

PROJECT PROPOSAL

Project Members : Alekya Kumar, Prachi Sawant, Trupti Jadhav

Project Topic : Dating website Recommender System

Data Set : The dataset and its details can be found at the below link.

<http://www.occamslab.com/petricek/data/>

Problem Description:

The aim of the project is to build a recommender system that enables users to find and introduce themselves to new personal connections over the Internet. The datasets contains 17,359,346 anonymous ratings of 168,791 profiles made by 135,359 LibimSeTi users as fetched on April 4, 2006.

All ratings are contained in the file "ratings.dat" and are in the following format:

UserID, ProfileID, Rating

- UserID is user who provided rating
- ProfileID is user who has been rated
- UserIDs range between 1 and 135,359
- ProfileIDs range between 1 and 220,970 (not every profile has been rated)
- Ratings are on a 1-10 scale where 10 is best (integer ratings only)
- Only users who provided at least 20 ratings were included
- Users who provided constant ratings were excluded
- Size ~250MB.

User gender information is in the file "gender.dat" and is in the following format:

UserID, Gender

- Gender is denoted by a "M" for male and "F" for female and "U" for unknown
- Size ~2MB

Input variables: The ratings.dat file has UserID, ProfileID, Rating as input variables. While gender.dat has UserID, ProfileID.

Analysis Methods:

Basic analysis on the dataset and implementing collaborative filtering for building the recommender system.