

- 1、自然语言处理常用模型种类
- 2、马尔科夫模型
 - 2.1特点
 - 2.2状态初始化概率和状态转移概率
 - 2.3参数估计-极大似然
- 3、隐马尔科夫模型
 - 3.1观测状态与隐状态
 - 3.2隐状态初始化概率、隐状态转移概率和观测状态生成概率
 - 3.3三个基本问题
 - 给定模型参数，如何计算一个观测状态序列的概率？
 - 给定模型参数和观测状态序列，如何而计算一个最可能的隐状态序列？
 - 给定观测序列集合，如何从中估计模型参数？
- 4、EM算法估计模型参数
 - EM算法介绍

1、自然语言处理常用模型种类

生成式模型： 根据山羊的特征，来算出这只羊是山羊的概率有多大；再根据绵羊的特征，算出这只羊是绵羊的概率为多大。再判断在哪个里面的概率更大，就属于哪一种。如Naive Bayes、Hidden Markov Model、LDA

判别式模型： 直接得到判别边界。比如说要判断一只羊是山羊还是绵羊，直接根据这只羊的特征，分别求出其是山羊和是绵羊的概率，哪个大就属于哪一种。如Linear Regression、SVM、KNN

非线性模型： SVM的非线性核函数、Neural Network

2、马尔科夫模型

2.1特点

马尔科夫模型属于典型的生成式模型，适用于处理时间序列数据。

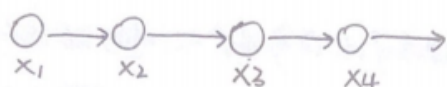
马尔科夫链的主要性质是“无记忆性”，即下一状态的概率分布只能由当前状态决定。

2.2状态初始化概率和状态转移概率

马尔可夫模型的训练参数就是状态初始化概率分布和状态转移概率分布。

对于一阶马尔科夫链，给出如下例子，给定第一天的天气情况，用来判断接下来几天的天气情况：

一阶马尔科夫链(=马尔可夫模型)：



$S = \{s_1, \dots, s_N\} \Rightarrow$ 状态集合.

$X = x_1, \dots, x_t, \dots, x_T \Rightarrow$ 观测状态序列.

以晴天、雨天、阴天三种天气情况为例，由上一天的天气情况对后一天的天气情况进行预测。 $N=3$.

定义状态初始化概率： $\pi_i = P(X_1 = s_i), 1 \leq i \leq N$.

↓

非负性，归一性

定义状态转移概率： $a_{ij} = P(X_t = s_j | X_{t-1} = s_i), 1 \leq i, j \leq N$

($t-1$ 时期状态为 i ， t 时期状态为 j)

现给出初始化概率：

晴	阴	雨
0.4	0.3	0.3

$\pi_1 = 0.4$

$\pi_2 = 0.3$

$\pi_3 = 0.3$

状态转移概率分布：

	晴	阴	雨
晴	0.4	0.5	0.1
阴	0.3	0.4	0.3
雨	0.1	0.5	0.4

连续四天下雨的概率： $P(\text{rainy, rainy, rainy, rainy}, \theta)$ θ : 模型参数.

$$= P(X_1 = s_3, X_2 = s_3, X_3 = s_3, X_4 = s_3; \theta)$$

$$= P(X_1 = s_3) \times P(X_2 = s_3 | X_1 = s_3) \times P(X_3 = s_3 | X_2 = s_3) \times P(X_4 = s_3 | X_3 = s_3)$$

$$= \pi_3 \times a_{33} \times a_{33} \times a_{33} = 0.3 \times 0.4 \times 0.4 \times 0.4 = 0.0192.$$

https://blog.csdn.net/wpxin_42515907

一阶马尔科夫性只能表示当前变量于前一个变量的关系，为了表达当前变量与更早变量的关系，需要引入高阶马尔科夫性。

2.3 参数估计-极大似然

此外，在这个模型里，状态初始化概率和状态转移概率是需要学习的参数，在现实情况中，每天的天气变化会产生各种不同的观测数据，利用这些历史数据来自动估计模型参数，需要用到的方法是**极大似然估计**。

使用马尔科夫模型计算观测状态的概率：

X 是观测状态序列， $x = x_1, \dots, x_t, \dots, x_T$ ， θ 是模型参数。

$$P(x; \theta) = \prod_{t=1}^T P(x_t | x_1, \dots, x_{t-1}) \approx P(x_1) \times \prod_{t=2}^T P(x_t | x_{t-1})$$

