Natural Language Process

分词和HMM

方法及原理介绍

目录

1 分词的定义

- 词是自然语言中最小的有意义的构成单位。汉语文本是基于单字的，汉语的书面表达方式也是以汉字作为最小单位的，词与词之间没有显性的界限标志，因此分词是汉语文本分析处理中首先要解决的问题之一。
- 对于大多数汉语处理系统来讲，第一步要进行的就是识别这些隐含的词语边界，即添加合适的显性的词语边界标志使得所形成的词串反映句子的本意。这个过程就是通常所说的分词。

2 分词的难点

2.1 如何识别未登录词

- 常见的未登录词有如下几类：1)专有名词，包括中文人名(如“朱镕基总理”)、地名(如“綦江县”)、机构名称(如“杭州娃哈哈集团公司”)、外国译名(如“克鲁普总统”)、时间词（如“2022年5月26日”）；2)重叠词，如“高高兴兴”、“研究研究”；3)派生词，如“一次性用品”；4)与领域相关的术语，如“高斯分布”。
- 一个鲁棒性的系统必须能很好的识别未登录词，但是到目前为止，现有的方法都很难完美。

2.2 如何廉价的获取语言学知识

- 一方面，目前还没有一个可以利用的大规模的汉语分词语料，而人工加工一个大规模的分词语料是一种耗费很大的工作；
- 另一方面，任一汉字对间都可能是一个词语边界，而且分词直接面对的是词，参数空间巨大，特别是高阶模型，目前还没有适用于分词的完全有效的无监督的参数学习方法

2 分词的难点

2.3 词语边界歧义

- 词语边界歧义指的是对于一个给定的汉语句子或汉字串，有多种词语边界划分形式。汉语词语边界歧义包括组合歧义和交叉歧义。
- 组合歧义是不同的组合方式。如句子“以/我/个人/的/名义/”和“他/一/个/人/在家/”的“个人”是一个组合歧义字段。
- **交叉歧义还可细分为真歧义和伪歧义**。真歧义指存在两种或两种以上的可实现的切分形式，如句子“必须/加强/企业/中/国有/资产/的/管理/”和“中国/有/能力/解决/香港/问题/”中的字段“中国有”是一种真歧义；而伪歧义一般只有一种正确的切分形式，如“建设/有”、“中国/人民”、“各/地方”、“本/地区”等。
- 在这些歧义中，伪歧义字段的切分结果是上下文无关的，一般仅依据字段内部的信息如词频或字间互信息就可正确切分伪歧义字段，而真歧义字段或组合歧义字段的结果依赖于它所处的上下文环境，因而正确处理真歧义字段，常常需要更多的信息，特别是上下文信息。

2 分词的难点

- 2.4 分词效率问题

- 大多数分词系统只注重识别准确率，而忽视了识别速度。有些应用系统，如机助翻译系统，其实时性能要求较高，要求分析算法对输入句子能做出迅速准确的处理。

- 对于给定的输入句子，其可能的切分词串数量与句子长度成指数关系，因为在理论上句子中的任何一个汉字串都可以成为一个词。已被证明，最坏情况下的穷举搜索算法实际并不可行。贪心算法虽然能避免组合爆炸，但它不能保证输出结果最佳。可见，识别算法的效率在实时性应用系统中地位非常重要。

3 分词的方法

3.1 正向最大匹配分词

- 最大正向匹配（FMM）的基本思想是：假设自动分词词典中的最长词条所含汉字个数为I，则取被处理材料当前字符串序数中的I个字作为匹配字段，查找分词词典。若词典中有这样的一个I字词，则匹配成功，匹配字段作为一个词被切分出来；如果词典中找不到这样的一个I字词，则匹配失败。匹配字段去掉最后一个汉字，剩下的字符作为新的匹配字段，进行新的匹配，如此进行下去，直至切分成功为止。即完成一轮匹配切分出一个词，然后再按上面的步骤进行下去，直到切分出所有词为止。

- 例如现有短语“计算机科学和工程”，假设词典中最长词为7字词，于是先取“计算机科学和工”为匹配字段，来查找分词词典以匹配这个字段，由于词典中没有该词，故匹配失败，去掉最后一个汉字成为“计算机科学和”作为新的匹配字段，重新匹配词典，同样匹配失败，取“计算机科学”作为新的匹配字段，来匹配词典，由于词典中有“计算机科学”一词，从而匹配成功，切分出第一个词“计算机科学”。同样的方法还可以切分出后续的词。

- 这种方法进行分词的时候，对上文提到的交叉歧义和组合歧义没有什么好的办法。因为对组合歧义才说，通常他都会作为一个分词单位，如“市场中国有企业才能发展”这个例句中，按照正向最大匹配分词方法，切分方法为“市场/中国/有/企业/才能/发展/” 我们可以看到，在这个例句中，有两个分词错误，分别为“中国/有”这个交叉歧义和“才能”这个分词的组合歧义。

3 分词的方法

3.2 反向最大匹配分词

- 与FMM方法相对应的方法是反向最大匹配分词方法，也称为BMM方法。它的分词过程与FMM方法相同，不过是从句子(或文章)末尾开始处理，每次匹配不成功时去掉的是前面的一个汉字。

