Knowledge Discovery in Spatial Data by Means of ILP

Luboš Popelínský

Masaryk University Brno and CTU Prague, Czechia

Email: `popel@fi.muni.cz`

Motivation

- Inductive logic programming and inductive query languages
- Description of (maybe) inexactly defined geographic objects

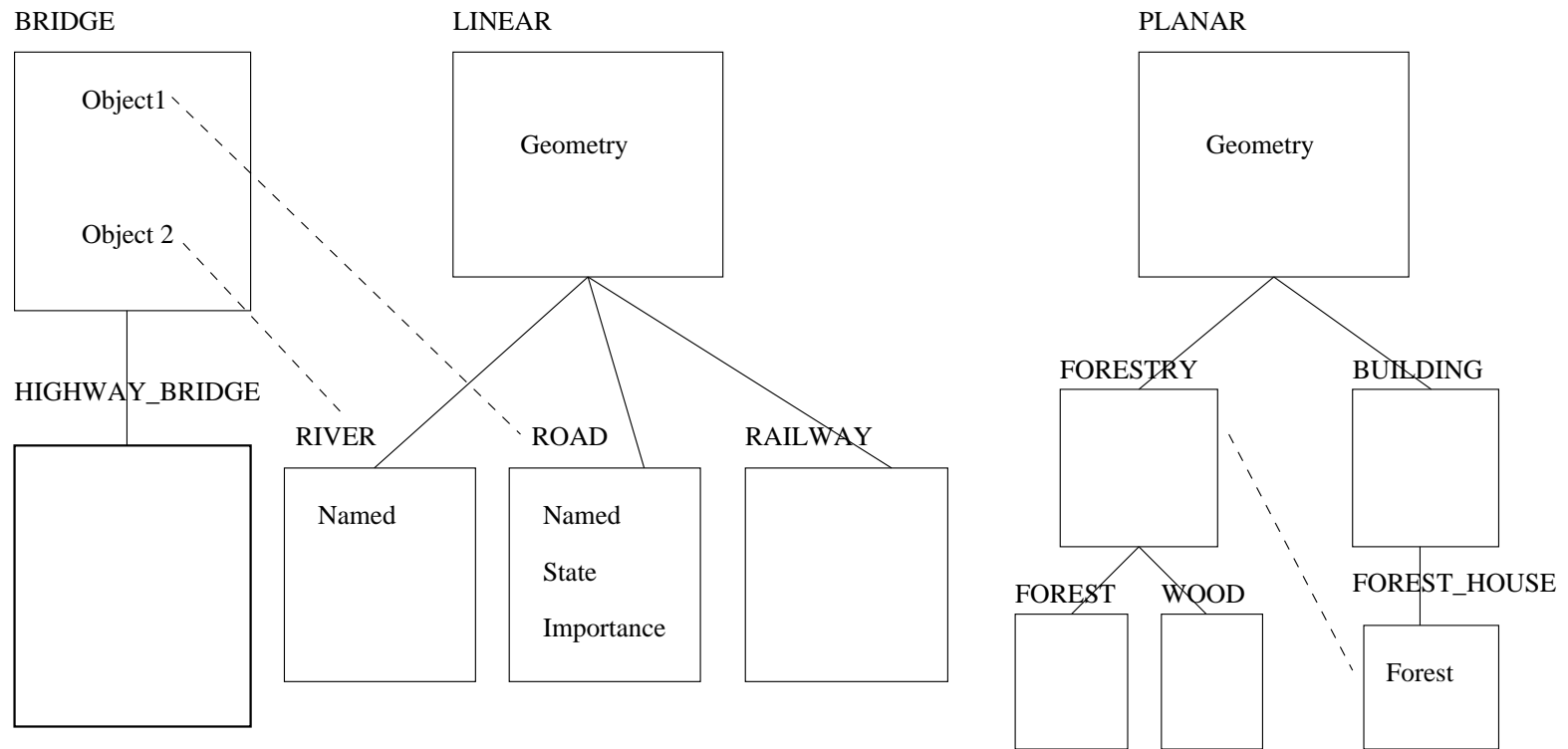# Knowledge Discovery in Spatial Data by Means of ILP

