ABSTRACT

A plethora of collaborative filtering algorithms have been proposed in related literature. Due to the dynamic and changing parameters of the various application contexts, careful testing and parameterization has to be carried out before an algorithm is finally deployed in a real setting. This paper investigates how a previously proposed tool for simulated testing of collaborative filtering algorithms, called the Collaborative Filtering Simulator (CollaFiS), can be expanded to an e-science environment for researchers so that it runs over a digital research infrastructure. More specifically, a survey of design options for neighborhood-based collaborative filtering systems is carried out in order to illustrate the variety of requirements that need to be met. Then, a number of usage scenarios that could be supported by an e-science environment for collaborative filtering research are presented. Three example dataset cases are used to illustrate how the new version of CollaFiS can support research and experimentation on collaborative filtering algorithms using different data and various parameter options. Overall, this paper showcases how e-science environments and infrastructures may facilitate the research activities of people working on recommender systems.

INTRODUCTION

Plenty of evaluation studies of recommender systems (Breese et al., 1998; Deshpande & Karypis, 2004; Herlocker et al., 2002; Papagelis & Plexousakis, 2005) indicate that careful testing and parameterization has to be carried out, before a recommender system is finally deployed in a real setting. On the other hand, many recommenders usually remain at a design or prototyping stage of development. Testing methods and tools that may support their systematic implementation and evaluation in the context of real-life applications are limited. Experimental testing for recommender systems could be greatly facilitated by e-science environments that can also give to researchers the computational, memory and storage power that large scale simulations require.

Towards an e-Science Environment for Collaborative Filtering Researchers

Nikos Manouselis, Agro-Know Technologies, Athens, Greece
Giannis Stoitsis, Agro-Know Technologies, Athens, Greece

Keywords: Collaborative Filtering, Collaborative Filtering Simulator (CollaFiS), e-Science, Neighborhood-Based Algorithms, Recommender Systems

DOI: 10.4018/ijdls.2014010104
(http://duineframework.org) that can be used to set up, parameterize, and evaluate a recommender system. In most cases they are software libraries and frameworks rather than providing a graphical user interface for researchers and features for research preparatory tasks such as handling the rating datasets. A software environment that provides both a graphical interface for researchers, as well as handles multi-criteria rating datasets was proposed by Manouselis and Costopoulou (2006b). It was a first attempt to create an integrated web application for performing experiments with ratings data sets for recommender systems using a graphical interface for parameterising the algorithms, but it was implemented using ad hoc and hard coded versions of the algorithms and not providing a complete coverage of use cases that a collaborative filtering researcher would need. In addition, it was hosted by a typical web server that did not have the computational, memory and storage capacity to carry out the variety and complexity of simulations that such experiments require. This is where the existing toolkits and frameworks could provide an opportunity to enhance such tools and evolve them into integrated e-science environments that can fulfill the requirements of researchers and developers of recommender systems.

This paper investigates how such an old fashioned and outdated tool, called the Collaborative Filtering Simulator (CollaFiS), can evolve into a modern software environment that may facilitate the work of researchers and run over a digital research infrastructure. First, it reviews the variety of design options that exist for neighborhood-based collaborative filtering in order to illustrate the range of parameterization requirements that need to be met. Then, a number of typical usage scenarios are presented, such as importing a data set, transforming a data set, visualizing a data set, and generating synthetic data. Then a new version of CollaFiS that has been enhanced to support these usage scenarios is presented, and results of working with three sample datasets are shown - a dataset from the Mendeley platform (http://dev.mendeley.com/datachallenge/), a dataset from a portal for European schools (http://lreforschools.eun.org), and a dataset from an agricultural education portal (http://organic-edunet.eu).

**BACKGROUND**

More than two decades ago, Malone et al. (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 et al., 2001). More specifically, they distinguished:

1. **Cognitive filtering** systems 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 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 the second category. They addressed two shortcomings of text-based systems (Konstan, 2004):

- 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.
- 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...
OpenDLib: A Digital Library Service System
www.igi-global.com/chapter/opendlib-digital-library-service-system/19864?camid=4v1a