Impact of Tweets on Box Office Revenue: Focusing on When Tweets are Written

Hyunmi Baek, Joongho Ahn, and Sehwan Oh

This study investigates the impact of tweets on box office revenue. Specifically, the study focuses on the times when tweets were written by examining the impact of pre- and post-consumption tweets on box office revenue; an examination that is based on Expectation Confirmation Theory. The study also investigates the impact of intention tweets versus subjective tweets and the impact of negative tweets on box office revenue. Targeting 120 movies released in the US between February and August 2012, this study collected tweet information on a daily basis from two weeks before the opening until the closing and box office revenue information. The results indicate that the disconfirmation that occurs in relation to the total number of pre-consumption tweets for a movie has a negative impact on box office revenue. This premise suggests that the formation of higher expectations of a movie does not always result in positive results in situations where tweets on perceived movie quality after watching spread rapidly. This study also reveals that intention tweets have stronger effects on box office revenue than subjective tweets.

Keywords: Tweet, Twitter, electronic word of mouth, eWOM, social media, Expectation Confirmation Theory, ECT.

Manuscript received July 24, 2013; revised Oct. 22, 2013; accepted Nov. 5, 2013.

Hyunmi Baek (corresponding author, lotus1225@hanyang.ac.kr) is with the College of Communication and Social Sciences, Hanyang University, Ansan, Rep. of Korea.

Joongho Ahn (jahn@snu.ac.kr) is with the Graduate School of Business, Seoul National University, Seoul, Rep. of Korea.

Sehwan Oh (sehwano@snu.ac.kr) is with the College of Business Administration, Seoul National University, Seoul, Rep. of Korea.

I. Introduction

With the rapid spread of social media, various changes are taking place across all sectors of society. To investigate the powerful influence of electronic word of mouth (eWOM), we conducted a study on the impact of tweets from Twitter — a representative social media service. Twitter, an online SNS and microblog service, enables users to share tweets that are up to 140 characters in length. Based on the context of Twitter, studies on the impact of tweets have been conducted [1]–[4]. However, the scope of published studies on the impact of tweets is still broad, and these studies appear relatively fragmented.

The first goal of this study is to investigate the relationship between tweets and box office revenue by focusing on the time when tweets were written. To perform a more in-depth analysis of the impact of tweets on sales, this study considers pre-consumption tweets (the number of tweets before movie release) and post-consumption tweets (the number of tweets after movie release) as influential factors on box office revenue and investigates the interaction between the two from the perspective of Expectation Confirmation Theory (ECT). According to [5], expectancy disconfirmation is a factor affecting consumer satisfaction levels. In this respect, this study focuses on how box office revenue is influenced by the difference in tweet volume of tweets containing movie-related messages before release and tweets containing movie-related messages after release. This study also investigates the impact of negative tweets on box office revenue and the moderating effect of tweet type (intention tweets versus subjective tweets) on the relationship between volume of tweets and box office revenue. For this study, we collected daily tweets and box office revenue information on 120 films. We also conducted text mining to classify each tweet as either irrelevant, intention, positive, neutral, or negative.

The research domain "movies" was selected for two reasons. First, seeing a movie is an experience whereby it is only possible to have a limited understanding beforehand of how it is going to be prior to making the decision to purchase a ticket to see it. This type of experience is an example of an "experience good" [6], and in such cases eWOM is considered an important prerequisite information source. Second, eWOM and revenue information about movies are relatively easy to obtain. This is in stark contrast to most other genres, where revenues are usually unknown — for example, book sales were reversely calculated through sales rankings in a previous study [7]. Among various social media, we chose Twitter for our analysis for several reasons. First, Twitter is widely used by consumers and thus has emerged as a representative of social media. Second, before movie releases, more eWOM are generated on Twitter than any other social media.

The findings suggest that the formation of appropriate expectations for movies is highly significant. We also find that intention tweets have a stronger impact on box office revenue than subjective tweets. These findings can contribute toward the development of an eWOM management strategy of a company.

The remainder of this paper is organized as follows. Section II reviews the theory foundation. Section III presents the research model and hypotheses development. Section IV describes the research methodology, including data collection and text mining. Section V explains the results of the empirical analysis from actual tweet data. Finally, section VI discusses the results and contributions of the study.

II. Theoretical Foundation

1. Impact of Tweets on Product Sales

Several studies on the impact of eWOM continue to be conducted on the various channels of social media [1]–[4], [7]–[12]. The research on the impact of eWOM in online review sites [7]–[11] has been conducted since the very early stages of eWOM studies and continues to be so with the appearance of new kinds of social media, such as blogs [12] and Twitter [1]–[4]. In this study, we look at the impact of eWOM on box office revenue by utilizing tweets obtained from Twitter, which has received much attention due to the increase in the number of its subscribers.

Twitter, an online SNS and microblog service, was launched in March 2006. In social media, users can see their connections with others or show their social networks to others. Twitter enables users to share tweets or microblogging postings that are limited to 140 characters in length. In March 2012, Twitter reported that its network had 140 million active users and that it hosted 340 million tweets a day.

