Demand Forecasting When Customers Consider, Then Choose

Srikanth Jagabathula, Ying Liu

Demand Forecasting When Customers Consider, Then Choose

Motivation

It is essential for a firm to accurately understand its customers preferences – what they like and dislike. Such understanding allows the firm to tailor its product offerings and marketing messages to the needs of its customers. Firms base this understanding on the observations of which products have been purchased in response to the products offered. Though the observations provide information about what the customers *like*, they do not immediately show what the customers *dislike*. Customers do not necessarily dislike the products they did not purchase despite being available because they often do not consider some of the products on offer. This may be because of lack of familiarity or awareness of these products or lack of fit to their current needs. Consequently, it is only reasonable to assume that the products that were considered but not purchased are disliked, whereas the products that were not considered may in fact be liked and purchased at a future time. The challenge of course is that the set of products considered – or what is called the *consideration set* – is usually latent and not readily observed. While existing techniques are available to identify consideration sets through detailed questionnaires, such techniques are expensive and not scalable.

Motivated by the above issues, we study the problem of inferring consideration sets of customers from sales transaction data, which comprise information on purchased products and the offer set at the time of the purchase. We suppose that the transactions are aggregated across all the customers. Such data are readily available in practice, typically collected from a firm's point-of-sale (POS) systems. As a concrete application for our problem, we take the view of a canonical retailer focused on a single category of mutually substitutable products. The techniques we develop are however more broadly applicable in the numerous applications in which firms are interested in identifying their competing products and services: what hotels did the customer consider before choosing a particular one, what cars did the customer consider before renting a particular one, and finally what ads did the customer consider before clicking on a particular one. Therefore, a tool to accurately predict consideration sets can clearly have a broad impact. Consideration sets have been gaining attentions in the recent operations management literature (see [1, 2, 3]).

The key challenge to inferring consideration sets from limited data, such as purchase transactions, is the associated computational burden. In general, the consideration set can be any subset of the offer set and identifying the "best" fitting consideration set requires searching over exponentially (in the number of products) many subsets. Existing approaches have focused on specific threshold-based screening rules, which specify that products whose features cross the thresholds are considered, and proposed ad hoc techniques to efficiently estimate the thresholds. These approaches are valid when the relevant features that drive customer behavior can be reasonably identified and measured. However, in several practical instances, determining the features that drive the choice behavior is challenging, especially when the product space is large and changing.

For the above reasons, we propose a principled heuristic to identify consideration sets directly in the product space, without assuming access to product features. The algorithm we propose makes several approximations to navigate the exponentially large space of subsets. In order to arrive at these approximations, we establish connections of our formulation to the work on fitting classification trees, which allows us to leverage the decades long work on practical techniques for estimation that can scale to large datasets.

Model and Algorithm

We consider the setting in which a firm is offering products to its customers from a universe of n products. The firm has collected historical sales transaction data by offering a sequence of offer sets and observing the sales for each product as a function of the offer set. Customers make choices according to a two-stage model. Each customer is described by a subset of products, called the consideration set, and a preference list of the alternatives. When offered a subset of products, they first *consider* the subset of available products that are present in their consideration sets and then *choose* the most preferred product from the considered products. Because neither the consideration set nor the preference list of the customers is observed, we suppose that a customer samples a consideration set from a distribution over all the subsets called the *consideration probability mass function (PMF)* and a preference list from a distribution over all possible preference lists, called the *preference PMF*.

In order to be able to estimate the parameters of the model from available data, we make the following additional assumptions. We assume that population of customers consists of K types with each type k described by the consideration set D_k. We suppose that the different types of customers are sufficiently separated so that the consideration sets $D_1, D_2, \ldots, D_K$ form a partition of the product universe; that is, the subsets D_ks are mutually exclusive and collectively exhaustive. The parameter α_k denotes the fraction of customers belonging to type k. Customers sample preference lists according to a multinomial logit (MNL) model with parameters $v_1, v_2, \ldots, v_n$. Our estimation procedure must estimate the parameters D_ks, α_ks, and v. Note that K is not pre-determined, and we shall estimate it from the data.

We use the method of maximum likelihood estimation (MLE) to estimate the parameters of the model. If the consideration sets are given, then estimating v is equivalent to fitting an MNL model, which can be done efficiently. However, because D_ks are latent, we propose an EM-style iterative algorithm (MLE Tree) that alternatively estimates the latent D_ks and then fits an MNL model to estimate v, until a stopping condition is met.

MLE Tree Algorithm. The algorithm starts with arbitrary initial estimates of the model parameters v. In each subsequent iteration, it first infers the latent consideration sets and then fits the MNL model to update the estimates of v. To infer the consideration sets, we construct a partition by recursively splitting the universe of products as follows.

We start with the entire set and split it into two parts so that the log-likelihood is increased. Inspired by the random forest, we do not search over all the splits on a node, but we randomly compare m splits and select the one with the highest log-likelihood values. We then recursively split each part again into two parts to further improve the log-likelihood. The process continues until the log-likelihood no long increases or the number of nodes exceeds some pre-determined threshold. The above process may be visualized as a binary tree with each split corresponding to a node and the two parts obtained from the split as the children. The leaves of the tree then correspond to the inferred consideration sets. We then update v's by fitting an MNL model. We repeat this and grow the trees sequentially until a stopping condition is met.

Results. We tested our algorithm on synthetic transaction data consisting of five hundred purchase instances that are randomly generated. The number of products is seven and the choice is made using an MNL model. We tested the algorithm on data sets from twenty ground truth models. Our results show that for 95% of the data sets, the optimal partitions could be achieved by the MLE Tree algorithm.

We also tested our algorithm on a real-world data set that contains the sales of ketchup bottles from a supermarket store, with 28 universal product codes in total. We split the data into training data corresponding to fifteen weeks, and test data corresponding to ten weeks. We used standard MNL as a benchmark and then measure their performances in terms of the root mean square error (RMSE) score in predicting aggregate market shares in the test data. Our results indicate that MLE Tree reduces the RMSE by 14.8% (60.78 versus 51.80).

References

[1] Srikanth Jagabathula and Paat Rusmevichientong. A nonparametric joint assortment and price choice model. *SSRN*, 2013.

[2] Srikanth Jagabathula and Gustavo Vulcano. A model to estimate individual preferences using panel data. *Submitted to Management Science*, 2015.

[3] Jun Li and Serguei Netessine. Who are my competitors? let the customer decide. *Working Paper*, 2012.