Markov Logic Network

Markov property

- 无记忆性，下一状态的概率分布今取决于当前状态（未来只和现在相关，和过去无关），即

  $$p(x_{n+1}=s_{n+1}|x_n=s_n,x_{n-1}=s_{n-1},...,x_1=s_1)=p(x_{n+1}=s_{n+1}|x_n=s_n)$$

- 明天的天气，下一时刻车站人数
- 不严谨，但计算高效

Markov Chain

- 定义
  - 一组随机变量 $X=\{X_1,X_2,...X_n\}$，每个随机变量的取值都在可数集内$X_i=s_i$，$s_i\in S$,如果随机变量的条件概率满足马尔可夫性质

    $$p(x_{n+1}=s_{n+1}|x_n=s_n,x_{n-1}=s_{n-1},...,x_1=s_1)=p(x_{n+1}=s_{n+1}|x_n=s_n)$$

    则X成为马尔可夫链，可数集S称为状态空间。

Markov Network

- 定义
  - 马尔可夫网是用来刻画一组随机变量 $X=\{X_1,X_2,...X_n\}$联合分布的模型，它由一个无向图G和一组势函数组成。
    - 每个随机变量对应G上的一个节点
    - G上的每个团都有一个势函数，势函数是从团的状态集到正实数集的映射
- 联合概率分布

  $$P(X=x)=\frac{1}{Z}\prod_k\phi_k(x_{\{k\}})$$

  其中ϕ_k就是势函数，$x_{\{k\}}$表示第k个团的状态（就是团中每个节点（随机变量）的取值），Z是归一化常数，$Z=\sum_x\prod_k\phi_k(x_{\{k\}})$,表示所有状态的下的团势能乘积和

  - 对数形式

    $$P(X=x)=\frac{1}{Z}exp(\sum_j\omega_i f_i(x))$$

    其中 ω_i为第 i个 团 的 权 重， f_i是 团 特 征 函 数
- 马尔可夫性质
  - 从联合概率分布公式可以看出，某个点的概率分布只和它的相邻节点相关，即只和包含它的团的状态相关。因为这个点相邻节点的取值会影响到包含它的团的特征函数的取值。

Markov Logic Network

- 一阶逻辑

Syntax

variables	$x, y, z, \ldots$
constants	$a, b, c, \ldots$
functions	$f, g, h, \ldots$
terms	variables, constants, or n-ary function applied to n terms as arguments
predicates	$p, q, r, \ldots$
atom	$\top$, $\bot$, or n-ary predicate applied to n terms
literal	atom or its negation

Syntax cont.

formula	literal, application of a logical connective $\{\neg, \wedge, \vee, \rightarrow, \leftrightarrow\}$ to formulae, or application of a *quantifier* to a formula

- Quantifiers
  - Existential: $\exists x.\ F[x]$
    "there exists an x such that $F[x]$"
  - Universal: $\forall x.\ F[x]$
    "for all x, $F[x]$"

ground term:闭项，基本项，不包含变量的项 $f(t_1, t_2)$

ground atom:闭原子，谓词参数是闭项的原子 $R(f_1(t_1), f_2(t_1, t_2))$

ground formula:闭公式，原子是闭原子的公式 $R_1(f_1(t_1), f_2(t_1, t_2)) \wedge R_2(f_1(t_1), f_2(t_1, t_2))$

- 马尔可夫逻辑网

定义 1[37]. Markov 逻辑网 L 是一组二元项(F_i,ω_i),其中,F_i 表示一阶逻辑规则,ω_i 是一个实数.这组二元项(F_i,ω_i)与一组有限常量集 $C=\{c_1,c_2,\dots,c_n\}$一起定义了一个 Markov 网 $M_{L,C}$:

(1) L 中的任意闭原子(ground atom)都对应了 $M_{L,C}$ 中的一个二值节点.若此闭原子为真,则对应的二值节点取值为 1;若为假,则取值为 0.

(2) L 中的任意闭规则(ground formula)都对应着一个特征值,若此闭规则为真,则对应的特征值为 1;若为假,则特征值为 0.并且,这个特征值的权重为二元项中该规则 F_i 对应的权重ω_i.

