Communicating Machine Learned Choices to E-commerce Users

N Seshadri, G Singh, J House, M Nathan, N Parikh
EBay Inc.
{nseshadri, gysingh, juhouse, munathan, nparikh}@ebay.com

Abstract

When a user researches a product on eBay's marketplace, the number of options available often overwhelms that user's capacity to evaluate and reach a purchase decision with confidence. To ease the user experience and to maximize the value a buyer gets from our marketplace, we have machine learned filters – each expressed as a set of attributes and values, that we hypothesize will help in framing and advancing their purchase journey. The introduction of filters to simplify and shortcut the customer journey presents challenges with the UX design. Here are our findings from the user experience research and how we have integrated research findings into our models.

INTRODUCTION

eBay is a marketplace where sellers list their products and buyers shop. eBay facilitates this exchange but does not sell or buy on its own inventory.

In typical retail purchases, decisions shoppers make tradeoffs between products with distinct feature sets. As a marketplace, eBay shoppers have to evaluate both product-need fit AND ALSO evaluate tradeoffs between offers (item condition, seller service level, etc.). This extra step might yield additional value vs standard retail outlets to shoppers who find an offer that best fits their need and budget, but adds friction to all shopping journeys, including those of shoppers who are less price-sensitive and more sensitive to incremental shopping friction/time/effort.

Product listings (aka item offers) can be filtered different ways on eBay's marketplace for example by condition. Condition filtering is valuable for products with inventory depth across multiple conditions (Fig#1). However, this approach does not work for many products in eBay's long-tail catalog. Some products only have live listings in one or two conditions. Sometimes users prefer to buy only in one condition, and evaluating items of other conditions is unnecessary. In either scenario, presenting a condition-based filter is of little value.

One approach to surface more valuable filters is to generate filters on attributes with observed value given a specific product and inventory context. After we generate filters, we uncover challenges communicating them in the user experience and attempt to address those challenges.

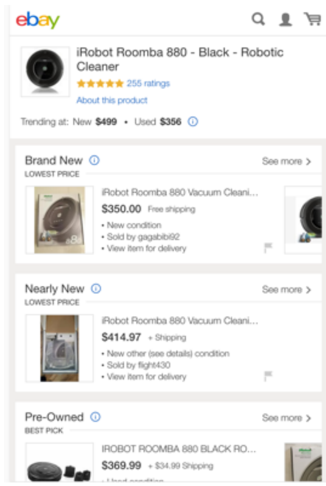

Fig 1. Grouping by conditions

Machine Learned (ML) Filters

We have analyzed historical transactional (past purchase), and behavioral data of eBay marketplace shoppers and identified that there is inherent value placed by users for different attributes associated with the products and listings. These attributes themselves can be classified into global or local.

Global attributes are those that are common across products, such as condition, selling format (fixed price or auction), etc. Through our data, we identified that users' preference for these attributes change from product to product/ eBay has captured some of these preferences through our recommendations on our seller's listing flow as well as buyer's search patterns. Our data shows that these preferences have implicit value associated that can be exploited to improve relevance and framing of listings that we highlight to buyers.

Local attributes are specific to a small subset of products that are either replace a global attribute (for example "rating" in baseball cards as a substitute for "condition"), or attributes that are specific to a subset of products (for example the OS version on an Android phone). Many of these attributes are unique to eBay's listings and have potential to provide differentiation value if presented as filters.

Relative Value

We define the term relative value of an attribute for a product as the value that buyers are willing to pay to purchase a listing having that attribute vs an otherwise identical listing. We have measured this value from our purchase history of a product by comparing price between two groups of listings of the same product – one with the attribute in consideration and the second group without.

The value difference between listings of a product for an attribute a1 is

$$R(a1) = (P(a1,v1) - P(a1,v2)) * 100 / (P(a1,v1)$$

Where,

a1 is attribute for which relative value needs to be computed

v1, v2 values of attributes compared

R is relative value

P is a function of price (ex: median price)

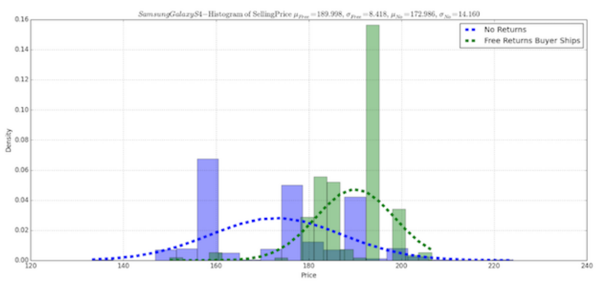

Fig 2. Histogram of Price of Galaxy S4 sold on eBay

In fig 2, there was a price difference between a returnable Samsung Galaxy S4 and a non-returnable one. Given the history of transactions for both groups, we conclude that relative value could be of interest to users in their decision making process.

