Chapter II
Overview of Design Options for Neighborhood-Based Collaborative Filtering Systems

Nikos Manouselis
Informatics Laboratory, Agricultural University of Athens, Greece

Constantina Costopoulou
Informatics Laboratory, Agricultural University of Athens, Greece

ABSTRACT

The problem of collaborative filtering is to predict how well a user will like an item that he or she has not rated, given a set of historical ratings for this and other items from a community of users. A plethora of collaborative filtering algorithms have been proposed in related literature. One of the most prevalent families of collaborative filtering algorithms are neighborhood-based ones, which calculate a prediction of how much a user will like a particular item, based on how other users with similar preferences have rated this item. This chapter aims to provide an overview of various proposed design options for neighborhood-based collaborative filtering systems, in order to facilitate their better understanding, as well as their study and implementation by recommender systems' researchers and developers. For this purpose, the chapter extends a series of design stages of neighborhood-based algorithms, as they have been initially identified by related literature on collaborative filtering systems. Then, it reviews proposed alternatives for each design stage and provides an overview of potential design options.

INTRODUCTION

About two decades ago, Malone, Grant, Turbak, Brobst, and Cohen (1987) provided an overview of intelligent information sharing systems, identifying a fundamental categorization of systems that support access to highly dynamic information resources (Belkin & Croft, 1992; Baudisch, 2001; Hanani, Shapira, & Shoval, 2001). More specifically, they distinguished:
Overview of Design Options for Neighborhood-Based Collaborative Filtering Systems

1. **Cognitive filtering** systems such as the ones that characterize the contents of an information resource (shortly referred to as an *item*) and the information needs of potential item users, and then use these representations to intelligently match items to users; and

2. **Sociological filtering** systems such as the ones that are working based on the personal and organizational interrelationships of individuals in a community.

Early information sharing systems belonged to the first category and were based on text-based filtering, which works by selecting relevant items according to a set of textual keywords. **Collaborative filtering** systems were first introduced as representatives of the second category. They addressed two shortcomings of text-based systems (Konstan, 2004):

1. The often overwhelming number of on-topic items (ones that would be all selected by a keyword filter) which has been addressed by the introduction of further evaluating the items based on human judgment about their quality, and

2. The issue of filtering non-text items which has been addressed by judging them on subjective criteria such as human taste.

In general, the problem of collaborative filtering is to predict how well a user will like an item that he has not rated (also called “evaluated” in the rest of this chapter), given a set of historical ratings for this and other items from a community of users (Herlocker, Konstan, & Riedl, 2002; Adomavicius & Tuzhilin, 2005). The problem space can be formulated as a matrix of users vs. items (or user-rating matrix), with each cell storing a user’s rating on a specific item. Under this formulation, the problem refers to predicting the values for specific empty cells (i.e., predict a user’s rating for an item).

A plethora of collaborative filtering algorithms have been proposed in related literature. The most popular approach is the adoption of traditional neighborhood-based techniques (Cover & Hart, 1967), which are appropriately adapted to suit collaborative filtering needs and to produce a rating prediction based on the ratings of a relatively small number of neighbors with preferences similar to the ones of the targeted user. Different aspects of neighborhood-based algorithms are explored by various researchers, leading to a wide variety of proposed design options for a collaborative filtering system. On the other hand, since the study of Herlocker et al. (2002), there has not appeared in the literature an overview of the various proposed design options and how they can be considered for implementation in an integrated manner.

In this direction, this chapter aims to provide an overview of various proposed design options for neighborhood-based collaborative filtering systems, in order to facilitate their better understanding, as well as their study and implementation by recommender systems’ researchers and developers. For this purpose, the chapter builds upon a series of design stages for neighborhood-based collaborative filtering algorithms, as they have been initially identified by related literature (Herlocker et al., 2002; Vozalis & Margaritis, 2003; Sarwar, Karypis, Konstan, & Riedl, 2000) and enhanced by our study (Manouselis & Costopoulos, 2006a). Then it reviews proposed alternatives for each design stage and provides an overview of potential design options.

More specifically, the chapter is structured as it follows: In the next section, the background of this study is presented, by introducing neighborhood-based algorithms for single-attribute collaborative filtering and outlining seven generic stages of a neighborhood-based algorithm. Each stage is further discussed, and methods for its implementation are reviewed. We next provide a synopsis of the reviewed design options, identify possible further extensions, and present the conclusions of this study. The final section outlines directions for future research.
Related Content

Feature Subset Selection Using Ant Colony Optimization for a Decision Trees Classification of Medical Data
[www.igi-global.com/article/feature-subset-selection-using-ant-colony-optimization-for-a-decision-trees-classification-of-medical-data/210061?camid=4v1a](www.igi-global.com/article/feature-subset-selection-using-ant-colony-optimization-for-a-decision-trees-classification-of-medical-data/210061?camid=4v1a)

Context Window Based Co-occurrence Approach for Improving Feedback Based Query Expansion in Information Retrieval

Comparing Different Sparse Matrix Storage Structures as Index Structure for Arabic Text Collection
[www.igi-global.com/article/comparing-different-sparse-matrix-storage/74784?camid=4v1a](www.igi-global.com/article/comparing-different-sparse-matrix-storage/74784?camid=4v1a)

Very Low Bit-rate Video Coding
[www.igi-global.com/chapter/very-low-bit-rate-video/30763?camid=4v1a](www.igi-global.com/chapter/very-low-bit-rate-video/30763?camid=4v1a)