



Universiteit Utrecht

[Faculty of Science
Information and Computing Sciences]

An Introduction to LeGo

From **L**ocal **P**atterns to **G**lobal **M**odels workshop,

2008

Arno Knobbe



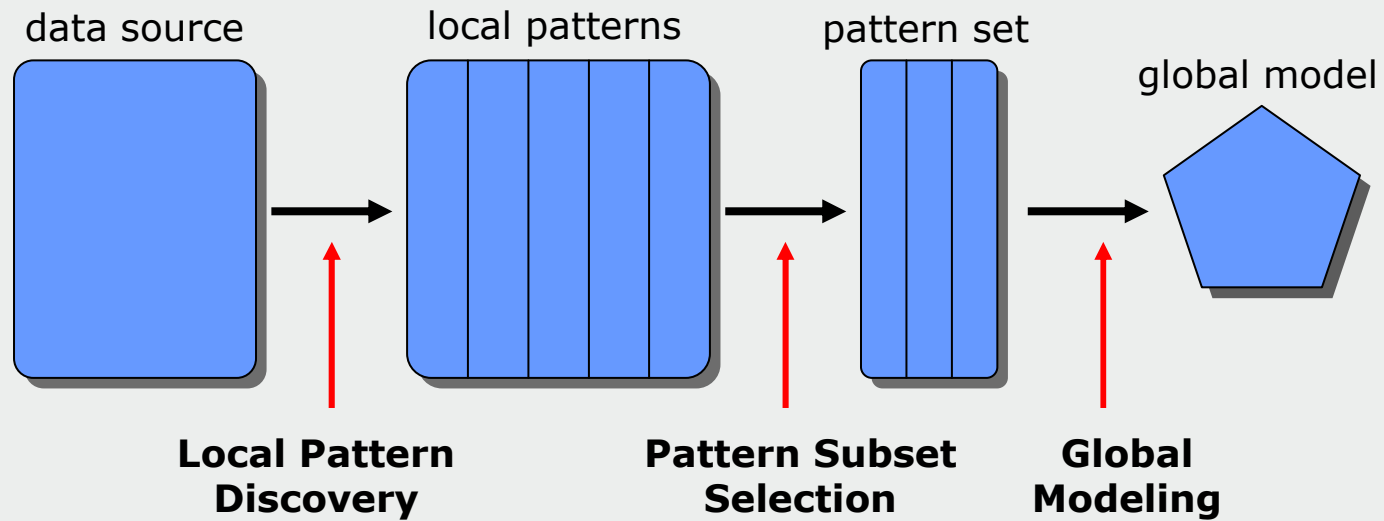
Motivating Ideas

Discussion at Dagstuhl seminar *Parallel Universes & Local Patterns* (2007):

- Easy to find more patterns than you can handle
- Patterns are 'local': half-products or building blocks
- End-users will be interested in 'global' solution
- Many *ad hoc* solutions to this problem



the LeGo phases



Local Pattern Discovery

- Like Feature Construction
- Any Pattern Discovery algorithm
- Exploratory, and often exhaustive, search for dependencies
- Patterns
 - unsupervised: frequency, entropy, ...
 - supervised: information gain, X^2 , WRAcc, ...
- Patterns are binary features: ignore syntax



Pattern Subset Selection

- Like Feature Subset Selection
- Drastically reducing the number of patterns (5–50)
- Removing redundancy between patterns (not top- k)
- Goal:
 - reporting key patterns: human inspection
 - improving global model by removing redundant features

- see Background Talk, Jilles Vreeken, 10:45 - 11:00



Global Modeling

- Inducing global models from patterns
- Deal with overlapping and conflicting patterns/rules
- 'Early attempts'
 - voting
 - probabilistic
 - covering: CBA, CPAR, APRIORI-C
- Patterns = features: any inductive algorithm will do
 - binary, few, non-redundant, predictive

- see Background Talk, Johannes Fürnkranz, 13:45 - 14:00



LeGo as a framework

It's not ...

- a guaranteed process for an optimal global model
- a technical solution (too many options)
- an entirely new idea

It is ...

