Building Optimized and Hyperlocal Product Assortments: A Nonparametric Choice Approach*

Vivek F. Farias[†], Srikanth Jagabathula[‡], Devavrat Shah[§]

Abstract

We consider the problem of "hyper-localizing" product assortments at a fashion retailer—that is, customizing the offerings to the particular preferences of customers visiting the store, so that customers can easily find the products that fit their tastes and purchase more. To make this decision, the firm must accurately predict the demand for each style at *each* store—a challenging task because of large variety and the small number of purchases per customer. To address this challenge, we propose: (a) a nonparametric choice modeling technique that uses purchase transactions tagged by customer IDs to build distributions over preference lists of products that are *personalized* to each customer and (b) an optimization framework that uses the predictions from our choice models to optimally allocate merchandise to different stores subject to inventory and dollar budget constraints. We implemented our methods at a large US fashion retailer with about \$3B in annual revenue and approximately 300 stores. In a controlled experiment, our methods resulted in additional 7% revenue growth (approx. \$200M profit impact) over the current method. We present the implementation details and the specific challenges (both technical and managerial) posed for assortment planning by fashion retail and the ways we addressed them.

*This document describes a methodological framework at a conceptual level for the problem of assortment optimization. Parts of the described technology have been patented (Shah et al., 2015, 2016). The described impact of the technology is from an implementation (based on proprietary technologies) by our industry partner Celect, Inc. (`http://celect.com`).

[†]Sloan School of Management and Operations Research Center, MIT, email:vivekf@mit.edu

[‡]Leonard N. Stern School of Business, NYU, email:sjagabat@stern.nyu.edu

[§]Department of Electrical Engineering and Computer Science, MIT, email:devavrat@mit.edu

1. Introduction

1.1. Business context

Apparel retailing has truly become a multi-channel experience. The advent of technology has significantly changed the retailing landscape with consumers increasingly shopping through a retailer's online and mobile channels. Despite these growing sales from the newer channels (both online and mobile), 94% of all retail sales have come from the traditional brick-and-mortar channel. Recognizing the significance of the physical channel, "online first" ecommerce start ups such as Birchbox, Warby Parker, Nasty Gal, etc. are establishing physical locations. These trends certainly point to a multi-channel future in which a single retailer interacts with the consumer through multiple channels.

The result of the multi-channel shopping behavior is that customer's expectations have been altered: they now expect a "Total Retail" experience, with the same level of customization and personalization in the physical channel as in the online channel. As a result, it has become increasingly important for a retailer to solve the omni-channel puzzle: offer a consistent shopping experience to the customer.

1.2. The assortment decision and it's significance

A key piece of the omni-channel puzzle is building an optimized and hyperlocal product assortment in each brick-and-mortar store. A hyperlocal product assortment is the product offering that is customized to the particular preferences of the customers who visit each physical store, so that customers immediately find products that fit their tastes, just like in an online store. Determining the "best" hyperlocal product assortment involves solving the following decision problem.

For each store location, and given a dollar budget, the firm must decide the amount of inventory dollars to allocate to each product. The "products" may be of any granularity: categories (such as jackets/blazers, dress shirts, knit tops, handbags, etc.), brands (Calvin

Klein, DKNY, Vince Camuto, IZOD, etc.), or even individual SKUs (with specific color and size combinations). To make the decision, the firm has access to store-level inventory and purchase transactions across all the stores from previous years. The transactions are tagged with individual customer ids. Beyond that, the firm may not have any individual-level demographic data (such as age, gender, education, etc.). The data are similar to the panel data that are commonly used in the marketing literature. However, because the data are organically collected, they are noisy, sparse (with very few transactions per customer), and missing key entries (such as customer demographics).

The above decision problem is commonly referred to as "the assortment problem." It is a complex problem to solve, requiring accurate demand predictions for the next season. For instance, making the decision involves answering complex questions, such as: (a) Michael Kors (MK) handbags were sold out at the Minneapolis store this season. What is the right allocation (or true demand) for the next season? (b) MK handbags were never sold at the Lancaster, PA, store. Should we introduce them into the store this season? If so, how well will they sell? (c) A new brand is under consideration for being introduced into a set of stores. What will be the impact (on overall sales and the sales of the existing brands) of the introduction?

The assortment problem is also an important problem with far-reaching implications for both the (operational) cost and demand of a firm's profit equation. Misallocation of inventory dollars results in excess leftover inventory for the unpopular items, increasing the inventory costs, and stock outs of popular items, resulting in lost demand and unsatisfied customers.

Recent trends in retail have only further increased the significance of the assortment decision, both from the operational and customer experience perspective. From an operational perspective, a consequence of the rise of the online and mobile commerce is the changing nature and role of the offline brick-and-mortar stores. On the one hand, offline stores are becoming fewer, smaller, and more impactful, effectively reversing the trend over the last three decades (during which time offline stores doubled in size). Further, newer formats such as pop up stores and mobile trucks are becoming increasingly common. The result is that the assortment decision must be made with an increasing number of (space) constraints, making a miss more costly: if only 10 items can be carried, then a miss on even one item can result in significant inventory costs and lost demand. On the other hand, from a customer's perspective, offline stores are becoming important for providing a complete shopping experience (Celect, LLC, 2016) to customers with their roles changing from "sales centric" to " customer centric." The in-store experience provided by the firm is becoming a key competitive weapon for a firm in addition to its offered selection and prices, and a key piece of the in-store experience is providing hyper local assortments, personalized/customized to the particular tastes of the customers at that store. In other words, the assortment carried can have a huge impact on the experience provided to the customer and, hence, their satisfaction and (long-term) spend.

1.3. Current practice

Apparel retailers broadly follow the following decision-making process. A year is typically divided into two major seasons: Fall-Winter (FW) and Spring-Summer (SS). Each season, the firm must decide the products and the corresponding quantities it will carry at each of its physical stores. The exact decision process used varies from firm to firm, but broadly consists of the following three steps: (a) first, the firm selects the brands, and the styles within the brands, it will carry the next season; (b) then, it decides the corresponding order quantities for each brand-style combination; and (c) finally, it allocates the inventories of the brand-style combinations to each of its stores. These decisions are made subject to the firm's seasonal "Merchandise Financial Plan," which is the dollar budget the firm has assigned for each category for that season. Given this budget and the sales trends identified from the previous season's transactions, the firm uses judgment and expertise/intuition to identify the brands and styles that will sell well next season and allocates them to the different stores. We illustrate this process with a simple hypothetical example for women's handbags. In the first step, the firm decides to carry Michael Kors (MK) and Chaps brands within women's

