Predicting Purchase Behaviors from Social Media

Yongzheng Zhang
eBay Inc.
2065 Hamilton Ave
San Jose, CA, USA 95125
ytzhang@ebay.com

Marco Pennacchiotti
eBay Inc.
2065 Hamilton Ave
San Jose, CA, USA 95125
mpennacchiotti@ebay.com

ABSTRACT

In the era of social commerce, users often connect from e-commerce websites to social networking venues such as Facebook and Twitter. However, there have been few efforts on understanding the correlations between users' social media profiles and their e-commerce behaviors. This paper presents a system for predicting a user's purchase behaviors on e-commerce websites from the user's social media profile. We specifically aim at understanding if the user's profile information in a social network (for example Facebook) can be leveraged to predict what categories of products the user will buy from (for example eBay Electronics). The paper provides an extensive analysis on how users' Facebook profile information correlates to purchases on eBay, and analyzes the performance of different feature sets and learning algorithms on the task of purchase behavior prediction.

Categories and Subject Descriptors

K.4.4 [**Computers and Society**]: Electronic Commerce; H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*Information networks*; H.3.5 [**Information Storage and Retrieval**]: Online Information Services; I.2.6 [**Artificial Intelligence**]: Learning

Keywords

E-commerce, social networks, social media, social commerce, recommender systems

1. INTRODUCTION

In recent years, many e-commerce companies such as Amazon and eBay have been moving into the social media space by allowing users to connect to social networking sites (e.g. Facebook and Twitter). The main strategic goal for social media interaction is to provide users with a more engaging and social experience, thus increasing user retention and adoption. More importantly, social media is often seen as a means to rejuvenate the user base and attract younger "social-savvy" users. Typical features unlocked by social media include the possibility of sharing purchase activities with friends, and tools such as friends gifting applications and chats.

When users connect from an e-commerce site to social media for the first time, they often agree to share with the e-commerce company basic information such as their demographics and personal interests (e.g. Facebook "likes"). However, e-commerce companies have not fully developed technologies to leverage this information to improve important features such as purchase behavior prediction and product recommendation. Social media information could also help solve the *cold start* problem, i.e., providing an engaging and personalized experience to brand new users. When a new user comes, traditional prediction and recommendation algorithms cannot in fact be applied, as no past information about the user is available.

In this paper we claim that social media information provides sufficient knowledge to predict, to a certain extent, the user's purchase behaviors. For example, a Facebook user who "likes" the Facebook pages of many fashion brands is more likely to purchase fashion products than a user who mostly likes car accessories.

1.1 Problem Statement

We deal with the problem of predicting the purchase behaviors of social media users who have unknown history on an e-commerce website (*cold start*). More in detail, we aim at predicting which product categories (e.g. electronics) the user will buy from by using solely information derived from the social network. Such a predictive system would help in several practical scenarios, including:

- build a *cold start* recommender system, by providing high-level recommendations to social media users who connect for the first time to an e-commerce website;

- improve existing product recommendation engines, by providing category-level priors that can guide the recommender system to find domains of interest for the user;

- provide e-commerce companies with tools for targeted social media campaigns.

We instantiate the problem by choosing Facebook as the social network, and eBay as the e-commerce website[1]. E-commerce companies such as Amazon and eBay ask users to share only a small portion of their personal social media information, in order to protect the privacy of the users and

---

[1] eBay, among many other websites, uses the "Facebook Connect" API to let users connect from eBay to Facebook.

of their friends. For example, both Amazon and eBay do not ask users to share the postings and comments on their social accounts.

In order to keep our task realistic, we therefore focus on the restricted set of Facebook information that users are typically required to share, namely demographic information (age, gender) and the pages that the users have liked on Facebook[2]. In the rest of the paper we will use interchangeably the terms "liked pages" and "likes" to indicate Facebook pages for which a user has expressed a preference by clicking on the *Like* button.

Note that Facebook pages are organized in categories. For example *Beatles* is part of the *Musician/band* category, and *Starbucks* of *Food/beverage*. We will make use of category information throughout the paper. As regards the set of product categories we want to predict, we use the 35 "eBay meta-categories" (or "eBay categories" for brevity) which form the first and most general layer of the eBay product taxonomy. Examples of meta-categories are *Books* and *Home & Garden*. The list of 35 eBay meta-categories and the full taxonomy are available at: http://hub.shop.ebay.com.

This paper has two main goals. The first goal is to explore if users' Facebook information is correlated with the eBay categories from which the users buy. The second goal is to leverage existing correlations to build machine learning algorithms that predict users' purchase behaviors from their Facebook information.

1.2 Main Contributions

This paper provides the following main contributions:

- We provide an extensive analysis of user purchase behaviors and Facebook information over a dataset of 13,619 users. We show that a subset of Facebook features correlates with purchase behaviors with statistical significance.

- We build and evaluate various machine learning models to predict which product categories from which a user will buy, by using Facebook information alone. We show that the prediction task can be successfully solved with promising accuracy.

To our knowledge this is the first study that systematically analyzes the correlations between social media profiles and purchase behaviors on e-commerce websites.

The rest of the paper is organized as follows: Section 2 reviews relevant previous work. Section 3 reports statistics of our dataset. In Section 4 we answer a fundamental question: do users have specific focus when they buy online, or do they exhibit an unpredictable behavior? In Sections 5 and 6, we analyze if specific information from social media, namely demographic data and Facebook likes, correlates to user behaviors on e-commerce websites. In Section 7 we describe and evaluate different algorithms for predicting user purchase behaviors from social media information. Finally, Section 8 concludes our work and describes future research directions.

---

[2]In this study we do not use the list of friends, as this information is not included in the scope of eBay's Facebook Connect.

2. RELATED WORK

2.1 Product Prediction and Recommendation

Recommender systems have been widely studied in the past [26, 5, 27, 16, 8] and applied to various domains. E-commerce is a typical application for these systems, where they are used to predict or recommend product purchases. Two main techniques are most used: collaborative filtering and content-based.

