Domain Aware Markov Logic Networks

Happy Mittal, Ayush Bhardwaj — Dept. of CSE, IIT Delhi

Vibhav Gogate — Dept. of CS, UT Dallas

Parag Singla — Dept. of CSE, IIT Delhi

INTRODUCTION

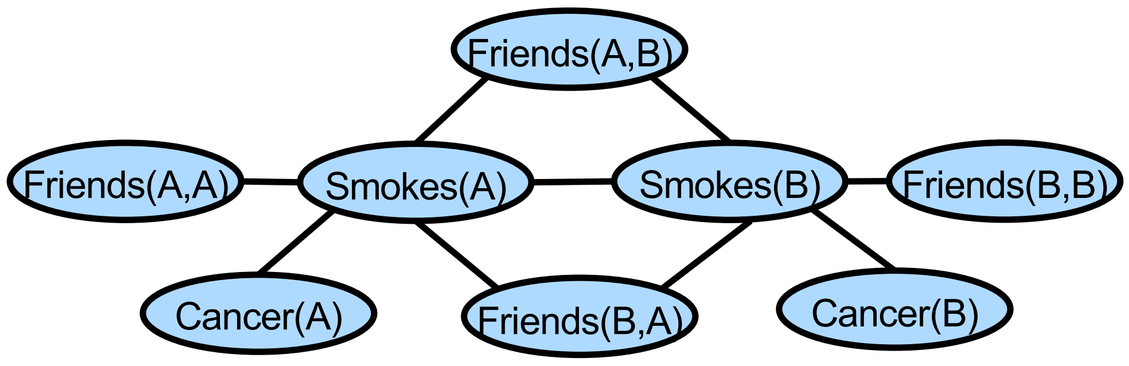
Markov Logic Networks (MLN) [Richardson & Domingos, 2006]

- A Markov Logic Network (MLN) is a set of weighted first order formulas.
- When a world violates a formula, it becomes less probable, but not impossible.
- Together with a set of constants, it defines a Markov network.

1.5 $\forall x\ Smokes(x) \Rightarrow Cancer(x)$

1.1 $\forall x, y\ Smokes(x) \wedge Friends(x,y) \Rightarrow Smokes(y)$

$\Delta x = \Delta y = \{A, B\}$



- **Joint Probability :**

y : query atoms
x : evidence atoms

$$P(y|x;w) = \frac{1}{Z_x} \exp\left(\sum_i w_i\, n_i(x,y)\right)$$

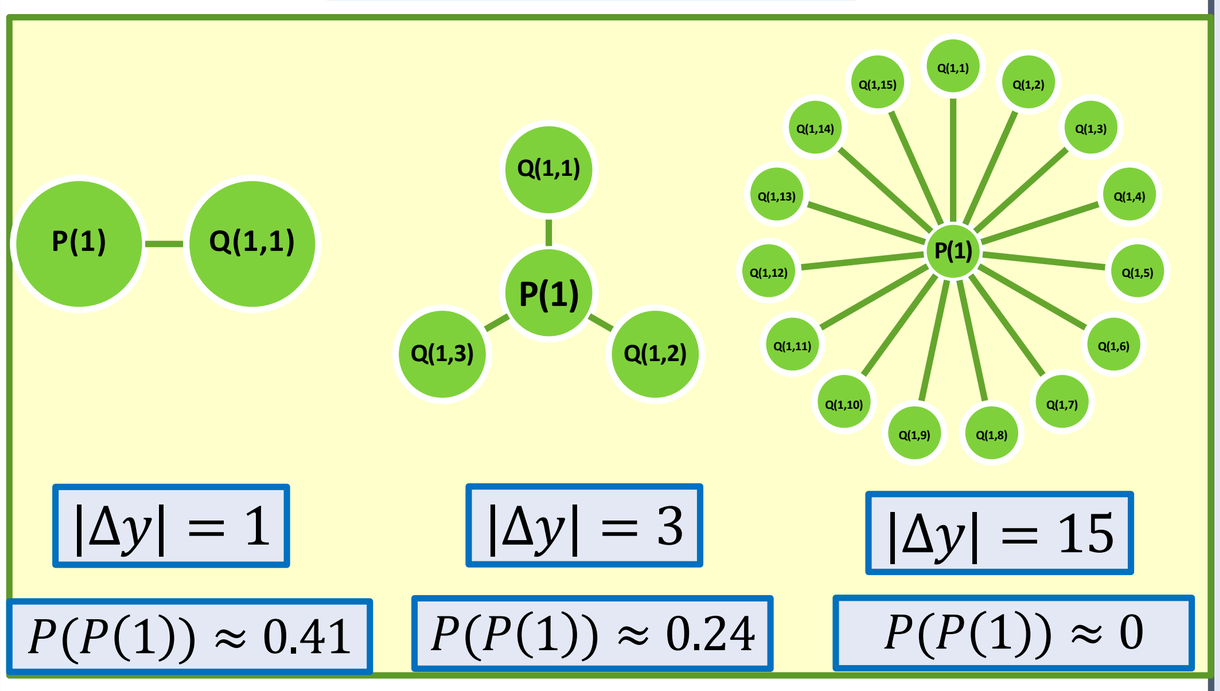
w_i: Weight of Formula i

$n_i(x,y)$: No. of true groundings of formula i

MOTIVATION

- Training : Small size domain
  - Unavailability of labelled datasets.
- Testing : Large size domain
- Observation : Extreme Marginal Probabilities

$$1.0 : P(x) \Rightarrow Q(x,y)$$



Use number of connections to adjust the weights.

