Personalized Digital TV Content Recommendation with Integration of User Behavior Profiling and Multimodal Content Rating

Hyoseop Shin, Minsoo Lee, and Eun Yi Kim

**Abstract — *This paper presents the novel development of an embedded system that aims at digital TV content recommendation based on descriptive metadata collected from versatile sources. The described system comprises a user profiling subsystem identifying user preferences and a user agent subsystem performing content rating. TV content items are ranked using a combined multimodal approach integrating classification-based and keyword-based similarity predictions so that a user is presented with a limited subset of relevant content. Observable user behaviors are discussed as instrumental in user profiling and a formula is provided for implicitly estimating the degree of user appreciation of content. A new relation-based similarity measure is suggested to improve categorized content rating precision. Experimental results show that our system can recommend desired content to users with significant amount of accuracy.*[1].

Index Terms — digital TV, content recommendation, user profiling, content rating, classification, keyword-based rating

I. INTRODUCTION

The advent of digital television offers us unprecedented versatility of content. As the amount of accessible information increases manifold, users are faced with significant problems when it comes to finding the right content at the right time. However, from the end user's point of view, the usefulness of services is measured largely by the suitability of the content provided. Hence, numerous information systems and applications are being developed to address the need for content customization. Recommendation technology is implemented as part of centralized systems such as Internet-based personal EPG (electronic program guide) services or distributed environments including Digital TV appliances. Cost-effective DTV appliances offer constrained hardware resources and limited collaboration facilities. This inevitably narrows down the range of applicable content recommendation technologies. In this paper, we present a new system for recommending a limited list of ranked TV content items to the user. The system is intended for embedding in DTV appliances and implements a content-based filtering approach. Our design decouples the two basic processes involved in personalized EPG generation: user profiling and content analysis.

The purpose of user profiling is to generate a complete and consistent set of user preferences, which would evolve with change in user's interests and inclinations. Existing recommendation systems achieve this either by using explicit user feedback and preference setting or by implicitly observing certain characteristics of user behavior with regard to content. With explicit preference setting, the user is expected to enter detailed registration and preference data, which requires time and cognitive effort and impedes overall system usability. Implicit preference tracking overcomes user reluctance with regard to preference indication and provides consistent long-term user interests data because it operates constantly without being intrusive.

Known systems typically analyze few aspects of user behavior; in most cases, only the choice of a particular content item for playback or tuning to some channel of broadcasting is taken into consideration. More advanced implementations attempt to make use of simple temporal characteristics of user behavior, such as playback time. It must be understood, however, that the relationship between playback duration and the actual viewing time is in practice rather indirect. It is not uncommon for a TV set to continuously stay on but unattended or just create a background when nobody is actually watching. The limited scope of observed behaviors affects the quality of user profile detrimentally, and our system addresses this problem by providing a framework for analyzing more complex user behaviors that have become possible with the introduction of interactive television systems.

Content rating process is a means of estimating the likelihood of user endorsement of a particular content item. Typically, a user is presented with a limited list of content items that have been rated highly against that user's profile. This list is often referred to as custom electronic program guide(EPG). Existing systems use keyword-based and/or relation-based similarity measures to rank content items against a user's profile. It is clear that keyword-based approach alone, efficient as it is for text filtering, is of limited use with audiovisual content, which is not typically accompanied by comprehensive and lengthy textual descriptions. Relation-based approach implies the existence of

[1] This work was supported by the Technology Infrastructure Foundation Program funded by the Ministry of Commerce, Industry and Energy, South Korea.

Hyoseop Shin is with the Advanced technology fusion Department, School of Internet & Multimedia Eng., Konkuk University, Seoul, South Korea (e-mail: hsshin@konkuk.ac.kr).

Minsoo Lee is with the Computer Science and Engineering Department,, Ewha Womans University, Seoul, 120-750 Korea. Corresponding author. (e-mail: mlee@ewha.ac.kr).

