The AAAI 2017 Spring Symposium on
Designing the User Experience of Machine Learning Systems
Technical Report SS-17-04

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 shopper researches a product on eBay's marketplace, the number of options available often overwhelms their capacity to evaluate and confidently decide which item to purchase. To simplify the user experience and to maximize the value our marketplace delivers to shoppers, we have machine learned filters—each expressed as a set of attributes and values—that we hypothesize will help frame and advance their purchase journey. The introduction of filters to simplify and shortcut the customer decision journey presents challenges with the UX design. In this paper we share findings from the user experience research and how we integrated research findings into our models.

Introduction

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

In typical retail purchase decisions, shoppers make tradeoffs between products with distinct feature sets. As a marketplace, eBay shoppers have to evaluate both product feature sets AND ALSO evaluate tradeoffs between offers (item condition, ship cost, seller quality, 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, even those of shoppers who would pay more to avoid the friction/time/effort of evaluating offer tradeoffs.

Product listings (aka item offers) are typically discovered through search on eBay. Search results can be filtered in different ways for example by condition (Fig#0). Our area of interest, however, is a new experience on eBay: a Product Page that represents inventory from a single product. Within this more constrained context, condition filtering is valuable when inventory exists for multiple conditions (Fig#1). However, condition filtering is less valuable when i) inventory is thin, (e.g. in the "long tail" of eBay's catalog), or ii) a shopper has given eBay a hint that she is only interested in items from one condition.

One approach to surface the most valuable filters in context is to generate filters on attributes where we can observe value from historic data, given a specific product and inventory context. Upon generating these filters, we uncovered challenges communicating them in the user experience and attempted to address those challenges.

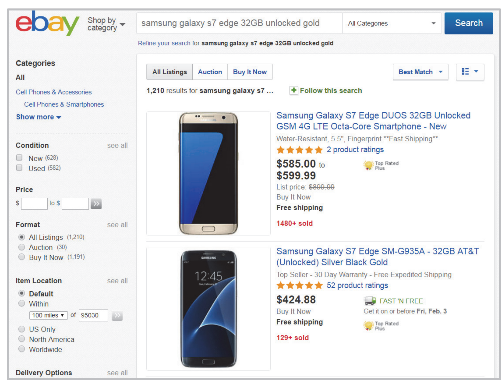


Fig 0. Condition filters on desktop Search Results Page

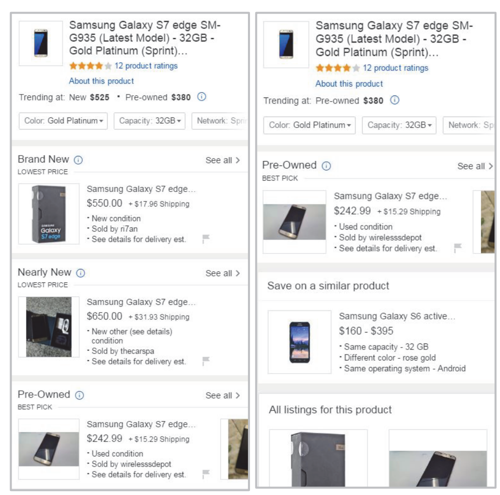


Fig 1. Condition grouping on mobile Product Pages with both deep (left) and shallow (right) inventory availability

Machine Learned (ML) Filters

We 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. We classified the attributes themselves 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 user preference for these attributes changes from product to product. We captured some of these preferences by examining the common attributes sellers include when listing an offer for a product that successfully sells. We also identified important attributes based on 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 because they modify products post-manufacturing, and historic data indicates high 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 a second group without the attribute.

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 is computed
v1, v2 values of attributes compared
R is relative value
P is a function of price (ex: median price)

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.

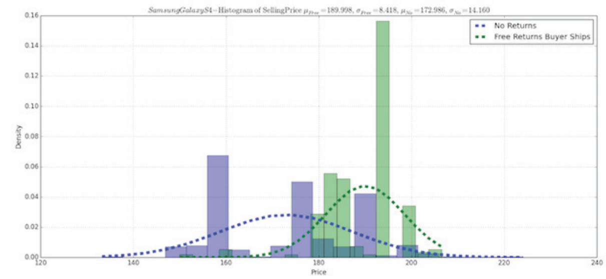


Fig 2. Histogram of Price of Galaxy S4 sold on eBay

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 in their purchase journey 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 opted for Median Absolute Deviation (MAD) based outlier detection.

We considered nineteen different global attributes such as condition, sale type, seller rating, return acceptance, etc. For each attribute in our consideration set, we evaluated relative value for each product. For higher precision, we applied conservative 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 is available to compute relative value with confidence.

Our next step after shortlisting attributes per product was to consider grouping one or more attributes in constructing a single 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. Other filters are more complex and are defined by a group of attributes. For example, a filter of "free returns" is a collection of three different attributes - return availability, zero restocking fee and free return shipping.

Spectrum of Choice

Value Filters

Having identified attributes of importance – either observed through relative value or through behavioral signals—our hypothesis was 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 are constrained by the 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 cohesive 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 employment 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. because they find it overwhelming to compare values for a multiple dimension at the same time. Instead, buyers typically compare values for a single dimension at a time and then aggregate results. In this way, they can place value on the each dimension and resolve tradeoffs to a single winning choice. 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 choice of filters is impacted by eBay's broader platform experience. In our first design launched to 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, which puts additional burden on the choice and presentation of filters.

Within this context, we chose 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.

We 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 identified earlier.

