Optimal Targeting through Uplift Modeling:

Generating higher demand and increasing customer retention while reducing marketing costs.

www.portraitsoftware.com

A white paper by Portrait Software™

December 2006

Portrait SOFTWARE™

Abstract

Most direct marketing targets the wrong people.

It wastes money by focusing effort on many who will not react positively, and some who may even react negatively, while neglecting others who would respond favorably if targeted. This paper shows how a revolutionary new modeling technique – uplift modeling – can be used to optimize targeting to maximize the returns from direct marketing. It achieves this by predicting, at an individual level, the *change* in behavior likely to result from a particular marketing intervention.

Uplift Modeling: An Overview

We need to target customers whose behavior we can change, rather than taking credit for outcomes that would have occurred even without our intervention.

This paper discusses the use of a new kind of modeling – uplift modeling – to improve the profitability and effectiveness of direct marketing spend. We start from five observations about targeted marketing aimed at demand generation (cross-selling, up-selling and deep-selling). These are:

1. Direct marketing works better for some customers than others.
2. Direct marketing has negative effects on some customers, reducing their level of spend.
3. The customers who spend most when targeted are *not* necessarily the best targets for marketing: some of those whose spend is high when targeted would spend as much, *or more*, without treatment.
4. Not all purchases by targeted customers are caused by our direct marketing. Some would have bought anyway. To measure the *incremental impact* or *uplift* in sales resulting from a specific, targeted action we need to employ control groups. However, this is merely a post-campaign assessment of impact: it does not tell us how to target better in the future.
5. The standard approach to targeting involves modeling only the treated group from a previous campaign. This allows us to estimate each customer's conditional *probability of purchase if targeted*, but not the *change* in that probability resulting from our targeted action. Mathematically, such "response" models predict

$$\text{Prob (purchase} \mid \text{treatment)} \qquad (1)$$

(the *probability of purchase given treatment*) while our returns are based on uplift, defined as

$$\text{Prob (purchase} \mid \text{treatment)} - \text{Prob (purchase} \mid \text{no treatment).} \qquad (2)$$

In light of the points above, it is clear that traditional methods of targeting are suboptimal, based as they are on so-called "response" models that actually predict conditional purchase probability[1] (equation 1).

Ideally, we should model the difference in behavior between the treated and control groups. This would allow us to predict the *change* in each customer's purchase probability when treated (equation 2). We call such models *uplift models*.

In this paper, we show real-world examples of how uplift models generate higher returns on marketing spend, often from lower targeting volumes. This constitutes a multiple win, because sales increase while targeting costs are reduced. Additionally, where incentives are used, the cost of these is usually significantly reduced because they are not "wasted" on customers who would purchase anyway. Customer satisfaction also tends to improve as customers who dislike being contacted are less likely to be targeted.

The same shift in approach can also bring even larger benefits in the area of customer retention, where similar considerations apply, and where it is distressingly common for retention activity to drive up attrition rates. Adoption of uplift modeling involves a move from targeting on the basis of estimated attrition probability to an approach based on savability. We illustrate that such a change can have a very significant impact on the bottom line, which is natural given the counterproductive effect of spending money to drive customers away.

[1] whether weighted by spend or not

Who Can Benefit from Uplift Modeling?

Uplift modeling requires control groups and is of greatest benefit in competitive markets where customers are subject to many influences. It can be used to reduce costs, to increase revenues and to enhance customer satisfaction and retention.

Uplift modeling allows businesses to optimize their targeting of customers in such a way as to maximize the return on investment of marketing spend or to optimize some alternative goal. The benefits from this in the areas of demand generation and customer retention can include:

- Reducing the number of customers required to achieve a given level of business stimulation and thus reducing costs.
- Increasing the level of business generation achieved for any given level of spend.
- Lowering customer dissatisfaction by decreasing the level of negatively received material.
- Enhancing understanding within the business of the effectiveness of various kinds of marketing spend.
- Eliminating many or all of the negative effects associated with mis-targeted campaigns.
- Increasing customer retention.

