ARTICLE

<https://doi.org/10.1057/s41599-020-00578-9>

OPEN

The impact of online review helpfulness and word of mouth communication on box office performance predictions

Sangjae Lee¹ & Joon Yeon Choeh²

While electronic word-of-mouth (eWOM) variables, such as volume and valence have been posited in previous studies to consistently affect product sales, there is a lack of studies on the different contexts and outcomes that affect the importance of eWOM variables. In order to fill this gap, this study attempts to use the helpfulness of reviews and reviewers as moderators to predict box office revenue, comparing the prediction performances of business intelligence (BI) methods (random forest, decision trees using boosting, the *k*-nearest neighbor method, discriminant analysis) using eWOM between high and low review or reviewer helpfulness subsample in the Korean movie market scrawled from the Naver Movies website. The results of applying machine learning methods show that movies with more helpful reviews or those that are reviewed by more helpful reviewers show greater prediction performance, and review and reviewer helpfulness improve the prediction power of eWOM for box office revenue. The prediction performance will improve if the characteristics of eWOM are likely to be combined to contribute to box office revenue to a greater extent.

¹ College of Business Administration, Sejong University, Seoul 05006, Korea. ² Department of Software, Sejong University, Seoul 05006, Korea.
✉email: sangjae@sejong.ac.kr

Introduction

Box office revenue represents a crucial indicator for assessing the success of a movie (Liu et al., 2016; McKenzie, 2008). The effective forecasting of box office revenue levels is crucial to decrease market risk, increase the competitiveness of the movie industry, and facilitate the advancement of movie-related markets.

Previous electronic word-of-mouth (eWOM) literature posited that there exists an empirically validated relationship between eWOM variables and product sales. For instance, eWOM has an effect on the customer purchase decisions (Chen et al., 2015). In addition, eWOM affects repurchase intentions while having a negative effect on perceived risk (Liang et al., 2018). Due to the enhanced role of social networks for information sharing, eWOM has a greater effect on purchase decisions (Liang et al., 2011). WOM variables such as valence and volume are strong determinants of product sales in different product contexts (Chintagunta et al., 2010; Elwalda et al., 2016; Rui et al., 2013; Zhu and Zhang, 2010).

Predicting box office revenue is challenging as doing so should consider the nonlinear effects of factors that affect box office revenue, including eWOM as well as movie-related variables (Zhang et al., 2009). Several aspects motivated this investigation of the influence of eWOM on movie performance outcomes.

First, the study intends to use the helpfulness of reviews and reviewers as moderators to predict box office revenue and to apply business intelligence (BI) methods to compare the prediction power of eWOM between high and low review or reviewer helpfulness subsamples. Despite earlier work that focused on the effect of eWOM on movie performance outcomes (Chintagunta et al., 2010; Rui et al., 2013) or the effect of eWOM on review helpfulness (Lee and Choeh, 2016; Siering et al., 2018), the prediction performance with regard to box office revenue using eWOM as affected by the helpfulness of reviews and reviewers has rarely been studied. There is no reported study on the use of the helpfulness of reviews and reviewers as a moderator influencing the prediction performance of box office revenue using various BI methods, such as decision trees and the *k*-nearest neighbor and discriminant analysis methods. This study employs these three machine learning methods given the goal of utilizing widely used non-statistical and statistical learning methods to examine how they differ in terms of prediction performance, as well as a combination of classifications to provide an operational and fast means of obtaining better solutions to optimize predictive measures (Shmueli et al., 2016). The goal of our study is to bridge the gap in previous studies on the use of the helpfulness of reviews and reviewers as a moderator influencing the aforementioned prediction performance by conducting a subsample analysis of movies with high and low review or reviewer helpfulness levels.

Helpfulness represents review effectiveness, i.e., a middle measure influencing the subsequent purchasing intentions and sales of a product (Wu, 2017). The helpfulness of reviews can have an influence on the relationship between eWOM and movie performance (Lee and Choeh, 2018). Helpfulness can represent diverse review quality levels and provide predictive power (Yu et al., 2012). Given that the helpfulness of reviews and reviewers can be regarded as qualities of eWOM variables, this study intends to compare the prediction performance of various machine learning methods using eWOM data between two subsamples of movies which have high and low review or reviewer helpfulness levels. Our study considers each movie as a unit of analysis and provides an average review of reviewer helpfulness for each movie.

Second, the study intends to test the explanatory power of eWOM variables, in this case volume and valence, which are commonly examined in the literature in cases where movies are

grouped into subsamples with high and low review or reviewer helpfulness. The effects of eWOM on movie performance revenue across high and low review or reviewer helpfulness are rarely investigated. The study examines whether these eWOM variables have explanatory power with regard to box office revenue when the movie samples are divided into high and low review or reviewer helpfulness subsamples.

Third, this study extends previous literature on eWOM for movies given its use of an international market, in this case Asia. Further, there is a lack of study using international movie markets other than that of the US in the box office prediction literature (Kim et al., 2015; Liu et al., 2016). This study uses movies distributed in the Korean movie market. Box office revenue predictions using eWOM in international movie markets will provide insights in the box office prediction literature, as the effect of eWOM can differ due to cultural issues.

Theoretical background

Previous studies have addressed the factors that affect box office revenue, finding that eWOM and movie characteristics both have effects. The characteristics of eWOM include the valence (rating) of the review, the depth (length) of the review, the volume (the number of) of reviews, the sentiments in the reviews (proportion of positive and negative reviews), and review extremity levels (Du et al., 2014; Lee and Choeh, 2018; Rui et al., 2013). Movie characteristics include the star or director power, revenue of similar movies, nationality, genre, opening times, and sequels.

