Visual Intents vs. Clicks, Likes, and Purchases in E-commerce

Riku Togashi
Waseda University, Mercari, Inc.
Tokyo, Japan
rtogashi@acm.org

Tetsuya Sakai
Waseda University
Tokyo, Japan
tetsuyasakai@acm.org

ABSTRACT

In product-to-product search and recommendation, the product image often plays a pivotal role for the user to determine the relevance of that product. The present study investigates the relationship between the users’ visual intents (in terms of colour, texture and material, and design) and the amount of user feedback (namely, clicks, likes, and purchases) using real product data and crowdsourcing. Through the analysis, we found that visual relevance (i.e., relevance of a target product with respect to a particular visual intent) correlates with the amount of user feedback, and that visual relevance can be the cause of user feedback.

CCS CONCEPTS

• **Information systems** → **Information retrieval**; **Query intent**; **Image search**; *Recommender systems*.

KEYWORDS

click behaviour; e-commerce; user feedback; visual intent

ACM Reference Format:
Riku Togashi and Tetsuya Sakai. 2020. Visual Intents vs. Clicks, Likes, and Purchases in E-commerce. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ’20), July 25–30, 2020, Virtual Event, China.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3397271.3401293

1 INTRODUCTION

In e-commerce, a product data record consists of a product name, category, description, price, product image, and so on. Among these components, the product image plays a pivotal role when the user determines the relevance of a product to their needs. Since the user’s need may be expressed in terms of a product image they have selected, the challenge for e-commerce sites is to determine the relevance of target product images given a product image query. That is, to provide effective product-to-product image search.

Due to the richness of a product image query, however, it is generally difficult for a search engine to understand the user’s intent behind it. The user intent may reflect specific visual requirements, such as “I prefer the colour of this hat over the other” and “I like the design of this product but not this particular colour.” We shall refer to these viewpoints such as colour and design as *visual intents*, and the relevance of a target product to the product query with respect to a particular visual intent as *visual relevance*. Understanding visual relevance is particularly important for product categories such as fashion.

The present study explores the relationship between visual relevance and user feedback (namely, clicks, likes and purchases) in the context of product-to-product image search in e-commerce. To this end, we collect user feedback from search logs of an e-commerce site, as well as visual relevance assessments from crowdsourcing.

ACM ISBN 978-1-4503-8016-4/20/07…$15.00
https://doi.org/10.1145/3397271.3401293

2 RELATED WORK

Analysis of search query logs has provided significant insights for understanding user behaviour in e-commerce search and recommendation. Sondhi et al. [7] proposed a taxonomy of product queries by analysing e-commerce search logs and demonstrated that queries fall under five categories. Each category is associated with a distinct user search behaviour. Carmel et al. [1] analysed query logs of voice assistant supporting shopping, namely, voice product search. Users purchase seemingly irrelevant items in exploration mode, given irrelevant offers based on past-purchases or on popularity.

Recently, visual information has been leveraged to optimise product recommender in e-commerce. Shanker et al. [6] proposed a visual search and recommender system for e-commerce based on image retrieval. Whereas the method proposed by Shanker et al. does not leverage any resource except product images, Chen et al. [2] proposed an explainable visual recommendation model by incorporating implicit feedback and textural reviews. McAuley et al.[4] proposed a method based on deep learning to simultaneously deal with the textual and visual information for heterogeneous graph embedding.

To the best of our knowledge, the present study is the first to investigate the relationship between visual relevance and user feedback (i.e., clicks, likes, and purchases) in the context of product-to-product image search in e-commerce.

3 DATA AND EXPERIMENTAL SETTINGS

3.1 Definition of Visual Intents

This study focuses on the fashion category as we expect visual intents to play a major role in product-to-product image search. We consider the following three visual intents for determining the relevance of a target product image to a given product image query:

Colour We expect colour to be a critical factor for the e-commerce user to determine the relevance of a recommended product to the query product image.

Texture and Material Texture refers to patterns on the product surfaces, such as striped, dotted, and plaid. Material refers

Figure 1: Query product image and target products.

to whether the product surface is leather, knit, metal, etc. but this is also considered as texture in image processing.

Design This refers to detailed features of each product type, such as tops, bottoms, hats, wallets, bags, and accessories. For example, for the tops category, design may refer to sub-categories such as short sleeved T-shirts.

We assume that the relevance of a target product image with respect to each visual intent to be binary. Hence, with the above three visual intents, the overall relevance of a target image can be considered to be on a 4-point scale.

