

Content Based Filtering

QUESTION 1:

Steps to consider when utilizing the reviews and User information:

1. Choose the right and useful features, eliminating junk. In our data, item description, age, gender, avg history of spend etc could form important feature, which involves information about the user personality and habits and about the product.
2. Encode the data into write format. There are many categorical and text fields in our chosen data. It is best to convert them to Numerical form (unless using Tree based model, which can handle categorical data). One hot encoding or labelling each have down sides, it is best, though time consuming to convert the categorical and text data into their embeddings. For text data, little pre=processing will be required to eliminate useless terms by either using tf-idf or manually selecting based on the context.

Tf-Idf removes the most and least occurring words which might not be useful in analysing the gist of the sentence.

3. Recommending Content: It involves making a prediction about how likely the user is going to like the recommended item. It can be found by one of the mentioned methods:

- Similarity/Distance-Inverse based
- Matrix Factorization
- Deep Learning approach

In all these approaches, we find how close the embedding vectors of the reviews features are to that of the other in the feature space, for all items.

4. Once we find the items similar to each other, we analyse the particular choice of the user, based on the User based feature embeddings. There are many ways to build preference profile for the User. The simplest way is to take the weighted mean of User's rating and the Tf-Idf vector representation of the respective movies. Any of the similarity metric between content vector of the items matching the one rated highly by the User could get the items to recommend. Pick the top best ones.

QUESTION 2:

Emadri should consider both the models, in order to reach out for providing the best experience to its customers.

For those customers which are regular and Emadri has enough data about their past deals to know their preferences, Content Based methos is the best

In those cases when Emadri makes a new deal with a new firm or individual, it is quite unlikely to be able to analyse the choices of those customers as there are not enough data, but yet we want to make our customer not only satisfy but also happy so that the deal is larger the next time. In those cases, Collaborative Filtering comes to the rescue. We can analyse the Users who has similar choices as this customer based and recommend items, and keep recording their reaction to that deal, thus preparing data for our Content based Filtering Approach later.