Probabilistic Constraint Logic Theories

Marco Alberti[1] Elena Bellodi[2] Giuseppe Cota[2] Evelina Lamma[2]
Fabrizio Riguzzi[1] Riccardo Zese[2]

Dipartimento di Matematica e Informatica – University of Ferrara

Dipartimento di Ingegneria – University of Ferrara
name.surname@unife.it

August 30, 2016

Outline

Motivations

Inference Problem

- Probabilistic logic models are gaining popularity due to their successful application in a variety of fields
- They usually require expensive inference procedures
- Many proposals to achieve tractability: Tractable Markov Logic, Tractable Probabilistic Knowledge Bases and fragments of probabilistic logics
  - They limit the form of sentences

Learning Problem

- Learning from entailment presents tractability problems.
  - The coverage problem consists in checking whether an atom follows from a logic program.

Integrity Constraints: a Possible Solution

- If logic theories are sets of integrity constraints and examples are interpretations
  - coverage problem consists in verifying whether the constraints are satisfied in the interpretations
  - the constraints can be considered in isolation: the interpretation satisfies the constraints iff it satisfies all of them individually
    $\rightarrow$ the learning from interpretation setting offers advantages in term of tractability
- Moreover...
  - they are useful for system verification or in the problem of checking whether a systems behaviour is compliant to a specification

Probabilistic Inference

- In Probabilistic Logic Programming (PLP) the distribution semantics is one of the most successful approaches.
  - The probability distribution over normal logic programs (worlds) is extended to queries and the probability of a query is obtained by marginalizing the joint distribution of the query and the programs
- Performing inference requires an expensive procedure that is usually based on knowledge compilation
  - ProbLog [De Raedt et al., 2007] and PITA [Riguzzi and Swift, 2011, Riguzzi and Swift, 2013] build a Boolean formula and compile it into a Binary Decision Diagram (compilation procedure is #P)

Probabilistic Constraint Logic Theories

- We consider a probabilistic version of sets of integrity constraints similar to distribution semantics
  - each integrity constraint is annotated with a probability
  - a model assigns a probability of being positive to interpretations
- Differently from PLP approaches under the distribution semantics
  - computing the probability of the positive class given an interpretation in a PCLT is logarithmic in the number of variables
  - PCLTs define a conditional probability distribution over a random variable C representing the class, given an interpretation

Syntax

A Constraint Logic Theory (CLT) T is a set of integrity constraints (ICs) C of the form

$$L_1, \dots, L_b \rightarrow A_1; \dots; A_h \tag{1}$$

where

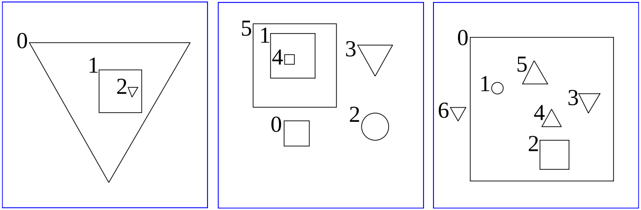
- $L_1, \dots, L_b$ is a conjunction of logical literals called *body*
- $A_1; \dots; A_h$ is a disjunction of atoms called *head*

We may also have a background knowledge B on the domain which is a normal logic program that can be used to represent domain-specific knowledge

Semantics

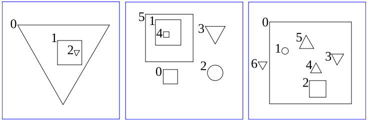
- CLTs can be used to classify Herbrand interpretations by considering a model $M(B \cup I)$ which follows the Prolog semantics
  - I is interpreted as the set of ground facts true in $M(B \cup I)$
  - $M(B \cup I)$ can contain new facts derived from I using B
- Given an interpretation I, a background knowledge B and a constraint C
  - we can ask whether C is true in I given B
    - $M(B \cup I) \models C$, if for every substitution θ for which $Body(C)$ is true in $M(B \cup I)$, there exists a disjunct in $Head(C)$ that is true in $M(B \cup I)$

