Analytics and Visualization of Big Data

Fadel M. Megahed

Lecture 18: Recommendation Systems

AUBURN UNIVERSITY
SAMUEL GINN
COLLEGE OF ENGINEERING

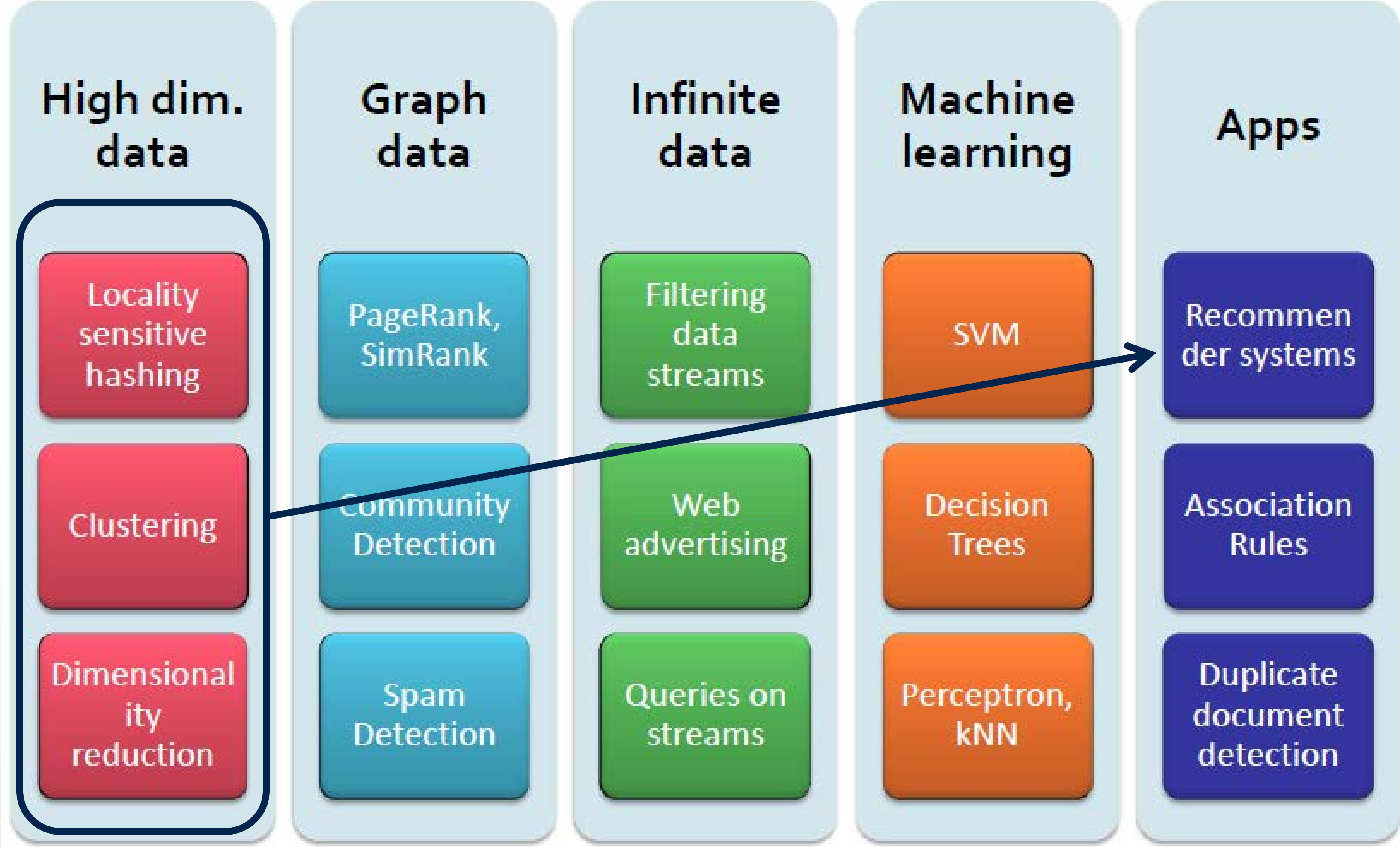
Department of Industrial and Systems Engineering

Spring 13

Spring Break Refresher: Course Objectives

- Explain the basics behind the hardware and software needed for “big data” analytics.
- Analyze high-dimensional data.
- Develop visualizations that makes the data “sing”☺.
- Describe the components of successful search engines.
- Mine the web using structured and unstructured data.
- Train algorithms that can be used to extract new knowledge from data.