RELATED WORK

- Identifying the problem :
  - [Poole et al 2014] : Characterized the cases of extreme marginals.
- Solving the Problem :
  - Adaptive MLNs [Jain et al 2010]
  - Relational Marginal Problems : Theory and Estimation [Kuzelka et al 2017

DA-MLNS

Number of Connections

$$F = [P_1, P_2, \ldots, P_m]$$

Formula — *Predicates*

$$c_j = \max\left(1, \prod_{x \in var(P_j)^-} |\Delta x|\right)$$

c_j: NumConnections of jth Predicate

$var(P_j)^-$: Variables in F not in P_j

Example

$w: S(x) \wedge F(x,y) \Rightarrow S(y)$

$\Delta x = \Delta y = \{1,2,\ldots,10\}$

c_1=10, c_2=1, c_3=10

$s = 10$

Scaling Factor

$$s = \max_{1 \le j \le m} c_j$$

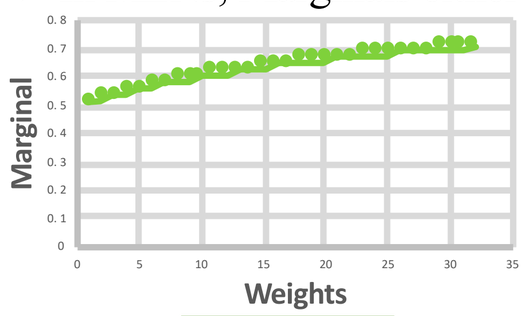
Probability Distribution

$$P(y|x;w) = \frac{1}{Z_x} \exp\left(\sum_i \frac{w_i}{s_i} n_i(x,y)\right)$$

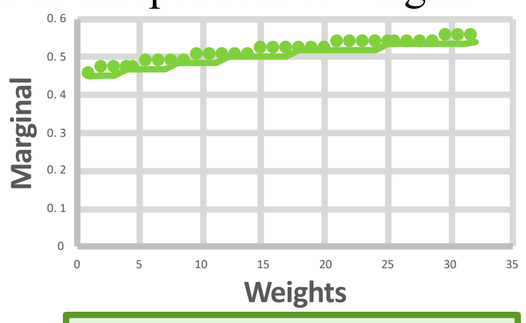
CHARACTERIZATION OF MLNS & DA-MLNs

Formula Type	Dom	$P(P(1)=1)$	
		MLN	DA-MLN
$P(x) \vee Q(x,y)$	Large $\lvert\Delta y\rvert$	**1.0**	exp(-w/2)
$P(x) \vee Q(y)$	$\lvert\Delta y\rvert = \lvert\Delta x\rvert$	0.75	Depends on w
$P(x) \vee Q(y)$	$\lvert\Delta y\rvert > \lvert\Delta x\rvert$	**1.0**	Depends on w
$P(x) \vee R(x,y) \vee Q(y)$	$\lvert\Delta y\rvert > \lvert\Delta x\rvert$	**1.0**	Depends on w
$P(1) \vee Q(y) \vee R_1 \vee \cdots \vee R_m$	$\lvert\Delta y\rvert > \lvert\Delta x\rvert$	$1-\frac{1}{2^m}$	Depends on w

- In MLNs, Marginals either extreme or independent of weights.



$P(x) \vee Q(y)$



$P(1) \vee Q(y) \vee R_1 \vee \cdots \vee R_m$

EXPERIMENTS & RESULTS

Friends & Smokers

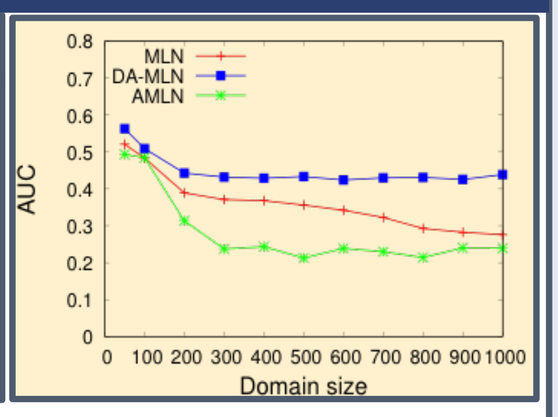
Formulas :

$S(x) \Rightarrow C(x)$

$S(x) \wedge F(x,y) \Rightarrow S(y)$

Data Generation :

- Communities of Smoker & Non-Smoker.
- Smokers have high chances of having Cancer.



IMDB

Formulas :

$Act(x) \wedge Dir(y) \wedge Mov(m,x) \wedge WU(x,y) \Rightarrow Mov(m,y)$

$Act(x) \wedge Dir(y) \wedge Mov(m,y) \wedge WU(x,y) \Rightarrow Mov(m,x)$ (and 6 more)

Dataset :

- https://www.kaggle.com/PromptCloudHQ/imdb-data/data.