Additionally, there is the potential to use these methods in other areas, such as optimizing the management of actions based on behavioral credit scores, optimizing collections targeting and optimizing response to customer complaints.

There are four qualification criteria important for evaluating the potential impact of uplift modeling:

- Control groups: Uplift modeling absolutely requires the systematic use of randomized control groups, so is only of benefit to businesses that already use, or are prepared to adopt this approach.
- Size of customer base: Businesses with large customer bases will tend to benefit disproportionately more than those with fewer customers, because reasonable volumes are needed to model this second-order effect. Though dependent on response rates, uplifts and the ratio of treated to control group sizes, we would not normally expect the technique to work well on campaigns targeting fewer than about 100,000 people.
- Influences: Uplift modeling is of particular benefit in situations in which many factors influence a customer's behavior. Companies with a largely single-channel, single-communication route to customers, in less competitive markets, will benefit less from uplift modeling than those with many channels and communication paths and greater competition.
- Overall level of uplift: Portrait believes that the approach of uplift modeling is correct even when the overall level of uplift is high and there are no obvious negative effects involved in a campaign. In these circumstances there are alternatives, albeit suboptimal ones. In circumstances where campaigns show little overall uplift, or negative overall uplift, other than a complete change of approach, there is no known alternative to uplift modeling.

Example 1: Deep-Selling

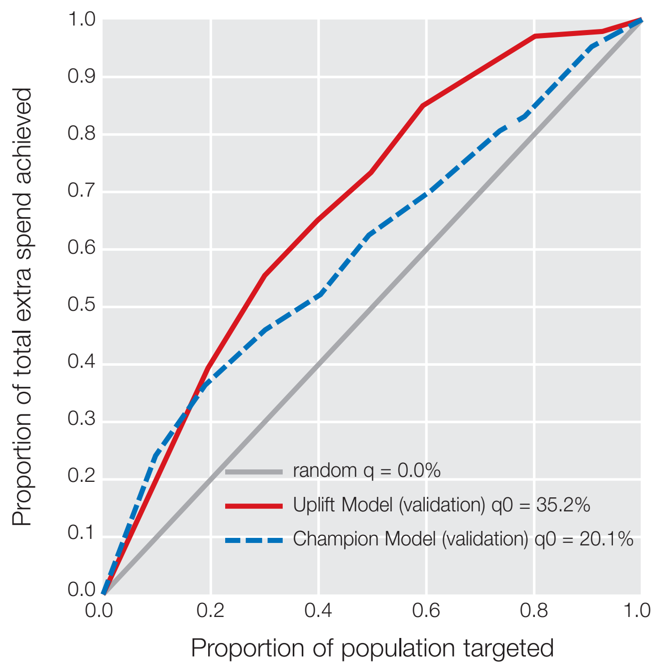
An uplift model for a catalog mailing allowed the *incremental* spend to be raised by 15% for a fixed-size target group.

A retailer used a catalog mailing to drive greater spend activity amongst active customers – an example of deep-selling.[2] The customers were selected on the basis of a conventional "response" model – the so-called "champion" model – built on data from customers mailed in a similar previous campaign. Approximately 100,000 of those ranked as likely high "responders" by the model were targeted, and approximately 50,000 were held back as a control group. On average, spend increased by $8 per head among those mailed. An uplift model was then built, using data from the same campaign as that used to construct the champion model. In contrast to the champion model, however, the uplift model used information about the control group as well as the mailed group.

Graph 1 shows a gains chart for "uplift in spend".[3] As with a conventional gains chart, the horizontal axis shows the proportion of the populated target. Customers are sorted by score, with those having the best expected outcomes on the left. However, whereas the vertical axis of a conventional gains chart shows the proportion of total spend achieved when targeting a given proportion of the population, this *uplift gains* chart shows the proportion of total *incremental* spend achieved. (This is assessed by comparison with the control group.)

As is the case with a conventional gains chart, the diagonal line shows the result of random targeting. Under this scenario, targeting any x% of the customers should yield the same x% of the total incremental revenue. This acts as a baseline, and any useful score will yield a curve that is "bowed" above this diagonal.

