

Combining Segmentation and Targeting Modeling: A Winning Formula

Aaron Lai, Howell Kiung, and Oiching Yeung, InfoAtlas, Inc., Oakland, CA

ABSTRACT

Customer Segmentation is used to improve customer experience based on his/her demographics or lifecycle. The idea behind demographic segmentation is to provide a comprehensive solicitation strategy based on their needs. Many companies, such as Acxiom, have rushed to develop sophisticated segmentation programs for this. They need to analyze all attributes in their demographic databases. Targeting Model is a technique that ranks the response likelihood of customers/prospects. In this paper, we will explain how did we utilize both technique to significantly improve response rate, as well as a better understanding of the customers or prospects. Proc Fastclus and Proc Logistics will be used.

INTRODUCTION

Some facts about predictive analytics software¹

- Sales in 2004: \$2.2B
- Sales of total core analytics market: 25%
- Project growth rate (compounded annual growth 2003-2008) through 2008: 8%

Do you want to be part of it?

Many database marketers / modelers consider segmentation/clustering and target modeling are two different tools: the former one is used to put customers into different market segments (a marketing term) or cluster (a statistics term) while the latter one is used to identify whom will be responsive to our marketing offer. Targeting model is a technique that ranks the response likelihood of customer/prospects. In this paper, we will explain how did we utilize both techniques to significantly improve response rate, as well as a better understanding of the customer or prospects. We will use a typical example (a consumer finance product) to illustrate our case.

CUSTOMER SEGMENTATION

Very often market segmentation is based on qualitative research via market research, focus group, or other questionnaire-based research. Alternatively, some people will attack this problem with demographic data from census, together with behavioral data from POS (point-of-sales) or mail/online-order data, to form segments. The idea behind demographic segmentation is to provide a comprehensive solicitation strategy based on customers' needs. For example, many companies such as Acxiom and Claritas, have rushed to develop sophisticated segmentation programs for this. One of the biggest problems they are facing is the number of attributes they have to analyze and the number of segments needed. An attribute is a particular piece of information about that particular customer. For instance, age is an attribute, number of minutes that a customer stay in a shop is an attribute, how much a customer spends in a single shopping is also an attribute. Many segmentation program will result in too many segments. It may be quite confusing to the marketers. In Acxiom's Prosonicx system, there are 70 segments; in the P\$YCLE system provided by Claritas, there are 42 segments. Therefore, it is not a small task to get it right.

SEGMENTATION LOGIC AND FORMULATION

We begin by developing of a new customer segmentation logic that leverage the following information:

- Age of the Customer
- Marital Status
- Home Ownership
- Present of Children
- Annual Income
- City vs. Rural

Each of the elements above are segmented into sub-groups:

¹ IDC, January 2004, quoted in Computerworld, June 13, 2005

| Age | Children | Marital Status | Income | Homeownership? | City vs. Rural |
|----------|----------------|----------------|--------------|----------------|----------------|
| 18 to 25 | New Born | Single | Low | Yes | City |
| 26 to 30 | Elementary | Married | Low Middle | No | Near City |
| 31 to 35 | High School | Divorced | Middle | Both | Suburban |
| 36 to 40 | College | Partner | Upper Middle | | Small Town |
| 41 to 45 | No Children | | Affluent | | Rural |
| 46 to 50 | Grand Children | | Well-off | | Farm Land |
| 51 to 55 | | | Retired | | |
| 56 to 60 | | | | | |
| 61 to 70 | | | | | |
| 71 plus | | | | | |

To perform the needed segmentation, we have used the PROC FASTCLUS to develop our segmentation logic, and used PROC CLUSTER to refine the hierarchy. There are altogether 7 high-level clusters/major groups, 35 mid-level cluster/sub-groups, and 125 micro-segments. Each of these seven high-level clusters is used to build customized marketing message. The segments we have identified are:

1. Young & Well Educated — those are young college-educated young people (either single or couple with no kids) who enjoy life and refined things.
2. Young & Less Educated — those are high school graduates who do not have a lot of free time and care most about immediate gratifications or needs.
3. Last Life Cycle — those are senior people with planning for passing wealth and late-year arrangement.
4. Early Family — those are relatively young couples or partners with small kids or going to have kids in near future. They will tend to care most about children-related matters.
5. Late Family Household — those are couples or divorced people who have grown-up children. They would start planning for their own activities.
6. Mid-Age Business Worker — those are mid-career office workers. They are single and educated. As a result, they can afford buying more expensive and sophisticated things such as collectable.
7. Mid-Age Hardhat — those are single workers in a more low level jobs. They have limited spending power and usually spend money on those with immediate needs.

