Spring 6-28-2016

PREDICTING PRODUCT RETURN RATE WITH “TWEETS”

Yi Ding
National University of Singapore, yiding@comp.nus.edu.sg

Haifeng Xu
National University of Singapore, xu-haif@comp.nus.edu.sg

Bernard C. Y. Tan
National University of Singapore, btan@comp.nus.edu.sg

PREDICTING PRODUCT RETURN RATE WITH "TWEETS"

Yi Ding, School of Computing, National University of Singapore, Singapore, yiding@comp.nus.edu.sg

Haifeng Xu, School of Computing, National University of Singapore, Singapore, xu-haif@comp.nus.edu.sg

Bernard C. Y. Tan, School of Computing, National University of Singapore, Singapore, btan@comp.nus.edu.sg

Abstract

This study classifies posts into four distinct topics and uses their sentiment values to predict product return rate at e-commerce websites. The results reveal that the sentiments of posts related to e-commerce company news (i.e., objective posts) are negatively related to product return rate. On the contrary, the sentiments of social network posts related to product use, purchase and service experiences (i.e., subjective posts) are positively related to product return rate at a focal e-commerce website. The paper contributes to product return research as well as social network prediction research. Practitioners may use the method to predict product return rates using social network posts.

Keywords: product return rate, social network prediction, Latent Dirichlet Allocation, VAR

1 INTRODUCTION

Product returns have been deemed as a major inconvenience for companies' supply chain management and drain on overall profits (Anderson et al. 2009). In the traditional retailing industry, product return rate reaches as high as 8.6% (National Retail Federation 2013). It costs manufacturers and retailers more than $100 billion each year, resulting in an average loss of about 3.8% of their profit (Blanchard 2007). To make matters worse, product return rate is even higher in the maturing online markets due to the higher product uncertainty (Hong and Pavlou 2014). It is reported the return rate ranges from 10% to 30% for different product categories in the e-commerce websites (Internet Retailer 2013). As product returns erode firms' overall profitability, e-commerce companies are seeking solutions to predict consumers' return behavior and tackle the problem (Banjo 2013).

Regarding consumer behavior prediction, previous literature has revealed the predictive power of user-generated content in social networking sites (SNSs). Online users can post product details in SNSs in order to get others' advises. They may also write reviews about products (or services), which are from a certain e-commerce website. These exposures probably affect their friends' perception of the website and influence their purchase (or even return) behavior subsequently. Extant literature has well documented the predictive power of social networks on product sales. For example, Goh et al. (2013) have investigated the impact of user-generated content in social media on driving clothes sales. For the experience goods, Rui and Whinston (2011) have used the posts in Twitter to predict movie sales. Similarly, Dhar and Chang (2009) have indicated the role of content in blogs in predicting music sales. However, little light has shed on the product return rate prediction.

Given the importance of product return and the current inaccurate return rate prediction, this study sets out to investigate the issue by incorporating a new aspect of social media. Posts in social media may provide a way to understand product return. For example, a user complaining about the service of a certain e-commerce platform disseminates a signal of bad return service, which may deter other customers' product return decisions. Previous literature has suggested the prediction accuracy rate will be improved when dividing posts into different genres (Goh et al. 2013; Ding et al. 2014). Different topics can have different impacts on consumer behavior. However, the existing literature about product returns does not consider the influence of different topics of social network posts. This motivates us to ask: *what is the role of "tweets" of different topics in predicting product return rate?*

To address this research question, we collaborate with one of the largest e-commerce websites in China to gain the return rate at a daily base. The other part of the data is the website-related posts in Microblog[1]. We crawl all posts mentioning about the name of the focal e-commerce website at a daily base. To gain a better understanding of the role of online posts in return prediction, we first differentiate the crawled posts according to their topics. Unlike most previous studies which predefined the post categories, we let the post topics emerge from the data using Latent Dirichlet Allocation (LDA) technique and K-means clustering approach. After these, all the posts will be clustered into groups with the same theme. We then conduct sentiment analysis for each post in each cluster. The sentiment value indicates the online user's attitude while writing the post (Turney 2002). Once we have sentiment values of all the posts related to the website, we conduct vector autoregressive model (VAR) to reveal the role of "tweets" in predicting product return in e-commerce website. Specifically, we conduct the analysis in each cluster in order to see the effects of posts with different topics.