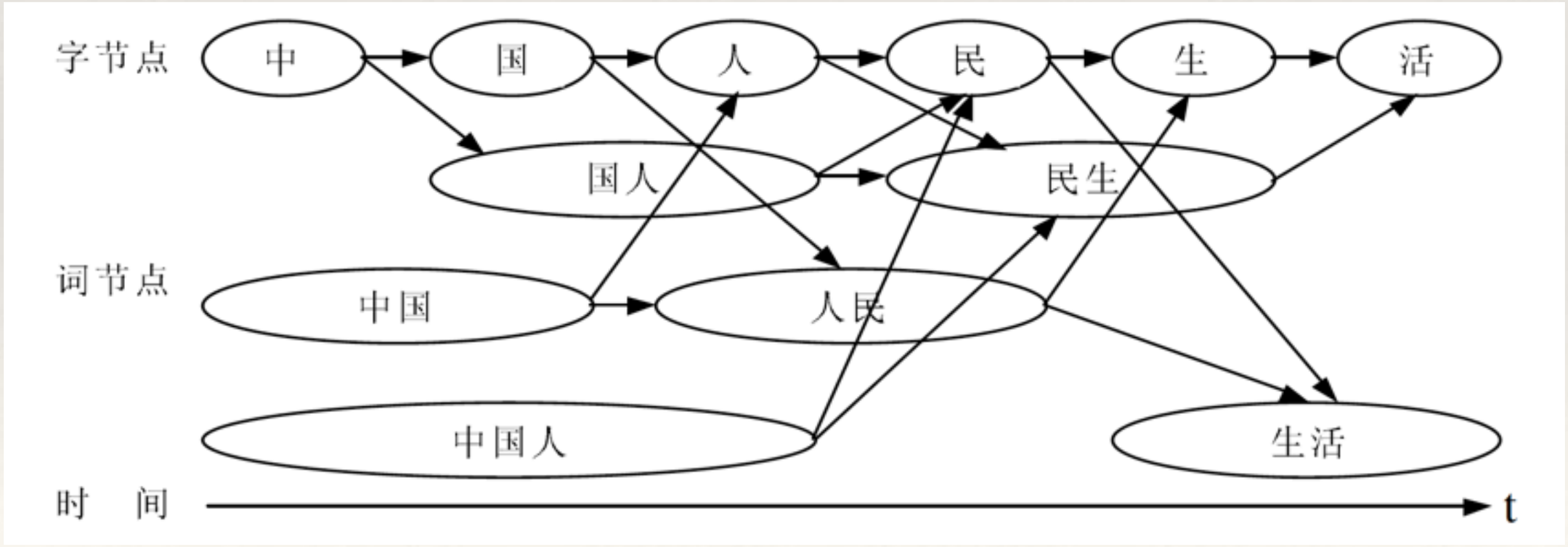
- 如“计算机科学和工程”，首先取“计算机科学和工程”作为匹配字段来匹配分词词典，由于词典中没有该词，故匹配失败。去掉最前面的一个汉字，即取“算机科学和工程”作为新的匹配字段，进行匹配，同样的匹配失败，……，最后，取“工程”作为匹配字段，来匹配分词词典，由于分词词典中有“工程”一词，则匹配成功，切分出第一个词“工程”。

- 例如上面方法中“市场中国有企业才能发展”这个例句中，按照反向最大匹配分词方法，切分方法为“市场中/国有/企业/才能/发展/” 我们可以看到，在这个例句中，有一个分词错误，那就是才能这个分词的组合歧义，应该切分为 “才/能”才可以。

3 分词的方法

- 3.3 基于统计的词网分词

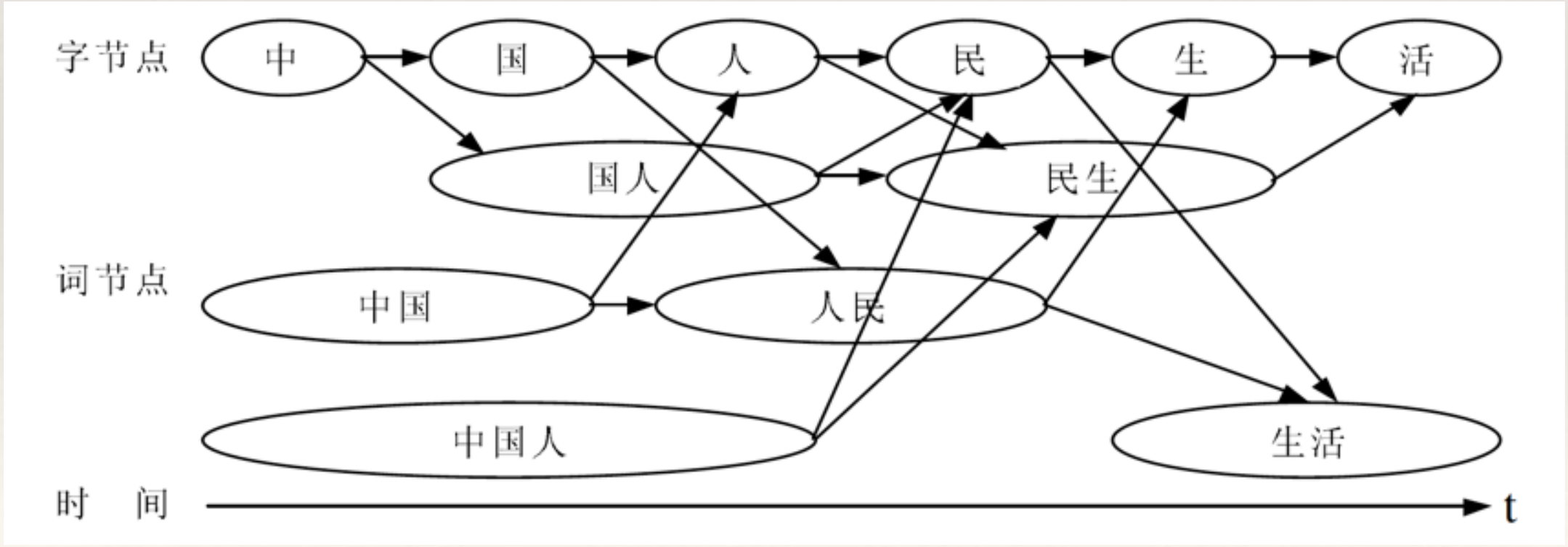
- 与基于词网格分词的第一步是候选词网格构造：利用词典匹配，列举输入句子所有可能的切分词语，并以词网格形式保存。实际上，词网格是一个有向无环图，它蕴含了输入句子所有可能的切分，其中的每一条路径代表一种切分。如下图“中国人民生活“的切分词网：



3 分词的方法

- 3.3 基于统计的词网分词

- 词网分词的第二步是计算词网格中的每一条路径的权值，权值通过计算图中每一个节点（每一个词）的一元统计概率和节点之间的二元统计概率的相关信息。然后根据图搜索算法在图中找到一条权值最小的路径，对应的路径即为最后的分词结果。



4 马尔可夫模型

- 马尔可夫模型描述
- 存在一类重要的随机过程：如果一个系统有N个状态S_1, S_2，…，S_N, 随时间的推移，该系统从某一状态转移到另一状态。系统在时间t 的状态记为q_t。系统在时间t 处于状态S_j(1≤j≤N) 的概率取决于其在时间1, 2, …, t-1 的状态，该概率为：

$$P(q_t = S_j \mid q_{t-1} = S_i, q_{t-2} = S_k, \ldots)$$

4 马尔可夫模型

- 假设1
- 如果在特定情况下，系统在时间t的状态只与时间t-1的的状态相关，则该系统构成一个离散的一阶马尔可夫链:

$$P(q_t = S_j \mid q_{t-1} = S_i, q_{t-2} = S_k, \ldots) =$$
$$P(q_t = S_j \mid q_{t-1} = S_i) \quad \ldots \qquad (4.1)$$

