

Enhancing Product Search by Best-Selling Prediction in E-Commerce

Bo Long, Jiang Bian, Anlei Dong, Yi Chang
Yahoo! Labs
Sunnyvale, CA 94089
{bolong, jbian, anlei, yichang}@yahoo-inc.com

ABSTRACT

With the rapid growth of E-Commerce on the Internet, online product search service has emerged as a popular and effective paradigm for customers to find desired products and select transactions. Most product search engines today are based on adaptations of relevance models devised for information retrieval. However, there is still a big gap between the mechanism of finding products that customers really desire to purchase and that of retrieving products of high relevance to customers' query. In this paper, we address this problem by proposing a new ranking framework for enhancing product search based on dynamic best-selling prediction in E-Commerce. Specifically, we first develop an effective algorithm to predict the dynamic best-selling, i.e. the volume of sales, for each product item based on its transaction history. By incorporating such best-selling prediction with relevance, we propose a new ranking model for product search, in which we rank higher the product items that are not only relevant to the customer's need but with higher probability to be purchased by the customer. Results of a large scale evaluation, conducted over the dataset from a commercial product search engine, demonstrate that our new ranking method is more effective for locating those product items that customers really desire to buy at higher rank positions without hurting the search relevance.

Categories and Subject Descriptors

H.3.3 [Information Systems]: Information Search and Retrieval

General Terms

Algorithms, Experimentation, Measurements

Keywords

E-Commerce, Product Search, Best Selling Prediction, Transaction History

1. INTRODUCTION

As E-Commerce comprises one of the fastest growing segments on the Internet, online product search has recently become a viable method for seeking customers' desired product items. Similar

to general purpose Web search, product search engines allow customers to submit keyword based queries and return to them with a list of product items, within which customers can pick the items they desired to purchase online. Most of product search engines today are built based on the relevance model from classic information retrieval theory [5]; or, some others use variant of faceted search [7, 1, 6] to facilitate browsing of product search. However, the mechanism underlying the process of locating product items that customers really desire to purchase is quite different from that of retrieving product items of high relevance to customers' queries. For example., customers usually attempt to find a best deal that can satisfy their demands rather than simply seek the product items relevant to their search queries. Due to this critical gap between the product search and general purpose Web search, it has made necessity to design a new ranking method that can identify customers' most wanted deal beyond retrieving product items relevant to customers' queries.

Some today's product search engines provide secondary ranking facilities, e.g. ranking criteria based on the volume of sales or customer ratings, to re-rank the search results. Although this approach has already taken into account customers' opinions in terms of volume of sales or ratings, it still yields a few shortcomings. First, it hardly considers the dynamics of customers' willing to purchase specific product items. For example, some product items used to be the best-selling ones. But, after a new generation of this product has been released recently, customers prefer to purchase the new generation even it does not have larger volumes of sales than the old generation. Second, for products with a low purchase frequency or sparse transaction records, existing secondary ranking criteria, such as volumes of sales or customer ratings, may not be effective enough to reflect the customers' preference. Moreover, the larger variety of products can indicate significant variations of transaction patterns, however, existing approaches rarely take into account the heterogeneity of products in terms of transaction patterns. As a consequence, existing approaches simply based on secondary ranking criteria lack the capability of deriving customers' preference. An alternative technique [4] succeeded in enhancing product search by incorporating expected utility theory from economics, while its major limitation lies on the difficulty to explicitly model the dynamics of customers' preference on products.

To address these challenges, in this paper, we propose a new ranking framework which introduces dynamic best-selling prediction to enhance the product search. Specifically, we first develop an effective algorithm to predict the dynamic best-selling, i.e. the volume of sales, for individual product item based on its transaction history. This algorithm is quite robust to the sparse transaction records. Then, we propose a new ranking model for product search by integrating the best-selling prediction into the relevance.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.

Copyright 2012 ACM 978-1-4503-1156-4/12/10 ...\$15.00.

