Content based Method

Introduction:
A content-based recommender works with data that the user provides, either explicitly (rating) or implicitly (clicking on a link). Based on that data, a user profile is generated, which is then used to make suggestions to the user. As the user provides more inputs or takes actions on the recommendations, the engine becomes more and more accurate.

Definition:
Based on a description of the item profile and user's profile preferences

Item Profiles

- For each item, create an item profile
- Profile is a set (vector) of features

  Movies: author, title, actor, director…

  Text: Set of "important" words in document

User profiles and Prediction

- User profile possibilities:

  Weighted average of rated item profiles

  Variation: weight by difference from average rating for item
- Prediction Heuristic:

  Given user profile *x* and item profile *i*, estimate

In movie description example, they used overview and tagline as description
In our amazon beauty product, we will use item description + item category
Result we would like to look at: `title` column
Columns we need: item `description, categories`

Steps:

1. Create item profile
   - Find item attributes (several **combinations**)
     - `Categories + Brand`
     - `Description`
     - `Title`
   - average or weighted attributes of above
2. Create user profile
3. Calculate similarity using different algorithms
   - TF/IDF: TF-IDF Vectorizer, calculating the Dot Product

     – Encodes text documents in multi-dimensional Euclidean space

1. weighted term vector

– TF: Measures, how often a term appears (density in a document)

1. Assuming important terms appear more often
2. Normalization has to be done in order to take document length into account

– IDF: Aims to reduce the weight of terms that appear in all documents

- Pairwise cosine similarity
  – Based on the angle between the vectors
- Machine learning methods

3. Results
   use train test split
4. Evaluations
   - Pros
     - We Do Not need profile of other users, which means no cold-start or sparsity problems
     - Able to recommend to users with unique tastes
     - Able to provide explanations

   - Cons
     - What result do we look at, like accuracy score or…
     - Finding the appropriate features is hard; Unable to exploit quality judgments of other users
     - When make recommendations for new users, there exist problem about build user profile
     - Never recommends items outside user's content profile, people might have multiple interests