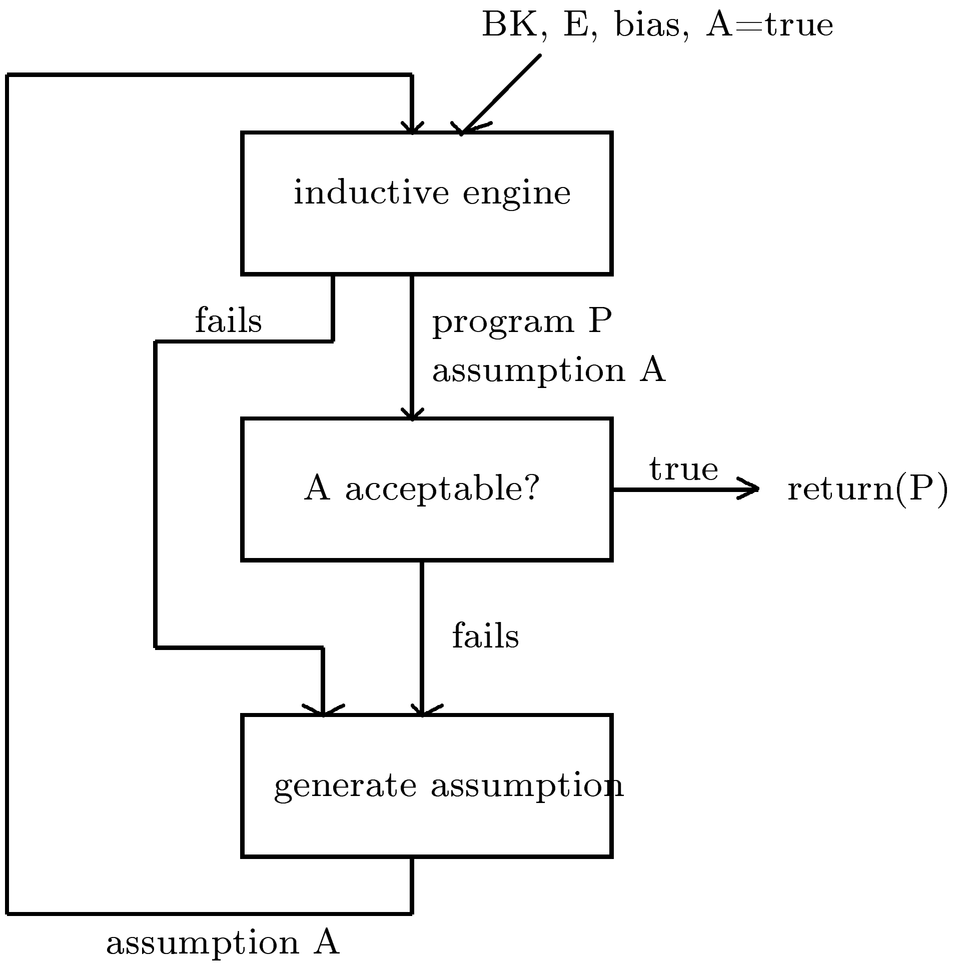
Efficient relational learning from sparse data

Luboš Popelínský
Knowledge Discovery Group
Faculty of Informatics, Masaryk University in Brno, Czechia
popel@fi.muni.cz
http://www.fi.muni.cz/kd

Relational learning - learning in first-order logic
Exact learning - learning from exact data
Sparse data - not more than 5 training examples
Generate&test top-down algoritms - from the most general hypothesis

Assumption-based learning



Generic algorithm of assumption-based learning

Given:

domain knowledge BK, *example set* E, *bias*,
assumption $A = true$
inductive engine I, overgeneral program P
function f, that computes an assumption A
acceptability module AM

```
1. Call $I$ on $BK \cup P, E \cup A, bias$.
     • if $I$ succeeds resulting in program $P^{'}$
           then call $AM$ to validate the assumption $A$.
           if $A$ is accepted then return($P^{'}$) else go to (2).
     • else go to (2).
2. Call $f$ to generate a new assumption $A$. If it fails, return(fail) and stop else go
   to (1).
```

