Recommendation System for HBB TV: Model Design and Implementation

Alexandra Posoldova, Alan Wee-Chung Liew
School of Information and Communication Technology
Griffith University, Gold Coast campus
QLD 4222, Australia
alexandra.posoldova@griffithuni.edu.au
a.liew@griffith.edu.au

Abstract— Hybrid broadcast and broadband (HBB) television is a new platform that opens many possibilities for new services. Recommendation system offers a personalized service that suggests items of interest according to user preference. Nowadays, the number of available programs is so large that one cannot realistically have a real time overview. Recommendation engines were developed to solve the problem of information overload, and save time and effort when looking for appealing content. In this paper, we present model design and implementation of a recommendation system for HBB TV. To explore user preferences and make predictions, an enhanced Naïve Bayes model for rating prediction is designed. The model uses a set of features to predict user rating based on past observation. The recommendation system presented in this paper is flexible and robust enough to handle a sparse data set with very few records of feature description. Experiments performed on a Yahoo movie data set indicated the promising performance of our approach.

Keywords— Recommendation systems, Bayesian network, content-based filtering, Yahoo movie data set, HBB TV, intelligent systems

I. INTRODUCTION

Hybrid broadcast and broadband (HBB) television combines traditional TV broadcast and internet transmission. The catch-up TV is a typical well-known service offered by this platform. This is, however, only a sneak peak of the possibilities this environment brings. Every user can have a profile with personal settings. Not only the broadcast content, but all content can be personalized in order to enhance user experience.

With an increasing amount of information available, there is a need to somehow prioritize and sort it in order to get fast access to relevant content. Recommendation systems solve the issue of information overload and target users’ needs. They have become very popular in recent years and have been designed for various applications. Two main approaches, content-based and collaborative filtering, have been widely implemented in recommendation engines. With the increasing popularity of social networks, community-based filtering has also increasingly been applied in recommendation systems. For this research, we adopted the content based approach because of the readily available data set.

In this paper, we present a content-based recommendation system using content information about an item to explore user preferences. The issue typical for content-based filtering is the sparse nature of data. An item can be described by a number of features. However, not all of them are relevant or they can be missing for some entries. Also, as the number of features describing an item increases, the performance of a model can sometime decrease. Accordingly, only significant features shall be considered in order to get the best possible results.

The data set in the case of content-based filtering often comprises categorical data. Therefore, numerical approaches cannot be applied straightforwardly. One option is to transfer categorical information into a vector space using some method of assignment. Another approach is to use probabilistic representation like graphical models. Graphical models can provide us with insight into relationships between features used to describe content. In this work, we use Bayes network to express feature dependences and compute recommendation. This paper presents an enhanced Naïve Bayes content-based recommendation system able to predict user rating and make recommendation accordingly. To deal with sparse data, our algorithm incorporates nonlinear weighting and missing data replacement to minimize the impact of singularities and a sparse data set.

The paper is organized as follows. First we introduce work related to our research. The next section is dedicated to implementation and description of recommendation engine application in HBB environment. Then we present our recommendation model based on Bayes principle. This will be followed by data set description and experiments. We performed two tests and compared performance of our model with an existing approach. In the first test, the model predicts user rating based on past observation. In the second test, the content with the highest rating is recommended. In both cases we used precision to evaluate performance of our approach and an existing method. Finally, we provide a summary in the last section.
II. RELATED WORK

A few recommendation systems have been designed for digital television. Collaborative context tagging-based prediction of user preferences is described in [1]. However, we would like to use content information about program already available by broadcast service provider via electronic program guide. The idea to personalize program offer can be found in [2], where user receives a daily Internet-based personalized TV listing.

The actual implementation of recommendation system for digital TV is shown in [3]. This paper presents end-to-end personalized service offering A/V content based on MPEG-7 standard.

Content-based recommenders based on graphical models have been proposed in [4]. Some designs are context aware [5, 6]. In the case of HBB television, contexts such as transmission condition, device specification, and time aspect should be considered to choose the feature set which is the most appropriate as designed in [7]. Then the best match with program offer will be found.

Bayes network model similar to our approach can be seen in [8], where canonical weighted sum (CSW) is used to compute relationships between items, features and users. Canonical weights are generally considered as unstable from one sample set to another [9]. Moreover, this approach only performs well if we have more than 10 entries for every feature. This criterion is not always satisfied for the data set we use in our experiments. Although the model proposed in this work can incorporate both content and collaborative filtering, here we only considered the content-based part and used features as evidence to predict rating.

