A Study on Recommendation System Using Data Mining Techniques for Large-sized Music Contents

대용량 음악콘텐츠 환경에서의 데이터마이닝 기법을 활용한 추천시스템에 관한 연구

Yong Kim*, Sung-Been Moon**

ABSTRACT

This research attempts to give a personalized recommendation framework in large-sized music contents environment. Despite of existing studies and commercial contents for recommendation systems, large online shopping malls are still looking for a recommendation system that can serve personalized recommendation and handle large data in real-time. This research utilizes data mining technologies and new pattern matching algorithm. A clustering technique is used to get dynamic user segmentations using user preference to contents categories. Then a sequential pattern mining technique is used to extract contents access patterns in the user segmentations. And the recommendation is given by our recommendation algorithm using user contents preference history and contents access patterns of the segment. In the framework, preprocessing and data transformation and transition are implemented on DBMS. The proposed system is implemented to show that the framework is feasible. In the experiment using real-world large data, personalized recommendation is given in almost real-time and shows acceptable correctness.

초 록

본 연구는 대용량 음악콘텐츠환경에서 개인화 추천 서비스를 위한 기반구조의 제공을 위하여 시도되었다. 추천서비스를 위한 기존의 많은 연구와 상용프로그램에도 불구하고 대규모의 쇼핑몰들은 개인화 추천서비스와 실시간으로 대용량의 데이터를 처리할 수 있는 추천시스템을 필요로 하고 있다. 이를 위하여 본 연구에서는 데이터마이닝 기술과 새로은 패턴매칭 알고리즘을 제안하고 있다. 콘텐츠 주제분야에 대한 이용자의 선호도를 이용한 이용자 분할을 위하여 군집화 기법이 사용되었다. 다음으로는 군집화를 통하여 생성된 분할된 이용자 그룹에서 개별 이용자의 콘텐츠에 대한 접근 패턴의 추출을 위하여 순차패턴 마이닝기법을 적용하였다. 최종적으로 각각의 이용자 군집의 콘텐츠 접근 패턴과 콘텐츠 선호도에 기반한 제안된 추천 알고리즘에 의해 추천이 이루어진다. 이러한 추천을 위하여 기반구조와 함께, 전처리과정과 원본 데이터의 형식변환이 데이터베이스에서 수행되어진다. 본 연구에서 제안하고 있는 기반구조의 적절성을 보여주기 위하여 제안된 시스템을 구현하였다. 실제 이용자에 의해 이용된 데이터를 실험에 적용하였으며, 해당 실험에서 추천은 실시간으로 이루어졌으며 추천결과에 있어서는 적절한 정확성을 보여주고 있다.

Keywords : personalization, recommendation system, sequential patterns, clustering, data mining
개인화, 추천시스템, 순차패턴, 군집화. 데이터마이닝

* Senior Researcher, KT Infra Research Lab. (yongkim@kt.co.kr)
** Professor, Dept. of Library and information Science, Yonsei Univ. (sbmoon@yonsei.ac.kr)

■ Received : 25 May 2007
■ Accepted : 19 June 2007

1. Introduction

With the rapid growth of authoring tools and information technologies (IT), the amount of information published with an electronic format and the number of users who access information on cyber space are growing at a tremendous rate. To satisfy users' information needs and provide efficient information services in libraries amidst information overflow, information providers like libraries are building electronic libraries and archiving centers. The potentially infinite and growing number of available online information sources makes it increasingly difficult for information seekers to quickly find relevant information. New services are urgently needed on the Internet to prevent users from being drowned by the flood of available information. Libraries and information centers have tried various information services to solve these problems. As an exemplary attempt, search engines such as Alta Vista, Google and Yahoo were developed, but they did not radically satisfy what users want because users should spent their time to search and decide what they want from all the retrieved information. As an important and alternative method, Selective Dissemination of Information (SDI) service traditionally provided in libraries was considered. Also, Customized Information Service (CIS) as a more advanced and active service based on user profiling information is beginning to emerge. CIS based on push technology is similar to traditional SDI service which periodically provides information that coincides with user' s interests and preferences. Some of libraries and information centers currently provide several types of CIS like the "MyLibrary" or "My Menu" services. In a study of its usefulness, 75% of the library users surveyed were satisfied with the service (Kim and Koo 2002). However, the CIS still cannot solve the problem in extracting and providing the exact information the user needs from the tremendous amount of available information. The CIS also has the limitation of not being able to correspond with the changing nature of users' profiles. Because of those limitations in satisfying user' s information need, libraries and information centers are currently considering Personalized Information Service (PIS) as an important and essential information service to analyze users information-seeking behaviors and needs. PIS can be provided with various technologies such as information retrieval technology, push technology, web agents, and so on. Also it aims to satisfy user' s needs by providing information matched to their profile which generally includes demographic information and private preference inputted by the information users and information using behavior extracted with data mining techniques. Various service organizations currently offer PIS as an adapted and more advanced form of CIS.

