



西安交通大学
XI'AN JIAOTONG UNIVERSITY

IAIR Est. 1986

Institute of
Artificial Intelligence
and Robotics



人工智能学院
College of Artificial Intelligence, XJTU

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

《高级机器学习》第三章

贝叶斯分类与概率图模型

Bayesian Classification & Probabilistic Graphical Models

魏平

西安交通大学人工智能学院

人工智能与机器人研究所

为什么要使用概率

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

□ 目标分类：输入图像 x , 求其类别 y

x



which city?

$$y = \begin{cases} LA \\ SH? \\ BL \end{cases}$$

方式 ①

$$y = f(x) = LA$$

方式 ②

$$g(y = LA) = 0.70$$

$$g(y = SH) = 0.18$$

$$g(y = BL) = 0.12$$

方式 ③

$$g(y) = \frac{1}{Z} \exp\{-E(y)\}$$

$$E(y = LA) = 0.01$$

$$E(y = SH) = 0.85$$

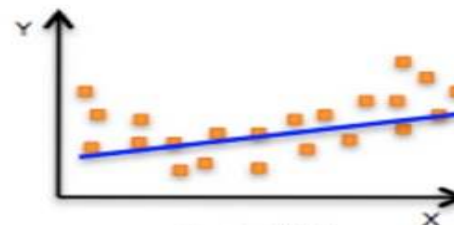
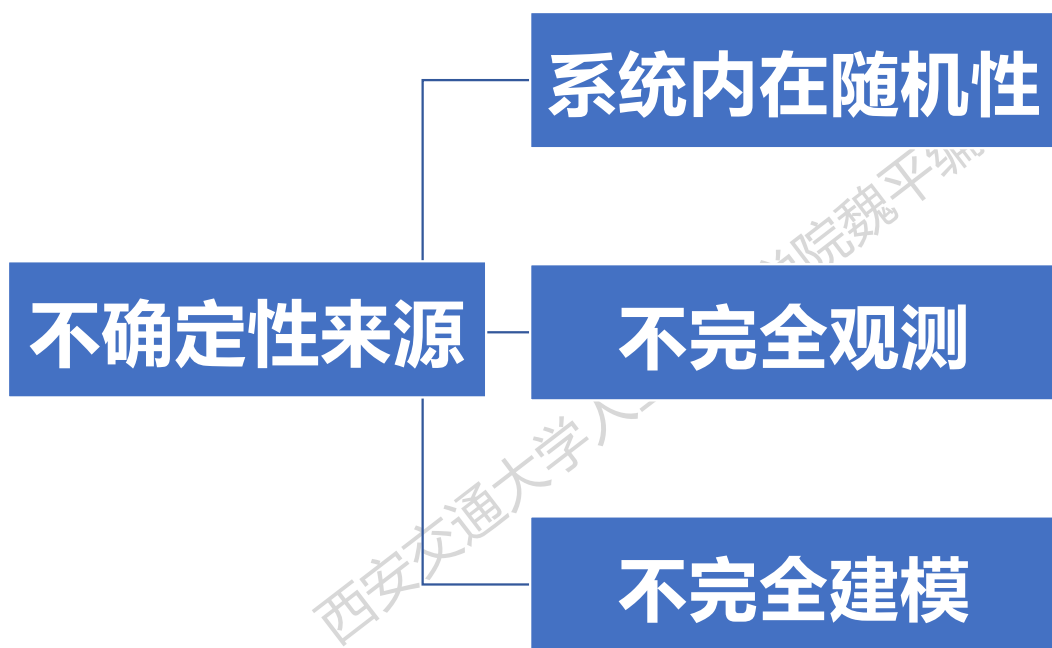
$$E(y = BL) = 3.5$$

→ $y = LA$

为什么要使用概率

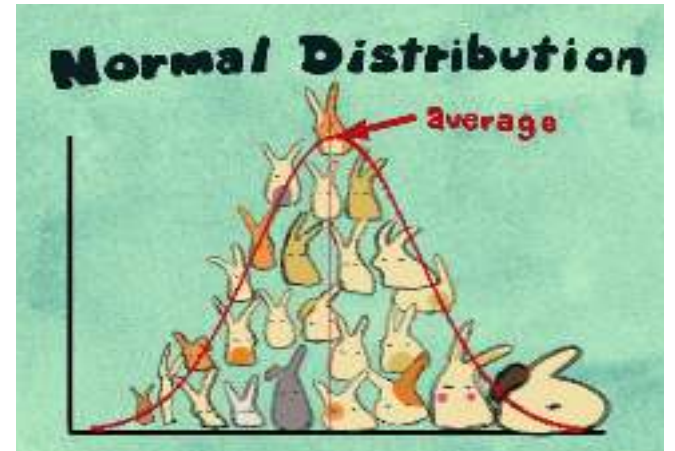
西安交通大学人工智能学院魏平编写。课程资料，请勿外传

- 世界上的万事万物充满不确定性，确定性是不确定性的特例，人工智能和机器学习要处理不确定量以及随机量



为什么要使用概率

- 概率是对一个或多个事件发生可能性大小的度量，概率论提供了量化不确定性以及导出新的不确定性的方法，即提供了一套形式化规则，根据这些规则，在给定一些命题为真的可能性后，可以计算其他命题为真的可能性
- **频率概率：一个事件发生的概率为 p ，即在无限次实验中，有 p 的比例发生该事件**
- **贝叶斯概率：概率表示对事件的一种置信度 (belief) ， 1表示非常肯定发生， 0表示非常肯定不会发生**
- 信息论提供了量化概率分布中不确定性的总量的方法
- **在人工智能领域，概率理论的两种用途**
 - **利用概率法则进行AI系统推理**
 - **利用概率理论分析AI系统行为**



概率论的基本原理

□ 对称原理

$$p(x, y) = p(y, x)$$



$$p(a, b, c) = p(c, a, b)$$

$$p(a, b, c) = p(a, c, b)$$

□ 乘法原理

$$p(x, y) = p(y|x) p(x)$$



$$p(a, b, c) = p(a|b, c)p(b, c)$$

$$p(a, b, c) = p(a, c|b)p(b)$$

$$p(a, b, c|d) = p(c|a, b, d)p(a, b|d)$$

□ 加法原理

离散变量 $p(x) = \sum_y p(x, y)$



连续变量 $p(x) = \int p(x, y) dy$

$$p(a, b) = \int p(a, b, c) dc$$

$$p(a) = \int p(a, b, c) db dc$$

CONTENTS



西安交通大学人工智能学院魏平编写。课程资料，请勿外传

IAIR Est. 1986
Institute of
Artificial Intelligence
and Robotics



人工智能学院
College of Artificial Intelligence, XJTU

- **贝叶斯决策与贝叶斯分类**
- **有向图与无向图一般概念**
- **马尔可夫随机场**
- **隐马尔可夫模型**
- **条件随机场**
- **图模型推理**
- **图模型应用例子**

贝叶斯定理

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

$p(y)$: 先验概率 prior probability

$p(y|x)$: 后验概率 posterior probability

$p(x|y)$: 似然函数 likelihood function

posterior \propto likelihood \times prior



托马斯·贝叶斯
Thomas Bayes,
1702-1761



Judea Pearl

贝叶斯分类器

□ 估计后验概率 $p(y|\mathbf{x})$ 是核心问题

- 方法一：直接建模和估计 $p(y|\mathbf{x})$ ，**判别式模型**，如决策树、神经网络、SVM
- 方法二：建模和估计联合分布 $p(\mathbf{x}, y)$ ，**生成式模型**

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}, y)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|y)p(y)}{p(\mathbf{x})}$$

$p(y)$: 先验概率: 样本空间中各类样本的先验分布

$p(y|\mathbf{x})$: 后验概率: 所求目标概率

$p(\mathbf{x}|y)$: 似然函数: 样本特征属性的联合概率



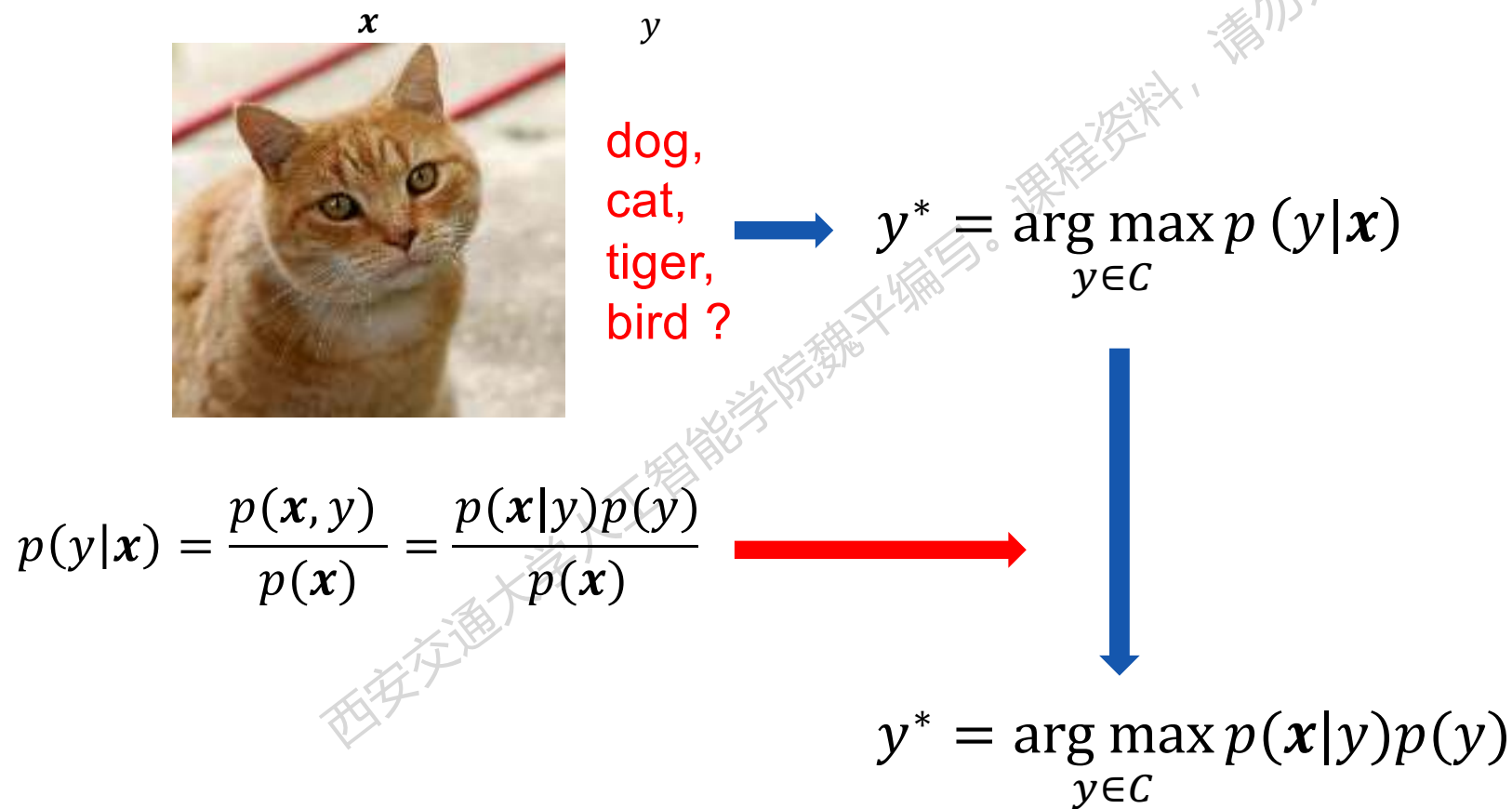
\mathbf{x}

dog,
cat, $\longrightarrow p(y|\mathbf{x})$?
tiger,
bird ?

y

例1-目标识别

- 贝叶斯定理的强大作用是提供了知识转化和推理的框架



例1-目标识别

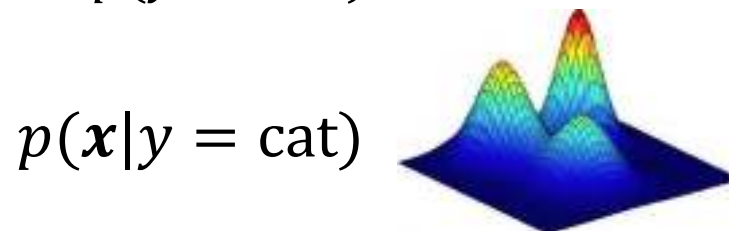
西安交通大学人工智能学院魏平编写。课程资料，请勿外传

$$y^* = \arg \max_{y \in C} p(\mathbf{x}|y)p(y)$$

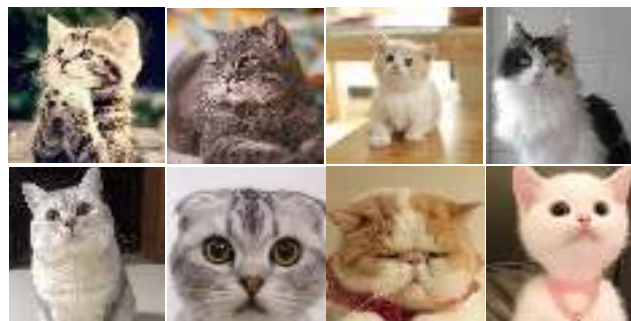
$p(y)$: 先验概率
样本空间中各类
别的先验分布



$$\begin{bmatrix} p(y = \text{dog}) \\ p(y = \text{cat}) \\ p(y = \text{tiger}) \\ p(y = \text{bird}) \end{bmatrix} = \begin{bmatrix} 0.15 \\ 0.65 \\ 0.15 \\ 0.05 \end{bmatrix}$$



$p(\mathbf{x}|y)$: 似然函
数，样本特征属
性的联合概率



例2-视频中人的注意力推理

清华大学人工智能学院魏平编写。课程资料，请勿外传



locate the dispenser



check the cup's status



check the book's status

$$p(\mathbf{X}, \mathbf{l}, \mathbf{Y}) =$$

$$\underbrace{p(l_1)p(y_1|l_1)}_{\text{加阐技荣}} \prod_{t=1}^{\tau} \underbrace{p(\phi(\mathbf{x}_t) | l_t, \mathbf{y}_t)}_{\text{寮深空}} \prod_{t=2}^{\tau} \underbrace{p(l_t | l_{t-1})}_{\text{寮深}} \prod_{t=1}^{\tau} \underbrace{p(\phi(\mathbf{x}_t) | \mathbf{y}_t, l_t)}_{\text{寮襟响喉}} \prod_{t=2}^{\tau} \underbrace{p(\mathbf{y}_t | \mathbf{y}_{t-1}, l_t, l_{t-1})}_{\text{寮襟}}$$

l_t 隐意图

y_t 注意力

\mathbf{x}_t 视频帧

ψ_t 意图特征

ϕ_t 注意力特征

$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_\tau)$ 输入视频数据

$\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_\tau)$ 注意力序列

$\mathbf{l} = (l_1, \dots, l_\tau)$ 隐意图序列

例2-视频中人的注意力推理

上海交通大学人工智能学院魏平编写。课程资料，请勿外传



Input RGBD Video



Reference RGB Video

朴素贝叶斯分类器(naïve Bayes)

- 建模或估计样本属性的联合概率 $p(\mathbf{x}|y)$ 是一个困难问题
- 朴素贝叶斯分类：假设所有属性相互独立

假设 \mathbf{x} 有 d 个属性， x_i 为 \mathbf{x} 的第 i 个属性

$$p(y|\mathbf{x}) = \frac{p(y)}{p(\mathbf{x})} \prod_{i=1}^d p(x_i|y)$$

- D_y 表示训练集 D 中 y 类样本组成的集合；对离散属性，令 D_{y,x_i} 是 D_y 中在第 i 个属性上取值为 x_i 的样本组成的集合，则

$$p(x_i|y) = \frac{|D_{y,x_i}|}{|D_y|}$$

朴素贝叶斯分类器(naïve Bayes)

- 对连续属性，可考虑概率密度为高斯函数

$$p(x_i|y) = \frac{1}{\sqrt{2\pi}\sigma_{y,i}} \exp\left\{-\frac{1}{2\sigma_{y,i}^2} (x_i - \mu_{y,i})^2\right\}$$

- 例子：用朴素贝叶斯分类器判断是否“好瓜”

编号	色泽	根蒂	敲声	纹理	脐部	触感	密度	含糖率	好瓜
测 1	青绿	蜷缩	浊响	清晰	凹陷	硬滑	0.697	0.460	?

西安

半朴素贝叶斯分类器(Semi-naïve Bayes)

- 朴素贝叶斯分类器假设所有属性独立，该假设“过强”，无法刻画属性之间的联系
- 对属性独立假设进行一定程度的放松，由此产生“半朴素贝叶斯分类器” (semi-naïve Bayes classifiers)
- 半朴素贝叶斯分类器最常用的一种策略：“独依赖估计”(One-Dependent Estimator,简称ODE)，假设每个属性在类别之外最多仅依赖一个其他属性，即

$$p(y|\mathbf{x}) \propto p(y) \prod_{i=1}^d p(x_i|y, par_i)$$

par_i 为属性 x_i 依赖的属性，称为 x_i 的父属性

CONTENTS



西安交通大学人工智能学院魏平编写。课程资料，请勿外传

IAIR Est. 1986
Institute of
Artificial Intelligence
and Robotics



人工智能学院
College of Artificial Intelligence, XJTU

- 贝叶斯决策与贝叶斯分类
- 有向图与无向图一般概念
- 马尔可夫随机场
- 隐马尔可夫模型
- 条件随机场
- 图模型推理
- 图模型应用例子

视频中人的注意力推理

西安交通大学人工智能学院魏平编写。课程资料，请勿外传



locate the dispenser



check the cup's status



check the book's status

$$p(\mathbf{X}, \mathbf{l}, \mathbf{Y}) =$$

$$\underbrace{p(l_1)p(y_1|l_1)}_{\text{加闹技荣}} \prod_{t=1}^{\tau} \underbrace{p(\phi(\mathbf{x}_t) | l_t, \mathbf{y}_t)}_{\text{寮深空}} \prod_{t=2}^{\tau} \underbrace{p(l_t | l_{t-1})}_{\text{寮深}} \prod_{t=1}^{\tau} \underbrace{p(\phi(\mathbf{x}_t) | \mathbf{y}_t, l_t)}_{\text{寮襟响喉}} \prod_{t=2}^{\tau} \underbrace{p(\mathbf{y}_t | \mathbf{y}_{t-1}, l_t, l_{t-1})}_{\text{寮襟}}$$

l_t 隐意图

y_t 注意力

\mathbf{x}_t 视频帧

ψ_t 意图特征

ϕ_t 注意力特征

$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_\tau)$ 输入视频数据

$\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_\tau)$ 注意力序列

$\mathbf{l} = (l_1, \dots, l_\tau)$ 隐意图序列

概率图模型

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

- 人工智能应用系统中的随机变量**多且复杂**，如何分析这些变量之间的关系、建立智能系统的模型是一个挑战性的问题
- 概率图模型提供了直观简便的方式分析随机系统的结构和变量之间的关系
- 概率图模型提供了洞悉系统性质、创造新模型的方式
- 概率图模型为复杂学习和推理计算提供了可视化工具



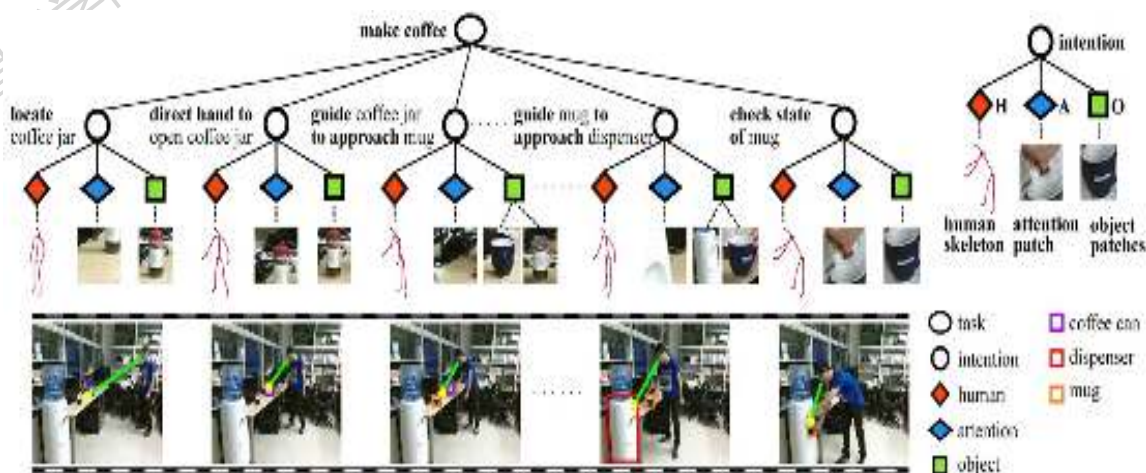
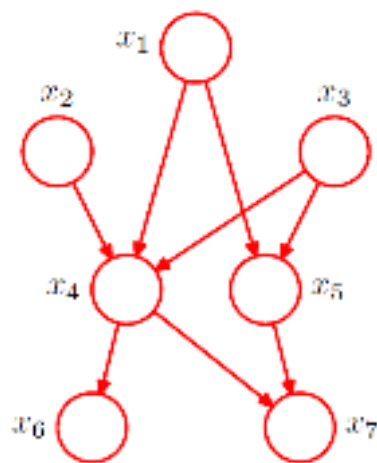
Judea Pearl
UCLA, 2011 图灵奖



