Analyzing Direct Marketing Campaign Performance Using Weight of Evidence coding and Information value through SAS Enterprise Miner Incremental Response Modeling Node

Abstract:
Data Mining and predictive models are extensively used to find the optimal customer targets so as to maximize the return on investment. Direct marketing techniques target all the customers who are likely to buy regardless of the customer classification. In a real sense this mechanism couldn’t classify the customers who are going to buy even without a marketing contact, there by resulting in a loss on investment. This paper focuses on Incremental Lift modeling approach using Weight of Evidence Coding and Information Value followed by Incremental Response and Outcome model Diagnostics. This model identifies the additional purchases that would not have taken place without a Marketing campaign. Modeling work was conducted using combined model. The research work is carried out on a Travel Center data, which identifies the increase in average response rate by 2.8% and the number of fuel gallons by 244 when compared with the results from the traditional campaign, which targeted everyone. This paper discusses in detail about the implementation of ‘Incremental Response’ node to direct the marketing campaigns and it’s Incremental Revenue and Profit analysis.

Incremental Lift Modeling Approach
This paper presents the motivation behind using Incremental Response node to target the group of customers who are persuadable instead of spending resources on others. Complete work is carried out on a Travel Center data. Incremental Response node requires a Treatment variable, a response target that is a binary with 1 being a treatment group and 0 a control group and the Target response variable. Here, the Outcome target is the number of gallons the customer fuels in a specific marketing campaign.

In general, the fundamental approach of handling categorical variables in modeling is the dummy variable coding and this result in a situation called curse of dimensionality and may cause over fitting of data. So, the best way to work with a target categorical variable is to perform the WOE (weight of Evidence) coding. WOE and IV (Information Value) have been extremely useful in variable reduction and results in variables, which got high predictor importance.

Weight of Evidence & Information Value
The target variable is the customer response. let’s code it as ‘Y’ and assume Y=1 as responded and Y=0 as not responded. The predictor variables are categorized into mutually exclusive bins, so the weight of evidence for each bin is going to be the log of the customers who responded to the ones who didn’t respond.

Weight of Evidence can be calculated as
Whereas in Incremental Response Modeling, we got Treatment and Control groups and the Net Weight of Evidence (NWOE) is calculated in the same way, but with the inclusion of both the Treatment and Control group responses.

\[
NWOE = \log \frac{P(X = x_i|Y = 1)_T}{P(X = x_i|Y = 0)_C}/ \frac{P(X = x_i|Y = 1)_C}{P(X = x_i|Y = 0)_C}
\]

Weight of Evidence analyzes the predictive power of a variable with respect to the targeted outcome but the Information value gives the overall predictive ability of the variables being considered and this helps in comparing the predictive importance of one with that of the other variables.

\[
IV = \sum_i (P(X = x_i|Y = 1) - P(X = x_i|Y = 0)) \cdot WOE_i
\]

Information value for Incremental Response Modeling is going to include the Treatment and Control groups in its calculation to determine the predictive importance of variables.

**Initial Exploratory Analysis**

The data, which is being used, is of a Travel center and comes from different tables of SQL Server databases, which are very huge containing 40 to 50 tables in each database. All the data is extracted with respect to a specific campaign, which is being implemented in a conventional method of A/B Testing. SAS codes are written to blend the data as required and a final data set is of 31 variables and 4,117 records. The figure 1 shows the variable summary giving the roles and measurement levels of all the variables involved.