In this viewpoint, personalized recommendation services can play an important role in many applications

because it is too expensive and not time-efficient for all users to learn and search about all possible alternatives. Depending on the specific application setting, users may be online shoppers, information seekers or an organization seeking a specific expertise. Personalized marketing and service-processing mechanisms have recently attracted significant industry interest in e-commerce areas including online shopping. Personalized recommen dation services include a service that organizes service menus which are similar to the MyYahoo! service provided by Yahoo.com, a method to recommend suitable contents matched with user preferences in e-commerce, a method to provide banner advertisements that the user is likely to be interested in, and so on. Because the personalized recommen dation service emerged from e-commerce areas, the term "personalization" can be widely used and applied in various online service areas. However, even though PIS was initially developed in e-commerce, it can be applied to many service areas.

As mentioned above, libraries need to provide appropriate and information matched with user information needs under information overflow environment. PIS can be the best to solve the problems with which libraries are faced. To provide more efficient and real information service for users, this study analyzes existing recommendation methods and proposes a more useful and efficient recommendation method to provide PIS for information users in libraries. In addition to the many various and new types of information resources including audio, video, and digitalized test files which libraries are processing and managing, the amount of sources is rapidly increasing. With these changes in mind, this study proposes, implements and evaluates a framework for a recommendation method.

Based on using data ming and IR techniques, the purpose of this paper is to extract a list of contents recommended from a large amount of transaction records.

The proposed recommendation system is to carry out a personalized recommendation while operating in real time. In order to process the data in large scale, sequential pattern and clustering technique of data mining has been used. Based on the above techniques, a recommendation algorithm for a personalized recommendation in real-time is proposed. The implemented system is composed of a clustering module, a newly developed pattern matching module, a pre-processing module and data transformation and transition module.

In the experiment using real-world large data, personalized recommendation is given in almost real-time and shows acceptable correctness.

2. Related Works

2.1 Recommendation system

Academic issues on personalized recommendation service were originally announced in mid-1990 (Hill et al 1995;

Rensnick et al 1994). The initial stage just focused on recommendation. The issues on recommendation have been widely studied in the areas of IR, data mining and e-commerce. The personalized recommen dation is best known for its use on e-commerce web sites, where users' interests and behavior are used to generate lists of recommended items. Since the service was introduced, a variety of recommendation methods have been developed to increase the accuracy and usefulness of recommen dation services (Balabanovic and Shoham 1997; Basu 1998). The collaborative filtering method among those methods is the most successful recommendation method that has been used in applications such as recommending web pages, movies, articles and digital contents. However, even though the collaborative filtering method, which identifies neighbor users who have common interests and recommends their preferred contents to a target user, has shown good performance, it has exposed two major limitations which are scarcity and scalability problems (Claypool et al 1999; Sarwar et al 2000; Kim 2005). Scarcity problem means that the number of contents with ratings is very small compared to the number of contents that need to be rated because typical collaborative filtering requires explicit non-binary user ratings for similar contents. As a result, collaborative filtering-based recommendations can't accurately compute the neighbor users and identify contents to be recommended. The second is related to scalability. A process to find neighbor users usually requires a very long computation time that grows linearly with both the number of users and the number of contents. Therefore, the collaborative filtering-based recommendation has serious scalability problems in the millions of users and contents environment. Recent studies have suggested data mining technology as an enabler to overcome the problems associated with collaborative filtering since it will reduce the need for obtaining subjective user ratings or registration-based personal preferences (Mobasher et al 2000). Web log data are richly detailed compared to off-line user data and click stream in a certain web site provides information essential to understand user behavior patterns on a web site such as what contents they downloaded, used or bought(Burke 2002; Linden, Smith, and York. 2003). Through analyzing such information (i.e. web usage mining), it is possible to make more accurate analysis of a user's interests or preferences across all contents than analyzing their purchase records only. Furthermore, mining association rules from click stream provides rich and interesting relationships or associations among contents, compared to the conventional mining association rules from purchased records. Nevertheless, the existing researches have not given a formal way for capturing individual user's preferences or associations among contents through web usage mining.