Recently, several studies have examined the impact of tweets from Twitter on box office revenue [1]–[4]. Reference [1] predicted opening-week box office revenue by using the tweet rate during the pre-release week (the number of tweets referring to a particular movie per hour). The study revealed that explanatory power is higher when considering sentiments presented in such tweets. Reference [3] indicated that the total number of such tweets has a positive effect on box office revenue. The study suggested that movie-related tweets with many followers have a stronger impact on box office revenue.

2. ECT

Since its appearance in the early 1970s, ECT has become an essential theory for the study of consumer satisfaction and dissatisfaction. According to ECT [5], levels of consumer satisfaction are affected by expectancy disconfirmation, which is measured as the difference between expectation and performance. In other words, consumers tend to have expectations of a product or service prior to purchasing it and later compare the performance of the product or service with their expectations after using it. Consumer expectation acts as a comparison standard to compare the performance of a particular service or product. If the level of expectation before consumption is consistent with the perceived performance after consumption, consumers are then thought to be satisfied with the particular product or service. If the perceived performance of a product or service is higher than expected, then positive disconfirmation will occur, but if the perceived performance is lower than expected, then this results in the formation of negative disconfirmation. If the perceived performance is the same as expected, consumers form simple confirmation of expectation. According to this theory, if expectancy confirmation forms, then satisfaction ensues. The causal relationship between expectancy disconfirmation and satisfaction, as well as the direct effect between expectation and satisfaction, appears significant regardless of products, situations, or research methods. Figure 1 shows the ECT according to [5].

After [13] investigated the continuous use of online banking services, a number of researchers conducted studies on the continuous use of information systems (ISs) using ECT [13]–[19]. In the IS field, ECT is combined with various theories and other factors to explain the continuous use of ISs by users. Reference [13] examined the continuous use of online banking services based on ECT and the Technology Acceptance Model. Reference [15] proposed that self-efficacy and confirmation are

http://dx.doi.org/10.4218/etrij.14.0113.0732

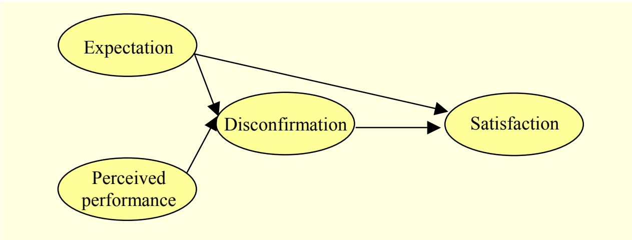

Fig. 1. Expectation Confirmation Theory.

factors contributing to the continuous use of the Internet. Reference [16] suggested perceived usefulness, attitude, concentration, subjective norm, perceived behavior control, and confirmation as explanatory factors that affect the consistent use of e-learning. Reference [17] investigated the value of including playfulness in ECT in explaining the continued use of a website. Reference [18] explored customer satisfaction with a website for long-term customer retention in e-commerce by using ECT. Reference [19] suggested perceived ease of use, perceived usefulness, perceived playfulness, confirmation, and satisfaction as explanatory variables to explain the consistent use of mobile Internet services.

While various factors are considered from different perspectives and research areas, the common finding is that the consistent intention to use is determined by the confirmation between expectation and personal experience with ISs. In the same context, this study verifies the hypothesis that confirmation of expectation, which draws from expectation (pre-consumption tweets) and perceived performance (post-consumption tweets), will have a positive impact on box office revenue.

III. Hypotheses Development

1. Impact of Intention Tweets on Box Office Revenue

Reference [3] showed that the effect of pre-consumption eWOM on box office revenue is actually larger than post-consumption eWOM. The study claimed that pre-consumption eWOM is generally about intentions or plans to purchase a product, whereas post-consumption eWOM is usually about experience toward a product after consumption.

Viewers consider subjective tweets as references prior to purchase. However, intention tweets express not only the purpose of purchase of the writer, but also influence the consumption behavior of people who have read such intention tweets. Therefore, the current study posits the following hypothesis:

Hypothesis 1. The volume of intention tweets will have a stronger impact on box office revenue than subjective tweets.

2. Impact of Negative Tweets on Box Office Revenue

Previous studies suggest that eWOM valence, which refers to customer opinions (positive or negative) about product quality [20], is likely to have an impact on product sales [3], [8]–[9], [21]–[22]. Positive eWOM fascinates new consumers and thus increases sales, whereas negative eWOM might hinder trust in a product that consumers want to purchase [23]–[24]. Hence, eWOM valence has an effect on sales. Using Twitter, [1] and [3] analyzed the impact of eWOM valence on product sales. The studies revealed that explanatory power is higher when considering sentiments presented in tweets. Reference [3] also showed that tweets containing negative sentiments about a movie have a negative impact on box office revenue. Reference [25] demonstrated that negative eWOM more strongly influences product sales than positive eWOM; thus, confirming negativity bias. Therefore, this study focused on the influence of negative tweets on box office revenue rather than that of positive tweets. Therefore, the present study posits the following hypothesis:

Hypothesis 2. The ratio of negative tweets to entire tweets will have a negative impact on box office revenue.

3. Impact of Pre- and Post- Consumption Tweets on Box Office Revenue

Previous studies examined that pre-consumption eWOM has a significant impact on opening-week box office revenue [1], [9], [22], [26]. Reference [1] revealed that the tweet rate during pre-release week (that is, the number of tweets referring to a particular movie per hour) could significantly affect the opening-week box office revenue. Reference [9] presented a model to predict opening-week box office revenue based on online reviews. Reference [26] revealed that a three-week pre-release Internet buzz, such as trailer views, message board comments, and votes of desire, has a significant impact on the opening-week box office revenue. Hence, the current study posits the following hypothesis:

Hypothesis 3. The volume of pre-consumption tweets will have a positive impact on opening-week box office revenue.

