Introduction to Inductive Logic Programming

Tomáš Horváth

About me

- assistant at ICS, UPJŠ in Košice
- PhD. student of prof. Peter Vojtáš, Prague
- research interests concerns ILP
  - ordinal classification
  - fuzzy ILP
  - Generalized Annotated Programs induction
  - user preferences, profiles

The presentation

- Inductive Logic Programming (ILP)
  - (Multi) Relational Data Mining method
  - Machine Learning + Logic Programming
  - complex data structures
    - medicine, genetics, chemistry, economic ...
- Goals
  - give basic knowledge on ILP
  - describe our results in this field

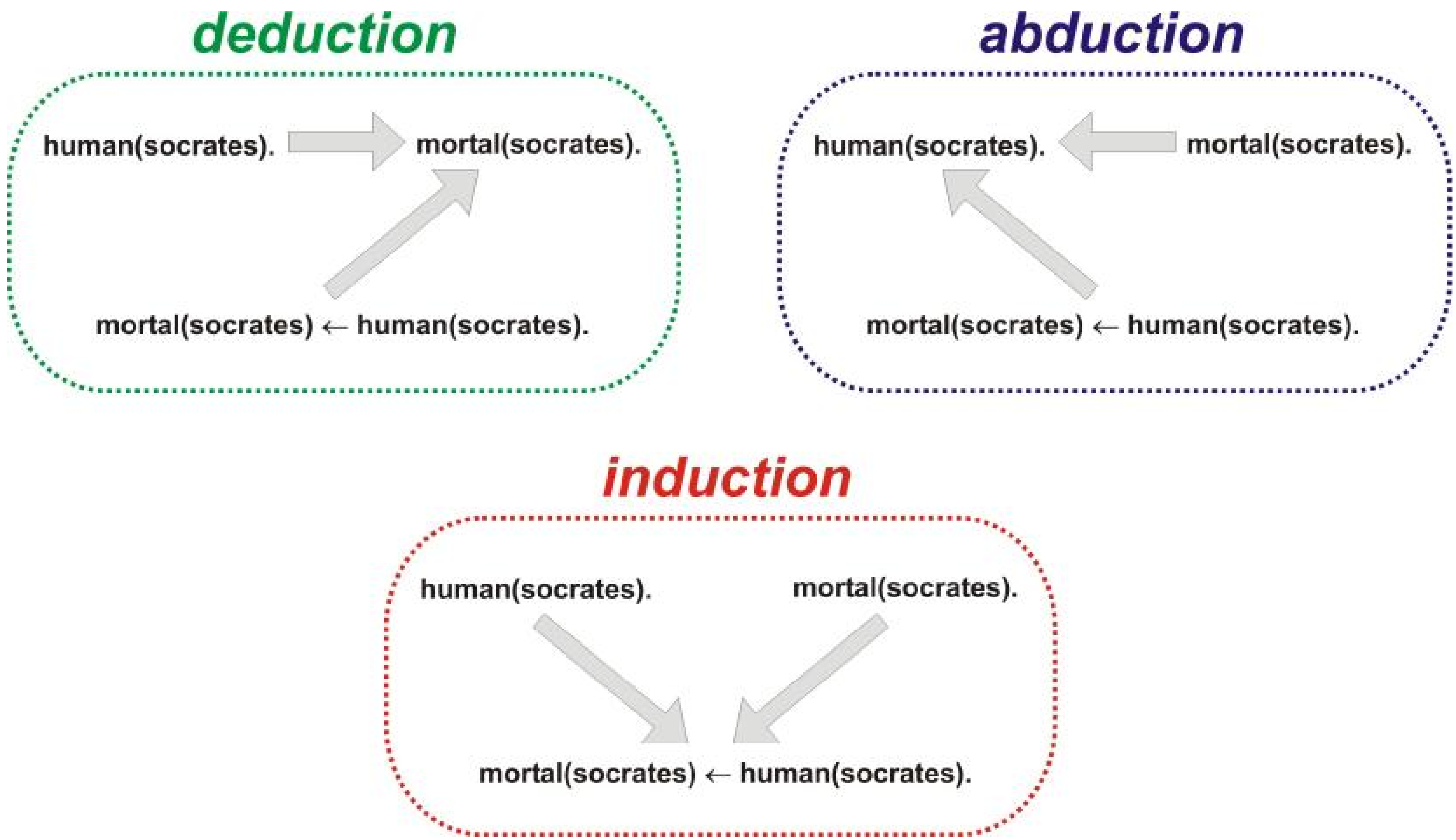
Outlines

- Basic concepts
- ILP techniques
  - refinement graphs (FOIL)
  - inverse resolution (CIGOL)
  - relative least generalization (GOLEM)
  - inverse entailment (ALEPH)
- Applications
- Future directions of ILP
- Our research

Several forms of reasoning

- (Background) Knowledge
  - Socrates is a human
- Observations (Examples)
  - Socrates is mortal
- Theory (Hypothesis)
  - IF X is a human THEN X is mortal

Several forms of reasoning



Several forms of reasoning

- Deduction (Abduction)
  - if the theory and background knowledge (examples) are true then the examples (background knowledge) are also true.
- Induction
  - an induced theory from given examples and background knowledge need not be true in case of other examples or background knowledge not used in the induction process

