南京大学
NANJING UNIVERSITY

自然语言处理

序列化标注

吴震

南京大学人工智能学院
南京大学自然语言处理研究组

2023年4月

目录

01

背景知识

BACKGROUND

问题一

南京大学 NANJING UNIVERSITY

- 问题描述
  - 你有一个住得很远的朋友，他每天跟你打电话告诉你他那天做了什么。你的朋友仅仅对三种活动感兴趣：公园散步，购物以及清理房间。他选择做什么事情只凭天气（晴天、下雨）。你对于他所住的地方的天气情况并不了解，因此决定根据他每天的活动情况来推测其所在地的天气情况。

状态：晴天、下雨

观测值：散步、购物、清理房间

观测序列：散步、购物、散步、清理房间、散步 → 状态序列：.....

问题二

南京大学 NANJING UNIVERSITY

- 问题描述
  - 最近一个赌场的老板生意不顺，他发现有位大叔在自己的赌场玩得一手好骰子，总能赢钱，几乎战无不胜。根据多年的经验，老板怀疑大叔使用了“偷换骰子大法”。老板是个冷静的人，看这位大叔也不是善者，不想轻易得罪他，又不想让他坏了规矩。正愁上心头，这时候进来一位名叫HMM的炼金术士，告诉老板他有一个很好的解决方案：不用近其身，只要在远处装个摄像头，把每局的骰子的点数都记录下来，然后运用其强大的数学功力，用这些数据推导出：
    - 该大叔是不是在出千?
    - 如果是在出千，那么他用了几个作弊的骰子？ 还有当前是不是在用作弊的骰子。
    - 这几个作弊骰子出现各点的概率是多少?

状态：正常骰子，作弊骰子1，作弊骰子2,...

观测值：骰子的点数

问题三

南京大学 NANJING UNIVERSITY

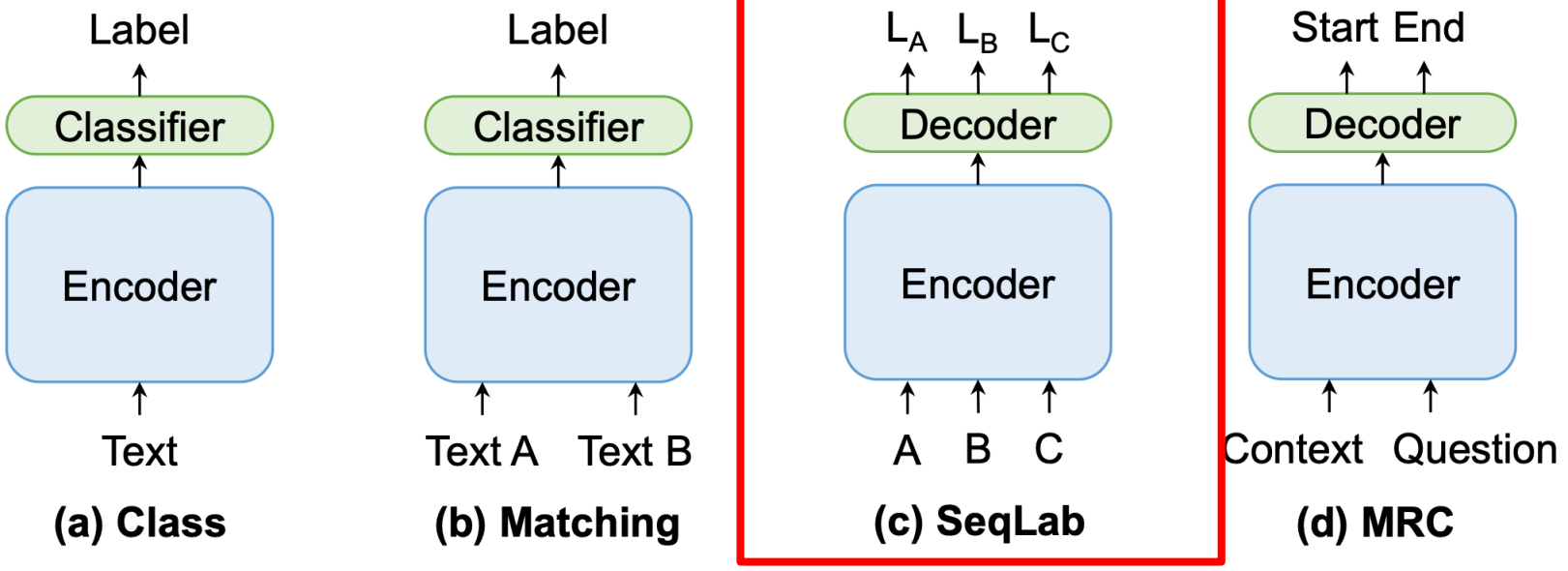
- 智能拼音输入法

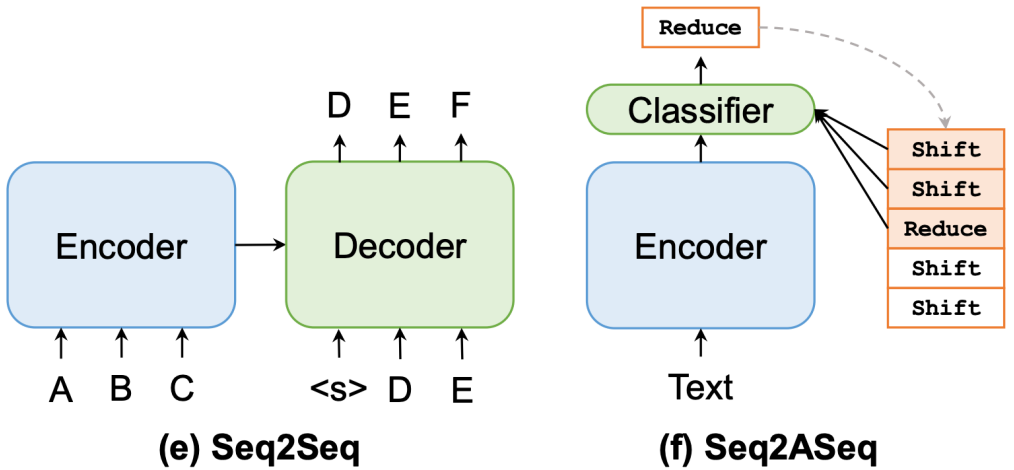
观测序列：nan jing da xue ren gong zhi neng xue yuan

状态序列：南 京 大 学 人 工 智 能 学 院

问题总结：对于一个观测序列，如何知道观测序列背后对应的状态序列？

自然语言处理中典型的任务形式





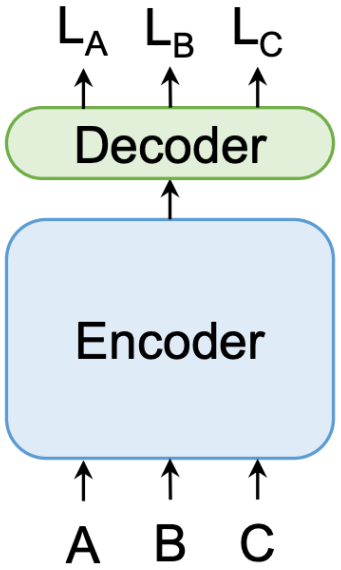
序列化标注 (SEQUENCE LABELING)