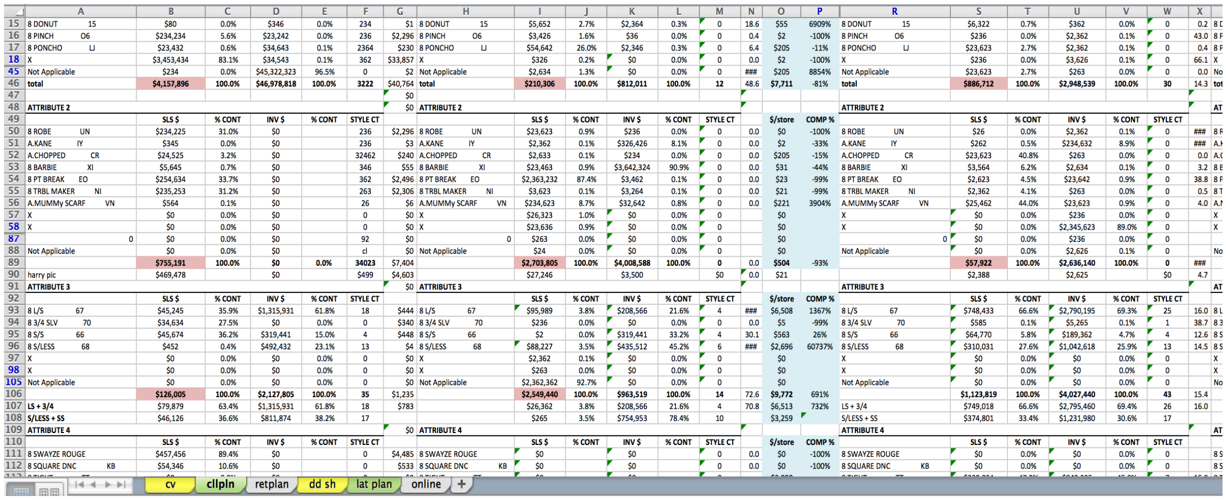

Figure 1: A typical spreadsheet decision support tool used by a retailer to make the assortment decision. Image credit: Celect, Inc (`http://celect.com`)

handbags for the next season. Then, based on its dollar budgets and sales trends, the firm decides to purchase \$60M worth of MK handbags and \$40M worth of Chaps handbags for the season. The firm has three store locations: Downtown Chicago, IL (CH); Lancaster, PA (LN); and Minneapolis, MN (MP). So, based on the characteristics of the customers at each of its stores, the firm decides to allocate \$40M and \$20M of its MK handbags to CH and MP stores, respectively, and \$20M of its Chaps handbags to CH and \$10 each to LN and MP stores, respectively.

The current practice for making this decision mainly relies on managerial expertise and intuition, coupled with trends identified from the sales transactions data. Figure 1 shows a typical spreadsheet decision support tool used by a retailer to make the assortment decision. Using the data collected in the spreadsheet, the firm analyses the past sales trends to forecast demand for the next season. For instance, the firm may forecast a weekly demand of 300 platform sandals this year because the demand was 200 last year and 100 the year before. These sales trends are then combined with macro fashion trends and expert judgment to make allocation decisions. For example, the firm may decide to increases its spend on "Athleisure wear"[1] because fashion trends fashion suggest that it will become big this year.

[1] Athleisure is a fashion trend in which clothes designed athletic activities is worn in other (casual) settings.

1.4. Limitations of current practice and challenges to improve

The biggest limitation of the existing manual process is the high risk of missing expensive bets. Particularly, in deciding the allocation of inventory dollars for the next season, the firm is making expensive bets on what will and will not sell well in each of its stores. A "miss" may either be an overstock or an under-stock of a product. Overstock occurs when the firm overestimates the demand for a product, resulting in high inventory and operational costs. It also ties up inventory dollars, leaving fewer dollars to invest in products that will indeed sell well. Under-stock occurs when the firm underestimates the demand for a product, resulting in lost sales and poor customer experience. Importantly, both misses carry immense opportunity costs, in the short-term as well as in the long-term. The opportunity cost comes from the lost sales because of not carrying products that would have been popular, which will result in lost revenue not only in the current season, but also in the future because the customer may defect to a competitor.

The risk of missing a bet is highest when a particular brand was not carried at the store last season. For instance, consider a retailer who carries Michael Kors (MK) handbags at its downtown Chicago stores but does not carry them at the Lancaster, PA store. Is it profitable for the firm to introduce MK handbags at the Lancaster store this season? MK is considered an "city" brand, so managerial expertise suggests that it will not be profitable to introduce MK handbags into the Lancaster store, which is close to Amish country. Because MK was never offered at Lancaster, there is no readily available data to suggest otherwise. Because the manager's compensation depends on meeting sales targets, they typically never introduce MK at the Lancaster store. However, analysis by our industry partner and subsequent implementation, revealed that this was a missed bet. As far as purchasing MK handbags goes, there are certain customers at the Lancaster store that are similar to those at the downtown Chicago store. Therefore, not introducing MK at Scranton is a missed opportunity. Of course, this is one brand and one store. The magnitude of potential opportunities across hundreds of stores and brands is immense.

To mitigate the loss of potential profit due to the existing decision process, managers must take advantage of an automated decision support system. However, there are two main challenges: (a) process-related and (b) fundamental technical/methodological. Because managers are used to the current process, they must be convinced that the recommendations produced by an automated tool should be followed, even though some of the recommendations ran counter to their intuitions. To accomplish this, our industry partner built a simple graphical interface (as described in detail below) to allow the managers to visually drill down into the recommendations. Theses recommendations fell into three classes: (a) similar to the manager's own recommendations, (b) different from the manager's recommendations but in line with the manger's intuition, and (c) different from the manager's recommendations and counter to the manager's intuition. As expected, the third type of recommendations are the hardest to provide. Mangers' confidence in the decision support tool was gained through the first two types of recommendations, which allowed them to implement the third type of recommendations.

But beyond the process-related challenge, there is a more fundamental methodological challenge, addressing which goes beyond managerial expertise. Specifically, using the example above, how does one figure out that MK handbags will sell well in the Lancaster, PA, store, when given the information that they have sold well in the downtown Chicago store? What do the data tell us? Because MK handbags were never carried at the Lancaster, PA, store, for predicting whether they will sell well, we must necessarily correlate data from across all the stores. The common approach for dealing with this challenge in the industry is to cluster stores by customer demographics and the size of the market, with the understanding that stores in the same cluster receive similar assortments. While this approach is certainly data-driven, the resulting clustering may be too coarse because it ignores subtle variations in customer purchase patterns: some of the customers at the Lancaster, PA, store may be similar to the customers at the downtown Chicago store while the others may be similar to those at the Orlando Park, IL, store. As a result, clustering is making inefficient use of the existing information and leaving money on the table. A more systematic and fine-grained way for making these correlations is to model the data generation process at each store using a choice model and leverage existing methodologies to make predictions. However, as detailed below, the predictive accuracies of existing techniques are insufficient for these purposes, necessitating a new development of choice models with a focus on predictive accuracy. We describe these techniques next.

2. Related work

The assortment optimization problem has received much attention in the operations management and the revenue management literature. The canonical problem involves finding the revenue/profit maximizing subset of products from a large universe. The optimal assortment must trade off losing revenues from not including low-revenue products with gaining revenues from inducing the switch to high-revenue products. The assortment problem becomes interesting because of the *product substitution* behavior, where customers substitute to an available product (say, a dark blue shirt) when her most preferred product (say, a black one) is not offered. Traditional models in operations and revenue management ignored substitution and assumed product demand to be independent of the offer set. Under this assumption, the assortment decision was driven by considerations such as inventory costs, space, seasonality, joint replenishment, etc. Since then, the significance of substitution effects has been recognized and the academic literature on assortment planning has been driven by considerations of product substitution. Initial work focused on the multinomial logit model and various parametric substitution models in which the pairwise substitution rates between products was parametrized. See Kök et al. (2008) for an excellent review of the existing substitution-based models. More recently, the focus has shifted from substitution models to using Random Utility Maximization (RUM)-based choice models to capture product substitution. Because choice models are rooted in utility theory, they provide a systematic way to capture product substitution. This work has considered popular parametric choice models (such as the multinomial logit (MNL), nested logit (NL), latent-class multinomial logit (LC-MNL) models) and exploited the resulting structure of the revenue function to obtain tractable (approximation) algorithms. See Jagabathula (2014) for an overview of the existing technical results.

