

INTERACTIVE LEARNING OF PATTERN RANKINGS

Vladimir Dzyuba, Matthijs van Leeuwen, Siegfried Nijssen, Luc De Raedt

Pattern mining

Pattern mining is concerned with discovering *interesting* regularities in data, with a strong emphasis on comprehensible *descriptions*.

It can be seen as an automated data-driven *generation* of hypotheses. Algorithmically, it is a combinatorial search problem.

Description attributes								Target
A1	A2							
F	3							F
F	4							T
T	10							T
T	12							T
T	11							T
T	7							F
F	4							F

Description: $A1=T$ Size = 4 data records
Quality: $P(\text{Target}=T | A1=T) = 3/4 > P(\text{Target}=T) = 4/7$

Standard algorithms are hard to use for domain experts

Obtaining genuinely interesting results is challenging in practice. For example, many algorithms require apriori selection of a pattern quality measure, which is a non-trivial task even for data mining experts.

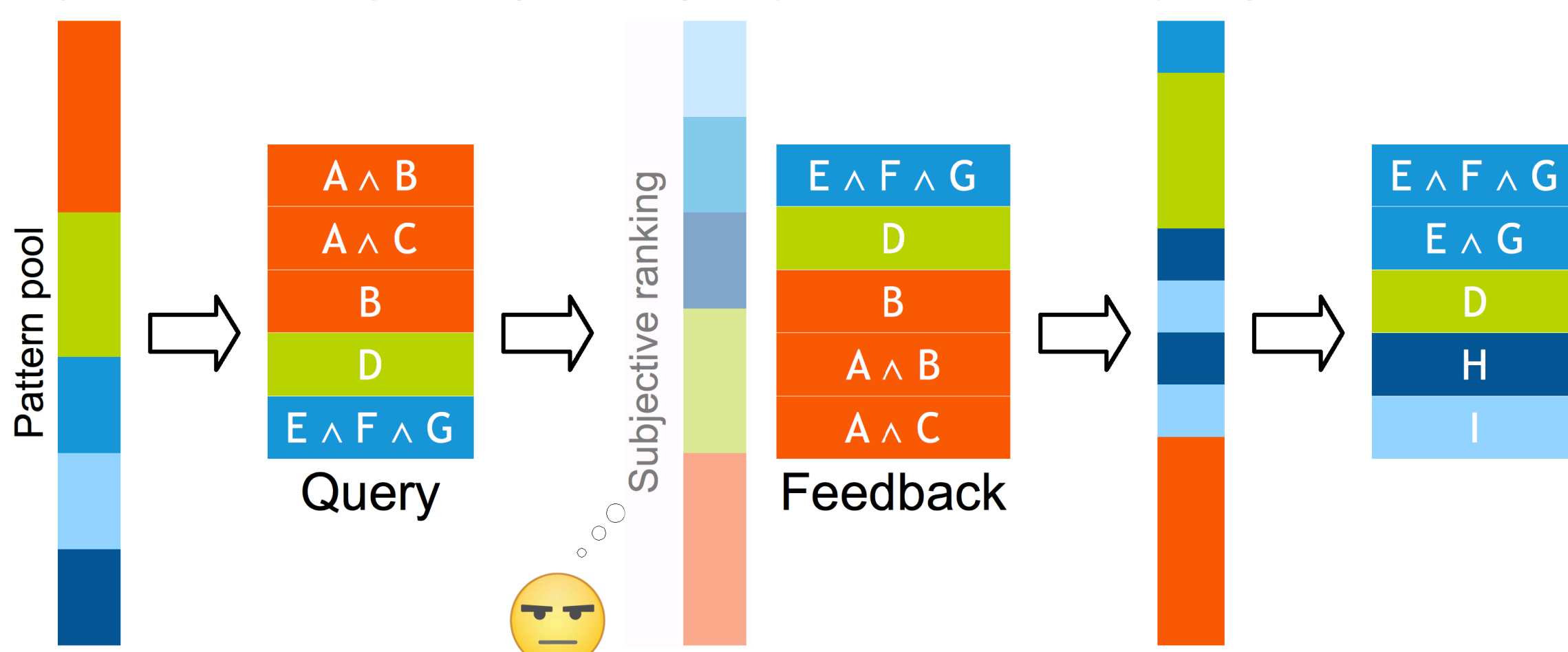
Few solutions are available, e.g. manual post-processing of results or parameter tuning, and they require considerable effort and/or data mining expertise.

