1-2014

How Can Consumer Preferences Be Leveraged for Targeted Upselling in Cable TV Services?

Bing Tian DAI
Singapore Management University, btdai@smu.edu.sg

HOW CAN CONSUMER PREFERENCES BE LEVERAGED FOR TARGETED UPSELLING IN CABLE TV SERVICES?

Bing Tian, Dai
Research Scientist, Living Analytics Research Centre, Singapore Management University, *Singapore*
80 Stamford Road, Singapore 178902
btdai@smu.edu.sg

ABSTRACT
Internet TV has attracted a significant amount of attention from the conventional cable TV service providers, by providing customized TV programs at preferred time slots. The cable TV service providers are seeking to retain their customers by giving them a better experience: by understanding their customers' preferences and upselling them the right products to cater to their interests. It is not easy to understand customer preferences though, since customers are not able to watch channels to which they have not subscribed. This makes it difficult to predict what they will like to watch, as a result. In this paper, I discuss my ongoing research on TV viewership behavior. I model customer preferences using a technique called latent Dirichlet analysis (LDA), by considering channel viewing behavior as a similar process of article generation. Customer preferences over unsubscribed channel are calculated from the LDA model. My model achieves better prediction performance as a result. I also present a quantitative case study to show that the observed channel viewing behavior makes sense.

KEYWORDS: Cable TV, customer targeting, data mining, retail telecom services, viewership patterns

1. INTRODUCTION

Watching TV is the most basic entertainment activity observed in every household. There are two ways of broadcasting in conventional TV services: radio frequency broadcasting and signal delivery through coaxial cables. The latter is referred to as *cable TV services*. Due to technological advances in the telecommunication industry, many services providers have started to offer cable TV services, or a combination of cable TV, telephone, and Internet services as *triple-play services*. In the case of Verizon (2011), a leading U.S. telecom firm, its annual report states that its total number of cable TV subscribers increased from 2.7 million in 2009 to 4.2 million in 2011, while the number of voice connections decreased by 4 million during that period. This shows that cable TV services have become a major force in the telecom industry recently.

As the Internet has become a more convenient platform for signal transmission, people have started to take advantage of new ways to get conventional TV content via the Internet. This is called *Internet TV services*. It begins with the duplication of content that already has been broadcast on conventional TV. Internet users are able to download and watch many TV programs as a result. A new direction in Internet TV services includes the creation of unique content for this delivery channel. Many consumers have become used to acquiring and sharing content online this way.

Internet TV has attracted many new customers recently, with growth of about 6.9% in 2012 (Friedman 2013). Four things have made Internet TV popular: it supports conventional TV entertainment viewing; it also allows viewers to watch anything that they want; they also can watch TV anytime; and they can watch TV anywhere. With this environmental change in the cable TV business, identification of customer preferences and offering the right services have become important issues in the cable TV business. Cable TV services providers offer hundreds of different channels in bundles to their customers. Yet this variety makes it difficult for customers to select what they really want.

Moreover, because TV channels represent *information goods* (Chuang and Sirbu 1999), customers usually will only realize the value of the channels they have purchased access to after they watch them. For that reason, understanding viewer preferences and predicting their behavior have been treated as difficult tasks in the media industry (Davenport and Harris 2009). The standard revenue model in telecom services usually relies on a monthly subscription fee, and increasing *average revenue per user* (ARPU) while achieving customer retention are the main goals in this kind of business (Iyengar et. al. 2007). Identifying each customer's unique preferences provides useful information for developing an upselling strategy that is tailored to the needs of existing customers. Under the threat of Internet TV, it is also important for customer retention to find out whether customers are satisfying with the current services.

To address this issue, I seek to identify customer preferences based on the current TV subscription and viewing behavior at the household level of analysis. This will enable me to recommend strategies to upsell cable TV channels based on observations of household-level viewing preferences. More specifically, I ask: Which channels and bundles will customers want to subscribe to next, if the prices are the same, based on their current preferences? There are always two issues in upselling: identifying customer preferences and figuring out what prices they are willing to pay. Customers may choose some channels that they like but they also may not be willing to pay very much. More preferred channels usually come with a higher price though, which makes the analysis of preferences more difficult. In this work, I will set aside the issue of pricing, and look at customer preferences. In particular, I will argue that it is important to study personal preferences to understand customers better, and then to plan upselling strategies around what can be learned about them.

In Section 2, I will review the background and challenges related to personalization and targeted upselling. Section 3 gives an overview of recommender systems concepts. In Section 4, I introduce my methodology, topic modeling. I also offer an analogy between topic modeling and customer preference modeling to motivate my analytical approach. Section 5 explains how to obtain preferences over unsubscribed channels and bundles from my model, and I apply it to real data in Section 6. I will discuss my additional work-in-progress at the Living Analytics Research Centre in Section 7.

2. TARGETED UPSELLING OF CABLE TV SERVICES

2.1. BACKGROUND

Like conventional TV services, Internet TV is also able to provide live content, and Internet users are able to choose whatever they want to watch. For cable TV services, this is called *video on demand*, and customers are required to pay extra to acquire this service, often on a pay-per-view basis. For Internet TV though, this service is often free. Services providers put some of their previously broadcasted TV programs online, so that their customers can catch up with viewing them. The platforms for Internet TV viewing have become very flexible, so Internet users can also watch Internet TV anywhere and anytime. For example, smartphone users can watch TV through phone applications that have become available for when they are traveling to and from work – something that conventional cable TV service providers have not yet begun to do.

