中文分词-02

OutLine

马尔可夫模型

隐马尔可夫模型（HMM）

中文分词-02

马尔科夫模型

- 每个状态只依赖之前有限个状态
  - N阶马尔科夫：依赖之前n个状态
  - 1阶马尔科夫：仅仅依赖前一个状态
    - p(w1,w2,w3,……,wn) = p(w1)p(w2|w1)p(w3|w1,w2)……p(wn|w1,w2,……,wn-1)
    - =p(w1)p(w2|w1)p(w3|w2)……p(wn|wn-1)
    - 例如：
    - p(w1=今天，w2=我，w3=写，w4=了，w5=一个，w6=程序)
    - =p(w1=今天)p(w2=我|w1=今天)p(w3=写|w2=我)……p(w6=程序|w5=一个)

马尔科夫模型

$$\xrightarrow{\pi_{S_1}} S_1 \xrightarrow{a_{S_1,S_2}} S_2 \xrightarrow{a_{S_2,S_3}} S_3 \xrightarrow{a_{S_3,S_4}} \cdots\cdots \xrightarrow{a_{S_{T-1},S_T}} S_T$$

- 参数
  - 状态，由数字表示，假设共有M个
  - 初始概率，由π_k表示

$$\pi_k = P(S_1 = k) \quad k = 1, 2, \ldots, M$$

  - 状态转移概率，由$a_{k,l}$表示

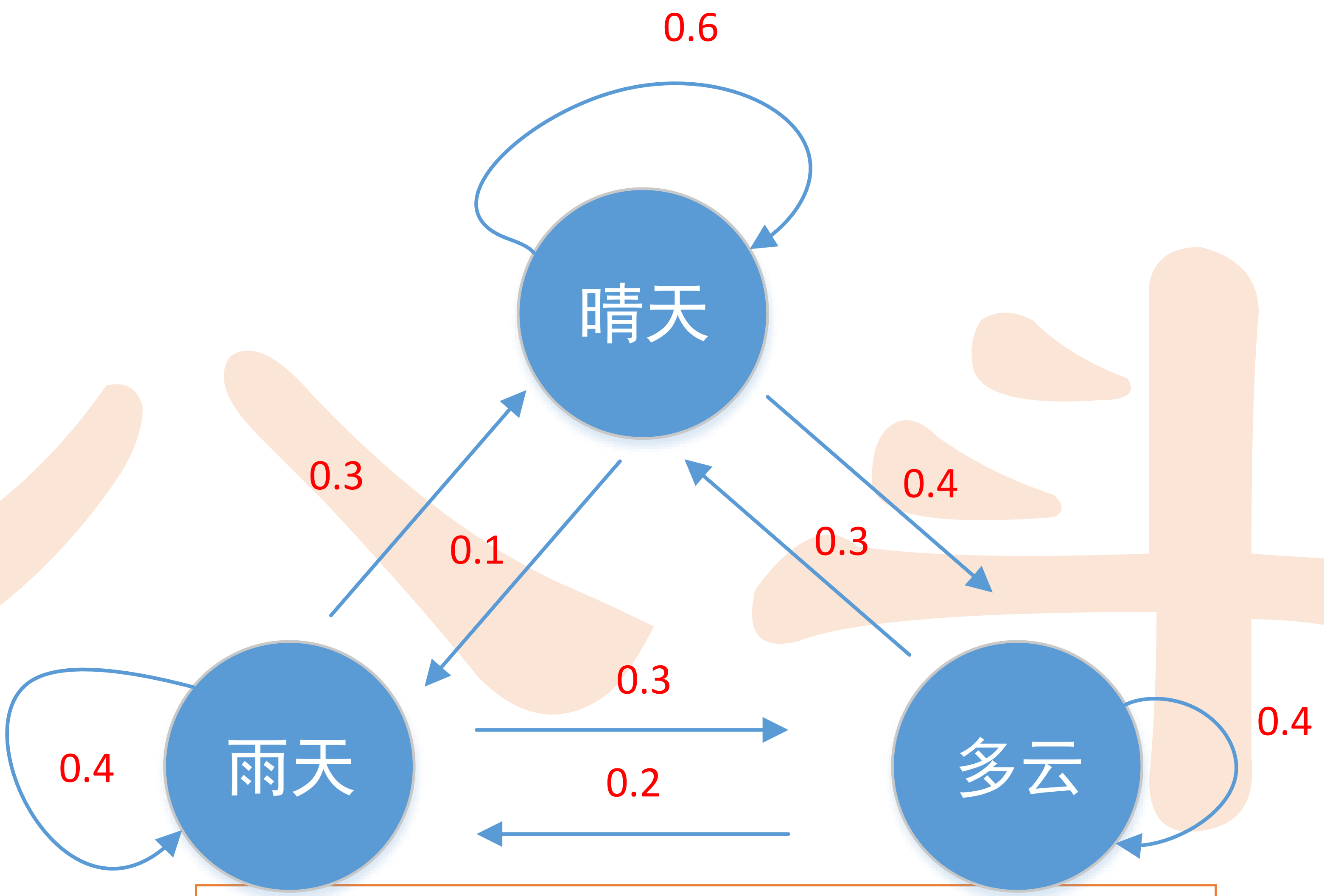
$$a_{k,l} = P(S_{t+1} = l \mid S_t = k) \quad k,l = 1, 2, \ldots, M$$

