Received 3 April 2017
Revised 19 June 2017
10 July 2017
12 August 2017
Accepted 20 November 2017

Predicting motion picture box office performance using temporal tweet patterns

Najafi Hossein
Department of Computer Science and Information Systems, University of Wisconsin – River Falls, River Falls, Wisconsin, USA, and
Darryl W. Miller
Department of Management and Marketing, University of Wisconsin – River Falls, River Falls, Wisconsin, USA

Abstract

Purpose – The purpose of this paper is to investigate temporal tweet patterns and their effectiveness in predicting the financial performance of a movie. Specifically, how tweet patterns are formed prior to and after a movie's release and their usefulness in predicting a movie's success is explored.
Design/methodology/approach – Volume was measured and sentiment analysis was performed on a sample of Tweets posted four days before and after the release of 86 movies. The temporal pattern of tweeting for financially successful movies was compared with those that were financial disappointments. Using temporal tweet patterns, a number of machine learning models were developed and their predictive performance was compared.
Findings – Results show that the temporal patterns of tweet volume, length and sentiment differ between "hits" and "busts" in the days surrounding their releases. Compared with "busts" the tweet pattern for "hits" reveal higher volume, shorter length, and more favourable sentiment. Discriminant patterns in social media features occur days in advance of a movie's release and can be used to develop models for predicting a movie's success.
Originality/value – Analysis of temporal tweet patterns and their usefulness in predicting box office returns is the main contribution of this research. Results of this research could lead to development of analytical tools allowing motion picture studios to accurately predict and possibly influence the opening night box-office receipts prior to the release of the movie. Also, the specific temporal tweet patterns presented by this work may be applied to problems in other areas of research.

Keywords Forecasting, Social media, Natural language processing, Sentiment analysis, Machine learning, Box office

Paper type Research paper

Introduction

A large body of literature (Zeithaml *et al.*, 1993; Money *et al.*, 1998; Godes and Mayzlin, 2004; Cheung and Thadani, 2012; points to the influence of Word-Of-Mouth communications (WOM) in influencing consumer purchase intention. Traditionally, WOM has been very difficult to observe and measure (East *et al.*, 2008). Marketing researchers have used experiments and surveys in trying to assess the impact of WOM on consumer behaviour and a product's popularity and potential for success in the marketplace (e.g. Bearden and Teel, 1983; Christiansen and Tax, 2000; Harrison-Walker, 2012). However, the former method lacks external validity and the latter is subject to response bias. Today, however, researchers and practitioners are leveraging the large volumes of readily available social media data and advanced techniques in data analytics (Lamberton and Stephen, 2016 present a detailed review). Other research points to how this user-generated Electronic-Word-Of-Mouth (eWOM) from online communities can drive sales (Brown, *et al.*, 2007; Stephen and Galak, 2012; Goh *et al.*, 2013; Jansen *et al.*, 2009; Rui *et al.*, 2013; Chen *et al.*, 2015).

One of the most popular social media channels is Twitter, the microblogging social network through which one can read and send short text messages of up to 140 characters,

International Journal of Intelligent Computing and Cybernetics
Vol. 11 No. 1, 2018
pp. 64-80

1756-378X
DOI 10.1108/IJICC-04-2017-0033

known as “tweets”. Tweets composed by users can be viewed by any number of their “followers”. With its 288 million users generating 500 million tweets a day (Twitter, 2015), Twitter provides a rich bank of soft data. Users of Twitter convey their opinions, feelings, views and general thoughts about many topics including products and services. These tweets are available for download via a continuous data stream through Twitter’s application programming interfaces (APIs). As a result, Twitter has gained popularity among researchers and practitioners as a viable data source for forecasting and predicting many events. A fundamental hypothesis is that the information content of the tweets combined with their volume, location and timing can be used to predict the likelihood and magnitude of future events. Specifically, the use of data from social media channels, such as Twitter, as substitute for traditional WOM communications sought for business intelligence is a promising area of research. Tweets about products and companies may offer a real-time measure not only of consumer awareness but perhaps purchase intention and rate of brand adoption as well.

The annual revenue generated by the motion picture industry was over \$11 billion in 2016 (comScore, 2017). Revenue per movie, however, varies significantly from one film to another. Given the high production and marketing costs along with the high variance in the success rate of movies, forecasting box office revenue as a means to mitigate risk has gained popularity in recent years. Risk mitigation pertains to both strategic and tactical decisions in motion picture production, distribution and exhibition (Eliashberg *et al.*, 2006; Spann *et al.*, 2009; Sawhney and Eliashberg, 1996). For example, on the basis of reasonably accurate forecasts, studio executives could adjust their marketing communications strategies by increasing advertising to support pictures forecasted to have a smaller than desired turnout. Alternatively, they might divert marketing resources from a film that forecasts clearly indicate will be a “turkey” toward those films that forecasts indicate to be more promising. Theatre complex managers might benefit from short-term opening night forecasts in determining theatre capacity, human resource and concession supply decisions necessary to service demand.

