Identifying Interestingness in Fashion E-commerce using Pinterest Data

Nazneen Rajani*
Dept. of Computer Science
University of Texas at Austin
Austin, TX 78712-1188, USA
nrajani@cs.utexas.edu

Khashayar Rohanimanesh
eBay Research Labs
2145 Hamilton Ave.,
San Jose, CA, USA
krohanimanesh@ebay.com

Eduardo Oliveira
eBay Research Labs
2145 Hamilton Ave.,
San Jose, CA, USA
eoliveira@ebay.com

ABSTRACT

Fashion is the fastest growing category in online shopping. However, research in finding "interestingness" in online fashion has been restricted to the domain of computer vision. This paper aims to address the effectiveness of using a product's textual description for showcasing fashion in terms of its attractiveness, i.e. the ability to draw consumer's attention, interest, and in turn their engagement. Pinterest is an online social media platform that allows users to pin products on their boards. Past research on Pinterest users shows that majority of these are things that the user "wants" or "needs". Thus it is reasonable to assume that all pins related to fashion on Pinterest are "interesting". Our system uses this assumption to discover interesting fashion items on a popular e-commerce website, eBay. We propose a data driven approach for extracting interestingness based on textual data from pins and applying it to eBay items. We evaluate our results using crowdsourcing by comparing them to Pinterest. We obtain an inter annotator agreement of $\kappa = 0.43$ for evaluating our classifier's performance on women's shoes. Evaluating the results using topic model analysis shows that *heels, wedges, converse, leather shoes, platform, pump* and *nike running shoes* are ranked highly positive for interestingness by our classifier.

1. INTRODUCTION

Popular e-commerce platforms like Amazon and eBay sell millions of products of different brands across various categories. With no physical products to inspect, consumer's are forced to decide based purely on the description and pictures provided [14]. The clutter and risk involved does not just make the consumer's decision difficult but may also put him/her off. Thus it is extremely important for online market places to ease the consumer's decision by linking him/her to the item he/or she wants or needs. Traditional recommender systems have been able to address this problem to a certain extent. To enhance these systems, few trust-aware recommendation methods have also been proposed in the recent past [2, 15, 16, 17, 20]. However, personalized recommender systems have certain limitations. The important one is restricting the user to choices that are similar to his/her's past choices or to those of the user's social circle. The question is if it is possible to find a middle ground between helping individuals make good choices and personalized recommendations that narrow an individual's choices. One possible solution is recommending products that a lot of people find interesting. As pointed out by Silberschatz and others, interestingness of discovered patterns are dependent on objective measures - those that depend on the structure of the pattern, as well as subjective measures - those that depend on the user examining the pattern [22]. An item that is attractive to someone may not be liked by someone else. Even though there are certain attributes and qualities that may make a product attractive to individuals, discovering interesting items in the fashion category is difficult. Interestingness in fashion is obscure and not well defined. In this paper we explore the possibility of finding objective measures of interestingness for items in the fashion category sold on eBay and evaluate using a subjective measure as discussed in [22].

Pinterest is an online social media platform that allows users to share images along with descriptions. On Pinterest, individuals can pin images from any websites or local files and repin or like images from other Pinterest users' collections. Pinterest users do not just share images with their followers, they collect images of things they like on the internet [10]. Some users browse e-commerce sites to find the products they like on Pinterest, for purchasing them. A recent market survey showed that a higher proportion of Pinterest users click through to e-commerce sites, and when they go there, they spent significantly more money than people who come from sites like Facebook or Twitter, thus Pinterest is of great interest to online retailers [12]. This research on Pinterest leads us to an important finding that pins of items on Pinterest are interesting to plenty of consumers which in turn leads to greater revenue for e-commerce companies resulting in a win-win situation. Recent study suggests that the majority of Pinterest users are women, in particular, 80% of United States users are women [12, 23]. Statistical research on Pinterest data shows that women concentrate their pinning activity on fewer topics. Women's fashion alone accounts for 10% of content gener-

