Learning Within-Session Budgets from Browsing Trajectories

Diane Hu, Raphael Louca, Liangjie Hong
Etsy, Inc.
Brooklyn, NY
{dhu,rlouca,lhong}@etsy.com

Julian McAuley
UC San Diego
San Diego, CA
jmcauley@ucsd.edu

ABSTRACT

Building price- and budget-aware recommender systems is critical in settings where one wishes to produce recommendations that balance users' preferences (what they like) with a model of purchase likelihood (what they will buy). A trivial solution consists of learning *global budget* terms for each user based on their past expenditure. To more accurately model user budgets, we also consider a user's *within-session budget*, which may deviate from their global budget depending on their shopping context. In this paper, we find that users implicitly reveal their session-specific budgets through the sequence of items they browse within that session. Specifically, we find that some users "browse down," by purchasing the cheapest item among alternatives under consideration, others "browse up" (selecting the most expensive), and others ultimately purchase items around the middle. Surprisingly, this mixture of behaviors is difficult to observe globally, as individual users tend to belong firmly to one of the three segments. To model this behavior, we develop an interpretable budget model that combines a clustering component to detect different user segments, with a model of segment-specific purchase profiles. We apply our model on a dataset of browsing and purchasing sessions from Etsy, a large e-commerce website focused on handmade and vintage goods, where it outperforms strong baselines and existing production systems.

CCS CONCEPTS

•Information systems → Content ranking; Personalization;

ACM Reference format:
Diane Hu, Raphael Louca, Liangjie Hong and Julian McAuley. 2018. Learning Within-Session Budgets from Browsing Trajectories. In *Proceedings of Twelfth ACM Conference on Recommender Systems, Vancouver, BC, Canada, October 2–7, 2018 (RecSys '18),* 5 pages.
DOI: 10.1145/3240323.3240401

1 INTRODUCTION

In modeling users' behavior and activities, recommender systems have to trade-off a variety of complex and competing objectives. What does the user like, and what are they currently interested in (preferences and context)? What can they afford and how profitable is the transaction (budgets and profit)? Several of these questions have been considered from different angles within recommender systems research. For example, user budgets have been considered from many angles, from budgets and price sensitivity, to dynamic/algorithmic pricing. The goals of these works are often at odds with each other: Recommending budget-compatible items may compromise profitability; recommending budget *in*compatible items may decrease user engagement and satisfaction.

In this paper, we aim to study a few of these questions holistically through the lens of session-based recommendation. We use production data from Etsy, a large online retailer of handmade and vintage items, many of which are "one-of-a-kind") and naturally exhibit a high degree of variance with regard to price. This means users are often uncertain of their "willing-to-pay" price, and determine this quantity gradually as they navigate alternatives.

Thus, our goal is to develop a session-based model for budget-aware recommendation, which tracks the trajectory of users' within-session browsing choices in order to determine their budget targets. In particular, we find that users' click sequences are highly indicative of their (local/session-based) willing-to-pay prices. Naïvely, one might hypothesize a few simple models for this dynamic, e.g. (1) users browse items in the vicinity of the amount they are willing to pay; (2) users select the cheapest item that matches their preferences; or (3) users select the most expensive item among those they can afford. In fact, we find that none of these specific behaviors predominates, but that users follow a mixture of such patterns. The surprising property we exploit (which leads to significant performance gains) is that users who follow one strategy in their browsing behavior continue to do so in subsequent visits to the site (e.g. users who select the cheapest item that matches their preferences tend to follow the same behavior for future purchases). We evaluate our system against alternative methods, including Etsy's live production algorithm, where our proposed method is shown to improve purchase prediction accuracy.

2 RELATED WORK

Budget Estimation, Promotions, and Dynamic Pricing. One classical notion that has been considered in e-commerce is the idea of *price sensitivity*, which has been mentioned as a potential direction in a classic survey [18], and more recently has been applied to real-world settings like online promotions and recommendation of grocery transactions [8, 9, 20, 21]. The above works are related to ours in their incorporation of price and budget features, but are generally concerned with customizing promotion strategies rather than learning the concept of a user budget as we do here.

