Introducción a *Markov Logic Networks*

Franco M. Luque

Lógica de primer orden

- Una *first-order knowledge base (KB)* es:
  - conjuntos de nombres de constantes, funciones (+aridad) y predicados (+aridad)
  - y un conjunto de fórmulas en lógica de primer orden con estos elementos.
- Problema de inferencia: ¿Vale F para todo mundo posible de KB?
- Pasamos las fórmulas a *conjunctive normal form (CNF)*. Las variables libres implícitamente estan cuantificadas con $\forall$.
- Suponemos entonces que cada fórmula es una disjunción de literales.

Probabilisticalización del problema

- Asignamos a cada fórmula F_i un peso w_i (esta será la definición de MLN).
- Ahora los mundos posibles serán más o menos probables dependiendo de cuántas fórmulas de la KB respeten.
- Problema de inferencia: ¿Cuál es la prob. de F_1 dado que vale F_2?:

$$P(F_1|F_2) = \frac{\sum_{x \in \chi_{F_1} \cap \chi_{F_2}} P(X = x)}{\sum_{x \in \chi_{F_2}} P(X = x)}$$

- Falta definir $P(X = x)$, a donde x es un mundo posible.

Formalización

- Tenemos $L = \{(F_i, w_i)\}$ (la MLN) y las cttes. $C = \{c_1, \dots, c_{|C|}\}$
- Suposiciones:
  1. Nombres únicos: cada ctte. es un elemento distinto.
  2. Dominio cerrado: todo elemento es alguna constante.
  3. Funciones conocidas: se conocen las funciones.
- Ahora un mundo posible x es una "interpretación" de los predicados. Hacemos $x = (x_1, \dots, x_n)$, a dónde cada x_i indicará el valor booleano de un predicado aplicado a una tupla de cttes. (esto es, un *ground atom*).
- El problema se convierte en modelar la distribución conjunta de un vector de variables aleatorias binarias.

Más formalización

- Un *grounding* de una fórmula F es la fórmula asignando valores concretos (cttes.) a las variables.
- Dada una fórmula F y un mundo posible x, podemos contar cuántos *groundings* de F dan verdaderos suponiendo x.
- Modelamos

$$P(X = x) = \frac{1}{Z} exp(\sum_i w_i n_i(x))$$

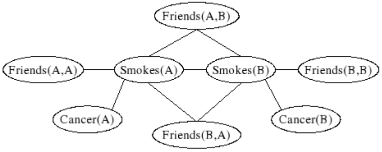
a donde $n_i(x)$ es la cantidad de *groundings* verdaderos de F_i en x (y $Z = \sum_x exp(\sum_i w_i n_i(x))$).

Markov Networks

- El modelo propuesto es una instancia de *Markov Networks*.
- Para trabajar con el modelo definimos un grafo G con un nodo por variable X_i, es decir, por *ground atom*.
- Cada *grounding* de una fórmula conecta todos los nodos de los *ground atoms* que lo componen.
- La colcha de Markov o *Markov blanket* de un nodo X_i son sus nodos vecinos, es decir, los nodos que aparecen con X_i en algún *grounding* de alguna fórmula de la MLN.

Bue, esta bien, un ejemplo:

- Predicados Sm (fuma), Ca (tiene cancer), Fr (son amigos):
  1. $F_1 = \forall x\ Sm(x) \Rightarrow Ca(x)$, $w_1 = 1,5$.
  2. $F_2 = \forall x \forall y\ Fr(x,y) \Rightarrow (Sm(x) \Leftrightarrow Sm(y))$, $w_2 = 2,2$.
- En CNF:
  1. $F_1 = \neg Sm(x) \vee Ca(x)$, $w_1 = 1,5$.
  2. $F_2 = \neg Fr(x,y) \vee Sm(x) \vee \neg Sm(y)$, $w_2 = 1,1$.
  3. $F_3 = \neg Fr(x,y) \vee \neg Sm(x) \vee Sm(y)$, $w_3 = 1,1$.
- Constantes $C = \{Anna, Bob\}$.



Recordamos el problema de inferencia

- Sean F_1 y F_2 conjunciones de *ground literals*.
- Problema de inferencia: ¿Cuál es la prob. de F_1 dado que vale F_2?:
$$P(F_1|F_2) = \frac{\sum_{x \in \chi_{F_1} \cap \chi_{F_2}} P(X = x)}{\sum_{x \in \chi_{F_2}} P(X = x)}$$
- Esto no se puede calcular ni a palos (NP-completo), aunque sea todo finito. Tampoco $P(X = x)$.
- Se puede aproximar muestreando X tales que cumplen F_2 y viendo cuántos cumplen también F_1 (método de Monte Carlo).