2.2 Data mining

Data mining is a non-trivial process of

identifying valid, novel, potentially useful, and an ultimately understandable pattern in data. Typically, the applications involve large-scale information banks such as data warehouse (Shardanand and Maes 1995).

There are many techniques in data mining, but the sequential pattern and clustering technique were used in this study (Weng and Liu 2004).

Sequential pattern mining is defined as finding a complete set of frequent subsequences in a set of sequences. Thus, it, for example, allows analysis of user buying patterns during separate visits in a market basket problem (Sarwar et. al 2001).

A sequence is an ordered list of itemsets(S).

$S=\langle s_1,s_2,s_3,\ldots,s_n\rangle$ where an itemsets is a non-empty set of items.

Generally, sequential pattern $\langle s_1,s_2,s_3,\ldots,s_n\rangle$ can be represented as the form of . $s_1,s_2,s_3,\ldots,s_n$. A sequence(a) $\langle a_1,a_2,a_3,\ldots,a_n\rangle$ is contained in another sequence b : $\langle b_1,b_2,b_3,\ldots,b_n\rangle$ if there exists integers $1\le i_1\langle i_2\langle\ldots\langle i_n\le m$, such that $a_1\subseteq b_{i1}$, $a_2\subseteq b_{i2}$, ... , $a_n\subseteq b_{in}$. (Choi, Lee and Lee 2005; Rensnick 1994)

The support for a sequence is defined as the fraction of total users who support this sequence. A user supports a sequence 'S' if 'S' is contained in the user-sequence for this user.

Clustering means identifying data clusters or segments that have similar properties from random data users (Srikant and Agrawal 1996). In this study, k-means clustering technique is used for clustering users who have similar preferences. The technique would be useful for grouping neighbor users who have similar preference on music contents.

3. The Proposed Personalized Recommendation System

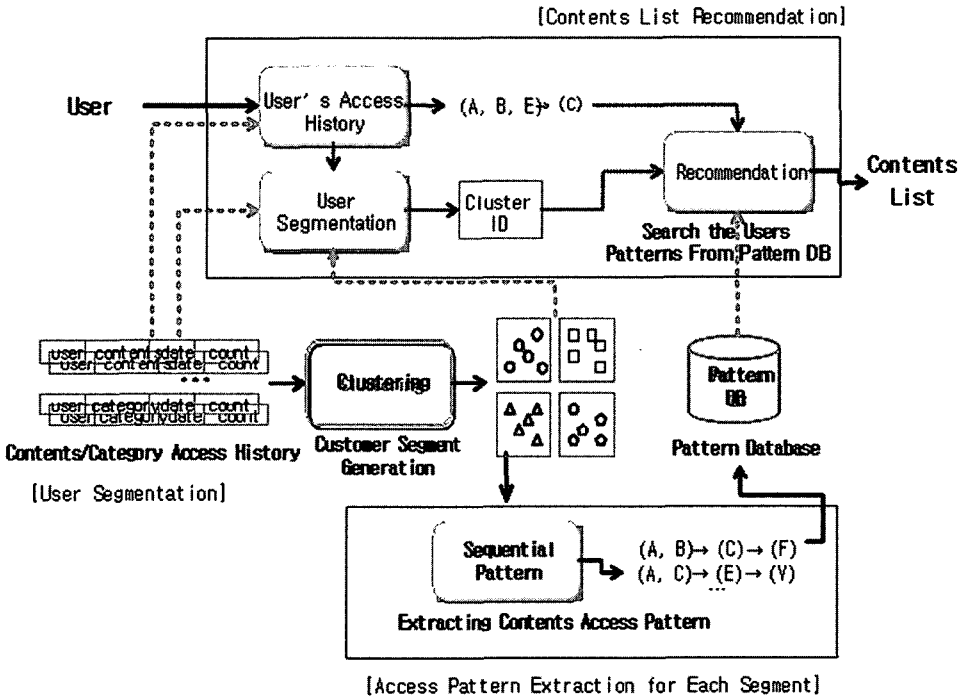
3.1 System overview

This study aims to increase recommen dation accuracy in large scale digital contents and users environment and solve problems occurred in the existing recommendation methods. To achieve the goals above, this study does not use users' demographic information such as sex, age, address, and contents attribute such as genre, director, singer, etc but use contents access history showing user behavior patterns on contents preference as a major consideration. Currently there are so many contents types such as audio, video, image, text and so on created in various areas. Also, the number of digital contents and users in web sites including electronic library, which provides digital contents, is comparatively a large scale size compared to past. Because of it, it is difficult to apply existing recommendation methods used for small or medium-sized web sites to current big web sites providing large-sized contents such as Amazon, e-Bay and so on. Also, the existing methods are based on user's explicit feedback to analyze user preference. But, this assumption can be applied to recommendation because