Most of this initial work has been motivated by applications to the airline/hotel revenue management and retailers with reasonably long product cycles and stable demand. Examples include grocery (supermarket) retailers, electronics retailers (such as Best Buy), apparel retailer carrying fashion basics, etc. Because the demand and assortment offerings are stable over time, parametric models fit to transactions data can provide reasonably accurate demand predictions. Therefore, the focus of the research has been on solving the decision problem while accounting for the impact of inventory decisions and supply chain structures.

However, the above characteristics do not apply to a fashion apparel retailer in which products have very short life cycles and the demand for the next season is highly uncertain. For these settings, making accurate demand predictions (and not necessarily the optimization problem) becomes the critical and challenging step. In addition, there are few reported instances of academic approaches on assortment planning being successfully implemented in practice. In fact, Kök et al. (2008) observe that there is a gap between the approaches studied in the academic literature and those that are implemented in industry (such as electronics retailer Best Buy, book and music retailer Borders, Indian jeweler Tanishq, and Dutch supermarket chain Albert Heinjn). Exceptions include Caro and Gallien (2010), who describe a successful implementation of academic work for inventory management at Zara, and Ferreira et al. (2015), who describe a successful collaboration between academia and Rue La La (a flash sales site) for carrying out price optimization. None of these, however, focus on the assortment problem or incorporate choice models.

While more recent work in operations has focused on incorporating choice models into assortment planning, as mentioned, the focus has been primarily on solving the decision problem as opposed to estimating the model parameters. On the other hand, much work has been done in the marketing literature on estimating the choice model parameters from panel data. This work extends all the way back to the seminal work of Guadagni and Little (1983), in which they fit a MNL model to household panel data on the purchases of regular ground coffee, and which has paved the way for choice modeling in marketing using scanner panel data; see Chandukala et al. (2008), Wierenga (2008) for a detailed overview of choice modeling using panel data in the area of marketing. The existing estimation techniques are either Bayesian or likelihood-based, and much of the work has focused on how various panel covariates (such as customer demographics) influence individual choice making.

Because of the rigor provided by choice models, our demand model is also based on choice models. Our modeling framework and estimation approach, however, deviate from much of the existing work in marketing in two ways. First, the popular way to account for heterogeneity in customer tastes is to use assume a hierarchical model in which it is assumed that the parameters of each individual are sampled from a distribution that is a function of the customer demographic variables. This allows for capturing heterogeneity while allowing for efficient parameter estimation by pooling together data of all the customers. However, in our application, although purchase transactions are tagged by the customer id, customer demographic data are not available. Second, most of the existing approaches in marketing are parametric, i.e., the model size does not grow with the amount of data. These methods are designed to assess the impact of various features (both individual and product) on consumer choice. Instead, our focus is on designing methods that optimize predictive power. Therefore, we design a nonparametric method in which the mode size grows with the number of customers and products in the data set. Such nonparametric methods have been found to have better predictive power than the traditional parametric methods (see Farias et al. (2013)).

3. Problem formulation and technical details of the proposed solution

We now describe the precise decision problem. We focus on a firm that will be selling J products, indexed $1, 2, \ldots, J$, through S stores in the next season. The firm will sell product j at price r_j and has Q_j units of the product in inventory at the beginning of the season. The firm must allocate its merchandise to the S stores with the objective of maximizing its revenues, subject to store-level budget constraints (described below). In particular, for each j and s, the firm must decide the quantity q_{js} of product j that will be allocated to store s.

To make this decision, the firm must predict the demand for the next season for each product j, at each of the stores it is offered. Naturally, if the firm expects to sell more of product j at store s when compared to store t, then it must allocate a larger quantity of product j to store s when compared to store t. However, the demand next season is not known. To predict the demand, we have access to the individual-level sales transactions data from the previous season.

Therefore, to solve the decision problem, we design two modules: (a) demand predictor and (b) optimizer. The demand predictor uses historical transactions data to provide accurate demand predictions for each product, at each of the stores it is offered. The optimizer operates on the predictions from the demand predictor and determines the optimal allocation of order quantities across the stores, while respecting the firm's business constraints (budget constraints, brand constraints, capacity constraints, etc.).

The critical challenge is in obtaining reliable demand predictions from the given data. Therefore, we first describe the optimizer assuming access to demand predictions and then describe how we obtain the demand predictions.

Let $D_{js}(\boldsymbol{q}_s)$ denote the predicted demand for product j at store s as a function of the vector of allocated product quantities $\boldsymbol{q}_s = (q_{1s}, q_{2s}, \ldots, q_{ns})$. Due to substitution behavior of customers, we expect the demand for product j to be dependent on which other products are offered; therefore, $D_{js}(\cdot)$ is a function of the quantities of all the products at store s. Let c_{js} denote the cost for carrying a unit of inventory at store s. Given this, the decision problem of the firm can be formulated as the following optimization problem:

$$
\begin{aligned}
\max_{\boldsymbol{q}\geq\mathbf{0}} \quad & \sum_{s=1}^{S}\sum_{j=1}^{J} r_j \cdot q_{js} & \\
\text{subject to} \quad & \sum_{s=1}^{S} q_{js} \leq Q_j, \ \text{ for } j = 1, 2, \ldots, J & \text{[inventory]} \\
& q_{js} \leq D_{js}(\boldsymbol{q}_s), \ \text{ for } j = 1, 2, \ldots, J \text{ and } s = 1, 2, \ldots, S & \text{[demand]} \\
& \sum_{j=1}^{J} c_{js} \cdot q_{js} \leq B_s, \ \text{ for } s = 1, 2, \ldots, S & \text{[budget].}
\end{aligned}
$$

The objective function is the total revenue from all the stores for the next season. The *inventory* constraint ensures that the total allocation of j to all the stores is no more than the available inventory Q_j, the *demand* constraint restricts the allocation of j to store s to be less than the expected demand $D_{js}(\boldsymbol{q}_s)$, and the *budget* constraint captures the 'business' constraint imposed by the firm to ensure that the allocation to each store is within the allocated dollar budget. Additional business constraints such as dollar budgets for brands and other size and capacity constraints can be similarly introduced.

The 'demand' constraint is key in the above optimization problem. It is restricting the quantity of the product that is allocated to each store to be less than the *expected* demand for the product at that store. However, demand next season and, hence, the realized profit are uncertain. As a result, a more accurate formulation of the decision problem will maximize the *expected profit*, resulting in the stochastic program:

$$
\begin{aligned}
\max_{\boldsymbol{q}\geq\mathbf{0}} \quad & \sum_{s=1}^{S}\sum_{j=1}^{J} \mathbb{E}\left[r_j \min\left\{q_{js}, d_{js}(\boldsymbol{q}_s)\right\} - \tilde{c}_{js}(q_{js} - d_{js})^{+}\right] & \\
\text{subject to} \quad & \sum_{s=1}^{S} q_{js} \leq Q_j, \ \text{ for } j = 1, 2, \ldots, J & \text{[inventory]} \\
& \sum_{j=1}^{J} c_{js} \cdot q_{js} \leq B_s, \ \text{ for } s = 1, 2, \ldots, S & \text{[budget],}
\end{aligned}
$$

where the expectation in the objective function is with respect to the demand distribution, x^{+} denotes $\max\{x, 0\}$ for any number x, and $\tilde{c}_{js}$ the cost of the leftover inventory at the end of the season. The objective function is the expected total revenue *net* the cost of leftover inventory. Because stochastic programs are much harder to solve, we are instead solving the deterministic approximation in which the demand constraint is imposed only in expectation. Such deterministic approximations are commonly used in revenue management (RM) (Ciocan and Farias, 2012, Liu and Van Ryzin, 2008) to tractably solve complex stochastic programs. Further, note that we are approximating the aggregate demand over thousands/millions of customers over the next season (3–6 months) by its expectation. As a result, we expect the observed demand to tightly concentrate around its expectation, making our approximation very accurate, as long as we can predict the expected demand accurately.

