

A Personalized Recommendation System for Electronic Program Guide

Jin An Xu and Kenji Araki

Graduate School of Information Science and Technology,
Hokkaido University,
Kita 14 Nishi 9, Kita-ku, Sapporo, 060-0814 Japan
Phone & Fax: +81-11-706-7389
{xja, araki}@media.eng.hokudai.ac.jp

Abstract. This paper proposes an idea for constructing a personalized recommendation system for the Electronic Program Guide (EPG). This system would use a basic statistics method with feedback process to predict television programs. We have applied this method to personal prediction of online Internet Electronic Program Guide (IEPG). The system was found to have good accuracy and dynamically adaptive capability.

1 Introduction

With the rapid development of digital television, the digital broadcasting provides a more efficient way to deliver television than analogue transmissions. A multiplex can then be broadcast via a channel. This means that on a channel which would have carried only one program using analogue technology, it is now possible to broadcast six to eight programs for digital broadcasting. With the increasing number of channels being offered, one problem of digital broadcasting is how to offer a convenient and intelligent user interface. It becomes troublesome to select our favorite TV programs by hand, even though we can use EPG navigation systems or other applications.

In related works, some learning theories were used to architect personal recommendation system [1,2,3,4]. In our previous work, an approach using Inductive Learning with N-gram to predict user's habits and preferences, showed good dynamically adaptive capability in small data sets [5].

In this paper, we propose an approach with term extraction method for constructing a personal recommendation system for EPG. The idea is based on a primitive approach of the term extraction method, according to statistical analysis among a compound noun and its components picked up from the history of each user's watched programs, and then to unified feedback process. The objective is to develop a good intelligent user interface between each TV fan and his/her digital television.

This paper includes three sections as follows: Presentation of our system architecture, evaluation of the performance of the present system and a summary of this work.

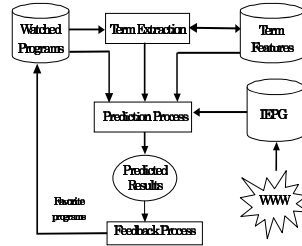


Fig. 1. Outline of the Procedure

2 Outline

The outline of our system consists of term extraction, prediction, feedback and IEPG extraction process, as shown in Figure 1.

Our idea of term extraction is to use bigrams. It is based on the algorithm of Nakagawa [6] et al., which considered both the score using simple noun bigrams as components of compound nouns to calculate score of compound nouns, and the nouns independent frequency to calculate the score of each extracted term. For more details see [6]. Here, we only focus on nouns (single-nouns and compound nouns) and unknown words to acquire the term features, .

The IEPG was extracted to XML format from Internet using XMLTV module [7]. IEPG data vector consists of start time, stop time, channel, title, description, Japanese category, English category and so on. In our system, we changed some contents of IEPG, such as time, day of week, channel names and so on.

In feedback process, we have taken into account two factors that cause the dynamics of personalization. The first step is to split the history of user watched programs by time to generate two terms vectors. For prediction of the user's favorite programs, a simple vector space modification model is used as a feedback method as $H' = H_{new} + \gamma \cdot H_{old}$, where, H' is the extracted terms vector for prediction, H_{old} means the extracted terms vector from user watched programs before two weeks ago. H_{new} means the extracted terms vector from the latest user watched programs of two weeks. γ is a coefficient. The second step is to create some tasks for each category according to user watched programs, elements of tasks are start time (including day of week), channel, title, Japanese category, English category and watched frequency. These tasks are extracted from the user watched programs and the frequency is acquired from user watched programs. We use these frequencies, to re-order the acquired categories for improving prediction precision.

The basic idea of prediction is to use extracted terms to evaluate each new TV program. The terms are extracted from watched programs according to the Concurrent Versions System (CVS) of Japanese TV programs. There are 14 categories in IEPG. Using these categories, we can acquire extracted terms of each category. In prediction, using the inner product method between the acquired

Table 1. The Data Size of Each Week

| Weeks | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---------------|----|----|----|----|----|----|----|----|----|----|----|----|
| Data Num. (A) | 30 | 28 | 30 | 27 | 30 | 28 | 28 | 30 | 26 | 28 | 27 | 28 |
| Data Num. (B) | 32 | 30 | 28 | 32 | 30 | 31 | 28 | 32 | 30 | 28 | 29 | 28 |
| Data Num. (C) | 28 | 32 | 30 | 30 | 28 | 32 | 30 | 32 | 32 | 30 | 28 | 28 |
| Data Num. (D) | 36 | 34 | 38 | 36 | 36 | 36 | 38 | 32 | 40 | 36 | 36 | 40 |

terms and the new programs, the credibility of each new program is calculated with the importance of each term. We use the logarithm of accumulation value of all matched terms as the credibility, and then, normalize these measures. We can give some threshold to control the number of recommendations.