users generally has a tendency not to perform extra works like evaluation of recommendation contents after completing using or purchasing contents. Also, even though users perform extra works like evaluation, using demographic information input by users contains critical problems for recommendation because users' criteria of evaluation for contents and products are individually different. That is, contents users have a tendency to input inaccurate information because of privacy and security issues. If a recommendation system recommends useless and inaccurate contents extracted with inaccurate information input by users, the recommendation would make users waste time and effort. Because of those problems above, this study uses users' usage frequency of contents that shows implicitly user preferences and interests as analyzing user contents using and searching behaviors. Analysis of preference based on contents usage frequency could be most accurate and efficient recommendation methodology.

The recommendation system based on the proposed method consists of 'collection and conversion module', 'learning module' and 'integration module'.

As used in preprocessing process, the user segmentation which has a role in collection and conversion module. The module collects web log data and converts it to a required format for this study as



〈Figure 1〉 The Proposed Recommendation Architecture

eliminating needless components of web log. For effective and efficient performance in calculating similarity among users and extracting of rules, pre-processing and learning works are performed in batch processing while the recommendation is conducted in real-time processing.

The learning module extracts user's contents usage pattern. The recommen dation module selects the final recom mended contents.

〈Figure 1〉 shows the recommendation system architecture used in this study.

As shown in the 〈Figure 1〉, user's implicit responses after recommending a certain contents would be considered as user feedback in learning process. It is an effective key function to increase accuracy of recommendation because user profile can be updated.

The overall structure of proposed system is suggested in 〈Figure 1〉.

It performs the recommendation with the following three procedures of user segmentation, contents access pattern extraction and contents list recommen dation. In the first two procedures, clustering and sequential pattern techniques were used.

Contents recommendation is a real-time procedure that produces the contents access history record and recommendation algorithm from the segment where the user belongs to.

A mass data processing part is performed through batch work process using user segmentation and pattern extraction. An individual recommendation process is performed by utilizing the previous learning result.

3.2 User modeling

User modeling utilizes only usage patterns, ignoring user's demographic information. A user's contents usage data should be extracted from log in the pre-processing stage. The extracted data in 〈Table 1〉 shows a general structure. The data are processed into the form of 〈Table 2〉. It contains four different information such as who, what kind of categories, what kind of contents, and when accessed. This information is processed into the form of 〈Table 2〉, it is individual access histories. In this process, access time is used in date

〈Table 1〉 Web Log Structure

User ID	Category ID	Contents ID	Access Time

User ID	Contents ID	Access Time	Access Count

〈Table 2〉 Access Log Structure

User ID	Category ID	Access Time	Access Count

unit for effective data processing and reducing the number of data processing. This structure is used in all three of the recommended processes.

3.3 Segmentation

It is necessary to extract the neighbor in order to make a recommendation. Clustering based on the preference regarding user's contents category is used to perform extraction of neighbors in the large scale users environment.

The general methodology can be described as following:

First, for all categories, preference on it is summed up to the parental or ancestral preference to the level 2 of the contents category tree (Choi, Lee and Lee 2005).

Second, in the level 2 categories, the preference value is collected in some predefined duration (14 days), then converted to feature vectors of users. Because a lots of log data are usually collected in a days we collect log data of 14 days as an appropriate size for this study.

Third, user clustering is produced using k-means clustering algorithm on the feature vectors.

3.4 Patterns of user contents accessing behavior

In the process of contents access pattern extraction, it extracts sequential patterns from the contents with high access frequency based on sequential pattern mining. To do it, lots of contents access information of each user in user segment over a period of time (i.e. 7 days) should be accumulated. Then, it extracts sequential contents access pattern from accumulated data and saves it into sequential DB.

The sequential contents access pattern S is defined as following.

$S=(P_1,P_2,\ldots,P_n)$, P is an item set

The following functions are used to get the part of a sequential pattern.

$Head(S_i)=S_1,S_2,\ldots,S_{n-1}.$, $Tail(S_i)=S_n$

The Sequential DB, which saves the sequential pattern extracted from user clustering, can be defined as follows.