Spring Break Refresher: Analytics Based on Data Type



Overview of Topics Covered in Chapter 9

A Model for Recommendation Systems

- The Utility Matrix
- Physical vs. Online Worlds
- Applications
- Populating the Utility Matrix

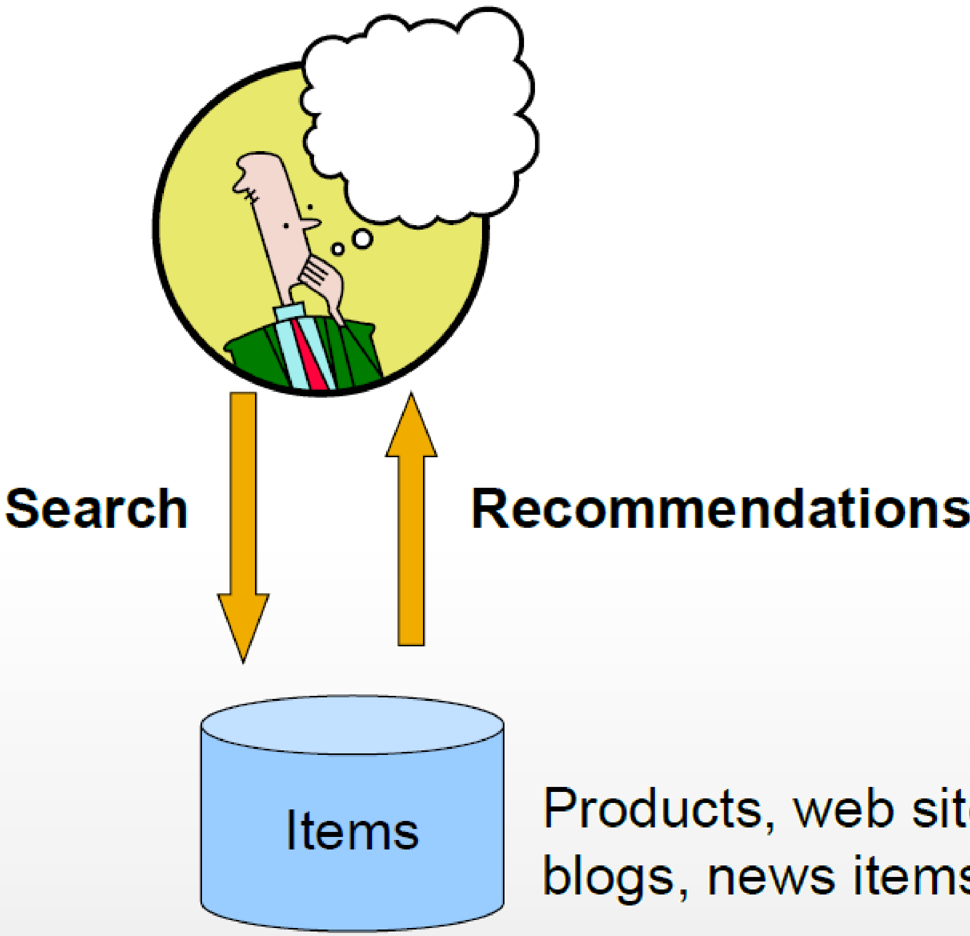
Content-based Recommendations

- Item Profiles
- Discovering Features of Docs
- User Profiles
- Recommending Items Based on Content

Collaborative Filtering

- Discussing its Main Idea (Only)

Recommendation Systems



Products, web sites, blogs, news items, …

Examples:

amazon.com
PANDORA
StumbleUpon
del.icio.us
NETFLIX
movielens
helping you find the *right* movies
last.fm
the social music revolution
Google News
YouTube
XBOX LIVE

Source: Jure Leskovic, Stanford CS246, Lecture Notes, see http://cs246.stanford.edu