Several studies suggested that eWOM regarding experiences with purchased products or services has a significant impact on product or service sales [1], [3], [7]–[8], [10]–[12], [26]. Reference [7] revealed that the number of consumer reviews has a significant positive impact on sales. Reference [8] examined the impact of online reviews on book sales from the biggest book retailer sites, such as Amazon.com and Barnesandnoble.com. They found that the eWOM volume had an impact on marketability and ranking in both book-selling sites. Reference [10] reported that volumes of online reviews can have significant explanatory power on box office revenue.

Reference [12] cited a causal connection between the number of blogs and box office revenue. Reference [3] noted that the number of tweets has a positive effect on box office revenue. Therefore, the present study posits the following hypothesis:

Hypothesis 4. The volume of post-consumption tweets will have a positive impact on box office revenue after opening week.

As mentioned previously, the present study will verify the hypothesis that disconfirmation of expectancy, which is measured as the difference in pre- and post-consumption tweet volume, affects box office revenue according to ECT. The foundation of this theoretical framework is the ECT put forward by [5]. The current study proposes the following research model: expectation and expectancy disconfirmation, which is drawn from expectation and perceived quality, affect box office revenue.

Reference [5] stated that the difference in the score between pre- and post-purchase attitudes is significantly related to post-exposure satisfaction. In the online review context, [27] revealed that eWOM arising from expectation-performance disconfirmation has an impact on box office revenue. The study used pre- and post-consumption evaluations posted on Korean movie sites as measures of movie expectation and movie quality, respectively. Therefore, the current study hypothesizes that expectancy disconfirmation (that is, the difference in pre- and post-consumption tweet volume) is negatively related to box office revenue. Thus, this study posits the following hypothesis:

Hypothesis 5. Expectancy disconfirmation (that is, the difference in pre- and post-consumption tweet volume) will have a negative impact on box office revenue.

Fig. 2. Screenshot of Twitter search results.

Table 1. Data obtained from Twitter.

Data collected		Definition	Instrumentation
Movie level	Volume	Number of tweets per day for a movie	Numerical value (scale)
Tweet level	Contents	Contents of tweet	Textual description
	Date	Date when tweet was written	Numerical value (scale)
User level	Follower number	Number of followers	Numerical value (scale)

IV. Research Methodology

1. Data Collection

This study collected tweet and box office revenue information on 120 films, released in the US between February and August 2012, as daily panel data (weekly sales information). Data collection began two weeks before and continued up until the end of the movie release period. Following [3], the present study used the open application programming interface of Twitter to develop a program to collect daily tweets and user information for the target movies. Previous research used data from Twitter to investigate the impact of eWOM on box office revenue [1], [3]. Figure 2 shows a screenshot of Twitter search results. Table 1 presents data obtained from Twitter.

Previous studies used data from BoxOfficeMojo.com to investigate the impact of eWOM on box office revenue [1], [3], [9]–[10], [12]. The current study also collected data on film-specific weekly revenue, genre, MPAA rating, and the presence of a sequel from BoxOfficeMojo.com.

For this study, the independent variables are tweet volume and valence information. Tweet volume represents all tweet messages and retweeted messages that contained a keyword, or more than one keyword, relating to a movie's title. The control variables are general movie characteristics, such as genre, MPAA rating, and the presence of a sequel. Reference [9] analyzed studies to summarize factors affecting box office revenue — namely, star power, movie genre, MPAA ratings, media advertising, timing of release, and professional critic reviews. Reference [22] considered movie genre, the number of screens, and MPAA rating as the factors that affect box office revenue. Reference [28] considered movie genre, sequel, MPAA ratings, running time, and production time as the factors that affect box office revenue. Therefore, the present study

http://dx.doi.org/10.4218/etrij.14.0113.0732

Table 2. Variables used in research model.

Variables		Data source	Definition
Independent variables	Weekly tweet number	Twitter	Number of tweets per week for a movie (intention/positive/neutral/ negative/total tweet number)
Control variables	Genre	BoxOfficeMojo	Genre of a movie
	MPAA rating		MPAA rating of a movie
	Sequel		Whether or not a movie is part of a series
Dependent variables	Weekly revenue	BoxOfficeMojo	Weekly box office revenue for a movie in US ($)

considered movie genre, MPAA ratings, and sequel as control variables. Dependent variables include weekly box office revenue. Table 2 presents the variables used in this study.

