open.michigan

Author(s): Rahul Sami, 2009

Citation Key

for more information see: http://open.umich.edu/wiki/CitationPolicy

Use + Share + Adapt

{ Content the copyright holder, author, or law permits you to use, share and adapt. }

PD-GOV **Public Domain – Government**: Works that are produced by the U.S. Government. (USC 17 § 105)

PD-EXP **Public Domain – Expired**: Works that are no longer protected due to an expired copyright term.

PD-SELF **Public Domain – Self Dedicated**: Works that a copyright holder has dedicated to the public domain.

ZERO **Creative Commons – Zero Waiver**

BY **Creative Commons – Attribution License**

BY-SA **Creative Commons – Attribution Share Alike License**

BY-NC **Creative Commons – Attribution Noncommercial License**

BY-NC-SA **Creative Commons – Attribution Noncommercial Share Alike License**

GNU-FDL **GNU – Free Documentation License**

Make Your Own Assessment

{ Content Open.Michigan believes can be used, shared, and adapted because it is ineligible for copyright. }

PD-INEL **Public Domain – Ineligible**: Works that are ineligible for copyright protection in the U.S. (USC 17 § 102(b)) *laws in your jurisdiction may differ

{ Content Open.Michigan has used under a Fair Use determination. }

FAIR USE **Fair Use**: Use of works that is determined to be Fair consistent with the U.S. Copyright Act. (USC 17 § 107) *laws in your jurisdiction may differ

Our determination **DOES NOT** mean that all uses of this 3rd-party content are Fair Uses and we **DO NOT** guarantee that your use of the content is Fair.

To use this content you should **do your own independent analysis** to determine whether or not your use will be Fair.

Lecture 13: Manipulation; Privacy

SI583: Recommender Systems

The Influence Limiter: Key Ideas

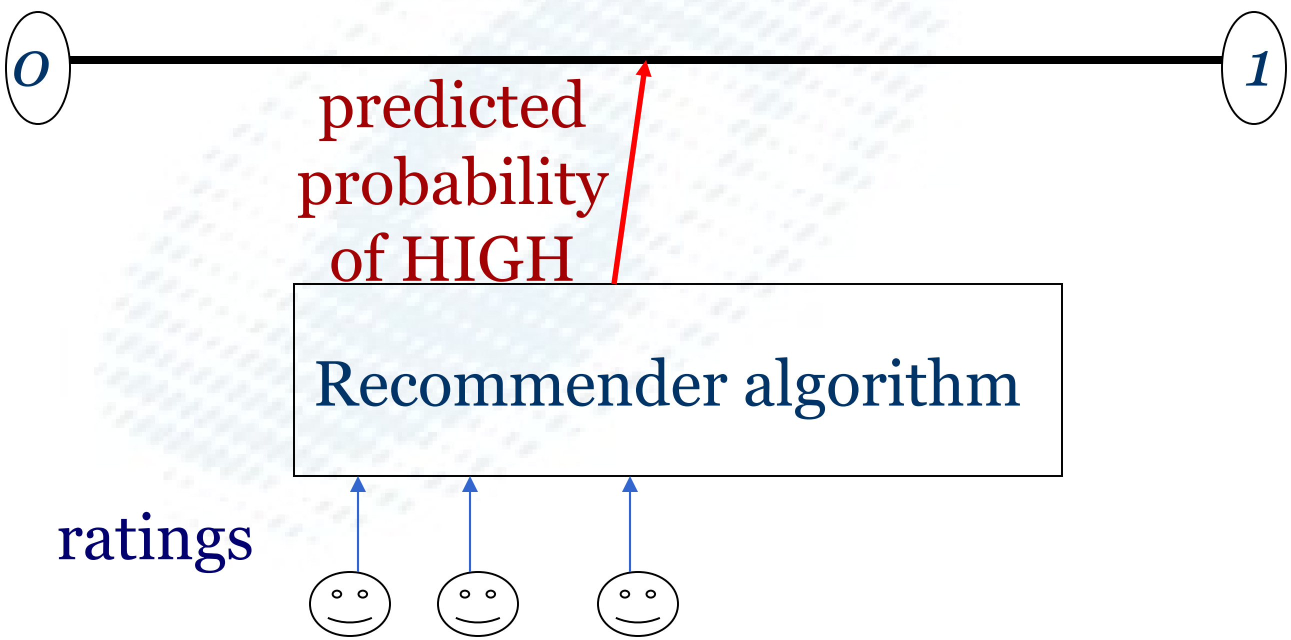
[Resnick and Sami, Proceedings of RecSys '07 conference]

- Limit *influence* until rater demonstrates *informativeness*
- *Informative* only if you're the first to provide the information

