Relevance Constrained Re-ranking in Sponsored Listing Recommendations

Zhen Ge
zhge@ebay.com
eBay Inc
New York, USA

Wei Zhou
wzhou6@ebay.com
eBay Inc
San Jose, USA

Jesse Lute
jlute@ebay.com
eBay Inc
New York, USA

Adam Ilardi
ailardi@ebay.com
eBay Inc
New York, USA

ABSTRACT

Advertising (Ad) revenue is a major revenue source for many technology and e-commerce companies; most of the revenue optimization research has been around third party display ads or Cost-Per-Click based first party ads. This paper discusses the Cost-Per-Action ad product at eBay; and the challenge of balancing ad revenue and relevance. We proposed a new measurement that uses Kullback–Leibler (KL) divergence to both optimize ad revenue and improve buyer experience in item recommendations. KL divergence is adopted in the re-ranking algorithm as a constraint for revenue optimization and it is solved by a greedy grid search algorithm. In addition, we are able to approximate KL divergence with inventory based features, and that simplified a full greedy search operation to a regression. Overall, we designed and A/B tested three different approaches, all of them showed significant improvement over the baseline. Through effective re-ranking, we showed that we can achieve significant revenue gain in a sponsored listing recommendation system, even without making any improvement on conversion estimation. We launched one of the implementations to production that yielded more than 12% revenue lift with minimum impact on user experience.

CCS CONCEPTS

• **Information systems** → *Sponsored search advertising*; **Content match advertising**.

KEYWORDS

Advertising, Recommendation, Revenue optimization, Relevance

ADKDD'21, 2021, Virtual
© 2021 Association for Computing Machinery.
ACM ISBN 123-4567-24-567/08/06. . . $15.00
https://doi.org/10.475/123_4

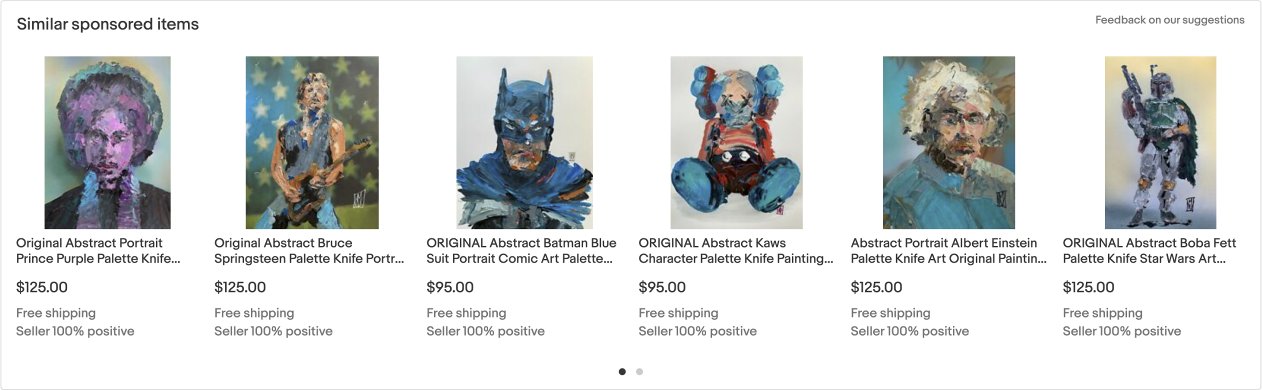
ACM Reference Format:
Zhen Ge, Wei Zhou, Jesse Lute, and Adam Ilardi. 2021. Relevance Constrained Re-ranking in Sponsored Listing Recommendations. In *Proceedings of ADKDD (ADKDD'21)*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Advertising, especially promoted/sponsored listings ("sponsored" and "promoted" are used interchangeably throughout the paper) has grown into a major revenue source at eBay in the past couple of years. The Promoted Listing program at eBay is a Cost-Per-Action (CPA) based system. When sellers sign up for the program, they set an ad rate for their listings. Unlike Cost-Per-Click (CPC) products, these sellers will only be charged if their listings are sold. This puts the sellers at no risk, and eBay as a platform is motivated to show buyers more relevant listings, that are more likely to convert (sell), other than to show simply high ad revenue listings.

