Recommender Systems
Filtering Approaches

Professor Robin Burke

Spring 2019

Thanks to Yong Zheng, IIT, for some materials

Outline

- Assignments
  - Homework 4
  - Project progress
  - Scuiz III
- Context-aware recommendations
  - Context splitting
  - Context modeling
    - Tensor factorization
    - Factorization machines
- Network-based recommendation
  - Intro

Homework 4

- Due Thursday
- What questions?

Project progress report

- Due today

Rest of quarter

- Week 12
  - Finish up context-aware topic
  - Recommendations in networks
- Week 13
  - Learning to Rank
- Week 14
  - Fairness in recommendation
- Week 15
  - Presentations I
- Week 16
  - Presentation II
  - Posters
- 5/5: Final report

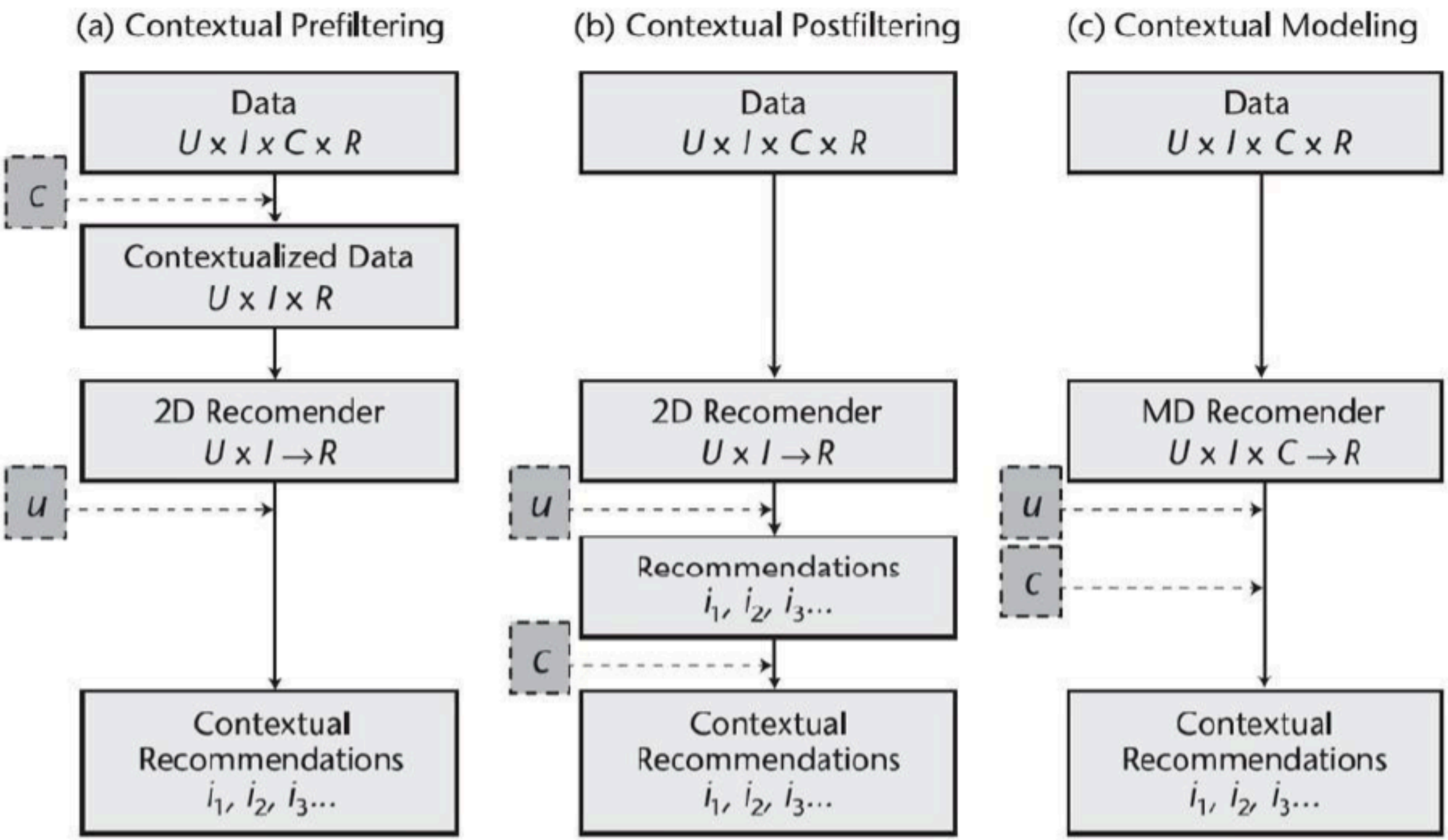
Note: No Homework 5

Context

- Context modulates preferences
- What is preferred in one context
  - Might not be preferred in another
- Aspects of a profile acquired over a long period
  - Might contain a conglomeration of actions taken in different contexts
  - Good recommendation might require dis-entangling them
- In practice, this works
  - In some domains, you get much better accuracy
  - If you can find the right contexts
  - And the right data

Algorithms

- Most research has been done on integrating context into collaborative filtering



Context-aware recommendation requires associating contextual information with ratings. This may be difficult because:

- A. Context information may be privacy sensitive
- B. Requiring users to supply context information adds overhead to supplying ratings
- C. Inferring context when ratings are implicit is inherently imperfect and noisy
- D. It is difficult to know what contextual features will be important and at what level of granularity
- E. All of the above

Context-aware Splitting

- Use context to selectively increase sparsity
- Example
  - Item splitting
    - The ratings for some items may be highly correlated with location

At Cinema

At Home

At Swimming Pool

Item splitting

User	Item	Location	Rating
U1	M1	Pool	5
U2	M1	Pool	5
U3	M1	Pool	5
U1	M1	Home	2
U4	M1	Home	3
U2	M1	Cinema	2

Location dimension has a big influence

M11: Pi + Pool

M12: Pi + Elsewhere

Item splitting: issues

- Which dimensions to use for splitting?
- What are the optimal values for splitting?

