

# 时间序列预测的指令微调

陈钊杰  
专业:计算数学

May 30, 2023

# 目录

- 1 建模的基本想法
- 2 语言模型的指令学习
- 3 时间序列预测的指令微调

# 建模目标:具有一定泛化能力的时间序列预测模型

## 具体任务:

- 在单个数据集中通过指令微调使得最终效果变好.
- 此模型能够接受多种指令微调的数据集合,并在这些数据集上的预测有比较好的表现.
- 此模型在接受一个未知的数据集也能作出相应的预测.

# 指令微调的文献-谷歌FLAN模型

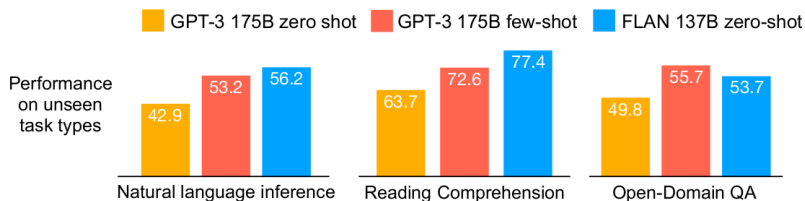
- 文章内容:

- 什么是指令微调
- 指令微调的优点
- 指令微调的缺点

# 什么是指令微调(语言模型)

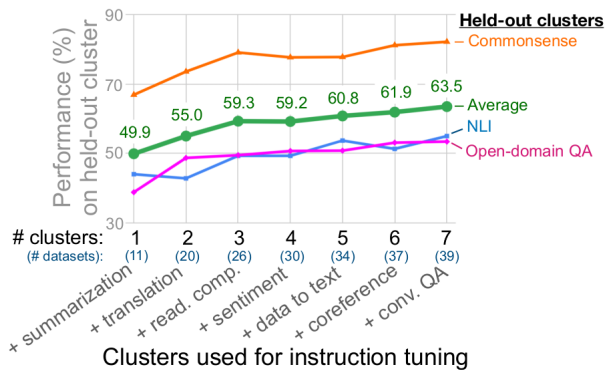
- **fine-tuning**: 需要对原始数据集中每一个数据进行标记, 训练. 这就需要所有的数据训练库, 比如情感分类, 我要通过对语言库中所有的语句进行标记, 这显然消耗巨大.
- **prompting**: 根据单个任务给一个特定标记的数据集, 用这个特定标记的数据集来训练模型, 其中这个标记的数据集规模可能远小于语言库中的数据集. 促使模型能够完成此任务.
- **instruction tuning** 是通过在多个任务上添加提示, 得到指令微调的数据集进行微调模型, 能在一个未见过的任务上有良好的表现.

# 指令微调的优点



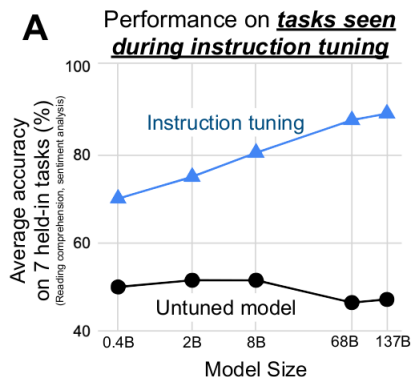
1. 对于未见过的任务, FLAN模型的效果最好, 甚至超越比他参数更多的gpt-3模型.

# 指令微调的优点



2.从图表中可以看出,当指令微调的任务越多,最终模型在完成没见过的任务上表现越好.

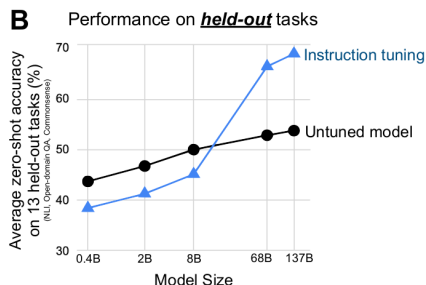
# 指令微调的优点



3.从图表中可以看出,在可见的任务上,使用指令微调的效果比不使用要好很多.而且随着数据数据的变多效果会更加好.



# 指令微调的缺点



1.从图中可以知道,在完成未知任务的时候,只有在模型参数超过10B的时候,FLAN模型才比不微调的模型效果好

原因:可能是在训练过程中,这个指令微调用掉了许多的参数.如果参数不够多的,指令微调就会用掉大量的参数,导致最终模型过拟合,在其他数据集上的效果反而变得更差.

# 指令微调的缺点

## 2. 不同任务的表现不一样

- 在一些任务上很有用:translation(翻译)
- 在一些任务上没啥用:coreference resolution(指代消解)

在这里指代消解就是将代词替换成具体指代的内容,这个就像是完形填空,和预处理的任务很相似,所以增加instructions反而显得冗余.

