Incremental Response Modeling Based on Segmentation Approach Using Uplift Decision Trees

Sankara Prasad Kondareddy[1(✉)], Shruti Agrawal[2],
and Shishir Shekhar[3]

[1] Specialist, Fidelity Investments, Bangalore, India
k.shankarprasad@gmail.com
[2] M.SC, Indian Institute of Technology Kanpur, Kanpur, India
Shruti.agrawal271@gmail.com
[3] Lead, Fidelity Investments, Bangalore, India
shishir.shekhar@gmail.com

Abstract. Data mining methods have been successfully used in direct marketing to model the behavior of responders. But these response models do not take in account, the behavior of customers who would take an action irrespective of marketing action. Redundant marketing communications sometimes annoy the customer and reduce the brand value of the company. Accurate targeting of customers also reduces direct marketing cost. Incremental response modeling aims to predict the behavior of customers who respond positively only in the case of marketing. In this paper, we propose a two-step approach for incremental response modeling. In the first step, we segment the data using uplift decision trees using traditional and modified divergence metrics. Then, in the second step we use the standard incremental response modeling methods. Experiments on real world direct marketing campaign data showed that the proposed method outperforms the uplift decision trees.

Keywords: True-lift modeling · Net lift · Uplift decision trees

1 Introduction

Direct marketing cost is one of the biggest contributors of overall marketing cost. Most companies use data mining tools to develop predictive models to identify customers who are likely to respond to direct marketing campaigns. But improper use of these models may lead to unexpectedly low response rates. This is mainly because many of the customers get annoyed due to repetitive phone calls, emails, etc. Companies thus want to target their customers precisely to retain the brand value and improve customer satisfaction.

Traditionally, response models try to predict the likelihood of response for a particular customer, given a marketing action. These models do not account for likelihood of response irrespective of marketing action. Incremental response modeling as shown in Fig. 1 (also known as uplift modeling or differential response modeling or true uplift modeling) removes this drawback and works towards choosing the customers

effectively. Incremental response models are used to identify the customers where there is a true uplift [5]. We want to identify customers who respond to the direct marketing campaign positively and remove the customers from the targeting list who will take an action irrespective of marketing or not.

		Respond if targeted	
		Yes	No
Respond if not targeted	Yes	Spontaneous response (Targeting not required here)	Don't target at all (customers getting annoyed, leading to negative response)
	No	Customers to be targeted	No effect of targeting (only increasing the cost to the company)

Fig. 1. True response uplift.

In true uplift modeling, historical treatment and control campaign datasets are used to build a predictive model. The treatment dataset consists of customers to whom the marketing communication was sent and the control set are the ones who are not reached out but kept aside for measurement purpose. Still, we get some response from control set which is called as natural response. The response from our treatment set consists of natural response and uplift in response due to campaign as shown in Fig. 2.

TREATMENT DATA	CONTROL DATA
Response= Yes (Natural Response)	Response =Yes (Natural Response)
Response= Yes (Uplift due to campaign)	Response=No
Response = No	

Fig. 2. Potential incremental response [10].

This uplift due to campaign is of importance to marketing division. For a particular customer, true uplift can be predicted by modeling the difference between probability of response, given treatment and probability of response, given no treatment (control). The customers with positive uplift can then be reached out. Standard classification models cannot account for the control set and thus cannot be used directly. In this work, we tried to model uplift using decision trees [1] based on various modified splitting criteria. We propose a model which constructs an uplift decision tree and then applies

logistic regression based uplift models on the observations at each terminal node to get the prediction. Experimental results showed that the proposed method outperforms various other uplift modeling techniques available.