2. Text Mining for Tweet Analysis

A tweet contains various kinds of writings but does not offer valence information as a numerical value. Hence, we constructed irrelevant, intention, and sentiment classifiers by using text mining. Based on these criterions, five further classifications of tweets were made — namely, irrelevant, intention, positive, neutral, and negative [3]. In this analysis, irrelevant tweets represent tweets that contain keywords of movie titles but are not relevant to those movies. Intention tweets are tweet messages that include a user's intention to go to the movie mentioned in the tweet. In addition, we classify positive tweets to be where a user evaluates one of the target movies as satisfactory, and vice versa for negative tweets. Finally, we classified neutral tweets to be messages containing either URLs or objective opinions of users that related to the target movies. Text mining primarily utilizes lexicons to extract manually the features from tweets and secondarily uses a Naïve Bayesian model. In terms of a tweet composed of 140 characters, a classification that only uses a Naïve Bayesian model is less accurate; thus, hybrid text mining (lexicon-based approach + Naïve Bayesian model) is used in the study. For this, we made several lexicons (lexicons for stop/relevant /irrelevant/non-intention/intention/neutral/negative/positive words) for the movie domain.

Naïve Bayesian classifier refers to a simple probabilistic classifier based on applying Bayes' theorem with strong (naïve) independence assumptions [29]. This is a methodology for assigning a certain document to a particular category based on the document's features.

$$P(C \mid F) = \frac{P(F) \times P(F \mid C)}{P(C)}. \quad (1)$$

In (1), F represents a feature of a document and C represents a class that belongs to a category. The probability of belonging to a particular class C given the existence of a particular feature F is represented by $P(C|F)$. Therefore, $P(C|F)$ is the value to be found through a Naïve Bayesian model using (1).

In this study, the order of text mining to classify tweets is as follows. Figure 3 represents the classifiers made for this study.

a. Filtering

① Unnecessary words are removed utilizing lexicons for stop words[1)] [30].

② Keywords of movie titles are replaced with a "query term."

b. Irrelevant classifier

① Relevant tweets are classified utilizing lexicons for relevant words.

② Irrelevant tweets are classified utilizing lexicons for irrelevant words.

③ Among the remaining sets, excluding the tweets fixed in *b.*-① and *b.*-②, relevant tweets and irrelevant tweets are classified utilizing a Naïve Bayesian model.

c. Intention classifier

① The sets classified as relevant tweets are classified through *b.*

② Non-intention tweets are classified utilizing lexicons for non-intention words.

③ Intention tweets are classified utilizing lexicons for intention words.

④ Among the remaining sets, excluding the tweets fixed in *c.*-② and *c.*-③, intention and non-intention tweets are classified utilizing a Naïve Bayesian model.

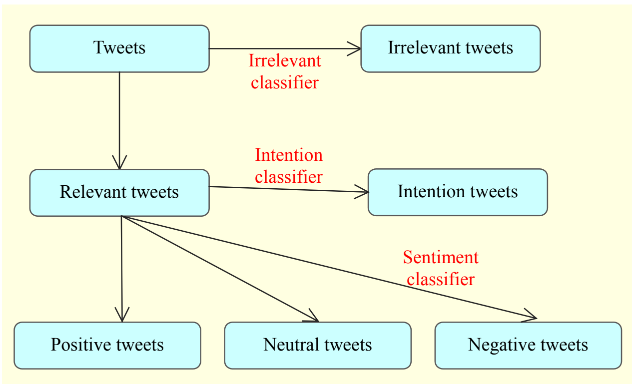

Fig. 3. Tweet classifiers.

1) Stop words are words that are filtered out prior to text mining (ex. the, is, at etc.)

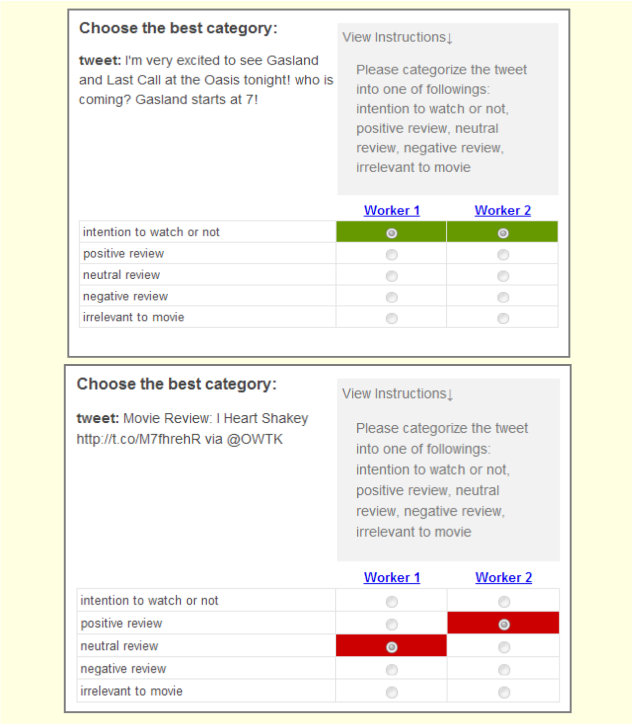

Fig. 4. Manual classification of tweets.

Table 3. Accuracy of tweet classifiers.

Classifier		Accuracy (%)
Irrelevant classifier		96
Intention classifier		86
Sentiment classifier	Positive	84
	Neutral	73
	Negative	89

d. Sentiment classifier

① The sets classified as non-intention tweets are classified through *c*.

② Neutral tweets are classified utilizing lexicons for neutral words.

③ Negative tweets are classified utilizing lexicons for negative words.

④ Positive tweets are classified utilizing lexicons for positive words.

⑤ Among the remaining sets — excluding the tweets fixed in *d*.-②, *d*.-③, and *d*.-④ — positive, neutral, and negative tweets, are respectively classified utilizing a Naïve Bayesian model.