With the emergence of Internet TV, cable TV services providers need to rethink how to retain their customer base and TV viewing market share. However, the current way that cable TV services providers acquire their customers is passive: they usually try to upsell their services to specific consumers who express the need for them. They also launch promotion campaigns to catch the attention of their customers with the hope that some will be persuaded to upgrade.

2.2. PERSONALIZATION AND TARGETED UPSELLING

Internet TV and modern online marketing strategies target groups of Internet users. One of the reasons is that the number of Internet users is huge and it is difficult to target at all of them at once in a meaningful way. Instead, it is better to look at some targeted groups of Internet users, and make marketing strategies that aim to improve the monetization of their consumption. Consider Internet TV as an example. When Internet users watch videos on YouTube (www.youtube.com) or Youku (www.youku.com), there are also given *personalized recommendations* for other videos to watch next. Such personalization is difficult to implement on cable TV though: most home TV sets are not interactive enough to permit this. Also, unlike the way that Internet users choose what they want to watch, cable TV viewers are passive viewers of TV programs: most do not get any recommendations from cable TV service providers that they can react to. They have to look at the program tables themselves, and then decide which channels to switch to on their own. There have been several studies related to personalized TV program listing services (Smyth and Cotter 2000, Martinez et. al. 2006. Lee et. al. 2009), but most of them are based on customers' self-reported preferences.

For cable TV service providers to retain their customers, they have to understand them better than ever before. So instead of launching promotions for all customers, they need to specialize their promotions to what different customer segments will see, and make marketing strategies for different groups in the same ways Internet TV providers are using now. By doing this, cable TV service providers can continue to create value for their customers, and encourage them to purchase more services. This process is called *targeted upselling*. In telecom services, customer usage history is treated as valuable information to generate new revenue through targeted upselling strategies (Kim 2002).

The key to success with targeted upselling is to give customers channels and entertainment services bundles that they like. To do targeted upselling, cable TV service providers have to learn about what their customer like or do not like. Unlike conventional promotion campaigns, targeted upselling increases the willingness-to-pay of specific customers, which makes them more satisfied. The provider's marketing strategies will be tailor-made for them.

2.3. OBSERVED VIEWING BEHAVIOR AND CUSTOMER PREFERENCES IN CABLE TV SERVICES

The most direct way to obtain information on customer preferences is by directly surveying customers themselves about what they like and what they do not like. This is not attractive though since most customers never bother to reply to such surveys. In addition, they often do not have a full picture of what programs are available to them among all the different TV channels they have access to.

Customer preferences can be observed from their behavior though: for example, how they change channels, and what programs they choose to watch. Someone who spends a lot of time watching movie channels will be a big fan of movies. Someone who is interested in entertainment shows will be observed to switch between different channels that have entertainment programs. Thus, a possible way to capture preferences is to data mine them from the observed behavior of cable TV customers.

With appropriate permission and data access, I have been able to observe the following kinds of behavior associated with the use of cable TV set-top boxes that cable TV services providers install in their customers' homes: (1) time spent on each channel or program; (2) channel and program switching behavior; and (3) usage behavior involving the remote control handset. There may be differences in customer behavior at the channel level and the program level, depending on the information that is given to the customer. For example, with access to Electronic Program Guide (EPG) data from a cable TV services provider, it is possible for a researcher to know the programs that are being viewed in different households. This helps to identify whether customers have switched from other channel, or there is continuous viewing of previous programs in a giving viewing period.

As channels and programs usually come with genre information, the time that customers spend on them indicates their interests over all of the possible genres that are available and to which they have access. Not only the time they spend on channels matters, what also matters is the patterns that they exhibit when they switch from one channel to another, or from one program to another. By studying the "digital traces" of consumer behavior in big data sets, such as are available at the Living Analytics Research Centre of Singapore Management University, I will be able to know how people switch channels and move from one genre to another.

Another interesting behavior is customer use of the remote control handset. Viewers can either browse one channel, or select a channel by directly entering a channel number, or select something to watch through the EPG. Each of these actions is associated with a different kind of remote control handset usage behavior. They give some hints about how much a customer likes the selected channels or programs he or she watches.

2.4. THE ROLE OF AVAILABILITY AND BUNDLE CONSTRAINTS

There are two challenges that make understanding customer preferences difficult. The first is *availability constraints*. Each customer's access is restricted by the household's subscription to different channels. He or she does not have access to channels to which he or she has not subscribed. As a result, there is not any information available for household's preferences for unsubscribed channels.

This is different from not consuming these channels at all though. There is only a clear indication of what channels customers do not like: we can get this by observing what channels they subscribe to but do not watch. As a result, customers are not really able to state their preferences about unsubscribed channels, since they are not able to consume channels to which they have not subscribed. Yet both scenarios show that the customer has spent zero time some of these channels. So customer behavior needs to be understood differently due to availability constraints.

The second issue is *bundle constraints*. In cable TV services, some channels are often grouped together and sold to customers. These groups of channels are called *bundles*. When customers choose some channels for their subscription, they must choose an entire bundle, which is likely to include channels that they do not like.