On eBay's marketplace, Promoted Listings can be seen on search result page and item listing page. This paper only discusses the latter. On the item listing page, there is one primary listing (Fig. 2) that we call seed item; there are also a few strip placements on the page that recommend similar or related items to the seed item. These placements can be non-sponsored (i.e. organic) or sponsored. We discuss the ranking in a sponsored placement (Fig.1).

The motivation behind this paper is to provide good buyer experience as well as to provide an effective promotion platform for the sellers. For example, if sellers raise their ad rate on high quality items, we want to make sure that those items get reasonable ranking boosts. We call this ad rate sensitivity. On the contrary, if sellers raise the ad rate on low quality (e.g. overpriced) items, we shouldn't give the items the same level of boost because it will hurt buyer experience. This paper proposed a re-ranking algorithm for promoted recommendation which relies on local inventory/recall to control ad rate sensitivity. It introduced a variable in our final ranking function that improves the ranking by controlling the relative entropy (KL divergence) between revenue ranked list and conversion ranked list. We can dynamically adjust ad rate sensitivity for each recommendation to balance relevance and revenue. This paper intends to fill some of the gap in the literature



Similar sponsored items

Original Abstract Portrait Prince Purple Palette Knife...
$125.00
Free shipping
Seller 100% positive

Original Abstract Bruce Springsteen Palette Knife Portr...
$125.00
Free shipping
Seller 100% positive

ORIGINAL Abstract Batman Blue Suit Portrait Comic Art Palette...
$95.00
Free shipping
Seller 100% positive

ORIGINAL Abstract Kaws Character Palette Knife Painting...
$95.00
Free shipping
Seller 100% positive

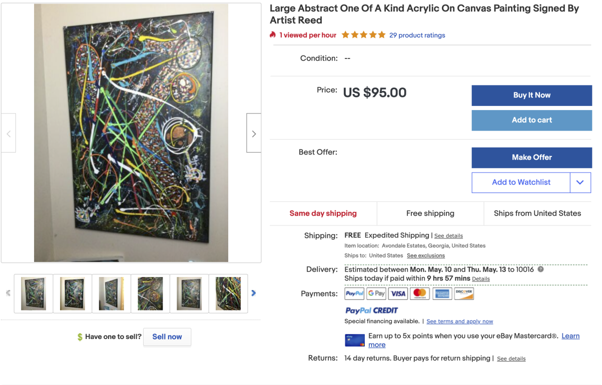
Abstract Portrait Albert Einstein Palette Knife Art Original Paintin...
$125.00
Free shipping
Seller 100% positive

ORIGINAL Abstract Boba Fett Palette Knife Star Wars Art...
$125.00
Free shipping
Seller 100% positive

Figure 1: An impression of similar listing placement

for CPA type sponsored products. We provided empirical evidence that, 1). KL divergence can be used as a quality measurement for a re-ranked list; 2) And when it's applied as a global standard for controlling for good buyer experience, it actually adjusts local ranked list's relevance individually; 3). We showed that KL divergence can be estimated through local inventory based features. 4). It's easy to implement this re-ranking approach in most of the recommendation systems to balance relevance and revenue. Because it does not require any changes in the conversion estimation, as it is completely decoupled from it.

The algorithms are built with real impression data and later tested online and subsequently launched to production.



