Buy It Again: Modeling Repeat Purchase Recommendations

Rahul Bhagat
Amazon.com Inc.
Seattle, WA
rbhagat@amazon.com

Srevatsan Muralidharan
Amazon.com Inc.
Seattle, WA
srevatsa@amazon.com

Alex Lobzhanidze
Amazon.com Inc.
Seattle, WA
alexlob@amazon.com

Shankar Vishwanath
Amazon.com Inc.
Seattle, WA
shavis@amazon.com

ABSTRACT

Repeat purchasing, i.e., a customer purchasing the same product multiple times, is a common phenomenon in retail. As more customers start purchasing consumable products (e.g., toothpastes, diapers, etc.) online, this phenomenon has also become prevalent in e-commerce. However, in January 2014, when we looked at popular e-commerce websites, we did not find any customer-facing features that recommended products to customers from their purchase history to promote repeat purchasing. Also, we found limited research about repeat purchase recommendations and none that deals with the large scale purchase data that e-commerce websites collect. In this paper, we present the approach we developed for modeling repeat purchase recommendations. This work has demonstrated over 7% increase in the product click through rate on the *personalized recommendations page* of the Amazon.com website and has resulted in the launch of several customer-facing features on the Amazon.com website, the Amazon mobile app, and other Amazon websites.

CCS CONCEPTS

• **Information systems** → **Personalization**; **Recommender systems**; **Electronic commerce**;

KEYWORDS

Personalization, Recommender systems, E-commerce, Repeat purchases

ACM Reference Format:
Rahul Bhagat, Srevatsan Muralidharan, Alex Lobzhanidze, and Shankar Vishwanath. 2018. Buy It Again: Modeling Repeat Purchase Recommendations. In *KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 19–23, 2018, London, United Kingdom.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3219819.3219891

1 INTRODUCTION

The Amazon.com website and the Amazon mobile app are popular destinations that customers use to purchase a wide variety of products (e.g., books, clothing, groceries, etc.) and to use a variety of other services (e.g., music listening, video streaming, etc.). Recommendations are shown throughout these Amazon experiences to help customers with their shopping needs. These recommendations help customers discover new products (or content) and help them complete their shopping missions. They are both personalized (e.g., Recommended for you) and non-personalized (e.g., Customers who bought this item also bought).

Majority of the recommendations on Amazon.com and other websites today are built around the following core idea: Given that a customer has purchased (or viewed or rated) a specific set of products, can we recommend other similar products that they are likely to purchase [15, 17]. The algorithms that power these recommendations are generally based on collaborative filtering [17] and websites often have many customer-facing features that make these recommendations using various strategies [4, 9, 15]. More recently, deep learning based methods have been explored to generate these recommendations [3, 18].

However, there is another aspect of recommendations that has been explored much less. The idea is the following: Given that a customer has purchased a specific set of products, can we recommend them products from their purchase history that they are likely to purchase again. We call these **repeat purchase recommendations**. The repeat purchasing phenomenon is certainly prevalent for consumable products (e.g., toothpastes, diapers, cat food, etc.). But our analysis shows that repeat purchase behavior is present even in other categories like electronics, computers, etc. (e.g., HDMI cables, memory cards, etc.). Similar repeat behavior also exists in scenarios such as music listening where customers are likely to listen to their favorite songs again and again, where the behavior can be called repeat listening. In this setting, timely recommendations of such 'deemed' repeat purchasable products (or repeat listen for the music scenario) may serve as reminders and can provide immense value to customers. Crucial to the success of this recommendation strategy is to predict if a customer is likely to repeat purchase a product, and if so, when is the right time to recommend it to them.

To explore this specific aspect of recommendations, we did a deep dive into academic literature to learn if other researchers have worked on modeling repeat purchase recommendations. While we could not find much work in the area of recommendations, we found some good references for research on related problems in the area of marketing science. Marketing science literature has looked at the problem of modeling repeat purchase behavior with focus on

KDD '18, August 19–23, 2018, London, United Kingdom

ACM ISBN 978-1-4503-5552-0/18/08.
https://doi.org/10.1145/3219819.3219891

predicting aggregate customer behavior and predicting long term impact of a customer's repeat purchase behavior. A few different models have been proposed in that area. For example, the Negative Binomial Distribution (NBD) model [6, 10], models the repeat buying behavior of customers for consumer brands (e.g., brands of shampoos). Given the repeat purchasing behavior of a sample set of customers over an extended time period (e.g., 12 weeks), the goal is to predict how many customers will repeat buy the same brand 0 times, 1 time, 2 times, etc. in an upcoming fixed time period (e.g., four weeks). Other models that try to model similar problems are the Erlang-2-Gamma model [2] which looks at brand prediction, the Pereto-NBD [16] which looks at finding customers who are likely to do business again with a company based on past transaction, the Beta-geometric-NBD (BG-NBD) [8] which focuses on predicting the life time value of a customer for a company, and some other models that focus on analyzing a company's customer base [7, 12].

More recently, Kapoor *et al.* [14] developed models for predicting the return time of a user coming back to a website based on Cox's proportional hazards model. Building on this work, Kapoor *et al.* [13] developed a semi-Markov model to predict the revisit time for familiar items by users. Their model accounts for latent psychological factors such as sensitization and boredom involved in repeat consuming the same items. Further, they propose creation of a recommender based on their predictive model and provide offline metrics on the performance of their recommender on publicly available user activity datasets. In this context we should note that our problem and our approach are much different than the work in Kapoor *et al.* [13, 14].

Another related work that is worth citing here is that of Dey *et al.* [5] who introduced a model for making broader estimates of repeat purchases of products by customers over longer time durations. This model is based on an approach similar to the NBD model [6] and assumes that subsequent repeat purchases of products by customers are not correlated with each other. Since their model generates a broad estimate similar to the NBD model [6], Dey *et al.* [5] note that their model should be used for improving the time sensitivity of recommendations generated by existing recommender systems and not for generating standalone repeat purchase recommendations. While our goal and the models discussed in this paper are both different than the approach presented in Dey *et al.* [5], this is one work we found in the area of recommendations that looks at repeat purchasing of products. In the context of this work, we should note that our models (patent pending) were developed independently and prior to the publication of work by Dey *et al.* [5].