This study is expected to have both theoretical and practical contributions. First, it extends the existing literatures about product returns through incorporating the aspect of social media. Second, the study adds to literatures about social media prediction (Asur and Huberman 2010; Rui and Whinston 2011). For practitioners, this study serves as a powerful tool for them to

[1] www.weibo.com: analogous to Twitter

predict the future return rate with social media posts. With the division of post categories, the tool is equipped with more accurate underlying predictive power.

2 CONCEPTUAL BACKGROUND

2.1 Product Return

Previous studies about product returns mainly focused on return policy optimization (Davis et al. 1998; Hess et al. 1996; Padmanabhan and Png 1997). For example, Hess et al. (1996) revealed separated non-refundable charges (e.g., shipping and handling fees) could be used to profitably control product returns in catalog retailing. Following this, Padmanabhan and Png (1997) found a strategic effect of return policies on retail competition and highlighted its profitability implications for a manufacturer. Davis et al. (1998) employed an analytical model to identify the potential causes for variation among firms' return policy. They found that a low-hassle return policy existed when the products were durable, or they were possible for cross-selling, or a high salvage value was included in the returned products. Wood (2001) divided customer return process into two stages of order decision and, upon receipt, keep-or-return decision in remote purchase environments. Their three experiments have shown that return policy leniency could minimize the inherent consumer risk and provide higher quality perception to customers. Moreover, Anderson et al. (2009) empirically revealed the value of return option in different product categories. For example, the opportunity to return unwanted products in women's footwear was valued much more than that for men's tops. Using the model, they stimulated how different return policies influence firms' profits. More recently, Bonifield et al. (2010) also found the correlation between perceived quality of e-retailers and product return leniency. The effect held in non-consumable products rather than consumable products.

Another stream of product return studies shed light on how purchase process affects customer return decisions. Hess and Mayhew (1997) used a hazard model to break out the effects of product categories and prices customers purchased. Their model provided insights for predicting when product return would occur after purchase. Bechwati and Siegal (2005) introduced a framework drawing on research in consumer choice, consumer memory and attitude stability to predict how customer pre-decisional process affected their post-purchase return behavior. Using two experiments, their results showed that the likelihood of product return was contingent on the amount and nature of pre-choice cognitive responses. Similarly, Petersen and Kumar (2009) examined the effect of customer buying behaviors on product returns. Specifically, they found that products purchased as gifts from a new shopping channel and as sales items were less likely to be returned. In contrast, products bought during holidays or from a new category were more likely to be returned. Both streams neglected the potential influence from customers' social network friends, which is the gap we are to bridge in this study.

2.2 Utility Theory of Product Return

To determine a consumer's return propensity, Anderson et al. (2009) has provided an approach using utility theory. Particularly, the utility of returning an item, $U(return)_{it}=-R_{it} = -MC_{it} - PC_{it}$, is in the form of return cost and is always negative. Return cost consists of two parts, i.e. monetary cost and psychic cost. Monetary cost denotes shipping or mailing cost charged for the return. Examples of psychic cost can be time and hassle investment in returning (e.g., appeal for product return with the after-sale service department).

The utility of keeping a product composes of three parts, i.e., $U(keep)_{it}=\mu_{it}+\Psi_i+\varepsilon_{it}$. The first term μ_{it} is a deterministic utility, which contains marketing variables of prices and promotional information. It determines customer mean utility level. The second term Ψ_i is product fit, which measures the degree to which product attributes match customer preferences (Hong and Pavlou 2014). For example, for apparels, product fit can be size fit or sensory related fit like color or fabric texture (Anderson et al. 2009). Product fit uncertainty before purchase is identified to be highly related with customer product return decision (Hong and Pavlou 2014). The third term

εit captures time-varying shocks for customer preferences. In a sum, whether a customer keeps a product depends on the net utility when comparing keep utility to return utility, UKit=μit + Ψi + Rit + εit. If UKit <0, a customer would choose to return the product.

