

Joint Optimization of Profit and Relevance for Recommendation Systems in E-commerce*

Raphael Louca*, Moumita Bhattacharya*, Diane Hu, Liangjie Hong

Etsy, Inc.

New York, U.S.A

{rlouca,mbhattacharya,dhu,lhong}@etsy.com

ABSTRACT

Traditionally, recommender systems for e-commerce platforms are designed to optimize for relevance (e.g., purchase or click probability). Although such recommendations typically align with users' interests, they may not necessarily generate the highest profit for the platform. In this paper, we propose a novel revenue model which jointly optimizes both for probability of purchase and profit. The model is tested on a recommendation module at Etsy.com, a two-sided marketplace for buyers and sellers. Notably, optimizing for profit, in addition to purchase probability, benefits not only the platform but also the sellers. We show that the proposed model outperforms several baselines by increasing offline metrics associated with both relevance and profit.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

Revenue Optimization, Recommendation systems, e-commerce.

1 INTRODUCTION

In recent years, online e-commerce platforms such as Amazon, Ebay, and Etsy have seen tremendous growth. Unlike traditional brick and mortar stores, such platforms do not manufacture, store, or source products, rather they operate as a two-sided marketplace between buyers and sellers, facilitating a convenient and safe transaction process. In exchange, they collect a percentage of the transaction amount as a fee. Because of the large selection of products available, such platforms rely predominantly on recommendation systems to help users find items that appeal to their tastes and interests. Traditionally, these recommendation systems focus on optimizing for relevance by predicting the purchase or click probability of an item. This relevance-centric approach manifests itself in increased conversion rates. However, it does not explicitly maximize the profit generated for the platform, or the sellers. Thus, the question is, how

can we design recommendation systems that jointly optimize for relevance and profit?

Summary of Contributions: In this work, we propose a novel revenue model, which optimizes both probability of purchase and profit. Specifically, we show that the proposed model can make strategic recommendations by surfacing items that are both relevant to users and profit-maximizing for Etsy. To the best of our knowledge, no current studies have jointly optimized for both objectives, although a few have directly optimized just for profit [3, 4, 8]. In addition to several well-studied metrics, we propose two new metrics to evaluate the efficacy of the revenue model. Our results show that the model achieves statistically significantly higher (p -value < 0.05) performance compared to multiple baselines.

2 RELATED WORK

Because of the propensity to optimize for user relevance when designing recommendation systems, only a few works thus far have proposed methods that optimize the profit generated for the e-commerce platform by the recommendation system [2–4, 8]. One such study by Chen et al. [2] propose a simple profit-aware recommendation system, where candidate items are ranked in decreasing order of expected profit. The expected profit of a candidate item is computed by simply multiplying the probability of purchase of said item with its price. In our work, we observed that this approach tends to rank items according to decreasing order of price (cf. Section 4). In another study, Das et al. [3] propose an optimization problem, which maximizes the expected profit subject to constraints, which ensure that the similarity (as defined by the Dice or Jaccard measure) between the vector of ratings of recommended items and the user's true rating vector is less than a certain threshold. Essentially, the authors develop a model that maximizes the vendor's expected profit while maintaining a level of "trust" with the customer.

In a separate line of work, Lu et al. [4] propose a dynamic model that takes into account a variety of factors including prices, valuations, saturation effects, and competition amongst products to recommend items. Their work is orthogonal to ours as the model finds a recommendation strategy that maximizes the expected total revenue *over a given time horizon*. In all of these studies, however, it is assumed that the e-commerce platform has access to a model that optimizes for relevance and yields either a set of purchase probabilities or a 'true' rating vector for each user. Our work is different in that it proposes a model that jointly optimizes for both relevance and profit.

*Copyright 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Presented at the RMSE workshop held in conjunction with the 13th ACM Conference on Recommender Systems (RecSys), 2019, in Copenhagen, Denmark.

*These authors have equally contributed to this work.

3 METHODS

In this section, we propose a novel objective function that optimizes both the likelihood of purchase as well as profit. Below, we describe the revenue model, the baselines, and the evaluation metrics.

