Assortment Optimization under General Choice

Vivek Farias • Srikanth Jagabathula • Devavrat Shah∗

1. Introduction

We consider the static assortment optimization problem – a central decision problem faced by operations managers that has been widely studied in the literature. The decision deals with finding the assortment of products (from a larger universe of products) that maximizes the expected revenue subject to a constraint on the size of the assortment. More concretely, if \( N = \{1, 2, \ldots, N\} \) denotes a set of \( N \) products, our goal is to solve

\[
\arg \max_{|M| \leq C} R(M),
\]

where \( R(\cdot) \) is the revenue subroutine that returns the expected revenue from offering each assortment \( M \subset N \), and \( C \leq N \) constrains the size of the assortments that can be offered. We need two components to solve the decision problem: (1) the revenue subroutine \( R(\cdot) \) that uses historical sales transactions to predict expected revenues \( R(M) \) from each assortment \( M \), and (2) an optimization algorithm that uses the subroutine to find the optimal assortment. In this work, we focus on the design of the optimization algorithm and assume we are given access to a revenue subroutine \( R(\cdot) \). Such a subroutine can be designed using any of the existing parametric\(^1\) or nonparametric\(^2\) approaches.

Finding the optimal assortment in general requires an exhaustive search over all assortments of size at most \( C \) needing \( O(N^C) \) calls to the revenue subroutine. This is prohibitive for large values of \( N \) or \( C \). In order to design more efficient algorithms, most existing approaches impose parametric structures on the revenue subroutine. There are two main issues with the existing algorithms:

1. Even after the imposing a reasonable structure, the problem remains computationally hard. The only exception to this rule is the multinomial logit (MNL) model for which efficient optimization algorithms have been proposed. For more complex models like the nested logit (NL) and mixed logit, the assortment optimization problem has either been proved to be computationally hard or believed to be hard.

2. The algorithms (both exact and approximate) that have been proposed so far are tailored to specific choice structures. As a result, they can’t be readily used with more general revenue subroutines.

We address the above issues in this work by proposing a simple and computationally efficient general optimization algorithm that

1. can be used with any general revenue subroutine, and

2. finds the optimal assortment efficiently when the revenue subroutine assumes the MNL structure.

We also show through an empirical study that our optimization algorithm: (a) finds good approximations for structures more complex than MNL model and (b) allows the use of very general nonparametric methods in predicting revenues resulting in much better approximations to the optimal assortment. Before we present the details of our optimization algorithm and summarize our main results, we do a brief survey of the most relevant literature.

Relevant literature. Most existing approaches impose parametric structure on the underlying choice model used for predicting revenues and then exploit this structure to propose efficient algorithms. The two most parametric structures that have been considered are the multinomial logit (MNL) model and the mixture of MNL (MMNL) models. The problem is well-understood in the context of the MNL model. Specifically, two main results have been obtained: (1) When the problem is uncapacitated, or equivalently when \( C = N \), it has been shown that the search can be restricted to at most \( N \) so called profit-ordered assortments (see Talluri and van Ryzin [2004]); (2) When the problem is capacitated \( (C < N) \), Rusmevichientong et al. [2010a] show that the optimal assortment can be found with a complexity of \( O(N^C) \) (they also

\[^{*}\text{VF is affiliated with MIT Sloan, SJ with NYU Stern, and DS with EECS, MIT. email: vivekf@mit.edu, sjagabat@stern.nyu.edu, devavrat@mit.edu}}\]

\[^{1}\text{Parametric approaches include fitting choice models like multinomial, nested, and mixed logit models.}}\]

\[^{2}\text{Nonparametric approaches include a recent method proposed by Farias et al. [2010]}}\]
propose a more complicated implementation that can reduce complexity to \(O(C \log N)\). These two results provide a complete understanding of problem in the context of the MNL model. Unfortunately, beyond the MNL model very little is known. For the MMNL model, Rusmevichientong et al. [2010b] show that even the uncapacitated decision problem with only two mixture components is NP-complete. For the NL model, Rusmevichientong et al. [2009] show that a slightly more general version of the static assortment problem, in which the weighted size (with each each product having weight \(c_i > 0\)) must be less than \(C\), is NP-complete; they propose a PTAS by reducing the problem to a sum-of-rationals problem. In summary, imposing reasonable structures doesn’t alleviate computational hardness of the problem, except in the case of the MNL model.

