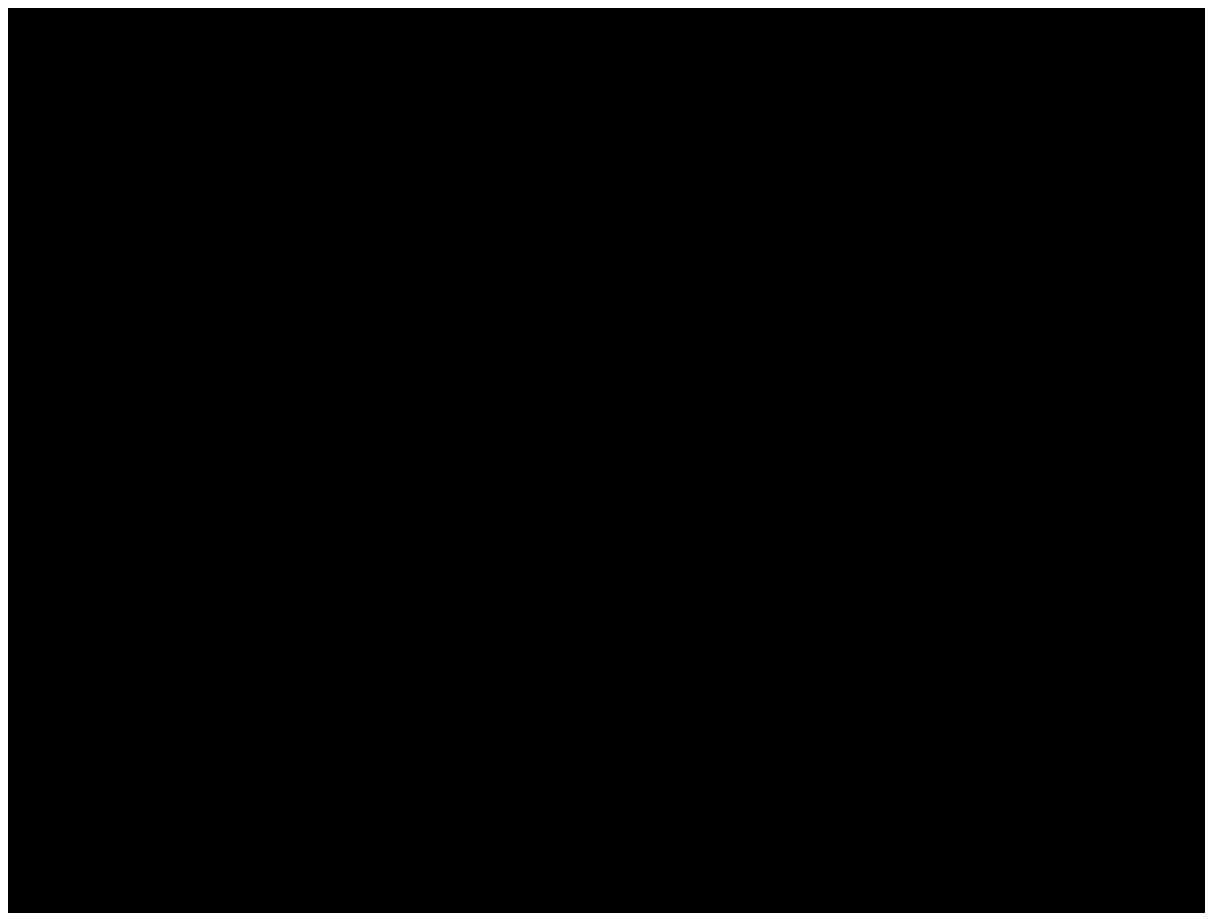


模式识别

HMM: Hidden Markov Model
隐马尔科夫模型

吴建鑫
南京大学计算机系, 2017

连续手写识别



<http://123.3tom.com/xhzd/flash.php?s=C07717179W.swf&h=%CB%E3>

目标

- ✓ 掌握随机过程和马尔科夫性质的基本概念
- ✓ 掌握隐马尔科夫模型（离散观测值）的应用条件和
相关推理算法
- ✓ 了解隐马尔科夫模型（离散观测值）的学习算法
- ✓ 提高目标
 - 进一步能通过独立阅读、了解HMM的实际应用
 - 进一步能通过独立阅读、了解基本的图模型
graphical model的概念、belief propagation
(BP) 算法

