实验 2. 隐马尔科夫模型实践

MG1733098，周华平，zhp@smail.nju.edu.cn

2017 年 11 月 29 日

综述

隐马尔科夫模型 (hidden Markov model, HMM) 是可用于标注问题的统计学习模型，描述由隐藏的马尔科夫链随机生成观测序列的过程，属于生成模型。

本实验首先针对一个已经训练好的 HMM，实现维特比算法，通过动态规划的思想对模型进行推断；其次，针对参数未知的 HMM，通过 Baum–Welch 算法对数据进行学习与训练，其中我们需要实现 HMM 的前向和后向算法。

实验一.

维特比算法实际是用动态规划解隐马尔科夫模型预测问题，即用动态规划求概率最大路径 (最优路径)。这时一条路径对应着一个状态序列。

根据动态规划原理，最优路径具有这样的特性：如果最优路径在时刻 t 通过节点 i_t^*，那么这一路径从节点 i_t^* 到终点 i_T^* 的部分路径，对于从 i_t^* 到 i_T^* 的所有可能的部分路径来说，必须是最优的。因为假如不是这样，那么从 i_t^* 到 i_T^* 就有另一条更好的部分路径存在，如果把它和从 i_t^* 到达 i_T^* 的部分路径连接起来，就会形成一条比原来的路径更优的路径，这是矛盾的。依据这一原理，我们只需从时刻 $t = 1$ 开始，递推地计算在时刻 t 状态为 i 的各条部分路径的最大概率，直至得到时刻 $t = T$ 状态为 i 的各条路径的最大概率。时刻 $t = T$ 的最大概率即为最优路径的概率 P^*，最优路径的终结点 i_T^* 也同时得到。之后，为了找出最优路径的各个节点，从终结点 i_T^* 开始，由后向前逐步求得节点 $i_{T-1}^*, \dots, i_1^*$，得到最优路径 $I^* = (i_1^*, i_2^*, \dots, i_T^*)$。这就是维特比算法。

首先导入两个变量 δ 和 ψ。定义在时刻 t 状态为 i 的所有单个路径 $(i_1, i_2, \dots, i_t)$ 中概率最大值为

$$\delta_t(i) = \max_{i_1, i_2, \dots, i_{t-1}} P(i_t = i, i_{t-1}, \dots, i_1, o_t, \dots, o_1|\lambda), \qquad i = 1, 2, \dots, N$$

由定义可得变量 δ 的递推公式：

$$\begin{aligned}\delta_{t+1}(i) &= \max_{i_1, i_2, \dots, i_t} P(i_{t+1} = i, i_t, \dots, i_1, o_{t+1}, \dots, o_1|\lambda) \\ &= \max_{1 \le j \le N} [\delta_t(j) a_{ji}] b_i(o_{t+1}), \qquad i = 1, 2, \dots, N; t = 1, 2, \dots, T-1\end{aligned}$$

定义在时刻 t 状态为 i 的所有单个路径 $(i_1, i_2, \ldots, i_{t-1}, i)$ 中概率最大的路径的第 $t-1$ 个节点为

$$\psi_t(i) = \underset{1 \leq j \leq N}{\operatorname{argmax}}[\delta_{t-1}(j) a_{ji}], \qquad i = 1, 2, \ldots, N$$

维特比算法

输入：模型 $\lambda = (A, B, \pi)$ 和观测 $O = (o_1, o_2, \ldots, o_T)$

输出：最优路径 $I^* = (i_1^*, i_2^*, \ldots, i_T^*)$

(1) 初始化

$$\delta_1(i) = \pi_i b_i(o_1), \qquad i = 1, 2, \ldots, N$$
$$\psi_1(i) = 0, \qquad i = 1, 2, \ldots, N$$

(2) 递推。对 $t = 2, 3, \ldots, T$

$$\delta_t(i) = \max_{1 \leq j \leq N}[\delta_{t-1}(j) a_{ji}] b_i(o_i), \qquad i = 1, 2, \ldots, N$$
$$\psi_t(i) = \underset{1 \leq j \leq N}{\operatorname{argmax}}[\delta_{t-1}(j) a_{ji}], \qquad i = 1, 2, \ldots, N$$

(3) 终止

$$P^* = \max_{1 \leq i \leq N} \delta_T(i)$$
$$i_T^* = \underset{1 \leq i \leq N}{\operatorname{argmax}}[\delta_T(i)]$$

(4) 最优路径回溯。对 $t = T-1, T-2, \ldots, 1$

$$i_t^* = \psi_{t+1}(i_{t+1}^*)$$

求得最优路径 $I^* = (i_1^*, i_2^*, \ldots, i_T^*)$。

实验二.

