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-
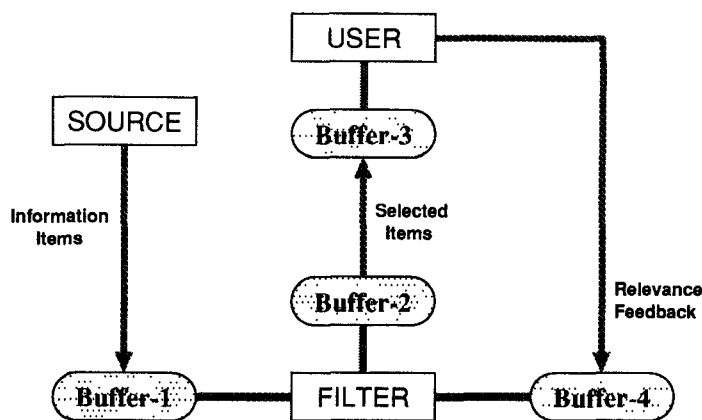
Figure 1. Generalized architecture of Information Filtering System

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.

Foltz took this approach of using similarity among documents for information filtering. He reported in [2] that using latent semantic indexing (LSI) for filtering news articles using articles pre-selected by a readers as the reader's profile can perform significantly better than filtering based on simple keyword matching on filtering performance.

Although their results are interesting, we felt that we need to study further on in following two points before directly employing the mechanism to judge similarity among documents in an information filtering system.

3.1 Semantical v.s. Syntactical Filtering

First point is the relevance of LSI upon "first impression" decision making of whether a document is interesting or not. Foltz claims that user preference will be similar for articles that are semantically similar. However, according to our experience, we all feel that when we read Japanese texts, we are not performing a deep understanding of each article upon making the decision of whether the article is important to us or not, but doing a pattern matching task looking for major keywords or combination of the keywords, and judging article's overall syntactical structure. Success in machine translation of Japanese into English text that uses only the syntactical information [11] partially justify this assumption.

It should be argued that people with languages using ideograms as mother tongue behave differently from those with languages using phonograms as mother tongue. It is natural that our assumption applies to those with languages using ideogram as a mother tongue, but we also know that there is a reading technique for phonogramic language text that picks up words diagonally from a page without considering the page's semantical structure, and still captures the semantical structure of the page. This means that certain variation of keyword search may work for phonogramic language texts also.

It is not clear that Foltz's result using LSI was compared to simple keyword count or using similarity between documents approach seen in the WAIS system. If information filtering using the syntactical similarity between documents performs as well as that based on LSI approach, it might be considered as our first choice for filtering mechanism because of its low processing overhead compared to the LSI method. It is needless to say that LSI approach deals more deeply with semantical content of documents, and when compared to "syntactical similarity" approach that only deals with one dimensional measure that is the number of keywords appeared commonly in two documents, it should perform better. But if two approach behaves almost the same, we claim that simple syntactic similarity is better for efficiency reasons. This is especially true for language texts that have no clear word break in a sentence, since breaking sentences into words require numerous machine cycles and we just can't afford them.

3.2 Profile Acquisition and User Feedback

Another is the user feedback issue, as discussed in the previous section. As Foltz pointed out in [2], in the conclusion of the paper, that we need some mechanism that reflects temporal transitions in interests of readers into the preference database. He also pointed out that his method of having users rate each article requires a lot of effort on the part of the user. That is, we need some mechanism to acquire and feedback user's preference to the profile.

Again, from our experience of reading news articles in Japanese text, as we reject uninteresting articles using our first impression of the articles, we can easily assume that articles which took considerable amount of time to read can be treated as potentially interesting articles. If we can determine whether a reader is interested in an article or not by measuring the time to read it, we might be able to capture the readers profile automatically. Foltz, although very vaguely, also pointed this out.

4 Profile Acquisition and Feedback by User Behavior Monitoring

Based on the second observation in the last section, we conducted an experiment that is similar to Foltz's experiment, to investigate the dominating factors that affects time to read an article. In this experiment, in addition to having users rate each article, we setup a system so that we could measure the time spent for each article.