More generally, we should note that our specific goal is to create standalone repeat purchase recommendations which is different from the goals in the work described above. Also, in the context of generating standalone repeat purchase recommendations, both our intuition and the observed shopping data for consumables products such as paper towels, toilet papers, etc. on Amazon requires us to develop models that carefully capture the time correlations between subsequent repeat purchases of products by customers. This is because the time at which a customer is likely to repeat purchase such a product depends on when they last purchased it and how quickly they run out of it. This is another difference between our work and the other work discussed above.

In this paper, we present the various models we developed to introduce and launch repeat purchase recommendations on the Amazon.com website leading to over 7% increase in the product click through rate on its *personalized recommendations page*. The remainder of the paper is organized as follows: Section 2 describes our modeling approach; Section 3 describes the different models we developed for repeat purchase recommendations; Section 4 presents an analytical comparison of our different models; Section 5 and Section 6 summarize our offline and online experiments respectively; and Section 7 presents our conclusion and future work.

2 MODELING APPROACH

Before we begin discussing our models for repeat purchase recommendations, we want to discuss some key intuitions that informed our modeling decisions. In this section, we first present these intuitions and then we formally define our problem.

2.1 Modeling Intuition

As we think about modeling repeat purchase recommendations, a few approaches come to mind. One approach is to rank repeat purchase recommendations of a customer in the descending order of the number of times they repeat purchased the products. Intuitively, this makes sense since if a customer repeat purchased a specific product multiple times, we can expect them to do so again. However, for scenario's such as diaper purchases, where customers are likely to make frequent purchases of diapers within a window of time and none outside the window of time, the diapers might still rank high in the recommendations even though they are not timely relevant to them anymore. This would lead us to think of a time-decay based model where the repeat purchase score is decayed based on some pre-specified half-life (or lives). But this approach has a problem too: such a time-decay based model would assign the highest score to a product right after the repeat purchase has been made thus increasing its rank in the recommender. In contrast, the need for a customer to repeat buy a product immediately after their purchase of the product is likely to go down. Hence, this approach is counter-intuitive as well, at least for consumable products. Thus, it is crucial to model the time based relevancy of products to customers while modeling their repeat purchase recommendations.

This brings us to the second approach, which is to assume that repeat purchasing of products by customers is a periodic phenomenon. Intuitively, this makes sense as well. If we think about a product like the protein bar, we can expect a certain rate of consumption (e.g., one bar per day) and thus correspondingly expect a certain periodicity of repeat purchase (e.g., one box every two weeks).

A third approach and a natural extension would be to assume that both the above factors, i.e., the number of times a customer repeat purchased a product and their repeat purchase periodicity play an important role (and there are perhaps other factors as well like the product category, etc.) in the problem of repeat purchase recommendations. In this paper, we will explore a few such approaches.

2.2 Problem Formulation

We formally define the problem of repeat purchase recommendations as follows: Given a customer's history of product purchases (including repeat purchases), we would like to estimate the probability of the customer repeat purchasing a product as a function of time from their last purchase of that product. In other words, say a customer C_j purchased a product A_i k times in the past with time intervals $t_1, t_2, t_3 \ldots t_k$, we would like to estimate the purchase probability density

$$P_{A_i}(t_{k+1} = t | t_1, t_2, t_3, \ldots t_k) \tag{1}$$

A hidden assumption that is made while writing Equation 1 is that the purchase events of different products are independent of each other.

A second assumption that we make is that Equation 1 is decomposable into two major components:

$$P_{A_i}(t_{k+1} | t_1, t_2, t_3 \ldots t_k) \approx Q(A_i) \times R_{A_i}(t_{k+1} | t_1, t_2, t_3, \ldots t_k, A_i = 1) \tag{2}$$

where the first term in the r.h.s. of Equation 2 $Q(A_i)$ is the repeat purchase probability of a customer buying a product a $(k+1)^{th}$ time given that they have bought it k times and the second term R_{A_i} is the distribution of t_{k+1}, conditioned on the customer repurchasing that product; indicated by $A_i = 1$.

A third simplifying assumption we make is that the time distribution $R_{A_i}(t_{k+1} | t_1, t_2, \ldots t_k)$ is $\approx R_{A_i}(t | t_1, t_2, t_3, \ldots, t_k)$. In view of clarity, it is worth mentioning that while $\int_0^\infty R_{A_i}(t)dt = 1$; integral $\int_0^\infty P_{A_i}(t)dt \leq 1$.

3 REPEAT PURCHASE RECOMMENDATIONS MODELS

Based on the framework described in Section 2.2, we present various models for repeat purchase recommendations.

3.1 Repeat Customer Probability Model

The first model that we consider is a simple time independent frequency based probabilistic model that uses aggregate repeat purchase statistics of products by customers. For each product A_i, we compute its repeat customer probability (RCP) as shown below:

$$\mathrm{RCP}_{A_i} = \frac{\text{\# customers who bought product } A_i \text{ more than once}}{\text{\# customers who bought the product } A_i \text{ at least once}} \tag{3}$$

The simplifying assumption we make is that $P_{A_i}(t_{k+1} | t_1, t_2, t_3 \ldots t_k)$ is approximately given by RCP_{A_i}, i.e., we assume:

$$P_{A_i}(t_{k+1} | t_1, t_2, t_3 \ldots t_k) \approx Q(A_i) \approx \mathrm{RCP}_{A_i} \tag{4}$$

and ignore the time factor altogether, i.e., we assume that R_{A_i} is a fixed constant r for all products A_i. Additionally, to ensure that the quality of repeat purchase recommendations are good, a threshold is enforced on RCP_{A_i} such that only the products that satisfy Equation 5 are deemed *repeat purchasable* and vice-versa.

$$\mathrm{RCP}_{A_i} > r_{\text{threshold}} \tag{5}$$

Finally, recommendations are generated by considering all the repeat purchasable products previously bought by customers and ranking them in the descending order of their estimated probability density $P_{A_i}(t)$ at a given time t using Equation 4. As noted previously, the RCP model is a simple probabilistic model and it is our first model for repeat purchase recommendations. Given the model's simplicity, we treat the RCP model as our baseline that subsequent models should improve upon.

