

作业 4

Jiaji Liu

2023 年 5 月 19 日

---

## 理论部分

### 1 单选题 (15 分)

1.1 B

1.2 D

1.3 B

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。](#)

解. (1)  $S_1$ : 畅销,  $S_2$ : 滞销, H: 高档, M: 中档, L: 低档。

初始状态空间的概率分布  $\pi = (0.5, 0.5)$ 。

状态转移概率矩阵 A 为:  $A = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix}$ , 即  $P(S_1|S_1) = 0.6$ ,

$P(S_2|S_1) = 0.4$ ,  $P(S_1|S_2) = 0.5$ ,  $P(S_2|S_2) = 0.5$ 。

给定状态下观察值的概率分布 B:  $P(H|S_1) = 0.4$ ,  $P(M|S_1) = 0.4$ ,

$P(L|S_1) = 0.2$ ,  $P(H|S_2) = 0.2$ ,  $P(M|S_2) = 0.3$ ,  $P(L|S_2) = 0.5$ 。

(2)

$$\delta_1(1) = \pi_1 b_1(O_1) = 0.2 \quad \varphi_1(1) = 0$$

$$\delta_1(2) = \pi_2 b_2(O_1) = 0.1 \quad \varphi_1(2) = 0$$

$$\delta_2(1) = \max\{\delta_1(1)P(S_1|S_1), \delta_1(2)P(S_1|S_2)\}b_1(O_2)$$

$$= \max\{0.2 \times 0.6, 0.1 \times 0.5\} \times 0.4 = 0.048$$

$$\varphi_2(1) = 1$$

$$\delta_2(2) = \max\{\delta_1(1)P(S_2|S_1), \delta_1(2)P(S_2|S_2)\}b_2(O_2)$$

$$= \max\{0.2 \times 0.4, 0.1 \times 0.5\} \times 0.3 = 0.024$$

$$\varphi_2(2) = 1$$

$$\delta_3(1) = \max\{\delta_2(1)P(S_1|S_1), \delta_2(2)P(S_1|S_2)\}b_1(O_3)$$

$$= \max\{0.048 \times 0.6, 0.024 \times 0.5\} \times 0.2 = 0.00576$$

$$\varphi_3(1) = 1$$

$$\delta_3(2) = \max\{\delta_2(1)P(S_2|S_1), \delta_2(2)P(S_2|S_2)\}b_2(O_3)$$

$$= \max\{0.048 \times 0.4, 0.024 \times 0.5\} \times 0.5 = 0.0096$$

$$\varphi_3(2) = 1$$

$$P^* = \max\{\delta_3(1), \delta_3(2)\} = 0.0096$$

$$q_3^* = 2, \quad q_2^* = \varphi_3(2) = 1, \quad q_1^* = \varphi_2(1) = 1。$$

因此, 最优状态序列  $q_1^* q_2^* q_3^*$  为  $S_1, S_1, S_2$ , 即“畅销, 畅销, 滞销”。  $\square$

## 3 编程作业报告

### 3.1 完成基于 GPT 的文本生成任务代码程序

#### 3.1.1 完成多头自注意力机制的计算

补全代码得到：

---

```
def forward(self, x, attn_mask=None):
    batch_size, seq_len, dim = x.shape
    q, k, v = self.q_layer(x), self.k_layer(x), self.v_layer(x)
    q = q.view(batch_size, seq_len, self.num_head, self.head_dim)
    k = k.view(batch_size, seq_len, self.num_head, self.head_dim)
    v = v.view(batch_size, seq_len, self.num_head, self.head_dim)
    q = q.transpose(1, 2)
    k = k.permute(0, 2, 3, 1)
    v = v.transpose(1, 2)
    attn = torch.matmul(q, k) / math.sqrt(self.head_dim)
    if attn_mask is not None:
        attn = attn.masked_fill(attn_mask, float('-inf'))
    attn = F.softmax(attn, dim=-1)
    attn = self.attn_drop(attn)
    out = torch.matmul(attn, v)
    out = out.transpose(1, 2).reshape(batch_size, seq_len, self.num_head
        * self.head_dim)
    result = self.proj_layer(out)
    result = self.proj_drop(result)
    return result, attn
```

---

#### 3.1.2 完成 GPT 的前向计算

补全代码得到：

---

```
def forward(self, word_idx, targets=None):
    batch_size, seq_len = word_idx.shape
    # >>> TODO 2: complete the forward process of GPT
    pos = torch.arange(seq_len, dtype=torch.long, device=word_idx.device)
    token_embed = self.word_token_embedding(word_idx)
    pos_embed = self.word_pos_embedding(pos)
    x = self.drop(token_embed) if self.no_pos else self.drop(token_embed
        + pos_embed)
    attn_mask = torch.ones((seq_len, seq_len),
        device=word_idx.device).byte()
```