The Utility Matrix

- In RS, there are two classes of entities:
  - Users
  - Items

- The data itself is represented as a **utility matrix**

- The **utility matrix** is typically very sparse

	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	4			5	1		
B	5	5	4				
C				2	4	5	
D		3					3

A Side Note: The Characteristic Matrix

- Properties of the Matrix:
  - Rows = elements of the universal set
  - Columns = sets
- 1 if and only if the token is a member of the set
- Column similarity is the Jaccard similarity of the corresponding sets (rows with value 1)
- Typical matrix is sparse

Element	S_1	S_2	S_3	S_4
a	1	0	0	1
b	0	0	1	0
c	0	1	0	1
d	1	0	1	1
e	0	0	1	0

Physical vs. Online Worlds

Physical World

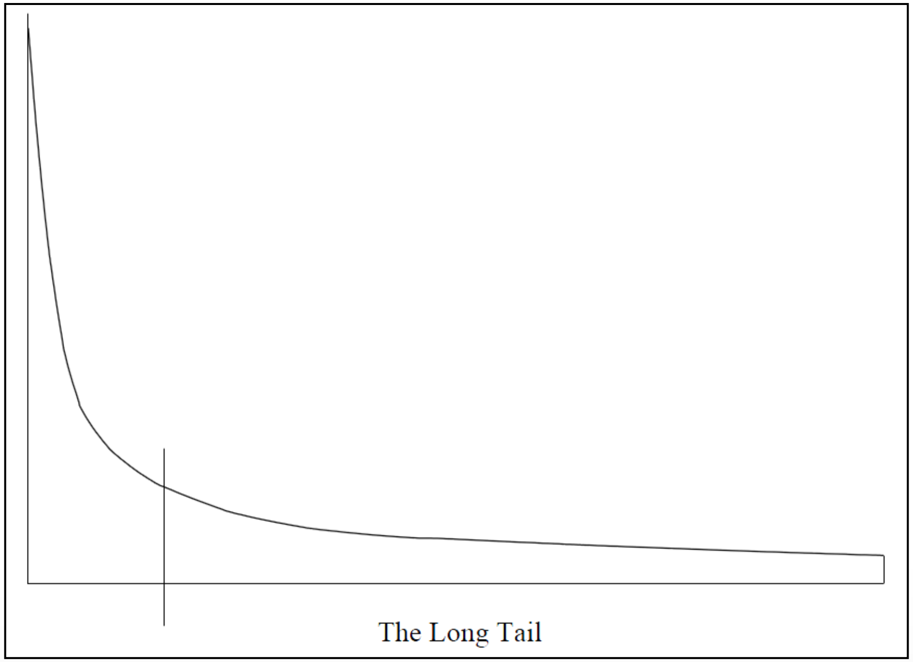
- Limited by shelf space
- Choice of products is based on aggregate #s
- Recommendation systems are fairly simple; recommend hits

Online World

- Essentially unlimited
- Choice of products is ~ infinite
- Recommendation systems can be tailored to a single customer

The Properties of the Online World → New Paradigm in Marketing and New Opportunities (How?)

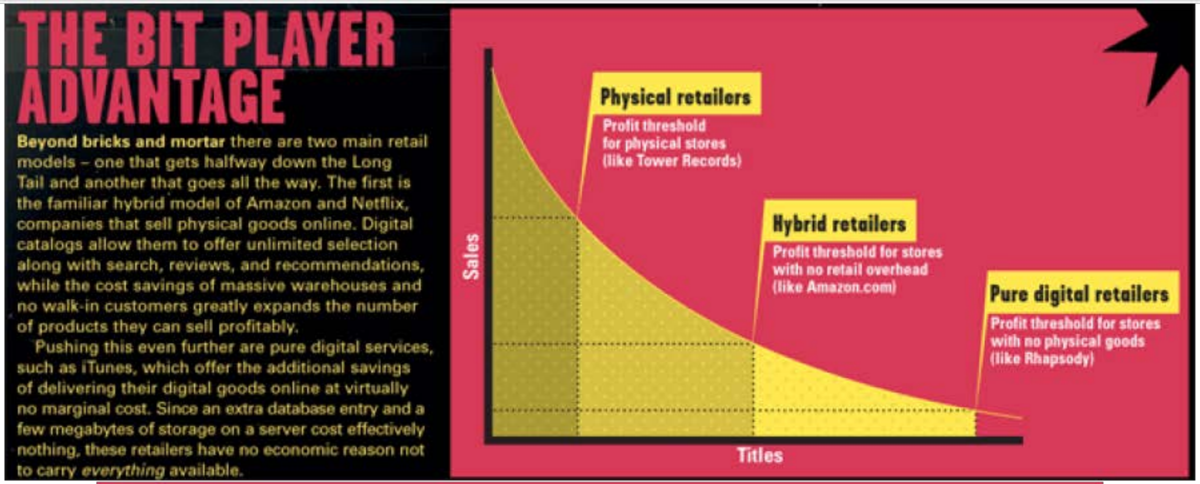
The Long Tail: A Discussion of the Wired Article