Markovian

随机过程stochastic process

- ✓ A stochastic process $\{X(t), t \in T\}$ is a **collection** of random variables. That is, for each $t \in T$, $X(t)$ is a random variable. The index t is often interpreted as time and, as a result, we refer to $X(t)$ as the *state* of the process at time t .
 - The set T is called the *index* set of the process.
 - When T is a countable set \dots a discrete-time process.
 - If T is an interval of the real line, \dots a continuous-time process.
 - The **state space** (状态空间) of a SP is defined as the set of all possible values of that random variables $X(t)$ can assume.
- ✓ A SP is a family of random variables that describes the evolution through time of some (physical) process.

Markov Property

✓ Curse of dimensionality

- $P(X_2|X_1)$ 需要多少存储空间才能指定?
- $P(X_3|X_2, X_1)$ 需要多少存储空间才能指定?
- $P(X_t|X_{1:t-1})$ 需要多少存储空间才能指定?
 - $N^t!$

✓ Markov Property 马尔科夫性质

- 限定: $P(X_t|X_{1:t-1}) = P(X_t|X_{t-1})$, 含义是?
- 无记忆性memoryless
- 这个假设有效吗?
- 好处是什么?

Andrey Markov

http://en.wikipedia.org/wiki/Andrey_Markov

Retrieved Jan 15 2014



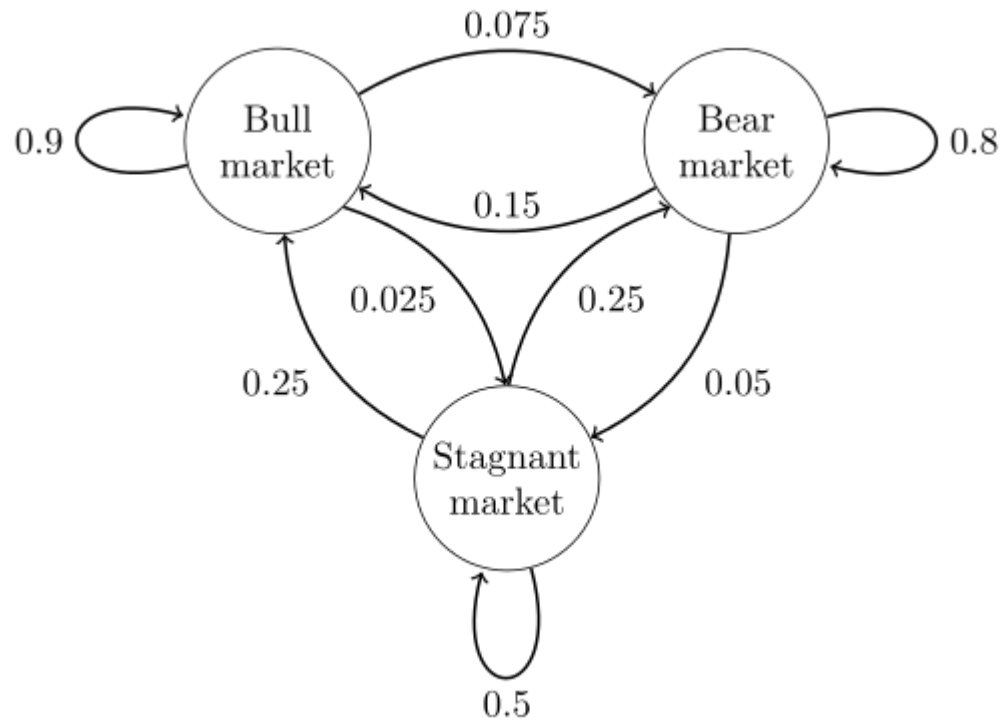
- Chebyshev–Markov–Stieltjes inequalities
- **Hidden Markov model**
- Gauss–Markov process
- Hidden Markov model
- Markov blanket