Other models try to estimate users' budgets or incorporate price-related features for the purpose of increasing purchase rates [4, 5], which is a similar goal of our proposed model. For example, Chen *et al.* [1] investigated various price-related features in order to model the outcome of Amazon's 'Buy Box' selection algorithm, and Zhao

RecSys '18, Vancouver, BC, Canada

DOI: 10.1145/3240323.3240401

et al. [22] developed a laboratory auction setup to collect user's WTP (willing-to-pay) data. Blei *et al.* [3] proposed a sequential model of market baskets that includes interpretable components that model purchase behavior, including item preferences, popularity, seasonality, and price. The main differentiation of our model is our contribution of learning *session-based* budget models that are revealed by browsing trajectories.

Session-Based Recommendation. Session-based recommendation models seek to model activities within and across sessions, as opposed to learning long-term *global* preference models from historical data. A few recent approaches include [19] and [23]. The goals of these papers are similar in that they look at 'micro-behavior' revealed by users' actions within a short time period, but different from ours in that they are not concerned with budget dynamics. Beyond session-based models, there are several approaches that combine both implicit and explicit data (e.g. click data vs. purchase data) as we do here [6, 7, 11–14]. However, these works are also generally not concerned with price or budget dynamics.

3 MODEL

The goal of an e-commerce recommender system is to provide a personalized ranking of items to users in a way that may increase the purchase rate and profitably for the site. To do so, we make use of the popular Bayesian Personalized Ranking (BPR) framework [17] for optimizing a user-specific ranking order where higher ranked items have higher purchase probabilities compared to lower ones. In the BPR framework, we construct a training set of triples x_{uij}, containing an item i that was purchased by user u and an item j that was not. The optimization criteria for BPR tries to maximize the difference between preferred (purchased) items and non-preferred items via the following objective,

$$\sum_{(u,i,j)} \ln \sigma(x_{uij}) - \lambda_\Theta \|\Theta\|^2, \quad \text{where } x_{uij} = x_{ui} - x_{uj}. \tag{1}$$

Here, Θ is a parameter vector, λ_Θ is a regularization coefficient, and σ is the sigmoid function. This formulation merely requires that we define a scoring function x_{ui} which indicates user u's preference for purchasing item i (i.e., the higher x_{ui}, the more likely user u will be to purchase item i). Before proposing our own solution, we first consider simple variants of x_{ui} that form our model's individual components and also act as baselines for comparison.

3.1 User Preference

We build our model on top of the classical BPR+MF scoring function, which models a user's preference toward an item as

$$x^{(1)}_{ui} = \beta_u + \beta_i + \langle \gamma_u, \gamma_i \rangle. \tag{2}$$

Here, the preference is modeled as an inner product between k-dimensional latent vectors γ_u and γ_i, which encode how well user u's latent preferences are aligned with item i's attributes. The user bias term, β_u, captures each user's tendency to purchase, while the item bias β_i can model each item's overall purchase popularity.

3.2 Global Budget Preference

The price of an item and how closely it matches a user's budget presumably play a significant role in determining whether a user will eventually purchase. A simple model to capture this phenomenon consists of defining a *global budget* preference, indicating how much (in some unit of currency) each user is willing to spend on a typical item. This is modeled as a matching function between the user's global budget, b_u, and the cost of the item c_i they are considering:

$$x^{(2)}_{ui} = \exp(-\omega(b_u - c_i)^2). \tag{3}$$

The goodness of match is represented by the difference between the user's budget and the cost of the item, scaled by a coefficient ω and passed through an exponential function. The coefficient ω is a learned, scalar variable that controls the 'steepness' of the penalty. Roughly speaking, the above formulation assumes that users' expenditures are approximately normally distributed where b_u and ω control the mean and variance, respectively. We also considered asymmetric matching functions where a larger penalty was given when the cost of the item was larger than the user's global budget; however this did not significantly improve the model.