Previous eWOM studies posited that eWOM variables have an influence on consumer decision-making processes with regard to purchases (Chen et al., 2015). eWOM is positively associated with repurchase intention along with perceived value while it exerts a negative influence on perceived risk (Liang et al., 2018). eWOM has a greater effect on purchase decisions, as consumers tend to focus on communicating information on social networks with acquaintances (Liang et al., 2011). eWOM variables such as valence and volume have been posited in previous studies to affect product sales consistently, while there are different contexts and outcomes that affect the significance of these variables (Chintagunta et al., 2010; Duan et al., 2008; Elwalda et al., 2016; Zhu and Zhang, 2010). For instance, social reference systems, such as Facebook “likes” (FBLs), enhance sales in social commerce, leading to variations across products and deals (Lee et al., 2015). Further, review sentiments, influenced by positive or negative terms, also exert a direct influence on product sales (Hu et al., 2014).

Previous studies have suggested statistical methods or methods based on machine learning for box office predictions. Examples include regression models (Elliott and Simmons, 2008) or support vector regressions (Kim et al., 2015; Liu et al., 2016), neural networks (Delen et al., 2007), and Bayesian networks (Lee and Chang, 2009).

Due to the volatility of movie box office revenue levels (Liu et al., 2016; Zhang et al., 2009), it is necessary to use a sample that promotes the forecasting accuracy of box office revenue. The prediction performance will improve if the characteristics of eWOM are likely to be combined to contribute to box office revenue to a greater extent. For instance, review helpfulness is one of the eWOM variables that can interact with other eWOM variables to affect box office revenue (Lee and Choeh, 2018). Thus, the prediction performance is likely to be affected if the sample is extracted such that it exhibits higher review helpfulness.

Research model

Online reviews will affect sales if customers depend on reviews greatly when deciding on a purchase (Zhu and Zhang, 2010). The

important qualities of online reviews to promote consumers' dependence include comprehensiveness, timeliness, and relevance (Cheung et al., 2008); factual information (Papathanassis and Knolle, 2011); and the latest information (Gretzel et al., 2007; Hu et al., 2014). Review helpfulness represents the quality of eWOM in terms of the textual and source characteristics of eWOM, and this variable can enhance consumers' reliance on online reviews, even influencing sales (Lee and Choeh, 2016; Hu et al., 2014). These helpfulness aspects represent the review's total perceived information utility (Li et al., 2013) and the perceived usefulness of the review—the extent that consumers believe a review to be helpful for promoting decisions to purchase (Zhou and Guo, 2017).

Helpfulness is linked to review effectiveness; i.e., it is an intermediate measure affecting subsequent measures such as purchase intentions and sales of the product (Wu, 2017). Thus, review helpfulness can have an effect on the relationship between eWOM and movie performance (Lee and Choeh, 2018). Online reviews are highly diverse in quality and in how they provide predictive power (Yu et al., 2012). Given that the perception of helpfulness is a prerequisite with regard to alternative evaluations for purchasing decisions (Hong et al., 2017; Mudambi and Schuff, 2010) and considering that the helpfulness of reviews and reviewers can be considered as qualities of eWOM variables, our study suggests that eWOM variables with greater levels of helpfulness of reviews and reviewers provide greater prediction accuracy when they are used as input variables in machine learning methods. Thus, it can be tested as to whether eWOM variables with higher textual and source quality levels can improve prediction performance outcomes when they are adopted as independent variables for machine learning methods. Accordingly, this study suggests the following hypotheses.

Hypothesis 1: Movies with greater review helpfulness have a greater prediction power of eWOM for current box office revenue levels compared to movies with lower review helpfulness.

Hypothesis 2: Movies with greater reviewer helpfulness have a greater prediction power of eWOM for current box office revenue levels compared to movies with lower reviewer helpfulness.

Figure 1 suggests the research framework and Fig. 2 presents the procedure used in this study. In order to test the prediction power of eWOM on movie performance, this study undertakes a multiple regression analysis to determine adjusted *R* square values to compare high and low reviews or reviewer helpfulness

subsamples. This study uses four machine learning methods, random forest, decision trees using boosting, *k*-nearest neighbor, and discriminant analysis, to compare prediction accuracy levels in terms of classification errors for each of the four machine learning methods. The procedure of the study is presented in Fig. 2. This study uses two advanced decision tree methods to improve performance outcomes: random forest and decision trees using boosting. For the random forest method, this study fits a classification tree to multiple random samples which are sampled with replacement from a training sample and combines the reductions from individual trees (10 trees in total) to obtain improved predictions. For the decision tree using boosting method, the study generates multiple random samples (10 samples, by sampling with replacement from the training sample) so that misclassified observations (or observations with large prediction errors) have higher probabilities of selection, based on procedures suggested in Shmueli et al. (2016). For the *k*-nearest neighbor method, this study chooses the most similar movie ($k = 1$) from the training sample and uses the revenue class of the most similar movie.

Prediction methods

This study converts each box office revenue level for 3 weeks into a categorical variable according to whether the revenue is in the top 20% of the revenue distribution (1 or 0); this discretization approach for a dependent variable has been applied in conjunction with machine learning methods for the classification of box office revenue or other types to make many learning algorithms faster and more accurate for both users and experts, and discretization approaches are easier to explain, understand, and use (Delen et al., 2007; Liu et al., 2002), as this study assumes that finding movies with results in the top 20% of box office revenue is of great concern to movie producers and distributors as opposed to predicting revenue accurately. Further, as the purpose of this study is to compare the prediction performance outcomes between high and low review or reviewer subsamples and not to forecast revenue levels accurately, the study intends to use the categorical variable of box office revenue in relation to machine learning methods.