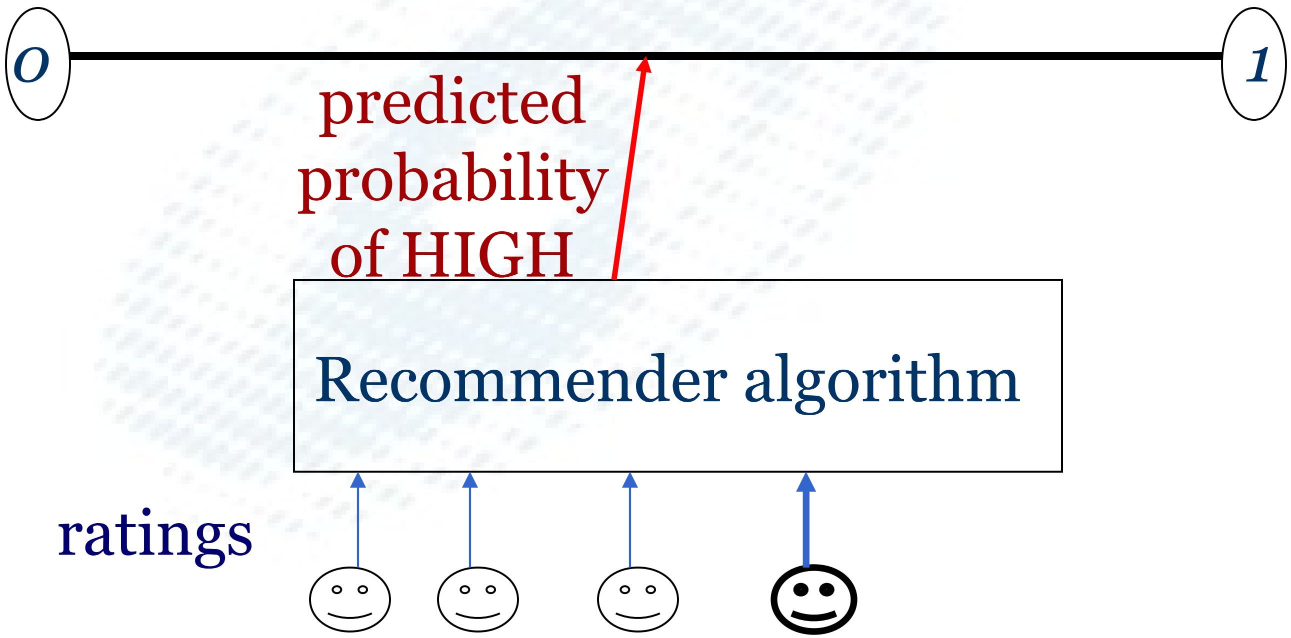
Results we *cannot* achieve

- Prevent any person J from manipulating the prediction on a single item X.
  - Cannot distinguish *deliberate manipulation* from *different tastes* on item X
- "Fairness", ie., two raters with identical information get exactly the same influence, regardless of rating order.
  - Cannot distinguish second rater with identical information from an informationless clone.

