Targeting 101

Gaurav Sood

June 22, 2018

Targeting Economics

Say that there is a company that makes more than one product. The company can run an ad in one of its products about the one or more other products it produces that a user doesn't use. Should it consider targeting—not showing the same ad to all users? There are six aspects to consider:

1. **Opportunity Cost:** Could the company make more profit by showing an ad for something else?

2. **Cost of Showing an Ad to an Additional User:** The cost of serving an ad; it is close to zero in the digital economy.

3. **Cost of Worse Product:** An ad for an irrelevant product lowers the user's welfare. (The magnitude of the reduction depends on how disruptive the ad is and how irrelevant it is.) As a result of seeing an irrelevant ad in the product, the user likes the product less.

4. **Poisoning the Well:** Showing an irrelevant ad means that people are more likely to skip whatever ad you present next. It reduces your ability to monetize future ads.

5. **Cost of Not Learning About the Relevant Product Sooner and Investment in Learning About an Irrelevant Product**: the cost of not learning about a product they could use sooner. Plus the investment a user makes in learning about a product that is not relevant to them.

6. **Profits:** On the flip side of the ledger are expected profits. What are the expected profits from showing an ad? If you show a user an ad for a relevant product, they may not just buy and use the other product, but may also become less likely to switch from your stack. Further, they may even proselytize your product, netting you more users.

Targeting Math

Say a company makes k different products and mantains a list of all potential customers. To keep things simple, let's assume that we can show each potential customer just one ad. And that showing them an ad costs nothing to the company or the user. Again, for simplicity, let's assume that customers won't buy the product if they don't see an ad for it. Given this simple setup, if the goal is to maximize profit, what is the optimal targeting strategy?

We need a little more notation before we can answer the question. Let i track which of the k products we are referring to. And let's denote the profit we make from selling the ith product by w_i. So, the profits from the k products are $w_1, ..., w_i, ...w_k$. For each user, there is an unobserved probability that they will buy a product (relevant or not) if shown an ad for it $p(buy|ad)$. If j iterates over the n users, we denote the true unknown underlying probability user j will buy product i if shown an ad for it by $p_{true_{ij}}$. And let's denote the probability the user will buy the product if shown an ad that we estimate from data by $p_{est_{ij}}$.

If the company has only one product to sell ($k = 1$), the optimal strategy is to target everyone. If the company has more than one product to sell, the company has to decide about which product to pitch to whom. To simplify, let's assume that the company has only two products to sell. (With minor amendments, the math can be generalized to cases where the company has more than two products to sell.) In the absence of data, the optimal strategy is to pitch everyone the product that yields greater profit $max(w_1, ..., w_i, ...w_k)$. And the most profit the company can make is:

$$\max \begin{Bmatrix} \sum_j w_1 * p_{true_{1j}} \\ \sum_j w_2 * p_{true_{2j}} \end{Bmatrix} \tag{1}$$

If we have data, we can generally do better. If the company has two products to sell, a customer can either buy $i = 1$ or $i = 2$ or buy nothing at all. So, for each person, using data, we can generate two numbers $p_{est_{1j}}$ and $p_{est_{2j}}$. (These probabilities won't sum to 1. Neither product may be relevant to the user. Or both may be.) And using the two probabilities, we can find out: a) what product to pitch to each person, b) the order in which to pitch (if running ads is costly).

First, pitch the product for which $w_i * p_{est_{ij}}$ is the highest. Second, the order to pitch is based on how big $w_i * p_{est_{ij}}$ is. The estimated profit if we use targeting is:

$$\sum_j \max(w_1 * p_{est_{1j}}, w_2 * p_{est_{2j}}) \tag{2}$$

Calculating the Benefits of Targeting

When may the number in 2 be greater than 1? The logic is pretty simple. Say the 2nd product yields more profit. The no targeting estimate is thus, $\sum_j w_2 * p_{true_{2j}}$. For any user j where $w_1 * p_{est_{1j}} > w_2 * p_{true_{2j}}$, targeting estimate swaps out the numbers. If this isn't true for any user, the targeting estimate is the same as no targeting estimate.

To the extent that estimated probabilities are negatively correlated with the actual probabilities, we would do worse.

Since we never have the true probabilities, the only way to estimate the benefit of targeting is to compare random targeting to model based targeting. Pick fixed size samples using random sampling and using random targeting. And compare the profits across both.

User Welfare

Until now, we have talked about profit maximizing targeting, assuming cost to the user being zero. So let's add concerns about user welfare to the mix.

Let's call the cost that a user j bears when they see an irrelevant ad for product i as r_{ij}. Accounting for cost imposed on the user changes our targeting strategy. Let's start by adapting the calculation we do in 2 for each user:

$$t_j = \max(w_1 * p_{est_{1j}} - r_{1j}, w_2 * p_{est_{2j}} - r_{2j}) \tag{3}$$

For all users where t_j is positive, we target them when the ad for the appropriate product. When $t_j \leq 0$, we don't target the users, and hence get 0 profit. So the profit under targeting is:

$$\sum_j \left\{ \begin{matrix} t_j, & \text{if } t_j > 0 \\ 0, & \text{if } t_j \leq 0 \end{matrix} \right\} \tag{4}$$

But we know little about r. So let's refine our intuition about it. The cost is likely inversely related to the probability a user will buy the product when shown an ad for it. It is reasonable to assume that r is positively correlated with $1 - p_{true_{ij}}$. We don't know what that number is as the user utility probably depends on things other than relevance but it is a good start. But what is the rough magnitude of r?

Pricing Pain

In epidemiology, there is a measure called Number Needed to Treat (NNT): how many people do we need to treat to prevent one bad outcome. The corresponding metric for a relatively nondisruptive ad, an ad you need to dismiss to continue, can be: how many people need to click "dismiss" to make one sale. No one thinks that this number is a million. No one thinks it is 1. It is probably in the thousands. So one estimate of pain is .0001–.001 cent.

Estimating the cost of not finding out about the product sooner is hard to do. But it is some version of pro-rated consumer surplus. And you can estimate it using some discontinuties in prices.

Estimating the cost of investment in learning about an irrelevant product only exists for users who invest time to try out a product, read it and eventually abandon it. We could tally the time the person uses a product and then multiply it with minimum wage to get a lower bound of that cost.