ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Probabilistic Logic Programming

Federico Chesani

DISI

Department of Informatics – Science and Engineering

Notice

These slides are largely an adaptation of existing material, including slides by Fabrizio Riguzzi on Data Mining and Analytics. I am especially grateful to Fabrizio for letting me access and use his own material.

Sharing: a copy of these slides can be downloaded from the servers of the University of Bologna and stored for personal use only. Please do not redistribute.

Brief recap and motivations

Prolog is a rule-based programming language that allows to represent and to reason upon some knowledge

- Fully-fledged programming language
- It has a logic-based semantics
- Backward reasoning
- It supports modelling paradigms such as Constraint Logic Programming
- Through meta-interpreters, the interpretation of the knowledge can be easily customized, e.g.:
  - Supporting the "confidence" of a rule, like it was in Mycin
  - Supporting the generation of explanations
  - Supporting the extension of the KB (lemma generation)

Brief recap and motivations

Prolog is a rule-based programming language that allows to represent and to reason upon some knowledge…

What about Uncertainty and Probabilistic Reasoning?

Why should we care?

Rule-based representation of knowledge is a powerful method. However, Prolog is based on "crisp", two-value logic…

We might want to enjoy the best of the two worlds of rules and uncertainty/probabilistic

Before starting... resources!

Two major systems are available, with equivalent semantics

Logic Programs with Annotated Disjunctions - LPAD

- http://cplint.eu/
  - Inference
  - Parameter Learning (EMBLEM algorithm)
  - Structure Learning (SLIPCOVER algorithm)
- Available as a library for SWI Prolog

ProbLog

- https://dtai.cs.kuleuven.be/problog/
  - Inference
  - Parameter Learning (LFI-ProbLog algorithm)

Probabilistic Logic Programming

Probabilistic Logic Programming

The approach can be traced up to a seminal work by Sato at the ICLP 1995 conference about the **Distribution Semantics**

A PLP defines a probability distribution over normal logic programs, called instances or possible worlds or simply **worlds**.

The distribution then is extended to a joint distribution over worlds and interpretations (or queries)

The probability of a query is then obtained from this distribution

Taisuke Sato, A Statistical Learning Method for Logic Programs with Distribution Semantics, In Proceedings Of the 12th International Conference On Logic Programming (ICLP'95), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.17.4408

Several PLP languages…

Many languages have been proposed over the Distribution Semantics…

… they differ in the way they define the distribution over logic programs

- Probabilistic Logic Programs [Dantsin RCLP91]
- Probabilistic Horn Abduction [Poole NGC93], Independent Choice Logic (ICL) [Poole AI97]
- PRISM [Sato ICLP95]
- **Logic Programs with Annotated Disjunctions (LPADs)** [Vennekens et al. ICLP04]
- ProbLog [De Raedt et al. IJCAI07]

LPAD

In LPAD, the head of a clause is extended with disjunctions, and each disjunct is annotated with a probability

Example: we sampled that

- in "people with flu", they also "sneeze in 70% cases";
- in "people with hay fever", they also "sneeze in 80% cases".

```
sneezing(X):0.7 ; null:0.3 :- flu(X).
sneezing(X):0.8 ; null:0.2 :- hay_fever(X).
```

Syntactic restriction: **`null`** does not appear in the body of any rule.

LPAD

What if we know something else?

Example:

- in "people with flu", they also "sneeze in 70% cases";
- in "people with hay fever", they also "sneeze in 80% cases".
- Bob has a flu
- Bob suffers of hay fever

```
sneezing(X):0.7 ; null:0.3 :- flu(X).
sneezing(X):0.8 ; null:0.2 :- hay_fever(X).
flu(bob).
hay_fever(bob).
```

Query: will Bob sneeze?

LPAD – probability distribution over rule heads

```
sneezing(X):0.7 ; null:0.3 :- flu(X).
sneezing(X):0.8 ; null:0.2 :- hay_fever(X).
flu(bob).
hay_fever(bob).
```

Each rule has a probability distribution over its head !