2 Literature Review

Despite the importance of the topic, surprisingly there is not much work done on uplift modeling. Basically, there are two ways of building uplift models. The first method is called two-model approach. It is based on creating different classification models on treatment set and control set separately, to predict probability of response for each set. Each of the treatment and control dataset is divided into training and test data. Models are constructed separately using standard machine learning classifiers on each training data. The difference of the two predicted probabilities from treatment and control model gives us the predicted uplift. Models are then used to predict probability of response given treatment and probability of response given no treatment (control) on the test data (consisting of treatment as well as control data) for evaluating the model. If uplift follows different pattern than the probabilities of response in treatment and control then, this approach may not perform well.

The second modeling technique predicts the uplift directly from treatment and control dataset using a single model. The complete dataset is divided into training and test data, each containing both treatment and control. A single model is developed using the training data and further evaluated on test data. Some of the papers that addressed decision tree based models for uplift prediction are [2, 3, 7–9, 11]. [7] presented the first paper which explicitly discussed the topic of uplift modeling. [7] provided various illustrations relating to uplift based on real world data and came up with modified decision tree based algorithm which was further detailed in [8]. [3] used splitting criterion based on incremental response rate ($\Delta\Delta P$) for construction of decision trees. This algorithm did not account for population size and thus leading to unwanted weightage on small populations with more uplift. In [9], trees are constructed on information theory based splitting criterion which is more in line with modern machine learning algorithms. The divergence measure (Euclidean or Kullback-Leiber) between the distribution of treatment and control response rate is used as a splitting criteria. Also, they have presented generalized versions for calculation of gain, normalization factor and used variance based pruning technique to construct tree. [11] further accounted for multiple treatment case. They also confirmed through experiments that decision trees based uplift model show significant improvement over previous uplift modeling techniques.

The regression based technique for uplift modeling were also been addressed in literature [4–6]. [5] developed a model by first adding interaction terms between independent variable and treatment-control flag variable. He then created model on the dataset and used it with great success for years in marketing at Fidelity Investments. In [12], two model based approach is followed using logistic regression. [4] proposed a model based on net weight of evidence. He also proposed the modified version of Information value for variable selection for uplift modeling. Most papers in literature used artificially created data to carry out their experiments. In this paper, we have

proposed a two-step approach for incremental response modeling and used real time direct marketing campaign datasets to carry out our experiments.

The paper is further arranged into following sections. Section 3 describes the notation and definition of various divergence measures used. Section 4 presents the approach of work, along with the new proposed method. Section 5 shows the results of various experiments done. Section 6 concludes the paper with the result of proposed method. Section 7 provides the scope for future research.

3 Notations and Definitions

Let us introduce and formalize the notation used in the following sections. $Y \in \{1, 0\}$ denotes the response variable for a marketing campaign, which is binary (1 for response = yes and 0 for no), flag $F \in \{1, 0\}$ denotes the binary variable which is 1 for treatment input and 0 for control input and $X_1, \ldots, X_m \in \mathbb{R}$ denotes the set of independent variables. $P^T(.)$ and $P^C(.)$ denote the probability with respect to treatment and control dataset respectively. We want to predict the uplift as difference in predicted probabilities between treatment and control group i.e. $P(Y = 1|F = 1, X) - P(Y = 1|F = 0, X)$, which is denoted as $P^T(Y = 1|X) - P^C(Y = 1|X)$. We used Laplace correction while estimating the probabilities. For any test A of a numeric variable X at split value v at a node, let a_l and a_r denote the two outcomes $(X \leq v)$ and $(X > v)$ of the test and N(left) and N(right) denote the corresponding number of outcomes for each. As we have only considered binary splits, for any categorical variable also, we will have correspondingly two outcomes a_L and a_R for any test of split. Also, N = N(left) + N(right) denote the number of observations at the node where split is to occur.

4 Proposed Method

In our work, we tried various new splitting criteria for uplift decision tree along with those presented in [11], on real world direct marketing campaign data. The problem with decision tree based approach is that the predicted posterior uplift values will not be a true representation of the actual uplift, because the terminal nodes will have one value as predicted posterior uplift.