General ILP task

- Given
  - Background Knowledge ***B***
  - Examples ***E***
    - Positive e^+
    - Negative e^- (sometimes not used in the learning process)
- Find
  - Hypothesis ***H***, such that
    - covers all positive examples (*completeness*)
    - covers non of the negative examples (*consistency*)
    - a complete and consistent hypothesis is *correct*

Normal setting (predictive)

- Representations
  - example *e* – <u>definite clause</u> (fact)
  - background knowledge *B* – definite program
  - hypothesis *H* – definite program

- H *covers* e w.r.t. B if
  - *(H È B) |= e*

Normal setting (predictive)

- Positive examples
  - { daughter(mary,ann), daughter(eve,tom) }
- Negative examples
  - { daughter(tom,ann), daughter(eve,ann) }
- Background Knowledge
  - { mother(ann,mary), mother(ann,tom), father(tom,eve), father(tom,ian), female(ann), female(mary), female(eve), male(ian), male(tom), parent(X,Y)←mother(X,Y), parent(X,Y) ← father(X,Y) }
- Hypotheses
  - { daughter(X,Y) ← female(X), parent(Y,X) }
  - { daughter(X,Y)←female(X), mother(Y,X); daughter(X,Y)←female(X), father(Y,X) }

Non-monotonic setting (descriptive)

- Representations
  - example *e* – <u>Herbrand interpretation</u>
    - often just positive examples
  - background knowledge *B* – definite program
  - hypothesis *H* – definite program

- H *covers* e w.r.t. B if
  - *H is true in the least Herbrand model* $M(B \cup E)$

Non-monotonic setting (descriptive)

- Examples
  - { mother(lieve,soetkin), father(luc,soetkin), parent(lieve,soetkin), parent(luc,soetkin), male(luc), female(lieve), female(soetkin), human(lieve), human(luc), human(soetkin) }
  - { mother(blaguna,sonja), father(veljo,saso), father(veljo,sonja), parent(blaguna,saso), parent(blaguna,sonja), parent(veljo,saso), parent(veljo,sonja), male(veljo), male(saso), female(blaguna), female(sonja), human(veljo), human(saso), human(blaguna), human(sonja) }
- Empty background knowledge
- Hypothesis
  - { parent(X,Y)←mother(X,Y); parent(X,Y)←father(X,Y); mother(X,Y)∨father(X,Y)←parent(X,Y); ←mother(X,Y),father(X,Y); human(X)←female(X); human(X)←male(X); female(X)∨male(X)←human(X); ←female(X),male(X); female(X)←mother(X,Y); male(X)←father(X,Y); human(X)←parent(X,Y); human(Y)←parent(X,Y); ←parent(X,X) }

Non-monotonic setting (descriptive)

- Examples
  - { class(fix), worn(gear), worn(chain) }
  - { class(sendback), worn(engine), worn(chain) }
  - { class(sendback) ,worn(wheel) }
  - { class(ok) }
- Background knowledge
  - { replaceable(gear), replaceable(chain), not_replaceable(engine), not_replaceable(wheel) }
- Hypothesis
  - { class(sendback)$\leftarrow$worn(X), not_replaceable(X) }

Non-monotonic setting (descriptive)

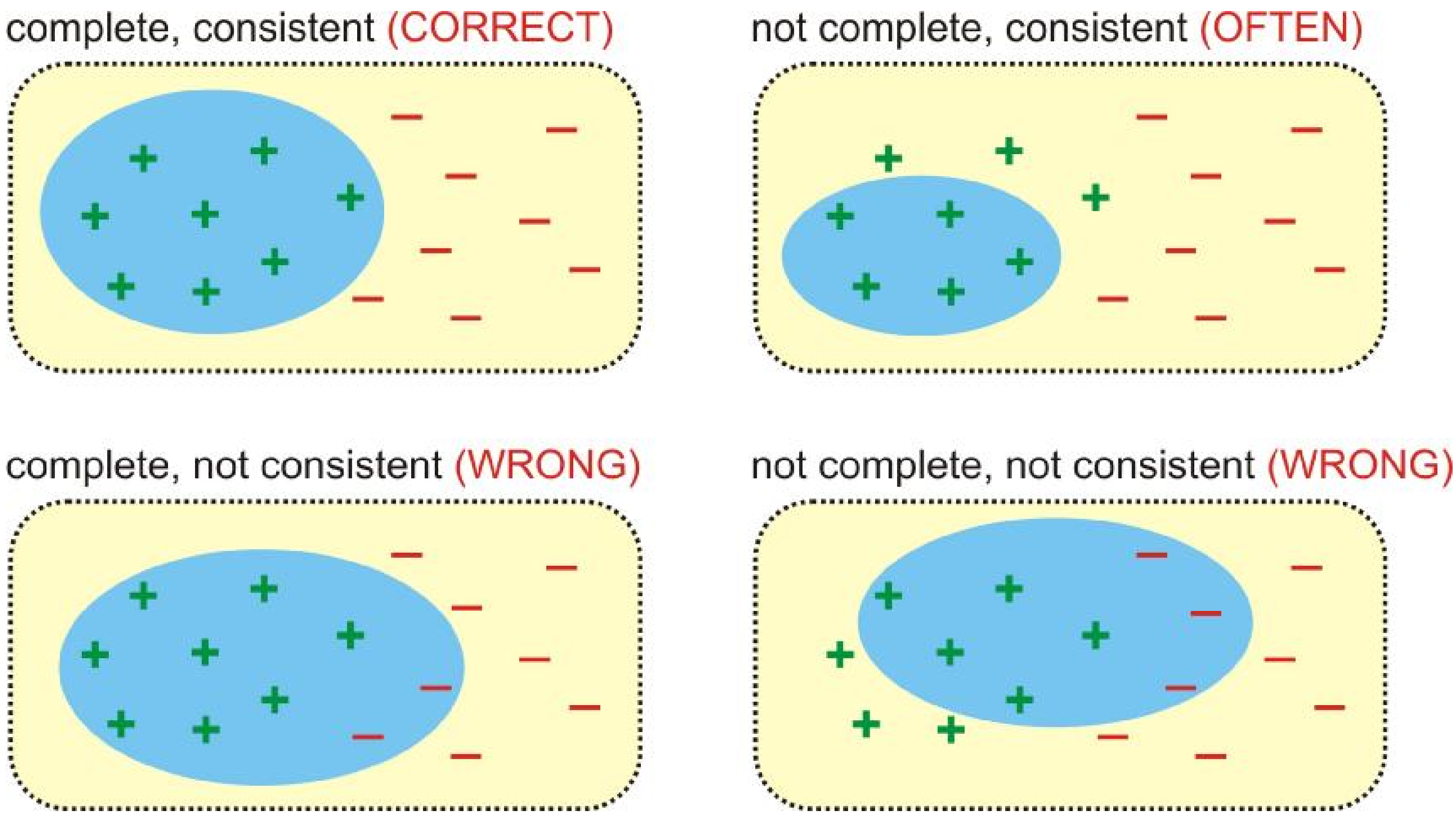
- Positive examples
  - { daughter(mary,ann), daughter(eve,tom) }
- Negative examples
  - { daughter(tom,ann), daughter(eve,ann) }
- Background Knowledge
  - { mother(ann,mary), mother(ann,tom), father(tom,eve), father(tom,ian), female(ann), female(mary), female(eve), male(ian), male(tom), parent(X,Y)←mother(X,Y), parent(X,Y) ← father(X,Y) }
- Hypotheses
  - { daughter(X,Y) ← female(X), parent(Y,X) }
  - { ←daughter(X,Y), mother(X,Y); female(X)←daughter(X,Y); mother(X,Y)∨father(X,Y)←parent(X,Y) }