To test the classifier accuracy, 2,000 movie tweets were manually classified into irrelevant, intention, positive, neutral, and negative. Amazon Mechanical Turk is utilized for manual labeling (see Fig. 4). Two native speakers performed cross checking; only cases in which identical classification occurs are utilized for manual labeling data. Among these data, 1,700 movie tweets were used for training and of these only 300 were used for accuracy testing.

Results of the automated program used to classify all tweets suggested that the irrelevant classifier showed 96% accuracy, the intention classifier 86% accuracy, the positive tweets 84% accuracy, the neutral tweets 73% accuracy, and the negative tweets 89% accuracy (see Table 3).

V. Results

1. Descriptive Statistics

Figure 5 shows the average tweet volume trend of the 120 target movies used in this study. On average, the first week of film release shows the largest number of tweets; then the number of tweets tends to decline after film release.

Figure 6 shows the percentage of intention, positive, neutral,

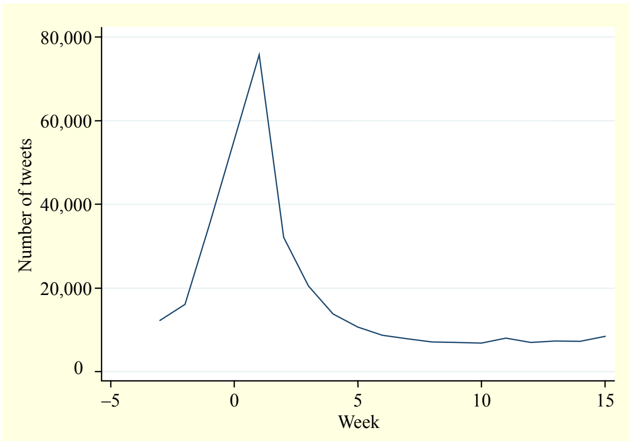

Fig. 5. Average tweet volume of all 120 target movies.

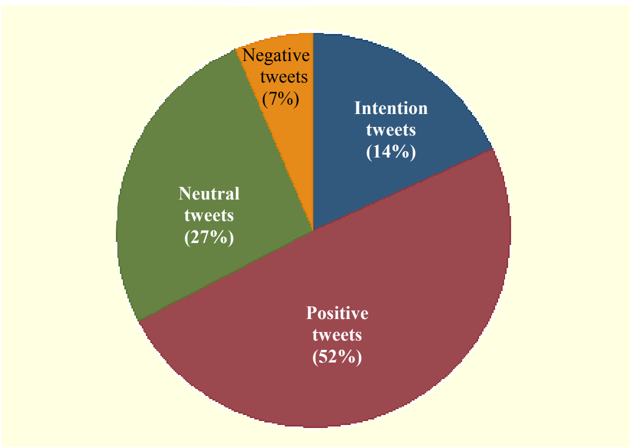

Fig. 6. Percentage of all tweets by type.

http://dx.doi.org/10.4218/etrij.14.0113.0732

Table 4. Number of daily tweets and average audience size per tweet.

	Intention tweet	Positive tweet	Neutral tweet	Negative tweet
No. of daily tweets	308	1,126	577	146
Avg. audience size per tweet	560	692	3,293	683

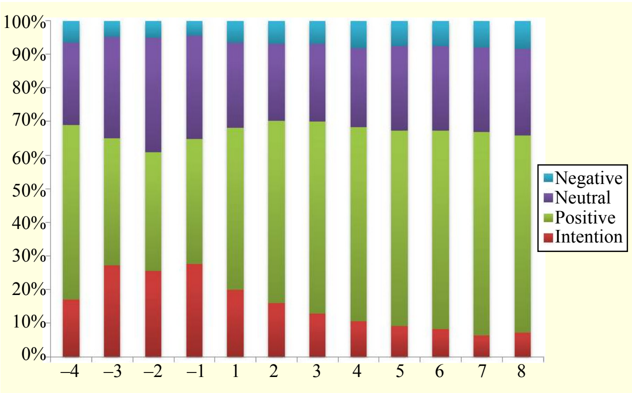

Fig. 7. Weekly proportions of tweet types.

and negative tweets of the entire film-related data. Positive tweets comprised 52% of all tweets, negative tweets 7%, intention tweets 14%, and neutral tweets 27%.

Each film had a daily average of 2,157 related tweets. Among these, 1,126 were positive, 146 were negative, and 308 were intention tweets (see Table 4). Based on the size of the audience per tweet, an average of 3,293 individuals per neutral tweet was found to be the highest audience-per-tweet rate among all types (see Table 4). The size of the audience is the number of followers of users who wrote the tweets.

Figure 7 presents the weekly proportions of the tweet types: intention, positive, neutral, and negative. Before film release, intention tweets accounted for a significant proportion, but then this proportion gradually decreases after film release. Meanwhile, subjective tweets (positive tweets + negative tweets) tended to increase gradually after film release.

2. Impact of Intention Tweets on Box Office Revenue

To investigate the tweet type that was the most influential on box office revenue (that is, intention tweet or subjective tweet), Least Squares Dummy Variable (LSDV) analysis is performed. Consequently, intention tweets are found to be statistically more influential on the weekly box office revenue than subjective tweets, which supported Hypothesis 1 (see Table 5).