- 这里规则的权重w表示满足规则世界和不满足规则世界的差异性，ω越大，差异越大。
  - 比如 $A \to B$, 满足该规则的世界有{(T,T),(F,T),(F,F)}, 不满足的世界有{(T,F)}, ω越大，表示取到满足世界的可能性越大，不满足的可能性越小（但不代表不可能）。
    - 当然，如果$\omega \to \infty$,则不满足世界的可能性为0，表示不可能发生。和强逻辑是等价的
- **马尔可夫逻辑网其实就是生成马尔可夫网的模板**
- 马尔可夫逻辑网结合常量集合C生产的马尔可夫网结构:
  - 任意闭原子对应一个节点
  - 当两个节点表示的闭原子出现在同一规则中，则他们之间有条边
    - 因此出现在同一闭规则的闭原子节点构成一个团，团的权重就是规则的权重
- 举例

Table 1 A simple example of Markov logic network[37]

表 1 一个简单的 Markov 逻辑网实例[37]

Proposition	First-Order logic	Clausal form	Weight
Smoking causes cancer.	F_1:$\forall x, Sm(x) \Rightarrow Ca(x)$	$\neg Sm(x) \vee Ca(x)$	1.5
If two people are friends, either both smoke or neither does.	F_2:$\forall x \forall y, Fr(x,y) \Rightarrow (Sm(x) \Leftrightarrow Sm(y))$	$\neg Fr(x,y) \vee Sm(x) \vee \neg Sm(y)$	1.1
		$\neg Fr(x,y) \vee \neg Sm(x) \vee Sm(y)$	1.1

给定常量集合C={A,B}，则马尔可夫网$M_{L,C}$。可以看出，网络大小和参数最多谓词的参数数量，常量集大小呈指数相关，会非常大。

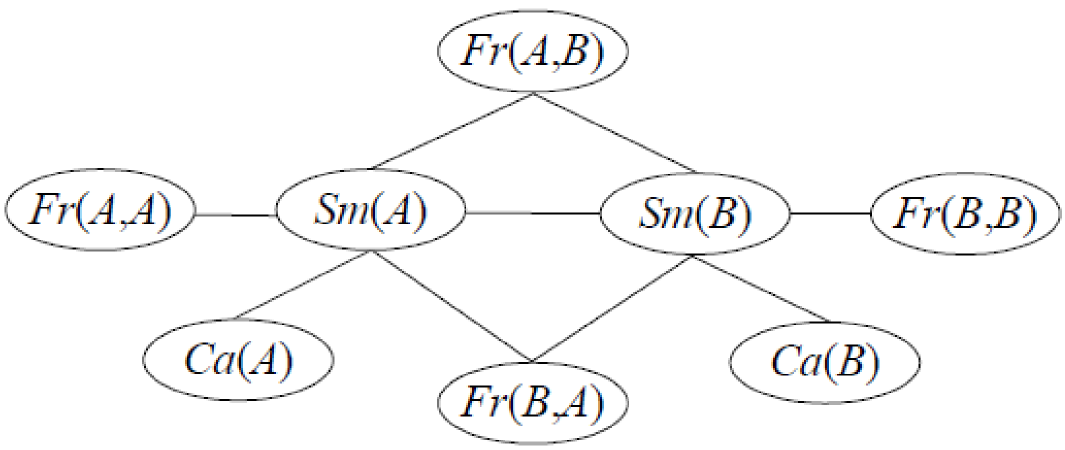

Fig. 1 An example of ground Markov network[37]

- 联合概率分布

$$P(X=x)=\frac{1}{Z}\exp\left\{\sum_i \omega_i n_i(x)\right\}=\frac{1}{Z}\prod_i \phi_i\left(x_{\{i\}}\right)^{n_i(x)}$$

其中ω_i表示规则(公式)F_i的权重，$n_i(x)$表示公式F_i在状态x下，为真的闭公式的个数

举例说明和一阶逻辑的关系

L = {($\forall x R(x) \rightarrow S(x)$,$\omega$)}, C={A}.

权重学习

类似于线性回归中的梯度下降算法（$y = wx + b$学习(w, b)）

x为试验样本

$$\frac{\partial}{\partial w_i} \log P_w(X=x) = n_i(x) - \sum_{x'} P_w(X=x')\, n_i(x')$$

推理

- 最大可能性问题

最大可能性问题是概率图模型推理中的重要内容,其基本过程可表述为:给定证据变量集 x,求变量集 y 最可能处于的状态,即求

$$\max_y P(y|x) \tag{4}$$

在 Markov 逻辑网中,结合公式(3),则问题转化为求

$$\max_y \sum_i w_i n_i(x, y) \tag{5}$$

其中,w_i 表示从句 i 的权重,$n_i(x,y)$表示从句 i 的闭从句的真值数量.

因此,计算最大可能性问题可转换为典型的最大化带权的可满足性问题,即寻找一组变量的赋值,使得所有

- 条件概率

则 F_2 取值为真时,规则 F_1 取值为真的概率为

$$P(F_1 | F_2, L, C) = \frac{P(F_1 \wedge F_2 | M_{L,C})}{P(F_2 | M_{L,C})} = \frac{\sum_{X \in \theta_{F_1} \cap \theta_{F_2}} P(X = x | M_{L,C})}{\sum_{X \in \theta_{F_2}} P(X = x | M_{L,C})}$$