Approaches that evaluate user log data and EPG (Electronic Program Guide) can be found in [10]. The authors computed recommendation value for key words and celebrities as sum over ratios of the number of times the feature appears in a viewer watched history to overall occurrence in the data set. For the recommendation, they evaluated 100 programs and chose one. Then the model’s performance was measured using precision as the number of correct recommendations to overall number of watched movies for the given user.

The method looks very simple. However, the application requires lots of computation as the algorithm has to search through hundreds of programs plus a test set for every feature in order to find the number of its occurrences. This makes their method high in computation cost and slow.

III. HYBRID BROADBAND AND BROADCAST TELEVISION

Hybrid broadband and broadcast television is a new platform combining two types of transmission. This new way of watching TV brings new challenges as well as opportunities for service development. It allows users to have personal profiles and to adjust TV settings to their convenience. A user can be identified by camera [11] or microphone [12]. Furthermore, a camera can be used to perform gesture recognition and personal gestures can be also created [13]. Therefore, HBB television offers a truly personalized approach. The recommendation system is another personalized service enhancing user experience in HBB TV.

The block diagram of the recommendation engine is shown in Figure 1. It comprises 5 basic parts. The profile learner performs training and re-training of the model. Once the recommendation system is trained, personal preferences are stored in a user profile. As HBB television combines two different transmissions, data collection and unification needs to be considered. This is, however, beyond the scope of this paper. If a user is willing to provide explicit feedback, such as like/dislike, rating or text comments, immediate response of a viewer to the recommended content can be obtained and the model can be re-trained and adjusted to his/her needs [14]. If giving feedback bothers a user, it can be done implicitly. For example, a user’s actions or emotions can be used to estimate viewer watching experience [15]. Finally, the collected data and user profile are used to make recommendations. Our research focuses on exploring viewer preferences in the training phase and making recommendations accordingly.

![Fig. 1. Recommendation engine as block diagram designed for HBB television.](image-url)
A. Recommendation System as HBB TV Application

The recommendation engine belongs to broadcast-related application group as it is associated with broadcast service and events. The application starts automatically as the TV terminal is turned on. According to the HBB standard, the “red button” icon indicates that the application is available to the user. Considering a standard remote control, when the red button is selected, the recommendation system application displays the full user interface. Unless the user sets it otherwise, the autostart broadcast related application will not display full screen automatically. In general, there are three states of this type of applications:

1. the “Red Button” notification is displayed to notify the user that the application is available,
2. no user interface display, the application is running on background,
3. full user interface display with list of recommended items.

The red button is used to switch between stages. If the remote control is equipped with the EXIT button, this can be used to terminate the application. Channel change event can also be used to start the application or to notify the user about availability of the recommendation service. All these incidents are recorded in the AIT - Application Information Table. Note that the recommendation system can also suggest items transmitted via broadband.

The service can be implemented in the terminal by manufacturer or can be downloaded via broadband and access to data is via broadband or broadcast transmission. In this case, the broadcast transmission consists of DVB - Digital Video Broadcasting network providing the hybrid terminal with broadcast content, application data, and signaling information. The application data and stream events are transferred by DSM-CC – Digital Storage Media – Command and Control carousel and provided to Runtime Environment where the application is executed. This environment has access to channel list information and the EIT - Event Information Table, both important for the recommendation system. The functional components of a hybrid terminal are illustrated in Figure 2. [16].

![Functional components of a hybrid terminal.](image)

The recommendation engine uses a user’s watching history to train the model. The signaling is used to capture these past events. Standard TS 102 809 [17], formerly designed for Digital Video Broadcasting – DVB, defines the application signaling and its transport via broadcast of HTTP.

B. Terminal Capabilities Used by Recommendation System

User preferences are observed according to his/her past behavior. In other words, based on a content watched and liked or disliked by a user, the recommendation system can extract the viewer watching pattern. In order to record the history of watched content, the system needs to track user actions. User actions can be captured as key events and can be requested by the application. The list of key events available for application, as defined in [14], and relevant remote control buttons is shown in Table 1.

<table>
<thead>
<tr>
<th>Remote control button</th>
<th>key event</th>
</tr>
</thead>
</table>

Fig. 2. Functional components of a hybrid terminal.