Graph 1: Demand Generation



Uplift Results Summary

Proportion of population targeted	50%	80%
Incremental spend per head – champion model	$10.24	$8.50
Incremental spend per head – uplift model	$11.84	$9.70
Extra spend per head gained from using uplift model	$ 1.60	$1.20
Percentage increase by using uplift model	16%	14%

The blue line (dashed) shows the result of targeting with the "champion" model and the red line (solid) shows the effect of targeting with the uplift model.[4] Up to about 20%, the two models perform similarly. Thereafter, the models diverge. For example, if 50% are targeted, the uplift model manages to identify customers delivering about 16% more revenue than the champion model – approximately $11.84 against $10.24 per head. And if 80% are targeted, the uplift model manages to retain 97% of the incremental spend ($9.70 per head) against only 85% ($8.50 per head) from the champion model.

Clearly, the uplift model is significantly better at identifying customers for whom marketing spend generates a positive return.

[2] While cross-selling tries to sell new products to customers, and up-selling aims to drive customers to upgrade, deep-selling simply tries to increase the frequency or size of their transactions.

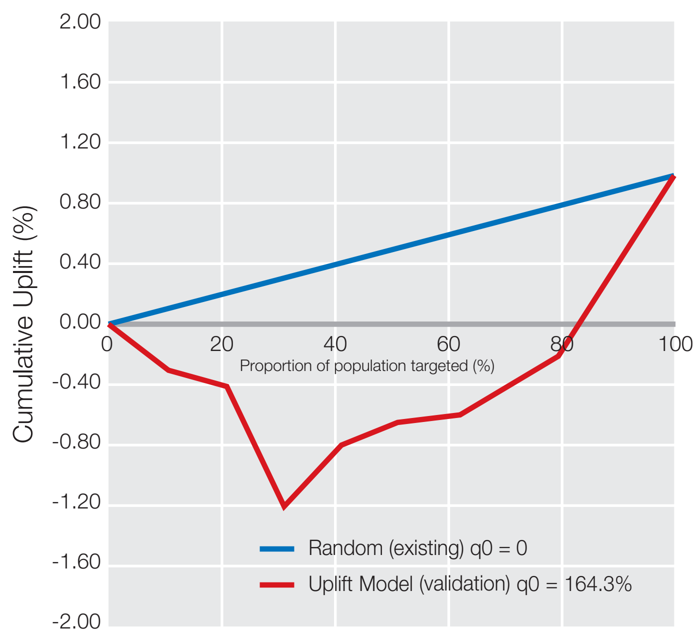
[3] Because the results shown here are from real companies, some results have been systematically rescaled to protect client anonymity and confidentiality.

[4] Both results are shown for validation data, which is also known as "holdout" or "test" data.

Example 2: Retention

A mobile phone operator ran a counterproductive retention mailing that increased churn[5] from 9% to 10%. An uplift model finds a 30% sub-segment which, if targeted alone, will allow an overall reduction in churn from 9% to 7.8%.

Graph 2: Churn Reduction



Uplift Results Summary

Strategy	Churn rate	Annual saving
Target everyone	10.0%	$ 0
Target no one	9.0%	$4,000,000
Target right 30%	7.8%	$8,800,000

A mobile phone company was experiencing an annual churn rate of approximately 9% in a segment. It targeted the entire segment with a retention offer, holding back only a control group. The net result was an *increase* in churn to 10% among the treated group, while the churn rate in the untreated group remained at 9%. Obviously, this is the exact opposite of the desired effect, but we have witnessed this phenomenon repeatedly. It seems that retention interventions often backfire because they variously remind customers of their ability to terminate, provide a catalyst to help overcome inertia and annoy customers through intrusiveness.

Clearly, one way to improve the situation is to stop this retention offer entirely. However, there was a strong belief within the business that the offer was valid and did work for some groups of customers. Also, no more successful approach had been identified. An uplift model was therefore built to try to identify a sub-segment within which the treatment was effective.