Michael Jordan
UC Berkeley

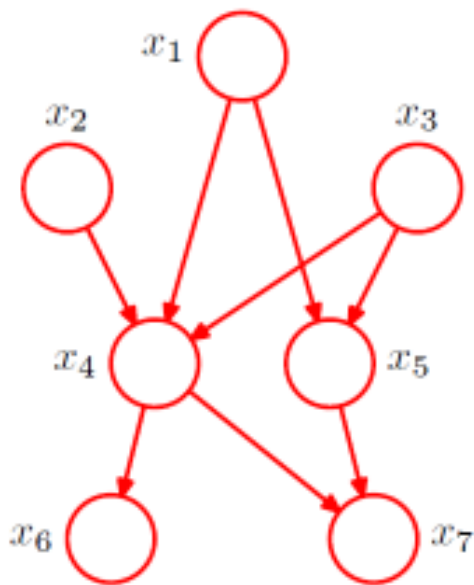
概率图模型

- 概率图包含节点和边，每个节点表示一个（组）随机变量（一维或多维），边表示变量之间的概率关系
- 概率图以可视化的方式表达了多个随机变量之间的依赖关系，**每个概率图是图中所有变量的联合分布**
- 概率图分为有向图和无向图；无向图中的边无方向，有向图中的边有明确的由一个节点到另一个节点的方向



有向图 (Directed graphical models)

- 有向图中的边有明确的由一个节点到另一个节点的方向，也称作**贝叶斯网络 (Bayesian Networks)**
- 一个有向图中所有变量的联合分布为每个节点以其所有父节点为条件的概率分布乘积



$$p(x_1, \dots, x_K) = \prod_{k=1}^K p(x_k | \text{par}(x_k))$$

$$p(x_1, \dots, x_7) = p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3) \\ \cdot p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)$$

有向图 (Directed graphical models)

一般表达

$$p(x_1, \dots, x_K) = p(x_K | x_1, \dots, x_{K-1}) \dots p(x_2 | x_1) p(x_1)$$

有向图表达

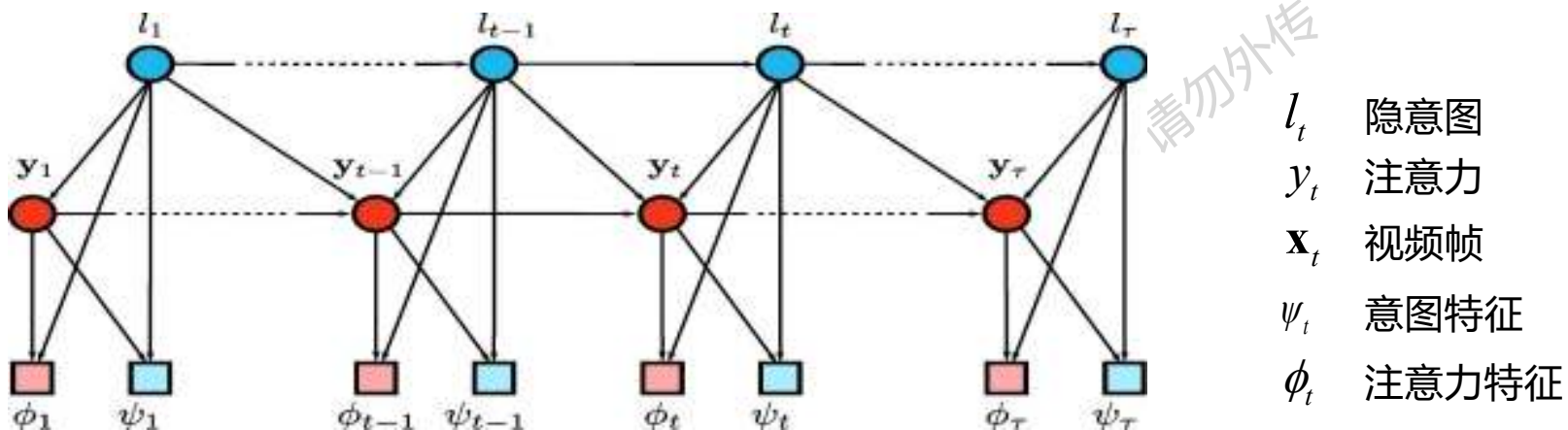
$$p(x_1, \dots, x_K) = \prod_{k=1}^K p(x_k | \text{par}(x_k))$$

以上两种表达的区别是什么、关系是什么：

一般表达是由概率原则得到的任何情况都正确的分解表达，其假设所有变量之间都有关系；有向图表达是在某种具体图结构下的表达，其不同变量之间引入了独立性假设

例子：视频中人的注意力推理

西安交通大学人工智能学院魏平编写。课程资料，请勿外传



$$p(\mathbf{X}, \mathbf{l}, \mathbf{Y}) =$$

$$\underbrace{p(l_1)p(y_1 | l_1)}_{\text{加阐技荣}} \prod_{t=1}^{\tau} \underbrace{p(\phi(\mathbf{x}_t) | l_t, y_t)}_{\text{寮深空}} \prod_{t=2}^{\tau} \underbrace{p(l_t | l_{t-1})}_{\text{寮深}} \prod_{t=1}^{\tau} \underbrace{p(\psi(\mathbf{x}_t) | y_t, l_t)}_{\text{寮襟响嗷}} \prod_{t=2}^{\tau} \underbrace{p(y_t | y_{t-1}, l_t, l_{t-1})}_{\text{寮襟}}$$

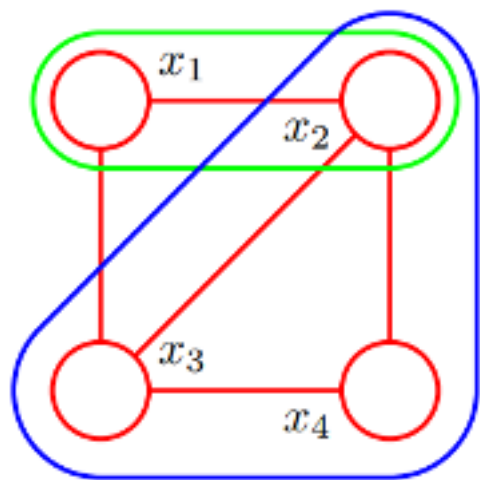
$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_\tau)$ 输入视频数据

$\mathbf{Y} = (y_1, \dots, y_\tau)$ 注意力序列

$\mathbf{l} = (l_1, \dots, l_\tau)$ 隐意图序列

无向图

- 无向图中连接节点的边没有方向，也称作马尔可夫随机场(Markov Random Field)
- 团(clique): 无向图中的节点的子集，其中任何两个点之间都有边相连
- 最大团(maximal clique): 一个团再增加图中任何一个其他节点都不成为团，则该团为最大团



两节点团: $\{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_4\}, \{x_4, x_2\}, \{x_1, x_3\}$

最大团: $\{x_1, x_2, x_3\}, \{x_2, x_3, x_4\}$

非团: $\{x_1, x_2, x_3, x_4\}$

无向图

- 无向图的联合概率分布可以写作图的**最大团**的**势函数**(potential function)的乘积

$$p(\mathbf{x}) = \frac{1}{Z} \prod_C \psi_C(\mathbf{x}_C)$$

C : 最大团; \mathbf{x}_C : 团 C 中的变量集合; $\psi_C(\mathbf{x}_C) \geq 0$: 势函数(potential function)

Z : 配分函数(partition function), 归一化常数: $Z = \sum_{\mathbf{x}} \prod_C \psi_C(\mathbf{x}_C)$

- 若限制 $\psi_C(\mathbf{x}_C) > 0$, 则可以定义 $\psi_C(\mathbf{x}_C) = \exp\{-E(\mathbf{x}_C)\}$, 则 $E(\mathbf{x}_C)$ 称作**能量函数**(energy function), 该指数表达称为**玻尔兹曼分布**(Boltzmann distribution)
- 联合分布定义为势函数的积, 则总能量是每个最大团能量的和

条件独立性 (Conditional Independence)

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

- 两个随机变量**相互独立(independent)** $a \perp b$

$$p(a, b) = p(a)p(b)$$

- 两个随机变量**条件独立(conditional independent)** $a \perp b | c$ a 和 b 关于 c 条件独立

$$p(a|b, c) = p(a|c) \quad p(b|a, c) = p(b|c) \quad p(a, b|c) = p(a|c) p(b|c)$$

条件独立性的重要意义在于利用它能：

- (1) 简化模型结构和变量关系
- (2) 简化模型学习和推理时的计算

有向图条件独立性：D分离方法 (D-separation)

- 如何根据有向图的结构判断条件独立？ - D分离方法
- 三个变量的基本结构 1 — 尾到尾节点 *tail-to-tail*

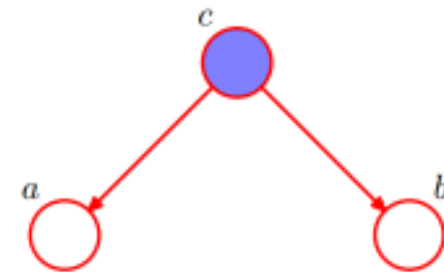
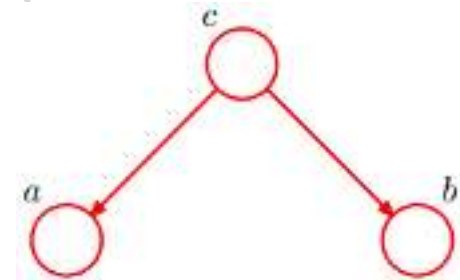
$$p(a, b, c) = p(a|c)p(b|c)p(c)$$

a, b 独立性不满足

$$p(a, b) \neq p(a)p(b)$$

a 和 b 关于 c 条件独立性满足

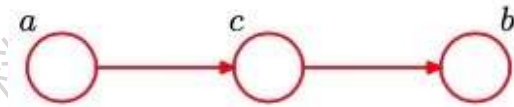
$$p(a, b|c) = \frac{p(a, b, c)}{p(c)} = p(a|c)p(b|c)$$



有向图条件独立性：D分离方法 (D-separation)

- 三个变量的基本结构 2 — 头到尾节点 *head-to-tail*

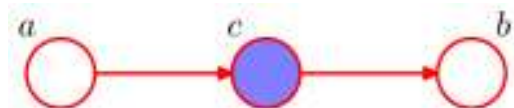
$$p(a, b, c) = p(a)p(c|a)p(b|c)$$



a, b 独立性不满足

$$p(a, b) \neq p(a)p(b)$$

a 和 b 关于 c 条件独立性满足



$$\begin{aligned} p(a, b|c) &= \frac{p(a, b, c)}{p(c)} = \frac{p(a)p(c|a)p(b|c)}{p(c)} = \frac{p(c)p(a|c)p(b|c)}{p(c)} \\ &= p(a|c)p(b|c) \end{aligned}$$

有向图条件独立性：D分离方法 (D-separation)

- 三个变量的基本结构 3 — 头到头节点 *head-to-head*

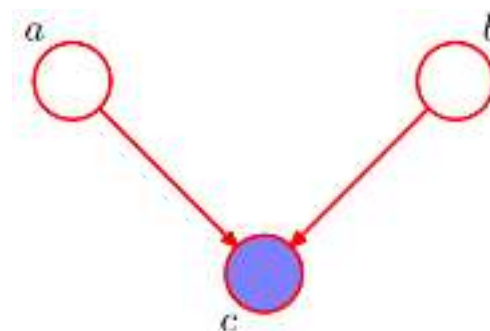
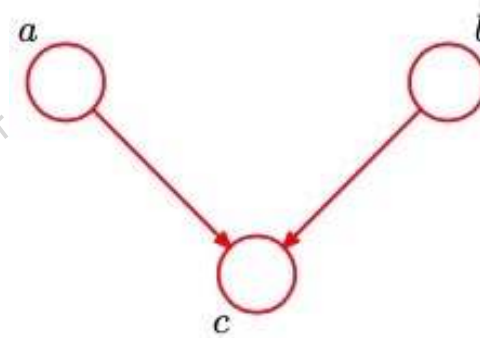
$$p(a, b, c) = p(a)p(b)p(c|a, b)$$

a, b 独立性满足

$$p(a, b) = p(a)p(b)$$

a 和 b 关于 c 条件独立性不满足

$$p(a, b|c) \neq p(a|c)p(b|c)$$



有向图条件独立性：D分离方法 (D-separation)

□ 有向图中判断节点集合条件独立性 — D分离方法

□ 设 A, B, C 是一个有向图中的三个节点子集，判断 $A \perp B | C$ ，考察从 A 中每一个节点到 B 中每一个节点所有的路径。如果任何一个路径包含一个节点满足如下任意一个条件，则称这条路径被阻隔(blocked)：

① 路径上的箭头以**头到尾**或**尾到尾**的方式交汇于这个节点，且这个节点在集合 C 中

② 箭头以**头到头**的方式交汇于这个节点，且这个节点和它的所有后代都不在集合 C 中

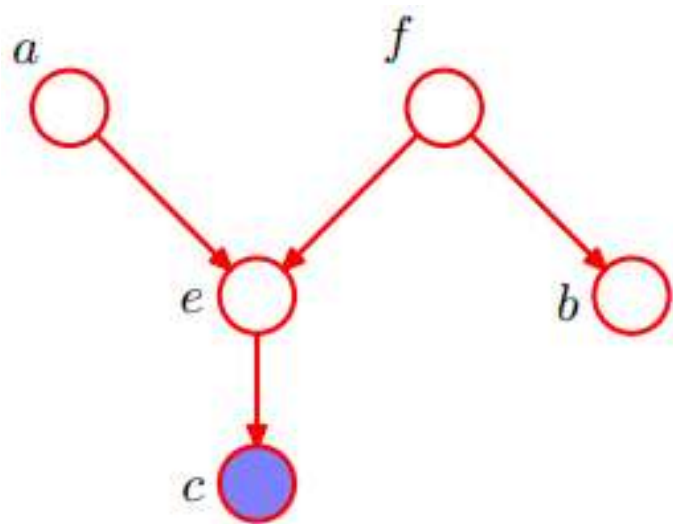
□ 若所有的路径都被阻隔，则称 C 把 A 从 B 中d-分离开，且所有变量上的联合概率分布满足 $A \perp B | C$

如果存在 x 到 y 的一条有向路径，则称 y 是 x 的后代

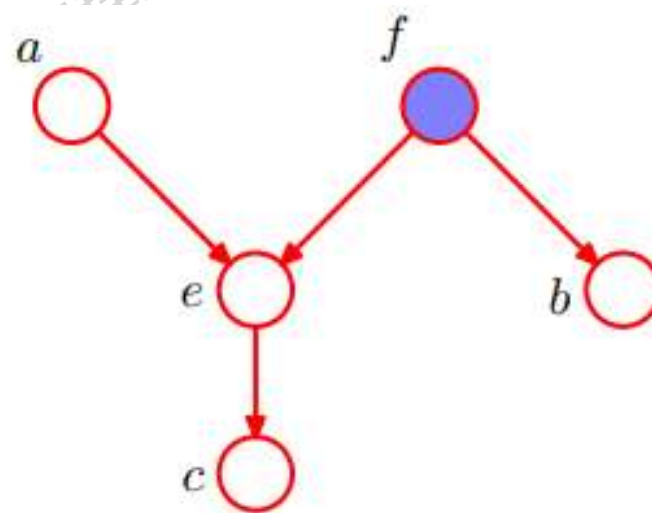
有向图条件独立性: D分离方法 (D-separation)

□ 有向图中判断 $A \perp B|C$, A 到 B 的**每一条路径都**:

- ① 有一个**头到尾节点**或**尾到尾节点**在 C 中, 或
- ② 有一个**头到头节点**, 它及其后代都不在 C 中



$a \perp b|c$ 不满足



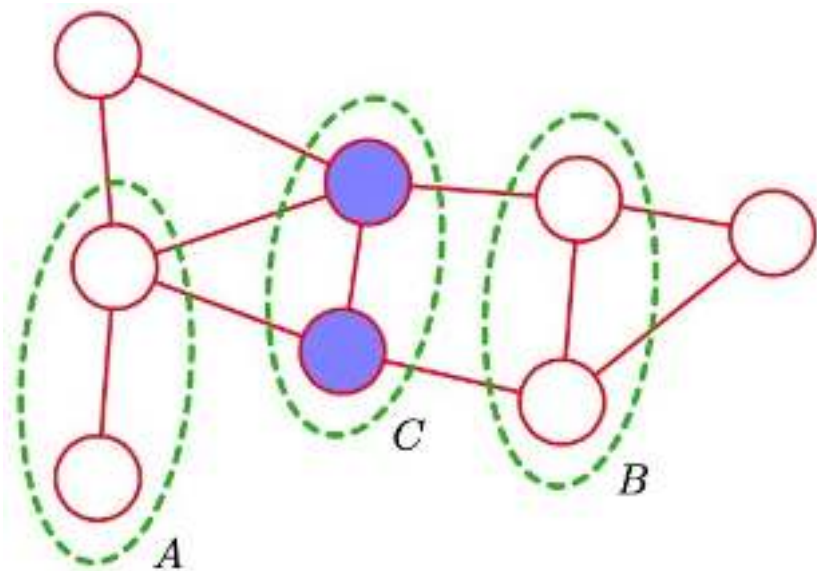
$a \perp b|f$ 满足

无向图条件独立性：条件移除方法

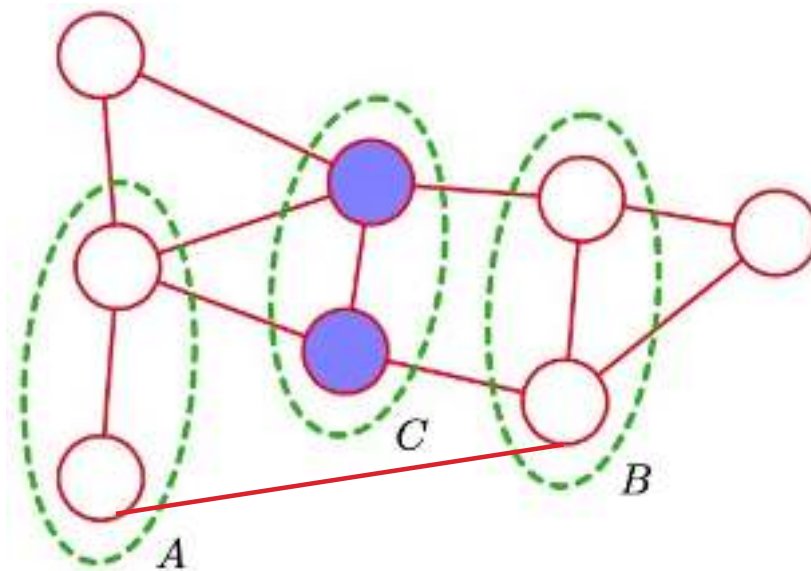
西安交通大学人工智能学院魏平编写。课程资料，请勿外传

□ 无向图中判断 $A \perp B|C$ 是否成立，采用**条件移除法**：

- ① 移除 C 中所有的节点以及连接到这些节点的边
- ② 判断是否存在连接 A 中任一节点和 B 中任一节点的一条路径，若不存在，则 $A \perp B|C$ 成立，否则，不成立



$A \perp B|C$ 成立， $A \perp C|B$ 不成立



$A \perp B|C$ 不成立

无向图条件独立性：条件移除方法

□ 无向图中判断 $A \perp B | C$ 是否成立，采用**条件移除法**：

① 移除 C 中所有的节点以及连接到这些节点的边

② 判断是否存在连接 A 中任一节点和 B 中任一节点的一条路径，若不存在，则 $A \perp B | C$ 成立，否则，不成立

□ 设 x_i 和 x_j 是任意无向图中不直接相连的两个节点， $\mathbf{x}_{\setminus\{i,j\}}$ 是图中除 x_i 和 x_j 的节点集，则

$$p(x_i, x_j | \mathbf{x}_{\setminus\{i,j\}}) = p(x_i | \mathbf{x}_{\setminus\{i,j\}})p(x_j | \mathbf{x}_{\setminus\{i,j\}})$$