While the objective function and the remaining constraints are linear functions of the decision variables and, hence, are tractable, the demand $D_{js}(\cdot)$ is a non-linear function of the decision variables because substitution causes the demand to be a function of the entire offer set. Several existing techniques can be used to approximately solve the optimization problem. Our industry partner employs a proprietary formulation of the optimization problem we have outlined, and uses proprietary algorithms to solve these problems. However, other techniques in the literature may also be used. One such technique is based on the approach described in Jagabathula (2011). Broadly, the difficulty in solving the optimization problem arises because $D_{js}(\boldsymbol{q}_s)$ is a set function, depending on the offer set $M(\boldsymbol{q}_s)\colon \{j : q_{js} > 0\}$ at the store. In fact, as shown below $D_{js}(\boldsymbol{q}_s)$ is *only* a function of the offer set $M(\boldsymbol{q}_s)$. Now, suppose the offer sets $M(\boldsymbol{q}_s)$ at each of the stores are known. Then, $D_{js}(\boldsymbol{q}_s)$ is determined and the above optimization problem becomes a linear program and can be solved efficiently to obtain the optimal revenue. However, the offer sets at the stores are not known. So, we use the local search heuristic described in Jagabathula (2011) to search over the space of offer sets to obtain an approximate solution.

3.1. Personalized demand predictions

We now discuss how we use historical sales transactions data to predict the next season's expected demand $D_{js}(\boldsymbol{q}_s)$, when given the order quantities $\boldsymbol{q}_s$. We will first describe the data we have access to, the model and the estimation technique, and then detail how our methods differ from the existing techniques in marketing, illustrating the novelty in and the necessity for our methods.

Data. The firm provides point-of-sales (POS) transactions and the weekly inventory data. The POS data log information on all the sales transactions that occurred at all the stores. Each transaction contains information on the product that was purchased, a hashed identifier for the customer who made the purchase, the time of purchase, the store id where the purchase occurred. The weekly inventory data contain information on the set of products that were stocked at each store, every week.

To feed the data into our demand predictor, we process them as follows. We first extract the individual customers indexed $i = 1, 2, \ldots, I$ who made purchases in the data set. For each individual i, we combine the corresponding sales transactions with the inventory data to obtain observations encoded as the multiset $\mathcal{O}_i = \{(j_{i,1}, M_{i,1}), \ldots, (j_{i,T_i}, M_{i,T_i})\}$, where the observation $(j_{i,t}, M_{i,t})$ denotes that customer i purchased product $j_{i,t}$ in the purchase instance t, when the offer set in the store was $M_{i,t}$. For purchase transaction t, the purchased item is obtained from the POS data and the offer set is obtained from the inventory data as the set of products that were stocked at the store where and in the week when the purchase was made. Because $\mathcal{O}_i$ is a multiset, multiple purchases of the same item from the same offer set appear multiple times, so there is no loss of information.

We make the following remarks. First, note that the data we have provides us information on not only what the customer purchased, but also what the customer *did not* purchase. Broadly, in each purchase instance, the customer is making a choice of which product to purchase from the set of products that are on offer at the store. Therefore, the purchase of product $j_{i,t}$ when the offer set is $M_{i,t}$ implies not only that the customer likes product $j_{i,t}$ but also that the customer prefers $j_{i,t}$ over all the other products that were offered in $M_{i,t}$. Our processing captures this by encoding the transaction as the tuple $(j_{i,t}, M_{i,t})$, which implies that product $j_{i,t}$ was chosen from $M_{i,t}$. Note that, strictly speaking, we want $M_{i,t}$ to be the set of the products the customer *considered*, as opposed to the set that was offered. However, the inventory data provide information only on what was stocked (or offered). Consequently, we approximate the consideration set by the offer set, which is a good approximation, especially because most of the purchases are considered purchases from a familiar category (such as handbags).

Second, note that our processing does not directly encode the store id for each purchase instance. However, most customers (modulo a few exceptions) purchase from only a single store. Therefore, instead of keeping track of the store id for each purchase instance, we encode the store information by keeping track of the set of customers who shop at each store.

Finally, for most customers in the data set, we don't have access to customer demographic information (such as age, gender, household income, etc.) because the data were never collected (because the customer either refused to provide the information or, more likely, the existing systems are inadequate to capture this information).

Model. We now describe how we make demand predictions. For each store s, our objective is to predict demand when given the quantity vector $\boldsymbol{q}_s$. To explain our approach, let's fix a store s. Let M denote the set of products that are being carried at the store this season i.e., $M = \{j : q_{js} > 0\}$. Our objective is to predict the number of customers who will purchase each product $j \in M$ during the next season. A natural starting point to make the prediction is the sales data from the previous season at the store. Suppose M' was the offer set at the store previous season and n_{js} was the sales count for product $j \in M'$. If $M = M'$, then applying a seasonal growth factor to the previous season's sales can provide reasonably accurate predictions. Such an approach is reasonable if the firm is selling fashion basics and when the assortments at the stores do not change substantially from season to season. However, for a fashion apparel retailer, very often M' and M are significantly different.

When the offer sets are different, the demand prediction problem becomes non-trivial. If a new product is introduced into the store, then it will cannibalize the demand of existing products because of which predicting the demand of the existing products is also not straightforward. To deal with this challenge, we build a nonparametric choice model to make individual-level predictions of the likelihood of the purchase of each product, given the offer set M.

Our model operates at the individual customer-level. Specifically, consider a customer i who makes purchases from store s. Because very few customers make purchases from multiple stores, we assume that each customer can be associated with a single store. Let $g_i(j, M)$ denote the probability that customer i will purchase product j next season, when the offer set is M. Assuming we have access to $g_i(\cdot, \cdot)$, we make the prediction:

$$D_{js}(\boldsymbol{q}_s) = \alpha_s \cdot \left[\sum_{i\colon i \text{ purchases at store } s} g_i(j, M) \right],$$

where α_s is a seasonal growth factor applied to scale the total number of customers from the previous season to the current season.

In order to make personalized demand predictions, we use a nonparametric choice modeling approach. Broadly, our approach consists of two steps: (a) fitting a population-level nonparametric choice model to pooled purchase transactions from all the customers; and (b) using the observations from each customer to determine a 'personalized' choice model. Next, we elaborate the two steps.

