TEL AVIV UNIVERSITY אוניברסיטת תל-אביב

Improved Recommendations via (More) Collaboration

Rubi Boim **Haim Kaplan** **Tova Milo** **Ronitt Rubinfeld**

Goal

Multi-organization / domain collaboration for improved recommendations

Difficulties

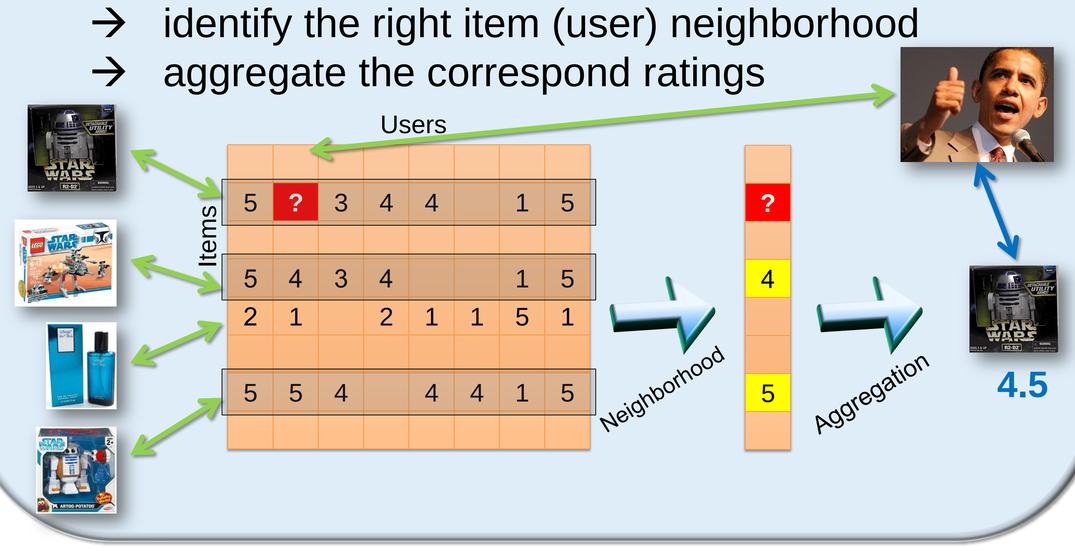
Impractical to centralize:
- → CF algorithms do not scale linearly
- → Constant updates
- → Privacy

Distributed environment:
- → Bandwidth (expensive to send the entire vector)

Standard CF (Centralized)

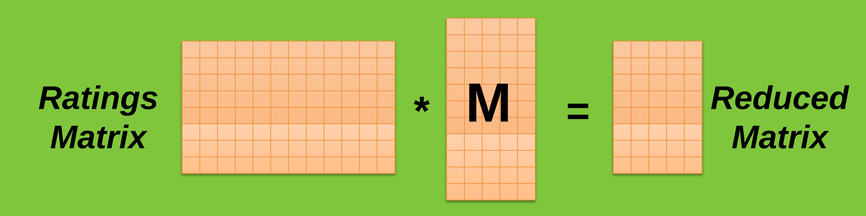
- → Item-based
- → User-based

Main idea:
- → identify the right item (user) neighborhood
- → aggregate the correspond ratings



Solutions

Dimension Reduction



Difficulties: standard reduction techniques (SVD, etc.) needs to see all data to compute **M**

- Reduction by 'random independent matrix' (*Ailon)*
- Can't be applied with "Shared Users" nor Pearson

Dimension Selection

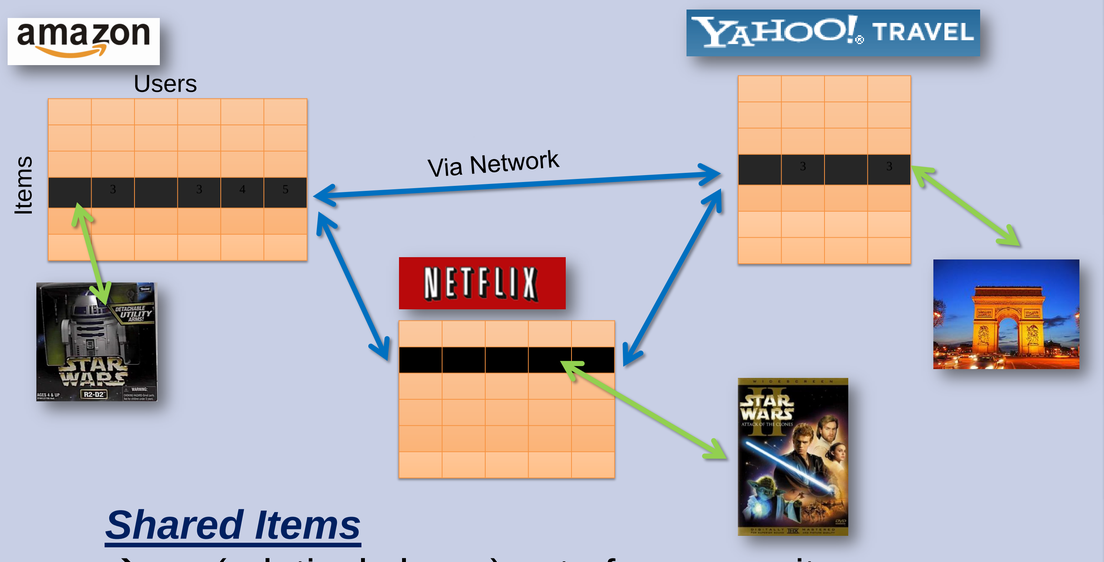
Manually choose the best representatives

