Application of recommendation techniques for Brazilian Portable Interactive Digital TV

Elaine Cecília Gatto and Sergio Donizetti Zorzo
Computer Science Departament
Federal University of São Carlos – UFSCar
São Carlos, São Paulo, Brazil
{elaine_gatto, zorzo}@dc.ufscar.br

Abstract—Before the possibility of new services and programs, and consequently more content available provided by Brazilian digital TV system, the users face difficulties to select their favorite programs. The Recommendation Systems become a tool to solve these difficulties besides improving the interactivity between the user and the digital television by filtering information and personalizing the content offer. Surveys herein present two personalization techniques, one using vector model and the other using data mining. The aim was to observe and evaluate how these data filtering techniques can be used and implemented in a recommendation system inside Digital TV context in cell phones.

Keywords- Personalization, Recommender Systems, Middleware Ginga, Cosine, Apriori.

I. INTRODUCTION

Digital television has created new services, products, contents, channels and business models. The Brazilian Digital TV System [1,2] allows fixed and portable reception, high quality of audio and video, as well as interactivity, creating different contents for users of fixed or portable IDTV (interactive digital television).

The EPG [3, 4, 5] – offered by the broadcasting stations – is a useful tool for users for it helps choosing a TV program which they want to watch. However, the increase in EPG content is necessary with the inclusion of new channels and so, due to the amount of information, the user finds difficulties and wastes time choosing programs. Therefore, recommendation systems can contribute for a better use of Digital TV whether in residences, in groups or individually, in a cell phone, for example.

This paper is divided as follows: Section I contains the introduction, Section II compares IDTV in residences and IDTV for cell phones, section III comments some related works, Section IV briefly presents our recommendation system, Section V explains methodology and tests, Section VI presents the analysis and results and finally, section VII presents the conclusion.

II. COMPARING IDTV IN RESIDENCES AND IDTV FOR CELL PHONES

The use of IDTV for cell phones will quickly boom due to the increasingly quantity of these devices surpassing television sets in Brazil, when cell phones with IDTV are available to population. Thus, some differences between IDTV for residences and for cell phones can be noticed.

IDTV standard adopted in Brazil calls full-seg the fixed devices like set-top-box, and one-seg, devices like cell phones, miniTVs, PDAs, etc. In residences, the IDTV is used by all residents while in the cell phone it is normally used by only one user, the owner of the device.

Another characteristic is the size of the display. In residences, the IDTV television sets have screens bigger than 30", where is possible to have a more flexible development, presentation and displaying of the content. However, cell phones screens are smaller than 10" requiring a higher effort in development to display the content on the screen avoiding image pollution and confusion to the user.

An exceptional characteristic in this environment is that IDTV for cell phones can be seen anywhere and anytime. On the other hand, IDTV viewing period in residences can be longer than in cell phones which are used in situations of waiting and displacement.

IDTV in cell phones can use already existent 2G/3G net architecture, and 4G in the future, as a return channel, making interactivity possible in this environment before occurring in IDTV.

The middleware adopted in Brazil has national technology and is called Ginga. Ginga-NCL and Ginga-J declarative and imperative portions of the middleware are necessary for full-seg devices. For one-seg devices, only Ginga-NCL declarative portion is required. There is a reference implementation of the middleware for full-seg devices. For one-seg devices, this reference implementation is not available yet, but are working in this middleware development, as PUC-RIO and UFES (Symbian e Android). [6]

The users of these devices need special attention due to current characteristics of this environment like processing power, storage capacity and battery.

III. RELATED WORK

There are many works involving recommendation systems for IDTV for set-top-box and more recently for portable devices. This section presents three recent works about recommendation systems for IDTV.

In [7] the recommendation system fits the systems with content-based filtering category, using text mining. The system uses a simple interface with the user and accepts a natural language as text entry as well as four values which reflect user preferences for comedy, action, horror and erotic. First, the system extracts texts and then searches for emotions in the text and the distances among themes are calculated. Finally, an index is calculated for each entry and a list of programs organized by this index is returned.