Collaborative Filtering Methods. The underlying assumption of collaborative filtering methods is that users who carry similar characteristics will tend to like similar products. Users are typically represented in a vector space which summarizes their characteristics (e.g. demographics, purchased products, review scores assigned to products). These systems suggest new products to a user by selecting a set of products that similar users have bought or reviewed in the past, but the user has not [30]. Sarwar et al. [28] apply nearest neighbor collaborative filtering for recommending purchases and predicting movie ratings, showing that dimensionality reduction techniques solve scaling issues on large-data without losing accuracy. Some studies [27, 16] have focused on the problem of *sparsity*. In Section 3 we will show the sparsity problem in social media and in the following ones the methods to tackle it.

Item-Item collaborative filtering, on the other hand, makes product recommendations directly based on users' past behaviors on e-commerce websites. Sarwar et al. [29] propose a system that models item to item relationships and demonstrate that such systems deliver good quality recommendations in sparse data situations. The item-item method is also promising for incremental modelling and has big performance gain over user-user modelling.

Content-based Methods. In contrast to collaborative filtering methods, content-based methods often utilize the vast overload of information on the web, such as product reviews, customer opinions, and social media (e.g. blogs, tweets) to directly make product recommendations. In recent years, such systems have become more popular. Wietsma and Ricci [33] propose a system that structures product reviews to help rate and recommend products in a mobile decision aid scenario. Aciar et al. [2] develop a recommender system based on consumer product reviews. The authors apply text mining techniques to extract useful information from review comments and then define an ontology to translate the review information into a form suitable for utilization by a recommendation system. Sen et al. [31] use users' tag preferences to predict movie preferences and demonstrate that such content-based systems are more effective than collaborative filtering methods.

Our system explicitly integrates collaborative filtering techniques in the feature selection phase. Our selection method for Facebook categories/likes/n-grams is inspired by model-based approaches to collaborative filtering discussed among others by Breese et al. [4].

2.2 Recommendation on Social Media

Recommendation on social media is a fairly new topic. Most work focuses on suggesting interesting content items (e.g. URLs, pictures, posts) or new friends. Social media recommender systems differ from classical ones in that they

often leverage existing social relations to boost the recommendations. Two main techniques are most used: content-based and community-based.

Content-based methods are most popular even though some interest has focused on collaborative filtering [12]. These systems assume that the content that users share on the social network reflects their own interests, thus recommending new items that are similar to their own shared content [25, 24]. Abel et al. [1] represent users by a frequency vector of the hashtags and entities that they mention in their tweets. A similar vector is built for URLs shared in Twitter by analyzing the content of their links. Users are then recommended URLs whose vector is most similar to theirs. Similarly, Chen et al. [6] represent users with the bag-of-words of all terms of their tweets, and the URLs with the terms of the tweets that mention them. Unlike these methods, in our dataset social content (e.g. Facebook posts) is not available.

Community-based methods make the assumption that the content coming from a user's friends or authoritative users is more likely to be interesting for a user than the rest. This assumption is usually combined with content-based and collaborative filtering approaches. Yan et al. [34] select personalized tweets using random-walks on a graph that joins the user's friends and tweets. The walk is bootstrapped by selecting with higher probability the tweets whose content is similar to what was previously posted by the user. Jamali and Ester [15] also use a random-walk approach, but they integrate collaborative filtering instead of content-based techniques to estimate the probability of a graph transition. Jiang et al. [18] combine user topic preferences and social influence into a probabilistic matrix factorization model, recommending items that are both topically similar to the previous user's preferences and authored by people who are trusted by the user. Community-based systems show better performance with respect to purely content-based and collaborative filtering methods. In our work we cannot use community-based techniques since we do not have access to the social graph of eBay users.

2.3 Social Networks and Purchase Behaviors

Some research has investigated the broader topic of how social network influences users in their purchases. Bhatt et al. [3] empirically demonstrate that a user's friends exercise "peer pressure": if friends widely adopt a product, the user is more likely to buy it. Guo et al. [11] study the trading dynamics on the e-commerce social network Taobao. They show that buyers are more likely to purchase from sellers that friends in their network have already bought from (information passing). They prove that when a buyer has to decide from which seller to buy a product, the social network has a bigger influence on the decision than the sellers' ratings and the price of the product. Similar findings are found in [13, 14].

All these studies assume that the e-commerce site and the social network are integrated into one platform (e.g. Taobao), or that the e-commerce site has holistic access to the social network. To the best of our knowledge, our work is the first to focus on a more practical and common scenario, where the system (e.g. e-commerce companies such as Amazon and eBay) has limited access to social network information.

Table 1: Example of User Information.

Name	Anonymous
Gender	Male
Age group	35-44
Facebook likes (*Category*)	Beatles (*Musician/band*) iPhone 5 (*Electronics*) Starbucks (*Food/beverage*) Walt Disney Studios (*Movie*)
eBay purchases (*Meta-category*)	iPhone 4S (*Electronics*) Beatles T-shirt (*Clothing*) Beatles Mug (*Collectibles*)

Table 2: Statistics of Our Dataset.

Users	13,619
Facebook categories	214
Facebook pages	1,373,984
Facebook likes	4,165,690
eBay categories	35
eBay purchases	628,753

3. FACEBOOK-EBAY DATASET

The dataset we use in this study comes from a database of eBay's "Facebook connect" users. It contains a random sample of 13,619 anonymized eBay users who connected to Facebook between June and August 2012. We excluded users under 18 years of age and those who have no Facebook likes or have not made any purchase on eBay in 2012. For each user, the dataset stores the following information:

- Basic demographic information obtained from Facebook, including age and gender;

- Facebook likes and their categories;

- A list of items purchased on eBay from January to August 2012 (item name and category).

An example of user information collected in our dataset is presented in Table 1.

Basic statistics of our dataset are reported in Table 2.