*This author was an intern at eBay Research Labs during this work ated across all categories on Pinterest [7]. These findings motivated our research on discovering interesting products for women's handbags and women's shoes on a popular e-commerce platform, eBay. Not all pins are tangible products that can be purchased, in fact top two categories on Pinterest according to a recent study are "food and drink" and "DIY crafts" [7]. Our research focusses on addressing this question of discovering "interestingness" among eBay items for women's handbags and shoes categories. In this paper we empirically show that every eBay product has a certain degree of *Pinterestingness* associated with it. We then propose a data driven approach of classifying products as being interesting or not. The classifier's output are ordered from the most interesting item to the least interesting item based on the SVM score. Although we do not use images for learning interestingness, we use images for evaluating our results. We evaluate our results by comparing the items to Pinterest's directly. We obtain concrete results for the women's shoes and handbags categories on eBay and find products that are comparable in interestingness to Pinterest pins in those categories even though we do not use images for our computation.

2. RELATED WORK

To the best of our knowledge there has been no past work in the area of discovering "interestingness" in fashion by mining textual data from Pinterest. Gilbert et al. in their paper give statistical overview of Pinterest. Their study shows that there are four verbs that set Pinterest apart from Twitter: *use, look, want* and *need* [12]. These words suggest that a lot of pins are things that users would like to own and thus strengthens our assumption that items on Pinterest tend to be attractive. Chang et al. in a recent work studied the distribution of content across Pinterest categories, the extent to which users specialize and share similar interests [7].

There has been some past work on fashion interestingness using images in the domain of computer vision. Di et al. propose a model that promotes visual attractiveness by incorporating presentation efficacy and user preference. They obtain qualitative results to improve user engagement on top of relevance [8]. However their work heavily relies on computer vision for identifying interestingness in the fashion category. Past research has shown the importance of images to buyers when purchasing on one of the largest e-commerce site, eBay [1, 3, 13]. There has also been research that suggest images to be the most influential risk-reducing factors for online shopping [4]. Although we use only text to deduce if an item on eBay is interesting, the importance of images led us to evaluate our results based on item images. Recent research by Di and others indicates positive evidence that images help increase buyer's attention and trust, and thus importance of higher quality images [9].

3. IMPLEMENTATION

3.1 Amazon Mechanical Turk

For preliminary analysis we used Amazon Mechanical Turk (AMT) [5], a crowdsourcing platform for annotating women's handbags on eBay. For each item, we display its title along with the image and ask the question, "Does this handbag catch your attention?". The turker had to select one of the 3 responses, "yes", "no" or "not a handbag". Each item was annotated 5 times by different turkers. We also provide instructions in which we elaborate on our question by asking it in different ways, "Would you show this handbag to someone?" or "Would you pin it on your Pinterest?". Turkers were paid 5 cents for each annotation and since every item is labeled by 5 different turkers, we spent 25 cents on labeling each item. The idea was to obtain an annotated set of eBay items for interestingness. This would allow us to create the gold standard/ground truth dataset for our problem. Since the question is very subjective and our benchmark for annotation is Pinterest, we identified a smaller set of 20 interesting test items and planted it for quality control. Only the judgments of turkers that labeled those smaller set of test items as positive are taken into account.

The results we obtained from the labeling experiment had a very high variance even after having a qualifying test set and restricting turkers from the US only. Several handbags had all the 3 possible responses from turkers and thus did not produce a majority response. Out of our initial batch of 60 handbags, only 50% of them had 3 or more workers agreeing on a label. These results made us realize that obtaining unbiased labeled data for our task is very difficult. There are several reasons for this bias: the turker may unintentionally compare the handbag to the most recent handbag he/she has seen and thus make a biased judgement, the turker may be a spammer i.e. he/she does not pay attention to the task and randomly selects responses. This analysis indicates that constructing a ground truth dataset for interestingness on eBay data is not feasible. We therefore decided to design a system that does not rely on explicitly labeled data.

