A User-Tunable Approach to Marketplace Search

Nish Parikh Neel Sundaresan
eBay Research Labs
2065 Hamilton Ave,
San Jose CA 95125.
1(408)376-2000
{nparikh,nsundaresan}@ebay.com

ABSTRACT

The notion of relevance is key to the performance of search engines as they interpret the user queries and respond with matching results. Online search engines have used other features beyond pure IR features to return relevant matching documents.

However, over-emphasis on relevance could lead to redundancy in search results. In document search, diversity is simply the variety of documents that span the result set. In an online marketplace the diversity in the result set is represented by items for sale by different sellers at different prices with different sales options. For such a marketplace, in order to minimize query abandonment and the risk of dissatisfaction to the average user, several factors like diversity, trust and value need to be taken into account. Previous work in this field [4] has shown an *impossibility result* that there exists no such function that can optimize for all these factors. Since these factors and the measures associated with the factors could be subjective we take an approach of giving the control back to the user.

In this paper we describe an interface which enables users to have more control over the optimization function used to present the results. We demonstrate this for search on eBay - one of the largest online marketplaces with a vibrant user community and dynamic inventory. We use an algorithm based on bounded greedy selection [5] to construct the result set based on parameters specified by the user.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *information filtering.*

General Terms

Algorithms, Design.

Keywords

eCommerce, search, diversity, relevance, trust, value

WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0637-9/11/03.

1. MOTIVATION AND RELATED WORK

Typically, online shopping sites provide users with an option to enter search keywords and then display a subset of the inventory matching the user's search. The sites would want to show the products which are likely to get bought and try to show the best results on the limited "real-estate" of the web page based on some ranking functions. How to show the best possible results is a problem that has been extensively studied in the Information Retrieval domain [7]. Relevance denotes how well a retrieved document or set of documents meets the information need of the user. Concept of relevance in the context of eCommerce is discussed in [6].

Optimizing for relevance can lead to redundancy in the retrieved result set [3]. Hence, promoting diversity in order to reduce redundancy is another goal that search engines must address. Diversity is important even in the physical world and has been studied and described in different fields of sciences in literature [10]. Diversification is important in all arenas be it ecosystems, flora, fauna, food, human languages, stock investing, bank asset management or even, selection of models to form an ensemble learning system.

In commerce the measure of strength of the store is the diversity of inventory available for sale. Online shopping has increased in popularity due to the large availability of online inventory.

For eCommerce, other factors that come into play are trust and value. For online retailers like Walmart and Target the trust or reputation associated with the products is that of the online retailer itself as they control and manage all the inventory sold on their site. However, in a multi-buyer to multi-seller marketplace like eBay as there are many sellers selling different items, the trust factor associated with every item is different. Some sellers are more professional than others, some sellers ship faster than others, some sellers have proven themselves for a long time through good customer service and hence can be trusted more than others and so on. If the same product is being sold for the same price by two different sellers with different reputations, buyers would surely want to buy it from the seller with a better reputation. The importance of reputation in online systems has been described in [2]. Thus, the trust value associated with a product is based on the reputation of the seller who is selling the product.

In today's online world, buyers have a lot of choice. A compelling value proposition is something that many buyers look for, before completing their purchase and hence it is no surprise that sites like Woot[1] and Slickdeals[2] are gaining popularity.

[1] http://www.woot.com

Given two exactly same items with the same trust factor, buyers are more likely to buy the one with a lower price. Thus, value proposition provided by an item to the buyer is a function of the price of the item.

Providing results specific to individual users is particularly important because different users expect different information given the same query. Some users are willing to pay a higher price for guaranteed quality and exceptional service from sellers. There is another class of users who always look for deals and might be willing to buy from non-professional sellers or liquidation inventory or gently used inventory if they get a great deal. There are users who use specific queries to describe what they are looking for, but there are others who use generic queries to search items. For example there are users (e.g. harry potter aficionados) who type in "harry potter" to look for collectibles, books, DVDs, t-shirts and other items. In these cases they expect to see a more diverse item-set and not only the best selling DVDs and books which might be deemed more relevant. For an online marketplace with multiple formats, for the same query, one user might want to see auction items and another one might want to see fixed price items. Also someone might be interested in used items, whereas someone else in new items.