Description	Cover	Size	Sensitivity	Specificity
$A1 = T$	3+ 1—	4	3 / 4	2 / 3
$A2 > 5$	3+ 1—	4	3 / 4	2 / 3
$A1 = T \wedge A2 > 9$	3+ 0—	3	3 / 4	1
$A1 = F$	1+ 2—	3	1 / 4	1 / 3
$A2 < 9$	1+ 3—	4	1 / 4	0

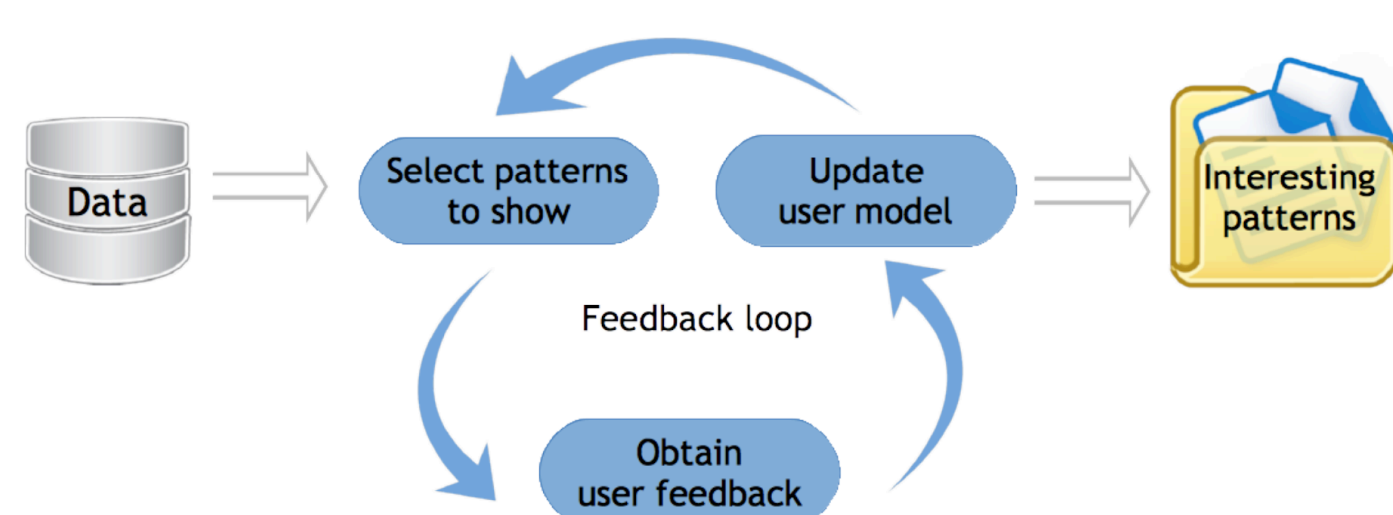
Objective quality measures might not capture subjective interestingness

Proposed solution: User perspective

Key idea: learn subjective pattern quality measures from easy-to-provide feedback.