使用极大似然估计从观测数据中自动获取最优模型参数：

$\theta = \operatorname{argmax}_{\theta} \{L(\theta)\}$ 。有 D 个样本的观测集 $D = \{x^{(d)}\}_{d=1}^D$ ，第 d 个样本。

$$\text{其中 } L(\theta) = \sum_{d=1}^D \log P(x^{(d)}; \theta) = \sum_{d=1}^D \left(\log P(x_1^{(d)}) + \sum_{t=2}^{T^{(d)}} \log P(x_t^{(d)} | x_{t-1}^{(d)}) \right)$$

(对数似然函数) (D 个样本的似然函数之和)

其中，模型参数 θ 需要满足两个条件： $\sum_{x \in S} P(x) = 1$ ， $P(x) = \sum_{x' \in S} P(x | x') = 1$ 。

↓

↓

某个时期下所有状态概率之和为 1； x 之后时期各状态概率之和也为 1。

Obviously. 这是求一个约束条件下的极值，需要用拉格朗日算法。

$$J(\theta, \lambda, r) = L(\theta) - \lambda \left(\sum_{x \in S} P(x) - 1 \right) - \sum_{x \in S} r_x \left(\sum_{x' \in S} P(x | x') - 1 \right), r_x \text{ 与状态 } x \text{ 相对应}.$$

$$\textcircled{1} \frac{\partial J(\theta, \lambda, r)}{\partial P(x)} = \frac{\partial \sum_{d=1}^D \log P(x_1^{(d)})}{\partial P(x)} - \lambda \quad (P(x) \text{ 是初始状态概率，} \lambda \text{ 是与 } P(x) \text{ 相关的拉格朗日乘子}).$$

$$\textcircled{2} \frac{\partial J(\theta, \lambda, r)}{\partial P(x | x')} = \frac{\partial \sum_{d=1}^D \sum_{t=2}^{T^{(d)}} \log P(x_t^{(d)} | x_{t-1}^{(d)})}{\partial P(x | x')} - r_x = \frac{1}{P(x | x')} \sum_{d=1}^D \sum_{t=2}^{T^{(d)}} \delta(x_{t-1}^{(d)}, x) \delta(x_t^{(d)}, x') - r_x$$

$$\text{记 } C(x, x', D) = \sum_{d=1}^D \sum_{t=2}^{T^{(d)}} \delta(x_{t-1}^{(d)}, x) \delta(x_t^{(d)}, x')$$

(在 D 个样本中， x' 跟在 x 之后出现的次数)

$$\text{上面的 } \textcircled{1} \text{ 可进一步简化为 } \frac{\partial J(\theta, \lambda, r)}{\partial P(x)} = \frac{1}{P(x)} \sum_{x'} C(x, x', D) - \lambda. \quad \delta(a, b) = \begin{cases} 1, & a=b \\ 0, & a \neq b \end{cases}$$

$$\text{记 } C(x, D) = \sum_{d=1}^D \sum_{t=2}^{T^{(d)}} \delta(x_{t-1}^{(d)}, x).$$

$$\Rightarrow \frac{\partial J(\theta, \lambda, r)}{\partial \lambda} = \sum_{x \in S} P(x) - 1. \quad \frac{\partial J(\theta, \lambda, r)}{\partial P(x)} = 0 \Rightarrow P(x) = \frac{C(x, D)}{\lambda}. \quad \Rightarrow \lambda = \sum_{x \in S} C(x, D)$$

$$\frac{\partial J(\theta, \lambda, r)}{\partial \lambda} = 0 \Rightarrow \sum_{x \in S} P(x) = 1. \quad \boxed{\text{状态初始化概率公式}} \Rightarrow P(x) = \frac{C(x, D)}{\sum_{x \in S} C(x, D)}$$

$$\Rightarrow \frac{\partial J(\theta, \lambda, r)}{\partial r_x} = \sum_{x' \in S} P(x | x') - 1. \quad \sum_{x' \in S} P(x | x') = 1. \quad \Rightarrow P(x | x') = \frac{C(x, x', D)}{\sum_{x' \in S} C(x, x', D)} \quad \boxed{\text{状态转移概率公式}}$$

log.csdn.net/weixin_42515907

于是，进一步说明。

给观测序列 $\mathcal{O} = \{x^{(d)}\}_{d=1}^D$ 。

状态初始化概率的估计公式为： $\pi_i = \frac{\sum_{d=1}^D \delta(x_1^{(d)}, s_i)}{D}$

状态转移概率的公式为： $a_{ij} = \frac{\sum_{d=1}^D \sum_{t=2}^{T^{(d)}} \delta(x_{t-1}^{(d)}, s_i) \delta(x_t^{(d)}, s_j)}{\sum_{d=1}^D \sum_{t=2}^{T^{(d)}} \delta(x_{t-1}^{(d)}, s_i)}$

