清华大学电子工程系 **媒体与认知** 课堂 2 2022-2023 学年春季学期

作业 4

卜丞科

2023年6月3日

理论部分

- 1 单选题 (15 分)
- 1.1 B
- 1.2 D
- 1.3 <u>B</u>
- 1.4 A
- 1.5 C
- 2 计算题 (15 分)
- 2.1 隐含马尔可夫模型的解码

某手机专卖店今年元旦新开业,每月上旬进货时,由专卖店经理决策,采用三种进货方案中的一种:高档手机(H),中档手机(M),低档手机(L)。

当月市场行情假设分为畅销 (S_1) 和滞销 (S_2) 两种。畅销时,三种进货方案的概率分别为 0.4, 0.4, 0.2; 滞销时,三种进货方案的概率分别为 0.2, 0.3, 0.5。

某月份市场行情为畅销,下一个月份为畅销和滞销的概率分别为 0.6 和 0.4;某月份市场行情为滞销,下一个月份为畅销和滞销的概率分别为 0.5 和 0.5。

开业第一个月市场行情为畅销和滞销的可能性均为 0.5。

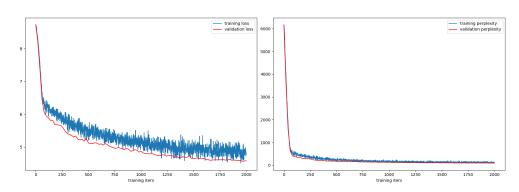
- (1) 如果我们采用隐含马尔可夫模型 (HMM) 对该专卖店进货环节建模,请写出 HMM 对应的参数 $\lambda = \{\pi,A,B\}$ 。
- (2) 在第一季度中,采购业务员执行的进货方案为"高档手机,中档手机,低档手机",即观测序列为 H, M, L。请利用 Viterbi 算法推测前三个月的市场行情。

$$(1)$$
 $\pi = [0.5, 0.5]$
 $A = [0.6, 0.47]$
 $B = [0.4, 0.4]$
 $0.5 = [0.4, 0.2]$
 $0.5 = [0.4, 0.4]$
 $0.5 = [0.5, 0.5]$
 (2) 6 , $(1) = [0.5, 0.4] = [0.2]$
 6 , $(2) = [0.5, 0.4] = [0.4]$
 9 , $(1) = [0.5, 0.4] = [0.4]$
 6 , $(2) = [0.2, 0.4] = [0.4]$
 6 , $(2) = [0.2, 0.4] = [0.4]$
 6 , $(2) = [0.2, 0.4] = [0.4]$
 6 , $(2) = [0.2, 0.4] = [0.4]$
 6 , $(2) = [0.4] = [0.4]$
 6 , $(2) = [0.4] = [0.4]$
 6 , $(3) = [0.4] = [0.4]$
 6 , $(4) = [0.4] = [0.4]$
 6 , $(4) = [0.4] = [0.4]$
 6 , $(4) = [0.4] = [0.4]$
 6 , $(4) = [0.4] = [0.4]$
 6 , $(4) = [0.4] = [0.4]$
 6 , $(4) = [0.4] = [0.4]$
 6 , $(4) = [0.4] = [0.4]$
 6 , $(4) = [0.4] = [0.4]$
 6 , $(4) = [0.4] = [0.4]$
 6 , $(4) = [0.4] = [0.4]$
 6 , $(4) = [0.4] = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $(4) = [0.4]$
 6 , $($

3 编程作业报告

3.1 模型的训练与测试

使用默认参数训练中文文本数据集(全宋词), loss 及困惑度变化如下图:



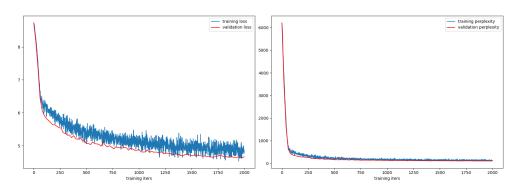
默认配置下生成样本(节选):

指定初始文本生成样本 (节选):

3.2 探究位置编码和残差连接在模型中的作用

3.2.1 关闭位置编码训练

训练后的 loss 函数和困惑度变化为:



生成的文本节选为:

```
++沁园春(送别)

青秋霭西风,天自有酒知闲。对梅谁宜约。笑语行下三洲。

春心无计。水晶帏、多少离离梦。归梦一年晚。

+++谒金缕兰花(次韵)

楚堂凝望。正月淡寒。玉楼斜时去。翠丝轻衣。风细雨小。

闲记取当枝先知。一夜雨悲愁。风凰山路隔。

++日取儿(第一)

一点流水生月。春寒应是处。人人知。千目送归去。

一年年老。应更须著真。尽匆匆匆得。到江西风。

++七君归。且爱甘州

春信与山路。青亭春老。明月无人。酒醒来无语。不堪怜。

++七君归。且爱甘州

春信与山路。青亭春老。明月无人。酒醒来无语。不堪怜。

++大人在几度。醉著个人生。

秋风如此。明月破尽芳菲。

++四楼中天上语春。绿丝竹柳,花明双鬟垂红。

++柳色微霞袖香。轻步小春半。

翠衾佳人寄柳枝。小窗长杨柳绕。□□□中。流初过了,燕泥尘生无处。

+++调金林小雨。金石。正风残云幕

月如今暮。烟波绕江湖,一夜高歌头。细被轻雾花。

西风催得酒醒。燕子风暖。又是闲关愁。别后还是春晚。

+++贺新郎(送西、中游)

绿窗几番风雨雨。又似黄边柳。双燕。芳草草。绿阴清,夜深深春信,曾倚楼。

小池前好道、空深户。十分明朝雨。欲到春风不在。

+++念奴娇
     清平乐年年,都是醉狂,人在危栏。向春色新啼。金花明月月,夜前都闲。

信人人误玉,江南门著。但明斗一声尘。

+诉衷情似。花上归

花暗映,几时花时。暗缕风流水。时花落花容。

尽尽雨还惊。雪成。断肠断肠留。对花如今。

+诉阑干欲与今。泪调歌(题)

食何处。有酒酒还人。一抹青山影。

二珠朱槛,白水东风雨。几度情吹阑时。

寄吟。谁知心情。一笑目断肠。雨转回思量。

+浣沙

减字木兰龙(梅)
         1773。世紀·日高。

元沙

或字木兰花(梅)

主天色,花尽芳草草。月落花红妆。月斜低。秋色落日寒。朱户。一枝清明。

鬼空。如许。把断肠人、双入啼。天欲别路。

满楼(长亭)

礼回首。冷落花长。却在人间绿。离恨只。相思归去。

岩人物水,花行。

小冷。秋面晚春风冷。人相思今,谁知。

人间利,一朝归来。平生一水。

·浣溪沙沙(画楼)

提春寒,似多年年年少。天然故。休要人秋。不似当时节。

聚天应笑,几年年少,今夕阳。月浸红红泪未忍。

玉箫鼓。都随菊夜新妆。香雾转惊。月浸凝辉。浅晓妆半。纵是飞絮、照溪湾。

·忆旧时令(自题赵赠作)

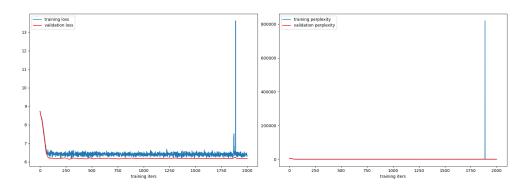
皮满。步池塘东流。初摇、春阴水,梦阑干倚阑。

兰院碧。数声里、
```

由此可见,关闭位置编码虽然对于 loss 和困惑度的下降影响不是很大,两者依然可以降到和没有关闭时相近的水平。但是生成的文字语义比较混乱,有的地方完全无法理解在说什么,并且经常出现连续使用相同的字的现象,生成文本的质量有较为明显的下降。由此可见,位置编码可以使机器学习到不同词应该用在文本中的什么位置,使得生成的文本语序,逻辑都更加通顺自然。

3.2.2 关闭残差连接训练

训练后的 loss 函数和困惑度变化为:



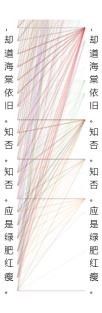
生成的文本节选为:

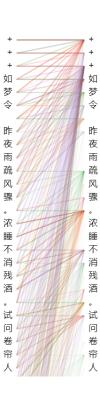
```
(test) E:\桌面\\媒体与认知\hw4\hw4\python sample.py — ckpt_path workdirs/quansongci_no_res sample from workdirs/quansongci_no_res\latest.pth data/quansongci | hotala/quansongci | hotala/quanso
```

由此可见,关闭了残差连接以后 loss 值和困惑度明显比之前大,并且在训练过程中还有冲击峰值。并且发现生成的文本根本无法阅读,几乎和乱码一样,这可能是因为出现了梯度消失等问题。由此可见,残差连接可以避免梯度消失,极大地提高模型的训练效果。

3.3 可视化

选用第二层观察可视化:





通过可视化结果,我们可以清楚地观察到注意力系数能够帮助模型有效地 提取上下文中的关键信息,从而生成更加符合正常逻辑的文本序列。此外, 注意力系数还有助于模型理解文本的语义和结构,在处理相反或相同语义 的词语和标点符号时表现出色,总而言之注意力系数为模型和文本之间建 立了一种动态而灵活的关系,这种关系可以根据不同任务和模型进行优化 调整,从而有效提高模型对文本的理解和生成能力。此外,多头注意力机 制使得注意力系数能够表示不同子空间和不同层次的文本信息,这对于处 理多样性和复杂性的文本非常有帮助。