# 对时间数据进行

## 模型选择

- 使用的函数序列测试,故使用的时间预测模型是DLinear,使得效果更好,速度更快.
- 测试序列的总长度是30000,21000个数据用作训练,3000个数据用作验证,6000数据用作测试,数据是按顺序选取.
- - 1 seq\_len:60
  - 2 pred\_len:10

# 面向时间序列的指令微调的数据集

## 构建常用函数的一个时间序列

- 构建了一个如下的函数时间序列:

$$\begin{array}{cccc}
 f1 : x & f2 : \sin(x^2 + 2) & f3 : \cos(x) & f4 : \sin(x) \\
 1 & \sin(1^2 + 2) & \cos(1) & \sin(1) \\
 \vdots & \vdots & \vdots & \vdots \\
 n & \sin(n^2 + 2) & \cos(n) & \sin(n) \\
 \vdots & \vdots & \vdots & \vdots
 \end{array}$$

- 对于一个时间序列,我们一般关注其中某个维度的信息,我们这边更关注 $\sin(x)$ 即第四个维度上的预测,模型最终的mse,mae是计算第四个维度的预测误差.

## 计算结果:

$$MSE : 0.2742$$

$$MAE : 0.2513$$

# 面向时间序列的指令微调的数据集

## ● 添加合适的指令微调

- $\sin(x)$  是一种周期函数, 所以简单想法就是增加一个周期性函数, 来达到对这个提示的效果:
- 构建了一个如下的微调的函数时间序列:

$f1 : x$	$f2 : \sin(x^2 + 2)$	$f3 : \cos(x)$	$f4 : \sin(x)$	$f5 : x \% 4$
1	$\sin(1^2 + 2)$	$\cos(1)$	$\sin(1)$	$1 \% 4$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n$	$\sin(n^2 + 2)$	$\cos(n)$	$\sin(n)$	$n \% 4$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

## ● 最终计算结果

$MSE : 0.2321$

$MAE : 0.2112$

# 面向时间序列的指令微调的数据集

## 错误的提示会带来更大的误差

- 把提示改成一个非周期的函数  $f5: \frac{\sin(x^2+2)+\cos(x)+\sin(x))}{3}$
- 构建了一个如下的微调的函数时间序列:

$$\begin{array}{ccccc} f1 : x & f2 : \sin(x^2 + 2) & f3 : \cos(x) & f4 : \sin(x) & f5 : \frac{\sin(x^2+2)+\cos+\sin}{3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

## 最终计算结果

$MSE : 0.2935$

$MAE : 0.3110$

## ● 对于未知特征的序列如何提示

- $\sin(x)$ 我们是知道他有周期性的,但是实际的序列的性质我们是未知的,这个时候要如何通过提示?

序列自己本身就是一种很好的提示!

- 构建了一个如下的微调的函数时间序列:

$f1 : x$     $f2 : \sin(x^2 + 2)$     $f3 : \cos(x)$     $f4 : \sin(x)$     $f5 : \sin(x)$

$\vdots$     $\vdots$     $\vdots$     $\vdots$     $\vdots$

## ● 最终计算结果

$MSE : 0.2214$

$MAE : 0.2016$

## ● 对于未知特征的序列如何提示

- 还可以考虑删除序列中相关性比弱的序列,比如这边删除那个相关性比较弱的序列 $\sin(x^2 + 2)$
- 构建了一个如下的微调的函数时间序列:  
 $f1 : x \quad f3 : \cos(x) \quad f4 : \sin(x)$   
 $\vdots \quad \quad \quad \vdots$

## ● 最终计算结果

$MSE : 4.566e - 06$

$MAE : 0.00201$



## ● 总结

- 可以通过增加一列目标值相同的序列,比如上面的 $\sin(x)$ 来对数据起到微调作用.
- 通过计算一个向量各个维度的序列之间的相关系数来考虑删减某一列.

## ● 下一步计划

- 思考如何在同一个模型里面,尝试使用指令微调对多种不同的数据集进行训练,能够使得模型能够在已知数据集上通过微调后有更好的表现.

## ● 问题

- ① 对于多种不同数据集合,最终对于每种序列的预测维度都是指定的吗?
- ② 时间序列和自然语言处理模型之间是有显著差异的。  
比如对于语言模型而言,

- 情感分类问题的提示可以是:这部电影很有趣,这个电影是[好看的]
- 翻译任务的提示可以是:这部电影很有趣,这个电影的英文表示是[movie]
- 去解决没有遇到过的问题比如是代词指代:这部电影很有趣,它的剧情很好,这个它的指代是[这部电影]

未见的任务,还是和已有的任务有一定的关联性的,但对于两个不同的时间序列而言,两者之间是完全没有关系的,比如:

- 一个是任务是预测序列 $\log(x)$
- 另一个是任务是预测序列 $\sqrt{x}$
- 未见的任务是预测序列 $\sin(x)$

未见序列 $\sin(x)$ 和前面已有的两个序列都没关系。  
所以可能对最终预测数据集合是有一定要求的。

# 谢谢老师和同学的聆听!