Behavioral Signals

Behavior analysis of users in their purchase flow generates a new set of signals with potential value in shopper decision making process. We analyzed shopper journeys with click/view trails and identified shopping patterns in which specific attributes are consistent or "sticky", and other attributes commonly occurring when the customer makes a quick, "impulsive" decision to purchase. We hypothesize that we may be able to amplify these behaviors by highlighting the sticky and impulsive attributes.

Stickiness:

- Attributes for which buyers stick to a specific value or range significantly more often to than a random chance would dictate.

Impulsive:

- Attributes that significantly correlated with impulsive transactions (Short view trail before purchase).

Attribute Selection for Filters

Our analysis was done for the cataloged products (with at least one valid identifier such as UPC) in our marketplace. We looked into six months of purchase history as well user's behavioral click stream data for millions of products across multiple countries including US, UK, Australia and Germany.

First, outlier data such as misclassified listings had to be removed. Our early versions of outlier detection were using standard statistical techniques including standard deviation, mean and median price. Later, we picked the Median Absolute Deviation(MAD) based outlier detection.

We considered nineteen different global attributes such as condition, sale type, seller ratings, etc. For each attribute in our consideration set, we evaluated relative value for each product. For higher precision, we applied high thresholds for data sufficiency – presence of sufficient data for each attribute within each product that we considered. A similar approach was applied on behavioral data to identify attributes that are sticky and/or impulsive for each product.

Behavior signals complement our relative value based approach when i) relative value is not evident for an attribute but that attribute is important from behavioral perspective or ii) insufficient transactional data to compute relative value with confidence.

Next step after shortlisting attributes per product is grouping to use one or more attributes in constructing a filter. Some filters are simple reflection of the attributes. For example, "charity" is a simple attribute filter that is on listings where the proceeds go towards a charity. There are also complex filters that are defined by a group of attributes. For example, a filter of "free returns" is a collection of three different attributes - return availability, restocking fee and shipping.

SPECTRUM OF CHOICE

Value Filters

We have identified attributes of importance – either observed through relative value or through behavioral signals. Our hypothesis is that presenting these attributes in some form of filters would enable users to get to the listing that best meets their needs quicker. The next big question we had to answer is how do we use this information to improve the user experience? We identified a list of UX-specific challenges:

1. Filter Naming: How do we communicate filters as understandable and compelling Filter Titles?

2. Filter Overlap: How do we communicate that filter results are not mutually exclusive?
3. Filter Heterogeneity: How do we communicate why eBay is displaying unrelated filter sets in close proximity?

Filter Naming

ML identified filters can include one or more attributes (global or local) and we need to identify each of these filter with a human relatable name. For example, for products where users prefer buying new condition and want to have the flexibility of returns and are weary of overseas shipping, we titled a theme intended to meet their needs "Hassel Free". One must note that such a name needs to consider the context – presence of other filters and their names as well. Hence, filters shown to the users should not be selected independently but as a set.

Filter Overlap

ML based recommendation systems are the norm for online marketplaces like eBay. Sophisticated models that ML practitioners would like to present should be easy to communicate to the users. Early engagement of user studies is one way to understand the challenges ML system would have to encounter. We at eBay learnt through user studies and incorporated changes to our approach in identifying, ranking the smart filters for our service to increase user engagement.

Filter Heterogeneity

Buyers are used to comparative view – new vs used, returnable vs non-returnable, etc. as it usually fits their natural shopping framework. They can place value on the choices that they are looking at to make tradeoffs easily. In the case of intelligent filters identified, the most valuable mix of filters are not always homogenous. For instance, our data could suggest that users who are interested in brand new items do not care about the sellers' rating or item origin. However, they may place more value on these attributes when considering used items. The items framed by these filters are not easy to directly compare. In this case, the value based smart filters may make the decision process more complex.

USER EXPERIENCE RESEARCH

Design Constraints

Our choices of filters are impacted by platform. In the mobile devices, we present a vertical list of filters with the top pick for each filter. We also enable an easy to browse selection of listings viewable on swipe on mobile devices. However, on desktops, additional real estate leads to more elements competing for the users' attention putting additional burden on the choice and presentation of filters.

We have chosen to show three filters, vertically stacked in mobile and horizontally arranged on desktop. Filters were ranked based on a number of factors including relative value and engagement metrics.

