Demystifying the Path to Purchase: A State Space Modeling Approach

Nachiketa Sahoo, Chrysanthos Dellarocas, Shuba Srinivasan

School of Management, Boston University

Abstract

In this study we explore the effect of customer characteristics and marketing activities on changes in customers' online and offline shopping behavior. We build a state space model using a longitudinal dataset covering a broad set of interaction points between customers and a North American Specialty Retailer. The model finds a small number of states of customers including "actively shopping", "idle" (from the retailer's perspective), and "online-browsing and offline shopping" states. Despite the prevalence of "idle" state we find a significant number of state transitions. Such state transitions are explained using multinomial logistic regression with several customer specific variables, such as age, gender, distance to store; and several marketing variable such as email campaigns sent to the customer, catalogs sent to the customer, number of promotions available and whether it is a holiday. We find that Holidays have one of the largest effects, at 58% increase in probability, on customers' switch to actively-shopping mode. However, Effects of Catalogs and Email marketing events are quite modest at 1.5% and 0.7%. People who are middle-aged (35—54) have a higher probability of moving to active shopping mode than both the younger (18—34) and older (>55) customers. Similarly, we find that Female customers have a higher chance of directly moving into one of the more active shopping states from the idle state than their male counterparts.

Demystifying the Path to Purchase: A State Space Modeling Approach

Nachiketa Sahoo, Chrysanthos Dellarocas, Shuba Srinivasan

School of Management, Boston University

Introduction

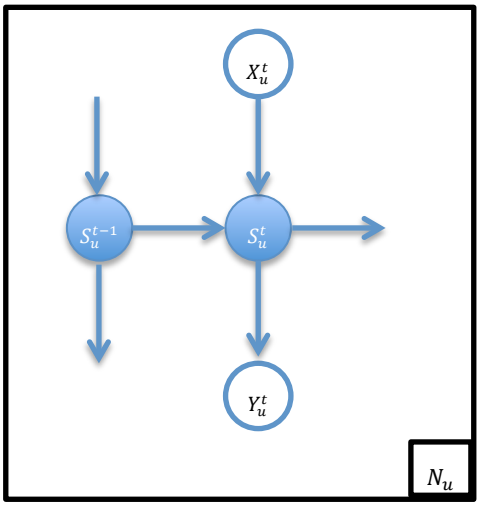
Customer behaviors are always changing. Not all behaviors are equally valuable to a firm. E.g., customer purchases can bring in profits where as bad reviews about a product or service can dissuade others from buying, hurting the bottom line. How to respond to such changing customer attitudes is of immense interest to the firms. The availability of large quantities of micro-level longitudinal data has opened up the possibility of studying patterns of user behaviors. Although, there are a number of dynamic models for aggregate measurements of customer behaviors, the study of user-level behavior changes and exploration of factors that affect them in the retail context are relatively rare. This work presents one such study. The goal of the current research is to identify temporary customer behaviors from a variety of data collected from CRM systems and examine the factors that explain changes in such user behaviors.

Data description

The dataset for this study is collected from a large North American Specialty Retailer. It includes data on marketing activity of the firm, customers' website activities, customers' purchases, as well as their post purchase behavior, such as product returns and product reviews. The interaction between the firm and the customers were tracked from July 2010 to June 2012. The dataset used in this study included approximately 24.8K customers.

Model

The state space model is described in Figure 1 using Bayesian Network.



S_u^t: The state of customer u in time period t. This is a latent variable.

X_u^t: The variables used to explain the probability of switching from S_u^{t-1} to S_u^t

Y_u^t: The output variables used to characterize user behavior in a week. It contains the observed online and offline behavior of the customer.

Figure 1 State space model of changing customer behavior, presented using Plate notation.

The length of each time period was set to one week. The output variables (Y_u^t) are measurements of various behaviors of the customer u in a given week t, such as the number of products browsed, number of searches conducted on retailer website, number of purchases made online and offline. Since these observations are non-negative numbers; they are modeled using state specific Gamma distributions.