- a good recipe for dealing with already-mined patterns
- a common frame of reference
- a way of discouraging *ad hoc* solutions
- an inspiration for new methods and research questions

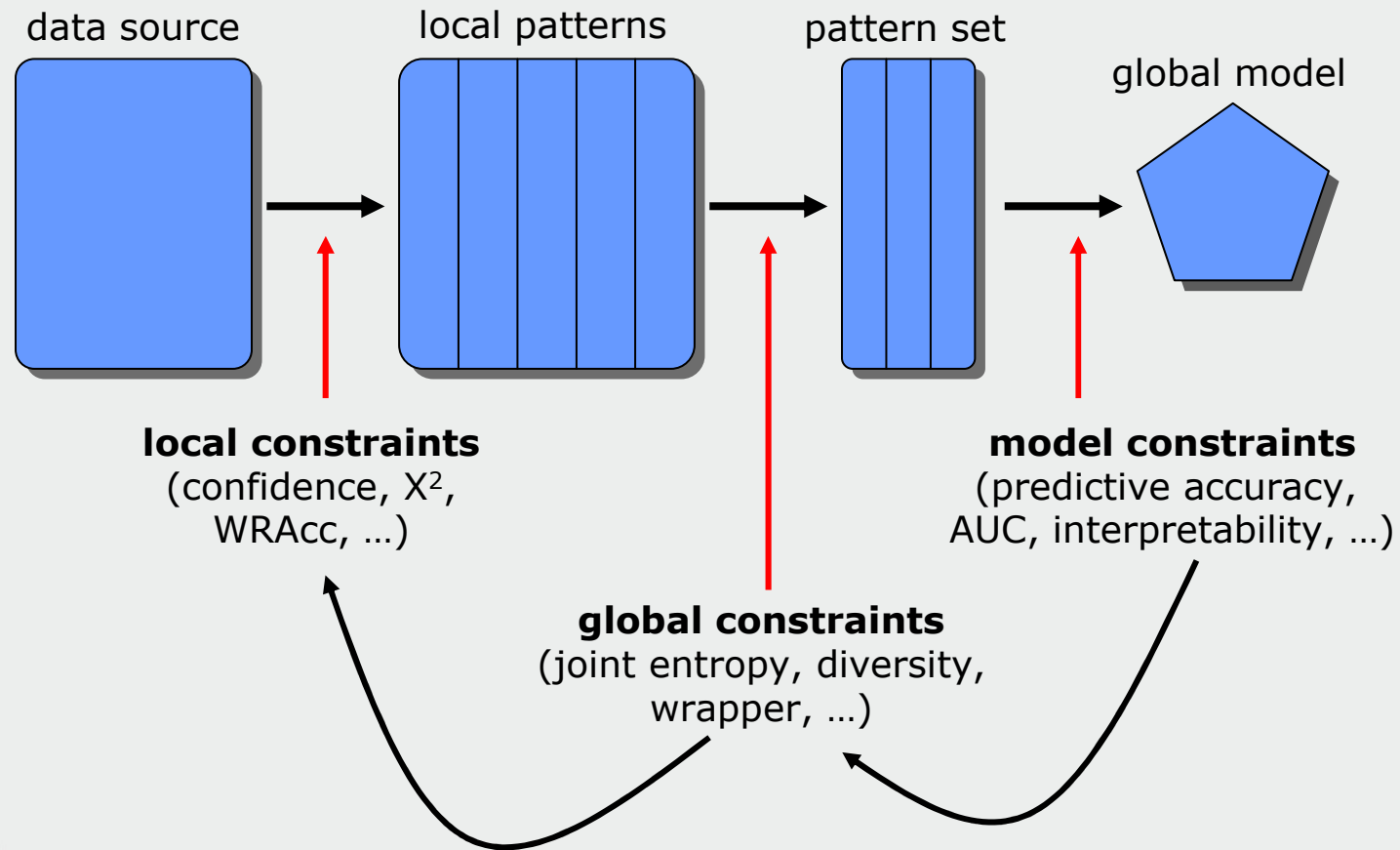


Open issues in LeGo

- Efficient Pattern Subset Selection
- Complexity trade-offs, distribution of work
 - a lot of LPD, some PSS, little GM ?
 - some LPD, no PSS, a lot of GM ?
- Finding patterns that are general-purpose
- Finding patterns that suit a particular global need
 - back-propagation of constraints
- ...
- etc.



Constraints back-propagated



Conclusions