In ProbLog, probability distributions are over facts.

LPAD – Worlds

```
sneezing(X):0.7 ; null:0.3 :- flu(X).
sneezing(X):0.8 ; null:0.2 :- hay_fever(X).
flu(bob).
hay_fever(bob).
```

Worlds will be obtained by selecting one atom from the head of every grounding of each clause:

1. ground the program
2. For each atom in each head, choose to include it or not

LPAD – Worlds

Grounding

```
sneezing(X):0.7 ; null:0.3 :- flu(X).
sneezing(X):0.8 ; null:0.2 :- hay_fever(X).
flu(bob).
hay_fever(bob).
```

```
sneezing(bob):0.7 ; null:0.3 :- flu(bob).
sneezing(bob):0.8 ; null:0.2 :-
                         hay_fever(bob).
flu(bob).
hay_fever(bob).
```

LPAD – Worlds

Generating the worlds:

```
sneezing(bob):0.7 ; null:0.3 :- flu(bob).
sneezing(bob):0.8 ; null:0.2 :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

W_1

```
sneezing(bob) :- flu(bob).
sneezing(bob) :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

W_2

```
null :- flu(bob).
sneezing(bob) :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

W_3

```
sneezing(bob) :- flu(bob).
null :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

W_4

```
null :- flu(bob).
null :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

LPAD – Distribution Semantics over worlds

Given a clause C and a substitution $\boldsymbol{\theta}$ such that C$\boldsymbol{\theta}$ is ground, it is defined:

- Atomic choice: selection of the i-th atom of the head of C for grounding C$\boldsymbol{\theta}$: (C, $\boldsymbol{\theta}$, i)
- Composite choice $\boldsymbol{\kappa}$: consistent set of atomic choices
- Probability of a composite choice $\boldsymbol{\kappa}$:

$$P(\boldsymbol{\kappa}) = \prod_{(C,\boldsymbol{\theta},i)\in\boldsymbol{\kappa}} P(C,i)$$

LPAD – Distribution Semantics over worlds

- Selection $\boldsymbol{\sigma}$: a total composite choice (one atomic choice for every grounding of each clause)
- A selection $\boldsymbol{\sigma}$ identifies a logic program $w_{\boldsymbol{\sigma}}$ called world
- Probability of a world is then defined as

$$P(w_\sigma) = P(\sigma) = \prod_{(C,\boldsymbol{\theta},i) \in \sigma} P(C, i)$$

LPAD – Probability of worlds

```
sneezing(bob):0.7 ; null:0.3 :- flu(bob).
sneezing(bob):0.8 ; null:0.2 :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

$P(w_1) = 0.7x0.8$

```
sneezing(bob) :- flu(bob).
sneezing(bob) :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

$P(w_2) = 0.3x0.8$

```
null :- flu(bob).
sneezing(bob) :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

```
sneezing(bob) :- flu(bob).
null :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

$P(w_3) = 0.7x0.2$

```
null :- flu(bob).
null :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

$P(w_4) = 0.3x0.2$

LPAD – Distribution Semantics: probability of a query

Given a ground query Q and a world w:

$$P(Q|w) = \begin{cases} 1 & if\ Q\ is\ true\ in\ w \\ 0 & otherwise \end{cases}$$

$$P(Q) = \sum_{w} P(Q,w) = \sum_{w} P(Q|w)P(w) = \sum_{w \models Q} P(w)$$

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

LPAD – Probability of worlds

```
sneezing(bob):0.7 ; null:0.3 :- flu(bob).
sneezing(bob):0.8 ; null:0.2 :- hay_fever(bob).
flu(bob).
hay_fever(bob).
:- sneezing(bob).
```

```
sneezing(bob) :- flu(bob).
sneezing(bob) :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

$P(w_1) = 0.7x0.8$

