LeGo 2008: From Local Patterns to Global Models

Participants Survey

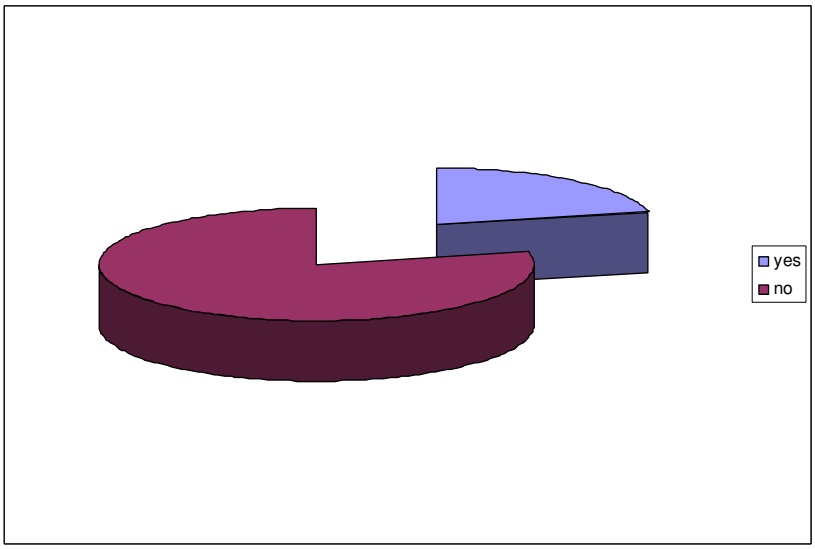
Prior to the LeGo 2008 workshop held on September 2008 in Antwerp, a survey was held amongst the contributors and potential participants of the workshop. The aim of the survey was to gather the opinion of researchers working in LeGo-related fields, on a number of open research questions. Furthermore, we were interested in the different directions this line of research should and would develop, and additionally whether post-workshop publications and future workshops could contribute to that.
Below, you will find an overview of the responses to the different questions of the survey. These results were also presented at the workshop. A total of 14 researchers responded. Although this is a moderate number, we still feel that the results show some interesting trends that may be of benefit to the community.

Arno Knobbe
Johannes Fürnkranz

---

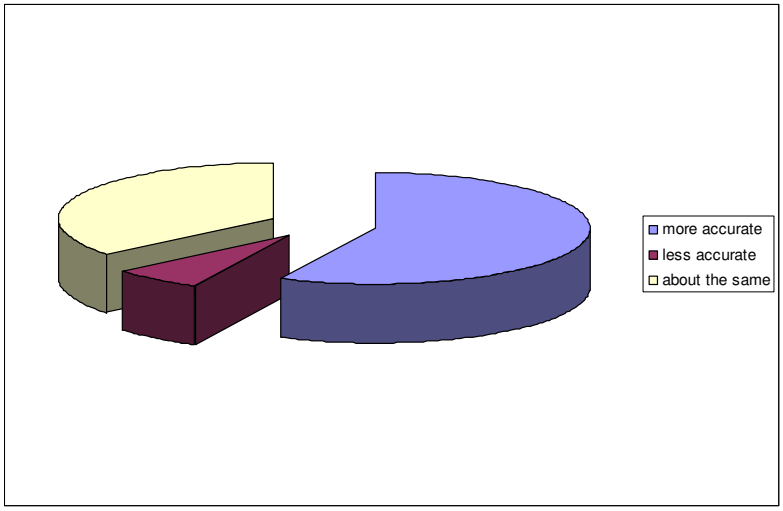
1. The Global Modeling phase can be straightforward. All the complexity and representational power is located in the Pattern Discovery phase.