3.2 Textual Feature Based Approach

We analyzed the text titles of women's handbags from both eBay and Pinterest using a fashion vocabulary comprising of adjectives used to describe fashion, for example chic, classic, vintage, timeless[1]. We found that around 50% of 10,781 Pinterest handbags had some fashion vocabulary word in their title compared to 33% of 664,282 eBay handbags. Pinterest was thus more expressive compared to eBay, which led us to explore features that could capture expressiveness in text.

Mikolov et al. show how to train distributed representations of words and phrases with the skip-gram model making them more expressive [19]. We use their *Word2Vec* model to train on large amount of Pinterest and eBay titles across *all* categories. We used approximately 734K pins and 2.2M eBay items for training. It took around 100 minutes to train the *Word2Vec* model using 200 dimensions and restricting minimum word count to 10. We used this trained model for generating features for pins and items in our dataset as follows: for each pin or item, we sum the word vectors for every word in its title and normalize by the total number of words. The final vector of 200 dimensions contributes 200 features.

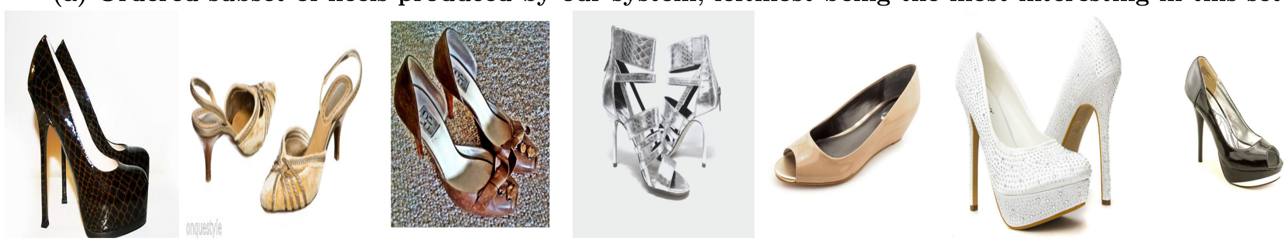
Apart from *Word2Vec* we also use popularity of a product as a feature. The interestingness problem can be also thought of as ranking items according to their attractiveness and we use popularity score as a proxy for interestingness. We define popularity score for a Pinterest pin as follows. Popularity Score(PS) measure for a pin x among a set of pins X is

$$PS(x) = \frac{\#comments(x) + \#likes(x) + \#repins(x)}{\sum_{x \in X} \#comments(x) + \#likes(x) + \#repins(x)}$$

The sum of scores for all pins in a set is 1 as per our definition. The popularity score as a feature enables the classifier to discriminate between popular pins and average pins.

[1] http://myvocabulary.com/word-list/fashion-vocabulary

(a) Ordered subset of heels produced by our system, leftmost being the most interesting in this set



(b) Subset of random heels on eBay

Figure 1: Comparing our classifier's output on subset of heels to a random sample of heels on eBay

For measuring popularity of an eBay item, parameters such as reviews and number of times an item is purchased are available. However, an item sold by different sellers may have different reviews based on the sellers credibility and other such factors. An item sold by different sellers at different cost would lead to noisy popularity score. Thus to overcome these problems for eBay items, we replace the popularity score by a feature that gives uniform score to every item. The idea behind using a uniform score is that every eBay item is equally likely to be interesting. The popularity score for an eBay item is, therefore, $1/N$ where N is the total number of items in our set and thus sum of scores for all items also sums to 1. In total we had 201 features for every Pinterest pin and eBay item in the dataset (200 *Word2Vec* features and 1 popularity score feature). Our experimental results show that popularity score boosts our classification accuracy.

Based on our assumption of interestingness, we assume all pins to be positive and all eBay items to be negative from attractiveness standpoint for classification purposes. Our approach then is to find Pinterest like items in eBay data by using a classifier. The idea being that an eBay item classified as positive has Pinterest like traits and is potentially interesting. The intuition behind this idea is that if a classifier learns well to discriminate between eBay and Pinterest instances then false positive produced by such a classifier are Pinterest-like eBay items. As we discuss in Section 4, we have much lesser Pinterest data compared to eBay data. Our classification approach thus comprises of training a model that learns from a small dataset of a particular category. The objective is to use the trained category specific model for finding interesting items on a much bigger dataset than used for training. However, this is only possible if the classifier can learn to discriminate between eBay and Pinterest instances for binary classification and the false positives are indeed attractive or Pinterest-like.