This study uses random forest, decision trees using boosting, the *k*-nearest neighbor method, discriminant analysis for the classification of box office classes. This study adopts these three machine learning methods as the intention here is to use both extensively used non-statistical and statistical learning methods to

Fig. 1 Research model.

Fig. 2 Procedures of this study.

examine how they differ in terms of their prediction performance, and a combination of classifications to provide an operational and rapid means of obtaining better solutions to optimize predictive measures (Shmueli et al., 2016). These methods are used to examine the differences in prediction performance outcomes between high and low reviews and reviewer subsamples.

In random forests and decision trees, multiple decision tree models are trained and combined using multiple random trees with different tree structures or multiple sampling with replacement of the training sample on the same tree structure (Shmueli et al., 2016). Random forest and decision tree methods with boosting are ensemble methods in which the most popular class is chosen from among multiple classifications as suggested by multiple weak learner models. There are no pruning conditions or limits when growing trees. The random forest method uses randomly selected features among independent variables to suggest multiple classifications. This study sets an integer value as a bootstrapping random number seed. Decision trees using boosting suggest multiple classifications using random sampling with replacement so that misclassified observations have higher probabilities of selection. The steps in the decision tree method using boosting include: (1) fitting a model to the training data, (2) drawing a sample from the data so that misclassified observations (or observations with the largest prediction errors) have higher probabilities of selection, and (3) fitting the model to the new sample. For random forest and decision tree methods using boosting, 10 different learner models are combined to test the ability of this strategy to predict box office revenue.

The k -nearest-neighbor regression method suggests the most similar movie ($k = 1$) in the training data set for each record of the target data set. Then, based on the classes to which those closest records belong, we assign a class of box office to the record that is to be classified. Discriminant analysis is a technique which creates discriminant functions for producing a classification score, and the record in target data set is suggested as class i if the i th classification score has the greatest value.

Data

The eWOM data for our study were crawled from the Naver Movies website (see <http://movie.naver.com/>). The Naver Movies website was selected because Naver is the most visited portal site in Korea (InternetTrend, 2020). The revenues and power data are obtained from KOFIC (see <https://www.kofic.or.kr/>). This study targeted movies released between January of 2014 and May of 2016. Overall, 1798 movies are included in the final sample, with collection taking place in June of 2016. For the selected motion pictures, box office ticket sales data, eWOM and movie-specific data were crawled between one week prior to the release date and 3 weeks after the release date. The study chooses independent variables from previous eWOM studies (e.g., Chintagunta et al., 2010; Duan et al., 2008; Mudambi and Schuff, 2010), which are significant in many cases (Table 1).

For each of the three machine learning methods utilized in this study, three dependent variables are suggested: the weekly box office earnings for the first, second and third week after the release of the movie, based on literature pertaining to box office reports or predictions in order to consider time-varying values of box office revenue (Chintagunta et al., 2010; Duan et al., 2008; Kim et al., 2015; Liu et al., 2016).

The dependent variable was discretized into two classes according to whether it belongs to the top 20% of the revenue distribution of the sample (1 if within the top 20%, 0 otherwise). The dependent variable is box office revenue and the predictors are six eWOM variables and seven movie-related variables which are used as control variables. All movie-related variables are of the

non-time-varying type. The variables adopted in the study are presented in Table 2. The high and low helpfulness subsamples are formed into separate samples according to the average helpfulness value. For instance, the average value of helpfulness in the prerelease stage is 0.589, and we created high or low review helpfulness subsamples according to whether the helpfulness value exceeded 0.589.

Results and discussions

This paper utilizes linear and non-linear models for box office predictions. In order to predict box office revenue at week t , the study uses the eWOM variables at week $t-1$. Tables 3–8 present the multiple regression analysis results using the six eWOM variables of the average number of reviews, average review rating, average review extremity level, average length of a review in words, average number of emotional reviews, and average number of positive reviews. The control variables include star power, awards, sequels, the timing of the release, genre (one binary variable representing drama), and nationality (two binary variables representing Korea and US). In total, 1798 movies are used in the multiple regression analysis.

The results of the multiple regression analysis (from Tables 3–8) show that the adjusted R square value is higher in the high review or reviewer helpfulness subsample than in the low review or reviewer subsample. The adjusted R square decreases as the release date moves further into the past. This shows that while the explanatory power of eWOM becomes weaker over time after the release, the explanatory power of eWOM in the high review or reviewer subsample is greater than that in the low review or reviewer subsample. Thus, eWOM variables provide greater explanatory power when the helpfulness values of reviews and reviewers are higher. Accordingly, the study shows the explanatory power of the eWOM variables of volume and valence, which have been investigated extensively in the literature when movies are divided into high and low review or reviewer helpfulness subsamples, which is insightful given that studies of the effect of eWOM on movie performance revenue between those with high and low review or reviewer helpfulness are nearly nonexistent. This indicates that review and reviewer helpfulness are crucial moderating factors increasing the explanatory power of eWOM for box office revenue.

The review volume (the average number of reviews) provides a consistent effect on revenue across all subsamples. The average number of reviews has a consistently significant effect on movie performance in both the high and low reviewer or reviewer subsamples for the first, second and third weeks after the release of the movie. This supports earlier work which noted a positive influence of eWOM on movie performance (Duan et al., 2008; Chintagunta et al., 2010). The effects of the average review rating and the average review extremity level are partially significant. Further, regarding the review helpfulness subsamples, the effect of the average review rating is negative for the first week after the release of the movie and positive for the second and third weeks after the release. For the reviewer helpfulness subsamples, the effect of the average review rating becomes negative for the first week after the release of the movie and positive for the third week after the release. This shows that rather than the review volume, the review valence has less of an effect on box office outcomes, a finding consistent with earlier work (Duan et al., 2008). In addition, the effect of review ratings becomes positive over passes after the release of the movie, which validates the effect of the valence as found in previous studies, which posit that valence exerts effects on box office outcomes (Chintagunta et al., 2010).