Eun Yi Kim is with the Advanced technology fusion Department, School of Internet & Multimedia Eng., Konkuk University, Seoul, South Korea (e-mail: eykim@konkuk.ac.kr).

Manuscript received April 19, 2009

some predefined hierarchy of content types (e.g., drama, news, sports, music, etc.) The problem with relation-based approach is to estimate the degree of content appropriateness when the type of content item does not exactly match that on the user's profile.

Known systems use relation-based similarity measures considering either depth or path length. Depth-based similarity methods [1] compare two concepts by estimating the depth of the common parent in the hierarchy of concepts. Path-length methods [2] are based on counting links between nodes in a semantic network. Each of these pure methods ignores a significant part of information represented by the tree of content types and relevant to the relationship of tree nodes. Depth-based methods ignore the actual distance between the nodes whose similarity is being measured. Path-based methods, on the opposite, ignore the depth of the nodes, and hence, their degree of generality. Moreover, as the number of possible content type hierarchies is unlimited, and multiple existing hierarchies are unbalanced, leaf nodes within a hierarchy occupying different levels, an idea of an adjustable similarity measure suggests itself. This work describes a new relation-based similarity measure called Descendant Nesting Similarity Measure (DNSM) that uses both depth and distance information, and at the same time is normalized by the number of levels in a given hierarchy sub-tree.

This paper includes a summary of related work in section 2. An architecture overview of the content recommendation system suitable for digital TV appliances is supplied in section 3. Section 4 suggests the framework for analyzing user behaviors in interactive television systems designed to improve implicit preference extraction. Section 5 describes our content rating scheme which includes DNSM, the proposed relation-based similarity measure designed to facilitate categorized content rating. Section 6 describes the experimental results of our proposed system. And our conclusions are provided in section 7.

II. Related Work

There has been a great amount of research effort targeted at content recommendation recently and various approaches have been suggested.

Personalized interactive broadcasting system introduced by Ahn et al. [3] uses ATSC PSIP and MPEG-4 technologies for EPG generation and content recommendation. Content descriptions are matched with a locally stored database of user preferences. Personalized Television Listings (PTV) system described by P. Cotter and B. Smyth [4] is a client-server Internet-based solution that provides intelligent personalized TV guides. The system integrates user profiling, case-based reasoning, content-based and collaborative filtering techniques.

Barbieri et al. [5] present a personal TV receiver with storage and retrieval capabilities that provides user-friendly video browsing tools. The receiver exploits metadata obtained from content provider or generated locally by keyword analysis to derive genre and sub-genre taxonomy for content-based filtering. Explicit and stereotypical user profiling is implemented. This system is different from our approach in that it is focused on automatic content classification and does not implement implicit profiling. Thus, user preferences are static and the system relies on the user to navigate classified content listings and manually adjust their preferences when recommendations are not satisfactory.

A Multi-Agent TV Recommender described by Kurapati et al. [6] encapsulates explicit user preferences, explicit feedback and implicit preferences for content-based recommendation. The system offers two alternative implicit recommender agents: one based on Bayesian approach and another on Decision Trees. However, the system is designed for broadcast TV, not PVR environment. Hence, implicit recommendation is based solely on whether the user has watched content item or not. As discussed in section 4 of this paper, we are using an advanced 2-dimensional classification of user behaviors. Besides that, both Bayesian models and Decision Trees used for recommendation imply a resource-intensive stage of training. Unlike said recommender, our suggested architecture uses keyword information in addition to genres, channels and viewing time for content recommendations.

Difino et al. [7] describe a multi-agent system for personalized EPG generation. The system implements a probabilistic approach to content-based filtering. Our system improves on this approach by decoupling user profiling and content recommending activities. Thus, all recommendation agents match content description data against an integral user profile. In our system, different types of user agents analyze inherently different aspects of content description (namely, structured and unstructured) as opposed to different sources of user preferences.