We propose a two-step model for incremental response modeling as shown in Fig. 3. In the first step, we build an uplift decision tree using traditional and modified divergence measure as splitting criteria and then in the second step we build uplift models at each of the terminal node using traditional uplift regression methods. This is equivalent to segmenting the complete data (treatment & control) using uplift decision trees and building uplift models in each of the segments. This is same as using clustered weighted modeling approach for uplift modeling. In this paper, we used two-model approach in the second step. We use a logistic model for treatment and control dataset at each terminal node separately and predicted uplift probability is calculated as the difference in individual predicted probabilities.

In the first step, we generated the uplift tree up to some significant depth, so that there are enough observations at the terminal nodes to build two-model logistic regression. For applying logistic regression, variable selection was first done using net information value [10] of variables and then, removing some of those variables while checking for multi-collinearity. As the logistic regression model is created on treatment and control samples separately at each terminal node, we constrained the number of observations that can be present in treatment and control set at a node after each split in the uplift decision tree. The difference between the predicted probabilities of two models on treatment and control dataset at each terminal node is the predicted posterior uplift probability. Applying logistic regression on terminal nodes is expected to give better results as it assigns posterior probabilities to each input observation that comes at terminal node, rather than giving the same value to entire node. As expected, this proposed model outperformed all the decision tree based models.
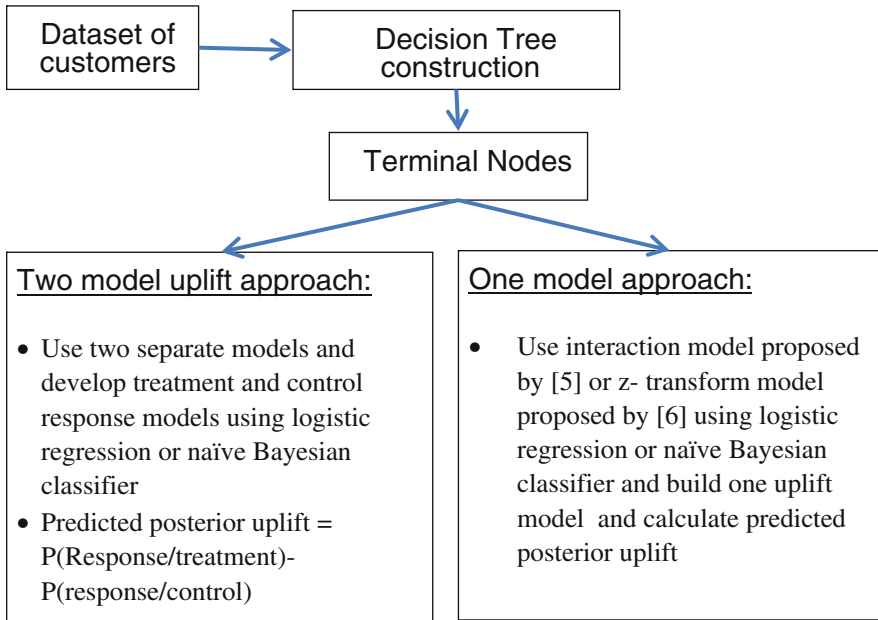
Fig. 3. Proposed hybrid model or two step model.

We implement the decision tree as described in [9] using various other splitting criteria along with Euclidean and Kullback-Leibler divergence measures.

For any split test A at a node, gain using divergence measure D is defined as follows:

$$D_{gain}(A) = D_{gain}(P^T(Y) : P^C(Y)/A) - D_{gain}(P^T(Y) : P^C(Y)) \qquad (1)$$

Where

$$D_{gain}(P^T(Y) : P^C(Y)/A) = \frac{N(Left)}{N} D[P^T(Y/a_L) : P^C(Y/a_L)] + \frac{N(Right)}{N} D[P^T(Y/a_R)$$
$$: P^C(Y/a_R)]$$

(2)