- 定义：给定一个观测序列作为输入，输出是一个标记序列或状态序列。
- 目标：建立一个模型，使它能够对观测序列给出对应的标记序列。

输入：观测序列 $X = (x_1, x_2, \ldots, x_n)$

序列化标注

输出：标记序列 $Y = (y_1, y_2, \ldots, y_n)$



以词性标注为例

词性

- 词性又称词类，是词汇的一个基本的语法属性。
- 反映了词在句子中的语法功能和意义。
- 语言学界对词性的数量、性质和普遍性进行了大量的争论
  - 封闭类
  - 开放类

词性

- 封闭类 (closed class，function words，每类词数有限)
  - Determiners (a/an, the, …)
  - Pronouns (this, that, …)
  - Prepositions (at, in, …)
  - Conjunctions (and, but, …)
  - Auxiliary verbs (do, does)
  - Particles (if, not, …)
  - Numerals (one, two, …)

词性

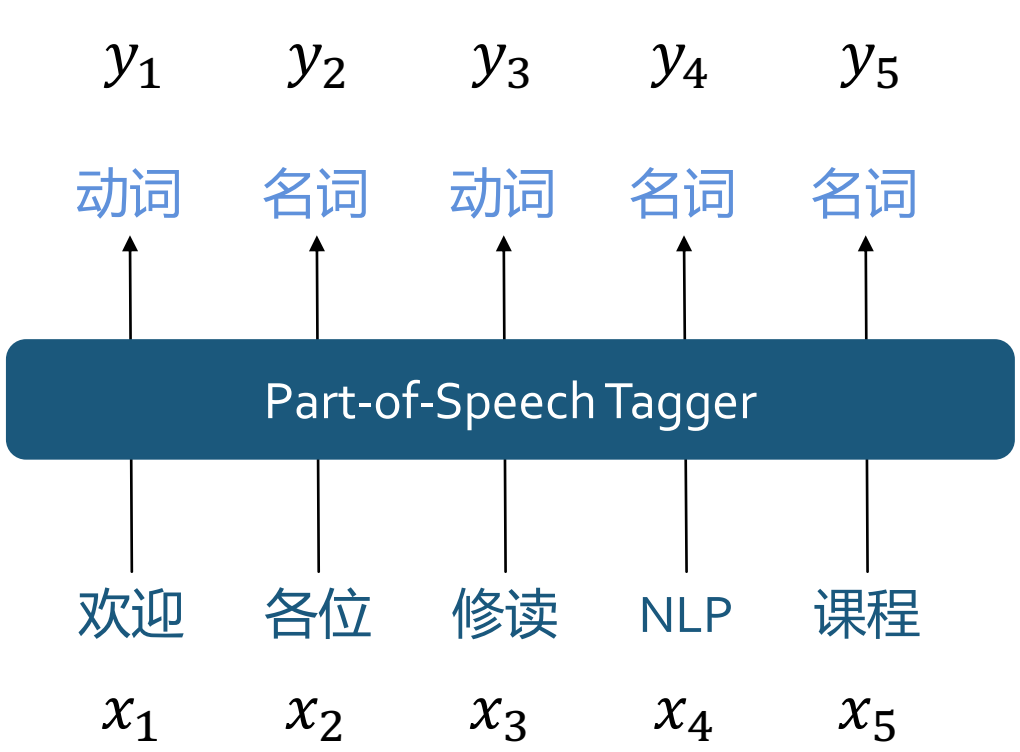
- 开放类（open class，每类词数不限）
  - Nouns
    - 句法上：可作物主、可有限定词、有复数形式
    - 语义上：人名、地名和物名等
  - Verbs
    - 句法上：作谓语、有几种词形变化
    - 语义上：动作、过程（一系列动作）
  - Adjectives
    - 句法上：修饰Nouns等
    - 语义上：性质
  - Adverbs
    - 句法上：修饰Verbs等
    - 语义上：方向、程度、方式、时间

PENN树库的词性集合

Tag	Description	Example	Tag	Description	Example
CC	coordin. conjunction	*and, but, or*	SYM	symbol	+,%, &
CD	cardinal number	*one, two*	TO	"to"	*to*
DT	determiner	*a, the*	UH	interjection	*ah, oops*
EX	existential 'there'	*there*	VB	verb base form	*eat*
FW	foreign word	*mea culpa*	VBD	verb past tense	*ate*
IN	preposition/sub-conj	*of, in, by*	VBG	verb gerund	*eating*
JJ	adjective	*yellow*	VBN	verb past participle	*eaten*
JJR	adj., comparative	*bigger*	VBP	verb non-3sg pres	*eat*
JJS	adj., superlative	*wildest*	VBZ	verb 3sg pres	*eats*
LS	list item marker	*1, 2, One*	WDT	wh-determiner	*which, that*
MD	modal	*can, should*	WP	wh-pronoun	*what, who*
NN	noun, sing. or mass	*llama*	WP$	possessive wh-	*whose*
NNS	noun, plural	*llamas*	WRB	wh-adverb	*how, where*
NNP	proper noun, sing.	*IBM*	$	dollar sign	*$*
NNPS	proper noun, plural	*Carolinas*	#	pound sign	*#*
PDT	predeterminer	*all, both*	"	left quote	' or "
POS	possessive ending	*'s*	"	right quote	' or "
PRP	personal pronoun	*I, you, he*	(	left parenthesis	[, (, {, <
PRP$	possessive pronoun	*your, one's*	)	right parenthesis	],), }, >
RB	adverb	*quickly, never*	,	comma	,
RBR	adverb, comparative	*faster*	.	sentence-final punc	. ! ?
RBS	adverb, superlative	*fastest*	:	mid-sentence punc	: ; ... – -
RP	particle	*up, off*			

词性标注 (PART-OF-SPEECH TAGGING)

- 定义：给一句话中的每个词 (word)标注上词性 (Part-of-Speech)



1 名词	5 代词	9 量词
2 动词	6 介词	10 助词
3 形容词	7 连词	11 感叹词
4 副词	8 数词	12 拟声词

一个以义为纲的词汇分类体系

——《现代汉语分类词典》*

洪桂治　苏新春

为什么需要词性标注？

- 为很多现实任务提供必要的信息
- 句法分析
  - 在对句子进行句法分析前需要知道每个词的词性
- 信息抽取
  - 帮助识别命名实体、关系
- 机器翻译
  - 帮助多义词进行更好的上下文翻译

词性标注的挑战

南京大学 NANJING UNIVERSITY

- 兼类词
  - 一个词具有两个或者两个以上的词性
  - 英文的Brown语料库中，10.4%的词是兼类词。例如：
    - The back door
    - On my back
    - Promise to back the bill
  - 汉语兼类词，例如：
    - 把门锁上　　买了一把锁
    - 他研究xx　　他的研究工作...
    - 由于缺少词形变化，汉语的兼类词更多！

词性标注语料库

- Brown Corpus：语料来自于美国英语出版物上的文本，共500篇，每篇大约2000个单词，合计100万词（1961）
- WSJ：语料来自于华尔街日报，合计100万词（1989）
- Switchboard：语料来自于电话对话文本，合计200万词（1990-1991）

Battle-tested/NNP industrial/JJ managers/NNS here/RB always/RB buck/VB up/IN nervous/JJ newcomers/NNS with/IN the/DT tale/NN of/IN the/DT first/JJ of/IN their/PP$ countrymen/NNS to/TO visit/VB Mexico/NNP ,/, a/DT boatload/NN of/IN samurai/NNS warriors/NNS blown/VBN ashore/RB 375/CD years/NNS ago/RB ./.