In [8] the main aim of the system is substituting the common content by a personalized and adapted content in a more attractive way for the user. Therefore, this system accepts and allows TV reception either through broadcast, or multimedia streaming. The system uses explicit collection – when using for the first time it is necessary to inform the preferences – and also implicit collection – user's actions in the device are monitored, stored and sent to the server. The personalized content– chosen based on preferences – is sent to the user's portable device by the Server in order to be previously stored before being exhibited.

The ZapTV [9] developed for DVB-H standard allows the user to create his own content, offering aggregated value services as multimodal access (Web and Cell phones), return channel, video note, personalized sharing and distribution of content. Besides the technology provided by DVB-H, ZapTV comprehends other technologies as TV-Anytime, Technologies emerging from Web 2.0 and involved in the Semantic Web. The main functionalities of ZapTV include a social net, personalized content broadcasting (implicit or explicit recommendation), thematic channels diffusion planning (age-group, genre or specific theme), client application and transmission of the electronic programming guide. ZapTV seeks to improve the recommendation using an intelligent personalization mechanism which matches information filtering with semantic logic processes and it was based on the principles of participation and sharing between Web 2.0 users, so that the creation, sharing, classification and note of content make the search for content easier.

Our recommendation system is in the portable device, and the inclusion of servers in Brazilian IDTV architecture for cell phones is not necessary to provide recommendation and, consequently the need for remote communication is also not necessary, avoiding that the user pay for the data traffic in the net in order to receive recommendation or send his data, and thus, protecting the user's data privacy.

IV. SYSTEM RECOMMENDATION

Our recommendation system aims to facilitate the IDTV user's routine by interacting with a simple interface which provides content of preference without spending so much time to find it.

The process starts when the user turns on the TV in the cell phone. The user history data collected is submitted to information filtering based on content in order to find the user's profile. Data resulting from this process are formatted. The user profile is stored in a database with the date and time of the generation. With the user profile updated it is possible to look in the EPG for compatible TV programs and which are going to be transmitted around the current time, providing a list of these programs. This list is also stored in a data base with the date and time of generation.

The recommendations are presented to the user and those required are stored with the user history. During the time the IDTV for cell phone is turned on, all programs viewed by the user are stored in the database which has the user history. This process is repeated every time the user turns on the TV. Ginga-NCL middleware has a layer for resident applications responsible for exhibition, other layer for the common center responsible for offering several services and a last layer regarding the protocols stack.

The recommendation system is considered as an element in Ginga-NCL architecture, in Ginga Common Core, due to the need to continue using data locally and also use Tunner libraries – in order to obtain information about the channels tune – ESInformation – in order to obtain information about EIT table generating the EPG – and Context Manager – to obtain system information.

As GINGA-NCL middleware is mandatory in Brazil for portable devices, the recommendation system was planned, designed and modeled according to Brazilian rules which refer to portable devices, thus meeting these devices needs. More details on our system can be obtained in [10].

V. METHODOLOGY AND TESTS

In order to test the system, particularly the personalization algorithm, it is necessary to obtain user's viewing data and also the EPG. This data was provided by IBOPE [11] and undergone a process of treatment almost totally manual in order to be adjusted to the tests. Data refers to a 15-day period of monitoring in 6 houses which have open TV.

Many technologies have been arising aiming to identify behavior patterns and use them in personalization. The recommendation systems operation is found on these techniques and the most used are the Collaborative Filtering and Content-Based Filtering which includes several algorithms for each one. A recommendation system can use only one technique or two together, becoming a hybrid system.

In order to study, analyze and choose an algorithm to be used in our system, two data filtering algorithms were tested.

The tests were performed in three steps. In the first step, tests were performed with Apriori [12] algorithm from Weka tool. In the second step, the forecast method was used, applying Cosine as measure of similarity. The third step was to compare the results and the operation with both algorithms, analyzing the facilities and difficulties, especially for the implementation.