Fitting a population level choice model. Instead of making any parametric assumptions, we model the preferences of the population of customers through a distribution λ defined over the space of $n!$ preference/rank lists of n items, so that $\lambda(\sigma)$ is the probability assigned to ranked list σ. Each ranked list σ specifies a preference ordering over the products with $\sigma(j)$ denoting the preference rank of product j and lower ranked being products preferred over higher ranked products. When given an offer set M of products, the customers make choices by sampling a preference list σ according to distribution λ and then choosing from the offered items, the most preferred item according to σ i.e., product j such that $\sigma(j) < \sigma(j')$ for all $j' \in M$ and $j' \neq j$. We implicitly assume that the no purchase option is present in every ranked list, so the customer may leave without making a purchase.

The above nonparametric model is also referred to as the rank-based choice model. It subsumes the rich class of Random Utility Maximization (RUM) models, of which popular choice models such as multinomail logit (MNL), nested logit (NL), latent-class MNL (LC-MNL), and mixed logit models are special cases. This is because an RUM model specifies a distribution over product utility vectors and supposes that each customer samples a utility vector and chooses the offered product that has the highest utility. Because each utility vector induces a preference ordering and only the preference order matters in so far the choice is concerned, the rank-based choice model subsumes the RUM family of models.

In order to learn the distribution λ, we use the pooled purchase transactions from all the customers. Let $M_1, M_2, \ldots, M_T$ denote the collection of the unique offer sets present in the data, i.e., $\bigcup_{i=1}^{I} \bigcup_{t=1}^{T_i} \{M_{i,t}\}$. Let $y_{j,t}$ denote the fraction of times product j was purchased when the offer set was M_t i.e.,

$$y_{j,t} = \frac{\sum_{i=1}^{I} \sum_{\tau=1}^{T_i} \mathbb{1}[j = j_{i,\tau}, M_{i,\tau} = M_t]}{\sum_{i=1}^{I} \sum_{\tau=1}^{T_i} \mathbb{1}[M_{i,\tau} = M_t]},$$

where the numerator counts the number of times product j was purchased when M_t was offered and the denominator counts the number of times M_t was offered.

The observed choices may be interpreted as lower order marginals under the distribution λ. Specifically, it follows from our model description that the choice probability that product j will be purchased from offer set M_t is equal to $\sum_{\sigma} \lambda(\sigma) \cdot \mathbb{1}[\sigma, j, M_t]$, where we abuse notation to let $\mathbb{1}[\sigma, j, M_t]$ denote the indicator variable that takes the value 1 when j is the most preferred product under σ among the products in M_t, i.e, $\mathbb{1}[\sigma, j, M_t]$ is defined to be $\mathbb{1}[\sigma(j) < \sigma(j') \ \forall \ j' \in M_t, j' = j]$. Given this (and assuming that there are no model misfit or finite sample errors), we can relate the population level model λ to the observations through the system of linear equations:

$$y_{j,t} = \sum_{\sigma} \lambda(\sigma) \cdot \mathbb{1}[\sigma, j, M_t] \;\forall\; j \in M_t, t = 1, 2, \ldots, T \iff \boldsymbol{y} = A \cdot \boldsymbol{\lambda},$$

where $\boldsymbol{y} = (y_{j,t})_{j \in M_t, t=1,\ldots,T}$ denotes the vector of observed choice fractions and $\boldsymbol{\lambda}$ denotes the $n! \times 1$ vector of preference distribution. We index each entry of $\boldsymbol{y}$ by j, t and each entry of $\boldsymbol{\lambda}$ by the corresponding ranking σ. A is the $L \times n!$ matrix consisting of $0 - 1$ entries, where $L = \sum_{t=1}^{T} |M_t|$, with the entry corresponding to the row j, t and column σ equal to $\mathbb{1}[\sigma, j, M_t]$. Thus, the relation between the observed choices and the underlying model can be represented in a compact form as $\boldsymbol{y} = A \cdot \boldsymbol{\lambda}$.

With the above formulation, the task of model learning reduces to inferring a joint distribution, λ, from the corresponding lower-order marginals $\boldsymbol{y}$. This formulation presents two key issues: (a) the system of linear equations $\boldsymbol{y} = A \cdot \boldsymbol{\lambda}$ is an under-determined system with multiple solutions, and (b) the vector $\boldsymbol{\lambda}$ entering into the equations has a dimension of $n!$, raising concerns of computational tractability.

The first issue is arising because our model specification is very general, resulting in 'set identification' of the models that are consistent with the data. Because the observed data are not sufficient to pin point a model, we need to introduce parsimony into our formulation. The popular way to introduce parsimony is to parametrize the distribution λ by imposing a structure that allows us to express the choice probabilities as a compact function of a small set of parameters. For example, the MNL model is described by n parameters and a LC-MNL model with K latent classes is described by Kn parameters. The issue with this approach is that the 'right' model structure depends on the amount of data; for example, the MNL model may be found to be a good fit to the data, but as more choice observations are added, it may be deemed to be under-fitting. In other words, because the number of parameters is fixed, the model 'complexity' does not scale with the data, resulting in either under-fit and over-fit issues. Because of this, these models tend to under-perform nonparametric models in terms of predictive accuracy (Farias et al., 2013).

To deal with the above issue, we use the sparsity, or the support size, of distribution λ as a measure of parsimony, and pick the sparsest distribution λ from among the set of models consistent with the data. Sparsity is a natural candidate to measure the complexity of a distribution and has found recent success in the area of compressed sensing, which spans the areas of signal processing, coding theory, and streaming algorithms (Candès et al., 2006, Donoho, 2006). It also naturally arises from our formulation $\boldsymbol{y} = A \cdot \boldsymbol{\lambda}$ because it follows from Carathéodory's theorem that a distribution of sparsity no more than $L+1$ is needed to describe the data vector $\boldsymbol{y}$. Using sparsity as a criterion also makes our model nonparametric. This is because, under appropriate technical conditions, Farias et al. (2013, Theorem 1) shows that for "almost all" data vectors $\boldsymbol{y}$, the sparsest distribution $\boldsymbol{\lambda}$ that is consistent with $\boldsymbol{y}$ has a sparsity of L or $L+1$. In other words, with sparsity as the criterion, the 'complexity' of the resulting model will scale with the dimension L of the data. We note that an alternate criterion of 'worst-case' revenues was suggested in Farias et al. (2013) to pick a distribution. For a given offer set M, the criterion picks the distribution that yields the worst-case revenue from among all the distributions consistent with the data. However as shown by Farias et al. (2013, Theorem 1), both criteria are similar in the sense that they both pick distributions that are sparse.

The second issue with our formulation is the computational challenge arising from the large dimension of $\boldsymbol{\lambda}$. We deal with this issue by developing optimization-based techniques. In particular, because the set of distributions are described by a system of linear equations, they can be equivalently represented as a bounded polyhedron with each permutation corresponding to an extreme point. To see this, let A_σ denote the column of the A matrix corresponding to permutation σ and define the polytope $\mathcal{P} = \text{conv}\,\{A_\sigma\colon \text{ for all permutations } \sigma\}$, where $\text{conv}(\mathcal{S})$ is the convex hull of the collection of points in $\mathcal{S}$. Now, because $\boldsymbol{y} = A \cdot \boldsymbol{\lambda}$, the vector $\boldsymbol{y}$ must belong to the polytope $\mathcal{P}$, which implies that the set of models consist with $\boldsymbol{y}$ correspond to the set of all convex decompositions of $\boldsymbol{y}$ in terms of the extreme points of $\mathcal{P}$. Based on this correspondence, Farias et al. (2010) reduces the problem of finding a sparse distribution to solving optimization problems on the polytope $\mathcal{P}$. Therefore, the computational complexity of our problem depends on our ability to obtain efficient descriptions of the polytope $\mathcal{P}$. For that, we exploit the structure of the data vector $\boldsymbol{y}$. The papers Farias et al. (2010, 2013) present various approaches to obtain either exact or approximate descriptions of the polytope $\mathcal{P}$ to solve our problem in a tractable fashion.