“/“ From/IN the/DT beginning/NN ,/, it/PRP took/VBD a/DT man/NN with/IN extraordinary/JJ qualities/NNS to/TO succeed/VB in/IN Mexico/NNP ,/, ”/” says/VBZ Kimihide/NNP Takimura/NNP ,/, president/NN of/IN Mitsui/NNS group/NN ’s/POS Kensetsu/NNP Engineering/NNP Inc./NNP unit/NN ./.

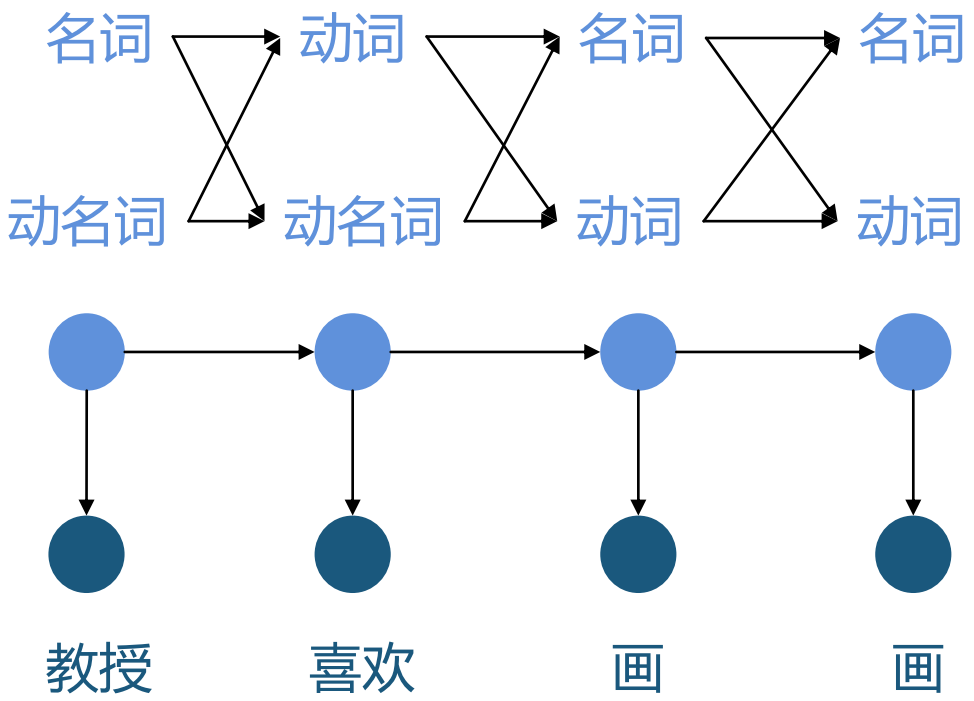
02

基于统计学习的序列化标注

STATISTICAL LEARNING-BASED SEQUENCE LABELING

决定一个词词性的因素

- 从语言学角度：由词的用法以及在句中的语法功能决定
- 统计学角度：
  - 和上下文的词性（前后词的标注）相关
  - 和上下文单词（前后词）相关



共有16种可能

隐藏状态序列Y

观测序列X

词性标注

- 词性标注：给定句子X，求句子对应的词性序列Y

$$\text{argmax}_Y P(Y|X) = \text{argmax}_Y \frac{P(Y,X)}{P(X)}$$

$$= \text{argmax}_Y \ P(Y,X)$$

$$= \text{argmax}_Y \boxed{P(Y)P(X|Y)}$$

隐含马尔可夫模型

Hidden Markov Model，HMM

词性标注

- 词性标注：给定句子X，求句子对应的词性序列Y

$$P(\text{名词 动词 动词 名词} \mid \text{教授 喜欢 画 画})$$

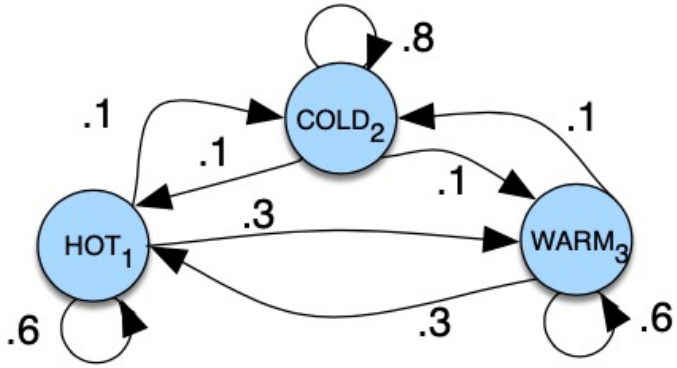
$$= P(\text{名词 动词 动词 名词 教授 喜欢 画 画}) / P(\text{教授 喜欢 画 画})$$