以三个关于天气情况的观测序列作为依据来得到模型的状态初始化概率分布和状态转移概率分布。

①. 晴, 阴, 雨

②. 晴, 阴, 晴.

③. 雨, 阴.

} 3个历史观测序列.

=>

晴	阴	雨
2/3	0	1/3

状态初始化概率分布

=>

	晴	阴	雨
晴	0/2	2/2	0/2
阴	1/2	0/2	1/2
雨	0/1	1/1	0/1

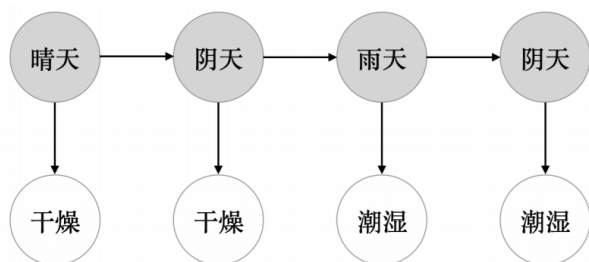
状态转移概率分布.

3、隐马尔科夫模型

3.1观测状态与隐状态

天气的阴晴可以通过地面的潮湿程度或者是空气的湿润程度等其他因素来预测，这些借以预测的因素被我们称之为是观测状态，而实际的天气情况则是隐状态。

下图左边是一条隐马尔科夫链的传导过程，右图是一个**观测状态生成概率分布**。



	干燥	潮湿
晴天	0.8	0.2
阴天	0.6	0.4
雨天	0.3	0.7

这种情况下引入**隐马尔科夫模型**

3.2 隐状态初始化概率、隐状态转移概率和观测状态生成概率

隐马尔可夫模型是一个典型的概率图模型，在隐马尔可夫模型中，有三个主要的参数，分别表示为：

隐马尔可夫模型：

$G = \{0, \dots, m\}$ ：观测状态集合。

$S = \{s_1, \dots, s_N\}$ ：隐状态集合。

$x = x_1, \dots, x_t, \dots, x_T$ ：观测状态序列。

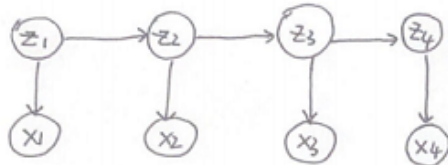
$z = z_1, \dots, z_t, \dots, z_T$ ：隐状态序列。

二阶马尔可夫模型表示为：

$$P(x, z) = \sum_{\theta} P(x, z, \theta)$$

$$= \sum_{\theta} P(z_1) \times P(x_1 | z_1) \times \prod_{t=2}^T P(z_t | z_{t-1}) \times P(x_t | z_t)$$

$$\theta \begin{cases} P(z_1) | z_1 \in S, \text{ 隐状态初始化概率. } \pi_i = P(z_1 = s_i), 1 \leq i \leq N \\ P(z_t | z_{t-1}) | z_t, z_{t-1} \in S, \text{ 隐状态转移概率. } a_{ij} = P(z_t = s_j | z_{t-1} = s_i), 1 \leq i, j \leq N \\ P(x_t | z_t) | x_t \in G, z_t \in S, \text{ 观测状态生成概率. } b_{jk} = P(x_t = o_k | z_t = s_j), 1 \leq j \leq N, 1 \leq k \leq m. \end{cases}$$