马尔科夫模型实例

- 天气
  - 状态定义
    - {晴天， 雨天， 多云}
  - 状态转移概率$a_{k,l}$
    - P(晴天|雨天)， P(雨天|多云)
  - 初始概率π_k
    - P(晴天)， P(雨天)， P(多云)

中文分词-02

马 尔 科 夫 模 型 实 例

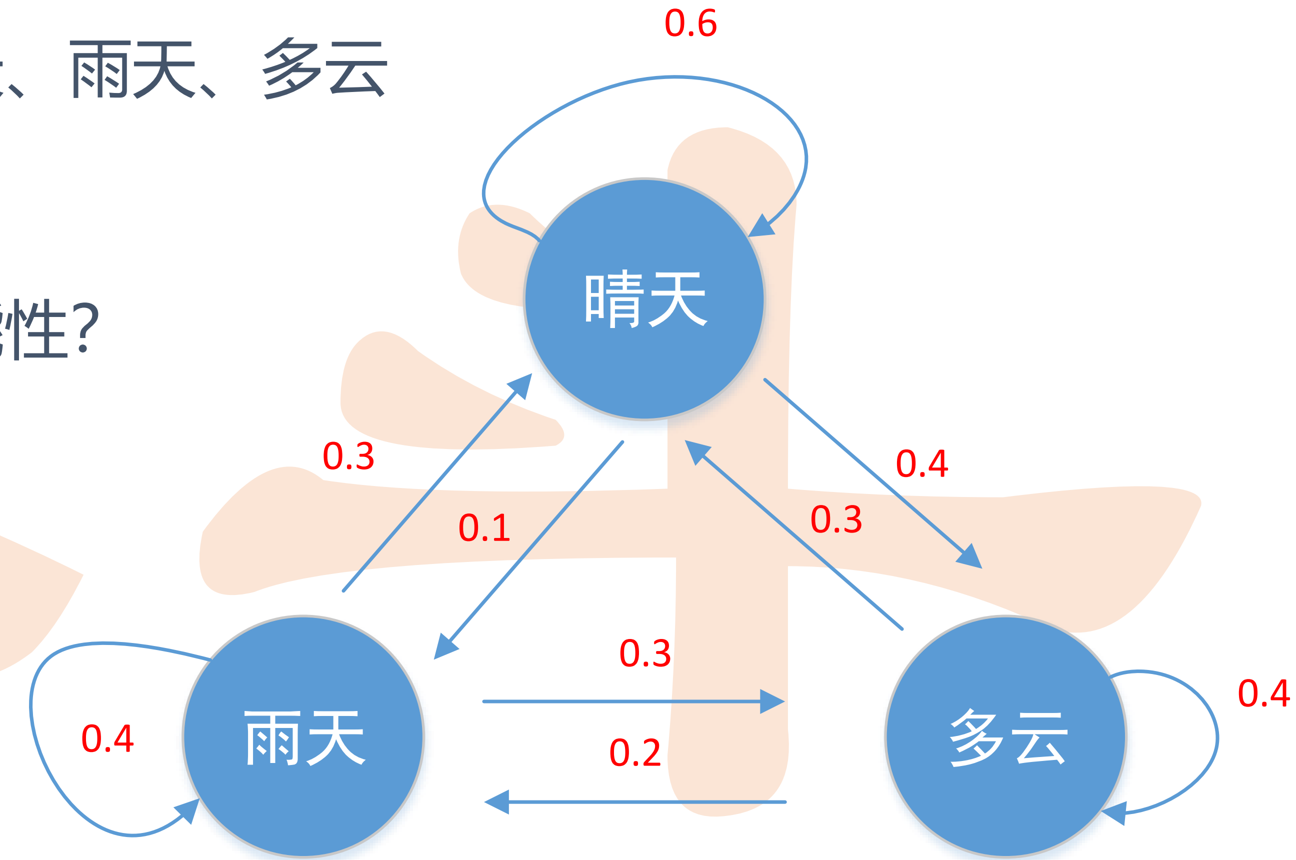


马尔科夫模型参数估计

- 最大似然法（策略+算法）
  - 状态转移概率$a_{k,l}$
    - P(St+1=l|St=k)=l紧跟k出现的次数/k出现的总次数
  - 初始概率π_k
    - P(S1=k)=k作为序列开始的次数/观测序列总数

马尔科夫模型应用

- 天气预测
  - 前几天天气情况：晴天、晴天、雨天、多云
  - 接下来一天的天气预计怎样?
  - 接下来三天天气都是晴的可能性?



小 结

- 马尔科夫模型是对一个序列数据建模，但有时我们需要对两个序列数据建模
