Recommendation Engine Assignment

LAB - 2

Kamal Nandan | MBD-O1 | 28/02/2018

CB – SIMPLE UNARY

- We calculate the User profile by doing a sumproduct of questions in a category and the up/down vote made by the respective user.
- After calculating the user-profile, we find the prediction for that particular user by using the following formula in excel sheet: =SUMPRODUCT($B$28:$K$28,B4:K4)/SUMPRODUCT(SQRT(SUMPRODUCT($B$28:$K$28,$B$28:$K$28)),SQRT(SUMPRODUCT(B4:K4,B4:K4)))
- The top 5 predictions/recommendations for the respective users are highlighted in yellow.

CB – UNIT WEIGHT

- We calculate the unit weight for each question in a category in a separate table (at the bottom), then we use this table and find the sumproduct of weights for each/question in that category with the up/down vote made by the respective user. We do so by calculating the proportion of each category in the respective question rather than just the no. of coccurrences
- After calculating the user-profile, we find the prediction for that particular user by using the following formula in excel sheet:
- =SUMPRODUCT($B$28:$K$28,B35:K35)/SUMPRODUCT(SQRT(SUMPRODUCT($B$28:$K$28,$B$28:$K$28)),SQRT(SUMPRODUCT(B35:K35,B35:K35)))
- The top 5 predictions/recommendations for the respective users are highlighted in yellow.
- Other than the unit weights, all the other procedures are more or less the same as simple unary.

CB-IDF

- We calculate the unit weight for each question in a category in a separate table (at the bottom); we later replaced this with inverse of sum of all the categories, but since I had already calculated it, I am leaving it there in the excel sheet
- We calculate the DF
- We calculate the IDF
- We calculate the inverse of sum of all the categories for a question

- Then we calculate the user's profile by doing the sum-product of questions in each catergory, the users' up/down vote for the respective questions, and the inverse of sum sum of all categories, as calculated above.
- Finally we calculate the prediction in the same manner as we have done in case of simple – unary.
- The top 5 predictions/recommendations for the respective users are highlighted in yellow.

SWITCHING HYBRID:

- In case of switching hybrid, we may need to switch from personalized content based to non-personalized one, in case of cold-start issues. If the user rated any question in a particular category, then the personalized content based recommendation will be given. Otherwise the mean/average (of all the other users profile who rated the question in the particular category will) for the respective topic will be provided as a recommendation.
- The top 5 predictions/recommendations for the respective users are highlighted in yellow.

HYBRID CHALLENGE:

1. In case of hybrid challenge too, we will more or less follow the same approach as we have been following in the switching hybrid approach. But in this case, we may play smart and think of interaction effect by clubbing together pairs of topics or may be more topics which may amplify or diminish the combined weight. Also, other than the above. But for simplicity, for now we will only consider pairs of topics. Also, instead of switching from personalized recommendation to non-personalized, in case of cold-start problem, this approach will
2. The base for the Hybrid challenge will be TFIDF algorithm with 2 differences. First, the input for prediction calculation will include also pairs of topics, not just the topics alone. It seems reasonable as sometimes the combination of different topics can create synergetic effect, or, otherwise, decrease the meaning of combined variables. Second, instead of switching from one method to another in order to solve the cold start problem, this algorithm will assign different weights for CB-IDF values in user profile and for average values according to how many times each topic has already occurred in questions user rated.