White Paper

Uplift Modeling: Making Predictive Models Actionable

by Mike Thurber
Lead Data Scientist

April 2017

ELDER RESEARCH
DATA SCIENCE & PREDICTIVE ANALYTICS

Uplift Modeling: Making Predictive Models Actionable

Table of Contents

Predictive models typically estimate the likelihood of future events, such as whether it will rain tomorrow or which customers are most likely to "churn" by cancelling their phone contract. In the case of the weather, we do not expect to change it; we just want to know how to adapt. However, the goal for most use cases is to be more proactive; we want to understand what action to take to change the outcome in a favorable way. (See the examples in Figure 1) In these cases *prescriptive*, not just predictive, analytics is required. The return on investment comes directly from knowing the impact of alternative treatments. By knowing the impact of each treatment, resources can be targeted where they will be most effective and withheld where they will have negligible effect or worse, have a negative effect. This great objective of data science, to intelligently drive day-to-day business decisions based on data, is the purview of uplift modeling. This white paper will explain what uplift modeling is and why it can be much better than directly modeling the outcome.

Predictive Models vs. Prescriptive Uplift Models

There are many applications of predictive modeling where the outcome is predicted as advice only to a human decision maker, and no action is directly taken automatically from the model result. An example is workload prioritization. For example, in the telecom industry we can predict which customers are most likely to churn (cancel their contracts). In healthcare we can predict which patients are most likely to recover. For universities or charitable organizations, we can predict which prospective benefactors are most likely to donate.

Sometimes this is sufficient. For example, if the outcome of our weather prediction is that it is likely to rain, we take an umbrella. Since we can't change the weather we can only be better prepared for it.

But where we can, we aim to *influence* the outcome one way or another. Will a live agent offering the phone customer a contract upgrade decrease their likelihood to churn? Will soliciting a fund raising prospect with a flyer in the mail improve their chances of making a donation? Will offering a moving bonus increase the likelihood that a desirable candidate will accept our employment offer? The most common example where uplift modeling has taken hold is in retail marketing where the goal is to predict not the likelihood of a customer buying, but what can be done to *increase* the likelihood of them making a purchase.

Use Case	Targeted Outcome	Treatment
Outdoors	Weather	Bring Umbrella
Phone Customer	Not Churn	Offer Upgrade
Patient	Recover	Treat
Voter	Vote	Message
Donor	Donate	Solicit
Prospect	Join	Invite
Candidate	Hire On	Pay Moving Bonus
Inmate	Not Relapse	Coach
PSTD Veteran	Not Self-harm	Counsel
Gas well	Not Shut-in	Insulate
Retail Customer	Buy	Offer Sale

Figure 1. Predictive modeling examples

The salient knowledge sought is the impact of the treatment, not the estimate of the outcome. For instance, would you rather spend campaign dollars trying to persuade your most loyal supporters (those with the highest probability of "buying"), or on the voters who will be swayed the most by an additional engagement? Simply predicting the expected outcome is not sufficient to optimize your use of money and resources. A few elections ago I was determined to vote for a particular candidate, who meanwhile, kept filling my mailbox with campaign material. Even though my publicly available data should have demonstrated that I was already a sure vote they could count on, they wasted many glossy flyers on me.[1]

Uplift is Not Directly Measurable

Uplift modeling is also known as incremental modeling, treatment effects modeling, true lift modeling, or net modeling. Uplift is the increase in likelihood of the outcome *with* the treatment as compared to the outcome *without* the treatment. We can't observe this difference, or causal effect, directly, but must infer it from an experiment.

Eric Siegel's book *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die* devotes a chapter with excellent case studies showing why it is important to have uplift modeling in your data science tool kit and to use it appropriately. It is very helpful to visualize a 2x2 matrix, as shown in Figure 2, with four categories of people (say) to be classified, as: (a) Persuadable, (b) Sure Thing, (c) Do-Not-Disturb, and (d) Lost Cause as shown in figure 3.