Note that we assumed that there was no model mis-specification or finite sample errors in equation the empirical fractions $y_{j,t}$ to their corresponding expected values. In the presence of these errors, the system of linear equations $\boldsymbol{y} = A \cdot \boldsymbol{\lambda}$ may be infeasible. In these cases, we use the following method-of-moments estimator:

$$\min_{\boldsymbol{\lambda} \geq 0, \|\boldsymbol{\lambda}\|_1 = 1} \|\boldsymbol{y} - A \cdot \boldsymbol{\lambda}\| + c \cdot \|\boldsymbol{\lambda}\|_0,$$

where $\|\cdot\|$ may be either the ℓ_1 or ℓ_2 norm, $\|\cdot\|_0$ is the ℓ_0 pseudo-norm that counts the number of non-zero elements in the vector, and $c > 0$ is the scalar penalty term that balances the trade-off between fit to the data and the complexity of the model. We solve this optimization problem by fixing a value of $\varepsilon > 0$ and solving the following constrained optimization problem:

$$\min_{\boldsymbol{\lambda} \geq 0, \|\boldsymbol{\lambda}\|_1 = 1} \|\boldsymbol{\lambda}\|_0 \text{ subject to } \|\boldsymbol{y} - A \cdot \boldsymbol{\lambda}\| \leq \varepsilon.$$

If we choose the ℓ_1 norm for $\|\boldsymbol{y} - A \cdot \boldsymbol{\lambda}\|$, then the constraint space becomes a polyhedron. The techniques described above immediately extend to this modified polyhedron, which we use to obtain a sparse distribution. Because the value the best value of ε is not known, we perform a binary search over the values of ε. Finally, we tune the value of c through cross-validation.

The output of this step is a distribution λ that describes the preferences of the population of the customers. Farias et al. (2013) details a case study in which they apply this technique to data from a major US automaker and show that it outperforms standard parametric models by about 20% (on standard relative error metrics) in accurately predicting conversion rates at various dealerships across the country.

Personalizing the choice model. Given the the population-level model λ, we obtain personalized demand predictions by computing a choice model that is personalized to each individual i. Specifically, suppose the λ has support size K with $\sigma_1, \sigma_2, \ldots, \sigma_K$ comprising the support. To derive the personalized choice model for each individual i, we suppose that the population consists of K types of customers with type k described by preference list σ_k and comprising a proportion $\lambda(\sigma_k)$ of the population. Now given the observation set $\mathcal{O}_i$ for individual i, proceed as follows. For each preference list k, we compute a similarity weight $w_{ik} > 0$ that measures how similar the individual's observations are to preference list k. Then, we compute the choice model λ^i personalized to individual i as

$$\lambda^i(\sigma_k) = \frac{w_{ik} \cdot \lambda(\sigma_k)}{\sum_{k'=1}^{K} w_{i,k'} \cdot \lambda(\sigma_{k'})} \text{ for } k = 1, 2, \ldots, K.$$

A consequence of the above definition is that λ and λ^i have the same support. Now, given the personalized choice model λ^i, we make the choice probability prediction:

$$g_i(j, M) = \sum_{k=1}^{K} \lambda^i(\sigma_k) \cdot \mathbb{1}[\sigma_k, j, M].$$

Several similarity metrics in the context of preference lists have been studied in the literature. The most popular similarity metric is the one that is based on the Kendall-Tau distance function. This distance function measures the distance between two ranked lists σ and π as the total number of pairwise disagreements: $d(\sigma, \pi) = \sum_{j<j'} \mathbb{1}[(\sigma(j) - \sigma(j') \cdot (\pi(j) - \pi(j')) < 0]$. Correspondingly, the similarity between the two ranked lists can be computed as $e^{-\theta \cdot d(\sigma, \pi)}$ for some 'concentration' parameter $\theta > 0$. This similarity metric can be extended to choice observation (j, M) by extending the distance function. Specifically, define the distance between the observation (j, M) and the ranked list σ as

$$d((j, M), \sigma) = \sum_{j' \in M, j' \neq j} \mathbb{1}[\sigma(j') < \sigma(j)] - \sum_{j' \in M, j' \neq j} \mathbb{1}[\sigma(j) < \sigma(j')],$$

i.e., the distance measures the number of disagreements and subtracts the number of agreements. This definition is a generalization of the standard Kendall-Tau distance function, which can be noted by verifying that the it yields the transformed Kendall-Tau distance $2 \cdot d(\sigma, \pi) - \binom{n}{2}$ when applied to total orderings. The similarity w_{ik} can now be computed as

$$w_{ik} = \exp\left(-\theta \cdot \sum_{t=1}^{T_i} d((j_{i_t}, M_{i,t}), \sigma)\right).$$

Although, not discussed here, the above method for computing personalized choice probabilities can be seen as approximate Bayesian inference. For that λ is viewed as the prior distribution and w_{ik} defined above becomes an approximate posterior probability that customer i is of type k when type k is described by a Mallows (Mallows, 1957, Marden, 1995) model with mode σ_k and concentration parameter θ.

Now, given the personalized predictions, we compute the demand:

$$D_{js}(\boldsymbol{q}_s) = \alpha_s \cdot \left[\sum_{i:\, i \text{ purchases at store } s} g_i(j, M)\right], \text{ where } M = \{j : q_{s,j} > 0\},$$

where α_s is the seasonal growth factor at store s. We tuned the regularization parameters λ_c by way of cross-validation. The seasonal growth parameters were estimated by applying standard time-series techniques on the aggregated sales data.

3.2. Transferability

Our method by is design nonparametric and data-driven, so it can be applied with little expert input. The method by itself does not use any context that is specific to a particular retailer. In addition, the optimization framework is designed to accommodate various business constraints, so that it is portable across a variety of retailers. Further, this technology has been patented (Shah et al., 2015, 2016). In addition, our technology received the following press coverage: Matyszczyk (2011), Taylor (2011).

4. Implementation challenges

A variant of our methodology was implemented by Celect, Inc. (details below) and applied as part of its assortment product offering at one of its clients. In implementing our solution, there were two key challenges: (a) scaling the methods to large data sets and (b) getting the managers to adopt our solution.

Data challenge. The implementation of our method faced the "big data" challenge. A typical large-size apparel retailer has a few million Stock Keeping Units (SKUs), few tens of millions of customers, and hundreds of stores. Data are available in the form of purchase transactions (both online and offline), customer browsing patterns, website clicks, ratings, etc. This results in a large volume of data that is high-dimensional. The raw data from a single retailer is of the order of tens to a few hundreds of gigabytes.

Further, the above choice observations result in a high-dimensional decision problem over a large volume of data. To put the dimensions of the problem in perspective, the famous Netflix Prize challenge (Netflix.com, 2009) dealt with making personalized predictions from 100M ratings provided by about 480K users for about 17K movies. In contrast, our demand prediction engine deals with making personalized demand predictions for millions of users over a few million of products. In addition, instead of the cardinal observations in form of ratings, our observations are ordinal in the form of choices. Consequently, existing fast matrix factorization techniques don't directly apply to our setting.

