

Deliverable 1

Project Overview

Pink Team

James Doolan (Project Manager)
Katharine Cooney (Communications)
Liam Creagh (Business)
Shuyu Huang (UX)
Kang Li (Developer)

June 16, 2015

1 Introduction

1.1 Overview

Our project is a content-based news recommender system, which will make recommendations based on social media behaviour. User data from Twitter (tweets, followed users, retweets), and Reddit (comments, upvotes, downvotes, subreddit subscriptions etc.) will be collected and used to provide recommendations.

We will initially use a simple method of generating recommendations, and we are currently researching several more sophisticated approaches (see Section 1.3). In addition to the content-based method we are currently utilising, we also hope to incorporate basic implicit feedback, *i.e.* click-throughs, and some functionality to allow users to remove certain articles from their feed.

1.2 Motivations

Our method of recommendation circumvents several problems inherent in recommender systems, *e.g.* the “cold start” problem, where the system cannot provide accurate recommendations if there are few ratings to begin with. Several studies in this area have shown that users are unlikely to supply explicit ratings [1], and in some cases, that users’ declared interest did not strongly predict their actual interest.[2]

1.3 Related Work

During our research into this area, we found many projects that attempt to address similar issues. One that we found particularly interesting was *Who Likes What?*, a web application

which uses Twitter's Lists functionality to determine latent user-topic relationships[3]. (The application is available at <http://twitter-app.mpi-sws.org/who-likes-what/>). We intend to explore these ideas more fully after the first prototype has been completed.

1.4 Success Criteria

Kapanipathi, Pavan, et al.[3] state that "one of the best ways of evaluating the quality of inferred interests is through direct human feedback". They use volunteers to provide feedback on their recommendations by presenting them with a choice of recommendation lists from unidentified sources. The users are then asked them to rank the lists on accuracy and completeness. We propose to use this methodology to assess our recommendations. The success criterion would be the percentage of users who ranked our recommendations as number 1. We still need to determine what listings we would be comparing to.

1.5 Technology

We will be using Django and Python to build our web application. The application will be backed by a PostgreSQL database. Article content will be obtained by parsing RSS feeds, following the urls to the articles and parsing their HTML.

2 Minimum Viable Product

2.1 Features

Our minimum viable product will be a condensed form of the full version. The main features of the MVP are as follows:

- User data will initially be gathered from Twitter only, in order to familiarise ourselves with the process of integrating with an API. We will move towards including other sources of user profiles (Reddit, Facebook) at a later stage in the project.
- Content data will be obtained through the Reuters RSS feed. We felt this was a good place to start, as we can obtain a wide variety of news from a respected source. We will eventually include other content sources.
- Basic signup and login functionality; the MVP will have a "connect with Twitter" option. Reddit connection will be added at a later stage.
- Very simple method of recommendation; we will take the most common words from the user's Twitter and compare with the articles' content, recommending the articles with the highest occurrence of those terms.

2.2 Desired Outcome

The main objectives for our MVP are as follows:

- Lay down the foundation for our application, *i.e.* basic functionality, leaving room for expansion and improvements in the later stages of the project
- Familiarise ourselves with new concepts and material (*e.g.* Twitter API, Django, natural language processing *etc.*)
- Produce some kind of recommendation, the quality of which is not expected to be high.
- Work together as a team using agile methodologies.

3 Development Methodology

3.1 Scrum

Our team will be following the Scrum methodology in this project, and we have decided to use Trello to organise our product and sprint backlogs. We have daily meetings and coding sessions, and keep each other apprised of our progress.

References

- [1] Doychev, Lawlor, et al. "An Analysis of Recommender Algorithms for Online News", *CLEF*, 2014.
- [2] Lavie, Talia, et al. "User attitudes towards news content personalization." *International journal of human-computer studies* 68.8 (2010): 483-495.
- [3] Bhattacharya, Parantapa, et al. "Inferring user interests in the twitter social network." *Proceedings of the 8th ACM Conference on Recommender systems*. ACM, 2014.