User	Item	Loc	Rating
U1	M1	Pool	5
U2	M1	Pool	5
U3	M1	Pool	5
U1	M1	Home	2
U4	M1	Home	3
U2	M1	Cinema	2

Transformation

User	Item	Rating
U1	M11	5
U2	M11	5
U3	M11	5
U1	M12	2
U4	M12	3
U2	M12	2

Impurity criteria

- Standard statistical question
  - Are the distributions of ratings in context A different from context B?
- Possible criteria
  - T-test
  - Z-test
  - Chi-square
  - Information gain

More splitting

- User splitting
  - Same idea
    - But we divide user profiles by context
  - $U1 \Rightarrow U11, U12$
    - User rates differently in one context vs another
- Combined user / item splitting
  - Split both

Splitting approaches

User	Item	Rating	Time	Location	Companion
U1	T1	3	Weekend	Home	Friend
U1	T1	5	Weekend	Cinema	Girlfriend
U1	T1	?	Weekday	Home	Family

(a) by Item Splitting

User	Item	Rating
U1	T11	3
U1	T12	5
U1	T11	?

(b) by User Splitting

User	Item	Rating
U12	T1	3
U12	T1	5
U11	T1	?

(c) by UI Splitting

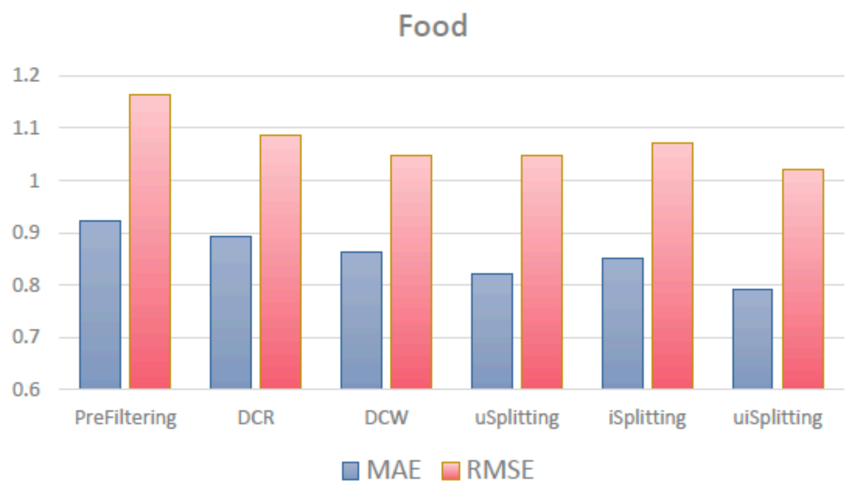
User	Item	Rating
U12	T11	3
U12	T12	5
U11	T11	?

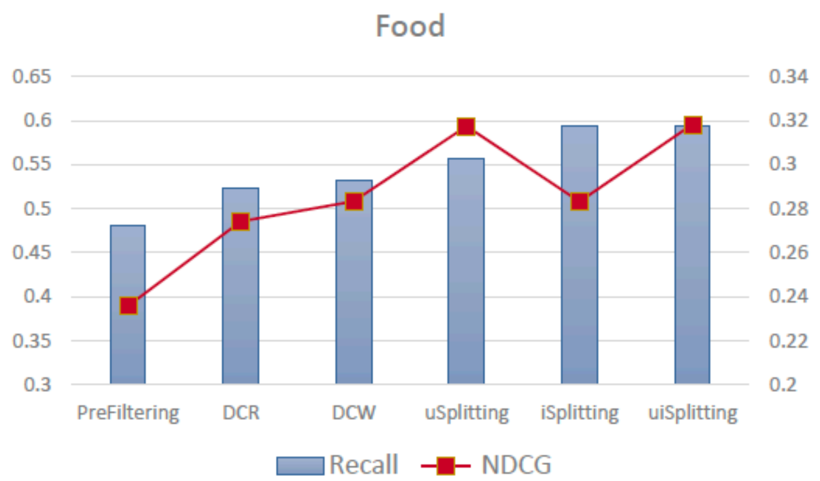
After splitting

- Looking generally for just one split
  - Otherwise too much sparsity
- Apply normal collaborative filtering methods
- Typically matrix factorization
  - To compensate for introduced sparsity

Example results

Japan Food Data: 6360 ratings given by 212 users on 20 items within 2 context dimensions





A recommender system for fashion uses item splitting where the context is the season of the year when a purchase was made. Assume that the purchases are evenly distributed throughout the year. The increase in sparsity due to item splitting would be:

- A. 2x because a single split is chosen, doubling the number of "items"
- B. 4x because there are four seasons.
- C. You can't tell because the correct split depends on which division of the data gives the highest impurity value