4 马尔可夫模型

- 假设2
- 如果只考虑公式（4.1）独立于时间t的随机过程，即不动性假设，状态与时间无关，那么：

$$P(q_t = S_j \mid q_{t-1} = S_i) = a_{ij}, \quad 1 \le i, j \le \mathrm{N} \quad \ldots \qquad (4.2)$$

- 该随机过程称为马尔可夫模型。其中a_{ij}表示转移概率。

4 马尔可夫模型

- 在马尔可夫模型中，状态转移概率a_{ij}必须满足下面的要求:

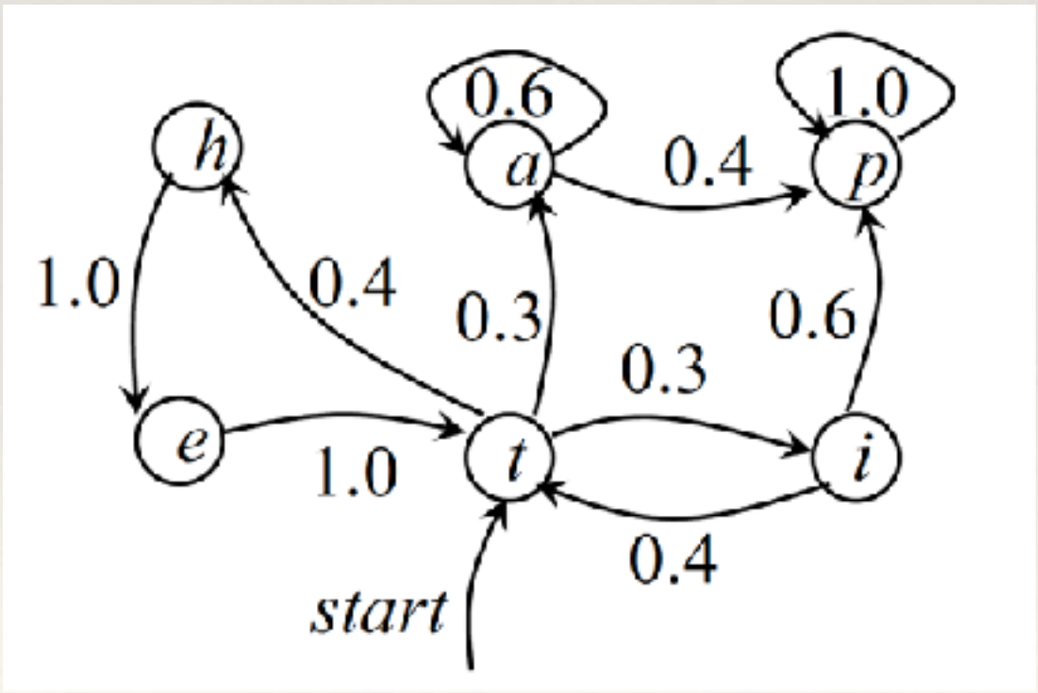
$$a_{ij} \geq 0$$

$$\sum_{j=1}^{N} a_{ij} = 1 \quad \ldots \qquad (4.3)$$

- 马尔可夫模型可以视为随机有限状态自动机，该有限状态自动机的每一个状态转换过程都有一个相应的概率，该概率表示自动机采用这一状态转换的可能性。

4 马尔可夫模型

- 马尔可夫模型链可以表示成状态图（即转移弧上有概率的非确定的有限状态自动机）。
- 零概率的转移弧省略。
- 每个节点上发出弧的概率之和为1



4 马尔可夫模型

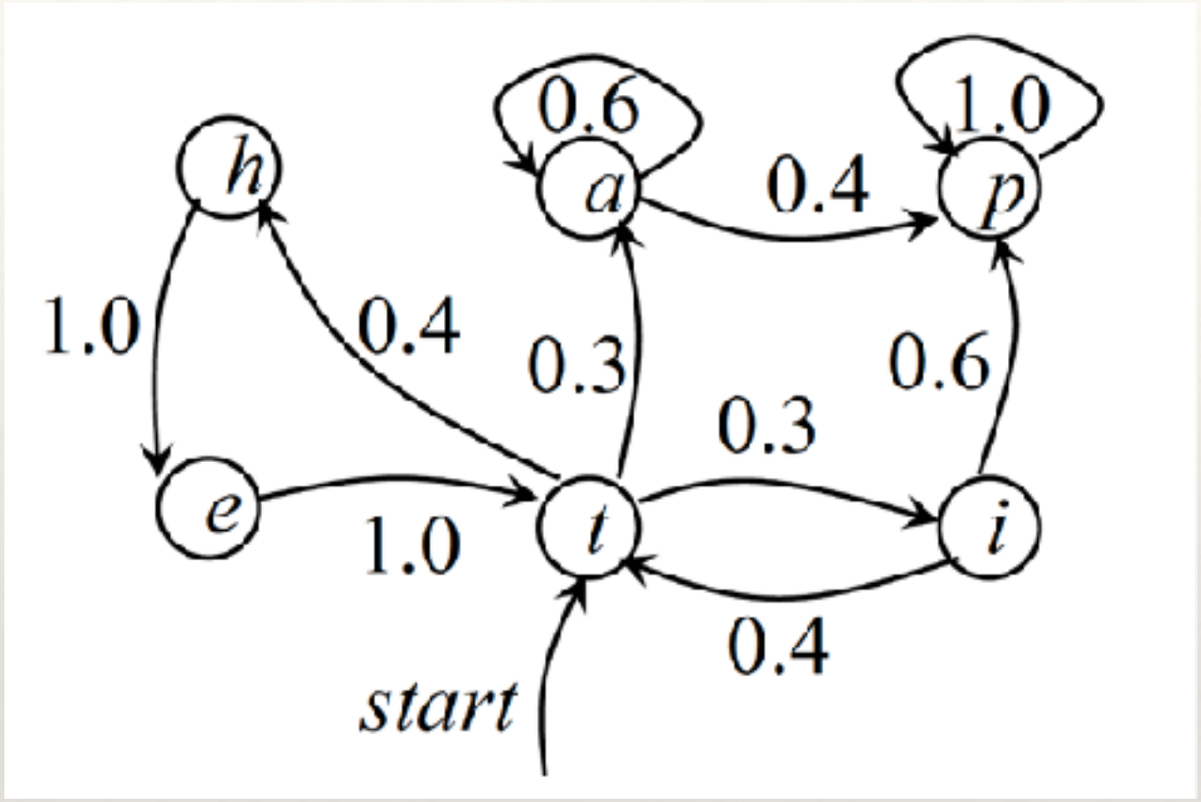
- 状态序列$S_1,\ldots,S_T$的概率：

$$P(S_1,\ldots,S_T)=P(S_1)P(S_2\mid S_1)\ldots P(S_T\mid S_1,S_2,\ldots,S_{T-1})$$

$$=P(S_1)P(S_2\mid S_1)\ldots P(S_T\mid S_{T-1})$$

$$=\pi_{S_1}\prod_{t=1}^{T-1}a_{S_tS_{t+1}}\ldots\ldots(4.4)$$

4 马尔可夫模型



$$P(t,i,p) = P(S_1 = t)P(S_2 = i \mid S_1 = t)P(S_3 = p \mid S_2 = i)$$
$$= 1.0 \times 0.3 \times 0.6$$
$$= 0.18$$