Predictive vs. Descriptive ILP

- Predictive
  - Learn a reason why positives are positives and negatives are negatives
  - You know what You are looking for, but you don’t know what it looks like.
  - Separate examples and background knowledge
  - <u>often used</u>

- Descriptive
  - Find something interesting about the data
  - You don’t know what You are looking for
  - all background knowledge about an example is incorporated in this example

Completeness and Consistency



Specialisation vs. Generalization

- $C \models D$
  - D is a specialisation of C
  - C is a generalization of D

- if $C \not\models e$ then $D \not\models e$
- if $D \models e$ then $C \models e$

The general ILP algorithm

- Input: E^+, E^-, B
- Output: H

- begin
  - initialize H
  - repeat
    - if H is not consistent specialize it
    - if H is not complete generalize it
  - until H is not correct
  - output H
- end

Subsumption Theorem

- cover relation “|=“
  - hard to implement
  - not decidable
  - need a framework to solve this problem

- subsumption
  - a clause C subsumes a clause D (*C ³ D*) if *($q) Cq Í D*
    - $C=p(X)\leftarrow q(a),r(Y) = \{p(X),\neg q(a),\neg r(Y)\} \geq \{p(b),\neg q(a),\neg r(c),\neg s(Z)\} = p(b)\leftarrow q(a),r(c),s(Z)=D$ for $\theta=\{X/b, Y/c\}$
  - if C ≥ D then C |= D (the converse does not hold)
    - $C=P(f(X))\leftarrow P(X)$, $D=P(f^2(X))\leftarrow P(X)$