3 Experimental Results

As mentioned above, the system based on our proposed approach was developed for experimentation to investigate its validity. In our experiments, we use open data to test the performance of our system. We adopted periodic training to our system. The training data is incremented on weekly basis.

Our experiment datasets were collected based on daily life of four graduate students of engineering over a period of about three months and the total number of data was 1,486. The data size of every week is shown in Table 1. The value of γ was given 0.5, the optimizing value γ will be investigated in future experiments. In order to keep the starting state constant for each user, the file of watched programs and the term features always started from an empty initial state.

We just use rank 10 of predicted results to evaluate precision per day because time used by everyone to watch TV is limited, and TV watchers will judge the prediction results. Figure 2 shows weekly precision of our proposed method.

As a comparison, we performed other experiments, without the feedback method. As reference work, other experiments were performed using a keyword input type TV recorder, named a TV channel server, which is produced by M corporation (Japan).

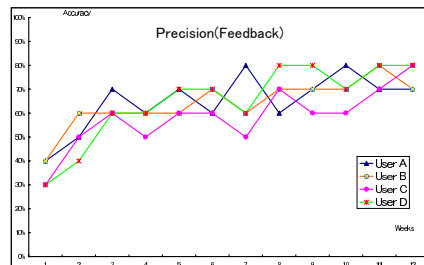


Fig. 2. Precision of Term Extraction with Feedback

In Figure 2 we plot the performance of our proposed method. It is indicated that the average precision is about 65.0% (User A), 64.2% (User B), 58.3% (User C) and 65.0% (User D). With the results of our comparison experiments, the accuracy was improved about 7.0% for both user A and B, 7.5% for user D, and improves by 3.3% for user C, than the method of without feedback.

The average precision of M corporation's TV channel server is about 46.1% (User A), 32.3% (User B), 44.9% (User C) and 40.0% (User D). Comparatively, our proposed method is more effective.

4 Conclusions and Future Work

In this paper, an approach for predicting users' favorite TV programs is described. The system has proven to have better performance than the conventional keyword input method. The system can provide simpler user interface with enhanced functions to TV watchers because they just need to select the recommendation programs rather than to consider any keywords for selection.

Our goal is to develop a personalized system of TV program recommendations with adaptive capability. We will try to test the performance and put our system to practical use for a computer or a digital television.

Experiments on a large scale including languages other than Japanese will be done in the near future to further verify the accuracy of the present system based on term extraction method. Other algorithms, such as reinforcement learning, SVM and neural networks will also be used for comparison.

References

1. Anna B., John Z., and Kaushal K, Improving Ease-of-Use, Trust and Accuracy of a TV show Recommender. *Proc. of 2nd International Conf. on Adaptive Hypermedia and Adaptive Web Based Systems: Workshop on Personalization in Future TV (AH)*, Universidad de Malaga, Spain, 2002, 1-10.
2. L. Ardissono, C. Gena, P. Torasso, F. Bellifemine, A. Chiarotto, A. Difino, B. Negro, Personalized Recommendation of TV Programs. *Lecture Notes in Artificial Intelligence. AI*IA 2003: Advances in Artificial Intelligence*, Pisa, Italy, 2003, 474-486.
3. P. Johansson, Natural language interaction in personalized EPGs. *Proc. of Workshop notes from the 3rd International Workshop on Personalization of Future TV*, Johnstown, Pennsylvania, USA, 2003, 27-31.
4. M. van Setten, M. Veenstra, A. Nijholt, Prediction strategies: Combining prediction techniques to optimize personalization. *Proc. of TV-02: 2nd Workshop on Personalisation in Future TV*, Location Malaga, Spain, 2002, 29-37.
5. J. Xu, T. Itoh and K. Araki, Action Prediction Method Using Recursive Different and Common Parts Extraction Method with N-gram. *Journal of Human Interface Society, Vol.7, No.1, Japan*, 2005, 55-68.
6. Nakagawa H., Mori T., A Simple but Powerful Automatic Term Extraction Method. *Proc. of 2nd International Workshop on Computational Terminology, COLING-2002 WORKSHOP*, Taipei, 2002, 29-35.
7. <http://membled.com/work/apps/xmltv/>.