Running Example: Bongard Problems



- Bongard Problems consist of a number of pictures, some positive and some negative
- Aim: learning a description which correctly classify the most figures
- The pictures contain different shapes with different properties (small, large, . . .) and different relationships between them (inside, . . .)
- Each picture can be described by an interpretation

Running Example: Bongard Problems



$$I_{leftpict} = \{triangle(0), large(0), square(1), small(1), inside(1,0), triangle(2), inside(2,1)\}$$

With the background knowledge B:

$$in(A,B) \leftarrow inside(A,B).$$
$$in(A,D) \leftarrow inside(A,C), in(C,D).$$

$M(B \cup I_{leftpict})$ contains $in(1,0)$, $in(2,1)$ and $in(2,0)$.
Given the IC $C_1 = triangle(T), square(S), in(T,S) \rightarrow false$
C_1 is false in $I_{leftpict}$, true in $I_{centrpict}$ and false in $I_{rightpict}$

Syntax

A Probabilistic Constraint Logic Theory (PCLT) T is a set of probabilistic integrity constraints (PICs) C of the form

$$p_i :: L_1, \ldots, L_b \rightarrow A_1; \ldots; A_h \tag{2}$$

where

- $L_1, \ldots, L_b \rightarrow A_1; \ldots; A_h$ is an IC
- p_i is a real value in $[0, 1]$ which defines its probability

We may also have a background knowledge B

Semantics

- A PCLT T defines a probability distribution on ground constraint logic theories called worlds
  - for each grounding of each IC, we decide to include or not the grounding in a world with probability p_i
  - we assume all groundings to be independent
  - similar to the notion of world in ProbLog where a world is a normal logic program.
- The probability of a world w is given by the product:

$$P(w) = \prod_{i=1}^{m} \prod_{C_{ij} \in w} p_i \prod_{C_{ij} \notin w} (1 - p_i)$$

where m is the number of PICs.

- Given an interpretation I, a background knowledge B and a world w, the probability $P(\oplus|w, I)$ of the positive class is
  - $P(\oplus|w, I) = 1$ if $M(B \cup I) \models w$
  - 0 otherwise.

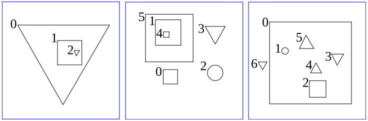
- The probability $P(\oplus|I)$ of the positive class is the probability of I satisfying a PCLT T given B. From now on we always assume B as given and we do not mention it again.

$$P(\oplus|I) = \sum_{w \in W} P(\oplus, w|I) = \sum_{w \in W} P(\oplus|w, I)P(w|I) =$$

$$\sum_{w \in W, M(B \cup I) \models w} P(w)$$

- The probability $P(\ominus|I)$ of the negative class given an interpretation I is the probability of I not satisfying T and is given by $1 - P(\oplus|I)$.

Running Example: Bongard Problems



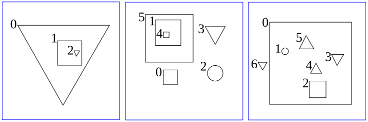
$$I_{leftpict} = \{triangle(0), large(0), square(1), small(1), inside(1,0), triangle(2), inside(2,1)\}$$

With the background knowledge B:

$$\begin{aligned} in(A,B) &\leftarrow inside(A,B). \\ in(A,D) &\leftarrow inside(A,C), in(C,D). \end{aligned}$$

$M(B \cup I_{leftpict})$ contains $in(1,0)$, $in(2,1)$ and $in(2,0)$.
Given the IC $C_1 = 0.5 :: triangle(T), square(S), in(T,S) \rightarrow false$
There are two different instantiations for the IC C_1 $\rightarrow$ four possible worlds

Running Example: Bongard Problems



Four possible worlds $\{\emptyset, \{C_{11}\}, \{C_{12}\}, \{C_{11}, C_{12}\}\}$

- for the first two of them $M(B \cup I_l) \models w_i$
- $P(\oplus | I_{leftpict}) = P(w_1) + P(w_2) = 0.25 + 0.25 = 0.5$

In the central picture there are four different instantiations for $C_1 \rightarrow$ 16 worlds

- $I_{centrpict}$ is verified in all of them (constraint is never violated)
- $P(\oplus | I_{centrpict}) = 1$.