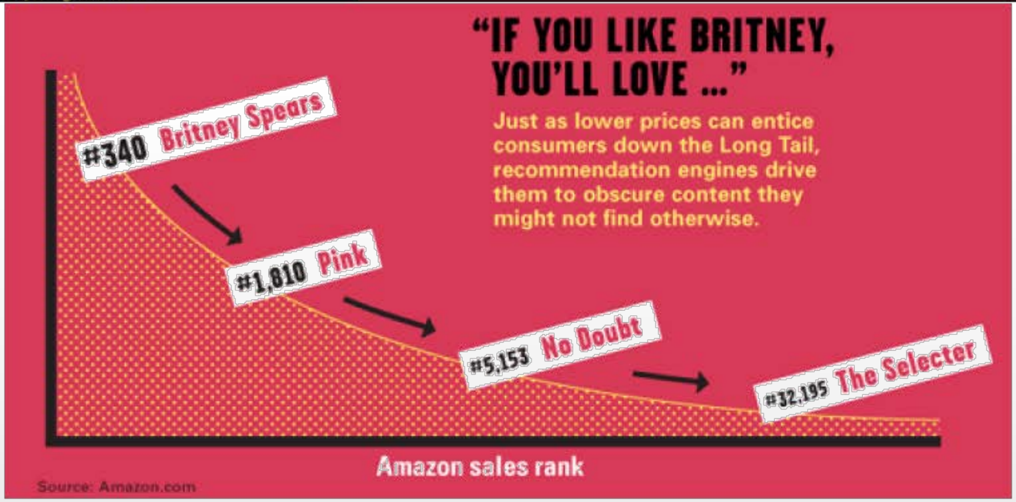


The long tail: physical institutions can only provide what is popular, while on-line institutions can make everything available

- **More choice necessitates better filters**
  - Recommendation engines
  - How Into Thin Air made Touching the Void a bestseller: http://www.wired.com/wired/archive/12.10/tail.html

The Long Tail: A Teaser





Read http://www.wired.com/wired/archive/12.10/tail.html to learn more!

Applications of Recommendation Systems

- Traditional Applications
  - Product Recommendations
  - Movie Recommendations
  - News Article Recommendations

- IE Applications
  - ????
  - ????

Key Issues in Recommendation Systems

1. **Gathering "known" ratings for matrix**

2. **Extrapolate unknown ratings from known ratings**
   - Mainly interested in high unknown ratings

3. **Evaluating extrapolation methods**

1) Gathering Ratings

- **Explicit**
  - Ask people to rate items
  - Doesn’t work well in practice – people can’t be bothered
- **Implicit**
  - Learn ratings from user actions
    - E.g., purchase implies high rating
  - What about low ratings?

2) Extrapolating the Empty Cells in the Utility Matrix

Key problem: matrix U is sparse

- Most people have not rated most items
- **Cold start:**
  - New items have no ratings
  - New users have no history

Three approaches to recommender systems:

- Content-based
- Collaborative
- Hybrid

Our Focus in Class

Overview of Topics Covered in Chapter 9

A Model for Recommendation Systems

- The Utility Matrix
- Physical vs. Online Worlds
- Applications
- Populating the Utility Matrix

Content-based Recommendations

- Item Profiles
- Discovering Features of Docs
- User Profiles
- Recommending Items Based on Content

Collaborative Filtering

- Discussing its Main Idea (Only)