When the availability and bundle constraints are combined, understanding customer preferences becomes more difficult. First, their consumption behavior will be truncated since customers do not have access to unsubscribed channels. Comparing customers on their subscribed channels does not give any information about those channels to which they are not subscribed. Second, due to bundle constraints, customers may spend time on not-so-preferred channels to which they are subscribed han channels that they really might prefer but to which they are not subscribed. The time spent they spend on these channels probably will not be proportional to their preferences though. This is interesting since my more general intuition is that customer time spent on each channel and program should be proportional to how much the customer enjoys viewing them, as a basis for mapping out their preferences. In my research, I consider the household level, which is an aggregate of its members' preferences.

Thus, the availability and bundle constraints are important for understanding customer preferences. Considering them leads me to suggest two reasons for customers not to watch a channel: either because they have not subscribed to it, or because they do not like it. On the other hand, customers may watch a channel either because they like it, or because it is included in a subscribed bundle, and a more preferred channel is not available to them.

3. RECOMMENDER SYSTEMS

Recommender systems are a class of systems that predict ratings or consumption behavior involving customers and products or services (Melville and Sindhwani 2010, Koren and Bell 2011). For example, at online shopping website Amazon.com, while users are viewing products online, they receive recommendations based on their past purchasing history or on their current browsing session. In general, recommender systems deal with a big matrix with rows corresponding to the users and columns corresponding to the products. Each entry of the matrix is a rating, for example, a number between 1 and 5 that the user gives to the product at that particular column. An entry with value 0 indicates unrated product, and the value is to be predicted by a recommender system. A subclass of recommender systems is adoption systems with binary values, 0 or 1. Here, 0 means that the user has viewed the corresponding product before, but did not adopt, and 1 means the user has adopted the product. In such adoption systems, the prediction task is also to predict which products a customer will adopt.

Recommender systems are intended to capture the interests of individual users, and they can be directly applied to targeted upselling. I next will introduce some basic techniques that underlie the construction of most recommender systems.

3.1. COLLABORATIVE FILTERING

Collaborative filtering is a basic technique for recommender systems (Su and Khoshgortaar 2009, Koren 2009). It makes use of customers with similar tastes (Desrosiers and Karypis 2011), so that the consumption behavior of one customer can be inferred from the consumption behavior of other similar customers. It is usually carried out in two steps. First, similarities among customers are calculated according to their consumption history. We can either indicate pairs of similar users or associate a similarity score to each pair of users. In the second step, we then evaluate the rating between a user and a product that he or she has not rated before. This is computed based on the ratings of users who are similar to him or her, by taking an average weighed by the similarity score.

Recall that due to availability and bundle constraints, the similarities among people's behavior patterns for watching TV are not very accurate. So, when applying such inaccurate similarities to understand customer preferences for unsubscribed channels, the results will not be very accurate either.

3.2. MATRIX FACTORIZATION

Another branch of recommender systems is based on *matrix factorization* (Salakhudinov and Mnih 2007, Koren et. al. 2009, Mairal et. al. 2010). As I noted in the beginning of this section, recommender systems deal with a big matrix of ratings. With n users and m products, the dimensionality of the matrix is $n \times m$. Matrix factorization decomposes the big matrix into two smaller matrices, with dimensionalities $n \times k$ and $k \times m$. Each row (a vector of k reals) of the first matrix represents a user while each column (another vector of k reals) of the second matrix represents a product. The predicted rating is the inner product of the vector representing the user and the vector representing the product.

A drawback of matrix factorization techniques is that the results are not readily interpreted. People cannot really explain what are vectors of k values stand for, and it is difficult to extract knowledge from the factorized matrices to support applied managerial decision-making.

3.3. CLUSTERING

The third approach for recommender systems is based on clustering (Ungar et. al. 1998). In cable TV services, customers usually show very strong and consistent watching behavior in a specific segment (Chang et. al. 2012). Customer segments give information on the aggregate behavior of all users in the cluster. Such aggregated behavior can then be used for predicting channel preferences. Again, due to the availability and bundle constraints, the customers segments that are obtained also will not be that convincing. In fact, the clustering method often groups customers with the same subscriptions together, which fail to capture their interests in unsubscribed channels.

All these different recommender systems can be used for targeted upselling, but extra attention must be given to understanding TV customer preferences. In the next section, I will introduce a new method for this purpose based on topic modeling in data mining.

4. MODELING CONSUMER PREFERENCES

I will next present why customer preference modeling can be understood through an analogy for why customer preference modeling can be handled with a *topic model* (Hofmann 1999). A *topic* in the context of recommender systems is similar to the definition of a *genre* in cable TV services, which represents groupings of channels according to some underlining similarities. For example, channels of the same language can be considered as one genre or topic, and the same goes for movie channels.

4.1. LATENT DIRICHLET ALLOCATION

The *latent Dirichlet allocation* (LDA) (Blei et. al. 2003, Hoffman et. al. 2010, Blei 2013) model is useful for *generating text* in a relevant *document or article space*. (See Figure 1 and Table 1.)