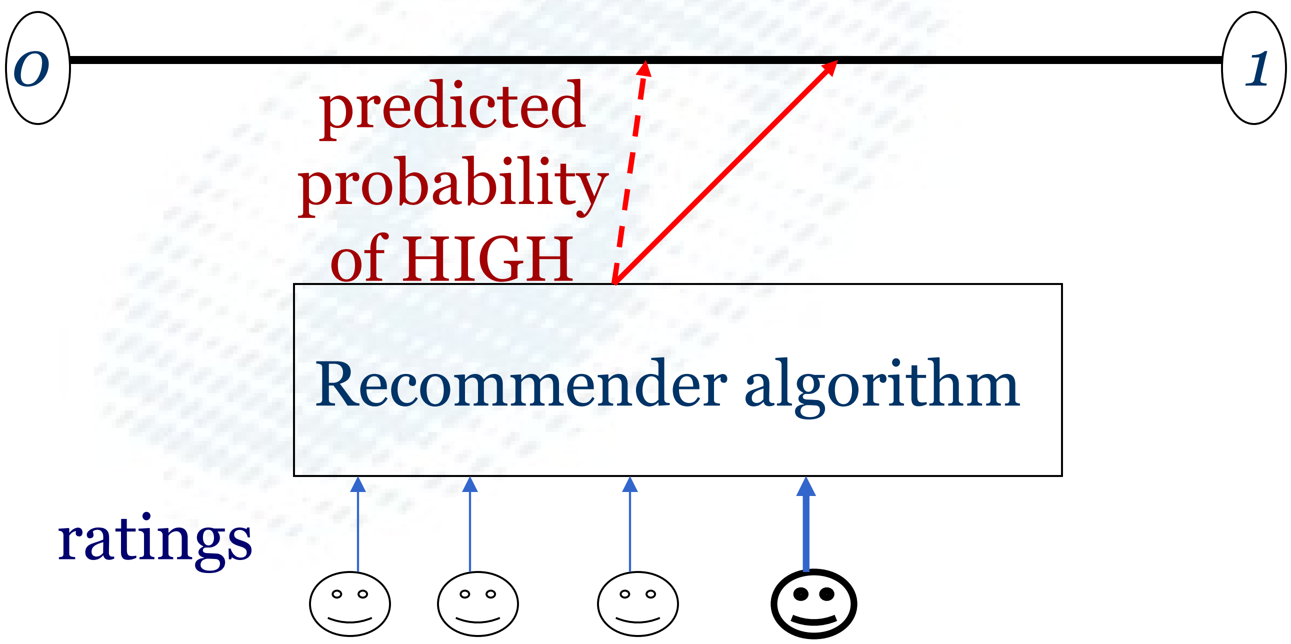
Predictions on an Item: A Dynamic View



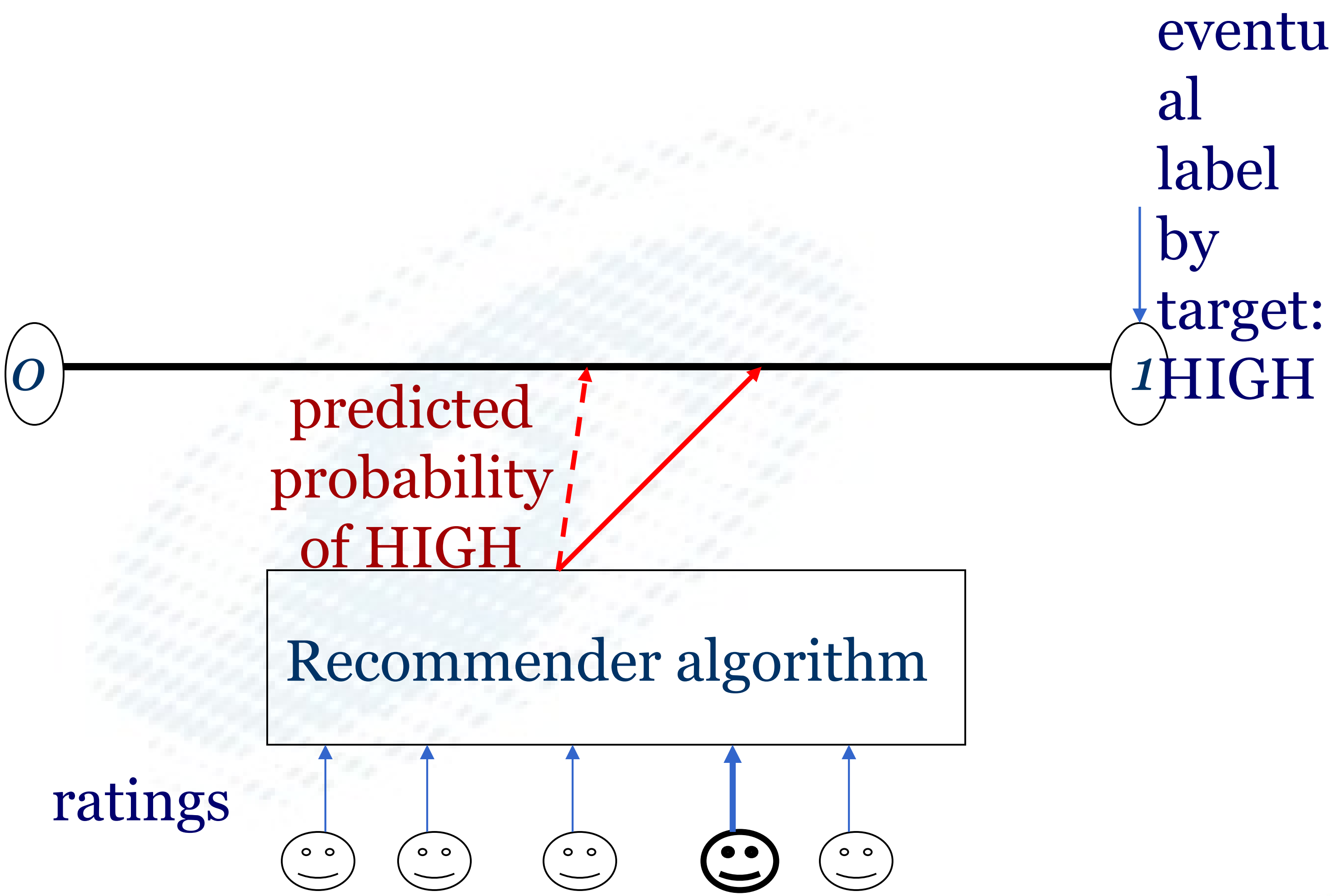
Predictions on an Item: A Dynamic View



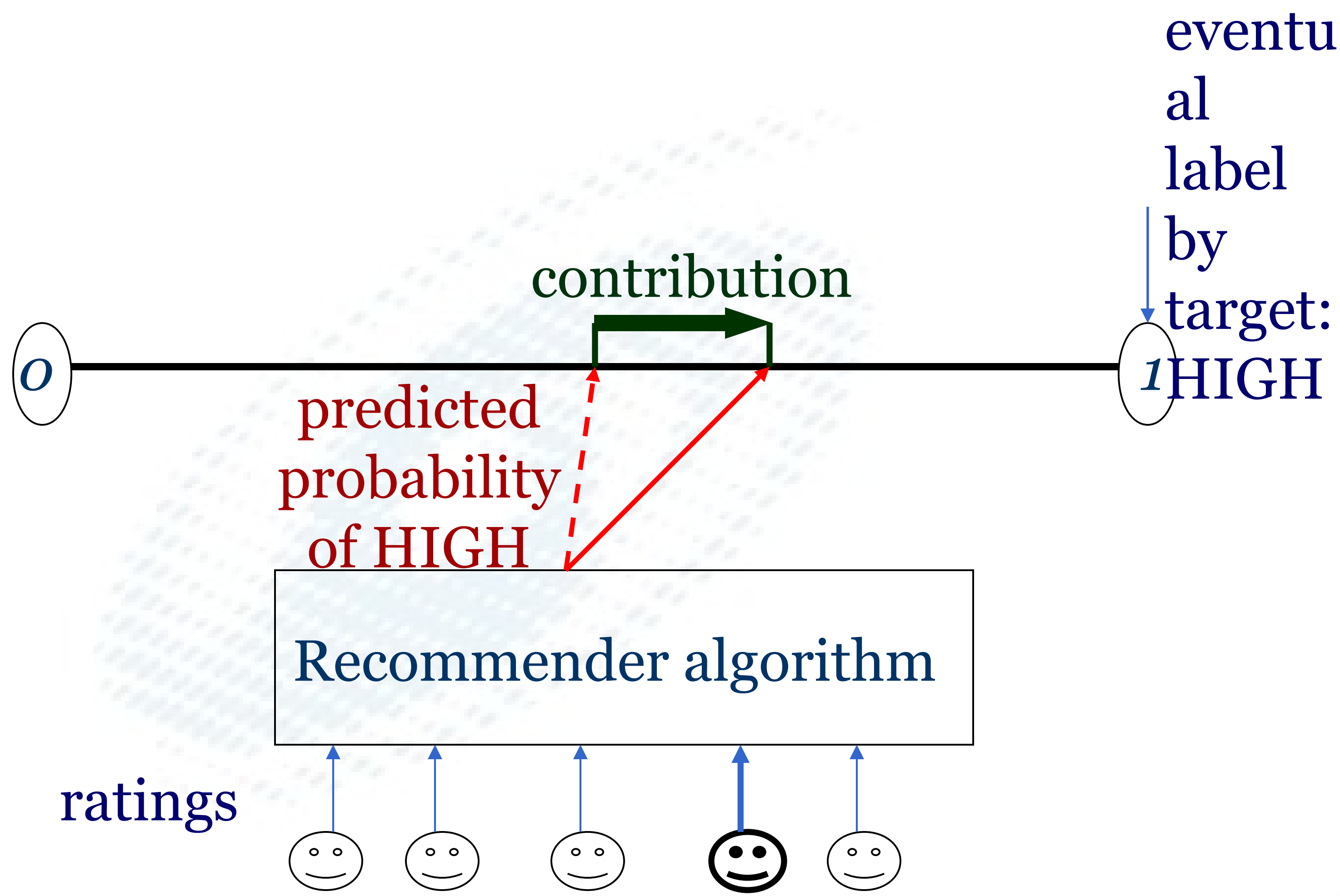
Predictions on an Item: A Dynamic View



Predictions on an Item: A Dynamic View

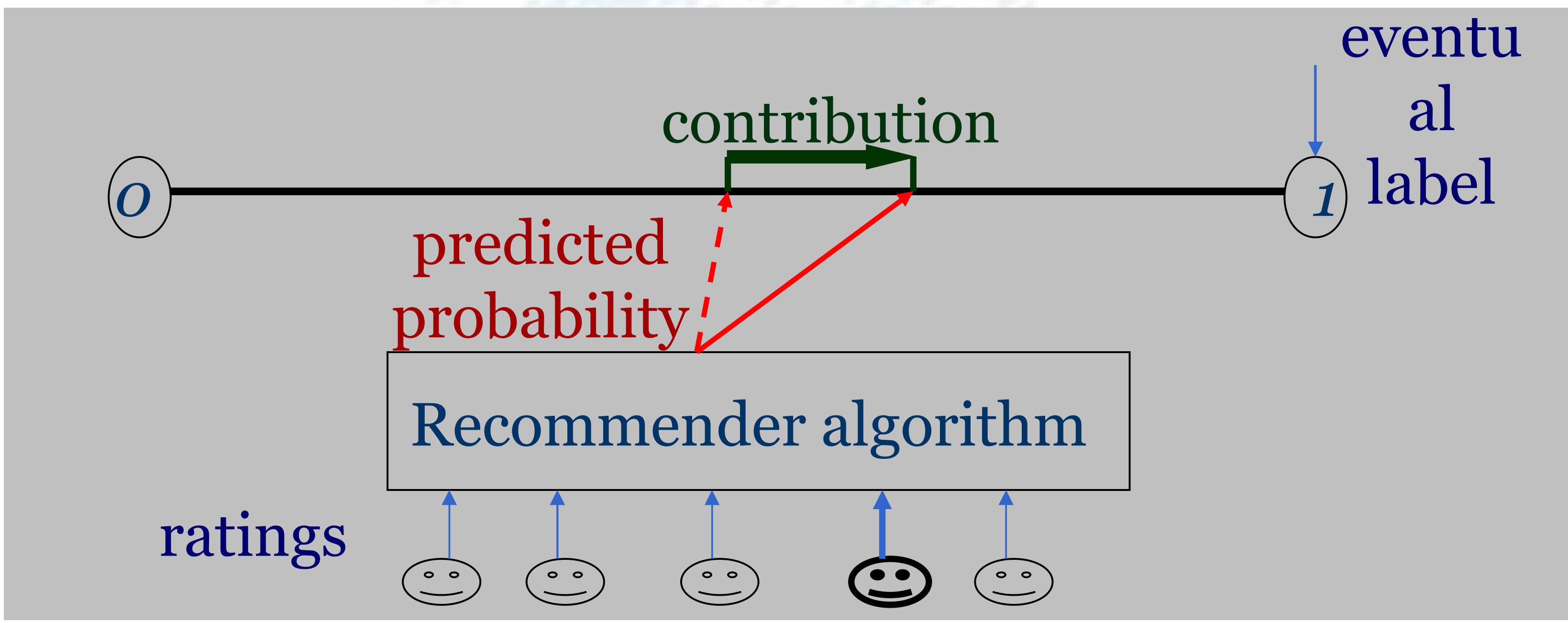