Understanding user intent is key to designing an effective ranking system in a search engine. In the absence of any explicit knowledge of user intent, search engines try and diversify results to improve user satisfaction. In such a setting, the search engine would like to trade-off relevance for diversity in the results. This is discussed in [4]. In [4], the authors also develop a set of natural axioms that a diversification system is expected to satisfy, and then show that no diversification function can satisfy all the axioms simultaneously.

In [3], the authors present a method for combing query-relevance with information-novelty. They describe a Maximal Marginal Relevance (MMR) criterion which strives to reduce redundancy while maintaining query relevance in re-ranking retrieved documents. They describe that some users may prefer to drill down on a narrow topic, and others a panoramic sampling bearing relevance to the query and hence argue that a user-tunable method like MMR would be beneficial. They measure relevance and novelty independently and provide a linear combination as a metric.

As the web world has evolved, ability of web sites to adapt to particular needs of users has also evolved. Customizable pages which give user more control over the presentation of a page in terms of templates, skins, colors or placements are becoming more commonplace. Personalized pages where the users can choose which gadgets show up on the page are also being increasingly offered by web sites. Letting the users take control of what kind of results they would like to see on a search page is also not uncommon. How Google lets users tune one of the factors in ranking is shown below in Figure 1.

As we see, in eCommerce search, the results need to be optimized along several dimensions. There has been work in the field of visualizing multi-variate data. The M-Cube: A Visualization Tool for Multi-dimensional Multimedia Databases is described in [8]. Radar Charts [9] are used in various domains where there is a need to visualize multi-dimensional data. Many video games tend to use radar charts as an aid to user input, where users can choose along multiple dimensions of player attributes or Avatar qualities.

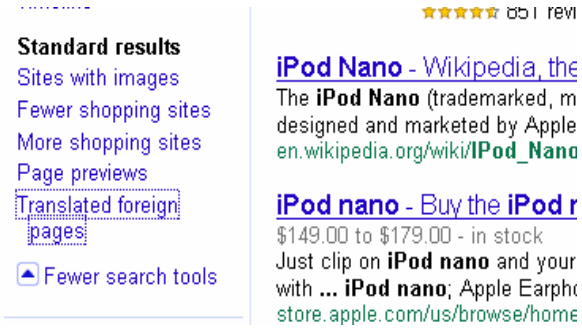

Figure 1 Portion of a screen capture showing Google Search Results for query "ipod nano". The user is given an option to tune the search results to see more or less shopping sites.

In this paper we describe a system which is motivated by the MMR [3] approach. In response to a query, the system matches eBay inventory. Then, the matches are ranked according to a linear combination metric of Relevance, Diversity, Trust and Value. However, the parameters controlling the impact of each factor are user-tunable. To the best of our knowledge, none of the leading eCommerce search engines have an option to let the users tune the ranking algorithm based on such dimensions. To let the users tune these parameters along 4 dimensions, we use an interface similar to that is used in video games when users want to select player types based on attributes along various dimensions. Computing an optimal set based on given parameters for Relevance, Diversity, Trust and Value is an intractable problem. How to efficiently compute diverse query results in structured online shopping search is described in [1]. However, eBay inventory is unstructured and we use a bounded greedy selection algorithm similar to the one described in [5] as a practical solution.

The architecture for our system is described in section 2. The algorithm powering the system is described in section 3 and the user interface for tuning parameters in section 4.

2. SYSTEM ARCHITECTURE

A high level block diagram description of our system is shown in Figure 2.

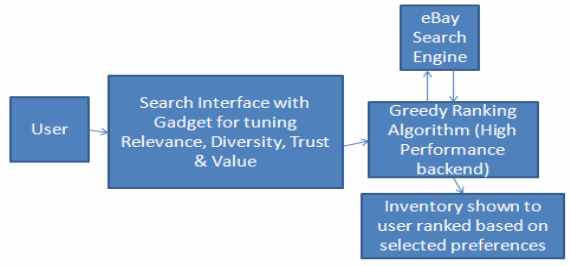

Figure 2 User interacts with a customizable search interface. The interface has a gadget that can be used to tune parameters. The parameters are sent back to backend along with the query.

[2] http://slickdeals.net/

The backend fetches a set of 2000 relevant results from eBay search engine, ranks them using the input parameters greedily and displays them back to the user.

3. ALGORITHM

We use a greedy algorithm based on a linear combination metric to sort and display inventory to the user.