5 隐马尔可夫模型-描述

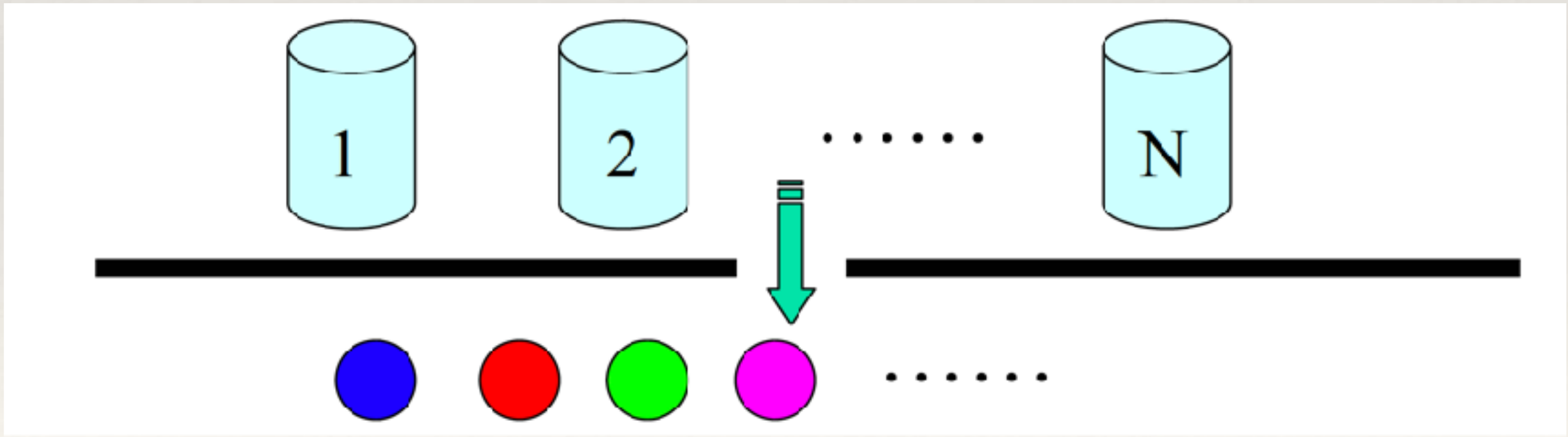
5.1 描述

- 隐马尔可夫模型是一个双重随机过程，我们不知道具体的状态序列，只知道状态转移的概率，即模型的状态转换过程是不可观察的（隐藏的），而可观察的事件的随机过程是隐藏的状态转换过程的随机函数。

5 隐马尔可夫模型-举例

- 5.2 举例

- 假设有N 个袋子，每个袋子中有M 种不同颜色的球。实验员根据某一概率分布选择一个袋子，然后根据袋子中不同颜色球的概率分布随机取出一个球，并报告该球的颜色。对局外人：可观察的过程是不同颜色球的序列，而袋子的序列是不可观察的。每只袋子对应HMM 中的状态；球的颜色对应于HMM 中的状态的输出。



5 隐马尔可夫模型-组成

- **5.3 组成**
- (1) 模型中的状态数为N (袋子的数量)
- (2) 从每一个状态可能输出的不同的符号数M(不同颜色球的数目
- (3) 状态转移矩阵：状态转移矩阵A=a_{ij}（a_{ij}为实验员从一只袋子转向另外一只袋子取球的概率）。

$$\begin{cases} a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i), \\ a_{ij} \geq 0, \\ \sum_{j=1}^{N} a_{ij} = 1 \end{cases} \qquad (5.1)$$

5 隐马尔可夫模型-组成

- 5.3 组成

- (4) 生成概率矩阵。从状态S_j观察到某一特定符号v_k的概率分布矩阵为：$B=b_j(k)$。其中$b_j(k)$为实验员从第j个袋子中取出第k种颜色的球的概率。那么：

$$\begin{cases} b_j(k) = P(O_t = v_k \mid q_t = S_i), \\ b_j(k) \geq 0 \\ \sum_{k=1}^{M} b_j(k) = 1 \end{cases} \qquad (5.2)$$

5 隐马尔可夫模型-组成

- 5.3 组成

- (5) 初始状态的概率分布：初始状态的概率分布 $\pi = \pi_i$，其中，

$$\begin{cases} \pi_i = P(q_1 = S_i), 1 \le i \le N \\ \pi_i \ge 0 \\ \sum_{i=1}^{N} \pi_i = 1 \end{cases} \tag{5.3}$$

为了方便，一般将HMM 记为：$\mu(A,B,\pi)$ 用以指出模型的参数集合。

5 隐马尔可夫模型-观察序列生成

- **5.4 观察序列的生成**
- 给定模型 $\mu(A,B,\pi)$，观察序列 $O=O_1,\ldots,O_T$产生的步骤如下：
- （1）初始化t为1
- （2）根据初始状态概率分布 $\pi=\pi_i$选择一初始状态$q_1=S_i$；
- （3）根据状态S_i 的生成概率分布$b_i(k)$, 输出$O_t=v_k$；
- （4）t = t+1, 如果t < T, 重复步骤(3) (4)，否则结束。

5 隐马尔可夫模型-三个问题

- 5.5 三个问题

- 问题1：在给定模型 $\mu(A,B,\pi)$ 和观察序列 $O=O_1,\ldots,O_T$ 的情况下，怎样快速计算 $p(O|\mu)$ ？

- 问题2：在给定模型 $\mu(A,B,\pi)$ 和观察序列 $O=O_1,\ldots,O_T$ 的情况下，如何选择在一定意义下“最优”的状态序 $Q=q_1,\ldots,q_T$，使得该状态序列“最好地解释”观察序列。

- 问题3：给定一个观察序列 $O=O_1,\ldots,O_T$，如何根据最大似然估计来求模型的参数值?即如何调节模型 $\mu(A,B,\pi)$ 的参数，使得 $p(O|\mu)$ 最大?