3.3 Local Budget Preference

A user may have a general, global budget based on past shopping experiences, but a user's budget may deviate from session to session, depending on the context of their shopping intent. In our setting where goods are often unique, a user may not have a sense of what they are willing to pay until they have browsed through several candidate items and observed their price ranges.

To incorporate this purchase decision-making behavior into our model, we take into account the prices of all items viewed thus far in a user's shopping session that fall in the same category as the purchased item (as to only consider competitive alternatives). We then seek to learn an in-session, or "local" budget preference term that relates the price of the considered item to the prices of all items seen thus far:

$$x^{(3)}_{ui} = \sum_{q=1}^{Q} \psi_{uq} \langle X_q, \rho_i \rangle. \tag{4}$$

Here, we learn a simple, factorized representation where X should uncover Q distinct patterns (or *purchase profiles*) of how users respond to the prices of seen items within a session (e.g. by purchasing the cheapest, or most expensive item, for example), and ψ represents a user-specific membership to each of these patterns/profiles.

More specifically, let c_i denote the cost of the target item i, and P the number of discretized price buckets to be considered. The cost of previously seen items, relative to item i are encoded in ρ_i as a $P-$dimensional, one-hot-encoded vector that indicates roughly which percentile c_i falls into relative to the prices of the previously viewed items that belong to the same category as item i. For example, if $P = 10$, and c_i falls into the first decile of all previously viewed items, then the first element of the vector will equal to one. The learned patterns of purchase behavior are represented by X, a $(P \times Q)$ matrix that uncovers Q different within-session *purchase profiles* of purchase behavior relative to prices of previously clicked listings. Finally, ψ_u is a Q-dimensional vector that learns user u's preference for each of the Q profiles. To make the vector probabilistically interpretable, it is passed through a softmax function such that all elements in a user's vector add up to 1: $\psi_{u,q} = \frac{\exp \psi'_{uq}}{\sum_{q'} \exp \psi'_{uq'}}$.

This vector indicates the amount of membership each user has in belonging to each of the Q latent purchase profiles.

3.4 Model Learning and Optimization

Each component of the scoring function described previously can be learned individually or jointly, depending on the application. For joint modeling, components may be combined in many ways. For our application, we chose to jointly model components using an additive, linear combination because of its computational simplicity and ease of interpretation. The combined form of our scoring function is as follows:

$$x_{ui} = \beta_u + \beta_i + \langle \gamma_u, \gamma_i \rangle + \alpha_u \exp\left(-\omega\,(b_u - c_i)^2\right) + \phi_u \sum_{q=1}^{Q} \psi_{uq} \langle X_q, \rho_i \rangle. \tag{5}$$

Note that we introduce additional importance weights, α and ϕ, which learn how important the local and global (respectively) budget terms are for each user. The above scoring function fits naturally into the BPR framework and is optimized via Stochastic Gradient Descent (on Eq. 1). Our system is implemented in *Tensorflow* using the Adagrad Optimizer. We describe our procedure for generating negative samples (x_{uj}) in the following section.

4 EXPERIMENTS

In this section, we describe implementation details as well as evaluation methods for quantifying the effectiveness of our method.

4.1 Dataset

Our dataset contains raw transaction data collected from Etsy spanning over 4 months. In order to gain insight into purchase predictability, we focus our efforts on signed-in users that have made at least 3 purchases in the last 4 months, and items that have been purchased at least twice. We also eliminate all sessions that do not include at least one purchase and hold out the last purchase from each user for purchase prediction evaluation.

