

Metafilter.com Post Recommender

www.metafilterpostrecommender.com
github.com/tomasbielskis/metafilterpostrecommender

MetaFilter (MeFi) is a community weblog where users post links to content found on the web and engage in discussions about it.

MeFi harbors a wealth of content far exceeding what is possible for an individual user to explore. Popular posts get a lot more exposure than the equally deserving ones hiding in the long tail. This self-fulfilling quality of popularity is an evil that we ought to fight!

Two main goals

- Quantity: increase the number of users that can benefit from my methodology compared to collaborative filtering based on favorite ratings
- Quality: improve the quality of the recommendations compared to the base case of picking random posts

Process

1. Scrape all content from the site and parse all posts and comments.
2. Combine with publicly available MeFi metadata.
3. Natural language processing: a) stemming all text, b) converting words into vectors based on term frequency (TF) and inverse document frequency (IDF), c) non-negative matrix factorization (NMF) and latent dirichlet allocation (LDA).
4. User preference feature engineering:
 - Sources of signal on user tastes derived from the available data:
 - Posts written and favorited by the user
 - Comments written and favorited by the user
5. Identify the posts that have features closest to the user preferences by using a cosine similarity matrix.

Tech Stack

- Python libraries: numpy, pandas, sklearn, nltk, multiprocessing, flask
- MongoDB, AWS S3, EC2, Elastic Beanstalk

Results

- Recommendations for 32,000 MeFi users:
 - 4 times better than the reach of a pure collaborative filtering recommender
- Outperformed the random recommendations on the test set of posts
 - 100 times better recall score
 - 5 times better precision