Probabilistic Inductive Logic Programming

Joana Côrte-Real
jcr@dcc.fc.up.pt

CRACS & INESC TEC
Faculty of Sciences
University of Porto

2nd Breast Cancer Workshop
7th April 2015

Overview

Section 1

Introduction

Inductive vs Deductive Reasoning

Deductive Reasoning derives new rules or facts from a pre-defined set of rules (and other background knowledge).

Inductive Reasoning can learn a rule from examples and a set of facts which describe the example (or background knowledge)

Rock-paper-scissors

Rule

```
beats(Round,PlayerA,PlayerB) :-
    plays(Round,PlayerA,rock),
    plays(Round,PlayerB,scissors).
```

Facts (BK)

```
plays(1,ines,rock).
plays(1,joana,scissors).
```

Example

```
beats(1,ines,joana).
```

Deduction in Rock-paper-scissors

Rule	`beats(Round,PlayerA,PlayerB) :-`
	`  plays(Round,PlayerA,rock),`
	`  plays(Round,PlayerB,scissors).`

+	
Facts (BK)	`plays(1,ines,rock).` `plays(1,joana,scissors).`
=	
Example	`beats(1,ines,joana).`

Induction in Rock-paper-scissors

Rule

```
beats(Round,PlayerA,PlayerB) :-
    plays(Round,PlayerA,rock),
    plays(Round,PlayerB,scissors).
```

=

Facts (BK)

```
plays(1,ines,rock).
plays(1,joana,scissors).
```

+

Example

```
beats(1,ines,joana).
```

Main characteristics of ILP

Inductive **L**ogic **P**rogramming

- induces rules which explain examples and BK
- based on Logic Programming (Prolog)
- machine learning technique
- can be used for prediction and/or description
- also used to interface with experts of other areas of knowledge

Section 2

Probabilistic ILP

What is it?

Probabilistic Inductive Logic Programming (PILP)

- was introduced by De Raedt and Kersting in 2004.
- is an extension of ILP which can deal with uncertainty.
- BK and examples can be annotated with probabilities.

Motivation

better models many fields of knowledge produce data with an inherent uncertainty

dataset size using probabilities to describe data can greatly reduce the dataset size and useful information can still be extracted

unknown values in cases where the full data is not known, information can still be used efficiently in the computation of a rule, for instance, by adding values from the literature

privacy compressing data this way can be used to protect private sensitive data

Semantics of Probability

Probabilities can be interpreted as

statistics statistical data such as frequencies, marginal distributions, ...

confidence or the degree of belief in a statement. This can be used to interface with human experts.

adaptation from a different type of data such as a numerical attribute.

PILP in Rock-paper-scissors

R

```
beats(Round,PlayerA,PlayerB) :-
        plays(Round,PlayerA,rock),
        plays(Round,PlayerB,scissors).
```

=

BK

```
plays(1,ines,rock).
plays(1,joana,scissors).
```

```
0.7::plays(ines,rock);
0.3::plays(ines,scissors).
0.5::plays(joana,rock);
0.5::plays(joana,scissors).
```

+

E

```
beats(1,ines,joana).
```

```
0.35::beats(ines,joana).
```

SkILL - a stochastic inductive logic learner

SkILL is a (our) PILP engine. It

- takes as input probabilistic BK and examples and can generate several rules.
- addresses efficiency issues by taking advantage of a probabilistic search space.
- has been used to analyze medical data, among other datasets.

SkILL - results for metabolism

SkILL was applied to a metabolism dataset adapted from ILP and these are the results:

	Search Strategies	
	(RMSE, σ)	(PAcc, σ)
SkILL	(0.616, 0.063)	(0.661, 0.045)
SkILL+pruning	(0.581, 0.099)	**(0.663, 0.045)**
Aleph	(0.656, 0.047)	

Section 3

PILP applied to Breast Cancer

ILP applied to breast cancer data

This rule shows that a breast cancer case is malignant IF:

1. A is classified as BI-RADS 5 and had a mass present in a patient who was between the ages of 65 and 70 and had two prior mammograms (B, C)
2. prior mammogram (B) had no mass shape described and no punctate calcifications
3. prior mammogram (C) was classified as BI-RADS 3

```
is_malignant(A):-
    'BIRADS_category'(A,b5),
    'MassPAO'(A,present),
    'Age'(A,age6570),
    previous_finding(A,B),
    'MassesShape'(B,none),
    'Calc_Punctate'(B,
        notPresent),
    previous_finding(A,C),
    'BIRADS_category'(C,b3).
```

PILP and non-definitive biopsies

Non-definitive biopsies:

- are biopsies whose result is not definitive - meaning that they are not obviously malignant or benign.
- represent 5% to 15% of all biopsies.
- only a small fraction of non-definite biopsies (10-20%) have in fact a malignant finding confirmed after the procedure.
- cause 35,000 to 105,000 patients to undergo excision in the US.

Preliminary work

1. explore medical data about non-definite biopsies
2. add medical literature values to the BK - representing statistical knowledge - such as the probability that a tumour is malignant given that its margin is spiculated in 90%.
3. use experts confidence in malignancy as a class variable so as to analyze their mental models.

Section 4

Conclusion

Summing up

Advantages of PILP:

- uses a language which is easy to interpret for experts from other areas of knowledge
- very concise classifiers
- great representative capacity: can represent relations and uncertainty

BUT:

- presents efficiency issues
- data may not be easy to adapt

Future work

1. gather more/better literature values for BK
2. use experts' confidence as a feature to build a classifier
3. consider other ways to annotate data with probabilities (such as the degree of malignancy)

Thank you

Joana Côrte-Real