This new model aims at ranking those product items, which are not only relevant to the user's query but with higher probability to be purchased by the user, into the higher positions. To evaluate our new ranking method, we collect a large scale dataset of search and transaction logs from a commercial product search engine. Experimental results demonstrate that our new ranking method is more effective for locating product items that users really desire to buy at higher rank positions without hurting the search relevance.

The specific contributions of this paper include: (1) Introduction of a new ranking framework for enhancing product search by using dynamic best-selling prediction. (2) Development of an effective algorithm to predict the dynamic volume of sales for individual product items based on transaction history.

2. ENHANCED RANKING FOR PRODUCT SEARCH

In this section, we will explore a new ranking framework for product search based on best-selling prediction. In particular, we will start with introducing the new general ranking framework, followed by concrete discussion on the major steps of the framework in sequence.

2.1 Ranking Framework

Ranking models are at the core of an effective product search system. To maximize customers' satisfaction when using product search, the ranking model should succeed in ordering the product items that customers really want to buy at top positions. Traditional ranking model based on relevance can help customers find the product items relevant to the search queries, but may not be effective to retrieve product items that customers are really willing to purchase. Fortunately, the volumes of sales or best-selling for individual product items provide good indication of customers' willing to purchase the corresponding products. Therefore, we propose a new ranking framework based on best-selling prediction, which consists of three major steps:

Step-1: Computing the relevance between product items and customers' queries by using traditional relevance model.

Step-2: Best-selling prediction for individual product items based on the transaction history.

Step-3: Combining the relevance scores with predicted best-selling scores to generate the ranking.

This new ranking framework can aggregate the relevance signals and customers' willing to purchase together so as to boost customers' satisfaction in product search. In the rest of this section, we will discuss the details of each step in sequence.

2.2 Relevance Model

In the first step, we target at computing the relevance between each product item and the customer's query submitted to product search. This step aims at finding those product items that are relevant to the customer's need. We can employ classic learning-to-rank approach to address this problem. Specifically, we first extract a set of features to describe the correlation between the query and the product items, such as the textual similarity between the query and title/description of product items or the textual similarity between the query and user-generated reviews of product items. After feature extraction, we can apply any popular learning-to-rank algorithm to learn the ranking model for computing the relevance between the query and the product item.

In our work, we use the Gradient Boosted Decision Tree (GBDT) algorithm [2] to learn the ranking model. GBDT is an additive regression algorithm consisting of an ensemble of trees, fitted to

current residuals, gradients of the loss function, in a forward step-wise manner. It iteratively fits an additive model as

$$f_t x = T_t(x; \Theta) + \lambda \sum_{t=1}^T \beta_t T_t(x; \Theta_t)$$

such that certain loss function $L(y_i, f_T(x+i))$ is minimized, where $T_t(x; \Theta_t)$ is a tree at iteration t , weighted by parameter β_t , with a finite number of parameters, Θ_t and λ is the learning rate. At iteration t , tree $T_t(x; \beta)$ is induced to fit the negative gradient by least square. That is

$$\hat{\Theta} = \arg \min_{\beta} \sum_i^N (-G_{it} - \beta T_t(x_i); \Theta)^2$$

where G_{it} is the gradient over current prediction function

$$G_{it} = \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f=f_{t-1}}$$

The optimal weights of trees β_t are determined by

$$\beta_t = \arg \min_{\beta} \sum_i^N L(y_i, f_{t-1}(x_i) + \beta T(x_i, \theta))$$

2.3 Best-Selling Prediction

Although the relevance model in the first step can hardly indicate customers' willing to purchase one specific product item, the volumes of sales or best-selling for this product item provide a strong signal about whether customers desired to buy this item. Therefore, in this step, we aim at predicting the volume of sales for each product items based on transaction history.

However, there are several challenges for best-selling prediction. First, it is uneasy to build one prediction model for a large amount of product items with large variations of transaction patterns. Second, it is quite challenging to predict volumes of sales for product items with a low purchase frequency or sparse transaction records. Moreover, the product search system requires efficient dynamic feature computation for updating the prediction model frequently.