3.1 Revenue Model

Suppose that we are given training data collected from user sessions at Etsy.com. Each training instance $i = 1, \dots, m$ is described by feature vector $x_i \in \mathbf{R}^n$, and a label $y_i \in \{-1, 1\}$ indicating whether the corresponding recommended item has been purchased. We assume that each Bernoulli random variable y_i (random, that is, before we observe the results) can be modeled by a logistic regression model, where

$$\mathbf{prob}[y_i = 1 | x_i; w] = \sigma(w^\top x_i + b) = 1/(1 + \exp(-(w^\top x_i + b)))$$

and $\sigma : \mathbf{R}^n \rightarrow \mathbf{R}$ is the sigmoid function. Traditionally, the objective is to find a maximum likelihood estimate of the model parameters (w, b) , which requires solving the following convex optimization problem:

$$\underset{w \in \mathbf{R}^n, b \in \mathbf{R}}{\text{maximize}} \quad \ell(w, b) := - \sum_{i=1}^m \log(1 + \exp(-y_i(w^\top x_i + b))). \quad (1)$$

This objective function, however, does not explicitly maximize for profit. This naturally points in the direction of designing a custom objective function that yields parameters that trade-off between optimizing for probability of purchase and for profit. Let π_i denote the price of item i . The expected revenue generated by a set of m recommended items is given by

$$\begin{aligned} \rho(w, b) &:= \sum_{i=1}^m \mathbb{E}[\pi_i y_i] = \sum_{i=1}^m \pi_i (2\mathbf{prob}[y_i = 1 | x_i; w] - 1) \\ &= \sum_{i=1}^m 2\pi_i \sigma(w^\top x_i + b) - \pi_i, \end{aligned}$$

where the last equality follows by the fact that y_i is a Bernoulli random variable that takes values in $\{-1, 1\}$. This gives rise to the following optimization problem

$$\underset{w \in \mathbf{R}^n, b \in \mathbf{R}}{\text{maximize}} \quad \ell(w, b) + \mu \rho(w, b), \quad (2)$$

whose objective is to find parameters that fit the data (via $\ell(w, b)$) while maximizing the expected revenue (via $\rho(w, b)$). Here, $\mu \geq 0$ is a hyperparameter of the model that controls the tradeoff between the two objectives. Because this is a maximization problem and $\pi_i \geq 0$ for all i , the model in (2) will find parameters (w, b) that increase $\sigma(w^\top x_i + b)$ for higher-priced items i while ensuring that said parameters are able to explain the data. It is to be noted that the log-likelihood function $\ell(w, b)$ is concave in (w, b) and can therefore be maximized. The expected revenue term $\rho(w, b)$, however, is a weighted sum of sigmoid functions, which is known to be nonconvex [7]. Therefore, the solution to problem (2) is only guaranteed to be locally optimal. In our experiments, we use interior point methods [1] to obtain a solution for (2).

Once the optimal parameters are learned, we use them to rank a set of candidate items. In particular we consider two rankers, one

is based on probabilities and the other on the expected revenue as follows:

- (1) **Raw Ranker (RR)**: Ranks a set of K items according to increasing value of $x_i^\top w^* + b^*$, $i = 1, \dots, K$. Because the sigmoid function is an increasing function of (w, b) , the RR is equivalent to a probability ranker, which ranks items according to increasing value of $\sigma(x_i^\top w^* + b^*)$.
- (2) **Expected Revenue Ranker (ER)**: Ranks a set of K items according to increasing value of $\pi_i \cdot \sigma(x_i^\top w^* + b^*)$, $i = 1, \dots, K$.

