QUIZ 1: Classical recommender systems

Let's test what we know about classical recommendation methods. Think before you answer!

Email address *

Q1. Content-based (CB) filtering (Mark all that are true) *

- [x] Content-based filtering requires information about available items
- [x] Content-based filtering recommends items with similar content to the user's past selections
- [ ] Content-based filtering requires user ratings
- [ ] Content-based filtering is affected by popularity bias
- [x] Most CB-recommendation methods originate from Information Retrieval

Q2. In a TFIDF algorithm, IDF stands for “Inverse Document Frequency.” What does that mean? *

- ○ IDF means that more popular documents are given a lower weight so they don’t get recommended too often
- ◉ IDF means that terms count for more in a profile or query if they don’t appear in very many documents
- ○ IDF means that the optimal number of times for a term to appear in a document is one. Any more is penalized by the algorithm
- ○ IDF is a technique that attempts to recommend each document equally often
- ○ I have no idea

Q3. Identify the correct statements related to Collaborative Filtering *

- ○ The problem of collaborative filtering is to predict how well a user will like an item that he has not rated given a set of historical preference judgments from a community of users
- ○ Predict the opinion the user will have on the different items
- ◉ Recommend the ‘best’ items based on the user’s previous likings and the opinions of like-minded users whose ratings are similar
- ○ None of these

Q4. Which of the following algorithm is less expensive and scalable? *

- ○ User-Based Collaborative Filtering
- ◉ Item-to-Item Collaborative Filtering
- ○ User-to-Item Collaborative Filtering
- ○ None of these

Q5. User-Based Collaborative Filtering methods (Mark all that apply) *

- [x] Are based on user's similarity only
- [x] In such methods complexity grows linearly with the number of customers and items
- [x] Suffers the problem of sparsity of recommendations on the data set
- [ ] None of these

Q6. Item-to-Item Collaborative Filtering methods *

- [x] It is more scalable than User-Based Collaborative Filtering methods
- [x] Amazon.com uses this method
- [ ] Scales independently of the catalog size or the total number of customers
- [x] Acceptable performance by creating the expensive similar-item table offline
- [ ] None of these

Q7. Collaborative filtering suffers from the problem of "Data sparsity" because *

- [x] The user-item matrix used for collaborative filtering could be extremely large and sparse, which brings about the challenges in the performances of the recommendation
- [ ] The new item problem limits the content-based recommendation
- [x] Cold start problem: As collaborative filtering methods recommend items based on users' past preferences, new users will need to rate sufficient number of items to enable the system to capture their preferences accurately and thus provides reliable recommendations
- [x] When new items are added to system, they need to be rated by substantial number of users before they could be recommended to users who have similar tastes with the ones rated them

Q8. When would it not make sense to use item-item collaborative filtering vs user-user? *

- (●) When there are many more items than users
- ( ) When new users appear often in the system
- ( ) When the user has not rated all items
- ( ) Never

Q9. The main tasks of collaborative filtering can be categorized into two parts. Identify the correct statement related to each part *

- [x] Look for users who share the same rating patterns with the active user (the user whom the prediction is for)
- [x] Use the ratings from those like-minded users found in step 1 to calculate a prediction for the active user
- [ ] Build an item-item matrix determining relationships between pairs of items
- [ ] Build an user-user matrix determining relationships between users

Q10. Collaborative filtering suffers from the problem of "Diversity" because *

- [ ] Due to diversity in product range, system does not perform well
- [x] As collaborative filtering systems recommend products based on sales or ratings, they cannot usually recommend products with limited historical data
- [x] It creates a rich-get-richer effect for popular products, leading to less diversity
- [ ] Sales may vary from item to item and user to user

Q11. Identify correct use(s) of collaborative filtering *

- [x] News reader services on web
- [x] Recommend interesting or popular information as judged by the community
- [ ] Can be used for user profiling and helps the site recommend content that submitted by other users similar to your taste, or you find interesting, as well as topics that you usually vote up and follow
- [ ] Services like Google News and Netflix are typical example of collaborative filtering based media

Q12. Which of these statements is true specifically of a situation where there are more items (or products) than users in a ratings matrix? *

- [x] An item-based recommender algorithm is less likely to be a sensible choice
- [ ] The user-user algorithm should use cosine similarity instead of correlation
- [ ] Dimensionality reduction will probably not work
- [x] Content-based algorithms are more likely to be effective
- [ ] I have no idea

Q13. Collaborative Filtering (Mark all that are true) *

- [ ] Collaborative Filtering requires information about items the user has rated
- [x] Collaborative Filtering produces good results in many cases
- [x] Collaborative Filtering is affected by a user cold-start problem
- [x] Pearson Correlation is most frequently used in neighborhood-based Collaborative Filtering

Q14. What's the difference between a recommendation and a prediction? *

- [x] Only a prediction involves an explicit estimate of "how much you like it."
- [x] Recommendations only come in lists -- say top 5 or top 10
- [ ] Predictions focus on whether you'll like it in the future, while recommendations are about liking in the past
- [ ] Recommendations come from another user; predictions come from the computer
- [ ] I have no idea

Q15. User-based CF (Mark all that are true)

- [ ] User-based CF is said to be model-based
- [ ] The advantage of user-based CF is scalability

1st ACM Summer School

- [x] The advantage of user-based CF is simplicity
- [x] The disadvantage of user-based CF is sparsity

Q16. Item-based CF (Mark all that are true) *

- [x] Item-based CF looks at item similarity
- [x] Item-based CF improves scalability over user-based CF
- [x] Item-based CF can be calculated offline
- [ ] Item-based CF does not have a cold-start problem

Q17. Why is item-item more amendable to pre-computation than user-user? (Mark all that are true) *

- [x] When there are many more users than items, item similarities are fairly stable, while user similarities can change rapidly as users interact with the system
- [ ] Because the user we're predicting for may not have rated all of the neighbor items in user-user CF. We need enough potential neighbors to be able to find k of them among the user's ratings

Q18. Which of these is an example of a product association recommender? *

- [x] Amazon.com's "people who bought this" recommender
- [ ] MovieLens' "Top Movies for You" recommender
- [ ] The Zagat Guide's scores for food, service, and decor.
- [ ] The Entree Restaurant Guide's "find me a restaurant that's cheaper" recommender
- [ ] I have no idea

Q19. Which of these is NOT an example of an implicit rating? *

- [ ] The number of times you listen to a song
- [x] The fact that you labelled the movie Star Wars with five stars
- [ ] The fact that you printed out an article from an online newspaper.
- [ ] The fact that purchased a Waring Blendor on Amazon.com
- [ ] I have no idea

Q20. Which of the following is a correct example of collaborative filtering *

- [ ] Sharing mail filtering rules with your friends to help filter out junk email
- [ ] Building and using a profile of music genres you like to play music for you and your friends
- [x] Choosing songs to listen to based on how much other people with similar tastes like those songs.
- [ ] Building a list of users from whom you want to reject e-mail
- [ ] I have no idea

Q21. Which of the following features adds the most value to a user-user collaborative filtering algorithm? *

- [x] Weighting the neighbors by the correlation with the target user
- [ ] Decreasing the weight of neighbors that have few items in common
- [ ] Capping the neighborhood size at a modest number such as 50-200.
- [ ] Using normalized ratings and de-normalizing to the target user's rating distribution
- [ ] I have no idea