即 x_i 和 x_j 不会出现在联合分布 $p(\mathbf{x})$ 分解式的同一个因子中

无向图和有向图的关系

- 无向图和有向图是可以相互转换的；将有向图变换为无向图，在很多情况下，可以将有向边变换为无向边



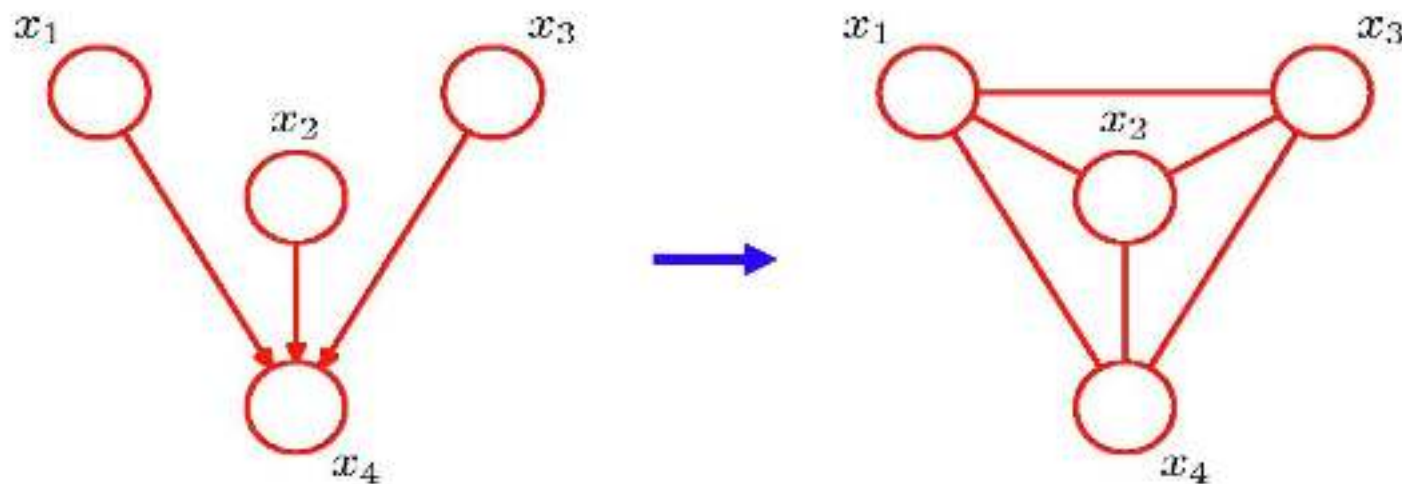
$$p(\mathbf{x}) = p(x_1)p(x_2|x_1)p(x_3|x_2)\cdots p(x_N|x_{N-1})$$

$$p(\mathbf{x}) = \frac{1}{Z} \psi_{1,2}(x_1, x_2) \psi_{2,3}(x_2, x_3) \cdots \psi_{N-1,N}(x_{N-1}, x_N)$$



无向图和有向图的关系

- 某些情况下，需要在共同的父节点之间加入额外的边，该过程称为moralization，转化得到的无向图称为 moral graph



$$p(\mathbf{x}) = p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)$$

CONTENTS



西安交通大学人工智能学院魏平编写。课程资料，请勿外传

IAIR Est. 1986
Institute of
Artificial Intelligence
and Robotics



人工智能学院
College of Artificial Intelligence, XJTU

- 贝叶斯决策与贝叶斯分类
- 有向图与无向图一般概念
- 马尔可夫随机场
- 隐马尔可夫模型
- 条件随机场
- 图模型推理
- 图模型应用例子

马尔可夫

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

□ 俄国数学家、统计学家，随机过程领域的开创者

- 在一系列事件中，某一给定事件发生的概率只取决于以前刚刚发生的那一事件

□ 广泛应用于自然科学和社会科学

e.g. 计算机视觉与机器学习

网页搜索排序算法

语音识别与语言处理

经济增长与衰退



安德雷·安德耶维齐·马尔可夫

1856 -1922



切比雪夫

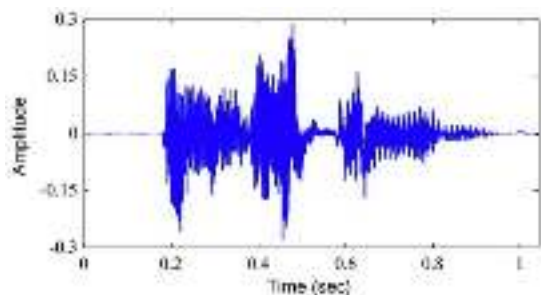


李雅普诺夫

马尔可夫模型

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

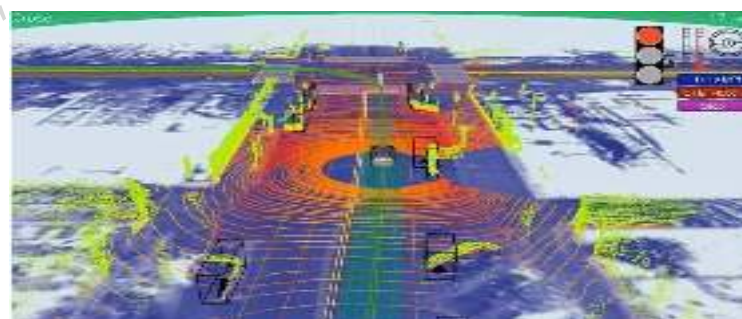
- ❑ 马尔可夫模型主要用来模拟和分析物理现象中的空间、时间或上下文关联作用
- ❑ 空间或上下文关联作用可用无向图来形式化表达，时间关联作用可用有向图表达



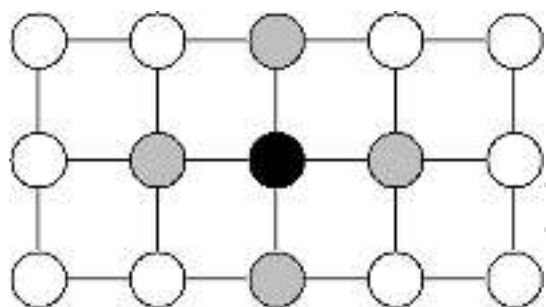
语音识别



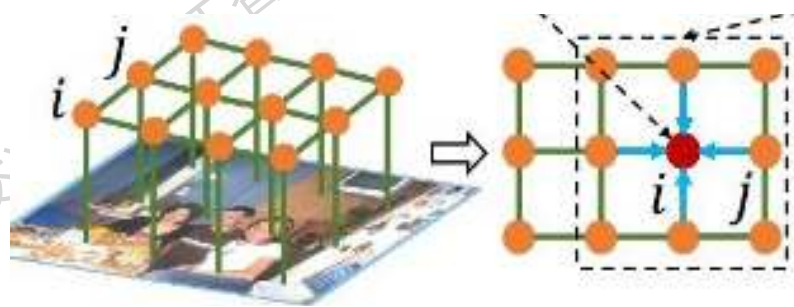
基因翻译



智能体定位



MRF



图像分割



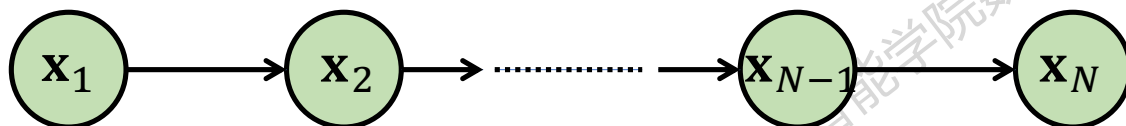
图像去噪

马尔可夫链模型

- $\mathbf{x}_1, \mathbf{x}_2 \dots \mathbf{x}_N$ 是一个时间观测序列，它们的联合分布为

$$p(\mathbf{x}_1, \mathbf{x}_2 \dots \mathbf{x}_N) = p(\mathbf{x}_1) \prod_{n=2}^N p(\mathbf{x}_n | \mathbf{x}_1, \dots, \mathbf{x}_{n-1})$$

- 一阶马尔可夫链模型 (First-order Markov Chain)



$$p(\mathbf{x}_n | \mathbf{x}_1, \dots, \mathbf{x}_{n-1}) = p(\mathbf{x}_n | \mathbf{x}_{n-1})$$

当前时刻的状态独立于除上一时刻状态的之前所有状态

$$p(\mathbf{x}_1, \mathbf{x}_2 \dots \mathbf{x}_N) = p(\mathbf{x}_1) \prod_{n=2}^N p(\mathbf{x}_n | \mathbf{x}_{n-1})$$

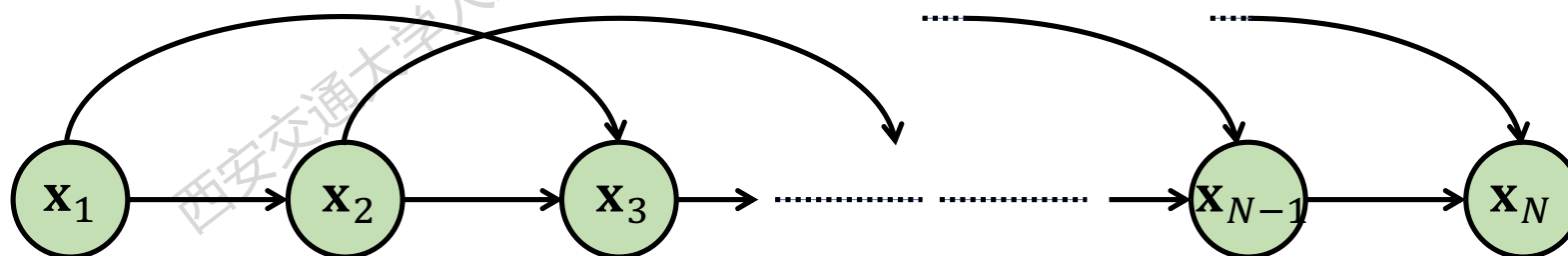
马尔可夫链模型

二阶马尔可夫链模型 (Second-order Markov Chain)

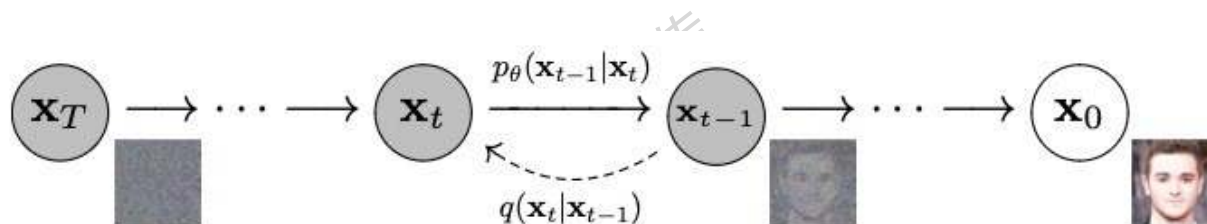
$$p(\mathbf{x}_n | \mathbf{x}_1, \dots, \mathbf{x}_{n-1}) = p(\mathbf{x}_n | \mathbf{x}_{n-1}, \mathbf{x}_{n-2})$$

$$p(\mathbf{x}_1, \mathbf{x}_2 \dots \mathbf{x}_N) = p(\mathbf{x}_1) p(\mathbf{x}_2 | \mathbf{x}_1) \sum_{n=3}^N p(\mathbf{x}_n | \mathbf{x}_{n-1}, \mathbf{x}_{n-2})$$

当前时刻的状态独立于除上两个时刻状态的之前所有状态



例：扩散模型 Diffusion Models



□ 逆向过程 reverse process

$$p_{\theta}(\mathbf{x}_{0:T}) := p(\mathbf{x}_T) \prod_{t=1}^T p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t), \quad p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_{\theta}(\mathbf{x}_t, t), \boldsymbol{\Sigma}_{\theta}(\mathbf{x}_t, t))$$

□ 前向过程 forward process 或 扩散过程 diffusion process

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1}), \quad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

随机场

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

- 随机场是被某种空间上的变量索引的随机变量的集合



随机场基本概念

□ 位点空间 Site Space

有限元素集合 $S = \{s_1, s_2, \dots, s_m\}$ 表示位点的集合，例如时间集合、空间坐标集合等

□ 相空间 Phase Space

有限元素集合 $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 表示变量的取值空间，例如类别、几何属性等

□ 随机场

位点空间 S 中每一个位点从相空间 Λ 中取值的随机变量的集合是空间 S 上的一个随机场

$$\mathbf{x} = \{x(s) | s \in S, x(s) \in \Lambda\}$$



邻域系统

- 一个定义在位点集合 S 上的邻域系统(neighborhood system)定义为

$$N = \{N_s | s \in S\}$$

N_s 是位点 s 的邻域，即所有与 s 相邻的位点的集合

- 一个位点的邻域定义为与此位点距离小于一个阈值的位点的集合

$$N_s = \{s' \in S | d(s, s') \leq \alpha, s' \neq s\}$$

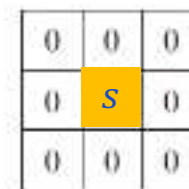
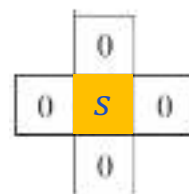
- 邻域系统的两个重要性质

- 一个位点不属于自己的邻域，即 $s \notin N_s$
- 相邻关系是相互的，即 $s \in N_{s'} \Leftrightarrow s' \in N_s$



“海内存知己，天涯若比邻”

“门锁帘垂月影斜，翠华咫尺隔天涯”



5	4	3	4	5
4	2	1	2	4
3	1	s	1	3
4	2	1	2	4
5	4	3	4	5

马尔可夫随机场定义

□ 马尔可夫随机场定义为: 对一个随机场 $\mathbf{x} = \{x(s) | s \in S, x(s) \in \Lambda\}$, 若满足:

$$\forall s \in S, \quad p(x(s) | \mathbf{x}(S - s)) = p(x(s) | \mathbf{x}(N_s))$$

则该随机场为关于邻域系统 N 的马尔可夫随机场

- 含义: 一个事件发生的概率只取决于相邻的事件

- 符号: $p^{(s)}(\mathbf{x}) \stackrel{\text{def}}{=} p(x(s) | \mathbf{x}(S - s)) = p(x(s) | \mathbf{x}(N_s))$ 称为位点 s 的马尔可夫局部特性

□ 一个马尔可夫随机场由以下五个要素决定

① 位点空间 $S = \{s_1, s_2, \dots, s_m\}$

② 相空间 $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$

③ 邻域系统 $N = \{N_s\}$

④ 局部特性 $\{p^{(s)}\}_{s \in S}$

⑤ 概率表达 $p(\mathbf{x})$

如果我们定义一个邻域系统使得任意位点 s 的邻域是整个位点空间, 那么任意随机场都成为MRF。然而, 具有实用价值的马尔可夫随机场, 往往都是邻域系统相对较小的MRF

吉布斯分布

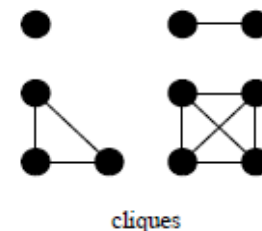
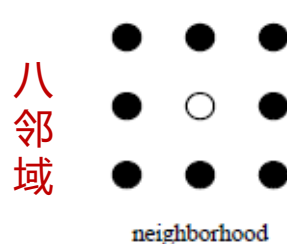
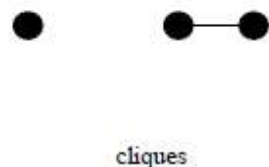
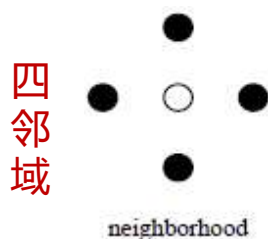
□ 吉布斯1902年提出了吉布斯分布(Gibbs distributions)模拟粒子之间的物理作用

□ N 为位点空间 S 上定义的邻域系统 N ，定义在 S 上的随机场 \mathbf{x} 是一个吉布斯随机场，当且仅当 \mathbf{x} 服从吉布斯分布

$$p(\mathbf{x}) = \frac{1}{Z} e^{-\frac{1}{T}U(\mathbf{x})}$$

● 能量函数: $U(\mathbf{x}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{x})$ ● 配分函数: $Z = \sum_{\mathbf{x}} e^{-\frac{1}{T}U(\mathbf{x})}$ ● 温度: T

□ $V_c(\mathbf{x})$ 是定义在团 c 上的吉布斯势函数。任意单位点 s 是一个团；元素数目大于1的子集 $c \in S$ ，若 c 中任意两个不同位点都相邻，则 c 是一个团。所有团的集合用 \mathcal{C} 表示



吉布斯与马尔可夫的等价性

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

- 一个定义在 (S, Λ, N) 上的马尔可夫随机场的联合概率分布 $p(\mathbf{x})$ 是一个吉布斯分布，即

$$p(\mathbf{x}) = \frac{1}{Z} e^{-U(\mathbf{x})}$$

$U(\mathbf{x})$ 为能量函数，由定义在 S 和 N 上的吉布斯势函数 $\{V_c(\mathbf{x}) | c \in \mathcal{C}\}$ 产生

- 能量函数的一种常见定义方式

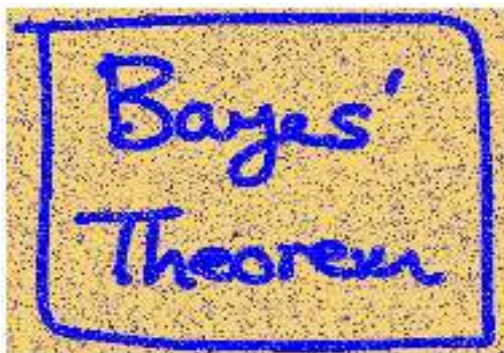
$$U(\mathbf{x}) = \underbrace{\sum_{s \in C_1} V_1(\mathbf{x}_s)}_{\substack{\text{unary term} \\ \text{一元项}}} + \underbrace{\sum_{\{s, s'\} \in C_2} V_2(\mathbf{x}_s, \mathbf{x}_{s'})}_{\substack{\text{binary relation} \\ \text{二元关系}}} + \underbrace{\sum_{\{s, s', s''\} \in C_3} V_3(\mathbf{x}_s, \mathbf{x}_{s'}, \mathbf{x}_{s''})}_{\substack{\text{triple relation} \\ \text{三元关系}}} + \dots$$

无向图应用实例：图像去噪

上海交通大学人工智能学院魏平编写。课程资料，请勿外传



原始图像



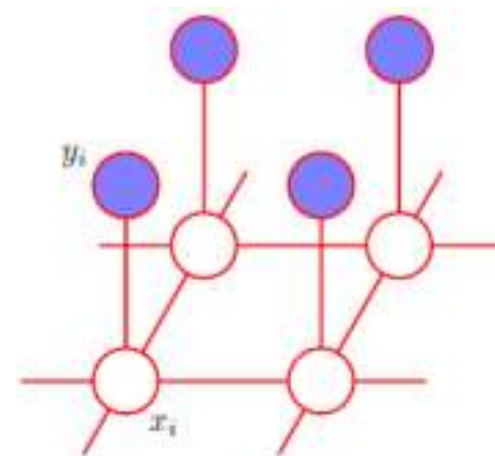
噪声图像y



还原图像x

噪声图像 $y = \{y_i | i = 1, \dots, m\}$, $y_i \in \{+1, -1\}$

还原图像 $x = \{x_i | i = 1, \dots, m\}$, $x_i \in \{+1, -1\}$



Ising Model

西安交通大学人工智能学院魏平编写。课程资料，请勿外传



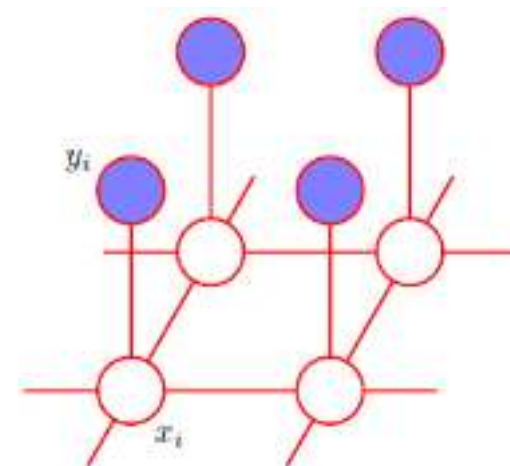
原始图像



噪声图像 y



还原图像 x



- (\mathbf{x}, \mathbf{y}) 上可以定义一个MRF

$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{Z} \exp\{-E(\mathbf{x}, \mathbf{y})\}$$

$$E(\mathbf{x}, \mathbf{y}) = h \sum_i x_i - \beta \sum_{\{i,j\}} x_i x_j - \eta \sum_i x_i y_i$$