As in the previous example, Graph 2 is a gains chart for uplift. The horizontal axis shows the proportion of the population targeted, while the vertical axis shows the resulting increase in total churn. (The goal, of course, is to find a *negative increase*, i.e. a reduction in churn.) The mailing was untargeted, so the diagonal line (blue) shows the effect of random targeting of customers. Of course, with such random targeting, extra customers are lost in proportion to the number treated. However, the red line (concave) shows the effect of targeting on the basis of the uplift model. The results are striking.

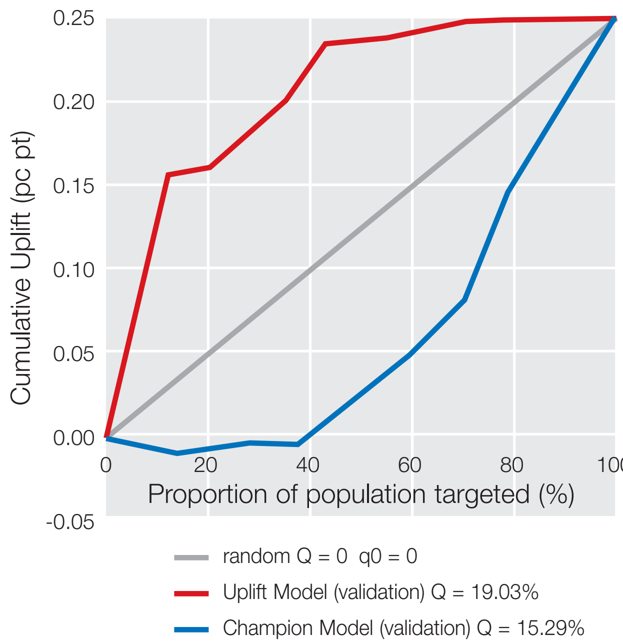
The model shows that retention activity was effective for about 30% of the customers: if only the 30% identified as "most savable" by the model are treated, churn across the *entire* segment falls by 1.2 percentage points, from 9% to 7.8%, i.e. over 13% fewer customers churn. Compared to targeting everyone, churn actually reduces from 10% to 7.8%—a proportionate fall of 22%. The segment contained approximately 1 million customers, so using industry-standard ARPU of $400/year, the financial impact of moving to uplift modeling, for one segment alone, is as shown in the results summary above.

[5] "Churn" is the term mobile phone companies tend to use to refer to customer loss or attrition

Example 3: Cross-selling High-value Products

The value of a new banking product is often high enough to justify cross-sell mailings, even when these generate additional sales in only one in a thousand cases. However, comparatively small segments are often responsible for most or all of the extra sales generated. We have used uplift models to identify such segments, resulting in increases in campaign profitability by up to five times.

Graph 3: Banking Cross-Sell



Portrait is not alone in having concluded that direct marketing is best targeted by predicting uplift. We have been approached by a number of large, successful banks who have themselves realized this, and in some cases have tried numerous techniques for modeling uplift, without much success. In every case, Portrait's approach to uplift modeling has far exceeded the performance of the best alternative.

A typical scenario involves cross-selling activity aimed at increasing product holding. The value of many banking products is high, so that even an increase in product take-up as low as a tenth of a percentage point can provide a positive return on investment for mailings. However, we have shown that with appropriate targeting, we can usually achieve between 80% and 110% of the same incremental sales while reducing mailing volumes by factors ranging from 30% to 80%. Because the banks in question have themselves attempted to model uplift, they typically have historical data allowing full longitudinal validation[6] of results.

One of the complicating factors in these scenarios is that the uplift from the campaign is usually significantly smaller than the natural purchase rate of the product being promoted – typically by a factor of five to twenty. It is therefore not unusual to see a purchase rate in the control group of around 1%, and in the treated group of 1.1%. However, drivers of the base purchase rate are often quite different from those of the incremental purchases resulting from the campaign. Because of this, non-uplift approaches to targeting are often doomed to failure, and sometimes actually perform less well than random targeting.