**Variable Summary**

<table>
<thead>
<tr>
<th>Role</th>
<th>Measurement Level</th>
<th>Frequency Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>NOMINAL</td>
<td>1</td>
</tr>
<tr>
<td>INPUT</td>
<td>BINARY</td>
<td>8</td>
</tr>
<tr>
<td>INPUT</td>
<td>INTERVAL</td>
<td>8</td>
</tr>
<tr>
<td>INPUT</td>
<td>NOMINAL</td>
<td>7</td>
</tr>
<tr>
<td>INPUT</td>
<td>ORDINAL</td>
<td>1</td>
</tr>
<tr>
<td>REJECTED</td>
<td>NOMINAL</td>
<td>3</td>
</tr>
<tr>
<td>TARGET</td>
<td>BINARY</td>
<td>1</td>
</tr>
<tr>
<td>TARGET</td>
<td>INTERVAL</td>
<td>1</td>
</tr>
<tr>
<td>TREATMENT</td>
<td>BINARY</td>
<td>1</td>
</tr>
</tbody>
</table>
The main purpose of the Incremental Response modeling is to know the customers who cause to achieve the True Lift for the Marketing Campaign. This paper explains in detail about the Combined True Lift Modeling Implementation using SAS Enterprise Miner.

**Two Sample TTEST**

The TTest assesses whether the means of two groups are statistically different from each other. In our context, we can determine whether the average response rate and average number of gallons are different between Treatment and Control groups.

\[
\begin{align*}
H_0: & \text{ Average Response rate of Treatment group is equal to Control group} \\
H_1: & \text{ Average Response rate of Treatment group is not equal to Control group}
\end{align*}
\]

SAS Enterprise Guide is used to run the TTest with default options at a 5% significant level. Response is being assigned as Analysis variable and Promotion being the classification variable. TTest provides two different methods to determine the significance based on the equality of variances. Figure 2 shows the Equality of Variances, P value is greater than the significance level, which means that we don’t reject the Null Hypothesis and we go with Equal Variances.

\[
\begin{array}{c|c|c|c|c|c}
\hline
\text{Method} & \text{Num DF} & \text{Den DF} & \text{F Value} & \text{Pr > F} \\
\hline
\text{Folded F} & 2141 & 1974 & 1.09 & 0.0525 \\
\hline
\end{array}
\]

**Figure 2. F Test**

Now that, we determined to go with Equal Variances (Pooled Method), Figure 3 shows the t-value and significance values. Figure 3 show that P-value is greater than the significance level, which means that we don’t reject the null hypothesis. This says that Average response rate is not significantly different between the Treatment and Control groups.

\[
\begin{array}{c|c|c|c|c|c}
\hline
\text{Method} & \text{Variances} & \text{DF} & \text{t Value} & \text{Pr > [t]} \\
\hline
\text{Pooled} & \text{Equal} & 4115 & 1.11 & 0.2675 \\
\text{Satterthwaite} & \text{Unequal} & 4109 & 1.11 & 0.2666 \\
\hline
\end{array}
\]

**Figure 3. TTest Procedure**

Let’s do the TTest for Average gallons, which determines whether average gallons are significantly different between Treatment and Control groups.

\[
\begin{align*}
H_0: & \text{ Average gallons of Treatment group is equal to Control group} \\
H_1: & \text{ Average gallons of Treatment group is not equal to Control group}
\end{align*}
\]
As we did before, we choose the method based on F-test, which is significant here. So, we choose unequal variances. Figure 4 shows the results from the T test. Here, the P-value using Satterthwaite method is less than the significance level, which is 0.05. This leads to rejection of Null Hypothesis, which concludes Average gallons of Treatment, and Control groups are significantly different.

| Method    | Variances | DF   | t Value | Pr > |t| |
|-----------|-----------|------|---------|------|---|
| Pooled    | Equal     | 4115 | 4.47    | <.0001 |
| Satterthwaite | Unequal | 4007.4 | 4.46 | <.0001 |

Figure 4. T Test Procedure

Imputation

The data being extracted got missing values for few variables and are imputed using Impute node with the default options of Mean for the Interval variables and Count for the class variables and Unique type of Indicator variables are created. The numbers of missing values are relatively small for all the variables except for two variables, one being the ‘Billing card’, which identifies the type of card being used for purchase and the other one, which deals with the discount offer being provided to a customer.

Figure 5. Imputations Property Panel
In any modeling project, it is essential to proceed for an honest assessment, which improves the credibility of the model being built. So, the data is partitioned to training and validation sets with a ratio of 50% each. The default properties are being used in the properties panel. At this point, the data is in shape as required and is connected to the Incremental Response Node. Changing the properties customizes the model and a combined model is generated.