It seems possible that analysis of social media volume and content may provide some predictive intelligence for decision making. Numerous studies have shown that tweets can be used to predict such diverse large-scale events as market movements (Bollen *et al.*, 2011; Kumar, 2014; Loughlin and Harnisch, 2014; Si *et al.*, 2014), earthquakes (Sakaki *et al.*, 2010), election results (Barbera and Rivero, 2015; Bermingham and Smeaton, 2011; DiGrazia, *et al.*, 2013; Shi *et al.* 2012), infectious disease outbreaks (Culotta, 2010; St Louis and Zorlu, 2012; Paul and Dredze, 2011), and national revolutions (Howard *et al.*, 2011).

Some recent studies have examined the mining of tweets to predict box-office returns (Asur and Huberman, 2010; Dhanawade, *et al.*, 2016; Jain, 2013; Jangid *et al.*, 2016, Mulay *et al.*, 2016; Najafi and Miller, 2015, 2016, 2017). Others (e.g. Oghina *et al.*, 2012) have combined features extracted from multiple social media channels, including Twitter and YouTube, to improve the results. These works generally investigate the revenue generated by a newly released movie as it compares with the volume and content of the tweets about that movie. For example, Asur and Huberman (2010) used 2.89 million tweets referring to 24 different movies released over a period of three months, to investigate whether they could accurately predict the box office revenue generated by each movie in its opening weekend. They demonstrated that the rate at which tweets are created and the rate of positive sentiment over the rate of negative sentiment about a movie outperformed market predictors. Their results show that the rate of tweets per day could explain nearly 80 per cent of the variance in movie revenue prediction. Likewise, Jain (2013) used sentiment analysis of tweets to predict a movie’s financial success. He manually labelled tweets to create a training set, and trained a classifier to sort the tweets into positive, negative, neutral and irrelevant sentiment groups. Based upon the level of each of these, he predicted likely box office success by classifying each movie as "hit", "flop" or "average". Singh *et al.* (2013) demonstrated further improvements with respect to predicting movie success by exploring the use of "Adverb + Verb" and "Adverb + Adjective", and proposed a feature-based heuristic scheme.

Other researchers extended theses results by looking at the influence of other popular social media channels on the box office revenue. Oghina *et al.* (2012) combined statistics extracted from Twitter data with statistics such as views, number of comments, number of favourites and number of likes/dislikes from the trailer clip on YouTube. They showed that the fraction of the number of likes and dislikes on YouTube, combined with textual features from Twitter, led to the best performing model, with strong agreement with the observed ratings and high predictive performance. Oh *et al.* (2017) studied the effect of social media from the Consumer Engagement Behaviour (CEB) perspective. They reported that although CEB on Facebook and YouTube positively correlated with box-office gross revenue, that same effect was not observed on Twitter. Baek *et al.* (2017) classified popular social media channels (Twitter, Yahoo!Movies, YouTube and blogs) from the perspective of their impact based on actual eWOM data. They demonstrated that Twitter is the most influential channel in the initial stage of a movie's opening, contributing its predictive power to high immediacy (strong awareness through its push mode) and high diffusibility (rapid spread via retweets). Also, they demonstrated that the influence of online reviews is lower in the initial stage of a movie's opening.

Najafi and Miller (2015) compared a traditional consumer survey method of predicting the opening-night box-office returns with a tweet-based approach involving analysis of tweet content (sentiment) and volume. They showed, for tweets posted within a short-term before release, the social media approach performs similarly to the traditional survey techniques. In a follow-up study, they showed that this social media analysis significantly predicted long-term success of a movie and was as good as the traditional survey-based (Najafi and Miller, 2016). In a later study Najafi and Miller (2017) confirmed that tweet volume was a better predictor than was tweet content. These studies dealt with behavioural intention formation and/or tweeting behaviour measured for a short period only prior to the opening night release of each film. Divakaran *et al.* (2017) extended the previous studies by investigating pre-launch prediction of a movie's success as opposed to its performance based on data after its release. They demonstrated that levels of variables, such as awareness, word-of-moth, expectations, and adoption intention, in online communities for an upcoming movie have an independent direct effect on the movie's success.

Previous studies have used a snapshot and/or an aggregated value of the tweet parameters and have ignored the tweet patterns that develop over a longer time period. This research investigates the relationship between movie box office success and the pattern of tweeting from prior to and after a movie's release date. It extends the prior work by using these temporal tweet patterns for predicting movie box office success.

Method

Sample of motion pictures and their hit/bust classification

A sample of 86 motion pictures scheduled for release between October 2013 and December 2014 was drawn from the listings available at Comingsoon.net and comprised genres designated for wide audiences (drama, comedy, action/adventure, etc.) and wide release in terms of the number of movie theatres. Animated films and those targeted at children were not included in the sample.

