清华大学电子工程系

媒体与认知 课堂 2

2021-2022 学年春季学期

作业 5

程书鹏 2022 年 5 月 5 日

理论部分

1 单选题（15 分）

1.1 B

1.2 D

1.3 B

1.4 D

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) $\pi = [0.5 \quad 0.5]$

$$A = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.4 & 0.4 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

(2) $O = [O_1 \quad O_2 \quad O_3] = [H \quad M \quad L]$

$\delta_1(1) = \pi_1 \cdot b_1(O_1) = 0.5 \times 0.4 = 0.2$, $\delta_1(2) = \pi_1 \cdot b_2(O_1) = 0.5 \times 0.2 = 0.1$

$\psi_1(1) = \psi_1(2) = 0$

$\delta_2(1) = \max\{\delta_1(1) \cdot a_{11}, \delta_1(2) \cdot a_{21}\} \cdot b_1(O_2) = \max\{0.2 \times 0.6, 0.1 \times 0.5\} \times 0.4 = 0.048$

$\psi_2(1) = 1$

$\delta_2(2) = \max\{\delta_1(1) \cdot a_{12}, \delta_1(2) \cdot a_{22}\} b_2(O_2) = \max\{0.2 \times 0.4, 0.1 \times 0.5\} \times 0.3 = 0.024$.

$\psi_2(2) = 1$

$\delta_3(1) = \max\{\delta_2(1) \cdot a_{11}, \delta_2(2) a_{21}\} b_1(O_3) = \max\{0.048 \times 0.6, 0.024 \times 0.5\} \times 0.2 = 5.76 \times 10^{-3}$

$\psi_3(1) = 1$

$\delta_3(2) = \max\{\delta_2(1) a_{12}, \delta_2(2) a_{22}\} b_2(O_3) = \max\{0.048 \times 0.4, 0.024 \times 0.5\} \times 0.5 = 9.6 \times 10^{-3}$

$\psi_3(2) = 1$

$P^* = \max\{\delta_3(1), \delta_3(2)\} = \delta_3(2) = 9.6 \times 10^{-3}$, $q_3^* = 2$

$\therefore q_2^* = 1$, $q_1^* = 1$ $\therefore$ 市场行情最可能为畅销、畅销、滞销.

图 1: 2.1 解答

2.2 循环神经网络的长时相关性建模能力

对序列中的长距离相关信息进行建模是涉及序列的任务中十分重要的一点，例如在阅读理解任务里，题目和正文中的关键词可能相距很远，这就需要模型具备足够好的长距离相关信息建模能力。传统 RNN 在训练时存在梯度消失问题，较远的误差无法得到有效传递，因此学习长距离相关信息时面临较大挑战，在本题中我们对传统 RNN 难以学习长距离相关信息的问题进行一个简单的讨论。

对 RNN 的计算过程进行简化，考虑一个暂不采用激活函数以及输入 x 的 RNN:

$$\boldsymbol{h}_t = U\boldsymbol{h}_{t-1} = U\left(U\boldsymbol{h}_{t-2}\right) = ... = U^t\boldsymbol{h}_0$$

其中 U^t 为 t 个 U 矩阵连乘。若矩阵 U 存在如下特征值分解：

$$U = Q\Lambda Q^\top$$

其中 Q 为单位正交矩阵（每一列为模长为 1 的特征向量），$Q^\top$ 为 Q 的转置，Λ 为特征值对角矩阵，则上述的 RNN 计算过程可表示为：

$$\boldsymbol{h}_t = Q\Lambda^t Q^\top \boldsymbol{h}_0$$

本题目包含以下三个问题：

(1) 假设某一特征值 $\lambda_i < 1$，当时刻 t 增大时，Λ^t 中第 i 行 i 列的值会怎样变化？

(2) 假设 $\boldsymbol{h}_0 = \boldsymbol{q}_i$，其中 $\boldsymbol{q}_i$ 为 U 矩阵的第 i 个特征向量（即 Q 的第 i 列），设 $\mathcal{L}$ 为目标函数计算出的 loss。试验证：