The sequential DB saving the Squential contents access pattern that is extracted from user clustering 'i' is $SeqDB_i$

With assumption above, if there are 'k' number of user clusterings, then also be 'k' number of SeqDB that contains different sets of sequential patterns.

The methods to extract SeqDB based on each user segment are followed by:

First, it groups user-contents access patterns in each date for a user cluster(i).

Second, it performs sequential pattern mining on the contents information of all the users in the user segment to gets the results of sequential patterns of user clustering, and saves them in $SeqDB_i$.

3.5 Recommendation using the proposed method

Using sequential DBs for each user clustering with clustering and sequential

Input : User(u)
Output : 'p' Unit of contents
begin
1. Extract contents access pattern S_u of a User(u) from access history.
2. Decide User segment 'i' that a User(u) is included in.
3. for (each sequential pattern S_t of $SeqDB_i$)
if (Include(Head(S_t), S_u)
then Extract the contents list from Tail(S_t)
4. Sort the contents list based on support, and select the top 'p' of contents.
end.

〈Figure 2〉 Recommendation Method

pattern, the proposed recommendation algorithm that performs the personalized recommendation in real-time is presented in 〈Figure 2〉.

4. Experiment

4.1 Test data and environment

The system is implemented in Intel PentiumTM 4, 2.8 GHz, 512MB, and MS Windows 2003TM by C++ and STL (Standard Template Library).

In preprocessing step, Weblog analysis tool and SQL Server 2000 are used for data transformation. The equipment used for implementation of each module is shown in 〈Table 3〉.

The experimental data are selected from the log of the music streaming service website served in Korea that has more than 4 million users and 0.6 million music contents. For sampling users and contents used for learning and evaluation step, 72 million weblogs for 14 days are collected. The collected log data are transformed into the required format in this study.

More detailed information including the

〈Table 3〉 System Environment

Module	Developing Tool
Preprocess, Data Transform	SQL Server 2000TM
Segmentation, Pattern Extraction	Visual C++TM
Recommendation	Visual C++TM

〈Table 4〉 Experimental Data

	Duration and days of log	Number of users	Number of contents
Learning data	36,007,553 (7days)	430,510	191,893
Test data	36,066,981 (7days)	338,150	193,617
Total Number of member in the web site		About 4,000,000	
Total Number of contents provided in the web site		About 650,000	
Nmber of daily users access		About 1,000,000 / day	

〈Table 5〉 Access History Data Size

Contents Access	About 9,000,000
Category Access	About 3,000,000

size of web site for this study is shown in 〈Table 4〉. Also, access history on contents and categories is shown in 〈Table 5〉.

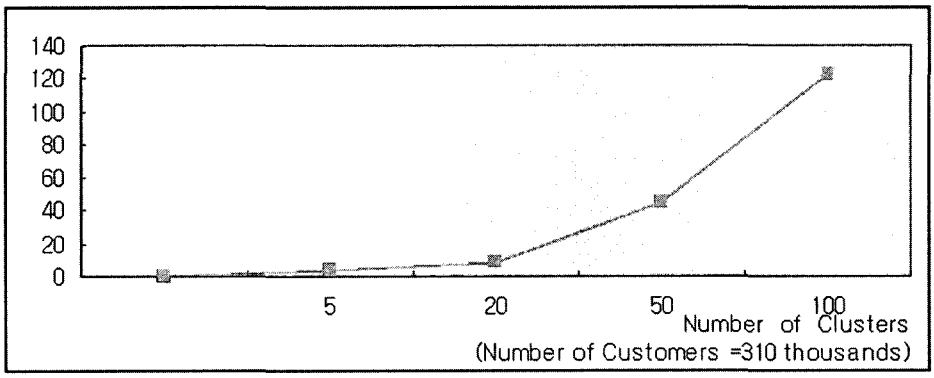
4.2 The results of execution time

The running time of the clustering is presented in 〈Figure 3〉.

