Scalable and Repeatable Extrinsic Evaluation for Pattern Discovery Systems

Mario Boley, Maike Krause-Traudes, Bo Kang, Björn Jacobs
University of Bonn & Fraunhofer IAIS
mario@realKD.org
Recently at Q&A time...

Q: This looks interesting, but is this really what *users* would want?

A: Well, I guess in order to really confirm that, we would need to test this *somehow* with real users.

Q: Yep, agreed. Thank you.
Extrinsic evaluation can support ultimate value of contributions

Photograph courtesy Dorothy Fragaszy (voices.nationalgeographic.com)
Extrinsic means: “not depending on theory used for development cycle”
Poll among ECMLPKDD authors: *half* skipped potentially useful studies

High costs are dominant reason for skipping on “study opportunity”

- No added benefit: 5%
- Unclear how to recruit participants: 55%
- High costs of conducting study: 98.33%
- Insecurity of outcome and acceptance: 15%

% of “yes”-respondents
High costs are dominant reason for skipping on “study opportunity”

- No added benefit of user study over automatized/formal evaluation: 5%
- Unclear how to recruit suitable group of participants: 55%
- Cost of developing study design: 46.6%
- Cost of embedding contribution in accessible UI: 40%
- Cost of organizing actual study: 63.3%
- Cost of evaluating results: 15%
- Insecurity of outcome and acceptance by peers: 15%
Creedo’s major contributions are...

• Allows definition of **reusable study designs**

• Elements focus on **scalable evaluation in application context**

• **Automatizes** process
A study is a process for providing evidence in favor or against...

Hypothesis:

“Users can solve a certain class of analysis tasks better with a specific target system than with other control systems.”
A study is a process for providing evidence in favor or against...

Hypothesis:

“Users can solve a certain class of analysis tasks better with a specific target system than with other control systems.”

Example:

“Users can discover a set of interesting patterns faster using a FORSIED-based association discovery process than when using a conventional* association discovery process.”

*based on a static interestingness measure that is oblivious to prior and gained knowledge
Data analysis systems are represented by Creedo analytics dashboards.
Algorithms can be integrated via the realKD library
Creedo *tasks* bridge formal abstraction and application context

1. Introduction

In this paper, we tackle the important problem of discovering interesting patterns from a given input dataset.

...
Creedo *tasks* bridge formal abstraction and application context
Creedo *tasks* bridge formal abstraction and application context
User perspective on task are natural language instructions

Analysis task instructions

As a new employee of the Data Science Department of the government of The Plain, you have to get familiar with the socio-economics status of your country. Go on and use our data mining tool to discover key phenomena from The Plain’s socio-economic data. The data consists of the socio-economic records of 1000 representative inhabitant samples of your country.

The data mining tool will propose statements about the data, and measures associativity among the statements. Such information is summarized in graphic representations (patterns) like the figure below:

- Sec.1 states whether the statements are positively or negatively associated.
- Sec.2 lists the considered statements, along with the frequency (proportion of inhabitants) that each individual statement holds true.
- Sec.3 visualizes the difference between the expected frequency (blue bar) and the actual frequency (red bar) of the statements. The larger the difference the stronger the positive/negative association is.
Creedo *tasks* bridge formal abstraction and application context
Creedo *tasks* bridge formal abstraction and application context
Creedo *tasks* bridge formal abstraction and application context
Task also defines elementary attributes of results.
All measurements can be aggregated to *system performance measures*.
All measurements can be aggregated to *system performance measures*
All measurements can be aggregated to *system performance measures*.
All measurements can be aggregated to *system performance measures*

\[ a \rightarrow \text{avg}\{t(x): \hat{c}(x) \geq \tau, x \in R_a\} \]
Assignment logic can control biases and balance confidence
Assignment logic can control biases and balance confidence.
Assignment logic can control biases and balance confidence
Creedo organizes study process

**Design Time**

- Study Owner defines System Variants + Analysis Tasks + Performance Metrics + Assignment Logic

**Execution Time**

- Creedo assigns Participant to work on Tasks using Analysis System producing Results
- Creedo assigns Evaluator to evaluate Result using Metric producing Result Evaluation

**Conclusion Time**

- Creedo aggregates Result Evaluations and presents Extrinsic System Evaluations to Study Owner
Yes, we can

mario@realKD.org