Net Lift Modeling vs. Propensity Modeling for Skewed Data

Zhen Zhang[1], Lei Zhang[2], Kendell Churchwell[1], James Veillette[1]
[1]C Spire, 1018 Highland Colony Parkway, Ridgeland, MS 39157
[2]Mississippi State Dept. of Health, 570 E. Woodrow Wilson Dr., Jackson, MS 39216

Abstract
Propensity modeling has been extensively used in telecommunication companies to optimize marketing outcomes in cross sell and up sell campaigns, retention tactics, recruiting strategies, etc. Uplift modeling, on the other hand, is less familiar territory due to its complexity[1]. Net lift models are reported as superior in terms of maximizing return on investment by some practitioners[2,3], while others cautioned on its trade-offs and limitations[4]. Little was reported, however, on the comparison of these two techniques with respect to skewed data. Our research shows that for highly skewed data, while the net lift model produced much improved incremental sales rates compared to the traditional propensity model, the propensity model outperformed the net lift model in terms of number of incremental sales, due to its much larger segments.

Key Words: net lift model, propensity model, skewed data, campaign segment, lift, incremental sales.

1. Introduction

In today's marketing, data mining has become an essential tool to marketers to find the optimal customer targets for marketing campaigns such as cross sell, up sell or customer retention. Of the data mining tool kit, propensity modeling has been the most established application to predict who is likely to respond to a marketing campaign. Net lift modeling is also gaining momentum as businesses seek to maximize return on investment (ROI).

1.1 Propensity Modeling
Propensity modeling is also referred to as response modeling. The success of a direct marketing campaign is determined by response rate. Propensity modeling aims to improve response rates by identifying prospects that are more likely to respond to a campaign. A propensity model generates an estimate of the likelihood of response for each customer. To build a propensity model, prior campaign data is collected, along with customer features such as tenure, transaction history, pattern, etc., and positive response is marked as target. Various algorithms are available for building a propensity model. Popular algorithms include logistic regression, neural network and tree algorithms. Once a propensity model is built, a new list of customer data with their corresponding features is fed through the model to generate a propensity score – the estimate of likelihood to respond. Customers are then ranked according to the propensity score and a campaign segment is selected from customers with highest scores. It is easy to see that the response

rate from this model generated segment is likely to be higher than a segment randomly selected. The difference between these two rates is referred to as the model lift.

The quality of a propensity model is typically evaluated using a cumulative gain chart (Figure 1).

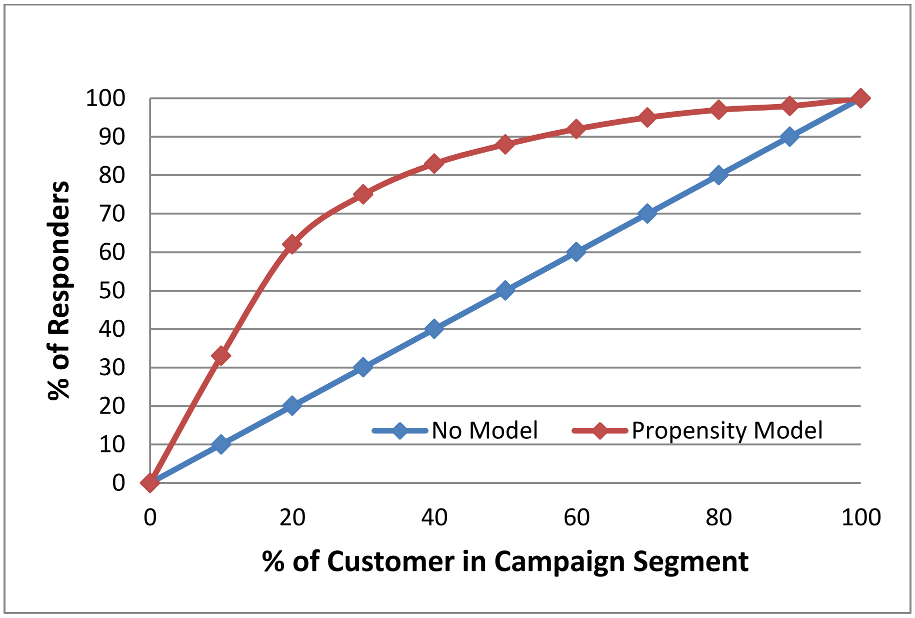

Figure 1: A cumulative gain chart shows the benefit of increased response in campaign segments generated by a propensity model.

The target of propensity modeling is "response" itself, with no discrimination of whether or not this customer is a self-selected responder (customers that would make a purchase without an incentive offer), or campaign-persuaded responder. Therefore, when evaluating campaign results, it is not convincing to credit all gain of responders to the campaign efforts. Of the model lift shown in Figure 1, a portion of the lift may be the self-selectors. Consequently, campaign effectiveness is undermined due to efforts and dollars spent on self-selectors.

1.2 Net Lift Modeling

In an effort to enhance campaign effectiveness and profitability, some practitioners turn to a net lift modeling approach[2, 5, 6]. Net lift modeling typically sees three categories of customer for a particular marketing campaign: self-selectors, campaign-persuaded responders, and non-response. Net lift modeling chose the campaign-persuaded responders as its target and thus aims at improving incremental sales rate. A net lift model produces a conditional probability of response. It has been reported that net lift model applications render a better ROI compared to propensity models[7,8].