- Markov chain Monte Carlo
- Markov decision process
- Markov's inequality
- Markov information source
- Markov network

- Markov number
- Markov property
- Markov process
- Subjunctive possibility

Markov Chain马尔科夫链

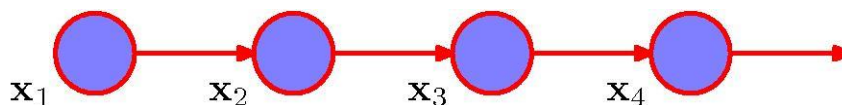
- ✓ Markov chain (discrete-time Markov chain or DTMC)



http://en.wikipedia.org/wiki/Markov_chain

可视化和形式化

✓ 可视化: (PRML Fig. 13.3)



- 注意填充的变量表示观察值（即随机变量值已知）
- ✓ 那么，如何形式化定义DTMC？需要哪些量？
 - 系统初始化Initialization: $P(X_1)$ 或者 $X_1 = x_1$
 - Transition probability: $A = P(X_{t+1}|X_t)$
 - 还需要别的吗？
 - 两次运行结果会一样吗？

转移概率矩阵

✓ Transition probability matrix 转移概率矩阵

- A 是一个 $N \times N$ 的矩阵
- $A_{ij} = P(X_t = j | X_{t-1} = i)$
- 行和为1!

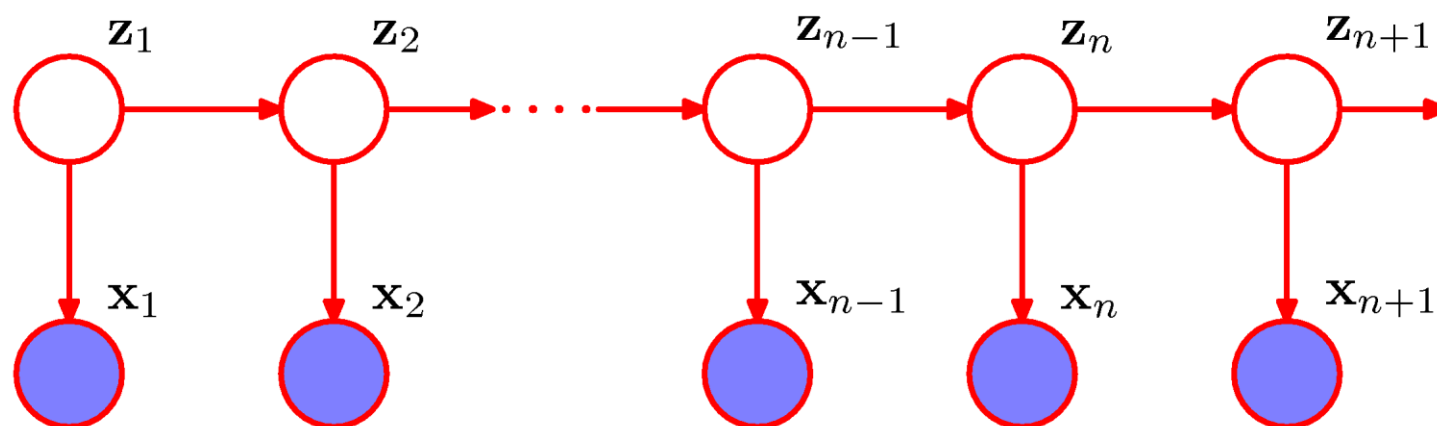
$$P = \begin{bmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}.$$

- ✓ 如果运行足够久 ($t \rightarrow \infty$)，那么 X_t 的分布在很多情况下将稳定下来，叫 Stationary distribution，记为 π
 - $\pi = A\pi$
- ✓ Google PageRank 的简化!