From our preliminary analysis we find that customer behaviors changes from week to week. One question of managerial interest is what are the factors that explain such change? To investigate this, we model state transition as a

multinomial logistic regression with the following explanatory variables: number of marketing communications sent to customer (catalogs, e-mails), age-group of the customer (young, middle aged, old), gender, distance to nearest store, whether the week is a holiday, and number of promotions available in the week.

The parameters of the model are estimated using a version of Expectation Maximization algorithm used to fit Hidden Markov Models (HMM). The primary extension is the estimation of logistic regression parameter for state transition model instead of the fixed transition probability matrix for HMM. Note that this extension allows the effective transition probability matrix for each consecutive pairs of weeks for each user to be different depending on the values of the explanatory variables.

Results

States →	1	2	3	4	5
Browse	27.26	27.80	0.00	21.76	10.67
Searches	2.08	0.00	0.00	2.09	0.00
Store Pur	0.64	0.58	0.21	0.93	0.40
Online Pur	0.98	1.60	0.00	2.14	0.00
State prob	1.8%	3%	88%	0.85%	6.7%

Table 1 State description through average activities in them

Trans	1	2	3	4	5
1	0.0180	0.0159	0.8869	0.0054	0.0738
2	0.0066	0.0218	0.8860	0.0058	0.0798
3	0.0008	0.0851	0.7725	0.0524	0.0892
4	0.0180	0.0159	0.8869	0.0054	0.0738
5	0.0119	0.1080	0.8701	0.0087	0.0013

Table 2 The base state transition probability

State descriptions

The descriptions of the states and the state transition matrices are given in Tables 1 and 2. From the Table 1 the states of the customers can be interpreted as follows.

State 1 represents a generally active customer who is browsing products online, searching for products on company website, make some purchases in store and more purchases online.

State 2 is another state of the customer with high rates of purchase, however, little online searching. This represents a customer who is locating products through some other means, browsing products online, and buying them either in store or online.

In contrast, state 3 has almost no online activity, and smallest rate of in store purchase activity. This can be thought of as the idle state. In any random week this is the most likely state of a customer.

The state 4 is similar to state 1, however it has twice as much purchase activity. This makes state 4 the state with most purchase activity. However, this is the rarest state for a customer to be in.

State 5 represents a state with moderate online browsing and in store purchase activity.

State transitions and their explanations

The base level of state transition probability is computing using the parameters of the logistic regression for a Young Male customer, receiving average number of marketing communications, living at a distance from store that is average of all the customers, during a non-holiday week when the number of promotions available is average of the number of promotions available on all the weeks in the data collections period.

Transition from the idle state (state 3)

Examination of the transition probability matrix for the base case reveals that users leave the idle state either by moving into the state 5 (moderately active) or via state 2 (actively shopping without searching). The coefficients of the multinomial regression model (Figure 2) suggest that receiving a catalog in the week is associated with 1.5% less chance of staying idle; receiving an E-mail is associated with 0.7% less probability of staying in the idle state.

Middle-aged customers have highest chance of leaving idle state is via state 2 (shopping state), where as older customers have higher chance of leaving state 3 via state 5 (moderate browsing). We also find that the Female customers have a higher chance of leaving idle state via state 2 or 4—both with high rate of purchases.

Unsurprisingly, holidays have a large effect on moving people from idle state to shopping state. The probability of switching from idle state to the state 2, and active shopping state, increases by 58% if it is a holiday week.[1]

Transition from moderately active state (state 5)

The state 5 is interesting because it acts as a route out of the idle state. However, a customer in state 5, although has an 11% chance of going to a shopping state, has 87% chance of going back to the idle state. Therefore, a question of practical interest is what are factors that can influence customers' switching one way or the other? Note that state 5 is a state where the customers have least probability of staying in. We again see effects of age, gender, distance, which are similar to the ones we found in the previous paragraph.

Transition to the state 4

The state 4 is the one in which a customer is making most purchases. However, it is one of the rarest states. So, a natural question of interest is how customers get into this state. From the state transition matrix we can see that the highest inflow into the most active state occurs from state 3, or the idle state! If we examine the coefficients of multinomial transition model we see that the variables that is most associated with this switch is occurrence of a holiday. There is a 52% increase in the probability of this state transition.

