组会汇报

陈钶杰
专业:计算数学

November 21, 2023

目录

这周主要做的事情

1. 阅读了MLC模型的代码
2. 修改了训练数据集的大小,测试模型的泛化能力对于训练数据量的要求,探索模型在少量数据的情况下模型的泛化能力. 可以说明一下具体修改数据量的过程
3. 对比chatglm模型的任务泛化能力的测试

MLC模型的一些特点说明

- MLC模型相比传统训练的模型的特殊之处在于训练的时候,对于每个训练数据都给出了一个相应的规则.如图:

```
*SUPPORT*
IN: tufa lug gazzer zup OUT: PURPLE GREEN PURPLE GREEN
IN: kiki lug blicket OUT: RED BLUE
IN: kiki dax OUT: BLUE BLUE BLUE BLUE
IN: gazzer lug tufa dax OUT: GREEN PURPLE GREEN PURPLE GREEN PURPLE GREEN
IN: tufa dax OUT: GREEN GREEN GREEN GREEN
IN: blicket OUT: RED
IN: gazzer lug tufa OUT: GREEN PURPLE
IN: blicket kiki OUT: RED BLUE
IN: tufa blicket zup OUT: GREEN RED GREEN RED
IN: tufa lug gazzer OUT: PURPLE GREEN
IN: tufa lug blicket OUT: RED GREEN
IN: tufa kiki lug blicket zup OUT: GREEN RED BLUE GREEN RED BLUE
IN: tufa zup OUT: GREEN GREEN
IN: tufa OUT: GREEN

*QUERY*
IN: blicket dax OUT: RED RED RED RED
IN: kiki zup OUT: BLUE BLUE
IN: gazzer tufa OUT: PURPLE GREEN
IN: gazzer dax OUT: PURPLE PURPLE PURPLE PURPLE
IN: gazzer OUT: PURPLE
IN: kiki OUT: BLUE
IN: gazzer kiki gazzer gazzer blicket OUT: PURPLE BLUE PURPLE PURPLE RED
IN: kiki lug tufa OUT: GREEN BLUE
IN: kiki zup zup OUT: BLUE BLUE BLUE BLUE
IN: kiki lug kiki OUT: BLUE BLUE

*GRAMMAR*
gazzer -> PURPLE
 blicket -> RED
 kiki -> BLUE
 tufa -> GREEN
 x1 zup -> [x1] [x1]
 u1 lug u2 -> [u2] [u1]
 x1 dax -> [x1] [x1] [x1] [x1]
 u1 x1 -> [u1] [x1]
```

```
*SUPPORT*
IN: blicket fep lug OUT: BLUE PINK
IN: zup OUT: YELLOW
IN: lug tufa OUT: BLUE BLUE
IN: lug fep zup fep lug OUT: BLUE YELLOW BLUE
IN: gazzer tufa OUT: PURPLE PURPLE
IN: gazzer zup fep blicket kiki lug OUT: BLUE PURPLE PINK YELLOW
IN: blicket fep lug tufa OUT: BLUE BLUE PINK
IN: zup fep lug gazzer lug kiki zup OUT: YELLOW BLUE PURPLE BLUE YELLOW
IN: lug kiki lug OUT: BLUE BLUE
IN: zup tufa OUT: YELLOW YELLOW
IN: blicket gazzer tufa OUT: PINK PURPLE PURPLE
IN: lug OUT: BLUE
IN: zup tufa kiki zup OUT: YELLOW YELLOW YELLOW
IN: blicket fep zup tufa OUT: YELLOW YELLOW PINK

*QUERY*
IN: blicket OUT: PINK
IN: zup tufa kiki lug OUT: BLUE YELLOW YELLOW
IN: blicket blicket OUT: PINK PINK
IN: gazzer blicket tufa kiki zup OUT: YELLOW PURPLE PINK PINK
IN: gazzer OUT: PURPLE
IN: blicket tufa kiki blicket OUT: PINK PINK PINK
IN: zup tufa kiki blicket OUT: PINK YELLOW YELLOW
IN: blicket tufa OUT: PINK PINK
IN: blicket fep lug blicket blicket tufa OUT: BLUE PINK PINK PINK PINK
IN: blicket lug fep zup tufa OUT: PINK YELLOW YELLOW BLUE

*GRAMMAR*
zup -> YELLOW
 gazzer -> PURPLE
 blicket -> PINK
 lug -> BLUE
 x1 kiki u1 -> [u1] [x1]
 u1 fep x1 -> [x1] [u1]
 u1 tufa -> [u1] [u1]
```

探究MLC模型泛化能力

- 探究数据量对MLC模型泛化能力的影响
  1. 通过使用1000,10000,100000个txt数据文件分别探究模型泛化能力的比较.
  2. 对每个数据量进行训练,并进行泛化性测试,得到如下表的结果准确率.
- MLC模型与chatglm的泛化能力对比
  1. MLC模型和chatglm模型使用相同大小的数据集,均使用10000个txt文本数据集(chatglm模型没有使用额外的GRAMMAR提示信息)

Table: 准确率结果(模型)

模型及参数	准确率(验证集)	准确率(泛化)
data:1000,nepochs:10	7.14%	0%
data:1000,nepochs:100	7.79%	0%
data:1000,nepochs:200	9.14%	0.5%
data:10000,nepochs:10	7.0%	0.0%
data:10000,nepochs:100	100%	57.2%
data:10000,nepochs:200	100%	72.2%
data:100000,nepochs:10	100%	71.7%
chatglm,data:10000	-	3.21%

结论

1. chatglm语言模型在逻辑任务上的泛化表示不好,而且其中的3%完成的只是最基本的一对一问题.
2. MLC模型在足够多epochs条件下,能够得到很强的泛化性,基本上和chatglm喂养相同的数据集,MLC模型对于泛化任务能够达到70%的准确率,而chatglm只有3%,几乎就是没有泛化能力.
3. MLC模型可以使用更少量的数据集,训练时使用更多的epochs也可以达到不错的准确率.
4. 从代码中可知MLC模型结构比较简单,在单种任务上有很不错的泛化效果,但是对于多种任务的泛化能力上表现还是有待提高!

下一步计划

1. 做其他的实验,确保他的这种框架在类似任务上同样有效.
2. chatglm的prompt改成和MLC模型中类似的提示数据集.

谢谢老师和同学们的聆听!