Figure 1 reports the distribution of gender and age groups in our dataset. We notice a prevalence of women (60% of all users) and people aged between 25 and 44 (55% of all users). In Section 5 we will explore if users in different demographic groups have distinctive purchase behaviors.

Figure 2 reports the distribution of Facebook likes for users, i.e., it indicates how many users have liked a given number of pages. The function is approximately power law with only a few outlier fluctuations, meaning that most users like few Facebook pages, and few users like many pages (the median is 152 likes). While not surprising, this indicates that our task is inherently hard: for most users we have to rely on scarce Facebook information for predicting their purchase behaviors.

Figure 3 reports the distribution of likes for Facebook pages, i.e., it indicates how many pages have a given number of likes. The function follows a perfect power law, showing that the majority of Facebook pages have few likes and only a few pages receive many likes (the median is 1 like). The fact that users' likes are so sparse poses a great challenge for our prediction task when likes are used as features.
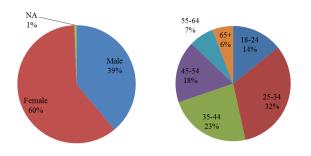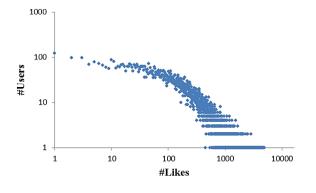
Figure 1: Distribution of user demographics.

Figure 3: Distribution of likes for pages.

Figure 2: Distribution of likes for users.

Figure 4: Distribution of purchases for users.

As regards user behaviors on eBay, the distribution of purchased items is also power law, as shown in Figure 4, indicating that most users tend to buy a limited number of items.

Figure 5 reports the distribution of purchases by meta-category. The distribution is highly skewed: more than 50% of all purchases come from the top five meta-categories. The *Clothing, Shoes & Accessories* category alone accounts for 17.5% of all purchases. In the context of our study this means that a system that selects the most popular meta-categories as a prediction of where a user will buy, would achieve a good degree of accuracy. We will have more on this in the experimental section. The median value of purchases per category is 8,316; the average is 17,964.

4. FOCUS ON PURCHASES AND LIKES

The first important question that we want to answer in this study is: are users focused when they buy online? One extreme hypothesis is that a user is completely unfocused, i.e., the user likes to buy randomly across categories. On the other end, the user may have a few favorite categories from which majority purchases are made.

The former hypothesis depicts a chaotic world where it is impossible to predict user behaviors and provide recommendations. In the context of this paper, we hope our world to be more organized than this.

To answer the above question, let us represent with $P(u)_k$ the ranked probability with which a user u buys from the kth favorite category. This rank is obtained by first estimating the probability $P(u, e)$ of a user u buying in each category e, and by successively ranking the probabilities:
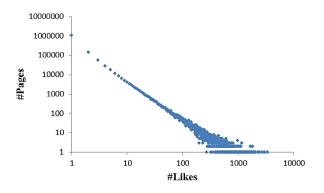
$$P(u, e) = \frac{purc(u, e)}{purc(u, E)} \qquad (1)$$

where $purc(u, e)$ is the number of purchases of u in category e, and E is the set of all 35 eBay meta-categories. For example if a user buys 3 items from one category and 2 from another, we have: $P(u)_1 = 0.6$ and $P(u)_2 = 0.4$.

To have an estimation of purchase focus we average the $P(u)_k$ across all users U. We thus obtain the probability distribution for the event of an average user buying in the top k ranked category:

$$P(U)_k = \frac{1}{|U|} \cdot \sum_{u \in U} P(u)_k \qquad (2)$$

The probability mass function for the distribution is reported in Figure 6, where categories are ordered by rank k.

The hypothesis of a chaotic world where a user buys randomly from different categories would be proved if the distribution was fitted by a uniform distribution. To check the fit, we apply the Kolmogorov-Smirnov (K-S) goodness-of-fit test. The result of the test shows that the hypothesis is rejected. As expected, users do not buy randomly.

We repeat the K-S test to check what continuous distribution best approximates the purchase distribution. The best fit is provided by a Gamma distribution ($\Gamma(0.625, 1.322)$ with the D-statistic of 0.19).
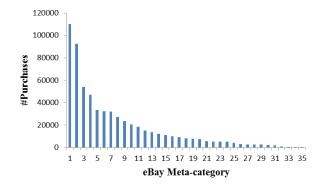
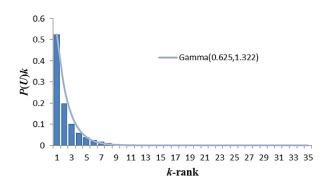**Figure 5: Number of purchases in all eBay meta-categories.**

Figure 6: Distribution of purchases in eBay meta-categories.

The shape of the distribution indicates that users are very focused in their purchase behaviors. Figure 6 shows that more than 50% of the times an average user buys from the first preferred category and 20% of the times from the second preferred category. The top 3 categories collectively account for about 85% of a user's purchases.

Another important question is: do users express specific interests on Facebook, i.e., do they like specific categories of pages? Similarly to what we do for eBay categories, we answer this question by checking the hypothesis that Facebook users like pages from random Facebook categories.

We build the probability distribution for the event of an average user liking a Facebook category f using the same procedure used for eBay categories but replacing e with f. The mass function (not reported for space limitation) fits a Gamma distribution that is less steep than the Gamma approximating eBay categories. We again reject the chaotic world hypothesis by running the K-S test on a uniform distribution. On average a Facebook user's most favorite category accounts for 19% of all liked pages, the second about 11%. Facebook likes spread out to more categories with respect to eBay purchases, though users appear to be quite focused also on Facebook.

Overall, the results provided in this section prove that users express strong personal interests on Facebook and are highly focused when purchasing online. One important question remains open. Is there a correlation between interests and purchases, i.e., do users purchase what they like on Facebook? If a correlation exists we could leverage Facebook likes to predict what users will likely purchase. We will answer this question in Section 6.