Subsumption Theorem

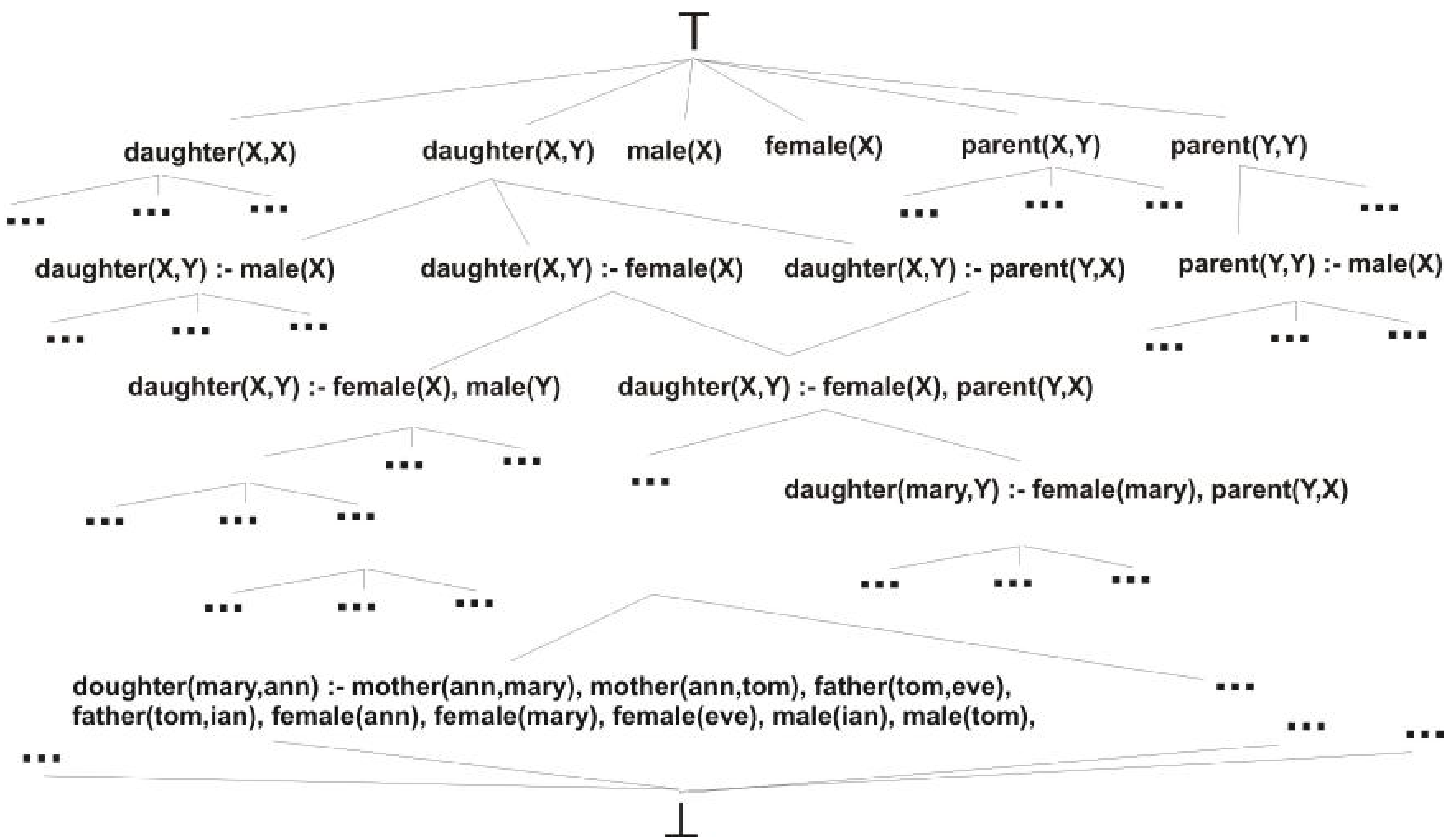
- SLD-refutation theorem
  - Let Σ is a set of Horn clauses. Then Σ is unsatisfiable iff $\Sigma \mid\!-_{sr} \Box$.
- SLD-Subsumption theorem
  - Let Σ is a set of Horn clauses and C a Horn clause. Then $\Sigma \models C$ iff $\Sigma \mid\!-_{sd} C$.
- SLD-refutation theorem and SLD-Subsumption theorem are equivalent.

- $\Sigma \mid\!-_{sr} C$ if there exists an SLD-resolution of C from Σ.
- $\Sigma \mid\!-_{sd} C$ if there exists an SLD-resolution of a clause D from Σ such that $D \geq C$ (D subsumes C)

Hypothesis space

- Space of (all) Horn clauses *H*
  - ordered by subsumption
- for every finite set $S \subseteq H$ there exists a greatest specialisation of S in *H*
- for every finite set $S \subseteq H$ there exists a least generalisation of S in *H*
- *H* ordered by $\geq$ is a lattice
  - $\perp$ - bottom element
  - T – top element

Hypothesis space



Hypothesis space

- Large space of all hypotheses
  - need for a space of acceptable hypotheses
    - language bias
- Refinement operator $r : H \rightarrow H$
  - determine the hypothesis space (refinement graph)
  - specialisation operator
    - $r(C)=D,\ C \models D$
      - applies a substitution θ to C
      - adds literal to the body of C
  - generalisation operator
    - $r(C)=D,\ D \models C$
      - applies an inverse substitution θ^{-1} to C
      - removes literal from the body of C

