8-2012

Consumer Micro-behavior and TV Viewership Patterns: Data Analytics for the Two-way Set-top box

Ray M. CHANG
Singapore Management University, mrchang@smu.edu.sg

Robert J. KAUFFMAN
Singapore Management University, rkauffman@smu.edu.sg

Insoo SON
Korea University

Consumer Micro-Behavior and TV Viewership Patterns: Data Analytics for the Two-Way Set-Top Box

Ray M. Chang
Singapore Management University
mrchang@smu.edu.sg

Robert J. Kauffman
Singapore Management University
rkauffman@smu.edu.sg

Insoo Son
Korea University
insoo114@gmail.com

ABSTRACT

Consumer behavior patterns related to home digital media use are changing due to technological innovations. We examine them in the presence of two-way cable television (CATV) set-top boxes. They permit viewers to change channels, switch to the Internet, and order paid programming, among other functions. We focus on micro-level data that are generated from consumer CATV viewing behavior. We capture clickstreams of channel-changing behavior when consumer use a remote control handset to interact with the set-top box in their home. We explore a variety of data analytics results that characterize patterns of consumer channel-switching behavior, as a basis for suggesting different clusters of observed behavior. We also probe the explanatory elements that give rise to what we see.

Categories and Subject Descriptors

H.m [**Information Systems, Miscellaneous**]: Cable TV, consumer behavior, cluster analysis, data analytics, entertainment.

General Terms

Management.

Keywords

Cable TV, consumer behavior, cluster analysis, data analytics, entertainment, preferences, TV channels, viewing patterns.

1. INTRODUCTION

A two-way set-top box provides an interface for the provider and the CATV system in a household that consists of a forward path and a return path. For the CATV service provider, the two-way set-top box stores data about the interactions that occur when a viewer uses a remote control. Examples include turning a set-top box on and off, and switching channels. It can transmit the user's CATV viewership information on the return path to the provider. This process yields a large of amount of data representing viewership records. Such data are generated in real-time, and create digital traces of micro-level consumer viewing habits in unprecedented detail compared to what has been available in the past. Micro-level data of this sort support discovery of new business knowledge and detailed consumer insights that have not been recognized or empirically validated.

What kinds of patterns are valuable to focus on? The literature from social science and marketing related to TV viewership behaviour suggests three things. The first is viewing duration, with channel-switching behavior as its complement [3]. The second is the heterogeneous viewing patterns for different members of a household [4]. And last, the different patterns of viewership aligned with the different service bundles and levels of willingness to pay that consumer express [2].

International Conference on Electronic Commerce '12, August 6-8, 2012, Singapore Management University, Singapore.

CATV viewers demonstrate specific behavioral patterns, based on their typical activities while watching TV, such as channel surfing, switching channels when ads come on, or switching from news shows to other news shows. It is possible to capture details of all of these activities as a stream of digital data, which continues until the consumer turns the TV off. From their observed choices, we can understand a TV viewer's behavior.

A way to evaluate is to use the total time a person spends watching channels or channel types during a time period, such as a day or a month. The resulting data represent the expressed preferences of the consumer. When we aggregate TV viewers across many set-top boxes, it is possible to assess whether there are distinctive clusters of channel preferences among viewers.

2. DATA

Our data were obtained in a setting with high ethnic diversity, and represent a single month from October 1 to 31, 2011. They cover greater than 100,000 two-way set-top boxes, resulting in about 50 million return path data (RPD) records. A single RPD record consists of four fields: a Set-Top Box ID, a Channel ID, a Start TimeStamp, and an End TimeStamp. These fields tell us about TV viewers in terms of which set-top box was involved, as well as which channel they were watching, and from what times start to finish that they watched TV. CATV service providers offer almost two hundreds of program channels. They are divided into eight channel types, representing program genres. Examples include movies, and sports. Others include assorted channels bundled for targeted viewers. In our setting, this includes channel types localized to the TV viewership.

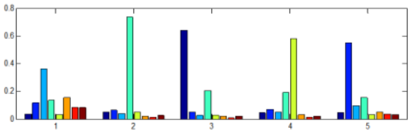
To work with the data, we prepared a master table, with rows for specific set-top boxes in the data sample, and columns for the amount of time a household watched each TV channel type via the set-top box during the one-month period. For each set-top box, we summed the time spent viewing each channel type to create a time matrix. We further computed percentage of total time each set-top box was used to watch each of the channel types. We summed the entries in each row to yield the total time the set-top box was used during the month, and then divided each row entry by this value to percentage matrix. This viewing time percentage vector gives set-top box use for viewing the channel types in a month.