5. DEMOGRAPHIC DIFFERENCES

This section and the following explore possible correlations between Facebook information and eBay purchases. In Section 7 we will leverage these correlations to build algorithms for predicting purchase behaviors. We start by focusing on demographic information available on Facebook, and later explore the use of the list of liked pages.

A large body of work has studied demographic distinctions in the area of e-commerce. Among others, Garbarino and Strahilevitz [10] and Chiu et al. [7] show that women have a significantly higher sensitivity to risk when buying online. Kau et al. [20] indicate that "on-off shoppers" (people who explore online but buy offline) are prevalently teenagers; "comparative shoppers" (people comparing product features before buying) tend to be males in their twenties; offline shoppers are mostly people over 40. Differences in product adoption across different demographics have been observed, among others, in online movie rentals [21], music downloads and mobile data services [32], though these studies were all limited either in scope or in the size of the data.

Building on previous work, we first analyze if women and men tend to buy from different eBay meta-categories. In order to do so, we compute for both genders the percentage of users who buy in each category. For example about 70% of women in our dataset buy items from the *Clothing, Shoes & Accessories* category, while only 45% of men do so.

For each category, we carry out a *t*-test between women and men to verify if the difference in percentage is statistically significant. The results of the test show that women buy significantly more than men in 10 categories with a statistical significance of $p = 0.99$. The most female-polarized categories are *Jewelry & Watches*, *Crafts*, and *Clothing, Shoes & Accessories*. Men buy significantly more than women in 16 categories, with the most polarized being *Toys & Hobbies*, *Collectibles* and *Sports Memorabilia*. For the remaining 9 eBay meta-categories we do not observe any significant difference. These results show that purchase behavior strongly varies across genders.

Differences across age groups are less strong. For example, in only 10 categories is there a significant difference between age groups 25-34 and 45-54. In general we observe that young people (25-34) tend to be prevalent in *Fashion*, while older people (45+) are prevalent in *Collectibles* and *Books*.

The overall demographic study suggests that gender and age are important signals for predicting the purchase behaviors of social media users. In the experimental section we will quantitatively analyze their effective value.

For the sake of completeness we also study gender and age differences on Facebook. Similarly to purchase behaviors, we note that different demographic segments tend to like different types of pages. Females are prevalent in liking *Clothing* and *Health/beauty* pages, while males prevail in *Electronics* and *Sports*. Young users like more *Actors & Directors* while older people are prevalent in liking *Politicians*.

It is worth noting that these results refer to our dataset of 13K Facebook-connected eBay users, and may not generalize to the general population of eBay users or to the whole e-commerce spectrum.

Table 3: Examples of Correlated Categories.

eBay Category	Facebook Category	χ
Computers/Tablets	Computers/technology	52.0
Computers/Tablets	Software	51.9
Music	Record label	95.5
Music	Musical Instrument	67.1
Travel	Bags/luggage	7.9
Travel	Book Genre	5.9
Jewelry & Watches	Jewelry/watches	63.6
Jewelry & Watches	Health/beauty	13.4
Cell Phones	Telecommunication	67.2
Cell Phones	Electronics	46.1

Figure 7: Percentage of eBay categories that have a given number of highly correlated Facebook categories according to the χ^2 statistic.

6. CORRELATION BETWEEN SOCIAL MEDIA INTERESTS AND PURCHASES

In this section we study the correlation between Facebook categories and eBay meta-categories, and check if there are Facebook categories that are highly predictive of eBay meta-categories. For example we would expect that users who like many *Fashion* pages are likely to buy items in the *Clothing, Shoes & Accessories* meta-category.

Let us define two categorical variables $\mathcal{F}$ and $\mathcal{E}$. $\mathcal{F}$ is defined on the sample space of users, and associates each user to the set of Facebook categories that the user likes at least once. $\mathcal{E}$ associates each user to the eBay meta-categories from which the user has bought at least once.

We compute the correlation between Facebook and eBay categories by applying the Pearson's chi-square test [23] on $\mathcal{E}$ and $\mathcal{F}$. The chi-square test checks if the null-hypothesis that two random variables are independent (i.e. not correlated) is true or not. The result of our test is a strong rejection of the null hypothesis with confidence $p = 0.95$.

This result is encouragingly suggesting that the set of Facebook categories may be predictive of purchase behaviors. However, the test is generic and does not directly indicate which specific Facebook category f is highly correlated to which eBay meta-category e.

We therefore compute the Pearson's chi-square test on single (e, f) events (i.e., we test on a 2×2 contingency table). Table 3 reports the obtained correlations for some eBay meta-categories. For all the pairs reported in the table the null hypothesis that they are independent is rejected with confidence p=0.99.

Figure 7 shows the percentage of eBay categories (y-axis) that have a given number of highly correlated (either p=0.99 or p=0.95) Facebook categories (x-axis). As the figure shows all eBay categories have at least one highly associated Facebook category, while only 15% of eBay categories have 30 or more correlated Facebook categories at p=0.99. The median number of correlated Facebook categories across all eBay categories at the p=0.99 level is 19. The median number of correlated Facebook categories at the p=0.95 level is 35.

These results are very promising. The large number of discovered correlations suggests that eBay categories may be predicted by looking at the Facebook categories liked by the user. However, some eBay categories are inherently hard to predict. For example, *Real Estate*, *Art* and *Everything Else* have respectively only 4, 5 and 6 correlated Facebook categories. This may not be sufficient to correctly support a predictive algorithm for these specific eBay meta-categories. The reason for such low correlations is twofold. First, some
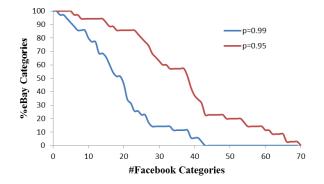
eBay categories correspond to concepts that are not popularly liked on Facebook (e.g. not many people like Real Estate companies). Second, some categories are too broad and vague to establish correlations (e.g. *Everything Else* and *Art*).