TARGET MODELING

The basic tool of target modeling is a statistical technique called logistic regression. In classical regression, we assume that the dependent variable (the attribute that we want to predict) is a continuous variable (i.e. it will have 1, 1.1, 1.253 ... etc. instead of just 1, 2, 3 ... etc.); in logistic regression, or any variation of it, assumes that the dependent variable is discrete (i.e. yes/no, 1/0, black/white ... etc.). In this particular case, we use the simple logistic regression to predict whether a prospect will respond to our campaign. The predicting attributes we have used fall into the following four categories: credit bureau data, demographic, relationship history, and past mailing history.

Some of the attributes in our model are:

| Credit Bureau | Demographic | Relationship History | Past Mailing History |
|---------------|-------------|----------------------|----------------------|
| Inquiries | Income | Number of Products | Times Mailed |
| Delinquency | Online | Type of Products | Past Response |
| Balances | | Profit | Tenure |
| Payments | | | |
| Trends | | | |
| Ratios | | | |
| Usage | | | |

RESPONSE MODEL FORMULATION

We have used the stepwise option of the PROC LOGISTIC to implement our model. The usual approach is to build models based on the acquisition channel: direct mail, email, walk-in/ATM, and inbound/outbound call. Each model groups population into 10 subgroups (deciles). The top six to eight deciles are mailed based on the response rate and dollars return per cell. In a direct mail campaign, we often need to try different combination of factors, ranging from the color of the envelope to the list-broker used. By assigning different tracking codes to different cells, we can find out what kind of marketing programs are working or not.

In this case, we use the clustering techniques to group our customers into different segments, and then use the logistic regression technique to build a more effective scorecard. As a result, we have 70 cells (seven segments * ten deciles). We believe that acquisition channel is less relevant now because of the spread of new technology such as Internet.

MODEL OBJECTIVES / LEARNINGS

Using the combination of segmentation and target modeling, we are able to expand the mailing universe and increase response rate due to customized message. We have developed the following matrix to help us making mailing decision.

| Return on Investment (ROI) | | Lifecycle Segments | | | | | | |
|----------------------------|------------------------------|-----------------------|-----------------------|-----------------|--------------|-------------|-------------------------|-----------------|
| | | Young & Well Educated | Young & Less Educated | Last Life Cycle | Early Family | Late Family | Mid-age Business Worker | Mid-age Hardhat |
| Response Deciles | Deciles 1 (Most Responsive) | +++ | +++ | + | + | ++ | ++ | + |
| | Deciles 2 | +++ | ++ | + | | ++ | ++ | + |
| | Deciles 3 | ++ | ++ | + | | + | + | + |
| | Deciles 4 | + | + | + | | + | + | |
| | Deciles 5 | + | + | - | - | | | |
| | Deciles 6 | | | | - | - | | - |
| | Deciles 7 | | | | - | - | | - |
| | Deciles 8 | - | - | | -- | -- | | -- |
| | Deciles 9 | -- | -- | - | --- | --- | | -- |
| | Deciles 10 (Less Responsive) | --- | --- | -- | ---- | ---- | | -- |

Every company should have its own cost of capital or hurdle rate to assess the economic benefit of doing a direct mail campaign, as indicated by the calculation of ROI (Return on Investment). The above table illustrates the use of ROI to make mailing decision. The “+” and “-“ symbol shows how high the ROI for a specific cell. We should only mail to those cells under shades because only those cells give you positive economic value.

CONCLUSION

Overall, lifecycle segments have proven to be an effective booster to target modeling. That additional segmentation information can help marketers to refine their messages so that customized offers could be made. We have “segmented” target models consistently outperform pure response model, or pure segmentation model. In our experience, the marginal lift is about 10% to 15% for major segments.

An additional advantage of combining segmentation and target modeling is that we can factor in our response history. If your company name is mostly associated with the sub-prime market, prime market customers may not be that receptive to your mailing campaign, even if you are able to provide a competitive offer. With historical response rate incorporated into to your segmentation model, you could model this part of consumer preference directly. You may also discover some more cost-effective approach to business development.

ENDNOTE

Aaron is an advisor to InfoAtlas while both Howell and Oiching belong to the technical team of business analytics solution development.

This is a paper prepared for WUSS 2005.

CONTACT INFORMATION

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

Author: Howell Kiung

Company: InfoAtlas, Inc.

Address: 1441 Franklin Street, Suite 204, Oakland, CA 95817, USA

Email: howell@infoatlas.com

Web: www.infoatlas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® Indicates USA registration.

Other brand and product names are trademarks of their respective companies.