Next Steps

The presented work is part of a larger ongoing investigation of the effect of Paid, Owned and Earned media on different types of customer interaction. One important type of the Owned medium is customer review and ratings, since users are increasingly relying on online customer reviews to make purchase decisions. We are currently investigating the factors that affect generation of Owned media using the model presented here.

[1] Since coefficients of a multinomial regression model represent changes in log odds ratio for each unit change in variable it has to be appropriately transformed to get changes in probabilities. Small coefficients can be used directly since $\log(1 + x) \approx x$, when x is small.

To -->	State 1	State 2	State 3	State 4	State 5		Variable definitions
From State 1	-1.4061	-1.533	2.4993	-2.6061	0	1	
	-0.001	0	-0.0034	-0.0035	0	Catalog	Number of catalogs received
	-0.0003	-0.0003	-0.0019	0	0	Email	Number of Emails received
	0.0966	0.229	0.2397	0.3347	0	MiddleAged	Is between 35 and 54?
	0	0.2578	0.312	0.3941	0	Old	Is older than 55
	0	-0.1656	-0.4209	-0.2287	0	Female	Is Female?
	0	0	-0.0003	0.0002	0	Distance to store	Distance to nearest store in miles
	0	0	-0.0316	0	0	Holiday Week	Is this a holiday week?
	0.0022	0	-0.0006	0.0026	0	Number of promotions	Total number of promotions available in this week
From State 2	-2.4927	-1.2917	2.4176	-2.6224	0	1	
	0.0006	-0.0026	-0.0027	-0.0053	0	Catalog	
	0	0	-0.0016	0	0	Email	
	0.2612	0.5735	0.2514	0.3798	0	MiddleAged	
	0	0.5751	0.3011	0	0	Old	
	-0.1779	0	-0.6264	0.373	0	Female	
	0	0	-0.0002	-0.0173	0	Distance to nearest store	
	0	0.4634	-0.1283	0.174	0	Holiday Week	
	0.0096	-0.0134	0	-0.0085	0	Number of promotions	
From State 3	-0.7101	0.4086	2.2068	-0.4901	0	1	
	0	-0.3651	-0.0156	-0.0099	0	Catalog	
	0	-0.0186	-0.0068	-0.0063	0	Email	
	0	0.8346	0.4636	0.6882	0	MiddleAged	
	0	0.382	0.9469	0.762	0	Old	
	0	0.7312	0.2169	0.3147	0	Female	
	0	-0.0093	-0.0022	-0.0135	0	Distance to nearest store	
	0	0.5414	0.0803	0.4166	0	Holiday Week	
	0	-0.0358	0.002	-0.0272	0	Number of promotions	
From State 4	-1.4061	-1.533	2.4992	-2.6062	0	1	
	-0.001	0	-0.0034	-0.0035	0	Catalog	
	-0.0003	-0.0003	-0.0019	0	0	Email	
	0.0967	0.2289	0.2402	0.3345	0	MiddleAged	
	0	0.2567	0.3118	0.3937	0	Old	
	0	-0.1651	-0.4209	-0.2286	0	Female	
	0	0	-0.0003	0.0002	0	Distance to nearest store	
	0	0	-0.0313	0	0	Holiday Week	
	0.0022	0	-0.0006	0.0026	0	Number of promotions	
From State 5	2.2725	4.5219	6.5883	2.0897	0	1	
	-0.0116	-0.0241	-0.0199	-0.0378	0	Catalog	
	-0.0025	-0.0097	-0.0065	-0.0265	0	Email	
	0.4356	0.5815	0.4559	0.2596	0	MiddleAged	
	0.2486	0.518	0.9702	0.5187	0	Old	
	0.2033	0.2627	0.5303	0.6242	0	Female	
	-0.0007	-0.0057	-0.0014	0	0	Distance to nearest store	
	0.206	0.1947	0.2368	0.3454	0	Holiday Week	
	-0.0497	-0.0658	-0.0561	-0.0698	0	Number of promotions	

Figure 2 The coefficients of multinomial logistic regression for state transition. Only the coefficients at p-value < 0.05 are shown.