To test our hypothesis we divide our implementation into two tasks. The first one is to verify that the aforementioned features are good discriminators between Pinterest and eBay data by training and testing a classifier on them. The second task is to verify that the false positives obtained by the Pinterest-eBay classifier are more interesting items than those classified as negative. For the first task we verify the performance of the classifier by training it on equal number of pins and eBay items and thereafter testing on a smaller equal number of pins and items. A good recall would indicate that the classifier has successfully learnt discriminating features and is able to retrieve most of the positive items. Thus eBay items falsely classified as being positive by such a classifier would imply that those items have traits of being in the positive class and thus are potentially interesting. We discuss the details of the first task in Section 5. For our second task we verify the interestingness of these false-positive eBay items by using various evaluation measures. Since its hard to evaluate interestingness using text only, we use the item's image along with its title. As discussed before, interestingness of an item is subjective and hard to annotate, thus we evaluate it in a relative sense. The details of evaluation are discussed in Section 6.

4. DATA

For all our experiments we use the data from Pinterest and eBay. We discuss each of the sources separately because they differ widely in terms of collection and pre-processing. As mentioned before, the focus of our research is to study what makes women's fashion interesting and therefore we only consider women's handbags and women's shoes categories for our experiments.

eBay		Pinterest	
Shoes	Handbags	Shoes	Handbags
785,014	664,282	8,729	10,781

Table 1: Number of women's shoes and handbags for eBay and Pinterest used in our experiments

4.1 eBay

Items sold on eBay are grouped into several categories and sub-categories. Our data comes from *Women's Shoes* and *Women's Handbags* & *Bags* categories on eBay. For each item, we extract the item's title along with its image. All eBay data was collected for a period between October

1st to October 15th 2014. Many items on eBay are sold by multiple sellers and therefore we removed all duplicate items based on their titles after the preprocessing. For the preprocessing step, we remove the item's price if it is in the title and eBay specific terms - new, used, NIB(new in box), NWT(new with tags), NWOT(new without tags). We filter out all items that have titles with less than 4 characters after preprocessing. Table 1 gives the size of eBay data for both categories after preprocessing.

4.2 Pinterest

We crawled Pinterest for 2 months and retrieved all pins returned for the search query - "women's handbags" and "women's shoes". The crawler fetches the image, URL and title for each pin as well as the number of re-pins, comments and likes. All data from Pinterest was passed through a pipeline for preprocessing and reducing noise. The first step in the pipeline is to filter the pins that are not products based on the pin's title and source URL. Many of the pins on Pinterest are ideas and not products. This step eliminates such pins because they are not useful for our task. Thereafter the second step is to verify the pin's category. For this step each pin is mapped to an existing eBay category using words in its title and source URL and filtering those that do not get mapped to any category. Finally we also eliminate pins with popularity score of zero i.e. they do not have any comments, likes or repins. The Pinterest preprocessing pipeline heavily uses NLP tools, inference rules and several heuristics and is not the focus of this paper. Also like the eBay items, we remove URL if any and Pinterest specific words from the titles of the pin - "pinterest", , "pin", "want" and "need". We also filter pins that have less than 4 characters in their title after preprocessing. Table 1 gives the size of Pinterest data for both categories.

Apart from preprocessing pins crawled from Pinterest, the pipeline also has a component that maps the pins identified as products to a matching eBay item using only textual data. The reason we do not use this component for identifying interesting items is that it has very low accuracy. Out of 600 pin to item matched products obtained from this component, only approximately 25% were correct matches when evaluated using Amazon Mechanical Turk using 5 assignments per task, even though turkers were instructed to consider versions of items with different colors and different camera angles to be the same product. We note that the task was very straight forward and the judgments were found to be trustworthy. Although we do not use the matched products in our computation, we do use the output of the matching component to evaluate our results because these eBay items are those that matched some pin. We call such eBay items as "meBay" because they are matched products on Pinterest. We will use this definition in Section 6 for evaluation.

