Available online at www.sciencedirect.com
ScienceDirect
Electronic Commerce Research and Applications 6 (2007) 139–145
Electronic Commerce Research and Applications
www.elsevier.com/locate/ecra

Common structure and properties of filtering systems

Junichi Iijima *, Sho Ho

Graduate School of Decision Science and Technology, Tokyo Institute of Technology, Tokyo, Japan

Available online 11 December 2006

Abstract

Recommendation systems have been studied actively since the 1990s. Generally, recommendation systems choose one or more candidates from a set of candidates through a filtering process. Methods of filtering can be divided into two categories: collaborative filtering, in which candidates are chosen based on choices of other persons whose interests or tastes are similar, and content-based filtering, in which items are chosen based on the profile or action history of the recommendee. However, these methods share the same structure in the sense that both of them recommend items based on relevance degrees of items and references, as well as relevance degrees between the recommendee and each reference. Most discussions about recommendation systems focus on the methods of choosing recommended candidates; few focus on foundational concepts of recommendation conditions that systems must satisfy, and problems that current systems have compared with these conditions. In this paper, recommendation systems are reconsidered from the viewpoint of multi-criteria decision making. Conventional filtering methods (e.g., collaborative filtering and content-based filtering) are formulated as linear weighted sum type recommendation systems. Several properties of linear weighted sum type recommendation systems are identified and formulated from the viewpoint of voting.

Keywords: Social choice theory; Collaborative filtering; Content-based filtering; Recommendation system

1. Introduction

As an information and communication technology (ICT) application that meets the growing needs of "personalization" in our advanced information society, recommendation systems have been studied actively since the 1990s. ICT-enabled recommendation systems (e.g., online shopping systems that recommend products based on customer profile and history of customer actions; enterprise knowledge portals that send necessary information in a timely manner to each employee according to specialty and position) are infiltrating various aspects of our life. On the other hand, due to the rapid growth of ICT, an enormous amount of information that exceeds the capability of human information processing is now being distributed via various networks. To handle the flood of information, recommendation systems that effectively collect and choose information based on objectives and preferences of users are becoming indispensable.

Early studies of ICT-enabled information recommendation systems include Tapestry [1], GroupLens [2], and Fab [3]. The term "collaborative filtering" was first used in 1992 by Goldberg et al. in their paper that introduced an information distribution system named Tapestry. In Tapestry, users can set rules such as "if Joe and Bill receive a message, then I would like to receive that message, too" and filter messages based on the rules. The collaborative filtering in Tapestry was conducted semi-automatically. In the late 1990s, GroupLens, which conducts collaborative filtering automatically, was developed. GroupLens is a system for collaborative filtering of online news in which the relative degree of users is calculated based on their rating of articles, and articles are recommended based on the rate given by highly relative users. Meanwhile, hybrid recommendation systems that integrate content-based filtering and collaborative filtering have also been developed. Fab, which recommends WWW pages to users, is a typical hybrid recommendation system. In the Fab system, users

* Corresponding author.
E-mail address: iijima@me.titech.ac.jp (J. Iijima).

1567-4223/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.elerap.2006.11.002

receive items that score highly against the user profile, as well as items which are rated highly by a user with a similar profile [3].

Generally, recommendation systems choose one or more candidate from a set of candidates through a filtering process. Methods of filtering used in recommendation systems can be divided into two categories. One is collaborative filtering, in which items are chosen based on the choices of other persons whose interests or tastes are similar to those of the *recommendee* (who receives the recommendations) to some degree. Examples of collaborative filtering systems include GroupLens [2], RINGO [4], and Jester [5]. Another method is content-based filtering, in which items are chosen based on the profile or action history of the recommendee. Examples of content-based filtering system include the movie selection system proposed by Alspector et al. [6], MetaSEEk [7], and the book recommend system developed by Mooney and Roy [8]. These two filtering methods are based on different ideas. That is, in collaborative filtering, an item will be chosen based on the choices of other persons whose interests or tastes are similar to those of the recommendee, while an item will be chosen based on the profile or action history of the recommendee in content-based filtering.