5 隐马尔可夫模型-问题1

- 5.6 前向算法
- （1）在给定模型 $\mu(A,B,\pi)$ 和观察序列 $O=O_1,\ldots,O_T$ 的情况下，快速计算 $p(O\mid\mu)$
- 对于给定的状态序列 $Q=q_1,\ldots,q_T$：

$$P(O\mid\mu)=\sum_Q P(O,Q\mid\mu)=\sum_Q P(Q\mid\mu)P(O\mid Q,\mu)...(5.4)$$

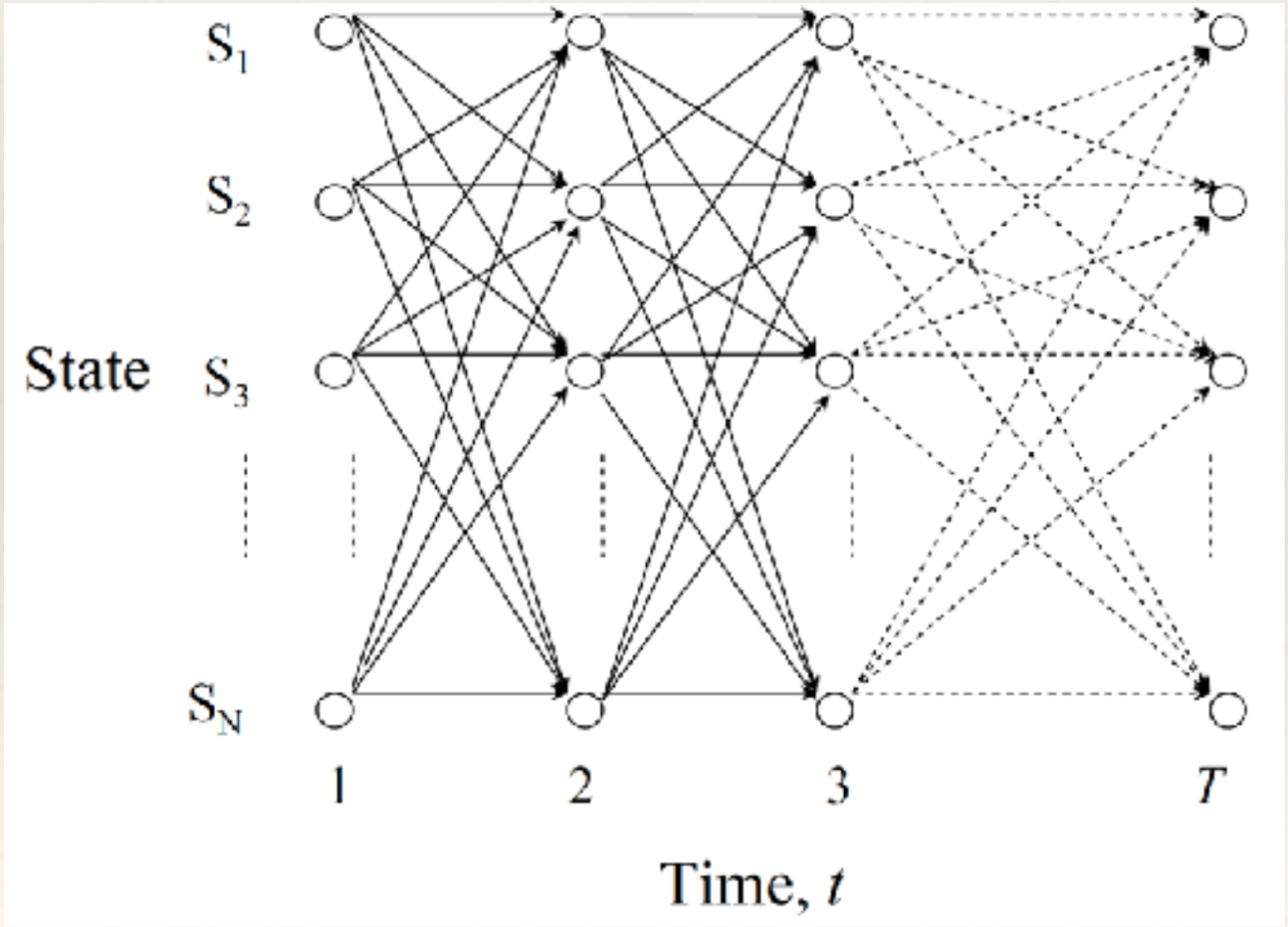
$$P(Q\mid\mu)=\pi_{q_1}a_{q_1q_2}a_{q_2q_3}...a_{q_{T-1}q_T}...(5.5)$$

$$P(O\mid Q,\mu)=b_{q_1}(O_1)b_{q_2}(O_2)...b_{q_T}(O_T)...(5.6)$$

5 隐马尔可夫模型-问题1

- 5.6 前向算法

- 如果模型 $\mu(A,B,\pi)$ 有N个不同的状态，时间长度为T，那么有 N^T 个可能的状态序列，搜索路径为指数级。



5 隐马尔可夫模型-问题1

- **5.6 前向算法**
- 解决办法：动态规划，前向算法。
- 基本思想：定义前向变量

$$\alpha_t(i) = P(O_1,\ldots,O_t,q_t = S_i \mid \mu)...(5.7)$$

- 如果能够高效地计算公式（5.7），那么就可以高效的计算 $p(O \mid \mu)$，因为 $p(O \mid \mu)$ 是在所有状态 q_T 下，观察到序列 $O = O_1,\ldots,O_T$ 的概率。