Let $P^T(Y) = [a_1 = P^T(Y = 1), a_0 = P^T(Y = 0)]$ & $P^C(Y) = [a_1 = P^C(Y = 1), a_0 = P^C(Y = 0)]$. The divergence measures defined for probability distributions $P^T(Y)$ and $P^T(Y)$ are defined in (3), (4), (5), (6), (7) and (8).

$$\text{Euclidean divergence measure ED} = \sum_i (a_i - b_i)^2$$

(3)

$$\text{Hellinger divergence measure HD} = \text{SQRT}[\sum_i (a_i - b_i)^2]$$

(4)

$$\text{Kullback - Liebler divergence measure KL} = \sum_i \left(a_i Log\left(\frac{a_i}{b_i}\right) \right)$$

(5)

$$\text{J - divergence measure JD} = \sum_i \left((a_i - b_i)Log\left(\frac{a_i}{b_i}\right) \right)$$

(6)

$$\text{Modified Euclidean divergence measure MED} = (a_1 - b_1)^2$$

(7)

$$\text{Modified J - divergence measure MJD} = (a_1 - b_1)Log\left(\frac{a_1}{b_1}\right)$$

(8)

For modified Euclidean Divergence and modified J divergence, we have incorporated only the $P(Y = 1)$ while calculating the divergence. The normalization factors considered to penalize the uneven splits are shown in (9) and (10).

$$I(A) = H\left(\frac{N^T}{N}, \frac{N^C}{N}\right) * KL\left(P^T(A) : P^C(A)\right) + \frac{N^T}{N} H\left(P^T(A)\right) + \frac{N^C}{N} H\left(P^C(A)\right) + 0.5$$

(9)

$$J(A) = G\left(\frac{N^T}{N}, \frac{N^C}{N}\right) * D\left(P^T(A) : P^C(A)\right) + \frac{N^T}{N} G\left(P^T(A)\right) + \frac{N^C}{N} G\left(P^C(A)\right) + 0.5$$

(10)

Where H(A) denotes the entropy function that is $H(A) = -\sum_i a_i * log(a_i)$ and $G(A) = \sum_i a_i(1 - a_i)$ for probability distribution $A = (a_1, \ldots, a_n)$. Our final splitting criterion is given in (11). Label to the leaf node (predicted posterior uplift probability) is given as in (12).

$$\frac{ED_{gain}}{J(A)}, \frac{HD_{gain}}{J(A)}, \frac{KL_{gain}}{I(A)}, \frac{JD_{gain}}{I(A)}, \frac{MED_{gain}}{J(A)}$$

(11)

$$\text{Label value to leaf node} = P^T(Y = 1/L) - P^C(Y = 1/L) \tag{12}$$

Where $P^T(Y = 1/L) \& P^C(Y = 1/L)$ denote treatment and control target class distributions at leaf node L. Some modified versions based on changes in normalization factors and computation of gain for each split were also tried out, but they did not perform very well. We found that the euclidean divergence and modified euclidean divergence measure performed best on the considered dataset.

$$a = P^T(X \le v/Y = 1) * P^C(X \le v/Y = 0) \tag{13}$$

$$b = P^T(X \le v/Y = 0) * P^C(X \le v/Y = 1) \tag{14}$$

$$c = P^T(X > v/Y = 1) * P^C(X > v/Y = 0) \tag{15}$$

$$d = P^T(X > v/Y = 0) * P^C(X > v/Y = 1) \tag{16}$$

Using (13), (14), (15) and (16), Modified Net weight of evidence (NWOE) and Modified Net Information Value (MNIV) as splitting criterion is defined as follows:

$$\text{NWOE for left node } NWOE_{Left} = \log\left(\frac{a}{b}\right) \tag{17}$$

$$\text{NWOE for right node } NWOE_{Right} = \log\left(\frac{c}{d}\right) \tag{18}$$

$$MNIV = (a - b) * NWOE_{Left} + (c - d) * NWOE_{Right} \tag{19}$$

We also implemented decision tree based on modified net information value [10] but it could not outperform the measures based on euclidean divergence. Along with two separate logistic regression model approach, we also tried a logistic regression model based on z transformation (single linear model) as described in [6], but it performed very poorly on the dataset considered.

