Corpus for Customer Purchase Behavior Prediction in Social Media

Shigeyuki Sakaki*†, Francine Chen*, Mandy Korpusik*‡, and Yan-Ying Chen*

*FX Palo Alto Laboratory, Inc., Palo Alto, California, USA
†Fuji Xerox Co., Ltd., Yokohama-shi, Kanagawa, Japan
‡CSAIL Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
E-mail: sakaki.shigeyuki@fujixerox.co.jp, chen@fxpal.com, korpusik@mit.edu, yanying@fxpal.com

Abstract

Many people post about their daily life on social media. These posts may include information about the purchase activity of people, and insights useful to companies can be derived from them: e.g. profile information of a user who mentioned something about their product. As a further advanced analysis, we consider extracting users who are likely to buy a product from the set of users who mentioned that the product is attractive.
In this paper, we report our methodology for building a corpus for Twitter user purchase behavior prediction. First, we collected Twitter users who posted a want phrase + product name: e.g. "want a Xperia" as candidate want users, and also candidate bought users in the same way. Then, we asked an annotator to judge whether a candidate user actually bought a product. We also annotated whether tweets randomly sampled from want/bought user timelines are relevant or not to purchase. In this annotation, 58% of want user tweets and 35% of bought user tweets were annotated as relevant. Our data indicate that information embedded in timeline tweets can be used to predict purchase behavior of tweeted products.

Keywords: purchase behavior, corpus for machine learning, micro-blogging

1. Introduction

Recently, many companies have tried obtaining insights into possible customers from Twitter. An example of such an application is inferring user profile information (Ikeda et al., 2013; Sakaki et al., 2014; Taniguchi et al., 2015), for use in marketing and in targeted advertising. Another application is to identify people who are likely to buy a company's products. By identifying prospective buyers, companies can remove barriers to purchasing their products, such as informing customers about product features through advertisements, offering coupons, and introducing shop or sales people. In addition, companies can estimate future sales from the expected number of users.
In Twitter, there is an abundance of tweets indicating that the owners of the tweets want something (e.g. "I want an iPhone", "I plan to buy a Nexus"). Figure 1 shows an example of a portion of the tweet timeline of a Twitter user who wants an iPhone. This user posted two tweets (want tweets) indicating interest in an iPhone, and four tweets later, he posted a tweet announcing that he actually bought an iPhone. As in the example timeline, we expect that purchase behavior prediction is available by utilizing text information around want tweets.

Figure 1: Timeline of tweets by a Twitter user who bought an iPhone. (most recent tweet on top)

We created a tweet corpus for use in research on automatically predicting whether a Twitter user will buy a product that was mentioned in their past tweets. We automatically collected English tweets related to purchase behavior and then manually annotated whether a user purchased a product. In creating this corpus, there were a number of challenges, including defining methods for identifying users who might buy a product and in getting reliable judgments of user purchase behavior. In this paper, we report our methodology for annotating Twitter user data by human annotators judging purchase behavior. We collected the timeline tweets of Twitter users who posted about wanting or buying products. For these tweet data, we performed two types of annotation: whether or not a Twitter user eventually bought a target product, and whether or not a tweet that mentions a product is related to purchase activities. Finally, we obtained two kinds of corpora: 1) a bought-or-not annotation corpus and 2) a relevant-or-not annotation corpus. In the following sections, we will explain our data collection, annotation policy, and observed insight into whether there are any signs in tweets of Twitter users who are going to buy something.

2. Prior Work

In research about customer purchasing, there are some studies about recommender systems. A popular approach used by recommendation systems is collaborative filtering for identifying users similar to a target user based on their purchase history, and then recommending products that similar users have already bought but the target user has not (Schafer et al., 1999; Sarwar et al., 2000). However, in this approach, it is required that a target user has bought something before. Furthermore, it is difficult to recommend

products that customers rarely buy i.e. car, smart phone, camera, game console. To solve such a problem, Zhang & Pennacchiott (2013) developed a "cold start" system to predict product categories a Facebook user will buy from, using Facebook profile information, gender, age, and which pages he/she "likes". Sen et al. (2009) implemented a content-based movie recommender system capable of "cold start" by using preference tags that customers labeled movies within in a movie review service. In contrast to these works, our goal is to infer the future purchase behavior of a customer who is interested in a product. This requires differentiating between two kinds of users: a user who is just curious about a product and a user who is likely to buy a product.