Large Abstract One Of A Kind Acrylic On Canvas Painting Signed By Artist Reed

1 viewed per hour ★★★★★ 29 product ratings

Condition: --

Price: US $95.00

Best Offer:

Same day shipping | Free shipping | Ships from United States

Shipping: FREE Expedited Shipping | See details
Item location: Avondale Estates, Georgia, United States
Ships to: United States See exclusions

Delivery: Estimated between Mon. May. 10 and Thu. May. 13 to 10016
Ships today if paid within 9 hrs 57 mins Details

Payments: PayPal CREDIT
Special financing available. | See terms and apply now
Earn up to 5x points when you use your eBay Mastercard®. Learn more

Returns: 14 day returns. Buyer pays for return shipping | See details

Figure 2: Item listing page

2 RELATED WORK

Previous research on revenue optimization has largely focused on estimating conversion rate [2, 4, 6, 8], then re-rank the document collection by expected revenue — conversion rate multiplied by cost. Most research on ad revenue optimization is around second-price auction, there haven't been many addressing CPA and CPC ads. Liang et al. [5] at Etsy discussed revenue optimization by incorporating price as part of the optimization function; Zhu et al.'s research discusses CPC based ad revenue optimization [10] and incorporating relevance constraint into revenue optimization to balance between revenue and user experience [9]. These methods combine two objectives, conversion and revenue into one model. Such models can be unstable during training, especially in our case, conversion is labeled 0 or 1, but ad revenue is unbounded. In addition, as discussed in Zhu et al.'s research, the balancing parameter is incorporated in the loss function and optimized globally. They observed a clear trade-off between revenue and accuracy. Our paper keeps the conversion estimation and the revenue ranking as two separate stages. And we are able to improve revenue without hurting relevance through local context based re-ranking.

3 INVENTORY BASED RE-RANKING

3.1 Background

Like most recommendation systems, eBay's "similar sponsored items" recommendation (Fig.1) on the item listing page has two main stages: recall and ranking [1]. The primary listing on the item page is called seed item, and the recalled items are from seed item's coview-ed items and similar titled items. For the ranking stage, sponsored listings go through two rounds of ranking: 1. conversion estimation, 2. expected revenue. Firstly, the conversion estimation for CPA product is purchase through rate (PTR), or probability of sale p. p is estimated by a learning-to-rank model and a calibration layer. We train a Gradient Boosted Tree (GBT) model with LambdaMART, using features such as listing context (e.g. title, price, item attributes), listing history (e.g. sale count, seller rating) and user preferences (e.g price preference). The raw model scores are then scaled with Platt calibration [7] to produce probability scores. Secondly, we rank promoted items by expected revenue through below ranking function:

$$score = \underbrace{p \cdot c}_{\text{Organic Revenue}} + w \cdot \underbrace{p \cdot b}_{\text{Ad Revenue}} \quad (1)$$

This ranking function is a weighted sum of expected organic revenue and ad revenue. The expected organic revenue is conversion estimation p multiplied by selling cost, and the expected ad revenue is conversion estimation multiplied by ad rate b. w is ad revenue weight, it's used to balance the two and thus influence the final ranking. When w is 1, the revenue ranking function is reduced to regular expected revenue. Organic revenue and purchase count are highly correlated and we consider overall purchase count as a measurement for recommendation relevance. This paper discusses the trade-off between ad revenue and purchase count.

When setting an ad rate, sellers tend to raise their listing price if they set a higher ad rate, to make up for their profit margin. If these overpriced listings are in the better slots and do not sell, it's lost opportunities for other sellers and less desired shopping experience for buyers. Since we use weight w to control the ad rate sensitivity in the final ranking, w has been set relatively low in the past to mitigate the risk of bad shopping experience. However, low ad revenue weight means sellers would have to set a very high ad rate to see a change in their ranking. It could impede sales velocity. Typically, as ad revenue weight goes up globally, the recommendation becomes more sensitive to sellers' ad rate, we rank items with high ad rate much higher and thus tend to recommend more expensive listings. Ad revenue gets a small lift through higher priced items; however, relevance of the recommendation drops, and the overall purchase count and purchase through rate drop. Fig.3 shows an earlier experiment where we tested a range of ad revenue weight from 0.1 to 3. The left y-axis is ad revenue per impression and the right y-axis is purchase count per impression. We see that on a global scale, there is a clear trade-off between ad revenue and purchase count, as the weight w increases, ad revenue increases but purchase count drops, which indicates that we make less relevant recommendations. Similar trade-off is also observed in Zhu et al.'s paper.