$$\propto P(\text{名词 动词 动词 名词 教授 喜欢 画 画})$$

$$= P(\text{名词 动词 动词 名词}) \, P(\text{教授 喜欢 画 画} \mid \text{名词 动词 动词 名词})$$

马尔可夫过程

- 马尔可夫链
  - 描述在状态空间中，从一个状态到另一个状态转换的随机过程。



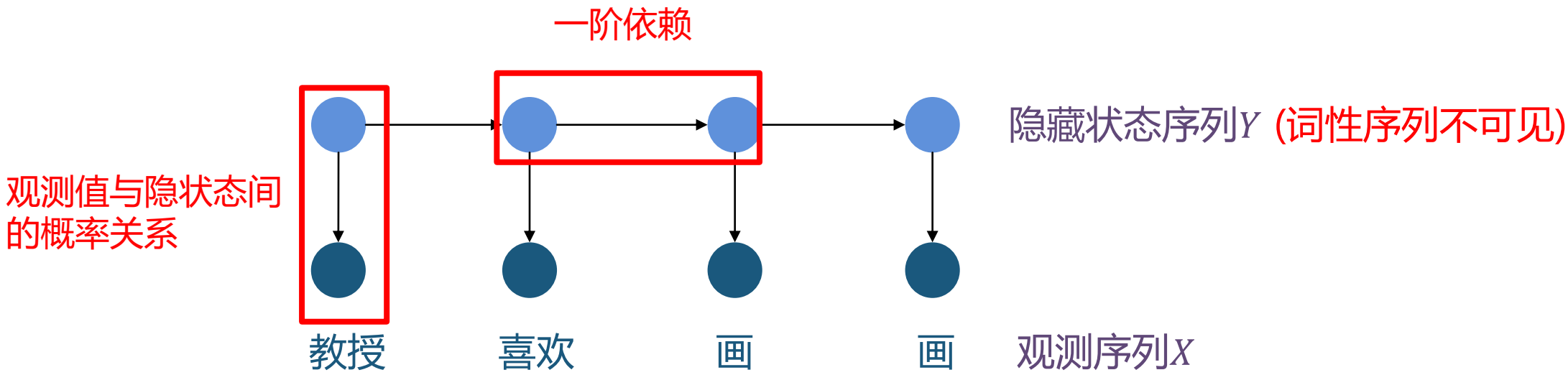
天气状态的马尔可夫链

- 马尔科夫假设
  - 马尔可夫链在任意时刻 t 的状态只依赖于它在前一时刻的状态，与其他时刻的状态无关

$$P(y_t|y_1,\cdots,y_{t-1}) = P(y_t|y_{t-1})$$

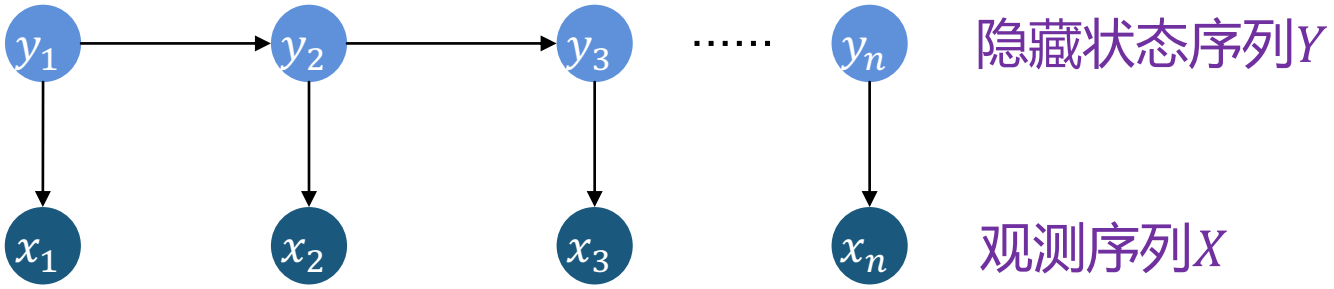
隐含马尔可夫模型 (HMM)

- HMM是一阶马尔可夫链的扩展
  - 状态序列不可见（隐藏）
  - 隐藏的状态序列满足一阶马尔可夫链性质
  - 可见的观察值与隐藏的状态之间存在概率关系



隐含马尔可夫模型 (HMM)

- 序列化标注的统计学模型
  - 描述了由隐马尔可夫链随机生成观测序列的过程，属于生成模型。
- 时序概率模型
  - 描述由一个隐藏的马尔可夫链随机生成不可观测的状态随机序列，再由各个状态生成一个观测值，从而产生观测序列的过程。



$$P(Y,X) = P(Y)P(X|Y)$$

隐含马尔可夫模型 (HMM)

- 计算$P(Y)$：

$$P(Y) = P(y_1, y_2, \dots, y_n)$$

$$= \prod_{t=1}^{n} P(y_t | y_1, \dots, y_{t-1})$$

- 马尔可夫假设：
  - 描述从一个状态到转换另一个状态的随机过程。该过程具备“无记忆”的性质，即当前时刻状态的概率分布只能由上一时刻的状态决定，和更久之前的状态无关。

$$P(y_t | y_1, \dots, y_{t-1}) = P(y_t | y_{t-1})$$

P(名词 动词 动词 名词)=P(名词) * P(动词|名词) * P(动词|动词) * P(名词|动词)

隐含马尔可夫模型 (HMM)

- 计算$P(X|Y)$：

$$P(X|Y) = P(x_1, x_2, \dots, x_n | y_1, y_2, \dots, y_n)$$

$$= \prod_{t=1}^{n} P(x_t | x_1, y_1, \dots, x_{t-1}, y_{t-1}, y_t)$$

- 观测独立性假设
  - 任意时刻的观测值只依赖于该时刻的马尔可夫链的状态，与其他观测及状态无关

$$P(x_t | x_1, y_1, \dots, x_{t-1}, y_{t-1}, y_t) = P(x_t | y_t)$$

P(教授 喜欢 画 画 | 名词 动词 动词 名词)=P(教授|名词) * P(喜欢|动词) * P(画|动词) * P(画|名词)

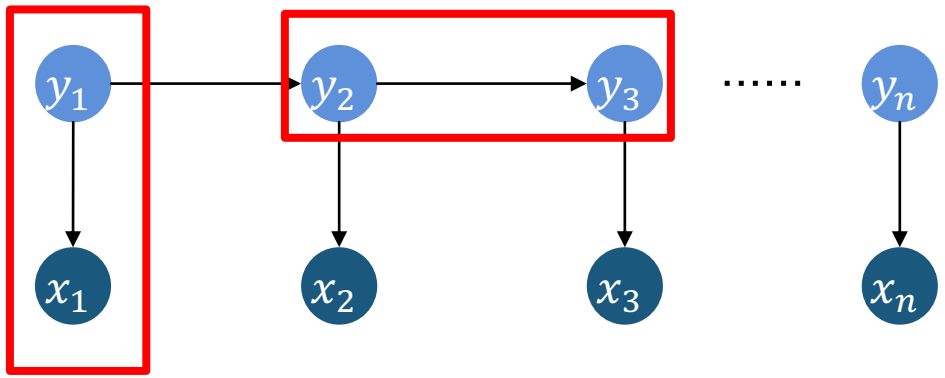
隐含马尔可夫模型 (HMM)

- 计算$P(Y,X)$ ：

$$P(Y,X) = P(Y)P(X|Y)$$

$$= P(y_1, y_2, \dots, y_n)P(x_1, x_2, \dots, x_n|y_1, y_2, \dots, y_n)$$

状态转移概率



发射概率

$$= \prod_{t=1}^{n} P(y_t|y_1, \dots, y_{t-1}) \prod_{t=1}^{n} P(x_t|x_1, y_1, \dots, x_{t-1}, y_{t-1}, y_t)$$

$$= \prod_{t=1}^{n} P(y_t|y_{t-1}) \prod_{t=1}^{n} P(x_t|y_t)$$

$$= \prod_{t=1}^{n} P(y_t|y_{t-1})P(x_t|y_t)$$

状态转移概率　发射概率

隐含马尔可夫模型 (HMM)

- 计算$P(Y,X)$ ：

P(名词 动词 动词 名词 , 教授 喜欢 画 画)

=P(名词) * P(动词|名词) * P(动词|动词) * P(名词|动词) *P(教授|名词) * P(喜欢|动词) * P(画|动词) * P(画|名词)

