

小作业二 HMM模型

截止时间 2021 11月 8, 23:59 之前
可用 2021 10月 25 15:40 之后

得分 10

提交 一份上传文件

文件类型 pdf

作业要求

本次作业要求同学们用已训练的HMM模型对中文句子做分词。具体要求：

1. 请各位同学把句子“xxx是个好同学”用viterbi算法算出句子的分词结果，其中“xxx”替换为自己的名字。
2. 分别用前向、后向算法计算这个句子的概率。
3. 提交报告格式为pdf。

模型文件说明

- 模型文件已经上传：[hmm_parameters.pkl](#)。
- 该文件使用python的pickle库保存，同学们可以使用下面的代码读取。

```
import pickle
with open("hmm_parameters.pkl", "rb") as f:
    hmm_param = pickle.load(f)
```

- 用pickle读取数据的结果是一个字典，其中包含start_prob、trans_mat和emission_mat三个numpy矩阵，分别对应起始概率矩阵、转移概率矩阵和发射概率矩阵。
- 模型隐状态集合为{B, I}，对应状态编号0、1。
- 观测状态为“字”，可以使用python内置函数ord获取字的编号。