3.2 Collecting User Feedback

To investigate the relationship between visual relevance and user feedback (i.e., clicks, likes, and purchases), we first collected real user feedback using a product-to-product search engine of a real e-commerce website, Mercari[1]. Figure 1 shows the user interface of the search engine: given a query product, the search engine result page (SERP) is returned, which displays target product images in a three-column grid design with infinite scrolling. By clicking on one of the target images, the user can access a detailed product page, where they can provide further feedback (i.e., likes and purchases).

Clearly, whether the user provides feedback on a target product depends on what the search engine returns. To minimise this *presentation bias*, we simply ranked the target products by freshness when returning the SERPs to the user. Moreover, as we are interested in product-to-product search within the same product category, we filtered out target products that belong to a category different from that of the query product.

From the search log, we randomly selected query-target image pairs while ensuring that each target image was displayed in the SERP for the query at least once. As a result, we obtained 40,962 query-target image pairs with clicks, and 8,900 pairs without clicks.

3.3 Collecting Relevance Assessments

To explore the relationship between user feedback and visual relevance, we collected visual relevance assessments for the query-target image pairs described above using crowdsourcing[2]. Each crowd worker was given a query image and a target image, and was asked to provide a binary relevance assessment for each visual intent, namely, colour, texture/material, and design. 513 workers participated, and each image pair was judged independently by three assessors. The final visual relevance for each intent was determined by a majority vote. Finally, a 4-point scale overall graded relevance was assigned to each image pair based on the number of relevant visual intents. Thus, we have obtained a dataset comprising $40,962 + 8,900 = 49,862$ query-target product image pairs with both user feedback and visual relevance assessments.

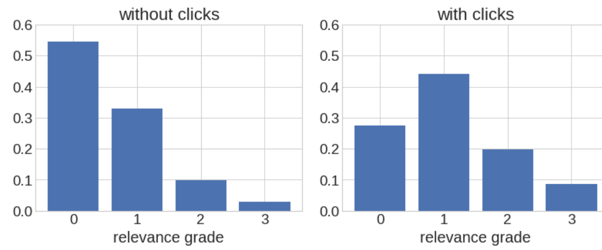

Figure 2: Distributions of visual graded relevance of judged product pairs with/without clicks

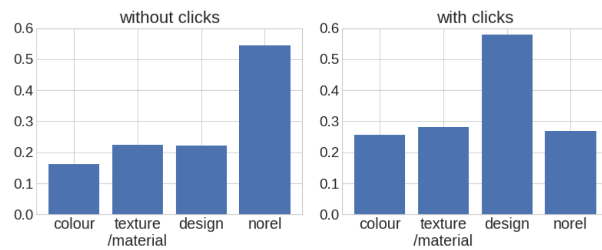

Figure 3: Distributions of visual intents of judged product pairs with/without clicks

4 VISUAL RELEVANCE VS. USER FEEDBACK

4.1 Correlation between Visual Relevance and User Feedback

First, we verified our hypothesis that clicked products tend to be more relevant than non-clicked ones. Figure 2 shows the distribution of the overall visual relevance grades for target products with and without clicks, respectively, which confirms our hypothesis. The mean relevance grade over the clicked products is 1.09; that over the non-clicked products is 0.609. The difference is statistically significant according to a Brunner-Munzel test ($p < 2e - 16, \rho = 0.655$).

Figure 3 also verifies the above hypothesis but by looking at the distribution of products over visual intents rather than relevance grades. Note that while the "norel" (nonrelevant) products and visually relevant products are mutually exclusive, the three visual intents are not; that is, a product can be relevant with respect to multiple intents. It can be observed that while "norel" is the major category for products without clicks, design is the most frequent visual intent for products with clicks.

In summary, clicked products do tend to be more relevant than non-relevant ones, and design is the prominent visual intent for the former.

We then considered a more general hypothesis that relevant products receive more user feedback (clicks, likes, and purchases)

[1] https://www.mercari.com/jp
[2] https://crowdsourcing.yahoo.co.jp/

Table 1: Effect of visual relevance on the amount of user feedback. The columns of 'Relevant'and 'Nonrelevant' are the average rate of user feedback. Each statistical test is conducted through an unpaired t-test.

User Feedback	Relevant	Nonrelevant	p-value	Hedge's g
Clicks	.614	.337	$< 2e-16$	.597
Likes	.105	.074	.016	.028
Purchases	$2.19e-3$	$9.73e-4$	.011	.027