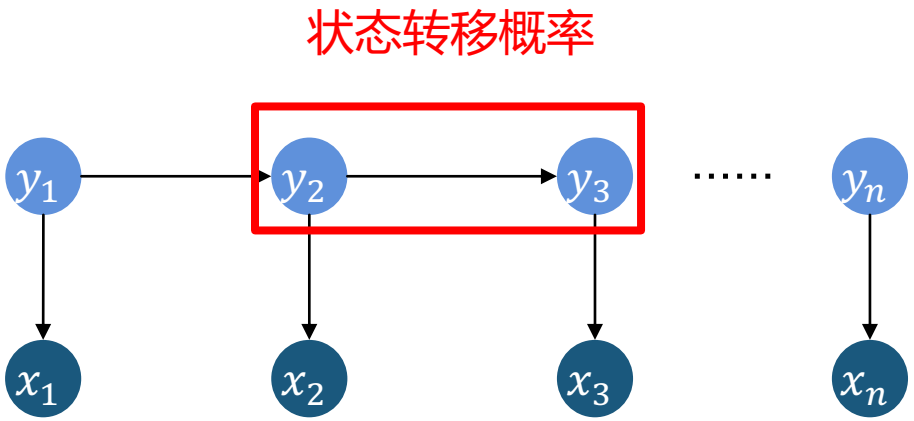
隐含马尔可夫模型 (HMM)

- 状态集合$\mathbb{Q} = \{q_1, q_2, \dots, q_Q\}$，观测值集合$\mathbb{V} = \{v_1, v_2, \dots, v_V\}$
  - Q和V分别表示状态数量和观测值数量

- $Y = (y_1, y_2, \dots, y_n)$是长度为 n 的状态序列，$X = (x_1, x_2, \dots, x_n)$是对应的观测序列
  - $y_t \in \mathbb{Q}$ 是一个随机变量，代表一个可能的状态值
  - $x_t \in \mathbb{V}$ 是一个随机变量，代表一个可能的观测值

隐含马尔可夫模型 (HMM)

- 状态转移概率矩阵$\mathbf{A}$：表示状态之间的转移概率
  - 其中$a_{i,j} = P(y_{t+1} = q_j | y_t = q_i)$，表示在 t 时刻处于状态 q_i 的条件下，在 $t+1$ 时刻转移到 q_j 的概率　　P(动词|名词)

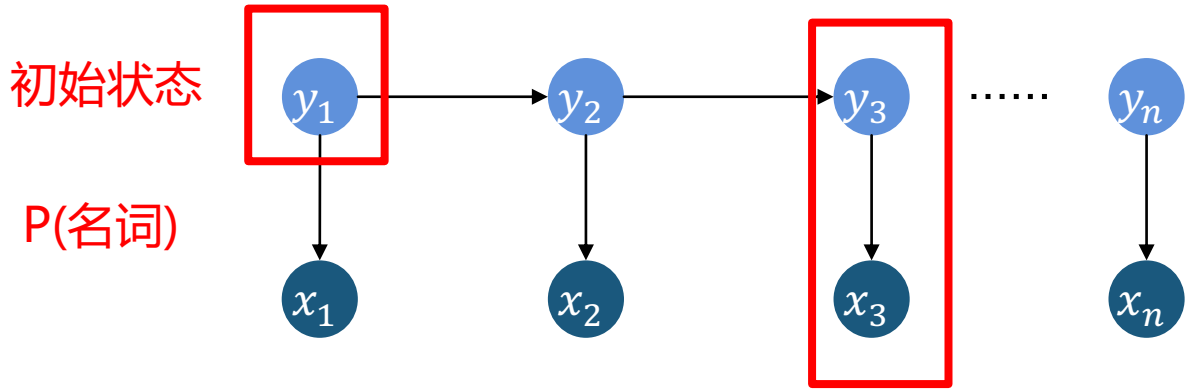


$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,Q} \\ a_{1,2} & a_{2,2} & \cdots & a_{2,Q} \\ \vdots & \vdots & \vdots & \vdots \\ a_{Q,1} & a_{Q,2} & \cdots & a_{Q,Q} \end{bmatrix}$$

隐含马尔可夫模型 (HMM)

- 发射概率矩阵**B** ：表示某个状态下生成某个观测值的概率
  - 其中 $b_j(k) = P(x_t = v_k | y_t = q_j)$ ，表示 t 时刻处于状态 q_j 的条件下生成观测值 v_k 的概率

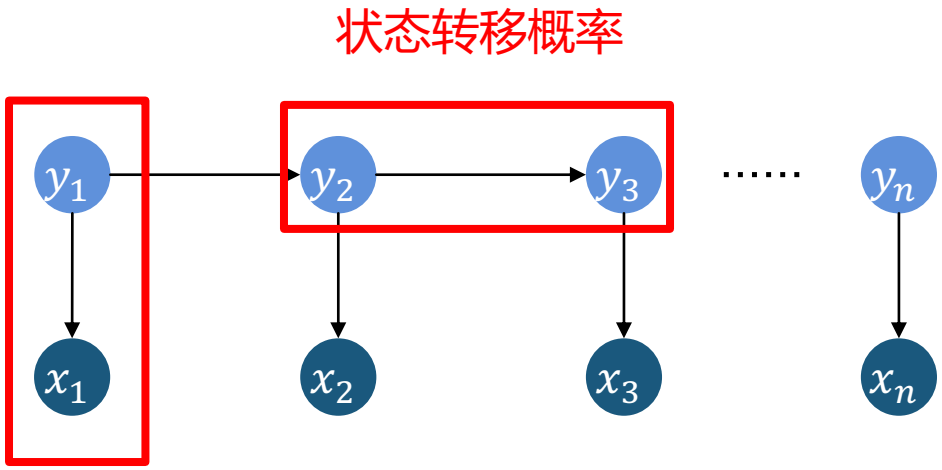
$$\mathbf{B} = \begin{bmatrix} b_1(1) & b_1(2) & \dots & b_1(V) \\ b_2(1) & b_2(2) & \dots & b_2(V) \\ \vdots & \vdots & \vdots & \vdots \\ b_Q(1) & b_Q(2) & \dots & b_Q(V) \end{bmatrix}$$



- 初始状态概率π ：$\pi = (\pi_1, \pi_2, \dots, \pi_Q)$
  - $\pi_i = P(y_1 = q_i)$ 表示开始时刻 $t = 1$ 时处于状态 q_i 的概率
  - $\sum_{i=1}^{Q} \pi_i = 1$

隐含马尔可夫模型 (HMM)

- 隐马尔可夫模型由初始状态概率 π 、状态转移矩阵 $\mathbf{A}$ 、以及发射概率矩阵 $\mathbf{B}$ 决定。一个隐马尔可夫模型可用三元符号表示：$\lambda = (\mathbf{A}, \mathbf{B}, \pi)$
  - 初始状态概率 π 和状态转移矩阵 $\mathbf{A}$ 确定了隐藏的马尔可夫链，生成了不可观测的状态序列；
  - 观测概率矩阵 $\mathbf{B}$ 确定了如何从状态生成观测值，与状态序列一起确定了如何产生观测序列。



$$P(Y, X) = P(Y)P(X|Y)$$

$$= \prod_{t=1}^{n} P(y_t|y_{t-1})P(x_t|y_t)$$

词性标注的HMM模型定义

- HMM：$\lambda = (\mathbf{A}, \mathbf{B}, \pi)$
- 状态集合$\mathbb{Q}$：预先定义的词性标签集
- 观测值集合$\mathbb{V}$：词表集合
- 状态转移概率矩阵$\mathbf{A}$：词性之间的转移概率
- 发射概率矩阵$\mathbf{B}$：某个词性生成某个词的概率
- 初始状态概率π：以某个词性作为开始状态的概率

