Treating Cold Start in Product Search by Priors

Parth Gupta, Tommaso Dreossi, Jan Bakus, Yu-Hsiang Lin, Vamsi Salaka
[guptpart|dreossit|jbakus|yuhsianl|vsalaka]@amazon.com
Amazon Search
Palo Alto, California, US

ABSTRACT

New products in e-commerce platforms suffer from cold start, both in recommendation and search. In this study, we present experiments to deal with cold start in search by predicting priors for behavioral features in learning to rank set up. The offline results show that our technique generates priors for behavioral features which closely track posterior values. The online A/B test on 140MM queries shows that treatment with priors improves new products impressions and increased customers engagement pointing to their relevance and quality.

KEYWORDS

cold start, product search, learning to rank

ACM Reference Format:
Parth Gupta, Tommaso Dreossi, Jan Bakus, Yu-Hsiang Lin, Vamsi Salaka. 2020. Treating Cold Start in Product Search by Priors. In *Companion Proceedings of the Web Conference 2020 (WWW '20 Companion), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3366424.3382705

1 INTRODUCTION

Learning to rank (LTR) models rely on several features to rank documents for a given query. Many LTR features are based on users' interactions with documents such as impressions, clicks, and purchases [3, 5]. We call these features *behavioral features*. Ranking models are trained to optimize user engagement, and therefore, such behavioral features tend to be the most important training signals. However, new and tail products that do not have user engagement lack behavioral features and hence are ranked as irrelevant, which in turn further excludes them from catching user engagement. It takes time for them to gather enough behavioral signals to show up at their fair ranking position. This leads to the causality dilemma: No behavioral data causes poor ranking which in turn results in new products having a reduced likelihood of accruing behavioral data. This phenomenon is referred to as cold start problem and poses serious concerns from bad customer experience to lost revenue opportunity.

There are several studies related to cold start in search for web search. For example, research on recency ranking focus on identifying recency sensitive queries and specifically training ranking models for them, e.g., breaking-news queries, weather information, etc. [1, 2]. Temporal IR deals with ranking during different time of the day [4]. However, product search is mainly dominated by transactional queries where new products need to surface for both existing generic queries (exploration) and specific ones (exploitation).

Our approach is to predict "prior" values for behavioral features for new products. These priors are initial values of behavioral features at the time of new product's introduction to the search index. Priors give two-fold benefits: i) they let us incorporate perceived importance for the product as certain degree of confidence and ii) they increase the exploration of new products so they can accumulate behavioral data. For example, a high prior value for a click-related behavioral feature describes the algorithm's confidence in the product's ability to receive clicks. We present offline and online experiments to further explain this.

2 APPROACH

In relevance ranking, behavioral features are often important because they encode user engagement, which is the target of learning to rank systems [3, 5]. The behavioral features, along with lexical and semantic features, are used to measure relevance (r) between query (q) and product (d) as $P(r|q, d)$. However, only products that already allured user engagement have informative behavioral features. The solution to the cold start problem should be a good exploration mechanism with the flexibility to adapt to the new user data.

We propose to estimate the prior values of the behavioral features for new products based on their attributes, such as brand, type, color, artist, author, etc. Precisely, we consider attributes such as the number of clicks the brand received historically, or the number of times the movies of an actor were watched in the past. We formulate the prior estimation as a machine learning problem where we learn a model that predicts the posterior values of the behavioral features using the attributes of the products. The posterior values predicted

WWW '20 Companion, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7024-0/20/04.
https://doi.org/10.1145/3366424.3382705

by the model are then used as the prior of the corresponding behavioral features. At ranking time, we substitute the initial behavioral features values with their prior values. After a certain time window, the new products receive user engagement, and we replace the collected prior values by the posterior ones. If the prior values are good estimates for the posterior values, then cold start can be alleviated. Our regression model for the prior estimation is defined as: $P_i^j = f(\vec{x}_i)$, where i is a product, $\vec{x}_i$ is the vector of attributes of i, j is a behavioral feature, and P_i^j is the prior for the behavioral feature j of the product i.

