Assignment 1 Due Date – 21st August, 11:30 AM 10 marks

Use the Movielens dataset 100K and 1M dataset.

The dataset is already segregated into 5 parts. You will use 4 parts for training and 1 part for testing. Say, you use sets 1, 2, 3, 4 for training, then you use 5 for testing. Next time, you will use 1, 2, 3, 5 for training and set 4 for testing, and so on. This is 5 fold cross validation.

The metric for testing accuracy is normalized mean absolute error (NMAE).

1. Use user-based model for predicting ratings. _____ (6)

2. Use item-based model for predicting ratings. _____ (+1)

3. Use significance weighting and show variations by changing the number of neighbours _____ (+1)

4. Use variance weighting _____ (+1)

5. Use neighbourhood selection and show variation with changing threshold and changing number of neighbours _____ (+1)

For each of 1-5 you have to generate one table reporting results of 5 fold cross validation. For questions 3 and 5, you additionally have to show graphs (for variations)

Bonus marks (+2) for students who will implement the combined user-item approach.

Link http://siplab.tudelft.nl/sites/default/files/sigir06_similarityfusion.pdf