We setup following hypothesis at beginning of the experiment:

- Users spend longer time to read interesting information than not-interesting ones.

In addition to the hypothesis, we decided to see if there are other factors which may affect the time spent for an article, such as the followings.

- Time spent for an article increases in proportion to size or length of the article.

- Time spent for an article is affected by the readability of the article, i.e., denseness of the article, in some degree.

- Time spent for an article is affected by a size of a back log that a user have to process, e.g. if the user have large quantity of unread articles, time spend for each article get shorter.

In the following, we will first describe how the experiment was carried out, followed by result of analyzing the collected data.
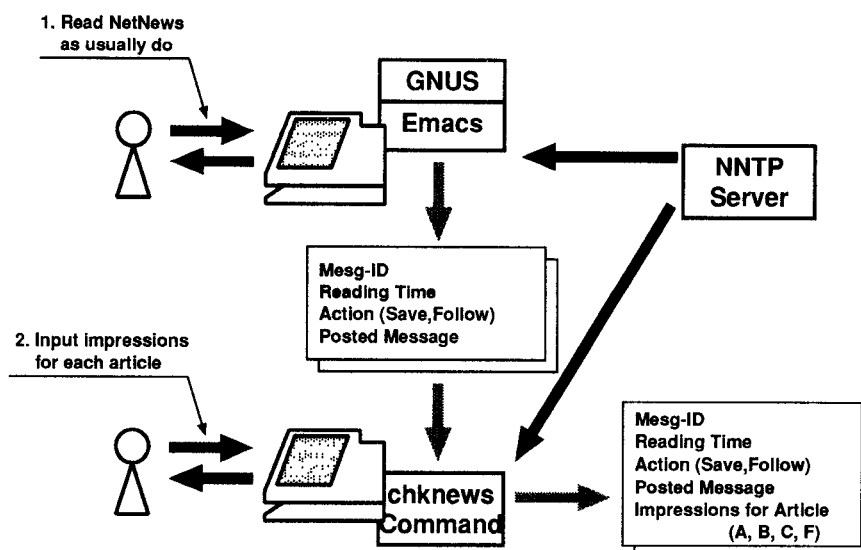
Figure 2. The Experimentation System

4.1 Experimentation

4.1.1 Subjects of the Experiment

This experimentation is aimed at NetNews (we focused on fj.* newsgroups, which is written mostly in Japanese) user, which is a popular Internet information service. Subjects of our experiment were recruited from actual user of NetNews by the NetNews itself, and 8 people volunteered as subjects. These 8 subjects read total of 8,000 articles during a period of 6 weeks.

4.1.2 Experimentation System

Figure 2 illustrates over all setting for the experiment. We modified GNUS, a widely used NetNews reader based on GNU Emacs, to record following information for each article that were read by subjects of this experimentation.

1. Message identifier of the article.

2. Time spend for the article (How much time was spent for the article).

3. Action for the article (The article was saved or follow-up was posted).

4. Copy of posted messages if the subject post new articles.

After a GNUS session is over, subjects rate each articles they read in the last session using **chknews** command. The **chknews** command uses the record of message identifiers to display contents of each article on users terminal and interviews subjects about impression of the article. Then it merges the records described above and the impressions about articles, and send them into a data collecting server using ordinal e-mail messages.

The impressions of article was to be answered in four grade of 'A', 'B', 'C' or 'F' where 'A' meaning very interesting, 'B' meaning interesting, 'C' meaning neither interesting nor not-interesting, and 'F' meaning not-interesting.

We requested subjects to behave as close to usual news reading session as possible during the experimental GNUS session, but also requested the followings to avoid abnormalities in the resulting data.

- During the GNUS session, subjects must wholeheartedly read articles and should not do other things such as leaving the terminal a while to get a cup of coffee or read newly arrived e-mail messages.

- During the period of experiment, subjects should read NetNews everyday as possible.