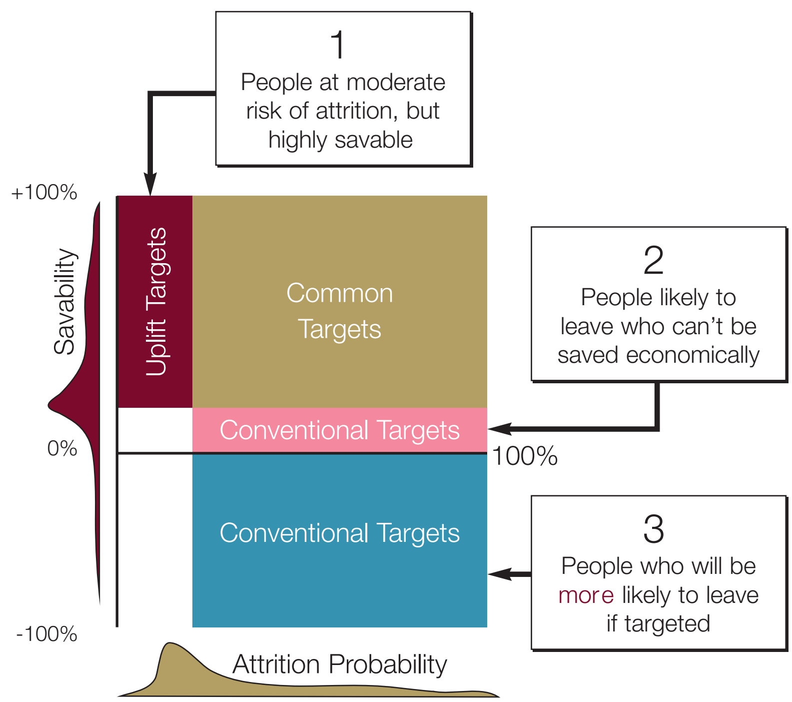
Graph 3 shows an example of one such campaign. Here, the net effect of the campaign was to increase the uptake of the product by a quarter of a percentage point. However, the uplift model shows that over 60% of the increase in sales comes from just 10% of the targeted population, 90% comes from 40% of the population and 99% comes from 70% of the population. Notice also that the "champion" model produced substantially worse results than random targeting – in fact, in this case, reversing the ranking would have been very much more effective than using its actual output. This suggests that this campaign is being effective in stimulating demand from the very people who tend not to purchase without intervention.

[6] Models are often validated by dividing a single historical dataset into two random subsets, building on one and measuring the quality of the predictions on the other. With a longitudinal validation, data from two different time periods is used. Models are built using data from the older time period and validated using data from the more recent time period. This constitutes a stronger validation because it verifies that the models identify patterns that are stable over time and really can be used to improve targeting in real-world situations.

Comparison of Targeting Strategies

Conventional models wastefully target some customers whose outcomes would be almost as good *or better* if not treated. Uplift models identify an extra group of customers whose outcomes can be profitably improved but who are overlooked by conventional models because the absolute level of their outcome falls below some arbitrary threshold.

Graph 4: Uplift vs. Conventional Targeting



A good way to get a feel for the difference between the approach encouraged by uplift modeling and conventional targeting is to consider the "swap sets" – customers who are targeted under one approach but not the other. Revisiting retention, the fundamental contrast is between targeting customers at high risk of attrition and targeting those whose likelihood of attrition can be materially reduced.

Graph 4 shows the probability of attrition[7] on the horizontal axis, and savability – decrease in churn likelihood when treated vertically. We can identify three swap sets.

1. Uplift models target a set of people whose probability of attrition is not huge, but who can be made much less likely to leave[8] (segment 1, top left, burgundy). Retention activity can be very effective here. In the context of demand-generation, the corresponding people are those whose probability of purchase when targeted is not high, but who are unlikely to purchase at all without inducement.
2. The second swap set consists of people who are at considerable risk of attrition, but are unlikely to be saved by our action (segment 2, thin pink strip above the horizontal axis). Trying to retain them costs more than the likely return. In a demand-generation setting, the corresponding people are good customers, who spend money, but whose behavior is largely unaffected by our actions. Uplift models discourage targeting these groups while conventional models recommend them.
3. The third swap-set is segment 3 (blue, below the axis), which is targeted using the conventional approach, but not an uplift approach. This consists of people whose outcomes are worse if treated, so we actually spend money to drive customers and revenue away in this segment. There seems no argument at all for such expensive and counterproductive activity.