III. System Overview

The proposed system (see Fig. 1) consists of two cooperating subsystems: user profiling subsystem and content rating subsystem. The user profiling subsystem accumulates various kinds of preferences for each user registered on the system. Thus, a user profile is essentially a database containing keywords, genres, channels, languages, etc., preferred by the user. It is important that some preferences, such as those pertaining to genre and channel, are context-related, i.e. they may depend on temporal viewing characteristics expressed in terms of weekday and time interval values. User profiles incorporate both data provided by the user explicitly and data inferred from the history of viewing behavior. As soon as the descriptions of new content items become available, the content rating subsystem can evaluate them by comparing with existing user profiles.

The user profiling subsystem includes three collaborating profilers that participate in building an accurate and evolving user model:

1. Explicit Profiler manages preferences explicitly set by the user via DTV user interface.

2. Stereotypical Profiler uses a set of pre-defined user stereotypes to fill in any missing preferences. Hence, the user may choose to just select some stereotype instead of manually setting every preference value.

3. Implicit Profiler allows the system to dynamically adjust to the user's changing interests by analyzing her/his watching behavior. Activity Logging Service supplies user action information.

User Profile Manager module of the user profile subsystem maintains extensive data provided by each of the profilers. A set of user preferences is stored in a relational database.

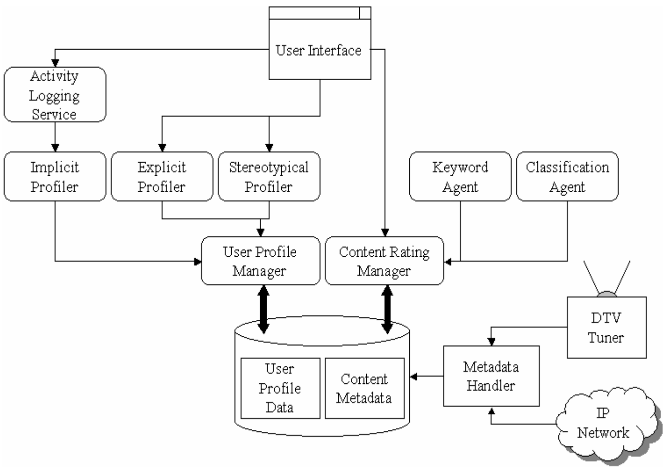

Fig. 1. Recommender components.

The content rating subsystem consists of two agents and Content Rating Manager. One function of Content Rating Manager is to extract and store a unified subset of content description data that would be sufficient for recommendation purposes. Metadata Handler receives raw description data from disparate sources including PSIP Event Information Tables (EITs) and TV-Anytime Content Description Metadata. Another function is to supply user profiles and unified content information to personalized user agents for processing. Ratings offered by agents are merged and made available, upon request, to the DTV user interface modules for recommended content items presentation.

Content recommendation is based on the ratings predicted by the following agents:

1. Classification Agent analyzing formal content classifications including genre, language and channel data. This agent is designed to cope with hierarchical classifications.

2. Keyword Agent analyzing loosely formatted or textual content descriptions provided by ATSC PSIP or TV-Anytime metadata. Keyword Agent performs content rating by calculating correlation coefficients between user profile and content description metadata for each user and content item.

Fig. 2 demonstrates the constituent parts of suggested content recommendation process. Two activities run in parallel to provide the information for content recommendation: user profiling and content metadata extraction. User profiling process involves user behavior logging, evaluating content items the user has acted upon and finally building the user profile using the descriptions of highly rated content. Further information regarding implicit preference extraction is provided in section 4.