3.2 Baselines

We compare the revenue model (2) with the following baselines:

- **Logistic Regression (LOR)**: Obtained from the revenue model by setting $\mu = 0$.
- **Weighted Logistic Regression (WLOR)**: A variant of LR, where purchased items are weighted by their price. In particular, let $\mathcal{P} = \{i \mid \text{item } i \text{ is purchased}\}$. WLR is formulated as

$$\begin{aligned} \underset{w \in \mathbf{R}^n, b \in \mathbf{R}}{\text{maximize}} \quad & - \sum_{i \in \mathcal{P}} \pi_i \log(1 + \exp(-y_i(w^\top x_i + b))) \\ & - \sum_{i \notin \mathcal{P}} \log(1 + \exp(-y_i(w^\top x_i + b))). \end{aligned}$$

- **Linear Regression (LIR)**: We consider a linear regression model where the label y_i of item i is equal to the profit generated by item i . More precisely, $y_i = \pi_i \mathbf{1}_{i \in \mathcal{P}}$. In linear regression, the optimal parameters are chosen to minimize the squared error between predictions and labels, i.e.,

$$\underset{w \in \mathbf{R}^n, b \in \mathbf{R}}{\text{minimize}} \quad \sum_{i=1}^m (x_i^\top w + b - y_i)^2.$$

The above optimization problem admits a closed form solution.

3.3 Evaluation Metrics

We use the following metrics to evaluate the performance of the proposed model and baselines. Let r be a ranking such that $r_1 \geq r_2 \geq \dots \geq r_K$. Also, let $\pi = [\pi_1, \dots, \pi_K]^\top$ be the vector of prices of items $1, \dots, K$.

- **Profit@k**: Given a position $k \in \mathbf{N}_K$ and a ranking r , the profit@k is defined as the profit generated by the k highest ranked items. More precisely, it is given by $\sum_{i=1}^k \pi_{r_i} \mathbf{1}_{r_i \in \mathcal{P}}$, where $\mathbf{1}$ is the indicator function and \mathcal{P} is the set of purchased items.
- **Average Price (AP) @k**: Given a position $k \in \mathbf{N}_K$ and a ranking r , the AP@k is defined to be equal to the average price of the k highest ranked items. It is given by $\sum_{i=1}^k \pi_{r_i}$.
- **Price-Based Normalized Discounted Cumulative Gain @k (P-NDCG) P-NDCG@k** is defined as NDCG@k [5], where the gain of item r_i is equal to $\pi_{r_i} / \|\pi\|$.
- **Area Under the Curve (AUC)**: AUC [6] is the area under the receiver operating characteristic curve, and it can be interpreted

Model	Ranker	μ	AUC	Metric / Metric@k								
				k = 1			k = 2			k = 3		
				P-NDCG	Profit	AP	P-NDCG	Profit	AP	P-NDCG	Profit	AP
LOR	RR	–	0.6050	0.6133	5.085	26.85	0.6411	8.6661	54.28	0.6688	11.6011	81.74
	ER	–	0.5399	0.9249	4.995	94.93	0.9383	8.6768	139.93	0.9462	11.5808	176.37
WLOOR	RR	–	0.6060	0.6125	5.2294	26.58	0.6406	8.8396	54.05	0.667	11.6070	81.19
	ER	–	0.5423	0.9192	5.1003	94.68	0.9352	8.8176	139.71	0.9433	11.6731	176.18
LIR	RR	–	0.5741	0.6191	4.9624	27.86	0.6474	8.2064	58.15	0.676	1087.91	86.36
Rev.	RR	1	0.6049	0.6194	5.1321	27.17	0.6479	8.7705	55.36	0.6754	11.7397	82.18
		1E2	0.6257	0.6527	5.5684	36.35	0.6828	9.4290	66.81	0.7077	12.3220	127.37
		1E4	0.6266	0.6510	5.5602	36.32	0.6804	9.3726	68.75	0.7054	12.2788	127.37

Table 1: Performance of evaluation metrics for the baseline and revenue models. The price-based NDCG (P-NDCG), profit, and average price (AP) metrics are shown for different values of k . For the revenue model, results are shown for three different values of the hyperparameter μ . For each metric, the highest performance level is shown in boldface.

as the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. We use AUC to measure relevance.