2.3 Sentiment Analysis and Social Media Prediction

In order to use social network posts to predict online consumers' return rate, sentiment analysis is the most frequently used approach to extract comprehensive information from the posts. Sentiment analysis (i.e., opinion mining) refers to the particular process of identifying and extracting subjective information (Rui and Whinston 2011). Pang and Lee (2008) developed an overall framework of sentiment analysis. Following their work, Chung and Tseng (2010) developed a BI system which related customer review content with their overall numerical ratings from Amazon. Similarly, Dang et al. (2010) studied sentiment classification of customer online reviews and proposed a new lexicon-enhanced method for sentiment classification. Moreover, some sentiment analysis tools are gaining popularity recently, such as OpinionFinder and SentiWordNet (Liu et al. 2010). Liu et al. (2010) applied the sentiment analysis tools to investigate the influence of online word-of-mouth on movie success.

Using sentiment analysis technique, several researchers began to use social netowrk data to do predictions (Asur and Huberman 2010; Rui and Whinston 2011). For instance, Asur and Huberman (2010) utilized the chatters in Twitter.com to demonstrate the effect of social media content on real-world outcomes. They found social media could be used to predict real-world outcomes after they did a cross-sectional study of 24 movies. Rui and Whinston (2011) designed a new BI system using content from Twitter.com to predict box-office sales. Bollen et al. (2011) analyzed the text content of daily Twitter feeds and used a Self-Organizing Fuzzy Neural Network (SOFNN) to predict DJIA closing values. In this study, we are going to use social network posts for product return prediction.

2.4 Signaling Theory

Previous scholars have begun to use signaling theory to explain the influence of social network posts on users' behavior. Signaling theory proposes that observable attributes of an entity serve as a signal of quality which can change stakeholders' perceptions, especially in situations of information asymmetry (Sanders and Boivie 2004; Spence 1973). In the seminal study, Spence (1973) demonstrated how a job applicant used the education degree as a signal of her quality to potential employers in the job market. Quality here refered to the unobservable capacity of signaler to meet the needs of observing outsiders, which depended on the investigation context (Connelly et al. 2011). In this study, quality is considered to be factors influencing consumers' product return decisions.

Moreover, it has been noted that in social media, signals can not only broadcast by companies; rather, online users play an important role in emanating signals in forms of word-of-mouth(Aggarwal et al. 2012; Fombrun and Shanley 1990). We are going to use these user generated signals to predict product return rate for e-commerce websites.

3 HYPOTHESES DEVELOPMENT

Based on utility theory and signaling theory, we develop our research hypotheses. Referring to the mainstream social networks of Twitter and Facebook, there are mainly two types of content people generate about a certain e-commerce company (i.e., objective vs. subjective) (Ghose and Ipeirotis 2011). One is relatively objective portraying factual information such as news about the focal e-commerce website. The other one is relatively subjective portraying users' emotions such as consumers' comments of purchasing experience, service experience, product use (Ghose and Ipeirotis 2011). A wide recognized wisdom is that negative comments are detriment to e-commerce website (Huang and Chen 2006; Tybout et al. 1981; Wyatt and Badger 1984), because it deters customer purchases and contributes to possible returns. However, this might not always be true (Berger et al. 2010).

For posts in the objective category, positive comments give a signal of good firm reputation, which may reduce the product return rate. For example, when a user posts a positive post on their social network page about a certain e-commerce website, this emits signals of their satisfaction and trust in the website to the network friends. This kind of signals can improve other customers' perception of the website, which decreases their propensity to return products bought from the website. Moreover, a positive piece of firm news in social network may confirm customers' purchase decision, enhancing the product keeping utility. Hence, we hypothesize that

H1: Sentiment values of objective posts about an e-commerce company in social media are negatively related to product return rate.

In contrast, negative comments about subjective experience (e.g., service related posts) may have a contrary role. A negative subjective review might give a signal of high psychic return cost, which lowers the return probability (Anderson et al. 2009). For example, when a customer has the intention to return products, her friends' complaints about the hassle return process may give her a signal that the return procedure is troublesome. Therefore, negative subjective reviews may lower the return rate by deterring those customers who are hesitated to return. In addition, negative customer subjective posts in social media may decrease consumer's expectation about the purchased the product. That is, consumers may purchase it because of the low price or special promotion instead of the high product quality. Under this circumstance, even if consumers are not so satisfied with the product, they may not return it. Hence, we hypothesize that:

H2: Sentiment values of customer subjective posts in social media are positively related to product return rate.

