Paper 2043 -2014

Managing Opt-Out Risk

Carol Li, Gilt Groupe, New York, NY, USA

ABSTRACT

Email is an important marketing channel for digital marketers. Marketers can stay connected with the subscribers and continue to engage them with relevant content as long as they are still subscribed to email communication.

Every marketing email received can be an opportunity to transact for one user. At the same time, it can be an opportunity to unsubscribe for another user. Monitoring the unsubscribe activity is important as it speaks to the health of the email program. Is the content relevant? Are emails going out at the right time, right cadence? Even when unsubscribe rate is being maintained at a steady stage, knowing who are the users with high opt-out risk will allow marketers treat them a bit differently and keep them staying longer, thus optimizing the overall contact strategy.

INTRODUCTION

The opt-out risk models were developed using logistic regression in SAS®. This paper first describes the model development process. Then, it gives an example of how the model is being used in combination with estimated life time value for frequency test.

GATHERING PREDICTORS

Predictors were sourced from four main areas: customer purchase behavior, email activity, traffic data, and user demographic information. All information was summarized at the user level, and “cleaned” for model development. PROC VARCLUS was used to eliminate redundant variables:

```
proc varclus data=&InputData maxeigen=1 short;
 var &PredictorVar;
run;
```

These are basic specifications for PROC VARCLUS. MAXEIGEN specifies that when choosing a cluster to split, VARCLUS should choose the cluster with the largest second eigenvalue, provided that its second eigenvalue is greater than the MAXEIGEN= value, which is 1 in this example. SHORT suppresses display of the cluster structure, scoring coefficient, and intercluster correlation matrices. (SAS 9.1.3 online documentation SAS/STAT). Below is part of the VARCLUS output that tells you number of clusters created and which cluster each variable is assigned to.

Output Sample

Cluster	Variable	R-squared with Own Cluster	R-squared with Next Closest	1-R**2 Ratio
Cluster 1	Variable 1	0.8264	0.1148	0.1962
	Variable 2	0.8264	0.0867	0.1901
Cluster 2	Variable 3	0.6805	0.1076	0.3580
	Variable 4	0.6202	0.0371	0.3945
	Variable 5	0.2090	0.0220	0.8087
	Variable 5	0.4479	0.0104	0.5579
Cluster 3	Variable 6	0.7723	0.0516	0.2401
	variable 7	0.7723	0.0036	0.2285

Output 1. Output from a PROC VARCLUS Statement

A variable selected from each cluster should have a high correlation with its own cluster and a low correlation with other clusters (Logistic Regression Modeling, pages 56-57). VARCLUS significantly cuts down the number of predictors and adds efficiency to the model development process.

OPT-OUT RISK MODEL

Model target, opt-out event, was defined looking at the users who had been receiving marketing emails 30 days ago, but unsubscribed from email communication as of current state. The target can be recoded into binary variable, where 1 represents opt-out and 0 represents non opt-out. Logistic regression (PROC LOGISTIC) is used to model the probability to opt-out (or unsub, which is short from “unsubscribe”):

```
proc logistic data=Dev_data descending;
 model unsub = &SelectedVar / selection=stepwise sle=0.01;
run;
```

The linear logistic model has the form

$$logit(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta' x$$

where α is the intercept parameter and β is the vector of parameters. By default, PROC LOGISTIC models the probability of lower response level. Specifying DESCENDING in the PROC statement LOGISTIC models the probability of higher response level. SLE specifies the significance level of the score chi-square for entering an effect into the model in the FORWARD or STEPWISE method (SAS 9.1.3 online documentation SAS/STAT). Below are partial result outputs:

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	0.00883	0.0409	0.0467	<.0001	
Variable 1	1	-0.5217	0.00663	6193.9255	<.0001	-0.2428
Variable 2	1	-0.942	0.0356	700.4325	<.0001	-0.1572
Variable 3	1	-0.245	0.0189	167.8597	<.0001	-0.0636
Variable 4	1	0.1803	0.0165	119.589	<.0001	0.045
Variable 5	1	-0.5073	0.0241	443.0036	<.0001	-0.1109
Variable 6	1	-0.3945	0.0236	279.0381	<.0001	-0.0798
Variable 7	1	0.3506	0.0241	212.2804	<.0001	0.0563

Association of Predicted Probabilities and Observed Responses

Percent Concordant	66.3	Somers' D	0.36
Percent Discordant	30.3	Gamma	0.373
Percent Tied	3.4	Tau-a	0.021
Pairs	9658620	c	0.68

Output 2. Partial Model Output from PROC LOGISTIC

To evaluate the quality of model, p-value for each variable should be less than 0.05, which confirm the variable coefficient does not equal to zero. We also look for a model with reasonable c stat, which range from 0.5 to 1. Higher c stat indicates a better model prediction.