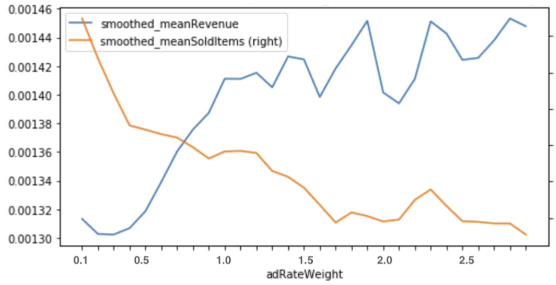


Figure 3: Average ad revenue & Average item sold count vs. Ad revenue weight

The ad revenue and item sold count (i.e. purchase count) are averaged over all impressions.

3.2 Dynamic ad revenue weight optimization (DARWO)

eBay is a marketplace with billions of listings. From high inventory categories such as clothing, handbag, cell phone case, to categories such as smartphones, TVs, and other common electronic devices, to the rare collectible coins, baseball cards, and vintage jewelry. There is an imbalance of inventory, listing quality, as well as competitiveness among different categories. For example, the cell phone case category is a much more competitive category than collectible coins, which means there are a lot more sellers with similar quality listings to compete for the same impression slot, as opposed to collectible coins.

It's important for CPA ad products to recommend relevant items because the action of purchase is a much rarer event than click. Showing the wrong recommendations loses the chance to show potentially convertible items, which may have lower ad rate, but still eventually contribute to ad revenue. Setting one fixed ad revenue weight w for all ad recommendations is not the optimal way for either buyer's shopping experience or sellers who promote their items. For some impressions, we may want to lower ad rate sensitivity by tuning down ad revenue weight because otherwise we could boost very low quality items that have very low chance of selling; for other impressions, we may have an abundance of good quality items in the recall set, high ad rate sensitivity would help sellers' sales velocity when they list items with higher ad rate and the items get better ranking. We adopted the below measurement to control for ranking quality for each impression independently.

3.2.1 Kullback–Leibler divergence constraint. When re-ranking listings by revenue, we make the assumption that the PTR scores from the machine learning model is the best purchase probability estimation for the item and the original ranking by PTR represents the most relevant recommendations. Although there are many arguments to be made that the predicted conversion probability and the true relevance are different[3], this paper applies them interchangeably. To compare two ranked lists, Jaccard Similarity and NDCG are often used. Jaccard Similarity calculates the overlap set of the top K items between two ranked lists. The overlap is based on listing Ids and it doesn't consider the positions. It's not ideal for our use case because 1. rank position matters; 2. We identify relevant recommendations based on PTR scores, not fixed sets of listings. Listing A and B are interchangeable in the ranked list if they have the same PTR score. NDCG considers position, it's often used to evaluate a ranked list with labeled relevance. Our labeled relevance only includes purchase (1) and no purchase (0), but the ranked list has K slots and most of them have no purchases. Since we are interested in the ranked list's PTR score changes but not the exact sequence of listing Ids, we decided to use the actual PTR score instead of listing rank position. Given a set of ranking candidates S from the recall stage, the machine learning model ranks them and we take top K. K is the number of slots in the placement, which are the listings that are visible to users. Each slot is denoted as k. The top K PTR score list is identified as P. Then with Eq.(1), the entire candidate set S gets re-ranked by revenue, which rearranges the PTR scores, and the new top K item list is denoted as Q. We then look at how the new revenue ranking Q diverged from the most relevant ranking.

$$D_{kl}(P||Q) = \sum_{k \in K} P(k) log(\frac{P(k)}{Q(k)}) \tag{2}$$

Essentially, given that P is the most relevant ranking, we want to measure and control how Q (re-ranked by revenue) diverges from P given the variable w from Eq.(1). For example, with K=6, if we have two different w's, w_i and w_j (PTR scores are shown unnormalized):

P: PTR(0.5, 0.45, 0.4, 0.3, 0.2, 0.15)
Q_{w_i}: PTR(0.45, 0.5, 0.3, 0.4, 0.15, 0.1)
Q_{w_j}: PTR(0.1, 0.5, 0.45, 0.3, 0.4, 0.2)
$D_{kl}(P||Q_{w_i}) = 0.021$
$D_{kl}(P||Q_{w_j}) = 0.258$

Ranking $Q(w_j)$ is less desirable because it introduces significant changes to users' experience.