There are several different places where personalized recommendations are offered on Etsy. To collect the triplet data, we treat each item i purchased through a recommendations module as a positive example while the top ranked, non-purchased item, j, in the same module is treated as the negative sample. This ensures that the positive item was explicitly preferred over the negative one by the user. Basic dataset statistics are as follows:

#Users	#Items	#Sessions	#Train	#Test	#Purchases
6,362	16,136	8,417	20,400	2,390	11,395

4.2 Evaluation Methodology

We aim to evaluate the effectiveness of the proposed model based on its purchase predictability and ease of interpretation. To quantify purchase predictability, we use the popular AUC (Area under the ROC curve) metric:

$$AUC = \frac{1}{N} \sum_u \frac{1}{\mid E(u) \mid} \sum_{(i,j) \in E(u)} \delta(x_{ui} > x_{uj}), \tag{6}$$

where $E(u)$ denotes the set of all triplets x_{uij} that are to be evaluated for user u. To show that the proposed method improves upon existing methods, we denote our method as **BPR-BUDGET** (Eq. 5) and compare its AUC to the AUC of individual components of our proposed method as well as other competitive baselines:

- **BPR-BUDGET-LOCAL**: The proposed budget model including the latent terms (Eq. 2) and the local term (Eq. 4) only
- **BPR-BUDGET-GLOBAL**: The proposed budget model including the latent terms (Eq. 2) and the global term (Eq. 3) only
- **MF**: Traditional Matrix Factorization using Gradient Descent [10]
- **BPR-MF**: Matrix Factorization with BPR ranked loss [17]
- **FM**: Factorization machines with binary loss [16]
- **LOG-REGRESSION**: Logistic Regression, predicting binary purchase outcome using features discussed below [15]
- **GBDT**: Gradient Boosted Decision Trees predicting binary purchase outcome using features discussed below [2]
- **PRODUCTION**: The model used on Etsy's user recommendation modules.

For feature-based baseline models, we use 3 features that are equivalent to the content features used in the proposed model: (1) the cost of item i, (2) the item's "seen item vector" (equivalent to ρ), and (3) the historical, average purchase price of each user.

Table 1: AUC results for the proposed model (and its ablation variants) compared other baselines

Model	AUC
BPR-BUDGET	0.770
BPR-BUDGET-LOCAL	0.748
BPR-BUDGET-GLOBAL	0.750
MF	0.687
BPR-MF	0.708
FM	0.694
LOGISTIC-REGRESSION	0.644
GBDT	0.713
PRODUCTION	0.704

4.3 Experimental Results

All AUCs obtained on the test dataset using the methods discussed in the previous section are reported in Table 1. From the results, all three proposed variations outperform existing baselines, with the combined scoring function (BPR-BUDGET) achieving the highest purchase prediction accuracy. Most hyper-parameters were chosen based on cross-validation. In particular, we found that setting $Q = 3$ and $P = 4$ led to high purchase predictability and interpretability as shown in Figure 2.

As discussed, one of the main contributions of our model, in addition to purchase predictability, is its ease of interpretation. Here, we visualize some of the representations learned by our model that give insight as to what user budgets are and how that impacts different categories across the site. In Figure 1, we look at how the learned user budgets are distributed across top-level categories by computing the average budget for all users who purchased each category. Users with the largest budgets purchased most from "Weddings" and "Purses & Handbags," while users with the smallest budgets purchased most in "Craft Supplies" and "Paper/Party Supplies." This

Table 2: Detailed statistics from users assigned to each of $Q = 3$ purchase profiles, labeled with segment interpretation.

Stat	"Cheap"	"Mixed"	"Expensive"
Avg Purchase Count	2.84	2.53	2.37
Avg Purchase Price	\$12.27	\$12.40	\$17.32
Med Purchase Price	\$5.65	\$6.75	\$9.00

result is consistent with average purchase prices in each category, as well as notions of 'luxury' vs. 'utility' goods.

We also give insight as to how users behave according to their purchase profile membership. Recall from Section 3.3 that this is a Q-dimensional vector where the i^{th} element indicates the degree to which the user exhibits behavior from the i^{th} purchase profile. When $Q = 3$ (Figure 2), this results in profiles that roughly correspond to a "cheap," "mixed," and "expensive" segments. For the sake of visualization, we assign each user to the profile that they had the highest probability for and report some brief statistics related to the items purchased by users assigned to each of the 3 profiles in Table 2. Here, we see that users who belong to the more "expensive" segment tend to spend more on individual items, but purchase fewer things, while users in the "cheaper" segment spend significantly less on individual items, but purchase slightly more items. Relatedly, Figure 3 shows a histogram of purchased item costs in intervals of \$5, for users in each purchase group. The plot shows that between the \$0 − \$10 range, there are more purchases by users in the cheaper purchase segment, but for items that are \$15 and beyond, the more expensive segment tends to dominate.

