实验 2. 隐马尔科夫模型实践

MF1733014，高少华，cuggsh@163.com

2018 年 4 月 3 日

综述

隐马尔科夫模型 (Hidden Markov Model, HMM) 是一种有向图模型，它用来描述一个含有隐含未知参数的马尔可夫过程，在语音识别，自然语言处理等领域有广泛的应用。以下对该模型的简要说明，参考自文献 [?] 的相关章节。

图 1: 隐马尔可夫模型的图结构

如图 1 所示，在隐马尔可夫模型中，存在两组变量，其中 $\{y_1, y_2, ..., y_n\}$ 为状态变量，通常是隐藏的，不可被观测的，也称作隐变量。隐变量的取值通常是有 N 个可能取值的离散空间 $S = \{s_1, s_2, ..., s_N\}$。另一组 $\{x_1, x_2, ..., x_n\}$ 为观测变量，可被直接观测到, 其取值可以是离散的，也可以是连续的，这里假设其取值也是一个离散空间 $O = \{o_1, o_2, ..., o_M\}$。在任意时刻，观测变量的取值仅依赖于隐变量，即 x_t 由 y_t 确定，与其他的隐变量及观测变量无关。同时，t 时刻的隐变量 y_t 仅依赖于 $t-1$ 时刻的隐变量 y_{t-1}，即系统下一时刻的状态只与当前时刻的状态有关。则可得所有变量的联合概率分布为

$$P(x_1, y_1, ..., x_n, y_n) = P(y_1)P(x_1|y_1)\prod_{i=2}^{n} P(y_i|y_{i-1})P(x_i|y_i) \tag{1}$$

确定一个一个隐马尔可夫模型还需以下三组参数：

- 状态转移矩阵：模型在各个状态间转换的概率，记作 $A = [a_{ij}]_{N\times N}$, 其中 a_{ij} 表示在任意时刻从状态 s_i 转换到状态 s_j 的概率。
- 发射矩阵：模型在当前状态获得各个观测值的概率，记作 $B = [b_{ij}]_{N\times M}$, 其中 b_{ij} 表示在任意时刻从状态 s_i 获取观测值 o_j 的概率。
- 初始分布：模型在初始时刻各状态出现的概率，记作 $\pi = (\pi_1, \pi_2, ..., \pi_N)$, 其中 π_i 表示模型在初始时刻，状态为 s_i 的概率。

隐马尔可夫模型有三个基本问题：

- 已知模型参数 $\lambda = [A, B, \pi]$ 和某一特定的观测序列 $\{x_1, x_2, ..., x_n\}$，求最后时刻各个隐状态的概率分布，推测最有可能的观测值 x_{n+1}。

- 已知模型参数 $\lambda = [A, B, \pi]$ 和某一特定的观测序列，求中间时刻各个隐状态的概率分布，换言之，就是根据观测序列推测隐藏的模型状态。通常使用 viterbi 算法解决，实验一将实现该算法。
- 已知某一特定的观测序列，调整模型参数 $\lambda = [A, B, \pi]$，使得该序列出现的概率最大。即，训练模型使其能最好地描述观测数据。通常可以使用极大似然法或是 Baum-Welch 算法，来估计模型参数，在实验二、三将实现 Baum-Welch 算法的关键步骤。

实验一. viterbi 算法

viterbi 算法是一种动态规划算法，它用于寻找最有可能产生观测序列的维特比路径——隐状态序列，特别是在马尔可夫信息源上下文和隐马尔可夫模型中。该算法的伪代码描述如算法 1 所示 [?]，其中输入及输出参数的说明见综述部分，$delta[i, j]$ 保存在 i 时刻获取到观测状态 s_j 的概率的最大值，$phi[i, j]$ 用于记录路径。

Algorithm 1 viterbi 算法

Input: O, S, π, X, A, B

Procedure: Y

for each state s_i **do**
    $delta[i, 1] \leftarrow \pi \cdot B_{ix_1}$
    $phi[i, 1] \leftarrow 0$
end for
for $i \leftarrow 2, 3, ..., T$ **do**
    for each state s_j **do**
        $delta[j, i] \leftarrow \max_k (delta[k, i-1] \cdot A_{kj} \cdot Bjx_i)$
        $phi[j, i] \leftarrow arg \max_k (delta[k, i-1] \cdot A_{kj} \cdot Bjx_i)$
    end for
end for
$Y[T] \leftarrow arg \max_k (delta[k, T])$
for $i \leftarrow T-1, ..., 2$ **do**
    $Y[i] \leftarrow phi[Y[i+1], i+1]$
end for
return Y

实验二. Forward 算法

前向概率 [?]：给定隐马尔可夫模型参数 $\lambda = [A, B, \pi]$，到 t 时刻为止，观测序列为 $o_1, o_2, ..., o_t$ 且状态为 s_i 的概率，记作：

$$\alpha_t(i) = P(o_1, o_2, ..., o_t, y_t = s_i | \lambda)$$

可以递归求得前向概率 $\alpha_t(i)$，Forward 算法描述如下：

Algorithm 2 Forward 算法

Input: $O, \lambda = [A, B, \pi]$

Procedure: α

for each state b_i **do**
    $\alpha_1(i) \leftarrow \pi_i b_i(o_1)$
end for
for $t \leftarrow 2, 3, ..., T$ **do**
    for each state b_i **do**
        $\alpha_t(i) = [\sum_{j=1}^{N} \alpha_{t-1}(j) a_{ji}] b_i(o_t)$
    end for
end for
return α

实验三. Backward 算法

Algorithm 3 Backward 算法

Input: $O, \lambda = [A, B, \pi]$

Procedure: β

for each state b_i **do**
    $\beta_T(i) \leftarrow 1$
end for
for $t \leftarrow T-1, T-2, ..., 1$ **do**
    for each state b_i **do**
        $\beta_t(i) = [\sum_{j=1}^{N} \beta_{t+1}(j) a_{ji}] b_i(o_{t+1})$
    end for
end for
return β

后向概率 [?]：给定隐马尔可夫模型参数 $\lambda = [A, B, \pi]$，在 t 时刻状态为 s_i 的条件下，从 $t+1$ 时刻到 T 的部分观测序列为 $o_{t+1}, o_{t+2}, ..., o_T$ 的概率，记作：

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, ..., o_T | y_t = s_i, \lambda)$$

可以递归求得前向概率 $\beta_t(i)$，Backward 算法描述如算法 3 所示。

结果

程序运行结果如图 2 所示。

图 2: 运行结果

参考文献