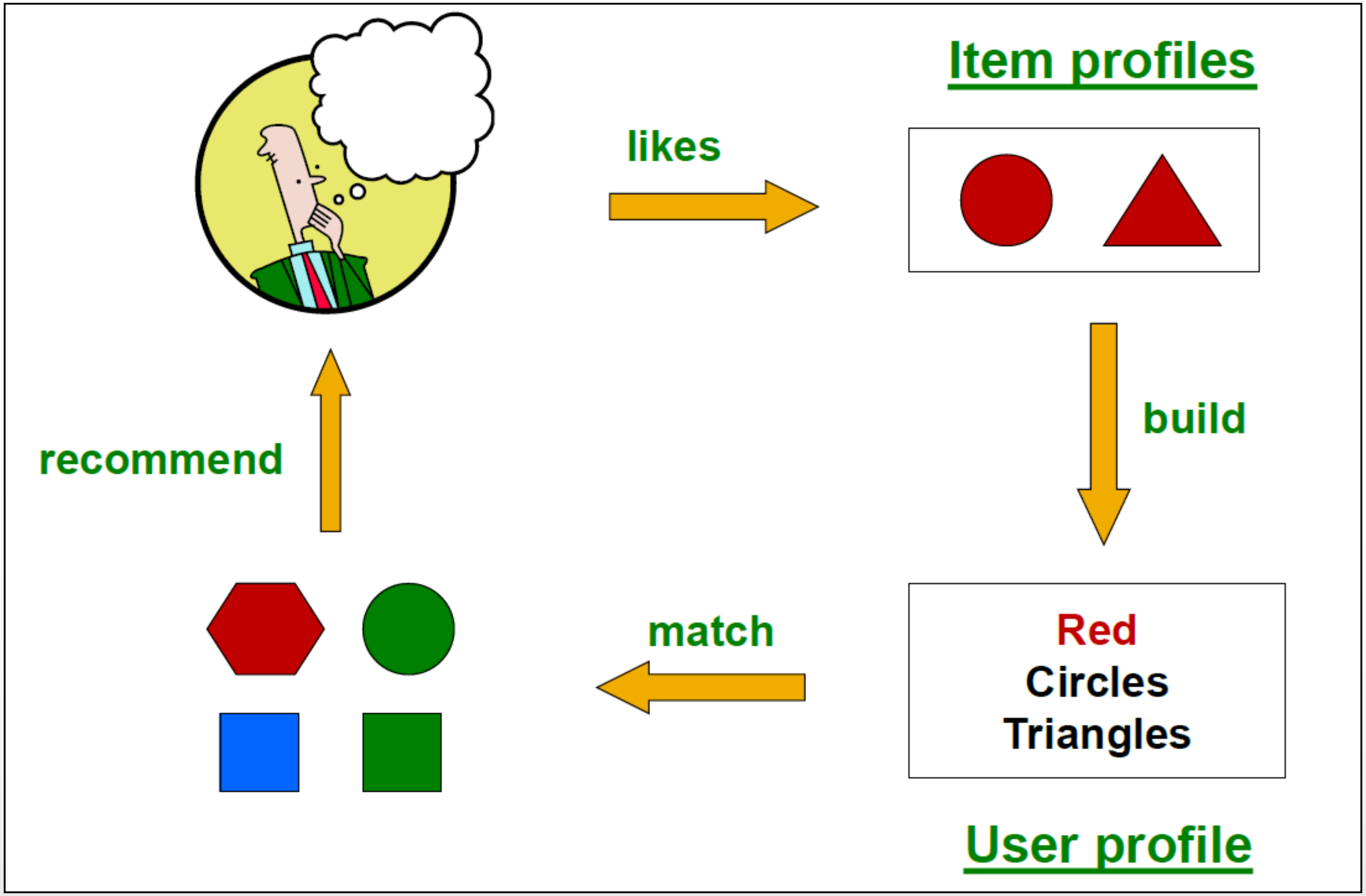
Content-Based Recommendations

Main idea: Recommend items to customer x similar to previous items rated highly by customer x

Example:

- **Movie recommendations:** Recommend movies with same actor(s), director, genre, …
- **Websites, blogs, news:** Recommend other sites with "similar" content
- **Images: How do you compare images?**