These results show that for experience goods such as movies, less extreme reviews may have a greater effect on movie

Table 1 Descriptive statistics and correlations among variables.

(a) Award, sequel, timing of release			
	Class	Count	Percent of count
Award	Yes	1764	98.1
	No	34	1.9
Sequel	Yes	1766	98.2
	No	32	1.8
Timing of release	Yes	1106	61.5
	No	692	38.5
Actor power	Maximum	Mean	Standard deviation
	2551.25 million (dollars)	103.41	224.51

(b) Mean sales of box office in each genre (dollars)			
Genre	Mean	N	Percent
Drama	2.39 million	823	45.8
Comedy	3.08 million	219	12.2
Melodrama/romance	1.89 million	197	11
Action	5.71 million	123	6.8
Animation	2.12 million	118	6.6
Documentary	0.27 million	116	6.5
Horror	1.35 million	73	4.1
Thriller, mystery	3.13 million	67	3.7
Crime	6.11 million	26	1.4
Science Fiction	15.76 million	11	0.6
Art	3.89 million	7	0.4
Family	0.92 million	7	0.4
Fantasy	9.22 million	5	0.3
History	14.70 million	3	0.2
Unclassified	3.99 million	2	0.1
War	11.57 million	1	0.1
Total		1798	100.0

(c) Nationality of movies			
Nationality	Frequency	Percent	
Korea	715	39.8	
United States	484	26.9	
Japan	160	8.9	
France	110	6.1	
England	81	4.5	
Others	248	13.8	
Total	1798	101.0	

(d) eWOM				
	N	Minimum	Maximum	Std. deviation
Average number of reviews at prerelease	1798	1.00	84371.00	4101.75947

Table 1 (continued)						
(d) eWOM						
	N	Minimum	Maximum	Mean	Std. deviation	
Average number of reviews at week 1	1776	1.00	9834.00	120.3170	370.18362	
Average number of reviews at week 2	1680	1.00	3810.00	44.4970	128.19458	
Average review rating at prerelease	1798	1.00	10.00	7.3482	1.42108	
Average review rating at week 1	1776	1.00	10.00	7.9141	1.50521	
Average review rating at week 2	1680	0.00	10.00	8.2851	1.57850	
Average review extremity at prerelease	1798	1.06	6.93	2.4715	0.90184	
Average review extremity at week 1	1776	0.07	6.93	2.4260	0.89608	
Average review extremity at week 2	1680	0.07	7.93	2.3210	0.94080	
Average length of a review in words at prerelease	1798	5.33	39.38	19.9016	5.61789	
Average length of a review in words at week 1	1749	2.00	62.00	18.4766	6.65295	
Average length of a review in words at week 2	1344	2.00	72.00	17.1065	6.81981	
Average number of emotional reviews at prerelease	1790	1.00	4.55	2.3780	0.53974	
Average number of emotional reviews at week 1	1720	1.00	10.00	2.2145	0.74633	
Average number of emotional reviews at week 2	1294	1.00	6.50	2.0949	0.73655	
Average number of positive reviews at prerelease	1790	0.25	2.70	1.2319	0.32712	
Average number of positive reviews at week 1	1720	0.00	8.00	1.2342	0.51221	
Average number of positive reviews at week 2	1294	0.00	4.29	1.2092	0.54656	
Helpfulness of review at prerelease	1786	0.00	1.00	0.5890	0.19312	
Helpfulness of review at week 1	945	0.00	1.00	0.5725	0.29752	
Helpfulness of review at week 2	673	0.00	1.00	0.6190	0.33551	
Helpfulness of reviewer	1783	0.17	1.00	0.5559	0.06263	
(e) Box office revenue						
	N	Minimum	Maximum	Mean	Std. deviation	
Box office revenue at the opening day	1798	0.00	4.09 million (dollars)	0.13 million	0.30 million	
Box office revenue at week 1	1798	14	28.27	1.13	2.68	
Box office revenue at week 2	1750	14	28.60	0.72	2.23	
Box office revenue at week 3	1591	0.00	25.89	0.43	1.66	
(f) Correlation among eWOM at prerelease						
	Average number of reviews at prerelease	Average review rating at prerelease	Average review extremity at prerelease	Average length of a words at prerelease	Average number of emotional reviews at prerelease	Average number of positive reviews at prerelease
Average number of reviews at prerelease	1	0.131(**)	-0.003	-0.061(*)	-0.034	0.003
Average review rating at prerelease	0.131(**)	1	-0.891(**)	0.021	-0.017	0.179(**)
Average review extremity at prerelease	-0.003	-0.891(**)	1	-0.071(**)	-0.010	-0.203(**)
Helpfulness of reviewer						
Helpfulness of review at prerelease						
Helpfulness of review at week 1						
Helpfulness of review at week 2						
Helpfulness of reviewer						