Our industry partner uses proprietary solutions to deal with the "big data" challenge. However, one may use existing techniques to make the estimation and optimization scalable. These technique rely on converting the computational problem into a graph-based problem. For instance, recent work by Negahban et al. (2012a,b) has shown that estimating the parameters of an MNL model can be converted into the problem of computing the stationary distribution of an appropriate Markov chain. For data sets observed in practice, the resulting graphs tend to be sparse. By exploiting this structure, one can use standard graph-partitioning techniques to make the computation 'map-reducable' and implementable on a cloud.

Adoption challenge. Because the automated decision support system implemented by our industry partner was different from the current practice, there was a challenge of getting the existing decision-makers to adopt the new system. Taking a cue from existing work (Leeflang et al., 2013, Little, 2004), the challenge was addressed by creating a user-friendly graphical interface that is simple to use and adapts in real-time based on user input. Figure 2 shows a screen shot of the interface, whose simplicity is in contrast with the current spreadsheet-based tool in Figure 1. Each box in the screen shot corresponds to a brand in a specific store. The color of the boxes represents the percent change in what our method recommends the retailer should spend in each of the brands. Dark blue represents a higher recommended investment, whereas dark red represents a sharp reduction in the investment. Yellow represents are new brands that the particular store was not carrying before, but our method recommends should carry now. The filters at the top allow the decision-maker to drill down to various levels of granularity (department, sub-department, brand, size, etc.) and the controls on the left hand side allow the decision-maker to add various constraints such as brand-level budget constraints, store-level budget constraints, etc. The tool offers several other features. A detailed demo is available at: `https://youtu.be/GUREmi4QUoYTo`.

The tool has eased the adoption of the recommendations generated by the system because it made it clear to the managers that the automated tools supplement – as opposed to replace – their decision process.

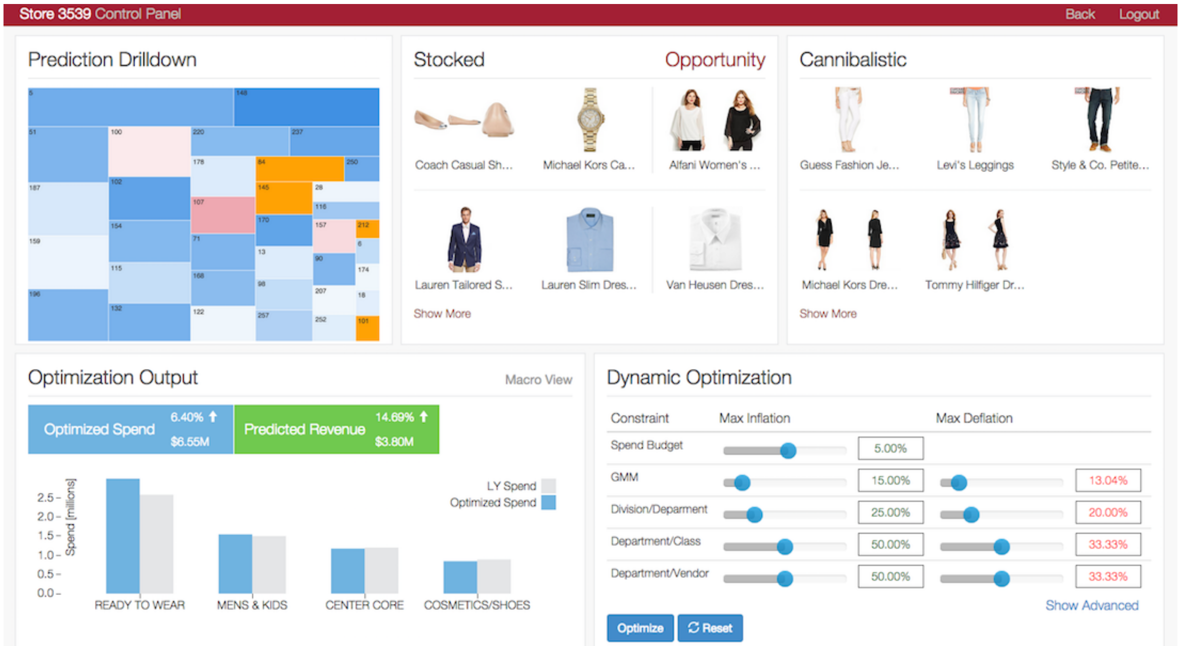

Figure 2: The graphical user interface that was designed by our industry partner to make the decision support system adaptive and simple to use and adapt; see main text for details. Image credit: Celect, Inc. (`http://celect.com`)

5. Impact of the methodology

5.1. Industry partner: Celect

Our industry partner is a Boston-based start up company called Celect. Celect is a predictive analytics platform focused on the retail sector. In fall 2014, Celect won the 2014 Demo God Award in the "Smart Data" category. Further, MIT CSAIL noted that Celect's underlying technology in their top 50 greatest innovations, along side notables like Akamai and technologies like encryption and robotics. Despite being a young company, Celect already boasts of several large clients such as Urban Outfitters, Zipcar, Bon-Ton, Anthropologie, and Free People.

5.2. Implementation and impact

Celect used our methodology at one of its large US retail clients. This large department store retailer needed a better way to personalize the customer experience and provide hyperlocal offerings at each of their physical stores. For starters, they needed to gain a better understanding of their customers buying patterns and preferences. This means generating choice models (and preferences) for every customer and transaction. The result of this is a precise understanding of similarities in customer buying patterns and deep insights into how products relate to each other in both positive and negative ways. Ultimately, this is to better understand what the optimized assortment is for a customer shopping at a particular location.

The retail client looked to Celect to optimize assortments across departments, classes, categories, and vendors. This is to ensure that stores are assorted based on accurate and predictable customer preferences and demand. The retailer recognized that the opportunity to optimize the assortment at the local store level had the potential to provide a much more meaningful impact on the customer experience and the bottom line.

Celect was able to very accurately identify brands and departments with high potential at specific stores. The financial impact of the Celect technology was assessed through a standard control experiment. The retailer chose the control stores and the test stores were selected based on similarities in purchase patterns. Celect technology was used to make assortment decisions at the test stores, whereas the retailer's existing techniques were used at the control stores. The experiment was done over a period of a quarter and the revenue figures were compared. To assess the incremental revenue impact, the standard 'double difference' methodology was used to account for differences both within and across the stores; essentially, the same store differences due to the same store trend were accounted for by subtracting lagging trends and then across store differences were accounted for by subtracting the control store lifts from the test store lifts, where the controls were picked by the retailer. Departments modified by Celect yielded an incremental 6.65% increase in revenue. This was against our predicted increase of 6.43%, which shows a high degree of accuracy.

With Celect in place, this retailer had a much more precise understanding of similarities in customer buying patterns. Insights were discovered about how products relate to each other along with the positive and negative impacts products may have on each other. For example, a certain category of high-end dresses was found to have a positive “halo” effect on a moderate counterpart. In this case, a 4% lift in high-end dresses accompanied a 35% lift on their moderate counterpart at one store. In another store, a 7% lift in high-end dresses accompanied a 46% lift on their moderate counterpart. Celect recommended increases spend in both of these categories at both stores – uncovering latent potential and resulting in immediate increase in sales and conversions across the board.

