Design and Implementation of Broadcasting and Television Decision Support System module for Personalized Programs Recommendation

Ya Gao, Fulian Yin, Jianping Chai
Information Engineering School
Communication University of China, CUC
Beijing, China
cuc_gaoya@163.com

***Abstract*—To solve the problem in the original and one-way industry chain of Broadcasting and Television that it is difficult for network operators to communicate with other members in the industry chain, a personalized program recommendation module is proposed as a typical module in broadcasting and television decision support system. This paper analyses the Clustering algorithm in data mining, and presents the personalized program recommendation module which is based on the Clustering algorithm. The module recommends personalized programs to TV users and provides effective decision support for network operators.**

Keywords-broadcasting and television decision support system; personalized program recommendation; data mining; Clustering algorithm

I. INTRODUCTION

Decision Support System (DSS) was brought forward first in the 1970s by American scientist Keen and Scott Morton. And it had a great development in the 1980s [1]. With the study and exploration of domestic and foreign experts and scholars, Decision Support System nowadays has developed to a novel DSS consisting of Data Warehouse (DW) [2]. On-line analytical processing (OLAP) and Data mining (DM).Its typical feature is getting information that can assist in making decisions from mass data [3].

Data mining (DM) is a technology that can extract valuable knowledge from a large amount of data. With the continuous improvement of the data mining technology, data mining has been more and more widely used in decision support technologies. This knowledge provides powerful support for decision making [4].

The personalized program recommendation module introduced in this article is a typical application of the data mining technology in the Broadcasting and Television Decision Support System (BTDSS).This module analyses viewing behavior of TV users making use of the mature clustering algorithm in the data mining technology, extracts users' personalized information from a large amount of viewing data and provides support for decision maker to recommend personalized programs to TV users.

II. FRAMEWORK OF BROADCASTING AND TELEVISION DECISION SUPPORT SYSTEM

Broadcasting and Television Decision Support System is a computer application system that can help every member in the industry chain of Broadcasting and Television to make decisions by analyzing TV users' need, using relevant data, model and knowledge, and by means of the man-machine interaction. Its important function is to provide decision support for government departments, program producers, program release agencies, network operators, advertisers and TV users.

Broadcasting and Television Decision Support System is a 4-base system that is based on Data base (DB), Model base (MB), Mode base (MB) and knowledge base (KB) [5]. As is shown in Fig. 1, its framework is constructed by access layer, service layer and data layer.

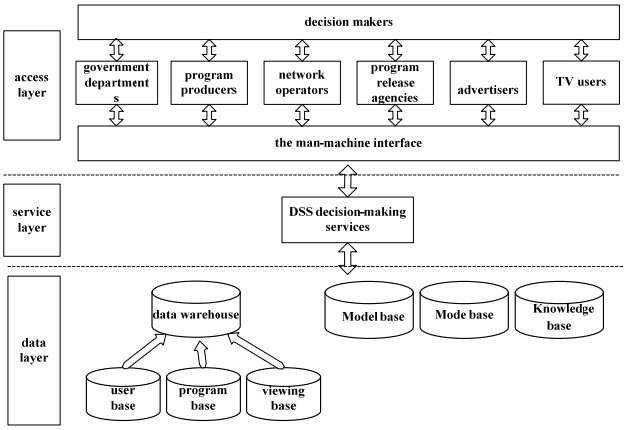

Figure 1. Framework of Broadcasting and Television Decision Support System

The top layer, access layer is the interactive platform for decision makers and BTDSS. Decision makers control the operation of the BTDSS through the man-machine interface, and the final results will be presented to the decision makers through the man-machine interface. The middle layer of the system is service layer that provides exclusive decision-making services for decision makers, this article will introduce the personalized program recommendation module in depth. At the bottom of the system is data warehouse. It stores and manages the user base, program base and viewing base in the way of Data Mart. Model base (MB), Mode base (MB) and Knowledge base (KB) store and manage the model, mode and knowledge in the system.

Published by Atlantis Press, Paris, France.