To address these challenges, we propose a transaction window based dynamic linear regression model for best-selling prediction. Traditional moving time window based approach does not work well due to the large transaction variation among the a large amount of shopping items. For example, given a time window such as 10 days, some items may have a lot of transactions and some items may have very few transactions. In our work, we propose a new concept of transaction window. Specifically, to predict the next daily transaction window, we use the data from last k transaction days. A transaction day for an item is defined as a day in which there is at least one transaction for this item.

In this approach, we take advantage of dynamic first derivative features. The first derivative feature for a item at i -th transaction day can be efficiently computed as follows:

$$x_i = \frac{2.0(s_i - s_{i-1})}{(s_i + s_{i-1})(d_i - d_{i-1})} \quad (1)$$

where x_i denotes the feature of i -th transaction day; s_i denotes the number of transactions at i -th transaction day; and, d_i represents the date for i -th transaction day, i.e. $d_i - d_{i-1}$ equals the number of days between i -th transaction day and $(i-1)$ -th transaction day.

After deriving dynamic first derivative features, we apply linear regression model to predict the number of transactions for the next

day by using a window of k transaction days as follow:

$$s_{k+1} = \left(1 + \sum_{i=1}^k \frac{\alpha_i x_i}{d_{k+1} - d_k}\right) s_k \quad (2)$$

where s_{k+1} is the number of transactions we predict for the day d_{k+1} ; x_i is the feature for the last i transaction days; α_i is the model parameter learned from the training data; and, s_k is the number of transactions at the most recent transaction day d_k .

2.4 Enhanced Ranking

After obtaining the relevance model and best-selling prediction model, we will combine them together to compute the final ranking score for each product items. In particular, given a customer's specific query, q , and a pool of candidate product items, $\{p_1, p_2, \dots, p_n\}$, we first apply the relevance model to compute the relevance scores for all product items, denoted as $\{\sigma_1, \sigma_2, \dots, \sigma_n\}$, then, we leverage best-selling prediction model to obtain the best-selling scores, denoted as $\{\eta_1, \eta_2, \dots, \eta_n\}$. Finally, we can compute the aggregated ranking score for each product item as follows:

$$S_i = \sigma_i + \beta \cdot \eta_i \quad (3)$$

where S_i represents the final ranking score of i -th product item; and, β is the model parameter that will be tuned based on cross-validation.

3. EXPERIMENTS

3.1 Data Preparation

To evaluate the performance of our new ranking framework for product search, we collect datasets from a commercial product search engine. We first collect a datasets to learn the best-selling prediction model. After that, we collect another dataset to test the performance of our enhanced ranking method. We also applied the ranking model trained for a text based search engine as the classic relevance model in our new ranking framework.

Dataset for learning best-selling prediction model: To obtain best-selling prediction model, we collect two years transaction history for all of over 35,000 product items based on logs of a commercial product search engine. This dataset records the number of transactions per day for each of the product items. We split the whole dataset into training, validation, and testing set.

Dataset for testing enhanced ranking: To evaluate the performance of our new ranking framework, we sample two datasets from the search logs of a commercial product search engine. Both of them are based on search logs of Aug, 2011 but of different types of queries, one type corresponding to baby products while the other for flag products. There are totally 211 queries for baby products as well as 200 queries for flag products. We collect the top-20 ranked product items for each query based on the classic relevance model. Each pair of $\langle \text{query}, \text{product item} \rangle$ was assigned with a relevance label, including (*excellent*, *good*, *fair*, or *bad*), by human editors. The search logs also record if the customers who submitted search queries purchased any product items from the search results.

Classic relevance model: For the classic relevance model in our new ranking framework, we applied the ranking model trained for a commercial text based search engine. In particular, this relevance model is trained using a dataset consisting of more than 30,000 pairs of $\langle \text{query}, \text{document} \rangle$, each of which was represented by textual features and was assigned with human judged relevance label.