```
null :- flu(bob).
sneezing(bob) :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

$P(w_2) = 0.3x0.8$

```
sneezing(bob) :- flu(bob).
null :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

$P(w_3) = 0.7x0.2$

```
null :- flu(bob).
null :- hay_fever(bob).
flu(bob).
hay_fever(bob).
```

$P(w_4) = 0.3x0.2$

- sneezing(bob) is true in w_1, w_2, w_3
- **P(sneezing(bob))** = 0.7x0.8 + 0.3x0.8 + 0.7x0.2 = **0.94**

LPAD – Example: Monty Hall Puzzle

- A player is given the opportunity to select one of three closed doors, behind one of which there is a prize.
- Behind the other two doors are empty rooms.
- Once the player has made a selection, Monty opens one of the remaining closed doors which does not contain the prize, showing that the room behind it is empty.
- He then asks the player if he would like to switch his selection to the other unopened door, or stay with his original choice.
- Does it matter if he switches?

LPAD – Example: Monty Hall Puzzle

http://cplint.eu/e/monty.swinb

```
% the prize is behind each door with probability 1/3
prize(1):1/3; prize(2):1/3; prize(3):1/3.

% Monty opens door 2 with probability 0.5 and door 3 with probability 0.5 if
the prize is behind door 1.
open_door(2):0.5 ; open_door(3):0.5 :- prize(1).

% Monty opens door 2 if the prize is behind door 3.
open_door(2):- prize(3).
% Monty opens door 3 if the prize is behind door 2.
open_door(3):- prize(2).

% the player keeps his choice and wins if he has selected a door with the
prize
win_keep:- prize(1).