[7] When *not* treated.

[8] Of course, everyone's probability of being saved is limited to their probability of leaving, so the disparity can never be huge. In reality, therefore, a portion of the top left part of this schematic is empty.

Response Codes, Incentives, Causation and the Critical Importance of Control Groups

Response codes do not prove that a customer purchase is caused by the associated marketing action. This is especially true when incentives are offered, as customers have a strong motive for using discount codes even when they would have purchased anyway. Only control groups allow reliable proof of causality.

Rational marketers want to understand which of their actions work best, and in what circumstances. One of the methods used to increase traceability is that of response codes, and their adoption has undoubtedly been of benefit. Unfortunately, however, response codes are not without their problems.

When a customer quotes a response code or uses a coupon, this clearly shows that he or she has been exposed to, and has taken some note of, the associated marketing intervention. Indeed, it is likely that a substantial proportion of transactions or other outcomes tied to a response code were caused by the intervention itself. However, in general it would be wrong to assume that *all* outcomes connected with a given response code are incremental.

Consider two situations. In the first, a little known company launches a new product in a new category. Its only marketing is a direct mail piece and there is no media coverage. In this situation it seems clear that all purchases *apparently* in response to the mailing are *actually* caused by the mailing, because there is no other way for a prospective customer to know about the product.

Contrast this with the situation in which a well-known credit card company, with a large spend on mass advertising, sends out a direct mail piece. The fact that a customer applies for a card using a mailed application form may mean that the mailing was crucial to that customer's recruitment; but it may not. There are many influences on that customer – including competitor behavior – and the mailed material may simply have been the easiest way to carry out a decision made on the basis of one or more of those other influences. As with the banking examples discussed earlier, we commonly see apparent "response" rates of 1% in this sort of situation, but uplifts of more like 0.1 percentage point (e.g. an account opening rate of 1.1% in a treated group against 1.0% in a control group).

The potential for being misled about the effectiveness of marketing is largest when a campaign involves a discount or other incentive. Most people will, if they are aware of a discount, use it even if they would have purchased the same product anyway. Attributing a response to a campaign on the basis of a discount coupon or response code associated with a discount, is therefore particularly dangerous, as there is a strong motivation for the customer to transact using that code even when there is absolutely no causality.

The conclusion, therefore, is that rigorous use of control groups is the only safe way to assess the true impact of a marketing campaign.[9]

[9] Ideally, two control groups should be used. The first, which we call the "mailing" or "action" control group, is the normal one – a randomly chosen group of people who meet the selection criteria but who are omitted from the campaign to allow the uplift to be assessed. The second, which we call the "targeting" control group, is a randomly-selected group of people who do *not* meet the targeting criteria but are treated as part of the campaign. This second form of control group is critical to allowing the effectiveness of the targeting – as opposed to the treatment action – to be assessed and improved.

Negative Effects

Negative effects are real and costly. They are most clearly seen in some retention activity, but can also be hidden in marketing campaigns that are successful overall. They both reduce revenue and increase costs, so merit special attention.

There is no question that certain marketing actions have negative effects on some customers. This is most clearly illustrated by the surprisingly common experience of mobile telecommunications operators running retention campaigns that demonstrably increase overall churn. These effects are actually quite easy to understand, as we shall explain below. It is less clear how often campaigns with a positive overall impact contain significant segments within which there is a negative effect. Where this does occur, it not only drives up campaign costs, but actually leads to a lower *gross* outcome than could be achieved by targeting a smaller volume.

It's not hard to understand why negative effects often occur in retention activity. The conventional approach first identifies customers at high risk of attrition. By and large, these are people who are already dissatisfied in some way. An intervention, particularly by phone, runs a high risk of the customer's responding by asking for immediate cancellation – bringing forward some attrition and crystallizing other cancellations that may otherwise not have happened at all because of customer apathy.

