FONDAZIONE BRUNO KESSLER

UNIVERSITY OF TRENTO - Italy
Department of Information Engineering and Computer Science

On Projectivity in Markov Logic Netorks

Sagar Malhotra[1,2] and Luciano Serafini [1]

[1]Fondazione Bruno Kessler
[2]University of Trento

Preliminaries: Probability Distribution

$\boldsymbol{X} = \langle X_1, ..., X_n \rangle$ is a boolean random variable and $P_{\boldsymbol{\theta}}^{(n)}(\boldsymbol{X})$ is a probability distribution

Preliminaries: Probability Distribution

$\boldsymbol{X} = \langle X_1, ..., X_n \rangle$ is a boolean random variable and $P_{\boldsymbol{\theta}}^{(n)}(\boldsymbol{X})$ is a probability distribution

An example of $P_{\boldsymbol{\theta}}^{(3)}(\boldsymbol{X})$

X_1	X_2	X_3	$P_{\boldsymbol{\theta}}^{(3)}(\boldsymbol{X})$
1	1	1	0.1
1	1	0	0.15
1	0	1	0.15
1	0	0	0.1
0	1	1	0.15
0	1	0	0.1
0	0	1	0.1
0	0	0	0.15

Preliminaries: Marginal Probability Distribution

We can obtain a Marginal Distribution on $\boldsymbol{X}' := \langle X_1, \ldots X_q \rangle$ as follows:

$$P_{\boldsymbol{\theta}}^{(n)} \downarrow [q](\boldsymbol{X}' = \omega') := \sum_{\omega' = \omega[1:q]} P_{\boldsymbol{\theta}}^{(n)}(\omega)$$

X_1	X_2	X_3	$P_{\boldsymbol{\theta}}^{(3)}(\boldsymbol{X})$
1	1	1	0.1
1	1	0	0.15
1	0	1	0.15
1	0	0	0.1
0	1	1	0.15
0	1	0	0.1
0	0	1	0.1
0	0	0	0.15

X_1	X_2	$P_{\boldsymbol{\theta}}^{(3)} \downarrow [2](\boldsymbol{X}')$
1	1	0.25
1	0	0.25
0	1	0.25
0	0	0.25

Mutations in Trentino

Let n be the population of Trentino, such that each person x has a label $i \in \{1 \ldots n\}$.

Goal: Surveying Trentino for Mutations

There is a genetic mutation M found in the population of Trentino. Any person, labelled i in Trentino, either has this mutation $(M(i) = 1)$ or not $(M(i) = 0)$.

Mutations in Trentino

Let n be the population of Trentino, such that each person x has a label $i \in \{1 \dots n\}$.

Goal: Surveying Trentino for Mutations

There is a genetic mutation M found in the population of Trentino.
Any person, labelled i in Trentino, either has this mutation $(M(i) = 1)$ or not $(M(i) = 0)$.
We want to model the distribution of mutations:

- A parametric probability distribution: $P_{\boldsymbol{\theta}}^{(n)}$
- Given the population of Trentino, we will estimate the parameters of $P_{\boldsymbol{\theta}}^{(n)}$

Mutations in Trentino: Formal Setup

We can model the space of possible populations of size n, by:

$$\Delta^{(n)} = \{0, 1\}^n$$

If we have $\omega \in \Delta^{(n)}$, such that $\omega[i] = 1$, then the person with label i has the mutation M.

Mutations in Trentino: Formal Setup

We can model the space of possible populations of size n, by:

$$\Delta^{(n)} = \{0,1\}^n$$

If we have $\omega \in \Delta^{(n)}$, such that $\omega[i] = 1$, then the person with label i has the mutation M.

A population of 3 people

- $\Delta^{(3)} = \{\langle 1,1,1\rangle, \langle 1,1,0\rangle, \langle 1,0,0\rangle, \langle 0,0,0\rangle \dots\}$
- $P_{\boldsymbol{\theta}}^{(3)} : \Delta^{(3)} \to [0,1]$

Mutations in Trentino: Binomial Distribution

A possible model :

$$P_{\boldsymbol{\theta}}^{(n)}(\omega) = \binom{n}{k} p^k (1-p)^{n-k}$$

where $\boldsymbol{\theta} = \{p\}$ and n is the population of Trentino.