5 Experimental Results

We have used two real time marketing campaign datasets to evaluate the performance of the uplift decision trees as proposed in [11] and the proposed method. Dataset 1 has overall uplift response rate of 4.2 % (population treatment response rate-control response rate) and Dataset 2 has 4.1 % overall uplift response rate. Since we have treatment and control groups, evaluation of the model performance is different from standard classification models procedure. For evaluating the uplift model, we followed the same approach as of [5] and this method is adopted by many industrial practitioners.

We first create the model predicted posterior uplift probability for each observation in validation sample. We use the decile report based on the predicted posterior uplift probability to evaluate the uplift model. The deciles are created by rank ordering all the

observations of the validation sample by predicted posterior uplift probability. In each decile, actual uplift is calculated as the difference in observed responses rates of treatment and control group. We checked the correlation of actual uplift (treatment response-control response) with predicted posterior uplift and also rank ordering of actual uplift. Since the application is from direct marketing, we have used the lift in uplift response (top decile uplift response rate/overall data uplift response rate) in the top decile while comparing different methods.

Table 1. Comparison of different uplift models.

| Method	Technique	Lift in uplift response rate	
		Dataset 1	Dataset 2
One step	Decision tree using euclidean divergence	2.72	1.49
One step	Decision tree using modified euclidean divergence	2.76	1.90
One step	Decision tree using net information value	1.88	0.76
Two step	Decision tree using euclidean divergence + logistic regression	3.30	1.98
Two steps	Decision tree using modified euclidean divergence + logistic regression	3.09	1.94

Table 1 shows the comparison of different uplift models. Uplift decision trees based on modified Euclidean Divergence is performing better than Uplift decision trees based on Euclidean Divergence. But the proposed two-step method is working better than the one step uplift decision trees.
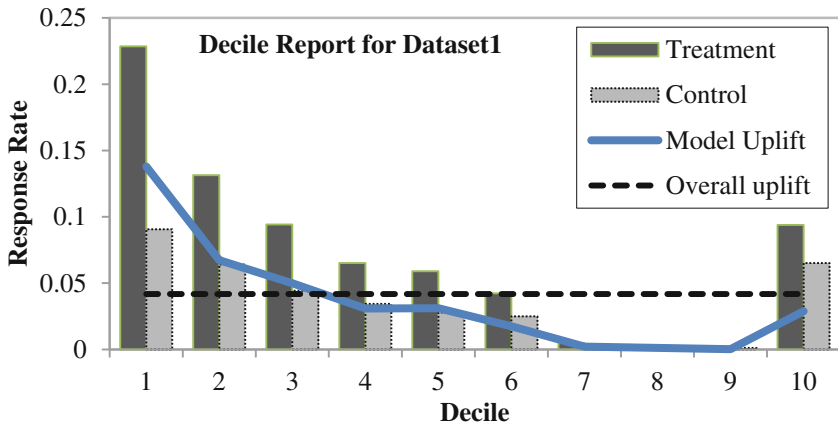
Fig. 4. Response rate by decile using two step method (decision tree using euclidean divergence + logistic regression) on validation dataset1.