A. I agree.
B. I disagree.



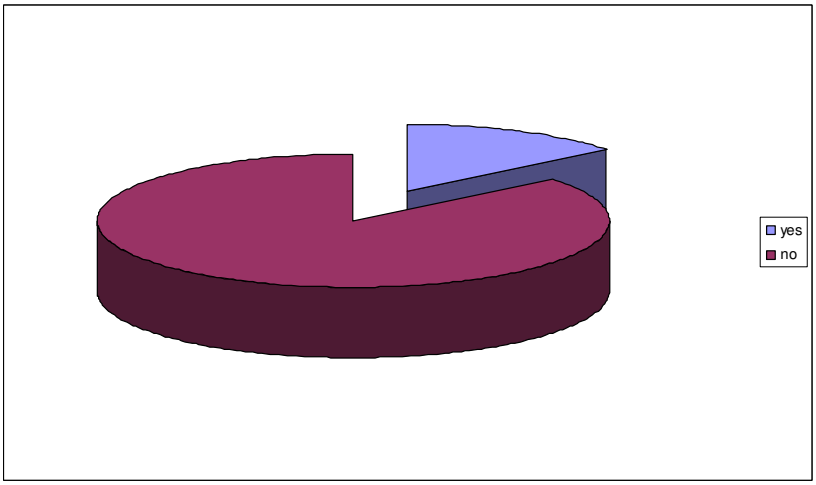
2. Pattern Subset Selection will make the final global model…

A. … more accurate, because the global model is not hampered by large numbers of redundant patterns/features.
B. … less accurate, because pattern selection results in loss of information.
C. … about as accurate as a model induced from all patterns.



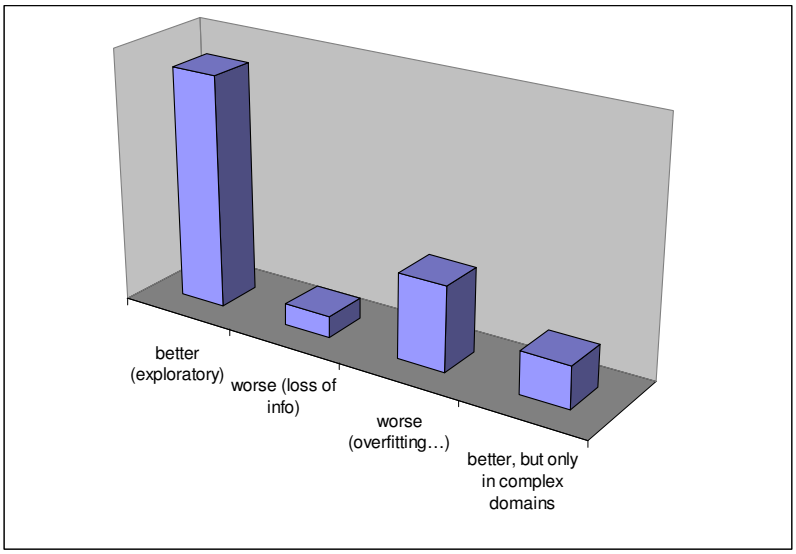
3. Pattern Subset Selection is only useful if a small set of important patterns is needed for human inspection.

A. I agree.
B. I disagree.



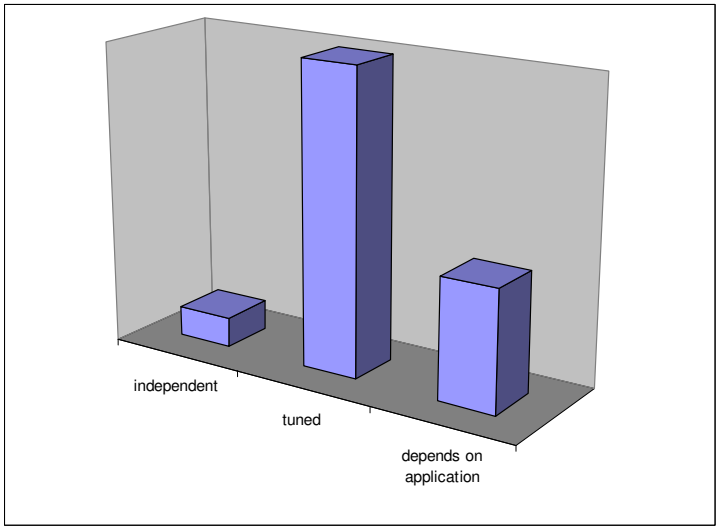
4. Compared to a traditionally induced model, a LeGo-based global model will be… (*multiple answers possible*)

A. … better, because of the exploratory (and often complete) search in the Pattern Discovery phase.
B. … worse because of the loss of information in binary patterns/features.
C. … worse because of the risk of overfitting and multiple testing implicit in the Pattern Discovery phase.
D. … better, but only in complex and structured domains, such as multi-relational and graphical.