The right picture has 8 different instantiations for IC $C_1 \rightarrow$ 256 worlds

- $I_{rightpict}$ is verified in only 32 of them
- $P(\oplus | I_{rightpict}) = 0.125$.

A Problem that Must Be Solved

Computing $P(\oplus|I)$ as seen before is impractical

The number of worlds is exponential in the number of instantiations of the ICs

A possible solution:

- we can associate a Boolean random variable X_{ij} to each instantiated constraint C_{ij}
  - if C_{ij} is included in the world X_{ij} takes on value 1
  - $P(X_{ij}) = P(C_{ij}) = p_i$
  - $P(\overline{X_{ij}}) = 1 - P(C_{ij}) = 1 - p_i$

- A valuation ν is an assignment of a truth value to all variables in $\mathbf{X}$.
    - One to one correspondence between worlds and valuations
    - ν can be represented as a set containing X_{ij} (C_{ij} is included in the world) or $\overline{X_{ij}}$ (C_{ij} is not included in the world) for each X_{ij}
    - ν corresponds with $\phi_\nu = \bigwedge_{i=1}^{m} \bigwedge_{X_{ij}\in\nu} X_{ij} \bigwedge_{\overline{X_{ij}}\in\nu} \overline{X_{ij}}$

$$P(\phi_\nu) = \prod_{i=1}^{m} \prod_{C_{ij}\in w} p_i \prod_{C_{ij}\notin w} (1 - p_i) = P(w)$$

Suppose a ground IC C_{ij} is violated in I

- The worlds where X_{ij} holds in the respective valuation are excluded from the summation of previous slide
- We must keep only the worlds where $\overline{X_{ij}}$ holds in the respective valuation for all ground constraints C_{ij} violated in I.

I satisfies all the worlds where the formula

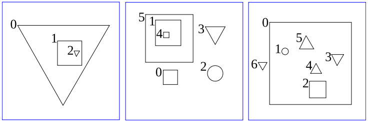
$$\phi = \bigwedge_{i=1}^{m} \bigwedge_{M(B \cup I) \not\models C_{ij}} \overline{X_{ij}}$$

is true in the respective valuations

$$P(\oplus|I) = P(\phi) = \prod_{i=1}^{m} (1 - p_i)^{n_i}$$

where n_i is the number of instantiations of C_i that are not satisfied in I.

Running Example: Bongard Problems



$C_1 = 0.5 :: triangle(T), square(S), in(T, S) \rightarrow false$

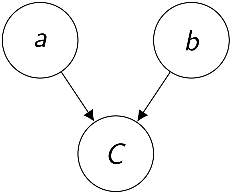
- In the left picture the body of C_1 is true for the single substitution $T/2$ and $S/1$ thus $n_1 = 1$ and $P(\oplus|I_{leftpict}) = 0.5$.
- In the central picture the body of C_1 is always false, thus $n_1 = 0$ and $P(\oplus|I_{centrpict}) = 1$.
- In the right picture the body of C_1 is true for three couples (triangle, square) thus $n_1 = 3$ and $P(\oplus|I_{rightpict}) = 0.125$.

Independence Assumption: an Example

PCLT can model any conditional probabilistic relationship between the class variable and the ground atoms.

Suppose you want to model a general conditional dependence between the class atom and a Herbrand base containing two atoms: a and b.

This dependence can be represented as



$P'(C \mid a, b)$		C	
a	b	$-$	$+$
0	0	$1-p_1$	p_1
0	1	$1-p_2$	p_2
1	0	$1-p_3$	p_3
1	1	$1-p_4$	p_4

where the conditional probability table has four parameters, $p_1, \ldots, p_4$, so is the most general.