- Subjects should read all articles in registered newsgroups, e.g. subjects should not skip articles without having a glance at them or issue "catch up" commands.

Note that we separated the reading session and the rating session for the same reason.

Table 1 summarizes collected data over the period of 6 weeks. In following sections, we use the terms 'useful' or 'interesting' for those article that subject rated 'A' or 'B', and 'un-useful' or 'not-interesting' for those rated 'F'.

4.2 Analyzing Behavior of NetNews Users

From the statistical analysis on the collected data, we concluded that a major factor that influence a time spent for an article is the preference of a user to the article. Figure 3 illustrates this observation. In this figure, number of articles and time spent to read them are plotted for three groups of articles, namely all articles, those rated interesting and

Table 1. Collected Data

Subject #	# of Session	# of Article for each impressions				
		A	B	C	F	Total
1	37	91	156	209	835	1291
2	68	30	451	969	90	1540
3	40	21	528	1218	27	1794
4	11	13	28	58	201	300
5	25	3	387	208	3	601
6	38	1	19	746	7	773
7	41	35	352	233	14	634
8	30	26	162	697	14	899
Total	290	220	2083	4338	1191	7832
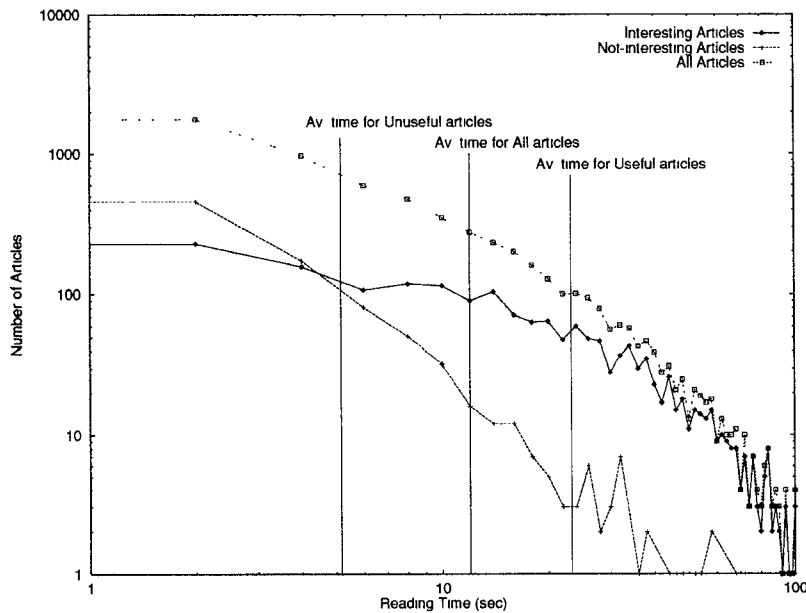
Figure 3. Distribution of Articles against Reading Time

those rated not-interesting. Other factors played only a very little role in affecting reading time of an article. The results from our analysis can be summarized as follows.

- There is a strong tendency to spend a long time to read articles those rated interesting.

- There is a tendency not to spend a long time to read articles those rated not-interesting.

- There was very low correlation of less than 0.08 between length of an article and a time to read the article. This probably indicates that not all articles are completely read, e.g. the decision of whether an article is interesting or not, especially in the not-interesting articles, are made by looking at first couple of dozens of lines.

- Readability of an article is not an important factor that affects its reading time. In our analysis, we used several metrics for denseness of characters in article, including average number of characters per line, ratio of blank lines in an article. We suspected that reading time become longer in proportion as denseness of the article become higher, but the results showed very low correlation between reading time and any of the metrics we used. The highest correlation we observed was 0.25, which was for average number of characters per line.

- Effect of back log (unread articles) was also very small. Correlation between number of unread articles and reading time is less than 0.18.