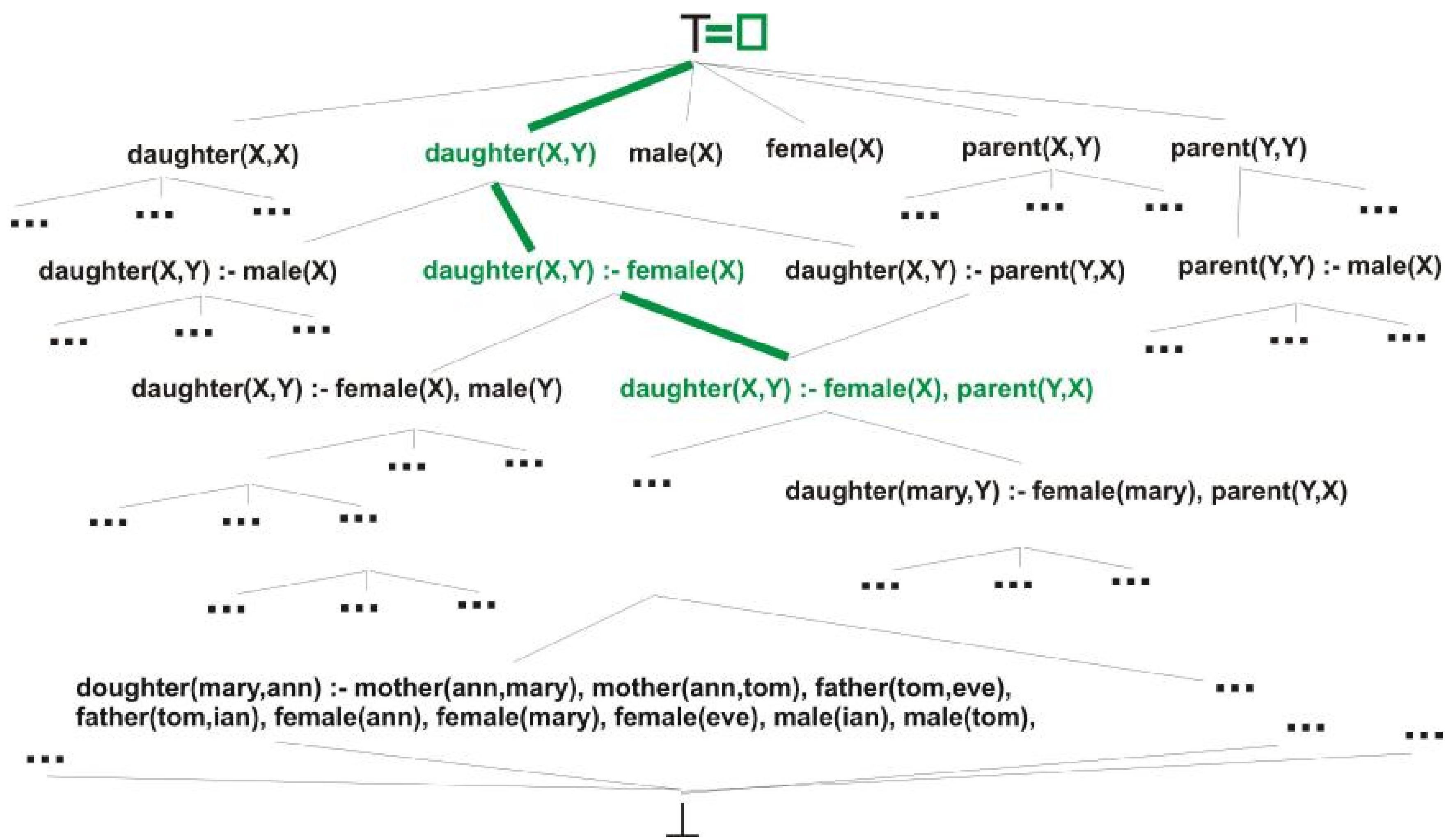
Outlines

- Basic concepts
- ILP techniques
  - refinement graphs (FOIL)
  - inverse resolution (CIGOL)
  - relative least generalization (GOLEM)
  - inverse entailment (ALEPH)
- Applications
- Future directions of ILP
- Our research

Searching refinement graphs

- top-down searching of refinement graph
- starting with T= □
- depth-first search
- implemented in system FOIL

Searching refinement graphs



Inverse resolution

- bottom-up approach
- applying inverse resolution to clauses
  - V-operators
    - absorption
    - identification
  - W-operators
    - intra-construction
    - inter-construction
- predicate invention
- not deterministic
- implemented in system CIGOL

