

Designing A Recommender System Based on Social Networks and Natural Language Processing

Fan Yang, Guangjun Su, Jiayue Wan

March 24, 2017

1 Problem Statement

The influence of social network and users' reviews play a critical role while we are making decision. However, they have rarely been considered in the traditional recommender system. In this project, we design a new paradigm of recommender system which incorporates these two important factors.

2 Dataset

In this section, we introduce the dataset that we use for this project. Our dataset is obtained from the Round 9 of the Yelp Dataset Challenge. This dataset includes information about local businesses in 11 cities across 4 countries, see Table 1.

Table 1: Dataset Overview

Country	Cities
U.K.	Edinburgh
Germany	Karlsruhe
Canada	Montreal and Waterloo
U.S.	Pittsburgh, Charlotte, Urbana-Champaign, Phoenix, Las Vegas, Madison, Cleveland

This challenge dataset is comprised of the following components:

- 4.1M reviews and 947K tips by 1M users for 144K businesses;
- 1.1M business attributes, e.g., hours, parking availability, ambience
- Aggregated check-ins over time for each of the 125K businesses;
- 200,000 pictures from the included businesses

For more information of the dataset, please visit https://www.yelp.com/dataset_challenge

3 General Idea

Our general idea is to incorporate the traditional recommender system with the social network and the natural language processing of users' reviews.

3.1 Mining Social Network

Traditional recommender systems, either content-based recommender systems or collaborative filtering, consider user-item, user-user, item-item similarities. None of them are concerned about the fact that people are very likely to be influenced by their friends in the social network. So, we feel the need of incorporating the social network into the recommender system.

For instance, when estimating the rating of user x on item i , we find similar users, and weigh the ratings of these users on item i not only according to the traditional similarity measures, but also a measure of how close these users are to user x . Different measures of closeness may be investigated, and optimal measure will be used.

3.2 Mining Review Text

Users' feedback comes in the form of ratings and reviews. Rating is the explicit measurement of user's fondness for a specific restaurant/bar/salon, and user's review is the best supplement which can help us explain why the user like or dislike, such as price, service, environment, and flavor etc.

We use Latent Dirichlet Allocation (LDA) to discover hidden labels/topics of each review text, and then construct a user profile and item profile. User profile and item profile can be applied to the content based recommender system, and also help to interpret the hidden factors and hidden topics of the Latent Factor models.

4 Performance Evaluation

We will measure the RMSE of our recommender system. As a baseline for comparison we will use traditional latent-factor recommender.

5 Expected Submission

This project is expected to make some improvement to the recommender system through combining the traditional recommender methods with social network analysis and natural language processing. By the end of the quarter, we plan to accomplish the following tasks:

1. to design and implement a new recommender system based on social network and natural language processing;
2. to write up a deliverable report and poster;

3. to build up a web application to vividly demonstrate how this recommender system work if time is permitted.

6 Biography

Fan Yang

1. Worked at a AI startup, and had experience in developing revenue management model;
2. Interested in big data mining and natural language processing;
3. Took CS 221, CS 246, and CS 224, and familiar with ML and NLP;
4. Proficient at Java, R, and Hadoop;
5. Want to learn more about the recommender system.

Guangjun Su

1. Worked at a AI startup, and had experience in developing revenue management model;
2. Interested in machine learning, artificial intelligence and big data;
3. Took CS 229 and CS 246 and have solid foundation of ML and big data processing;
4. Proficient at programming, like Java, Python, and SQL;
5. Want to apply what I have learned from courses into solving real world problem.

Jiayue Wan

1. Interested in machine learning and natural language processing
2. Took CS 246 in Winter 2017, familiar with machine learning and big data mining algorithms
3. Proficient at Java, Python; experience with Hadoop
4. Dedicated to applying the knowledge learned in CS 246 to solving real world problems