Table 1 (continued)									
(f) Correlation among eWOM at prerelease									
	Average number of reviews at prerelease	Average review rating at prerelease	Average review extremity at prerelease	Average length of a review in words at prerelease	Average number of emotional reviews at prerelease	Average number of positive reviews at prerelease	Helpfulness of review at prerelease	Helpfulness of reviewer	
Average length of a review in words at prerelease	−0.061(*)	0.021	−0.071(**)	1	0.904(**)	0.727(**)	0.094(**)	0.313(**)	
Average number of emotional reviews at prerelease	−0.034	−0.017	−0.010	0.904(**)	1	0.812(**)	0.069(**)	0.290(**)	
Average number of positive reviews at prerelease	0.003	0.179(**)	−0.203(**)	0.727(**)	0.812(**)	1	0.094(**)	0.285(**)	
Helpfulness of review at prerelease	−0.018	0.170(**)	−0.168(**)	0.094(**)	0.069(**)	0.094(**)	1	0.275(**)	
Helpfulness of reviewer	0.013	0.134(**)	−0.127(**)	0.313(**)	0.290(**)	0.285(**)	0.275(**)	1	
*p < 0.05; **p < 0.01.									

Table 2 eWOM and movie related variables.

Category	Variables	Description	Number of possible values
eWOM variables	Average number of reviews	Represents the average number of user reviews until the movie is released in a new market (time varying)	Real values
	Average review rating	Signifies the average review rating (time varying)	Real values
	Average review extremity	Shows the absolute value of the difference between the reviewers' ratings and the average of all user ratings (time varying)	Real values
	Average length of a review in words	Represents the length of a review in words (time varying)	Real values
	Average number of emotional reviews	Represents the average number of emotional reviews (time varying)	Real values
	Average number of positive reviews	Represents the average number of positive reviews (time varying)	Real values
	Helpfulness of review	Shows the helpfulness of review for each movie (time varying)	Real values
	Helpfulness of reviewer	Represents average helpfulness of reviewers for each movie (not time varying)	Real values
Movie-related variables	Star power	Indicates the total revenue earned by the actors casted in each movie	Real values
	Award	Dummies for academy winners/nominees	One binary variable
	Sequel	Indicates whether a movie is a sequel (value of 1) or not (value of 0).	One binary variable
	Timing of release	Indicates if movie is releases in holiday seasons (July or December). Binary variable for holiday starting.	One binary variable
	Genre	Represents the content category the movie most likely belongs to. A movie can be belonged to more than one content category and the category to which the movie most likely belongs to is chosen. Represents a binary variable for a drama genre among genres of movies (drama, melodrama/ romance, animation, comedy, action, documentary) to which our sample belongs to in the greatest proportions.	One binary variable for each drama genre
	Nationality	Represents two binary variables for two nationalities (Korea and US) among five nationalities of movies (Korea, US, England, France, others) to which our sample belongs to in the greatest proportions. For example, in case of France, the values for two binary variables are 0 and 0.	Two binary variables

Table 3 Regression analysis for box office revenue for the first week after release using review helpfulness subsample.

Variables	High review helpfulness subsample (917)			Low review helpfulness subsample (881)		
	Standard coefficient	t	Sig.	Standard coefficient	t	Sig.
Average number of reviews (prerelease)	0.623	25.740	0.000	0.598	22.992	0.000
Average review rating (prerelease)	−0.041	−0.906	0.365	−0.157	−2.861	0.004
Average review extremity (prerelease)	−0.082	−1.689	0.092	−0.188	−3.347	0.001
Average length of reviews in words (prerelease)	−0.131	−2.745	0.006	−0.060	−1.102	0.271
Average number of emotional reviews (prerelease)	0.044	0.773	0.440	−0.024	−0.364	0.716
Average number of positive reviews (prerelease)	0.029	0.785	0.433	0.037	0.869	0.385
Adjusted R square	0.638			0.567		
F	124.1			89.3		

performance than extreme reviews. For the review and reviewer helpfulness subsamples, the average review extremity level exerts a negative effect on revenue for the first week after the release of the movie. Thus, the negative influence of the review extremity level during this first week can be explained by evaluations of the movie showing two-sided comments representing good or poor opinions about the movie. This supports Mudambi and Schuff (2010) in that for experience goods such as movies, reviews with moderate ratings turn out to be more helpful than reviews with extreme ratings. Further, this is related to the mixed effect of the valence of reviews, which is negative for the first week and positive for the third week after the release of the movie.

In order to test hypotheses 1 and 2 further, this study applied four BI methods (random forest, decision trees with boosting, the *k*-nearest neighbor method, and discriminant analysis).

The effect of the sample size on the prediction performance should be removed for a fair comparison of the prediction performance between the high and low review or reviewer helpfulness subsamples. The numbers of movies in the high and low review or reviewer helpfulness subsamples is set to be identical to the minimum of the two subsamples. Thus, pairs of 881, 505, and 368 movies are ultimately sorted into the subsamples of high and low review helpfulness for predicting box office revenue levels at weeks 1, 2, and 3, respectively (see the numbers in parentheses in Table 9). For the subsamples of high and low reviewer helpfulness (which are not time-varying), the pair of 578 movies is included to predict box office revenue levels during weeks 1, 2, and 3 (see the numbers in parentheses in Table 10). The subsamples in three pairs are divided into two subsets: a training set and a validation set.

Table 4 Regression analysis for box office revenue for the second week after release using review helpfulness subsample.

Variables	High review helpfulness subsample (505)			Low review helpfulness subsample (1293)		
	Standard coefficient	t	Sig.	Standard coefficient	t	Sig.
Average number of reviews (first week)	0.386	10.822	0.000	0.506	20.780	0.000
Average review rating (first week)	0.037	0.711	0.477	0.087	2.190	0.029
Average review extremity (first week)	−0.010	−0.177	0.859	−0.014	−0.357	0.721
Average length of reviews in words (first week)	−0.015	−0.274	0.785	−0.053	−1.697	0.090
Average number of emotional reviews (first week)	−0.007	−0.102	0.919	−0.003	−0.088	0.930
Average number of positive reviews (first week)	−0.017	−0.352	0.725	0.006	0.174	0.862
Adjusted R square	0.518			0.431		
F	40.3			70.9		

Table 5 Regression analysis for box office revenue for the third week after release using review helpfulness subsample.

