

Content Recommendation and Filtering Technology

Jin Hatano[†], Kyotaro Horiguchi, Masahito Kawamori, and Katsuhiko Kawazoe

Abstract

We describe content recommendation and filtering technology for TV broadcasting that employs user preference information and usage history information. Usage history information can be collected and programming recommendation and filtering can be performed while protecting user privacy using a virtual user ID.

1. Need for recommendation and filtering

Terrestrial digital broadcasting began on December 1st of last year in Japan, and the number of channels available to users is expected to increase as the conversion to digital broadcasting progresses. The larger number of broadcasting channels is expected to result in an increase in the amount of programming provided by broadcasting stations and service providers, which will make the table-like displays of ordinary electronic program guides (EPG) impractically large. A way for users to quickly and easily find programs they would like to see is needed. Conversely, a means of filtering so that unwanted content is not displayed is also necessary.

In this article, we introduce content recommendation and filtering technology for implementing advanced EPGs and electronic content guides (ECGs) for sorting out broadcast-type content and on-demand content that matches the user's needs from a large amount of available programming. We also describe our Advanced EPG and ECG Control System, which was constructed using this technology.

2. Content searching service

The appending of various types of standard metadata to content as a means of quick and easy search-

ing for requested content in a large volume of content has been widely studied. The fusion of communication and broadcasting allows this mechanism to be applied to the field of broadcasting. Technology for cooperative broadcasting and communication provides the three types of searching services listed below.

1. Content and scene searching by free keyword and genre
2. Content searching based on user metadata
3. Content and scene searching based on program description metadata

The first type of searching is conducted on the basis of keywords and genre input by the user. The second type uses the content recommendation and filtering technology that we explain here in detail. The third type uses data content, which enables linking with the Internet and bi-directional transmission in conventional broadcasting. By supplying links to other video content, creators of video content enable various scenarios to be assembled by metadata searching. In the first and third types of searching, the set-top box (STB) can use simple object access protocol (SOAP), which conforms to the TV-Anytime [1] format to search LiteObject, an XML database developed by NTT Cyber Space Laboratories.

3. Four types of metadata

The implementation of content recommendation and filtering technology for content searching based

[†] NTT Cyber Solutions Laboratories
Yokosuka-shi, 239-0847 Japan
E-mail: hatano.jin@lab.ntt.co.jp

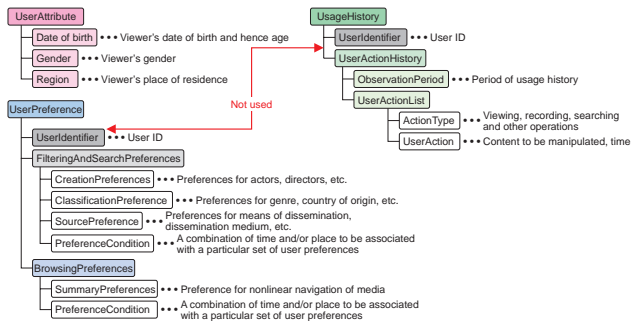


Fig. 1. User metadata schema.

on user metadata uses four types of metadata.

1. User attribute metadata
2. User preferences metadata
3. Usage history metadata
4. Program description metadata

The last three types of metadata all conform to TV-Anytime. User attribute metadata includes the user's age, sex, and place of residence. User preferences metadata can specify the user's preferred genre, keywords, and so on. Usage history metadata can describe actions taken by the user with respect to programs (viewing, recording, etc.) as well as the target programs. Program description metadata can specify detailed content information (title, genre, synopsis, cast, etc.). The first three types of user metadata are summarized in Fig. 1.

The user can select an STB freely from among any TV-Anytime standard STBs available from several vendors and can reuse the program and user metadata for different STBs and broadcasting centers that conform to these metadata standards. Moreover, the circulation of compatible metadata among vendors should lead to lower development and manufacturing costs and larger markets for service providers. Next, we explain how these kinds of metadata are being used.