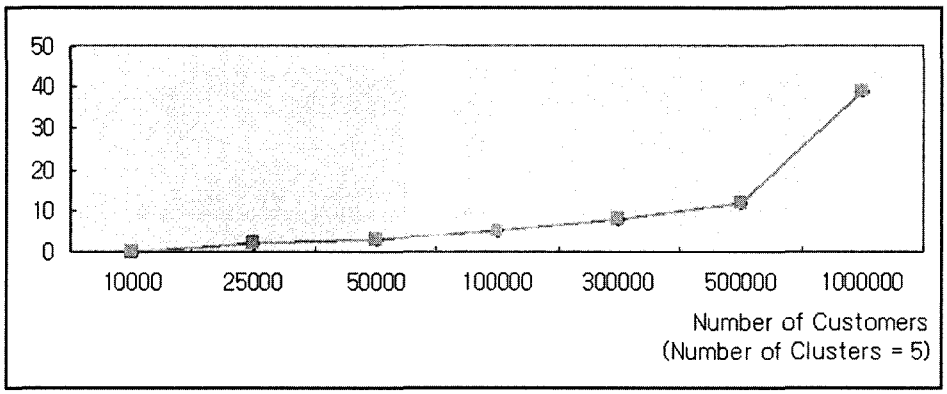
As increasing the number of users, the running time is also increased as shown in 〈Figure 4〉. Specifically, the running time for clustering is about 0.03 m/s per a user.

Cluster numbers (1 ~5) are assigned and the results of clustering is shown in 〈Table 6〉. The following examples are based on these typical results.

The running time of contents access pattern extraction on some clusters is



〈Figure 3〉 Execution Times by Cluster Number



〈Figure 4〉 Execution Times by User Number

presented in 〈Figure 5〉 and 〈Figure 6〉. The sequential pattern mining has the properties that running time and numbers of the result patterns are sensitive to the value of minimum support. To handle large capacity data, automatic support adjustment technique is used in order to keep the running time of sequential pattern mining in a limited time. Thereby this effect, it shows that the running time has been increased modestly while Minimum Support has increased.

The experiment was performed to assess the recommendation result as follows.

Step 1. Decide users segments based on the data in 7 days of the first half in the log data of 14 days, and extract sequential pattern.

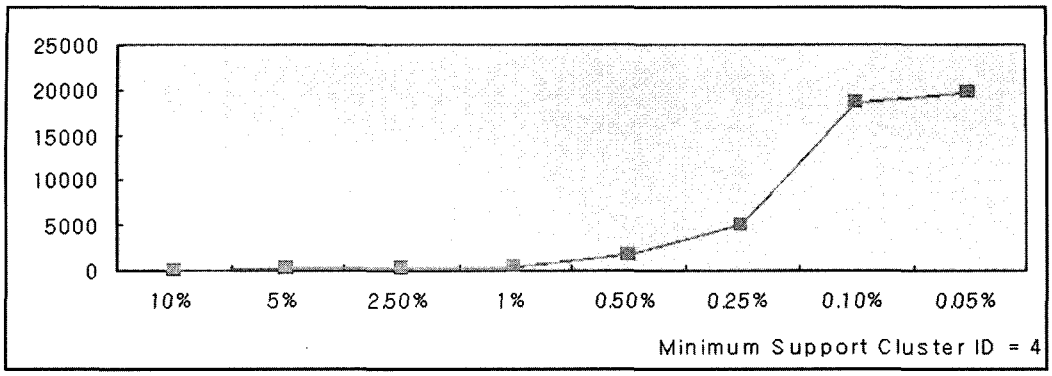
Step 2. Repeat following Level 3 & 4 on the users who have access history in 7 days of the last half. It is about 100,000 persons.

Step 2-1. Get five recommend contents from Recommendation Algorithm. The recommendation is only based on first half duration.

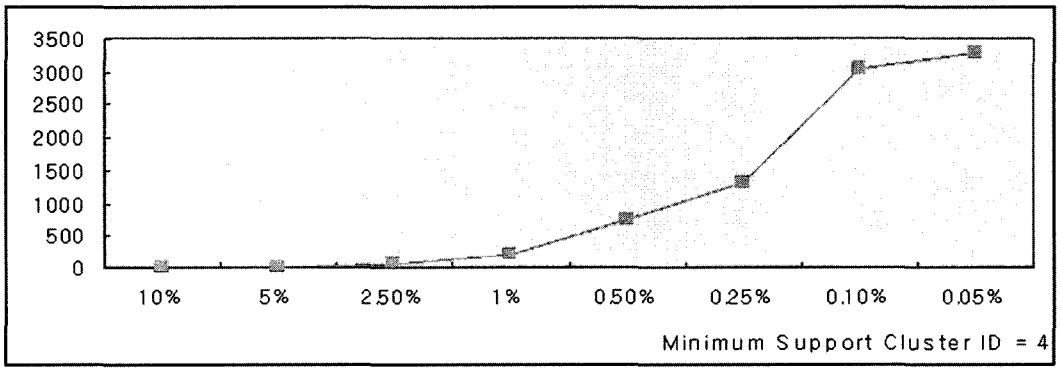
Step 2-2. Check if specified users have had accessed to the five recommend contents for 7 days of the last half.