III. THE CLUSTER ANALYSIS IN DATA MINING

Cluster analysis is one of the most commonly used technologies in data mining, and in this module, it is one of most important methods to analyze TV users' viewing behavior. The so-called clustering is the process of dividing the data into classes or clusters, so that the objects in a cluster are similar to each other, and objects in different clusters are dissimilar. As a branch of statistics, many clustering algorithms are widely used.

A. Main clustering algorithms

In general, the main clustering algorithms can be divided into 5 categories:

1) Method based on partition: object data are divided into K clusters, ensure that each cluster having at least one object and each object must belong to and only belong to one cluster. Clustering algorithms belonging to this class are as follows: K-means algorithm, K-mediods algorithm, PAM, CLARA, CLARANS etc.

2) Method based on hierarchical: decompose given object data by condensation or splitting method. Clustering algorithms belonging to this class are as follows: DIANA algorithm and AGNES algorithm.

3) Method based on density: as long as the density of " neighborhood " (number of objects or data points) exceeds a certain threshold, then continue to clustering. Clustering algorithms belonging to this class are as follows: DBSACN and OPTICS.

4) Method based on network: divide the data into finite unit of grid structures, and then continue to clustering in single unit. Clustering algorithms belonging to this class are as follows: STING and CLIQUE.

5) Method based on model: assume a model for each cluster, and search for data sets that can meet the model. Clustering algorithms belonging to this class are as follows: COBWEB and SOM.

The most famous and the most commonly used is partition method and the following will introduce K-means, and K-mediods [6].

B. K-means algorithm

K is the input parameter in K-means method, and N object data are divided into K clusters. The objects in a cluster have high similarity, but the similarity between clusters is low. The similarity of clusters is mean measure of objects in the cluster, and it can be seen as centroid or center of gravity. The process of K-means method is as follows[7]:

Input:

K: number of cluster

D: data set contains N objects

Output: a collection of K clusters

Method:

1) select any k objects as the initial cluster centers from D;

2) repeat

3) assign each object (again) to the most similar cluster according to the means of objects in clusters

4) update cluster means, that is calculating average of objects in each cluster;

5) until the criterion function convergence and no longer changes

The square error criterion function is usually used as criterion function, which is defined as follows:

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} \left| p - m_i \right|^2 \quad (1)$$

E is the square error sum of objects in the data set and the corresponding cluster centers, p represents a given object, m_i is the mean of cluster C_i.

C. K-mediods algorithm

The process of K-mediods is similar to K-means and their difference lies in: K-mediods takes use of an actual object to represent a cluster when K-means takes use of the average value of the object in a cluster. PAM（Partitioning Around Medoids）is one of the K-mediods methods first proposed. The process of PAM method is as follows:

Input:

K: number of cluster

D: data set contains N objects

Output: a collection of K clusters

Method:

1) Select any k objects as the initial representative objects from D;

2) Repeat

3) assign each remaining object (again) to the nearest cluster according to its distance between representative objects;

4) randomly select an object as O_{random} which is not the representative object

5) calculate the total cost S when update the representative object O_j with O_{random};

6) If S<0,then update O_j with O_{random}, and get a new data set with K clusters;

7) until the total cost S no longer changes

The total cost S is effected by the absolute error E, which is defined as follows:

$$E = \sum_{j=1}^{k} \sum_{p \in C_j} \left| p - o_j \right| \quad (2)$$

E is the absolute error sum of objects in the data set，p represents a given object，O_j is the representative object of cluster C_i . The quality of the clustering result is measured by the cost function. If the current representative object is replaced by other object, cost function will calculate the difference of absolute error. Total cost S is the cost sum of objects which are not representative objects. If S<0, E will

decrease, O_j can replace by Orandom; If S>0, the current O_j is acceptable, no change occurred in this iteration.

D. The choice of the clustering algorithm

K-means is chosen in this module based on the following three considerations:

- For a user in this system, the differences between him and the other users can be distinguished through appropriate measure. Distance measure used in the method based on partition is an intuitive and widely used measurement method. Therefore, K-means is feasible chosen in this module.
- K-means algorithm is simple and fast. The load of the system is usually very high, and K-means can reduce the additional system overhead as much as possible. K-means is chosen because the execution cost of K-means is lower than K-mediods and for this module it is important to find clusters quickly.
- K-means algorithm can handle large data sets in this system with relatively good scalability and high efficiency.