Predictions on an Item: A Dynamic View



Our approach

- Information-theoretic measure of contribution and damage
- Limit *influence* a rater can have had based on past *contribution*
- This limits *net damage* an attacker can cause



Our Model

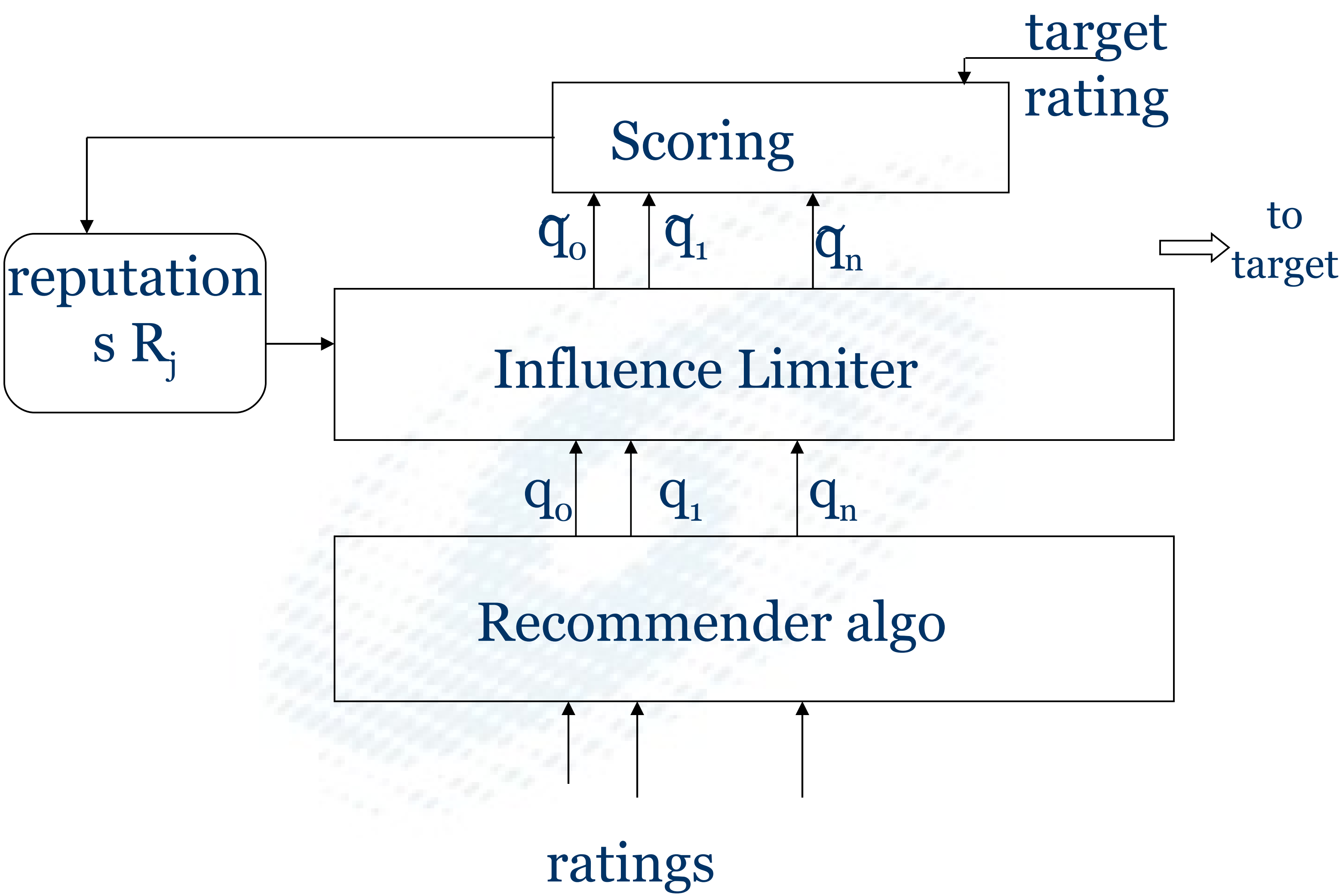
- Binary rating system (HIGH/LOW)
- Recommendations for a single target person
- Any recommender algorithm
- Powerful attackers:
  - Can create up to *n* sybil identities
  - Can “clone” existing rating profiles
- No assumptions on non-attackers:
  - Attacker’s sybils may form majority
  - Do not *depend* on honest raters countering attacks

