

组会汇报

陈钶杰
专业:计算数学

July 4, 2023

目录

- 1 论文:PromptCast时间序列预测
 - 文章中和当前工作相似的内容
 - 文章中得到的一些启发
- 2 代码调试工作
 - 结果展示
 - 实验结果分析
 - 下一步的计划

input,output形式

- 该文章中的输入和输出形式和我们的结构基本相同,但他的微调方式是Fine-tuning。

TABLE I
TEMPLATES FOR TRANSFORMING PISA-NUMERICAL TO PISA-PROMPT.

			Template	Example
CT	Input Prompt (Source)	Context	From $\{t_1\}$ to $\{t_{\text{obs}}\}$, the average temperature of region $\{U_m\}$ was $\{x_{t_1:t_{\text{obs}}}^m\}$ degree on each day.	From August 16, 2019, Friday to August 30, 2019, Friday, the average temperature of region 110 was 78, 81, 83, 84, 84, 82, 83, 78, 77, 77, 74, 77, 78, 73, 76 degree on each day.
		Question	What is the temperature going to be on $\{t_{\text{obs}+1}\}$?	What is the temperature going to be on August 31, 2019, Saturday?
	Output Prompt (Target)	Answer	The temperature will be $\{x_{t_{\text{obs}+1}}^m\}$ degree.	The temperature will be 78 degree.
ECL	Input Prompt (Source)	Context	From $\{t_1\}$ to $\{t_{\text{obs}}\}$, client $\{U_m\}$ consumed $\{x_{t_1:t_{\text{obs}}}^m\}$ kWh of electricity on each day.	From May 16, 2014, Friday to May 30, 2014, Friday, client 50 consumed 8975, 9158, 8786, 8205, 7693, 7419, 7595, 7596, 7936, 7646, 7808, 7736, 7913, 8074, 8329 kWh of electricity on each day.
		Question	What is the consumption going to be on $\{t_{\text{obs}+1}\}$?	What is the consumption going to be on May 31, 2014, Saturday?
	Output Prompt (Target)	Answer	This client will consume $\{x_{t_{\text{obs}+1}}^m\}$ kWh of electricity.	This client will consume 8337 kWh of electricity.
SG	Input Prompt (Source)	Context	From $\{t_1\}$ to $\{t_{\text{obs}}\}$, there were $\{x_{t_1:t_{\text{obs}}}^m\}$ people visiting POI $\{U_m\}$ on each day.	From May 23, 2021, Sunday to June 06, 2021, Sunday, there were 13, 17, 13, 20, 16, 16, 17, 17, 19, 20, 12, 12, 14, 12, 13 people visiting POI 324 on each day.
		Question	How many people will visit POI $\{U_m\}$ on $\{t_{\text{obs}+1}\}$?	How many people will visit POI 324 on June 07, 2021, Monday?
	Output Prompt (Target)	Answer	There will be $\{x_{t_{\text{obs}+1}}^m\}$ visitors.	There will be 15 visitors.

评估指标

- 引入了一个评估指标：缺失率，定义为

$$\frac{(n_{test} - n_{decoded})}{n_{test}} \times 100\%$$

这个 n_{test} 和 $n_{decoded}$ 分别表示测试的数量和测试集中正确解码的数量。

我在测试的时候，使用的精确度类比这个就是 $\frac{n_{decoded}}{n_{test}} \times 100\%$

- 文章使用了RMAE,MSE的评估指标。但是文章中进行了五次不同的随机种子运行，并报告了平均性能和标准差。这样做是为了考虑到深度学习方法在不同的随机初始化下可能产生不同的结果。通过进行多次运行并计算平均性能和标准差，可以提供更全面和可靠的评估。

文章的重点实验

- 他将10个语言模型和10个时间序列模型进行比对，但是语言模型是直接接受三组数据的训练，而时间序列模型是对每组数据进行相应的训练，最终得到了对比的结果。
- 测试这些模型的泛化能力:一共三组数据，用其中两组数据对模型进行训练，剩下的一组数据作为测试集合，来测试模型的精确度。最后发现语言模型在完成未见过的任务上要优于其他时间序列模型。
- 根据预测长度的对比实验，比如根据前面7天的数据预测后面1天，4天，7天的结果。根据文章的结果可得预测的长度越长，误差就越大，相同的预测长度下，历史数据越多，误差越小。

文章的重点实验

- 对于同一个问题，使用不同的提示方式，进行微调，结果也不同。其中提示C的性能更好，可能相比提示A,提示AB中没能把不同特征之间的关系描述出来。因此我之后可以尽可能使用鲜明对比的提示方式。

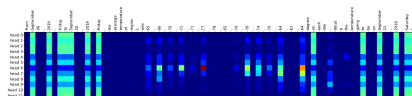
TABLE XI
EXAMPLES OF TWO TYPES OF PROMPTS UNDER THE MULTIVARIATE TIME SERIES SETTING.

