Chapter V

Information Filtering and Personalization Services

Chunxiao Xing
Tsinghua University, China

Chun Zeng
Tsinghua University, China

Zhiqiang Zhang
Tsinghua University, China

Lizhu Zhou
Tsinghua University, China

Abstract

Personalization service is becoming one of the core services in digital libraries, and an exciting and challenge research area. In this chapter, we analyze several key technologies and the related works in information filtering and personalized services, and then present a content-based personalized searching algorithm and a probabilistic model to represent user interests, which is more effective than the vector space model by the experiments. To solve the data sparsity and scalability problems in collaborative filtering, we present new methods for similarity computation and instance selection. The experiments show it is higher predicted precision.
and performance than the others. Based on the above research results, we design and develop a prototype, TH-PASS, which provides personalized searching and recommending services.

Introduction

With the rapid advancement of information technology and the knowledge economy, digital libraries (DLs) have emerged as the large scale and distributed information and knowledge environment and infrastructure to bring together collections, services, and people in support of the full life cycle of creation, dissemination, use, storage, and preservation (Xing et al., 2002).

To quickly and easily gather useful knowledge and to alleviate information overload problems, it has therefore become necessary to provide users with active and personalized service mechanisms that automatically extract only relevant information. To overcome the above challenge, one of the exciting and hot research areas is personalized active service that can route, recommend, rank and filter documents based on users’ interest profiles. This service is becoming one of the core services in the future DLs, and it relies on many established techniques applied in information retrieval (IR), information filtering (IF), user modeling and machine learning, and etcetera.

In this chapter, we mainly focus our attention on information filtering technology and its application on personalization service: (1) Exploring the basic problems of content-based filtering including feature selection, representation and revision of user profile; (2) Tracking, studying and representing users’ interests based on information filtering technology; (3) Analyzing the performance of several feature selection methods and comparing the different representation of user interests; (4) Researching the collaborative filtering to solve the data sparsity and scalability problems in collaborative filtering; (5) Presenting a class-based method for similarity computation and an approach for instance selection; (6) Comparing the performance of several collaborative filtering algorithms and giving the experiment results; (7) Building a prototype system to provide the users personalized filtering and notification service based on user modeling and profile learning.

The rest of the chapter is organized as follows. The second section gives a brief overview of related works. The third section discusses the representation of document and user interest, analyzes the methods of user interests’ revision, and provides a content-based filtering personalized service algorithm. The fourth section then provides the problem description in collaborative filtering algorithms, discusses our method for similarity measure, and describes the technique