  - 例如：
    - 机器翻译：源语言序列 <-> 目标语言序列
    - 语音识别：语音信号序列 <-> 文字序列
    - 词性标注：文字序列 <-> 词性序列
      - 写/一个/程序
      - Verb/Num/Noun

OutLine

- 马尔可夫模型
- 隐马尔可夫模型（HMM）

两 个 序 列

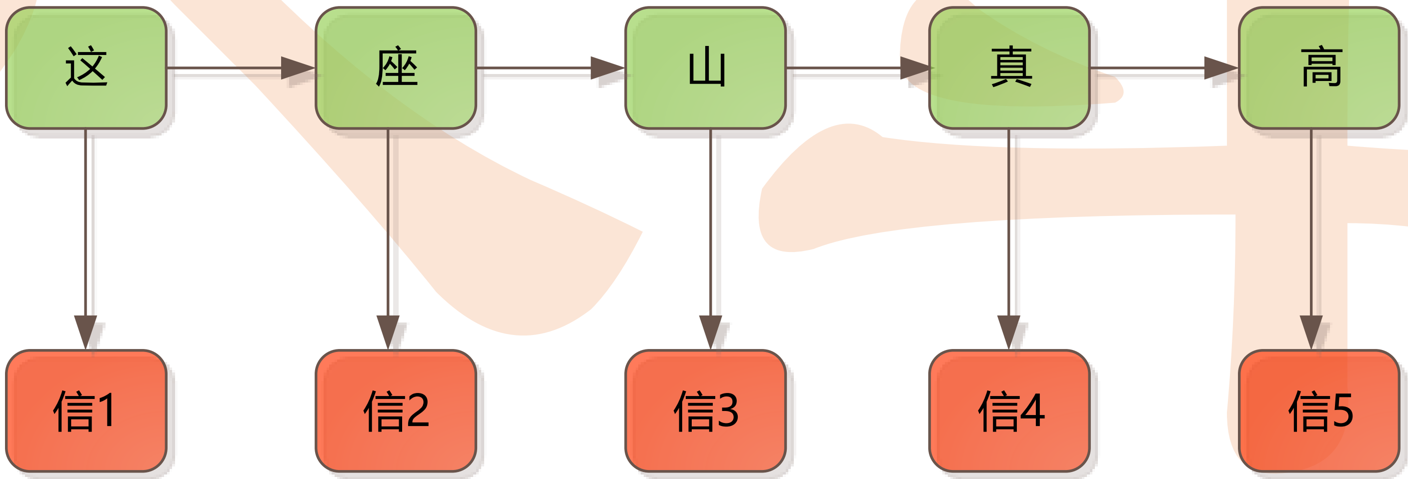
- 马尔科夫模型是对一个序列数据建模，但有时我们需要对两个序列数据建模
  - 例如：
  - 机器翻译：源语言序列 <--> 目标语言序列
  - 语音识别：语音信号序列 <--> 文字序列
  - 词性标注：文字序列 <--> 词性序列
    - 写 / 一个 / 程序
    - Verb / Num / Noun
  - 拼音纠错：原始文字序列 <--> 纠正过的文字序列
    - 自己的事情自己做
    - 自己的事情自已做

