Collaborative Filtering

Sunday, October 28, 2018 1:55 PM

2016:

1. **[Collaborative Filtering]**

    a. Two Algorithms used for Collaborative Filtering:
        i. K nearest neighbours
        ii. Matrix Factorization

    b. What problems are well suited for Collaborative Filtering?

Problem	**Well Suited**
Predicting missing values in a matrix of sensor readings from a town, where the columns correspond to sensors and the rows correspond to timestamps	Yes
Predicting the books that a user would want to read in a library, given historical records of the loans of all the users	Yes
Predicting the sentiment for a named object in a new tweet, given a large data set of annotated tweets that may not contain the named object	No
Predicting the current water level in a reservoir, given recent rainfall data and historical records of water measurements in the reservoir	No

2017:

2. **[Collaborative Filtering]**
Suppose that you are given a data set from Amazon in the form of a partially-observed matrix Y whose entry Yai represents the rating of customer a for a product i. You decide to try both the k-nearest-**neighbors algorithm and matrix factorization** to predict the values of unknown ratings Yai.
In k-nearest-neighbors, to predict the unbiased rating $Yai - \bar{Y}a$ where $\bar{Y}a$ is the average of all observed ratings by customer a, we use the weighted sum of the unbiased ratings of neighbors b which are nearest to a. These neighbors are ranked according to a cosine similarity function sim(a,b).

a. If we predict $Y_{ai} - \bar{Y}_a$ using $\sum_b w_b r_b$, which weights w_b and values r_b should we use?

A. $w_b = \frac{\mathrm{sim}(a,b)}{\sum_{b'} \mathrm{sim}(a,b')}$ and $r_b = Y_{bi} - \bar{Y}_b$

B. $w_b = \frac{|\mathrm{sim}(a,b)|}{\sum_{b'} \mathrm{sim}(a,b')}$ and $r_b = Y_{bi} - \bar{Y}_b$

C. $w_b = \frac{\mathrm{sim}(a,b)}{\sum_{b'} |\mathrm{sim}(a,b')|}$ and $r_b = \mathrm{sign}(\mathrm{sim}(a,b))\,(Y_{bi} - \bar{Y}_b)$

D. $w_b = \frac{|\mathrm{sim}(a,b)|}{\sum_{b'} |\mathrm{sim}(a,b')|}$ and $r_b = \mathrm{sign}(\mathrm{sim}(a,b))\,(Y_{bi} - \bar{Y}_b)$

A or D