The recommendation engine uses a user’s watching history to train the model. The signaling is used to capture these past events. Standard TS 102 809 [17], formerly designed for Digital Video Broadcasting – DVB, defines the application signaling and its transport via broadcast of HTTP.

B. Terminal Capabilities Used by Recommendation System

User preferences are observed according to his/her past behavior. In other words, based on a content watched and liked or disliked by a user, the recommendation system can extract the viewer watching pattern. In order to record the history of watched content, the system needs to track user actions. User actions can be captured as key events and can be requested by the application. The list of key events available for application, as defined in [14], and relevant remote control buttons is shown in Table 1.

<table>
<thead>
<tr>
<th>Remote control button</th>
<th>key event</th>
</tr>
</thead>
</table>

TABLE I. KEY EVENTS AND REMOTE CONTROL BUTTON

![Diagram showing functional components of a hybrid terminal.](image)
The buttons in Table 1. are compulsory for every HBB remote control. VK_* is preposition for key codes specific for KeyEvent interface and cannot be used in general code.

According to purpose of these buttons, we can divide them in three categories. The ENTER, OK, BACK and color buttons can be used for service navigation. Red button opens the application on full screen or puts it on background, as it was specified in the previous section. Function of other colors is not specified by the HBB standard. The numbers, up and down buttons are used to switch channels and build the watching history. The rest of the buttons can be used to estimate an implicit feedback.

Having the user watching history recorded, we can proceed to building the recommendation system.

IV. CONTENT BASED RECOMMENDATION ENGINE

Content based recommendation explores similarities between the features describing content in relation to viewer preferences. Then, based on content from past observation, the recommendation is made. As the content describing our data set is categorical, numerical approaches cannot be applied directly. Therefore, we chose, Bayes networks allowing probabilistic representation of data and providing clear insight into the relationship between content features [18].

An important part of a good recommendation system is its ability to build trust. In other words, if a user likes the item suggested by a system, the more likely the user is to use the system, and the better it can be customized and trained. Content-based filtering provides a transparent explanation to the user of how the recommendation works and so increases trust in the system [19].

V. ENHANCED NAIVE BAYES FOR RATING PREDICTION

Every program/item can be described by a set of features, such as actor, director, genre, producer, and other, where every feature can have a number of states. For example, for genre, the possible states are crime, drama, comedy, and others. For some feature classes, like actor, more than one feature can describe an item. As we said earlier, the data are categorical. Consequently, every feature class can be described by a multinomial distribution [20]. Suppose we have feature class F with S possible states \( F = \{ F_1, F_2, ..., F_s \} \). Then for M trials, represented by M items \( I = \{ I_1, I_2, ..., I_m \} \), the example of one feature class translated in 1ofN coding scheme can be expressed as following matrix (1):

\[
\begin{bmatrix}
F_1 & F_2 & F_3 & \cdots & F_s \\
I_1 & 0 & 0 & 1 & 0 \\
I_2 & 1 & 0 & 0 & 1 \\
I_m & 0 & 0 & 0 & 1
\end{bmatrix}
\]

Suppose we have a history of watched programs in the past with their content description and assigned user rating. This data is observable and is represented by grey nodes \( \{ f_1, ..., f_n \} \) in our graphical structure as shown in Figure 3.
The conditional probability that a user $u$ will give a particular rating $r_u^n$ to a program $p$ described by a set of features $f_1, f_2, \ldots, f_n$ [21] is (2):

$$p(r_u^n | f_1, \ldots, f_n) = \frac{p(r_u^n) \prod p(f_i | r_u^n)}{p(r_u) \prod p(f_i | r_u^n) + p(-r_u^n) \prod p(f_i | -r_u^n)}$$

(2)

As some combinations of rating for given feature appear only once in the user data set, it results in conditional probability equals to 1. This singularity, however, does not reflect the true user preference and has to be adjusted. On the other hand, if some feature appears in a number of entries in the user data set, it reflects his/her preference. Accordingly, we implemented a nonlinear weighting, using sine function, to give higher probability to observations with higher number of occurrences and penalize singularities (3).

$$p(f_m | r) = \frac{\sum (f_m, r)}{\sum f_m} \cdot \sin\left(\frac{\sum (f_m, r) \pi}{M/2}\right)$$

(3)

where $M = \{1, \ldots, m\}$ is the total number of observations in the training set. Note that we could not apply any resampling techniques to smooth out the distribution as our data is mostly categorical and numerical approaches cannot be used. This nonlinear weighting will smooth the distribution as desired and works also for small sample data.