than nonrelevant ones. we first obtained a balanced data set containing 8,900 products with clicks (randomly sampled from the aforementioned 40,962 products) and the 8,900 products without clicks. The $2 * 8,900 = 17,800$ products were then divided into relevant (to at least one visual intent) and nonrelevant, and the proportion of clicked products was computed for each group. The "clicked" row of Table 1 shows the results: it can be observed that relevant products are clicked substantially more often than nonclicked ones. We further examined the entire set of 40,962 clicked products as follows. First, from this set, we extracted all clicked products that were liked by the user that were liked by the user; this resulted in 1,760 products. This set was again divided into relevant and nonrelevant, and the proportion of liked products was computed for each group. The "Likes" row of Table 1 shows the results: again, relevant products are liked more often than nonrelevant ones, and the difference is statistically significant. Second, we also extracted purchased products from the 40,962 clicked ones; this resulted in 76 products. The "Purchases" row of Table 1 shows the proportion of purchased products for relevant and nonrelevant groups. It can be observed that relevant products are purchased more often than nonrelevant ones, and the difference is statistically significant. The above results show that visual relevance is positively correlated with clicks, likes, and purchases. Therefore, if we want to predict user feedback such as above for a given target product, modelling visual relevance is probably beneficial.

4.2 Predicting User Clicks Using Visual Relevance

In Section 4.1, we showed that visual relevance is positively correlated with user feedback, and argued that modelling visual relevance may be beneficial for predicting user feedback. In this section, we verify this hypothesis by focusing on click prediction. For our click prediction experiment, we leverage the 49,862 (40,962 clicked and 8,900 non-clicked) products described in Section 3.2 and perform an 8-fold cross-validation with 43,629 products as the training data (and 20% of it as validation data), with LightGBM [3] to construct the model. The base features we use for click prediction are (1) the price of the query product divided by that of the target product; (2) the price of the query product subtracted from that of the target product; (3) the status (sold out or on sale) of the query product; (4) whether the brand names of the query and target product are the same; and (5) the id of the category of the query and target product. In addition, to see if visual relevance information helps the click prediction task, we utilise the aforementioned manual visual relevance assessments as features. Note that, in practice, visual relevance assessments for target products are not available, and they need to be automatically estimated; the present study explores the best-case scenario where the estimated visual relevance for each visual intent is perfectly accurate.

Table 2: Effect of visual relevance on the mean of AUC in the prediction of users' click. The mean AUC is computed through 8-fold cross-validation.

Features	Mean AUC
Base	.9815
Colour	.9818
Texture/Material	.9816
Design	.9816
Colour + Texture/Material	.9818
Colour + Design	.9818
Texture/Material + Design	.9818
Colour + Texture/Material + Design	.9821

Table 2 shows the click prediction performances of LightGBM using different combinations of features, in terms of Mean AUC over the eight folds. It can be observed that the visual relevance features enhances the click prediction accuracy and that using all three visual relevance features on top of the base features achieves the best result. These results suggest that a good click prediction model should explicitly model visual features such as colour, texture/material and design. How to automatically estimate each visual relevance is left for future work.

4.3 Causal Effect of Visual Relevance on User Feedback

In this section, we investigate the causal effect of visual relevance (colour, texture/material, design) on user feedback (clicks, likes, and purchases), rather than their correlation. To this end, we utilise one-to-one propensity score matching (PSM) [5]. The propensity score in our setting is the conditional probability that a product pair is relevant with respect to a visual intent if the covariates for the product pair are given. For each combination of variables corresponding to visual relevance, we model the propensity score by utilising Random Forest. For the explanatory variables (covariates), we extract the base features described in Section 4.2. Table 3 lists the results of the best-fit model incorporating the base features. The 'Variables' column indicates the combinations of visual intents, and the 'Click', 'Like', and 'Purchase' columns show the average treatment effect of the visual intents on each of the user feedback types. The average treatment effect is the difference between expected rates of user feedback in the relevant and nonrelevant groups; here, for example, the relevant group for "Colour + Texture/Material" for the "Click" column consists of clicked products that are relevant to *both* Colour and Texture/Material, and the nonrelevant group consists of clicked product that are not.

For clicked products, the average treatment effects are statistically significant for all combinations of visual features. That is, all of these combinations induce clicks. Colour, design, and texture/material are all important features (in this order), and the combination of all three results in the largest effect.

For liked products, the average treatment effects are also statistically significant for all combinations of visual features. Texture/material is the most impactful variable in this case; it even outperforms combinations of multiple visual intents.

Table 3: Average treatment effect of visual relevance on each user feedback. † and ‡ indicate $p < 0.05$ and $p < 0.01$ in the paired t-test between relevant/nonrelevant groups, respectively.

Variables	Click	Like	Purchase
Colour	$.083^{‡}$	$.019^{†}$	$.010^{†}$
Texture/Material	$.046^{‡}$	$.050^{‡}$	$-.004$
Design	$.073^{‡}$	$.027^{†}$	$.022^{†}$
Colour + Texture/Material	$.084^{‡}$	$.045^{‡}$	.002
Colour + Design	$.104^{‡}$	$.037^{‡}$	$.023^{‡}$
Texture/Material + Design	$.087^{‡}$	$.034^{‡}$	$.020^{†}$
Colour + Texture/Material + Design	$.106^{‡}$	$.038^{‡}$	$.019^{†}$

For purchased products, the average treatment effects are statistically significant except for Texture/Material and "Colour + Texture/Material." That is, interestingly, while Texture/Material was important for liked products, it is not for purchased products. Design is the most important visual intent for purchases.

