A SVM-based Personal Recommendation System for TV Programs

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: +81-11-706-7389 Fax: +81-11-706-7389
{xja, araki}@media.eng.hokudai.ac.jp

Abstract

This paper presents a SVM-based prediction approach for constructing personal recommendation system for TV programs. We have applied Support Vector Machine (SVM) to personal prediction of online Internet Electronic Program Guide (IEPG). Our basic idea is to combine SVM and feedback processing into our system, using user-watched histories as retraining data, to realize personal predictions. We evaluate the precision by experiments with open data. The results show that the proposed polynomial kernel SVM system offers a statistically significant increase in performance compared to other method, and this system demonstrates good dynamically adaptive capability.

1   Introduction

Information and communications technology is an increasingly powerful tool participating in global markets for improving the delivery of basic services and enhancing local development opportunities. Especially, the rapid development of digital processing technology has resulted in more and more digital products in our life. The challenge, how to provide Artificial Intelligence to these digital products, for example, an intelligent digital television etc, has become a research point.

Today, a number of technical developments, such as satellite, cable and digital TV technology have resulted in an increasing number of available TV channels, hundreds of channels broadcast thousands of TV programs everyday. It is difficult and sometimes impossible to find what the user is looking for – the relevant information is often buried under thousands of TV programs, this problem is called information overload.

As we known, the Electronic Program Guide (EPG) is an especially important player on the digital television scene. It contains real-time information and future programs of the broadcasters. Using the EPG, viewer can request background and additional information about programs.

There have been several research projects around EPG recommendation system [1,2,3,4,5]. There are some problems for these works, such as the unsatisfied prediction accuracy, low recall and low dynamically adaptive capability for personal recommendations.

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 set [6].

In this paper, we propose a SVM-based method for constructing a personal recommendation system for EPG. The objective is to develop a good intelligent user interface for customers. This paper includes four sections as follows: Support Vector Machine, presentation of our system architecture, evaluation of the performance of the present system and a summary.

2   Support Vector Machine

SVM is based on statistical learning theory, which uses the principle of Structural Risk Minimization instead of Empirical Risk Minimization, as a supervised machine learning technique, developed by Vapnik [7]. The SVM technique finds a maximal margin separating hyperplane between two classes of data, and thereby creating the largest possible distance between the separating hyperplane and the examples on either side of it.

Two key elements in the implementation of SVM are the techniques of mathematical programming and kernel functions. By using kernel functions, we can construct a non-linear separating surface in the original feature space. Fortunately, such non-linear training does not increase the computational cost if the calculation of the kernel function is as cheap as the inner product. A polynomial function ($K(x,y) = (x \cdot y + 1)^d$) is popular in applications of SVMs.
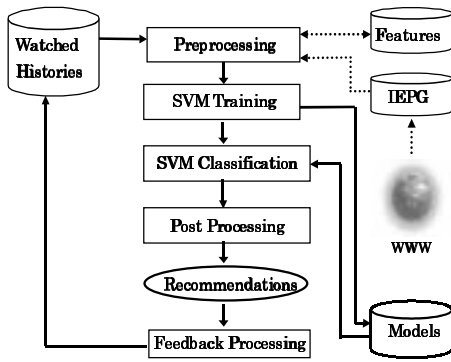
Figure 1. Outline of the Procedure

This paper discusses the performance of the linear and the polynomial kernel functions to construct classifiers for generating the EPG programs prediction models. These kernels provide good classification accuracy, are fast to learn and fast classifying new instance. We use polynomial kernel degree of 2 that have been very effective when applied to other tasks, such as natural language processing [8,9,10].

3 Outline

This section describes the procedure of our proposed method as shown in Figure 1, our system consists of Internet Electronic Program Guide (IEPG) extraction, preprocessing, SVM training, SVM classification, post processing and feedback processing.

3.1 IEPG Extraction

The internet is a globally distributed dynamic information repository that contains vast amount of digitized information, and more and more of such information is now available in multimedia forms.

In our system, TV programs are extracted to XMLTV [11] format as our IEPG database from internet using XMLTV module, from the URL of http://www.ontvjapan.com/program/, they are updated automatically using crontab command of Linux OS.

We use Japanese IEPG for our system. All of the extracted TV programs are classified automatically according to the Concurrent Versions System (CVS) of Japanese TV programs.

3.2 Features

One of the major problems of traditional SVM classifiers is the large number of features, which result in long processing time. To address this problem, we propose to represent each TV program by a limited number of the named entities and a part of extracted words as features. This reduces the dimensionality of the classifier and thus the number of training examples needed. The characteristics of program structure also can be incorporated into these features easily.

