DATA 643: Recommender Systems

Discussion 1: NY Times Recommender

Walt Wells, Summer 2017

The New York Times

Overview

The New York Times recommends articles and news to subscribers in browser and applications available on most major devices. To facilitate this, they have developed an intricate hybrid recommender system. Most of our observations below will be based on an article from August of 2015 (ancient!). It is difficult to determine the specifics of their current model or if it even differs at all from the process outlined above, but their current policies and broad descriptions of the system are available at: https://www.nytimes.com/recommendations

Model

Content Based Filtering

Relies on Item and User Profiles to make associations. Item profiles include = "the topics, author, desk and associated keyword tags of each article."

- **Upside:** User reads 10 articles tagged "politics", confident they'll want another tagged "politics".
- **Downside:** TFIDF, rare tags have a large effect in generating User Profiles.

Collaborative Filtering

Goal is to recommend based on what similar users have read.

- **Upside:** Can find interesting patterns for readers with unusual tastes. News Specific Downside = Tunnel Vision – reading preferences too narrowly clustered.
- **Downside:** Cold start problem.

Collaborative Topic Modeling

Based on a 2011 paper out of Princeton - Collaborative Topic Modeling for Recommending Scientific Articles; Wang, Blei.

The NYTimes article outlines an iterative recommendation system, based roughly on the above paper that is consistently updating based on new signals. It:

> "(1) models content, (2) adjusts this model by viewing signals from readers, (3) models reader preference and (4) makes recommendations by similarity between preference and content."

Model Content

Uses Latent Dirichlet Allocation (LDA) algorithm to perform text analysis of articles. Determines and weights topics.

Adjust Model Based on User Input

Since words are dependent on context, taking the step to adjust assumptions about the topic content based on the user profiles of those who read the content, helps improve the modelling of the topic.

Adjust User Profile

Based on content and item profiles, iteratively adjusts user profile. Allows for some noise (clicked on articles they didn't like, etc).

Privacy

"NYTimes Recommendations does make use of your viewing history on NYTimes.com, but only in order to build your profile and to make accurate suggestions for content we think you might be interested in. We do not share your viewing history or trends with other users or other sites. NYTimes Recommendations fully complies with the New York Times Privacy Policy." - https://www.nytimes.com/content/help/extras/recommendations/recommendations.html#recommendationsq10

References

- Building the Next New York Times Recommendation Engine, Alexander Spangher; August 11, 2015 - 11:27 am
- https://www.nytimes.com/content/help/extras/recommendations/recommendations.html
- Collaborative Topic Modeling for Recommending Scientific Articles; Wang, Blei