1.3 Skewed Data Issue

Modeling practitioners often encounter situations where the category of the target field is imbalanced, sometimes extremely skewed. Skewed data refers to the situation where classification categories are not approximately equal. The class that has the greater percentage is called the majority class, whereas the class that has the smaller percentage is called the minority class. For instance, churn prevention is of great interest in the telecommunication industries, yet the binary target field is often as skewed as having the minority class being less than 5%.

While many studies were done on modeling imbalanced data[9], little was reported on the comparison of propensity models vs. net lift models with respect to skewed data. Since in our modeling practices we deal with skewed data daily, we are motivated to find the best practices for skewed data. Results presented in this paper are generated with real data.

2. Method

2.1 Variable Selection

First, we applied cross-sampling univariate correlation analysis to filter out unstable variables, and then multivariate selection is done to finalize the model set.

2.2 Propensity Modeling Algorithm

Logistic regression was used to build the propensity model:

$$P\,(Y=1|X_1,\dots X_i \dots X_n) = \exp(\beta_0+\beta_1 X_i) \,/\, [1+ \exp(\beta_0+\beta_1 X_i)]$$

2.3 Net Lift Modeling Algorithm

We adopted Larson's probability decomposition technique[2] that combines two propensity models to produce the net lift score. The second model penalizes the gross propensity score according to the likelihood that a customer is a self-selector. Let P_1 be the probability of purchasing given treatment, P_2 be the probability of receiving treatment given a purchase, and let P_{NL} be the conditional probability of response, that is, probability of a customer's response being influenced by a campaign treatment, then

$$P_{NL} = P_1\,(2 - 1/\,P_2)$$

2.4 Statistical Tests

50 random samples were drawn from device sales data and previous device campaign data. Propensity scores and net lift scores were produced for all sample sets. Paired t tests were applied to compare sales rates: control vs. treated, and propensity segments vs. net lift segments.

3. Results

As shown in Table 1, both the propensity model and the net lift model generated segments showed significantly higher sales rates, compared to baseline, while the net lift model showed a larger lift.

The sales rate in the treated group was significantly higher than that of the control group; and this was true for both the propensity model and the net lift model generated segments, as shown in Table 2. This means that both models successfully picked up the prospects that could be persuaded to purchase by campaign treatment.

Incremental sales rate is defined as the difference between the sales rates of the control group and the treated group. In this study, the incremental sales rate for the net lift model nearly doubled that of the propensity model, 1.73% and 0.87% respectively. However, in terms of the net incremental sales count, propensity model produced a significantly larger number, 46 vs. 17, respectively, as shown in Table 3.

Table 1. Comparison of Sales Rate between a Model Segment without Campaign Treatment, and Base Line

Group	N	Mean	p-value
Propensity Model Segment	50	4.56%	<0.001
Baseline	50	2.72%	
Net Lift Model Segment	50	5.20%	<0.001
Baseline	50	2.72%	

Table 2. Comparison of Sales Rate between Control Group and Campaign-treated Group

Model	Group	N	Mean	p-value
Propensity	Treated	50	5.43%	<0.001
	Control	50	4.56%	
Net Lift	Treated	50	6.93%	<0.001
	Control	50	5.20%	

Table 3. Comparison of Incremental Sales Rate and Sales Count between Models

Group	N	Mean	p-value
Propensity Incremental Sales Rate	50	0.86%	<0.001
Net Lift Incremental Sales Rate	50	1.73%	
Propensity Incremental Sales Count	50	46	<0.001
Net Lift Incremental Sales Count	50	17	

4. Discussion

This study shows that net lift models can generate segments that result in much improved incremental sales rates compared to traditional propensity models, which confirms other researchers' reports. But the propensity model outperformed the net lift model by producing a larger net incremental sales count. This is because the net lift model generated a much smaller segment, which canceled out the improved incremental sales rate that the net lift model afforded.

The choice of modeling techniques is sensitive to domain as well as to the characteristics of the target. For extremely skewed data that results in a small target, propensity models may be the better choice for overall gain.

Acknowledgements

This work is supported by C Spire.

References

1. Uplift Modeling: The next big thing in predictive analytics. R. L. Mitchell, ComputerWorld.com, 2012.
2. Net Lift Models: Optimizing the Impact of Your Marketing Efforts. K. Larsen, 2010.
3. Uplift Modeling in Direct Marketing. P. Journal of Telecommunications and Information Technology. Rzepakowski and S. Jaroszewicz. , 2012.
4. Real-world Uplift Modeling with Significance-based Uplift Trees. N. J. Radcliffe and P. D. Surry. Stochastic Solutions White Paper, 2011.
5. Incremental Response Modeling Using SAS Enterprise Miner. T. Lee, R. Zhang, X. Meng and L. Ryan. SAS Global Forum 2013
6. Incremental Value Modeling. B. Hansotia and B. Rukstales. 2002 Journal of Interactive Marketing 16:35-46.
7. Identifying who can be saved and who will be driven away by retention activity. N. J. Radcliffe and R. Simpson. Stochastic Solutions Limited, 2007
8. Optimal Targeting through Uplift Modeling. A white paper by Portrait Software, 2006
9. The Art of Balancing. Z. Zhang, J. Croft and K. Churchwell. The 2014 Joint Statistical Meeting Proceedings, 2063-2069

Submitting author
Zhen Zhang, Ph.D., Department of Marketing, 1018 Highland Colony Parkway, Ridgeland, MS 39157, USA. Phone (601) 540-7157

E-mail: zzhang@cspire.com