

Mining Web logs for a Personalized Recommender System

Sutheera Puntheeranurak[†] Hidekazu Tsuji[‡]
 Graduate School of Engineering School of Information Technology and Electronics
 Tokai University[†] Tokai University[‡]

1. Introduction

As the Web rapidly grows, however, the number of matching pages increases at a tremendous rate when users use the search engine for finding some information. It is not easy for a user to retrieve the exact information he/she requires. In particular, browsing a web set is an expensive operation, both in time and cognitive effort. Recommender systems have then become valuable resources for users seeking intelligent ways to search through the enormous volume of information available to them. In this paper we propose a new framework based on web logs mining for building a personalized recommender system. At the core of personalization is the task of building a profile of the user. We have developed an approach that user's information learned from user's web logs data to construct accurate comprehensive individual profiles. One part of this profile contains facts about a user, and the other part contains rules describing that user's behavior. We use Web usage mining to derive the behavioral rules from the data.

2. Background

The aim of personalization is to select data whose content are most relevant to the user from a greater volume of information and to present them in suitable way for the user. Then collecting Web documents according to a user profile are not an easy task. A number of systems that incorporate AI techniques have been developed to support it. In general, these systems operate in five stages as shown in Figure 1:

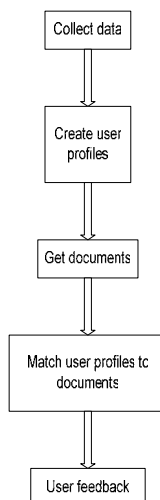


Figure 1. General operation in personalization system

- First, they collect data about the user's preferences based either on observations of user behavior during Web surf or on explicit user input.
 - Second, they extract data features, such as common words or phrases, and use them as input to an algorithm for creating user profiles.
 - In the third stage, the system attempts to discover documents similar to those that generated the profile. This task requires a Web search mechanism, such as a search engine.
 - Next, the system compares the newly discovered documents to the user's profile and presents the most relevant ones to the user.
 - Finally, the system process the user's relevance feedback to modify the profile.
- Then the process iterates from the third stage.

This process describes a personalization system. One main variation can occur in the third stage. However, recommender systems also differ at the second stage, in how they create the user profile. Then we present a framework of a personalized recommender system by using web logs mining to create the Meta profile as shown in the next section.

3. A framework of a Personalized Recommender System

The personalized recommender system based on Web logs mining consists of a set of interconnected agents. Figure 2 illustrates its main components and overall architecture. Each agent has a specific job and contributes through a common interface to produce a modular but integrated system.

The personalized recommender system is divided into two processing: offline processing and online processing. A classification agent is working in offline processing to produce the user's Meta profiles. The online processing includes a monitoring agent, a learning agent and a recommendation agent. The monitoring agent is used to capture the active user session and create user data file and web logs data. The recommendation agent is used to generalize recommendation list to user. The learning agent is responsible to update the user's Meta profiles according to user's behaviors or their user data, so as to adapt to the user's most recent preferences

Monitoring Agent

This agent includes registration module and Web log filtering module. For a registered user, the system records the following information: id, name, sex, city, province, country, born date, hobby. This information

is provided by the user in a web form at the time of registration. User's behavior is monitored by keeping web logs data. In addition to web access logs, our given input includes personal data on a subset of users, namely those who are registered to this system.

Although Web logs are potentially able to provide useful knowledge for making recommendations, the raw log data cannot be used before appropriate pre-processing. Therefore, we first convert raw Web logs into a set of user transactions before performing the log mining task. The approach contains three sequential steps: data cleansing; user session identification; and transaction identification.

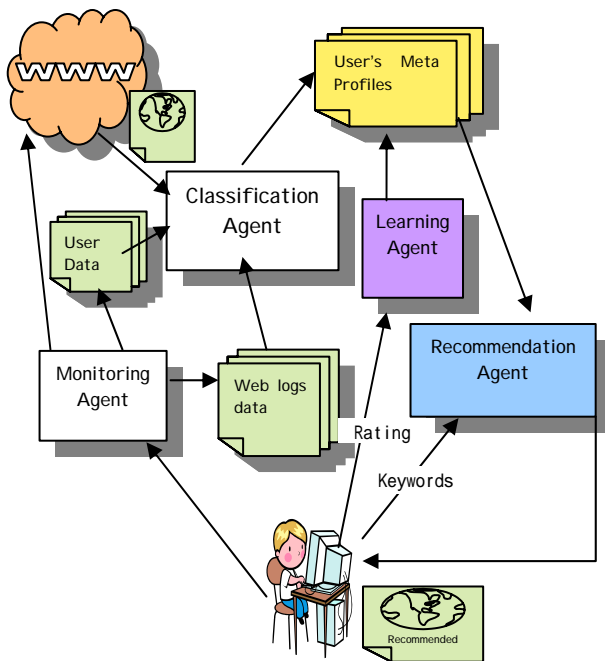


Figure 2. A personalized recommender system's framework

Classification Agent

After we have developed a data mart of Web logs specifically to support Web personalization analysis. A data set was collected from the crawling and parsing process. We use Web usage mining to extract the important information from Web pages and apply rule discovery methods individually to every user's Web logs data by using an association rules algorithm. The results are a set of rules in the form of "IF (preconditioned set of Web pages) THEN (post-conditioned Web page)." We keep then factual information's user and those rules in the user's Meta profile to describe the customer's behavior.

Learning Agent

The learning agent uses AI techniques to accomplish the difficult task of learning a user's interest. At the same time, if the factual user's information or user's behavior has changed, this agent will adapt the Meta profile. The learning agent also

presents all new information collected by recommendation agent. The user's positive or negative feedback on recommendation agent is used to modify the Meta profile.

Recommendation Agent

We propose two recommendation approaches: content-based filtering algorithm and collaborative filtering algorithm. This system recommends Web pages based on the similarity between the user's Meta profiles. To resolve the cold start problem, we first use the support vector machine to process content-based filtering for finding the recommended Web pages. When the user has received and looked over his/her recommendations, he/she are required to assign appropriate ratings. The user's ratings are forwarded to the learning agent. After we have received the user's rating, we can process the collaborative filtering by using clustering technique to find the recommended Web pages.

4. Conclusions

In this paper, we have proposed a framework of a personalized recommender system that based on the Web logs mining for making recommendations. It is proposed to improve the efficiency and effectiveness of the traditional information retrieval system. The personalized recommender system consists of three sequential steps:

- (1) data preparation of the Web logs data;
- (2) usage Web log mining; and
- (3) Generate the Web page recommendation.

We proposed three alternatives for identifying transactions from Web logs data and discussed an association rule approaches for making the user's Meta profile. Our recommendation framework identifies the recommended Web pages that relevant with the Meta profile's users. In the future studies, we will test and establish the proposed system using real life data. To demonstrate the potential usage of the framework, a system prototype for a personalized recommender system is designed and implemented as a form of a Web-site navigational assistant.

References

- [1] R.Baeza-Yates and B.Ribeiro-Neto, Modern Information Retrieval, Addison-Wesley, Reading, Mass., 1999.
- [2] S. Chakrabarti, Mining the Web, Morgan Kaufmann, San Francisco, 2003
- [3] R.Agrawal et al., Advances in Knowledge Discovery and data Mining, AAAIPress, Menlo Park, Calif., 1996.
- [4] W.Frakes and R.Baeza-Yates, Information retrieval-Data Structures and algorithms, Prentice-Hall, 1992