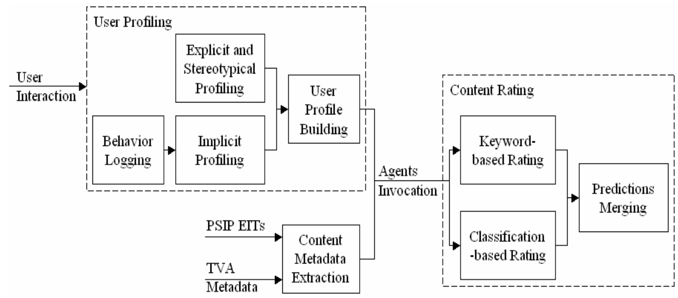

Fig. 2. Content Recommendation Process.

Content metadata extraction process comprises parsing PSIP Event Information Tables (EITs) for basic descriptions of upcoming content and then extending those with rich TV-Anytime Content Description Metadata (when available). ATSC Program and System Information Protocol for Terrestrial Broadcast and Cable (PSIP) standard is a collection of tables designed to allow navigation and access to each of the channels within digital TV broadcast and to provide event descriptions that give the user content information for browsing and selection. To facilitate the latter purpose, PSIP defines Event Information Tables (EIT). However, PSIP information alone is not enough for accurate content rating and may be augmented with TV-Anytime content description metadata when possible. TVA content description includes Program, Group and Credits Information Tables. These tables form a rich and structured description of TV programs and their groups including title, synopsis, genre, cast and keywords. Once parsed and matched with the corresponding PSIP data, TVA metadata allows for both classification-based and keyword-based content rating.

User profiles and content metadata enable the process of upcoming content rating, which consists of invoking keyword and classification agents to generate their predictions and then merging the predictions into a unified content rating. Top-rated content may then be recommended to the user. Please refer to section 5 of this paper for a discussion of classification-based content rating.

IV. IMPLICIT PROFILING

User profiling combines explicit and implicit techniques to build a comprehensive and evolving set of user preferences. This section focuses on the discussion of implicit profiling; explicit techniques and stereotypical approaches have been described in other works in detail.

Identifying user behavior is the cornerstone of implicit preference extraction. We propose two-dimensional modeling of user behaviors relating to DTV content in a way similar to the one suggested by Oard et al. [4]. User behaviors are classified into three categories in accordance with their purpose: *consume*, *store*, and *approve*. *Consume* category involves user actions relevant to content playback and navigation. It is clear that content viewing characteristics are indicative of user interest in content. *Store* category includes user activity directed towards content retention. Storing content items suggests the intention of future use and, besides that, demonstrates user's readiness to allocate certain system resources. *Approve* category groups behaviors evincing willingness to add value to content or keep track of it (e.g., by annotating, referencing or organizing it). Additionally, it resembles store category in that it may imply the possibility of use in the future.

Besides traditional TV program manipulation, modern DTV systems allow the user to manage portions of programs or collections of programs. Therefore, content can be naturally classified into four scopes: *video frame*, *program segment*, *entire program*, and *program group*. Viewing a program as a group of segments simplifies content navigation, allows customized playback patterns and provides opportunities for reducing bandwidth and storage requirements. On the other hand, groups of programs allow leveraging multiple programs at once. Some behaviors, such as *request info,* may be observed at a number of scopes, others, e.g. *bookmark*, belong to a single scope. Observing behaviors with regard to content scope provides more reliable evidence of the value the user ascribes to content item. For example, if the user spends most of the program viewing time examining a particular program segment, that segment's characteristics may be used for profiling, as opposed to those of the entire program. Table 1 shows user behaviors that in general are feasible to detect in a DTV system with recording capabilities (the actual list of behaviors depends on system functionality).

