CL2000 Tutorial
Applications of Inductive Logic Programming

Stephen Muggleton
Department of Computer Science
University of York

July, 2000

Overview

Part 1. Introduction to ILP

Part 2. Case studies

Part 3. Summary & new directions

Part 1

Introduction

Motivation

Large databases - genomics, natural language.

Science and industry knowledge-driven.

Procedural or declarative?

Declarative - logical representation.

Inductive logic programming (ILP).

Inductive logic programming

Background knowledge:	B
Examples:	$E = E^+ \wedge E^-$
Hypothesis:	H
Prior consistency:	$B \wedge E^- \not\models \Box$
Prior necessity:	$B \not\models E^+$
Posterior consistency:	$B \wedge H \wedge E^- \not\models \Box$
Posterior sufficiency:	$B \wedge H \models E^+$
	$B \wedge \overline{E^+} \models \overline{H}$

ILP variants

Categorised by deviation from "Standard" model which has:

1. Complete Background Knowledge (BK).
2. Correct BK.
3. Complete set of BK predicates.
4. Minimal set of BK predicates.
5. Noise-free data.
6. Pos+Neg examples.
7. Plentiful example set.
8. Best B,H,E are LPs.
9. Observational predicate learning.

Part 2

Case studies

EX definitions

eastbound(T) :- has_car(T,C), short(C), closed(C).

muta(M) :- atom(M,A,o,40,-0.389), bond(M,B,A,2), bond(M,C,B,1).

Case studies

Mutagenesis

British Telecom library

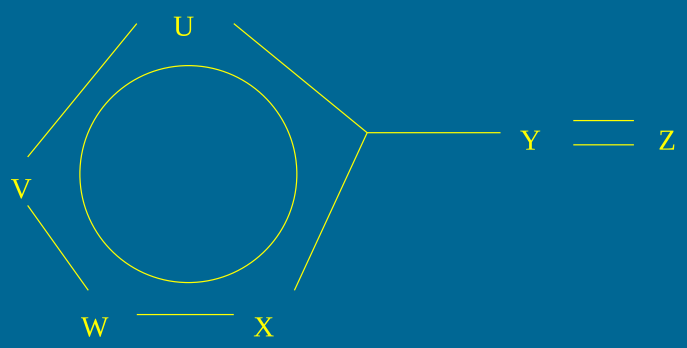
Metabolic Pathways

Mutagenesis

Related to carcinogenicity. Measured by Ames test.
Not always determinable empirically.

New structural alert (42 molecules):

Mutagenic if there exists a double bond
connected to a 5-aromatic ring.



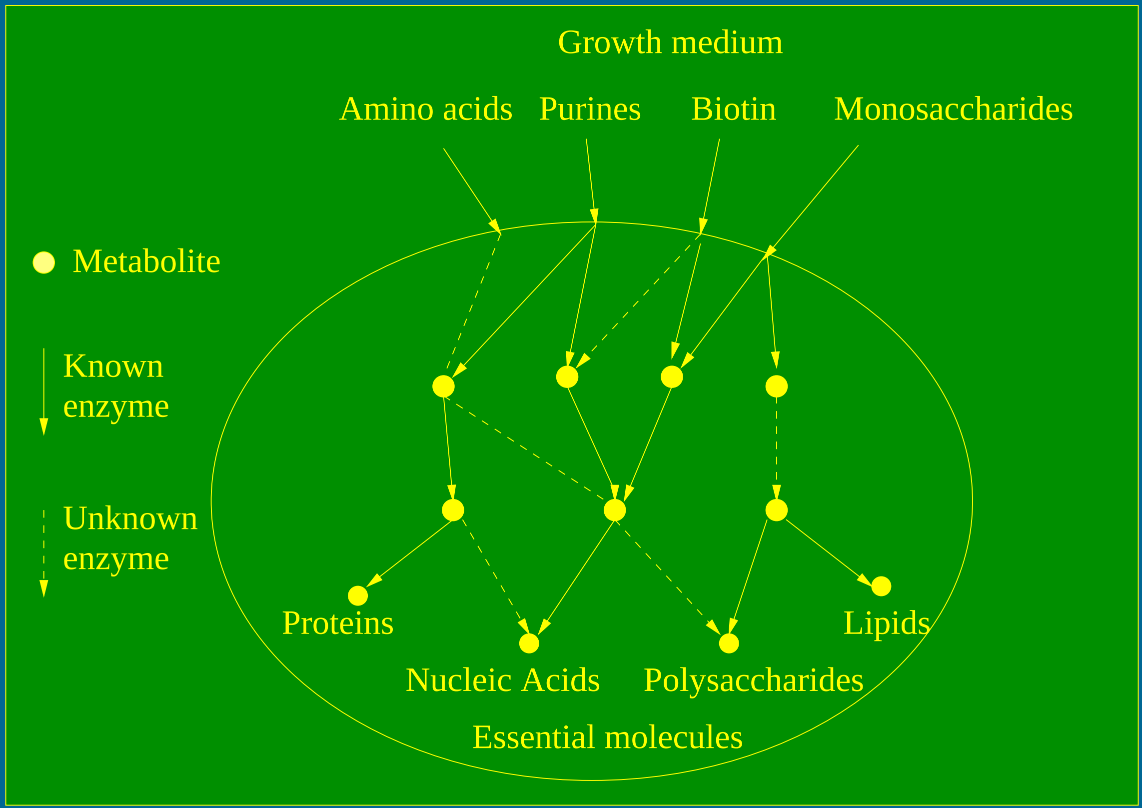
Progol	85.7%
BackProp	64.3%
Regression	66.7%
Default	69.0%