5. EXPERIMENTS

	Train		Test	
	Shoes	Handbags	Shoes	Handbags
Pinterest	8,000	10,000	729	781
eBay	8,000	10,000	729	781

Table 2: Training and test datasets used for shoes and handbags categories

In this section we discuss the details of interestingness as a classification task for the two categories of interest, women's shoes and women's handbags. As discussed earlier we have two classes, pins being the positive class and eBay items being the negative class. For each category, we train and test separate classifiers on the data for that category. We use LIBSVM's two class classifier with linear kernel for training and do a 10 fold cross validation [6]. The parameter C is tuned using cross validation. Table 2 gives the number of training and test data size for each class in each category. Since we have a much smaller size of Pinterest data, we are constrained to use an equal amount of eBay data for training to avoid unbalanced classes. The results obtained on training and testing the Pinterest-eBay classifier on each of the categories is given in Table 3. A high recall signifies that the classifier has learnt a good decision boundary between the positive and negative class. Although we treat all eBay items to be negative in our classification task, that is not the case in reality. In fact the eBay items with positive class attributes fall on the Pinterest side of the hyperplane and are thus classified as positive. If the number of such eBay items is large, it may significantly affect precision. We obtain a recall of 0.943 for the shoes category and a recall of 0.799 for the handbags category. The classifier does really well on retrieving positive class instances for the shoes category and not so well of the handbags category. As we see in Section 6 that indeed the shoes category gives us better results compared to the handbags category.

	Shoes	Handbags
Train Accuracy	0.920	0.999
Precision	0.835	0.825
Recall	**0.943**	**0.799**
F1	0.885	0.812
Accuracy	0.878	0.814

Table 3: Performance of 2 class SVM on Shoes and Handbags categories. Number of test instances for each category are given in parentheses

The classifier's performance on test data is really good compared to a random guessing classifier as baseline. Without the popularity score feature, the recall drops down approximately 2% for both the categories. Because we have very large eBay dataset we also experimented by training a one class SVM proposed by Schölkopf on only eBay data for each of the categories [21]. However, a significant number of pins were classified as negative and vice versa. We therefore did not go ahead with the one class approach.

Based on the results obtained by our classifier we infer that it is possible to train a model with simple features that discriminate between eBay items and Pinterest pins. The next step is to verify the misclassified eBay items. We therefore hypothesize that the eBay items misclassified as Pinterest's are outliers with positive class attributes and thus potentially interesting items. The trained two class model is therefore tested on a larger dataset comprising of eBay items only. We tested our shoes model on $775,014$ eBay women's shoes and the handbag model on $654,282$ eBay women's handbags. We obtained approximately 8700 false positives for the shoes category and around 8000 false positives for the handbag category. We then rank the misclassified eBay items from highly positive to less positive based on the classifier's score for that item. Thereafter we calculate the greatest drop in the score for the correctly classified Pinterest test data that we discussed before. This allows us to separate the highly interesting to average items and we

(a) Ordered subset of women's shoes produced by our system, leftmost being the most interesting in this set

(b) Subset of women's shoes on meBay

Figure 2: Comparing our classifier's output on subset of women's shoes to sample of women's shoes on meBay

were therefore left with only 2872 and 538 eBay items that were misclassified for the shoes and handbags categories respectively.

Although the classifier is trained using textual features, judging the results of the classifier based on only the titles of the items was very difficult so we evaluate our output using images along with the titles. As discussed earlier interestingness is subjective and has no concrete definition for the fashion category. Therefore it is extremely difficult to evaluate and quantify interestingness of an item. This led us to develop techniques to evaluate our classifier's output and thus our hypothesis. The details of evaluation are discussed in the next section.