Twitter provides a glimpse of some of the daily thoughts or activities of users. Adamopoulos & Todri (2015) improved the accuracy of their recommender system by using Twitter data to estimate personality traits of users who shared their Amazon purchases on Twitter. However, this data only applies to users who actually bought products while our problem statement also requires users who eventually didn't buy products mentioned in a past tweet. Most other recommender research is based on customer purchase histories and doesn't include any information of customer daily activities. Furthermore, most open data provided for recommender systems are either customer purchase history or movie reviews (RecSys, 2011).

Given the types of available data, we decided to create a corpus of tweets annotated with which users bought products and which users did not. Our approach is novel, applicable to cold start systems, and can be used by many more consumer companies since it does not need to collect information related to real user purchase logs for prediction.

3. Data Collection

We collected Twitter data by identifying tweets containing cue phrases. We defined "want" tweets and "bought" tweets and collected the tweets of those Twitter users: the former indicates a user who is curious about a product and the latter is a user who bought a product. First, we collected want/bought tweets by using text cue phrases (want/bought phrase + product name: e.g. "want a Xperia", "my new Canon 7d"). Second, we collected the timeline tweets of the people who posted want/bought tweets. The detailed data collection flow is explained in the following steps.

Defining want and bought phrases:

We created a set of regular expressions that may indicate that a user bought or wanted one of the products for the use of text cue phrases (Table 1).

Extracting product names from eBay pages:

Since people rarely discuss frequently bought products such as daily necessities e.g., shampoo, detergent, we focused on product categories that users buy only occasionally: mobile device, camera, and game console. We first identified a set of product names (i.e., models) for each product category from eBay listings. Similar names were merged, e.g., "iPhone 4", "iPhone5", and "iPhone 6s" were merged into "iPhone", resulting in 80 mobile device names, 146 camera names, and 14 game console names. Table 2 shows examples of product names.

Want Phrase	• should I buy a/an • should I go for a/an • should I upgrade to a/n • plan to buy a/an • want a/an • wanna
Bought Phrase	• bought a/an • bought a new • got a new • gotta new • my new • owner of a new • paid for a/an • splurged on a/an • sprang for a/an

Table 1. Want/bought phrases.

Mobile Device	blackberry, iphone, ipad, nexus 7, kindle fire
Camera	alpha nex, canon eos, fujifilm x-t1, go pro hero, nikon 1
Game Console	microsoft xbox, nintendo 3ds, nintendo wii, playstation, ps vita

Table 2. 5 examples of product names in each category.

Product Category	Want Tweet	Bought Tweet
Mobile Device	14892	34945
Camera	7406	20902
Game Console	28318	44281

Table 3. Number of collected want/bought tweets.

Collecting want/bought tweets:

We created a search query by combining a product name and a cue phrase. Tweets containing a bought or want expression for one of the eBay products were then collected using the Twitter search API.

Collecting timeline tweets of a want/bought tweet user:

The users associated with each of these tweets were identified from the tweet meta-data and their tweets around want/bought tweets were collected using the Twitter search and timeline APIs. We considered users identified from "bought" regular expressions to be candidate buy users, and users identified from "want" regular expressions to be candidate want users.

By executing the above steps, we obtained tens of thousands of want/bought tweets (Table 3). Examples of collected tweets are shown below.

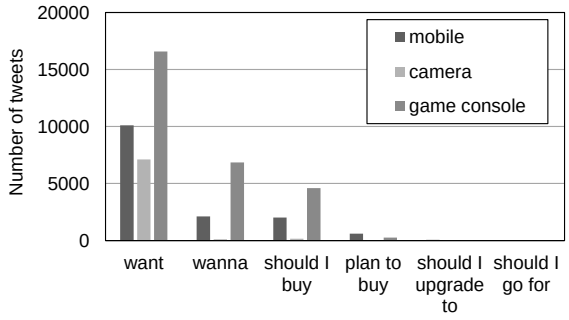

Figure 1. Frequently occurring phrases in want tweets of Table 3.

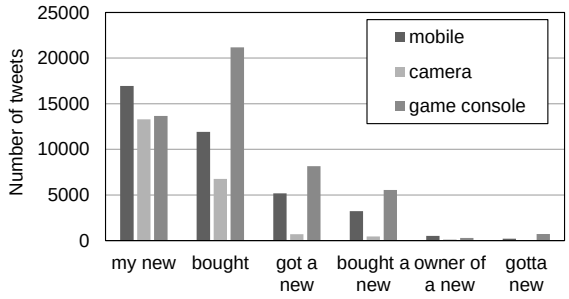

Figure 2. The top 6 most frequently occurring phrases in bought tweets of Table 3.