From the observation made above, we can conclude that preference of a user to an article is the dominating factor that affects time to spend reading it, meaning that it is possible, to a certain extent, to monitor the time spend for an article and know if the user is interested in the article or not. Consider for example, with some threshold time τ seconds, and treating articles which took longer than τ seconds to read as interested articles, and use these article as feedbacks, we can perform user feedback with significant precision. For example, our result indicates that with τ being 20 seconds, 30% of interesting articles can be retrieved with precision of 70%.

Table 2. % of Mutual Relation of Impression between Initial and Follow up Articles

Impression for Initial Articles	Impressions for Follow up Articles			
	Interesting	Not-interesting	Others	Total
Interesting (1447 articles)	489 (52.3%)	47 (5.0%)	400 (42.7%)	936 (100.0%)
Not-interesting (493 articles)	42 (10.5%)	271 (68.3%)	84 (21.2%)	397 (100.0%)

Table 3. % of Mutual Relation of Reading Time between Initial and Follow up Articles

Reading Time for Initial Articles	Reading Time for Follow up Articles			
	less than 10 sec	more than 20 sec	10 to 20 sec	Total
less than 10 sec (2151 articles)	1333 (77.4%)	138 (8.0%)	251 (14.6%)	1722 (100.0%)
more than 20 sec (1019 articles)	200 (35.1%)	219 (38.5%)	150 (26.4%)	569 (100.0%)

4.3 Effect of Relationship Between Articles on User Impressions

NetNews articles usually have a initial-follow relationship. During the analysis of our experiment, we examined mutual relation between impressions for initial article and it's follow up, and have found following interesting facts as shown in table 2.

This table can be interpreted as follows.

- Impressions for an initial article is inherited to follow up articles with high probability. That is, follow up articles of a not-interesting article are also not-interesting, and conversely, follow up articles of an interesting article are also interesting.

- It is a rare case that an impression for an initial article is reversed in follow up articles.

These observation suggest that if an initial article is considered to be un-useful or not-interesting, a news reader program can drop its follow up articles with very low probability of dropping potentially interesting articles.

We also examined a similar relationship between initial and follow up articles using time spent to read them, to see if they have any new indications. Table 3 shows mutual relationships between initial-follow relationship and users' behavior (reading time). From this table, we can conclude the followings.

- If a user spent short time (say, less than 10 seconds) reading an initial article, time spent for its follow up articles are also short (less than 10 seconds) with very high probability of almost 76%. This is easily expected from the above observation concerning the inheritance of impressions.

- Even if a user spent long time (say, more than 20 seconds) to read an initial article, time spent for its follow up articles vary (second row in table 3). That is, time spent for follow up articles to an initial articles having long read time are not necessarily long.

Let us examine the second point more carefully. Our result indicates that 26.5% within the 35.1% of the articles with short read time (less than 10 seconds) that are follow up to an initial article with long read time (more than 20 seconds) are rated interesting. This figure is statistically meaningful compared to the ratio of interesting articles within all articles with reading time less than 10 seconds, which is 14%. That is, follow up articles to an interesting initial article are potentially interesting even if its own reading time is short.

5 Information Filtering using Sub-String Indexing

In this section, we describe a proposed information filtering method, using the user behavior monitoring described in the previous section as user profile acquisition and feedback, and its simulation results using the collected data.

In this filtering method, we adopt following hypotheses:

- We are making a first approximation of whether an information item is interesting or not by the appearance or pattern of characters or keywords in the information item.

- There are some common or similar features in the pattern of characters or keywords within a collection of information items rated interesting, and in a collection of information items rated not-interesting also.

In traditional approach to filter information based on user model [8, 9, 12], it has been considered that deep understanding of semantical contents of the information items is a necessity. In these approach, it is proposed that information filtering system is told of users preference in a form of "user model". Then, upon arrival of an incoming information, the information is semantically analyzed and checked against the user model if the item fit the user's needs.

Our approach is quite different from the above. As already given as one of our hypotheses, it seems that we do not understand the details of information when we are getting information from information service mainly for casual use like NetNews, etc., and we are mostly judging from impressions of pattern of characters or keywords to determine if the given piece of information item is interesting or not. So, we adopted the sub-string indexing method that uses only a simple syntactical information as primitive method for evaluating pattern of characters.