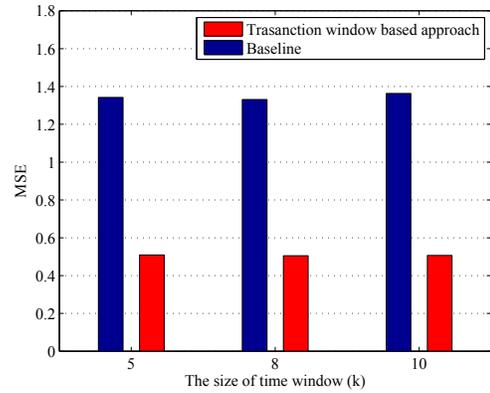


Figure 1: MSE for transaction window based dynamic linear regression model compared with the baseline model over varying sizes of time window.

3.2 Evaluation Metrics

To evaluate the performance of our new algorithm for best-selling prediction, we adapt the metric of Mean Square Error:

- **Mean Square Error (MSE):** The MSE reports the difference between predicted numbers of transactions and actual numbers of transactions, which can be calculated as

$$\text{MSE} = \text{E}[(\hat{s} - s)^2]$$

where \hat{s} denotes predicted number of transactions, and s is the actual number of transactions.

To evaluate the performance of our new ranking framework for product search, we adapt the Discounted Cumulative Gain to measure the relevance while we use number of selling at rank position 1 to measure customers' willing to purchase.

- **Discounted Cumulative Gain (DCG):** DCG has been widely used to assess relevance in the context of search [3]. For a ranked list of N items, we compute the following variation of DCG:

$$\text{DCG}_N = \sum_{i=1}^N \frac{G_i}{\log_2(i+1)}$$

where G_i represents the weights assigned to the label of the item at position i , e.g. in this paper, 7 for *excellent*, 3 for *good*, 1 for *fair*, and 0 for *bad*.

- **Number of selling at position 1 (Sell₁):** This metric is used to describe whether customers desire to purchase the product items ranked at top-1 position.

3.3 Experimental Results

3.3.1 Best-Selling Prediction

We first evaluate the accuracy of our transaction window based dynamic linear regression model for best-selling prediction. In this offline experiment, we employ a *replay* evaluation method which simulates the online learning procedure. Specifically, for one product item, to predict its volume of sales at one specific day, we use its transaction records from previous k transaction days as training set to learn the parameters of the best-selling model; then, we apply the model to predict its volume of sales at that day and compute the MSE between the predicted value and the actual volume of sales. The final MSE is computed as the average value over all product items and all of their transaction days.