Finally, in Figure 4 we analyze which top-level categories users from each purchase segment most frequently purchase from. To highlight the differences, we focus only on the cheap and expensive groups and show how each group's normalized purchase frequency differs positively or negatively from the average. From this graph, we can conclude that users from the expensive group purchased (relatively) more from "Home & Living," "Accessories," and "Toys and Games." On the other hand, users from the cheap group purchased significantly more from "Craft Supplies & Tools," "Art & Collectibles," and "Paper & Party Supplies."

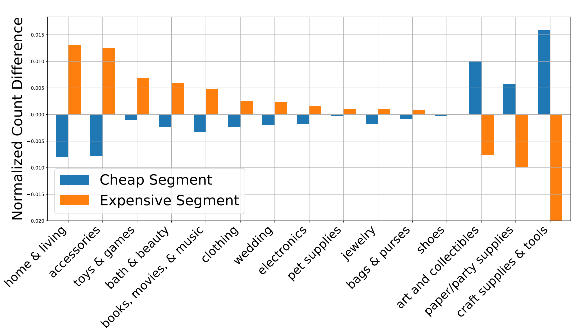

Figure 4: Differences in purchase frequency between the "cheaper" and "expensive" purchase segments, by top-level category.

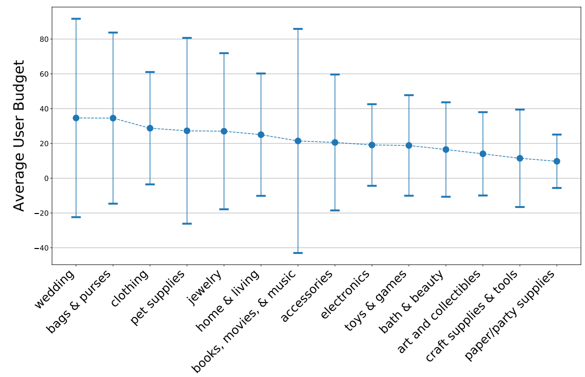

Figure 1: Average learned budgets (with standard deviation) for users who purchased in each top-level category, sorted in descending order from left to right.

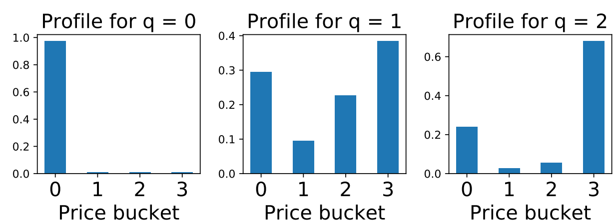

Figure 2: Purchase profiles learned with $P = 4$, and $Q = 3$ give fairly interpretable results where, from left to right, the profiles correspond roughly to users who purchase (a) the cheapest, (b) a mix, or (c) the most expensive item out of what they've seen so far.

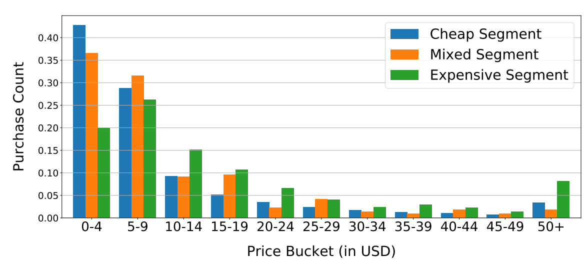

Figure 3: Histogram of purchased item costs from users assigned to each of the 3 purchase profiles.

5 CONCLUSION

In this paper, we proposed an interpretable model for session-based recommendation that accounts for the observation that users' short-term, within-session browsing behavior indirectly reveals their (within-session) willing-to-pay targets. Accounting for these dynamics substantially increases the purchase prediction accuracy. The learned representations from the model also gives insight as to what user budgets are and how this may affect their spending across different categories. We apply our model to a large production dataset of browsing and purchase data, where it outperforms other budget-aware methods as well as production systems.