- 图像去噪

$$\mathbf{x}^* = \max p(\mathbf{x}|\mathbf{y})$$

CONTENTS



西安交通大学人工智能学院魏平编写。课程资料，请勿外传

IAIR Est. 1986
Institute of
Artificial Intelligence
and Robotics



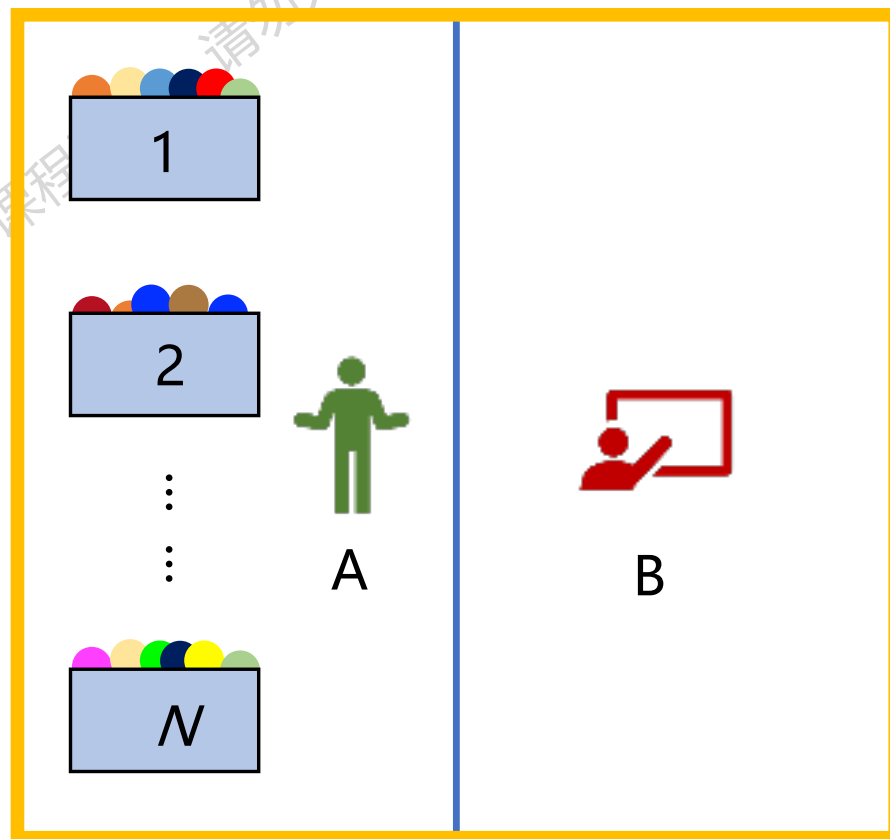
人工智能学院
College of Artificial Intelligence, XJTU

- 贝叶斯决策与贝叶斯分类
- 有向图与无向图一般概念
- 马尔可夫随机场
- **隐马尔可夫模型**
- 条件随机场
- 图模型推理
- 图模型应用例子

引例

□ 彩色球取样实验

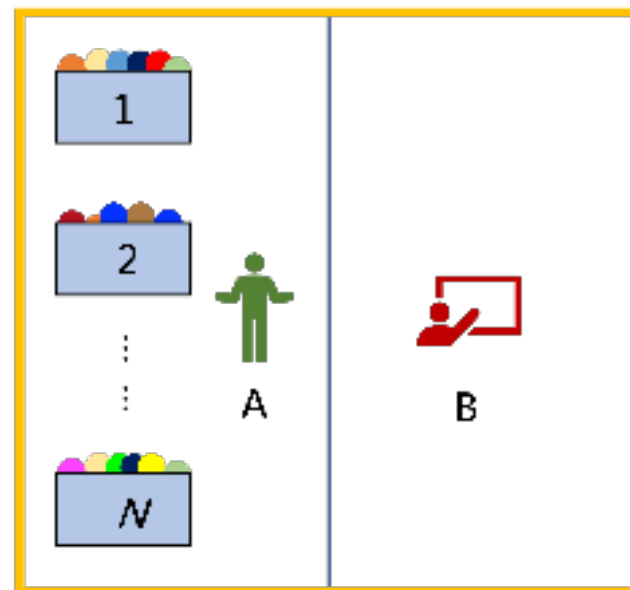
- 用不透光材料将一个房间分为两部分，在其中一边，A进行彩色球取样实验，在另一边，B记录实验结果
- 实验中有 N 个盒子，每个盒子含有大量 M 种不同颜色的球
- 实验从一个初始盒子开始，A从该盒子中随机抽取一个球，将球的颜色告诉B，然后把球放回它被取出的盒子里。除了球的颜色外，B得不到其它任何信息
- 接下来A会依照与当前盒子有关的一个随机过程选取一个新的盒子，重复上述随机抽取球的实验，并将球的颜色告诉B



引例

- 通过上述彩色球抽取实验，观测者得到一系列球的颜色记录。问题是：
 - 能否计算出观测到某一特定颜色序列的概率？
 - 给定特定的颜色观测序列，能否计算出最有可能的球被抽取出的盒子序列？
 - 给定多个观测序列，能否计算出两个盒子之间的转移概率，以及从每一个盒子中观测到特定颜色的概率？

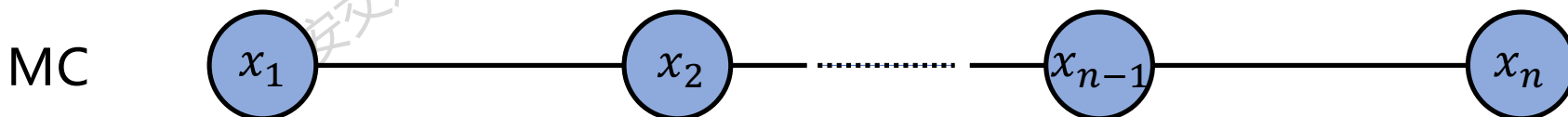
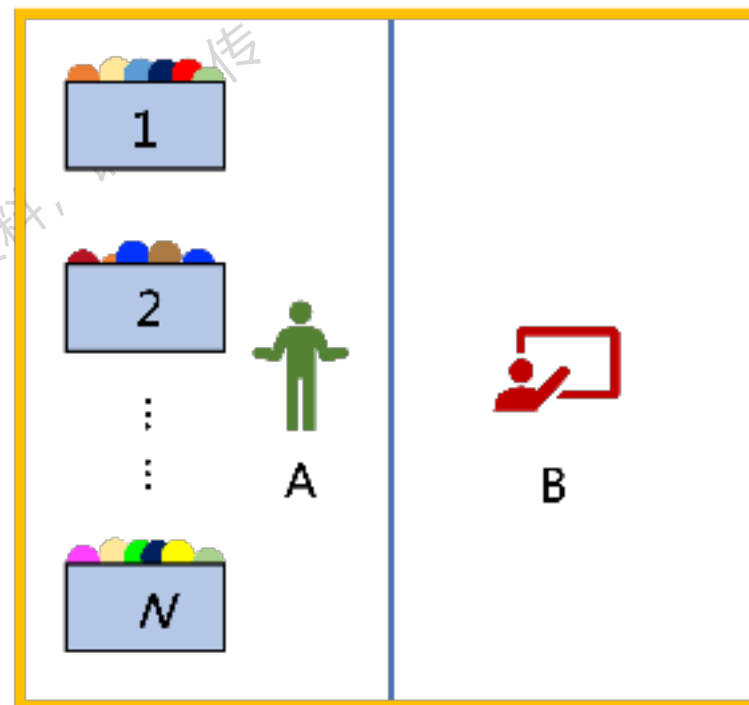
- 上述彩色球抽取实验中存在着两个相关的随机过程
 - 从被选到的盒子中随机抽取彩色球并观察其颜色，这是一个**可观测**的过程
 - 随机选择抽取彩色球的盒子，这是一个**隐藏**的过程



引例

❑ 马尔可夫链（随机场）并不足以模拟这一过程

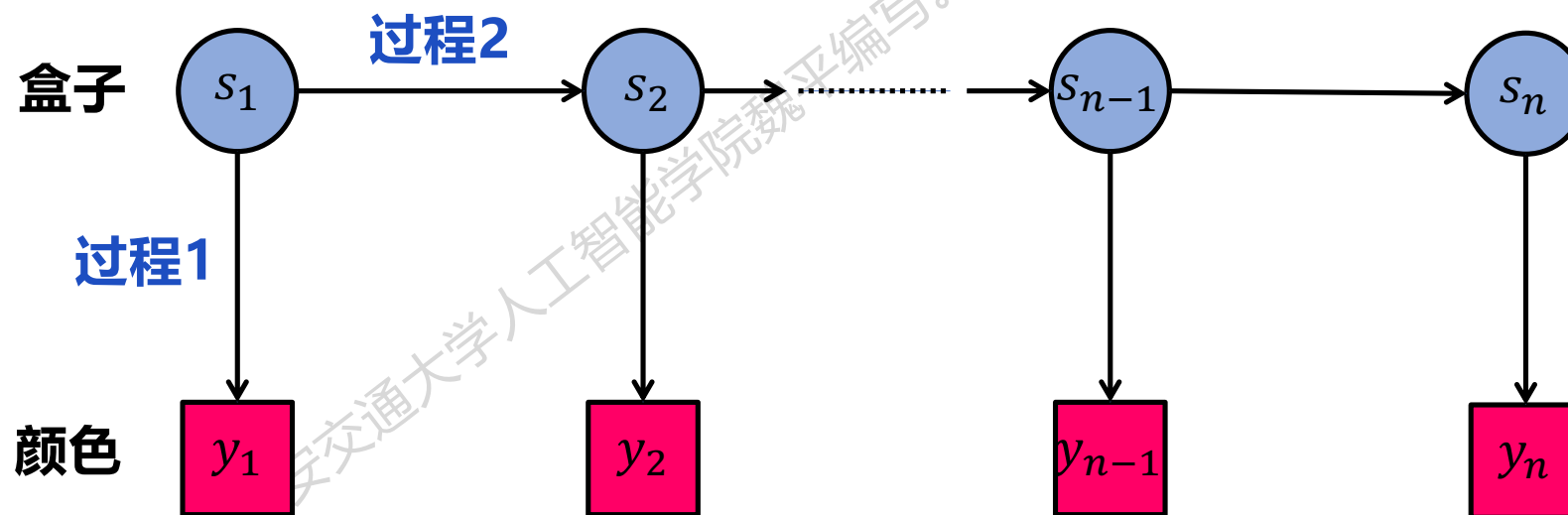
- 马尔可夫链的每一个状态都要对应于一个可观测的物理事件或对象，也就是上述例子中的彩色球颜色
- 因此使用马尔可夫链，我们只能模拟彩色球取样及观测的过程，而完全漏掉盒子选取这一隐藏过程



隐马尔可夫模型

□ 图模型建模

- **过程1: 可观测过程**, 从被选到的盒子中随机抽取彩色球并观察其颜色
- **过程2: 隐藏过程**, 随机选择抽取彩色球的盒子

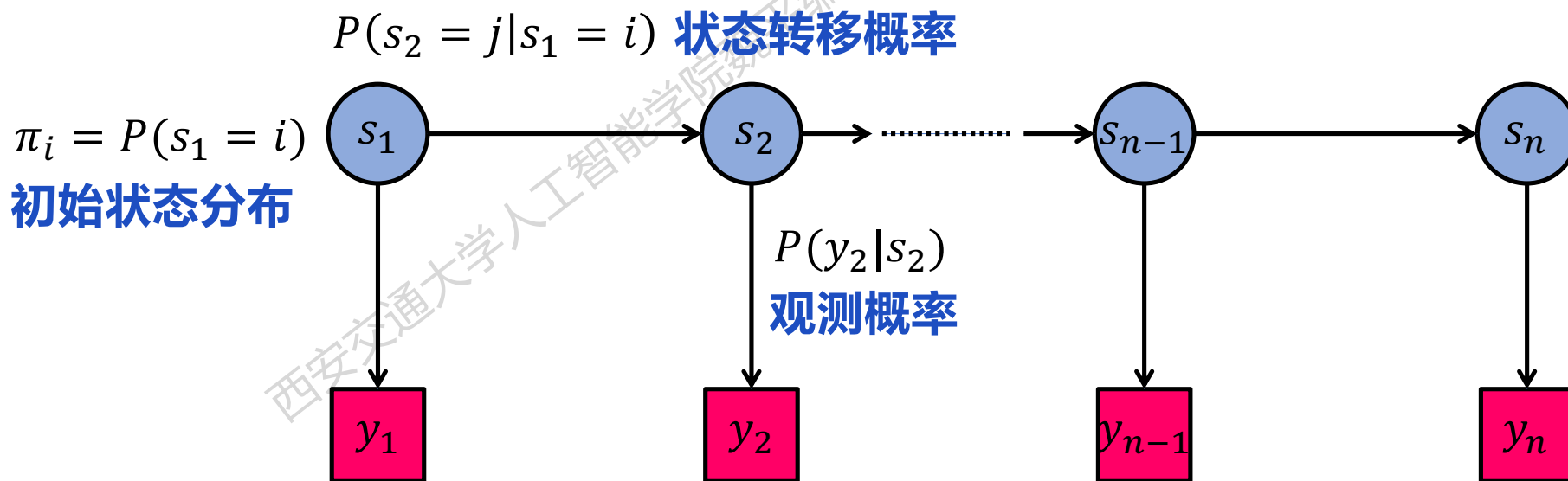


隐马尔可夫模型定义

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

□ HMM定义

隐马尔可夫模型是关于时序的概率模型，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而产生观测随机序列的过程。隐藏的马尔可夫链随机生成的状态的序列成为**状态序列** (state sequence)，每个状态生成一个观测，而由此产生的观测的随机序列称为**观测序列** (observation sequence)，序列的每一个位置可以看作是一个时刻



隐马尔可夫模型定义

□ HMM由以下五个组成要素定义

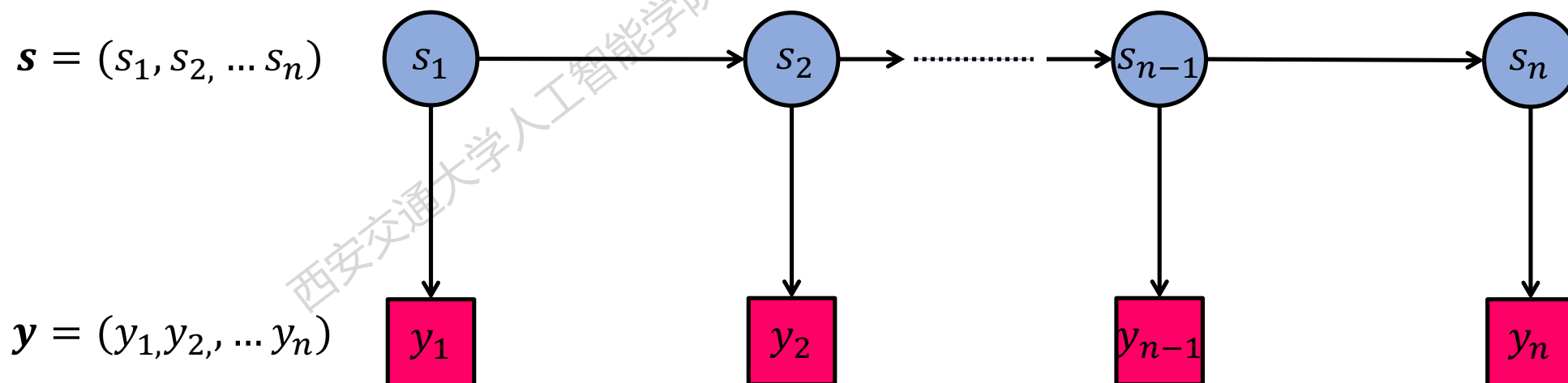
- ① **状态空间 S** : HMM的状态空间通常是可数的, 元素可表示为 i, j, k, \dots 。尽管HMM的状态是隐藏的, 但在实际应用中都有明确的物理意义。比如, 用HMM模拟上述彩色球抽取实验, 每个状态将对应一个包含 M 种颜色球的盒子
- ② **输出集/观测集 Y** : 元素 $y_i \in Y$ 对应系统的一个物理输出。在彩色球抽取实验中, 输出对应抽取到的球的颜色
- ③ **状态转移矩阵**: $P = \{p_{ij}\}$, $p_{ij} = P(j|i)$ ($i, j \in S$) 表示从状态 i 转移到 j 的概率
- ④ **输出概率/观测概率分布矩阵**: $Q = \{q_{sy}\}$, $q_{sy} = P(y|s)$ 是状态 s 下输出 y 的概率, $y \in Y$, $s \in S$
- ⑤ **初始状态分布**: $\pi = \{\pi_i\}$, $i \in S$, π_i 表示初始状态为 i 的概率

隐马尔可夫模型定义

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

□ HMM状态与观测的联合概率密度为

$$\begin{aligned} p(\mathbf{s}, \mathbf{y}) &= p(s_1, s_2, \dots, s_n, y_1, y_2, \dots, y_n) \\ &= p(s_1) \prod_{i=1}^n p(y_i | s_i) \prod_{i=2}^n p(s_i | s_{i-1}) = \pi_{s_1} q_{s_1 y_1} \prod_{i=2}^n p_{s_{i-1} s_i} q_{s_i y_i} \end{aligned}$$



HMM的三个基本问题

- HMM由五个要素决定：状态空间 S ，输出集 Y ，状态转移矩阵 P ，输出概率矩阵 Q ，初始状态分布 π ； S 和 Y 对应 HMM 的模型结构， $\lambda = (P, Q, \pi)$ 对应HMM的参数
- 当前不存在好的方法来自动估计模型结构，利用对问题的了解和直觉判断来设计HMM的结构；模型的参数从数据中自动学习
- HMM三个基本问题
 - **问题1**：给定模型参数 λ 和观测序列 $\mathbf{y} = (y_1, y_2, \dots, y_n)$ ，如何计算 $p(\mathbf{y}|\lambda)$ ？评估问题，计算给定模型生成某观测输出序列的概率，**eg. 行为识别**
 - **问题2**：给定模型参数 λ 和观测序列 $\mathbf{y} = (y_1, y_2, \dots, y_n)$ ，如何计算最可能生成观测序列 \mathbf{y} 的状态序列 $\mathbf{s}^* = (s_1, s_2, \dots, s_n)$ ，即 $\mathbf{s}^* = \operatorname{argmax} p(\mathbf{s}|\mathbf{y}, \lambda)$ ，**eg. 语音识别**
 - **问题3**：给定多个观测序列 \mathbf{y}_i ，如何找出最优模型参数集 λ ，使得对所有的 \mathbf{y}_i ， $p(\mathbf{y}_i|\lambda)$ 最大，**模型参数估计问题**

问题1的解决方法

- 给定观测序列 $\mathbf{y} = (y_1, y_2, \dots, y_n)$ 及参数 λ ，快速、高效计算出概率 $p(\mathbf{y}|\lambda)$
- 最直接的方法是利用加法原理，在所有可能的状态转移序列 s_1, s_2, \dots, s_n 上求和，从联合分布中消除 s_1, s_2, \dots, s_n 得到 y_1, y_2, \dots, y_n 的概率，从而求得 $p(\mathbf{y}|\lambda)$
- 该算法的复杂度为 $O(2n|S|^n)$ ($|S|$ 为状态空间 S 的状态数)，该复杂度几乎是无法计算的。比如当 $|S| = 5$ ， $n=100$ 时，需要复杂度为 $2 \times 100 \times 5^{100}$ 的计算量

$$\begin{aligned}
 p(y_1, y_2, \dots, y_n) &= \sum_{s_1, \dots, s_n} p(s_1, s_2, \dots, s_n, y_1, y_2, \dots, y_n) \\
 &= \sum_{s_1} \sum_{s_2} \dots \sum_{s_n} p(s_1, s_2, \dots, s_n, y_1, y_2, \dots, y_n) \\
 &= \sum_{s_1} \sum_{s_2} \dots \sum_{s_n} p(s_1) \prod_{i=1}^n p(y_i | s_i) \prod_{i=2}^n p(s_i | s_{i-1})
 \end{aligned}$$

问题1的解决方法：前向算法

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

$$p(y_1, y_2, \dots, y_n) = \sum_{i=1}^{|S|} p(y_1, y_2, \dots, y_n, s_n = i)$$

$$p(y_1, y_2, \dots, y_t, s_t) = p(y_1, y_2, \dots, y_t | s_t) p(s_t)$$

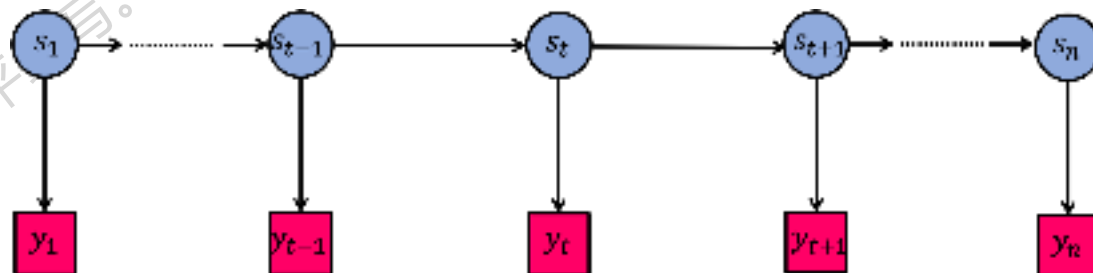
$$= p(y_t | s_t) p(y_1, y_2, \dots, y_{t-1} | s_t) p(s_t) = p(y_t | s_t) p(y_1, y_2, \dots, y_{t-1}, s_t)$$

$$= p(y_t | s_t) \sum_{s_{t-1}} p(y_1, y_2, \dots, y_{t-1}, s_{t-1}, s_t)$$

$$= p(y_t | s_t) \sum_{s_{t-1}} p(y_1, y_2, \dots, y_{t-1}, s_t | s_{t-1}) p(s_{t-1})$$

$$= p(y_t | s_t) \sum_{s_{t-1}} p(y_1, y_2, \dots, y_{t-1} | s_{t-1}) p(s_t | s_{t-1}) p(s_{t-1})$$

$$= p(y_t | s_t) \sum_{s_{t-1}} p(y_1, y_2, \dots, y_{t-1}, s_{t-1}) p(s_t | s_{t-1})$$



$$a_t = p(y_1, y_2, \dots, y_t, s_t)$$

$$a_t = p(y_t | s_t) \sum_{s_{t-1}} a_{t-1} p(s_t | s_{t-1})$$

问题1的解决方法：前向算法

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

- 定义前向变量 $a_t(i) = p(y_1, y_2, \dots, y_t, s_t = i)$ ，其中 $s_t = i$ 表示随机过程在 t 时刻的状态为 i ，前向变量是部分观测序列 y_1, y_2, \dots, y_t 与 t 时刻状态 i 的联合概率
- 前向变量 $a_t(i) = p(y_1, y_2, \dots, y_t, s_t = i)$ 按照下述流程递归计算