Variables	High review helpfulness subsample (368)			Low review helpfulness subsample (1430)		
	Standard coefficient	t	Sig.	Standard coefficient	t	Sig.
Average number of reviews (second week)	0.341	4.692	0.000	0.146	5.096	0.000
Average review rating (second week)	0.148	1.535	0.127	0.117	2.667	0.008
Average review extremity (second week)	0.084	0.827	0.410	0.040	0.871	0.384
Average length of reviews in words (second week)	−0.010	−0.112	0.911	−0.033	−0.871	0.384
Average number of emotional reviews (second week)	0.047	0.378	0.706	−0.030	−0.609	0.542
Average number of positive reviews (second week)	−0.065	−0.574	0.567	0.033	0.791	0.429
Adjusted R square	0.177			0.246		
F	3.912			26.106		

Table 6 Regression analysis for box office revenue for the first week after release using reviewer helpfulness subsample.

Variables	High reviewer helpfulness subsample (578)			Low reviewer helpfulness subsample (1220)		
	Standard coefficient	t	Sig.	Standard coefficient	t	Sig.
Average number of reviews (prerelease)	0.602	22.072	0.000	0.593	26.024	0.000
Average review rating (prerelease)	0.018	0.345	0.730	−0.159	−3.280	0.001
Average review extremity (prerelease)	−0.017	−0.320	0.749	−0.193	−3.850	0.000
Average length of reviews in words (prerelease)	−0.088	−1.991	0.047	−0.041	−0.868	0.386
Average number of emotional reviews (prerelease)	0.055	1.029	0.304	−0.021	−0.365	0.716
Average number of positive reviews (prerelease)	0.026	0.685	0.494	0.036	1.036	0.300
Adjusted R square	0.724			0.532		
F	117.5			107.0		

Table 7 Regression analysis for box office revenue for the second week after release using reviewer helpfulness subsample.

Variables	High reviewer helpfulness subsample (578)			Low reviewer helpfulness subsample (1220)		
	Standard coefficient	t	Sig.	Standard coefficient	t	Sig.
Average number of reviews (first week)	0.419	13.473	0.000	0.529	20.643	0.000
Average review rating (first week)	0.066	1.456	0.146	0.060	1.439	0.150
Average review extremity (first week)	0.021	0.434	0.665	−0.055	−1.293	0.196
Average length of reviews in words (first week)	−0.028	−0.629	0.529	−0.023	−0.675	0.500
Average number of emotional reviews (first week)	−0.005	−0.091	0.927	−0.022	−0.541	0.589
Average number of positive reviews (first week)	−0.011	−0.256	0.798	0.016	0.451	0.652
Adjusted R square	0.583			0.421		
F	58.5			64.8		

Table 8 Regression analysis for box office revenue for the third week after release using reviewer helpfulness subsample.

Variables	High reviewer helpfulness subsample (578)			Low reviewer helpfulness subsample (1220)		
	Standard coefficient	t	Sig.	Standard coefficient	t	Sig.
Average number of reviews (second week)	0.093	1.884	0.061	0.146	5.096	0.000
Average review rating (second week)	0.035	0.501	0.617	0.117	2.667	0.008
Average review extremity (second week)	−0.033	−0.452	0.651	0.040	0.871	0.384
Average length of reviews in words (second week)	0.055	0.802	0.423	−0.033	−0.871	0.384
Average number of emotional reviews (second week)	−0.001	−0.008	0.994	−0.030	−0.609	0.542
Average number of positive reviews (second week)	−0.031	−0.397	0.692	0.033	0.791	0.429
Adjusted R square	0.415			0.246		
F	15.6			26.1		

Table 9 Average prediction error among machine learning methods for high and low review helpfulness groups.

Box office revenue	Number of prediction error data used to compare	Random forests	Decision trees using boosting	k-nearest neighbor method	Discriminant analysis
<i>Week 1</i>					
Low helpfulness (881) ^a	30	0.120	0.131	0.152	0.179
High helpfulness (881)	30	0.077	0.089	0.127	0.148
t-value (p-value)		2.255 (0.023)	1.969 (0.054)	1.256 (0.214)	1.331 (0.188)
<i>Week 2</i>					
Low helpfulness (505)	34	0.233	0.250	0.235	0.287
High helpfulness (505)	34	0.111	0.093	0.136	0.113
t-value (p-value)		4.960 (0.000)	6.290 (0.000)	4.172 (0.000)	5.919 (0.000)
<i>Week 3</i>					
Low helpfulness (368)	37	0.345	0.381	0.355	0.309
High helpfulness (368)	37	0.259	0.309	0.360	0.247
t-value (p-value)		2.430 (0.018)	1.879 (0.064)	0.146 (0.884)	1.800 (0.076)

^aThe number in parenthesis represents the number of movies to be compared.

Table 10 t-test of the difference in prediction error among machine learning methods for high and low reviewer helpfulness groups.