Mine, Interact, Learn, Repeat

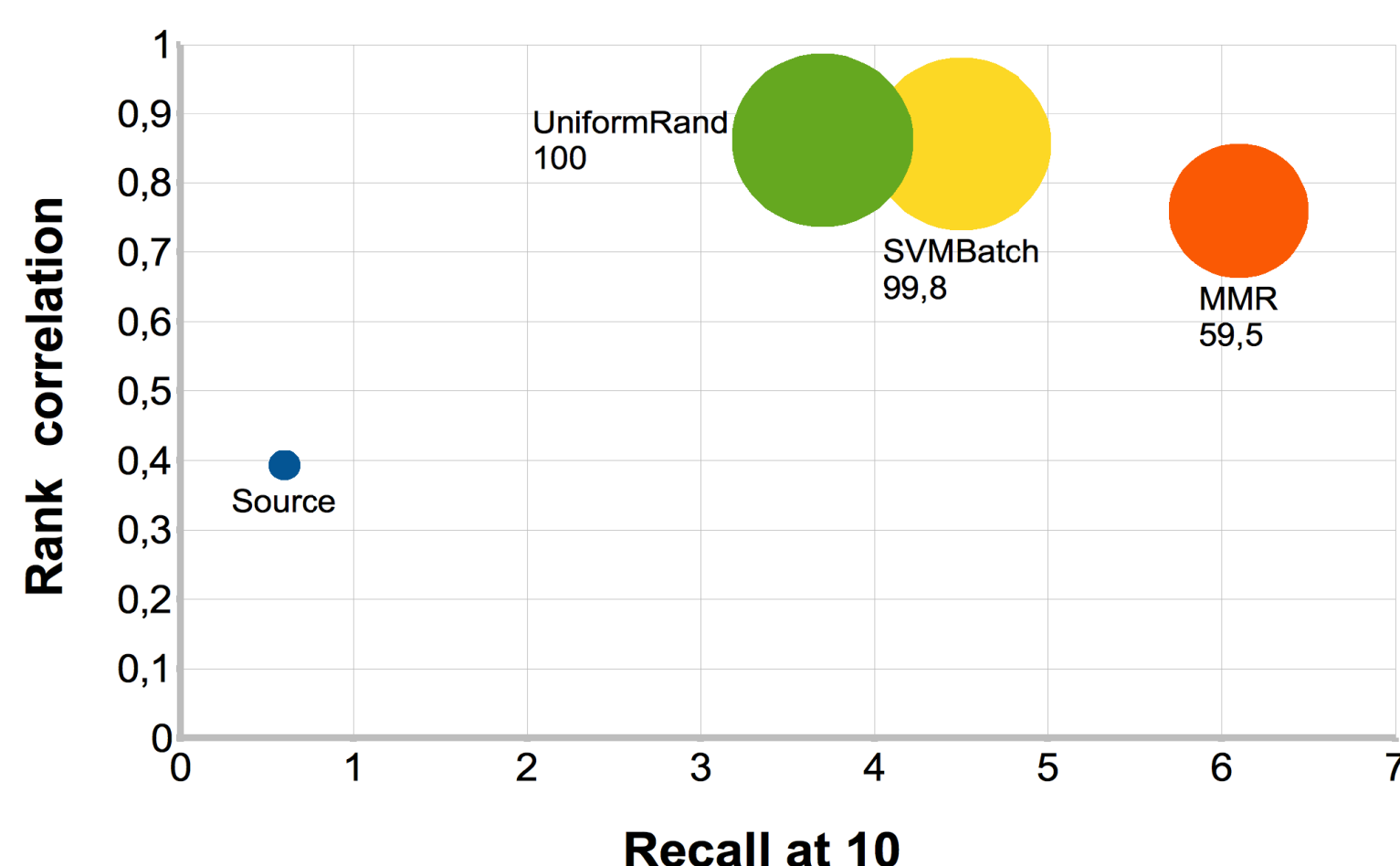


Research questions:

- Which feedback format to use?
- How to model & learn user knowledge/interests?
- Which patterns to show (to *query*)?