Domestic box office receipts along with a movie's corresponding production budget were used to estimate the success of a movie. Following Najafi and Miller (2017) a movie was designated as a "hit", if its net revenue was over 20 Million US dollar, and "bust" otherwise. This resulted in a sample of 36 hits and 50 busts.

Tweet collection

The timing of the collection of tweets for each film was coordinated with the release of each movie. Tweets about each movie were collected for a period of approximately four days prior and four days after its release. This period will be referred to as the sampling period. Each tweet was then cleaned and analysed for its information content as described below. A tweet feature vector was formed to represent the information content of each tweet. The feature vectors produced in this manner for each movie were then used to reveal the eWOM sentiment of the tweeting population about the corresponding movie. Figure 1 provides an overview of this process. Details of the process are described next.

Tweet processing

Sentiment analysis using natural language processing has been explored by researchers at many levels for many different types of documents (Esuli and Sebastiani, 2007; Hatzivassiloglou and McKeown, 1997; Kim, and Hovy, 2004; Pang and Lee, 2008; Yu and Hatzivassiloglou, 2003). However, given the limited length of tweets and their informal and acronym-filled nature, analysing them for sentiment is a challenging problem. Figure 2 provides an example of a tweet retrieved from the Twitter API and its common structures. Note the informal nature of the language and the use of Types of internet slang such as web acronyms and webisms. These examples and other cyberslang were commonly found in the tweet samples. Hashtags are used by users to group tweets on the same topic. For example, users who wished to tweet about the movie "Guardians of the Galaxy" would include the hashtag #GuardiansOfTheGalaxy in their tweets. Anyone searching for tweets related to this movie would be able to find them by using the corresponding hashtag. Hashtags are also used by users to punctuate statements or jokes. For example, "#I #srsly #enjoyed #this #movie" provides an exaggerated feeling of joy as expressed by the user. In this study we flag this practice as HasHashString. Users that come across an interesting tweet might retweet the original tweet. Retweeting is used when a user wishes to submit someone else's tweet. This is signified in the tweet by the keyword "RT". Often when retweeting, users add their own comment to the original tweet. For example, in the tweet "lol, can't wait RT @AvaCam :Guardians of the Galaxy is a must see", the "lol, can't wait" is a comment on the original tweet "Guardians of the Galaxy is a must see". This paper refers to this practice as a retweet comment or RTComment.

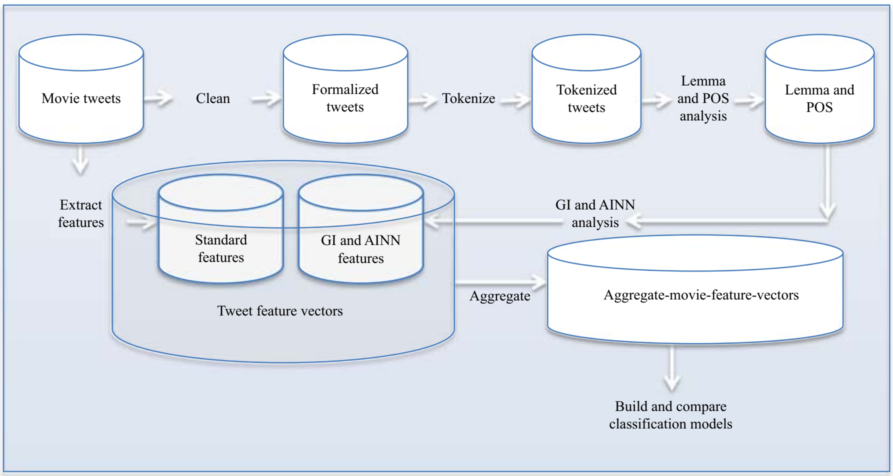

Figure 1. Overview of the movie tweet analysis process

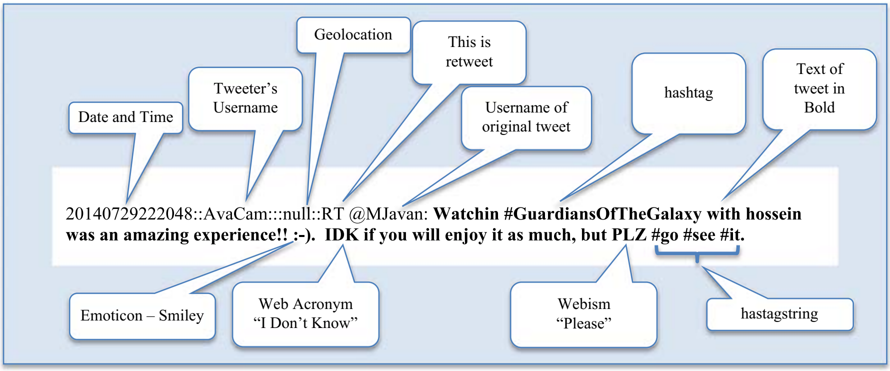

Figure 2.
Example of a tweet retrieved through the twitter API and its common elements