To promote a desired response we target the "a" population – those who are *Persuadable*. For all others, the treatment is wasteful or, for the *Do-Not-Disturbs*, it is actually counterproductive. Contacting the *Do-Not-Disturbs* may result in the customers acting in the exact opposite way that was intended – it can be perilous to "wake a sleeping dog".

Uplift modeling's objective is to find *Persuadables*. Of course, uplift modeling can apply to any modeled outcome, human or not, such as the effect of fertilizer on crop yields or sending email messages in political campaigns. Again, where traditional predictive modeling focuses on the outcome, uplift modeling focuses on the effectiveness of the treatment. Then, you can target resources on the cases that are likely to be positively impacted by the treatment.

Response if Treated	N	**Do-Not-Disturb** ***c***	**Lost Cause** ***d***
	Y	Sure Thing *b*	Persuadable *a*
		Y	N
		Response if <u>not</u> treated	

Figure 2. Uplift model matrix.

Estimating Uplift

2,000 Potential Churners	
Offer free upgrade	No offer
400 random test accounts	1,600 control accounts
8 churns	160 churns
392 non-churns	1,440 non-churns
R_T: 2% churn	R_C: 10% churn
U: Offer had -8% uplift	
Churn Odds Ratio: 5.4	
Odds Ratio 99% CL: 2.12 to 14.00	
Uplift 99% CL|R_C=10%: -5.0% to -9.2%	

Figure 3. Telecom Uplift model example.

Consider a telecom example of trying to prevent customer churn as shown in figure 3. The treatment is to offer an upgrade to a customer who is a potential churner. To perform uplift analysis, we conduct an experiment with 400 randomly selected test accounts to whom we offer a free upgrade, and a control group of 1600 accounts that receive no offer. (It is common to have a larger control group as it is less expensive). In this experiment, we record 8 churns in the group that received an offer, and 160 churns in the group that did not receive an offer. This means that there is a 2% churn in the experimental group (R_T) and a 10% churn in the control group (R_C). The offer has a -8% uplift (U):

$$\text{Overall Uplift } U = R_T - R_C = 2\% - 10\% = -8\%$$

The uplift in this case is negative because we are trying to avoid the target behavior rather than promote it.

For Uplift to be actionable in practice, we also need to know the treatment effect for each individual person uniquely, in addition to the general population. For example, my previous volume of online shopping may indicate that I am more persuadable to click on a particular advertisement than others in my same demographic group. Thus, we want to *model* how the attributes of a case impact the treatment uplift of that case. The way such a model is created in practice is as follows:

1) predict the outcome with the treatment applied (R_{Ti} in the telecom example),
2) predict the outcome without the treatment applied (R_{Ci} in the telecom example),
3) calculate the difference in the rates as the uplift ($Ui=R_{Ti}-R_{Ci}$), and
4) compute the upper and lower 95% confidence limits on Ui.

Once these values are calculated, individuals can be allocated to the four quadrants of the treatment effect matrix using these rules:

- If the confidence limits of the Incremental Uplift (U_i) includes zero, the treatment effect can be thought of as unknown and not significant. Regardless of treatment, the *Sure Things* have a high outcome likelihood and the *Lost Causes* have a low outcome likelihood.
- If the Incremental Uplift (U_i) is significantly greater than zero, the predicted outcome increases because of the treatment. These are the *Persuadables* if the outcome is positive.
- When the Incremental Uplift (U_i) is significantly less than zero, the predicted outcome is less likely because of the treatment. Traditionally, these are called the *Do-Not-Disturbs*.

Remember, of course, that this is a modeled estimate of a, b, c, and d, and not every persuadable individual will actually be persuaded by the treatment.

Achieving a Return on Marketing Investment

Model Score Decile	Response Rate	Lift
1	28.1%	3.4
2	17.3%	2.1
3	9.6%	1.2
4	8.4%	1
5	4.8%	0.6
6	3.9%	0.5
7	3.3%	0.4
8	3.4%	0.4
9	3.5%	0.4
10	0.1%	0

Figure 4. Retail targeting Uplift model example.