Figure 1: Latent Dirichlet Allocation (LDA)

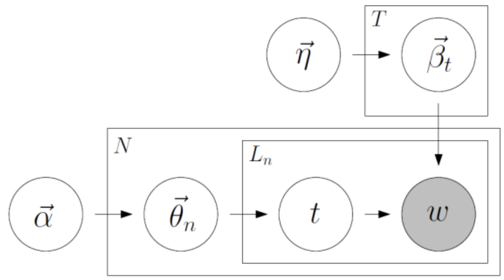


Table 1: LDA Modeling Notation

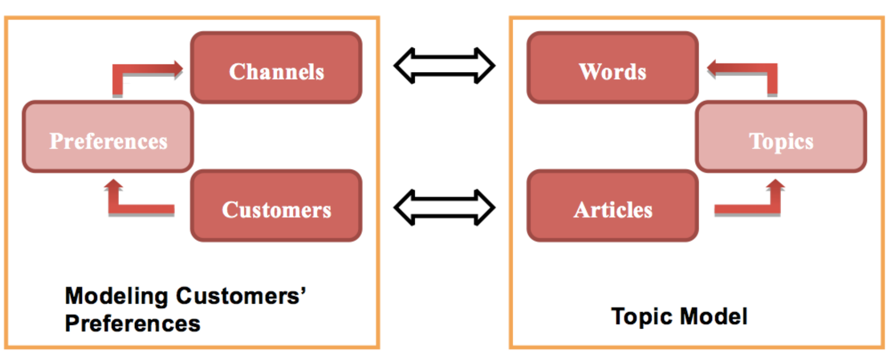
N	Number of documents
T	Number of topics
$\vec{\alpha}$	Topic prior
L_n	Length of document *n*
$\vec{\theta}_n$	Topic probability of document *n*
t	Chosen topic *t*
$\vec{\eta}$	Word prior
$\vec{\beta}_t$	Word probability of topic *t*
w	Observed word that is chosen

To generate an article from a set of *words*, I first generate a *topic probability* for a given article from a *Dirichlet prior*. For example, for the three topics, "news," "sports," and "technology," an article is generated by *topic prior* $\vec{\alpha}$, with the topic probability set given by {<news: 0.3>, <sports: 0.0>, <technology: 0.7>}. A topic is first generated according to the topic probabilities of the article being chosen. For the above probability distribution, there is a 30% probability that a person will choose the topic "news" and a 70% probability of choosing "technology." If the topic "technology" is chosen, a word is then chosen from another set of words associated with this topic, each with possibly different probabilities. The probability of choosing a word given a topic is determined by a *word prior* $\vec{\eta}$.

4.2. AN ANALOGY

Although topic models are used in the domain of document clustering and topic extraction, we can extend the technique to apply to cable TV content by the analogy shown in Figure 2. Note that in the topic model, generating words to represent each article is a two-step procedure. The generation of topics comes first, and then the generation of words that describe the topic. This is exactly the same as choosing a channel to watch.

Figure 2: Topic Modeling and Customer Preference Modeling – An Analogy



A customer first chooses a channel that is related to the household's preferences. For example, if he or she has preferences described by how the household tends to split up its time for viewing a set of genres or topics based on the probability set {<entertainment: 0.3>, <sports: 0.0>, <movies: 0.7>}, this will guide a viewer's choices for channels to watch. The viewer first will pick a preferred genre, say movies, and then will choose a channel to watch under the movie category. By thinking about how customers choose the channels they view as a two-step procedure as with the generation of words, we use an analogy between modeling customer preferences for cable TV viewing and the topic model.

Cable TV viewing customers are like the articles in the topic model, and preferences are like topics. In the topic model, an article is expressed as a probability distribution over topics. Likewise, customer preferences can also be expressed as a probability distribution over preferences. By the same token, choosing movies under a specific preference is the same as choosing a word under a specific topic. Therefore, preferences are like topics, and channels are like words in the analogy I have made. By using this analogy, I can apply the LDA topic model to model customer preferences.

4.3. MODELING CUSTOMER PREFERENCES BY LDA

The next issue to consider is how to extract the observed words from cable TV watching behaviors. Unlike words in articles, cable TV viewing sessions involve time spent on different channels. A *TV viewing session* is a viewing period of at least 15 minutes. Anything less than 15 minutes usually indicates that the viewer is switching channels, or figuring out whether a viewer wants to watch a specific TV program by learning what it is about. If we set a longer period of time for a TV viewing session, we lose fidelity in our observations of what viewers like to watch.

Figure 3 shows that an observation in the cable TV viewing context is not a set of words, but now a set of channels that a customer has viewed, as indicated by the shaded nodes. Table 2 provides the corresponding mathematical notation for the model.

Each channel *c* that a customer watches in a TV viewing session is observable in my research context. For example, I can observe that a customer has watched “Channel 234” as many as 66 times, and “Channel 235” only 8 times in a month. In the LDA model, the sequence of TV viewing sessions is not important. My method generates each topic *t* independently from the genre preference $\vec{\boldsymbol{\theta}}_{\boldsymbol{n}}$. Based on my previous example, a customers with genre preference {<entertainment: 0.3>, <sports: 0.0>, <movies: 0.7>}, and there are 100 observations from her over that month, so $\boldsymbol{L_n}$ = 100. For each of the 100 TV viewing sessions, a genre *g* is generated first. So there are 30 sessions under the “entertainment” genre, and 70 TV viewing sessions under the “movies” genre that are expected. For each genre, the model then chooses a channel as that channel which was viewed during that TV viewing session. This way of modeling customer preference is effective and this model captures appropriate information to identify the genres.