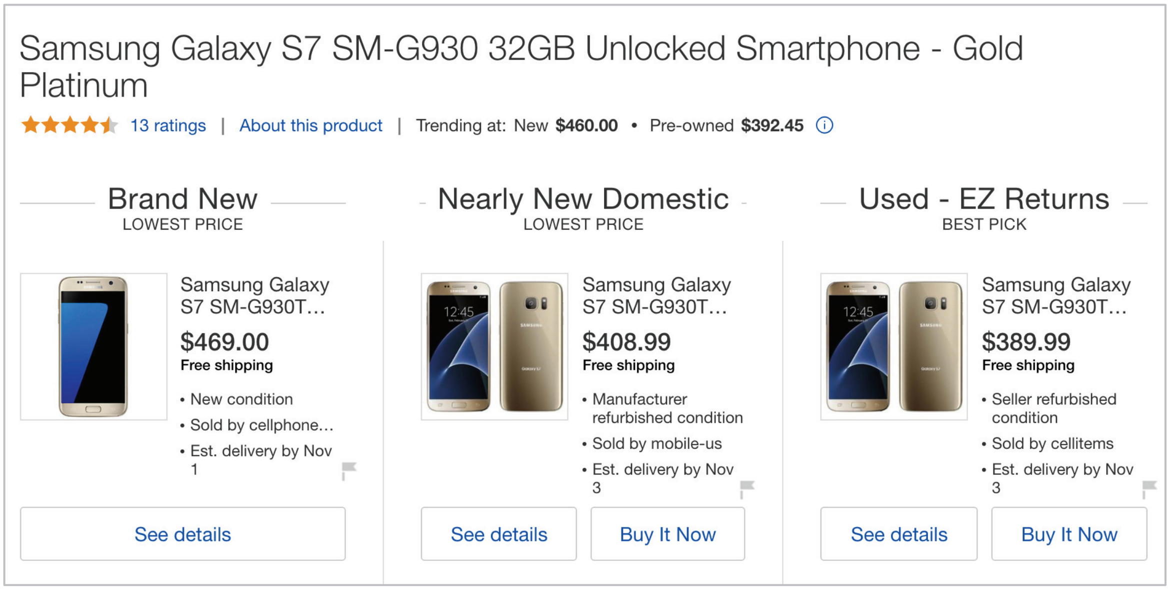
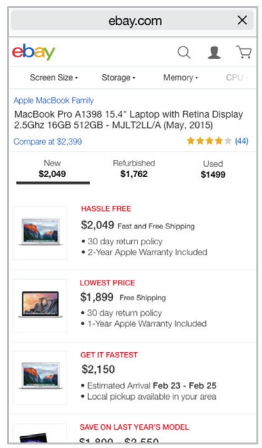


Fig 3. Compound filter titles in desktop setting



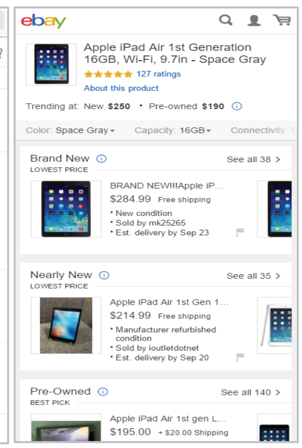


Fig 4. Filter naming (A) *Fig 5. Filter naming (B)*

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 our need to translate filter titles into languages with longer words, 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)
Example B: "descriptive" titles (split over two lines)

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 at a glance the inventory presented in one theme to inventory in the next

C. Users had concerns about the purpose of the filters, i.e., that these filters could be tied to advertising or promotion

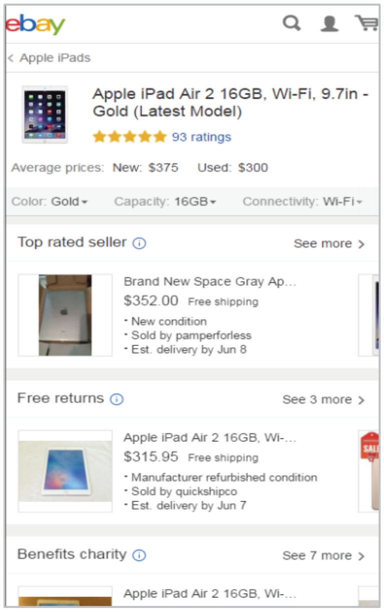


Fig 6. UER – Filter mix to gauge tolerance for non-heterogeneity

Next Steps

Filter Hierarchy

One overarching theme from our studies is that our users would like to see listings framed by condition above other attributes. Based on this finding, we have changed our filters to group by condition and are exploring options of mixing condition attributes with secondary ML value driven attributes as well.

Inventory Coverage

User study participants also expressed low confidence in our recommendations when the inventory covered using ML filters was smaller than that of search results. For example, when the value based filters are concentrated on one or two attributes, significant inventory may be left out. We addressed this concern by taking inventory coverage into consideration in our ML research.

Filter Ranking

Our initial approach of ranking of the filters was based on relative values supported by human curation. However, the user study findings lead us to include additional metrics such as inventory coverage and click through rate in our ranking algorithm.

Alternative Experience

We also added navigation links for shoppers to explore the entire inventory beyond our recommendation, which has helped us gain users trust in our recommendations. These links to "see all inventory" also provide easy access to listings not highlighted by our filter-sort formula, in support of cases where a shopper's 'version of perfect' went undetected by our analysis.

Personalization

In the future, we will personalize the selection of filters to present even more relevant listings, save our buyers time and alleviate some of the concerns that participants of our in-lab shopping experiment 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.

Acknowledgement

Dan Fain, David Goldberg, Victor Chong and project designers and developers.

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

Christophe Leys, Christophe Ley, Olivier Klein, Philippe Bernard, Laurent Licata; *Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median, Journal of Experimental Social Psychology, Volume 49, Issue 4, July 2013*