Theoretically, the new products should be explored with some probability and later the decision to update that probability should be data driven based on the exploration result. Priors provide a good mechanism for this exploration in LTR framework. In the most naive way, priors can follow uniform distribution but the success of this approach lies in ability to generate priors that tightly follow the actual posterior distribution.

3 EXPERIMENTS AND RESULTS

For evaluation our methods, we carried offline and online experiments. We give details of the model training and evaluation with results and analyses in the following sections.

Model Implementation. Due to nonlinear interaction among product attributes, we used gradient boosted trees for regression. The input to the regression model is a vector of the attributes of a product, and the output is the predicted behavioral feature value. We predicted priors for three behavioral features: i) click rate, ii) purchase rate, and iii) consume rate.[1] For the training data, we took around 3MM media products for which behavioral data are available.

Offline Prior Estimation. Here we measure how close the predictions of our priors are to the values of posteriors. The performance is measured in terms of goodness of fit (R-squared) and Pearson's correlation. The models were trained on historical data and evaluated on a disjoint test set. Model performances are reported in Table 1.

Table 1: Offline evaluation on 3MM media products across books, music, and video category for three behavioral features (click, purchase, and consume rates). Cells report Person's correlation and R-squared scores (above and below, respectively).

	Click rate	Purchase rate	Consume rate
Pearson's Corr.	0.9304	0.9357	0.9475
R-Squared	0.8993	0.8754	0.8978

Online Prior Effectiveness. The priors give us flexibility to abstract the cold start treatment out of the system architecture of LTR. The ranker is agnostic to where the behavior feature value is coming from, prior or posterior. For evaluation, we ran an A/B test using control (baseline) with no priors and treatment with priors. One benefit of evaluating with online test is that it alleviates the bias caused by lacking initial behavioral feature values. The offline historical data cannot tell us whether new products initialized with priors and warmed up in ranking receive more clicks or consumptions. This leads to the difficulty of estimating offline effect of priors on ranking metrics such as NDCG or MRR without explicit human judgment.

Table 2: Online results of A/B test with baseline (without priors) and control (with priors). The statistical significance is denoted by * (p-value < 0.05).

System	Impression	Click	Consumption
Priors	+3.23%*	+1.35%*	+5.42%*

The test was run for 4 weeks and the user behavior was tracked to see: i) if the new products are explored more, and ii) if higher exploration entails stronger customer engagement. The former is measured by the impression rate of new products, while the latter is measured in terms of click and consumption rate. Note that since the activities associated with new products are only a small portion of all activities, we collect only the search data attributed to new products. Without such filtering, the signal would be overwhelmed by the noise coming from the activities unrelated to the new products. Table 2 shows the results of our A/B test.

It can be noticed from Table 2 that incorporation of priors help surface higher number of new products. Furthermore, it instills positive customer behavior denoted by higher click and consumption rate. It means, the customer engaged significantly more (p-value < 0.05) with the new products in treatment compared to control.

4 REMARKS

We presented a framework to encode cold start treatment through priors of behavioral features. The offline evaluation suggests that ML system can accurately predict priors that closely track posterior values of new product features. This is work in-progress and initial results are encouraging. The A/B test on 140MM searches strongly suggest that the presented method impresses high number of new products which attracts significantly positive customer engagement.

REFERENCES

[1] R. Campos, G. Dias, A. M. Jorge, and A. Jatowt. Survey of temporal information retrieval and related applications. *ACM Comput. Surv.*, 47(2):15:1–15:41, Aug. 2014.

[2] A. Dong, Y. Chang, Z. Zheng, G. Mishne, J. Bai, R. Zhang, K. Buchner, C. Liao, and F. Diaz. Towards recency ranking in web search. In *WSDM 2010*.

[3] S. Jiang, Y. Hu, C. Kang, T. Daly, Jr., D. Yin, Y. Chang, and C. Zhai. Learning query and document relevance from a web-scale click graph. In *SIGIR 2016*.

[4] M. Shokouhi and K. Radinsky. Time-sensitive query auto-completion. In *SIGIR 2012*.

[5] W. Wu, H. Li, and J. Xu. Learning query and document similarities from click-through bipartite graph with metadata. In *WSDM 2013*.

[1] Consume rate defines a rate at which customers read, listen or watch a media product like digital book, music, or video.