From the initial exploration of the binary response target variable and the treatment promotion variable, we get to see that percentage of response in both the treatment and the control group is about 12.54%.
With the descriptive statistics performed, Incremental Response node is all set to go. After we ran this node, it gives all the necessary details for targeting specific customers. Looking over the Response outcome summary table, we got to see an average of around 130 gallons more in Treatment group than in Control with respect to Training data. The values of training and validation data clearly show that the fit of data is pretty good, as the data holds for validation.

<table>
<thead>
<tr>
<th>Treatment variable</th>
<th>Data Role</th>
<th>Number of Observations</th>
<th>Number of Response</th>
<th>Rate of Response</th>
<th>Average gallons</th>
<th>Total gallons</th>
<th>Promotion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promotion</td>
<td>Train</td>
<td>104</td>
<td>123</td>
<td>0.173566</td>
<td>1512.929</td>
<td>1938.615</td>
<td>189461.53</td>
</tr>
<tr>
<td>Response</td>
<td>Train</td>
<td>101</td>
<td>107</td>
<td>0.155941</td>
<td>1742.583</td>
<td>1752.136</td>
<td>178213.61</td>
</tr>
<tr>
<td></td>
<td>Validate</td>
<td>104</td>
<td>127</td>
<td>0.150806</td>
<td>1482.105</td>
<td>186570.40</td>
<td>186570.40</td>
</tr>
<tr>
<td></td>
<td>Validate</td>
<td>86</td>
<td>102</td>
<td>0.155689</td>
<td>1577.571</td>
<td>169822.41</td>
<td>169822.41</td>
</tr>
</tbody>
</table>

Till this point, everything looks same as the traditional models. Let’s jump in to discuss the real purpose of using this node.
**Response Model Diagnostics**

Figure 9 shows the incremental response for both the training and validation data. It seems the data holds out for validation, which is a good part. The pattern goes in a nicer fashion as we can observe; the first few deciles got higher values of response compared with the other ones. The predicted Increment in the first decile is about 11% compared to 10% in the observed increment. The difference between the predicted and observed increment should be small and it is of the same case here. The predicted increment comes through targeting just the persuadable and not counting on any other customers. So, this helped the company to design marketing campaign in such a way that it targets the first few deciles.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Data Role</th>
<th>Predicted Treatment</th>
<th>Predicted Control</th>
<th>Predicted Increment</th>
<th>Observed Control</th>
<th>Observed Treatment</th>
<th>Observed Increment</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Train</td>
<td>0.379647</td>
<td>0.262196</td>
<td>0.11745</td>
<td>0.232</td>
<td>0.3375</td>
<td>0.1055</td>
</tr>
<tr>
<td>10</td>
<td>Validate</td>
<td>0.352902</td>
<td>0.244091</td>
<td>0.108571</td>
<td>0.229891</td>
<td>0.303289</td>
<td>0.075435</td>
</tr>
<tr>
<td>20</td>
<td>Train</td>
<td>0.189561</td>
<td>0.157985</td>
<td>0.031566</td>
<td>0.171171</td>
<td>0.180951</td>
<td>0.009868</td>
</tr>
<tr>
<td>20</td>
<td>Validate</td>
<td>0.169779</td>
<td>0.144836</td>
<td>0.024143</td>
<td>0.2</td>
<td>0.152941</td>
<td>-0.04706</td>
</tr>
<tr>
<td>30</td>
<td>Train</td>
<td>0.148605</td>
<td>0.129158</td>
<td>0.015366</td>
<td>0.105981</td>
<td>0.195122</td>
<td>0.089431</td>
</tr>
<tr>
<td>30</td>
<td>Validate</td>
<td>0.130861</td>
<td>0.113853</td>
<td>-0.00698</td>
<td>0.178285</td>
<td>0.25</td>
<td>0.071705</td>
</tr>
<tr>
<td>40</td>
<td>Train</td>
<td>0.109121</td>
<td>0.118916</td>
<td>-0.00797</td>
<td>0.175926</td>
<td>0.154639</td>
<td>-0.02129</td>
</tr>
<tr>
<td>40</td>
<td>Validate</td>
<td>0.079233</td>
<td>0.092548</td>
<td>-0.01342</td>
<td>0.052174</td>
<td>0.044444</td>
<td>-0.00773</td>
</tr>
</tbody>
</table>

Figure 9. Incremental Response Modeling Statistics

As the model helps in predicting both the response and the outcome, it is of great importance to discuss the Incremental Outcome Model.