In this system, TV programs are semi-structured documents as XML, useful information often can be used to derive some important features of IEPG. Such characteristics of TV programs could possibly be applied to improve the performance of recommendations.

Features were extracted from collected IEPG. We extracted all of the named entities for TV programs, i.e. channel names, time information (day of week, start time, stop time), categories, etc. Moreover, we extracted the nouns and unknown words from the description of TV programs with ChaSen [12]. For all extracted nouns and unknown words, give a threshold according to the frequencies of lexicon words, ignore the lowest frequencies words to reduce dimensionality, then the features can be decided.

3.3 Preprocessing

The statistical approach aims to obtain an attribute-value representation of the textual content. This representation is usually called a bag of words (attributes) and describes a vector that contains attribute-weight pairs. The vector conception requires some tasks to be relied on, such as attributes identification, weight assigning and representation reduction [9].

There are two approaches to calculate the weight of the features:

Boolean: 0 and 1 values are used to represent, respectively, absence or presence of a term (attribute) on textual content;

Numeric: represents statistical metrics based on term frequency (tf) into the document. An approach [13], called $tf \cdot idf$, suggests the inclusion of the term frequency on all documents as a reverse measure of its capacity to represent specifically a document.

We are interested in which is the most efficient representation for the program prediction performance and the learning performance, boolean representation or $tf \cdot idf$ representation. Aiming to provide the correct programs representation to use on the SVM method, the preprocessing phase was divided on the following tasks.

(1). Remove XML tags.

(2). Extract all of the named entities and the words of TV programs with ChaSen.

(3). Calculate the weight of the features according to the *Boolean* approach or the $tf \cdot idf$ method.

(4). Label +1 to the watched histories for the positive class, label -1 to the rest of the old IEPG data for the negative class and label 0 to new programs for prediction.

3.4 Other Processings

In SVM training, the positive data and the negative data are used to generate the prediction models. Then, the new programs are being classified with generated models. In post processing, The system changes the classified results into XML format of original IEPG, sort these results according to the classified scores, day of week etc, and then, output these results to the user. In Feedback process, user judges the recommendations whether he/she likes the programs or not. These judgments are registered into the watched histories, and then they are used for the re-training as training data. This processing assures the dynamic adaptive capability for personal recommendations.

4 Experiments

As mentioned above, the systems based on our proposed approaches were developed for experimentation to investigate their validities.

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 one weekly basis.

Our experiment datasets were collected based on the actual life of each person, from five graduate students of engineering. This faculty generated total 2974 watched histories as our datasets that corresponded to a period of three months for each person.The data size of every week is shown in table 1. These data sets were labelled +1 for positive class, as the favorite programs of each user, the rest data of each week were labelled -1 for negative class, as the negative data for training SVM, the total nmber is 37701.

In order to keep the starting state constant for each user, the watched history and models always started from an empty initial state of the SVM-based system.

We use open data to evaluate prediction performance, the correct recommendations are judged from rank 60, and we use the collected history data of next week after the latest training data as the specified correct results. Precision (Recall) is defined as:

$$Precision = \frac{number\ of\ true\ correct\ results\ in\ rank\ 60}{number\ of\ specified\ correct\ results} \quad (1)$$

Figure 2 shows weekly precision, in case of using linear function with $Boolean$ feature vector (LB). Figure 3 shows weekly precision, in case of using linear function with $Numeric$ of $tf \cdot idf$ feature vector (LN). Figure 4 shows weekly precision, in case of using polynomial kernel function with $Boolean$ feature vector (PB). Figure 5 shows weekly precision, in case of using polynomial kernel function with $Numeric$ of $tf \cdot idf$ feature vector (PN).
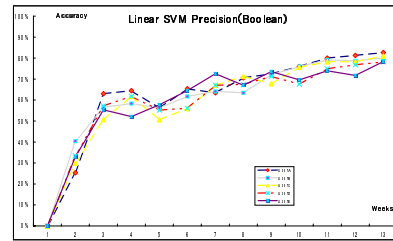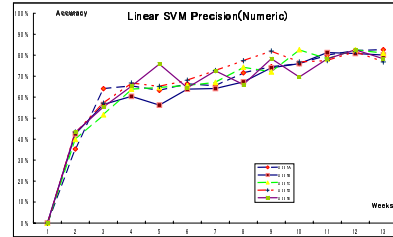

Figure 2. Precision of LB

Figure 3. Precision of LN

5 Results and Considerations

We have measured the performance of our proposed methods with linear SVM and polynomial function SVM. We have also compared test results to another baseline system, which also applied the personal recommendations on the same data sets. All these data sets are extracted from internet with XML format.

