BIM 476 DATA ACQUISITION AND PROCESSING
FINAL PROJECT

Due Date: Hand in your project including every required document by the last class.
Students can work in groups of two people.

Project Description

1. Study collaborative filtering (CF), memory- and model-based CF algorithms, and accuracy measures like mean absolute errors, normalized mean absolute errors, classification accuracy, F-measure, and so on.
2. Download the paper entitled An algorithmic framework for performing collaborative filtering (Herlocker et al.). Study the paper and the CF algorithm proposed by Herlocker et al. Please briefly discuss the methods they propose to increase the overall performance of CF.
3. Download Jester data set and study it. You can select another data set collected for CF purposes.
4. Prepare a report discussing the first three parts and submit it by the first midterm.
5. Using MATLAB, perform some experiments based on Jester to determine the optimum values of some parameters like number of users, various normalization techniques, and so on.
6. Apply some preprocess tasks to your data set: Fill missing values, normalize ratings, select task relevant data, transform ratings, and so on.
7. After determining optimum values of some parameters based on experimental results, write a program for the CF algorithm proposed by Herlocker et al. using a programming language that you choose.
8. Prepare a web site to generate both predictions for single items and top-*N* recommendations based on Jester or the data set that you chose.
9. Your web site should generate error messages as necessary.
10. Write a final report about your project and hand in it before your presentation.
11. Each group is expected to present their works in the last class.
12. If you need any help, please see the TA or the instructor.