Monte Carlo con cadenas de Markov (MCMC)

- Una forma de obtener una muestra es simulando una cadena de Markov $X^0, X^1, \ldots$ cuya dist. estacional sea $P(X = x)$.
- Hagamos $\pi_x = P(X = x)$ y $p_{xy} = P(X^{t+1} = y | X^t = x)$.
- Un teorema dice que si

$$\pi_x p_{xy} = \pi_y p_{yx}$$

  entonces π es la dist. estacional de la cadena p_{yx}.
- Suele ser fácil encontrar cadenas para muestrear cualquier $P(X = x)$.
- Un ejemplo de MCMC es el algoritmo de Metropolis-Hastings, y en particular el muestreo de Gibbs.

Muestreo de Gibbs: Definición

- Para obtener muestras de $P((X_1, \ldots, X_n) = (x_1, \ldots, x_n))$ usando la dist. condicional de cada variable.
- Es un MCMC. Algoritmo para obtener x^{t+1} de x^t:
    1. Elegir $j \in \{1, \ldots, n\}$ al azar.
    2. Elegir x_j con prob. $P(X_j = x_j | X_k = x_k^t, k \neq j)$.
    3. $x^{t+1} = x^t_{[j \leftarrow x_j]}$.
- Formalmente,

$$p_{xy} = \begin{cases} \frac{1}{n} P(X_j = y_j | X_k = y_k, k \neq j) & \text{si } \forall k \neq j : y_k = x_k \\ 0 & \text{c.c.} \end{cases}$$

Muestreo de Gibbs: Funciona

- Si $\pi_x = P(X = x)$, se puede probar que

$$\pi_x p_{xy} = \pi_y p_{yx}$$

- Esquema de la prueba: si $\forall k \neq j : y_k = x_k$,

$$\begin{aligned}
\pi_x p_{xy} &= P(X = x)\, \frac{1}{n}\, P(X_j = y_j | X_k = y_k, k \neq j) \\
&= P(X = x)\, \frac{1}{n}\, \frac{P(X = y)}{\sum_{z : z_k = x_k, k \neq j} P(X = z)} \\
&= \ldots \\
&= \pi_y p_{yx}
\end{aligned}$$

Uso de Gibbs para inferencia

- Dijimos
$$P(X = x) = \frac{1}{Z} exp(\sum_i w_i n_i(x))$$
- Tenemos que calcular $P(X_j = x_j | X_k = x_k, k \neq j)$.
- Ahora,
$$P(X_j = x_j | X_k = x_k, k \neq j) = P(X_j = x_j | B_j = b_j)$$
a donde B_j es la colcha de Markov de X_j.
- Es decir, X_j sólo depende de sus nodos vecinos en la *Markov Network*.

Uso de Gibbs para inferencia

- Sea F_j el cjto. de *groundings* de fórmulas que tienen a X_j.
- Si $f \notin F_j$, su valor de verdad no depende de X_j.
- Si $f \in F_j$, su valor de verdad **sólo** depende de X_j y B_j.
- Esquema de la prueba:

$$
\begin{aligned}
P(X_j = x_j | X_k = x_k, k \neq j) &= \frac{P(X = x)}{P(X = x_{[j \leftarrow 0]}) + P(X = x_{[j \leftarrow 1]})} \\
&= \frac{exp(\sum_i w_i n_i(x))}{exp(\sum_i w_i n_i(x_{[j \leftarrow 0]})) + exp(\sum_i w_i n_i(x_{[j \leftarrow 1]}))} \\
&= \ldots \\
&= \frac{exp(\sum_{f \in F_j} w_f f(x_j, b_j))}{exp(\sum_{f \in F_j} w_f f(0, b_j)) + exp(\sum_{f \in F_j} w_f f(1, b_j))} \\
&= P(X_j = x_j | B_j = b_j)
\end{aligned}
$$

Ok, un ejemplo:

- $X_j = Friends(B, A)$ y $B_j = (Smokes(A), Smokes(B))$.
- Calculemos $P(X_j = x_j | B_j = b_j)$ si $b_j = (1, 0)$.
- Los groundings en los que aparece $Friends(B, A)$ son:
  1. $F_2(B, A) = \neg Fr(B, A) \vee Sm(B) \vee \neg Sm(A)$, $w_2 = 1, 1$.
  2. $F_3(B, A) = \neg Fr(B, A) \vee \neg Sm(B) \vee Sm(A)$, $w_3 = 1, 1$.
- Luego
  1. $P(X_j = 0 | B_j = b_j) = \frac{exp(2{,}2)}{exp(2{,}2)+exp(1{,}1)}$
  2. $P(X_j = 1 | B_j = b_j) = \frac{exp(1{,}1)}{exp(2{,}2)+exp(1{,}1)}$