

An Advanced Organizer for Asynchronous Discussion Forums

Andrew K. Lui Reggie Kwan

*School of Science and Technology
Open University of Hong Kong
{alui, rkwan}@ouhk.edu.hk*

S.C. Li

*Department of Education Studies
Hong Kong Baptist University
sandyli@hkbu.edu.hk*

Abstract

This paper presents an advanced organizer called SmartTag as a solution to ease the effort in identifying relevant messages in asynchronous discussion forums. SmartTag includes two tools for solving the problem: first, a message organizer for tagging relevant messages for future references, and second, a message filter that computationally evaluates the relevance of new messages for a user based on the personal preferences of this and other users. SmartTag employs techniques in statistical text analysis, recommender systems and collaborative filtering.

1. Introduction

Asynchronous discussion forum is both a facility for communications and an archive of learning resources. Achieving the second purpose requires an effective means to identify and retrieve relevant messages. The consideration of relevancy involves significant cognitive effort in reading and interpreting message content. A more attractive approach is an automatic and adaptive process that can search a large amount of messages for relevant content according to personal preferences.

This paper presents a discussion forum organizer called SmartTag, deployed as a component in discussion forum software. SmartTag provides a facility to tag any existing message that is considered relevant. SmartTag infers personal profiles from the tagged messages, and then applies the profiles to screen and evaluate the relevancy of new messages. A statistical text analysis technique, known as Latent Semantic Analysis (LSA) [5], is used in evaluating the relevancy message content. SmartTag also employs techniques from recommender systems [3] and collaborative filtering [4] to improve accuracy.

Before we describe the design and implementation of SmartTag, we will briefly discuss the background of the problem. At the end of the paper we will suggest directions for future work.

2. Background

The nature of asynchronous discussion facilitates the creation of messages of higher learning and reusable value. The characteristics of predominately text interface, self-paced participation, and flexible access allows more in-depth analysis and reflections [1]. Instructors in particular have more incentives to put effort in composition because their messages can reach and influence every reader [2].

Although useful learning resources exist in asynchronous discussion forums, identification of the relevant resources is difficult. The key problem to overcome is the personal cognitive effort required to comprehend the content and assess the relevancy of ever-increasing amount of messages. One reason is simply that more online learners are acclimatized to asynchronous discussion and more messages are posted. The feasibility of catching up with all the messages is hampered by infrequent and irregular participation, which is typical of online learners. The problem is further compounded by the requirement to track the progress of several coexisting discussion threads of different themes.

Online learners need assistance to organize and to retrieve relevant messages for their learning tasks. Most discussion forums support the capability to sort messages by date, subject, author, read status, and other attributes, and the meta-information allows an indirect evaluation of message relevancy. A more direct content-based approach is supported by keyword search facility that identifies relevant messages based on text content. Better accuracy can be achieved with statistical text analysis techniques such as LSA. The notion of relevancy however involves personal judgment and preference. A better approach should include a feature of capturing the personal preference about some characteristics of relevant messages.

Recommender systems are computer applications that provide suggestions based on personal preferences [3]. These systems evaluate suggestions according to

personal profiles, which in our problem consist of characteristics of relevant messages. Personal profiles may be constructed explicitly or implicitly. The explicit method is often considered less attractive because user effort is required in indicating the personal preferences. On the other hand the implicit method, which tracks and infers personal preferences from user actions, is unlikely to match the accuracy.

Collaborative filtering is a technique that includes the personal preferences of others in providing suggestions [4]. The technique assumes that there exists more than one person sharing a similar preference. If both users A and B prefer w, x, and y, and if user A prefers z, then user B is likely to also prefer z.

3. An Asynchronous Discussion Organizer

SmartTag, the proposed asynchronous discussion organizer, makes the following assumptions about the relevancy of discussion messages. First, relevancy and similarity of two messages can be inferred from text statistics. Second, an existing message considered relevant can indicate the characteristics of relevant new messages. Third, there is more than one online learner sharing a similar personal preference over the relevancy of messages.

The principal components of SmartTag are described in the following minor sections.

3.1. Message Organizer

The Message Organizer allows users to define categories and then manually tag an existing message with a category (see Figure 1). Two purposes are served. First, the user can use the tag to identify a relevant message. Second, the tagged messages indicate the content that the user considers as relevant. This approach constructs personal profiles implicitly while reaching the accuracy of the explicit method.

3.2. New Message Filter (Individual)

Figure 2 shows the operation of the New Message Filter. The filter screens each new message and evaluates its relevancy. By referring to the text statistic characteristics of the messages of each category, the filter determines whether a new message is relevant, and which category it is relevant.

3.3. New Message Filter (Collaborative)

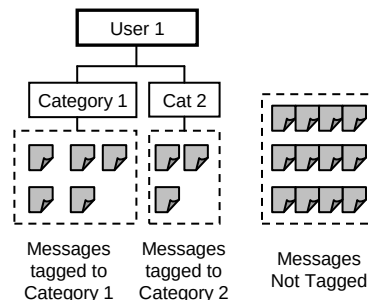


Figure 1. (Basic Message Organizer) Users can individually tag messages according to their own way of categorization.