---

```
attn_mask = torch.triu(attn_mask, diagonal=1)
attention_weights = []
for i in range(self.num_layer):
    x, attn = self.transformer[i](x, attn_mask)
    attention_weights.append(attn)
logits = self.language_model_head(self.norm(x))
# <<< TODO 2
```

---

## 3.2 训练/文本生成/可视化

### 3.2.1 模型的训练与测试

在中文文本数据集上训练，得到的 loss 以及 perplexity 的曲线如图 1所示，生成样本得到如图 2。

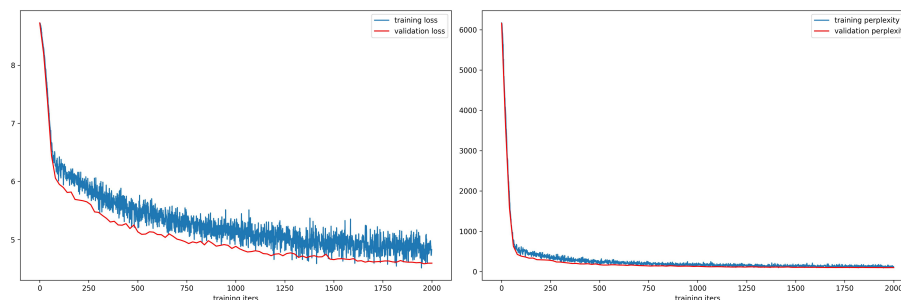


图 1: 中文文本数据集训练 loss 以及 perplexity

```
+++玉楼春（送赵宗父日）
日暮年春去花中。晓色相思处，江湖影。回首春情又。云飞絮晓。
玉楼清远，梦魂归去。断肠吹梦，何关无数。是东风月到新人。
+++减字木兰花
叶上池塘。曲屏风，楼上一声。翠屏春去。锦帐、金缕。
珠帘卷。玉娥相知。烟玉。月明年旧韵。算子当时情。
+++虞美人
从前愁不断。到客情难。不似旧时来故语。向取一声成晚，歌声断梦中。
南园春睡起月边。东风月破晴蟾。谁与杨柳花过。
```

(a) 质量较好的文本

```
+++西江月（寿周郎）
池林如今雪，烟波南。楚山无事凭阑。金光欲度楼中。
夜飞鸿短，不恐平生。应好梦魂。
为否春何神子，此游人，愁无限。不须知。
+++少年游（和李陈中）
故人相忆，不羡花风寒。有情情，何许香情处。花来无语，轻暗不能。
莫恨风吹。一夜月春寒，人静处，还是、不成无语。听得三年。
+++水调归（和花）
寒沙上归路，一点鸳鸯凝。正凉如事，数愁只许。梦来倚阑干，不知成柳下梅花。
+++水龙吟（秋雪）
清风起泪。几许桃花又。一枝风定，风痕暮。清明月下，月明月暖。
风流何处。笑写梨梢，只有谁能到。看灯花处住，云情。
+++如得媚
春来不肯醉青溪。人远来时候。绿树侵香正满窗，水摇歌。
昨生分，已是佳辰。
夜永，春风流雨。见好、笑语量消息。况解春风，苦醉随风，已须消。
+++水龙吟（席上三夕）
春风底，风吹半。故园春又晚，依隐、山花、应画角。不觉新迟，独见秦桥路，又月斜阳。
故人无寸笑，佳逢今好。莫唱华年何处，月月相思。一江湖底事，长安几个。
+++鹧鸪天（梅）
绿梧桐柳初满。更疏花对夜飞。记与与花间，不到尽风流。
自有关多好、一枝桃杏香花。待为君、寻寻语。与说与年、不知何计。
```