Figure 3: LDA for Modeling TV Customer Preferences

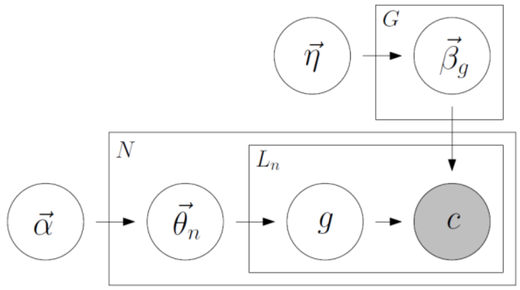


Table 2: LDA Modeling Notation for Modeling TV Customer Preferences

N	Number of customers
G	Number of genres
$\vec{\alpha}$	Genre prior
L_n	Number of watching sessions of customer *n*
$\vec{\theta}_n$	Genre probability of customer *n*
g	Chosen genre *g*
$\vec{\eta}$	Channel prior
$\vec{\beta}_g$	Channel probability of genre *g*
c	Observed channel that is chosen

5. DATA COLLECTION AND BUNDLE PREDICTION

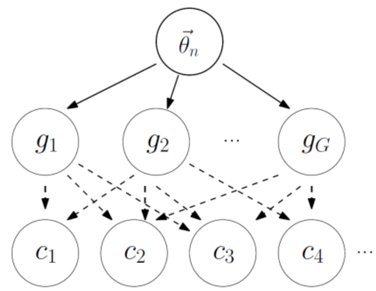
To validate the application of LDA to customer preference modeling, I collected an extraordinarily large amount of TV viewing history data from a leading cable TV services provider. (I am not able to disclose details about the provider due a binding non-disclosure agreement.) As there are prices associated with channel subscriptions, and different channels come with different prices, it is difficult to analyze channel subscription behavior without considering their prices.

On the other hand, bundle prices are much simpler. Let us assume that the provider offers a total of 7 *sub-bundles*, say sub-bundles b_1, b_2, b_3, b_4, b_5, b_6 and b_7, that can be flexibly included in the construction of a *bundle subscription*, *B.* We further assume that all customers have to subscribe to 3 bundles at a minimum. The price is fixed if the number of sub-bundles that a customer selects is the same. So the prices for the subscription $\{b_1, b_2, b_3, b_4\}$ and $\{b_1, b_3, b_5, b_7\}$ will be the same. Though the bundle subscriptions are different, their prices are identical. So when customers want to upgrade their bundle subscriptions, no matter which sub-bundle they choose, the extra price they pay will be the same. For example, if a customer has already subscribed to sub-bundles b_2, b_3, b_5, no matter which one he or she chooses to add from among sub-bundles b_1, b_4, b_6 or b_7, the price difference will be the same. This property enables sub-bundle prediction without considering prices. Next, I will show how to compute a customer's channel references.

5.1. COMPUTATION OF CHANNEL PREFERENCES

By deriving an LDA model from household customer viewing history data, for customer *n*, I can compute the genre preference $\vec{\theta}_n$, and channel probability distribution $\vec{\beta}_g$. Next is to compute the household's preferences over the set of unsubscribed channels. In general, the more a customer likes a channel, the more likely that the customer will choose this channel to watch. So a household's preference for a channel can be measured by the probability that the household will observe to choose that channel to watch. As I show in Figure 4, if channel c_2 is available to customer *n*, the probability of choosing c_2 can be from any of the *G* genres.

Figure 4: Channel Preference Computation in LDA



For example, the customer can choose g_1 or g_2, and then choose c_2. Since it is a two-step procedure, the probability of choosing c_2 through g_1 will be $\theta_{n,g_1} \times \beta_{g_1,c_2}$. Thus, the probability of customer *n* choosing channel *c* by will be $p(n,c) = \sum_{g=1}^{G} \theta_{n,g} \times \beta_{g,c}$, which is the addition of probabilities over all of the genres. With this equation, I can measure the preferences over all of the unsubscribed channels for each household customer. I will explain next how to infer the preferences over the sub-bundles from the preferences over channels.

5.2. COMPUTATION OF BUNDLE PREFERENCES

There are two different ways to think about how to compute customer preferences for subscribed cable TV program bundles.

Maximum value-based preference computation. Some customers construct subscription bundles, with sub-bundles for which there is one channel that they like very much. So even when there are some channels that a customer dislikes, the customer will still include that sub-bundle in the subscription bundle. Their valuation of the one channel may dominate their decision to include the sub-bundle. The customer's preference for this bundle is determined by the maximum preference over all the channels in the subscription bundle. Customer *n*'s preference for sub-bundle *b* is calculated based on all their preferences for channel *c* in sub-bundle *b* as $p(n,b) = \max_{c \in b} p(n,c)$.

Average value-based preference computation. Other customers create their subscription bundles from sub-bundles for which they generally like all of the channels that are included. There may not be one channel in the sub-bundle that a customer especially likes, but in general, all of the channels are good enough so that they together offer some degree of utility. A customer's preference for a sub-bundle like this will be determined by the average of the household's preferences over all channels in the bundle. Customer *n* 's preference of sub-bundle *b* is calculated based on the household's preferences over all of the channel *c* as $p(n,b) = \frac{1}{|b|} \sum_{c \in b} p(n,c)$, where $|b|$ is the number of channels in sub-bundle *b*.

