

实验2. 隐马尔科夫模型实践

MG1733099, 周天烁, tianshuo.zhou@smail.nju.edu.cn

2017 年 12 月 4 日

综述

隐马尔可夫模型 (Hidden Markov Model, HMM) 是统计模型, 它用来描述一个含有隐含未知参数的马尔可夫过程。其难点是从可观察的参数中确定该过程的隐含参数。然后利用这些参数来作进一步的分析, 例如语音识别, 词性标注, 时间序列分析等。一个基本的隐马尔可夫模型主要以下五个部分组成, 包括2个状态集合和3个概率矩阵:

- 1 隐含状态 S
- 2 可观测状态 O
- 3 隐含状态转移概率矩阵 A
- 4 观测状态转移概率矩阵 B
- 5 初始状态概率矩阵 π

通过指定状态空间 S 、观测空间 O 和一组概率矩阵, 就能确定一个隐马尔科夫模型, 通常用 $\lambda=[A, B, \pi]$ 来指代。给定一组观测序列, Baum Welch algorithm算法利用 EM algorithm和最大似然估计 (maximum likelihood) 来学习得到一组马尔科夫模型参数 λ 。该算法主要由Forward procedure、Backward procedure、Update三部分组成, 实验二三分别实现其中的Forward procedure和Backward procedure, 实验一则实现经典的维特比算法 (Viterbi algorithm) 来利用学习的参数进行股票的预测。

实验一.维特比算法

维特比算法采用动态规划思想, 给定隐马尔可夫模型 λ 和观测序列 O , 该算法估计出最大可能的隐藏序列 X 。具体实现参考维基百科页面伪代码。

实验二.实现Forward Algorithm

Forward Algorithm是Baum Welch algorithm中的重要一步, 用来估计给定某时刻的隐状态下最有可能的该时刻之前的观测序列。具体实现参考维基百科页面伪代码。

实验三.实现Backward Algorithm

Backward Algorithm是Baum Welch algorithm中的重要一步，和Forward Algorithm相反，该算法用来估计隐状态给的下的某时刻之后的最有可能的观测序列。具体实现参考维基百科页面伪代码。

附加任务 Do something EXTRA

利用以上实现的隐马尔科夫模型来进行自然语言处理的词性标注，具体实现参见POStag.py及其说明文件。