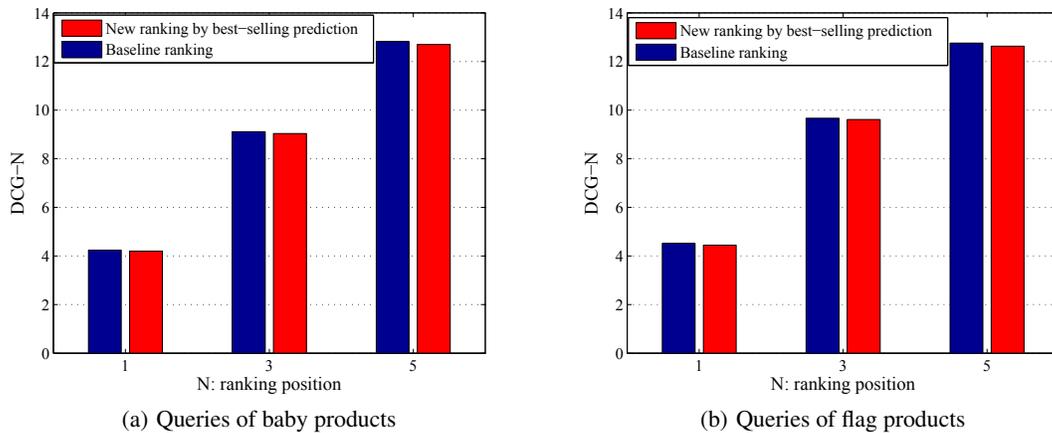


Figure 2: DCG_N for our new ranking method compared with the baseline ranking

In this experiment, we compare the performance of our transaction window based dynamic linear regression model with that of traditional moving time window based approach. Figure 1 demonstrates the MSE for transaction window based dynamic linear regression model compared with the baseline model over varying sizes of time window. From this figure, we can find that, for both transaction window based model and the baseline model, we can achieve the best accuracy when the size of time window is set as 8. More importantly, this figure shows that our new transaction window based model can reach better performance, with a gain over 62%, for best-selling prediction than the traditional time window based model.

3.3.2 Enhanced Ranking for Product Search

In this experiment, we seek to enhance the ranking performance of product search by combining the relevance score with the best-selling prediction score. As introduced in Section 2.2 and 2.3, we apply GBDT as the relevance model and use transaction window based dynamic linear regression model to predict best-selling.

Table 1: The growth rate of the number of transactions happened at rank position 1, for two types of queries, respectively, by using our new ranking framework over the baseline ranking which uses relevance model only.

	Queries for baby products	Queries for flag products
Growth rate	245.6%	149.3%

We test our new ranking method on the two types of sampled queries. After tuning parameters on a hold-out validation set, we set the boost coefficient as 0.05. Table 1 illustrates the growth rate of the number of transactions happened at rank position 1, for two types of queries, respectively, by using our new ranking framework over the baseline ranking which uses relevance model only. From these two figures, we can find that, on both types of queries, our new ranking framework can significantly boost the number of transactions happened at ranking position 1, indicating that our new ranking framework can more effectively retrieve those product items that customers are really willing to buy.

To assess the effects of our new ranking framework on the search relevance, we evaluate the ranking relevance conducted by our new ranking framework on two types of queries. When still setting the boost coefficient as 0.05, Figure 2 show the ranking relevance in

terms of DCG_i , on two types of queries, respectively, for our new ranking method compared with the baseline which uses pure relevance model without best-selling prediction. From these two figures, we can find that, although the ranking relevance of our new ranking framework declines a little compared to the baseline ranking, there is no big hurt for the relevance. It can imply that our new ranking framework can improve customers' satisfaction with almost no hurt on the relevance.

4. CONCLUSION

In this paper, we introduced a new ranking framework for enhancing product search based on dynamic best-selling prediction in E-Commerce. In particular, we first developed an effective algorithm to predict the dynamic volume of sales for each product item based on its transaction history. Then, we propose a new ranking model for product search, incorporating such best-selling prediction with relevance. A large scale evaluation, conducted over the dataset from a commercial product search engine, demonstrate that our new ranking method is more effective to retrieve those product items that customers really desire to buy with very little hurt on the search relevance. In the future, we will explore more aspects other than best-selling to represent customers' willing to purchase the product items.

5. REFERENCES

- [1] X. Chen, H. Wang, X. Sun, J. Pan, and Y. Yu. Diversifying product search results. In *Proc. of SIGIR*, 2011.
- [2] J. H. Friedman. Greedy function approximation: a gradient boosting machine. In *Annals of Statistics*, 2001.
- [3] K. Jarvelin and J. Kekalainen. Cumulated gain-based evaluation of ir techniques. In *ACM Transactions on Information Systems*, 2002.
- [4] B. Li, A. Ghose, and P. G. Ipeirotis. Towards a theory model for product search. In *Proc. of WWW*, 2011.
- [5] Z. Nie, J.-R. Wen, and W.-Y. Ma. Webpage understanding: beyond page-level search. In *Proc. of SIGMOD*, 2008.
- [6] Q. Peng, W. Meng, H. He, and C. Yu. Clustering e-commerce search engines. In *Proc. of WWW Alt.*, 2004.
- [7] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proc. of CHI*, 2003.