- Proved to be NP-Complete
- We provide Greedy algorithm (no guarantees but impressive results!)
- Speeds up computation by over 50%!

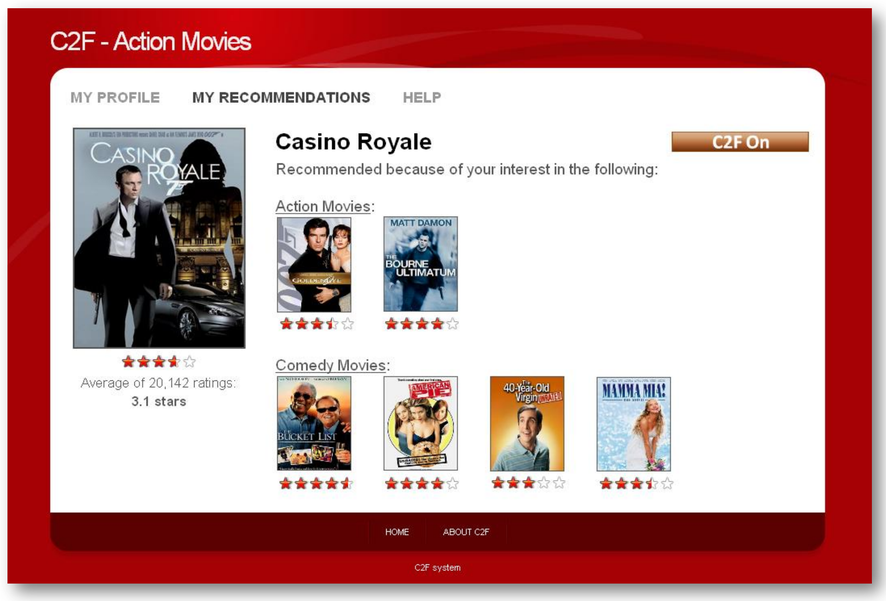
Distributed CF

A common denominator is required

Shared Users
- → a (relatively large) set of common users
- → example: same users shop at amazon, netflix, etc.

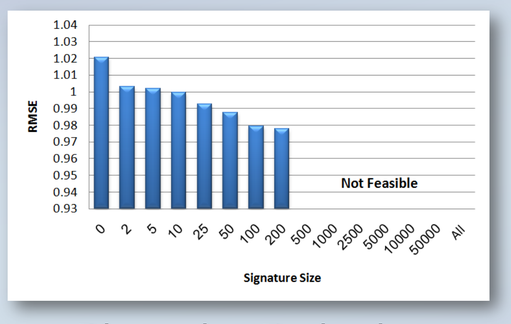


Shared Items
- → a (relatively large) set of common items
- → examples:
  -- Blockbuster branches in different countries
  -- Software mirror sites

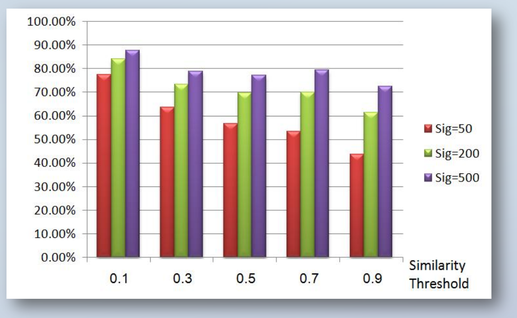


Experimental Results

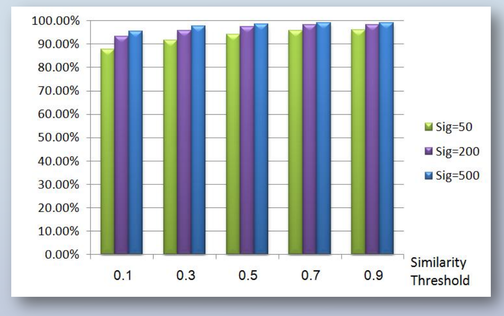
Dimension Selection



Dimension Reduction



DS Neighborhood - Precision



DS Neighborhood - Recall