Targeting 101

Gaurav Sood

June 22, 2018

Targeting Economics

Say that there is a company that makes more than one product. And users of any one of its products don't use all of its products. In effect, the company has a *captive* audience. The company can run an ad in any of its products about the one or more other products that a user doesn't use. Should it consider targeting—showing different (number of) ads to different users? There are five things to consider:

1. **Opportunity Cost:** If the opportunity is limited, could the company make more profit by showing an ad about something else?

2. **The 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. **The Cost of a Worse Product:** As a result of seeing an irrelevant ad in the product, the user likes the product less. (The magnitude of the reduction depends on how disruptive the ad is and how irrelevant it is.) The company suffers in the end as its long-term profits are lower.

4. **Poisoning the Well:** Showing an irrelevant ad means that people are more likely to skip whatever ad you present next. It reduces the company's ability to pitch other products successfully.

5. **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 maintains 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. (Technically, the profit can vary by person.) 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 opportunity to sell and only one product to sell ($k = 1$), the optimal strategy is to target everyone.

If the company has only one opportunity to sell but 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 on users that allow us to model their preferences, the optimal strategy is to pitch everyone the product that yields the greatest profit $max(w_1, w_2)$. 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 on users, we can generally do better. A customer can either buy $i = 1$ or $i = 2$ or nothing at all. So, for each person, using the 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.) Using the two probabilities, we can find out what product(s) to pitch to each person.

First, if we have one opportunity, pitch the product for which $w_i * p_{est_{ij}}$ is the highest. The estimated profit if we use a model is:

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

The actual profit may be higher or lower than the estimated profit (or for that matter the profit we get when we don't use targeting). The actual profit depends on the quality of the model. To the extent that estimated probabilities are weakly or negatively correlated with the actual probabilities, we could do worse.

But 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 $w_1 * p_{est,true_{1j}} > w_2 * p_{est,true_{2j}}$, targeting estimate swaps out the profit.

Long-term Costs

Until now, we have not assumed any costs. So let's add costs to the mix.

Let's call the cost that the company bears when user j sees an irrelevant ad for product i as r_{ij}. (The net cost is a sum of two costs: a. the cost of worse product, which may affect retention, etc., b. the cost of poisoning the well, which reduces persuadability of the next ad.) Accounting for these costs 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:

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

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.

Ideally, you want to estimate the costs empirically. The cost of poisoning the well can be estimated using the decomposition laid out by Hohnhold, O'Brien and Tang (2015).

Showing irrelevant ads can mean fewer people use the product and fewer people pay attention to the ads that are shown to them. We can measure these effects experimentally.

References

Hohnhold, Henning, Deirdre O'Brien and Diane Tang. 2015. Focusing on the Long-term: It's Good for Users and Business. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM pp. 1849–1858.