WiM

inductive engine $Markus^+$

- depth-first search
- automatic setting of bias
- multiple predicate learning
- 2nd-order schema may be employed

generator of assumptions

- choose the simplest positive example
- find its *near-miss*

acceptability criterion

- membership oracle

WiM: results

2 – 4 examples for learning most of ILP benchmark predicates (list processing, Peanova aritmetika)

learning from positive examples only; negative examples, if any, generated with WiM itself

max. 1 query to the user

less dependent on quality of examples

easy to use

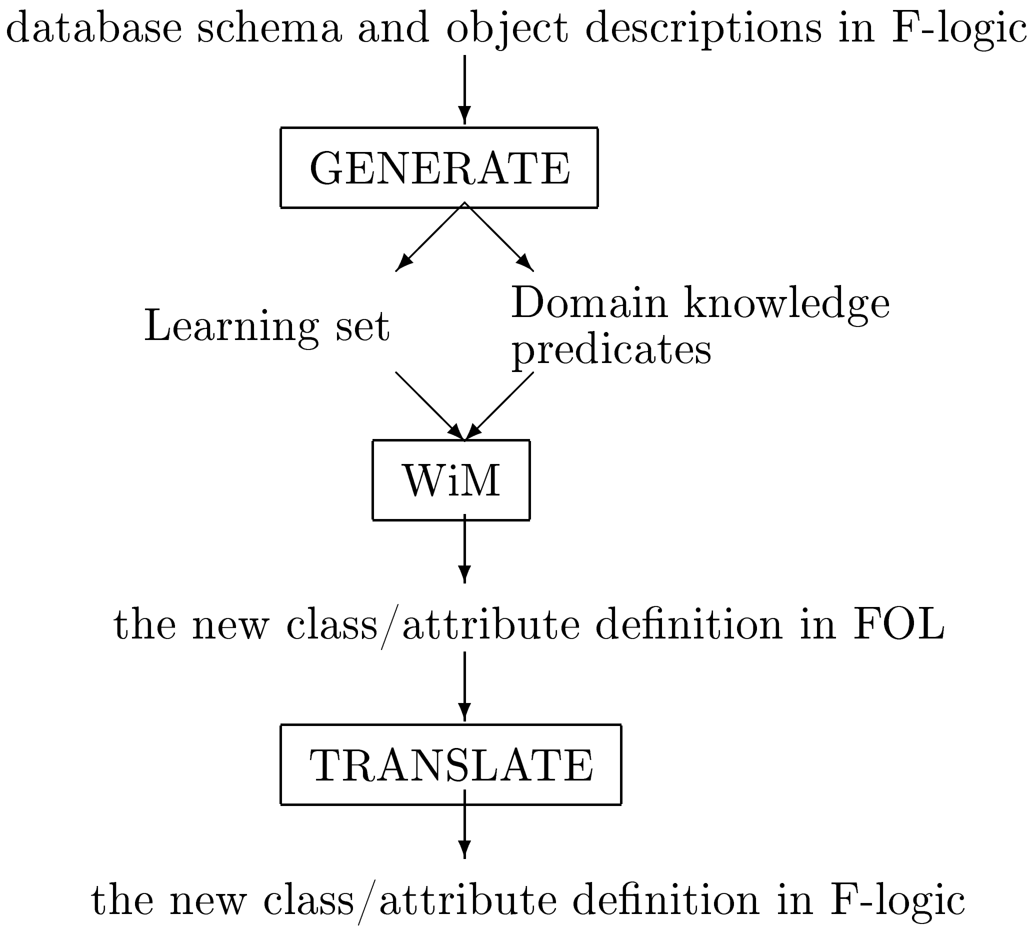
$CRUSTACEAN$, $SKILit$ a WiM : Randomly generated examples

	$CRUSTACEAN$		$SKILit$			WiM		
	2	3	2	3	5	2	3	5
member	0.65	0.76	0.70	0.89	0.95	0.80	0.97	0.97
last	0.74	0.89	0.71	0.72	0.94	0.76	0.89	0.94
append	0.63	0.74	0.76	0.80	0.89	0.77	0.95	0.95
delete	0.62	0.71	0.75	0.88	1.00	0.85	0.88	0.97
reverse	0.80	0.86	0.66	0.85	0.87	0.85	0.95	0.99