**Incremental Outcome Model Diagnostics**

When we look into the difference between the average gallons made by the treatment and control groups, it is of 130 gallons, which means a gallons lift of 130 in overall training data. Now that, we build the Incremental model, the Figure 10 shows the training and validation data average gallons. The first decile predicted increment gallons is 208 whereas the observed being 207, a slight difference. The point to note here is that, the predicted gallons are just from the specific persuadable customer group, which means that they account to True Lift.

The Travel center used to target all the set of customers and achieved a decent lift in gallons, but the present takes the lift in gallons to a higher level and there by the dollars. They targeted just the first few deciles to get more than the expected lift in gallons and dollars.

Figure 10. Incremental Model Outcome Diagnostics
The above two steps help in better prediction of the set of customers. We can even get to see the profitable deciles as a result of this node. Average Incremental Revenue shows that the first three deciles are profitable both in Training and validation datasets.

<table>
<thead>
<tr>
<th>Decile</th>
<th>Profitable</th>
<th>Average Incremental Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Train</td>
<td>Profitable</td>
<td>307,724</td>
</tr>
<tr>
<td>2 Train</td>
<td>Profitable</td>
<td>59,593</td>
</tr>
<tr>
<td>3 Train</td>
<td>Profitable</td>
<td>13,497</td>
</tr>
<tr>
<td>1 Validate</td>
<td>Profitable</td>
<td>38,025</td>
</tr>
<tr>
<td>2 Validate</td>
<td>Profitable</td>
<td>49,653</td>
</tr>
<tr>
<td>3 Validate</td>
<td>Profitable</td>
<td>5,077,033</td>
</tr>
</tbody>
</table>

Figure 11. Average Incremental Revenue

**Penalized NIV**

We looked through different aspects of the Incremental Response Node. As we discussed earlier about the NWOE and NIV values to determine the variable importance, it is now time to see the variables being selected by the model.

Honest assessment in predictive models is used to assess the stability and robustness of the data, which means that the predictive importance of the Training and Validation dataset shouldn’t differ much. If the validation dataset doesn’t hold with the training, the model is not robust. Variable, which lacks robustness, results in difference in NWOE of Training and Validation data set.

In order to minimize that, the NIV is adjusted with a penalty term, which takes in to account the difference of NWOE from training dataset and NWOE from validation dataset. This is termed as Penalized NIV

\[
PNI\text{V} = NIV - \text{penalty}
\]

Penalized NIV improves the predictive robustness of each variable with respect to the Target.

Figure 12 shows the Variables being selected by PNIV in our data. Seven variables are being selected as the most predictive variables sorted by Penalized NIV.
There is an increase in the average response rate from 10\% to 12.8\% using this model and increase in the average number of fuel gallons from 150 to 394 gallons is also being observed.

**Conclusion**

This Paper provides a different approach to Marketing Campaigns using Incremental Response Modeling and explains in detail on how the True Lift achieved with the help of this node is better than the traditional campaign Lift. It demonstrates the Weight of Evidence Coding and Information value, which helps in variable reduction and presents the high predictor variables. The traditional propensity model may be good at targeting customers, but it couldn't achieve the true responders of campaigns and this model helps the company to achieve that.

**References**


[http://www.socialresearchmethods.net/kb/stat_t.php](http://www.socialresearchmethods.net/kb/stat_t.php)

**Acknowledgement**

Dr. Goutam Chakraborty, Professor (Marketing) at Oklahoma State University and Management Consultant

**Contact**

Sravan Vadigepalli

Email: Sravan.vadigepalli@gmail.com