Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval

Masahiro MORITA
hiro@jaist.ac.jp

Yoichi SHINODA
shinoda@jaist.ac.jp

School of Information Science
Japan Advanced Institute of Science and Technology
15 Asahi-dai, Tatsunokuchi, Ishikawa 923-12 Japan

Abstract

Information filtering systems have potential power that may provide an efficient means of navigating through large and diverse data space. However, current information filtering technology heavily depends on a user's active participation for describing the user's interest to information items, forcing the user to accept extra load to overcome the already loaded situation. Furthermore, because the user's interests are often expressed in discrete format such as a set of keywords sometimes augmented with if-then rules, it is difficult to express ambiguous interests, which users often want to do. We propose a technique that uses user behavior monitoring to transparently capture the user's interest in information, and a technique to use this interest to filter incoming information in a very efficient way. The proposed techniques are verified to perform very well by having conducted a field experiment and a series of simulation.

1 Introduction

Recent developments in computers and computer networks interconnecting large numbers of systems have brought us convenience in many aspects, but have introduced a situation of "information overloading" also. As highly developed computer networks disseminate into our everyday life, the amount of information we create and exchange has grown by an order of magnitude. The amount is far more than a person can understand and manage, and the task of retrieving a piece of information from the magnitude of information is beginning to take considerable amount of time and effort.

A concept of *information filtering* [1] was introduced as a key technology to overcome this situation, but we are all aware that technologies such as user profile acquisition should be studied in detail before the information filtering technologies are put into any practical use.

In this paper, we propose a profile acquisition and user feedback technique to accumulate a user's preference for information, based on user behavior monitoring, as well as an information filtering technique using the acquired profile. The proposed techniques are based on an assumption that a user is performing a quick pattern matching task upon deciding if a piece of information is relevant to the user or not, before the user attempts to understand the semantics of the information any further. If this is all true, we can expect that information not of interest to the user is immediately rejected, and conversely, information of interest requires a noticeable amount of time to process.

In the following, we will briefly review the general architecture of information filtering systems and associated issues in section 2, followed by an explanation of our assumption in section 3. Section 4 describes the method and results of the field experiment we conducted with readers of NetNews, to prove the correctness of our assumption for the profile acquisition. Then, in section 5, the proposed information filtering technique using the acquired profile is described. A simulation of the proposed information filtering technique showed that the technique can retrieve relevant information for a user with sufficient precision and recall rates for casual source of information such as NetNews. Finally, in section 6, summaries of the proposals and results as well as remaining issues and directions for future research are presented.

2 Information Filtering Systems

In this section, we will give a brief explanation of information filtering systems and associated issues, based on our survey on information filtering technology [6].

In general, filtering information can be classified roughly into the following three categories[5], namely, *cognitive filtering* that does filtering by characteristics of information contents, *social filtering* that works on personal and organizational interrelationships of individuals in a community, and *economic filtering* that filters by cost-benefit assessments. While issues regarding social filtering are relatively well studied and economic filtering has the concrete measure of cost, cognitive filtering raises numerous issues which are directly or indirectly coupled with problems in natural language understanding technologies. We believe that cognitive filtering is very important, because it ac-

B. W. Croft et al. (eds.), *SIGIR '94*

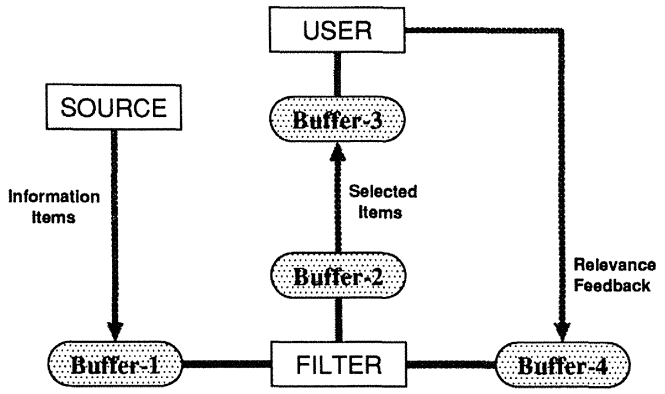

Figure 1. Generalized architecture of Information Filtering System

counts for humans' creative work by capturing potentially important information from casual information sources such as NetNews. From here on, we focus heavily on cognitive filtering systems.

Figure 1 shows a generalized architecture of information filtering systems [4]. This figure explains an information filtering system as consisting of seven logical units, namely the information source, the information filter, the user and four buffers. In this system, the source represents some descriptors of the information items to the filter, and the filter forwards to the individual user a subset of the items selected based on advance knowledge of the user's needs that is stored as the user's *profile*. The profile is like a reference database representing information needs or preferences of the user. Users may have the option of providing the filter with feedback indicating an extent to which the selection met their current information needs.

There are four buffers between the source, the filter, and the user as shown in the figure. These buffers are required in order to accommodate diverse information types or usage scenario types. For instance, buffer–4 stores the user's profile.

In our survey, we concluded that there are three important issues, listed below, that must be studied to realize any practical information filtering system.

- *Accumulation of user's preference to information*

  User's preferences must be accumulated in advance as a profile, and should be updated appropriately.

  Most of the previous work on cognitive filtering relied heavily on the user's active participation in profile accumulation and feedback. Profiles were often expressed in discrete forms such as a set of keywords possibly augmented with if-then rules, so that the user can enter and edit the profiles with an editor of some kind. However, we claim that it is generally difficult to express one's "interest", which is a non-discrete quantity, in the discrete way. Discrete profiles would result in incomplete description of the user's interest, which would lead to possible loss of potentially interesting information.

  Relying on the user's active participation in updating the profile database also have a problem in tracking temporal changes in the user's interest.

  We claim that a cognitive filtering system must be able to handle user's interest in a non-discrete fashion.

- *Representation of the profile*

  This issue is the direct product of the above issue. If a filtering system must deal with non-discrete information, the representation of the profile that stores the information will immediately be an issue.

- *Filtering mechanism*

  This issue, again, is the direct product of the first issue. Because a filtering system must deal with non-discrete profile and discrete target information, what will the best filtering technique be ?

We also point out that a practical cognitive filtering system should do its best not to filter out potentially important or stimulating information, to allow users of the system to have opportunity to be exposed to new information.

3 Information Filtering as Pattern Matching

The information filtering technique described in this paper is largely inspired by the advances in information retrieval using full-text search. In the WAIS system [3], search based on a concept of "similar to", that uses similarity among documents that uses number of words that appear commonly in the documents can be performed. This information retrieval technique based on the similarity of documents can be used as filtering mechanism in a information filtering system, if a set of documents that correctly reflect a user's preference to incoming information is given in advance.