

PERSONALIZED NEWS RECOMMENDER SYSTEMS

Machine Learning ITCS6156

Project proposal

This document contains a brief idea on news recommender system project, timeline, tasks and primary details of data.

Divya Kulkarni
Ketki Savle
Rakesh Harish
Yangqi Su

Introduction

Reading news online has become very popular as the web provides access to news articles from millions of sources around the world. However, a key challenge of news websites is helping users find the articles that fits reader interests. This not only includes articles that may be of interest to the general public, but also news recommendations that are tailored to each individual reader. This challenge becomes even more pressing since an individual's preferences may change with time. Our aim for this project is to build a personalized news recommender system that allows user to see the news as per his/her interest and works efficiently to suggest a user what they would like to read next.

Problem statement:

Recommending news articles is one of the most challenging recommendation tasks. The news domain differs from other domains in many ways. For example; the popularity and recency of news articles evolves quickly over time. So focusing on the recency issue becomes more challenging than it is in other domains. Furthermore, some news articles may share chronological connections such that the user may want to read the previous news items related to the one she has already read or she may want to be kept informed about the current topic. By only learning user preferences can lead to an unsatisfactory solution to news recommendation. This is because the user may want to read news articles that are not necessarily of her interest but rather of general importance. An example would be wanting to read the news about elections even if there isn't actual interest in politics. Adding to the complexity of the online learning problem is the sheer number of new articles published every hour.

Approach:

1. Data Collection and Preprocessing:

Gathering the news data from web and preprocessing the data for further analysis.

2. Content based filtering:

In content-based recommendations, the properties of items are used to make recommendations. Items which have similar properties with the user's previous preferences are recommended to the user. Bayesian framework could be used to compute the same.

3. Collaborative filtering:

In collaborative filtering, recommendations are done by using other people's preferences which are similar to the user's past preferences. This could be computed using classification algorithms.

4. Final Model:

Merge content and collaborative filtering and build personalized recommendation model.

Timeline:

- **Phase 1** Gathering user data for Yahoo news/ Google news
- **Phase 2** Cleaning and preprocessing the dataset. Visualize the information.
- **Midterm Report**
- **Phase 3** Build classification of information
- **Phase 4** Build user preference algorithm

Team Members:

Divya Kulkarni, Ketki Roplekar, Rakesh Harish, Yangqi Su

Roles and Responsibilities:

At this point of time since this is a learning phase for all group members, all tasks will be performed in collaboration.

Questions that we want to answer

- Will the recommender system be able to identify same topics that are named differently. For example, will Action Movie and Action film be treated as different items or same?
- Will the recommender system be able to identify changing user interests? For instance, a user may read news not because he/she interested in the topic but because she found it important.
- Will data sparsity impact the performance of the recommender system?

Expectations (what we will be learning)

- To be able to build an engine which classifies the content correctly.
- Learn the classification of real world data efficiently.
- Learn how recommender systems can be applicable to various fields like e-commerce, music, movies, travel, etc.
- Unlike the search engines recommender system, learn to bring the information to the users without any manual search effort.

Novelty

Most users either spend a lot of time trying to find an interesting article on a single website, or they just read the front-page news, both of which is not adequate. When we consider the plethora of news sources on the internet, one can spend plenty of time just reading the news. News recommender systems aim to give the most relevant article recommendations to users suited to their personal interests and preferences. Hence this idea is novel.