$$P(O \mid \mu) = \sum_{S_i} P(O_1 O_2 \ldots O_T, q_T = S_i \mid \mu)$$

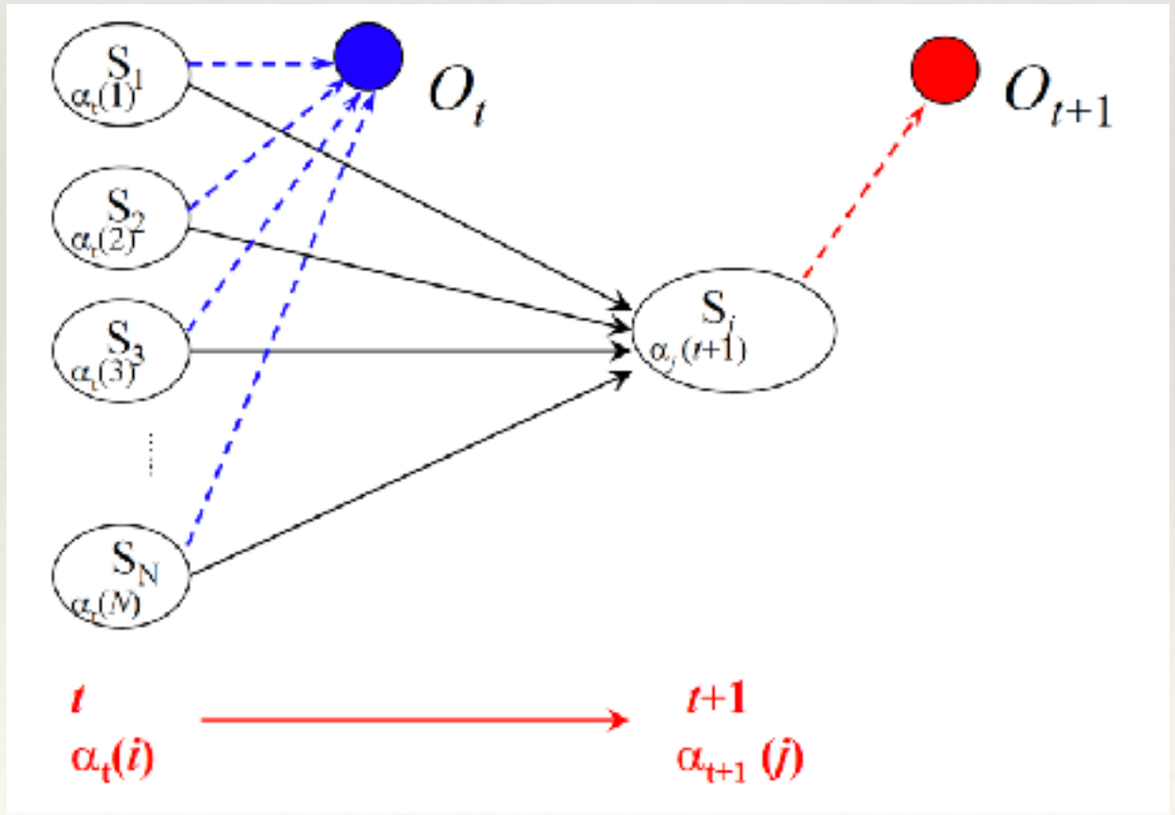
$$= \sum_{i=1}^{N} \alpha_T(i) \cdots (5.8)$$

5 隐马尔可夫模型-问题1

- 5.6 前向算法

- 利用动态规划计算 $\alpha_t(i)$： 在时间t+1 的向前变量可以根据时间t 的向前变量的值递归计算，即：

$$\alpha_{t+1}(j)=[\sum_{i=1}^{N}\alpha_t(i)a_{ij}]b_j(O_{t+1})...(5.9)$$



5 隐马尔可夫模型-问题1

- **5.6 前向算法**
- 前向算法的过程：
- （1）初始化：$\alpha_1(i)=\pi_i b_i(O_1),1\le i\le N$
- （2）计算 $\alpha_{t+1}(j)=[\sum_{i=1}^{N}\alpha_t(i)a_{ij}]b_j(O_{t+1}),1\le t\le T-1$
- （3）结束，输出 $P(O\mid\mu)=\sum_{i=1}^{N}\alpha_T(i)$

5 隐马尔可夫模型-问题1

- 5.6 前向算法
- 算法的时间复杂度：
- 计算每计算一个 $\alpha_t(i)$ 必须考虑从t-1 时的所有N 个状态转移到状态Si 的可能性，时间复杂度为O(N)，对应每个时刻t，要计算N 个向前变量：$\alpha_t(1)\alpha_t(2)...\alpha_t(N)$ ，所以时间复杂度为 $O(N)\times N = O(N^2)$ 又因t = 1, 2, …, T，所以向前算法总的复杂度为：$O(N^2T)$

5 隐马尔可夫模型-问题1

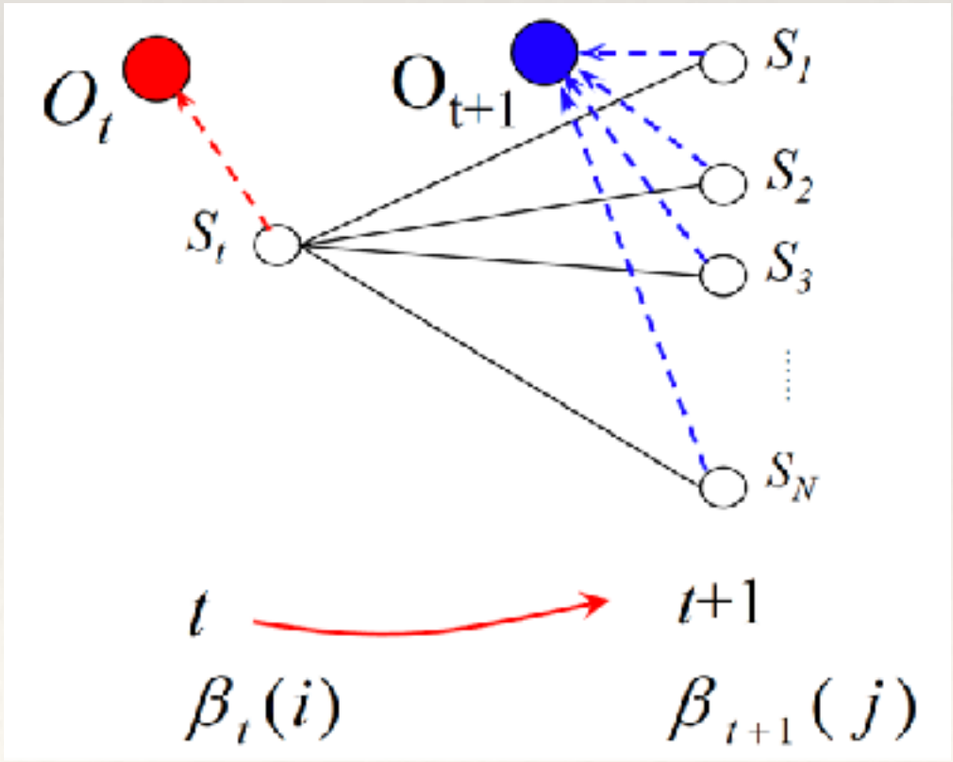
- 5.7 后向算法

- 定义后向变量 $\beta_t(i)$ 是在给定模型 $\mu(A,B,\pi)$ 和假定t状态为S_i的的条件下，模型输出观察序列 $O=O_1,\ldots,O_T$ 的概率：

$$\beta_t(i)=P(O_{t+1}O_{t+2}\ldots O_T \mid q_t=S_i,\mu)\ldots(5.10)$$

5 隐马尔可夫模型-问题1

- 5.7 后向算法
- 和前向变量的计算一样，可以使用动态规划来求解后向变量。
- 第一步的概率为：$a_{ij} \times b_j(O_{t+1})$
- 第二步的概率按后向变量的定义为：$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) ...(5.11)$