Plan of Action



Source: Jure Leskovic, Stanford CS246, Lecture Notes, see http://cs246.stanford.edu

Key: Generating Item (or User) Profile

- For each item, create an **item profile**

- Profile is a set (vector) of features
  - Movies: author, title, actor, director,…
  - Text: set of "important" words in document

- How to pick important features?
  - Usual heuristic from text mining is TF-IDF (Term frequency * Inverse Doc Frequency)
    - Term … feature
    - Document … item

- **Question: Now what?**

Pros: Content-based Approach

- **+: No need for data on other users**
  - No cold-start or sparsity problems
- **+: Able to recommend to users with unique tastes**
- **+: Able to recommend new and unpopular items**
  - No first-rater problem
- Can provide explanations of recommended items by listing content-features that caused an item to be recommended

Cons: Content-based Approach

- –: **Finding the appropriate features is hard** E.g., images, movies, music

- **–: Overspecialization** Never recommends items outside user's content profile
  - People might have multiple interests
  - Unable to exploit quality judgments of other users

- **–: Recommendations for new users** How to build a user profile?

Overview of Topics Covered in Chapter 9

A Model for Recommendation Systems

- The Utility Matrix
- Physical vs. Online Worlds
- Applications
- Populating the Utility Matrix

Content-based Recommendations

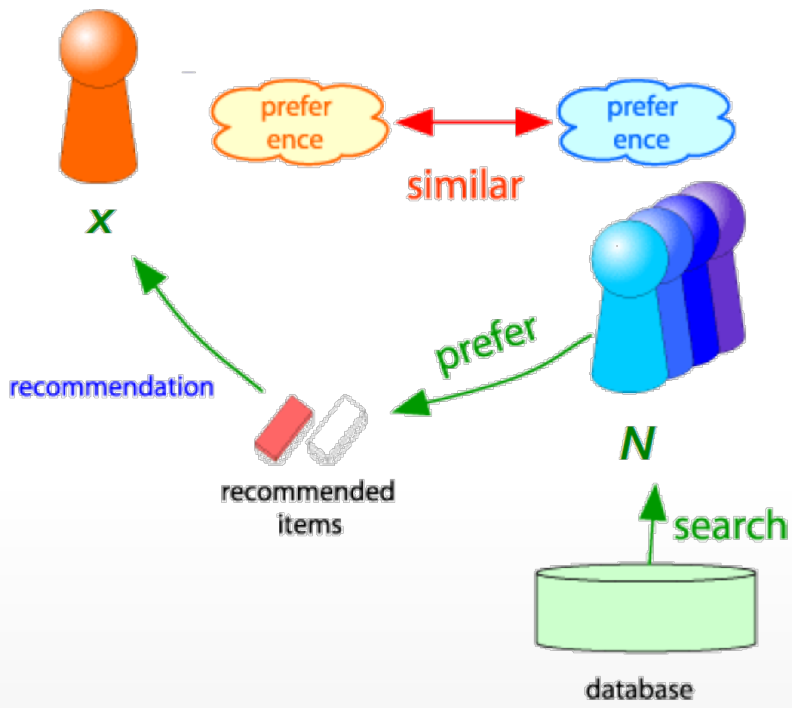
- Item Profiles
- Discovering Features of Docs
- User Profiles
- Recommending Items Based on Content

Collaborative Filtering

- Discussing its Main Idea (Only)

Collaborative Filtering

- Consider user x
- Find set N of other users whose ratings are "**similar**" to x's ratings
- Estimate x's ratings based on ratings of users in N



Pros/Cons: Collaborative Filtering

- **Works for any kind of item**
  - No feature selection needed
- **Cold Start:**
  - Need enough users in the system to find a match
- **Sparsity:**
  - Hard to find users that have rated the same items
- **First rater:**
  - Cannot recommend an item that has not been previously rated
  - New items, Esoteric items
- **Popularity bias:**
  - Cannot recommend items to someone with unique taste
  - Tends to recommend popular items

Practical Questions to think about: Evaluation

?

Analytics and Visualization of Big Data

Fadel M. Megahed

Lecture 18: Recommendation Systems

AUBURN UNIVERSITY

SAMUEL GINN
COLLEGE OF ENGINEERING

Department of Industrial and Systems Engineering

Spring 13