Product Category	Candidate Want User	Candidate Buy User
Mobile Device	2200	8425
Camera	1790	6778
Game Console	2574	7932

Table 4. Number of successfully collected user timelines.

Want tweet

- I *want a blackberry Q10* already.
- I *want a canon 5d* so bad. Why can't I have it?
- *should i buy a ps3!?* ehhhhh i dunno what to do my 360 is broke/gone to the wind.

Bought tweet

- Wooooo typing this off *my new iPhone!*
- packing for my china trip...I leave tomorrow and get back June 7! *Bought a Nikon D7000* to capture the memories.
- *Bought a PlayStation 4* yesterday.

Figures 1 and 2 show the frequency of want/bought phrases in each category. The most popular want phrase is "want", and the most popular bought phrases are "my new" and "bought". The frequency of commonly used phrases is not very different between categories, suggesting that we may be able to reuse the phrases for different products.

We collected thousands of timeline tweets for users who posted one of the collected want/bought tweets (Table 4). Because we couldn't access the timeline tweets of older want/bought tweets due to an API limitation, the number of users whose timeline tweets were successfully collected is smaller than that of want/bought tweets. The candidate want/bought users can be used for distant supervision training. For example, in an approach using distant supervision over phrases, Bollen et al. (2009) use the phrases e.g., "I feel" to extract a tweet indicates sentiment of a user.

4. Data Labeling

We hired an expert annotator to label whether a candidate want/bought user eventually bought a product or not. Annotation of candidate bought users consisted of two steps. First, we asked the annotator to label a candidate buy user as buy/not buy by examining their bought tweet as identified in the previous section. If the annotator labeled "not buy", then an extra annotation task is performed in which the annotator views all of the user's tweets to identify those that include: 1) a product name or a category name i.e., "mobile", "camera", "game" to identify product tweets and 2) a first person pronoun, i.e., "I", "my", or "me" in order to identify tweets which are related to a tweet owner. From these tweets, the annotator then judges whether or not that user really bought a product. The following set of tweets is an example of a user whose tweets require the annotator to perform the extra annotation task.

Bought tweet

- *My* new *HTC Desire* wallpaper. Wanna guess what this is? http://twitpic.com/xxxx

Product tweets

- Just gave *my* ancient PDA a hard reset and installed Opera *Mobile* 10 b
- Folks! *My* first ever tweet with my newly acquired *HTC Desire*.

In the case of the bought tweet above, the annotator can't judge whether bought tweet means the user actually bought a HTC desire or not. Then, the annotator performs the extra task and checks that the tweets include product names and a personal pronoun. In the above case, from the second product tweet, the annotator can know the user actually bought HTC Desire.

For the candidate want users, we implemented a single annotation step in which we showed the annotator all the tweets that satisfy the same two conditions used in the extra task during candidate bought user annotation. After checking the all product tweets, the annotator determines which buy/not buy label is appropriate. For the judgment of buy/not buy, we defined the following "gray areas" as buy: 1) The product must be considered new and was bought within the last week. 2) A user could: order a product to arrive within a week, say when the product is arriving, upload a video to YouTube about a review of the product, or trade the product soon after the purchase. These conditions did not indicate buy: considering a product, going to buy a product, being given a product, retweeting

User Type Label	Candidate Want User	Candidate Buy User
Buy	246	2503
Not buy	1227	303

Table 5. Corpus statistics showing the number of users associated with the collected tweets containing want or buy expressions that are labeled as buy and not-buy users.

Product Category	Candidate Want User	Candidate Buy User
Mobile Device	696/1500	503/1500
Camera	438/500	339/500
Game Console	325/500	40/500

Table 6. Data statistics for a random sample of tweets labeled as relevant in three product categories. E.g., 696 tweets in the random sample of 1500 tweets from candidate want users were annotated as relevant to purchase behavior.

another user who announced buying a product.

We labeled whether a random sample of tweets was "relevant" to purchase activities. Since users post various kinds of tweets, we believe that relevant/irrelevant information about one tweet will be useful for excluding noise in tweets. We also wanted to know whether our Twitter data includes many purchase relevant tweets. We extracted a random sample of tweets with want phrases and product names (*want* tweets) from the timeline tweets of candidate want/buy users. The expert annotator implemented the relevant/irrelevant annotations according to the following rules: considered buying is relevant, comparing two products to buy is relevant, wanting a product is relevant, retweeting another user is relevant.