HMM的基本问题

- 概率计算
  - 给定HMM模型 $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$ 和观测序列 $X = (x_1, x_2, \dots, x_n)$ ，计算观测序列 X 出现的概率 $P(X|\lambda)$　　P(教授 喜欢 画 画 | λ)
- 模型学习（参数估计）
  - 已知观测序列$X = (x_1, x_2, \dots, x_n)$ ，估计HMM模型$\lambda = (\mathbf{A}, \mathbf{B}, \pi)$的参数，使得该模型下观测序列的概率 $P(X|\lambda)$最大。　$\lambda = (\mathbf{A}, \mathbf{B}, \pi)$
- 预测（解码）
  - 已知HMM模型$\lambda = (\mathbf{A}, \mathbf{B}, \pi)$ 和观测序列 $X = (x_1, x_2, \dots, x_n)$ ，求该观测序列对应的最可能的状态序列 $Y = (y_1, y_2, \dots, y_n)$　　argmax_Y P(Y| 教授 喜欢 画 画, λ)

HMM的概率计算-直接计算

- 概率计算
  - 给定HMM模型 $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$ 和观测序列 $X = (x_1, x_2, \dots, x_n)$ ，计算观测序列 X 出现的概率 $P(X|\lambda)$

- 直接计算法
  - 枚举所有长度为n的状态序列，计算它们生成观测序列的概率并求和

$$P(X|\lambda) = \sum_{y_1, y_2, \dots, y_n} \pi_{y_1} \prod_{t=1}^{n} a_{y_t, y_{t+1}} b_{y_t}(x_t)$$

计算复杂度$O(n \times Q^n)$，不可行

HMM的概率计算-前向算法

- 定义前向概率
  - 给定HMM模型 $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$ ，定义到 t 时刻部分观测序列为$x_1, x_2, \ldots, x_t$ 且状态为 q_i 的概率为前向概率，记作：

$$\alpha_t(i) = P(x_1, x_2, \ldots, x_t, y_t = q_i|\lambda)$$

HMM的概率计算-前向算法

- 输入：隐马尔可夫模型$\lambda = (\mathbf{A}, \mathbf{B}, \pi)$，观测序列$X = (x_1, x_2, \dots, x_n)$
- 输出：观测序列概率 $P(X|\lambda)$
- 算法流程：
  - 初始化：

$$\alpha_1(i) = \pi_i b_i(x_1), \qquad i = 1,2, \dots, Q$$

  - 递推：

$$\alpha_t(i) = \left[\sum_{j=1}^{n} \alpha_{t-1}(j) a_{j,i}\right] b_i(x_t) \qquad i = 1,2, \dots, Q \quad t = 2, \dots, n$$

计算复杂度$O(n \times Q^2)$

HMM的参数估计

- 模型学习（参数估计）
  - 已知观测序列$X = (x_1, x_2, \ldots, x_n)$，估计模型$\lambda = (\mathbf{A}, \mathbf{B}, \pi)$的参数，使得该模型下观测序列的概率 $P(X|\lambda)$最大。

- 根据训练数据的不同，隐马尔可夫模型的学习方法也不同
  - 监督学习：训练数据包括观测序列和对应的状态序列，通过监督学习来学习隐马尔可夫模型。
  - 无监督学习：训练数据仅包括观测序列，通过无监督学习来学习隐马尔可夫模型。

HMM的参数估计—监督学习

- 假设数据集为$\mathbb{D} = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_N, Y_N)\}$，其中：
  - $X_1, \dots, X_N$为N个观测序列； $Y_1, \dots, Y_N$为对应的N个状态序列。
  - 序列X_k ，Y_k的长度为n_k。
- 估计转移概率 $a_{i,j}$
  - 设样本中前一时刻处于状态 q_i 、且后一时刻处于 q_j 的频数为 $A_{i,j}$ ，则转移概率 $a_{i,j}$ 的估计是：

$$a_{i,j} = \frac{A_{i,j}}{\sum_{u=1}^{Q} A_{i,u}}, \qquad i = 1,2, \dots, Q; \ \ j = 1,2, \dots, Q$$

HMM的参数估计—监督学习

- 估计转移概率 $a_{i,j}$
  - 设样本中前一时刻处于状态 q_i 、且后一时刻处于 q_j 的频数为 $A_{i,j}$ ，则转移概率 $a_{i,j}$ 的估计是：

$$a_{i,j} = \frac{A_{i,j}}{\sum_{u=1}^{Q} A_{i,u}}, \qquad i = 1,2,\dots,Q;\ \ j = 1,2,\dots,Q$$

$$a_{\text{动词,名词}} = \frac{A_{\text{动词,名词}}}{A_{\text{动词}}} = \frac{10471}{13124} = 0.797$$

HMM的学习问题—监督学习

- 估计观测概率 $b_j(k)$
  - 设样本中状态为 q_j 且其对应观测值为 v_k 的频数为 $B_{j,k}$，则状态为 q_j 并且观测值为 v_k 的概率 $b_j(k)$ 的估计为：

$$b_j(k) = \frac{B_{j,k}}{\sum_{v=1}^{V} B_{j,v}}, \qquad j = 1,2,\dots,Q;\ \ k = 1,2,\dots,V$$

$$b_{\text{动词}}(\text{画}) = \frac{B_{\text{动词,画}}}{B_{\text{动词}}} = \frac{4046}{13124} = 0.308$$

HMM的学习问题—监督学习

- 估计初始状态概率 π_i
  - 设样本中初始时刻（$t=1$）处于状态 q_i 的频数为 C_i ，则初始状态概率 π_i 的估计为：

$$\pi_i = \frac{C_i}{\sum_{j=1}^{Q} C_j}, \qquad i = 1,2,\dots,Q;$$

$$\pi_{\text{动词}} = \frac{C_{\text{动词}}}{\sum_{j=1}^{Q} C_j} = \frac{3728}{8429} = 0.442$$

HMM的模型预测（解码）

南京大学 NANJING UNIVERSITY

- 预测（解码）
  - 已知模型$\lambda = (\mathbf{A}, \mathbf{B}, \pi)$ 和观测序列 $X(x_1, x_2, \dots, x_n)$，求该观测序列对应的最可能的状态序列 $Y = (y_1, y_2, \dots, y_n)$