〈Table 6〉 Clustering Experiment Result

Cluster ID	Number of User	Usage Count
1	8,522	58,588
2	224,743	2,586,351
3	39,887	337,242
4	37,013	784,984
5	2,953	32,267



〈Figure 5〉 Execution Times by Minimum Support



〈Figure 6〉 Number of Pattern by Minimum Support

〈Table 7〉 Recommendation Execution Time

	Number of Recommended contents = 1	Number of Recommended contents = 5	Number of Recommended contents = 10
Number of Users recommended	Running time (sec)	Running time (sec)	Running time (sec)
100	0.032	0.031	0.046
1,000	0.407	0.437	0.438
10,000	4.61	4.593	5.109
100,000	41.782	43.172	44.657

Step 3. Calculate evaluation scales about 100,000 users of recommendation perfor mance.

In this experiment, the results of recommendation speed are as shown in 〈Table 7〉:

4.3 Performance evaluation

To evaluate performance of the proposed method, the latter half of total log data for 15 days are used. The number of log data is 36,066,981 usage data. But, because there are some duplicated usage cases used by the same user, we consider the duplicated usage cases in one contents as one time usage. As a result of it, total usage log is 11,832,631 usages. More detailed input data are mentioned in 〈Table 5〉.

For sampling users, 0.5% (1,668 users) of the total users in the test data are selected and the number of recommendation contents are fixed at 1, 5, 10 contents. For the effective measurements, Recall, Precision, F-Measure and Success Ratio are used. Among those measures, The two most commonly used measures of relevance are the ratios of recall and precision.

Success Ratio means that if a user uses more than one contents among the recommended contents, the recommen dation is successful. Therefore, the values used in success ratio are '1' and '0'.

Recall is the percentage of relevant contents actually retrieved from the total available pool of relevant contents in the total contents set. It measures how well the system is able to retrieve relevant contents.

$$Recall = \frac{what was obtained useful (Hits)}{total useful contents in total contents (Hits+Misses)}$$

Precision is the ratio of the relevant contents actually retrieved to the total set of retrieved contents, which may consist of both relevant and nonrelevant contents. It gives an indication of how well the system provides relevant contents from among those retrieved.

$$Precision = \frac{what was obtained and was considered useful (Hits)}{total of what you got from the total contents (Hits+Noise)}$$

Generally, Recall and Precision are in inverse proportion to each other. For more accurate measure, an adjustment method should be adopted. Lewis et. al (1994) proposed F-Measure to make up for the weak points in Recall and Precision. This study also adopts it.

$$F_{\beta} = \frac{(\beta^2+1) \cdot preision \cdot recall}{(\beta^2+1) \cdot (preision \cdot recall)} \; 2 \; \frac{preision \cdot recall}{preision+recall})$$

In this study, β is set as "1" considering same importance to Recall and Precision.

It is shown in 〈Table 8〉 that the recommendation performance evaluation scales are resulted in each recommendation experiment

The analysis result is shown as a low value, but it should be considered that these are different scales with general recommendation systems. Because the result is not the analysis on user's behavior when the recommendation result is shown to them, but it is the scales on user behavior prediction. It is considerably

predicted that 12% of users was concerned with more than one contents in recommen dation results. That is, as one of the measurements, Success Ratio means that if a user uses any recommended contents regardless of the number of the recommen ded contents, it can be considered as successful recommendation. It is a sort of criteria to evaluate success or failure of recommendation.

As shown in 〈Table 8〉, the proposed method has high performance in Success Ratio, compared to other measurements. Especially, as increasing the number of the recommended contents, it shows better performance in Success Ratio. The experiment result shows that contents recommended by the proposed method contains contents which have higher user preference.

In Precision measurement, one of the distinctive features in this study is for large scale users and contents. Therefore, it is comparatively difficult to recommend contents which accurately accord with user preference compared to in small-medium users and contents environments. We can easily expect that the performance in Precision is lower than in small-medium users and contents environments.

In Recall measurement, Recall in this experiment shows that the recommended contents contains contents in accordance with user preference as increasing the number of the recommended contents. As shown in 〈Table 8〉, performance of the proposed method in Recall is comparatively lower than in Precision.

In F-Measure which is a measurement to adjust Recall and Precision, it assigns 'β' in F-Measure to Recall or Precision as weight value. The range of 'β' is from 0 to 1. In this experiment, we equally assign '$\beta = 1$' to Recall and Precision.