概率图模型。

https://blog.csdn.net/weixin_42515907

3.3 三个基本问题

隐马尔可夫模型有三个主要的问题，包括如下：

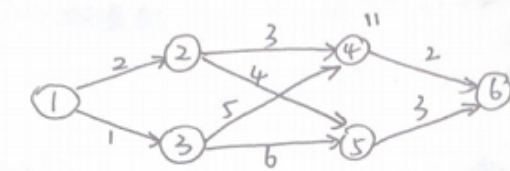
给定模型参数，如何计算一个观测状态序列的概率？

全枚举和动态规划

由于一个隐状态序列可能伴随着多个观测状态序列，随着传导过程的延申，可能会出现指数爆炸的情况。

由一个状态到另一个状态之间有多种不同的路径，计算观测状态序列概率就等价于计算所有路径之和。这个过程可以通过全枚举和动态规划算法来实现。

全枚举和动态规划的算法如下：



$1 \rightarrow 2 \rightarrow 4 \rightarrow 6$ 12
 $1 \rightarrow 2 \rightarrow 5 \rightarrow 6$ 24
 $1 \rightarrow 3 \rightarrow 4 \rightarrow 6$ 10
 $1 \rightarrow 3 \rightarrow 5 \rightarrow 6$ 18

(全枚举)

i	1	2	3	4	5	6
α_i	1	2	1	11	14	64

(动态规划)

$$\alpha_6 = w_{46} \alpha_4 + w_{56} \alpha_5$$

https://blog.csdn.net/weixin_42515907

前向概率计算和维特比算法

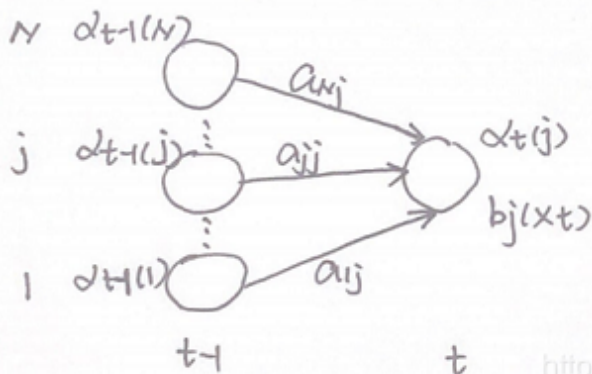
求出总的路径之和的过程就是一个计算前向概率的过程，对前向概率的定义如下：

观测状态 x_1, \dots, x_t 与第 t 个隐状态 S_i 的联合概率称为前向概率。

$$\alpha_t(i) = \pi_i b_i(x_1), 1 \leq i \leq N. \quad \dots \text{初始化}$$

$$\alpha_t(j) = \left(\sum_{i=1}^N \alpha_{t-1}(i) a_{ij} \right) b_j(x_t), 1 \leq j \leq N \quad \dots \text{递归}$$

$$P(x|\theta) = \sum_{i=1}^N \alpha_T(i) \quad \dots \text{终止}$$



https://blog.csdn.net/weixin_42515907

现在给出模型参数，要求计算观测状态序列“干燥、潮湿、干燥”的前向概率。

晴天	阴天	雨天
0.4	0.3	0.3

表6: 隐状态初始化概率分布

	晴天	阴天	雨天
晴天	0.4	0.5	0.1
阴天	0.3	0.4	0.3
雨天	0.1	0.5	0.4

表7: 隐状态转移概率分布

	干燥	潮湿
晴天	0.8	0.2
阴天	0.6	0.4
雨天	0.3	0.7

表5: 观测状态生成概率分布

计算过程如下:

$$\begin{aligned}
 P(X_1|Z_1) &= P(Z_1) \cdot P(X_1|Z_1) = 0.4 \times 0.8 = 0.32 \\
 P(X_1|Z_2) &= P(Z_2) \cdot P(X_1|Z_2) = 0.3 \times 0.6 = 0.18 \\
 P(X_1|Z_3) &= P(Z_3) \cdot P(X_1|Z_3) = 0.3 \times 0.3 = 0.09
 \end{aligned}$$

$P(X_1|Z_1)$ 表示第1时期隐状态为晴天时，观测状态为干燥的概率。

$$P(X_1, \theta) = \sum_z P(X_1, z_1, \theta)$$

$$\begin{aligned}
 P(X_2 = \text{潮湿}, Z_2 = \text{晴}, \theta) &= \sum_z P(Z_1) \cdot P(X_1|Z_1) \times P(Z_2|Z_1) \cdot P(X_2|Z_2) \\
 &= 0.4 \times 0.8 \times 0.4 \times 0.2 + 0.3 \times 0.6 \times 0.3 \times 0.2 + 0.3 \times 0.3 \times 0.1 \times 0.2 = 0.0382
 \end{aligned}$$

$$\begin{aligned}
 P(X_2 = \text{潮湿}, Z_2 = \text{阴}, \theta) &= \sum_z P(Z_1) \cdot P(X_1|Z_1) \times P(Z_2|Z_1) \cdot P(X_2|Z_2) \\
 &= 0.4 \times 0.8 \times 0.5 \times 0.4 + 0.3 \times 0.6 \times 0.4 \times 0.4 + 0.3 \times 0.3 \times 0.5 \times 0.4 = 0.1108
 \end{aligned}$$