There are over 40 thousand different categories at eBay, some categories are well stocked and have a lot of similar items (in price, title, selling history) than the categories in the long tail. Some seed items can have hundreds of items in the recalls (e.g. iphone), and these items on average would have much higher PTR scores than the long tail items. One important reason we chose KL divergence is because of its normalization nature. Since P and Q are probability distributions, the top K PTR scores are normalized. It's easy to see that the entropy of the top K PTR score does not change if all elements get multiplied by a constant factor. The magnitude of scores does not affect entropy. In a marketplace where each impression can have vastly different PTR scores, KL divergence will not suffer from the variability of PTR scores and can guarantee a better user experience for each recommendation.

3.2.2 Method 1: Greedy optimization.

With Eq.(1), given ad revenue weight $w \in \mathcal{W}$, we formulated the problem as:

$$\begin{aligned} Q^* &= \arg\max_{w \in W} f_{rev}(Q_w) \\ &\text{s.t.} D_{kl}(P||Q_w) \leq \theta_{KL} \end{aligned} \tag{3}$$

θ_{KL} is a constant. $f_{rev} : \mathbb{R}^k \to \mathbb{R}$ is the estimated ad revenue for ranking Q_w. Distribution Q_w represents the ranked list X^w, in which x_i represents the item in each ranking slot.

$$\begin{aligned} X^w &= [x_1^w, x_2^w, ..., x_k^w] \\ f_{rev}(Q_w) &= f_{rev}(X^w) \\ &= \sum_{r=1}^{k} a_r^w \cdot v_r \end{aligned} \tag{4}$$

a_r^w is the ad revenue for the item at slot r. v_r is the unbiased click through rate for slot r. v_r is separately estimated based on a previous exploration experiment. Eq.(3) maximizes revenue over all $\mathcal{W}$'s space with a relevance constraint θ_{KL}. θ_{KL} is to guarantee relevance of the recommendation for each impression. Q^* can be solved by grid searching in $\mathcal{W}$ to get the ranking list Q_w which meets the θ_{KL} limit and has the highest ad revenue. In practice, we set $\mathcal{W}$ to be in range [0, 7]. Empirically, we found that Q_w no longer varies when w is greater than 7. We perform the grid search for each individual recommendation, and each impression will have a different ad revenue weight w for its final ranking function (Eq.(1)).

3.2.3 Method 2: Regression.

When the recall size gets large, grid search can get expensive at inference time. So we reformulate the problem as a regression problem. Instead of searching for the right KL divergence between the final ranking PTR score list and the original PTR scores list, we tried to predict the KL divergence, based on some characteristics of an entire recall set and a given ad revenue weight. The hypothesis is that each recall set's quality is different, high quality impressions can have more item shuffling in the second round ranking without damaging the impression quality; and low quality impressions should stick to the first round PTR ranking results as closely as possible. We created below summary statistics as features that represent recall set quality: recall set size, and the maximum, minimum, mean, median, and standard deviation of all recall set item's price, PTR score, and ad rate. E.g. For PTR score, we have features: maxScore, minScore, meanScore, medianScore, and stdScore. For all these statistical summary based features, we also added interaction features to account for collinearity. Other context features include marketplace (identifies different countries) and category. All these features are calculated for a given recall set. Then we generate training examples by varying ad revenue weight w to produce different rank lists Q_w in order to calculate $D_{kl}(w)$ for top K:

$$D_{kl}(w) = f(recall\_set\_statistics, context, w) \tag{5}$$

Ordinary Least Squares Regression (OLS). We did some transformations on both features and target D_{kl} to facilitate a better fit for OLS. Recall set size, price and score related features are taken *log*. D_{kl} is transformed with Box-Cox method:

$$D'_{kl} = boxcox(D_{kl} + \epsilon, \lambda = -5) \tag{6}$$

$$D'_{kl} = \beta_0 + \sum_{i=1}^{n} x_i \cdot \beta_i \tag{7}$$

ϵ is a positive small value added to D_{kl} to make sure it's a positive value. β is the coefficient of the OLS regression, x is the recall set features. The fitted OLS model shows that our features can well explain the target variable D'_{kl} . The regression model resulted in an R^2 and adjusted R^2 value of 0.43. At the inference time, we calculate the recall set level features, and set a fixed θ_{KL} to calculate ad revenue weight for that impression(Eq.(8)).

$$w = \frac{\theta_{KL} - \beta_0 - \sum_{i=0, i \neq w}^{n} x_i \cdot \beta_i}{\beta_w} \tag{8}$$

In addition to OLS, we also tested a GBT regression model. For this ensemble based tree model, we swapped the target variable D'_{kl} and ad revenue weight w during training, to make predicting ad revenue weight at inferencing time feasible. However, we recognize that fitting regression models helps us to achieve the relevance goal but not directly optimize revenue, but the online experiments showed improvement in both fronts.

4 EXPERIMENTS

We want to increase ad rate sensitivity for sellers as well as to increase ad revenue. From the past experience we know that simply raising ad revenue weight can achieve that but leads to purchase drops (Fig.3). The goal is to dynamically set ad revenue weight more effectively so that we can increase ad revenue without hurting purchase count.

For the first experiment, we set the constraint θ_{KL} target at the median value of D_{kl} from the offline data; the predicted ad revenue weight distribution from the OLS model is shown in Fig.4. In production, negative predicted values are truncated at 0.

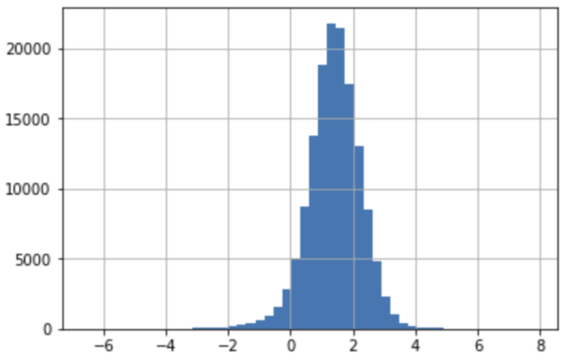


Figure 4: Ad Revenue Weight Distribution

Experiment 1. The first experiment is set up to test the hypothesis that dynamically changing ad revenue weight for each impression can improve the relevance of the recommendation. The A/B test is set up as follows (Table 1):

- Control (production): Fixed ad revenue weight at 0.25 for all impressions.
- Treatment 1: OLS based dynamic ad revenue weight (DARWO) variant.
- Treatment 2: Fixed ad revenue weight at 1.75. 1.75 is selected because it's the median value of predicted ad revenue weight from the OLS DARWO variant.