4 METHODOLOGY

4.1 Research Framework

Unlike previous studies which generally put all sentiment results as explanatory regressors for sales prediction (Asur and Huberman 2010; Bollen et al. 2011), we propose a new framework accounting for different categories of social media posts. Using LDA (Latent Dirichlet Allocation) technique and K-means clustering method, we are able to divide the original "tweets" to appointed numbers of topics. We assign the number of topics according to the thematic word distribution. Sentiment analysis gives sentiment values to each post. The hypotheses will be tested based on these sentiment values. Figure1 depicts the research framework.

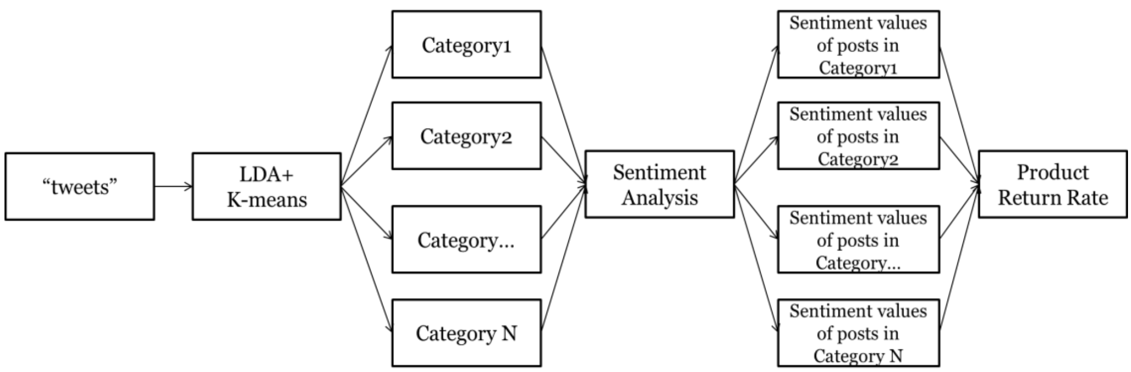

Figure1. Research framework

4.2 LDA (Latent Dirichlet Allocation)

LDA is a thematic words extraction technique. It is a generative model which allows sets of observations to be explained by unobserved groups explaining some parts of the data are similar (Blei et al. 2003). When the words are collected in forms of documents, then each document represents a mixture of topics, and each word in the document is attributed to one of the document's topic. The LDA model was introduced by Blei et al. (2003). In Blei et al. (2003) seminal work, they define a word is the basic unit of discrete data, defined to be an item from a vocabulary index by (1,…,V). They represent words using unit-basis vectors that have a single component equal to one and all other components equal to zero. A document is a sequence of N words in forms of w=(w1,w2,…,wN); and a corpus is a collection of M documents denoted by D=(w1,w2,…,wN).

LDA is a generative probabilistic model of a corpus. The basic idea for LDA is that each document denotes a mixture of latent topics, where each topic is characterized by a distribution of words. The process of LDA for each document w in a corpus D is as follows.

1. Choose $N \sim \text{Poisson}(\xi)$.

2. Choose $\theta \sim \text{Dir}(\alpha)$.

3. For each of the N words w_n:

    (a) Choose a topic $z_n \sim \text{Multinomial}(\theta)$;

    (b) Choose a word w_n from $p(w_n|z_n,\beta)$, a multinomial probability conditioned on the topic z_n.

Figure 2 denotes graphical model representation of LDA. The boxes are "plates" representing replicates. The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document.

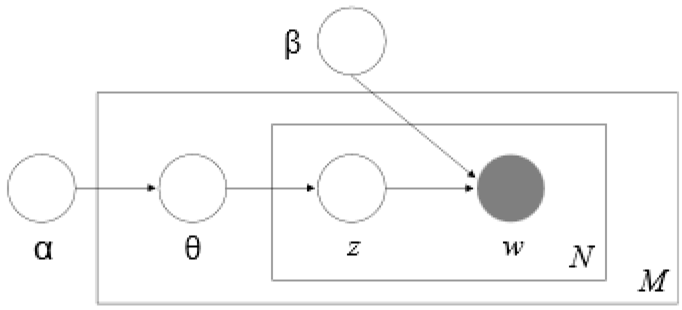

Figure2. Graphical model representation of LDA

4.3 K-means Clustering

K-means was first proposed by Lloyd (1982). It is a method of vector quantization, which is popular for cluster analysis in data mining. Basically, k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster (Lloyd 1982).