3. Impact of Negative Tweets on Box Office Revenue

Each of the 120 films studied received on average a negative

Table 5. Impact of intention and subjective tweets.

	Intention tweets		Subjective tweets	
	Coefficient	Standard error	Coefficient	Standard error
Weekly $\text{tweet}_{t\text{-}1}$	589.83***	24.38	194.92***	10.23
(Constant)	1819022***	396545.8	1259564***	443687.7
R^2	0.4788		0.3911	
N	1067		1067	
Chin's T-value	14.95			

Note: *$p < 0.05$, **$p < 0.01$, and *** $p < 0.001$.
Dependent variable: weekly box office revenue.

Table 6. Impact of negative tweets.

	Coefficient
Weekly proportion of negative $\text{tweets}_{t\text{-}1}$	–4.28e + 07**
(Constant)	6494042***
R^2	0.1669
N	1,067

Note: *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.
Dependent variable: weekly box office revenue.

tweet proportion of 4.3%. To investigate the impact of negative tweets on box office revenue, LSDV analysis was conducted. Consequently, we can conclude that box office revenue decreased by $42.8 million for every 1% increase in the proportion of negative tweets. This result supported Hypothesis 2 (see Table 6).

4. Impact of Pre-consumption Tweets on Opening-Week Revenue

To verify Hypothesis 3, we tested the impact of pre-consumption tweets on opening-week revenue by using hierarchical regression analysis. Hierarchical regression analysis is a type of multiple regression analysis that is used to analyze the comparative impact of a certain independent variable in which a researcher has interest. Table 7 presents the analysis results for verifying Hypothesis 3. In Model 1, we considered the control variables to be: the presence of a series, MPAA ratings, and movie genre. Consequently, presence of a series and movie genre (for example, action and adventure) appears to have an impact on box office revenue during opening week. In Model 2, weekly tweet volume, a week before film release, was entered as an independent variable. As a result, weekly tweet volume before film release had a significant positive impact on opening-week revenue (that is, Hypothesis 3 is supported). Moreover, even when movie

Table 7. Impact of pre-consumption tweets.

Independent variable	Model 1	Model 2
	Coefficient	Coefficient
Series	4.63e+07***	3.18e+07***
R & NC_7	5579749	–1226424
PG_13	1.08e+07	3776571
Science fiction	1.20e+07	9899739
Kid	2.87e+07	3.00e+07**
Thriller	1738099	1854565
Comedy	5633351	3163018
Drama	–786277.9	1984299
Action/adventure	6.87e+07***	2.15e+0*
Weekly $tweet_{t-1}$		293.1775***
R^2	0.484	0.845
R^2 change	0.484***	0.361***
N	120	108

Note: *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.
Dependent variable: opening-week box office revenue.

Table 8. Impact of post-consumption tweets.

Independent variable	Coefficient
Weekly $tweet_{t-1}$	120.5423 ***
R^2	0.6056
N	959

Note: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and FE2SLS.
Dependent variable: weekly box office revenue.

characteristics, such as movie genre, MPAA ratings, and the presence of a series, are considered, the explanatory power is 84.5%.

5. Impact of Post-Consumption Tweets on Box Office Revenue

To verify Hypothesis 4, we verified the impact of post-consumption tweets on box office revenue after the second week (see Table 8). Panel data analysis is utilized; thus, idiosyncratic characteristics associated with each movie were not considered separately. To verify the impact of tweets on box office revenue, a one-week lagged tweet volume is considered as an independent variable. According to the analysis, the impact of post-consumption tweets on box office revenue after the second week was significantly positive (that is, Hypothesis 4 is supported).

6. Impact of Expectancy Disconfirmation

To verify Hypothesis 5, we performed hierarchical regression analysis with total box office revenue as the dependent variable. The presence of a movie series, MPAA ratings, and genre were the control variables. The pre-consumption tweets and the difference between pre- and post-consumption tweets were the independent variables. For the analysis in this study, pre-consumption tweets pertained to the number of weekly tweets one week before film release. Post-consumption tweets denoted the number of average weekly tweets for five weeks after film release. When these variables were considered as independent variables, standardization values were used for analysis because direct comparisons between the number of pre- and post-consumption tweets were unreasonable.

Table 9. Impact of expectancy disconfirmation.

Independent variable	Model 1	Model 2	Model 3
	Standardized coefficient	Standardized coefficient	Standardized coefficient
Series	1.079***	0.770***	0.716**
R & NC_7	0.082	–0.040	–0.018
PG_13	0.261	0.104	0.095
Science fiction	0.227	0.174	0.160
Kid	0.964*	0.881***	0.872***
Thriller	0.024	0.004	–0.008
Comedy	0.101	0.094	0.110
Drama	0.017	0.074	0.070
Action/adventure	1.642***	0.566*	0.543*
Pre-consumption tweets		0.634***	0.781***
Disconfirmation of expectation			–0.720***
R^2	0.518	0.827	0.840
R^2 increase	0.518***	0.308***	0.013***
N	109	108	108

Note: *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.
Dependent variable: total box office revenue.
VIF (variance inflation factor) = 2.28.