The association techniques algorithms identify associations between the data registers which are related in some way. The basic premise finds elements which imply the presence of others in a same operation aiming at determining which are related. The association rules interconnect objects trying to show characteristics and tendencies. The association discoveries present trivial and non trivial association.

The data was adapted in order to be used in Apriori algorithm of Weka tool, that is, it was submitted to a pre-processing phase. CSV files correspondent to the user history were created from IBOPE data. For the implementation, it is not necessary that the data go through adjustments, as it will be collected in the correct format to be used.

In Weka, StringToNominal and NumericToNominal conversion filters were applied in some fields of the user history and the confidence of 0.9 and support of 0.1 were used as standards. The results obtained by Weka have been satisfactory, concluding that Apriori can be adopted since it is adequate to our system needs.

The Cosine [13] is a similarity measure, a forecast method which calculates the similarity between items and users, consults similar items to a given item and matches item content and user profile. The data also had to be adjusted to be used with Cosine. Database in MySQL was used with the EPG and the user history. From these two tables, it was possible to derive two more, one with the profile of the program watched by the user and other with the profile of genres.

It was necessary that the EPG passed through a modification which should also occur in the implementation. A new table was created, identical to the EPG table, but added with fields containing the genres names. According to the adjustment of the program in the genres, these fields were populated with 0 or 1, becoming a matrix.

From these tables was calculated the Cosine for the programs and genres, the profile and what could be recommended to the user. The results from the Cosine were also satisfactory confirming that this technique can be applied to the system for it can be adjusted to the system needs.

VI. ANALISYS AND RESULTS

During the tests, it was possible to note some particularities. Our system recommends contents based on the programs genres and our analyses were performed according to this standard. With Apriori algorithm, the data are collected in the correct format to be used. For the Cosine, the EPG needs to be changed to a matrix before starting the process of discovering profiles and recommendations.

The Apriori is able to mine only the historical view of the User, discovering your profile from the rules. To select the programs to be recommended, another technique should be used. The Cosine can discover the user profile and select the programs to be recommended. However, Apriori can discover other characteristics in the user history, for example, the "user remains in front of the TV more frequently at night, he likes to watch movies and watches TV more frequently on Monday".

The Cosine cannot discover these characteristics, but reaches our goal. In order to discover behaviors similar to the association rules, it is necessary to consult more complex the database.

Apriori output must be operated in order to give the correct user profile, that is, the rules must be understood, and that is very hard concerning implementation. The Cosine output is clearer; the result straightly reaches intended goal, allowing the output to be used without the need of a post-treatment.

Regarding the input, there is no need of treatment for Apriori, since all data will be used as they are collected. However, for the Cosine, whenever the EPG is updated, the table containing the EPG matrix must be changed according to the new EPG, becoming something hard to work.

Then simulated with the two techniques the process of delivery and acceptance of recommendations by calculating the percentage of correct and generating graphics. Figure 1 shows the percentage of correct Cosine, Figure 2 the percentage of correct Apriori and Figure 3 a comparison between both techniques for all households. The percentage was calculated by dividing the number of recommendations accepted (0 to 5) by the number of recommendations made (5).

The profile of the genres founded by both algorithms is similar. A sample of the profile can be seen in Table I. Although both techniques to cover the needs of the system, the Cosine is one that can be better utilized.

In a desktop, the feedback of the Cosine calculation is faster in relation to the feedback of Apriori association rules. However, further studies relating the use of these algorithms processes in a cell phone with IDTV is still not possible in Brazil. The time required to conclude the recommendation process varies according to the personalization technique to be used. In our tests and simulations, Cosine ends the process before Apriori.

Figure 1.   Accuracy Cosine all residence.
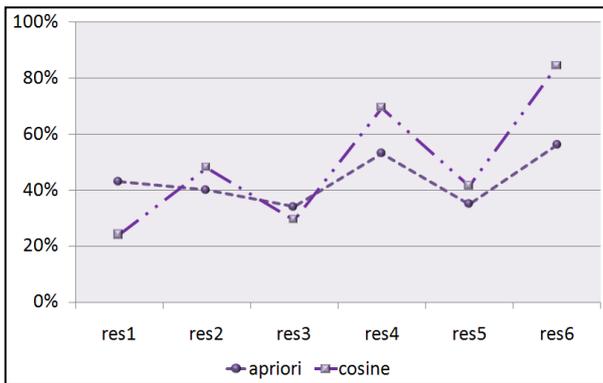
Figure 2. Accuracy Apriori all residence.