4 EXPERIMENTS AND DISCUSSION

In this section, we present offline experiments to evaluate the performance of the proposed revenue model (2). We use a training set consisting of implicit feedback data collected over a day from an item-to-item recommendation module that is placed on item pages at Etsy.com. An example of this module is shown in Figure 1. The training data is sampled so that 40% of the data are positive instances and the remaining are negative. Our feature set consists of item features (e.g., item purchase count) and cross features between the target and candidate items (e.g., tfidf similarity between the two items). We evaluate the model on next day’s data collected from the same recommendation module. It is to be noted that for the revenue model, we do not rank according to expected revenue (ER) because such ranking is meaningful only if the underlying model maximizes just for the likelihood of purchase (e.g., logistic regression model (1)).

In Figure 3, we plot the distribution of predictions returned by the optimal parameters (w, b) of problem (2) as function of μ . It is worth noting that for $\mu = 0$, the boxplot depicts the distribution of predictions for the logistic regression model. Compared to this distribution, we observe that for $\mu = 100, 10000$, the spread in the distribution of predictions induced by the optimal parameters increases while for $\mu = 1$ it decreases. The median (red line in boxplot) is observed to increase for all values of $\mu > 0$. This is expected since $\pi_i \geq 0$ and the revenue term is maximizing $\pi_i \text{prob}[y_i = 1 | x_i; w] = \pi_i \sigma(x_i^T w + b)$.

In Table 1, we observe that the proposed revenue model attains the highest AUC and profit@ k , among all other models, for all three values of k . In particular, we observe a 3.57% increase in AUC and 9.50% in profit@1 compared to the LR model using the raw ranker (LOR/RR). It is also worth noting that compared to LOR/RR, the

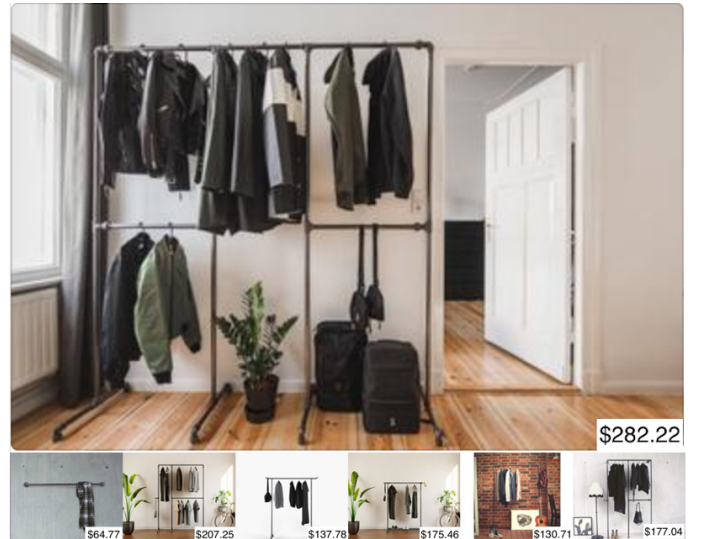


Figure 1: Target item (top) and a set of six recommended items (bottom).

proposed revenue model also increases P-NDCG@ k and AP@ k for all k by at least 3.57% and 23.08%. Therefore, the proposed model ranks relevant but high-priced items higher. Similar comparisons can be made between the revenue model and the WLOOR model using RR. In Table 1, we also observe that the LOR model using the expected revenue ranker (LOR/ER) attains the highest values for P-NDCG@ k and AP@ k . Thus, it favors higher-priced items. However, unlike our model, this model results in a 10.76% and 16.06% decrease in AUC compared to the LOR/RR model and the revenue model respectively.