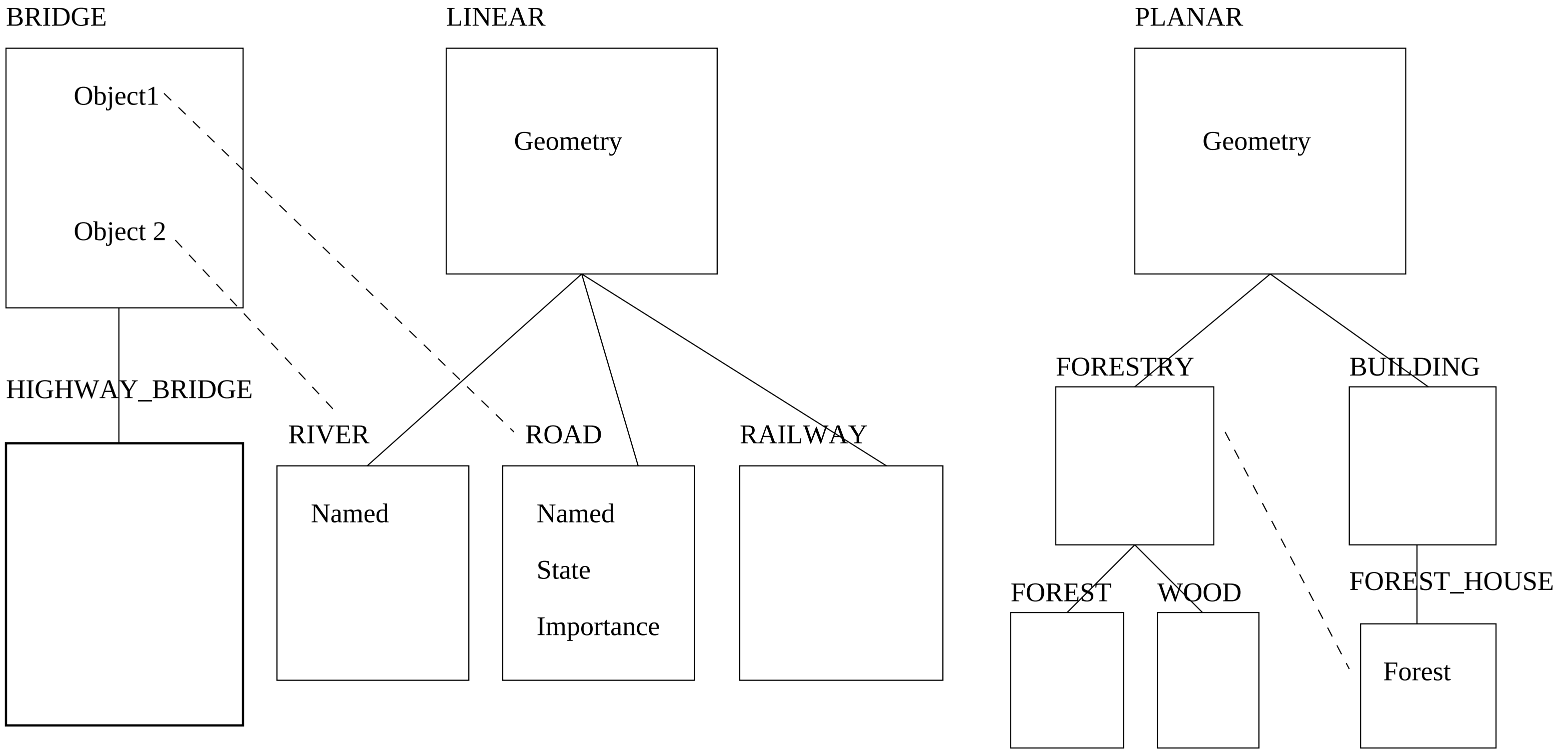
Randomly generated examples: Learning with assumptions

# pos.	2			3			5		
	bez	s	TP	bez	s	TP	bez	s	TP
last	0.885	0.896	6	0.906	0.934	7	0.932	0.971	8
delete	0.882	0.962	8	0.857	0.937	7	0.874	0.943	7
leq	0.380	0.703	0	0.527	0.795	4	0.572	0.932	9
length	0.540	0.659	0	0.692	0.816	1	0.728	0.956	4

DWiM schema


database schema and object descriptions in F-logic
GENERATE
Learning set
Domain knowledge predicates
WiM
the new class/attribute definition in FOL
TRANSLATE
the new class/attribute definition in F-logic

Spatial database schema



Inductive query language for mining in geographic data [PKDD'98]

extract characteristic rule
for bridge
from road, river.

extract discriminate rule
for forest
in contrast to wood
from point of view area.

```
bridge(X,Y):-
     road(X),roadGeometry(X,Z),
     river(Y),riverGeometry(Y,U),
     member(V,Z),member(W,U),W=V.

forest(F) :-
      geometry(F,GForest),
      area(GForest,Area),
      100 < Area.
```

extract dependency rule
for differentHouses
from forestHouse, forest, building
where building(B, GB),
    not forestHouse(B, F)
from point of view distance, less

```
differentHouses(FH,F,H) :-
      distance(FH,F,D1),
      distance(H,F,D2),
      D1<D2.
```