观察序列和隐藏序列

- 通常其中一个序列是我们观察到的，另一个隐藏的但也是我们先要寻找的
  - 通常把观察到的序列表示为O，隐藏的序列表示为S
    - 语音识别中我们的观测到的是声波信号，记为O，实际上表达的文字观测不到，记为S。而我们要找的是S，即“到底说了什么”。
    - 中文输入法：拼音是O，文字是S
    - 词性标注？拼写纠错？

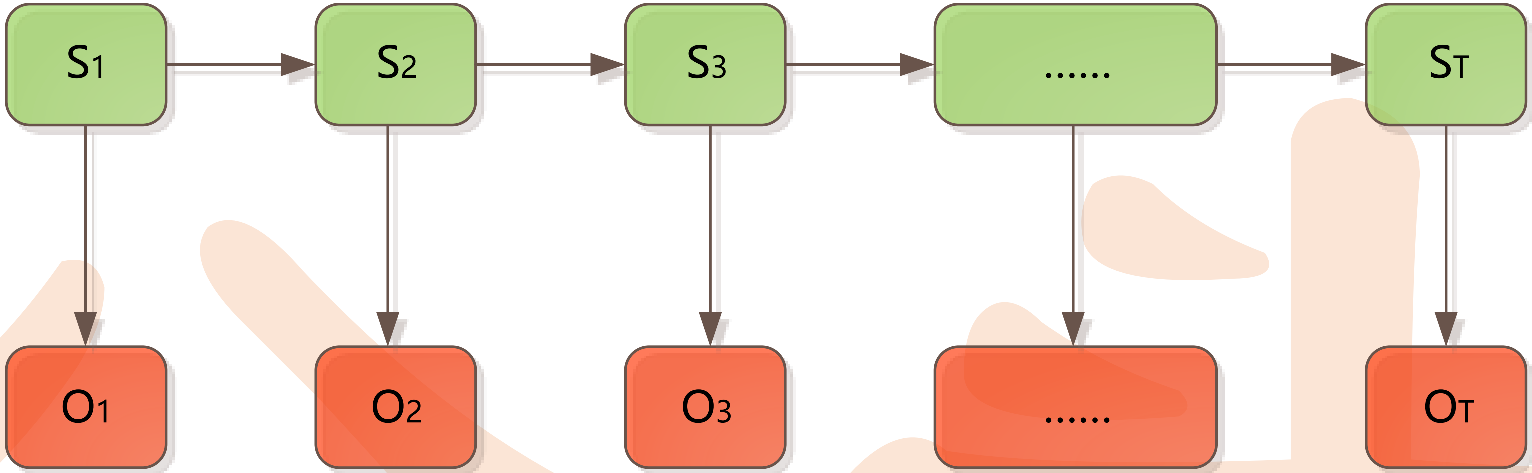
隐 马 尔 可 夫 模 型

- 观察序列O中的数据通常是由对应的隐藏序列数据决定的，彼此间相互独立
- 隐藏序列数据间相互依赖，通常构成了马尔科夫序列
  - 例如，语音识别中声波信号每段信号都是相互独立的,有对应的文字决定
  - 对应的文字序列中相邻的字相互依赖，构成Markov链



隐 马 尔 可 夫 模 型

- 观察和隐藏序列共同构成隐马模型



- O($o_1 o_2 \ldots o_T$)：观测序列，o_t只依赖于s_t
- S($s_1 s_2 \ldots s_T$)：状态序列（隐藏序列），S是Markov序列，假设1阶Markov序列，则s_{t+1}只依赖于s_t

隐马尔可夫模型

- HMM参数
  - 状态，由数字表示，假设共有M个
  - 观测，由数字表示，假设共有N个
  - 初始概率，由π_k表示
  - 状态转移概率，由$a_{k,l}$表示
    - $a_{k,l} = \mathrm{P}(S_{t+1} = l | S_t = k) \quad k, l = 1,2, \ldots, M$
  - 发射概率，由$b_k(u)$表示
    - $b_k(u) = \mathrm{P}(O_t = u | S_t = k) \quad u = 1,2, \ldots, N \; k = 1,2, \ldots, M$

我 们 的 参 数

- 初始概率
  - BEMS：位置信息
    - B（开头）
    - M（中间）
    - E(结尾)
    - S(独立成词)
  - 词性：
    - n 名词
    - nr 人名
    - ns 地名
    - v 动词
    - vd 副动词
    - vn 名动词

取概率的Log值

('B', 'mq'): -6.78695300139688,
('B', 'n'): -1.6966257797548328,
('B', 'ng'): -3.14e+100,
('B', 'nr'): -2.2310495913769506,
('B', 'nrfg'): -5.873722175405573,
('B', 'nrt'): -4.985642733519195,
('B', 'ns'): -2.8228438314969213,
('B', 'nt'): -4.846091668182416,
('B', 'nz'): -3.94698846057672,

我 们 的 参 数

- 转移概率

```
('B', 'ad'): {('E', 'ad'): -0.0007479013978476627,
              ('M', 'ad'): -7.198613337130562},
('B', 'ag'): {},
('B', 'an'): {('E', 'an'): 0.0},
('B', 'b'): {('E', 'b'): -0.06753917715798491,
             ('M', 'b'): -2.7286269787493125},
('B', 'bg'): {},
('B', 'c'): {('E', 'c'): -0.04442738163948101,
             ('M', 'c'): -3.136030746864676 6},
('B', 'd'): {('E', 'd'): -0.04677309521554972,
             ('M', 'd'): -3.0857425240950174},
```