(b) 质量较差的文本

图 2: 默认配置下生成样本

指定初始文本“+++ 水调歌头”得到：

```

+++水调歌头（腊韵）
今日过了，不把春事，春光无力。为君人共得，直应来、轻过了。
芳心曾见，相逢无奈。多少年公在处。莫教人相留。回首。何人人与酒，多少身。
+++水调曲（陈徊）
天山林木雪，东风有花东。江南山色，芳边柳外，山山无处。
多愁远古。且休忍来无。何是千情苦。谁遣舞双蛾。
+++西江月
锦。相思梦少。莫放花行，不奈金销，已得悠醒。
年年心暮不似此情。不知多酒杯头。未得人、归来似酒长。
+++山溪沙（用双词）
旧时人、少年年少。真个家山。望溪肠。小檐边，不恨闲人。
+++点绛唇（二里）
雪浓烟渺。绿阴渐。碧烟林。花阴好，香露微。帘卷烛尘肌。
露零寒，风如晓，已觉干点。几度芳早。一离情似梦，细鸟啼。
+++采桑子（和伯州父别）
中物天路。锦衾闲怯。芳草阴中。珠帘幕锁，何处不随云。故园春梦。
人在春迟，轻寒暮户。日风吹絮，轻寒新怨。斜窗里无情。
+++菩萨蛮
长江湖清客。冰靥声。翠屏春去处。空惊起梦回头。君与故楼关。
遥望烟相思。不堪旧韵。算子当时。从今朝明事，有幽时节。
+++洞仙游（十寿）
金门下、风露试，淡霏烟昏。可惜闲成，落落日西湖，月边依旧度。
身远江皋，江南，过水调，都凝别。明朝天气

```

(a) 质量较好的文本

```

+++水调歌头（寿昌父制）
九年年十海中。水中风雨雨。孤水回何处。人不省天涯。
新来何入梦，白星横空。莫他年。有情难寄。明年常自见松。
+++感州令（寿）
长叹行人是，朝日流霞里。人间满酒。多情无计，又不休。春去得更三千载。
+++阮郎仙
++渔家傲
水如孤鳞。露丛晴日雨催花。时节如今梦，画堂深院、几番谁知。
轻鸭又醒，画阁愁似。怕花梢青望似玉。
夜来未到。相思何处，只待玉阑思。犹今夜，乱心悄。
+++渔家傲
青山青眼锦阑干。半桥桥静处，花光湿，夜深山低唱。
万家乡万里，新浓。烟流水，红残照。依旧渐绿，青青梦远，凄凉月长。应相思此处。
+++江波
人在离情。都有情，云里、又天然苦。想今宵、更有、梅花别。
+++醉中措（东叔中）
平亭北，早晚云清绝，晴堤柳、高高雨。雪径罗衾，曾断空，又如愁别。
无多未去，谁知故人有，犹是东流。不说自思、犹归梦断。
+++好事近（声中）
秋到东风，又有东风、是春来别。落日暗，暮山千里。芳心如梦，不须唤。
深情好，情为语，离怀应愿。欢难觅，不定身然，歌声易得。休辞春又断，怕怕怜人，情多少。
+++水调歌头（感）
春来行处，夜远江头雨。春空岸、柳溪紫。又匆匆。一叶情不数

```

(b) 质量较差的文本

图 3: 指定初始文本生成样本

### 3.2.2 探究位置编码和残差连接在模型中的作用

关闭位置编码进行训练，得到的 loss 以及 perplexity 的曲线如图 4所示。

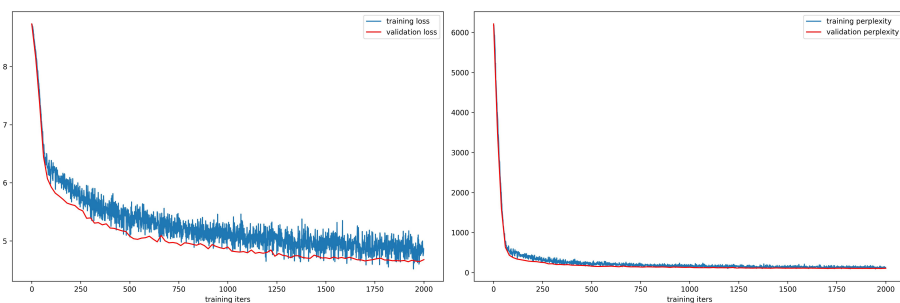


图 4: 关闭位置编码 loss 以及 perplexity

生成样本得到：

```
+++踏莎行
南州晓，暖云笺雪。春色不好。轻风回，莫是旧事，醉花红日空。酒年游多少年。更唱新思，便与行声。
+++相思临望中人（大处）
春心事，斜照秋风，一枝分花信。似笑怜凄凉，珠帷外，不来风与我愁。
玉蟾宫上水，无穷夜深处。东风和楚。
花影下月精神仙。泪珠尘吹。想西风细细。绿阴幽香。
+++菩萨蛮
凉意好月初秋。孤山中秋水，此怀
一似画船。小楼前回去。今日水长灯。
+++醉西湖山，醉中秋夜愁。
多情怀无客。不似碧窗。人间一梦，凭阑干处，唤起来人起。
笑与人心难瘦。算。何处何似古江南，犹恋花知。寄几回年，也为我，有意可怜。
+++菩萨蛮
++菩萨蛮（送春后）
晓风流里，无何如梦。玉奴娇痕，残日暮。归去不放花，几日飞来春去。东风雨。醉乡离情，别一声，未住风流。
月下寒浓。待得当时节。一愁来，相思佳人老。向桃花、人间少。
+++水调歌头（东东）
春霜风月淡，风冷，独是一梦重阳，重阳高昼。不是临佳人处，西湖、东山是，断一落花半。
春来犹记晓，风露稍归。为一枝移。到年王孙会，更有东风，花下月，醉里、寒食莺莺。
```