Result sets are created incrementally, by adding one item at a time to the set. Incremental addition of items to the result set is done in such a way, that at each iteration a so-far unselected item $I_i \in R \setminus D$ maximizing the following criterion is selected,

$$\left(\alpha.Rlv(I_i,Q) + \beta.\left(\frac{\sum_{I_j \in D}\left(1 - Sim(I_i,I_j)\right)}{|D|} \right) + \gamma.\tau(I_i) + \delta.\upsilon(I_i) \right).$$

Denoting the universal set of all available inventory as C, $R = IR(C,Q,\Theta)$ where R is the set of items retrieved, IR is the eBay search engine's information retrieval and relevance function, Q is the query and Θ is the maximum number of items retrieved by IR for any query Q. For our experiments, we used Θ = 2000. D is the subset of items in R already chosen to be a part of the result set. $R \setminus D$ is the set difference i.e. the set of as yet unselected items in R.

$Rlv(I_i,Q)$ is the relevance of item I_i to query Q as determined by eBay's relevance ranking function and $0 \le Rlv(I_i,Q) \le 1$.

For our system, we use seller data and eBay's Feedback system[3] to compute the trust score of any item $\tau(I_i)$, where the score is normalized such that $0 \le \tau(I_i) \le 1$ for any item I_i.

$\upsilon(I_i)$ is the value proposition provided to the buyer by item I_i. If the price of an item is less than the typical selling prices of similar items then the value proposition for the item is high and vice versa. The value proposition for a particular item is computed by comparing the item's price with other similar items that were bought previously and $0 \le \upsilon(I_i) \le 1$ for any item I_i.

$Sim(I_i,I_j)$ is the function that measures the similarity between two items I_i and I_j. On an eCommerce site like eBay, the similarity could be based on various factors like:

a) Are the items being sold by the same seller? b) Are the items in the same format? Auction, Fixed Price, Classified Listing etc. c) Do the items belong to a similar product line? d) Do the items offer similar shipping options or accept similar methods of payment? e) If a taxonomy is defined for the marketplace, then do the items belong to the same node in the taxonomy tree?

For our system, we use seller, format and titles of the items to compute similarity. Specifically,

$$Sim(I_i,I_j) = \alpha' Sim_s(I_i,I_j) + \beta' Sim_f(I_i,I_j) + \left(1 - (\alpha' + \beta')\right) Sim_i(I_i,I_j)$$

$Sim_s(I_{i,},I_j) = 1$ if I_i and I_j are being sold by the same seller and $Sim_s(I_{i,},I_j) = 0$ otherwise.

$Sim_f(I_{i,},I_j) = 1$ if I_i and I_j are of the same format (fixed price, auctions, classifieds etc) and $Sim_f(I_{i,},I_j) = 0$ otherwise.

Let T be the set of terms found in the title of an item I.

$$Sim_i(I_{i,},I_j) = \frac{|T_i \cap T_j|}{|T_i \cup T_j|}.$$

We learn from human judgments and choose $\alpha' = 0.2$ & $\beta' = 0.4$.

$\alpha,\beta,\gamma,\delta$ are coefficients which control the relative importance of different factors in sorting and the user is allowed to tune these coefficients using an interface described in section 4, where $0 \le \alpha,\beta,\gamma,\delta \le 1$ and $\alpha + \beta + \gamma + \delta = 1$.

4. INTERFACE AND IMPLEMENTATION

We created a search gadget that could be used by users to specify their preferences. We create a set of pre-defined profiles which can cater to the expectations of different users. Some profiles are balanced (for average users), some are heavily focused on Value (to help the people looking for deals), some on Trust (for people expecting guaranteed service) etc. The user has the option of selecting one of these predefined profiles. The coefficients are displayed in the form of a radar chart [9] as shown in Figure 3.

For users, who prefer not to use any of the pre-defined profiles, we also provide a custom interface based on radar charts [9] inspired by gaming interfaces similar to the one found in the FIFA Soccer Game made by Electronic Arts for the PS3[4]. For the custom interface, the user gets 100 points and can spend those points on any of the dimensions she is interested in, in any proportions she is comfortable with, thus leaving full control with the end user. A scenario in which a user is in the process of creating a custom profile to sort on using horizontal sliders is shown in Figure 4.