Luboš Popelínský
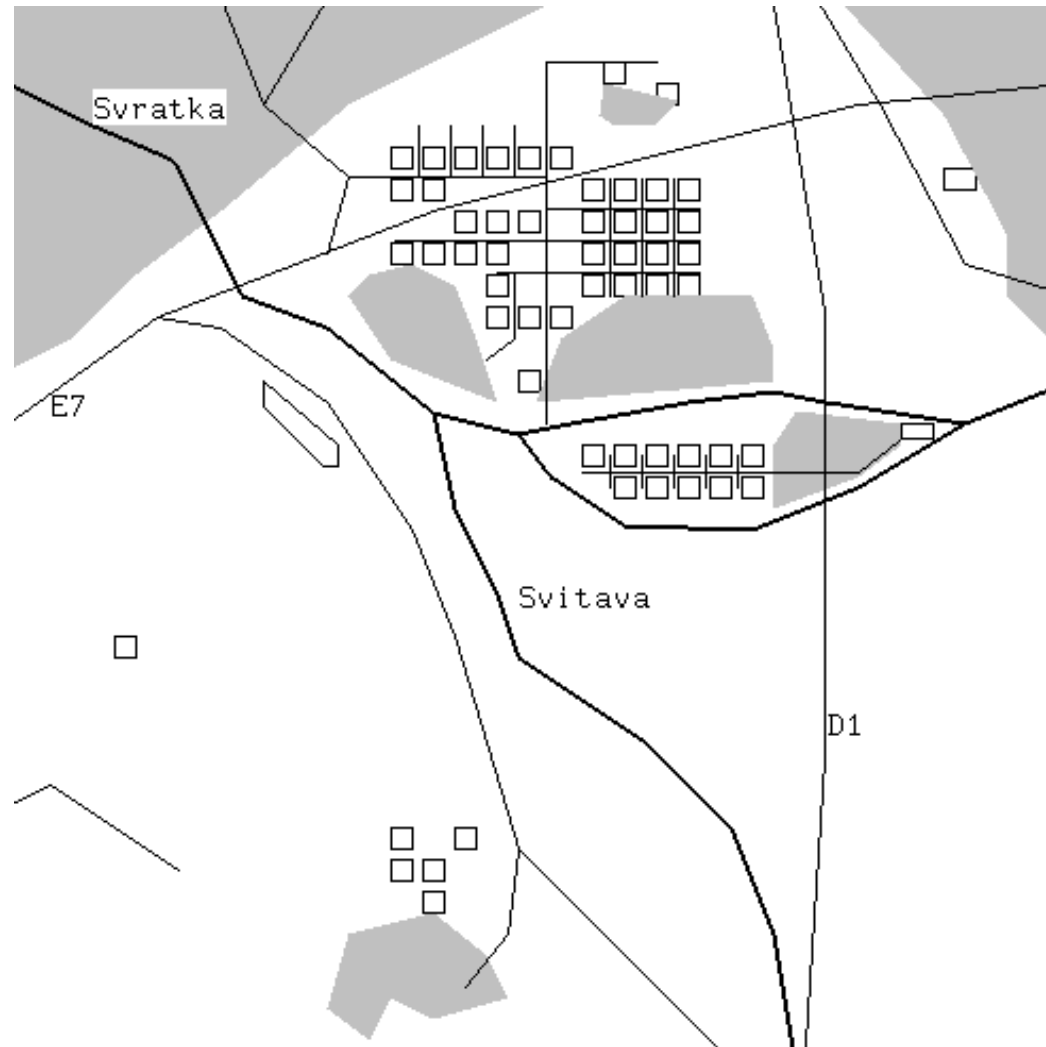
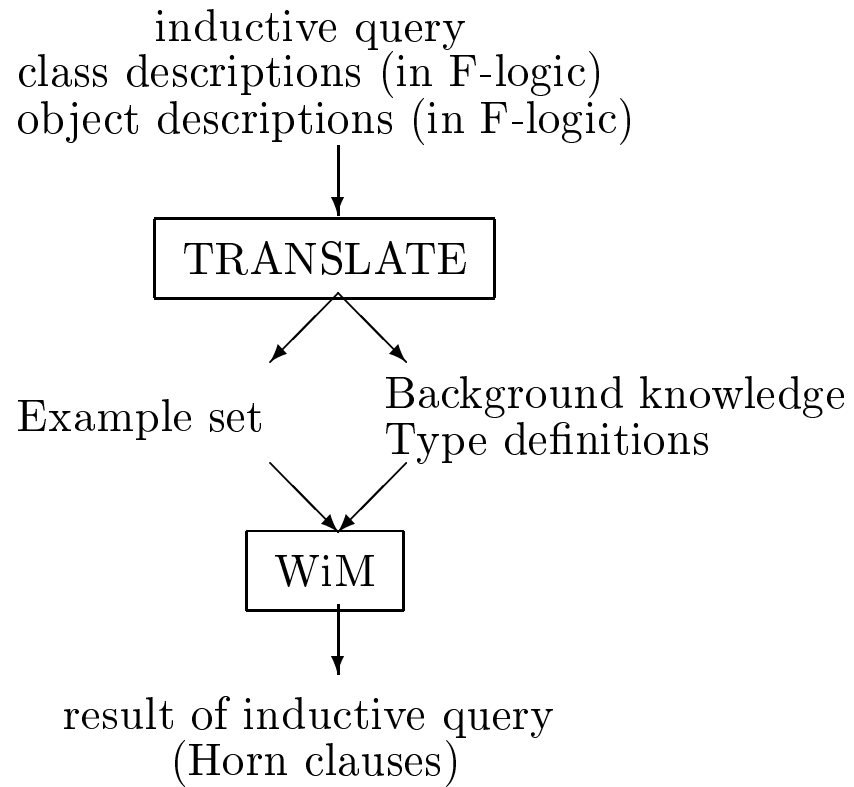*Masaryk University Brno and CTU Prague, Czechia*