During recent years, many researchers have explored and identified processing techniques that best fit this type of content (Bifet and Frank, 2010; Davidov *et al.*, 2010; Jansen *et al.*, 2009; Kouloumpis *et al.*, 2011; Pak and Paroubek, 2010; Socher *et al.*, 2011). Along the same lines and with the addition of some of our own initiatives, the tweets were processed to extract features with potential predictive power. These will be referred to as the standard features. For example, tweets with more than 50 per cent of their content in uppercase characters were flagged as shouting. Shouting is a way of exaggerating emotion in a tweet. Another way of amplifying statements is to use an unexpected repetition of letters. For example, when a user tweets "I looooooovvvvvvvveeeeeeee u", they are stressing the emotion of love by repeating the letters in the word. These tweets are flagged with the HasRepeat feature. Emoticons, such as smilies are also commonly used to express emotions. This study flags the presence of emoticons in the tweets through a number of features, at different levels of granularity. For example, HasHappyFace, HasSadFace, HasPositiveEmoticon and HasEmoticons are examples of such features. Table I contains the complete list of the standard features and their detail description.

Further tweet content analysis can be accomplished by using a commonly available lexical database of English. However, the informal language commonly used in tweets makes this challenging. For this reason, tweets need to be cleaned before further processing can take place. The main objective of the cleaning process is to transform the informal language used in tweets to a more formal English version. To accomplish this, it is first necessary to identify the informal elements of the tweet and then replace them with formal English. For example, the web acronym "IDK" is identified and replaced with "I don't know", while the webism "pls" is identified and replaced with "please". Through experimentation, other common informal language was identified. These included unexpected repeats (eg "loooovvvvveee" for "love"), foreign words (eg "hola" for "hello"), common misspellings (eg "uuu" for you), web acronyms (eg "lol" for "laughing out loud") and websims (eg "ppl" for "people"). This was done by running 100,000 randomly selected tweets through an English-language spell-checker. The misspelled words generated in this manner were sorted according to their frequency. The high-frequency misspelled words were then manually translated to their corresponding formal English version. This was then used as a dictionary to process and formalise the tweets. Table II summarises the tweet cleaning actions and their descriptions.

The formalized tweets were then tokenised (i.e. each word in the tweet is represented by a token) and processed through MorphAdorner (Socher *et al.*, 2011) for its lemma and part-of-speech (POS) analysis. The following is an example of this process as applied to a tweet:

> Original tweet: I am just the right amount of excited to see Guardians of the Galaxy *Tweet in lemma form:* I be just the right amount of excite to see guardians of the Galaxy *Tweet's POS:* pns11 vbm av-j dt j-jn n1 pp-f vvn p-acp vvi n2 pp-f dt n1

Feature	Description
HasURL	Set to 1 if the tweet contains one or more URLs; reset to 0 otherwise
HasRT	Set to 1 if the tweet is a retweet; reset to 0 otherwise
HasHT	Set to 1 if the tweet contains one or more hashtags; reset to 0 otherwise
HasSelfReflection	Set to 1 if the tweet has self-reflection of author such as "I'm", "I am", "I feel", "make me", etc; reset to 0 otherwise
IsRTComment	Set to 1 if the tweet is a comment about another tweet; reset to 0 otherwise. For example, "lol, can't wait RT @AvaCam :Guardians of the Galaxy is a must see", the "lol, can't wait" is a comment on the original tweet "Guardians of the Galaxy is a must see" from the user AvaCam
HasHashString	Set to 1 if the tweet contains a hashtag string of three or more; reset to 0 otherwise. For example, "#go #see #it" is a hashtag string with a length of three
MostlyUpper	Set to 1 if more than 50 per cent of the tweet is in uppercase letters; reset to 0 otherwise. Use of uppercase in tweets is generally an indication of the user shouting the tweet
HasWebAcronym	Set to 1 if the tweet contains any web acronyms; reset to 0 otherwise. For example web acronyms include "idc" (I don't care) and "jk" (just kidding)
HasWebism	Set to 1 if the tweet contains any webisms; reset to 0 otherwise. For example web acronyms include "pls" (please) and "yolo" (you only live once)
HasReplacement	Set to 1 if any of part of the tweet is replaced; reset to 0 otherwise. For example, webisms and web acronyms are replaced with the corresponding full text during the cleaning process. For example, "idc" is replaced with "I don't care", leading to hasReplacement set to 1
HasRepeats	Set to 1 if an unexpected repeats in English spelling exists in the tweet; resrt to 0 otherwise. For example, in "looooooooooove", the letter o is repeated unexpectedly. For the letter "o" the expectation is that it will not repeat more than twice in an English word. Examples of expected use include "book" or "pop"
HasProfanity	Set to 1 if the tweet contains any common profanity language; reset to 0 otherwise
HasShort	Set to 1 if the tweet contains any shortening; reset to 0 otherwise. Examples include "2day" (today) and "gtta" (have to)
HasHappyFace	Set to 1 if the tweet has happy faces emoticon such as :-) :) :o) :] : > =] =) :} :ↄ) :'-) :') and reset to 0 otherwise
HasSadFace	Set to 1 if the tweet has sad faces emoticon such as :-(:(:ↄ(:-[:[:{ :'-(:'(:(& ._. and reset to 0 otherwise
HasShockedFace	Set to 1 if the tweet has shocked faces emoticon such as o.o :-o - > and reset to 0 otherwise
HasDisappointedFace	Set to 1 if the tweet has disappointed faces emoticon such as :-| :| and reset to 0 otherwise
HasLaughingFace	Set to 1 if the tweet has laughing faces emoticon such as x-D xD X-D XD and reset to 0 otherwise
HasLoveExpression	Set to 1 if the tweet has a love expression such as <3 or xoxoxo and reset to 0 otherwise
HasFrustratedFace	Set to 1 if the tweet has frustrated faces such as >.< and reset to 0 otherwise
HasPositiveEmoticon	Set to 1 if the tweet has positive emoticons; reset to 0 otherwise. Happy faces, laughing faces and love expressions are considered positive
HasNegativeEmoticon	Set to 1 if the tweet has negative emoticons; reset to 0 otherwise. Sad faces, shocked faces, disappointed faces and frustrated faces are considered negative. HasEmotes Set to 1 if the tweet has any emoticons; reset to 0 otherwise
TweetLength	Set to the length of the tweet in number of characters