It is also the case that certain kinds of intervention run intrinsically high risks of negative responses for some customers. Intrusive communication mechanisms, like phone calls, are often received badly. Niche communications, designed to appeal to a specialist group, may well backfire on people who do not belong to that group, or do not conform to that group's stereotype. One creative's "ingenious idea" is another customer's "inappropriate communication".

One user of Portrait's Uplift Modeling reports that even in demand-generation applications (cross-selling, deep-selling and up-selling) there are almost always material negative effects in the last one or two deciles by uplift score, and on this basis would never advocate a 100% mailing, no matter how unarguably attractive the offer seems.

The financial impact of negative effects can be large, because they not only cause a campaign to generate fewer incremental sales or saved customers than it might: they also increase target volumes and costs.

Building Uplift Models

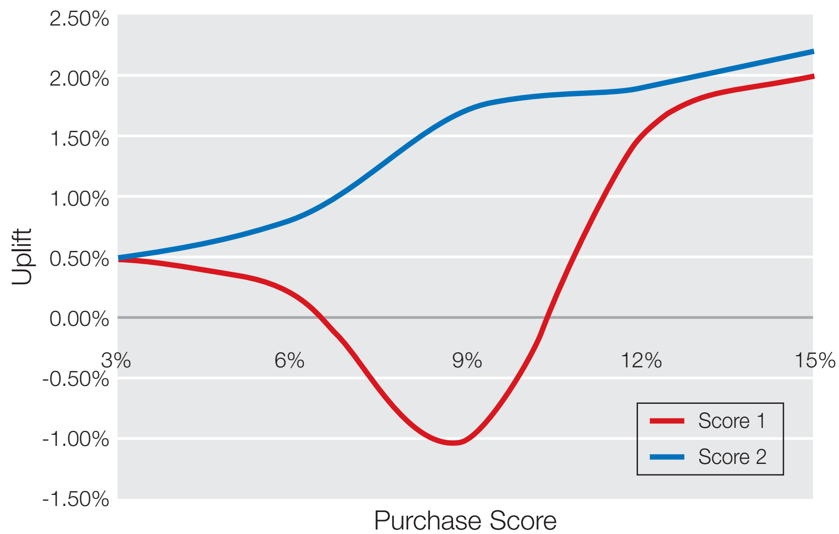
There are a number of ways to use conventional models to improve targeting for uplift. However, these approaches rarely match the performance of models whose goal during fitting is to predict uplift.

This paper argues that direct marketing activity should be targeted so as to maximize the desired change in behavior of customers for every unit of marketing spend. Once this has been accepted, there are a number of possible approaches. We discuss these in increasing order of sophistication.

1. *Target with a conventional model but keep a control group.* Needless to say, this is now standard practice, and does not constitute uplift modeling, but at least allows assessment of the true impact of marketing after the intervention. If post-campaign analysis measures uplift on a per-segment basis, variations may even allow improved targeting of future actions.

2. *Build a conventional model but assess for uplift.* A subtle variation involves building a conventional model, but then assessing uplift as a function of the score. If the uplift increases monotonically[10] with score, as is the case for Score 1 in Graph 5 (upper, blue), we would feel fairly comfortable targeting on the basis of the purchase score. If, however, it were more like Score 2 (lower, red), we might be more nervous, and would likely choose not to target customers with purchase scores between 7% and 10.5%.

Graph 5: Uplift as a function of Purchase Score



3. *Build two models and subtract.* Perhaps the most obvious approach is to build two models, one for the treated population and one for the control population. These can then be subtracted to give an estimate of uplift. This is a fairly common technique among practitioners who have realized the need for uplift modeling, and it sometimes works well. However, the fundamental weakness of this approach is that the two models are independent, each fitting a different population. Nothing specifically encourages the models to fit the *variation in uplift*. This is true at all stages of the modeling process. Different variables may drive uplift from purchase. Different interactions between variables may occur. Errors may occur in different places in the models, introducing variation in predicted uplift where there is none in reality. Quite simply, when we perform two independent regressions, we are not even *asking* the models to fit the difference in outcomes, so there is little reason to suppose that they will.