6. EVALUATION AND RESULTS

We propose 3 techniques to evaluate our hypothesis that false positive eBay items are Pinterest-like and thus interesting. We rank the classifier's misclassified output on eBay items from most positive to most negative so that the item that is highly Pinterest like, i.e. far away from the classifier boundary, appears before the one that is less positive and closer to the classifier boundary. Thereafter we create another set of items belonging to the same category using one of the following techniques:

1. **Random:** Random sampling of items belonging to the same eBay category without order.
2. **meBay items:** Random sampling of items from meBay. These are set of eBay items that some Pinterest pin matched and thus are interesting based on our assumption.

We use crowd sourcing to evaluate our results on the AMT platform. The classifier's output set is juxtaposed with the set of items obtained from one of the 3 aforementioned techniques. Turkers are then asked to select the set of items they think are more interesting or would catch their attention. Since the comparison is relative, the problem of unconscious comparison that we discussed earlier while labeling is eliminated and we obtain unbiased judgments. We restrict labeling by turkers belonging to the United States only. Figure 1 is a subset of the list of interesting items generated by the classifier when tested on *heels* subcategory under shoes, along with another random subset of heels from eBay. Figure 2 is a subset of interesting eBay shoes generated by the classifier along with a set of sample shoes from meBay. Since items on meBay are those that matched some Pinterest pin through the pipeline process discussed in Section 4.2, they are all interesting items based on our Pinterest interestingness assumption.

Using the aforementioned process we evaluated 2872 eBay shoes and 538 eBay handbags on AMT. Individuals were asked to select the more attractive and interesting set of items from among the two that were displayed in front of them. Seven items were shown at a time along with their title and image. Turkers had only two options and they had to select one of the two sets that they thought was more interesting. We use Fleiss' Kappa to measure inter-annotator agreement [11]. For the shoes category we obtained $\kappa = 0.43$ and $\kappa = 0.25$ for the handbags category on the entire evaluation set. The annotators had access to both the images and title of the item while judging for interestingness. However it is important to note that our classifier was trained using textual features only and had no access to the item's image. Fewer less complicated features made our system extremely fast and robust even on large data.

7. DISCUSSION

The "interestingness" system uses a trained *Word2Vec* model to extract textual features and then predicts the class of the eBay item using a trained SVM model. The SVM model is trained on smaller dataset and thus the training is extremely fast. Also since no images are used by the system, the entire process is very computationally inexpensive. We believe such a system is ideal for the very large datasets ubiquitous in industry especially in e-commerce. For the eBay women's shoes and handbags categories we ran our output through Mallet's topic modeling system [18]. We considered all the potential interesting items' titles as a single document for the topic model. Table 4 gives the top words from all the titles.

Shoes Topic	Handbag Topic
womens	bag
shoes	purse
size	handbag
heels	**leather**
leather	shoulder
wedge	**tote**
nike	**clutch**
platform	black
pump	brown
converse	satchel

Table 4: Top 10 topics for "interesting" women's shoes

The top 3 words are very common in most titles and do not convey much meaning but the other words give very concrete sense of what makes a shoe attractive to many buyers. We find that if a shoe has either *heels, wedge, leather, nike, platform, pump* or *converse* in its title then it is a potentially interesting item. Unsupervised technique such as topic modeling on Pinterest data does not give us good us results. This is mainly because Pinterest textual data is very noisy and most pins have no annotation or any description of what the pin is about. As mentioned in Section 5 for women's handbags category, our classifier does not obtain very high recall and consequently the inter annotator agreement on evaluation was also low compared to the women's shoes categories., Section 6 Therefore, we only show results for the women's shoes category in this paper. We infer that interestingness for women's handbag category is not well captured by our textual features and may require visual features or more sophisticated techniques.

8. CONCLUSION

Fashion is the fastest growing category in online shopping and there has been very little work on identifying interestingness in fashion. Most past research in online fashion has been restricted to the domain of computer vision. In this paper we showed that using a product's textual data is effective for identifying its attractiveness and ability to draw consumer's attention and interest.