IV. THE PERSONALIZED PROGRAM RECOMMENDATION MODULE

A. The theoretical basis of the personalized program recommendation module

The personalized program recommendation module is designed for network operators and TV users on the basis of TV users' viewing behavior analysis. TV users' viewing behavior analysis, is to find the watch habit of different users. This module is based on households' viewing preference ranking to a type of program (measured by viewing duration), and households are classified as several grades. According to the cost budget and promotion need, the network operators can recommend personalized programs to users selectively.

B. The algorithm of the personalized program recommendation module

As is shown in Fig. 2, it is the flow chart of the personalized program recommendation module.

After selecting the region, time and type of program, the system will give the households' viewing preference ranking to the selected type of program. There are two methods to classify the families which are artificial classification and automatic clustering. The decision maker can recommend personalized programs to TV users through the Message platform or Email platform. Main algorithms involved in this module are as follows:

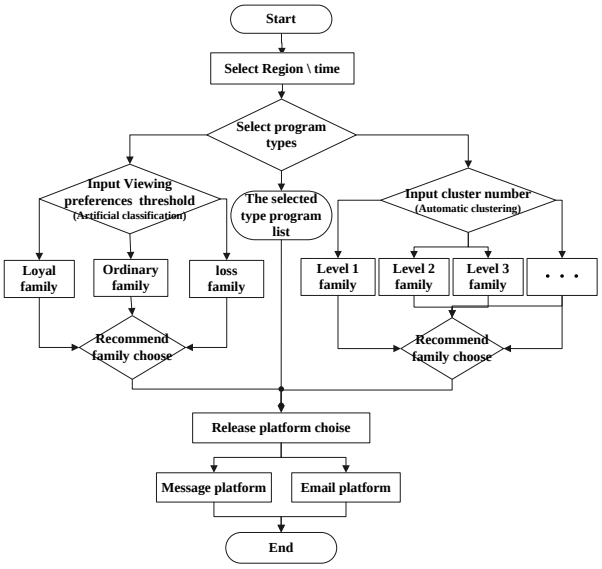

Figure 2. The flow chart of the personalized program recommendation module

1) Viewing preferences threshold algorithm: When the decision maker chooses artificial classification, he needs to put in two viewing preferences thresholds: a, b (a<b), according to his experience. When viewing preferences of a household (vwp) ≥b, this family is a loyal household, that is, this household always watches the selected type of program .When a< vwp <b, this household is an ordinary household. When vwp ≤ a, this household is a loss household, that is, this household seldom watches the selected type of program and has the most improvable space.

2) The application of K-means clustering algorithm: When the decision maker chooses Automatic clustering, he needs to put in the number of cluster. The system will call the K-means clustering algorithm , and all households' viewing preference values will be clustered. The system will calculate the center and proportion of every cluster, the clusters will be named level 1,2…households. Level 1 households are the most loyal to the selected type of program, and others' loyalty decline one by one.

V. DECISION MAKING EFFECT DISPLAY OF THE PERSONALIZED PROGRAM RECOMMENDATION MODULE

As is shown in Fig. 3, households' viewing preference values are shown in the form of bar chart in ascending order, so that the decision maker(s) can control the change trend more broadly. The classification results of households are shown in Fig. 4 and Fig. 5 in the form of pyramid graph, the closer to the top means that the faithful degree is higher. The result shows that loyal households are only a small part in the whole people, when the loss households are the most.

The user number in this system is large enough to ensure sufficient clustering data, and reflect the effect of the recommendation method. Based on the cluster theory, the higher the K is, the better the quality of clustering. But on the other hand, the higher the K is, the greater the system

overhead. So experience value, 3 is chosen as the cluster number.

