Estimating the Benefit of Targeting

Gaurav Sood*

December 15, 2018

What is the benefit of targeting? Why (and when) do we need experiments to estimate the benefits of targeting? And what is the right baseline to compare against?

I start with a `business casual` explanation, using examples to illustrate some of the issues at hand. Later in the note, I present a `formal` explanation to precisely describe the assumptions to clarify under what conditions targeting may be a reasonable thing to do.

Business Casual

Say that you have some TVs to sell. And say that you could show an ad about the TVs to everyone in the city for free. Your goal is to sell as many TVs as possible. Does it make sense for you to build a model to pick out people who would be especially likely to buy the TV and only show an ad to them? No, it doesn't. Unless ads make people less likely to purchase TVs, you are always better-off reaching out to everyone.

You are wise. You use common sense to sell more TVs than the guy who spent a bunch of money building the model and selling less. You make tons of money. And you use the money to buy Honda and Mercedes dealerships. You still retain the magical power of being able to show ads to everyone for free. Your goal is to maximize profits. And selling Mercedes nets you more profit than Hondas. Should you use a model to show some people

*Gaurav can be reached at: gsood07@gmail.com

ads about Toyota and other people ads about Honda? The answer is still no. Under likely to hold assumptions, the optimal strategy is to show an ad for Mercedes first and then an ad for Toyota. (You can show the Toyota ad first if people who want to buy Mercedes won't buy a cheaper car if they see an ad for a cheaper car first.)

But what if you are limited to only one ad? What would you do? In that case, a model may make sense. Let's see how things may look with some fake data. Let's compare the outcomes of four strategies: two model-based targeting strategies and two target-everyone with one ad strategies. To make things easier, let's assume that selling Mercedes nets ten units of profits and selling Honda nets five units of profit. Let's also assume that people will only buy something if they see an ad for their preferred product.

Table 1: Expected Outcomes From Different Targeting Strategies

Name	Income	Preference	All Mercedes	All Honda	Good Model	Bad Model
Brian	300k	Honda	0	5	5	0
John	500k	Mercedes	10	0	10	0
Matt	300k	Honda	0	5	5	5
Gui	400k	Mercedes	10	0	10	10
Gaurav	200k	Tesla	0	0	0	0

As Table 1 shows, showing everyone an ad for Mercedes yields a profit of 20 units and showing everyone an ad for Toyota yields a profit of 10 units. Now say you hire two different companies to develop models based on the one thing you know about the people—their income. One company develops a good model and another one, a bad model. The profit from the good and bad models is 30 and 15 units respectively. Thus, targeting with a good model generates an additional profit of 10 units compared to the best performing target-all strategy while targeting with a bad model yields you five fewer units of profit than the target-all strategy.

You are wise. You test the performance of different methods using an experiment and pick the winning horse. You make a ton of money. A few months later, however, you find out that because of some changes, you can no longer show ads for free. Simultaneously, you find out that your customers have access to more information. So it is no longer true that they will only buy something if they see an ad for their preferred product. For instance, Brian could find out about Honda on the Internet and buy it from your dealership.

If the goal is to maximize profit, would a targeting model help? To judge that, we first need to think about how to learn the targeting model. The aim is to maximize profit. But we don't know what Brian will do without seeing an ad. To learn which ad (or no ad) delivers the highest ROI, we need to run an experiment. And we have to use the data from the experiment to estimate a model for which ad (or no ad) delivers the greatest ROI for each person. Once we have a model, we can use it to show ads to people where the expected payoff is positive and show them the ad with the highest expected ROI.

Formal

To answer the questions posed at the top more precisely, we need to formalize a bit. Formalization is there so that we can be precise about our assumptions about the moving parts.

To make formalizations accessible, let's stick with the use case of sending out mailers. A company makes m total products and users of one of its products don't use all of its products. This gives the company an opportunity show ads about the products that the user is not using. Let's also assume that the company knows a lot about each of its users so it can build reasonable models if it chose to.

To formalize, it helps to have short stand-ins (symbols) for words—notation.

Let's use i to iterate over n total users. And let's use j to iterate over m total products. And let w_{ij} refer to the expected profits if sold product j to user i using a particular ad for the product. (The true profits are unknown and must be estimated. Technically, we have $w_{true_{ij}}$ and $w_{est_{ij}}$.) Let's also assume that there are two costs of showing an ad to a user—the cost of showing an ad, which is born by the company, and a hit to the product experience (the product that the user is using) which impacts the long-term profitability of the company. We expect the cost of showing an ad to be roughly constant across users and products so let's denote it with a constant c. We expect the cost born by the user to vary by how irrelevant the product is to them and how disruptive the ad is and how sensitive people are to disruption. Let's denote that cost by r_{ij}. (Once again, you would break it into two things. There is a true unobserved cost, and there is an estimated cost.) Finally, let's assume the probability the user will buy the product if we show them an ad for it to be $p_{true_{ij}}$. And let $p_{est_{ij}}$ denote the probability we estimate from data.

The optimal strategy for the company is to show ads to as many people for whom:

$$p_{est_{ij}} * w_{ij} - c - r_{ij} > 0 \tag{1}$$

If the company only gets to show an ad about one product, the optimal strategy is to take the maximum of $p_{est_{ij}} * w_{ij} - c - r_{ij}$ for each user and show ad j that provides the greatest profit.

But is this strategy any good? It depends on the quality of p_{est} (and how well we estimate w_{ij} and r_{ij}). To test how good p_{est} is, we still need to experiment. But before we experiment, we must think through the baseline to compare against.

Given the core points we discuss next generalize to more complex scenarios easily, let's ignore r_{ij} and w_{ij} for now.

Say the company is not forward thinking and cannot borrow money. This means that the company has a fixed budget for showing ads, which means that it can only show ads to a certain number of people. (This is not something that a company is generally faced with but say that it is.) Also, assume that the company only makes one product. You want to learn whether it is better to select users with Model A (which can be: select users randomly) or with Model B. For that, the experiment to run is as follows: draw a random sample of users and randomly assign them to treatment and control. Then posthoc, estimate what is the expected profit if you subset on users selected by Model A versus users selected by Model B. The problem, in effect, reduces to the measurement of heterogeneous treatment effects—are the effects of showing an ad greater (the baseline is organic adoption in the control group) among some people than others. Non-profits often face a problem analogous to this—should we invest money to develop a targeting model which we could then use to "optimally" allocate money or should we just allocate money randomly? But, as I note, this is not a problem that companies often face. (Note also that I have introduced a new cost to our setup: the cost of targeting, above and beyond the cost of serving an ad.)

If the company doesn't have a budget constraint and does not have numbers for r_{ij} and w_{ij} and wanted to measure the benefit of targeting, and was pitching just one product, the appropriate baseline is to compare the profit from reaching out to everyone than to a few people. (In this trivial case, reaching out to everyone will always win.)

Often, however, a company will be rational, able to borrow (and you can model borrowing costs in the equation to reflect the true c), and will have more than one product to pitch. Then the question is not about whether to show only some people ads but which ad(s) to show to which people. There, again, you should randomize, show ads based on one model to one group and based on another model to another group and measure the total profit in both groups. Another no-targeting baseline to beat is the maximum profit you can get by showing everyone an ad for one product or the other. Generally, however, we would use neither of these baselines. The baseline would be profit from business as usual, which would include some rules-based-targeting, etc.