(1) 初始化: $a_1(i) = \pi_i q_{iy_1} \quad 1 \leq i \leq |S|$

(2) 递归: $a_t(j) = \left[\sum_{i=1}^{|S|} a_{t-1}(i) p_{ij} \right] q_{jy_t}, \quad 2 \leq t \leq n, 1 \leq j \leq |S|$

- 计算 $p(y_1, y_2, \dots, y_n | \lambda)$

$$p(y_1, y_2, \dots, y_n | \lambda) = \sum_{i=1}^{|S|} a_n(i)$$

$$a_t = p(y_1, y_2, \dots, y_t, s_t)$$

$$a_t = p(y_t | s_t) \sum_{s_{t-1}} a_{t-1} p(s_t | s_{t-1})$$

前向算法复杂度

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

- 直接求和算法的复杂度为 $O(2n|S|^n)$ ，前向算法复杂度为 $O(n|S|^2)$ ，当 $|S| = 5$ ， $n = 100$ 时，直接求和算法复杂度为 $2 \times 100 \times 5^{100}$ ，使用前向算法复杂度为2500

$$\begin{aligned} & p(y_1, y_2, \dots, y_n) \\ &= \sum_{s_1} \sum_{s_2} \dots \sum_{s_n} p(s_1) \prod_{i=1}^n p(y_i | s_i) \prod_{i=2}^n p(s_i | s_{i-1}) \\ &= \sum_{s_1} \sum_{s_2} \dots \sum_{s_n} p(s_1) p(y_1 | s_1) \dots p(y_n | s_n) p(s_2 | s_1) p(s_3 | s_2) \dots p(s_n | s_{n-1}) \\ &= \sum_{s_1} p(s_1) p(y_1 | s_1) \sum_{s_2} p(y_2 | s_2) p(s_2 | s_1) \dots \sum_{s_n} p(y_n | s_n) p(s_n | s_{n-1}) \end{aligned}$$

问题2的解决算法

- 问题2：给定模型数 λ 和观测序列 $\mathbf{y} = (y_1, y_2, \dots, y_n)$ ，如何才能有效地计算出最可能生成观测序列 \mathbf{y} 的状态序列 $\mathbf{s}^* = (s_1, s_2, \dots, s_n)$ ，即 $\mathbf{s}^* = \operatorname{argmax} p(\mathbf{s}|\mathbf{y}, \lambda)$

注意到
$$p(s_1, s_2, \dots, s_n | y_1, y_2, \dots, y_n) = \frac{p(s_1, s_2, \dots, s_n, y_1, y_2, \dots, y_n)}{p(y_1, y_2, \dots, y_n)}$$

因此
$$\operatorname{arg max} p(\mathbf{s}|\mathbf{y}, \lambda) = \operatorname{arg max} p(s_1, s_2, \dots, s_n, y_1, y_2, \dots, y_n)$$

对所有 t
$$p(s_1, \dots, s_t, s_{t+1}, \dots, s_n, y_1, \dots, y_t, y_{t+1}, \dots, y_n)$$

条件独立

$$= p(s_1, \dots, s_t, y_1, \dots, y_t) \underbrace{p(s_{t+1}, \dots, s_n, y_{t+1}, \dots, y_n | s_1, \dots, s_t, y_1, \dots, y_t)}$$

$$= p(s_1, \dots, s_t, y_1, \dots, y_t) \underbrace{p(s_{t+1}, \dots, s_n, y_{t+1}, \dots, y_n | s_t)}$$

问题2的解决算法

因此

$$\begin{aligned} & \max_{s_1, s_2, \dots, s_n} p(s_1, \dots, s_t, s_{t+1}, \dots, s_n, y_1, \dots, y_t, y_{t+1}, \dots, y_n) \\ &= \max_{s_t, s_{t+1}, \dots, s_n} \left[p(s_{t+1}, \dots, s_n, y_{t+1}, \dots, y_n | s_t) \times \max_{s_1, s_2, \dots, s_{t-1}} p(s_1, \dots, s_{t-1}, s_t, y_1, \dots, y_t) \right] \end{aligned}$$

定义

$$\gamma_t(i) = \max_{s_1, s_2, \dots, s_{t-1}} p(s_1, \dots, s_{t-1}, s_t = i, y_1, \dots, y_t)$$

将其带入得到

$$\begin{aligned} & \max_{s_1, s_2, \dots, s_n} p(s_1, \dots, s_t, s_{t+1}, \dots, s_n, y_1, \dots, y_t, y_{t+1}, \dots, y_n) \\ &= \max_i \left\{ \max_{s_{t+1}, \dots, s_n} [p(s_{t+1}, \dots, s_n, y_{t+1}, \dots, y_n | s_t = i) \gamma_t(i)] \right\} \end{aligned}$$

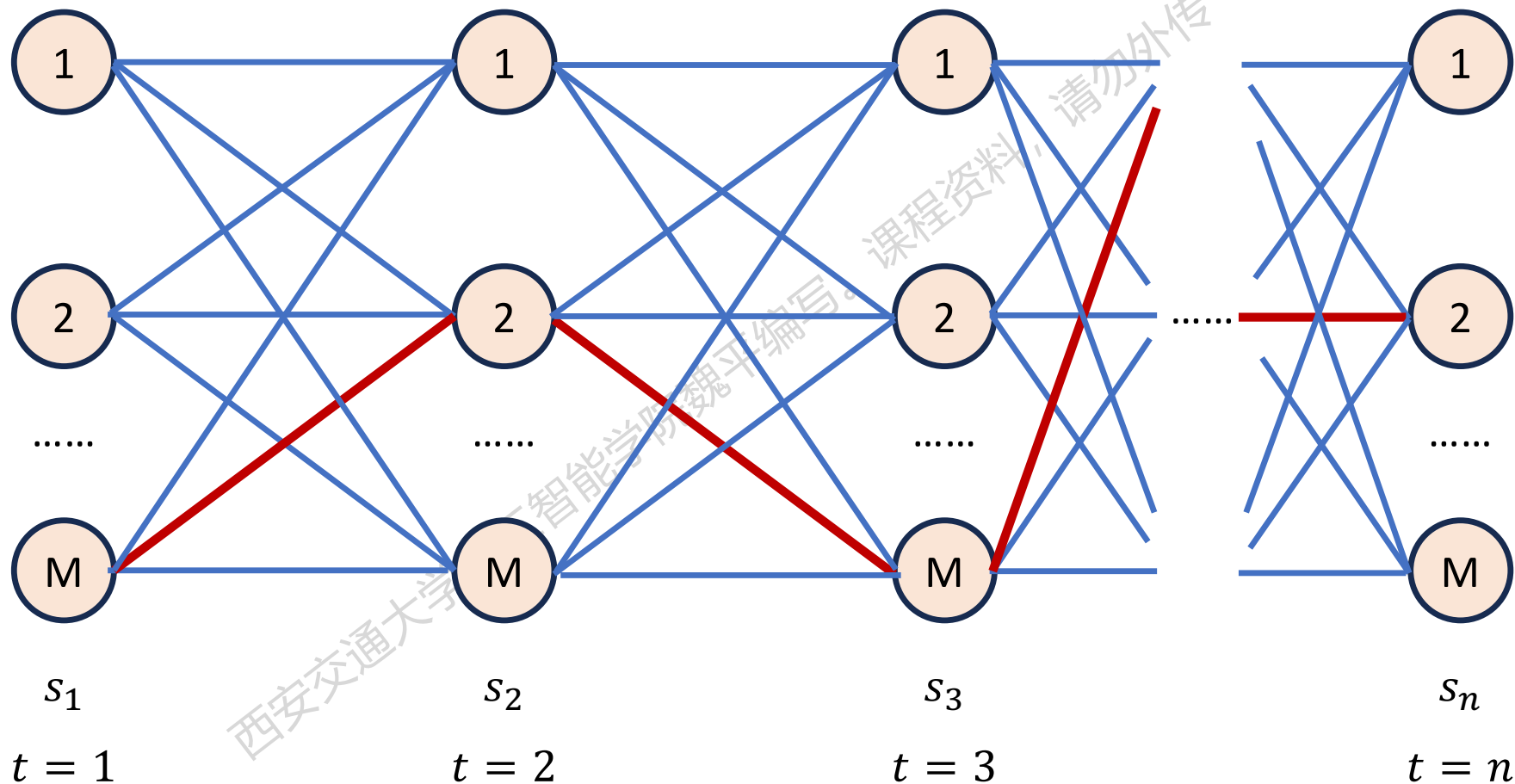
问题2的解决算法

- 这一等式告诉我们，我们可以通过以下三步找到最有可能的状态序列 $s_1, s_2, \dots, s_t, s_{t+1}, \dots, s_n$ ：
- ① 对 t 时刻的每一个状态 $s_t = i, i \in S$ ，找出 s_t 之前的最有可能的子序列 $s_1(i), \dots, s_{t-1}(i)$ ，这对应计算 $\gamma_t(i)$
 - ② 找出紧随 $s_t = i$ 之后的最有可能的子序列 $s_{t+1}(i), \dots, s_n(i)$ ，这对应计算 $\max p(s_{t+1}, \dots, s_n, y_{t+1}, \dots, y_n | s_t = i)$
 - ③ 找出状态 s_t 使得整个状态序列的概率最大

$$\max_i \left\{ \max_{s_{t+1}, \dots, s_n} [p(s_{t+1}, \dots, s_n, y_{t+1}, \dots, y_n | s_t = i) \gamma_t(i)] \right\}$$

维特比算法 (Viterbi algorithm)

西安交通大学人工智能学院魏平编写。课程资料，请勿外传



问题2的解决算法

□ $\gamma_t(i)$ 的计算也可以递归进行:

$$\begin{aligned} p(s_1, \dots, s_t, s_{t+1}, y_1, \dots, y_t, y_{t+1}) &= p(s_1, \dots, s_t, y_1, \dots, y_t) p(s_{t+1}, y_{t+1} | s_1, \dots, s_t, y_1, \dots, y_t) \\ &= p(s_1, \dots, s_t, y_1, \dots, y_t) p(s_{t+1}, y_{t+1} | s_t) = p(s_1, \dots, s_t, y_1, \dots, y_t) p(s_{t+1} | s_t) p(y_{t+1} | s_{t+1}) \end{aligned}$$

$$\begin{aligned} \gamma_{t+1}(j) &= \max_{s_1, s_2, \dots, s_t} p(s_1, \dots, s_t, s_{t+1} = j, y_1, \dots, y_{t+1}) \\ &= \max_{s_1, s_2, \dots, s_t} p(s_1, \dots, s_t, y_1, \dots, y_t) p(s_{t+1} = j | s_t) p(y_{t+1} | s_{t+1} = j) \\ &= \max_i \max_{s_1, s_2, \dots, s_{t-1}} p(s_1, \dots, s_t = i, y_1, \dots, y_t) p(s_{t+1} = j | s_t = i) p(y_{t+1} | s_{t+1} = j) \\ &= \left(\max_i \gamma_t(i) p(s_{t+1} = j | s_t = i) \right) p(y_{t+1} | s_{t+1} = j) \end{aligned}$$

□ 注意到 $\max_{s_1, \dots, s_n} p(s_1, \dots, s_n, y_1, \dots, y_n) = \max_i \gamma_n(i)$

维特比算法 (Viterbi algorithm)

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

□ Viterbi 算法: $\gamma_t(i) = \max_{s_1, s_2, \dots, s_{t-1}} p(s_1, \dots, s_{t-1}, s_t = i, y_1, \dots, y_t)$

① 初始化: $\gamma_1(i) = \pi_i q_{iy_1}, \quad A(1, i) = 0, \quad 1 \leq i \leq |S|$

② 递归: $\gamma_t(j) = \left(\max_i \gamma_{t-1}(i) p_{ij} \right) q_{jy_t}$

③ 终止: $A(t, j) = \arg \max_i [\gamma_{t-1}(i) p_{ij}], \quad 2 \leq t \leq n, 1 \leq j \leq |S|$

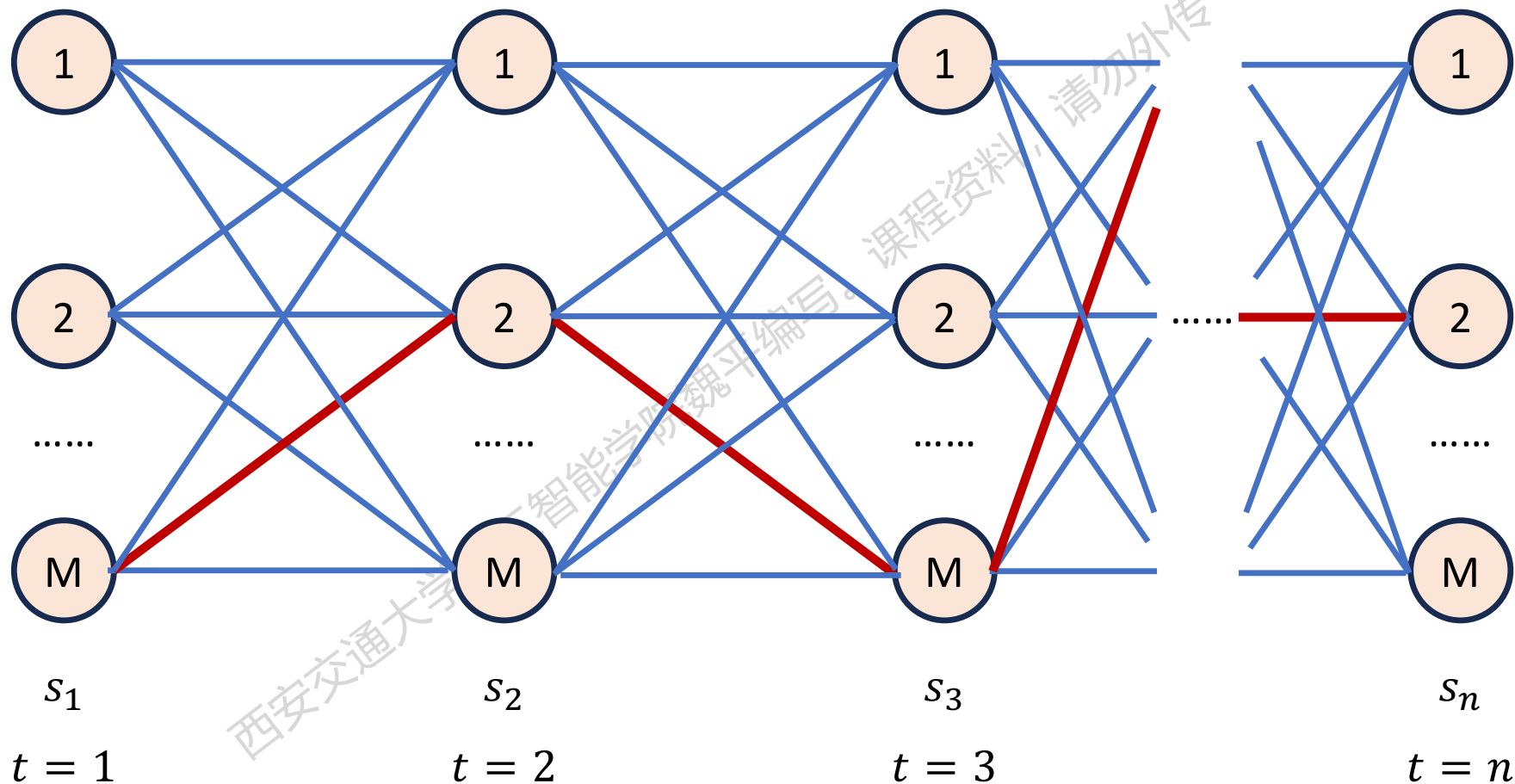
④ 回溯路径 (Path backtracking):

$$\max_{\mathbf{s}} p(\mathbf{s}, \mathbf{y}) = \max_i \gamma_n(i), \quad \mathbf{s}^* = \arg \max_i \gamma_n(i)$$

$$s_t^* = A(t+1, s_{t+1}^*), \quad t = n-1, n-2, \dots, 1$$

维特比算法 (Viterbi algorithm)

西安交通大学人工智能学院魏平编写。课程资料，请勿外传



Andrew J. Viterbi 安德鲁·维特比

- 维特比算法发明人
- 高通公司(Qualcomm)创始人
- CDMA之父
- UCLA和USC教授
- USC工学院冠名教授




A. J. Viterbi, "Error Bounds for Convolutional Coding and an Asymptotically Optimum Decoding Algorithm," IEEE Transactions on Information Theory, Vol. IT-13, No. 2, 1967, pp. 260-269

USC Viterbi
School of Engineering

[About](#) [Academics](#) [Admission](#) [Research & Innovation](#)

[Students](#) [Alumni](#) [Parents](#) [Corporate](#) [Giving](#) [Faculty](#) [Search](#)



 USC University of Southern California

问题3的解决算法

□ **问题3**：给定多个观测序列 y_i ，我们如何才能找出最优模型参数集 λ ，使得对所有的 y_i ， $p(y|\lambda)$ 最大

□ **EM算法 — Expectation Maximization**

标准的EM算法常常开始于初始参数值 θ' ，通过交替重复下列两个步骤直到 $p_{\theta'}(Y)$ 的值停止攀升

① **求期望值 (Expectation)**: 求得随机变量 $\log p_{\theta}(X, Y)$ 关于旧分布 $p_{\theta'}(X|Y)$ 的期望

$$\sum_x p_{\theta'}(x|y) \log p_{\theta}(x, y)$$

② **最大化(Maximization)**: 将该期望值看作是参数集 θ 的函数，最大化这一期望值

CONTENTS



西安交通大学人工智能学院魏平编写。课程资料，请勿外传

IAIR Est. 1986
Institute of
Artificial Intelligence
and Robotics



人工智能学院
College of Artificial Intelligence, XJTU

- 贝叶斯决策与贝叶斯分类
- 有向图与无向图一般概念
- 马尔可夫随机场
- 隐马尔可夫模型
- 条件随机场
- 图模型推理
- 图模型应用例子

概念

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

- 条件随机场是一种判别式无向图模型，对多个变量在给定观测值后的条件概率进行直接建模
- 观测数据 $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ ，标记数据 $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ ，条件随机场构建条件概率模型 $P(\mathbf{y}|\mathbf{x})$



$$\mathbf{x} = \{x_1, x_2, \dots, x_n\}$$



$$\mathbf{y} = \{y_1, y_2, \dots, y_n\}$$

概念

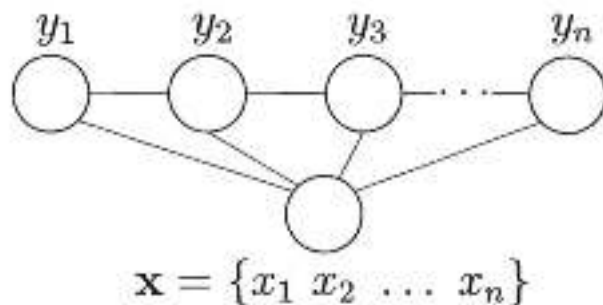
西安交通大学人工智能学院魏平编写。课程资料，请勿外传

- $G = \langle V, E \rangle$ 是一个无向图，节点与 \mathbf{y} 中元素对应， y_v 与节点 v 对应， $n(v)$ 表示邻接节点，若图 G 的每个变量 y_v 都满足马尔可夫性，

$$P(y_v | \mathbf{x}, \mathbf{y}_{V \setminus v}) = P(y_v | \mathbf{x}, \mathbf{y}_{n(v)})$$

则 (\mathbf{x}, \mathbf{y}) 构成一个条件随机场

- G 可以是任意结构，现实问题中常采用链式结构，即链式条件随机场 (chain-structured CRF)



条件概率定义

□ 条件随机场的概率 $P(\mathbf{y}|\mathbf{x})$ 定义为

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z} \exp \left\{ \sum_k \sum_{i=1}^n \mu_k s_k(y_i, \mathbf{x}, i) + \sum_j \sum_{i=1}^{n-1} \lambda_j t_j(y_{i+1}, y_i, \mathbf{x}, i) \right\}$$

$s_k(y_i, \mathbf{x}, i)$: 一元特征函数 μ_k : 一元特征权重

$t_j(y_{i+1}, y_i, \mathbf{x}, i)$: 二元特征函数 λ_j : 二元特征权重

$$Z = \sum_{\mathbf{y}} \exp \left\{ \sum_k \sum_{i=1}^n \mu_k s_k(y_i, \mathbf{x}, i) + \sum_j \sum_{i=1}^{n-1} \lambda_j t_j(y_{i+1}, y_i, \mathbf{x}, i) \right\}$$