Box office revenue	Number of prediction error data used to compare	Random forests	Decision trees using boosting	k-nearest neighbor method	Discriminant analysis
<i>Week 1</i>					
Low helpfulness (578) ^a	29	0.150	0.184	0.207	0.246
High helpfulness (578)	29	0.055	0.062	0.081	0.104
t-value (p-value)		4.857 (0.000)	5.175 (0.000)	5.205 (0.000)	4.923 (0.000)
<i>Week 2</i>					
Low helpfulness (578)	29	0.226	0.273	0.267	0.291
High helpfulness (578)	29	0.104	0.130	0.133	0.123
t-value (p-value)		4.466 (0.000)	6.343 (0.000)	5.774 (0.000)	5.019 (0.000)
<i>Week 3</i>					
Low helpfulness (578)	29	0.308	0.335	0.380	0.305
High helpfulness (578)	29	0.277	0.291	0.356	0.235
t-value (p-value)		1.157 (0.252)	1.450 (0.153)	0.696 (0.489)	2.246 (0.029)

^aThe number in parenthesis represents the number of movies to be compared.

N-fold cross validation of the sample and the three dependent variables of the box office revenue are utilized to investigate the stability of the comparison results. The subsamples with 881, 505, 368, and 578 movies were correspondingly divided into 30, 34, 37, and 29 subsets with 30, 15, 10, and 30 movies, respectively. When each of these subsets is entered as a validation sample in sequence, the remaining movies are entered as the training sample. Using each training sample, machine learning methods predict the target variables of the corresponding sample. Thirty

training and validation sample pairs are devised with 852 and 30 movies, respectively, to compare the prediction performance at week 1 between high and low review helpfulness. Here, 29 training and validation sample pairs are created with 549 and 29 movies, respectively, to compare the prediction performance at week 1 between high and low reviewer helpfulness. The prediction error is computed using the classification error. The predictions from the learned BI models were compared with the true values in the validation sample to produce the prediction errors,

and the prediction errors were then averaged across a number of validation samples.

The results in Tables 9 and 10 show that the average prediction errors are significantly lower in the high review subsample than in the low review or reviewer sample, except when kNN and discriminant analysis methods are used at week 1. For the reviewer helpfulness subsamples, the prediction error of the high reviewer helpfulness subsample is lower than that of the lower reviewer helpfulness subsample, except when the random forest, decision trees using boosting, and kNN methods are used at week 3. Even for insignificant differences in revenue at week 1 or 3, at least either the review or reviewer helpfulness subsamples show a difference in the prediction error for that specific week. This shows that the results using machine learning methods confirm the results of the multiple regression analysis, i.e., that eWOM with greater helpfulness of reviews and reviewers has greater explanatory power with regard to box office revenue.

Conclusions and implications

This study attempted to use the helpfulness of reviews and reviewers as moderators to predict box office revenue levels while applying machine learning methods (random forest, decision trees using boosting, the *k*-nearest neighbor methods, and discriminant analysis) to compare the prediction power of eWOM with regard to box office revenue between high and low reviews or reviewer helpfulness subsamples. The results of applying the machine learning methods show that movies having more helpful reviews or that are reviewed by more helpful reviewers show greater prediction performance outcomes than movies with less helpful reviews and reviewers.

Implication to researchers. Studies of the influence of eWOM on movie performance across high and low review or reviewer helpfulness levels are necessary, and this study provided insight by showing whether these eWOM variables have explanatory power when movie samples are divided into high and low reviews or reviewer helpfulness subsamples. Review volume has a consistent effect on revenue across all subsamples. It was found that eWOM variables with a greater extent of the helpfulness of reviews and reviewers provide greater explanatory power with regard to box office revenue. Review valence and review extremity levels have a partial influence on box office revenue, and the results show that for experience goods such as movies, less extreme reviews may exert a greater effect on movie performance than extreme reviews. Review and reviewer helpfulness are important moderating factors that improve the explanatory power of eWOM for box office revenue.

The study provides future research issues for researchers of eWOM in the movie industry. First, testing additional semantic eWOM variables such as the number of concepts in relation to reviews, writing styles or the proportion of negative reviews may improve our understanding of the influence of eWOM on movie performance. Second, the interaction effect between eWOM and helpfulness can be tested further in different contexts, such as in international markets. Third, the study model can also be applied to other products to validate the influence of the helpfulness of reviews and reviewers on the prediction power of eWOM. Fourth, additional experiments with other parameter settings (such as 10:90 or 50:50 ratios with regard to box office discretization) can be included to elaborate if there is sensitivity for specific parameter selections. Fifth, to improve the predictive power of the model, additional movie-related variables can be considered, such as MPAA ratings (or age restrictions), the number of screens, the number of screening days, and the production budget, among others. Finally, in order to evaluate the prediction

performance of the model, future studies can consider the use of various performance measures of prediction, such as the root mean squared residual in addition to the classification accuracy.

Implication to practitioners. Our study provides insights to those who practice forecasting box office revenue levels in that they compare prediction performance outcomes using the eWOM characteristics of volume, valence, and review extremity between high and low review or reviewer helpfulness subsamples using machine learning methods to examine the helpfulness of reviews and reviewers as a moderator influencing the prediction performance of box office revenue. Due to the volatility of movie box office revenue, it is necessary to use a sampling strategy that considers the helpfulness of reviews or reviewers to promote the forecasting accuracy of box office revenue. The prediction performance will improve if we create samples with greater values of the helpfulness of reviews or reviewers. Thus, in order to increase the prediction performance, the sample should be extracted such that it provides a higher level of review helpfulness.

This study provides insight into eWOM in international movie markets given its use of a sample of movies distributed in the Korean movie market. Given the relatively few studies of international movie markets other than the US in the box office prediction literature, this study extends the literature on eWOM as it pertains to movies by using an international market in Asia. The significant effect of the review volume confirms that the review volume remains an important eWOM variable for box office revenue in the Korean market.