Now, to make predictions of rating for a new content, we compute the probability of rating $r$ for a set of features $F = \{f_1, f_2, \ldots, f_n\}$. We can create the matrix $R_u$ of all possible ratings in the range $\{r_1, \ldots, r_m\}$ for a given combination of content features (4):

$$R_u = \begin{bmatrix} p(r_1 | f_1, \ldots, f_n) & \ldots & p(r_1 | f_1, \ldots, f_n) \\ \vdots & \ddots & \vdots \\ p(r_m | f_1, \ldots, f_n) & \ldots & p(r_m | f_1, \ldots, f_n) \end{bmatrix}$$

(4)

where $f_n^p$ is the feature of program $p$ from the feature class $f_n$.

Then we apply the maximum a posteriori estimation (MAP) method to choose the rating with the highest conditional probability for a given set of features as (5):

$$MAP[p(r_m | f_1^p, \ldots, f_n^p)] = \arg \max_{r_m} \{p(r_m | f_1^p, \ldots, f_n^p)\}$$

(5)

VI. DATA SET

We tested the model for rating prediction using the Movies user ratings of movies, and movie descriptive content information data set version 1.0 collected by the Yahoo Labs [22]. It contains all 11,915 movies/items with detail descriptions. The training set consisting of 211,231 ratings from 7,642 users, where the average number of ratings per user is 27.64. The user rating acquires values from 1 to 5. The average number of ratings per user is 4.39 and the average number of ratings per item is 4.26. All users have rated at least one item and all items have been rated by at least one user. This data set is complex and sparse. Moreover, some viewers have only a small number of observations in the training set and a small number of records in the testing set. As there is no time aspect in the data, we decided to combine training and test set and divided this combined set into three equal parts, where 2/3 of data represents training examples and the rest test examples.

We also reduced the number of features considered by our model from 34 to 18 as some features are duplicated, for example actors are defined by their names and ids. As we use a probabilistic representation, duplicated features will appear twice in our...
computation with the same probability value. This would not only increase the computational cost, but also affects the prediction results.

We use MATLAB R2013b to program the model. The missing information about a feature for some item is given by ‘‘N’’ in the Yahoo data set. Some features are completely empty for some users. We use 1 of N coding to create feature matrices with logical 1 if the feature describe some item and logical 0 if not.

VII. EXPERIMENTS

We evaluated our approach for rating estimation designed as a Naïve Bayes model, as illustrated in Figure 1. In the training phase, we computed joint and conditional probability distribution for every combination of user rating and feature class.

In the case of empty feature matrix, as we have no information about this feature, it should not contribute to the overall result. Therefore, we excluded this from further computation by replacing its conditional probability by a matrix of ones. If the feature matrix has just some elements empty, we cannot do the same as it still carries some information. In order to avoid multiplication by 0, we decided to replace the element with a small constant $c=0.001$ and normalize the rest of the matrix to keep the marginals unchanged. We computed the precision of our model as the number of correctly selected ratings $C$ to the number of all observations $K$ (6).

$$ P = \frac{C}{K} \quad (6) $$

We compared our results for rating prediction with results from a randomized training set and the canonical weighted sum approach presented in [10]. The system then made predictions according to the estimated rating. Items with the highest predicted rating were considered for recommendation.

A. Rating estimation experiment

We tested our Naïve Bayes recommendation model on 223 users from the Yahoo movie data base. The average number of entries in the training set was 22.8. The training set with the fewest observations comprised of 8 items and the biggest one had 852 observations. We estimated the user rating as the conditional probability that user will rate the program from 1 - 5 for a given combination of features. We studied how the constant that was used to replace missing values affects the prediction, so we tried $c=0.01$ and 0.001. In Table 1, we can see that the model performed slightly better with $c=0.001$, but it had no real significant impact.

<table>
<thead>
<tr>
<th>TABLE II. PRECISION OF RECOMMENDATION SYSTEM PREDICTING RATING OF USERS TESTED ON YAHOO MOVIE DATA SET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean P for c=0.001</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>0.5568</td>
</tr>
</tbody>
</table>

To justify the model’s performance and confirm that the rating prediction is better than chance, we randomized the training set of each user where we randomly picked items from the Yahoo movie data base and assigned a random rating from 1 - 5. Then, we performed the test using the original test set for every user. The results are presented in Table 2 and 3.

