

Leveraging Multi-Method Evaluation for Multi-Stakeholder Settings

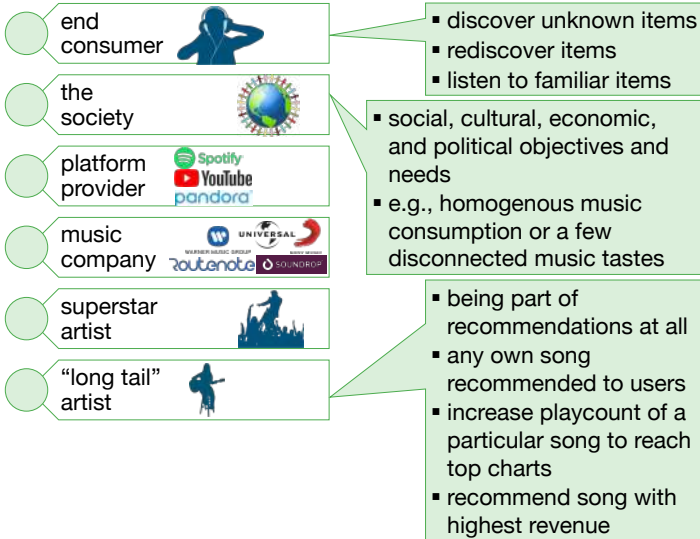
Rationale

There are always multiple stakeholders involved in recommendation settings.

- always multiple – possibly diverging – perspectives and goals
- all stakeholders' need to be considered in RS evaluation
- multiple evaluation methods and criteria have to be combined!

Example: Recommender systems in the digital music ecosystem

Various stakeholders' goals and preferences



Blind spots in single method evaluation



Offline evaluation

It shows that users' historic listening behavior can be simulated (e.g., high accuracy).

- Does the user want to listen to these familiar songs in future?
- Would the user be satisfied with the same number/proportion of unfamiliar songs?
- Is the user interested in discovering (more) new songs?
- ...

Online evaluation

It shows that users click or skip recommended songs; or stay on platform for longer/shorter than usually.

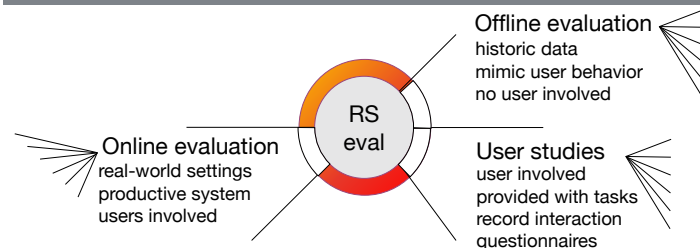
- Does the user want to listen to the recommended songs in future?
- Is the user is satisfied with the number/proportion of unfamiliar songs recommended?
 - wants more discovery
 - skipped songs did not meet preferences
 - not in the mood for unfamiliar songs
- Is the user interested in discovering (more) new songs?
- ...



Multi-method evaluation

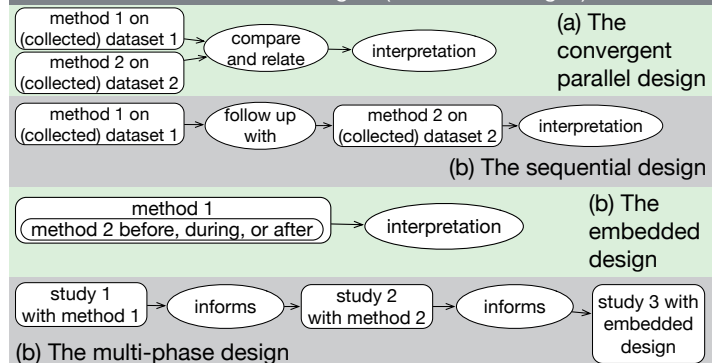
- Goal: integrated big picture of RS performance
- Combine several (quant. and/or qualitative) evaluation methods
 - To capture the same phenomenon from different angles
 - To capture diverse, but complementary phenomena
 - To resolve conflicting findings

Tradition of recommender systems evaluation



- Focus on one single perspective
- Incomplete picture: blind spots
- Small set of metrics; mostly from one perspective only
- Evaluation results may differ, e.g.,
 - user satisfaction does not always correlate with high recommender accuracy
 - offline evaluations of accuracy are not always meaningful for predicting relative performance of different techniques

Multi-method strategies (evaluation designs)



John W Creswell and Vicki L. Plano Clark. 2011. Designing and conducting mixed methods research. Sage Publications, Los Angeles, CA, USA.

Open Questions – Please discuss with us!

- Let's get concrete!
 - How can we implement multi-evaluations in practice?
 - What are the best practices?