Mutations in Trentino: Binomial Distribution

A possible model :

$$P_{\boldsymbol{\theta}}^{(n)}(\omega) = \binom{n}{k} p^k (1-p)^{n-k}$$

where $\boldsymbol{\theta} = \{p\}$ and n is the population of Trentino.

Ideally, we observe the entire population of Trentino, say $\omega \in \Delta^{(n)}$ and get the estimate for p as follows:

$$p_n^* = \underset{p}{\operatorname{argmax}} \quad P_{\boldsymbol{\theta}}^{(n)}(\omega)$$

Mutations in Trentino: Sampling

But we cant survey the entire Trentino. Hence, we survey (observe) the commune of Trento: $\omega' \sim \Delta^{(q)}$

Mutations in Trentino: Sampling

But we cant survey the entire Trentino. Hence, we survey (observe) the commune of Trento: $\omega' \sim \Delta^{(q)}$

We estimate $P_{\boldsymbol{\theta}}^{(q)} : \Delta^{(q)} \rightarrow [0, 1]$ given by a Binomial Distribution:

$$P_{\boldsymbol{\theta}}^{(q)}(\omega') = \binom{q}{k} p^k (1-p)^{q-k}$$

$$p_m^* = \underset{p}{\operatorname{argmax}} \quad P_{\boldsymbol{\theta}}^{(q)}(\omega')$$

Mutations in Trentino: Sampling

But we cant survey the entire Trentino. Hence, we survey (observe) the commune of Trento: $\omega' \sim \Delta^{(q)}$

We estimate $P_{\boldsymbol{\theta}}^{(q)} : \Delta^{(q)} \to [0, 1]$ given by a Binomial Distribution:

$$P_{\boldsymbol{\theta}}^{(q)}(\omega') = \binom{q}{k} p^k (1-p)^{q-k}$$

$$p_m^* = \underset{p}{\operatorname{argmax}} \quad P_{\boldsymbol{\theta}}^{(q)}(\omega')$$

We have $P_{\boldsymbol{\theta}}^{(q)}$, but what we wanted was $P_{\boldsymbol{\theta}}^{(n)}$!

Mutations in Trentino : Gneralising

SOLUTION: We just take $P_{\boldsymbol{\theta}}^{(q)}$ and we plug it's parameters in $P_{\boldsymbol{\theta}}^{(n)}$! i.e.

$$p_n^* \leftarrow p_q^*$$

Probabilsitic Consistency

After $p^*_n \leftarrow p^*_q$, does our model change it's mind about Trento ?

PROBABILSITIC CONSISTENCY

After $p_n^* \leftarrow p_q^*$, does our model change it's mind about Trento ?

i.e.

Is the marginal probability of ω' under $P_{\theta}^{(q)}$ same as $P_{\boldsymbol{\theta}}^{(n)} \downarrow [q](\omega')$?

PROBABILSITIC CONSISTENCY

After $p_n^* \leftarrow p_q^*$, does our model change it's mind about Trento ?

i.e.

Is the marginal probability of ω' under $P_\theta^{(q)}$ same as $P_{\boldsymbol{\theta}}^{(n)} \downarrow [q](\omega')$?

i.e.

Is the following True ?

$$P_{\boldsymbol{\theta}}^{(n)} \downarrow [q] = P_{\boldsymbol{\theta}}^{(q)}$$

Probabilsitic Consistency

After $p_n^* \leftarrow p_q^*$, does our model change it's mind about Trento ?

i.e.

Is the marginal probability of ω' under $P_\theta^{(q)}$ same as $P_{\boldsymbol{\theta}}^{(n)} \downarrow [q](\omega')$?

i.e.

Is the following True ?

$$P_{\boldsymbol{\theta}}^{(n)} \downarrow [q] = P_{\boldsymbol{\theta}}^{(q)}$$

Good News

For Binomial Distribution :

$$P_{\boldsymbol{\theta}}^{(n)} \downarrow [q] = P_{\boldsymbol{\theta}}^{(q)} \tag{1}$$

We call probability distributions with property (1) Projective.