In this study, we annotated a subset of the candidate users: for buy/not buy annotation, we extracted 2806 candidate buy users (mobile device: 1951 users, camera: 855 users) and 1473 candidate want users (mobile device: 1007 users, camera: 466). For relevant/irrelevant annotation, we extracted 1500 want tweets from each type of user about mobile device category, and 500 tweets from each type of user about camera and game console category. The numbers of tweets extracted from each user type and each product category are shown in the denominators in Table 6.

5. Annotation Results and Data Analysis

From the annotation results, we can examine how often users expressing one of our defined indicator phrases actually made a purchase. In Table 5 we observe that only a small percentage of users who indicate that they want a product by tweeting one of the phrases in Table 1 actually bought a product. In contrast, we note that many of the users expressing a buy phrase did buy a product.

To identify the most common terms used in actual bought tweets, we examined the bought tweets posted by 246 candidate want users who were annotated as having bought a product of interest. We observed that the most popular terms contained in the want users' bought tweets were "new", "got" and "bought" and that each term appeared 110 times, 64 times, and 24 times, respectively. A buy expression containing at least one of these three terms occurred in 67% of the tweets, indicating that the bought phrase list that we used in Section 3 captures the majority of buy tweets, as well as the three most popular "buy" indicator terms. The next two most frequently occurring terms were "ordered" and "get", which occurred 17 times and 14 times, respectively. In the next data collection phase, we plan to add "ordered" and "get" to the query phrases in Table 1 to improve coverage.

We observe in Table 6 that the percentage of relevant tweets varies depending on the product. Cameras had more relevant than not relevant tweets, while the opposite was true for mobile devices and game consoles. A sizeable number of tweets posted by candidate buy users were annotated as relevant, indicating that some Twitter users post purchase-relevant tweets before they buy these types of products. From this, we expect that purchase inference from text information can be performed for these users.

6. Future Work

We plan to build a machine learning classifier trained on our annotation data which infers purchase behavior of Twitter user. We will train a classifier using "buy" users as positive samples and "not buy" users as negative samples; the timeline tweets of each user will be used as features. Since information of whether a tweet is relevant or not is considered to be useful for purchase behavior prediction, the classifier will also learn the difference between relevant tweets and irrelevant tweets. We expect the classifier trained by two kinds of annotation data has ability to predict whether or not a user will buy a product that they have mentioned.

The cue phrases-based corpus creation methodology used in this study can be applied to other language and other labeling tasks by defining phrases. Using the corpus creation methodology in this paper and machine learning methods, we expect that various kinds of automatic labeling task on social media e.g., location detection of a user, activity detection of a user, will be possible.

References

Adamopoulos, P. and Todri, V. (2015). Personality-Based Recommendations: Evidence from Amazon.com. In *Proceedings of the 9th ACM International Conference on Recommender Systems.*

Bollen, J., Mao, H. and Zeng, X. (2011). Twitter mood predicts the stock market, *Journal of Computational Science* 2(1), pages 1-8.

Ikeda, K., Hattori, G., Ono, C., Asoh, H. and Higashino, T. (2013). Twitter User Profiling Based on Text and Community Mining for Market Analysis, *Knowledge Based Systems* 51, pages 35-47.

RecSys. (2011). ACM RecSys wiki CATEGORY:

DATASET, Available at: http://www.recsyswiki.com/wiki/Category:Dataset.

Sakaki, S., Miura, Y., Ma, X., Hattori, K. and Ohkuma T. (2014). Twitter User Gender Inference Using Combined Analysis of Text and Image Processing. In *Proceedings of the 3rd Workshop on Vision and Language*, pages 54-61.

Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. (2000). Analysis of Recommendation Algorithms for E-Commerce. In *Proceedings of the 22nd ACM International Conference on Electronic Commerce*, pages 158-167.

Schafer, J.B., Konstan, J. and Riedl, J. (1999). Recommender Systems in E-Commerce. In *Proceedings of ACM International Conference on Electric Commerce*, pages 158-166.

Sen, S., Vig, J. and Riedl, J. (2009). Tagommenders: Connecting Users to Items through Tags. In *Proceedings of the 18th International Conference on World Wide Web*, pages 671–680.

Taniguchi, T., Sakaki, S., Shigenaka, S., Tsuboshita, Y. and Ohkuma T. (2015). A Weighted Combination of Text and Image Classifiers for User Gender Inference. In *Proceedings of the 4th Workshop on Vision and Language*, pages 87-93.

Zhang, Y. and Pennacchiotti, M. (2013). Predicting Purchase Behaviors from Social Media. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1521–1532.