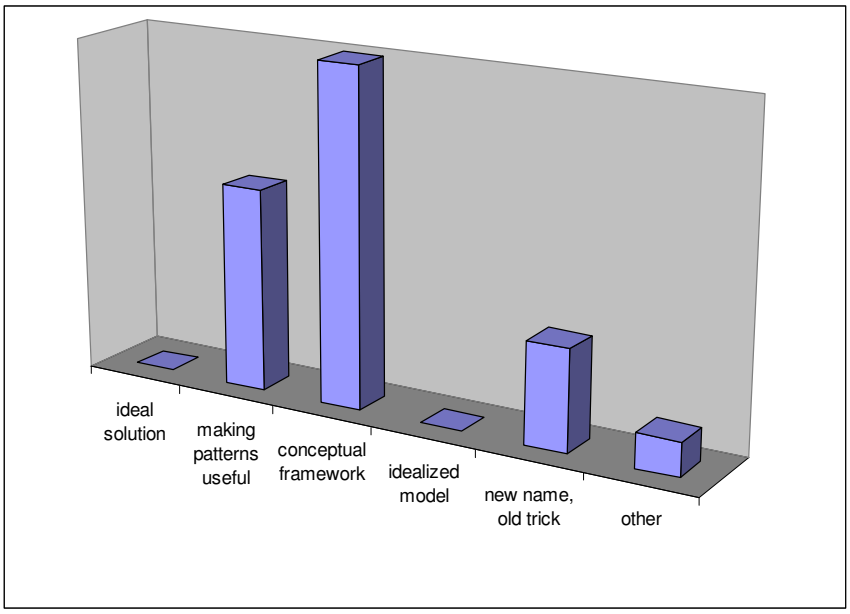
5. The workings and parameters of the three LeGo phases can best be…

A. … treated as independent. Greater flexibility allows for different algorithms to be plugged in.
B. … tuned to each other. By adapting the settings to the previous and next phase, optimal results can be achieved.
C. It depends on the application.



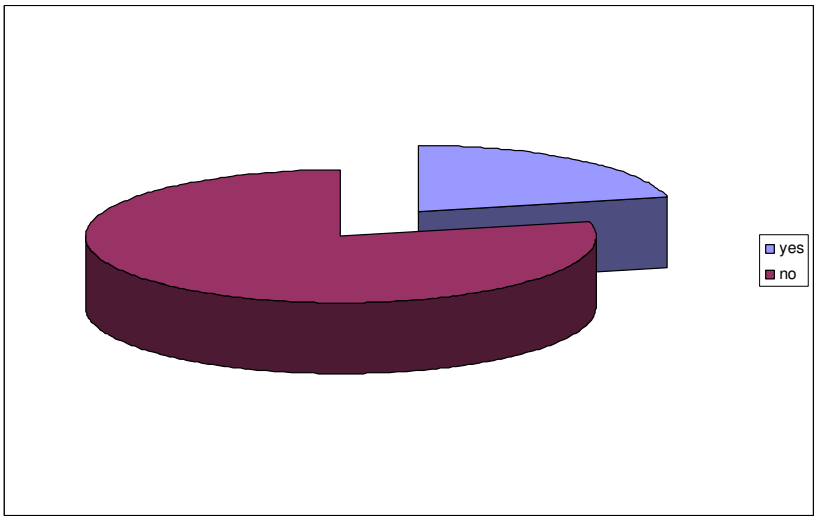
6. I view LeGo as… (*multiple answers possible*)

A. … the ideal solution for building good models
B. … a good approach for doing something useful with the patterns I have found.
C. … a conceptual framework for thinking and talking about different mining and modeling steps.
D. … an idealized and overly simplified model that doesn't really fit the complex applications I work with.
E. … a new name for an old trick.
F. Other. (*please specify*)