Based on the experiments we can say that the model with the original training set performs better by almost 40% compared to the model from the baseline random set. Because of that, we can say that our enhanced Naïve Bayes model can identify user preferences and predict the rating for new items. Compared with the CSW approach, although there was only a small improvement in mean prediction accuracy, the standard deviation of our model is much smaller than CSW. Thus, our method is highly competitive when compared with more sophisticated approaches such as canonical weighted sum.

| TABLE III. MEAN COMPARISON OF NAÏVE BAYES, RANDOMIZED SET, AND CSW APPROACH |
|-----------------|-----------------|----------------|
| Mean P for c=0.001 | Random mean P for c=0.01 | CSW mean P |
| 0.5568           | 0.1437           | 0.5507        |

TABLE IV. STANDARD DEVIATION COMPARISON OF NAÏVE BAYES, RANDOMIZED SET, AND CSW APPROACH
From a computational cost point of view, these two approaches are comparable. Once we compute the joint and conditional probabilities for the enhanced Naïve Bayes, the rating estimation is just the product of conditional probabilities. Similarly, for CWS, once the weights are computed, sum of weights and conditional probabilities give the rating prediction.

B. Item recommendation experiment

We computed the conditional probability of rating for a given feature set. The probability of assigning a rating from 1-5 to an item from the user test set is known. To simulate program offering, we randomly sampled from the Yahoo data set and estimated the rating for every item. The current program offering is set to 100 as in paper [10]. In the recommendation phase, the item with the highest rating is selected from the program offering. In case when more than one item was estimated to have the highest rating, the user can choose from multiple choices. The recommendation is accepted as correct if the item from the test set appears among the list of recommended items.

We compared our approach with the method that computes the recommendation value, RV, presented in paper [10]. They evaluated their model on 10 households from a set of 923. In order to make the comparison under the same settings, the same numbers of program offerings were used, which consisted of 99 randomly chosen items and one item from the test set. The number of times the recommendation was made was equal to the number of items in the test set.

We kept the program from [10] running for more than 21 hours to compute RV and the program returned results for 10 users. For our model, including the training phase, it took less than 10 minutes to compute recommendations for the same number of users. As expected, the method that computes RV is very slow. The results are shown in Table 4.

To allow our model to present multiple choices, the items within the top 5% of recommendation value are included in the recommendations offer. In both models, we consider the recommendation as correct if the item from the test set is included in the recommendation offer. As in the rating prediction experiment, the precision was used to evaluate the model’s accuracy.

<table>
<thead>
<tr>
<th>Enhanced Naïve Bayes</th>
<th>RV [8]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean P</td>
<td>0.7598</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>0.0713</td>
</tr>
</tbody>
</table>

From our experiments, our model is able to make recommendations with almost 76% precision. It clearly outperformed the RV approach by 16%. Moreover, if we compare the computational cost, the enhanced Naïve Bayes model is about 18 times faster (even with training phase included) than RV. Because of this, we consider our model more suitable for real time program recommendation than RV.

VIII. CONCLUSION

This paper has introduced implementation of recommendation system in HBB television according to specifications [14,15] considering signalization and user input needed to build viewer watching history. The recommendation engine designed for this purposes is represented by the enhanced Naïve Bayes approach. As we shown, this model is able to handle sparse data with singular occurrences. To compute dependences between nodes, the Bayes conditional probability for multiple variables was used. We applied nonlinear weighting to penalize singularities, and replaced missing data with a small constant to avoid multiplication by zero. Our model is able to discover user preferences and make predictions significantly better than chance. We further compared our approach to CSW. Based on our experiments, we showed that our model can compete with state-of-the-art content-based approach such as CSW. To simulate the recommendation process, we created a program offering from 99 items randomly selected plus one item from a user test set. We allowed the model to have multiple choices and compared our model to the RV approach of [8]. We can conclude that not only does our approach outperform RV, but it also works much faster.

REFERENCES


ETSI TS 102 809 (V1.1.1), “Digital Vioeo Broadcasting (DVB); Signalling and carriage of interactive application and services in Hybrid Broadcast/Broadband environments,” 2010.