TABLE I
OBSERVABLE USER BEHAVIORS FOR DTV CONTENT

	Video Frame	Program Segment	Entire Program	Program Group
Consume	Pause	Request info Play Pause Rewind Skip backward Fast forward Skip forward Stop	Request info Schedule to tune Tune Play Pause Rewind Skip backward Fast forward Skip forward Stop	Request info
Store	Capture still image Copy still image		Schedule to record Record Copy Delete	Schedule to record Record Delete
Approve	Bookmark Annotate	Add to favorites Annotate	Add to favorites Annotate	Add to favorites Annotate

Each of the observable behaviors or actions is assigned an experimentally adjusted weight coefficient. It is important that behaviors pertaining to different categories need to be evaluated in different ways. Generally, events of the *consume* group occur more frequently on longer content items and may be normalized by content item duration. On another hand, recording long programs is more costly in terms of system resources, prompting the user to favor short recordings. It may therefore be assumed that *store* operations on longer items are more important. In contrast, *approve* behaviors appear independent of content duration. For *play* action, the coefficient must be proportional to the duration of playback. Repetition of actions, such as multiple repetitive rewinds and playbacks of video fragments, increases content rating. Content deletion or skipping, on the contrary, decreases the assigned rating. The following simple expression can be used to estimate user appreciation degree R_d of a content item d :

$$R_d = \frac{\sum_{i=1}^{m} c_i}{T_d} + T_d \times \sum_{j=1}^{n} s_j + \sum_{k=1}^{p} a_k \qquad (1)$$

where

$c_i, s_j,\ a_k =$ weights of consume, store or approve actions i , j , k ;

$m, n, p =$ numbers of actions observed on content item d for categories *consume*, *store* and *approve* respectively;

$T_d =$ the duration of content item d .

The history of observed behaviors and related content information are stored until their quantity has reached a predefined threshold. After a sufficient number of content items have been estimated as shown above, newer items replace the oldest ones so that storage capacity is conserved. Keyword and classification information related to highly appreciated content is added to the user profile.

V. CONTENT RATING

The most accurate predictions can be obtained when using a combination of keyword- and classification-based rating. The results from both agents are merged to produce unified ratings.

A. Classification-based Content Rating

This section is dedicated to the classification-based approach. There are several types of classification preferences stored in the user profile. They include genre, channel and language preferences and their list may be extended. Content item information is related to the same classifications. The classification agent is proposed for classification-based content rating. It needs a relation-based similarity measure to make recommendations in accordance with available TV content classifications such as TV-Anytime genre taxonomy.

For example, let us consider the classification in Fig. 3. One of the known depth-based methods, MSCA (most specific common abstraction) [1], would assign the highest degree of correlation to a football match and a croquet program because their most specific common ancestor – team sports – is the deepest non-leaf member. However, from the fact that the user has specified football as her/his particular interest we would not normally infer that she/he would also appreciate croquet-related content. Conversely, it makes sense to assume that the user has selected a particular category to clearly distinguish it from its siblings. However, more generalized preference settings (e.g., team sports) are available for those with less specific interests. The same issue presents itself with a purely path-based approach [2]: two sibling nodes are considered almost as similar as parent and child. Besides this, both depth- and path-based measures are sensitive to density or sparseness of different classification segments. Nodes in richly structured parts of classification (such as sports in our example) are consistently judged less similar to one another than nodes in more sketchy domains (e.g. leisure/hobby).

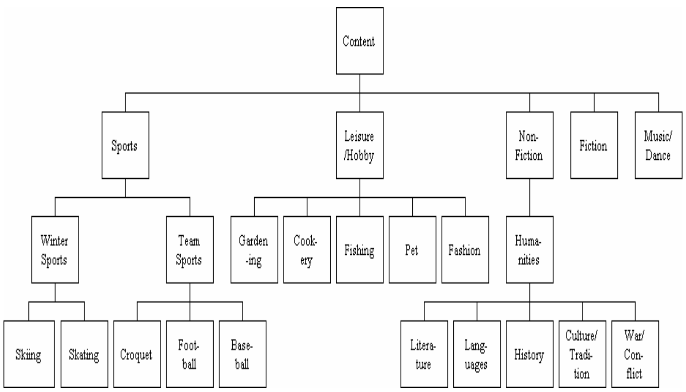

Fig. 3. A portion of TV-Anytime genre dictionary

To overcome these drawbacks, we suggest a new relation-based measure called Descendant Nesting Similarity Measure (DNSM). Like MCSA, this measure takes into account the ancestor depth. Besides that, descendant node's level of nesting is also considered. Both of these are normalized by the overall number of levels in their segment of hierarchy. Table II shows the proposed similarity measure.