Inverse resolution

absorption

$q \leftarrow A$ $\quad$ $p \leftarrow q,B$

$p \leftarrow A,B$

intra-construction

$q \leftarrow B$ $\quad$ $p \leftarrow A,q$ $\quad$ $q \leftarrow C$

$p \leftarrow A,B$ $\quad$ $p \leftarrow A,C$

identification

$p \leftarrow A,q$ $\quad$ $q \leftarrow B$

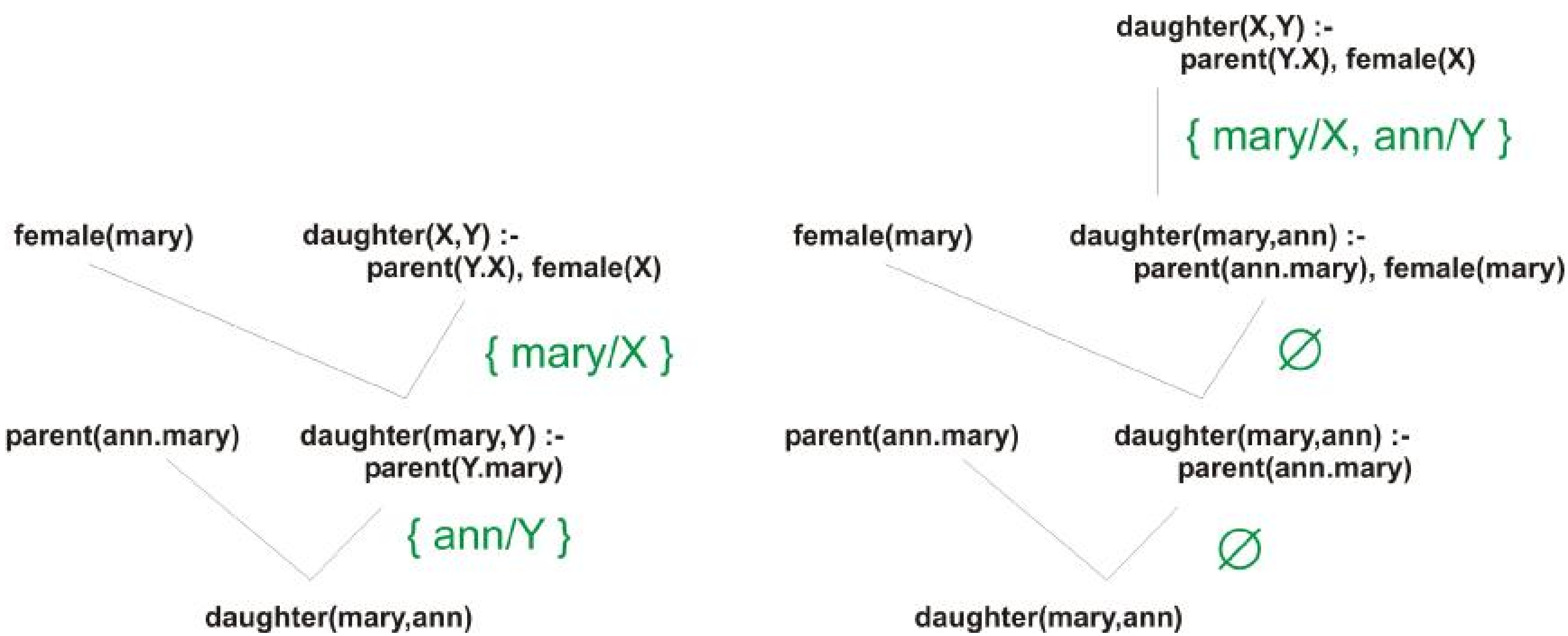
$p \leftarrow A,B$

inter-construction

$p \leftarrow r,B$ $\quad$ $r \leftarrow A$ $\quad$ $q \leftarrow r,C$

$p \leftarrow A,B$ $\quad$ $q \leftarrow A,C$

Inverse resolution



Relative least generalization

- $H \cup B \models e$
  - Let H consist of single clause C
  - $C \cup B \models e \Rightarrow C \models B \rightarrow e$
  - if e – atom, B – atoms then $e \leftarrow B$ is a Horn clause

- $C \geq_B D$ if $C \geq (D \cup \{\neg L_1, \ldots, \neg L_n\})$
- $LGS((D_1 \cup \{\neg L_1, \ldots, \neg L_n\}), \ldots, (D_m \cup \{\neg L_1, \ldots, \neg L_n\})$ is an $RLGS_B$ of $\{D_1, \ldots, D_m\}$ relative to $B=\{L_1, \ldots, L_n\}$ in *H*

- bottom-up approach
  - searches correct $LGRS_B$ of positive examples
- implemented in system GOLEM

Relative least generalization

- $RLGS_B$(daughter(mary,ann),daughter(eve,tom)) for B={female(mary), parent(ann,mary), female(eve), parent(tom,eve), female(ann)} is
- daughter($V_{m,e}$,$V_{a,t}$) $\leftarrow$ parent(ann,mary), parent(tom,eve), female(mary), female(eve), female(ann),parent($V_{a,t}$,$V_{m,e}$), female($V_{m,e}$), female($V_{m,a}$), female($V_{a,e}$).
  - if C\{L} covers at least as many positive examples and at most as many negative examples as C then the literal L is irrelevant
- after removing irrelevant literals we get daughter($V_{m,e}$,$V_{a,t}$) $\leftarrow$ parent($V_{a,t}$,$V_{m,e}$), female($V_{m,e}$), so daughter(X,Y) $\leftarrow$ parent(Y,X), female(X)

Inverse entailment

- $H \cup B \models e$
  - Let H consist of single clause C
  - $C \cup B \models e \Rightarrow B \cup \neg e \models \neg C$
  - $\neg\bot$ is a (possibly infinite) conjunction of ground literals which are true in every model of $B \cup \neg e$
  - $B \cup \neg e \models \neg\bot$
  - $\neg C$ is true in all models of $B \cup \neg e \Rightarrow \neg C$ contains a subset of $\neg\bot$
  - $B \cup \neg e \models \neg\bot \models \neg C \Rightarrow C \models \bot$
- top-down approach
  - searches for clauses which subsumes $\bot$
  - ability to have rules in background knowledge
- implemented in system ALEPH
  - language declarations