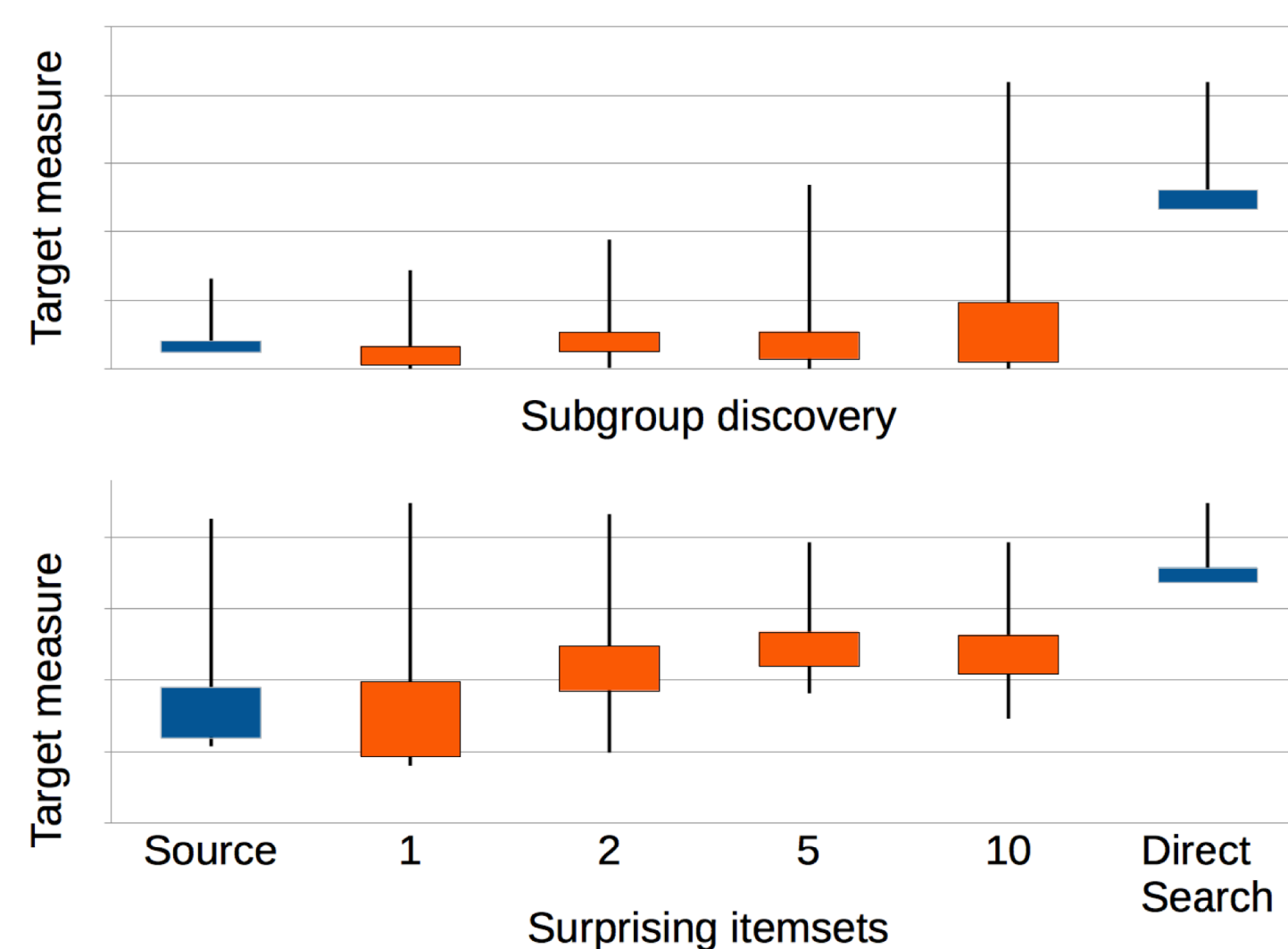
Towards interactive discovery of interesting patterns: Active preference learning of pattern ranking functions

- Ordered feedback: $\{A, B, C\} \rightarrow B > A > C$
- *Preference learning* for feedback generalization
- *Active learning* heuristics aim at minimizing user effort



Trade-off between active learning performance and user effort

In experiments, subjective user feedback was emulated using an objective quality measures, which were unknown to the learner, and queries were selected from pre-mined sets of patterns.



Generalization heavily depends on the source pattern set.

Future work

Selecting queries from a pre-mined set of patterns is restrictive. An obvious alternative is to mine queries directly from the pattern space. We intend to employ *pattern sampling* to obtain diverse and representative pattern sets for queries: we plan to build upon recent research on constraint programming for itemset mining and combine it with *SampleSearch*, a technique to sample solutions of CSPs.