Input Prompt A&B	From May 13, 2016, Friday to May 27, 2016, Friday, the PM2.5 was 32, 23, 20, 53, 82, 113, 133, 94, 64, 83, 20, 4, 18, 48, 57; PM10 was 32, 40, 57, 87, 90, 113, 133, 94, 64, 83, 20, 20, 18, 48, 103; and SO2 was 2, 3, 4, 11, 29, 27, 26, 33, 13, 14, 6, 2, 2, 6, 16 on each day. What are the pollutant values going to be on May 28, 2016, Saturday?
Output Prompt A	The pollutant values will be 3, 23, 2.
Output Prompt B	The PM2.5 will be 3. The PM10 will be 23. The SO2 will be 2.
Input Prompt C	From May 13, 2016, Friday to May 27, 2016, Friday, the PM2.5 was 32, 23, 20, 53, 82, 113, 133, 94, 64, 83, 20, 4, 18, 48, 57 on each day. What is the PM2.5 value going to be on May 28, 2016, Saturday? From May 13, 2016, Friday to May 27, 2016, Friday, the PM10 was 32, 40, 57, 87, 90, 113, 133, 94, 64, 83, 20, 20, 18, 48, 103 on each day. What is the PM10 value going to be on May 28, 2016, Saturday? From May 13, 2016, Friday to May 27, 2016, Friday, the SO2 was 2, 3, 4, 11, 29, 27, 26, 33, 13, 14, 6, 2, 2, 6, 16 on each day. What is the SO2 value going to be on May 28, 2016, Saturday?
Output Prompt C	The PM2.5 will be 3. The PM10 will be 23. The SO2 will be 2.

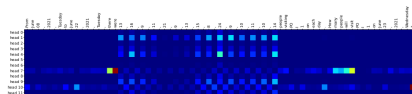
文章的重点实验

● 探索语言模型可以进行时间序列预测的原因?

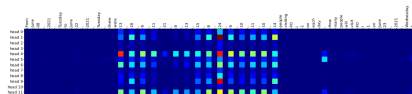
通过使用热力图，将提示中的每一个单词的重要性都展示出来了，可以看到越到后面几层，提示中的历史数据就变成最重要的信息了，这不仅说明了这个提示序列的重要性，还说明了这个语言模型确实用到了历史序列的信息，因此能够进行时间序列的预测。



(a) Case 1: CT sub-set.



(b) Case 2 (layer 4): SG sub-set.



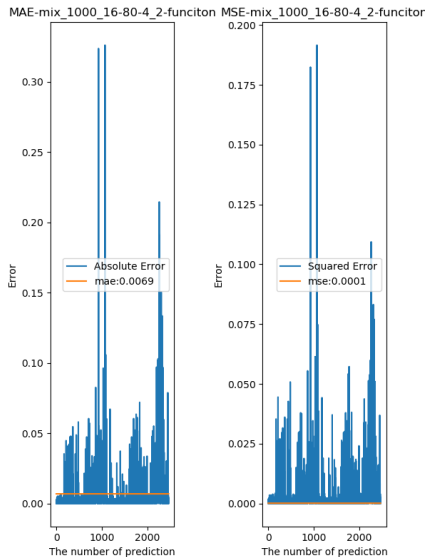
(c) Case 2 (layer 5): SG sub-set.

5. Visualizations of attentions in language models under the proposed PromptCast setting.

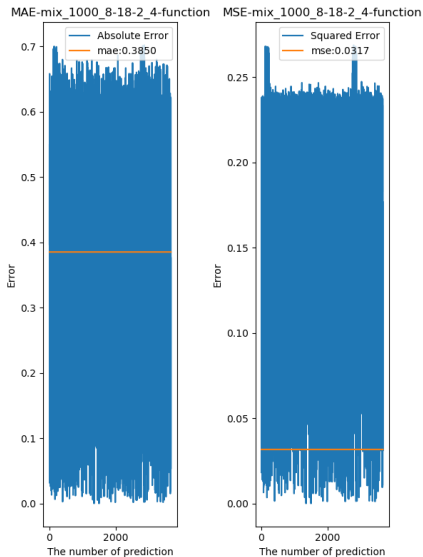
后续相关任务

- 需要做一个多模型比对实验。
- 选择合适的提示方式，参考文章中使用更加能够描述不同特征的提示，并且关于时间的提示只是给定了开始和结尾，可以省略大量的信息。
- 在评估RMAE和MSE指标的时候需要进行多次随机种子实验来保证测试的可靠性。
- 可以像文章中使用热力图的方式，将语言模型能够进行预测的内部原因挖掘出来。

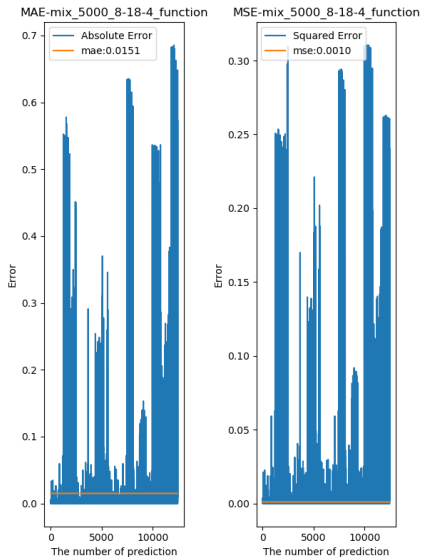
准确率:100%, 微调总数据量:40000, 两种不同序列和20种不同长度的组合, 每种类型1000



准确率:98.05%,微调总数据量:40000,四种不同序列和5种不同长度的组合,每种类型1000



准确率:99.95%,微调总数据量:200000, 4种不同序列和5种不同长度的组合,每种类型5000



根据数据的结论

- 从中可以看出，不同的预测长度对于精确度影响不大，但是不同种的序列对精确度的影响比较大。
- 但类型数量取1000的时候，对于不同种类的序列会使得最终误差较大，但是提高数量到5000，能够让误差明显的变小。

实现目标

- 尝试对未见过的序列进行预测，看看结果如何。
- 修改提示方式，寻找更好的描述不同的特征的提示。
- 多种模型的测试，参考PromptCast论文中，不仅要有语言模型进行预测，还要用时间序列的模型预测来进行结果比对。
- 尝试使用其中一篇文章中的MEGABYTE方法来处理长序列进行建模，以解决时间数据集中序列过长的的问题。

谢谢老师和同学的聆听!