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{h}_0} = \lambda_i^t \frac{\partial \mathcal{L}}{\partial \boldsymbol{h}_t}$$

(3) 对于更一般的 $\boldsymbol{h}_0$，由于 Q 中的特征向量构成一组完备正交基，可以将 $\boldsymbol{h}_0$ 分解为 Q 中不同特征向量的线性组合，即 $\boldsymbol{h}_0 = \sum_{i=1}^{n} k_i \boldsymbol{q}_i$。通过上述分析，请尝试解释传统 RNN 训练中的梯度消失现象，由此理解传统 RNN 对长距离相关信息建模的困难。

解：(1) Λ^t 中第i行i列的值为 λ_i^t，而 $\lambda_i<1$，因此随着t增大，该值会逐渐趋于0.

(2) ∵ $\frac{\partial L}{\partial h_0} = \frac{\partial L}{\partial h_t}\cdot\frac{\partial h_t}{\partial h_0}$

∴即验证 $\frac{\partial h_t}{\partial h_0}=\lambda_i^t$.

$$h_t = (g_1 \cdots g_n)\begin{pmatrix}\lambda_1^t & & \\ & \ddots & \\ & & \lambda_n^t\end{pmatrix}\begin{pmatrix}g_1^T\\ \vdots \\ g_n^T\end{pmatrix} g_i$$

$$=(\lambda_1^t g_1 \cdots \lambda_i^t g_i \cdots \lambda_n^t g_n)\begin{pmatrix}0\\ \vdots\\ g_i^2 \\ \vdots \\ 0\end{pmatrix}$$

又 g_i 为归一化的特征向量

∴ $g_i^2=1$

∴ $h_t = \lambda_i^t g_i$. 即 $\frac{\partial h_t}{\partial h_0}=\lambda_i^t$.

∴即验证了 $\frac{\partial L}{\partial h_0}=\lambda_i^t\cdot\frac{\partial L}{\partial h_t}$.

(3) ∵ $h_0=\sum_{i=1}^n k_i g_i$

∴ $h_t = \sum_{i=1}^n k_i\cdot\lambda_i^t g_i$

∴ $\frac{\partial L}{\partial h_t} = \sum_{i=1}^n \frac{\partial L}{\partial g_i}\cdot\frac{\partial g_i}{\partial h_t}$

$=\sum_{i=1}^n \frac{1}{k_i\lambda_i^t}\cdot\frac{\partial L}{\partial g_i}$

当 $|\lambda_i|<1$ 时，$t\to\infty$ 时 $\frac{\partial L}{\partial h_t}\to\infty$. 梯度爆炸

当 $|\lambda_i|>1$ 时，$t\to\infty$ 时 $\frac{\partial L}{\partial h_t}\to 0$. 梯度消失.

图 2: 2.2 解答

3 编程作业报告

3.1 验证网络变量输出尺寸正确

```
PS E:\Desktop\媒体与认知\第三周演示程序与教程> cd E:\Desktop\媒体与认知\第五次作业\hw5
PS E:\Desktop\媒体与认知\第五次作业\hw5> python network.py
The output size of model is correct!
```

