# Hybrid Systems and Evaluation

Email address *

Your name *

Q1. What is the difference between a weighted hybrid recommender and a mixed hybrid recommender? *

- ( ) A mixed hybrid is a weighted hybrid where each underlying recommender is given the same weight
- (•) Each recommendation from a weighted hybrid is based on combined scores of different recommenders; recommendations from a mixed hybrid can come from a single recommender.
- ( ) In a weighted hybrid, recommenders are run in parallel; in a mixed hybrid, the output of one is fed as in input into the other
- ( ) There is no difference. They are two names for the same type of hybrid.
- ( ) I have no idea

Q2. What is the difference between feature combination and feature augmentation as hybrid recommendation methods? *

- ( ) in feature combination, features from different recommendation sources are combined as input to the same method while in augmentation the output of one method is used as input features to another method
- ( ) There is no difference. They are two names for the same type of hybrid
- (•) In feature augmentation, features are added from one step to the other in a cascade mode, while in feature combination, new features are combined with the existing features so the number of features stays the same from step to step
- ( ) I have no idea

Q3. A recommendation list with high diversity will have a mix of highly scored and lower scored items near the top *

- (•) True
- ( ) False

Q4. In blending recommendations, which of the statements are correct? *

- [x] Blended recommender may be higher quality and more diverse than any individual ranker
- [x] Round-robin is a fair method with respect to source selection
- [ ] Greedy sampling displays highest quality items and represents all sources in the best way
- [x] Round-robin achieves good diversity but in the cost of quality because it does not consider the scores of the items

Q5. In blending recommendations, which of the statements are correct? *

- [x] Multinomial sampling better captures user click probability, because we can randomly draw from each source based on weights
- [x] Both multinomial sampling and round-robin do not consider item quality
- [ ] Multinomial provides the best user experience because it captures best what customers often like
- [ ] Greedy is a form of multinomial, where the weight of a source is set to the weight of each highest ranked item

Q6. Why would you use a different metric for evaluating prediction Vs. top-N recommenders (Match the correct answer) *

- ( ) Most of the singular values are almost always very close to zero, indication that not much is lost by reducing those dimensions
- ( ) Incorporating new ratings, users, or items into the model by using the existing factorization and updating/computing a feature vector from the ratings
- (•) Because you need different algorithms to compute predictions v. top-N recommendation
- ( ) The evaluation can only judge whether the returned items are among the rated/consumed/purchased ones. having too many other items just increases the number of desirable but not-yet consumed items, making it harder to tell whether the recommendations are good

Q7. What operational goals are needed to make good recommendations? (Mark all that are true) *

- [x] Accuracy of predictions
- [ ] Similarity to past user selections
- [ ] High predicted ratings
- [x] Serendipity
- [x] Diversity

Q8. When computing serendipity we depend upon a prior "primitive" estimate of obviousness and a determination of whether a recommended item is actually relevant. why do we need these measures? *

- (●) Because serendipity is measuring the degree to which an algorithm is giving recommendations for non-obvious but still relevant products or items
- ( ) Because ranking accuracy at the top of the list is weighted more heavily than accuracy further down the set
- ( ) Because it directly optimizes the recommender to pick good items rather than just score items accurately

Q9. When holding out (not putting in) ratings from a user's profile for evaluation what is the benefit of holding out the last ratings rather than holding out random ratings *

- ( ) It directly optimizes the recommender to pick good items rather than just score items accurately
- (●) It more accurately simulates the recommender's knowledge when the held-out ratings were given
- ( ) There is no impact

Q10. Which of these statements best explains how we perform an n-fold cross validation for getting a more accurate measure of the accuracy experienced by users in recommender systems? *

- [x] Divide the data set into n partitions, hold one partition out as the test data and train the recommender on the other partitions
- [ ] Divide the withheld user data into query data used for training and "test" data. Measure the accuracy of prediction for the test data from the query data for each user and average

Q11. Which of these is NOT likely to be an effective means for explaining user-user recommendations to a user, if your goal is to get the user to believe the recommendation? *

- [ ] Giving a brief overview of the data -- e.g., 20 ratings, 18 of them liked it.
- [ ] Giving a graph showing each neighbor's rating against your similarity with that neighbor and number of items rated in common
- [ ] Showing how often similar recommendations were correct in the past
- [x] Giving related supporting information about the user's preference for product features (e.g., movie actors, product brands)
- [ ] I have no idea

Q12. In some top-n evaluations, instead of considering all items, the recommender recommends from the useful items the user has rated/consumed/purchased plus a random subset of all items. Why is this useful? *

- ◯ Incorporating new ratings, users, or items into the model by using the existing factorization and updating/computing a feature vector from the ratings
- ◯ Because you need different algorithms to compute predictions v. top-N recommendation
- ◉ The evaluation can only judge whether the returned items are among the rated/consumed/purchased ones. having too many other items just increases the number of desirable but not-yet consumed items, making it harder to tell whether the recommendations are good
- ◯ To help take into account a user's emotional state or current mood, location and time and date

Q13. Which blending algorithm tends to maximize diversity? *

- ◯ Round robin
- ◉ Multinomial sampling
- ◯ Submodular diversifier
- ◯ Fair-and-balanced

Q14. Which of the following statements about recommendation diversity is FALSE? *

- ◯ Diversity is often in conflict with the quality of individual recommendations.
- ◉ A submodular utility function grows linearly with the number of items.
- ◯ Diversification is a good way to hedge against uncertainty in user intent
- ◯ Diversity is a property of both the attributes and the layout of the recommendations

Q15. Which of the following is true about displaying search results in a 2D grid? *

- ○ The user will tend to view the recommendations in a left-to-right, top-to-bottom order.
- ○ The number of ways to assign items to positions is O(N^2)
- ◉ Quadratic models help mitigate the combinatorial explosion of assigning items to slots
- ○ It is impossible to infer user scanning behavior without eye-tracking.