Source: Najafi and Miller (2015)

Table I. Standard features and their descriptions

Note that each word in the original tweet is converted to its lemma (i.e. its root word) and then further processed for its behaviour in-terms of syntax (i.e. POS). In this tweet, the word "I" has the lemma "I" which is a "1st singular subjective, personal pronoun", while the word "am" has the root "be" which is a "1st singular, "be".

The lemmas and POS from each tweet were then further processed to identify the tweet's sentiment. To accomplish this, the Harvard General Inquirer (GI), a lexicon that maps POS tagged words to their syntactic, semantic and pragmatic information was used. The GI is composed of almost 2,500 words. Each word is combined with its possible POS role and then

Cleaning action	Description
Retweet	Example: Original tweet = Tweet after the cleaning action Remove the RT @username from the tweet RT@hossein:That was a fun movie = That was a fun movie
Retweet comment	Remove the original tweet from the retweet lol, can't wait RT @AvaCam: Guardians of the Galaxy is a must see = lol, can't wait
Hashtags	Remove the # from the hashtags and convert to component words on capital letters #LifeIsGood = life is good
Unexpected repeats	Remove all unexpected repeats I Looooooooovvvvvvveeeeeeeed the movie = I loved the movie
Foreign words	Replace all commonly used foreign words with the corresponding English word hola mama, I love you = hello mother, I love you
Common misspellings	Replace common tweet misspelling with the corresponding English word uuuuu r nice = you are nice
Web acronyms	Replace web acronyms with the corresponding English word JK, see you soon = Just kidding, see you soon
Websims	Replace webisms with the corresponding English word yolo, enjoy it = you only live once, enjoy it
Shortenings	Replace common shortenings with the corresponding long version I h8 Guardians of the Galaxy = I hate Guardians of the Galaxy

Source: Najafi and Miller (2015)

Table II.
Tweet cleaning actions and their descriptions

mapped to 182 possible categories. For example, the word "hate" appears twice, once as a noun and once as adverb/conjunction/particle/preposition and it has the following categories assigned to it in each case:

HATE, Negativ, Ngtv, Hostile, Passive, Arousal, Power cooperation, power, *noun*

HATE, Negativ, Ngtv, Hostile, Passive, Arousal, emotion, sureness, *adverb/conjunction/particle/preposition*

Each GI category was then associated with a corresponding binary GI feature. The lemma and POS of each word in the tweet were checked against the words in each GI category. Where a match was found, the tweet was flagged with the corresponding GI feature. For example, the word "likely" appears in the "Causal" GI category. If the word "likely" appeared anywhere in the tweet, then the "Causal" feature was set for that tweet. In addition, each GI category was also mapped to a binary GI negated feature to track the cases where the expressed sentiment was negative. For example, "not likely" would set the feature "Not-Causal", instead of the feature "Causal".

The GI-Positive, GI-Negative and their negations were used in the following manner to form a GI sentiment score for the tweet:

$$\mathrm{Gi-Based-Sentiment} = (\mathrm{GI-Positive} + \mathrm{GI-Not-Negative})$$

$$-(\mathrm{GI-Negative}–\mathrm{GI-Not-Positive})$$

In addition to the GI Lexicon, AFINN lexicon database (Nielsen, 2011) was used as a secondary measure for the tweet's sentiment. AFINN is a list of English words rated for sentiment with an integer between −5 (negative) and +5 (positive). Each tweet was scanned and the AFFIN scores of its words were averaged to form an AFINN-Based sentiment feature value for the tweet. GI-Based features and AFINN-based feature were then combined to form an Average sentiment feature for the tweet.