Email: `popel@fi.muni.cz`

Outline

1. Inductive query language
2. Method & WiM
3. Examples
4. Discussion & Future research

BRIDGE

Object1

Object 2

HIGHWAY_BRIDGE

LINEAR

Geometry

RIVER

Named

ROAD

Named

State

Importance

RAILWAY

PLANAR

Geometry

FORESTRY

BUILDING

FOREST   WOOD

FOREST_HOUSE

Forest

Object-oriented database schema

Raw data

inductive query
class descriptions (in F-logic)
object descriptions (in F-logic)

TRANSLATE

Example set

Background knowledge
Type definitions

WiM

result of inductive query
(Horn clauses)

$GWiM$ **schema**

WiM

inductive learner

    efficient searching for the refinement graf

    shift of *syntactic* bias

generator of near-misses

oracles

needs from 2 to 4 examples for most of the ILP benchmark predicates (list processing)
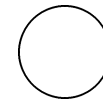
smaller dependency on the quality of the example set in comparison to some of ILP programs

has been tested both on good examples and on randomly chosen example sets.

Inductive language

extract characteristic rule

```
extract  < KindOfRule > rule
for < NameOfTarget >
from [ < ListOfClasses >]
[< Constraints >]
[ from point of view < Domain >]
```

○

extract discriminate rule

○ ○

extract dependency rule

○ ○

Discrimination of Forests and Woods

Find a difference between forests and woods from the point of view of area. *area* is the name of set of predicates like *area*(*Geometry*, *Area*).

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

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

Relation between Forests and Woods

Find a relation between forests and woods from the point of view of area. *area* is the name of set of predicates like *area*(*Geometry*, *Area*).

extract dependency rule
for forestOrWood
from forest, wood
from point of view area.

```
forestOrWood(F,W) :-
    geometry(F,GF),area(GF,FA),
    geometry(W,GW), area(GW,WA),
    WA<GA.
```

Discussion

1. The query language is quite powerful $\Rightarrow$ quite complex queries can be formulated.
   However, the price that user has to pay for is sometimes too big.

2. How to process large amount of data

Future research

- Interface to PostgreSQL object-relational DBMS

- Geographic domain knowledge