

Effects of Position Bias on Click-Based Recommender Evaluation

Katja Hofmann

Microsoft Research

Anne Schuth

University of Amsterdam

Alejandro Bellogin

Universidad Autónoma de Madrid
(work completed while at CWI)

Maarten de Rijke

University of Amsterdam

Overview

Evaluation of recommender systems has long focused on *explicit* feedback, where users rate items.

Much more data can be collected when considering *implicit* feedback, such as clicks.

What is the relationship between using explicit and implicit feedback for recommender system evaluation ?

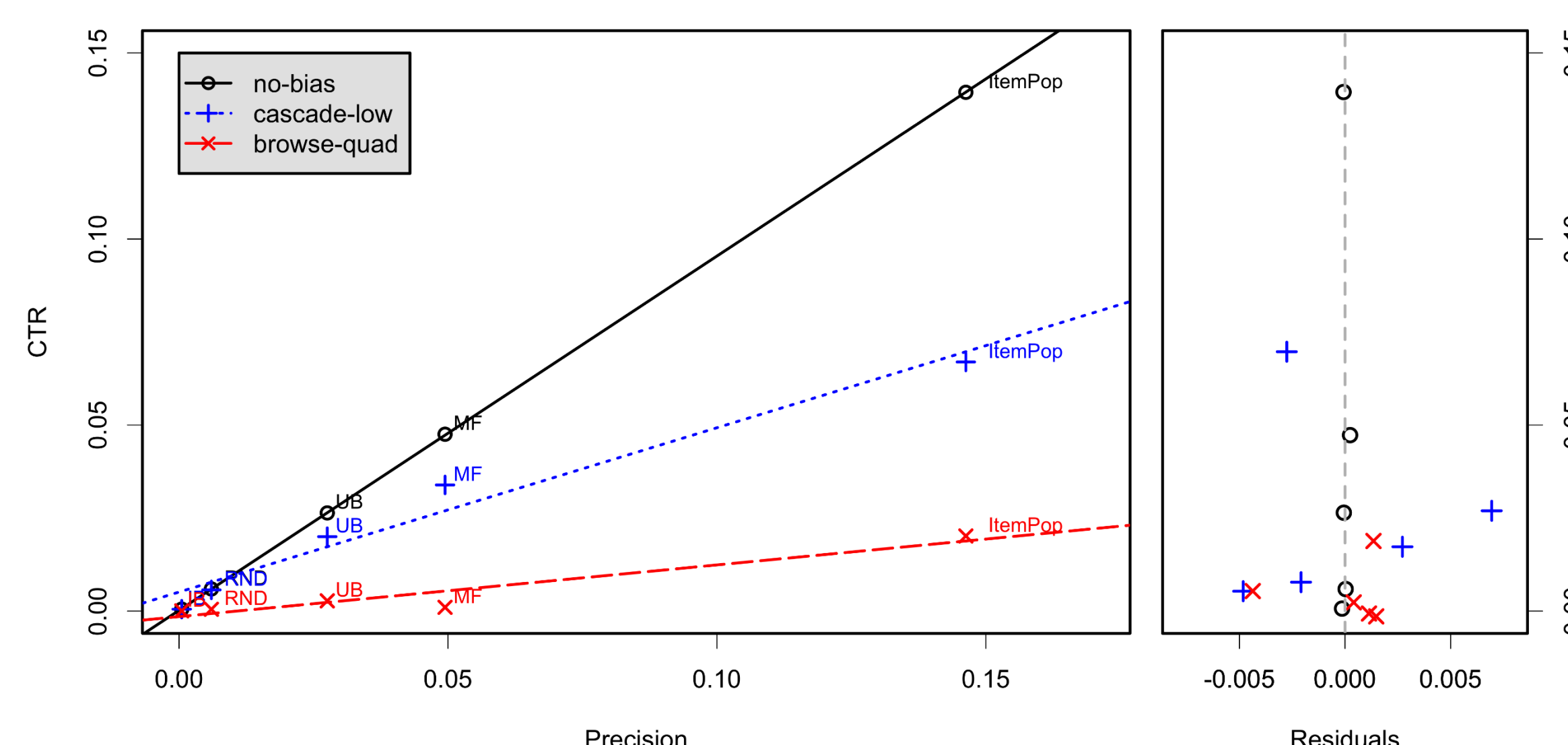
Experiments

- Given evaluation with explicit feedback, predict online performance (measured using implicit feedback).
- Compare recommender systems: random baseline (RND), popularity-based (ItemPop), item-based (IB), matrix factorization (MF), user-based (UB).
- Compare evaluation with explicit and implicit feedback:
 - explicit feedback** – nDCG and Precision on explicit labels
 - implicit feedback** – clickthrough rate (CTR) with user models: no-bias, examination and browsing models with logarithmic and quadratic discount (exam- $\{\log, quad\}$, browse- $\{\log, quad\}$), and cascade model with low and high stop probability (cascade- $\{\text{low}, \text{high}\}$).
- Tasks: (1) rank recommender systems and check agreement; (2) predict “online performance” (based on implicit feedback) from evaluation with explicit ratings on MovieLens 1M data (with 80-20 test / train split).

Results

- Implicit and explicit evaluation agree well when assumptions agree well (e.g., precision@10 and CTR with no-bias).
- Match between assumption on user behavior and explicit evaluation matters – if assumptions are violated, the wrong recommender system can be preferred.

Overview of results. Precision and CTR scores for the tested RS are shown for selected user models.



Open questions

- How to use explicit labels for RS development before implicit feedback can be collected?
- How to obtain guarantees on performance with implicit labels, given offline evaluation with explicit labels?
- How to deal with missing explicit labels?

Interfaces for collecting explicit and implicit feedback for different RS.