The standard-features, GI-Based features and AFINN-based feature were then concatenated to form a tweet feature vector for each tweet. Tweet feature vectors associated with each movie were then aggregated to form an aggregate movie feature vector. Average tweet length was computed as the average of the tweet length. For all other binary features, the aggregate was computed as the fraction of the tweets that had that feature set to a value of 1.

Sequential feature selection was then applied to the aggregated feature vectors to identify the most promising features. "Number of Daily Tweets" and "Average Daily Tweet Length" were identified as the strongest predictors. It is worth citing that strong correlation between these two features and a movie's success was observed and reported by Najafi and Miller (2015). Aggregated HasHT, HasSelfReflection, MostlyUpper, and HasWebAcronym were identified as the next most predictive features and they were combined to form a single feature which we will refer to as "Combined Feature Vector. Average Daily Sentiment" was also considered as predictive.

Number of daily tweets, average daily tweet length, combined feature vector and average daily sentiment were considered as the final features for predictive analytics. These will be referred to below as predictive features.

Classification

Training and validation sets were formed using the daily values of each of the final predictive features. Linear Kernel SVM (L-SVM), Gaussian Kernel SVM (G-SVM), K Nearest Neighbour (KNN), Binary Tree (BT), AdaBoost (AB), and Neural Network (NN) models were used for predictive analysis. Using ten-fold cross validation, performance of the models were measured and compared in terms of their precision, recall and F-measure.

Results

Temporal tweet patterns

Tweet feature values from all 86 movies were combined over the sampling period and the average and variance of each Predictive Feature was computed for each day of the period. Figures 3-6 show the plots of these results. The average number of daily tweets is highest on the release day of a movie independent of its success (see Figure 3). This is inline with

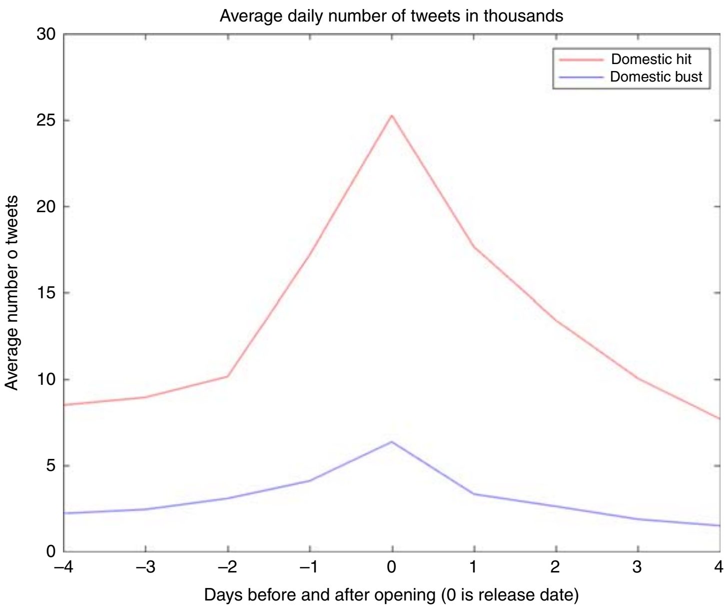

Figure 3.
Average number of daily tweets over the sampling period

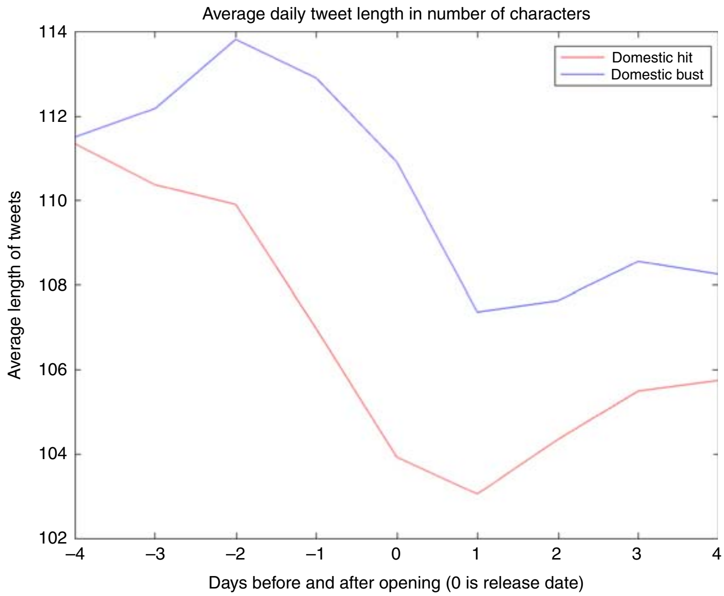

Figure 4.
Average daily tweet length in number of characters over the sampling period