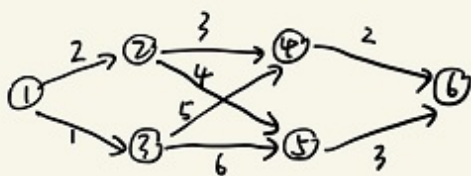
$$P(X_2 = \text{潮湿}, Z_2 = \text{雨}, \theta) = \dots = 0.0854$$

$$\begin{aligned}
 P(X_2 = \text{干燥}, Z_2 = \text{晴}, \theta) &= \sum_z P(Z_1) \cdot P(X_1|Z_1) \cdot P(Z_2|Z_1) \cdot P(X_2|Z_2) \cdot P(Z_3|Z_2) \cdot P(X_3|Z_3) \\
 &= 0.0382 \times 0.4 \times 0.8 + 0.1108 \times 0.3 \times 0.8 + 0.0854 \times 0.1 \times 0.8 = 0.0456
 \end{aligned}$$

$$\begin{aligned}
 P(X_2 = \text{干燥}, Z_2 = \text{阴}, \theta) &= \sum_z P(Z_1) \cdot P(X_1|Z_1) \cdot P(Z_2|Z_1) \cdot P(X_2|Z_2) \cdot P(Z_3|Z_2) \cdot P(X_3|Z_3) \\
 &= 0.0382 \times 0.5 \times 0.6 + 0.1108 \times 0.4 \times 0.6 + 0.0854 \times 0.5 \times 0.6 = 0.0637
 \end{aligned}$$

$$P(X_2 = \text{干燥}, Z_2 = \text{雨}, \theta) = \dots = 0.0214$$

$$\Rightarrow P(X, \theta) = \sum_z P(X, z, \theta) = 0.0456 + 0.0637 + 0.0214 = 0.1307 \quad (\text{干燥} \rightarrow \text{潮湿} \rightarrow \text{干燥})$$



\Rightarrow 最长路径长为 24.

最大路径为 $1 \rightarrow 2 \rightarrow 5 \rightarrow 6$

i	1	2	3	4	5	6
g_i	1	2	1	6	8	24
u_i	0	1	1	2	2	5

https://blog.csdn.net/weixin_42515907

维特比算法和前向概率计算方法一致，就是说维特比算法计算的是前向传播的最大路径概率即为最终概率，而前向概率计算方法需要找到最大路径之和为最终的概率。

给定模型参数和观测状态序列，如何而计算一个最可能的隐状态序列？

计算所有的隐状态序列可能性(前向预测方法和维特比算法)，然后从中挑选出一个概率最大的，即为最可能的隐状态序列。

给定观测序列集合，如何从中估计模型参数？

EM算法

4、EM算法估计模型参数

EM算法介绍

$$\hat{\theta} = \arg \max_{\theta} \{L(\theta)\}. \quad p(x, \theta) = \sum_z p(x, z; \theta).$$

\nearrow 观测状态.
 \nearrow 隐状态.

$$L(\theta) = \sum_{d=1}^D \log p(x^{(d)}; \theta) = \sum_{d=1}^D \log \sum_z p(x, z; \theta).$$

(由于隐状态序列是指数级的, 不同的隐状态序列可能产生相同的观测状态。) \Rightarrow EM算法提出

EM: (Expectation maximization) $\left\{ \begin{array}{l} x: \text{一组观测数据} \\ z: \text{未观测数据} \\ p(x, z; \theta): \text{观测数据与未观测数据的联合分布} \\ \theta: \text{模型参数} \end{array} \right.$

EM算法 $\left\{ \begin{array}{l} E \text{步: 确定对数似然的期望值: } Q(\theta | \theta^{old}) = E_{z|x, \theta^{old}} [\log p(x, z; \theta)] \\ M \text{步: 计算最大化该期望值的参数: } \theta^{new} = \arg \max_{\theta} \{Q(\theta | \theta^{old})\} \end{array} \right.$