7. PURCHASE BEHAVIOR PREDICTION

In previous sections we have shown that users have focused behaviors when buying online. We have also shown that basic information from Facebook (namely demographics and liked pages) have promising correlation with purchase behaviors. In this section we explore if that information is strong enough to predict purchase behaviors. Specifically, we build and evaluate a variety of machine learning models to solve the following ranking problem.

Problem Statement. Given a social media user u and a set of features derived from the user's social media account, we aim at producing a ranked list of product categories from a category set E, where higher rank is assigned to categories that the user is more likely to buy from.

7.1 Experimental Setup

7.1.1 Gold Standard

As the gold standard dataset for the experiments we use the 13,619 eBay users who connected to Facebook, as described in Section 3. For each user u the gold standard ranks categories by assigning to each category e the ranking score:

$$gsRank(u, e_i) = \frac{purc(u, e_i)}{\sum_{e \in E} purc(u, e)} \qquad (3)$$

establishing the rank :

$$e_i \succ e_j \iff gsRank(u, e_i) > gsRank(u, e_j) \qquad (4)$$

Categories with the same ranking score are considered ties. For example if a user buys 5 items in *Music*, 3 in *Crafts* and 0 in *Electronics*, the gold-standard ranking for the user will be : *Music* $\succ$ *Crafts* $\succ$ *Electronics*.

The ideal prediction algorithm should provide in output for each user a category ranking equivalent to the gold standard.

7.1.2 Evaluation Measures

To evaluate the prediction models we use the following measures :

Normalized Discounted Cumulative Gain (NDCG). For each user we define Discounted Cumulative Gain (DCG) [17] at position k as:

$$DCG_k = \sum_{i=1}^{k} \frac{w(i)}{log(i+1)} \qquad (5)$$

where $w(i)$ is the relevance weight of the category ranked in position i by the algorithm. We set the relevance weight as follows:

$$w(i) = \frac{purc(e_i)}{\sum_{e \in E} purc(e)} \qquad (6)$$

where $purc(e_i)$ is the number of items bought by the user in category e_i. We also define IDCG (ideal DCG) at position k as the DCG of the gold-standard at k. NDCG at position k is defined as $\frac{DCG_k}{IDCG_k}$.

Precision at Rank k (P_k). Given a position k in the predicted ranking for a given user, P_k is defined as:

$$P_k = \frac{\sum_{i=1}^{k} B(e_i)}{k} \qquad (7)$$

where $B(e_i)$ equals 1 if the user bought at least one item from category e_i and zero otherwise. We compute P_k for each position, until we reach the position at which the algorithm has retrieved all categories with $B(e_i) = 1$.

Note that we do not use any ranking correlation coefficient for our evaluation (e.g. Spearman or Kendall Tau). Given that we are solving a ranking problem, this choice may seem counterintuitive. However, in our case we are not interested in computing how similar two rankings are as a whole, but just in how good an algorithm is in catching the correct categories as early as possible. In this case, NDCG and precision at rank are more reliable measures.

We evaluate our ranking models using 10-fold cross validation in order to reliably compute statistical significance values. For each fold we use 90% of the users as training and 10% as testing. We compute the above measures for each fold by averaging the measures over all testing users.

7.1.3 Baseline and Learning Models

Baseline. A reasonable system that ranks categories according to their *popularity*, i.e., the number of users in the training set who has bought from the category.

Supervised Mapping. We also experiment with a simple supervised model. In the training phase, we build a bipartite graph where the left side nodes are Facebook categories and the right side nodes are eBay meta-categories. We draw an edge between a Facebook category f and an eBay meta-category e if there exists at least one user who likes a page in f and have bought an item in e. The weight of the edge is computed as:

$$w(f, e) = |f, e| \qquad (8)$$

where $|f, e|$ is the number of users who like at least one page in f and have bought from e. In the testing phase,

for each user u and eBay meta-category e we compute the ranking score $\sum_{f \in F_u} w(f, e)$, where F_u is the set of Facebook categories that user u likes at least once. The ranking score is used to produce the output ranking for each user.

Naive Bayes (NB) Classification. We use a standard Naive Bayes model which for each user-category pair predicts the probability that the user will purchase from the category. The algorithm returns the ranked list of categories for each user.

Logistic Regression (LR). We use LibLinear [9] to build a regression model for each eBay meta-category e, for a total of 35 models. For training, a user u is represented by a feature vector (features are described in Section 7.1.4), and the label is the ranking score $gsRank(u, e)$. During testing, for each user we collect the predicted scores for each category as produced by the 35 models, and rank the categories accordingly. The $L2$ regularization parameter is optimized on a subset of the training set.

Support Vector Machines (SVM) Classification. We use SVM^{light} [19] to build a SVM classification model for each eBay meta-category e. For training, positive examples are users who buy at least one item in e. An equal number of random negative examples is provided. During testing, for each unknown user SVM returns a confidence score [19] that we use for ranking[3]. SVM parameters are chosen by grid search on a subset of the training sets. We report results for a Radial Basic Function (RBF) kernel. Results for the linear kernel are comparable to or below RBF.

7.1.4 Experimented Features

We apply all our machine learning algorithms (Naive Bayes, Logistic Regression, and SVM classification) using various feature families. We group features in the following four families.

Demographics (D). We have shown in Section 5 that different gender and age groups tend to buy in specific eBay categories. It is therefore natural to use the demographic information as features for the learning algorithms.

We use a total of eight binary features to represent each gender (male or female) and age group (18-24, 25-34, 35-44, 45-54, 55-64, 65+), where the feature value is 1 if the user is of a given gender/age group, 0 otherwise.

Facebook Categories (F). This feature family includes 214 features, one for each Facebook category in the dataset. For each user u and Facebook category f the feature value is computed using TF-IDF [22] as follows:

$$tfidf(u, f) = \frac{like(u, f)}{\max_{f_i \in F} like(u, f_i)} \cdot \log \frac{|U|}{|(U, f)|} \qquad (9)$$

where $like(u, f)$ is the number of page likes by user u in category f, and $|(U, f)|$ is the number of users who like at least one page in category f.