5.1 Sub-String Indexing Method

In our information filtering method, we use sub-string indexing method to evaluate the pattern of characters. Here, sub-string (or n-gram) refers to a product of splitting a given string into a fixed length sub-parts allowing fixed length overlapping. Figure 5 gives an example of decomposing a Japanese string into sub-string of length 2 with overlapping length of 1.
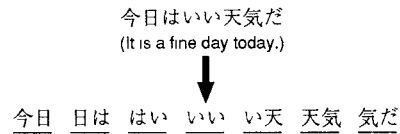
今日はいい天気だ
(It is a fine day today.)

今日　日は　はい　いい　い天　天気　気だ

Figure 4. Decomposition of a Japanese String into Sub-strings

Sato showed in [10] that best match information retrieval method using this sub-string indexing is very powerful and efficient. The basic idea of his method is to put Japanese texts in database with its sub-strings as keys, and a search is carried out using a query key, which also is decomposed into sub-strings, by searching a text that has highest number of matching sub-string keys. Figure 5 gives an example of score calculation.
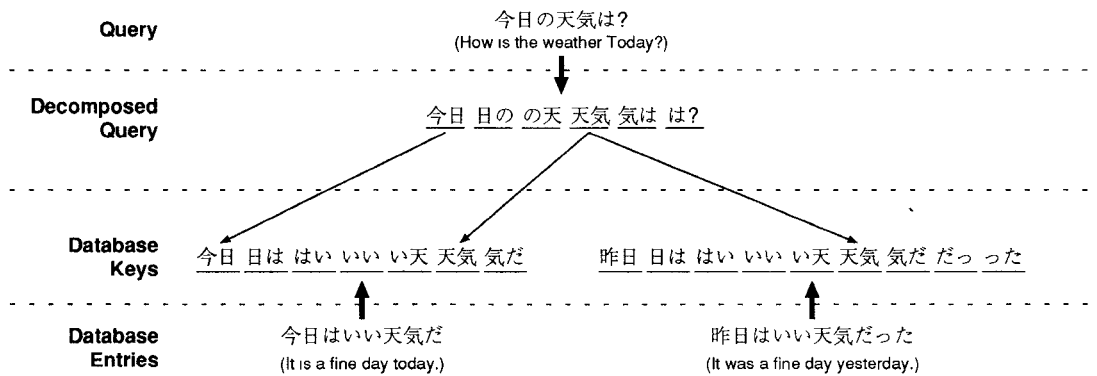
Query

今日の天気は?
(How is the weather Today?)

Decomposed Query

今日　日の　の天　天気　気は　は?

Database Keys

今日　日は　はい　いい　い天　天気　気だ　　　昨日　日は　はい　いい　い天　天気　気だ　だっ　った

Database Entries

今日はいい天気だ
(It is a fine day today.)

昨日はいい天気だった
(It was a fine day yesterday.)

Figure 5. Best Match Text Retrieval using Sub-string Indexing

Our filtering method is a variation of the best match retrieval method. When a new article is given, we calculate its score, or occurrence count of all sub-string of the article in a sub-string database containing sub-strings made by splitting articles determined to be interesting in the past. We then use some threshold score to determine if the new article is interesting or not. Similarly, to determine if a new article is not-interesting, we can carry out the same scoring procedure using those articles that are determined not-interesting in the past.
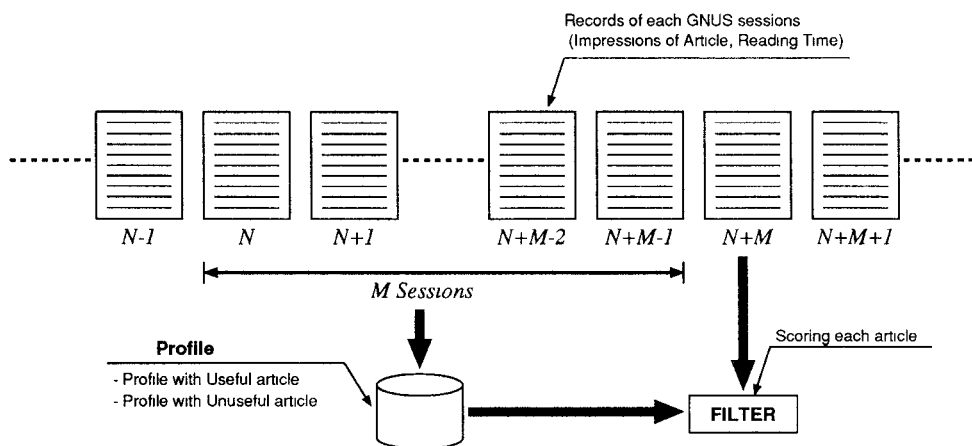
Records of each GNUS sessions
(Impressions of Article, Reading Time)