% the player switches and wins if the prize is behind the door that he has
% not selected and that Monty did not open
win_switch:- prize(2), open_door(3).
win_switch:- prize(3), open_door(2).
```

Case study: Assessing the fall risk of a subject

Introduction to the domain

- Subjects >65 years old experiment (unwanted) falls
- A fall may have huge impacts
  - Immediately, as a consequence of trauma
  - In general, since its consequences affect the quality of life of the subject, both at a physical and at a psychological level
- It is estimated that every year 30% of community dwelling adults aged >65 experiment a fall

- A number of studies (whose results are available as scientific publications) investigated the statistical relation between fall events and the exposure of the subject to risk factors (for falls)

Introduction to the domain

In the paper:

Deandrea S, Lucenteforte E, Bravi F, Foschi R, La Vecchia C, Negri E. Risk factors for falls in community-dwelling older people: a systematic review and meta-analysis. Epidemiology. 2010 Sep;21(5):658-68. doi: 10.1097/EDE.0b013e3181e89905. PMID: 20585256.

the authors perform a meta-analysis of reviews of papers about risk factors (w.r.t. fall events)

- "Single study" papers investigate the relation between a specific risk factor and falls, w.r.t. to a patient cohort
- Review papers "aggregate" the results of many single-study papers, summarizing the relation between a specific risk factor and falls
- Meta-analysis put together the findings about all the risk factors

Introduction to the domain

TABLE 3. Association of Sociodemographic Risk Factors With Falls in Community-dwelling Older People

Characteristic	All Studies No. Studies	All Studies Heterogeneity *P*	All Studies OR (95% CI)	Multivariate Analysis Only No. Studies	Multivariate Analysis Only Heterogeneity *P*	Multivariate Analysis Only OR (95% CI)	High Frequency of Fall Assessment Only No. Studies	High Frequency of Fall Assessment Only Heterogeneity *P*	High Frequency of Fall Assessment Only OR (95% CI)
Age (5-year increase)									
All fallers	18	<0.0001	1.12 (1.07–1.17)	8	0.0002	1.12 (1.05–1.19)	8	0.007	1.11 (1.05–1.17)
Recurrent fallers	15	0.009	1.12 (1.07–1.18)	6	0.0007	1.15 (1.00–1.32)	9	0.10	1.12 (1.07–1.18)
Sex (women vs. men)									
All fallers	22	0.004	1.30 (1.18–1.42)	7	0.003	1.28 (1.06–1.54)	12	0.22	1.37 (1.21–1.55)
Recurrent fallers	18	<0.0001	1.34 (1.12–1.60)	6	0.0002	1.68 (0.97–2.89)	12	0.0002	1.34 (1.08–1.68)
Living situation (alone vs. not alone)									
All fallers	11	0.44	1.33 (1.21–1.45)	1	NA	1.20 (0.69–2.08)	3	0.52	1.26 (1.04–1.53)
Recurrent fallers	9	0.43	1.25 (1.10–1.43)	1	NA	1.59 (1.00–2.52)	4	0.79	1.16 (0.98–1.38)
History of falls (yes vs. no)									
All fallers	18	<0.0001	2.77 (2.37–3.25)	12	0.002	2.92 (2.50–3.40)	9	0.35	2.79 (2.43–3.20)
Recurrent fallers	12	0.04	3.46 (2.85–4.22)	7	0.04	3.07 (2.31–4.08)	9	0.54	3.09 (2.63–3.63)
Physical activity (limitation vs. no limitation)									
All fallers	10	0.01	1.20 (1.04–1.38)	1	NA	0.70 (0.40–1.21)	7	0.008	1.22 (1.00–1.50)
Recurrent fallers									
Physical disability (yes vs. no)									
All fallers	9	<0.0001	1.56 (1.22–1.99)	4	0.0001	1.46 (0.85–2.52)	4	0.20	2.30 (1.55–3.43)
Recurrent fallers	8	<0.0001	2.42 (1.80–3.26)	2	0.02	2.63 (1.06–6.51)	6	0.22	2.24 (1.81–2.77)
Instrumental disability (yes vs. no)									
All fallers	6	0.03	1.46 (1.20–1.77)	2	0.70	1.25 (1.02–1.53)	1	NA	1.40 (0.92–2.14)
Recurrent fallers	4	0.002	2.04 (1.41–2.95)	0			1	NA	2.00 (1.35–2.96)
Body mass index (low vs. intermediate/high)									
All fallers	3	0.64	1.17 (0.93–1.46)	1	NA	1.04 (0.73–1.48)	1	NA	1.20 (0.85–1.70)
Recurrent fallers	6	0.55	1.03 (0.86–1.23)	1	NA	0.88 (0.59–1.31)	4	0.73	1.03 (0.84–1.27)
Education (low vs. intermediate/high)									
All fallers	7	0.01	1.01 (0.88–1.16)	1	NA	0.93 (0.76–1.13)	3	0.03	0.91 (0.62–1.32)
Recurrent fallers	8	0.001	0.81 (0.62–1.05)	2	0.66	0.87 (0.71–1.08)	4	0.47	0.71 (0.59–0.86)
Walking aid use (yes vs. no)									
All fallers	11	0.006	2.18 (1.79–2.65)	3	0.80	2.50 (1.80–3.47)	6	0.12	2.46 (1.91–3.15)
Recurrent fallers	6	0.009	3.09 (2.10–4.53)	1	NA	3.20 (1.70–6.01)	4	0.01	3.05 (1.87–4.95)

NA indicates not applicable.

Deandrea S, Lucenteforte E, Bravi F, Foschi R, La Vecchia C, Negri E. Risk factors for falls in community-dwelling older people: a systematic review and meta-analysis. Epidemiology. 2010 Sep;21(5):658-68. doi: 10.1097/EDE.0b013e3181e89905. PMID: 20585256.

Introduction to the domain

- 30 Risk factors are identified
  - Sociodemographic factors
  - Medical and psychological risk factors
  - Medication-related risk factors
  - Association of mobility and sensory risk factors

- Results are given in terms of Odds Ratios for each risk factor (plus confidence interval CI)

TABLE 6. Association of Mobility and Sensory Risk Factors With Falls in Community-dwelling Older People

	All Studies			Multivariate Analysis Only			High Frequency of Fall Assessment Only		
Characteristic	**No. Studies**	**Heterogeneity *P***	**OR (95% CI)**	**No. Studies**	**Heterogeneity *P***	**OR (95% CI)**	**No. Studies**	**Heterogeneity *P***	**OR (95% CI)**
Gait problems (yes vs. no)									
All fallers	5	0.54	2.06 (1.82–2.33)	3	0.32	2.06 (1.76–2.41)	2	0.17	2.02 (1.39–2.93)
Recurrent fallers	6	<0.0001	2.16 (1.47–3.19)	2	0.11	3.68 (1.87–7.22)	4	0.04	2.58 (1.79–3.74)
Vision impairment (yes vs. no)									
All fallers	15	<0.0001	1.35 (1.18–1.54)	6	0.01	1.21 (0.92–1.58)	7	0.61	1.51 (1.29–1.78)
Recurrent fallers	13	<0.0001	1.60 (1.28–2.00)	4	<0.0001	1.45 (0.83–2.53)	8	0.50	1.81 (1.58–2.08)
Hearing impairment (yes vs. no)									
All fallers	7	0.13	1.21 (1.05–1.39)	0			4	0.15	1.25 (1.03–1.51)
Recurrent fallers	8	0.36	1.53 (1.33–1.76)	0			5	0.28	1.50 (1.27–1.78)

NA indicates not applicable.

Deandrea S, Lucenteforte E, Bravi F, Foschi R, La Vecchia C, Negri E. Risk factors for falls in community-dwelling older people: a systematic review and meta-analysis. Epidemiology. 2010 Sep;21(5):658-68. doi: 10.1097/EDE.0b013e3181e89905. PMID: 20585256.

Introduction to the domain

Under some assumptions, and with some maths, odds ratios can be written as conditional probabilities:

P(fall | "gait problems") = 0.07726945577352498

P(fall | "diabetes") = 0.014788031645008427

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Introduction to the domain

Question:

given that (we know that) a subject is exposed to certain risk factors

can we compute the risk of falling (within a year) ?

Some practical issues:

1. Sometimes we might not know if a subject is exposed or not to a risk factor
2. How to estimate the exposure to a risk factor?

Idea: use LPAD to compute probability of risk

P(fall | "gait problems") = 0.07726945577352498

P(fall | "diabetes") = 0.014788031645008427