2. Algorithm and theoretical guarantees

We now describe the algorithm and summarize our theoretical guarantees. Since our setup is very general with little structure to exploit, we adopt the greedy method. However, a naive greedy algorithm – which starts with an empty set and at each stage greedily adds a product that results in the maximum increase in revenue until no more products can be added – fails even in the simple case of the MNL model (see Rusmevichientong et al. [2010a]). In order to overcome this issue, we allow for greedy “exchanges” in addition to greedy “additions.” Particularly, at every stage, we allow a new product to be either added (which we call an “addition”) to the solution set or replace an existing product (which we call an “exchange”) in the solution set; the operation at each stage is chosen greedily. The termination condition now becomes an interesting question. As in the naive implementation, we could terminate the process when addition or exchange no longer results in an increase in revenue. However, since we never run out of products for exchanges, the algorithm may take an exponential (in the number of products) number of steps to terminate. We overcome this issue by introducing a control parameter that caps the number of times a product may be involved in exchanges. Calling that parameter \(b\), we show that the algorithms calls the revenue subroutine \(O(N^2 b C^2)\) times for the capacitated problem.

Next, we state the guarantees for the algorithm. We consider the case when the underlying choice model used by the revenue subroutine is the MNL model. We can show that our algorithm, when run with \(b \geq C\), succeeds in finding the optimal assortment with \(O(N^2 C^3)\) calls to the revenue subroutine. Therefore, in the special case when the underlying choice model is the MNL model, our algorithm captures what is already known. It also provides a simpler alternative to the more complicated algorithm proposed by Rusmevichientong et al. [2010a]. We also prove that our algorithm is robust to noise. In particular, when the underlying model is MNL and the revenue predictions are known to within a factor of \(1 - \varepsilon\) of the true value, our algorithm finds an assortment with revenue that is within \(1 - f(\varepsilon)\) of the optimal value; here \(f(\varepsilon)\) goes to zero with \(\varepsilon\) and also depends on \(C\) and the parameters of the underlying model.

3. Empirical evaluation

In this section, we present the results from an empirical study. We conducted two studies to show that our algorithm: (a) finds good approximations for structures more complex than MNL model and (b) allows the use of very general nonparametric methods in predicting revenues resulting in much better approximations to the optimal assortment. For our studies, we started with a mixed logit model in which customers are assumed to belong to \(L\) different classes and within each class make choices according to a different MNL model. The resulting model is a discrete mixture of \(L\) different MNL models with appropriately chosen mixing weights. We generated a family of 1000 mixed logit models with \(L = 5\) customer classes as follows: for each model, choose the prices of the products uniformly at random (u.a.r) from the interval [100, 150], the mixing weights of the \(L\) classes u.a.r. from [0, 1] and normalize to sum to 1, the MNL weights for each class by choosing 4 products u.a.r and setting their weights u.a.r. from [0, 10] and setting the weights of the remaining products to 0.

For our first study, we used each of the above 1000 mixed logit models as the underlying ground truth for the revenue subroutines. We then used these revenue subroutines with our assortment optimization algorithm to find approximations to optimal assortments for all values of capacities \(C\). The results for \(N = 10\) and \(N = 15\) products are summarized in the first two columns of Table 1. It is immediate from our
results that in more than 98% of instances our algorithm found the optimal assortment. In the remaining less than 2% of instances the average optimality gap (OptGap) is about 3%. We conclude that our method is successful in producing good approximations to the optimal assortment when the underlying choice model is more complex than the MNL model.

For the second study, we considered a more realistic scenario where we don’t know the underlying choice model but know only sales transaction data. We simulated the aggregate transaction data for each of the 1000 mixed logit models as follows: for every pair of products \( i, j \) we simulated the purchases according to the mixed logit model of a large number of customers to compute the fraction of purchases of each of the products \( i, j \) when both are on offer. Given this transaction data, we estimated the optimal assortment in two ways: (1) force-fitted an MNL model to the transaction data and computed the exact optimal assortment under the MNL model, and (2) used the nonparametric method proposed by Farias et al. [2010] (which we call the ‘robust’ approach) as the revenue subroutine and computed the optimal assortment using the approximation algorithm proposed in this paper. The last two columns of Table 1 present the results. Figure 2 compares the histograms of OptGaps over suboptimal instances. It is evident from the results that the second method vastly outperforms the first method. We draw the following conclusion from this result. A priori it is not clear which of the two methods yields better performance. On the one hand, model misfit errors result in low revenue prediction accuracy when we force fit an MNL model; however, due to its simplicity, we can solve the optimization problem to optimality. On the other hand, the nonparametric method generates more accurate revenue predictions; whereas, the optimization problem can’t be solved to optimality due to its complexity. Our algorithm provides a way to approximately find the optimal assortment. In addition, our empirical results show that using a more complex method that produces more accurate revenue predictions and then using it to find an approximation of the optimal assortment yields better performance than fitting an incorrect model and solving the corresponding optimization problem to optimality.

References