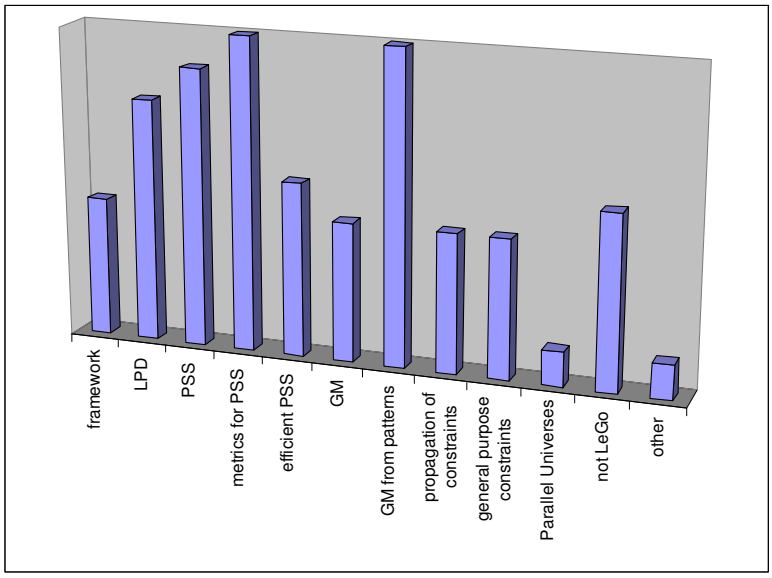
7. The so-called *covering-approach* (a.k.a. Separate & Conquer) is not an instance of Lego, as the patterns are not all discovered a priori.

A. I agree.
B. I disagree.



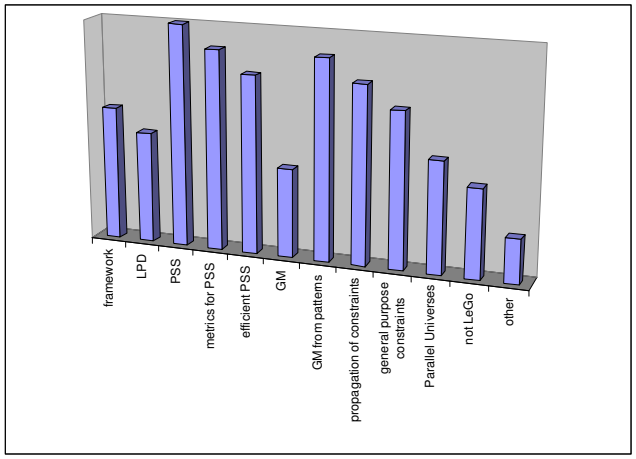
8. I will be working on… (*multiple answers possible*)

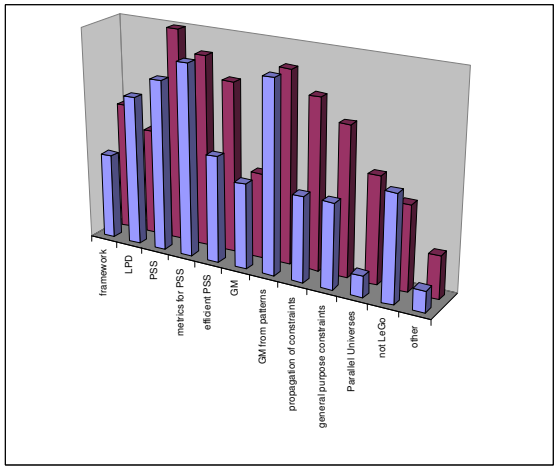
A. A framework for global modeling from local patterns.
B. Local pattern discovery.
C. Pattern subset selection.
D. Metrics for pattern subset selection.
E. Efficient implementations of pattern subset selection.
F. Global modeling.
G. Global modeling from sets of patterns.
H. Propagation of constraints: translating global constraints on the model into constraints of the local patterns.
I. General purpose constraints: finding local patterns that are useful for a variety of global models.
J. Parallel Universes.
K. Not-LeGo related subjects.
L. Other. (*please specify*)



9. We should be working on… (*multiple answers possible*)

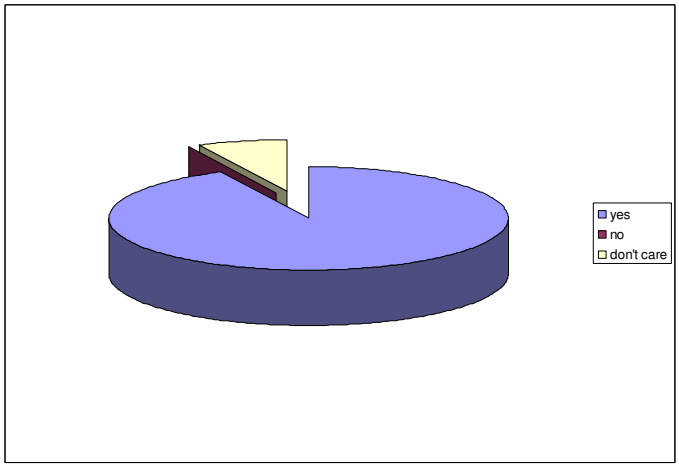
A. A framework for global modeling from local patterns.
B. Local pattern discovery.
C. Pattern subset selection.
D. Metrics for pattern subset selection.
E. Efficient implementations of pattern subset selection.
F. Global modeling.
G. Global modeling from sets of patterns.
H. Propagation of constraints: translating global constraints on the model into constraints of the local patterns.
I. General purpose constraints: finding local patterns that are useful for a variety of global models.
J. Parallel Universes.
K. Not-LeGo related subjects.
L. Other. (*please specify*)