```
fall : 0.079.
fall : 0.07726945577352498    :-   gait.
fall : 0.014788031645008427   :-   diabetes.
```

Better formulation: given S a list of risk factors

```
fall(S) : 0.079.
fall(S) : 0.07726945577352498    :-   member(gait, S).
fall(S) : 0.014788031645008427   :-   member(diabetes, S).
```

Domain issue 1: dealing with the unknown

```
fall(S) : 0.079.
fall(S) : 0.07726945577352498    :-  member(gait, S).
fall(S) : 0.014788031645008427   :-  member(diabetes, S).
```

Suppose the profile of our subject is S = [gait]. What does it mean exactly?

a) The subject suffers of gait problems; and
b) The subject does not suffer of diabetes

Prolog (and LPADs) adopts a CWA-based approach.

However, there are cases where we might not know if a subject, for example, suffers of diabetes or not… the physician might have some suspects, but no certain knowledge.

Domain issue 1: dealing with the unknown

Enrich the profile of our subject with more information:

```
S = [ (gait,t), (diabetes,f), (parkinson,u)].
```

Intended meaning:

- The subject suffers of gait problems
- The subject does not have a diabetes
- It is not known if the subject has parkinson

```
fall(S) : 0.014788031645008427   :-  check((diabetes,t), S).
check( (diabetes,t), S) :- member((diabetes,t), S).
check( (diabetes,f), S) :- member((diabetes,f), S).
```

Domain issue 1: dealing with the unknown by means of prevalence

- It is not known if the subject has diabetes

We might discover by the literature (or in statistical records) that around 10% of the population (aged 65) suffers indeed of diabetes. This is usually named as the prevalence of diabetes over the referenced population.