Infernce from sub sampled domains: Framework

Sample : $[m] \sim [n]$

Learn : $\boldsymbol{\theta}^* \leftarrow \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \quad P_{\boldsymbol{\theta}}^{(m)}$

Generalize : $\boldsymbol{\theta}^* \rightarrow P_{\boldsymbol{\theta}}^{(n)}$

Demographics of Trentino: Relational Case

Goal: Modeling Family Influence on Gene Mutation

- A unary predicate $M(x)$ denoting if x has a mutation or not
- A symmetric and irreflexive relation $R(x, y)$ denoting that x is relative of y

And the following features:

- $f_1(x) : M(x)$
- $f_2(x, y) : M(x) \wedge R(x, y) \rightarrow M(y)$

Probabilistic Model: Markov Logic Network

Given a population $\omega \in \Delta^{(n)}$, we express a Markov Logic Network as follows:

Probabilistic Model: Markov Logic Network

Given a population $\omega \in \Delta^{(n)}$, we express a Markov Logic Network as follows:

$$P_{\boldsymbol{\theta}}^{(n)}(\omega) := \frac{1}{Z} \times \exp(N_1(\omega)w_1 + N_2(\omega)w_2)$$

where,

- $f_1(x) : M(x)$
- $f_2(x,y) : M(x) \wedge R(x,y) \rightarrow M(y)$
- $N_1(\omega) := \sum_i 1_{\omega \models f_1(i)}$
- $N_2(\omega) := \sum_{i,j} 1_{\omega \models f_2(i,j)}$
- $Z = \sum_{\omega' \in \Delta^{(n)}} \exp(N_1(\omega')w_1 + N_2(\omega')w_2)$

Markov Logic Networks: Sampling

We cant survey entire Trentino !

Markov Logic Networks: Sampling

We cant survey entire Trentino ! Hence, we survey Trento ☺!

We get our model on Trento $P_{\boldsymbol{\theta}}^{(m)}$ ☺!!

Markov Logic Networks: Sampling

We cant survey entire Trentino ! Hence, we survey Trento ☺!

We get our model on Trento $P_{\boldsymbol{\theta}}^{(m)}$ ☺!!

We take the it's parameters $\boldsymbol{\theta}*$ and we plug them in $P_{\boldsymbol{\theta}}^{(n)}$! Hurrah ☺!

Markov Logic Networks: Consistency

With respect to the model of Trento i.e. $P_{\boldsymbol{\theta}}^{(q)}$: The Probability of a Alessandro having the mutation is $\mathbf{0.1}$

Markov Logic Networks: Consistency

With respect to the model of Trento i.e. $P_{\boldsymbol{\theta}}^{(q)}$: The Probability of a Alessandro having the mutation is $\mathbf{0.1}$

But we realise that this probability is $\mathbf{0.9}$ w.r.t $P_{\boldsymbol{\theta}}^{(n)} \downarrow [q]$

What must we believe ??? ☹

Markov Logic Networks: Projectivity

There is no objective way to come out of this situation in general

Markov Logic Networks: Projectivity

There is no objective way to come out of this situation in general

MLNs (and most models on relational data) are probabilistically inconsistent i.e. they are not projective !

Projectivity in MLNs

"In this work, our goal is to charectierize the set of conditions such that MLNs are *Projective*"

Projectivity in MLNs

"In this work, our goal is to charectierize the set of conditions such that MLNs are *Projective*"

As a first step, we provide these conditions for the two variable fragment i.e. MLN with features involving at most 2 variables

Revisiting Trentino

In Trentino, we had a unary predicate $M(c)$ for every person c. We had a binary relation $R(c,d)$ for every pair, and we have that R is symmetric and irreflexive.

Revisiting Trentino

In Trentino, we had a unary predicate $M(c)$ for every person c. We had a binary relation $R(c,d)$ for every pair, and we have that R is symmetric and irreflexive.