Compared to control, Treatment 1 largely raised ad revenue weight from 0.25 (as shown in Fig.4). Based on Fig.3, if ad revenue weight is raised for all impressions, we know that purchase count will drop and ad revenue will increase. To truly test if DARWO is more effective at setting ad revenue weight, we created the second treatment, fixing ad revenue weight at 1.75. We hope that setting it at the median value of OLS DARWO's predicted value, Treatment 2 can be used as a baseline comparison to Treatment 1. We expect both treatments to drop purchase count due to higher ad revenue weight. However, we expect treatment 1 to have less purchase impact due to the nature of controlling for impression quality. Table 1 shows the result for different marketplaces. US's test result didn't reflect the advantage of dynamic ad revenue weight; UK, AU and DE however, showed that dynamic ad revenue weight generated similar revenue lift to its counterpart (treatment 2), but resulted in less purchase drop. We plotted the average ad revenue and purchase count changes based on the dynamic ad revenue weight per impression (Fig.5). It shows that since we now set a global standard to guarantee good user experience, ad revenue weight is selected for each local impression to reflect inventory quality. There is no longer a clear trend up/down along ad revenue weight for either purchase count or ad revenue. Moreover, these two formerly inversely trending metrics now go closely together with a pearson correlation of 0.724 ($p < 10^{-6}$). Even though the OLS based DARWO dropped purchases compared to control, we launched it to production, considering it to be more efficient in balancing revenue and relevance (compared to Treatment 2). Meanwhile, we worked on implementing other DARWO variants.

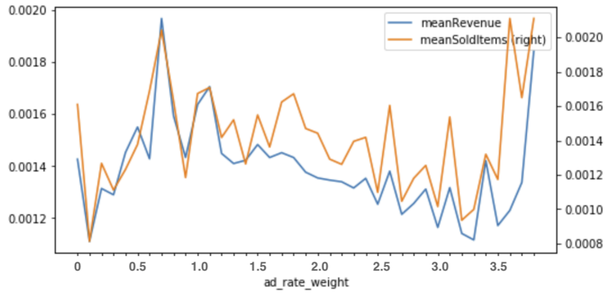


Figure 5: Average ad revenue & Average item sold count vs. Ad revenue weight (DARWO)

Experiment 2. In the second A/B test, we wanted to compare two different regressions and greedy optimization. We used OLS DARWO variant as control, the design is follows (Table 2):

- Control: OLS DARWO variant
- Treatment 1: GBT DARWO variant
- Treatment 2: Greedy optimization DARWO variant

All three variants have the same θ_{KL} constraint. The second A/B test shows that the GBT variant is a clear improvement over the simpler OLS variant. It lifted revenue without hurting purchase count, it even increased purchases in the US market. The Greedy variant lifted both revenue and purchases.

Overall, comparing to the previous production variant (control from Experiment 1), the compounded revenue lift from the GBT variant for all marketplaces is 12.6%, while the

[1] bold numbers are significant with $p<0.1$

Table 1: OLS DARWO vs. Fixed ad revenue weight by marketplace

		Ad Revenue	Purchase Count
US	treatment 1	**+3.81%**[1]	**-4.05%**
	treatment 2	**+5.33%**	**-5.07%**
UK	treatment 1	**+6.89%**	**-4.11%**
	treatment 2	**+5.81%**	**-6.55%**
AU	treatment 1	**+7.10%**	-1.97%
	treatment 2	**+8.30%**	**-3.11%**
DE	treatment 1	**+6.44%**	**-3.44%**
	treatment 2	**+5.38%**	**-4.68%**

Table 2: GBT DARWO vs. Greedy DARWO by marketplace

		Ad Revenue	Purchase Count
US	treatment 1	**+7.72%**	**+5.45%**
	treatment 2	**+8.70%**	**+8.53%**
UK	treatment 1	**+7.00%**	0.90%
	treatment 2	**+4.13%**	**+4.91%**
AU	treatment 1	**+7.45%**	+2.49%
	treatment 2	**+8.50%**	**+6.78%**
DE	treatment 1	**+4.67%**	+1.33%
	treatment 2	+2.40%	**+3.57%**

Table 3: Offline Purchase Ranking Comparison: Production, GBT DARWO and Greedy DARWO

	Mean Reciprocal Rank	NDCG@6	NDCG@12
Production	0.508	0.567	0.615
GBT DARWO	0.480	0.544	0.593
Greedy DARWO	0.516	0.576	0.620

negative impact on purchases is -1.8% (p<0.1). For Greedy variant, the compounded revenue lift for all marketplaces is 11.0%, and instead of dropping purchases, it actually improved purchase count by 2.5% (p<0.1). We subsequently decided to launch the GBT variant given its high revenue lift and simple implemention.