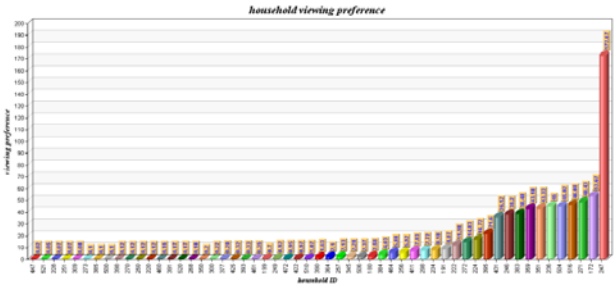

Figure 3. Households viewing preference

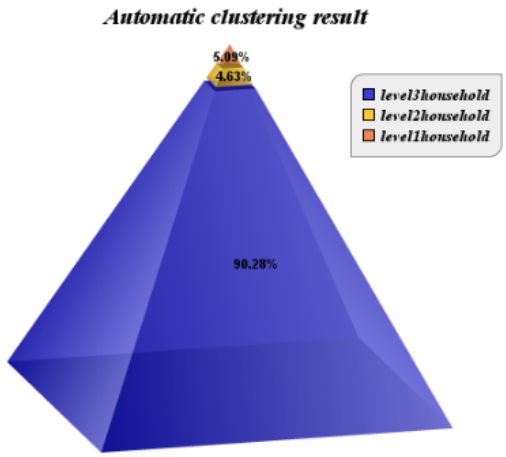

Figure 4. Automatic clustering result

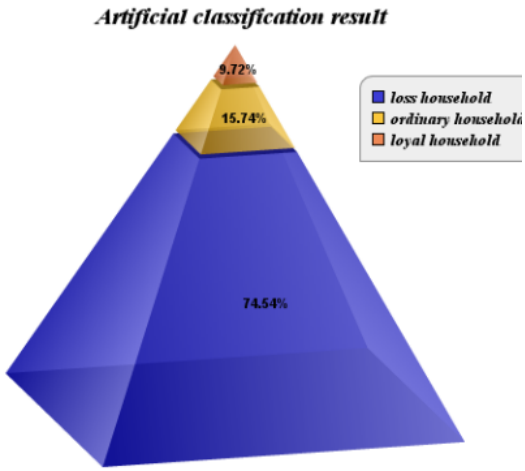

Figure 5. Artificial classification result

VI. CONCLUSIONS

Under the background of building the Broadcasting and Television Decision Support System (BTDSS) at full blast, this paper proposes a personalized program recommendation module which is based on the model of BTDSS. The Paper analyses the data mining algorithms involved, theoretical basis of the module and displays the decision making effect. The establishment of this module can recommend personalized programs to different TV users. It has very important meaning to the development of BTDSS.

REFERENCES

[1] Vo Thi Ngoc Chau and Nguyen Hua Phung, "A Knowledge-Driven Educational Decision Support System," Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), 2012 IEEE RIVF International Conference, Ho Chi Minh City, pp. 1-6.

[2] Shaohui Wang, "Application of Decision Support System in E-Government ,"2009.MASS '09.Wuhan,pp.1-4,Sept.2009.

[3] Fick, G., Sprague Jr., R.H., Decision support systems: issues and challenges, Pergamon Press, Oxford, 1980.

[4] Meijing Kong and Wanzhen Zhou, "The application stmy of the data mining technology in university DSS," Proceedqs of the Third International Conference on Machine Learning and Cybernetics, Shanghai, pp.26-29, August 2004.

[5] Xianyi Qian and Xianjun Wang, " A New Study of DSS Based on Neural Network and Data Mining," 2009.EBISS '09.Wuhan,pp.1-4,May 2009.

[6] Odukoya, O.H,Aderounmu,G.A and Adagunodo,E.R, "An Improved Data Clustering Algorithm for Mining Web Documents," CiSE,Wuhan,pp.1-8,Dec.2010.

[7] Hongwei Xie, Li Zhang, Li Zhang and Xueli Yu, "Application of K-means Clustering Algorithms in News Comments," ICEE.2010.942, Guangzhou,pp. 3759 – 3762, 7-9 May 2010.