Practically, given a set of n observations ($x1,x2,\ldots,x_n$), each of them in a d-dimensional vectors. K-means clustering aims to divide the observations into k sets ($k<=n$) $S=(S1,S2,\ldots S_k)$ to minimize the within-cluster sum of squares (WCSS):

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2$$, where μ_i denotes the mean point of S_i.

4.4 Sentiment Analysis

As we set the research context in China, software called ROST Emotion Analysis Tool specializing in emotion analysis for Chinese language is employed. The software is developed by researchers in Wuhan University.

ROST employs a lexicon-based approach to extract sentiment values from the posts. The lexicon contains two files, one including a list of positive words while the other including a list of negative words. The words have their corresponding emotional values based on their psychological definitions. ROST works as follows. It first identifies key words in each sentence. Then the software assigns the emotional values to each key word according to the lexicon. Finally, through imposing a proper weight to the words, the software calculates the emotional value of each sentence. The larger the value is, the more positive the sentence is[2].

4.5 VAR (Vector-Autoregression)

To analyze the dynamic relationships between sentiment values of posts in different topics and product return rate, we employed vector autoregression (VAR) which accounts for endogeneity and the dynamic interactions between antecedents and outcomes (Dekimpe and Hanssens 1999). VAR model treats all variables symmetrically and allows each variable to have an equation explaining its evolution based on the lags of all model variables (Sims 1990).

An unrestricted VAR model measures the evolution of a set of k endogenous variables over the same period (t = 1, ..., T) as a linear function of only their past values:

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + \varepsilon_t \text{; t=1,2,…,T}$$

i.e. $$\begin{pmatrix} y_{1t} \\ y_{2t} \\ \cdots \\ y_{kt} \end{pmatrix} = A_1 \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ \cdots \\ y_{kt-1} \end{pmatrix} + \ldots + A_p \begin{pmatrix} y_{1t-p} \\ y_{2t-p} \\ \cdots \\ y_{kt-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \cdots \\ \varepsilon_{kt} \end{pmatrix} \text{; t=1,2,…,T}$$

where p indicates the lag order; $A_1,\ldots,A_p$ are k*k coefficient matrix to be estimated and ε_t is the time series of white noise. We employed STATA 12.0 to conduct the analysis.

5 EMPIRICAL STUDY

5.1 Data Acquisition

The panel data set contains a span of one month (June 2012), which consists of two parts. The first part is product return rate of one of the largest e-commerce websites in China. The data is aggregated at daily level. It is calculated as the proportion of the amount of return orders in the number of total transactions in a certain day. The average return rate, which is calculated by the portion of returned transactions relative to all transactions within the month, is 5.62%.

The other part is the e-commerce website related “tweets” in Weibo during June 2012 in China. We crawled all tweets mentioning about the name of the e-commerce website at a daily base. Totally, 15949 “tweets” records are logged in the database. They will be the input for the return analysis through clustering and sentiment analysis. The distribution of post numbers is given in Table 1.

[2] http://blog.sciencenet.cn/blog-239936-276349.html

Date	Freq.	Percent	Date	Freq.	Percent
1-Jun	572	3.59	16-Jun	649	4.07
2-Jun	371	2.33	17-Jun	646	4.05
3-Jun	219	1.37	18-Jun	937	5.87
4-Jun	297	1.86	19-Jun	877	5.5
5-Jun	282	1.77	20-Jun	754	4.73
6-Jun	359	2.25	21-Jun	889	5.57
7-Jun	394	2.47	22-Jun	712	4.46
8-Jun	427	2.68	23-Jun	459	2.88
9-Jun	561	3.52	24-Jun	470	2.95
10-Jun	315	1.98	25-Jun	543	3.4
11-Jun	506	3.17	26-Jun	701	4.4
12-Jun	535	3.35	27-Jun	559	3.5
13-Jun	394	2.47	28-Jun	427	2.68
14-Jun	683	4.28	29-Jun	325	2.04
15-Jun	771	4.83	30-Jun	315	1.98
			Total	**15,949**	**100**