3.2 Aggregate Time Distribution Model

Our analysis shows that for most customers we only have a few repeat purchases for a specific product. But, for many products, we have a large number of customers who have repeat purchased those products. This leads to the idea of a time based model that uses aggregate repeat purchase behavior of a product across all repeat purchasing customers to determine its repeat purchase characteristics.

Specifically, our goal is to determine the distribution of repeat purchase time intervals (t) of a product across all of its repeat purchasing customers. To determine this distribution, we selected a random sample of products. For each product in this sample, we looked at all customers who had repeat purchased that product and obtained the mean repeat purchase time interval for each of the customers. Figure 1 shows the distribution of repeat purchase time intervals of a random consumable product. Both in view of scalability as well as simplicity, we went down the route of fitting the observed repeat purchase intervals to various well known parametric distributions such as log-normal and gamma distributions, whose parameters were determined by the maximum-likelihood principle. Empirically, we found that the log-normal distribution, as defined by Equation 6, had the best fit for most consumable products.

$$R_{A_i}(t) = \ln\mathcal{N}(t; \bar{\mu}_i, \bar{\sigma}_i) = \frac{1}{\sqrt{2\pi}t\bar{\sigma}_i}\exp\Big[-\frac{(\ln t - \bar{\mu}_i)^2}{2\bar{\sigma}_i^2}\Big], \; t > 0. \tag{6}$$

Figure 2 shows the QQ plot for log-normal distribution fit of repeat purchase time intervals and Figure 3 shows the distribution of log repeat purchase time intervals for the same consumable product as Figure 1. Thus, for every product that was deemed repeat purchasable, the parameters of the log-normal distribution are estimated in an empirical fashion [1] by fitting them to the repeat purchase time intervals (t) across all its repeat purchasing customers.

This leads to a simple recommendations model where R_{A_i} is estimated using Equation 6. The second assumption we make is that $Q(A_i)$ is a fixed constant q for all products A_i at any given time t. Additionally, as noted previously, only the products that satisfy Equation 5 are deemed *repeat purchasable* and vice-versa. Finally, recommendations are generated by considering all the repeat purchasable products previously bought by customers and ranking them in the descending order of their estimated probability density $P_{A_i}(t)$ at a given time t using Equation 2. We call this model the Aggregate Time Distribution (ATD) model. It should be noted that while simple, we are not aware of any previous work where the ATD model was used to generate repeat purchase recommendations.

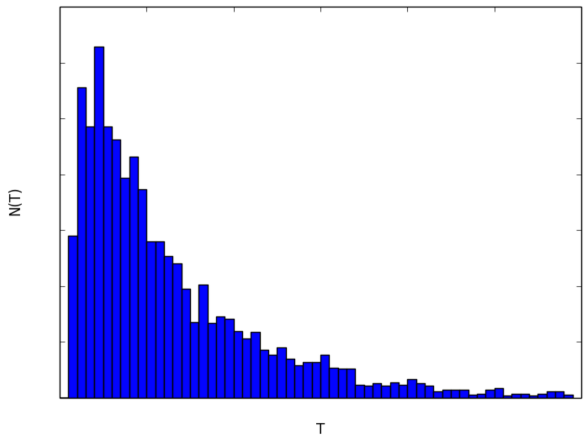

Figure 1: Distribution of repeat purchase time intervals (t) of a random consumable product across all its repeat purchasing customers

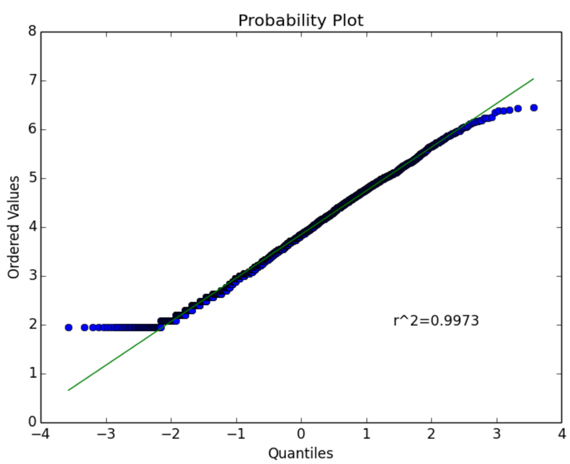

Figure 2: QQ plot for log-normal distribution fit of repeat purchase time intervals (t) of a random consumable product across all its repeat purchasing customers

3.3 Poisson-Gamma Model

In Section 1, we described several models from the marketing science literature that focused on modeling repeat purchase behavior in the context of predicting aggregate customer behavior and predicting long term impact of a customer's repeat purchase behavior. While none of these models were developed in the context of repeat purchase recommendations for individual customers, these models do have an inherent prediction model that can help in generating and ranking recommendations. The NBD model [6, 10], the seminal academic work in this field, is based on the following assumptions:

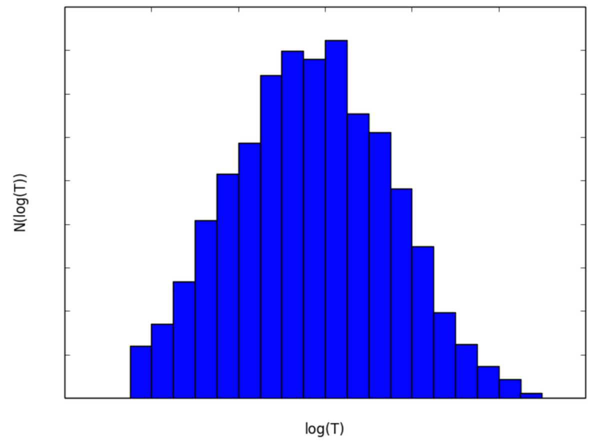

Figure 3: Distribution of log repeat purchase time intervals (t) of a random consumable product across all its repeat purchasing customers

(1) Assume that a customer's repeat purchases follow a homogeneous Poisson's process with repeat purchase rate λ. In other words, they assume that successive repeat purchases are not correlated with each other.
(2) Assume a gamma prior on λ, i.e., assume that λ across all customers follows a Gamma distribution with shape α and rate β.