Overview of Results

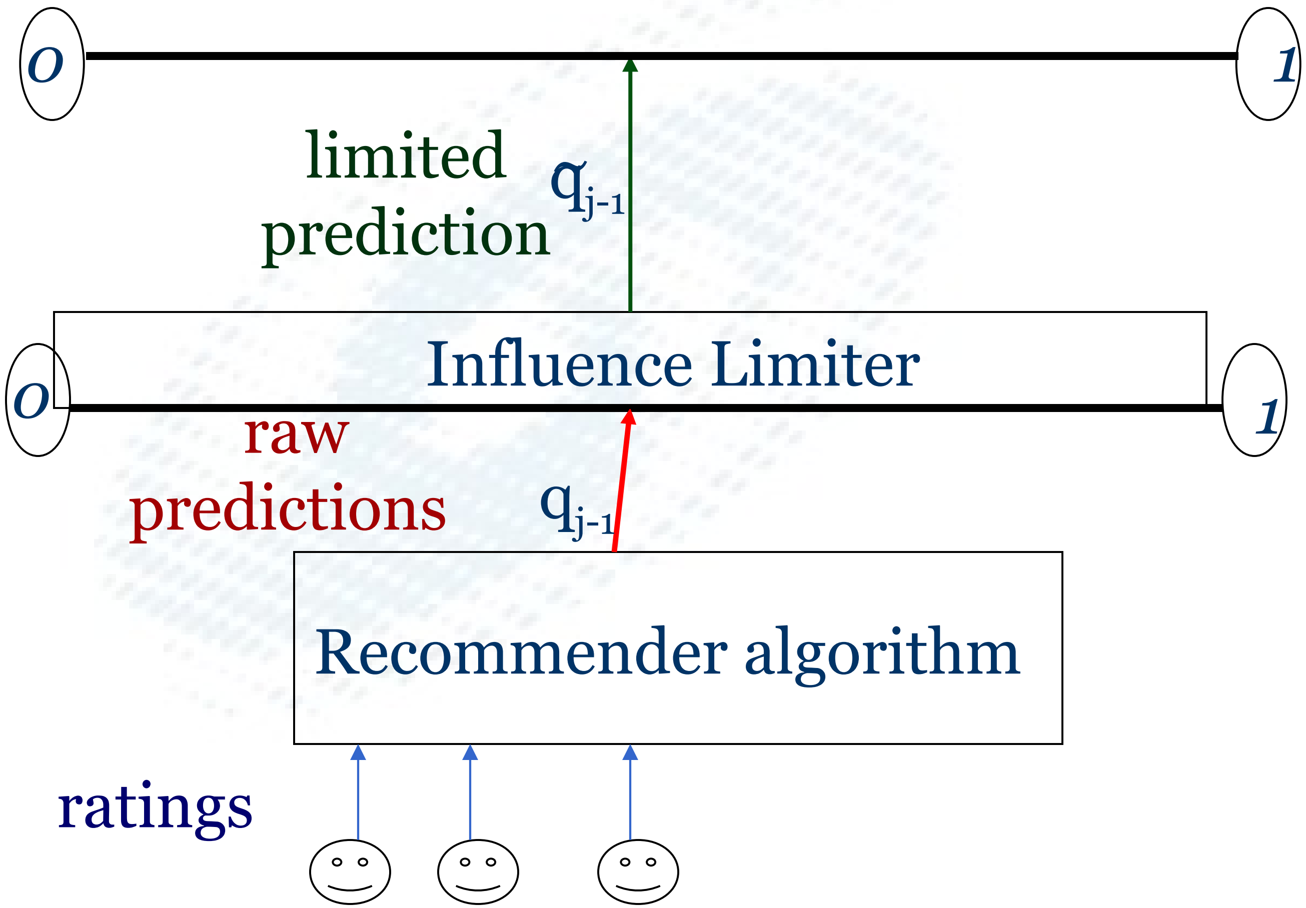
"Influence-limiter" algorithm can be overlaid on any recommender algorithm to satisfy (with caveats):