Table 1. The distribution of post numbers

5.2 Data Process and Results

Before using LDA, we used NLPIR to separate words within each post[3]. NLPIR is a professional system to separate words in Chinese sentences. It is widely used for Chinese nature language processing, with more than 200,000 users. Next, as discussed in the method part, we first used these 15949 tweets to extract the thematic words using LDA approach. We assigned the model to divide four topics according to the word distribution. According to LDA results, the four topics can be described as follows. The first category mainly talks about news of the focal company (5662 posts), which is relatively objective. The thematic words extracted comprise "names of other large e-commerce websites in China", "competition", "sales" and other related. The second category represents customer purchasing experience (3685 posts), which is relatively subjective. Key words like "price", "buying", "discounts", "saving" and etc. are emerged from the data. The third theme extracted is products use (3488 posts), which is relatively subjective. The words in this category consists of "notebooks", "books", "car", "product parameters", "function", "use" and etc. People may discuss about products of the certain e-commerce website in social networks. The final category of words emerge as those related to service experience (3114 posts), which is customer-specific and relatively subjective. Key words include "after purchase warranty", "return", "service representatives" and etc. Meanwhile, every post was assigned with a weight of each topic. Thus, each post is a 4-dimensional observation for the further clustering.

After LDA, we conducted k-means clustering for all the posts. Through minimizing the within-cluster sum of squares (WCSS), each post is assigned to a particular category divided by LDA. Then we input all the posts into ROST to get their sentiment values. All these values were averaged at a daily base. Moreover, we used "comment number" and "repost number" as control variables. Based on the above setting, the econometric model specifications are:

$$ReturnRate_t = \beta_1 EmotionNews_{t-1} + \beta_2 EmotionNews_{t-2} + \beta_3 EmotionPurchase_{t-1} + \beta_4 EmotionPurchase_{t-2} + \beta_5 EmotionProduct_{t-1} + \beta_6 EmotionProduct_{t-2} + \beta_7 EmotionService_{t-1} + \beta_8 EmotionService_{t-2} + \beta_9 Comment_{t-1} + \beta_{10} Comment_{t-2} + \beta_{11} Repost_{t-1} + \beta_{12} Repost_{t-2} + \beta_{13} Repost_{t-1} + \beta_{13} Repost_{t-2} + \varepsilon_t$$

[3] http://ictclas.nlpir.org/

The VAR results are showed in Table 2.

Vector Autoregression				
Sample: 3 - 30			No. of obs	28
Log likelihood = 5808.297			AIC	-389.807
FPE = 2.6e-174			HQIC	-384.702
Det(Sigma_ml) = 6.3e-197			SBIC	-373.107
	Coef.	Std. Err.	z	P>z
return				
L1.	0.012672	0.038489	**0.33**	**0.742**
avg_sentiment_company_news				
L1.	-0.00041	0.000148	**-2.75**	**0.006**
avg_sentiment_purchase_experience				
L1.	0.004015	0.000156	**25.73**	**0.000**
avg_sentiment_product				
L1.	-0.00707	0.000154	**-45.75**	**0.000**
avg_sentiment_service				
L1.	-0.01147	0.000549	**-20.90**	**0.000**
_cons	-0.06328	0.005326	-11.88	0.000

Table 2. VAR results

As can be seen from Table 2, the sentiment value of company news is significantly negatively related to product return rate. In contrast, the sentiments of product use, customer purchasing experience and service reviews are positively related to return rate. Hence, H1 and H2 are both supported.

6 DISCUSSION AND IMPLICATIONS

The results reveal that positive posts about company news in social network are negatively related to future product return rate. Positive news posts help build brand reputation, and increase customers' trust. After viewing positive news about the e-commerce company, customers tend to have more trust on the products purchased on the platform, and thus are less likely to return the products. In contrast, the sentiment values of posts related to product use, purchasing and service experiences are positively related to product return rate as we hypothesized. That is, if consumers have a high expectation of the purchased product from the online word-of-mouth, they are more likely to return the product, which is consistent with the utility theory of product return (Anderson et al. 2009).