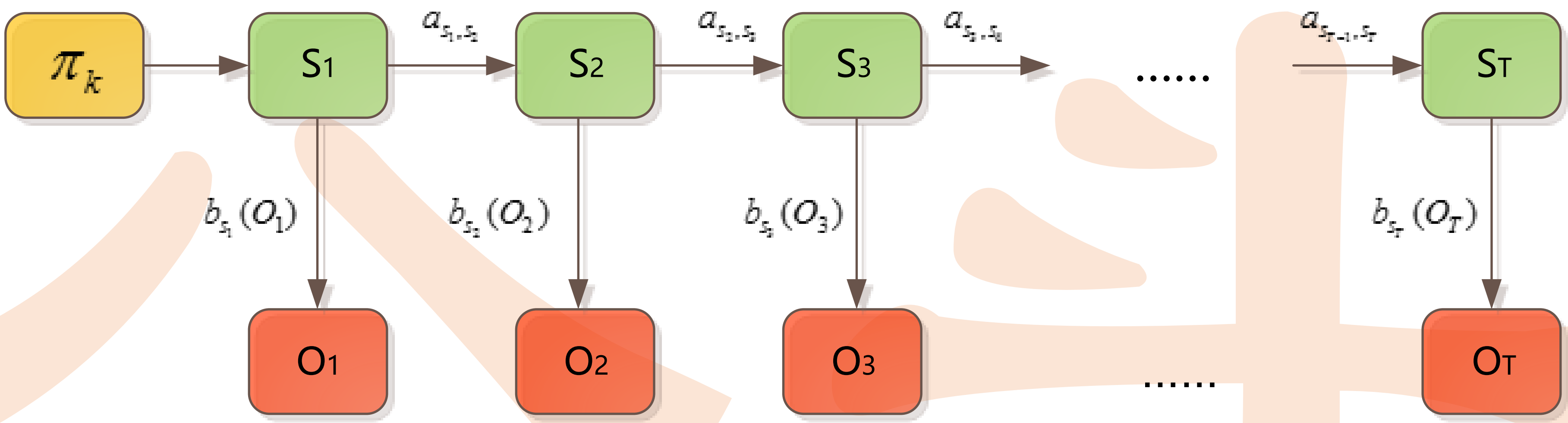
我 们 的 参 数

- 发射概率

```
('B', 'df'): {u'不': 0.0},
('B', 'dg'): {},
('B', 'e'): {u'呜': -2.5576660040704287,
            u'哇': -6.52795791762255,
            u'哈': -4.13006264482418,
            u'哎': -0.5265430396614004,
            u'啊': -1.8551290831606442,
            u'尚': -4.918520005188451},
('B', 'en'): {},
('B', 'f'): {u'一': -5.967972486230796,
            u'上': -3.548007352145394З,
            u'下': -4.27226731366774З,
            u'业': -5.628317434111854,
```