5 隐马尔可夫模型-问题1

- 5.7 后向算法
- 后向算法的过程：
- （1）初始化：$\beta_T(i)=1,1\le i\le N$
- （2）计算 $\beta_t(i)=\sum_{j=1}^{N}a_{ij}b_j(O_{t+1})\beta_{t+1}(j),T-1\ge t\ge 1,1\le i\le N$
- （3）结束，输出 $P(O\mid\mu)=\sum_{i=1}^{N}\pi_i\beta_1(i)$
- 整个算法的时间复杂度为 $O(N^2T)$

5 隐马尔可夫模型-问题2

- 问题2，如何发现“最优”状态序列
- 解释不唯一，关键是如何理解“最优”的状态序列?
- 第一种解释是：状态序列中的每个状态都单独地具有概率，即：对于每个t（$1 \le t \le T$），寻找q_t使得 $\gamma_t(i) = P(q_t = S_i \mid O, \mu)$ 最大。

$$\gamma_t(i) = P(q_t = S_i \mid O, \mu) = \frac{P(q_t = S_i, O \mid \mu)}{P(O \mid \mu)} \ldots(5.12)$$

- 公式（5.12）的分子表示HMM 的输出序列O，并且在时间t 到达状态i 的概率。

5 隐马尔可夫模型-问题2

- **问题2，如何发现“最优”状态序列**
- 分解过程：
- （1）HMM 在时间t 到达状态i, 并且输出 $O=O_1,\ldots,O_t$ 。根据前向变量的定义，实现这一步的概率为 $\alpha_t(i)$。
- （2）从时间t，状态Si 出发，HMM 输出 $O=O_{t+1},\ldots,O_T$，根据向后变量定义，实现这一步的概率 $\beta_t(i)$ 。于是：

$$P(q_t=S_i,O\mid\mu)=\alpha_t(i)\times\beta_t(i)...(5.13)$$

5 隐马尔可夫模型-问题2

- **问题2，如何发现“最优”状态序列**
- 公式（5.12）中的分母于时间t无关，所以分母：

$$P(O|\mu)=\sum_{i=1}^{N}\alpha_i(i)\times\beta_t(i)...(5.14)$$

$$\gamma_t(i)=\frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N}\alpha_t(i)\times\beta_t(i)}...(5.15)$$

- t时刻的最优状态为： $\tilde{q}_t=\arg\max_{1\le i\le N}(\gamma_t(i))$
- 但是，每一个状态单独最优不一定使整体的状态序列最优，可能两个最优的状态和之间的转移概率为0。

5 隐马尔可夫模型-问题2

- 问题2，如何发现“最优”状态序列
- 第二种解释：在给定模型和观察序列的条件下求概率最大的状态序列：

$$\tilde{Q} = \arg\max_Q P(Q \mid O, \mu) ...(5.16)$$

- 可以使用Viterbi算法动态搜索最优状态序列。

- Viterbi算法的定义：Viterbi 变量$\delta_t(i)$是在时间t 时，HMM 沿着某一条路径到达Si，并输出观察序列$O = O_1, \ldots, O_t$ 的最大概率：

$$\delta_t(i) = \max_{q_1, q_2, \ldots, q_{t-1}} P(q_1, q_2, \ldots, q_t = S_i, O_1 O_2 \ldots O_t \mid \mu) ...(5.17)$$

5 隐马尔可夫模型-问题2

- 问题2，如何发现“最优”状态序列
- 递归计算 $\delta_{t+1}(i)=[\max_j \delta_t(j)a_{ji}]b_i(O_{t+1})...(5.18)$
- （1）初始化：$\delta_1(i)=\pi_i b_i(O_1),1\le i\le N$，概率的最大的路径变量为$\psi_1(i)=0$
- （2）递归计算：

$$\delta_t(j)=\max_{1\le i\le N}[\delta_{t-1}(i)a_{ij}]b_j(O_t),2\le t\le T,1\le j\le N$$

$$\psi_t(i)=\arg\max_{1\le i\le N}[\delta_{t-1}(i)a_{ij}]b_j(O_t),2\le t\le T,1\le i\le N$$

- （3）结束

$$\tilde{Q}_T=\arg\max_{1\le i\le N}[\delta_T(i)]$$

$$\tilde{P}(\tilde{Q}_T)=\max_{1\le i\le N}\delta_T(i)$$

- （4）通过回溯得到路径（状态序列）：

$$\tilde{q}_t=\psi_{t+1}(\tilde{q}_{t+1}),t=T-1,T-2,...,1$$

5 隐马尔可夫模型-问题3

- 问题3，参数学习
- 递归计算，给定一个观察序列O，如何根据最大似然估计来求模型的参数值？即如何调节模型μ=(A, B, π) 的参数，使得P(O | μ) 最大？即估计模型中的π_i, a_{ij}, $b_j(k)$使得观察序列O 的概率P(O | μ) 最大。
- 如果产生观察序列O 的状态Q 已知，可以用最大似然估计来计算HMM 的参数：

$$\overline{\pi}_i = \delta(q_1, S_i)$$

$$\overline{a}_{ij} = \frac{\sum_{t=1}^{T-1} \delta(q_t, S_i) \times \delta(q_{t+1}, S_j)}{\sum_{t=1}^{T-1} \delta(q_t, S_i)}$$

- 上面的公式中，分母表示Q中从状态qi转移到状态qj的次数。分子表示Q中从状态qi转移到其它状态的总次数。其中，δ(x, y) 为克罗奈克(Kronecker)函数，当x= y 时，δ(x, y)=1,否则δ(x, y) = 0。

5 隐马尔可夫模型-问题3

- 问题3，参数学习
- 根据同样的原理，可以得到：

$$b_j(k)=\frac{\sum_{t=1}^{T}\delta(q_t,S_j)\times\delta(O_t,v_k)}{\sum_{t=1}^{T-1}\delta(q_t,S_j)}\quad\ldots(5.19)$$