Thus, the NBD model is Bayesian model where the evidence is distributed as a Poisson and the prior on λ is a gamma prior. Hence this is also called the Poisson-Gamma model (PG).

In the PG model, the parameters of the product-specific gamma distributions are estimated in an empirical fashion [1] by fitting them to the maximum likelihood estimates of the purchase rates of repeat purchasing customers. Then, a Bayesian estimate of the customer's repeat purchase rate is performed by combining the prior distribution with customer's own past purchase history using Equation 7

$$\lambda_{A_i,C_j} = \frac{k + \alpha_{A_i}}{t + \beta_{A_i}}, t > 0 \tag{7}$$

where α_{A_i} and β_{A_i} are the shape and rate parameters of the gamma prior of product A_i; k is the number of purchases of product A_i by customer C_j; and t is elapsed time between the first purchase of product A_i by customer C_j and the current time.

This leads to a recommendations model where R_{A_i} is assumed to be a poisson distribution where the rate parameter is estimated using Equation 7 and the probability mass is estimated using Equation 8

$$R_{A_i,C_j}(t) = \sum_{m=1}^{\infty} \frac{\lambda^m_{A_i,C_j} exp(\lambda_{A_i,C_j})}{m!}, t > 0 \tag{8}$$

where m is the number of expected future purchases. The second assumption we make is that Q_{A_i} is a fixed constant q for all products A_i at any given time t. This assumption is similar to the one made in the ATD model. Additionally, similar to the RCP and ATD models, only the products that satisfy Equation 5 are deemed *repeat purchasable* and vice-versa. Finally, recommendations are generated by considering all the repeat purchasable products previously bought by customers and ranking them in the descending order of their estimated probability density $P_{A_i}(t)$ at a given time t using Equation 2.

It should be noted, this method where the prior distribution is estimated from data is known as the *Empirical Bayesian method* which contrasts with the standard Bayesian methods where the prior distribution is assumed to be fixed before any data are observed. More specifically, the PG model is a *Parametric Empirical Bayes model* which can be considered an approximation of a fully Bayesian hierarchical model where the likelihood and prior take on simple parametric forms. This significantly simplifies the final model. It should also be noted that the Bayesian formulation ensures that we are able to combine a customer's own purchase behavior with the aggregate purchase behavior for a particular product which makes this model inherently *personalized*. Finally, it should be noted that while the PG has been used in past to solve different problems [6, 10], this is the first work that uses this model for generating standalone repeat purchase recommendations.

3.4 Modified Poisson-Gamma Model

The original PG model was developed in the context of predicting aggregate purchasing behavior. Thus while it can be used for making predictions for an individual customer, some of its assumptions are counter intuitive in terms of personalized recommendations.

Specifically, the homogeneous Poisson assumption may not be accurate across all product categories, especially for product categories such as consumables. This is so because from a theoretical standpoint, in a homogeneous Poisson's process, the probability of occurrence of events is a constant and is independent of time. An easy way to understand this is to realize that a Poisson's process is a limiting case of a sequence of Bernoulli processes in the limit of large N and small constant probability and is memoryless. For the case of customer purchase behavior in a number of product categories, this is not expected to be the case, since if a customer bought a product, their need for buying the same product immediately following their current purchase is small and with time this need changes (assuming they have a certain affinity to buy this product at all).

Further, even though a homogeneous Poisson's process is a constant rate process, the estimate of this constant rate as performed by Equation 7 does vary with time. Say, a customer repeat bought a product A_i k^{th} time t time units after its first purchase. The best estimate of the purchase rate just before the k^{th} repeat purchase (see Equation 7) is $(k-1+\alpha)/(t^{-}+\beta)$ and right after is $(k+\alpha)/(t^{+}+\beta)$. Note that the estimated purchase rates right after the purchase is larger than right before the purchase. While this is a reasonable estimate for an assumed 'constant' rate of purchase, using that for ranking the recommendations leads to a situation where a product's estimated purchase rate is largest right after its previous purchase. This is in direct contrast to what one would expect, realistically.

Another issue with the standard PG model is that while it models a products time-distribution, it does not incorporate the product's time-independent repeat customer probability, i.e., RCP into the model. It only uses RCP for filtering. However, RCP is an important time-independent signal that can be incorporated into the recommendations model itself. To address these issues, we propose a modification to the PG model and call the new model the Modified Poisson-Gamma model (MPG).

The MPG model makes the following assumptions:

(1) Assume that a customer's successive purchases are correlated and the repeat purchases follow a process that we call the Modified-poisson process. This process uses a single repeat purchase rate parameter λ and assumes that λ is dependent on the last time the customer repeat purchased that product. In this regard this process is distinct from the homogeneous Poisson's process assumed in the PG model.
(2) Assume a gamma prior on λ, i.e., assume that λ across all customers follows a Gamma distribution with shape α and rate β.

Now, similar to the PG model, the parameters of the product-specific gamma distributions are estimated in an empirical fashion [1] by fitting them to the maximum likelihood estimates of the purchase rates of repeat purchasing customers.