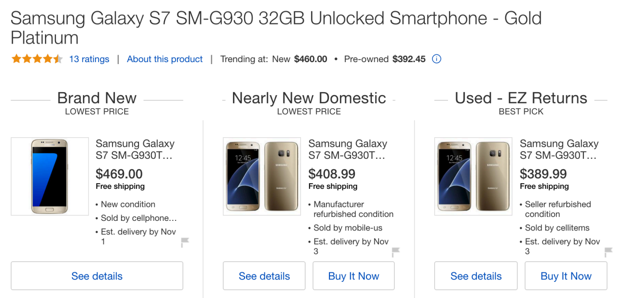

Fig 3. Filters in a desktop setting

We have conducted a series of user experience research to understand how the users react to the filters that we present. The study goals were aligned to the challenges that we have identified earlier.

Study Findings- Filter Names

The first steps in helping users find their perfect item is to translate the features chosen by our model into titles describing i) what filter was applied and ii) why it's relevant to the buyer. Due to mobile space constraints and need to translate into more verbose languages, we had 22 characters with which to craft each title in English. We developed themes in coordination with many functions (content, marketing, legal, etc.) and polled sellers to collect their feedback on filters and titles.

We then iterated on titles over five rounds of usability research in lab in US, UK, DE, and AU. We tested a variety of titles along a spectrum from "emotive and engaging" (example A) to "simple and descriptive" (example B).

We discovered:

A. Users overwhelmingly preferred simple titles
B. Item condition was the first reference frame most users locked onto
C. Users found longer titles, especially those with compound filters, were difficult to understand

As indicated in Example B, we landed on a launch design that broke up titles into two lines, so we have room to indicate both filter and sort.

Example A: "engaging" (titles in red)

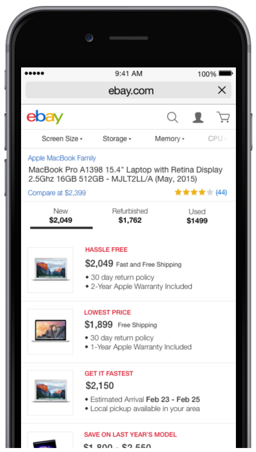

Fig 4. UER – Filter naming (A)

Example B: "descriptive" titles (split over two lines)

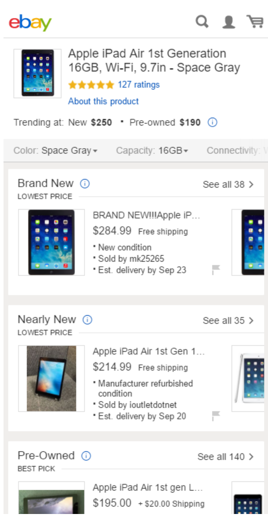

Fig 5. UER – Filter naming (B)

Study Findings – Filter Heterogeneity

Our next study was to find how well users react to the filters with name closely tied to the attributes in that filter. As in fig 5, we presented filters with the attributes that eBay users are familiar with from their past experience with our marketplace.

We observed:

A. Users found difficulty to find the condition of the listings presented
B. Users found these mixed set of filters to be confusing as they couldn't compare against each other
C. Users had concerns about the purpose of the filters, i.e., that these filters could be tied to advertising or promotion

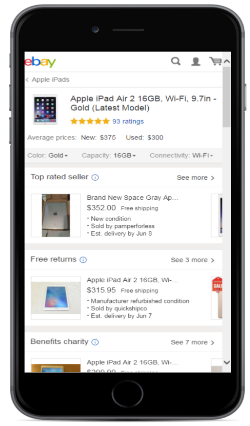

Fig 6. UER – Filter mix

NEXT STEPS

One overarching theme from our studies is that our users would like to see listings by condition. We have changed our filters to be grouped by condition. We are exploring options of mixing condition based and ML driven value based filters as well.

We have added navigation links to explore the entire inventory and has helped us gain users trust in our recommendations. This is an easy way of accessing listings not covered by filters presented.

Personalized selection of filters should help us present even more relevant listings to save our buyers time and alleviate some of the concerns that our in-lab shopping experiment participants expressed.

CONCLUSION

ML based recommendation systems are the norm for online marketplaces like eBay. Sophisticated models that ML practitioners would like to present should be easy to communicate to the users. Early engagement of user studies is one way to understand the challenges the ML system will encounter. eBay has learned through user studies and incorporated changes to our approach in identifying, ranking the smart filters for our service to increase user engagement.

References

Dong, T., Churchill, E.;, and Nichols, J. DIS. 2016. Understanding the Challenges of Designing and Developing Multi-Device Experiences.

Schwartz, B; *The paradox of Choice: Why More is Less*, 2005, Harper Perennial