The results shown in Table 1 are further supported by Figure 2, which shows the six candidate items ranked in the order that is generated by the LOR/RR (1st row), LOR/ER (2nd row), and revenue



Figure 2: Each row depicts the ranking of the candidate items generated by a given model. Specifically, the first and second rows show the ranking generated by the logistic regression model with the raw and expected revenue ranker, respectively. The third row shows the ranking generated by the revenue model with raw ranker and $\mu = 100$. At each row, the leftmost item is the highest ranked item. At each row, the blue dashed-line box denotes the purchased item.

model (3rd row). The item in the blue dashed-lined box is the item that was purchased. As shown in the figure, the revenue model is able to assign the highest rank to the purchased item while the LOR/RR and LOR/ER models rank that item last and second, respectively. It is worth noting that the revenue model ranks the second highest-priced item first, which is also the one being purchased. Therefore, our model generates a ranking that trades-off between optimizing for relevance and profit. We can also observe from the figure that the ranking obtained by the LOR/ER model sorts items

according to decreasing order of price, thus attaining the highest possible $AP@k$ and $P\text{-}NDCG@k$ for any k (i.e., $P\text{-}NDCG@k=1$, for all $k = 1, \dots, 6$). It is also straightforward to verify that in this example, the revenue model outperforms the LOR/RR model both in terms of $P\text{-}NDCG@k$ and $AP@k$ for all values of k .

5 CONCLUSION

In this preliminary study we propose a novel model that optimizes both profit and probability of purchase, while generating recommendations. We show that the recommendations produced by our model is able to increase profit for the platform while retaining high relevancy for users. In future work we plan to train our model on much larger datasets and assess its performance in the face of real user-traffic in Etsy.com by launching an online A/B experiment.

REFERENCES

- [1] Stephen Boyd and Lieven Vandenberghe. 2004. *Convex optimization*. Cambridge university press.
- [2] Long-Sheng Chen, Fei-Hao Hsu, Mu-Chen Chen, and Yuan-Chia Hsu. 2008. Developing recommender systems with the consideration of product profitability for sellers. *Information Sciences* 178, 4 (2008), 1032–1048.
- [3] Aparna Das, Claire Mathieu, and Daniel Ricketts. 2009. Maximizing profit using recommender systems. *arXiv preprint arXiv:0908.3633* (2009).
- [4] Wei Lu, Shanshan Chen, Keqian Li, and Laks VS Lakshmanan. 2014. Show me the money: dynamic recommendations for revenue maximization. *Proceedings of the VLDB Endowment* 7, 14 (2014), 1785–1796.
- [5] Christopher Manning, Prabhakar Raghavan, and Hinrich Schütze. 2010. Introduction to information retrieval. *Natural Language Engineering* 16, 1 (2010), 100–103.
- [6] Kevin P Murphy. 2012. *Machine learning: a probabilistic perspective*. MIT press.
- [7] Madeleine Udell and Stephen Boyd. [n. d.]. Maximizing a sum of sigmoids. ([n. d.]). http://www.stanford.edu/~boyd/papers/max_sum_sigmoids.html
- [8] Peng Ye, Julian Qian, Jieying Chen, Chen-hung Wu, Yitong Zhou, Spencer De Mars, Frank Yang, and Li Zhang. 2018. Customized Regression Model for Airbnb Dynamic Pricing. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 932–940.

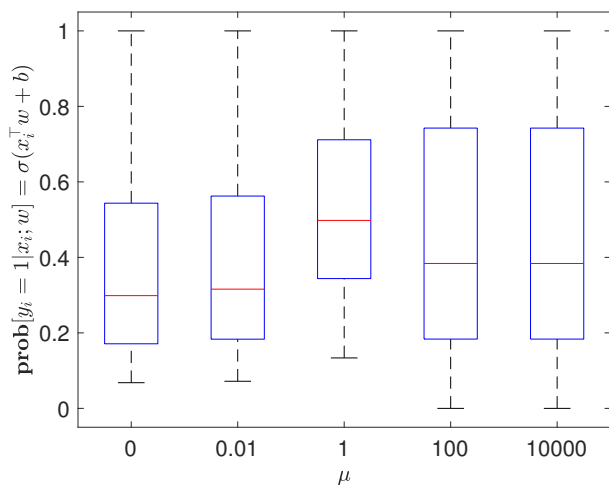


Figure 3: Distribution of predictions induced by the optimal parameters (w, b) of the revenue model (2) for different values of the hyperparameter μ .