Figure 3. Comparison Accuracy between Apriori and Cosine.

TABLE I. PROFILE OF GENRES

Cosine	Apriori
Journalism / Newscast	Variarity/Others
Variarity/Others	Journalism/Newscast
Infantile/Infantile	Infantile/Infantile
Soap Opera/Soap Opera	Soap Opera/Soap Opera
Series/TV Series	Journalism/Report
Reality Show/Series	Series/TV Series
Movie/Movie	Movie/Movie

VII. CONCLUSION

The system was designed according to Brazilian rules [14] for portable devices due particularly to the fact it is impracticable today to develop the system integrated to a middleware for portable digital television.

Surveys showed that although both algorithms meet our needs, we conclude that the algorithm Cosine can

have a better performance in recommendation systems for IDTV for cell phones, considering the processing time and the accuracy in the residence.

Future works can include collaborative filtering algorithms and also a new architecture using client-server, generating and offering other kind of personalization services for users.

REFERENCES

[1] "Sistema Brasileiro de Televisão Digital." Available in: http://sbtvd.cpqd.com.br/. Access in August 16, 2009.

[2] "Fórum do Sistema Brasileiro de Televisão Digital." Available in: http://www.forumsbtvd.org.br/. Access in August 17, 2009.

[3] "Electronic Programme Guide. Protocol for a TV Guide using electronic data transmission." ETSI standard ETS 300 707. Available in: http://webapp.etsi.org/workprogram/Report_WorkItem.asp?WKI_ID=3883. Access in August 18, 2009.

[4] ABNT NBR 15603-2D3-2007-Vc3-2009. "Digital Terrestrial Television – Multiplexing and service information (SI) – Part 3: Syntaxes and definitions of extension information of SI."

[5] "Service Information For Digital Broadcasting." ARIB standard STD-B10. Available in: http://www.dibeg.org/techp/techp.htm. Access in August 18, 2009.

[6] ABNT NBR 15606-2D2-2007-Vc3-2008. "Digital Terrestrial Television – Data coding and transmission specification for digital broadcasting – Part 2: Ginga-NCL for fixed and mobile receivers – XML application language for application coding."

[7] Bär, A. et al. "A Lightweight Mobile TV Recommender: Towards a One-Click-to-Watch Experience." In Proceedings 6th European Interactive TV Conference, p.142-147, Salzburg, Áustria, 03-04/07/2008.

[8] Uribe, S. et al. Mobile TV Targeted Advertisement and Content Personalization. In 16th International Workshop Conference on Systems, Signals and Image Processing, Chalkida, Greece, 18-19/06/2009.

[9] Solla, Alberto Gil et. al. "ZapTV: Personalized User-Generated Content for Handheld Devices in DVB-H Mobile Newtorks." In: Proceedings 6th European Interactive TV Conference, p.193-203, Salzburg, Áustria, 03-04/07/2008.

[10] Gatto, Elaine; Zorzo, Sergio. "Sistema de Recomendação para TV Digital Portátil Interativa Brasileira," in 8th International Information and Telecommunication Technologies Symposium. Florianópolis, Santa Catarina, Brasil. 09-11/12/2009.

[11] "IBOPE." Available in: http://www.ibope.com.br. Access in August 28, 2009.

[12] Witten, Ian H., Frank, Eibe. "Data Mining: Practical Machine Learning Tools and Techniques." Morgan Kaufmann, June 2005, 525 pages.

[13] Torres, Roberto. "Personalização na Internet." Novatec Editora. 2004. 158p.

[14] Normas Brasileiras para Televisão Digital. Available in: http://www.forumsbtvd.org.br/materias.asp?id=112. Access in March 2010.