However, to address the specific issues noted with the computation of a customer's personalized repeat purchase rate in the PG model, we propose a modification to Equation 7. The proposed change relies on the fact that we can estimate a customer's known mean repeat purchase time interval for a specific product based on their first and last purchase of that product. This is our best estimate for their mean repeat purchase time interval based on the observed data. So, ideally our model should assign the highest repeat purchase rate at the observed mean. Moreover, if the estimated purchase rate increased gradually till we reach the observed mean time interval and goes down gradually after that, that would match the expected customer repeat purchase behavior. We accomplish these by making some modifications to the purchase rate calculation for the PG model.

Specifically, we assume that t_{purch} is elapsed time interval between the first and last purchase of product A_i by customer C_j; t is the elapsed time interval between the last purchase of product A_i by customer C_j and the current time; and t_{mean} is the estimated mean repeat purchase time interval between successive purchases of product A_i by customer C_j. When $t < 2 * t_{mean}$ we estimate the purchase rate for the MPG model using Equation 9

$$\lambda_{A_i,C_j} = \frac{k + \alpha_{A_i}}{t_{purch} + 2 * |t_{mean} - t| + \beta_{A_i}}, t > 0 \tag{9}$$

and where α_{A_i} and β_{A_i} are the shape and rate parameters of the gamma prior of product A_i; k is the number of purchases of product A_i by customer C_j. Additionally, when $t \geq 2 * t_{mean}$, we estimate the purchase rate for the MPG model using Equation 7. This entails that we choose $2 * t_{mean}$ as the partition point after which the MPG model becomes the same as the PG model. This is because Equation 9 ensures that the purchase rate λ_{A_i,C_j} increases gradually from $t = 0$ to $t = t_{mean}$, achieves its peak value at $t = t_{mean}$, and decreases gradually from $t = t_{mean}$ to $t = 2 * t_{mean}$. This is what we would intuitively expect. At $t = 2 * t_{mean}$, the value for λ_{A_i,C_j} can be obtained by using either Equation 7 or Equation 9 and is exactly the same. So at this point and beyond, the MPG model becomes the same as using the PG model and the value of λ_{A_i,C_j} continues to decrease.

This leads to a recommendations model where the purchase rate parameter λ_{A_i,C_j} is estimated using Equation 9 when $t < 2*t_{mean}$ and using Equation 7 when $t \geq 2 * t_{mean}$. R_{A_i} is then assumed to be a poisson distribution and the probability mass is estimated using Equation 8. The second assumption we make is that we can estimate $Q(A_i)$ using its RCP_{A_i}, i.e., we assume Equation 4. This assumption is similar to that of the RCP model. Additionally, similar to all the previous models, only the products that satisfy Equation 5 are deemed *repeat purchasable* and vice-versa. Finally, recommendations are generated by considering all the repeat purchasable products previously bought by customers and ranking them in the descending order of their estimated probability density $P_{A_i}(t)$ at a given time t using Equation 2.

4 ANALYTICAL MODEL COMPARISON

In this section, we present an analytical comparison of the different repeat purchase recommendations models - RCP model, ATD model, PG model, and MPG model - presented in Section 3.

For the purpose of illustration, let us imagine that there is a customer who has purchased *paper towels* and *laundry detergent* on Amazon. Let us assume that these *paper towels* have an estimated repeat customer probability = 0.22, i.e., $\text{RCP}_{\textit{paper towels}} = 0.22$; most customers repeat purchase these *paper towels* once every 2 months; and this customer has purchased these *paper towels* 4 times in the last 12 months. Similarly, let us assume that this *laundry detergent* has an estimated repeat customer probability = 0.21, i.e., $\text{RCP}_{\textit{laundry detergent}} = 0.21$; most customers repeat purchase this *laundry detergent* once every 3 months; and this customer has purchased the *laundry detergent* 3 times in the last 12 months. Finally, let us also assume that this customer last repurchased the *paper towels* 1 week ago and they last repurchased the *laundry detergent* 3 months ago. Given this scenario, let us compare the behavior of the different models.

Firstly, the RCP model - which only uses the time independent aggregate repeat purchase behavior of customers - will use the RCP_{A_i} to rank a customer's repeat purchase recommendations. It will thus always rank *paper towels* higher than the *laundry detergent* in the customer's repeat purchase recommendations.

On the other hand, the ATD model - which uses the aggregate distribution of repeat purchase time intervals across all its repeat purchasing customers - will take into account the fact that at an aggregate level customers repeat purchase the *paper towels* every 2 months and the *laundry detergent* every 3 months. It will also take into account that the customer most recently purchased the *paper towels* 1 week ago and the *laundry detergent* 3 months ago. It will thus rank the *laundry detergent* higher than the *paper towels* in the customer's repeat purchase recommendations. It should be noted that this behavior that uses the time signal is desired not only because it makes intuitive sense, but also because the RCP and ATD models are built with the stated goal of generating standalone repeat purchase recommendations. In this scenario, ignoring the time signal and recommending a customer the same product that they repeat purchased recently will result in a sub-optimal customer experience.

Thirdly, the PG model - which uses the customer's own posterior purchase rate for products - will compute the posterior purchase rates for *paper towels* and *laundry detergent*. Given the illustrative data above, the posterior purchase rate estimated by the PG model for *paper towels* is $\approx 5/14 = 0.36$ and that of *laundry detergent* is $\approx 4/15 = 0.27$. It will thus rank the *paper towels* higher than the *laundry detergent* in the customer's repeat purchase recommendations. Similar to the RCP model, this is will be the case even though the customer last repurchased the *paper towels* 1 week ago, but has not repurchased the *laundry detergent* in the last 3 months. However, in comparison to both the RCP and ATD models, the PG model uses the customer's own personalized repeat behavior and thus its predictions will be more accurate when a longer time frame is considered.