Inverse entailment

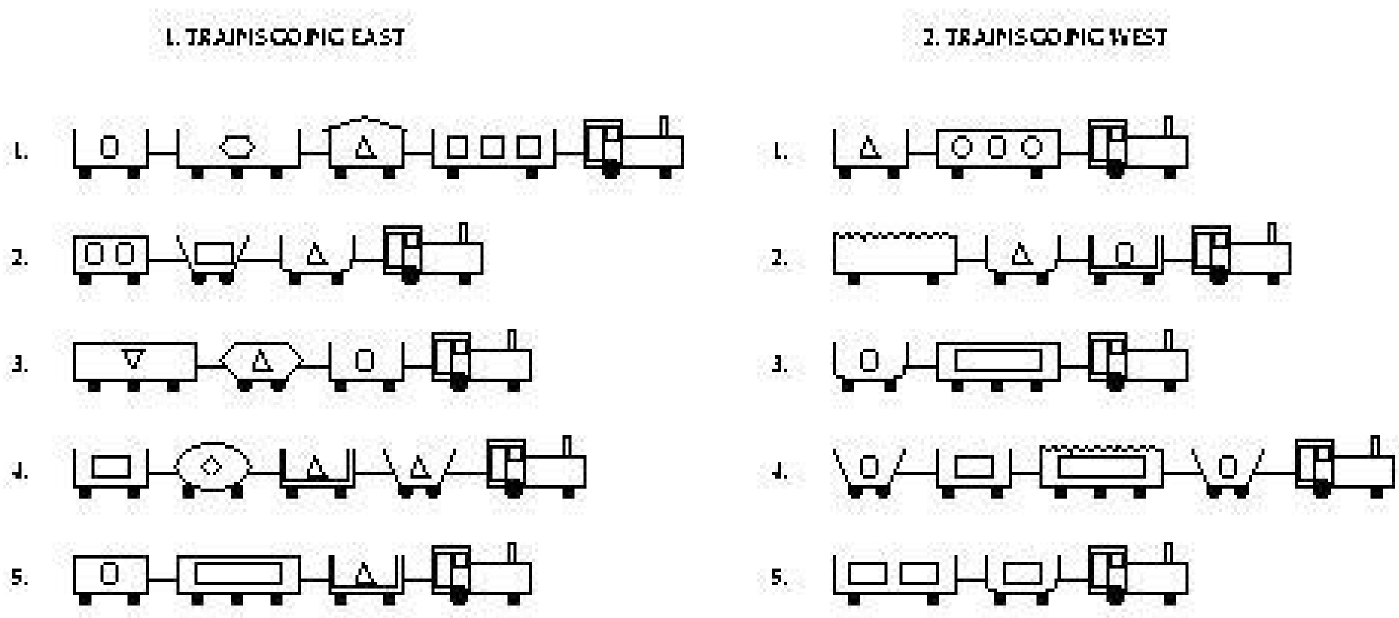
- $\bot_{daughter(mary,ann)}$ = daughter(A, B) :- mother(B, A), female(B), female(A), parent(B, A).
- $\bot_{daughter(eve,tom)}$ = daughter(A, B) :- father(B, A), female(A), male(B), parent(B, A).
- H = { daughter(A, B) :- female(A), parent(B, A). }

Outlines

- Basic concepts
- ILP techniques
  - refinement graphs (FOIL)
  - inverse resolution (CIGOL)
  - relative least generalization (GOLEM)
  - inverse entailment (ALEPH)
- Applications
- Future directions of ILP
- Our research

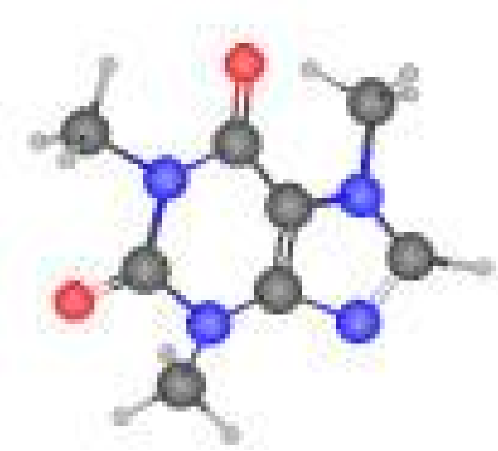
East-West trains

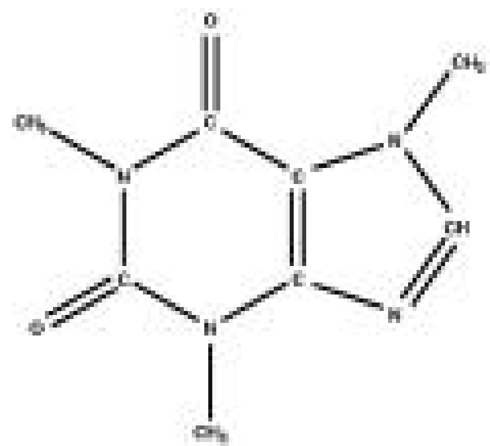
- what makes a train to go eastward?



Biology, Chemistry, ...

- what makes a molecule to be mutagenetic?





Other applications

- Engineering
  - finite element mesh design
  - detecting a traffic problem
- Natural language processing
  - learning language grammars
  - speech tagging
  - text categorisation
- Life Sciences
  - 3D protein structure
  - predicting carcinogenicity

Outlines

- Basic concepts
- ILP techniques
  - refinement graphs (FOIL)
  - inverse resolution (CIGOL)
  - relative least generalization (GOLEM)
  - inverse entailment (ALEPH)
- Applications
- Future directions of ILP
- Our research

Imperfection, imprecision

- Bayesian nets
- Probabilistic reasoning
- Fuzzy logic

Large data sets

- use of a power of databases
- distributed mining
- efficient choice of training set