图 3: 验证网络变量输出尺寸

由上图可知网络的输出变量尺寸正确，可进行下一步的训练和验证。

3.2 训练和验证

训练 40 轮完成后，得到可视化结果如下：

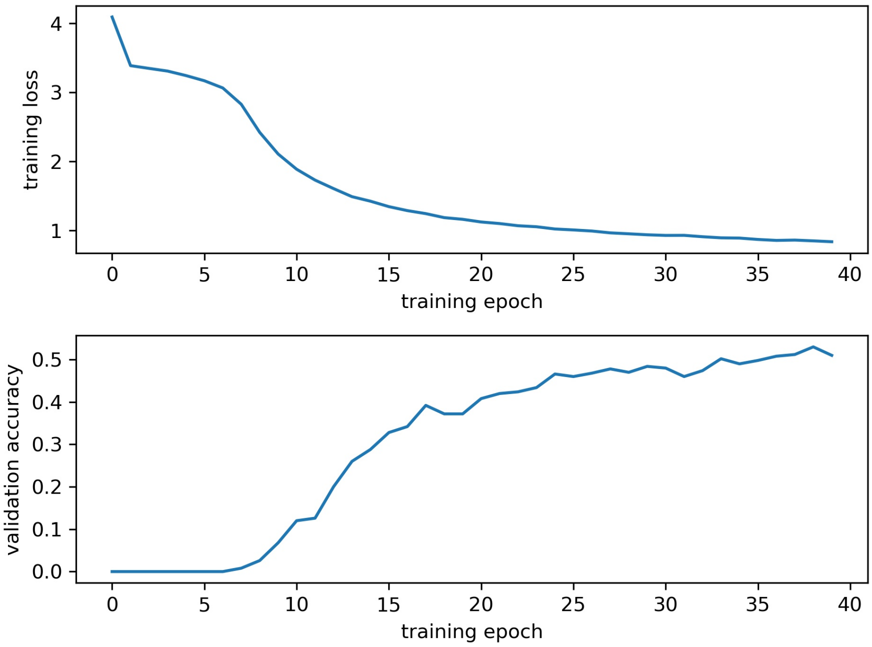

图 4: 训练集 loss 和验证集正确率

验证集最终正确率为 51.0%

```
Epoch [40/40] start ...
train loss = 0.838, validation word accuracy = 51.0%
[Info] model saved in models\model_epoch40.pth
loss and accuracy curves has been saved in loss_and_accuracy.jpg
```

图 5: 验证集最终正确率

根据结果，可能是由于数据量较小或者训练轮数不够，最终验证集正确率刚过 50%。同时，在训练 10 轮过后，loss 才下降到一个较低的水平，验证集正确率才开始快速提升，这可能是因为 10 轮训练过后模型才初步的有了字母识别的针对性。

3.3 预测新的文本图像

利用训练好的模型，输入以下图像：

图 6: 输入的图像

得到的预测结果如下：

```
PS E:\Desktop\媒体与认知\第五次作业\hw5> python main.py --mode predict --im_path data/my_own/a.png
[Info] Load model from models/model_epoch40.pth
prediction: parking
```

图 7: 预测结果

可见之前训练好的模型对这张文本图像的预测是正确的　　同时得到的可视化结果如下：

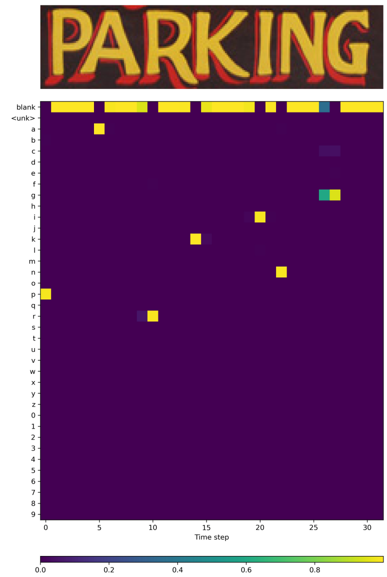

图 8: 预测图像的可视化结果

根据可视化结果，CTC 解码结果为 p–a–r–k–i–n–g–（空格数不严格，仅示意用），与课上所学的 CTC 相关知识吻合。

4 本次作业遇到的问题及解决方法

本次作业中，我遇到的主要问题是隐含哈尔科夫模型的前向变量和后向变量算法中的符号含义对应，在我仔细复习课件后才终于搞清楚各个符号的含义，从而顺利完成解答题 2.1。

5 意见及建议

在本次作业中，助教在课堂上第三小节的习题课讲解为我梳理了这两节课老师所讲的内容，在晚上的学习辅导交流活动中为我们示范了编程，这二者都对我的帮助很大，真心希望这项活动能保持下去。最后，再次感谢老师和助教的辛苦付出！