BT library system for Reverse Querying

- INSPEC library database.
- Simple keyword search too many hits.
- Evaluating interactive ILP-based ReQuire system.
- +ve and -ve examples of target.
- Relevant taxonomic background knowledge problem.

```
voice_over_ip(P) :-    has_code(P,C), telephony(C),
                       has_code(P,D), computer_communications(D).
```

Metabolic pathways



McCarthy's appearance/reality challenge

[5] < G >

[0] < C >

[4] < O >

[1] < A >

[3] < D >

[2] < T >

Background knowledge

```
observe(Pos,Ap,Act,Ap1) :-
    act(Act,Pos,Sit,Sit1), appears(Sit,Pos,Ap), appears(Sit1,Pos,Ap1).

appears(Sit,Pos,Ap) :- situation(Sit,Pos,Dial,Ap), letter(Ap),
    not(obscured(Sit,Dial,Pos)).

obscured(Sit,DialN,Pos) :-
    situation(Sit,Pos,DialM,Let), situation(Sit,Pos,DialN,_),
    DialM<DialN, letter(Let).
```

Examples

Example sequence - <0,Left>, <5,Right>

appears(0,0,c). appears(0,1,a). appears(0,2,t).

appears(0,3,d). appears(0,4,o). appears(0,5,g).

appears(1,0,a). appears(1,1,t).

appears(1,3,d). appears(1,4,o). appears(1,5,c).

observe(0,c,left,a). observe(5,c,right,g).

Integrity constraints

:- appears(S,P,L1), appears(S,P,L2), not(L1=L2).

:- situation(S,P,D,L1), situation(S,P,D,L2), not(L1=L2).

Hypothesis

situation(S,P,D,L) :- offset(S,D,O), offby(P,O,P1), dial(D,P1,L).

Part 3

Summary & new directions

ILP variants

Categorised by deviation from “Standard” model which has:

1. Complete Background Knowledge (BK).
2. Correct BK.
3. Complete set of BK predicates.
4. Minimal set of BK predicates.
5. Noise-free data.
6. Pos+Neg examples.
7. Plentiful example set.
8. Best B,H,E are LPs.
9. Observational predicate learning.

Summary

Name	Motivation	Deviation
Pharmacophores	EX	[2,6]
Mutagenesis	EX subset	[1,4,8]
PTE	EX + reuse	[1,7,8]
BT library	EX	[3,5,7]
Past Tense	EX	[1,4,6]
Folds	EX + novelty	[4,5,6,7]
Grammar learning	EX	[1,4,6,8]
Pathways	Non-observational	[1,2,3,4,5,9]
NeuroPeptides	EX	[1,2,3,4,6,8,9]

Application areas

A. Natural language. Background knowledge - existing large grammars, e.g. Cambridge CLE. Data source - corpora.

B. Genome data. Background knowledge - molecular biology. Data sources - PDB/SCOP and SwissProt.

C. Agents/control. Background knowledge - Dynamics, etc. Data sources - simulators eg. Flight, RoboCup and users.

Research issues

i. Hierarchically structured learning (A,B,C).

ii. Optimal querying/experimentation (A,B,C).

iii. Analysis of comprehensibility (A,B,C).

iv. Multi-agent representation (C).

v. Complex lambda-term structures (A).

vi. Relevance of background knowledge (A).

Computational logic

a. Integration of logic and probability/statistics (vi).

b. Integration of time/-complexity analysis into logic (i,ii,iii,iv).

c. Efficient higher-order representations (i,v,vi).

d. Representation of agency (iv).

Conclusion

- Representation of delarative knowledge.
- Use of background knowledge.
- Applications to scientific discovery and language learning.
- Various possible extensions of representation.

Bibliography

ILP general

Muggleton (1999), *Inductive Logic Programming: issues, results and the LLL challenge*, Artificial Intelligence, 114(1):283-296.

Muggleton and De Raedt (1994), *Inductive Logic Programming: Theory and Methods*, Journal of Logic Programming, 19,20, 629-679.

ILP and Scientific Discovery

Muggleton (1999), *Scientific Knowledge Discovery using Inductive Logic Programming*, Communications of the ACM, 42(11):42-46.