4. *Directly fit an uplift model.* Naturally, the most powerful uplift models are those that are optimized (fitted) on the basis of uplift. What is really needed is an approach in which the fitting process directly uses information from both the treated and the control populations and optimizes the fit of the model to the difference in behavior – the uplift. This is precisely the approach that Portrait's Uplift Modeling solution takes.

[10] i.e. if higher predicted outcomes always correspond to higher levels of uplift.

Portrait Software's Uplift Modeling Solution

Portrait Software markets the world's only uplift modeling software, and has unique experience in this field.

Quadstone, now part of Portrait Software, has pioneered uplift modeling, first releasing software containing such capabilities in 1998, and first publishing a paper on the approach in 1999. As we have worked with customers and gained more experience with the techniques, we have moved through three generations of approach, each progressively more sophisticated and powerful.

Portrait's approach begins by modifying standard, greedy tree-building algorithms so that they look at both a treated and a control population and directly model the difference. This involves modifying both splitting and pruning criteria, as well as the performance measure for the model itself.

Portrait is also acutely aware that theoretically desirable data volumes are often not available in one or other of the treated and control segments. (For roll-out campaigns, control groups are often small; for trial campaigns, treated volumes tend to be limited.) Moreover, even when overall outcome rates are high, overall uplift is often low, and sometimes zero or negative.
We have therefore added a large number of features to the software to allow modeling for small populations, small effects and high noise levels. We are not aware of any other company or group with comparable experience in this area, or any other software with packaged uplift modeling capability.

Portrait's Uplift Modeling solution is packaged software that can be used as a stand-alone capability, as an integrated component of the award-winning Quadstone System or as an integrated component of a SAS environment.

About Portrait Software™

Portrait Software™ is a global provider of Customer Interaction Optimization software that helps B2C organizations provide a great customer service, gain deep customer insight and sell more, profitably.

Portrait Software™ solutions combine analytical capabilities with a process-centric customer interaction management platform, providing organizations with the ability to deliver the most highly optimized customer processes across all sales and service channels.

Portrait Foundation™ is the only specialized Microsoft based Application Platform Suite that enables organizations to quickly build process centric, multi-channel interaction management applications.

Portrait Interaction Optimizer™ enables organizations to exploit customer insight, evaluate in real time the "best next action" and automate the follow-up processes to maximize the value of each and every customer interaction.

Portrait Customer Analytics™ delivers analytic solutions that allows large consumer organizations to explore and exploit customer insight. It combines multiple sources of data to graphically discover, model, and predict trends and relationships in customer behavior in a collaborative environment.

Our 200+ customers are organizations that lead the world's most demanding customer-intensive sectors, including Financial Services, Public Services, Telecoms and the Independent Software Vendors that support these markets. Our customers include KeyCorp, Wells Fargo Bank, Fidelity, Washington Mutual, Lloyds TSB, Merrill Lynch, Nationwide Building Society, Rainier Pacific Bank, Chelsea Building Society, British Telecom, China Automobile Association, T-Mobile, and Fiserv CBS.

*In December 2005, Portrait Software acquired Quadstone, a company recognized as best-of-breed for customer analytics. The acquisition provides Portrait with the ability to offer its customers real-time intelligence and decision management tools for marketing, sales and service applications. The Quadstone product suite has been integrated into the Portrait Collaborative Customer Analytics offering as well as other products in the Portrait Customer Optimization suite.

Contact Details

Europe

Portrait Software
The Smith Centre
The Fairmile
Henley-on-Thames
Oxfordshire, RG9 6AB, UK

Email: contactus@portraitsoftware.com
Tel: +44 1491 416600
Fax: +44 1491 416601

Americas

Portrait Software
125 Summer Street
16th Floor
Boston, MA 02110
USA

Email: contactus@portraitsoftware.com
Tel: +1 617 457 5200
Fax: +1 617 457 5299

Asia Pacific

Portrait Software
10 Faber Park
129101
Singapore

Email: contactus@portraitsoftware.com
Tel: +65 9752 1759
Fax: +65 6872 6764

www.portraitsoftware.com