- **Limited damage**: An attacker with up to *n* sybils can never cause net total damage greater than *O(1)* units of prediction error
- **Bounded information loss**: In expectation, O(log *n)* units of information discarded from each genuine rater in total.

Influence Limiter: Architecture

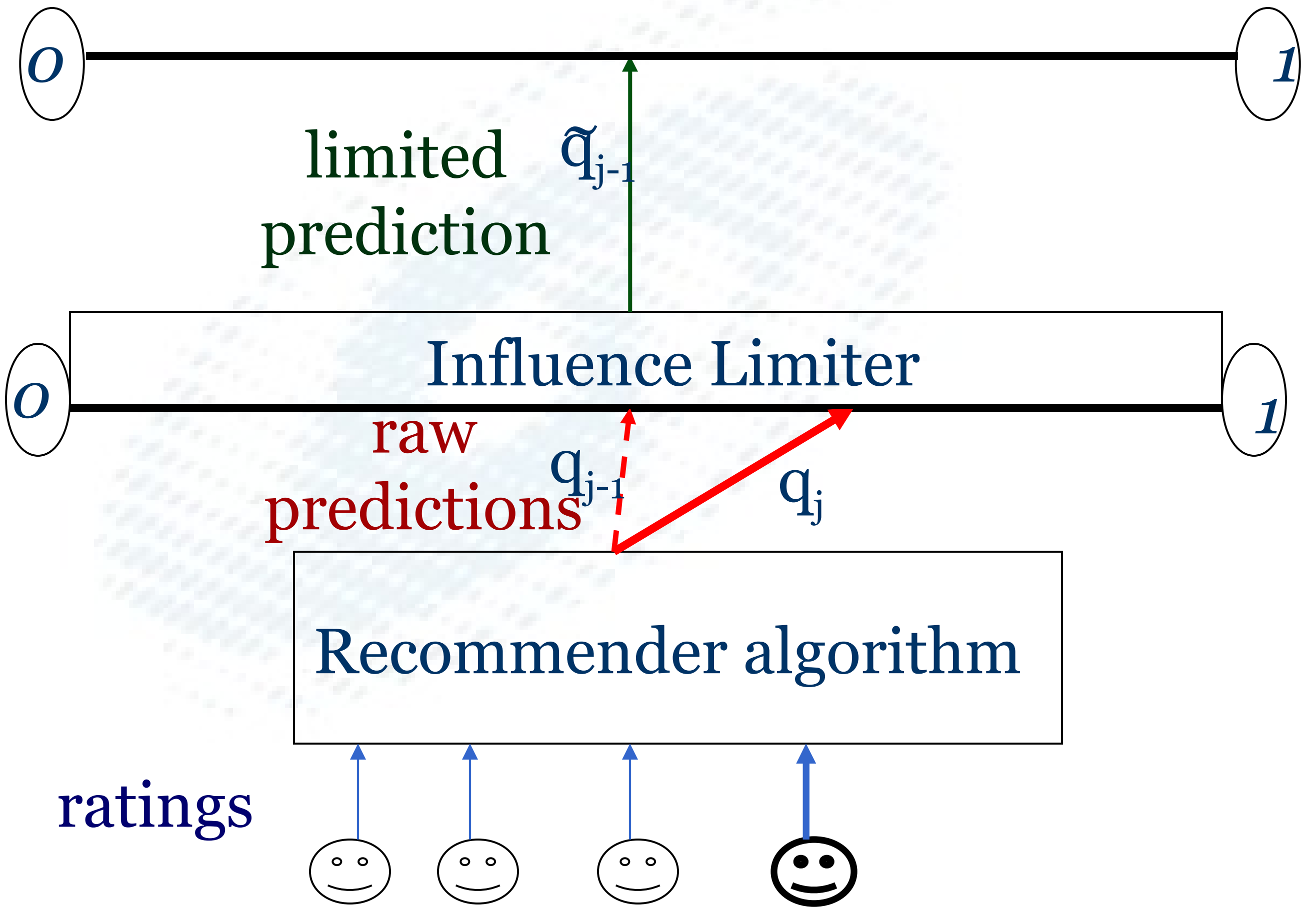


Influence Limiter Algorithm: Illustration



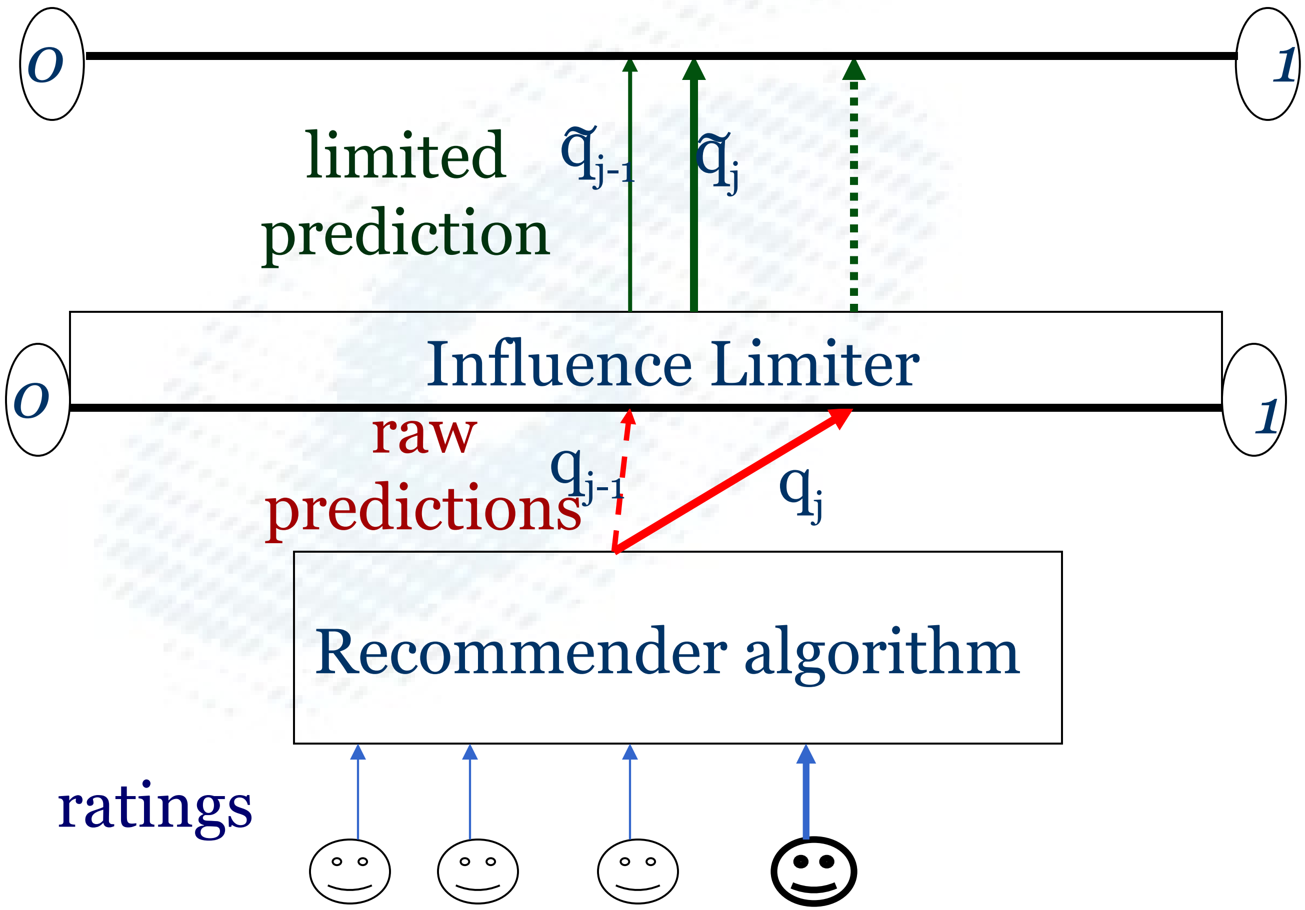
Influence Limiter Algorithm: Illustration

A rater with R=0.25 puts in a rating



Influence Limiter Algorithm: Illustration

A rater with R=0.25 puts in a rating



Manipulation: summary

- Increasingly important problem

- Range of techniques to defend:
  - Detecting and filtering attack profiles
  - Influence Limiter
  - Incentive schemes
  - Strong identity verificiation
  - Combinations of these methods

Privacy in Recommender Systems

- Privacy: your right to control dissemination of personally identifiable information

- Privacy hazards:
  - Monitoring behavior without user's consent
  - Persistent storage of information in cookies
  - Data leaks
  - Data leaks from *anonymized datasets*

Privacy-preserving CF [Canny]

- High-level idea: distributed computing of recommendations
  - User-specific information not available outside the user's computer
  - uses neat cryptographic protocols ("zero-knowledge" protocols) to compute an SVD

Review: Topics we have covered

- Eliciting ratings
- Using implicit ratings
- Collaborative Filtering methods
- Implementation/Architectures
- Evaluation of Recommenders
- Explanations; task-based evaluation
- Manipulation
- Privacy