We used Pinterest data as our benchmark for interesting items because pins have been established to be attractive. We proposed a data driven approach for extracting interestingness based on textual data from pins and applying it to eBay items. Our implementation for interestingness involved two steps, the first was to use a classifier for identifying Pinterest like eBay items and the second was to evaluate that the false positives from the classifier are actually interesting. For the evaluation task we used AMT and compared interesting eBay items extracted by our classifier to Pinterest matched eBay items and items. We obtained inter annotator agreement of $\kappa = 0.43$ for evaluating our classifier's performance on women's shoes. Because we do not use explicitly labeled data for training, our data driven approach for finding interesting items in the fashion category by training a 2 class eBay-Pinterest classifier can be trained on small amount of data and applied effectively on large datasets common in e-commerce applications.

9. REFERENCES

[1] P. Bajari and A. Hortacsu. The winner's curse, reserve prices, and endogenous entry: empirical insights from ebay auctions. *RAND Journal of Economics*, pages 329–355, 2003.

[2] P. Bedi, H. Kaur, and S. Marwaha. Trust based recommender system for semantic web. In *IJCAI*, volume 7, pages 2677–2682, 2007.

[3] E. M. Bland, G. S. Black, and K. Lawrimore. Determinants of effectiveness and success for ebay auctions. *The Coastal Business Journal*, 4(1):5–15, 2005.

[4] E. M. Bland, G. S. Black, and K. Lawrimore. Risk-reducing and risk-enhancing factors impacting online auction outcomes: empirical evidence from ebay auctions. *Journal of Electronic Commerce Research*, 8(4):236–243, 2007.

[5] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, 6(1):3–5, 2011.

[6] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[7] S. Chang, V. Kumar, E. Gilbert, and L. G. Terveen. Specialization, homophily, and gender in a social curation site: findings from pinterest. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 674–686. ACM, 2014.

[8] W. Di, A. Bhardwaj, V. Jagadeesh, R. Piramuthu, and E. F. Churchill. When relevance is not enough: Promoting visual attractiveness for fashion e-commerce. *CoRR*, abs/1406.3561, 2014.

[9] W. Di, N. Sundaresan, R. Piramuthu, and A. Bhardwaj. Is a picture really worth a thousand words?:-on the role of images in e-commerce. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 633–642. ACM, 2014.

[10] Z. Feng, F. Cong, K. Chen, and Y. Yu. An empirical study of user behaviors on pinterest social network. In *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on*, volume 1, pages 402–409. IEEE, 2013.

[11] J. L. Fleiss and J. Cohen. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*, 1973.

[12] E. Gilbert, S. Bakhshi, S. Chang, and L. Terveen. I need to try this?: a statistical overview of pinterest. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2427–2436. ACM, 2013.

[13] J. H. Gilkeson and K. Reynolds. Determinants of internet auction success and closing price: An exploratory study. *Psychology & Marketing*, 20(6):537–566, 2003.

[14] A. Haywood. Online auctions: User experience insights from ebay. *Chimera Working Paper*, 10:2006, 2006.

[15] H. Ma, I. King, and M. R. Lyu. Learning to recommend with social trust ensemble. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 203–210. ACM, 2009.

[16] H. Ma, H. Yang, M. R. Lyu, and I. King. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 931–940. ACM, 2008.

[17] P. Massa and P. Avesani. Trust-aware recommender systems. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 17–24. ACM, 2007.

[18] A. K. McCallum. Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu, 2002.

[19] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119, 2013.

[20] J. O'Donovan and B. Smyth. Trust in recommender systems. In *Proceedings of the 10th international conference on Intelligent user interfaces*, pages 167–174. ACM, 2005.

[21] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.

[22] A. Silberschatz and A. Tuzhilin. On subjective measures of interestingness in knowledge discovery. In *KDD*, volume 95, pages 275–281, 1995.

[23] C. Taylor. Women win facebook, twitter, zynga; men get linkedin, reddit. *Retrieved Feb*, 10:2013, 2012.