Semantic web

- XML, ...
- description logic, RDF

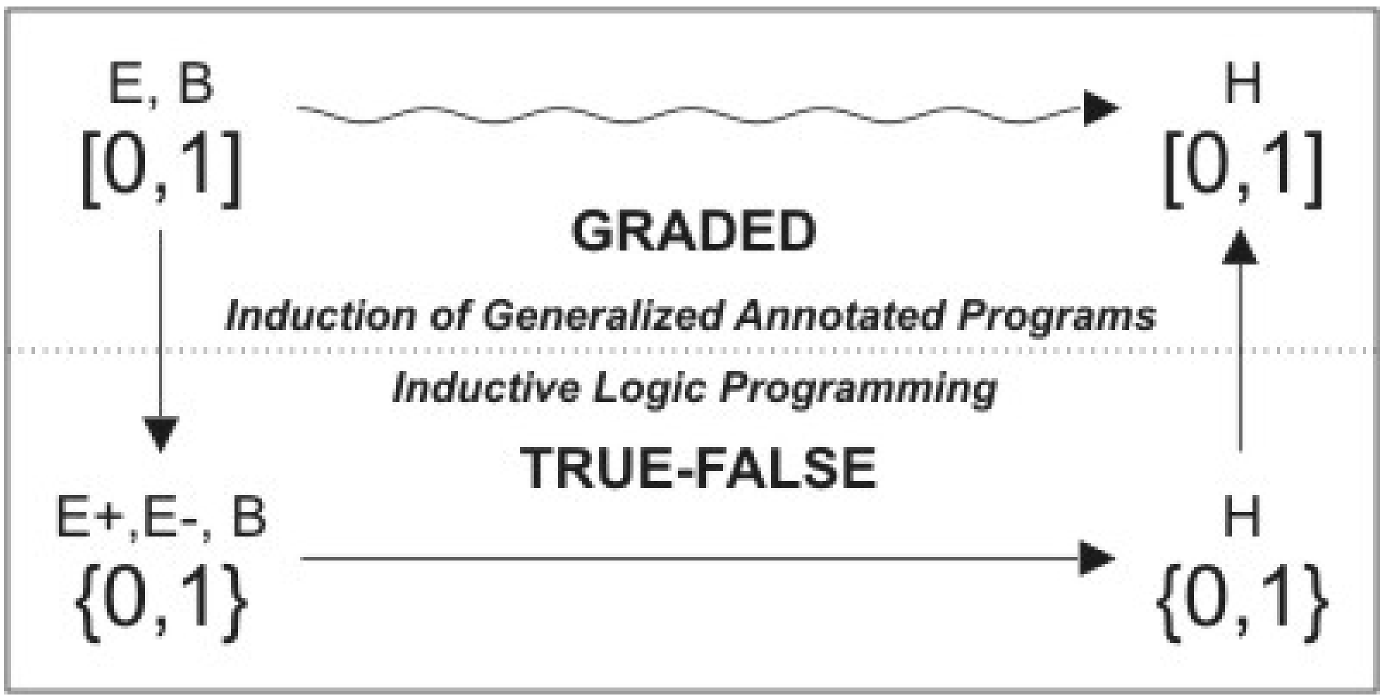
Outlines

- Basic concepts
- ILP techniques
  - refinement graphs (FOIL)
  - inverse resolution (CIGOL)
  - relative least generalization (GOLEM)
  - inverse entailment (ALEPH)
- Applications
- Future directions of ILP
- Our research

Inductive Generalized Annotated Programming

- $A:\rho(\mu_1, \ldots, \mu_k) \leftarrow B_1:\mu_1 \ \& \ \ldots \ \& \ B_k:\mu_k$
- the first approach to induce GAP programs
- convenient for ordinal classification problems
- equivalent to fuzzy logic programs
- better representation of the real-world as probabilistic ILP approaches
- multiple use of classical ILP system ALEPH with orderings in the background knowledge
- succesfully used in a Slovak project NAZOU
  - learning user preferences

Inductive Generalized Annotated Programming



Inductive Generalized Annotated Programming

Hotel name	Location	Price ($)
africa	centre	20
america	east	50
antarctica	west	80
australia	east	110
asia	west	50
europe	centre	80

Conference name	Location
icml	centre
ecml	East
ilp	West

Price ($)	TV of „cheap“
20	1.0
50	0.7
80	0.4
110	0.1

TV of „near“ from / to	centre	east	west
centre	1.0	0.7	0.4
east	0.7	1.0	0.1
west	0.4	0.1	1.0

TV of „approppriateness of a hotel for a conference“	icml	ecml	ilp
africa	1.0	1.0	0.7
america	1.0	1.0	0.1
antarctica	0.4	0.1	0.4
australia	0.1	0.1	0.1
asia	0.7	0.1	1.0
europe	0.4	0.4	0.4

Results of our algorithm:

A Hotel is appropriate for Conference with truth value at least 1 ***IF*** *its price is cheap with truth value at least 0.7* ***and*** *is located near the conference with truth value at least 0.7*

A Hotel is appropriate for Conference with truth value at least 0.7 ***IF*** *its price is cheap with truth value at least 0.7* ***and*** *is located near the conference with truth value at least 0.4*

A Hotel is appropriate for Conference with truth value at least 0.4 ***IF*** *its price is cheap with truth value at least 0.4* ***and*** *is located near the conference with truth value at least 0.4*

Every Hotel is appropriate for Conference with truth value at least 0.1.

attribute values ordering

cheap(A,X) :- le(X,Y), cheap(A,Y).
near(A,B,X) :- le(X,Y), near(A,B,Y).

le(0.1,0.4). le(0.4,0.7). le(0.7,1.0).

References

- http://www.cs.bris.ac.uk/~ILPnet2/
- Shan-Hwei Nienhuys-Cheng, Ronald de Wolf: *Foundations of Inductive Logic Programming.* Springer-Verlag, 1997, ISBN 3540629270.
- Nada Lavrač, and Sašo Džeroski. *Inductive Logic Programming: Techniques and Applications.* Ellis Horwood, New York, 1994.
- Proceedings of the Conference on Inductive Logic Programming (ILP), since 1990.

Thanks for Your attention.

Tomáš Horváth
horvath@ics.upjs.sk
http://ics.upjs.sk/~horvath