Table 9 presents the analysis results of the hierarchical regression model for verifying Hypothesis 5. In Model 1, the variables relating to movie characteristic were included as control variables. From this model, we concluded that films belonging to either a series or to the action/adventure-for-kids genre had a positive impact on box office revenue. In Model 2, we added pre-consumption tweets as an independent variable. As a result of the analysis, the pre-consumption tweets were found to have a positive impact on box office revenue. In Model 3, the difference in pre- and post-consumption tweet volume was added as an independent variable. The results showed that this difference was found to have a negative impact on box office revenue. This finding supported Hypothesis 5. Therefore, the greater the difference in pre- and post-consumption tweet volume, the lower the box office revenue.

VI. Conclusion

Initially, this study looked at the impact of the volume of tweets both before and after film release on box office revenue. As a result of our analysis, we found that box office revenue is proportional to the total volume of tweets. This study also indicated that the impact of expectancy disconfirmation (measured as the difference in pre- and post-consumption tweet volume) was negative in relation to box office revenue. The findings are expected to provide some suggestions on the time-periodic management of tweets for interested parties wishing to maximize product sales. For example, implications that the strategic management of tweets is not always the best way to increase the expectations of a specific movie have emerged. Conversely, the strategy of maximizing revenue by lowering expectations of a movie is also perhaps inappropriate and can lead to lost opportunities to increase revenue, as it is difficult to induce a willingness to see a movie into an audience once it has been released. In other words, forming appropriate expectations for a movie is highly significant.

We also analyzed the difference in the impact on box office revenue of the volume of subjective tweets versus the volume of intention tweets. The result revealed that the impact of intention tweets on box office revenue was three times more than that of subjective tweets. Furthermore, in terms of valence, the higher the proportion of negative tweets, the more this reduced box office revenue.

The contributions of our study are multifaceted. First, the study contributes to research methodology through its unique use of tweet classifiers within the movie domain. For this analysis, we made the tweet classifiers — irrelevant classifier, intention classifier, and sentiment classifier. We could analyze the effect of tweets on box office revenue by obtaining information on the contents of tweets as quantified variables through this process. Second, the study contributes to real-life practices. Based on ECT, this study shows that the expectancy disconfirmation (measured as the difference in pre- and post-consumption tweet volume) of a movie has a negative impact upon its sales. This finding may eventually help to recognize strategic directions for corporate management of tweets.

References

[1] S. Asur and B.A. Huberman, "Predicting the Future with Social Media," *Int. Conf. Web Intel. Intel. Agent Technol.*, Toronto, Canada, vol. 1, Aug. 31–Sept. 3, 2010, pp. 492–499.

[2] J. Bughin, "*Online Buzz and Searches as Telecoms Brand Performance Indicators*," Working Papers, Social Science Research Network, Aug. 2011. http://papers.ssrn.com/sol3/papers.cfm? abstract_id=1904328

[3] H. Rui, Y. Liu, and A. Whinston, "Whose and What Chatter Matters? The Effect of Tweets on Movie Sales," *Decision Support Syst.*, vol. 55, no. 4, Nov. 2013, pp. 863–870.

[4] X. Zhang, H. Fuehres, and P. Gloor, "Predicting Stock Market Indicators through Twitter, I Hope it is not as Bad as I Fear," *Procedia - Social Behavioral Sci.*, vol. 26, 2011, pp. 55–62.

[5] R.L. Oliver, "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," *J. Marketing Res.*, vol. 17, no. 4, Nov. 1980, pp. 460–469.

[6] P. Nelson, "Information and Consumer Behavior," *J. Political Economy*, vol. 78, no. 2, Mar.–Apr. 1970, pp. 311–329.

[7] P.Y. Chen, S. Wu, and J. Yoon, "The Impact of Online Recommendations and Consumer Feedback on Sales," *Int. Conf. Inf. Syst.*, Washington DC, USA, Dec. 12–15, 2004.

[8] J.A. Chevalier and D. Mayzlin, "The Effect of Word of Mouth on Sales: Online Book Reviews," *J. Marketing Res.*, vol. 43, no. 3, Aug. 2006, pp. 345–354.

[9] C. Dellarocas, X.M. Zhang, and N.F. Awad, "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *J. Interactive Marketing*, vol. 21, no. 4, 2007, pp. 23–45.

[10] W. Duan, B. Gu, and A.B. Whinston, "Do Online Reviews Matter? — An Empirical Investigation of Panel Data," *Decision Support Syst.*, vol. 45, no. 4, Nov. 2008, pp. 1007–1016.

[11] Y. Liu, "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *J. Marketing*, vol. 70, no. 3, July 2006, pp. 74–89.

[12] L. Qin, "Word-of-Blog for Movies: A Predictor and an Outcome of Box Office Revenue?," *J. Electron. Commerce Res.*, vol. 12, no. 3, 2011, pp. 187–198.

[13] A. Bhattacherjee, "Understanding Information Systems Continuance: An Expectation-Confirmation Model," *J. MIS Quarterly*, vol. 25, no. 3, Sept. 2001, pp. 351–370.

[14] A. Bhattacherjee and G. Premkumar, "Understanding Changes in Belief and Attitude toward Information Technology Usage: A Theoretical Model and Longitudinal Test," *J. MIS Quarterly*, vol. 28, no. 2, June 2004, pp. 229–254.