3. METHODS AND RESULTS

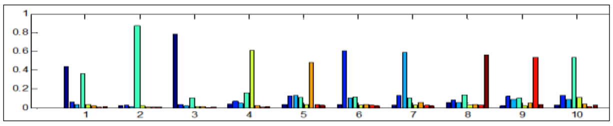
3.1. TV Viewership Patterns

To provide evidence for different viewing patterns, we adopted k-means clustering with Euclidean distance as its metric [1]. It forms each cluster as a hypersphere of arbitrary shape positioned around a centroid vector in higher-order dimensional space, based on the dimensions used to form the clusters. A centroid vector is a mean vector for all observations that belong to a cluster. The observations that surround the centroid show more or less variation in the values of the dimensions relative it.

By applying k-means clustering, we seek to establish the existence and characteristics of channel preference patterns among TV viewers in our data set. To address existence, we compared the clustering outputs between from the actual viewership in the CATV service provider's data set with another random machine-generated data set. This data set is of the same size and format, and was also formed into a percentage matrix, with the values of all rows summing to 1. See Figure 1.



(a) Five clusters



(b) Ten clusters

Figure 1. Clustering results based on viewership data

3.2. Viewership Clusters and Bundles

We did exploratory modeling work using a nested logit model to see if it would be possible to analyze program bundle choices for each of the clusters. We also estimated a multinomial logit model for the purposes of making an initial comparison based on a more naïve empirical.

In the CATV business, subscription fees are a major revenue source. To facilitate the choice of program channels, service providers usually offer different bundles of channels with monthly subscription plans at different price points. The bundles may be designed differently, for example, in terms of the basic channels, high-definition versus standard-definition broadcast quality, and the inclusion of premium contents. The latter often includes add-on channels to make the coverage of movies and sports richer. The choice of a specific bundle reflects a customer preference-driven willingness to pay.

Identifying which customer cluster is more likely to buy which program bundle enables a provider to evaluate viewing time-based customer clusters from a business value perspective. It provides a basis for preparing marketing promotions and bundle add-ons to create greater value with each customer cluster.

In our analysis, we used five different bundle options in increasing order in terms of price and number of channels included: from Bundles 1 to 5. Our preliminary estimates were obtained for a model with Bundle 3 as the base case. First, customers seem to have preferred program bundles with more diverse channels across all of the different customer clusters. This is indicated by the negative coefficients in the Bundle 1 and Bundle 2 columns of Table 2. The results suggest the growing need for more diverse content for customers in the market represented by our data. Customers who benefit the most from two-way set-top boxes seem to be those who watch a variety of program genres, and take advantage of the interactive features, such as DVR or pause-live capabilities.

Second, the clusters we obtained appear to affect bundle choices for TV viewership. For instance, Cluster 1 is more likely to represent Bundles 3 and 7. Households in Cluster 1 show a strong preference for program bundles offering more channel types. They are willing to pay for higher-priced bundles because they like to watch programs across a variety of channel types. Bundle 5 loads the highest on Cluster 4. This suggests that the cluster should include families with children. They may prefer bundles with a greater diversity, so each family member will enjoy the contents.

Third, program bundle choices also are linked to household dwelling type. Households in Cluster 4 tend to select less diverse program bundles, such as Bundles 1 and 2. So these households probably have a greater preference for some specific channels. Fourth, the variables Condo and House have negative coefficients for Bundles 1 and 2, and positive coefficients for Bundles 6 and 7. Compared to households living in apartments, condo and house dwellers have a stronger preference for the bundles that give them access to many channels.

4. ONGOING WORK

Our exploratory analysis suggests some significant relationships between the explanatory variables and the bundles chosen by customers, but the variance captured by the model is still relatively small. Contextual factors may affect customer choices of program bundles. Using the current analysis as a stepping stone, in our ongoing work we are assessing other contextual factors that may more clearly explain customer choices.

5. ACKNOWLEDGMENTS

This research is supported by the Singapore National Research Foundation under International Research Centre at Singapore Funding Initiative administered by the Interactive Digital Media Programme Office.

6. REFERENCES

[1] Bishop, C. M. 1995. Neural Networks for Pattern Recognition. Oxford University Press, Oxford, UK.

[2] Crawford, G.S., and Yurukoglu, A. 2008. The welfare effects of bundling in multi-channel television markets. Working paper, Dept. Economics, University of Arizona, Tucson, AZ.

[3] Dix, S., and Phau, I. 2010. Measuring situational triggers of television channel switching. Mktg. Intel. and Planning, 28, 2, 137-150.

[4] Wonneberger, A., Schoenbach, K., and van Meurs, L. 2009. Dynamics of individual television viewing behavior: models, empirical evidence and a research program. Communication Studies, 60, 3, 235-252.