HMM生成过程

- 先生成第一个状态，然后依次由当前状态生成下一个状态，最后每个状态发射出一个观察值

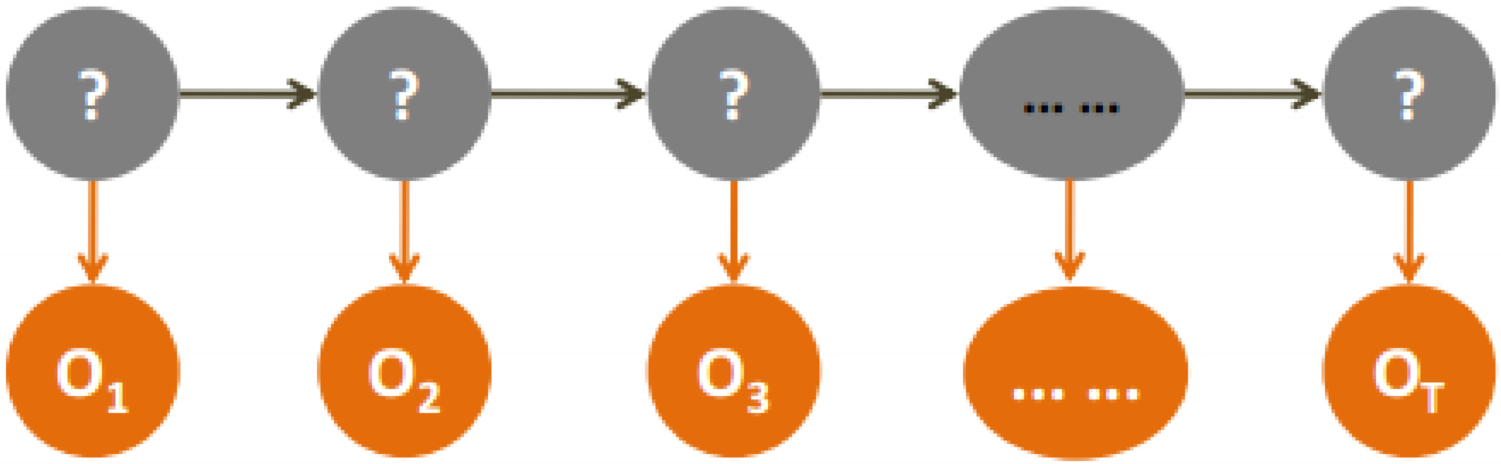


$$P(o_{1:t}, s_{1:t}) = P(s_1)P(o_1|s_1)P(s_2|s_1)\ldots\ldots P(s_t|s_{t-1})P(o_t|s_t)$$