[15] M.H. Hsu, C.M. Chiu, and T.L. Ju, "Determinants of Continued Use of the WWW: An Integration of Two Theoretical Models," *Ind. Manag. Data Syst.*, vol. 104, no. 9, 2004, pp. 766–775.

[16] M.C. Lee, "Explaining and Predicting Users' Continuance Intention toward E-learning: An Extension of the Expectation-Confirmation Model," *Comput. Educ.*, vol. 54, no. 2, Feb. 2010, pp. 506–516.

[17] C.S. Lin, S. Wu, and R.J. Tsai, "Integrating Perceived Playfulness into Expectation-Confirmation Model for Web Portal Context," *Inf. Manag.*, vol. 42, no. 5, July 2005, pp. 683–693.

[18] V. McKinney, K. Yoon, and F.M. Zahedi, "The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach," *Inf. Syst. Res.*, vol. 13, no. 3, Sept. 2002, pp. 296–315.

[19] J.Y.L. Thong, S.J. Hong, and K.Y. Tam, "The Effects of Post-Adoption Beliefs on the Expectation-Confirmation Model for Information Technology Continuance," *Int. J. Human-Comput. Studies*, vol. 64, no. 9, Sept. 2006, pp. 799–810.

[20] B. Gu, J. Park, and P. Konana, "The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products," *Inf. Syst. Res.*, vol. 23, no. 1, Mar. 2012, pp. 182–196.

[21] N. Hu, L. Liu, and J.J. Zhang, "Do Online Reviews Affect Product Sales? The Role of Reviewer Characteristics and Temporal Effects," *Inf. Technol. Manag.*, vol. 9, no. 3, Sept. 2008, pp. 201–214.

[22] J. Yang et al., "The Heterogeneous Effect of WOM on Product Sales: Why the Effect of WOM Valence is Mixed?," *European J. Marketing*, vol. 46, no. 11–12, Nov. 2012, pp. 1523–1538.

[23] F.F. Reichheld and W.E. Sasser, "Zero Defections: Quality Comes to Services," *Harvard Business Rev.*, vol. 68, no. 5, Sept. 1990, pp. 105–111.

[24] Y. Jung, "The Influence of Characteristics (Volume, Consensus, Informant) of Electric Word-of-Mouth on Customers' Choice," *Broadcast. Commun.*, vol. 14, no. 1, Mar. 2013, pp. 47–76.

[25] G. Cui, H.K. Lui, and X. Guo, "The Effect of Online Consumer Reviews on New Product Sales," *Int. J. Electron. Commerce*, vol. 17, no. 1, Sept. 2012, pp. 39–58.

[26] A. Versaci, "*A Theory and Evidence… You Talkin'to Me?*, *Using Internet Buzz as an Early Predictor of Movie Box Office*," M.S. thesis, New York University, NY, USA, 2009.

[27] K. Chang, "A Study on the Effect of Pre-opening Promotion and Post-Opening Word of Mouth in the Motion Picture Industry," *J. Marketing Manag. Res.*, vol. 13, no. 4, Oct. 2008, pp. 23–24.

[28] S. Moon, P.K. Bergey, and D. Iacobucci, "Dynamic Effects among Movie Ratings, Movie Revenue, and Viewer Satisfaction," *J. Marketing*, vol. 74, no. 1, Jan. 2010, pp. 108–121.

[29] P. Bhargavi and S. Jyothi, "Applying Naïve Bayes Data Mining Technique for Classification of Agricultural Land Soils," *Int. J. Comput. Sci. Netw. Security*, vol. 9, no. 8, Aug. 2009, pp. 117–122.

[30] A. Zaeri and M.A. Nematbakhsh, "A Framework for Semantic Interpretation of Noun Compounds Using Tratz Model and Binary Features," *ETRI J.*, vol. 34, no. 5, Oct. 2012, pp. 743–752.

Hyunmi Baek is an assistant professor in the Department of Information Sociology at Hanyang University, Ansan, Rep. of Korea. She earned her PhD in management information systems from the College of Business Administration, Seoul National University, Rep. of Korea, in 2013. She earned her MA in information technology management from Information and Communications University, Daejeon, Rep. of Korea and her BS from Pohang University of Science and Technology, Rep. of Korea. From 2003 to 2013, she worked for Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea. Her research interests include electronic word of mouth, social media, information adoption, and open collaboration.

Joongho Ahn is a professor of Information Systems at the Graduate School of Business, Seoul National University, Rep. of Korea. He earned his PhD from Stern School of New York University, USA and his MPA and BA from Seoul National University, Rep. of Korea. His areas of interests include business transformation, electronic commerce and inter-organizational information systems, strategic use of information technology in government and business, and IT governance. He was past president of the Korea Society of MIS and also past president and founder of the Korean Association of IT Governance. He was a member of the presidential council on National Information Strategy, Rep. of Korea.

Sehwan Oh is a PhD candidate at the College of Business Administration at Seoul National University, Rep. of Korea and a senior researcher at Korea International Trade Association, Seoul, Rep. of Korea. He earned his MS in information technologies from the School of Computer Science, Carnegie Mellon University, Pittsburgh, USA and his BA in economics from the College of Social Sciences, Seoul National University, Rep. of Korea. He has over ten years of industry experience in IS and international trade areas. His current research interests include electronic word of mouth, social media, and mobile commerce.

http://dx.doi.org/10.4218/etrij.14.0113.0732