Overall, this retailer saw an approximate 7% in-store revenue increase as a result of optimized merchandise assortments and allocation.

In addition, Celect received press coverage for the successful implementation of its technology at its clients: Levine (2015), Nanos (2015).

6. Key learnings and future work

There were several key learnings from the project. First, making accurate demand *predictions* is the biggest challenge to designing a practical assortment optimization solutions. It is especially true for fashion apparel retailers who sell short life-cycle products in an uncertain demand environment. It is also becoming increasingly true for more traditional retailers because of the increase in product variety and the corresponding decrease in product life times.

Second, while the volume of data generated and collected is large, the dimension of the corresponding decision space is also large, making the information available for predictions very sparse. For instance, even though we have a large volume of transactions data in our applications, there are tens of millions of customers and millions of products, resulting in only a few observations per customer. This makes the assortment problem a fertile ground for “big data” technologies that can take in large amounts of unstructured/semi-structured data and provides meaningful decisions from data. This requires sophisticated techniques that can efficiently extract meaningful signals from the sparse data and scale to handle large volumes of data.

Finally, the retail landscape is rapidly changing and it is predicted that retail will change more in the next five years than the last fifty (Fitch, 2015). There is an increasing convergence of the online and traditional brick-and-mortar channels, resulting in new challenges to assortment planning and providing easy access to new types of data. In addition, there have been cultural shifts with the retailers showing increasing willingness to adopt automated decision support systems.

All of the above provide a fertile ground for further academic inquiry from marketers, statisticians, and the machine learning community into the area of assortment planning to obtain solutions that can find successful practical application. Particularly, with practitioner willingness to adopt proposed solutions and availability of rich data, this is an opportune time for further academic study into (choice models based) demand models and estimation techniques that are geared towards predictive accuracy, but are also amenable to tractable optimization.

Future work: Our work opens the doors for future work on two fronts. From the technological standpoint, we can extend our assortment optimization framework to incorporate other practical aspects such as supplier effects (such as joint replenishment, lead time uncertainty, and dual sourcing), promotions (the impact of display and price promotions, personalized coupons, and non-monetary rewards such as loyalty points), and personalized assortments (either online or as a sales associate tool in brick-and-mortar stores). From the practical implementation standpoint, our methodology also applies to other aspects of assortment planning such as product design, deciding the styles to carry for the next season, and deciding the overall order quantities. With the experience and credibility gained from optimizing the allocation of inventory dollars, applying our methodology to these other aspects is a natural next step.

Acknowledgements

We acknowledge that the implementation we report in Sections 4 and 5 in this document was done by the team at Celect, Inc. We thank Akshay Anand, John Andrews, Juan Chaneton, Todd Harris (a special thanks for help with Section 5), Ying Zong Huang, Vin Sachidananda, Vig Sachidananda and Greg Yu.

References

Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

Felipe Caro and Jérémie Gallien. Inventory management of a fast-fashion retail network. *Operations Research*, 58(2):257–273, 2010.

Celect, LLC. Future of retail (e-book). `http://celect.com`, 2016. [Online; accessed: 03-June-2016].

Sandeep R Chandukala, Jaehwan Kim, and Thomas Otter. Choice models in marketing, 2008.

Dragos Florin Ciocan and Vivek Farias. Model predictive control for dynamic resource allocation. *Mathematics of Operations Research*, 37(3):501–525, 2012.

David L Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

Vivek F Farias, Srikanth Jagabathula, and Devavrat Shah. Sparse choice models. *arXiv preprint arXiv:1011.4339*, 2010.

Vivek F Farias, Srikanth Jagabathula, and Devavrat Shah. A nonparametric approach to modeling choice with limited data. *Management Science*, 59(2):305–322, 2013.

Kris Johnson Ferreira, Bin Hong Alex Lee, and David Simchi-Levi. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 2015.

Fitch. Retail 2020: Retail will change more in the next 5 years than the last 50. `http://www.slideshare.net/FITCH_design/retail2020-pdf`, 2015. [Online; accessed: 27-January-2017].

Peter M Guadagni and John DC Little. A logit model of brand choice calibrated on scanner data. *Marketing science*, 2(3):203–238, 1983.

Srikanth Jagabathula. *Nonparametric choice modeling: applications to operations management*. PhD thesis, Massachusetts Institute of Technology, 2011.

Srikanth Jagabathula. Assortment optimization under general choice. *New York University working paper*, 2014. URL http://ssrn.com/abstract=2512831.

A Gürhan Kök, Marshall L Fisher, and Ramnath Vaidyanathan. Assortment planning: Review of literature and industry practice. In *Retail supply chain management*, pages 99–153. Springer, 2008.

Peter Leeflang, Dick R Wittink, Michel Wedel, and Philippe A Naert. *Building models for marketing decisions*, volume 9. Springer Science & Business Media, 2013.

Barry Levine. Celect lands $5M to help stores figure out which sweaters you like best. http://venturebeat.com/2015/06/25/celect-lands-5m-so-it-can-help-stores-figure-out-which-sweaters-you-like-best, 2015. [Online; accessed: 27-January-2017].

John DC Little. Models and managers: The concept of a decision calculus. *Management science*, 50 (12_supplement):1841–1853, 2004.

Qian Liu and Garrett Van Ryzin. On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management*, 10(2):288–310, 2008.

C. Mallows. Non-null ranking models. *Biometrika*, 44(1):114–130, 1957.

J. Marden. *Analyzing and Modeling Rank Data*. Chapman and Hall, 1995.

Chris Matyszczyk. MIT prof: Netflix has its recommendations wrong. http://www.cnet.com/news/mit-prof-netflix-has-its-recommendations-wrong/, 2011. [Online; accessed: 27-January-2017].

Janelle Nanos. The hat whisperers: This boston startup tracks shopper's habits to help brick-and-mortar stores make sales. http://www.betaboston.com/news/2015/06/25/the-hat-whisperers-this-boston-startup-tracks-shopping-habits-to-help-brick-and-mortar-stores-makes-sales/, 2015. [Online; accessed 27-January-2017].

Sahand Negahban, Sewoong Oh, and Devavrat Shah. Iterative ranking from pair-wise comparisons. In *Advances in Neural Information Processing Systems*, pages 2474–2482, 2012a.

Sahand Negahban, Sewoong Oh, and Devavrat Shah. Rank centrality: Ranking from pair-wise comparisons. *arXiv preprint arXiv:1209.1688*, 2012b.

Netflix.com. Netflix prize. `http://www.netflixprize.com`, 2009. [Online; accessed: 27-January-2017].

D. Shah, V.F. Farias, S. Jagabathula, and A.T. Ammar. System and method for finding mood-dependent top selling/rated lists, December 1 2015. URL `https://www.google.com/patents/US9201968`. US Patent 9,201,968.

D. Shah, V.F. Farias, S. Jagabathula, and A.T. Ammar. System and method for providing personalized recommendations, February 2 2016. URL `https://www.google.com/patents/US9251527`. US Patent 9,251,527.

Colleen Taylor. Is it time to retire the 5-star rating system? `https://gigaom.com/2011/07/13/recommendation-engine-research/`, 2011. [Online; accessed: 27-January-2017].

Berend Wierenga. *Handbook of Marketing Decision Models.* Springer Science + Business Media, New York, January 2008. doi: 10.1007/978-0-387-78213-3.