Hence, we have that:

- For every person c, $\omega \models M(c)$ xor $\omega \models \neg M(c)$
- For every pair of persons (c,d), $\omega \models R(c,d)$ xor $\omega \models \neg R(c,d)$

Parametric Normal Form MLN

Any Markov Logic Network over the Trentino Example can be expressed as follows:

- $M(x) : s_1$
- $\neg M(x) : s_2$
- $M(x) \land M(y) \land R(x,y)$: w_1^{11}
- $M(x) \land M(y) \land \neg R(x,y)$: w_2^{11}
- $M(x) \land \neg M(y) \land R(x,y)$: w_1^{12}
- $M(x) \land \neg M(y) \land \neg R(x,y)$: w_2^{12}
- $\neg M(x) \land \neg M(y) \land R(x,y)$: w_1^{22}
- $\neg M(x) \land \neg M(y) \land \neg R(x,y)$: w_2^{22}

Parametric Normal Form MLN

$$P_\theta^{(n)}(\omega) := \frac{1}{Z(n)} \times \prod_{1\leq i\leq 2} (s_i)^{k_i} \prod_{1\leq i\leq j\leq 2} \prod_{1\leq l\leq 2} (w_l^{ij})^{h_l^{ij}}$$

- $k_1 := \left|\{c : \omega \models M(x)\}\right|$
- $k_2 := \left|\{c : \omega \models \neg M(x)\}\right|$
- $h_1^{11} := \left|\{(c,d) : \omega \models M(x) \wedge M(x) \wedge R(x,x)\}\right|$
- $h_2^{11} := \left|\{(c,d) : \omega \models M(x) \wedge M(x) \wedge \neg R(x,x)\}\right|$
- $h_1^{12} := \left|\{(c,d) : \omega \models M(x) \wedge \neg M(x) \wedge R(x,x)\}\right|$
- $h_2^{12} := \left|\{(c,d) : \omega \models M(x) \wedge \neg M(x) \wedge \neg R(x,x)\}\right|$

$\vdots$

Parametric Normal Form MLN

We can create such a parametric form for any MLN

$$P_\theta^{(n)}(\omega) := \frac{1}{Z(n)} \times \prod_{1 \le i \le u} (s_i)^{k_i} \prod_{1 \le i \le j \le u} \prod_{1 \le l \le b} (w_l^{ij})^{h_l^{ij}}$$

Necessary and Sufficient Conditions for Projectivity

An MLN is projective if :

$$\forall i, j : \sum_{l} w_l^{ij} = \sum_{l} w_l^{ij} = S$$

Why does it work ?

$$\forall i,j : \sum_l w_l^{ij} = \sum_l w_l^{ij} \implies Z(n) = (\sum_i s_i)^n \times (S)^{\binom{n}{2}}$$

Why does it work ?

If $Z(n) = (\sum_i s_i)^n \times (S)^{\binom{n}{2}}$, then the MLN reduces to:

$$P_\theta^{(n)}(\omega) = \prod_{1 \le i \le u} \left(\frac{s_i}{\sum_i s_i}\right)^{k_i} \prod_{1 \le i \le j \le u} \prod_{1 \le l \le b} \left(\frac{w_l^{ij}}{S}\right)^{h_l^{ij}}$$
$$= \prod_{1 \le i \le u} p_i^{k_i} \prod_{1 \le i \le j \le u} \prod_{1 \le l \le b} (w_{ijl})^{h_l^{ij}}$$

It can be easily proven that:

$p_i = P(c$ has the i^{th} property $)$

$w_{ijl} = P((c, d)$ has the l^{th} binary property — c and d have the i^{th} and $j^{th})$

PROJECTIVITY AND LEARNING

The maximum likelihood estimate is simply:

$$p_i = \frac{k_i}{n} \qquad w_{ijl} = \frac{h_l^{ij}}{k_i \times k_j}$$

Projectivity and Learning

The maximum likelihood estimate is simply:

$$p_i = \frac{k_i}{n} \qquad w_{ijl} = \frac{h_l^{ij}}{k_i \times k_j}$$

The marginal inference tasks like : $P(R(x,y) \wedge R(y,z) \rightarrow R(x,z))$ can be computed in constant time w.r.t domain cardinality!

Conclusion

Projectively is a gift but it’s also a curse (I guess)!