

ABSTRACT

Collaborative filtering systems use a database of user preferences to find users with similar preferences in order to recommend items to a given user. A central part of such systems, on which the quality of the recommendations depends, is the measure used to determine similarity of preferences. When a user asks for recommendations, the system uses the preferences of other similar users to predict the preferences of the current user for unpurchased or unseen items. In this thesis, a probabilistic distance measure approach is used in creating an algorithm that has desired properties of a good similarity measure. The distance between two users' preferences is defined as the probability that two randomly chosen items are ranked differently by the two users. The algorithm is empirically evaluated against a commonly used collaborative filtering algorithm based on the Pearson correlation coefficient. The exhaustive experiments show the behavior of the algorithms over various databases and show the optimal similarity neighborhood for each algorithm