$N-1$　　N　　$N+1$　　$N+M-2$　　$N+M-1$　　$N+M$　　$N+M+1$

M Sessions

Profile

- Profile with Useful article
- Profile with Unuseful article

Scoring each article

FILTER

Figure 6. Evaluating Method

5.2 Evaluation

We evaluated the proposed information filtering method using the sub-string index method by a simulation using the collected data.

Figure 6 illustrates show how the simulation was carried out. We took records of M consecutive sessions starting at session N from a subject's records of GNUS interaction and rating work, and made a profile database of articles rated interesting and not-interesting by the subjects. Then, articles in $N + M$ th session was scored using the profile database. To increase the samples, we shifted the starting session N within the entire recorded session.

Table 4. Number of Articles in Evaluation Data Set

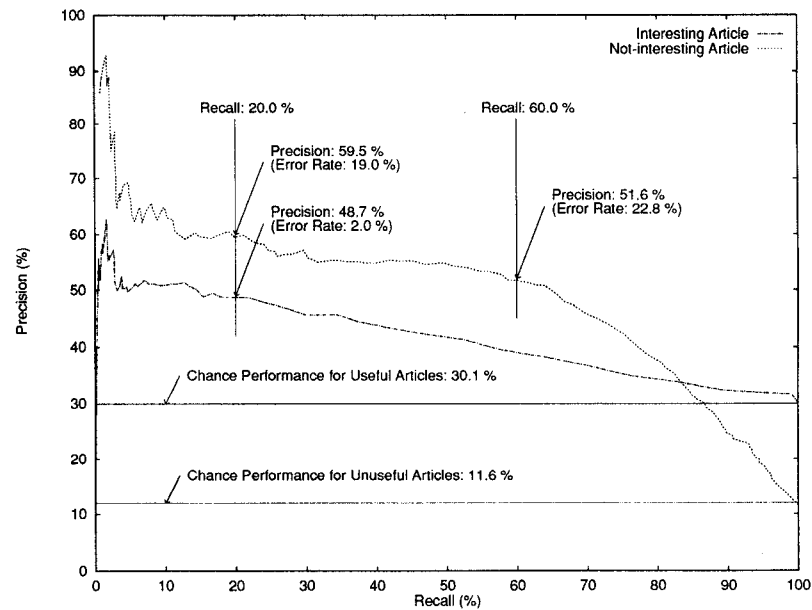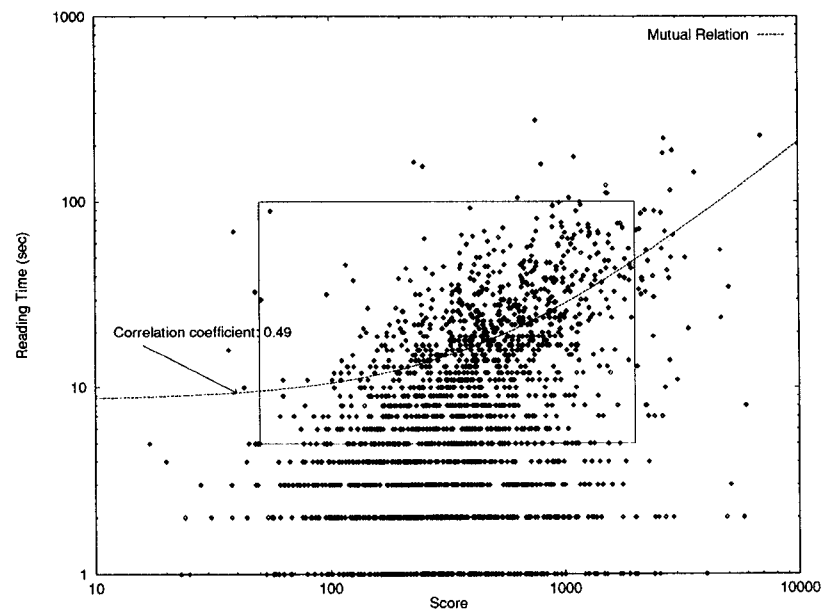|	# of Articles	Chance Performance
Rated Interesting	1956	30.1 %
Rated Not-interesting	756	11.6 %
Total	6511	