6. EXPERIMENTAL STUDY

Next, I will validate my model against data provided by a cable TV service provider. Again, I define a TV viewing session as an instance when a customer watches the same channel for at least 15 minutes. I consider only those customers with less than 100 TV viewing sessions in a month; inactive TV viewers are excluded for this experimental study.

6.1. EXPERIMENT DESIGN

Since the data only represent a snapshot of subscription bundle-based TV viewing sessions over one month, there were no changes in customer subscription bundles that I observed. Subscription bundle changes were simulated for this experiment. For each customer, I hid one sub-bundle at random from a household's subscription bundle. For example, if a customer subscribed to $\{b_1, b_2, b_3, b_4\}$, I randomize the hiding of sub-bundle b_4, and treated b_4 as a candidate sub-bundle that the customer could choose from. So, if the household subscribed to sub-bundles$\{b_1, b_2, b_3\}$ and I am able to analyze its TV viewing history, then the task is to predict which one sub-bundle the household will choose from $\{b_4, b_5, b_6, b_7\}$, for which it has not yet subscribed. If my model works, it will predict that the household will select sub-bundle b_4 as the most probable choice of its next sub-bundle to add. When I remove bundle sub-bundle b_4 from the household's subscription, I also must change the household's TV viewing history as well, since in the dataset, the customer's TV viewing history will span all $\{b_1, b_2, b_3, b_4\}$. I remove the TV viewing sessions that have channels that are included in sub-bundle b_4. Thus, the customer's viewing history will be truncated.

The cable TV service provider requires that all of its customers must subscribe to no fewer than 3 sub-bundles when they construct an overall subscription bundle. So I only selected customers with 4 or more sub-bundle subscriptions. I also removed one sub-bundle from the subscriptions so that these customers will still have at least 3 sub-bundles in their overall bundle subscriptions. With at least 100 TV viewing sessions that have at least 3 sub-bundles after one has been hidden, my data set included approximately 7,000 customers overall. The reason is that most customers only have subscription bundles that include 3 sub-bundles.

6.2. BASELINE METHOD AND PREDICTION ACCURACY

I will use a method called MaxComb as the baseline to compare my model against. MaxComb predicts the sub-bundle contents that yield the maximum number of customers who will adopt it, after combining it with their existing subscription bundles. For example, if a customer has subscribed to $\{b_1, b_2, b_3\}$, we will assess four options: $\{b_1, b_2, b_3, b_4\}$, $\{b_1, b_2, b_3, b_5\}$, $\{b_1, b_2, b_3, b_6\}$ and $\{b_1, b_2, b_3, b_7\}$. If the number of customers who subscribe to $\{b_1, b_2, b_3, b_4\}$ is more than the number of customers who subscribe to the other three combinations, then MaxComb will give b_4 as its primary prediction.

I next will compare my model with the MaxComb method. In LDA, there is a pair of Dirichlet priors, the genre prior $\vec{\alpha}$ and the channel prior $\vec{\eta}$. For simplicity, I set all dimensions of $\vec{\alpha}$ to be α, and all dimensions of $\vec{\eta}$ to be η. For each value pair α and η, I ran LDA 10 times under this parameter setting, since LDA is an algorithm that involves randomization. For each run, the number of correct predictions is the number of customers with an identical predicted bundle and hidden bundle. I divide this by the total number of customers to obtain the *prediction accuracy*. The prediction accuracy of each pair of parameters α and η is based on the average of the prediction accuracies over the 10 runs. My results are shown in Figures 5 and 6.

Figure 5: Prediction Accuracy of LDA versus MaxComb for Maximum Value-Based Preference

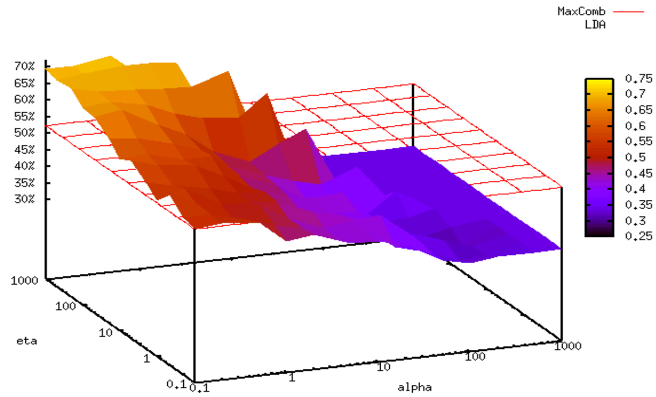
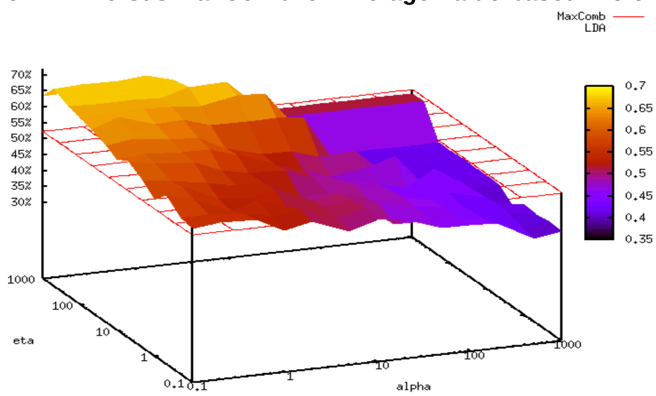


Figure 6: Prediction Accuracy of LDA versus MaxComb for Average Value-based Preference



The red grid shows the accuracy of the baseline method, MaxComb, which is about 52%. The colorful surface represents the performance of LDA, for different pairs of parameter settings. In Figure 5, I computed the preferences of unsubscribed sub-bundles based on the customers' maximum preference values. In contrast in Figure 6, the preferences of unsubscribed sub-bundles are computed based on the average preference values. In both figures, the best parameter settings for LDA significantly improved the prediction accuracy to approximately 70%, from about 52% for the MaxComb method, which does not consider the customer preferences.