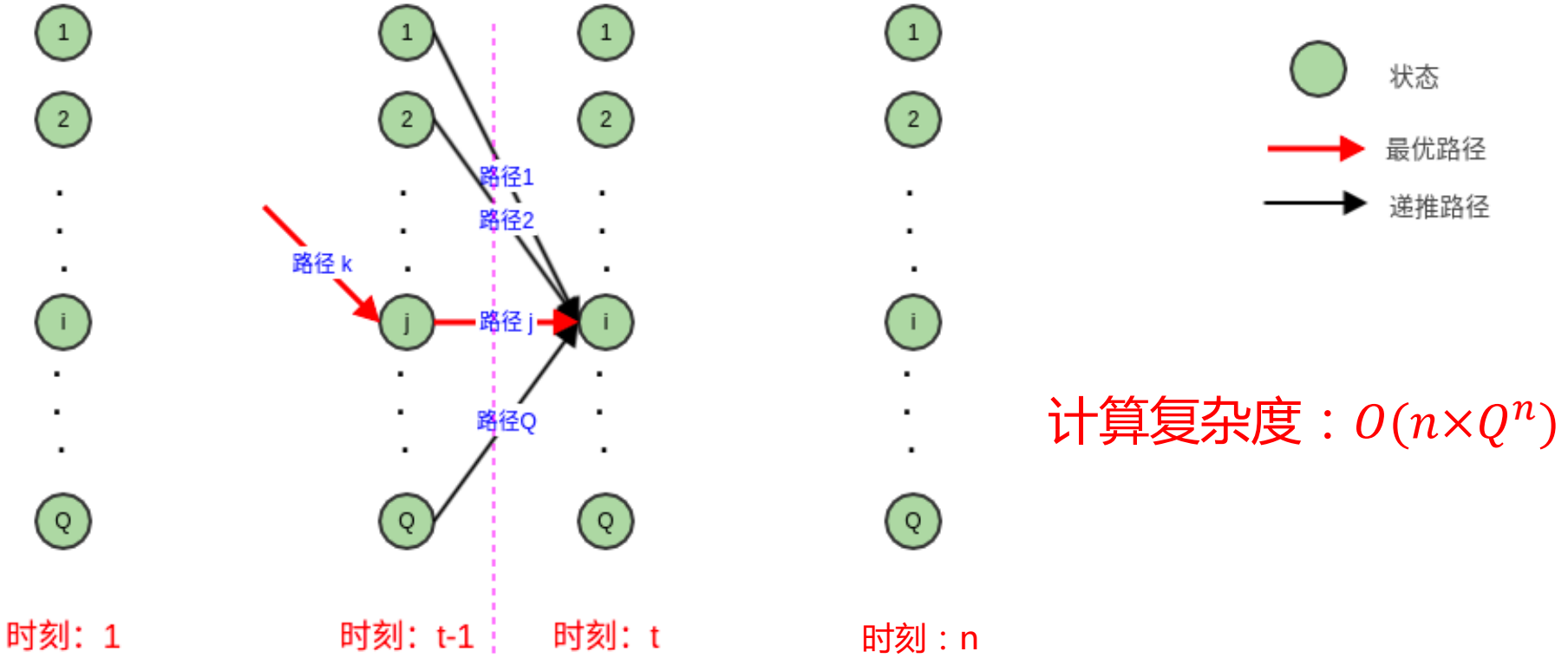
- 计算目标

$$\text{argmax}_{y_1,y_2,\dots,y_n} P(y_1, y_2, \dots, y_n, x_1, x_2, \dots, x_n) = \text{argmax}_{y_1,y_2,\dots,y_n} \pi_{y_1} \prod_{t=1}^{n} a_{y_t,y_{t+1}} b_{y_t}(x_t)$$

HMM的模型预测（解码）

- 计算目标

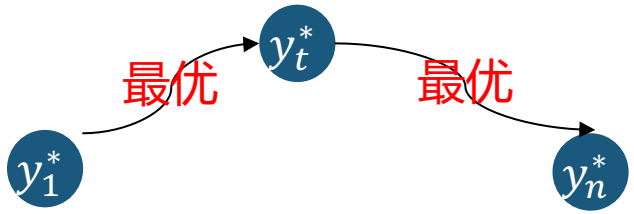
$$\text{argmax}_{y_1,y_2,\ldots,y_n} P(y_1, y_2, \cdots, y_n, x_1, x_2, \ldots, x_n) = \text{argmax}_{y_1,y_2,\ldots,y_n} \pi_{y_1} \prod_{t=1}^{n} a_{y_t,y_{t+1}} b_{y_t}(x_t)$$



计算复杂度：$O(n \times Q^n)$

HMM的模型预测—维特比算法

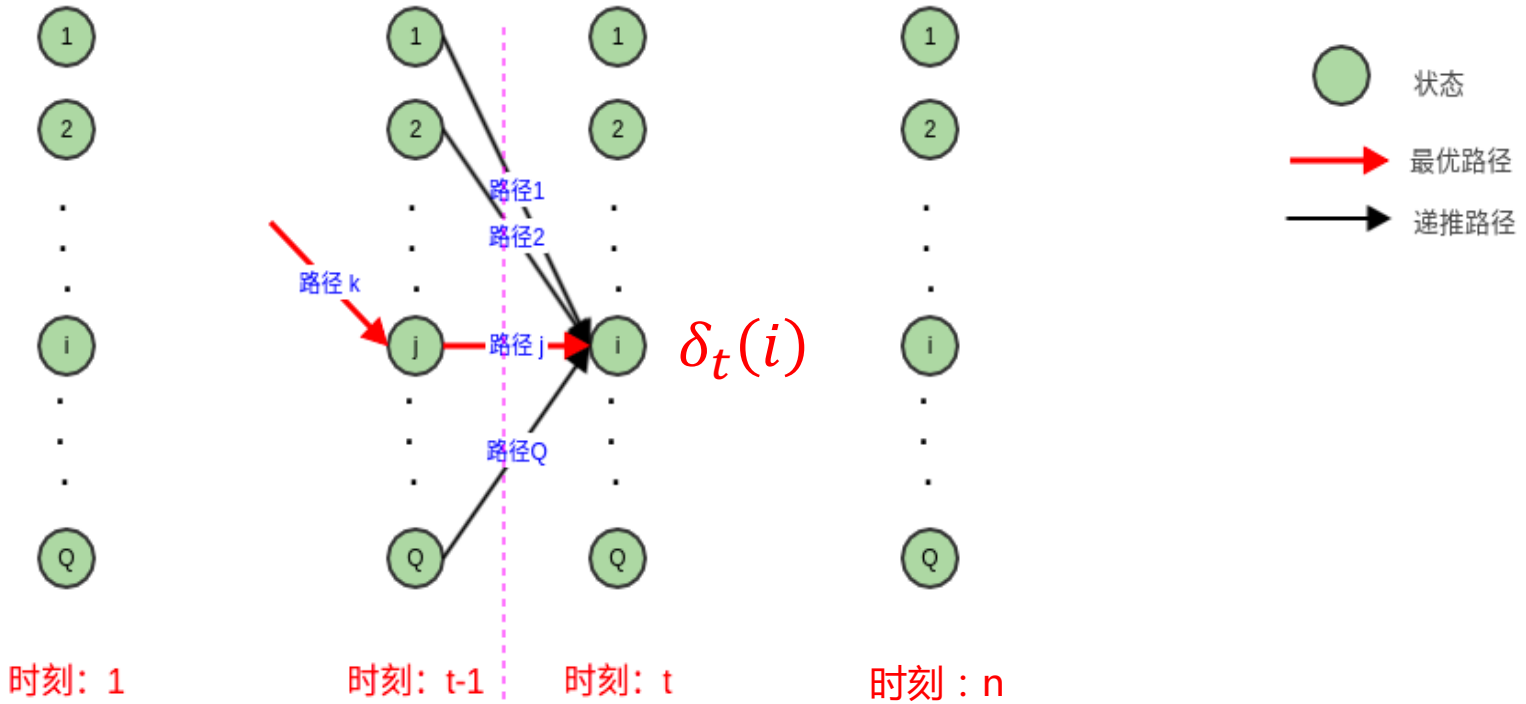
- 算法思想：最优子结构
  - 根据动态规划原理，最优路径具有这样的特性：如果最优路径在时刻 t 通过结点 y_t^* ，则这一路径从结点 y_t^* 到终点 y_n^* 的部分路径，对于从 y_t^* 到 y_n^* 的所有可能路径来说，也必须是最优的。