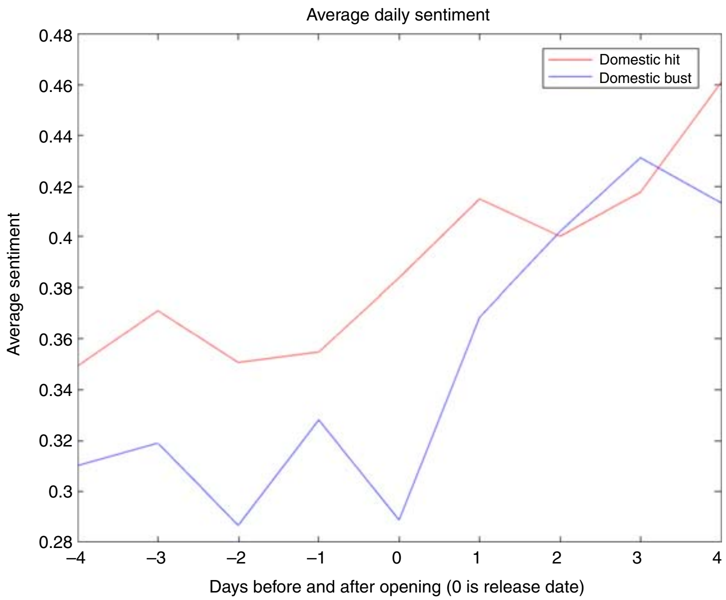

Figure 5.
Average daily sentiment over the sampling period

what was reported by Asur and Huberman (2010). However, it is noteworthy that the average number of daily tweets for busts is significantly lower than that for the hits over the whole sampling period. Base on this, one can conclude that it may be possible to predict opening night success of a movie by observing the number of tweets many days in advance of a movie's release date. It also may suggest that investment by movie

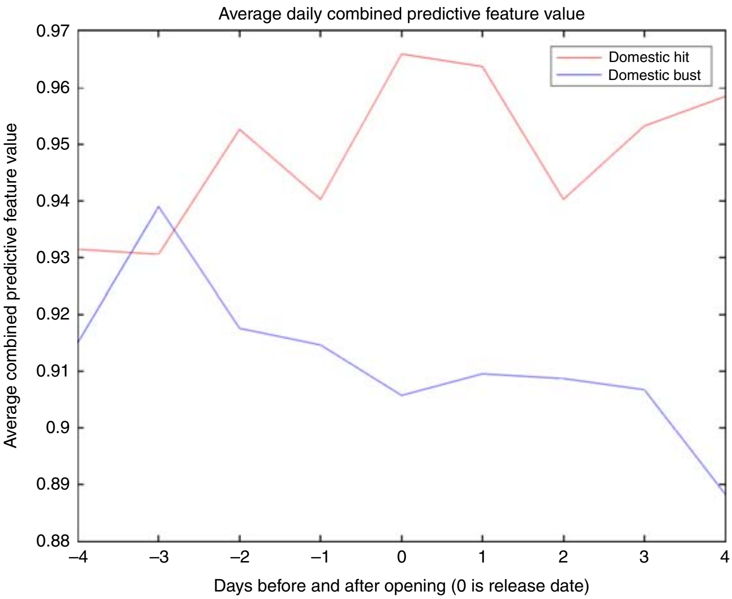

Figure 6. Average daily combined feature value over the sampling period

studios in increasing tweet volumes prior to release of a movie could impact its performance on the opening night.

Figure 4 demonstrates that the average daily tweet length for busts was consistently higher than that for the hits over the whole sampling period. This is consistent with Najafi and Miller (2015) who reported a significant inverse correlation between average tweet length and domestic net revenue suggesting that moviegoers who disliked a film generated longer tweets signalling dissatisfaction compared with the length of those communicating satisfaction. Such a result is consistent with research into dissatisfaction behaviour within the marketing literature (Oliver *et al.*, 1997).

The average daily sentiment of all 86 movies based on their hits and busts categorisation is presented in Figure 5. Here higher sentiment values represent more positive sentiment. Note that the sentiment of hits generally remain more positive than that for busts. Furthermore, the hits sentiment remains more positive than the busts Sentiment during the full period prior to and including the release date of movies. However, after the release date, the two sentiments tend to increase and converge. While previous research (i.e. Asur and Huberman, 2010) points to how polarity of tweets changes after the release of a move, correlation between a movie's success and overall sentiment of its tweets is reported to be insignificant (Najafi and Miller, 2015). The temporal pattern of the tweet sentiments may suggest otherwise.

