A Framework Supporting Collaborative Filtering for Internet Information*

Sukumal Imudom B. Clifford Neuman
Information Science Institute
University of Southern California
4676 Admiralty Way Suite 1001
Marina del Rey, CA 90292
http://gost.isi.edu/info/ariel
imudom@isi.edu bcn@isi.edu
(310)822-1511 (voice) (310)823-6714 (Fax)

Introduction

Over the past few years recommender systems have emerged to assist users in finding pertinent information on the Internet. These systems have the following limitations.

- They are designed to handle different types of information and applications using different recommendation formats. The recommendation content comes in two different forms: enumerated and free text. Although the enumerated form is simple to understand, its inflexibility limits its usefulness. The free text form allows evaluators to contribute a wider range of annotation semantics; however, it makes automatic processing of information difficult.
- The issues of determining recommendation credibility and of finding incentives for evaluators to provide recommendations are problems that need to be worked out.
- These systems lack a general business model. So far, the systems are free for the consumers or are supported by advertisers only.

The Framework

To address these limitations of current recommender systems, we propose a framework supporting *collaborative filtering* of any type of information. Our framework represents an annotation as an object with various attributes instead of as an unstructured file. This representation provides the ability to handle both enumerated and free text annotation structurally. The framework consists of three building blocks: annotation semantic categories, determination of the expertise of the annotators, and a charging mechanism. It allows application developers to build on top of the framework using an API. We plan to furthur integrate this framework with electronic commerce.

This research was supported by the Defense Advanced Research Projects Agency under NASA Cooperative Agreement NCC-2-539 (DIstributed VIRtual System - DIVIRS), Contract J-FBI-95-155 (Supporting Organizationally Accepted Practice - SOAP), and Contract J-FBI-95-204 (Global Operating Systems Technologies GOST). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the National Aeronautics and Space Administration, the Defense Advanced Research Projects Agency, or the U.S. Government.

Annotation Semantics

Other systems, like the Tapestry system (Goldberg *et al.* 1992) allows mail readers to say either they "Like" or "Hate" a message while the GroupLens(Resnick *et al.* 1994) allows readers to provide a numeric score, from 1 to 5. The evaluators cannot express any further opinions with such annotations. Using free text comments, evaluators can more generally express their thoughts about an object. Our framework allows both enumerated and free text annotation.

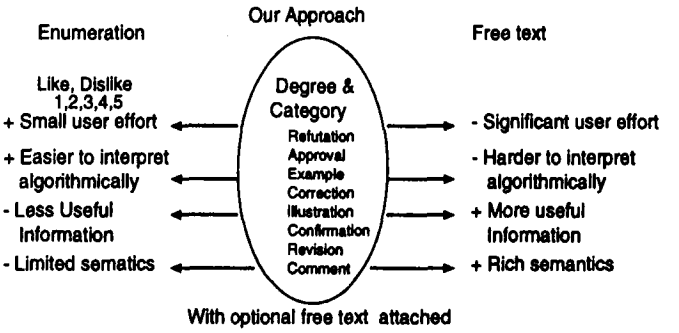

Figure 1: Annotation Contents

We define a set of annotation semantic categories to classify the free text annotations and provide structure. The annotation semantic set consists of Refutation, Approval, Example, Correction, Illustration, Confirmation, Revision, and Comment. Thus, the annotators can categorize their own annotations. Though it requires more work from the evaluators, it allows the automated processing of the annotations and reduces the processing complexity. A generic category is provided. Figure 1 depicts the spectrum of solutions with our approach in the middle.

Annotation Credibility

We introduce the notion of "expertise" to help identify experts in our system. People are experts in various fields. It is preferable that people who have knowledge of a subject be the ones who contribute evaluations. Annotations provided by experts are more valuable than those made by amateurs.We are investigating techniques to determine the expertise of the annotators. One of the techniques is *Cosine Similarity* (Salton & McGill 1983) comparing an evaluator's profile and an object's profile. The profiles of the evaluators are created from their resumes and their publications. The process of checking the validity of a person's documents may be done out-of-band during a registration phase. The expertise of evaluators can be used to help determine the credibility of the evaluations.

Charging Model

Up to this point, none of the annotation systems are equipped with a charging model which can enforce the evaluation integrity and prevent negative influences from advertisers. We investigate charging models as a possible solution to the cold-start problem. Charging provides an incentive to evaluators to provide evaluations. We are looking into a hybrid of the subscription and transaction models, which uses electronic credits as a reward for evaluators who make early and useful evaluations. The users in the system are required to pay an initial fee to receive the service. The user accounts are established and the credits are set. The consumers need to pay per each access of the evaluations. The evaluators earn credits for useful evaluations. The advantage of this model is that it does not predetermine who will be consumers or evaluators. The usefulness of the evaluations can be determined by the evaluator's level of expertise and the reader's votes. The proportion of the credits earned by each evaluator is determined by the level of expertise on the topic and is a function of the votes and the number of hits for the evaluation. Therefore, the annotators whose evaluations are more worthwhile will be paid more than those whose opinions are simply provided. A scenario of our credit transfer is shown below.

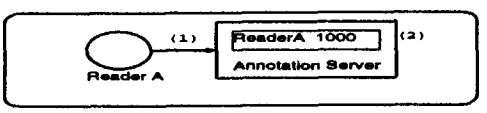

Figure 2: Account establishment

When subscribing to the service, reader A pays a subscription fee. The annotation Server establishes reader A's account and sets reader A's initial credits. See Figure 2.

Figure 3 shows the transaction steps taken by a reader retrieving an annotation. First, a reader requests an annotation from the annotation server. Next, the annotation server searches and retrieves the requested annotation. After sending the reader the requested annotation, the server decrements the reader's credits. For simplicity, we assume that the cost for accessing any annotations is 1 credit. However, a pricing policy can apply here to set the price of each annotation.

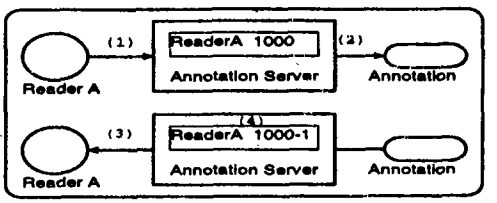

Figure 3: Reader transaction when retrieving an annotation

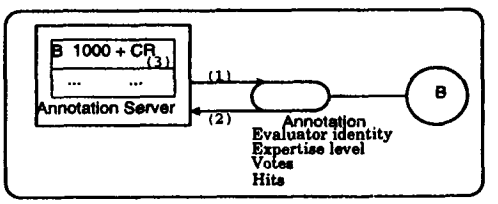

Figure 4: Transaction when paying an evaluator

The number of credits earned by an evaluator can be determined by the the evaluator's expertise and the reader's votes. Figure 4 shows the transaction steps taken when an annotation server pays an evaluator. From the annotation's attribute values, the number of credits paid to the evaluator B is calculated as a function of the number of votes and the number of hits.

References

Goldberg, D.; Nichols, D.; Oki, B.; and Terry, D. 1992. Using collaborative filtering to weave an information tapestry. *Communications of the ACM* 35(12):61–70.

Resnick, P.; Iacovou, N.; Suchak, M.; Bergstrom, P.; and Riedl, J. 1994. Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of ACM CSCW'94 Conference on Computer-Supported Cooperative Work*, Sharing Information and Creating Meaning, 175–186.

Salton, G., and McGill, M. J. 1983. *Introduction to Modern Information Retrieval.* McGraw-Hill Book Company.