给定隐马尔科夫模型 λ，定义到时刻 t 部分观测序列为 $o_1, o_2, \ldots, o_t$ 且状态为 q_i 的概率为前向概率，记作

$$\alpha_t(i) = P(o_1, o_2, \ldots, o_t, i_t = q_i | \lambda)$$

可以递推地求得前向概率 $\alpha_t(i)$ 及观测序列概率 $P(O|\lambda)$。

观测序列概率的前向算法

输入：隐马尔科夫模型 λ，观测序列 O

输出：观测序列概率 $P(O|\lambda)$

(1) 初值

$$\alpha_1(i) = \pi_i b_i(o_1), \qquad i = 1, 2, \ldots, N$$

(2) 递推。对 $t=1,2,\ldots,T-1$

$$\alpha_{t+1}(i)=[\sum_{j=1}^{N}\alpha_t(j)a_{ji}]b_i(o_{t+1}),\qquad i=1,2,\ldots,N$$

(3) 终止

$$P(O|\lambda)=\sum_{i=1}^{N}\alpha_T(i)$$

前向算法中，步骤 (1) 初始化前向概率，是初始时刻的状态 $i_1=q_i$ 和观测 o_1 的联合概率。步骤 (2) 是前向概率的递推公式，计算到时刻 $t+1$ 部分观测序列为 $o_1,o_2,\ldots,o_t,o_{t+1}$ 且在时刻 $t+1$ 处于状态 q_i 的前向概率。步骤 (3) 给出 $P(O|\lambda)$ 的计算公式，因为

$$\alpha_T(i)=P(o_1,o_2,\ldots,o_t,i_T=q_i|\lambda)$$

所以

$$P(O|\lambda)=\sum_{i=1}^{N}\alpha_T(i)$$

实验三.

给定隐马尔科夫模型 λ，定义在时刻 t 状态为 q_i 的条件下，从 $t+1$ 到 T 的部分观测序列为 $o_{t+1},o_{t+2},\ldots,o_T$ 的概率为后向概率，记作

$$\beta_t(i)=P(o_{t+1},o_{t+2},\ldots,o_T|i_t=q_i,\lambda)$$

可以用递推的方法求得后向概率 $\beta_t(i)$ 及观测序列概率 $P(O|\lambda)$。

观测序列概率的后向算法

输入：隐马尔科夫模型 λ，观测序列 O

输出：观测序列概率 $P(O|\lambda)$

(1) 初值

$$\beta_T(i)=1,\qquad i=1,2,\ldots,N$$

(2) 递推。对 $t=T-1,T-2,\ldots,1$

$$\beta_t(i)=\sum_{j=1}^{N}a_{ij}b_j(o_{t+1})\beta_{t+1}(j),\qquad i=1,2,\ldots,N$$

(3) 终止

$$P(O|\lambda)=\sum_{i=1}^{N}\pi_ib_i(o_1)\beta_1(i)$$

步骤 (1) 初始化后向概率，对最终时刻的所有状态 q_i 规定 $\beta_T(i)=1$。步骤 (2) 是后巷概率的递推公式。为了计算在时刻 t 状态为 q_i 条件下时刻 $t+1$ 之后的观测序列为 $o_{t+1},o_{t+2},\ldots,o_t$ 的后向概率 $\beta_t(i)$，只需考虑在时刻 $t+1$ 所有可能的 N 个状态 q_j 的转移概率，以及在此状态下的观测 o_{t+1} 的观测概率，然后考虑状态 q_j 之后的观测序列的后向概率。步骤 (3) 求 $P(O|\lambda)$ 的思路与步骤 (2) 一致，只是初始概率 π_i 代替转移概率。

参考文献

[1] 周志华. 机器学习. 清华大学出版社, 2016.

[2] 李航 et al. 统计学习方法. 清华大学出版社, 北京, 2012.