This study has several theoretical implications. First, it extends the existing utility theory and signaling theory by incorporating social media influence on product returns (Anderson et al. 2009; Spence 1973). Although the effect of social media marketing on product sales has been studied a lot in previous literature, this is the first study to investigate the influence of SNS posts on product return. Due to the rapid development of SNS usage and the huge loss caused by product returns, it is significant to examine the relationship between the online word-of-mouth and product return rate. Thus, it also enriches the literature of product return, as well as social media prediction research (Asur and Huberman 2010; Rui and Whinston 2011). Second, we divide the posts in SNSs into four categories (e.g., company news, product use, purchasing and service experiences). Based on this categorization, we examine the influence of each type of posts on product return rate. As previous studies usually regarded all posts in a single category, our study is a rare attempt to categorize SNS posts based on their contents and themes. The findings of different directions of the relationships imply that it is necessary and critical to categorize different types of SNS posts. Third, our study leverages machine learning techniques to help analyze the large SNS posts data. The similar machine learning methods used in this research can also be applied to similar context to solve some complex questions which cannot be addressed by human coders.

Furthermore, this study also has practical implications. First, we provide a new way to predict product return rate for e-commerce websites using social media posts. Specifically, the division of objective and subjective posts increases the prediction power. Second, the results suggest that the most efficacious way to lower return rate is to increase the e-commerce brand reputation. Thus more positive objective posts in social media, like positive company news, lead to lower return rate. For positive subjective posts in social media, although they may increase people’s intention to purchase the product, at the same time they also increase the risk of returning the product. Therefore, positive subject posts are like a double-edge sword. They may increase the sales and return rates simultaneously. Thus, it may not be an optimal strategy for practitioners to encourage making positive subjective posts in SNSs.

7 CONCLUSION

By using text mining techniques, such as LDA, K-means clustering and sentiment analysis, we first classify posts into four distinct topics and obtain the sentiment of each post. Then, use the sentiment values to predict product return rates based the categorization of the SNS posts. Our results reveal that the emotion of posts related to brand news, product and service have a negative effect on returns, while emotion of purchasing experience has a reverse effect. Our study has both theoretical implications and practical implications for researchers and practitioners. For researchers, we extend the social media and product return literature by leveraging utility theory and signaling theory to explain the relationship between SNS posts and product return rates. For practitioners, they can benefit from this study to have an advanced perception of product returns based on the contents in social media. Thus, the predictive power of product return rate is greatly enhanced through dividing social media post into distinct categories.

References

Aggarwal, R., Gopal, R., Gupta, A., and Singh, H. 2012. "Putting Money Where the Mouths Are: The Relation between Venture Financing and Electronic Word-of-Mouth," *Information Systems Research* (23:3-part-2), pp. 976-992.

Anderson, E.T., Hansen, K., and Simester, D. 2009. "The Option Value of Returns: Theory and Empirical Evidence," *Marketing Science* (28:3), pp. 405-423.

Asur, S., and Huberman, B.A. 2010. "Predicting the Future with Social Media," *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pp. 492-499.

Banjo, S. 2013. "Rampant Returns Plague E-Retailers." Retrieved Feb 9th, 2016, from http://www.wsj.com/articles/SB10001424052702304773104579270260683155216

Bechwati, N.N., and Siegal, W.S. 2005. "The Impact of the Prechoice Process on Product Returns," *Journal of Marketing Research* (42:3), pp. 358-367.

Berger, J., Sorensen, A.T., and Rasmussen, S.J. 2010. "Positive Effects of Negative Publicity: When Negative Reviews Increase Sales," *Marketing Science* (29:5), pp. 815-827.

Blanchard, D. 2007. "Supply Chains Also Work in Reverse," *Industry Week* (256:5), pp. 48-49.

Blei, D.M., Ng, A.Y., and Jordan, M.I. 2003. "Latent Dirichlet Allocation," *the Journal of machine Learning research* (3), pp. 993-1022.

Bollen, J., Mao, H., and Zeng, X. 2011. "Twitter Mood Predicts the Stock Market," *Journal of Computational Science* (2:1), pp. 1-8.