TABLE II
DESCENDANT NESTING SIMILARITY MEASURE

Category A	Category B	Assigned Rank
n -th level subcategory	m -th level descendant of A	$\dfrac{n}{\frac{m}{N}+N}$
n -th level subcategory	Non-descendant, non-ancestor of A	0
n -th level subcategory C	N/a	0
N/a	N/a	0

(N is the total number of levels in a given hierarchy segment)

This table is based on the following assumptions:

1. The greater the level of nesting, the greater the specificity of classifier;
2. The more levels between related classifiers, the less the correlation;
3. Unrated content can be ranked as neutral (0).
4. Siblings may be ranked as neutral or higher, depending on the organization of hierarchy. For example, in multilevel hierarchies sibling nodes tend to be more closely related. Obviously, sibling and unclassified content ratings are subject to experimental adjustment.

For our example, the DNSM value for football and croquet would be 0 because they are siblings. If the user has specified team sports ($n=3$) on her/his profile, then a football match ($m=1$) is assigned DNSM value 0.7. However, less specific profile settings result in lower similarity estimates.

B. Keyword-based Content Rating

The purpose of keyword-based relevance rating algorithm is to select content items based on the correlation between the loosely structured (textual) parts of content items descriptions and the user's keyword preferences. The proposed algorithm uses vector space approach to content rating as opposed to probabilistic approaches because it involves no model learning overhead. Overlap-based algorithms have been discarded as they are reported to generally demonstrate inferior performance [8,9].

Not all terms are equal in a given data set, and also with regard to the particular content item. There exist multiple techniques of term weighting. Content item-specific weighting makes no sense for brief descriptions of TV events. Hence, only inverse document frequency (IDF) weighting is required. IDF weighting is based on the assumption that the more times a term occurs in different items in a collection, the more poorly it discriminates between items. Given term weights, standard cosine correlation of vectors may be used to compute similarity measure or content item rank.

C. Merging Multiple Agents' Ratings

Please note that term extraction and weighting needs to be performed upon major updates of content descriptions only. However, term extraction may still be a time-consuming operation in DTV environment because stop list exclusion and stemming are necessary. Accordingly, the initial version of the system is going to take advantage of designated TVA content description metadata sections for keyword identification, hence eliminating the need for stop lists and stemming algorithms.

It has been demonstrated experimentally [10,11] that combined predictions of multiple agents are generally better than single agent's predictions. The purpose of multiple agents' prediction merging algorithm is to combine the strengths of each agent and to mitigate their weaknesses, producing consistent content ratings.

In our system, agents operate on two distinct types of preferences: classification and keyword-based. The relative precision of these preference types may change while the system is being used. It is likely that the initial incomplete user profile is going to eventually get more precise due to implicit preference observation. In coherence with this, it is possible that relative weights of various agents' predictions may change in time. Consequently, adaptive technology based on user feedback is likely to improve prediction quality.

However, collecting explicit user feedback may decrease system usability. On another hand, implicit feedback observation is difficult to interpret, not precise and may lead to false assumptions. This is why a simple fixed-coefficient merging algorithm is proposed:

$$R = \frac{\sum_{k=1}^{n}(a_k \times r_k)}{n} \tag{2}$$

where

$R =$ the resulting combined content item rating for a particular user;

$a_k =$ the weight coefficient assigned to agent k ;

$r_k =$ the content item rating predicted by agent k;

$n =$ the total number of collaborating agents.

Implicit user feedback inference and adaptive prediction merging are possible areas of algorithm improvement in the future.