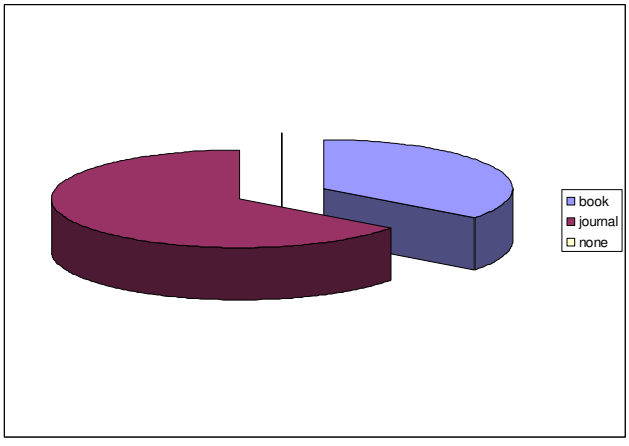
10. There should again be a LeGo workshop next year.

A. I agree. There is still a lot of research to be done in this area.
B. I disagree. Most of the problems have been solved.
C. I don't care either way.



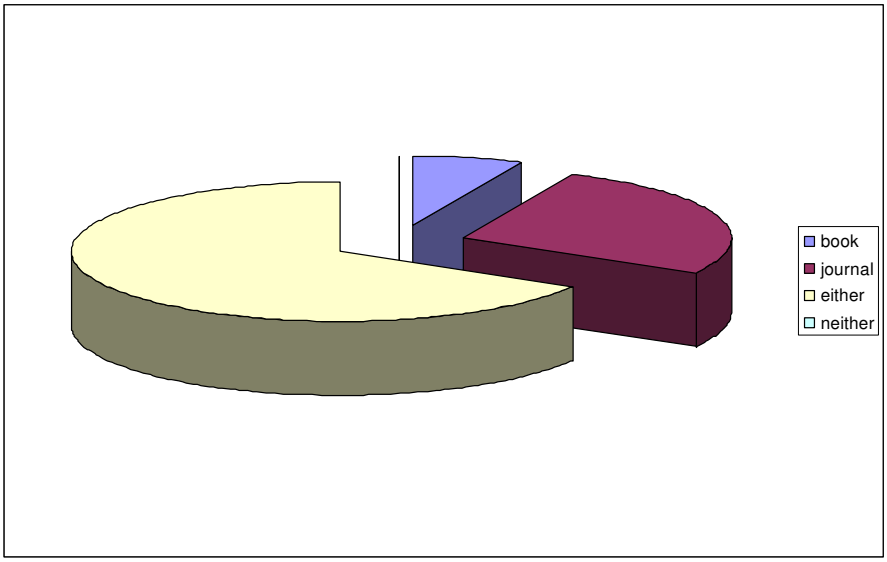
11. I think it is a good time for a publication on this subject (not necessarily limited to the workshop results).

A. Yes, an edited volume.
B. Yes, a journal special issue.
C. No.



12. I would be interested in contributing to a publication on this subject.

A. An edited volume.
B. A journal special issue.
C. Either.
D. Neither.



Interesting Patterns

Below are a number of association rules that were found by the Safarii Data Mining system. In the spirit of the LeGo framework, we have selected these rules from a larger collection of rules.

We should work on efficient PSS → LeGo model is better (exploratory)
(coverage 58%, lift 75% to 100%)

I will work on not-LeGo related → LeGo model is worse (overfitting)
(coverage 33%, lift 33% to 75%)

We should not be working on a framework → LeGo: putting patterns to use
(coverage 58%, lift 42% to 71%)

I will work on general-purpose constraints → LeGo is a conceptual framework
(coverage 33%, lift 67% to 100%)

I will work on global modeling from patterns → I will work on LPD
(coverage 58%, lift 50% to 85%)