



第9章：序列标注

课件制作：曲彦儒、吴宪泽、朱耀明、屠可伟
讲解人：曲彦儒、吴宪泽、朱耀明

问题定义

□ 已知:

- 标签集合 $Y = \{y^1, y^2, \dots, y^L\}$

□ 输入:

- 文本 $x = \{x_1, x_2, \dots, x_n\}$

□ 输出:

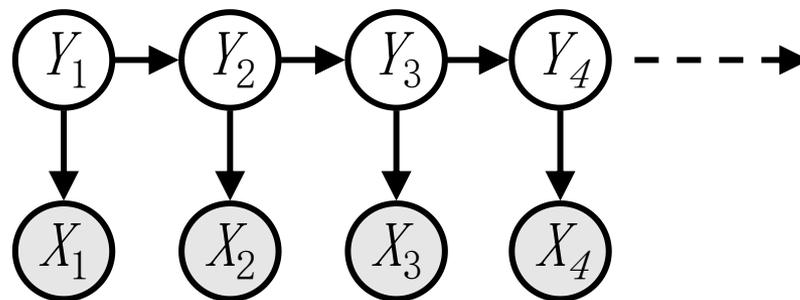
- 对 x 中每个单词 x_i 的一个预测标签 $y_i \in C$

9.2 隐马尔可夫模型

隐马尔可夫模型 (Hidden Markov Model, HMM)

图模型表示

- 状态 (Y) 代表标签
- 观测 (X) 代表单词, 每个观测对应一个状态

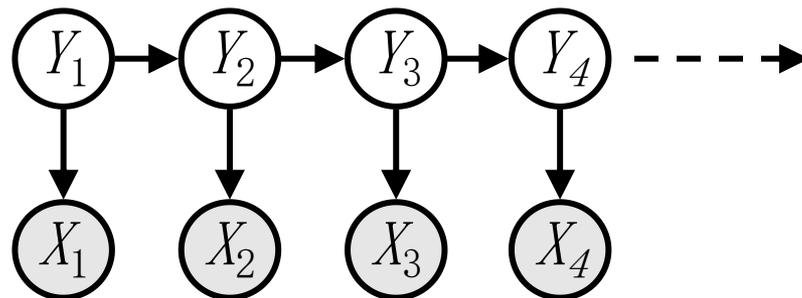


参数

- 转移 (Transition) 模型
 $P(y_t|y_{t-1})$: $N \times N$ 维矩阵
- 观测 (Emission) 模型
 $P(x_t|y_t)$: $N \times M$ 维矩阵
- 先验概率 $P(y_1)$: N 维向量, 可看作是从默认起始状 $Y_0=START$ 到 Y_1 的转移概率
- 另可假设最后一个状态 Y_n 会转移到默认终止状态 $Y_{n+1}=STOP$

9.2 隐马尔可夫模型

隐马尔可夫模型



- HMM建模了输入序列与标签序列的联合概率分布

$$P(x_1, \dots, x_n, y_1, \dots, y_{n+1}) = P(y_{n+1}|y_n) \prod_{t=1:n} P(y_t|y_{t-1}) P(x_t|y_t)$$

隐马尔科夫模型的解码

HMM解码

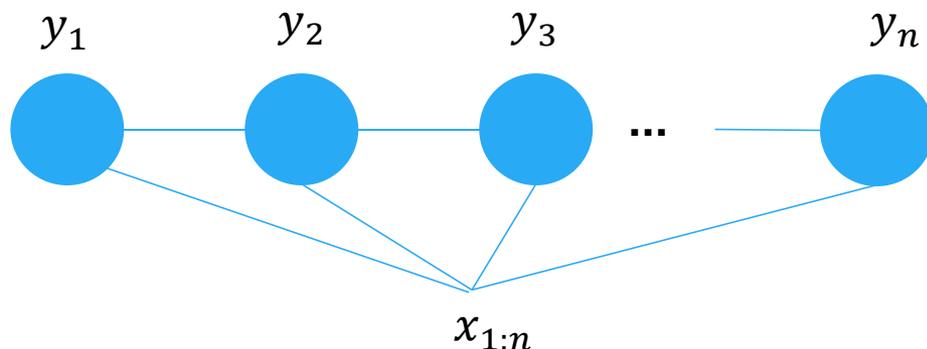
- 给定输入序列，找到概率最高的标签序列
- 理论上，只要枚举所有可能的标签序列，为每个标签序列计算联合概率，最后选择联合概率最高的标签序列就可以解码， $O(L^n)$
- 维特比（Viterbi）算法：动态规划， $O(nL^2)$

隐马尔可夫模型的学习

HMM学习

- 隐马尔可夫模型可以采用监督学习或无监督学习
- 监督学习：利用最大似然估计，最大化输入序列和标签序列的联合概率
- 无监督学习：利用最大期望值法来最大化输入序列的边际概率，该算法也被称作Baum-Welch算法

9.3 条件随机场



- 与隐马尔可夫模型类似，条件随机场（conditional random field, CRF）也会建模相邻标签之间的关系
- 条件随机场是一个无向图模型，使用打分函数来评价任意两个标签相邻的合理性
- 条件随机场的结构如上图所示

9.3 条件随机场

- 解码：与隐马尔可夫模型类似，条件随机场同样需要使用维特比算法进行动态规划求解
- 监督学习：条件随机场监督学习的一个常见优化目标是 minimized 人工标注的标签序列的负对数条件似然

9.4 神经序列标注模型

□ 神经softmax:

- 也被称为最大熵模型。它将输入序列通过一个循环神经网络或是Transformer编码器，得到每个输入元素的向量表示，随后输入一个线性分类器，得到对应的标签预测

□ 神经条件随机场:

- 将神经softmax模型中独立预测每个标签的线性分类器替换为条件随机场，由此得到神经条件随机场 (neural conditional random field)

总结

- 序列标注：
 - 问题定义
 - 隐马尔可夫模型
 - 条件随机场
 - 神经序列标注模型