(a) 质量较好的文本

```
+++东窗墨迹（文字）
晓风流里，无何如梦。玉奴娇痕，残日暮。归去不放花，几日飞来春去。东风雨。醉乡离情，别一声，未住风流。
月下寒浓。待得当时节。一愁来，相思佳人老。向桃花、人间少。
+++水调歌头（东东）
春霜风月淡，风冷，独是一梦重阳，重阳高昼。不是临佳人处，西湖、东山是，断一落花半。
春来犹记晓，风露稍归。为一枝移。到年王孙会，更有东风，花下月，醉里、寒食莺莺。
```

(b) 质量较差的文本

图 5: 关闭位置编码生成样本

可以看到，生成的文本质量略低于有位置编码时的情况，此时生成的文本有拼凑堆砌的嫌疑，不够流畅。

关闭残差连接进行训练，得到的 loss 以及 perplexity 的曲线如图 6所示。

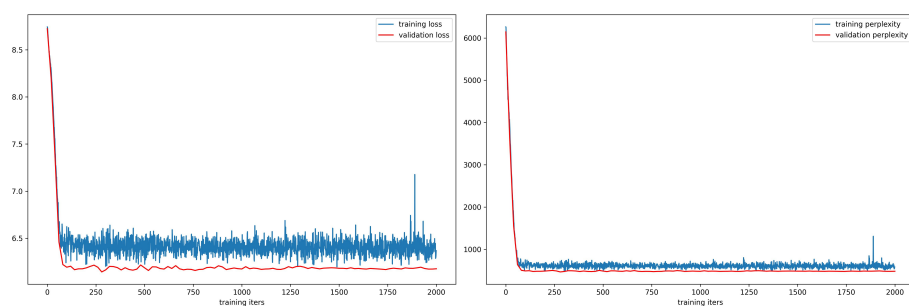


图 6: 关闭残差连接 loss 以及 perplexity

生成样本得到：

，  
。西山+何。，色，  
，，望今雪露。。南。见日。雨  
间不。光，。楼，风。。。黄烟  
，平下。应。。。千+。。春何。。（，游落天寒无人，+。三中  
犹春秋+倚。好。  
醉。，人残，。人，，溪花风寒  
波，，，+十香画水情花来  
到，轻仙+十。  
尽+风。。  
夜。春，到人++，还红白。  
光下柳，初满，断。玉为翠、是归为残花在水，，  
归花  
、。笑更+。尘，如事+。愁，。头。重。光+西雨），帘，春柳长无  
+。。），去中，雪香到清天起，。  
。雪，旧。一老风作，枝，+花清柳、十月，。莫。  
，舞++南。，。长  
。+。满+绿。+。处、意云情、。。。重香，，，，一  
事，青玉光人兰来。，，+。。正）窗，。。人。。生。。+下，白  
、春

图 7: 关闭残差连接生成样本

关闭残差连接后完全不能生成正常的文本。

### 3.2.3 可视化



图 8: 可视化注意力系数



运行 `attnvis.ipynb` 文件，得到如图 8 所示结果。可以看到，一般情况下，当前文字和邻近上文的注意力系数较大，和较远文字的注意力系数较小，说明当前文字主要由邻近上文决定。有些文字和较远的文字之间也存在较大的注意力系数，说明这两处文字存在较强的语义上的关系。另外，不同注意力头生成的注意力系数不同，让模型可以关注到不同语义尺度的特性。

## 4 总结

通过完成基于 Transformer 的 GPT 模型的构建、模型训练和文本生成任务，我深刻体会到了深度学习技术的强大和神奇。这种模型构建和训练的过程需要耗费大量的时间和精力，但我们可以用预训练的模型轻松地生成大量的文本，并且从中发掘到许多有趣的信息。这样的技术不仅可以用于文学艺术的创作，还可以在许多领域中发挥巨大的作用，促进科学技术的发展与进步。