---

[3]We also experimented with SVM^{rank}, as it appears to be a natural choice for ranking problems. However, the algorithm did not perform well in our experiment, probably due to the peculiar size of the problem (many users, few categories) and the lack of category-based features. We omit SVM^{rank} results for space reasons.

Facebook Likes (L). In addition to Facebook categories, we also experiment with features derived directly from the liked pages. The intuition is that category features may be too generic to capture useful correlations with the eBay categories that need to be predicted; or even worse, there may be no Facebook categories predictive of an eBay category. In such cases, page-level features may help. The values of these features is computed similarly to Facebook categories, i.e., by computing the $tfidf$ between users and likes.

This feature family includes all the 1.3 million pages liked by users in our dataset. Since the number of irrelevant features may be high, we perform feature selection before feeding the feature vectors to the machine learning algorithms. The feature selection strategy we use is Information Gain, since it has proved to be effective in many learning tasks, e.g. text categorization [35]. Information Gain computes the number of bits of information obtained for the prediction task from a new feature. The information gain of a like l is formally defined as follows:

$$
\begin{aligned}
IG(l) \quad = \quad & -\sum_{i=1}^{|E|} P(e_i) \log P(e_i) \\
& + P(l) \sum_{i=|E|}^{l} P(e_i|l) \log P(e_i|l) \\
& + P(\bar{l}) \sum_{i=1}^{|E|} P(e_i|\bar{l}) \log P(e_i|\bar{l}). \quad (10)
\end{aligned}
$$

where $|E|$ is the number of eBay categories; $P(e_i)$ is approximated by the fraction of training users who buy from category e_i and $P(l)$ by the fraction of users who like l; $P(e_i|l)$ is approximated by the fraction of users who like l and also buy from category e_i; and $P(\bar{l})$ is approximated by the fraction of users who do not like l.

For each unique like in our dataset, we compute its information gain and remove all likes whose information gain is less than a pre-defined threshold (5% of maximum IG). The underlying reasoning is that likes with high information gain are more useful for category prediction. Hence, the quality of a like feature is proportional to its information gain score, i.e., the higher the $IG(l)$ score, the better the feature is. Each user is represented using a feature vector of $tfidf$ values of top likes.

Using the eBay category *Clothing, Shoes & Accessories* as an example, the top 10 Facebook likes ranked by IG are: *Sephora, Victoria's Secret, Victoria's Secret Pink, Bath & Body Works, JustFab, Macy's, Coach, ShoeDazzle, eBay Fashion, MAC Cosmetics*. As we can see, the top likes are highly related to the *Clothing, Shoes & Accessories* category.

Facebook n-grams (N). We also experiment with n-grams ($n=1,2,3$) derived from individual Facebook page names, e.g. for the Facebook page *Boston Running Club* we will create a set of candidate n-grams: {*boston, running, club, boston running, running club, boston running club*}. Since we have 1.3 million Facebook pages, the number of derived n-grams will be even larger. We therefore perform feature selection also in this case, to choose the most informative unigrams, bigrams and trigrams. Each user is represented using a feature vector of $tfidf$ values of top n-grams.
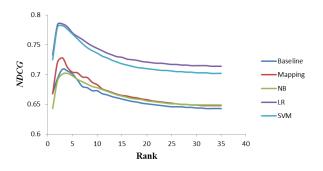

Figure 8: $NDCG_k$ distributions of algorithms using all four feature families.

7.2 Experimental Results

7.2.1 Algorithms

Table 4 reports the results of different algorithms using the complete set of features (demographics, Facebook categories, likes and n-grams) with feature selection. Figure 8 reports the trend of $NDCG$ at different rank levels, for all the experimented algorithms.

Logistic Regression and SVM significantly outperform[4] the baseline system at all rank levels in both precision and $NDCG$. The Mapping system and Naive Bayes show significantly lower accuracy.

In general the Baseline system has good performance. Predicting meta-categories by simply ranking by popularity proves to be a hard baseline to beat, as we would have expected from the statistics reported in Figure 5.

The Mapping algorithm performs slightly better than Baseline, but without statistical significance. Overall, the performances of the two algorithms are very similar. In order to better understand the reason for this behavior, we measure the similarity of the ranking produced by the two algorithms. We do this by computing the Jaccard similarity coefficient J on the set of top 7 ranked categories[5]. We obtain J=0.74, i.e., on average Baseline and Mapping share 5 out of the top 7 predicted categories. The reason for this high correlation is that the weight in Equation 8 promotes eBay categories that are very popular among users, similarly to what Baseline does.

Naive Bayes is the worst performing algorithm, showing performance below or very close to the baseline. A possible explanation is that Naive Bayes assumes feature independence, while the features derived from social media profiles are not necessarily independent of one another. For example, the category *Sports* and *Sport Teams* are highly dependent on each other. The Jaccard coefficient between Naive Bayes and Baseline is $J = 0.52$, showing that the Naive Bayes system is mildly correlated to Baseline, but not as much as Mapping.

The top performing systems, Logistic Regression and SVM, are far apart from all others. The good performance of SVM

---

[4]We calculate the 95% confidence interval for each evaluation measure using the sample obtained from the 10-fold cross validation. If the confidence intervals for two samples do not overlap, then there is a significant difference.

[5]On average the users in our dataset buy from 7 eBay meta-categories.

Table 4: Experimental Results of Different Algorithms Using All Feature Families.

Algorithm	P_1	P_2	P_3	P_5	P_7	$NDCG_1$	$NDCG_2$	$NDCG_3$	$NDCG_5$	$NDCG_7$
Baseline	0.668	0.547	0.513	0.454	0.451	0.668	0.694	0.709	0.701	0.680
Mapping	0.668	0.571	0.524	0.494	0.489	0.668	0.721	0.728	0.704	0.696
NB	0.643	0.560	0.502	0.477	0.469	0.643	0.690	0.701	0.698	0.688
LR	**0.733†**	**0.655†**	**0.628†**	**0.582†**	**0.565†**	**0.733†**	**0.784†**	**0.785†**	**0.770†**	**0.759†**
SVM	0.725†	0.653†	0.622†	0.570†	0.530†	0.725†	0.780†	0.782†	0.768†	0.752†

† indicates statistical significance at 0.95 level with respect to *Baseline*.

is expected. A large volume of previous work has already shown its superior classification power with respect to Naive Bayes and other basic approaches. As for the good performance of Logistic Regression, it indicates that using a regression approach to purchase prediction is a viable, promising direction.