Offline evaluation showed a similar pattern in purchase ranking metrics (Table 3). We compared GBT DARWO and Greedy DARWO variants with the previous production variant (control from Experiment 1). It shows that the Greedy variant makes more relevant recommendations than both the previous production variant and the GBT variant. The GBT variant has a slight negative impact on purchase compared to production.

5 CONCLUSIONS

In this paper, we introduced a simple yet highly effective re-ranking mechanism based on KL divergence between the re-ranked listing scores and the original ranked listing scores for each impression. We were able to closely control the ranked listing quality for each impression in spite of the high variability of inventory at eBay. We tested three different algorithms and all showed significant improvement compared to the baseline variant, which has been in production since the inception of the Promoted Listing program. With the launch of the new re-ranking algorithm, we are able to better promote sales velocity for some impressions by raising ad revenue weight while maintaining good buyer shopping experiences.

This ad hoc re-ranking stage is completely independent of the previous ranking or conversion stages, it is not limited to revenue re-ranking and can be easily applied to any recommendation systems with re-ranking needs.

REFERENCES

[1] G. Adomavicius and A. Tuzhilin. 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* 17, 6 (2005), 734–749. https://doi.org/10.1109/TKDE.2005.99

[2] Afroze Ibrahim Baqapuri and Ilya Trofimov. 2014. Using Neural Networks for Click Prediction of Sponsored Search. *CoRR* abs/1412.6601 (2014). arXiv:1412.6601 http://arxiv.org/abs/1412.6601

[3] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2016. Unbiased Learning-to-Rank with Biased Feedback. *CoRR* abs/1608.04468 (2016). arXiv:1608.04468 http://arxiv.org/abs/1608.04468

[4] Cheng Li, Yue Lu, Qiaozhu Mei, Dong Wang, and Sandeep Pandey. 2015. Click-through Prediction for Advertising in Twitter Timeline. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*. Association for Computing Machinery, New York, NY, USA, 1959–1968. https://doi.org/10.1145/2783258.2788582

[5] Wu Liang, Hu Diane, Hong Liangjie, and Liu Huan. 2018. Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. Association for Computing Machinery, New York, NY, USA, 365–374. https://doi.org/10.1145/3209978.3209993

[6] Quan Lu, Shengjun Pan, Liang Wang, Junwei Pan, Fengdan Wan, and Hongxia Yang. 2017. A Practical Framework of Conversion Rate Prediction for Online Display Advertising. In *Proceedings of the ADKDD'17 (ADKDD'17)*. Association for Computing Machinery, New York, NY, USA, Article 9, 9 pages. https://doi.org/10.1145/3124749.3124750

[7] Alexandru Niculescu-Mizil and Rich Caruana. 2005. Predicting Good Probabilities with Supervised Learning. In *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*. Association for Computing Machinery, New York, NY, USA, 625–632. https://doi.org/10.1145/1102351.1102430

[8] Ilya Trofimov, Anna Kornetova, and Valery Topinskiy. 2012. Using Boosted Trees for Click-through Rate Prediction for Sponsored Search. In *Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy (ADKDD '12)*. Association for Computing Machinery, New York, NY, USA, Article 2, 6 pages. https://doi.org/10.1145/2351356.2351358

[9] Yunzhang Zhu, Gang Wang, Junli Yang, Dakan Wang, Jun Yan, and Zheng Chen. 2009. Revenue Optimization with Relevance Constraint in Sponsored Search. In *Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising (ADKDD '09)*. Association for Computing Machinery, New York, NY, USA, 55–60. https://doi.org/10.1145/1592748.1592756

[10] Yunzhang Zhu, Gang Wang, Junli Yang, Dakan Wang, Jun Yan, Jian Hu, and Zheng Chen. 2009. Optimizing Search Engine Revenue in Sponsored Search. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '09)*. Association for Computing Machinery, New York, NY, USA, 588–595. https://doi.org/10.1145/1571941.1572042