REFERENCES

[1] Le Chen, Alan Mislove, and Christo Wilson. 2016. An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace. In *WWW*.

[2] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *KDD*.

[3] David M. Blei Francisco J. R. Ruiz, Susan Athey. 2017. SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements. In *Arxiv*.

[4] Yong Ge, Qi Liu, Hui Xiong, Alexander Tuzhilin, and Jian Chen. 2011. Cost-aware Travel Tour Recommendation. In *KDD*.

[5] Yunhui Guo, Congfu Xu, Hanzhang Song, and Xin Wang. 2017. Understanding Users' Budgets for Recommendation with Hierarchical Poisson Factorization. In *IJCAI*.

[6] Gawesh Jawaheer, Martin Szomszor, and Patty Kostkova. 2010. Comparison of implicit and explicit feedback from an online music recommendation service. In *HetRec*.

[7] Gawesh Jawaheer, Peter Weller, and Patty Kostkova. 2014. Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback. *ACM Transactions on Interactive Intelligent Systems* (2014).

[8] Yuanchun Jiang and Yezheng Liu. 2012. Optimization of online promotion: a profit-maximizing model integrating price discount and product recommendation. *International Journal of Information Technology & Decision Making* 11, 05 (2012), 961–982.

[9] Yuanchun Jiang, Jennifer Shang, Yezheng Liu, and Jerrold May. 2015. Redesigning promotion strategy for e-commerce competitiveness through pricing and recommendation. *International Journal of Production Economics* 167 (2015), 257–270.

[10] Y. Koren, R. Bell, and C. Volinsky. 2009. Matrix factorization techniques for recommender systems. *IEEE Computer* (2009).

[11] Nathan N Liu, Evan W Xiang, Min Zhao, and Qiang Yang. 2010. Unifying explicit and implicit feedback for collaborative filtering. In *CIKM*.

[12] Weike Pan, Nathan N Liu, Evan W Xiang, and Qiang Yang. 2011. Transfer learning to predict missing ratings via heterogeneous user feedbacks. In *IJCAI*.

[13] Denis Parra and Xavier Amatriain. 2011. Walk the talk: Analyzing the Relation between Implicit and Explicit Feedback for Preference Elicitation. In *UMAP*.

[14] Denis Parra, Alexandros Karatzoglou, Xavier Amatriain, and Idil Yavuz. 2011. Implicit feedback recommendation via implicit-to-explicit ordinal logistic regression mapping. In *CARS*.

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[16] Steffen Rendle. 2012. Factorization machines with libfm. *TIST* 3, 3 (2012), 57.

[17] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *UAI*.

[18] J Ben Schafer, Joseph Konstan, and John Riedl. 1999. Recommender systems in e-commerce. In *EC*.

[19] Trinh Xuan Tuan and Tu Minh Phuong. 2017. 3D Convolutional Networks for Session-based Recommendation with Content Features. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. ACM, 138–146.

[20] Panniello Umberto. 2015. Developing a Price-Sensitive Recommender System to Improve Accuracy and Business Performance of Ecommerce Applications. *International Journal of Electronic Commerce Studies* 6, 1 (2015), 1.

[21] M. Wan, D. Wang, M. Goldman, M. Taddy, J. Rao, J. Liu, D. Lymberopoulos, and J. McAuley. 2017. Modeling consumer preferences and price sensitivities from large-scale grocery shopping transaction logs. In *World Wide Web*.

[22] Qi Zhao, Yi Zhang, Daniel Friedman, and Fangfang Tan. 2015. E-commerce Recommendation with Personalized Promotion. In *Recsys*.

[23] Meizi Zhou, Zhuoye Ding, Jiliang Tang, and Dawei Yin. 2018. Micro behaviors: A new perspective in e-commerce recommender systems. In *WSDM*. ACM.