```
fall(S) : 0.014788031645008427   :-  check((diabetes,t), S).
check( (diabetes,t), S) :- member((diabetes,t), S).
check( (diabetes,f), S) :- member((diabetes,f), S).
check( (diabetes,t), S):0.109 ; check((diabetes,f), S):0.891 :-
        member((diabetes,u), S).
```

Domain issue 2: exposure to a risk factor

- Exposure to certain risk factors can be assessed immediately
  - E.g. either the subject has a diagnosis of diabetes, or not
- Some risk factors are defined vaguely, since physicians assume a certain background knowledge
  - E.g. "vision impairment" is a risk factor, but different physicians asses it in different ways:
  - visual stereognosis score <= 3, or
  - visual acuity 3 mt score <=5 (Monoyer scale), or
  - contrast sensitivity score <=16
- The criteria above are a typical example of expert knowledge, that should be mixed up with probabilistic reasoning

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Domain issue 2: exposure to a risk factor

```
fall(S) : 0.026906047778913072 :- check( ('vision impairment', t), S).

check( ('vision impairment', t), S) :- factor3( ('vision impairment', t), S).
check( ('vision impairment', f), S) :- factor3( ('vision impairment', f), S).
check( ('vision impairment', t), S) : 0.19 ;
        check( ('vision impairment', f), S) : 0.81 :-
                factor3( ('vision impairment', u), S).
```

Domain issue 2: exposure to a risk factor

Physicians assess visual impairments in different ways:

- visual stereognosis score <= 3, or
- visual acuity 3 mt score <=5 (Monoyer scale), or
- contrast sensitivity score <=16

```
factor3( ('vision impairment', t), S) :-
        member( ('visual stereognosis', N), S), number(N), N =< 3.
factor3( ('vision impairment', t), S) :-
        member( ('visual acuity 3 m', N), S), number(N), N =< 5.
factor3( ('vision impairment', t), S) :-
        member( ('contrast sensitivity', N), S), number(N), N =< 16.
```

Is it enough? Is it correct?

What if the subject has done all the three tests?

Domain issue 2: exposure to a risk factor

Different physicians assess visual impairments in different ways

```
factor3( ('vision impairment', t), S) :-
        member( ('visual stereognosis', N), S), number(N), N =< 3.
factor3( ('vision impairment', t), S) :-
        member( ('visual acuity 3 m', N), S), number(N), N =< 5.
factor3( ('vision impairment', t), S) :-
        member( ('contrast sensitivity', N), S), number(N), N =< 16.

factor3( ('vision impairment', u), S) :-
        member( ('visual stereognosis', u), S), !,
        \+ (factor3( ('vision impairment', t), S).
factor3( ('vision impairment', u), S) :-
        member( ('visual acuity 3 m', u), S), !,
        \+ (factor3( ('vision impairment', t), S).
factor3( ('vision impairment', u), S) :-
        member( ('contrast sensitivity', u), S), !,
        \+ (factor3(('vision impairment', t), S).
factor3( ('vision impairment', f), S) :-
        \+ (factor3('vision impairment', t), S) ),
        \+ (factor3( ('vision impairment', u), S) ).
```

Concluding remarks

- PLP is a powerful paradigm and reasoning tool for mixing up two different types of knowledge:
  - Probabilistic knowledge, e.g. statistically learned
  - Classical, logical knowledge, e.g. physician background knowledge

- Would it be possible to learn the probabilities? **Yes**
- Would it be possible to learn the structure of the rules? **Yes**