These figures allow me to further note that the performances of LDA decrease with an increase in α. The larger the value of α is, the stronger will be the belief that the prior $\vec{\alpha}$ controls customer preferences $\vec{\theta}_n$. As the values on all dimensions of $\vec{\alpha}$ are set at α, when this belief is stronger, all $\vec{\theta}_n$ generated from the prior $\vec{\alpha}$ will be more similar: the differences among different customer preferences will diminish. Hence, performance will not be as good.

The prediction accuracies show that LDA performs much better in comparison to the baseline method, MaxComb. So it is important to model customer preferences, so the cable TV services provider will understands its customer better, and so its targeted upselling will be more effective. In fact, not only does performance matter, but it also is important to remember that a model that learns from viewing behavior also be interpretable by managers.

6.3. VALIDATING CASE STUDY: IDENTIFYING THE GENRES DISCOVERED BY THE LDA MODEL

My next step is to illustrate the genres that my LDA model was able to discover. With parameter settings $\alpha = 0.2$, $\eta = 200$, and the number of genres $G = 10$, the 3 channels with the highest probabilities associated with each genre are identified by LDA. The original 7 sub-bundles include *Ethnic Entertainment* (*EthE*), *Ethnic Infotainment* (*EthI*), *Education* (*Edu*), *Entertainment* (*Ent*), *Kids* (*Kids*), *Lifestyle* (*Life*) and *News* (*News*). Each of the original 7 sub-bundles is represented by at least one genre, with the exception that there are two genres for the *Kids* channels, and three genres for the *Entertainment* channels. See Table 3.

Table 3: LDA Modeling Notation for Modeling TV Customer Preferences

News	Sky News (News), BBC Word News (News), CNN (News)
Ethnic Infotainment	TVBS Asia (EthI), ONE (EthE), CTI TV (EthE)
Entertainment HD	Star World HD (Ent), FOX HD (Ent), AXN HD (Ent)
Ethnic Entertainment	ONE HD (ThE), tvN (EthE), E City (EthE)
Education	BBC Knowledge (Edu), History (Edu), Discovery Channel (Edu)
Lifestyle	E! Entertainment (Life), Food Network Asia (Life), BBC Lifestyle (Life)
Entertainment Normal	Star World (Ent), FOX (Ent), AXN (Ent)
Kids Junior	Disney Junior (Kids), Nick Jr (Kids), BabyTV (Kids)
Kids	Disney Channel (Kids), Nickelodeon (Kids), Cartoon Network (Kids)
Music and Anime	MTV SEA (Ent), Animax (Ent), Channel [V] (Ent)

The two *Kids* genres cover children at different ages. Channels in the *Kids Junior* genre are more for younger children and toddlers, while channels in the main *Kids* genre are for older children or even teenagers. This is also the same for the *Entertainment* sub-bundle. Since some viewers just like watching MTV and anime TV, channels about *Music and Anime* are separated from the rest of the *Entertainment* channels.

Another interesting finding is that the *Entertainment HD* and *Entertainment Normal* genres contain almost the same set of channels, except that one is in high definition, and the other is in normal screen resolution. This shows that people who have HD subscriptions probably <u>never</u> watch the normal resolution channels with the same content. In addition, for households that do not have a subscription to the *Entertainment HD* sub-bundle, they will just keep watching the normal resolution channels. This is what created two genres with the same content but different screen resolutions. This shows very well that the concept of a *topic* in the original definition of LDA is very powerful: it is broader than the concept of *genre* and it is able to convey a more nuanced understanding for the corporate sponsor of its customers' behaviors. It also helps to validate the logic of LDA's application in this context. In closing, I note that an LDA topic in the cable TV viewing context can refer to a program genre, or to some other similarities that are shared among channels, such as channels of the same language, or channels with the same audience segment, or channels with the same levels of violence or coarse language, and so on.

7. CONCLUSION AND WORK-IN-PROGRESS

With the above results, I conclude that, first of all, it is important to model customer preferences in cable TV services, for the purpose of targeted upselling. It is also effective to use topic models like LDA to model customer preferences, and prediction accuracy can be improved by as much as 30% to 40% compared to when no preferences is taken into consideration.

I currently have work-in-progress to extend this research. I have not addressed the issue of price yet, and only have studied the customer preferences. There are many add-on channels that require individual subscriptions in cable TV services. Usually, these channels are not uniformly priced. So there is a need to study how customer preferences are related to prices. For example, would it be possible for me to identify *incremental willingness-to-pay* for add-on channels? If so, which channels would be accessible at different levels of customer willingness-to-pay? I will try to build a similar model to take prices into the consideration, and recommend both channels and sub-bundles to the customers. This problem will be interesting and valuable to solve. It will also create the potential for the corporate sponsor to upsell to more customers and generate higher ARPU, since it will be more likely for customers to add on some extra channels than extra sub-bundles.