CONTENTS



西安交通大学人工智能学院魏平编写。课程资料，请勿外传

IAIR Est. 1986
Institute of
Artificial Intelligence
and Robotics

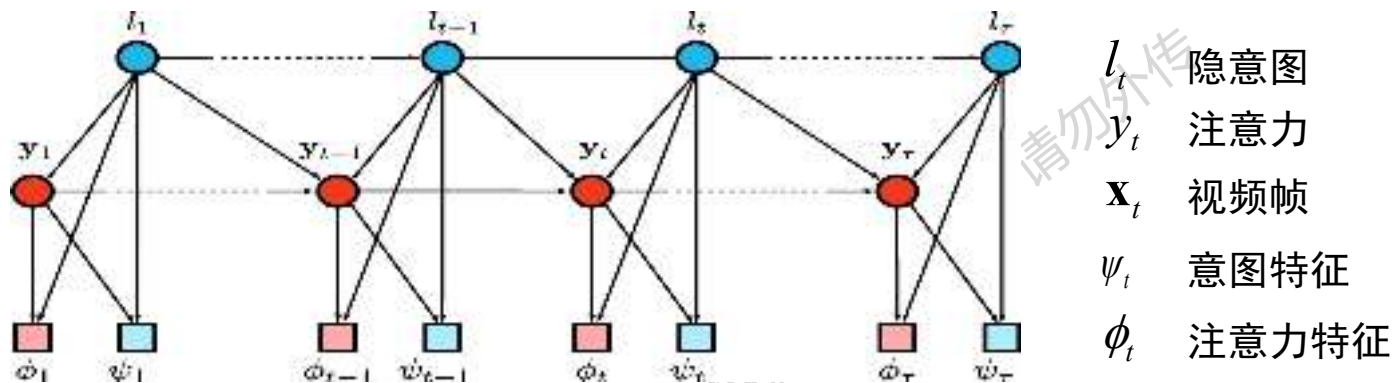


人工智能学院
College of Artificial Intelligence, XJTU

- 贝叶斯决策与贝叶斯分类
- 有向图与无向图一般概念
- 马尔可夫随机场
- 隐马尔可夫模型
- 条件随机场
- **图模型推理**
- 图模型应用例子

概念

西安交通大学人工智能学院魏平编写。课程资料，请勿外传



$$p(\mathbf{X}, \mathbf{l}, \mathbf{Y}) =$$

$$\underbrace{p(l_1)p(y_1|l_1)}_{\text{加阐技荣}} \prod_{t=1}^{\tau} \underbrace{p(\phi(x_t)|l_t, y_t)}_{\text{寮深空}} \prod_{t=2}^{\tau} \underbrace{p(l_t|l_{t-1})}_{\text{寮深}} \prod_{t=1}^{\tau} \underbrace{p(\psi(x_t)|y_t, l_t)}_{\text{寮襟响曠}} \prod_{t=2}^{\tau} \underbrace{p(y_t|y_{t-1}, l_t, l_{t-1})}_{\text{寮襟}}$$

问题：给定视频序列 \mathbf{X} ，求注意力序列 \mathbf{Y} ？

$$\mathbf{Y}^* = \arg \max p(\mathbf{Y}|\mathbf{X})$$

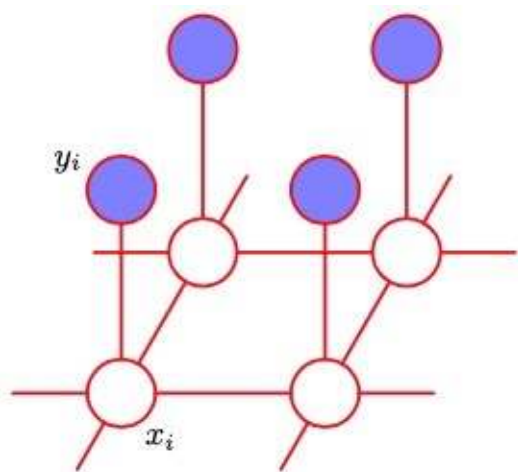
$\mathbf{X} = (x_1, \dots, x_{\tau})$ 输入视频数据

$\mathbf{Y} = (y_1, \dots, y_{\tau})$ 注意力序列

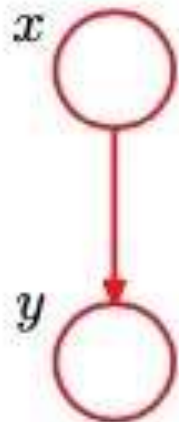
$\mathbf{l} = (l_1, \dots, l_{\tau})$ 隐意图序列

概念

- 在一个图模型中，推理是计算某些节点的边缘分布或以一些节点为条件计算其他节点的后验概率
- 在很多算法中，推理可以看作是局部信息(local messages)在图中的传播(propagation)



$$p(\mathbf{X}|\mathbf{Y})$$



$$p(x, y) = p(y|x)p(x)$$



$$p(x|y)$$

链上推理

西安交通大学人工智能学院魏平编写。课程资料，请勿外传



$$p(\mathbf{x}) = \frac{1}{Z} \psi_{1,2}(x_1, x_2) \psi_{2,3}(x_2, x_3) \cdots \psi_{N-1,N}(x_{N-1}, x_N)$$

N 个离散节点，每个节点有 K 个状态，则每个势函数是一个 $K \times K$ 表格，整体分布具有 $(N - 1)K^2$ 个参数

一个推理的目标是计算 $p(x_n)$ ，对一般 $p(\mathbf{x})$ 采取的方法：

$$p(x_n) = \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} p(\mathbf{x})$$

\mathbf{x} 具有 K^N 可能取值，上式计算的时间和空间复杂度为 N 的指数级

链上推理

西安交通大学人工智能学院魏平编写。课程资料，请勿外传



$$p(\mathbf{x}) = \frac{1}{Z} \psi_{1,2}(x_1, x_2) \psi_{2,3}(x_2, x_3) \cdots \psi_{N-1,N}(x_{N-1}, x_N)$$

$$p(x_n) = \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} p(\mathbf{x})$$

根据链的特点，可以采取加法乘法分配的方式，如

$$\sum_{x_N} \psi_{N-1,N}(x_{N-1}, x_N)$$

链上推理

西安交通大学人工智能学院魏平编写。课程资料，请勿外传



$$p(\mathbf{x}) = \frac{1}{Z} \psi_{1,2}(x_1, x_2) \psi_{2,3}(x_2, x_3) \cdots \psi_{N-1,N}(x_{N-1}, x_N)$$

对整个分布：

$$p(x_n) = \frac{1}{Z} \underbrace{\left[\sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \cdots \left[\sum_{x_2} \psi_{2,3}(x_2, x_3) \left[\sum_{x_1} \psi_{1,2}(x_1, x_2) \right] \right] \right]}_{\mu_\alpha(x_n)} \underbrace{\left[\sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \cdots \left[\sum_{x_N} \psi_{N-1,N}(x_{N-1}, x_N) \right] \right]}_{\mu_\beta(x_n)} \quad (8.52)$$

链上推理

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

可以通过递归方式计算传递的信息：

$$\begin{aligned}\mu_{\alpha}(x_n) &= \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \left[\sum_{x_{n-2}} \cdots \right] \\ &= \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \mu_{\alpha}(x_{n-1}).\end{aligned}$$

$$\begin{aligned}\mu_{\beta}(x_n) &= \sum_{x_{n+1}} \psi_{n+1,n}(x_{n+1}, x_n) \left[\sum_{x_{n+2}} \cdots \right] \\ &= \sum_{x_{n+1}} \psi_{n+1,n}(x_{n+1}, x_n) \mu_{\beta}(x_{n+1}).\end{aligned}$$

链上推理

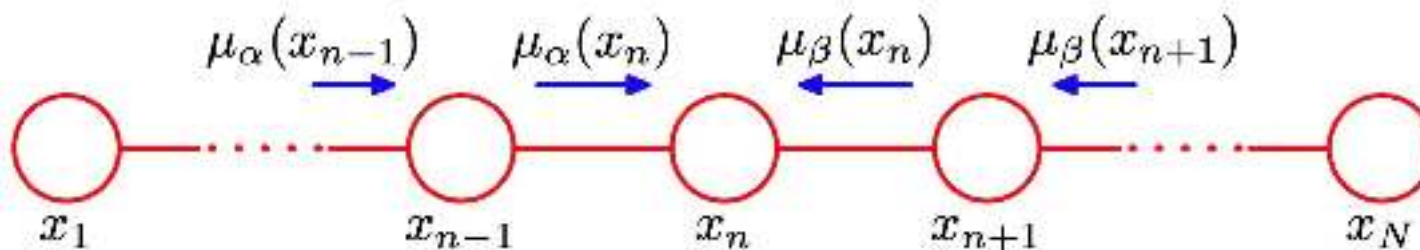
西安交通大学人工智能学院魏平编写。课程资料，请勿外传

$$p(x_n) = \frac{1}{Z} \underbrace{\left[\sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \cdots \left[\sum_{x_2} \psi_{2,3}(x_2, x_3) \left[\sum_{x_1} \psi_{1,2}(x_1, x_2) \right] \right] \right]}_{\mu_\alpha(x_n)} \underbrace{\left[\sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \cdots \left[\sum_{x_N} \psi_{N-1,N}(x_{N-1}, x_N) \right] \right]}_{\mu_\beta(x_n)} \quad (8.52)$$

从节点 x_{n-1} 传递给 x_n 的信息

$$p(x_n) = \frac{1}{Z} \mu_\alpha(x_n) \mu_\beta(x_n)$$

从节点 x_{n+1} 传递给 x_n 的信息



链上推理

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

□ 时间复杂度分析

方式1：直接积分

$$p(x_n) = \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} p(\mathbf{x})$$

\mathbf{x} 具有 K^N 个可能取值，总时间复杂度为 $O(K^N)$

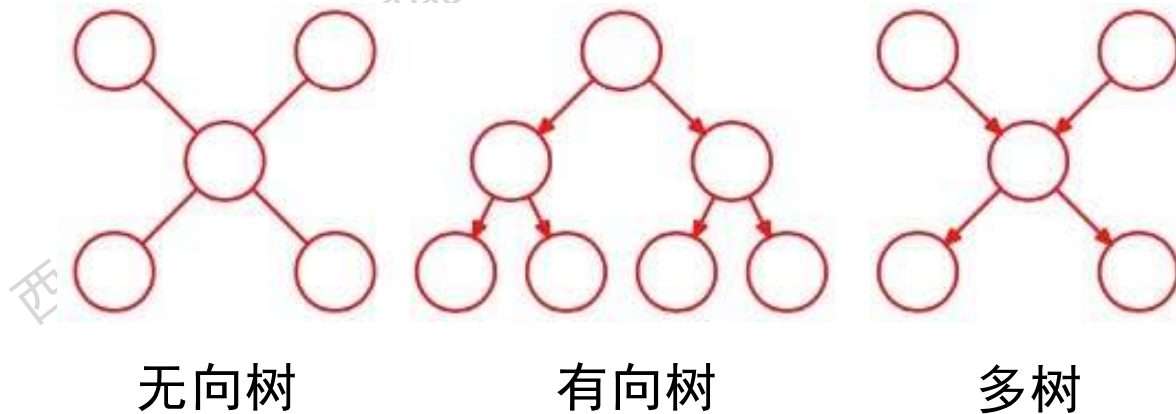
方式2：信息传播

$$p(x_n) = \frac{1}{Z} \mu_\alpha(x_n) \mu_\beta(x_n)$$

每一项有 K^2 个状态，总时间复杂度为 $O(NK^2)$

树(Trees)

- 无向图中，一个树是在任意两个节点之间有且只有一条路径的图；在有向图中，一个树包含唯一根节点，其没有父节点，其他节点都只有一个父节点
- 多树(polytree)：一个有向图中包含具有多个父节点的节点，但在任意两节点之间仍然只有一条路径（忽略箭头方向）
- 树和多树提供了一种可高效推理的结构



因子图(Factor Graphs)

- 无向图和有向图的联合分布都表达为多个因子的乘积，因子图是对这一形式的统一表达；在因子图上可以更好地对有向图、无向图进行统一的推理计算

$$p(\mathbf{x}) = \prod_s f_s(\mathbf{x}_s)$$

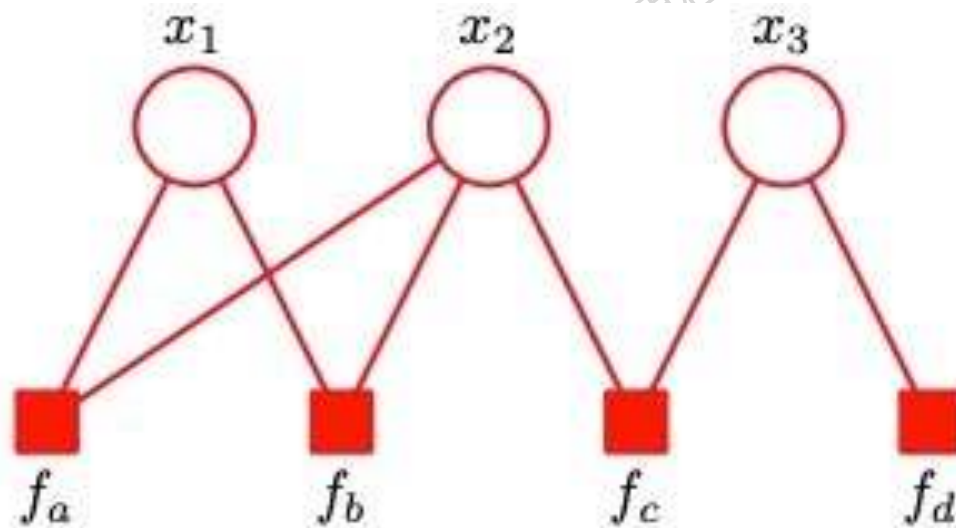
\mathbf{x}_s 是变量集合，因子 f_s 表示定义在某一个变量集合上的函数

- 因子图中包含两个类型的节点：变量节点和因子节点，变量节点和因子节点通过无向边相连
- 因子图中，同一类型的节点不直接相连

因子图(Factor Graphs)

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

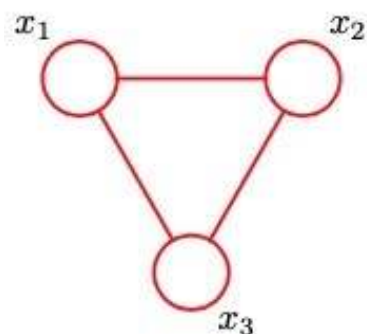
- 每一个变量节点对应一个随机变量，每一个因子节点对应一个函数，每个因子节点和其对应的所有变量节点之间有无向边相连，是与其相连的变量节点的函数



$$p(\mathbf{x}) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3)$$

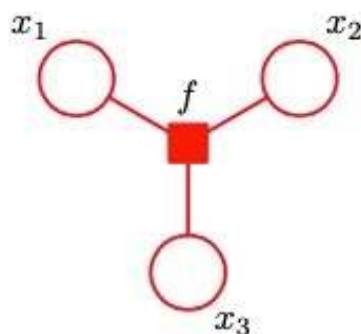
因子图(Factor Graphs)

- 无向图转化为因子图：节点对应，每个最大团上的势函数对应一个因子，常系数假设定义在一个空节点上的因子，多个因子图可对应于同一个无向图



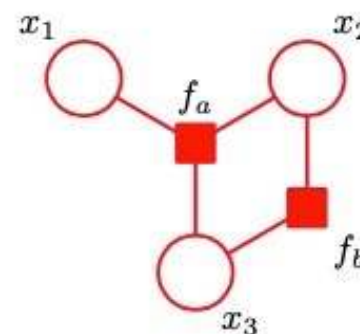
无向图

$$\psi(x_1, x_2, x_3)$$



因子图

$$f(x_1, x_2, x_3) = \psi(x_1, x_2, x_3)$$

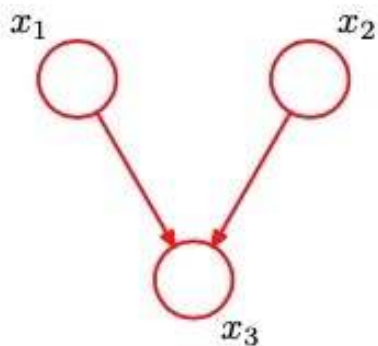


因子图

$$f_a(x_1, x_2, x_3) f_b(x_1, x_2) = \psi(x_1, x_2, x_3)$$

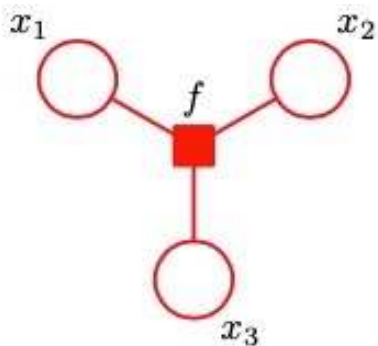
因子图(Factor Graphs)

- 有向图转化为因子图：节点对应，一个条件分布对应于一个因子，多个因子图可对应于同一个有向图



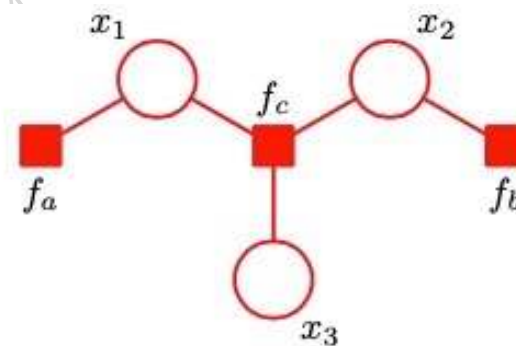
有向图

$$p(x_1)p(x_2)p(x_3|x_1, x_2)$$



因子图

$$f(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3|x_1, x_2)$$



因子图

$$f_a(x_1) = p(x_1)$$

$$f_b(x_2) = p(x_2)$$

$$f_c(x_1, x_2, x_3) = p(x_3|x_1, x_2)$$

Sum-Product算法

- Sum-product算法是在树结构图上进行精确推理(exact inference)的算法，用于求某些节点或节点集合的边缘分布；有向无环图上常用的置信传播(belief propagation)是Sum-Product算法的特例
- 问题：求一个变量的边缘分布

$$p(x) = \sum_{\mathbf{x} \setminus x} p(\mathbf{x})$$

- Sum-product算法的基本思想是将 $p(\mathbf{x})$ 用因子图表达式代替，交换求和和求积的顺序，实现高效算法

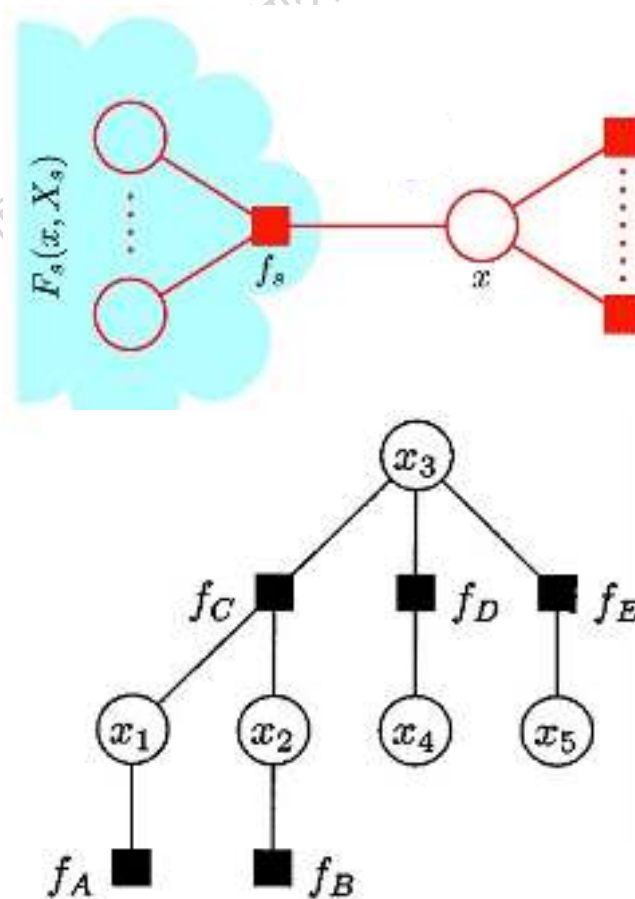
$$p(\mathbf{x}) = \prod_s f_s(\mathbf{x}_s)$$

Sum-Product算法

- 将 x 看作根节点，将 $p(x)$ 表达为因子图形式

$$p(\mathbf{x}) = \prod_{s \in \text{ne}(x)} F_s(x, X_s)$$

$\text{ne}(x)$ 是 x 的相邻因子节点的集合， X_s 是通过因子节点 f_s 与 x 相连的子树的所有变量节点的集合，由树结构可知 $X_i \cap X_j = \emptyset$ ， $F_s(x, X_s)$ 表示与 f_s 相关的分支中所有因子节点的乘积