4. Content filtering

A media asset management platform [2], which is a

system that performs situation adaptive metadata retrieval, is used to display an electronic program guide that matches the user's preferences. This retrieval technique can adapt information processing and display to individuals. It selects and displays content that matches user attributes such as age, sex, place of residence, and preferred genres from a large number of programs. In this technique, the set of content is managed as a tree directory structure in which the content is placed in hierarchical categories. The order of the content items and content categories can be adapted according to individual user attributes (Fig. 2). The attributes of individual users can be customized. For example, various attributes such as marital status or zodiac sign can be added.

To construct an EPG that displays content suitable for the user, we used this situation adaptive metadata retrieval technique to filter content according to user attributes. Content items are listed in descending order of relevance.

5. Two types of recommendation

There are two types of recommendation value: one is attached to each content item and the other is attached to each genre. For on-demand content such as video on demand (VOD) and for future planned broadcast-type content, recommendation evaluations are attached to content item by item. With on-demand content, the recommendation evaluation is based on

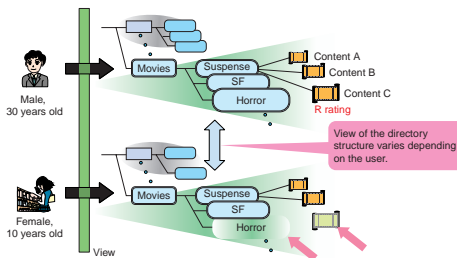


Fig. 2. Filtering with situation adaptive metadata retrieval.

the viewing rate, so that content with a high viewing rate is recommended in the EPG. For the types of content planned to be broadcast in future, the operator can edit the recommendation values, so the broadcasting station or the content service provider can decide what content to recommend to the user.

Broadcast content, such as digital TV programs, is transmitted only one time, except for re-runs, so the recommendation value calculated from the viewing rate of the content cannot be used to recommend future broadcasting content. Therefore, viewing rates for genre are used for recommending broadcast-type content. The genre is extracted from the content

metadata and used to determine what genre a particular user often views. Then, content for future broadcasting that belongs to the often-viewed genre is recommended.

6. Metadata circulation

The Advanced EPG and ECG Control System implements both the content recommendation and filtering functions explained above. In this system, the processing is done in the order shown in Fig. 3. The user attribute information is registered in the STB as initial settings. When the user views or records con-

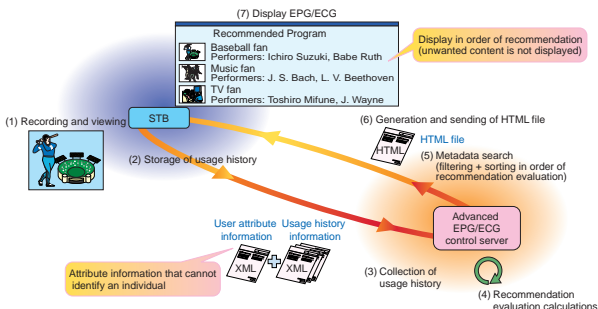


Fig. 3. Flow of recommendation and filtering.

tent, that information is stored as usage history information within the STB. In this system, a virtual user ID is generated from the user attribute information and the user preferences information. The virtual user ID and usage history information metadata are periodically sent to the Advanced EPG and ECG Control System's server. The virtual user ID and the metadata do not contain any information that identifies the user, so the usage history information needed for recommending content can be collected while protecting user privacy. The virtual user ID and usage history information stored in the server is used to calculate recommendation values for content items and genre. These values provide the basis of the sorting order for presenting the content items in the EPG display, thus implementing content recommendation. Three special features of the Advanced EPG and ECG Control System and the advantages they provide are described below.

1. The metadata collected by the STB conforms to the TV-Anytime specifications: Even if the STB is replaced, the same recommendation and filtering services can be received.
2. The usage history information and content recommendation values are managed on the broadcasting station side rather than at the STB: Recommendation services that reflect trends in content can be received.
3. The User cannot be identified from the virtual ID (recommendations are made for groups of users who have the same attributes and preferences): Usage history information can be collected without violating individual privacy.