According to these results, the average prediction performances of the polynomial kernel degree 2 are significantly better than other works. Using Polynomial function, the total average precision is 83.4% and 85.3% for the boolean feature vector and the numeric feature vector. From these experiment results, we can find the highest accuracies are all about 92.5% at the end of these experiments.

The linear SVM achieved the total average prediction rates of 63.6% and 68.8% for the boolean feature vector and the numeric feature vector. The total prediction accuracy of our baseline is about 65.0%.

In all these experiments, it is indicated that the precisions increase with increasing training data. This means the feedback processing is needed for improving an adaptive system.

In our experiments of SVM-based system, we just considered a two-class case classifier to predict user favorite programs. We think that in other cases, there is a need to consider the multi-class classification, such as the One-versus-Rest, this will be resolve in our future works.

Table 1. The collected Data Size of Each Week (Positive Data for SVM-based System)

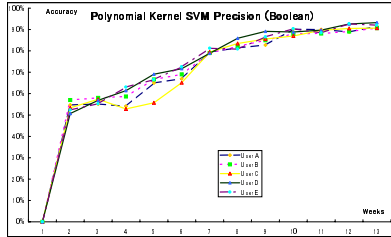| Weeks	1	2	3	4	5	6	7	8	9	10	11	12
UserA	48	47	46	48	52	52	51	51	50	45	48	48
UserB	52	50	48	52	50	51	48	52	50	48	49	48
UserC	48	52	50	50	48	52	50	52	52	50	48	48
UserD	48	54	48	56	46	48	48	52	50	48	46	50
UserE	50	52	48	52	48	48	50	50	50	48	48	52

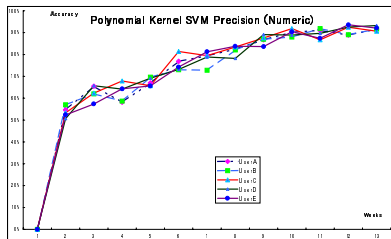

Figure 4. Precision of PB

Figure 5. Precision of PN

6 Conclusions and Future Works

This paper introduced a SVM-based approach for constructing a personal recommendation system for TV programs. According to the experiment results, our system expressed good accuracy. The feedback processing assured the high dynamically adaptive capability.

We will do some other experiments, in case of considering syntax analysis for features to improve performance. 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, neural network, will also be used for comparison.

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 computer or digital television.

References

[1] B. Anna,Z. John and K. Kaushal 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*, pages 1–10, 2002.

[2] L. Ardissono, F. Portis, P. Torasso. F. Bellifemine, A Chiarotto and A. Difino. Architec-ture of a system for the generation of personalized Electronic Program Guides. *Proc. UM 2001 Workshop On Personalization in Future TV, Sonthofen, Germany*, 2001.

[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*, pages 27–31, 2001.

[4] M. van Setten, M. Veenstra and A. Nijholt. Prediction strategies: Combining prediction techniques to optimize personalization. *Proc. of TV-02: 2nd Workshop on Personalisation in Future TV, Location Malaga, Spain*, pages 29–37, 2002.

[5] M. van Setten, M. Veenstra, A. Nijholt and B. van DijkM. Case-Based Reasoning as a Prediction Strategy for Hybrid Recommender Systems. *Proc. of Conf. Second International Atlantic Web Intelligence Conference (AWIC 2004), Cancun, Mexico*, pages 13–21, 2004.

[6] 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*, pages 55–68, 2005.

[7] Vapnik. The Nature of Statistical Learning Theory. *Springer-Verlag, New York*, 1995.

[8] C. Aone, M. Okurowski and J. Gorlinsky. Trainable Scalable Summarization Using Ro-bust NLP and Machine Learning. *Proc. of the17th COLING and 36th ACL*, pages 62–66,1998.

[9] T. Joachims. Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *Proc. of ECML*, pages 137–142,1998.

[10] M. Skowron and K. Araki. Effectiveness of Combined Features for Machine Learning Based Question Classification. *Special Issue of the Journal of the NLPS Japan on Question Answering and Automatic Summarization, Vol.12, (in printing)*, 2005.

[11] http://membled.com/work/apps/xmltv/

[12] Y. Matumoto. http://chasen.naist.jp/hiki/ChaSen/.

[13] G. Salton, C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management, 24*, pages 513–523,1988.