VI. Experimental Results

To verify the effectiveness of the proposed method, experiments were performed on recommending the TV programs. These programs are divided into three groups according to the objective which they are used: the data for training the linear regression, the data for testing the linear regression-based user profiling scheme, and the data for testing classification-based content rating scheme.

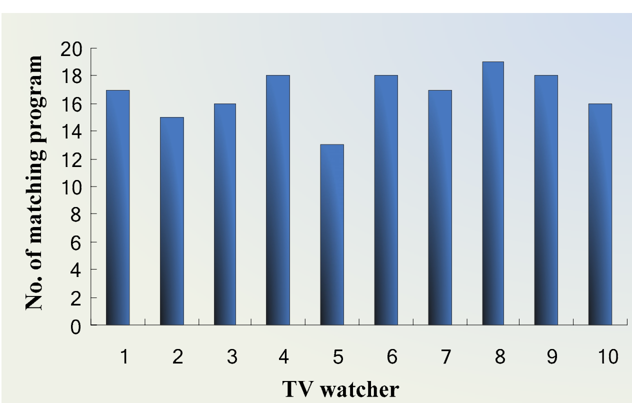

Fig. 4. Performance of the proposed method

For the evaluation of our proposed schemes, we employed 10 persons and let them watch the same TV programs. For randomly selected 100 programs, we let each person use the remote controller during watching the TV program, and then logged the user's activities on the TV remote controller. Based on the investigation, the data for the 50 programs were used for training the linear regression to estimate user's appreciation degree and the others were used for the test.

As shown in Fig. 4, the overall accuracy of our system is 83.5%. Moreover, 8 out of 10 persons said that they actually liked at least 15 TV programs out of 20 programs recommended by the implemented system. This result implies that the system can provide users with quite useful recommendations. And the results show that the content rating precision is enhanced enough by the proposed schemes.

VII. Conclusion

This paper presents an overview of the DTV appliance-oriented content recommendation system architecture. The system currently under development makes extensive use of TV-Anytime content description metadata as well as ATSC PSIP TV event information. This information, collected from versatile sources, provides opportunities for better user preference predictions.

We suggest two types of agents for content-based recommendation. The first type analyzes structured classifiers related to content items, such as the standard TV-Anytime genre classification. The second type operates on textual descriptions of TV content like title and synopsis. The proposed architecture is highly flexible in that it facilitates the storage of an exhaustive user preference data set and at the same time can be seamlessly extended with new agents. Moreover, the architecture is event-driven, so that the system selectively processes changed data on demand.

Further in this paper, we introduce the framework that categorizes potentially observable user behaviors specific to the DTV environment. The application of said framework allows more detailed analysis of user behavior, enhancing the implicit user profile quality. A new relation-based similarity measure called Descendant Nesting Similarity Measure (DNSM) is described. DNSM takes into account both node depth and node distance when rating categorized content items against a user profile, and adjusts itself to match particular content type hierarchies. Thus, content rating precision is enhanced.

TV usage research [12,13,14] demonstrates that television viewing is largely a family or social activity. Therefore, it is beneficial to be able to model groups of people watching together. However, modeling a group of television viewers is not considered in this paper and is seen as a possible future enhancement.

References

[1] J. Kolodner and C. Riesbeck, *Case-Based Reasoning*, Morgan Kaufman Pulishers, 1993.

[2] R. Rada, H. Mili, E. Bicknell, and M. Blettner, "Development and application of a metric on semantic nets," IEEE Transactions on Systems, Man, and Cybernetics, vol.19, no. 1, pp. 17-37, Feb. 1989.

[3] S. Ahn, Y. Cho, J. Choi, J. Kim, and C. Ahn, "Personalized and interactive broadcasting system," UM2001 Workshop on Personalization in Future TV, Sonthofen, Germany, July 2001.