The other problem that I have not yet fully addressed is related to customer channel switching behavior. My LDA model does not consider the sequence of the program descriptors (words) in an effective way yet. Its underlying assumption is that program genres (as the topics of the words) are drawn independently from the customer's TV viewing genre preference (the topic preference). This needs further investigation though. When a viewer switches to another channel, there will be a correlation between the channel the viewer already has watched and the next channel. So a model that represents both customer viewing and switching behaviors is necessary to capture such correlations.

The third interesting problem is about TV viewer identification within a household. A model that explains switching behavior assumes that the viewer watching the channel before switching and the viewer watching the channel after a switch occurs will be the same person. This may not be true in a household though. Households consist of multiple family members. When another family member comes, there is some likelihood that the household will switch to a channel that they like – representing their joint preferences, as opposed to the singular preferences of one family member. This kind of switching behavior is not captured in my current model. If such switching behavior caused by the participation of different viewers can be identified, I will be able to model the different individual preferences in the same household. This may also allow me to offer upselling strategies that are more micro-targeted, and not just focused on the overall preferences of the entire household.

There is a lot of interesting research work to do to address the business problem of targeted upselling in cable TV services from the perspective of recommender systems and big data analytics. I can tap into existing techniques that are helpful in making the recommendations for channel upgrades more accurate for the corporate sponsor, and I also am in a unique position to be able to create new techniques that will be interesting for my research setting, as well as much broader scientific interest. I am working toward the development of solutions that cable TV service providers can apply to improve their capabilities for customer relationship management, which will be beneficial to their business and provide a basis for potential commercialization of the related new algorithms and methods.

8. ACKNOWLEDGMENT

This research is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office, Media Development Authority (MDA).

9. REFERENCES

Blei, D.M., Ng, A.Y., and Jordan, M.I. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, January 2003.

Blei, D.M. Topic Modeling and digital humanities. *Journal of Digital Humanities*, 2(1), 2013.

Chang, R.M., Kauffman, R.J. and Son, I. Consumer micro-behavior and TV viewership patterns: Data analytics for the two-way set-top box. *International Conference on Electronic Commerce*, 2012.

Chang, R., Kauffman, R.J. and Kwon, Y.O. Understanding the paradigm shift to computational social science in the presence of bid data. *Decision Support Systems*, 2013, in press.

Chuang, J., and Sirbu, M. Optimal bundling strategy for digital information goods: network delivery of articles and subscriptions. *Information Economics and Policy*, 11, 2, 1999, 147-176.

Davenport, T.H., and Harris, J.G. What people want (and how to predict it). *Sloan Management Review*, 50, 2, 2009, 22-31.

Desrosiers, C., and Karypis, G. A comprehensive survey of neighborhood-based recommendation methods. *Recommender Systems Handbook* , 2011, 107-144.

Friedman, W. U.S. Internet TV users soaring. *MediaDailyNews*, April 2, 2013.

Hoffman, M.D., Blei, D.M., Bach, F.R. Online learning for latent Dirichlet allocation. *Neural Information Processing Systems*, 2010:856-864.

Hofmann, T. Probabilistic latent semantic indexing. *ACM SIGIR Conference on Information Retrieval*, 1999, 50-57.

Iyengar, R., Ansari, A., and Gupta, S. A model of consumer learning for service quality and usage. *Journal of Marketing Research*, 44, 4, 2007, 529-544.

Kim, S. Firm characteristics influencing the extent of electronic billing adoption: an empirical study in the U.S. telecommunication industry. *Telematics and Informatics*, 19, 3, 2002, 201-203.

Koren, Y. Collaborative filtering with temporal dynamics. *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2009, 447-456.

Koren, Y., Bell, R.M., and Volinsky, C. Matrix factorization techniques for recommender systems. *IEEE Computer 42*, 8, 2009, 30-37.

Koren, Y., and Bell, R.M. Advances in collaborative filtering. *Recommender Systems Handbook* , 2011, 145-186.

Lee, H., Lee, S., Kim, H., and Bahn, H. Personalized recommendation schemes for DTV channel selectors. *IEEE Transactions on Consumer Electronics*, 52, 3, 2006, 1064-1068.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11, 2010, 19-60.

Martinez, A., Arias, J., Vilas, A., Garcia-Duque, J., and Lopez-Nores, M. What's on TV tonight? An efficient and effective personalized recommender system of TV programs. *IEEE Transactions on Consumer Electronics*, 55, 1, 2009, 286-294.

Melville, P., and Sindhwani, V. Recommender systems, *Encyclopedia of Machine Learning*, 2010.

Salakhutdinov, R., and Mnih, A. probabilistic matrix factorization. *Neural Information Processing Systems*, 2007.

Smyth, B., and Cotter, P. A personalized TV listing service for the digital TV age. *Knowledge-Based Systems*, 13, 2, 2000, 53-59.

Su, X., and Khoshgortaar, T.M. A Survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009.

Ungar, L., Foster, D., Andre, E., Wars, S., Wars, F.S., Wars, D.S., and Whispers, J.H. Clustering methods for collaborative filtering, AAAI, 1998.

Verizon. 2011. Annual report. Available at www22.verizon.com/investor/DocServlet?doc= vz_ar_final_2011.pdf.