Finally, the MPG model - which takes into account the customer's own posterior purchase rate after adjusting for the last purchase time and the product's RCP_{A_i} - will compute the posterior purchase rates for *paper towels* and *laundry detergent*. Given the illustrative data above, the posterior purchase rate estimated by the MPG model for *paper towels* is $\approx 5/20 = 0.25$ and that of *laundry detergent* is $\approx 4/14 = 0.29$. It will also take into account the fact that $\text{RCP}_{\textit{paper towels}} \approx \text{RCP}_{\textit{laundry detergent}}$ and will thus rank the *laundry detergent* higher than the *paper towels* in the customer's repeat purchase recommendations. It is worth noting again that this behavior that uses both the time-dependent and time-independent signal is desired, especially when generating standalone repeat purchase recommendations, because ignoring the time signal and recommending a customer the same product that they repeat purchased recently will result in a sub-optimal customer experience.

5 OFFLINE EXPERIMENTS

In Section 4, we presented an analytical comparison of the different repeat purchase recommendations models - RCP model, ATD model, PG model, and MPG model - presented in Section 3. In this section, we present an empirical comparison of the same models based on the experiments we conducted to evaluate the quality of these models.

5.1 Data

We used a dataset containing customer purchases on Amazon as our training and test data. We held out one week of most recent customer purchases from this dataset for testing and used the last y years of purchases made prior to this week for training. A customer and their product purchase were considered as a repeat purchase in the test period (most recent one week) if and only if the customer purchased a product in the training period (y years before the test period) and also purchased the same product sometime in the test period.

Table 1: Lift in precision, recall, and nDCG for the ATD, PG, and MPG models at rank $m = 3$ as compared to the baseline RCP model

Models	Precision (%Lift)	Recall (%Lift)	NDCG (%Lift)
ATD	+8.22%	+5.56%	+5.70%
PG	+32.34%	+29.04%	+37.11%
MPG	**+36.51%**	**+30.33%**	**+39.42%**

5.2 Metrics

We used the industry standard evaluation metrics of precision and recall to evaluate the quality of our recommendations: precision measures the fraction of recommendations that are relevant and recall measures the fraction of relevant recommendations that have been retrieved. For recommendations, we look at recommendations at a specific rank m, and measure precision at rank m and also recall at rank m.

Precision and recall are set based evaluation metrics. But in the context of recommendations, ranking is also important. So we also use a third industry standard metric called Normalized Discounted Cumulative Gain (nDCG) to evaluate the quality of our recommendations. Specifically, nDCG measures the quality of ranking at a specific rank m.

5.3 Results

We trained each of our models offline using the training data mentioned above. We then computed the precision, recall, and nDCG at various ranks using the test data and computed the lift in these scores as compared to our baseline, i.e., RCP model. Table 1 shows the lift in precision, recall, and nDCG metrics at rank $m = 3$ for each of our models compared to the baseline RCP model. Additionally, Figure 4, Figure 5, and Figure 6 show the nature of graphs for precision, recall, and nDCG respectively for the RCP, ATD, PG, and MPG models at ranks $m = 1$ to 5.

These results show that, in terms offline metrics, all the our subsequent models are better than the baseline RCP. They also show that the RCP and ATD models are much worse than both the both the PG and MPG models. It should be noted here that both the PG and MPG models are Bayesian in nature which enables them to use a customer's personalized purchase signals in addition to the aggregate purchase behavior. This is likely the reason why they are superior to the RCP and ATD models which only use the aggregate purchase behavior. The results also demonstrate that the MPG model is best overall. One possible reason for the superiority of the MPG model is that in addition to the personalized signal, it combines the both time and time-independent signals for generating recommendations while also ensuring that its time-model is adapted for the recommendations use case.

6 ONLINE EXPERIMENTS

While offline metrics are informative, the true test of a recommendations model is online, where we can measure its impact on customer behavior in a production setting. To measure the value of repeat purchase recommendations in production setting, we built a

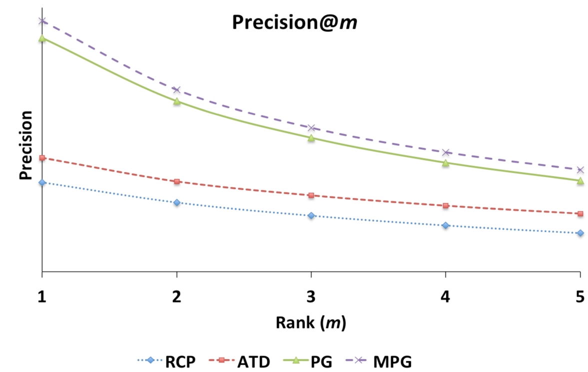

Figure 4: Precision for the RCP, ATD, PG, and MPG models at ranks $m = 1$ to 5

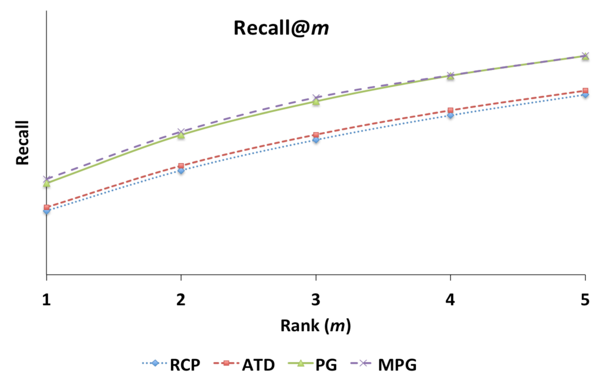

Figure 5: Recall for the RCP, ATD, PG, and MPG models at ranks $m = 1$ to 5

new feature called *Buy It Again* (BIA) that displays the standalone repeat purchase recommendations generated by our models. Figure 7 shows the BIA recommendations feature for an anonymous random customer on the *personalized recommendations page* of the Amazon.com website. We then tested the BIA feature and our models for generating BIA recommendations on the Amazon.com website through online A/B tests that were run for 14 days each. During the A/B tests, the traffic was divided randomly and equally between *control* and *treatment* resulting in tens of thousands of impressions per day for each and the models were updated once every day after midnight.