〈Table 8〉 shows the result of the performance evaluation based on F-Measure. The result is similar to it in Recall. It is because the 'β' value is equally assigned to Recall and Precision.

5. Conclusion

The goal of this study is to solve the inaccuracy problem and provides efficient and useful recommendation service in large scale users and contents environment. To achieve the goals, this study proposes a hybrid recommendation method which

〈Table 8〉 Recommendation Result

Number of Recommended contents	Precision	Recall	F-Measure	Success Ratio
1	3.7%	0.3%	0.5%	3.7%
5	1.9%	0.7%	1.0%	8.1%
10	1.5%	1.1%	1.3%	12.0%

accommodates advantages user clustering and data mining techniques. With the proposed method, we can partly solve inaccuracy and computational complexity in existing recommendation methods. The contribution of this study is to provide framework for personalized recommen dation service in large scale users and contents environment.

Handling about 0.76 million users and 72 million data in personal computer, segmentation and pattern extraction has been performed in several hours. As well, a personalized recommendation system is implemented that performs the recommen dation of 2,000 people per second at once. This framework can be applied to huge electronic library, which need to process and manage over millions of information and users.

To show the validity of the recommen dation system, some recommendation evaluation scales are examined. The value of recommendation accuracy factors was close to 1.5 % and the recommendation Success Ratio was close to 12%. The experiment data are different from existing studies in properties. As well, since it was performed in a new environment with the large data capacity, we could not compare mutually with existing studies about validity. However, it might be possible to commercialize the proposed system with a large capacity even though it was not a good result that accuracy or reappearance was just around 10 %.

In the future, recommendations concerning popularity or oldness of contents, or recommendation concerning about contents or category attribute should be considered.

References

Balabanovic, M. and Y. Shoham. 1997. "Fab: Content-based, collaborative recommendation". *Communications of the ACM*, 40(3): 66-72.

Basu, C., H. Hirsh and W. Cohen. 1998. "Recommendation as Classification Using and Content-based Information in Recommendation". *Proc. of the Fifteenth International Conference on Artificial Intelligence (AAAI-98)*, 714-720

Claypool, M., A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin. 1999. "Combining content-based and collaborative filters in an online newspaper". *Proc. of the ACM SIGIR Workshop on Recommender Systems.*

Hill, W., L. Stead, M. Rosenstein, and G. Furnas. 1995. "Recommending and evaluating choices in a virtual community of use". *Proc. of the*

ACM CHI' 95 Conference on Human Factors in Computing Systems. 194-201.

Choi, Hyun-Wha, Dong-Ha Lee and Jeon-Young Lee, 2005. "Multi-Level Linear Location Tree for Efficient Sequential Pattern Mining". *Key Engineering Materials*, 277(1): 369-374.

Kim, Hyun-Hee and N. Y. Koo. 2002. "A Study on the Design and Evaluation of the Model of MyCyber Library for a Customized Information Service". *Journal of the Korean Society for information Manage ment*, 19(2): 132-157.

Kim, Yong, S. B. Moon. 2005. "A Study on Development of Hybrid Personali zation Recommendation System based on Learning Algorithm". *Korean Journal of the Library and Information Science*, 39(3): 75-81.

Mobasher, B. H., T. L. Dai, M. Nakagawa, Y. Sun, and J. WIiltshire. 2000. "Discovery of aggregate usage profiles for web personalization". *Proc. of the Workshop on Web Mining for Ecommerce-Challenges and Opportunities*.

Rensnick P., N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. 1994. "GroupLens: An open architecture for collaborative filtering of NetNews". *Proc. of the ACM CSCW' 94 Conference on Computer-Supported Cooperative Work*, 175-186.

Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. 2000. " Analysis of recommendation algorithms for e-commerce". *Proc. of the ACM Conference on Electronic Commerce*, 158-167.

Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. 2001. " Item-based collaborative filtering recommen dation algorithms". *Proc. of the 10th International WWW Con ference*, 285-295.

Shardanand, U. and P. Maes. 1995. "Social information filtering: Algorithms for word of mouth". *Proc. of ACM CHI '95 Conference on Factors in Computing Systems*, 210-217.

Srikant, Ramakrishnan and Rakesh Agrawal. 1996. "Mining Sequential Patterns: Generalizations and Performance Improvements". *Proc. of the Fifth International conference on Extending Database Technology (EDBT ;96)*.

Weng, Sung-Shun and Mei-Ju Liu. 2004. "Feature-based recommendations for one-to-one marketing". *Expert Systems with Applications*, 26(4): 493-508.