HMM的预测问题—维特比算法

- 对于观测序列$X = (x_1, x_2, \dots, x_n)$
  - t 时刻状态为 q_i 且已观测序列为$x_1, x_2, \dots, x_t$ 的所有可能路径 $(y_1, y_2, \dots, y_t)$ 中概率最大值为：

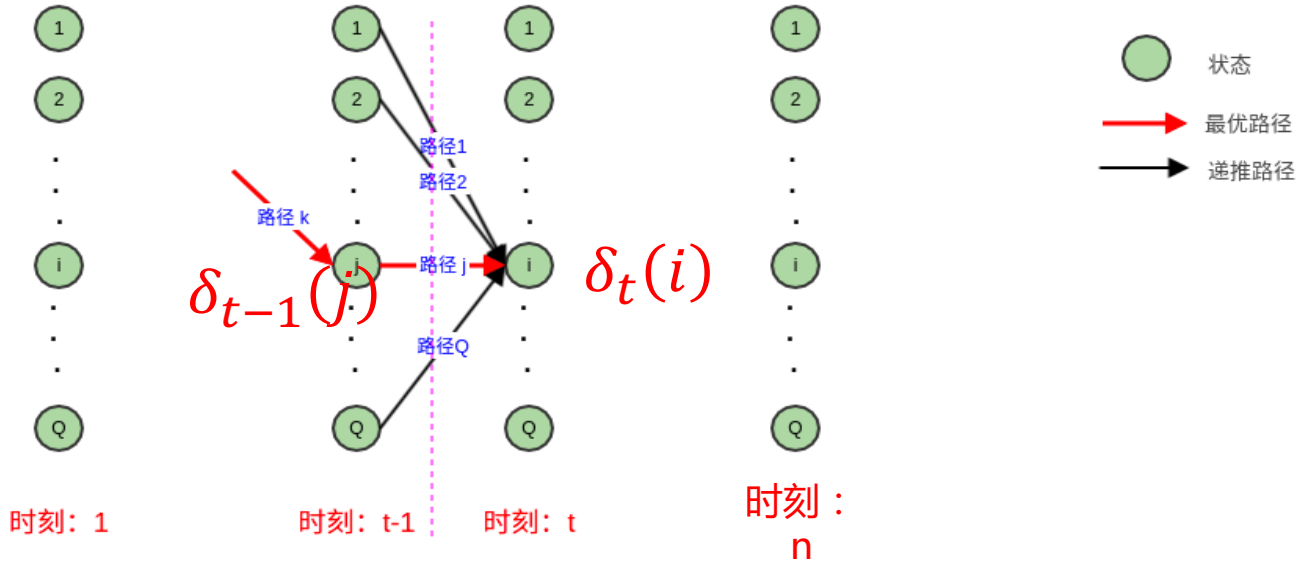
$$\delta_t(i) = \max_{y_1,\dots,y_{t-1}} P(y_1, \dots, y_{t-1}, y_t = q_i, x_1, \dots, x_t), \qquad i = 1,2, \dots, Q$$



HMM的预测问题—维特比算法

- 对于观测序列$X = (x_1, x_2, \dots, x_n)$
  - 得到变量 δ 的递推公式

$$\delta_t(i) = \max_{y_1,\dots,y_{t-1}} P(y_1, \dots, y_{t-1}, y_t = q_i, x_1, \dots, x_t) = \max_{1 \le j \le Q} \delta_{t-1}(j) \times a_{j,i} \times b_i(x_t)$$



- t 时刻状态为 q_i 的所有单个路径中概率最大的路径的第 $t-1$ 个结点为：

$$\Psi_t(i) = \text{argmax}_{1 \le j \le Q} \delta_{t-1}(j) a_{j,i} \ , \qquad i = 1,2, \dots, Q$$

HMM的预测问题—维特比算法

- 输入：隐马尔可夫模型$\lambda = (\mathbf{A}, \mathbf{B}, \pi)$，观测序列$X = (x_1, x_2, \dots, x_n)$
- 输出：最优的状态路径 $Y^* = (y_1^*, y_2^*, \dots, y_n^*)$
- 算法流程：
  - 初始化：$\delta_1(i) = \pi_i b_i(x_1), \Psi_1(i) = 0, \quad i = 1,2, \dots, Q$
  - 递推：$\delta_t(i) = \max_{1 \le j \le Q} \delta_{t-1}(j) \times a_{j,i} \times b_i(x_t)$

    $\Psi_t(i) = \operatorname{argmax}_{1 \le j \le Q} \delta_{t-1}(j) a_{j,i}$

    $i = 1,2, \dots, Q; \quad t = 2, \dots, n$
  - 终止：$P^* = \max_{1 \le i \le Q} \delta_n(i), \quad y_n^* = \operatorname{argmax}_{1 \le j \le Q} \delta_{t-1}(j) a_{j,i}$
  - 最优路径回溯：$y_t^* = \Psi_{t+1}(y_{t+1}^*), \quad t = n-1, \dots, 1$
  - 获得最优路径 $Y^* = (y_1^*, y_2^*, \dots, y_T^*)$。

HMM生成观测序列的过程

- 输入：隐马尔可夫模型$\lambda = (\mathbf{A}, \mathbf{B}, \pi)$和观测序列长度$n$
- 输出：观测序列$X = (x_1, x_2, \ldots, x_n)$
- 算法步骤：
  - 按照初始状态分布 π 产生状态 y_1
  - 令 $t = 1$ ，开始迭代。迭代条件： $t \leq n$。迭代步骤为：
    - 按照状态 y_t 的观测概率分布 $b_j(k)$ 生成观测值 x_t
    - 按照状态 y_t 的状态转移分布 $a_{i,j}$ 产生状态 y_{t+1}
    - 令 $t = t + 1$

HMM的遗留问题

- 由于观测独立性假设（任意时刻的观测只依赖于该时刻的马尔可夫链的状态），很难融入更多的特征（如上下文）以表示复杂的关系

- Label bias问题：由于马尔可夫假设使得在计算转移概率时做了局部归一化，算法倾向于选择分支较少的状态

南京大学
NANJING UNIVERSITY

Thank you！
Q&A