7. EPG services

The Advanced EPG and ECG Control System is divided into an application layer (AP) and a platform layer (PF), as shown in Fig. 4. The AP layer employs the MVC model (model, view, and controller). This system enables EPG services such as those shown in Fig. 5 to be implemented, and

the MVC model makes it possible to revise the AP layer to easily implement services in various forms.

8. Future development

For broadcasting that involves both push-type (content is sent by the broadcasting center) and pull-type (content is requested and downloaded by the user) services, we have applied content recommendation and filtering technology to search for desired content from among a huge amount of available content and discard unwanted content from among the large amount of incoming content.

In future work, we intend to continue researching new technology and services using content recommendation and filtering technology, such as targeting for commercials and e-flyers (electronic advertising circulars presenting sales information about content and related products).

References

- [1] <http://www.tv-anytime.org/>
- [2] R. Kataoka, N. Kotani, J. Hatano, T. Kitakaku, K. Mii, and H. Tanabe, "Media Asset Management Technology," NTT R&D, Vol. 51, No. 2, pp. 114-123, 2002 (in Japanese).

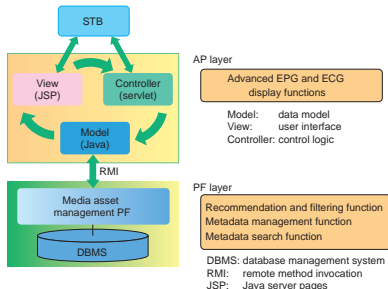
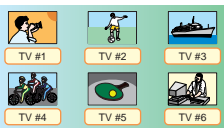


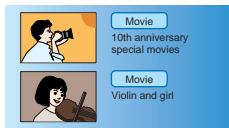
Fig. 4. System configuration diagram.

(1) Display of programs recommended by the broadcasting station



Programs that are recommended by the broadcasting station are displayed. A person at the broadcast station edits the recommendation evaluation of recommended broadcast programs for each user category.

(2) Display of recommended on-demand programs



On-demand programs that are being broadcast are listed in order of recommendation.

(3) Programs displayed by time and channel



Programs are listed by channel in order of time. Programs of preferred genre are emphasized in the display.

(4) Display of recommended genre programs



Programs of the preferred genres are displayed according to genre and programs of often-viewed genres are emphasized in the display.

Fig. 5. Examples of services.

**Jin Hatano**

Promotion Project 1, NTT Cyber Solutions Laboratories.

He received the B.S. and M.S. degrees in computer science from Keio University, Kanagawa in 1994 and 1996, respectively. In 2001, he joined NTT Cyber Space Laboratories, Kanagawa. His current research area is metadata and cooperative broadcasting and communication technology.

**Masahito Kawamori**

Senior Research Engineer, NTT Cyber Solutions Laboratories.

He joined NTT Laboratories in 1989. He has worked in research areas such as artificial intelligence, language processing, and interactive agents using speech recognition. His current research area is metadata and media delivery systems for broadcasting and broadband communications.

**Kyotaro Horiguchi**

Promotion Project 1, NTT Cyber Solutions Laboratories.

He joined NTT in 1994. Recently he has been engaged in the development of a Program Metadata Database System conforming to the TV-Anytime standard based on XML DBMS.

**Katsuhiko Kawazoe**

Senior Research Engineer, NTT Cyber Solutions Laboratories.

He received the B.E. and M.E. degrees in engineering from Waseda University, Tokyo in 1985 and 1987, respectively. Since joining NTT in 1987, he has mainly been engaged in R&D of radio communication systems, satellite communication systems, and the personal handy-phone system (PHS). His specialty is forward error correction systems. He is currently a co-chairman of the Association of Radio Industries and Businesses Working Group for Broadcasting Systems based on a Home Server. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and received the Young Engineer Award from IEICE in 1995.