- 上面的公式中，分母表示Q中从状态q_j输出符号v_k的次数，分子表示从状态q_j输出任意符号的总次数。

5 隐马尔可夫模型-问题3

- 问题3，参数学习
- EM算法
- 初始化时随机地给模型的参数赋值(遵循限制规则，如：从某一状态出发的转移概率总和为1)，得到模型μ_0，然后可以从μ_0 得到从某一状态转移到另一状态的期望次数，然后以期望次数代替公式(5.19)中的实际次数，便可得到模型参数的新估计，由此得到新的模型μ_1，从μ_1 又可得到模型中隐变量的期望值，由此可重新估计模型参数。循环这一过程，参数收敛于最大似然估计值。

5 隐马尔可夫模型-问题3

- 问题3，参数学习

- EM算法给定HMM 模型μ和观察序列O，那么，在时间t 位于状态Si，时间t+1 位于状态Sj 的概率：

$$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j \mid O, \mu)$$

$$= \frac{P(q_t = S_i, q_{t+1} = S_j, O \mid \mu)}{P(O \mid \mu)}$$

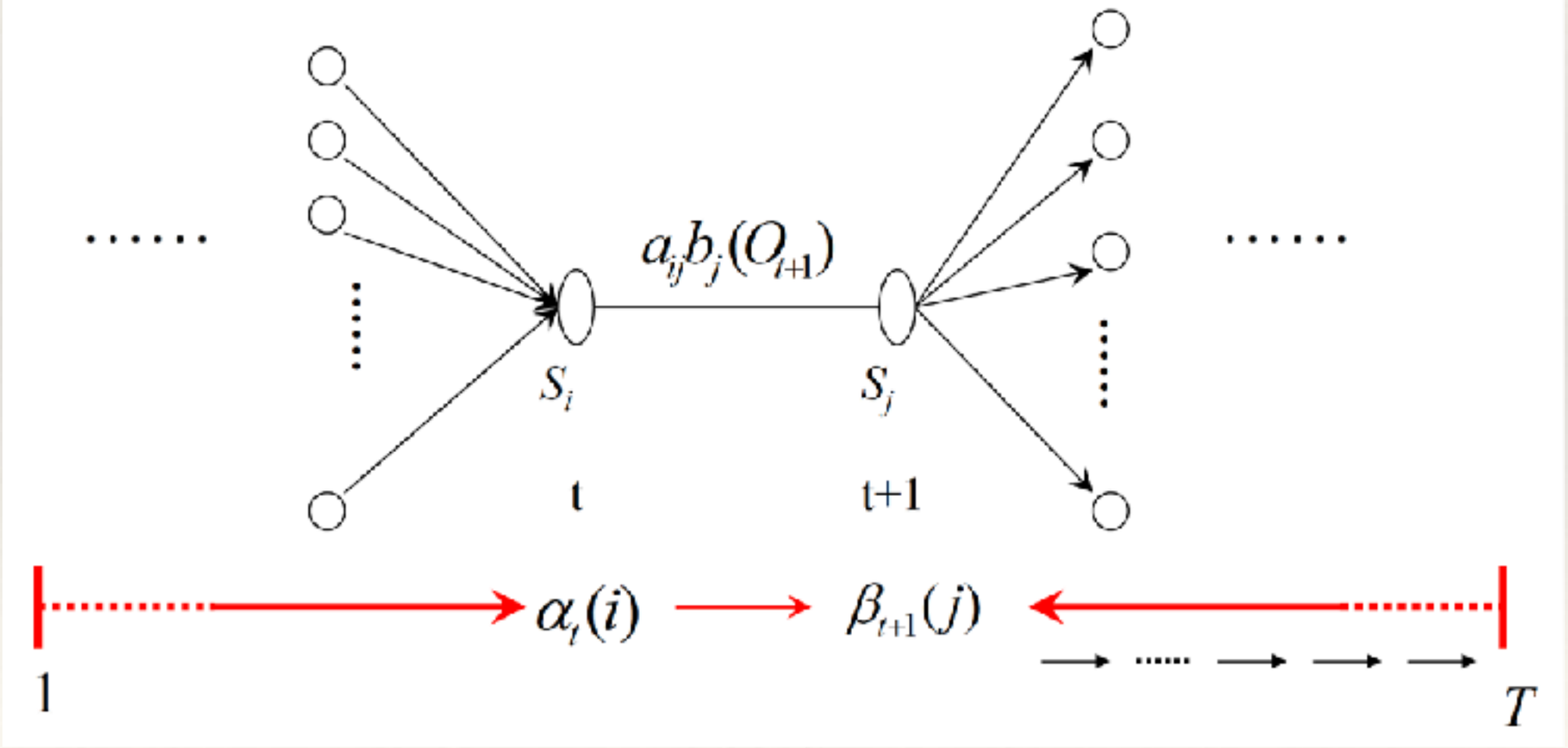
$$= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O \mid \mu)}$$

$$= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \quad ...(5.18)$$

5 隐马尔可夫模型-问题3

- 问题3，参数学习

- EM算法给定HMM 模型μ和观察序列O，那么，在时间t 位于状态S_i，时间t+1 位于状态S_j 的概率：



5 隐马尔可夫模型-问题3

- 问题3，参数学习
- 那么，给定模型μ 和观察序列，在时间t 位于状态Si 的概率为：$\gamma_t(i)=\sum_{j=1}^{N}\xi_i(i,j)$ 由此，模型μ 的参数可由下面的公式重新估计：
- （1）q1 为Si 的概率：$\pi_i=\gamma_1(i)$
- （2）$\overline{a}_{ij}=\dfrac{\sum_{i=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)}$
- （3）$\overline{b}_j(k)=\dfrac{\sum_{t=1}^{T}\gamma_t(j)\times\delta(O_t,v_k)}{\sum_{t=1}^{T}\gamma_t(j)}$

5 隐马尔可夫模型-问题3

- 问题3，参数学习
- Baum-Welch算法：
- （1）初始化：随机地给，a，b (k) 赋值，使得

$$\left\{\begin{array}{l}\sum_{i=1}^{N}\pi_i=1\\ \sum_{j=1}^{N}a_{ij}=1,1\le i\le N\\ \sum_{k=1}^{M}b_i(k)=1,1\le i\le N\end{array}\right\}$$

- （2）EM算法更新参数

6 总结

- 1 分词
- 2 MM
- 3 HMM
- 4 POS Tagging
- 5 Deep Learning