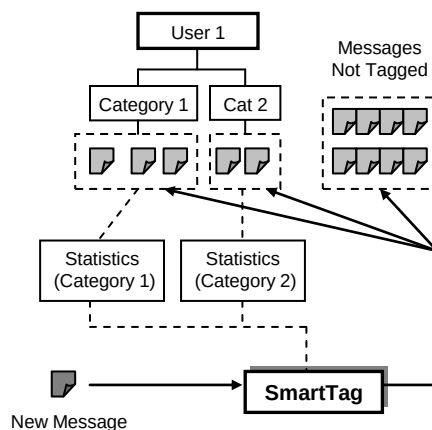


Figure 2. (New Message Filter - Individual) SmartTag finds the category that most resembles the statistical features of the new message and tags it accordingly (or leaves it untagged).

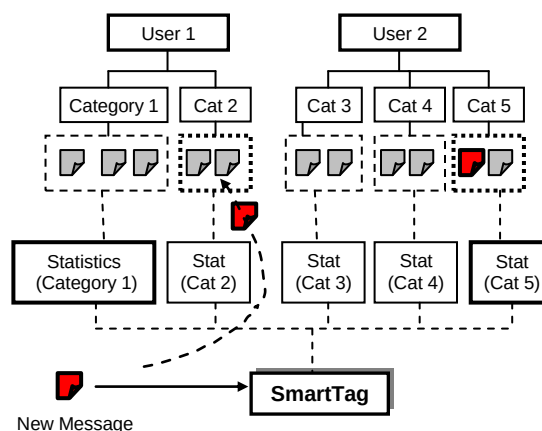


Figure 3. (New Message Filter - Collaborative) SmartTag finds the category that most resembles the statistical features of the new message and tags it accordingly (or leaves it untagged).

Figure 3 shows the operation of the New Message Filter when taking into consideration of the message categorization of other users. To evaluate the relevancy of a new message *M* to a user *A*, the New Message Filter searches for a category defined by other users that satisfies two conditions: the category contains *M* and it is similar to a category of user *A*. In this case, the New Message Filter considers *M* as relevant to the category of user *A*. Figure 3 illustrates an example, in which SmartTag is processing a new message for User 1. It discovers that User 2 has already tagged the message as *Category 5*. SmartTag then scans the categories of User 1 and finds out that *Category 2* contains similar statistics to *Category 5*. It then recommends the new message is relevant to *Category 2* to User 1.

3.4. Implementation

We implemented SmartTag as a part of a web-based learning platform. Asynchronous discussion forum is one of the components, and online learners can choose between a traditional forum interface and the SmartTag interface. A user of SmartTag is first required to create tag categories. Then the user can browse messages and attach a tag to any message considered relevant. The tagged messages can be identified easily for reference later.

The New Message Filter is implemented as an autonomous agent running in the background. For each active user, the agent scans new messages and evaluates their individual relevance. The agent tags relevant new messages with tags in yellow to indicate that the messages are agent-tagged. The user can replace them with their user-defined tags (the white tags) as a confirmation.

A key component in SmartTag is a comparator of the similarity of text messages, which is required in the following two tasks: finding the similarity between a new message and a group of messages in a category; finding the similarity between two groups of messages from two categories. The similarity comparator uses Latent Semantic Analysis (LSA) to evaluate text similarity [5]. Statistical text analysis techniques, including LSA, exploit the statistical characteristics like the word count in text documents. If two texts are similar in the statistics, then the two text documents are likely to be similar semantically. The core of LSA is a matrix for evaluating the similarity between training text documents and new text documents. Figure 4 below illustrates this process of building the comparator.

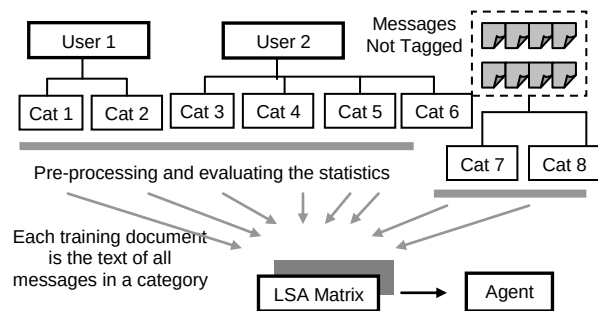


Figure 4. The construction of the content similarity comparator based on LSA.

4. Discussion and Conclusions

This paper describes a solution for the problem of identifying relevant messages in asynchronous discussions. The success of this solution lies in the active usage of the Message Organizer. In the future, we will implement several measures to encourage its use. Enforcing reciprocity is one way that allows users to use the New Message Filter only if they have used the Message Organizer to organize their messages. On the other hand, new users could need assistance to create new SmartTag categories. We will implement a new component that suggests SmartTag categories based on how other users create their categories.

5. References

- [1] Hülsmann, T., (2003). Texts That Talk Back - Asynchronous Conferencing: A Possible Form of Academic Discourse? In U. Bernath & E. Rubin (eds.), *Reflections on Teaching and Learning in an Online Master Program: A Case Study*, 75-120.
- [2] Taylor, J.C. (2001). Automating e-Learning: The Future of Higher Education. Keynote address delivered at the International Conference on Emerging Telecommunications Applications (ICETA), Kosice, Slovakia, October, 2001.
- [3] Resnick, P., Varian, H.R., (1997). Recommender systems. *Communications of the ACM*. 40(3), 56-58.
- [4] Paulson, P., and Tzanavari, A., (2003). Combining collaborative and content based filtering using conceptual graphs, in *Modeling with Words* (J. Lawry et. al. eds.), Springer-Verlag.
- [5] Landauer, T.K., et. al., (1998). An Introduction to Latent Semantic Analysis, *Discourse Processes*, 25 (2-3), 337-354.

Acknowledgement

The authors are grateful to Dr. Rex Sharman for his useful comments.