[4] P. Cotter and B. Smyth, "PTV: intelligent personalized TV guides," Twelfth Conference on Innovative Applications of Artificial Intelligence, pp. 957-964, Austin, Texas, Aug. 1-3 2000.

[5] M. Barbieri, M. Ceccarelli, G. Mekenkamp, and J. Nesvadba, "A personal TV receiver with storage and retrieval capabilities," UM2001 Workshop on Personalization in Future TV, Sonthofen, Germany, July 2001.

[6] K. Kurapati, S. Gutta, D. Schaffer, J. Martino, and J. Zimmerman, "A multi-agent TV recommender," UM2001 Workshop on Personalization in Future TV, Sonthofen, Germany, July 2001.

[7] A. Difino, B. Negro, A. Chiarotto, "A multi-agent system for a personalized electronic program guide," 2nd Workshop on Personalization in Future TV, Malaga, Spain, May 2002.

[8] W. B. Frakes and R. A. Baeza-Yates, *Information Retrieval: Data Structures and Algorithms*, Prentice-Hall, 1992.

[9] R.R. Korfhage, *Information Storage and Retrieval*, Wiley Computer Publishing, 1997.

[10] N. Good, J. B. Schafer, J. A. Konstan, A. Borchers, B. Sarwar, J. Herlocker, and J. Riedl, "Combining collaborative filtering with personal agents for better recommendations," Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99), pp. 439-446 , July 1999.

[11] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," Technical Report. MSR-TR-98-12, Microsoft Research, 1998.

[12] J. Masthoff, "Modeling a group of television viewers," 2nd Workshop on Personalization in Future TV, Malaga, Spain, May 2002.

[13] H. Hirakawa, Z. Xu, and K. Haase, "Inherited feature-based similarity measure based on large semantic hierarchy and large text corpus," Proceedings of the 16th conference on Computational linguistics, vol. 1, pp. 508-513, Copenhagen, Denmark, Aug. 1996.

[14] D. W. Oard, and J. Kim, "Modeling information content using observable behavior," Proceedings of the ASIST Annual Meeting, vol. 38, pp. 481-488, Washington, DC, Nov. 2001.

Hyoseop Shin is an associate professor at the department of internet and multimedia engineering of Kunkuk university in Seoul, Korea, since March 2005. He received his degrees of Ph.D., Master, and Bachelor at the department of computer science and engineering of Seoul National University, in 2002, 1996, 1994, respectively. He worked for Samsung Electronics from July 2001 to February 2005. During this period, he led several research projects related to TV-Anytime Forum standardization acitivities and development of relevant digital television software products. In 1999 and the first half of 2001, he visited the department of computer science of University of Arizona as a visiting student scholar and co-worked with the professor, Bongki Moon. These visits were financially sponsored by KOSEF and BK 21 foundation. In 2002, he won the best database paper award from the Korea Information Science Society. His current major research interests are web information retrieval and data mining.

Minsoo Lee is an associate professor at the department of computer science and engineering of Ewha womans university in Seoul, Korea, since March 2002. He received his Ph.D degree from the University of Florida, and his Master, and Bachelor at the department of computer science and engineering of Seoul National University, in 2000, 1995, 1992, respectively. He worked for LG Electronics from July 1995 to July 1996. And worked for Oracle Corporation in the US as a Senior Member of Technical Staff. During this period, he worked on developing business intelligence tools. His research interests include data mining, data warehouse, Web information infrastructures, and stream data processing.

Eun Yi Kim is an associate professor at the School of Internet and Media, Konkuk University, Korea. She was an invited professor of the Dept. of Computer Engineering Information and Science in the Kaya University. She has received a Ph.D. degree in Computer Engineering from the Kyungpook National University, Korea, 2001. And she had been awarded the degree of Master of Science in Engineering in Computer Engineering from the same university, 1999. Her research interests include computer vision, pattern recognition, and robot-computer interaction.