Bonifield, C., Cole, C., and Schultz, R.L. 2010. "Product Returns on the Internet: A Case of Mixed Signals?," *Journal of Business Research* (63:9), pp. 1058-1065.
Chung, W., and Tseng, T.-L.B. 2010. "Extracting Business Intelligence from Online Product Reviews: An Experiment of Automatic Rule-Induction," *International Conference on Information Systems* Kaminoyama, Japan.
Connelly, B.L., Certo, S.T., Ireland, R.D., and Reutzel, C.R. 2011. "Signaling Theory: A Review and Assessment," *Journal of Management* (37:1), pp. 39-67.
Dang, Y., Zhang, Y., and Chen, H. 2010. "A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews," *Intelligent Systems, IEEE* (25:4), pp. 46-53.
Davis, S., Hagerty, M., and Gerstner, E. 1998. "Return Policies and the Optimal Level of “Hassle”," *Journal of Economics and Business* (50:5), pp. 445-460.
Dekimpe, M.G., and Hanssens, D.M. 1999. "Sustained Spending and Persistent Response: A New Look at Long-Term Marketing Profitability," *Journal of Marketing Research* (36:4), pp. 397-412.
Dhar, V., and Chang, E.A. 2009. "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *Journal of Interactive Marketing* (23:4), pp. 300-307.
Fombrun, C., and Shanley, M. 1990. "What's in a Name? Reputation Building and Corporate Strategy," *Academy of management Journal* (33:2), pp. 233-258.
Ghose, A., and Ipeirotis, P.G. 2011. "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *Knowledge and Data Engineering, IEEE Transactions on* (23:10), pp. 1498-1512.
Hess, J.D., Chu, W., and Gerstner, E. 1996. "Controlling Product Returns in Direct Marketing," *Marketing Letters* (7:4), pp. 307-317.
Hess, J.D., and Mayhew, G.E. 1997. "Modeling Merchandise Returns in Direct Marketing," *Journal of Interactive Marketing* (11:2), pp. 20-35.
Hong, Y., and Pavlou, P.A. 2014. "Product Fit Uncertainty in Online Markets: Nature, Effects, and Antecedents," *Information Systems Research* (25:2), pp. 328-344.
Huang, J.H., and Chen, Y.F. 2006. "Herding in Online Product Choice," *Psychology & Marketing* (23:5), pp. 413-428.
Internet Retailer. 2013. "Reducing the Rate of Returns." Retrieved Apr 28th, 2015, from https://www.internetretailer.com/2013/05/29/reducing-rate-returns
Liu, Y., Chen, Y., Lusch, R.F., Chen, H., Zimbra, D., and Zeng, S. 2010. "User-Generated Content on Social Media: Predicting Market Success with Online Word-of-Mouth." IEEE Intelligent Systems.
Lloyd, S. 1982. "Least Squares Quantization in Pcm," *IEEE Transactions on Information Theory* (28:2), pp. 129-137.
National Retail Federation. 2013. "2013 Consumer Returns in the Retail Industry."
Padmanabhan, V., and Png, I.P.L. 1997. "Manufacturer's Return Policies and Retail Competition," *Marketing Science* (16:1), pp. 81-94.
Pang, B., and Lee, L. 2008. "Opinion Mining and Sentiment Analysis," *Foundations and trends in information retrieval* (2:1-2), pp. 1-135.
Petersen, J.A., and Kumar, V. 2009. "Are Product Returns a Necessary Evil? Antecedents and Consequences," *Journal of Marketing* (73:3), pp. 35-51.
Rui, H., and Whinston, A. 2011. "Designing a Social-Broadcasting-Based Business Intelligence System," *ACM Transactions on Management Information Systems (TMIS)* (2:4), p. 22.
Sanders, W., and Boivie, S. 2004. "Sorting Things Out: Valuation of New Firms in Uncertain Markets," *Strategic Management Journal* (25:2), pp. 167-186.
Sims, C.A. 1990. "Macroeconomics and Reality," *Econometrica* (48:1), pp. 1-48.
Spence, M. 1973. "Job Market Signaling," *The quarterly journal of Economics* (87:3), pp. 355-374.
Turney, P.D. 2002. "Thumbs up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews," *Proceedings of the 40th annual meeting on*

association for computational linguistics: Association for Computational Linguistics, pp. 417-424.
Tybout, A.M., Calder, B.J., and Sternthal, B. 1981. "Using Information Processing Theory to Design Marketing Strategies," *Journal of Marketing Research* (18:1), pp. 73-79.
Wood, S.L. 2001. "Remote Purchase Environments: The Influence of Return Policy Leniency on Two-Stage Decision Processes," *Journal of Marketing Research* (38:2), pp. 157-169.
Wyatt, R.O., and Badger, D.P. 1984. "How Reviews Affect Interest in and Evaluation of Time," *Journalism Quarterly* (61:4), pp. 874-878.