Overall, the results suggest that SVM and Logistic Regression make a much better use of the social features than Mapping and Naive Bayes. These two latter systems appear to be more influenced by the strong meta-category prior probabilities (Figure 5) than by the features themselves.

7.2.2 Feature Analysis

In the previous section we experimented with different algorithms using the full set of features: demographics, Facebook categories, likes and n-grams. In this section we report experimental results for our best performing algorithm, Logistic Regression, with the different feature families described in Section 7.1.4, both in isolation and in combination, in order to analyze how they contribute to the overall task[6].

Table 5 summarizes experimental results for the different feature families.

All feature families taken in isolation outperform the baseline (row 2-5 of Table 5). Demographic features (D) show the smallest improvement. However, results still indicate that simple demographic information easily available on social media, such as age and gender, can help significantly in the purchase prediction task. This is particularly important for those e-commerce applications that do not request social media users to share the complete list of likes.

All other individual feature families, i.e., Facebook categories (F), likes (L) and n-grams (N), significantly outperform D features. This is not surprising because these feature families provide much richer and more relevant information with respect to age and gender. Intuitively, it may often be the case that D features are subsumed by F, L and N. As a matter of fact, we showed at the end of Section 5 that the Facebook categories preferred by a user are usually correlated to the user's gender and age group.

Within the four individual feature families, F performs the best, indicating that social media profiles at the category level convey enough information for predicting users' purchase behaviors on e-commerce sites. However the small difference in performance of F with respect to L and N also suggests that F, L, and N mostly convey the same information. From one side this is expected, since all these three feature families are generated from the same source (the list

of users' likes). From the other side, we would have expected L and N to slightly outperform F, since they carry more fine-grained information. A closer analysis of the L and N feature families reveals that these features are often too sparse, thus limiting their prediction power. On the contrary, F features are general enough to provide generalization power across users. We leave as a future work the exploration of "middle-ground" features that have a degree of generalization in between F and L, such as clusters of likes automatically discovered using topic model techniques or clustering algorithms.

When the best individual feature family F is combined with other feature families (rows 6-12), we see a small additional gain in prediction quality. For example, when Facebook categories and likes are combined, P_1 goes up from 0.708 for F and 0.706 for L to 0.718. In general, the more feature families we use, the greater the gain in prediction quality. However, the gain in performance is very small. As already outlined in the previous paragraph, L and N come from the same source of F but have sparsity problems. Therefore, they do not carry new relevant information with respect to F.

It is finally worth mentioning that the dimensional space of Facebook likes and n-grams is much larger than that of Facebook categories. Hence, when computational cost is concerned, Facebook categories are more favorable.

Feature Selection. All results reported so far use Information Gain for selecting top likes and n-grams. To check the effect of feature selection, we run Naive Bayes and Logistic Regression on the whole set of features but without any feature selection[7]. Results show that both Naive Bayes and Logistic Regression perform worse when feature selection is not performed. For example, P_1 for Naive Bayes goes from 0.643 with feature selection down to 0.376 without feature selection and P_2 goes from 0.560 down to 0.392.

8. CONCLUSIONS

In this paper we study the relations and interactions between social media profiles and purchase behaviors on e-commerce websites. We demonstrate that there are significant correlations between social media information and online purchases. We also develop machine learning algorithms that use different feature families derived from Facebook profiles for solving the task of purchase behavior prediction in a "cold start" setting. Results on a large dataset of eBay users who connected to Facebook show that purchase behaviors can be successfully predicted using social media information alone.

---

[6]Logistic Regression performed better or comparably to other algorithms in all feature combinations. We therefore use only Logistic Regression to illustrate how different feature families perform in the prediction task.

[7]Running SVM with RBF kernel without feature selection would be computationally unfeasible.

Table 5: Experimental Results of Logistic Regression Using Isolate and Combined Feature Families.

Feature Sets	P_1	P_2	P_3	P_4	P_5	$NDCG_1$	$NDCG_2$	$NDCG_3$	$NDCG_5$	$NDCG_7$
Baseline	0.668	0.547	0.513	0.454	0.451	0.668	0.694	0.709	0.701	0.680
D	0.670	0.593	0.565	0.534	0.504	0.670	0.728	0.735	0.721	0.710
F	0.708	0.652	0.621	0.572	0.549	0.708	0.761	0.765	0.749	0.736
L	0.706	0.647	0.613	0.568	0.538	0.706	0.759	0.761	0.748	0.733
N	0.705	0.636	0.605	0.563	0.533	0.705	0.757	0.760	0.745	0.732
$F + D$	0.715	0.649	0.623	0.575	0.553	0.715	0.766	0.770	0.765	0.753
$F + L$	0.718	**0.657**	0.625	0.576	0.555	0.718	0.770	0.775	0.768	0.755
$F + N$	0.717	0.655	0.623	0.578	0.552	0.717	0.769	0.776	0.766	0.752
$F + D + L$	0.723	0.653	**0.634**	**0.586**	0.559	0.723	0.775	0.782	**0.771**	0.756
$F + D + N$	0.722	**0.657**	0.624	0.577	0.558	0.721	0.773	0.780	0.770	0.758
$F + L + N$	0.729	0.656	0.629	0.581	0.563	0.729	0.780	0.778	0.763	0.750
$F+D+L+N$	**0.733**	0.655	0.628	0.582	**0.565**	**0.733**	**0.784**	**0.785**	0.770	**0.759**

D: Demographics; F: Facebook categories; L: Top Facebook likes; N: Top n-grams(n=1,2,3)

In the future we will explore several research directions, including the following:

- We will go a step further and apply our system to predict behaviors at the subcategory level, e.g., *Women's Handbags* instead of the more generic parent category *Clothing, Shoes & Accessories*. We will go even deeper to verify the feasibility of predicting purchases directly at the product level: this would be considered as a full-fledge recommender system.