These results suggest that visual relevance of the target product can be the cause of user feedback, and that different visual intents are important for increasing different user feedback types (e.g. Likes vs Purchases).

5 DIVERSITY OF VISUAL INTENTS

So far, our experiments have shown that visual relevance and the amount of user feedback are correlated, and that visual relevance can be the cause of user feedback. However, we hypothesise that different users with the same product query may have different visual intents. This section examines this hypothesis. For the analysis, we sampled 2,328 query products with at least two clicked items in the search results. The total number of sampled product pairs is 10,871. We collected a new set of per-intent relevance judgements for this experiment using the procedure described in Section 3.2. In this section, we consider all seven combinations of the three visual intents (e.g., <colour>, <colour, design>, <colour, texture/material, design>) to define *visual intent diversity* for query q in terms of entropy as follows.

$$H_q = -\sum_i p_q(i) \log p_q(i) \tag{1}$$

where i denotes the i-th combination of visual intents, and $p_q(i)$ indicates the probability of i that can be obtained by normalising the histogram of visual intents of clicked products for q. The maximum value of the entropy is 1.946 and can be given if the visual intents distribute over the seven candidates uniformly. The minimum value is 0.0 and indicates that all of the clicked target products for a query product are relevant to only one of combinations of visual intents. Figure 4 shows the distribution of the visual intent diversity. From Figure 4, it can be observed that different product queries have a wide range of visual intent diversity scores. The mean is 0.643, and $568/2,328 = 24.4\%$ of the query products have a diversity score larger than 1.0. Since the results in Section 4.3 show that product relevance induces user feedback, *and* this result shows that the users' visual intents are diverse, it follows that product-to-product image search engines should diversify their SERPs to obtain more user feedback.

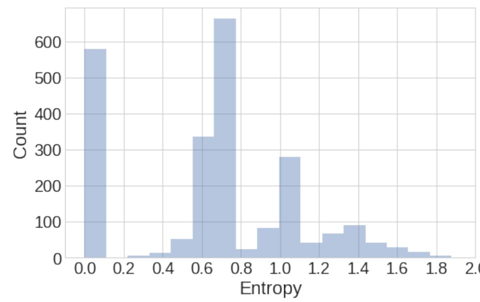

Figure 4: Distribution of the visual intent diversity for 2,328 query products

6 CONCLUSION

The present study investigated the relationship between visual intents in terms of colour, texture/material, design and user feedback in the form of clicks, likes, and purchases in the context of product-to-product image search for e-commerce. To this end, we collected real user feedback data from an e-commerce site and visual relevance assessments from crowdsourcing. Our main findings are follows.

- Visual relevance (in terms of colour, texture/material, design) positively correlates with user feedback (clicks, likes, and purchases), and visual relevance can actually cause user feedback.
- Different combinations of visual intents may lie behind the same product image query, and therefore search result diversification is probably a good strategy for maximising the amount of user feedback on the SERP.

In future work, we will consider session-level logs and consider situations that involve visual intents across multiple queries. Also, based on our click prediction results, we shall address the problem of estimating the visual relevance of a product image automatically.

REFERENCES

[1] David Carmel, Elad Haramaty, Arnon Lazerson, Liane Lewin-Eytan, and Yoelle Maarek. 2020. Why Do People Buy Seemingly Irrelevant Items in Voice Product Search? On the Relation between Product Relevance and Customer Satisfaction in eCommerce. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 79–87.

[2] Xu Chen, Hanxiong Chen, Hongteng Xu, Yongfeng Zhang, Yixin Cao, Zheng Qin, and Hongyuan Zha. 2019. Personalized Fashion Recommendation with Visual Explanations based on Multimodal Attention Network: Towards Visually Explainable Recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 765–774.

[3] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems*. 3146–3154.

[4] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. 2015. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 43–52.

[5] Paul R Rosenbaum and Donald B Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 1 (1983), 41–55.

[6] Devashish Shankar, Sujay Narumanchi, HA Ananya, Pramod Kompalli, and Krishnendu Chaudhury. 2017. Deep learning based large scale visual recommendation and search for e-commerce. *arXiv preprint arXiv:1703.02344* (2017).

[7] Parikshit Sondhi, Mohit Sharma, Pranam Kolari, and ChengXiang Zhai. 2018. A Taxonomy of Queries for E-commerce Search. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 1245–1248.