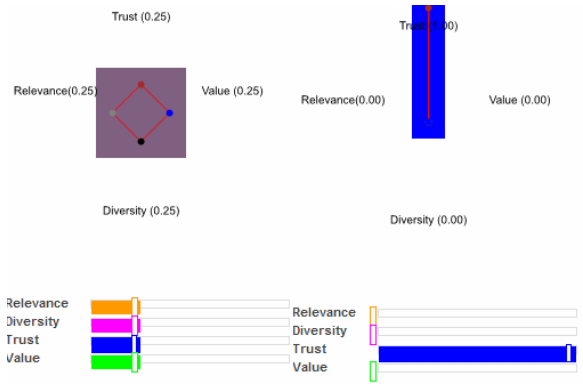

Figure 3 A gadget with pre-defined profiles. Profile on the left lets the user select a balanced ranking function which gives equal importance to Relevance, Diversity, Trust and Value. The one on the right emphasizes only on Trust.

[3] http://pages.ebay.com/services/forum/feedback.html

[4] http://www.ea.com/soccer/fifa/ps3

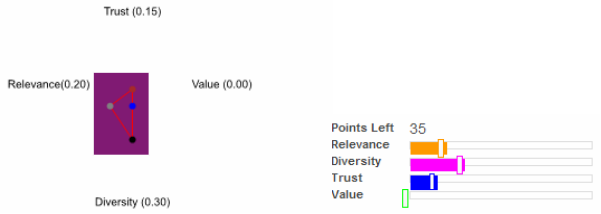

Figure 4 A custom profile designed by a user. The user has chosen to use 15 of the allotted points on Trust, 20 on Relevance and 30 on Diversity. She still has 35 points left to add to the dimension of her choice. Top portion of gadget is shown on left and bottom portion on right.

5. DEPLOYMENT AND EVALUATION

We deployed this system at http://labs.ebay.com/searchstudio/. We invited eBay users to use the system for their buying needs. The system was found to be useful for advanced searching needs. For exploratory users and users focused on domains like Collectibles, being able to see more diverse items was useful. For users, who were looking for a guaranteed professional service, fast delivery and hassle free buying, being able to tune heavily on Trust was appealing. For users trying to hunt for deals, being able to tune Value heavily was happily received. In general, we observed that regular eBay users were pleased to see an option which allowed them to have more control over what inventory set they were more likely to see on the search page.

6. CONCLUSIONS AND FUTURE WORK

We have demonstrated a search system where the user can tune the parameters used for optimizing the ranking and can hence have more control over the Relevance, Diversity, Trust and Value associated with the result set.

As part of future work we would like to analyze and quantify the impact of such a system on lifetime customer value and Net Promoter Score[5]. We would also like to analyze and see if the data gathered from usage patterns for this system by different segments of users can be used to improve the quality of personalization features.

[5] http://en.wikipedia.org/wiki/Net_Promoter

7. ACKNOWLEDGMENTS

We are thankful to Jack Shen and Long Hoang for helping with this research.

8. REFERENCES

[1] Vee, E., Srivastava, U., Shanmugasundaram, J., Bhat, P., Yahia, S.A. 2008. Efficient Computation of Diverse Query Results. ICDE (2008), 228-236.

[2] Sundaresan N. 2007. Online Trust and Reputation Systems. ACM Electronic Commerce (2007), 366-367.

[3] Carbonell J., Goldstein J. 1998. The use of MMR, Diversity-Based Re-ranking for Reordering Documents and Producing Summaries. SIGIR (1998), 335-336.

[4] Gollapudi S., Sharma A. An Axiomatic Approach for Result Diversification. Proceedings of the 18th international conference on WWW(2009). 381-390.

[5] Bradley K., Smyth B. Improving Recommendation Diversity. Proceedings of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science (2001). 85-94.

[6] Alonso O., Mizzaro S. Relevance Criteria for E-Commerce: A Crowd sourcing-based Experimental Analysis. SIGIR (2009). 760-761.

[7] Mizzaro S. Relevance: The whole history. JASIS 48(9) (1997). 810-832.

[8] Maximo A., Sabab M., Velho L. M-Cube: A Visualization Tool for Multi-dimensional Multimedia databases. Proceedings of Interaction (2009).

[9] Wikipedia. Radar Chart. http://en.wikipedia.org/wiki/Radar_chart

[10] Whittaker R. 1972. Evolution and Measurement of Species Diversity. Taxon. 21, 213-51