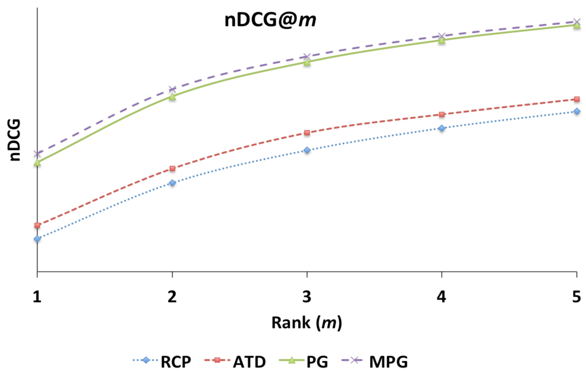

Figure 6: NDCG for the RCP, ATD, PG, and MPG models at ranks $m = 1$ **to** 5

6.1 Metrics

As noted previously, the success of recommendations on an e-commerce website is measured through an online A/B test. The goal of an A/B test is to measure the difference in customer engagement between the *control* and *treatment* sets and use that as the success-metric for the A/B test. One commonly used customer engagement metric in A/B testing is the click through rate (CTR). Thus, in our A/B tests below, we used the lift in CTR for products shown on the *personalized recommendations page* of the Amazon.com website between the *control* and *treatment* sets as the success-metric for our experiments. We should note that while we only report CTR here, we also tracked other important e-commerce metrics (e.g., products purchased) in our A/B tests and found these metrics to be positively correlated with CTR.

6.2 Results

As noted previously, we tested the BIA recommendations feature and our models for generating BIA recommendations on the Amazon.com website through online A/B tests. We ran these A/B tests on the *personalized recommendations page* of the Amazon.com website. Here is the summary of the A/B tests that we ran:

(1) Our first A/B test introduced the BIA recommendations feature on the *personalized recommendations page* of the Amazon.com website. In this A/B test, the recommendations in the BIA recommendations feature were generated by the ATD model. Since this A/B is akin to introducing the BIA recommendations feature on *personalized recommendations page* of the Amazon.com website, in this test, *control* did not have the BIA recommendations feature and the *treatment* showed the BIA recommendations feature. It should be noted that this is a unique but a valid A/B test, in which the *treatment* measures the value that BIA recommendations add while displacing other recommendations.

Table 2: Lift in CTR in *treatment* compared to the corresponding *control* in the online A/B tests on the *personalized recommendations page* of the Amazon.com website

Control	Treatment	CTR (%Lift)	p-value
Existing Recommendations	BIA Recommendations using ATD model	+7.1%	0.001
BIA Recommendations using ATD model	**BIA Recommendations using MPG model**	**+1.3%**	**0.015**

(2) We followed up our first A/B test with a second one. In this A/B test, we compared the relative quality of BIA recommendations generated by the ATD and MPG models on the *personalized recommendations page* of the Amazon.com website. For this A/B test, *control* showed the BIA recommendations feature using the recommendations generated by the ATD model and *treatment* showed the BIA recommendations feature using the recommendations generated by the MPG model.

We did not test the RCP and PG models online due to the negative customer feedback we received from internal beta customers about these model recommending products too early.

Table 2 summarizes the results of the online A/B tests. The table shows that the first A/B test, which introduced the BIA recommendations feature on the *personalized recommendations page* of the Amazon.com website using the ATD model, resulted in a 7.1% increase in the CTR for products on that page in *treatment*. Aside from the fact that this increase is statistically significant ($p < 0.01$), a 7.1% increase in CTR is considered a large difference for a website like Amazon.com where both *control* and *treatment* get tens of thousands impressions per day. This A/B test shows the relative value the BIA recommendations feature in comparison to the other recommendations on the *personalized recommendations page* of the Amazon.com website. It should be noted that during this A/B test, the BIA recommendations feature had the same customer experience as the other recommendations features on that page and displaying BIA recommendations often resulted in another recommendation feature not being shown on the page.

Table 2 also shows that the second A/B test, which compared the relative quality of BIA recommendations generated by the ATD and MPG models, results in a 1.3% increase in CTR for products on the *personalized recommendations page* of the Amazon.com website in *treatment*. This increase is also statistically significant ($p < 0.05$) and confirms the result from the offline experiments that the MPG model is significantly better than the ATD model.

7 CONCLUSION AND FUTURE WORK

Repeat purchase recommendations is an important area for e-commerce that has been relatively unexplored. Our current experiments with various repeat purchase recommendations models have shown positive results both in offline and online settings with the online experiments demonstrating over 7% increase in click through rate