$$= \pi_{s_1} \prod_{i=1}^{t-1} a_{s_i,s_{i+1}} \prod_{i=1}^{t} b_{s_i}(o_i)$$

HMM应用

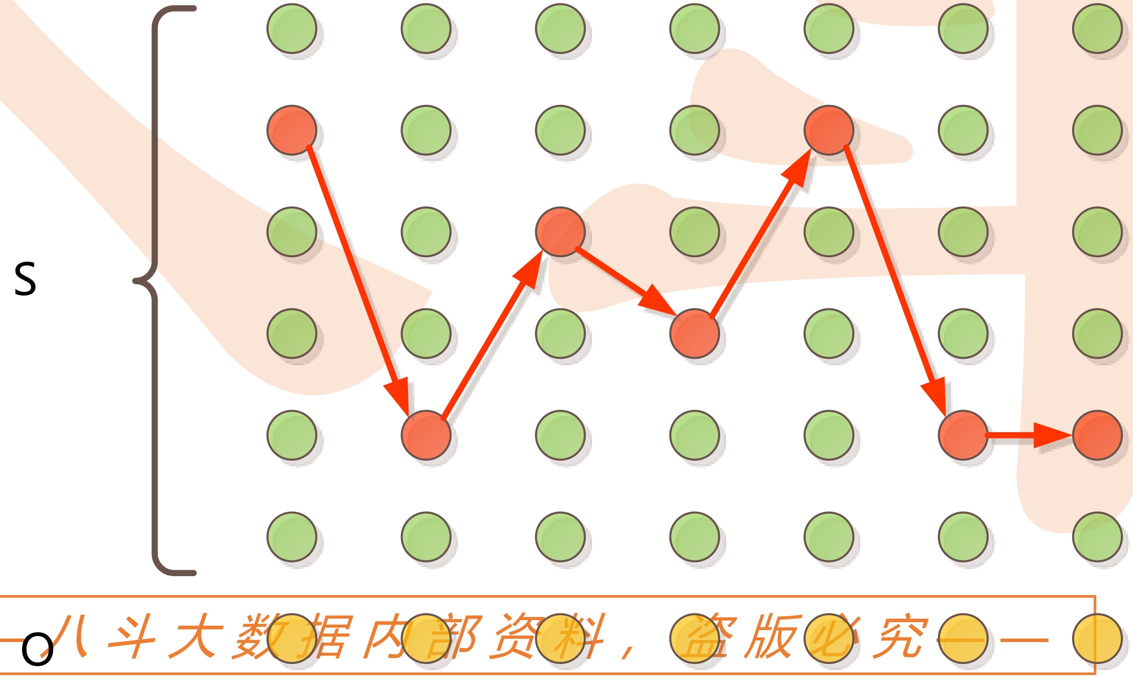
- 给定O，寻找最优的S



$$S^* = \arg\max_S P(S \mid O) = \arg\max_S \frac{P(S,O)}{P(O)} = \arg\max_S P(S,O)$$