Data availability

The datasets underpinning this study are available upon request from authors.

Received: 2 February 2020; Accepted: 20 August 2020;

Published online: 07 September 2020

References

- Chen C-H, Nguyen B, Klaus PP, Wu M-S (2015) Exploring electronic word-of-mouth (eWOM) in the consumer purchase decision-making process: the case of online holidays—evidence from United Kingdom (UK) consumers. *J Travel Tourism Market* 32(8):953–970
- Cheung CMK, Lee MKO, Rabjohn N (2008) The impact of electronic word-of-mouth: the adoption of online opinions in the online customer communities. *Internet Res* 18(3):229–247
- Chintagunta PK, Gopinath S, Venkataraman S (2010) The effects of online user reviews on movie box office performance: accounting for sequential rollout and aggregation across local markets. *Market Sci* 29(5):944–957
- Delen D, Sharda R, Kumar P (2007) Movie forecast Guru: a Web-based DSS for Hollywood managers. *Decision Support Syst* 43:1151–1170
- Du J, Xu H, Huang X (2014) Box office prediction based on microblog. *Expert Syst Appl* 41:1680–1689
- Duan W, Gu B, Whinston AB (2008) The dynamics of online word-of-mouth and product sales—an empirical investigation of the movie industry. *J Retail* 84(2):233–242
- Elliott C, Simmons R (2008) Determinants of UK box office success: the impact of quality signals. *Rev Ind Organ* 33:93–111
- Elwalda A, Lü K, Ali M (2016) Perceived derived attributes of online customer reviews. *Comput Hum Behav* 56:306–319
- Gretzel U, Yoo KH, Purifoy M (2007) Online travel review study: role and impact of online travel reviews. Laboratory for Intelligent Systems in Tourism, Texas A&M University. www.tripadvisor.com/pdfs/OnlineTravelReviewReport.pdf. Accessed 20 Aug 2012
- Hong H, Xua D, AlanWang G, Fan W (2017) Understanding the determinants of online review helpfulness: a meta-analytic investigation. *Decis Support Syst* 102:1–11
- Hu N, Koh NS, Reddy SK (2014) Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision Support Syst* 57:42–53

- Internetrend.co.kr (2020) Logger and Internet trend. BizSpring Inc. <http://www.internetrend.co.kr/trendForward.tsp>
- Kim T, Hong J, Kang P (2015) Box office forecasting using machine learning algorithms based on SNS data. *Int J Forecast* 31:364–390
- Lee K, Lee B, Oh O (2015) Thumbs up, sales up? The contingent effect of Facebook Likes on sales performance in social commerce. *J Manag Inf Syst* 32(4):109–143
- Lee KJ, Chang W (2009) Bayesian belief network for box-office performance: a case study on Korean movies. *Expert Syst Appl* 36(1):280–291
- Lee S, Choeh JY (2016) The determinants of helpfulness of online reviews. *Behav Inf Technol* 35(10):853–863
- Lee S, Choeh JY (2018) The interactive impact of online word-of-mouth and review helpfulness on box office revenue. *Manag Decision* 56(4):849–866
- Li M, Huang L, Tan C-H, Wei K-K (2013) Helpfulness of online product reviews as seen by consumers: Source and content features. *Int J Electron Commer* 17(4):101–136
- Liang LJ, Choi HSC, Joppe M (2018) Understanding repurchase intention of Airbnb consumers: perceived authenticity, electronic word-of-mouth, and price sensitivity. *J Travel Tourism Market* 35(1):73–89
- Liang TP, Ho YT, Li YW, Turban E (2011) What drives social commerce: the role of social support and relationship quality. *Int J Electron Commer* 16:69–90
- Liu H, Farhad H, Chew TL, Dash M (2002) Discretization: an enabling technique. *Data Min Knowl Discov* 6:393–423
- Liu T, Ding X, Chen Y, Chen H, Guo M (2016) Predicting movie Box-office revenues by exploiting large-scale social media content. *Multimedia Tools Appl* 75:1509–1528
- McKenzie J (2008) Bayesian information transmission and stable distributions: motion picture revenues at the Australian box office. *Econ Rec* 84(266):338–353
- Mudambi SM, Schuff D (2010) What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Q* 34(1):185–200
- Papathanassis A, Knolle F (2011) Exploring the adoption and processing of online holiday reviews: a grounded theory approach. *Tourism Manag* 32(2):215–224
- Rui H, Liu Y, Whinston A (2013) Whose and what chatter matters? The effect of tweets on movie sales. *Decision Support Syst* 55:863–870
- Shmueli G, Bruce PC, Patel NR (2016) Data mining for business analytics: concepts, techniques, and applications with XLMiner, 3rd edn. John Wiley & Sons, Inc., Hoboken
- Siering M, Muntermann J, Rajagopalan B (2018) Explaining and predicting online review helpfulness: the role of content and reviewer-related signals. *Decis Support Syst* 108:1–12
- Wu J (2017) Review popularity and review helpfulness: a model for user review effectiveness. *Decis Support Syst* 97:92–103
- Yu X, Liu Y, Huang JX, An A (2012) Mining online reviews for predicting sales performance: a case study in the movie domain. *IEEE Trans Knowl Data Eng* 24(4):720–734
- Zhang L, Luo J, Yang S (2009) Forecasting box office revenue of movies with BP neural network. *Expert Syst Appl* 36(3):6580–6587
- Zhou S, Guo B (2017) The order effect on online review helpfulness: a social influence perspective. *Decis Support Syst* 93:77–87
- Zhu F, Zhang X (2010) Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics. *J Market* 74:133–148

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.L.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020