Once the final model is developed, we validated the model using both in-time and out-of-time validation. We compared the Cumulative Gains Chart to evaluate the model lift and stability.

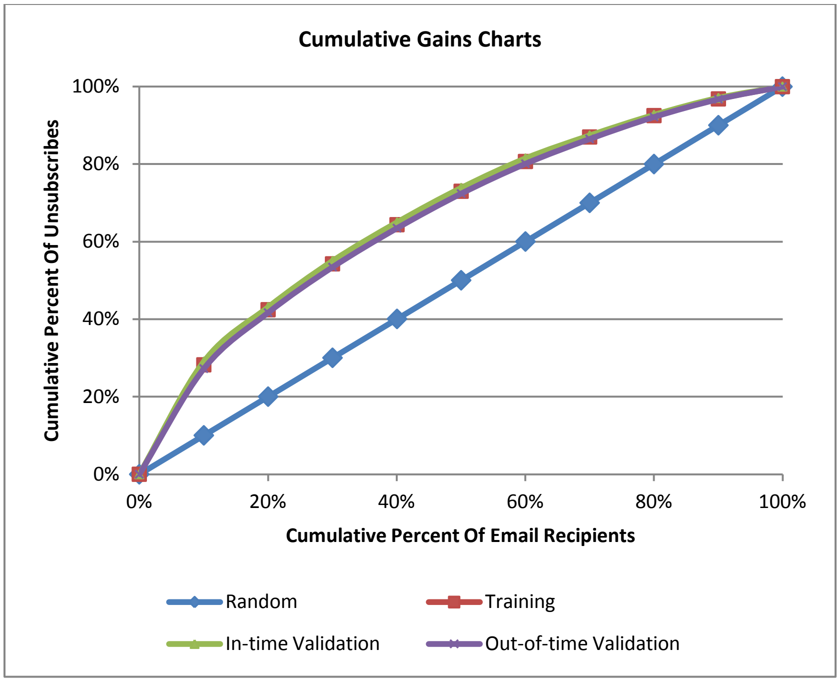

Figure 1. Cumulative Gains Chart

Based on the chart above, selecting the top 40% email recipients by LOGISTIC model score, we will capture 63% of the opt-out users. This is the group of users we want to test with different email tactics to improve unsubscribe rate. Our first attempt is to run an email frequency test on these users to see if reduced email frequency can help reduce their unsubscribe rate.

UTILIZING THE OPT-OUT RISK MODEL

Based on past email frequency test results, test users ended up with lower conversion rate, especially in the shorter term. Therefore, we revised our testing strategy by taking the following steps. First identify the email that generates the least amount of traffic and revenue per week. That email will be dropped for high opt-out risk users, so they are receiving one less email per week. Lastly, we overlay estimated life time value with opt-out risk score when defining target for test. This is an attempt to identify users for whom every email they received is not so much an opportunity to transact, but an opportunity to unsubscribe. We launched a frequency test for recipients with high opt-out risk, but low estimated life time value. As a result, the unsubscribe rate for the test users decreased by 5%, while the conversion rate stayed flat.

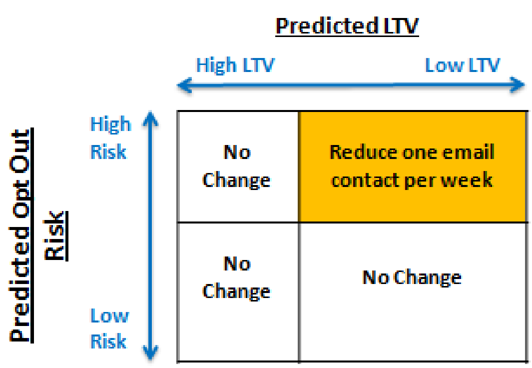

Figure 2. Overlaying predicted opt out risk with predicted life time value

In addition to frequency testing, this opt-out risk approach can also be used to guide audience selection for various incremental email activities. Proactively offer users the opportunity to adjust their email frequency, or remind them about other ways to stay in the loop, i.e., mobile, social channels.

CONCLUSION

The goal of developing an opt-out risk model is to improve email unsubscribe rate, and, ultimately, to maintain a healthier email program. Modeling is just the very first step of the task. For the model to be successful, it needs to be utilized to the fullest through the business strategies. We should combine the model as a targeting tool with other available tactics to continue engage users at the right time, right channel with relevant contents.

REFERENCES

SAS Institute Inc. 1999. Logistic Regression Modeling. Course Notes. Cary, NC: SAS Institute Inc.

SAS Institute Inc. SAS Online Documentation 9.1.3. Cary, NC, USA. 2003.
http://support.sas.com/onlinedoc/913/docMainpage.jsp

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Carol Li
Gilt Groupe
2 Park Avenue, 4th Floor
New York, NY, 10006
Email: caroll@gilt.com