In business it is always important to understand the return on investment for taking a course of action (applying a treatment). Uplift modeling enables you to estimate the expected return on treatment by summing the Incremental Uplift of those persuadable by treatment, which is the overall estimated treatment effect. Consider this retail targeting example from analyticbridge.com where a purchase propensity model output was used to generate a campaign direct mailing list. As shown in figure 4 the traditional predictive model was very accurate. There, the response rate of the highest decile (the top ten percent as defined by the model score) is 281 times that of the bottom decile—**a huge relative lift**.

The third column shows the lift of each decile over the base response rate of 8.2%, so the top three deciles are seen to have greater than average propensity to buy.

It is natural, at first, to use the predictive model directly, and want to promote the product to those predicted to be most likely to buy. (We will see the better way shortly.) Doing this, they mailed the promotion to (that is, treated) randomly selected persons in the first four deciles. They became the test group; and the rest, the control group, received no mailings. The top table in figure 5 compares the response rates for the two groups, and we see that the treatment was not helpful.

Model Score Decile	Test Group Response Rate	Control Group Response Rate	Incremental Response Rate
1	26.99%	27.90%	-0.91%
2	20.34%	20.90%	-0.56%
3	10.70%	10.04%	0.66%
4	8.90%	7.52%	1.38%
Deciles 1-4	**16.73%**	**16.59%**	**0.14%**

Uplift Model Score Decile	Test Group Response Rate	Control Group Response Rate	Incremental Response Rate
1	18.80%	12.90%	5.90%
2	7.80%	5.40%	2.40%
3	6.90%	4.50%	2.50%
4	4.30%	3.60%	0.70%
Deciles 1-4	**9.45%**	**6.60%**	**2.88%**

Figure 5. Top table shows retail targeting model results using a predictive model. The bottom table shows the Uplift model results.

In some cases, the control group bought more often, and overall, the response rate in the test group was only 0.14% higher. This is undoubtedly because many of those whom the model predicted were likely to buy were "sure things" and were going to buy anyway. The promotion did not effect any change worthy of its cost.

However, if we take the predictive model scores, and do the further work necessary to create an uplift model we can then rank prospects by their *uplift score*, and put each person in deciles by that score. Now, treating the top four uplift model deciles, as shown in the bottom table of figure 5, reveals an incremental response rate improvement of 2.88%. ***This return on marketing investment is 20 times better!***

Benefits of Uplift Analytics

Uplift analysis models the effect of treatment, rather than the outcome directly. If we know how likely something is already, and how likely we are going to be able to change it with a treatment, we can classify prospects as either "sure things", "persuadables", "lost causes", or "do not disturbs". This is extremely valuable as a way to get the most out of ones analytics investment.

---

[1] The past several close Presidential elections have been won by the campaign with the more intelligent use of voter data, and uplift analysis has played a large role.

About the Author

Lead Data Scientist Mike Thurber is an expert data analyst, comfortable with diverse data sources in diverse industries, and extracting relevant and valuable insights from available data. His modeling work ranges from predicting high payouts on long term care claims to identifying healthcare provider fraud to measuring the effect of Cesarean delivery on infant health. His broad experience managing a variety of analytic initiatives consistently generates business value through expert collaboration, data integration, insightful data analysis, statistical testing, and predictive modeling. Other examples include gleaning insights on how complex consumer choices impact sales, predicting profitability of prospective customers, calculating fraud and financial risk of many kinds, showing how call center interactions affect customer retention, forecasting recovery of losses due to loan default, modeling maintenance events on natural gas wells, and predicting propensity to make voluntary monetary donations. Mike earned a BS degree in Chemical Engineering from Brigham Young University and a Master's degree in Statistics from Virginia Commonwealth University.

www.elderresearch.com

ELDER RESEARCH
DATA SCIENCE & PREDICTIVE ANALYTICS

National Capital Region
2101 Wilson Boulevard
Suite 900
Arlington, VA 22201

855.973.7673

Headquarters
300 W. Main Street
Suite 301
Charlottesville, VA 22903

434.973.7673

Maryland Office
839 Elkridge Landing
Suite 215
Linthicum, MD 21090

855.973.7673