- We will extend our feature sets to include other types of features, such as clusters of likes automatically inferred using topic models, and the likes of the user's friends. Even though eBay does not ask users to share the likes of their friends for privacy reasons, these may be provided in other contexts, e.g., at Amazon.com.

- We will experiment with more sophisticated models integrating social graph information, when available.

- We will explore if and how Facebook information can improve existing product recommendation systems that rely on purchase history.

- We will integrate our prediction models into existing products: to offer, for example, a more personalized experience for social media users when they visit an e-commerce website; or to automatically redirect users to the vertical of interest when they access the website.

To the best of our knowledge our study is the first attempt to predict e-commerce behaviors using only social media information. It is our hope that it will help and inspire e-commerce companies to develop better and new recommendation engines that leverage social media features in a time when the issue of monetizing social media information is being more and more debated.

9. ACKNOWLEDGEMENTS

10. REFERENCES

[1] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. In *WebSci'11 Conference Proceedings*, 2011.

[2] S. Aciar, D. Zhang, S. Simoff, and J. Debenham. Recommender system based on consumer product reviews. In *WI'06 Conference Proceedings*, pages 719–723, 2006.

[3] R. Bhatt, V. Chaoji, and R. Parekh. Predicting product adoption in large-scale social networks. In *CIKM'10 Conference Proceedings*, 2010.

[4] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, 1998.

[5] R. Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interation*, 12(4):331–370, 2002.

[6] J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi. Short and tweet: experiments on recommending content from information streams. In *CHI'10 Conference Proceedings*, 2010.

[7] Y.-B. Chiu, C.-P. Lin, and L.-L. Tang. Gender differs: assessing a model of online purchase intentions in e-tail service. *International Journal of Service Industry Management*, 16(5):416–435, 2005.

[8] M. D. Ekstrand, J. T. Riedl, and J. A. Konstan. Collaborative filtering recommender systems. *Foundations and Trends in Human-Computer Interaction*, 4(2), 2011.

[9] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008.

[10] E. Garbarino and M. Strahilevitz. Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation. *Journal of Business Research*, 57(7):768–775, 2004.

[11] S. Guo, M. Wang, and J. Leskovec. The role of social networks in online shopping: information passing, price of trust, and consumer choice. In *ACM EC'11 Conference Proceeding*, pages 157–166, 2011.

[12] J. Hannon, M. Bennett, and B. Smyth. Recommending twitter users to follow using content and collaborative filtering approaches. In *ACM RecSys'10 Conference Proceedings*, pages 199–206, Barcelona, Spain, 2010.

[13] S. Hill, F. Provost, and C. Volinsky. Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science*, 21(2):256–276, 2006.

[14] R. Iyengar, S. Han, and S. Gupta. Do friends influence purchases in a social network? Working paper, Harvard Business School, 2009.

[15] M. Jamali and M. Ester. Trustwalker: a random walk model for combining trust-based and item-based recommendation. In *KDD'09 Conference Proceedings*, pages 397–406, 2009.

[16] D. Jannach. *Recommender Systems : an Introduction*. Cambridge University Press, 2011.

[17] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems*, 20(4):422–446, 2002.

[18] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, and S. Yang. Social contextual recommendation. In *KDD'12 Conference Proceedings*, pages 45–54, 2012.

[19] T. Joachims. *Learning to Classify Text Using Support Vector Machines*. Dissertation. Kluwer, 2002.

[20] A. K. Kau, Y. E. Tang, and S. Ghose. Typology of online shoppers. *Journal of Consumer Marketing*, 20(2):139–156, 2003.

[21] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[22] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[23] M. Nikulin. Chi-squared test for normality. In *Proceedings of the International Vilnius Conference on Probability Theory and Mathematical Statistics*, pages 119–122, 1973.

[24] M. Pennacchiotti and S. Gurumurthy. Investigating topic models for social media user recommendation. In *WWW'11 Conference Proceedings*, 2011.

[25] D. Ramage, S. Dumais, , and D. Liebling. Characterizing microblogs with topic models. In *AAAI ICWSM'10 Conference Proceedings*, 2010.

[26] P. Resnick and H. R. Varian. Recommender systems. *Special Issue of the Communications of the ACM*, 40(3), 1997.

[27] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor. Recommender systems handbook. *Media*, 54(3):217–253, 2011.

[28] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Analysis of recommendation algorithms for e-commerce. In *EC'00 Conference Proceedings*, 2000.

[29] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *WWW'01 Conference Proceeding*, 2001.

[30] J. Schafer, J. Konstan, and J. Riedi. Recommender systems in e-commerce. In *ACM EC'99 Conference Proceedings*, 1999.

[31] S. Sen, J. Vig, and J. Riedl. Tagommenders: connecting users to items through tags. In *WWW'09 Conference Proceedings*, pages 671–680, 2009.

[32] W. Shi, P. Wu, W. Zhou, and J. Chen. Gender differences in purchase intention on mobile data services. In *IEEE CSO'09 Conference Proceedings*, pages 773–777, 2009.

[33] R. T. A. Wietsma and F. Ricci. Product reviews in mobile decision aid systems. In *Proceedings of the PERMID*, pages 15–18, Munich, Germany, 2005.

[34] R. Yan, M. Lapata, and X. Li. Tweet recommendation with graph co-ranking. In *ACL'12 Conference Proceedings*, volume 1, 2012.

[35] Y. Yang and J. O. Pedersen. A Comparative Study on Feature Selection in Text Categorization. In *ICML'97 Conference Proceedings*, Nashville, TN, 1997.