

隐式马尔科夫模型

严春伟

2012 年 12 月 28 日

1 前言

隐马尔科夫模型 (hidden Markov model, HMM) 是一种统计学习模型，可以用于标注问题。隐马尔科夫模型在序列化的数据的处理中有比较成熟的应用，如：

- 语音识别
- 自然语言处理
- 生物信息
- 模式识别

本文将会论述隐马尔科夫的基本问题，如定义，解决问题，基本算法等，此外，本文还会论述 HMM 在著名的中文分词库 ICTCLAS 中的应用原理。

2 基本概念

2.1 定义

隐马尔科夫模型 (HMM) 描述一个隐藏的马尔科夫链生成不可观察的状态 (state) 随机序列，在由各个状态生成一个观测而产生观测随机序列的过程。

HMM 的确定因素：

1. 初始状态分布 π
2. 状态转移概率分布 A
3. 观测概率 B

隐马尔科夫的要素

1. N , 表示模型中的状态数目
2. M , 表示模型中每个状态不同的观察符号
3. A , 状态转移概率分布。 $A = \{a_{ij}\}$

$$a_{ij} = P(q_t = S_j | q_{t-1} = S_i) \quad (1)$$

4. B, 观察字符在状态 j 时的概率分布, $B = \{b_j(k)\}$, 其中

$$b_j(k) = P(v_k | q_t = S_j) \quad 1 \leq j \leq N, 1 \leq k \leq M \quad (2)$$