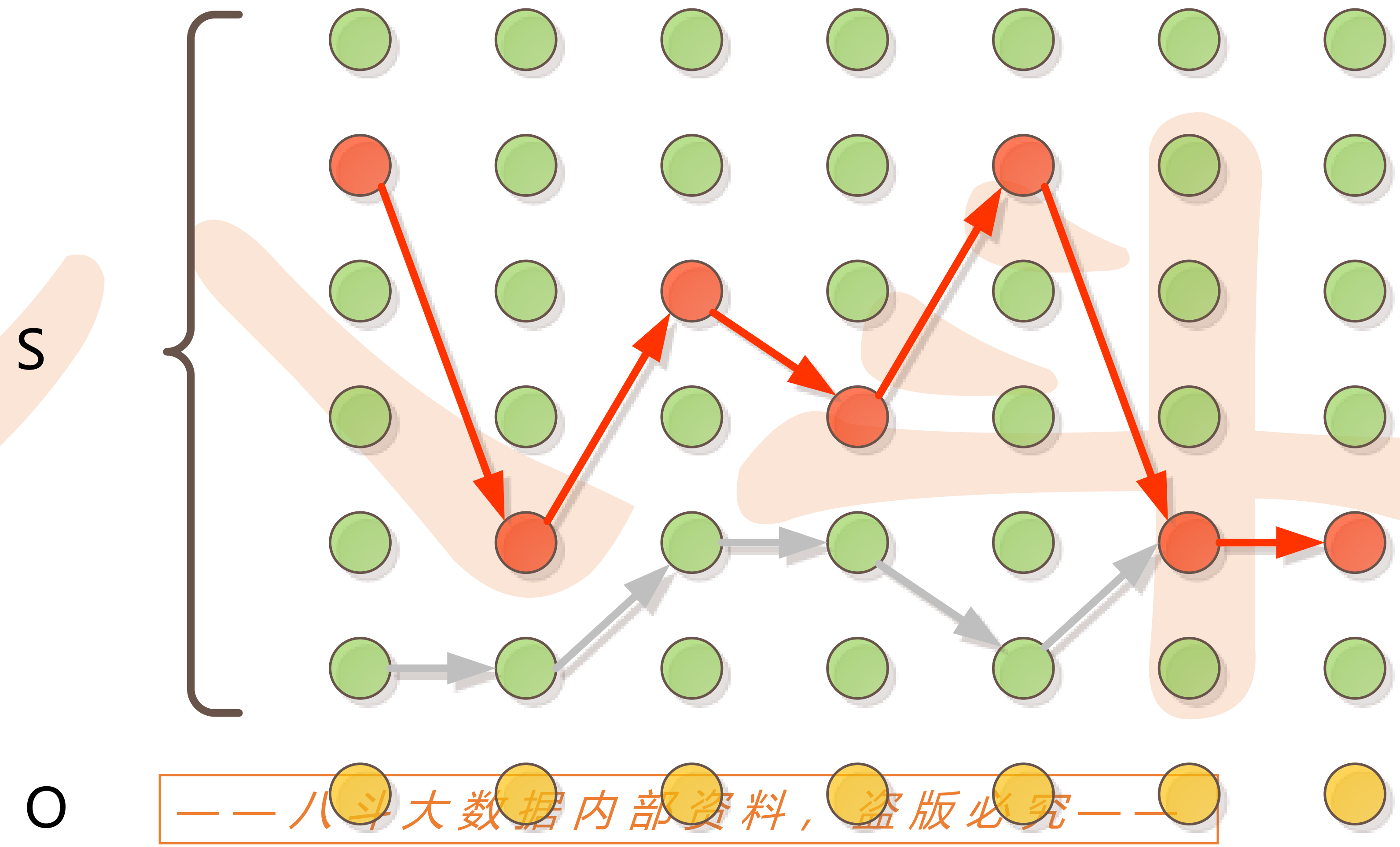
HMM应用

- 给定O，寻找最优的S
- 寻找一条最优的路径
- 如果比较所有路径：遍历所有的S，算出一个最大的，则时间复杂度是M^T，不可接受！



HMM应用-viterbi算法

- 动态规划，在t+1位置重用t的结果



Q&A

@八斗数据