Sum-Product算法

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

$$p(x) = \sum_{X_1} \dots \sum_{X_S} \prod_{s \in \text{ne}(x)} F_s(x, X_s) = \prod_{s \in \text{ne}(x)} \sum_{X_s} F_s(x, X_s)$$

$$= \prod_{s \in \text{ne}(x)} \mu_{f_s \rightarrow x}(x)$$

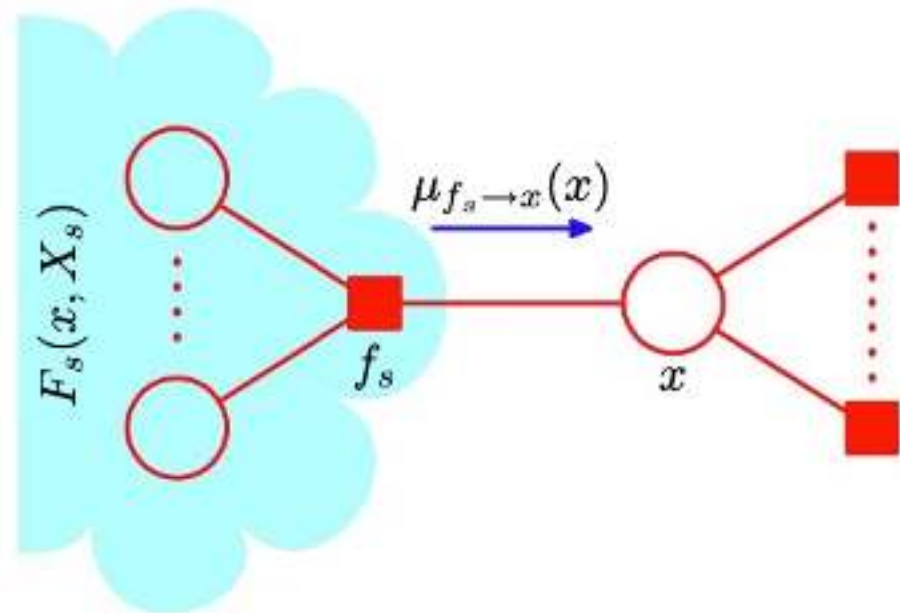
$$\text{记} \quad \mu_{f_s \rightarrow x}(x) = \sum_{X_s} F_s(x, X_s)$$

$\mu_{f_s \rightarrow x}(x)$ 可以看作是从因子节点 f_s 到变量节点 x 的信息 (message)，表示与变量节点 x 相连的第 s 个因子节点传给 x 的结果

Sum-Product算法

□ $p(x)$ 化为

$$p(x) = \prod_{s \in \text{ne}(x)} \mu_{f_s \rightarrow x}(x)$$

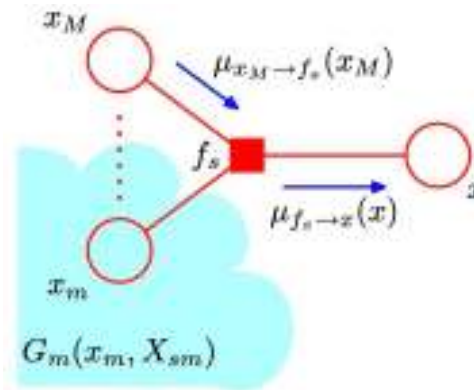


Sum-Product算法

- ❑ 假设因子节点 f_s 与节点集合 $\{x, x_1, \dots, x_M\}$ 相连接, f_s 接收 x_1, \dots, x_M 发来的信息, 经过运算得到 $\mu_{f_s \rightarrow x}(x)$
- ❑ 怎么样计算 $\mu_{f_s \rightarrow x}(x)$?
- ❑ 变量节点 x_i 与若干因子节点相连

$$F_s(x, X_s) = f_s(x, x_1, \dots, x_M)G_1(x_1, X_{s1}) \dots G_M(x_M, X_{sM})$$

$G_m(x_m, X_{sm})$ 是 f_s 分支中通过 x_m 连接到 f_s 的子树的概率



Sum-Product算法

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

$$\begin{aligned}\mu_{f_s \rightarrow x}(x) &= \sum_{X_s} F_s(x, X_s) \\ &= \sum_{X_s} f_s(x, x_1, \dots, x_M) G_1(x_1, X_{s1}) \dots G_M(x_M, X_{sM}) \\ &= \sum_{x_1} \dots \sum_{x_M} f_s(x, x_1, \dots, x_M) \prod_{m \in \text{ne}(f_s) \setminus x} \left[\sum_{X_{sm}} G_m(x_m, X_{sm}) \right] \\ &= \sum_{x_1} \dots \sum_{x_M} f_s(x, x_1, \dots, x_M) \prod_{m \in \text{ne}(f_s) \setminus x} \mu_{x_m \rightarrow f_s}(x_m)\end{aligned}$$

Sum-Product算法

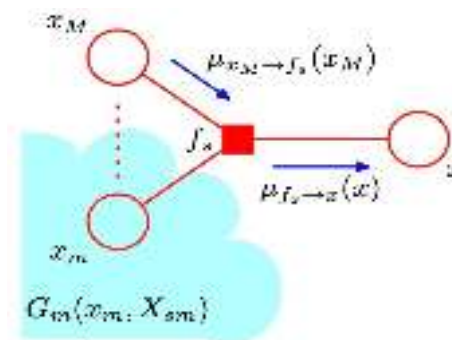
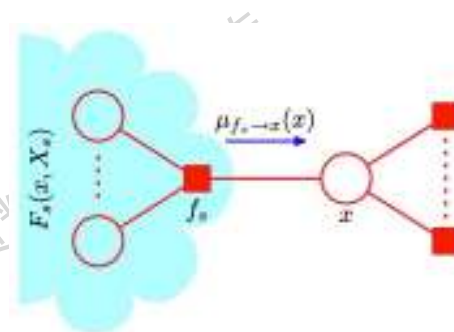
西安交通大学人工智能学院魏平编写。课程资料，请勿外传

$$p(\mathbf{x}) = \prod_{s \in \text{ne}(x)} F_s(x, X_s)$$

$$p(x) = \sum_{X_1} \dots \sum_{X_S} p(\mathbf{x}) = \prod_{s \in \text{ne}(x)} \mu_{f_s \rightarrow x}(x)$$

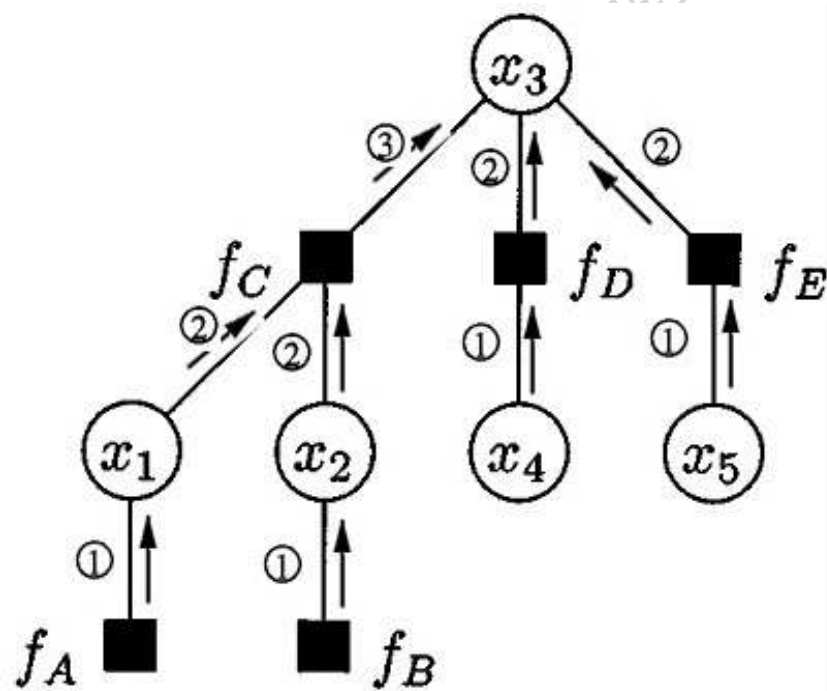
$$\mu_{f_s \rightarrow x}(x) = \sum_{x_1} \dots \sum_{x_M} f_s(x, x_1, \dots, x_M) \prod_{m \in \text{ne}(f_s) \setminus x} \mu_{x_m \rightarrow f_s}(x_m)$$

$$\mu_{x_m \rightarrow f_s}(x_m) = \sum_{X_{sm}} G_m(x_m, X_{sm})$$



Sum-Product算法

- 将要求的变量看作根节点，消息从每个叶子节点开始朝根节点传递，中间节点等到所有子节点的消息全部到达时才向父节点传递，直到传递到根节点停止



精确推理与近似推理

- 在没有环的图结构中，利用sum-product, max-sum, junction tree等算法可以实现精确推理 (exact inference)
- 在有环图中，一般很难实现精确推理，可以利用变分方法 (variational methods), 有环置信传播 (loopy belief propagation), 采样方法 (sampling) 或蒙特卡洛方法 (Monte Carlo) 等实现近似推理 (approximate inference)
- 图模型学习与推理是机器学习与人工智能领域一个长盛不衰的研究问题，要主动将图模型的方法与其他方法结合起来解决问题

CONTENTS



西安交通大学人工智能学院魏平编写。课程资料，请勿外传

IAIR Est. 1986
Institute of
Artificial Intelligence
and Robotics



人工智能学院
College of Artificial Intelligence, XJTU

- **贝叶斯决策与贝叶斯分类**
- **有向图与无向图一般概念**
- **马尔可夫随机场**
- **隐马尔可夫模型**
- **条件随机场**
- **图模型推理**
- **图模型应用例子**

事件和物体联合识别的4D时空交互

Ping Wei, Yibiao Zhao, Nanning Zheng, Song-Chun Zhu. Modeling 4D Human-Object Interactions for Joint Event Segmentation, Recognition, and Object Localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 39(6): 1165-1179, 2017.

问题描述

- 给定一个包含多个事件、多个物体的长RGBD视频



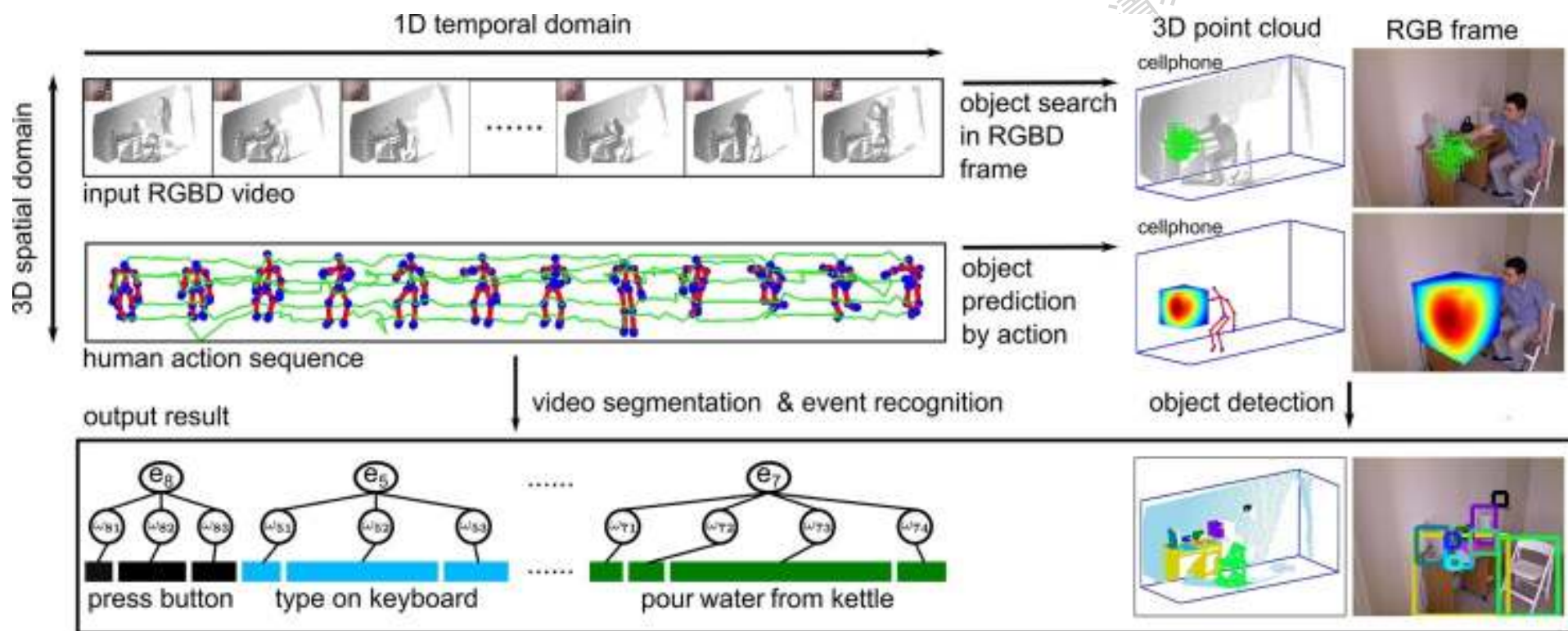
目标是：

- ① 分割视频序列
- ② 识别视频中每一帧三维点云中的物体
- ③ 识别视频中的事件

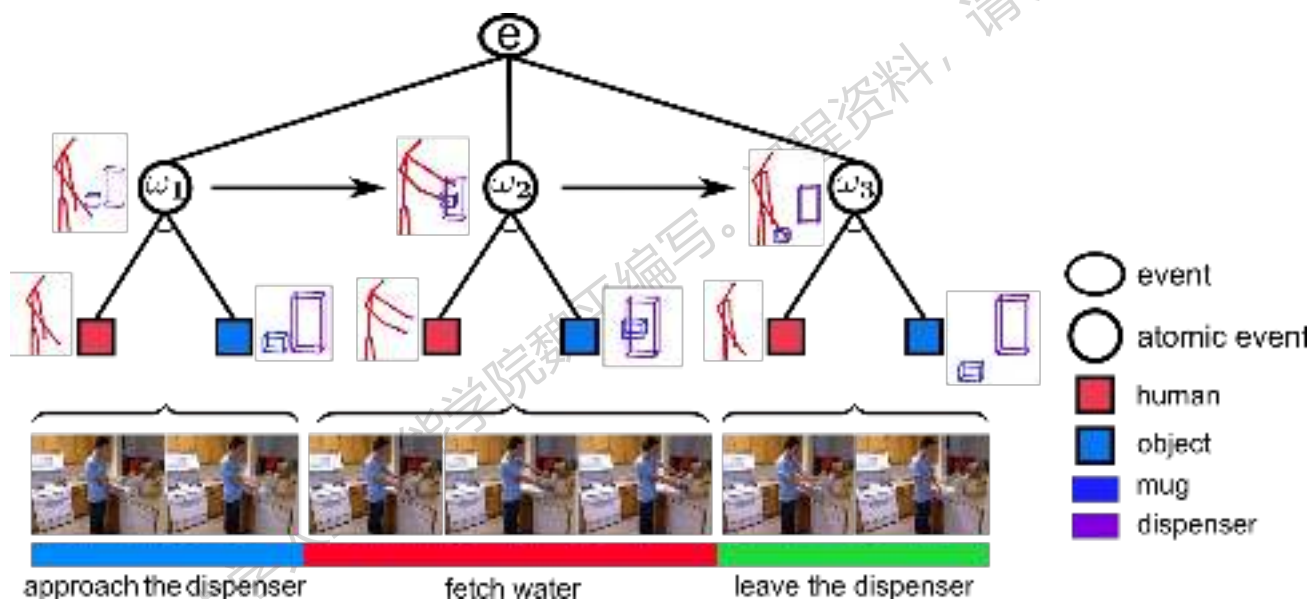
数学模型

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

4D时空交互模型



事件的分层随机图模型



四维交互时空的含义:

- ① 一维时域, 原子事件之间存在交互
- ② 三维空域, 人与物体在语义和几何相容性上存在交互

数学模型

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

□ 事件的分层随机图模型

视频序列: $V = (I_1, \dots, I_\tau)$

分层随机图: $G = \langle E, L \rangle$

(1) 根节点: 事件类别 E

(2) 视频帧解释结点序列: $L = (l_1, \dots, l_\tau)$ $l_t = (h_t, o_t, a_t)$

则分层随机图 G 解释视频 V 的能量为:

$$\text{En}(G | V) = \underbrace{\sum_{t=1}^{\tau} \Phi(I_t, l_t)}_{\text{鄙}\square\square\text{痰}} + \underbrace{\sum_{t=2}^{\tau} \Psi(l_{1:t-1}, l_t)}_{\text{崩}\square\text{蜉}\text{痰}}$$

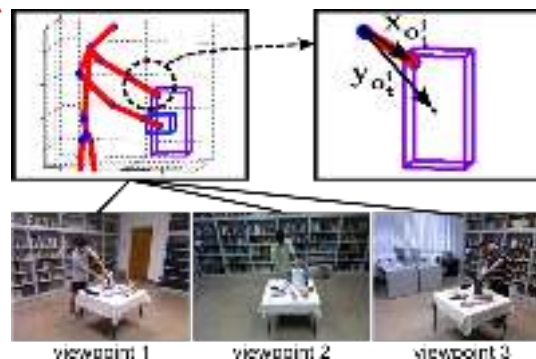
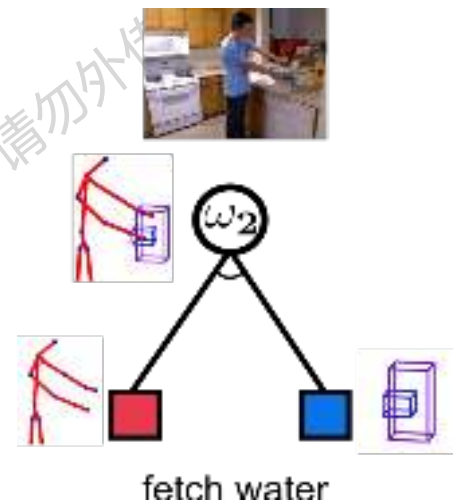
数学模型

3D空间中人与物的交互

$$\Phi(I_t, l_t) = \underbrace{\phi_1(a_t, h_t)}_{\text{尘殆摇拽}} + \underbrace{\phi_2(a_t, o_t, I_t)}_{\text{}} + \underbrace{\phi_3(a_t, h_t, o_t)}_{\text{尘-敞掺}}$$

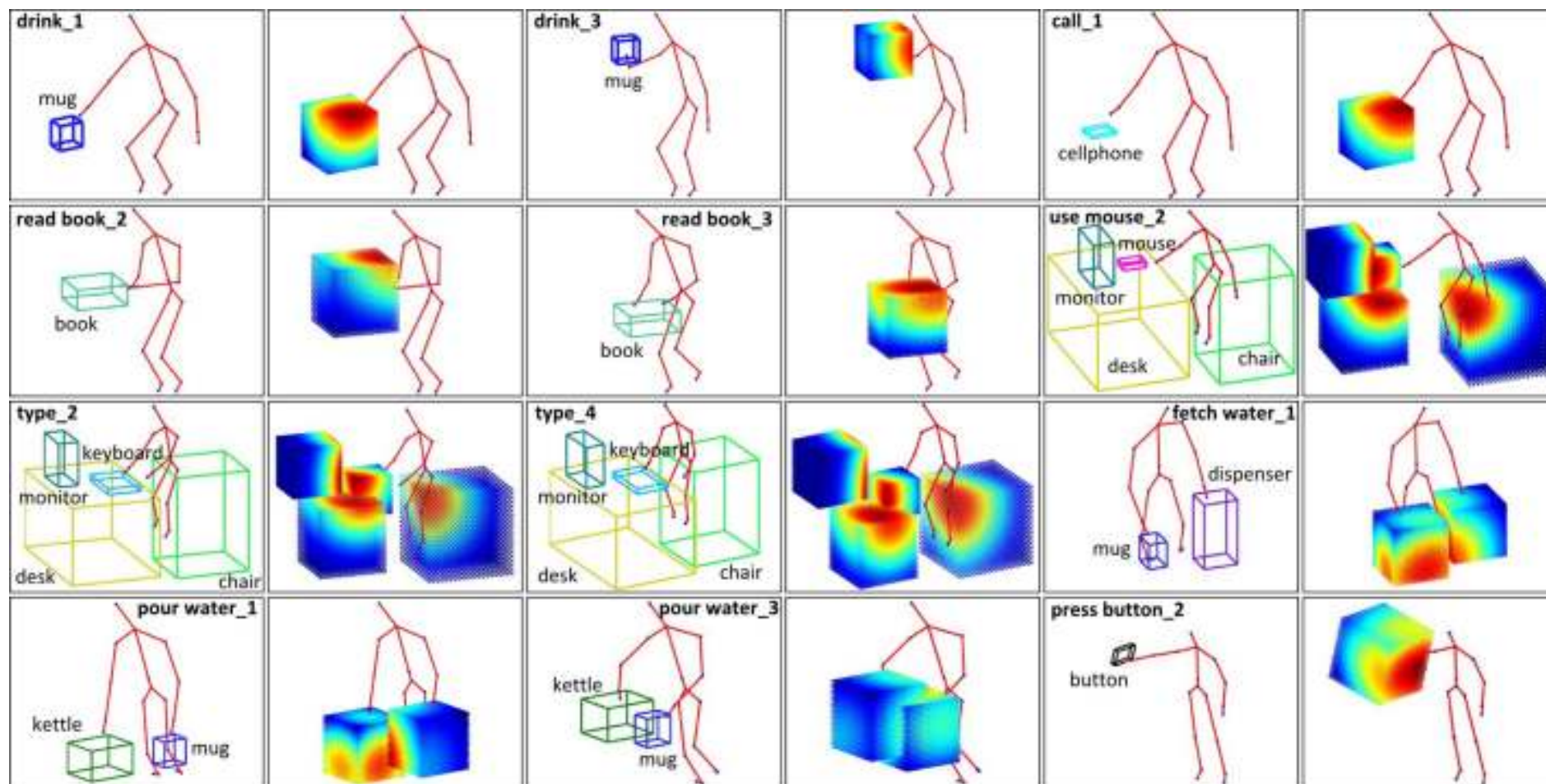
人-物交互模型

$$\phi_3(h_t, o_t | a_t) = -\frac{1}{n_t} \sum_{i=1}^{n_t} \ln N(y_{o_t^i}; W_{o_t^i}^{a_t} x_{o_t^i}, \Sigma_{o_t^i}^{a_t})$$