Figure 7. Precision curves of Filtering Information

Figure 8. Correlation between Estimated Usefulness
Scores and Reading Time

Table 4 shows dimension of data sets used in this evaluation. We varied the value of M in the range of 1, 2, 5, 10 and 20, and obtained best results with $M = 5$. Following discussions are based on this value of M.

Figure 7 shows precision-recall curves of predicting interesting and not-interesting articles, obtained by varying the threshold scores that determines whether the articles are interesting or not. It illustrates that the filtering for both interesting and not-interesting articles performed considerably better than their chance performance. For example, at the recall rate of 20%, precisions are 48.7% for interesting articles and 59.5% for not-interesting articles. Filtering of not-interesting articles perform well even at the high recall rate of 60% with precision of 51.6% against chance performance of 11.6%.

To further justify our hypothesis given at the beginning of this section, we examined the correlation between estimated usefulness scores and reading time for all articles. Figure 8 shows the correlation between these two values. This figure shows there is a strong correlation between them (correlation coefficient is 0.49), suggesting that the score calculated by our method can predict user's impression with very high precision.

5.3 User Feedback Methods and its Evaluation

We also carried out a similar simulation, using articles that are deemed interesting by monitoring user's behavior (reading time > 20 seconds) in the profile database, instead of articles by user's actual rating, and did the same prediction procedure using the score by sub-string indexing. With this simulation, although surprising but somewhat expected, the filtering method performed better than using a user's actual rating as the profile, in both precision and recall performance. We are interpreting this phenomenon as a result of user's mis-rating of articles. We consider this result as both interesting and encouraging, because the result suggests that articles having long read time has certain good characteristics as items in profile database.

However, we also found out that precision and recall performance degrade as we increase the degree of the feedback (increase M in figure 6). We call this phenomenon as *sub-string saturation*. When sub-string saturation occur, that is, when the profile database is saturated with too many sub-strings, any piece of information can obtain a high score. We are aware of the problem, and are still investigating the way to avoid the phenomenon.

6 Summary and Concluding Remarks

We conducted experiments to verify our assumption that (a) users' preference to NetNews articles are well reflected into the time spent to read these articles, and (b) users are carrying out simple syntactical pattern matching task upon deciding whether an article is interesting or not. Results of our experiment is very encouraging. Results presented in this report collectively suggest that our method does provide solutions for major issues of information filtering systems, namely profile acquisition/feedback without users' extra effort and filtering technique itself. Although we do not claim that the technique described here behave "semantically correct", e.g. it is not capable of picking up information that users actually do need, results presented in this paper indicates that our mechanism does perform as good as or little more than human users do in practice.

There are, however, several issues still waiting to be studied.

- *Improvement in filtering precision*

  As discussed in the last section, we observed the phenomena of sub-string saturation when we tried to increase the precision of filtering by increasing the degree of feedback. We must study further on to increase the precision without having sub-string saturation. We are looking at removing sub-strings that are obviously meaningless from the profile database.