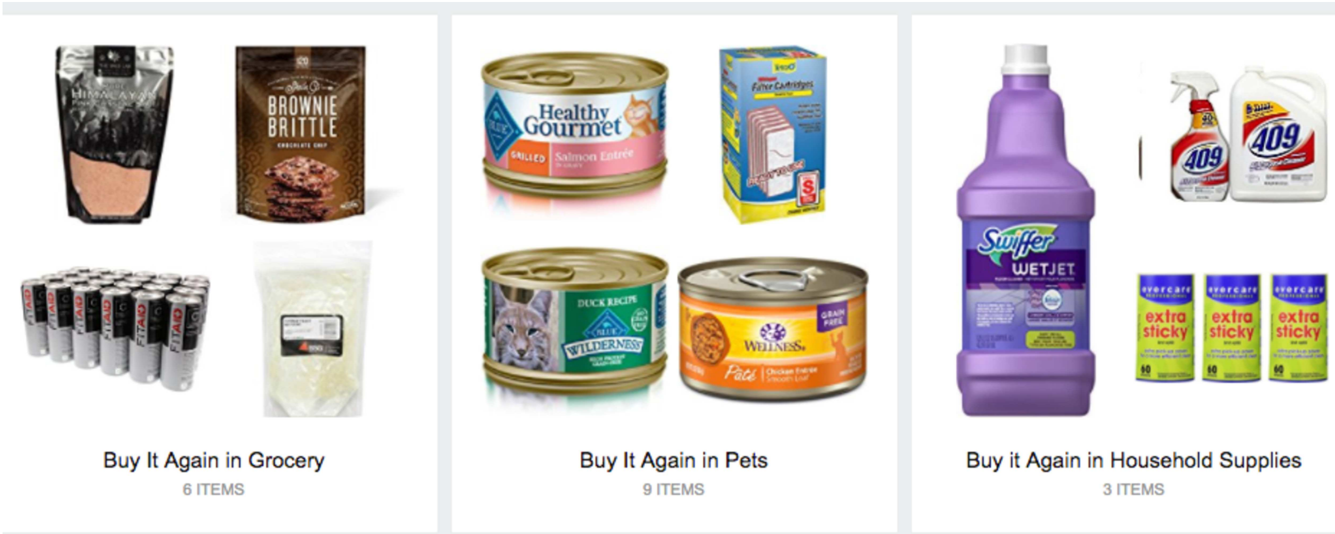

Figure 7: Buy It Again recommendations feature on the *personalized recommendations page* of the Amazon.com website

for products on the Amazon.com personalized recommendations page. Since the success of these experiments, the Buy it Again recommendations feature has been A/B tested successfully on various pages on the Amazon.com website and the Amazon mobile app including the Amazon home page. The feature has also been successfully A/B tested and launched on Amazon websites and Amazon mobile app in different countries (e.g., Germany, India, Japan, etc.).

In future, we plan to experiment with various other models. Firstly, we are planning to explore some recent models that are inspired by the NBD model [6], but which model some additional useful factors, like the BG-NBD model [8]. Secondly, we are exploring the use of customer behavior data, collected from the customer interactions with our widgets, to build supervised learning models like Logistic Regression [11] to help in improving the quality of our recommendations. These models make it easy to incorporate additional customer and product features that might help in improving both candidate selection and ranking for repeat purchase recommendations. Finally, we are also exploring the use of matrix factorization [11] and neural networks [11] for generating repeat purchase recommendations.

ACKNOWLEDGMENTS

We want to thank Srikanth Thirumalai, Alex Rosalez, Madhu Kurup, Jody Biggs, and Ron Whitman for supporting this work. We also want to thank Prateek Kotak, Vijai Mohan, John Lindsey, Quamrul Tipu, Jim Chan, and Murtaza Halai for their help and feedback.

REFERENCES

[1] G. Casella. 1985. *An Introduction to Empirical Bayes Data Analysis.* Vol. 39. The American Statistician 39(2): 83-87, 1985.

[2] Chattfield and Goodhardt. 1973. *A Consumer Purchasing Model with Erlang Inter-Purchase Times.* Vol. 68. Journal of the American Statistical Association.

[3] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16).* ACM, New York, NY, USA, 191–198. https://doi.org/10.1145/2959100.2959190

[4] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, and Dasarathi Sampath. 2010. The YouTube Video Recommendation System. In *Proceedings of the Fourth ACM Conference on Recommender Systems (RecSys '10).* ACM, New York, NY, USA, 293–296. https://doi.org/10.1145/1864708.1864770

[5] Suvodip Dey, Pabitra Mitra, and Kratika Gupta. 2016. Recommending Repeat Purchases Using Product Segment Statistics. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16).* ACM, New York, NY, USA, 357–360. https://doi.org/10.1145/2959100.2959145

[6] A.S.C Ehrenberg. 1959. *The Pattern of Consumer Purchases.* Vol. 8. Journal of Royal Statistical Society, Series C (Applied Statistics).

[7] P.S. Fader and B.G.S. Hardie. 2009. *Probability Models for Customer-Base Analysis.* Vol. 23. Journal of Interactive Marketing.

[8] P.S. Fader, B.G.S Hardie, and K.L.Lee. 2005. *Counting Your Customers the Easy Way: An Alternative to the Pereto/NBD model.* Vol. 24. Marketing Science.

[9] Carlos A. Gomez-Uribe and Neil Hunt. 2015. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Trans. Manage. Inf. Syst.* 6, 4, Article 13 (Dec. 2015), 19 pages. https://doi.org/10.1145/2843948

[10] G.L. Graham. 1969. *NBD Model of Repeat-Purchase Loyalty: An Empirical Investigation.* Vol. 6. Journal of Marketing Research.

[11] T. Hastie, R. Tibshirani, and J. Freidman. 2011. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Series in Statistics.

[12] D. Hoppe and U. Wagner. 2007. *Customer Base Analysis: The case of Central Variant of the Betageometric/NBD Model.* Vol. 3. Marketing – Journal of Research and Management.

[13] Komal Kapoor, Karthik Subbian, Jaideep Srivastava, and Paul Schrater. 2015. Just in Time Recommendations: Modeling the Dynamics of Boredom in Activity Streams. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15).* ACM, New York, NY, USA, 233–242. https://doi.org/10.1145/2684822.2685306

[14] Komal Kapoor, Mingxuan Sun, Jaideep Srivastava, and Tao Ye. 2014. A Hazard Based Approach to User Return Time Prediction. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14).* ACM, New York, NY, USA, 1719–1728. https://doi.org/10.1145/2623330.2623348

[15] Greg Linden, Brent Smith, and Jeremy York. 2003. Amazon.Com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing* 7, 1 (Jan. 2003), 76–80. https://doi.org/10.1109/MIC.2003.1167344

[16] D.C. Schmittlein, D.G. Morrison, and R. Colombo. 1987. *Counting your customers: Who are they and what will they do next.* Vol. 3. Management Science.

[17] Xiaoyuan Su and Taghi M. Khoshgoftaar. 2009. A Survey of Collaborative Filtering Techniques. *Adv. in Artif. Intell.* 2009, Article 4 (Jan. 2009), 1 pages. https://doi.org/10.1155/2009/421425

[18] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. 2015. Collaborative Deep Learning for Recommender Systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15).* ACM, New York, NY, USA, 1235–1244. https://doi.org/10.1145/2783258.2783273