数学模型

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

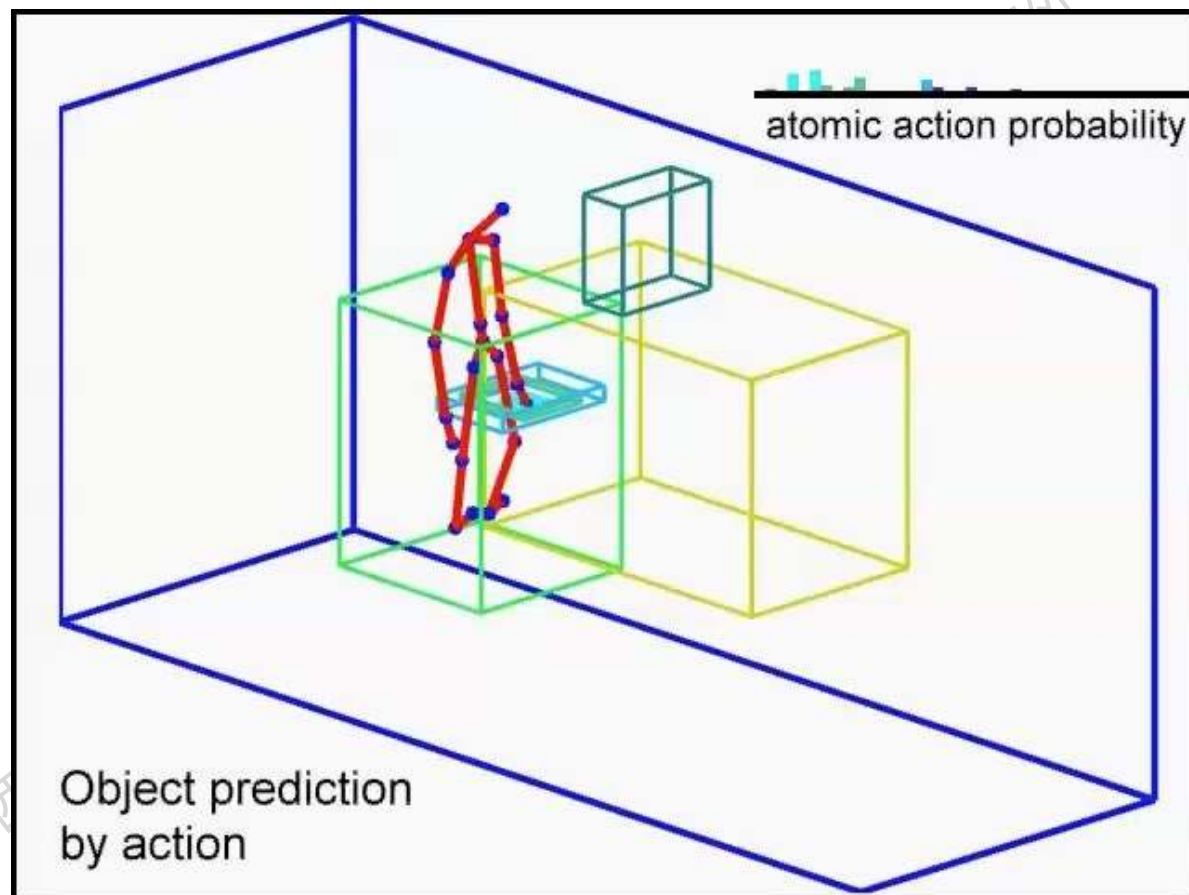


数学模型

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

□ 人-物交互模型：行为对物体的预测

‘pantomime’



模型推理

- 给定一个RGBD视频序列，推理的目的是用一个分层图序列解释视频，即分割视频、识别事件和物体

$$\mathbf{G}^* = \arg \min \varepsilon(\mathbf{G} | \mathbf{V}) \quad \varepsilon(\mathbf{G} | \mathbf{V}) = \sum_{q=1}^Q \text{En}(G_q | V_q)$$

- 动态规划束搜索推理 (Dynamic Programming Beam Search)

基本思想

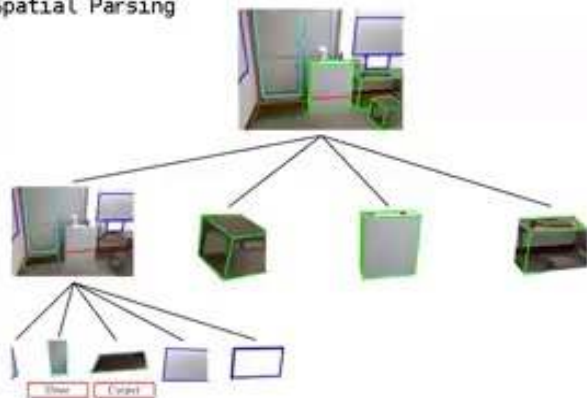
- ① 动态规划：基于对前面视频帧的解释，解释当前帧，循环向前
- ② 束搜索：保留概率较大的解释，作为解释下一帧的依据

模型推理

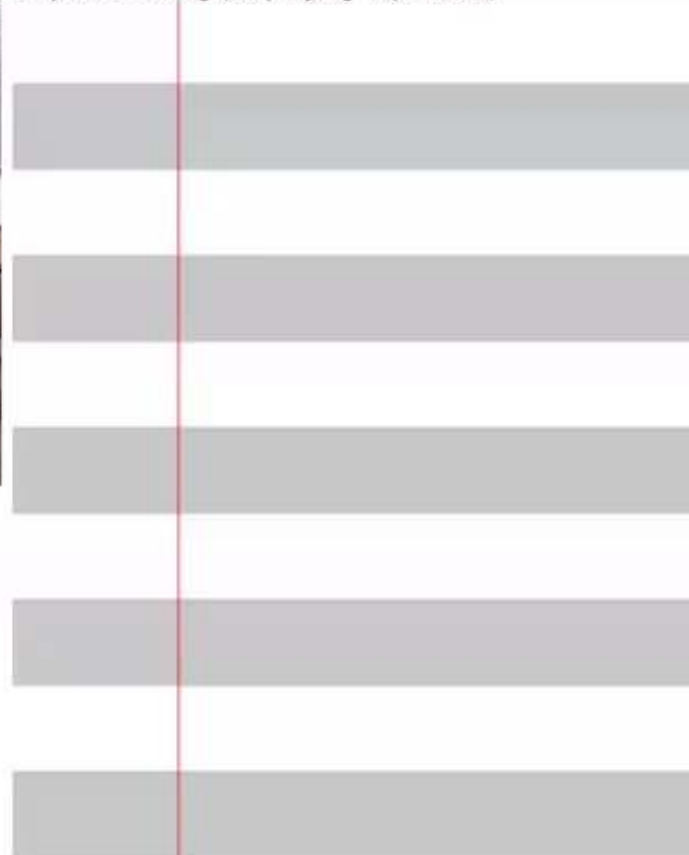
西安交通大学人工智能学院魏平编写。课程资料，请勿外传



Spatial Parsing



Temporal Parsing (Displaying Top 10/500)



Causal Parsing

Door Open

视频中人的注意力推断

Ping Wei, Dan Xie, Nanning Zheng, Song-Chun Zhu. Inferring Human Attention by Learning Latent Intentions. *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1297-1303, 2017.

问题背景

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

□ 人的三维注意力推断被广泛应用于许多人工智能领域



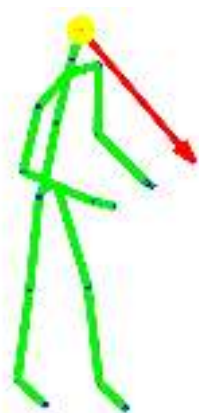
人与机器人协作
Human-Robot Collaboration



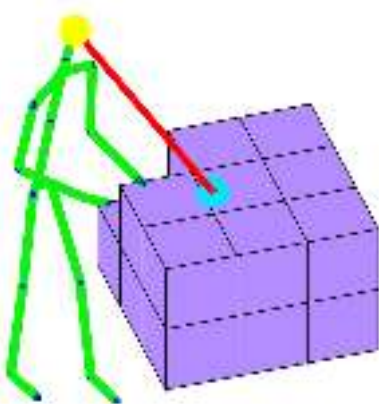
人的意识推理
Reasoning about Human Mind

问题描述

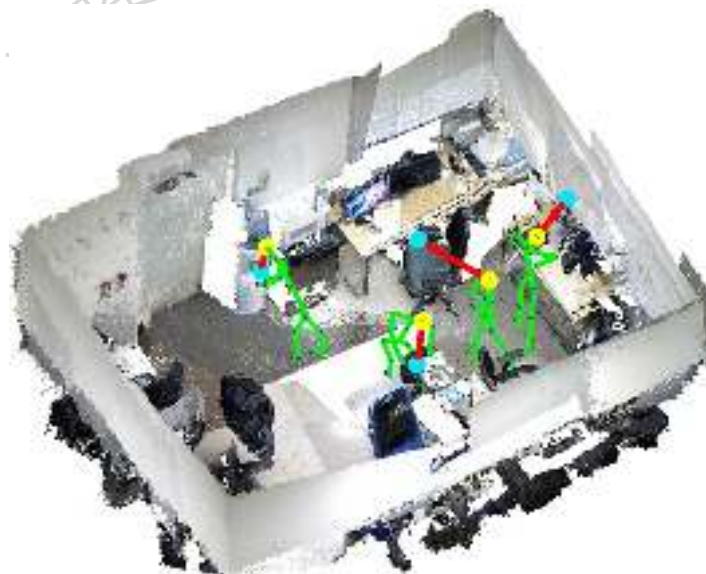
- 给定一个日常生活中的RGBD人体活动视频，我们要推断每一帧中这个人往哪里看
 - 三维注意力方向
 - 三维注意力目标



三维注意力方向



三维注意力目标



三维场景中的日常活动

基本思想

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

- 人的注意力是由人的意图驱动的和控制的
- 人的注意力推理是与人的意图联合优化的过程



locate the dispenser



check the cup's status



check the book's status

- 当人的意图是locate the dispenser, 他的注意力快速从桌子上扫描直到饮水机.
- 当人的意图是检查某种物体的状态时, 他的注意力将集中在目标物体上, 如check the cup' s status, check the book' s status.

数据特征

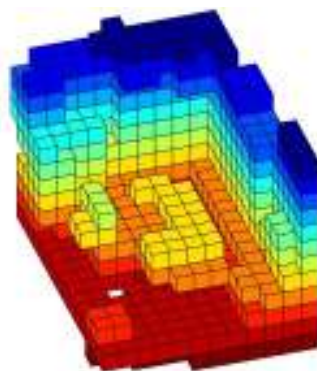
西安交通大学人工智能学院魏平编写。课程资料，请勿外传



场景点云



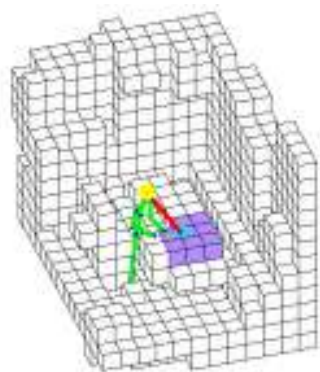
注意力方向



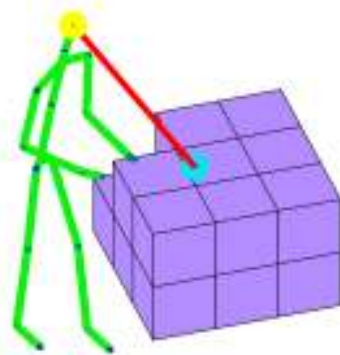
体素高度特征统计



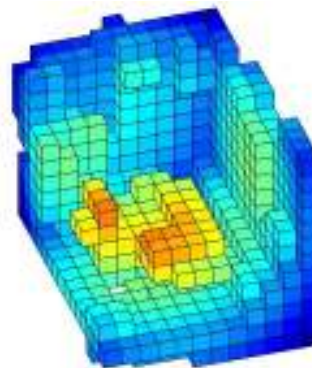
体素高度特征



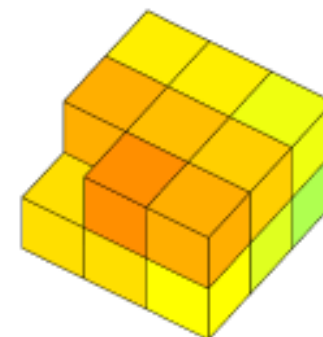
场景体素



注意力体素



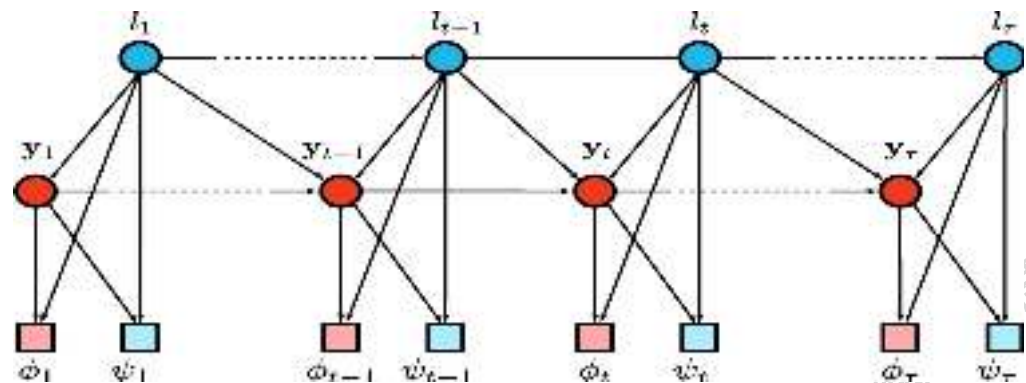
体素距离统计



体素距离特征

图模型

西安交通大学人工智能学院魏平编写。课程资料，请勿外传



- l_t 隐意图
- y_t 注意力
- x_t 视频帧
- ψ_t 意图特征
- ϕ_t 注意力特征

$$p(\mathbf{X}, \mathbf{l}, \mathbf{Y}) =$$

$$\underbrace{p(l_1)p(y_1 | l_1)}_{\text{加阉技荣}} \prod_{t=1}^{\tau} \underbrace{p(\phi(x_t) | l_t, y_t)}_{\text{寮深空}} \prod_{t=2}^{\tau} \underbrace{p(l_t | l_{t-1})}_{\text{寮深}} \prod_{t=1}^{\tau} \underbrace{p(\psi(x_t) | y_t, l_t)}_{\text{寮襟晌曠}} \prod_{t=2}^{\tau} \underbrace{p(y_t | y_{t-1}, l_t, l_{t-1})}_{\text{寮襟}}$$

$\mathbf{X} = (x_1, \dots, x_\tau)$ 输入视频数据

$\mathbf{Y} = (y_1, \dots, y_\tau)$ 注意力序列

$\mathbf{l} = (l_1, \dots, l_\tau)$ 隐意图序列

实验结果

西安交通大学人工智能学院魏平编写。课程资料，请勿外传



Input RGBD Video



Reference RGB Video

复杂视频中人的意图、注意力和任务联合推理

Ping Wei, Yang Liu, Tianmin Shu, Nanning Zheng, Song-Chun Zhu. Where and Why Are They Looking? Jointly Inferring Human Attention and Intentions in Complex Tasks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

一般性结论

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

□ 在理智的人的行为中，

- 人的注意力和行为都是由人的意图驱动和控制的
- 人的注意力和行为又反映了人的意图和目的



问题描述

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

□ 给定一个包含复杂任务的长RGBD视频

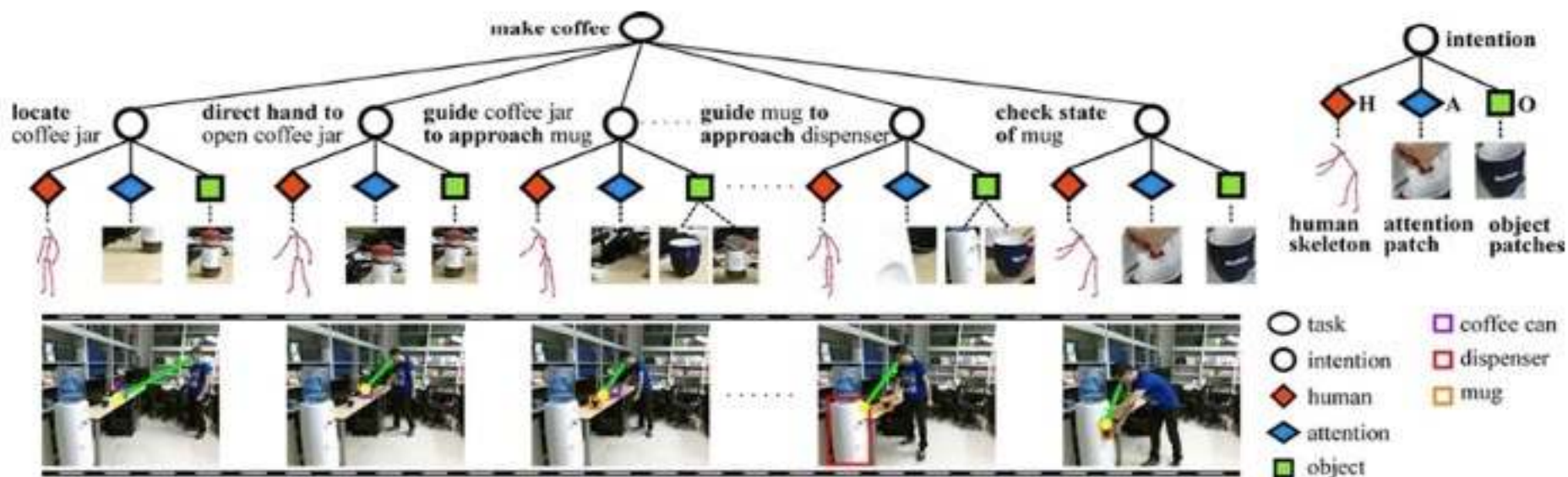


目标是：

1. 预测视频中人的意图
2. 计算视频中人的注意力
3. 识别视频中的复杂任务

图模型

西安交通大学人工智能学院魏平编写。课程资料，请勿外传



$I = \{I_t | t = 1, \dots, \tau\}$ 输入视频

$H = \{(h_t, x_t) | 1, \dots, \tau\}$ 人体特征

$L = \{l_t | t = 1, \dots, \tau\}$ 意图类别

$Y = \{y_t | t = 1, \dots, \tau\}$ 注意力

S 任务类别

$$\begin{aligned} \mathcal{E}(Y, L, S | I, H) = & \underbrace{\sum_{t=1}^{\tau} \Phi(\mathbf{h}_t, \mathbf{a}_t, \mathbf{o}_t, l_t)}_{\text{feature matching}} \\ & + \underbrace{\sum_{t=1}^{\tau} \Psi(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t, l_t)}_{\text{HAO geometric relation}} + \underbrace{\sum_{t=2}^{\tau} \Gamma(\mathbf{y}_{t-1}, \mathbf{y}_t, l_{t-1}, l_t)}_{\text{attention and intention transition}} \end{aligned}$$

演示1:意图行为状态联合推理

清华大学人工智能学院魏平编写。课程资料，请勿外传



RGB Video



RGBD Video

Jointly Inferring Human Minds from Videos

Human Action

Human Mind

Object Fluent

演示2: EyeKindle-助残阅读系统

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

Eye-Kindle

演示3: 人与机器人协作

西安交通大学人工智能学院魏平编写。课程资料，请勿外传

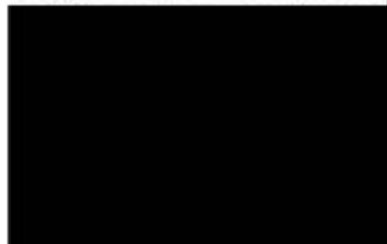
Baxter View



Left Wrist Camera



Right Wrist Camera



Hand-held Camera