The validation dataset decile reports using the proposed two step method: Decision Tree using Euclidean Divergence + Logistic Regression is shown in Figs. 4 and 5. Clearly we can see that the proposed method outperforms the other uplift decision trees.
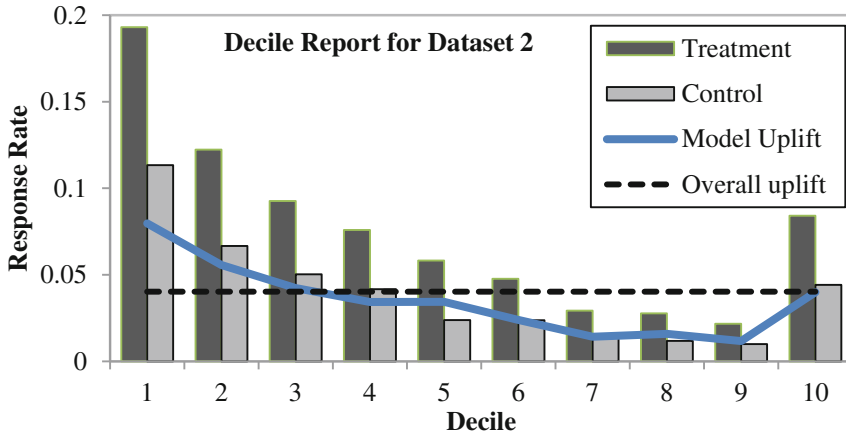
Fig. 5. Response rate by decile using two step method (decision tree using euclidean divergence + logistic regression) on validation dataset2.

6   Conclusion

Incremental response modeling or true lift modeling has got little attention in literature and is getting importance among marketing practitioners in recent years. We have implemented decision trees based classifiers for incremental response modeling on real world marketing campaign data. Out of various divergence measures used for splitting criterion, we found that those based on Euclidean measure give the best results. We proposed a model based on constructing a decision tree and then applying two-model logistic regression approach at each terminal node. Based on the experiments on real world marketing campaign data, we found that the two step model outperformed all decision tree based classifiers. Also tree based classifiers have been shown to outperform various other uplift models in literature. Hence, the proposed model holds significant importance in its area.

7   Future Work

The dataset consisted in each of the terminal nodes can be viewed as supervised clusters created from data. As we have tried logistic regression (two model based approach) on these clusters (dataset on terminal nodes), the future work can be based on trying different machine learning classification models for uplift modeling on these nodes.

References

1. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and Regression Trees. CRC Press, New York (1998)
2. Chickering, D.M., Heckerman, D.: A decision theoretic approach to targeted advertising. In: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, Stanford, CA, pp. 82–88 (2000)
3. Hansotia, B., Rukstales, B.: Incremental value modeling. J. Interact. Mark. **16**(3), 35–46 (2002)
4. Larsen, K.: Net lift models: optimizing the impact of your marketing. In: Predictive Analytics World (2011)
5. Lo, V.S.: The true lift model - a novel data mining approach to response modeling in database marketing. SIGKDD Explor. **4**(2), 78–86 (2002)
6. Jaskowski, M., Jaroszewicz, S.: Uplift modeling for clinical trial data. In: ICML 2012 Workshop on Clinical Data Analysis, Edinburgh, Scotland (2012)
7. Radcliffe, N.J., Surry, P.D.: Differential response analysis: modeling true response by isolating the effect of a single action. In: Credit Scoring and Credit Control VI, Edinburgh, Scotland (1999)
8. Radcliffe, N.J., Surry, P.D.: Real-world uplift modeling with significance-based uplift trees. White Paper TR-2011-1, Stochastic Solutions (2011)
9. Rzepakowski, P., Jaroszewicz, S.: Decision trees for uplift modeling. In: 2010 IEEE International Conference on Data Mining, Sydney, Australia, pp. 441–450 (2010)
10. Lee, T., Zhang, R., Meng, X., Ryan, L.: Incremental response modeling using SAS enterprise miner. In: Proceedings SAS Global Forum Conference, San Francisco, p. 96 (2013)
11. Rzepakowski, P., Jaroszewicz, S.: Decision trees for uplift modeling with single and multiple treatments. Knowl. Inf. Syst. **32**, 303–327 (2012)
12. Vansteelandt, S., Goetghebeur, E.: Causal inference with generalized structural mean models. J. Roy. Stat. Soc. B **65**(4), 817–835 (2003)