Hidden Markov Model

怎样在模式识别中发挥更大作用？

- ✓ 例如，在连续手写识别中，笔画stroke没有在DTMC里面用到，那么DTMC就没法用于连续手写识别
 - 想办法把笔画加进去！（PRML Fig. 13.5）
 - 状态是什么？



形式化

- ✓ $Q(q)$: 隐变量(hidden variable), 不可观测的状态
- ✓ N : number of states 状态数 $\{S_1, \dots, S_N\}$
- ✓ $O(o)$: 观察值(observations), 记为 O_1, O_2, \dots
- ✓ T : number of possible observation values 观察值数
- ✓ π : 初始化, $\pi_j = P(q_1 = S_j)$
- ✓ A : transition probability matrix, $A_{ij} = P(q_{t+1} = S_j | q_t = S_i)$
- ✓ B : emission probability 发出观察值的概率
 - $b_{S_j}(k) = P(o_t = k | q_t = S_j)$
 - 假设 B 不随时间变化, 当未知状态为 j 时观察到为 k 的概率
 - 那么, j, k 的取值范围是? B 的行和是?

HMM中要解决的问题

- ✓ 怎样设计状态？—— 自动学习？
- ✓ 怎样设计观察值？—— 根据问题的特点和实践反复设计
- ✓ 与具体问题无关的
 - 指定一个HMM需要的所有参数： $\lambda = (\boldsymbol{\pi}, A, B)$
 - 问题1：Evaluation估值
 - 问题2：Decoding解码
 - 问题3：Learning学习

Problem 1. Evaluation

✓ 输入

- 一个完全指定的HMM模型，即 $\lambda = (\pi, A, B)$ 已知
- 一个完全观测的输出序列 $O_1 O_2 \cdots O_T$ ，或 $\mathbf{O} = O_{1:T}$

✓ 输出

- $P(\mathbf{O}|\lambda)$ - 含义是？
- 在这个模型 λ 中观察到特定输出 \mathbf{O} 的概率

✓ 作用是？

- 可以看出此模型对该观察序列的成绩score
- 可以用来从多个模型中选择最适合的模型

Problem 2: Decoding

✓ 输入

- 一个完全指定的HMM模型，即 $\lambda = (\boldsymbol{\pi}, A, B)$ 已知
- 一个完全观测的输出序列 $O_1 O_2 \cdots O_T$ ，或 $\boldsymbol{O} = O_{1:T}$
- 某个标准criterion

✓ 输出

- 一个完全指定的隐变量序列 $X_{1:T}$ 的值

✓ 作用是？

- 如，语音识别中状态可能有实际意义（各音节）
 - 唯一吗？
- 可以用来观察模型结构，优化模型

Problem 3: Learning

✓ 输入

- 网络结构，状态数、输出数
- 若干观测序列 $\{\mathbf{O}\}$

✓ 输出

- 最优的参数 $\lambda = (\boldsymbol{\pi}, A, B)$ 使得 $P(\{\mathbf{O}\}|\lambda)$ 最大

✓ 作用

- 显而易见
- 最重要的问题
- 有时候一个足够长的观测序列就够了

进一步的阅读

✓ 如果对本章的内容感兴趣，可以参考如下文献

- HMM Tutorial:

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=18626&tag=1

- HMM software: HTK in <http://htk.eng.cam.ac.uk/>

- PRML 13.1, 13.2

- Graphical model: PRML chapter 8, 9, 10, 11, 12, 13

- Graphical model: “Probabilistic Graphical Models: Principles and Techniques” by Daphne Koller and Nir Friedman;

<http://mitpress.mit.edu/books/probabilistic-graphical-models>