A Near-Optimal Exploration Exploitation Approach for Assortment Selection

Shipra Agrawal, Columbia University
 Vashist Avadhanula, Columbia University
 Vineet Goyal, Columbia University
 Assaf Zeevi, Columbia University

In retail settings, the assortment of products selected by a retailer has significant impact on demand for individual products and resulting revenue. Choice models capture substitution effects among products by specifying the probability that a consumer selects a product given the offered set. In most cases, the choice model belongs to a family and historical sales data are used to estimate these parameters and then decide assortment. This approach becomes challenging in many e-commerce settings such as fast fashion retail, where many products have short selling seasons and historical data may not be a reliable indicator for future choices. We consider a stylized dynamic optimization problem that captures some salient features of this application domain. Our goal is to develop an exploration-exploitation policy that simultaneously learns from current observations and exploits this information gain for future decisions. In particular, we consider a cardinality constrained assortment selection problem under the Multinomial logit (MNL) model with N substitutable products and a no purchase option, specified by unknown parameters $v_0, v_1, \dots, v_N$. Under this model the probability that a consumer purchases product i when offered an assortment $S \subset \{1, \dots, N\}$ is given by,

$$p_i(S) = \begin{cases} \frac{v_i}{v_0 + \sum_{j \in S} v_j}, & \text{if } i \in S \cup \{0\} \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

and the expected revenue when assortment S is offered is given by

$$R(S) = \sum_{i \in S} r_i p_i(S) = \sum_{i \in S} \frac{r_i v_i}{1 + \sum_{j \in S} v_j}. \quad (2)$$

We have a specified time horizon T and can offer at most K products at each time step. If S^* is the optimal assortment, our goal is to offer a sequence of assortments, $S_1, \dots, S_T$ such that the cumulative expected revenues over said horizon is maximized,

A full version of this paper appears at: http://www.columbia.edu/~vg2277/dynamic_assort.MNL.pdf

V. Goyal is supported by the NSF grants CMMI 1201116 and CMMI 1351838. Author's addresses: S. Agrawal, V. Avadhanula, V. Goyal, and A. Zeevi, Columbia University, New York 10027. Email: {sa3305, va2297, vg2277, ajz2001}@columbia.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

EC'16, July 24–28, 2016, Maastricht, The Netherlands.

ACM 978-1-4503-3936-0/16/07.

<http://dx.doi.org/10.1145/2940716.2940779>

or alternatively minimize the *regret*, defined as

$$Reg(T) = \sum_{t=1}^T R(S^*) - E[R(S_t)], \quad (3)$$

where $R(S)$ is the expected revenue when assortment S is offered as defined in (2).

This problem can be formulated as a classical multi-armed bandit (MAB) problem by considering every assortment as an arm, however that would lead to exponentially many arms. Other authors (see [Chen et al. 2014]) have considered versions of MAB where one can play a subset of arms in each round and the expected reward is a function that depends on the subset played. However, their approach assumes that the reward for each arm is observable and generated independent of the other arms in the subset. This is not the case typically in retail settings, and in particular, in the MNL choice model, where the retailer observes only purchase decisions which depend on the assortment of products offered in a time step. [Sauré and Zeevi 2013], [Rusmevichientong et al. 2010] consider a similar problem of minimizing regret under MNL model when consumer preferences are unknown. Both works follow an *explore first and exploit later* approach, where a fixed set of assortments are explored to learn the MNL parameters to a desired (known) accuracy, and then this information is exploited for the rest of the selling period. However, their regret bound assumes a known constant “separability” parameter.

Our main contribution is an algorithm that interweaves exploration and exploitation and achieves a *parameter independent* regret bound of $O(\sqrt{NT} \log T + N \log^3 T)$. Here the Big-Oh notation is hiding only absolute constants. It is interesting to note that there is no dependence on the cardinality constraint K in our regret bound, despite the combinatorial complexity that is dictated by the relationship between N and K . We also establish a lower bound, by showing that any exploration-exploitation algorithm for the cardinality constrained assortment selection problem under the MNL model must incur a regret of $\Omega(\sqrt{\frac{NT}{K}})$ for $K < N$. This establishes that the performance of our algorithm is nearly optimal within a factor of $\sqrt{K}$.

Our algorithm is based on the upper confidence bound framework, also referred to as optimism in face of uncertainty. A main challenge in our problem is that the customer response on offering an assortment S depends on the entire set S , and does not directly provide an unbiased sample of demand for a product $i \in S$. We design a novel epoch based approach, where in every epoch, we offer an assortment S multiple times: a chosen S is offered repeatedly until a no-purchase happens. We show that on proceeding in this manner, the average number of times a product i is purchased provides an unbiased estimate of parameter v_i . Given these unbiased estimates, our algorithm uses UCB framework to maintain high confidence regions for each of the parameters. Another challenge is the computational complexity of finding the most optimistic assortment among exponentially many assortments. We use the properties of MNL choice model to provide a nontrivial reduction of this problem to a static assortment optimization problem, for which efficient optimization methods are available.

REFERENCES

- W. Chen, Y. Wang, and Y. Yuan. 2014. Combinatorial Multi-Armed Bandit and Its Extension to Probabilistically Triggered Arms. *CoRR* (2014).
- P. Rusmevichientong, Z. M. Shen, and D.B. Shmoys. 2010. Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations research* (2010).
- D. Sauré and A. Zeevi. 2013. Optimal Dynamic Assortment Planning with Demand Learning. *Manufacturing & Service Operations Management* (2013).