However, they share the same structure in the sense that both of them recommend items based on relevance degrees of items and references, as well as relevance degrees between the recommendee and each reference. Highly relevant candidates with highly relevant references are recommended to use filtering systems in both types.

A considerable amount of research has been conducted on recommendation systems, most of which propose concrete designs for software. Meanwhile, few studies have been made about relevant foundational concepts and the theories. Nevertheless, it is important to identify basic conditions that a system needs to meet to be a recommendation system. For example, a recommendation system should recommend x than y if everyone prefers x to y in collaborative filtering. Or it should protect a user to lead the result of recommendation to her intended direction by making a fraud preference declaration for certain candidates in collaborative filtering. Adding that, it is desirable for a recommendation system to recommend the first runner-up in the set of candidates if the most preferable candidates are removed.

Also it is necessary to check the various proposed recommendation systems to see if they meet these conditions. When we choose a candidate from several items based on multiple orders of preference, our action can be understood in the framework of voting. The mechanism of voting has been investigated for decades in the field of social choice theory. The correspondence between collaborative filtering and social choice theory has been explored by Pennock et al. [9]. Pennock et al. considered collaborative filtering as systems that combine preferences and focused on the concepts of universal domain (UNIV) and unanimity (UNAM) in social choice theory. Based on these concepts, several properties of collaborative filtering systems were identified [9]. However, some other important properties of recommendation systems which are related to the voting schemes in social choice theory were not addressed in Pennock et al.'s work.

In this paper, we introduce a new formulation of collaborative filtering and content-based filtering and put our focus on linear weighted sum type recommendation systems. Based on social choice theory, several properties of linear weighted sum type recommendation systems are discussed.

2. Common structure of filtering systems

Recommendation systems assume the preference of the recommendee in some way and choose one or more items from a set of items through a filtering process. Collaborative filtering systems choose items based on choices of other persons whose interests or tastes are similar to those of the recommendee to some degree. Content-based filtering systems choose items based on the profile or action history of the recommendee. Recommendation systems choose a set of recommended candidates out of a set of candidates based on a set of references (i.e., other users or various attributes in user profiles), the relevance degree between each reference and the recommendee, and the relevance degree between each reference and candidates.

The concept of recommendation problem is defined as follows:

Definition 1 (Recommendation problem). *A recommendation problem RP is a quadruple* $\langle C, R, \rho, \sigma\rangle$ *where*:

- C is the set of candidates $C = \{c_j | j = 1, \ldots, m\}$.
- R is the set of references $R = \{r_i | i = 1, \ldots, n\}$.
- $\rho : R \rightarrow \Re$ is a function that denotes the *relevance degree between reference and the recommendee*. Here we assume that $\rho(r) \geqslant 0$ for any $r \in R$, which means that no reference has a negative correlation with the recommendee.
- $\sigma : C \times R \rightarrow \Re$ is a function that denotes the *relevance degree between candidate and reference*. Here we assume that $\sigma(x,r) \geqslant 0$ for any $x \in C$, $r \in R$, which means that no reference has a negative correlation with any candidate.

Let us denote the set of recommendation problems by $\boldsymbol{RP}$ and let

$$\boldsymbol{C} = \{C_0 | (\exists \langle C, R, \rho, \sigma\rangle \boldsymbol{RP})(\boldsymbol{C_0} \subset \boldsymbol{C})\}.$$

In a recommendation problem, $\sigma(x,r)$ denotes the relevance degree of reference r, and candidate x. For example, in collaborative filtering, $\sigma(x,r)$ shows the rating that user r gives to candidate x, while in content-based filtering, $\sigma(x,r)$ denotes the value of feature r of candidate x. In the case of GroupLens [2], $\rho(r)$ is determined by the correlation between preference of the target user and users who serve as references, thus $\rho(r)$ may have a negative value, while another collaborative filtering system SPARS-J [10] uses the