成:生成多种可能,比如像上图中的多行thoughts.
  - ③ 思维评估:对生成的thoughts进行可能性分析.
  - ④ 搜索算法:选择合适的算法进行最优结果的搜索,比如这里选择BFS.

## 总结

- 提出了一种完全由LLM + 启发式搜索算法结合的TOT算法，其可以从多条解决路径，快速的找到最佳解决方法，可以解决的一些复杂的，GPT-4都表现差的一些任务。其主要由thought生成、thought评估、搜索算法组成，可以生成解决方案、对方案进行自我评估、同时可以通过回溯算法来延续之前的解决思路，利用剪枝算法过滤不可靠解决方案，提升找到最优解决路径的速度。
- TOT其各个部分都是高度模块化的，例如可以用不同的LM，不同的搜索算法来实现，通用性比较强，同时其对于每个任务thought的定义都不太一致，如何针对不同的任务设置更好的thought也比较重要，他这里提出了“不能太小”、“不能太大”的指导原则可以参考。
- LMs与经典人工智能方法的交叉是未来工作的一个重要的方向。

# 相关启示

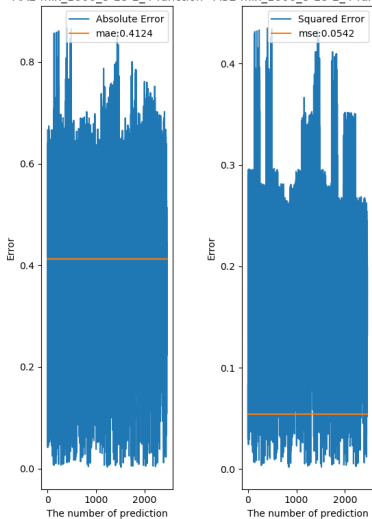
- 针对我们目前做的时间序列的提示:可以参考他的思维分解,将我们的问题直接归到时间序列预测问题.
- 在思维产生的时候可以是对于我们当前序列的一些更正.想法可以是不同的拟合方式,也可以是的时间序列预测的相关方法,选择相应最合适的方法.然后通过相应的评估(比如计算MAE,MSE)作为最终的判别条件.

新的提示在部分训练数据中出现乱码现象

[illegible]

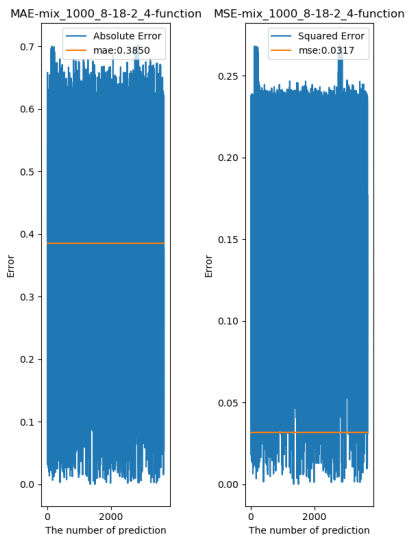
# new prompt(见过的任务)

MAE-mix\_1000\_8-18-2\_4-function MSE-mix\_1000\_8-18-2\_4-function

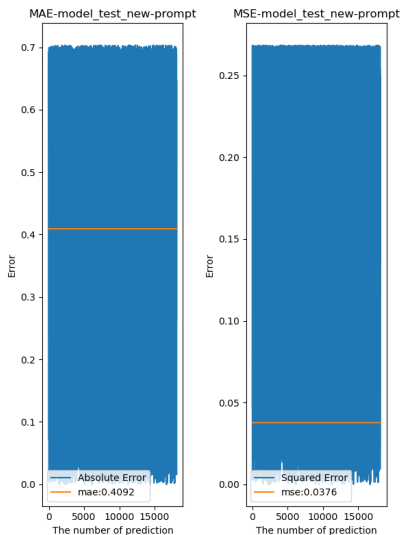




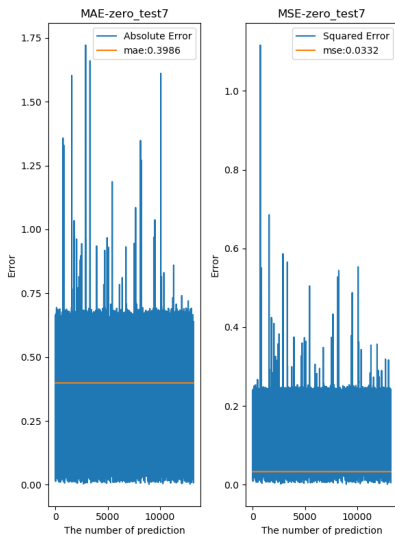
# old prompt(见过的任务)



# new prompt(泛化性能测试)



# old prompt(泛化性能测试)



# 实验结论

- 新的提示和旧的提示在没见过的任务上的性能上的测试以及在见过的任务上的测试基本一致,甚至旧的提示比新提示效果更好!从中可以看出修改提示的排列可能对模型最终的训练影响不大.
- 使用新的提示后,部分测试结果出现了乱码现象.
  - ① 使用chatglm2训练一个大型数据集时,在预测已经见过的任务上出现了乱码的现象.
  - ② 使用chatglm进行泛化测试的一个实验,也出现了乱码的情况.

# 下一步计划

- 是否还要使用新的提示方式?(可能更加简洁的提示效果会更好)
- 如何提高模型的泛化能力
  - ① 在通过修改提示的方式上目前效果不太好.
  - ② 通过迁移学习来提高泛化能力.
  - ③ 寻找提高模型泛化能力的相关论文?
- 进行时间序列的预测.

# 谢谢老师和同学们的聆听!