Independence Assumption: an Example

This model can be represented with the following PCLT

$$\begin{aligned} C_1 = 1 - p_1 &\quad :: \quad \neg a, \neg b \rightarrow \mathit{false} \\ C_2 = 1 - p_2 &\quad :: \quad \neg a, b \rightarrow \mathit{false} \\ C_3 = 1 - p_3 &\quad :: \quad a, \neg b \rightarrow \mathit{false} \\ C_4 = 1 - p_4 &\quad :: \quad a, b \rightarrow \mathit{false} \end{aligned}$$

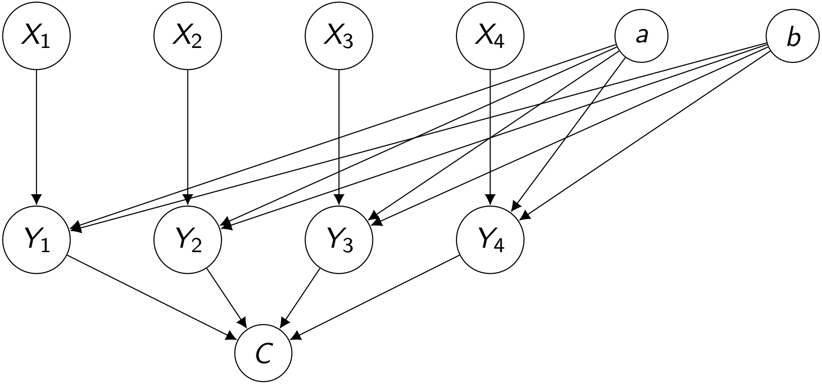
For example, the probability that the class variable assumes value $+$ given that a and b are false is

$$P(C = + | \neg a, \neg b) = 1 - (1 - p_1) = p_1$$

given interpretation $\{\}$ (only constraint C_1 is violated)

Independence Assumption: an Example

The Bayesian network above is equivalent to



- Boolean variable X_i represents whether constraint C_i is included in the world
- Boolean variable Y_i whether constraint C_i is violated

Independence Assumption: an Example

- The conditional probability tables for nodes X_is are

$$P''(X_i = 1) = 1 - p_i$$

- those for nodes Y_is encode the deterministic functions

$$\begin{aligned} Y_1 &= X_1 \wedge \neg a \wedge \neg b \\ Y_2 &= X_2 \wedge \neg a \wedge b \\ Y_3 &= X_3 \wedge a \wedge \neg b \\ Y_4 &= X_4 \wedge a \wedge b \end{aligned}$$

- that for C encodes the deterministic function

$$C = \neg Y_1 \wedge \neg Y_2 \wedge \neg Y_3 \wedge \neg Y_4$$

where C is interpreted as a Boolean variable with 1 corresponding to $+$ and 0 to -

Independence Assumption: an Example

It is possible to show that the probability distribution of this BN coincides with P for all the possible interpretations.

$\mathbf{X}$ variables are mutually unconditionally independent, showing that it is possible to represent any conditional dependence of C from the Herbrand base by using only independent random variables.

PCLT and Markov Logic Networks

- Similarly to MLNs, PCLTs encode constraints on the possible interpretations and the probability of an interpretation depends on the number of violated constraints
- MLNs encode the joint distribution of the ground atoms and the class, differently we concentrate on the conditional distribution of the class given the ground atoms
- Given a PCLT, it is possible to obtain an equivalent MLN with an equivalent probability distribution

Conclusions and Future Work

- Conclusions
  - We have proposed a probabilistic extension of constraint logic theories.
  - Under this extension the computation of the probability of an interpretation being positive is logarithmic in the number of falsified constraints.
- Future Work
  - The development of a system for learning such probabilistic integrity constraint
    - We will exploit Limited-memory BFGS for tuning the parameters and constraint refinements for finding good structures

THANKS FOR LISTENING

AND

ANY QUESTIONS ?

References I

De Raedt, L., Kimmig, A., and Toivonen, H. (2007).
ProbLog: A probabilistic Prolog and its application in link discovery.
volume 7, pages 2462–2467, Palo Alto, California USA.

Riguzzi, F. and Swift, T. (2011).
The PITA system: Tabling and answer subsumption for reasoning under uncertainty.
11(4–5):433–449.

Riguzzi, F. and Swift, T. (2013).
Well-definedness and efficient inference for probabilistic logic programming under the distribution semantics.
13(Special Issue 02 - 25th Annual GULP Conference):279–302.