- *Determination of uninteresting information by user behavior monitoring*

  Although our method of user behavior monitoring is highly capable of automatically determining interesting information, it does not perform as well for determining uninteresting information. As we observed that the disinterest of an initial article is inherited to its follow up articles stronger than interestingness of an initial article is inherited to its follow up articles, it is important to establish a way to determine if a certain information is uninteresting.

  In our experiment, due to the limitation in the GNUS news reader program, we were not capable of knowing if an article was completely read by a subject, or if time spent for each displayed page got shorter as the subject progress through the article. Since issuing a "next article" command or typing successive "next page" command are routinely used in our everyday NetNews life to skip uninteresting articles, monitoring these behaviors might bring a new result.

- *Effectiveness of our method for phonogramic language text*

  As discussed in section 3, it is interesting to examine if our method does perform as well for phonogramic language text. As also pointed out in section 3, we have a reason to expect our method to work as well. It is equally interesting to compare our techniques with filtering techniques using semantic processing such as LSI.

As needs for research on issues do remain, our approach seems to be quite promising as user friendly and cost effective way of information filtering, at least for ideogramic language text, especially those that are costly to carry out even a lowest level of semantical or grammatical analysis.

Acknowledgements

Our appreciation also goes to Professor Naohito Chino of Aichi Gakuin University, department of psychology, for his advice in statistical analysis of the collected data.

References

1. Nicholas J. Belkin and W. Bruce Croft. Information Filtering and Information Retrieval: Two Sides of the Same Coin? *Communications of the ACM*, 35(12):29–38, December 1992.

2. Peter W. Foltz. Using Latent Semantic Indexing for Information Filtering. In *Proceedings of the ACM Conference on Office Information Systems*, pages 40–47, April 1990. Boston.

3. Brewster Kahle. Wide Area Information Server Concepts. Thinking Machines (FTP:/wais/doc/wais-concepts.txt@quake.think.com), 1989.

4. Shoshana Loeb. Architecting Personalized Delivery of Multimedia Information. *Communications of the ACM*, 35(12):39–48, December 1992.

5. Thomas W. Malone, Kenneth R. Grant, Franklyn A. Turbak, Stephen A. Brobst, and Michael D. Cohen. Intelligent Information-Shareing System. *Communications of the ACM*, 30(5):390–402, May 1987.

6. Masahiro Morita. A Survey of Information Filtering Technology (In Japanese). JAIST Research Report IS–RR–93–9I, School of Information Science, Japan Advanced Institute of Science and Technology, 1993. FTP: /pub/jaist/doc/93/rr/is-rr-93-9i.ps@ftp.jaist.ac.jp.

7. Masahiro Morita. Information Filtering Based on User Behavior Analysis (In Japanese). Master's thesis, School of Information Science, Japan Advanced Institute of Science and Technology, February 1994.

8. Ashwin Ram. A Knowledge goals: A Theory of Interestingness. In *Proceedings of the 12th annual conference of the cognitive science society*, pages 206–214, July 1990. Cambridge, Mass.

9. Ashwin Ram. Natural Language Understanding for Inforamtion Filtering Systems. *Communications of the ACM*, 35(12):80–81, December 1992.

10. Satoshi Sato. CTM2: A Japanese-English Example-Based Translation Aid System using N-gram Index Method (In Japanese). JAIST Research Report IS–RR–93–6I, School of Information Science, Japan Advanced Institute of Science and Technology, 1993. FTP: /pub/jaist/doc/93/rr/is-rr-93-6i.ps@ftp.jaist.ac.jp.

11. Satoshi Sato and Takeshi Kawase. A High-Speed Best Match Retrieval Method for Japanese Text. *Submitted to COLING'94*, 1994.

12. Irene Stadnyk and Robert Kass. Modeling Users' Interests in Information Filters. *Communications of the ACM*, 35(12):49–50, December 1992.