The combined predictive feature values of the hits and busts over the sampling period is shown in Figure 6. As noted above, this feature combines Aggregated HasHT, HasSelfReflection, MostlyUpper, and HasWebAcronym into one feature. Note that the hits have higher values than the busts during the whole period and the gap between them increases with time. Although, we don't have a good explanation for this, we recognise its predictive potential and will include it for model development.

Hits and busts classification

The features described above were used to train a number of classifiers for predicting a movie's success (hit or bust). Performance of the models was compared using precision, recall and F-measure. Precision, recall and F-measures were computed using ten-fold cross validation averaged over ten trials.

Figures 7- 9 provide a summary of the performance of the six classifiers. Note that for a given day, the performance represents how well the model did in predicting a movie's success (i.e. hit or bust) by being trained on data from that day and all previous days in the sampling period. For example, results shown for day −1, indicate that feature values from days −4, −3, −2, and −1 were used to train the model.

All but one of the classifiers (Gaussian SVM) had precision, recall and F-Measures in the rage of 60-80 per cent. The Linear SVM was best with respect to precision and F-measure;

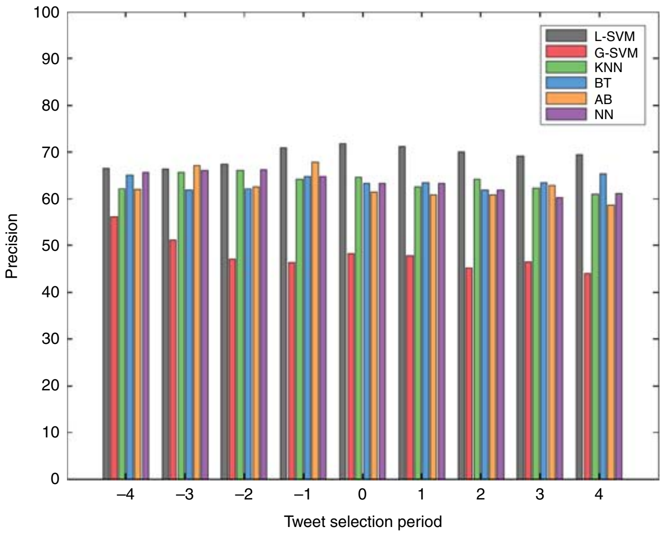

Figure 7.
Precision of classification models over the sampling period

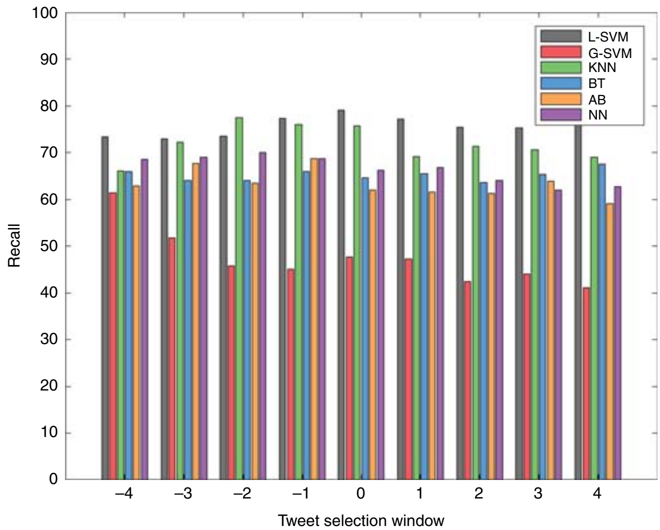

Figure 8.
Recall of classification models over the sampling period

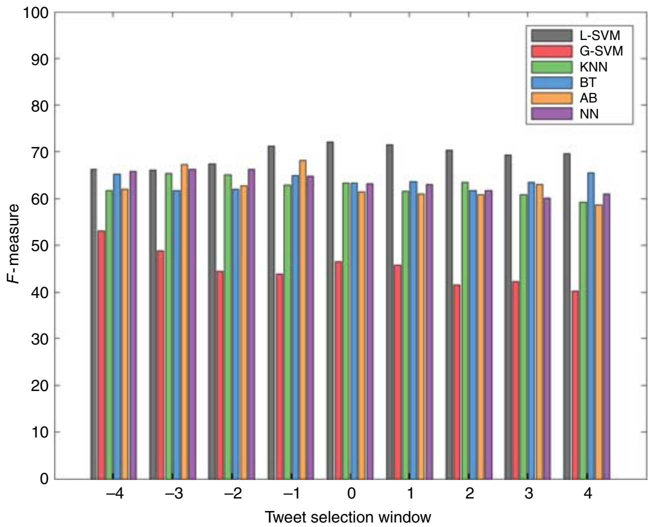

Figure 9.
F-Measure of classification models over the sampling period

while KNN did better or as well as the Linear SVM's performance on Recall. Looking at the Linear SVM's F-measure, note that the model's predictive performance improves with time until the release date and it flattens out after that. Remarkably even four days prior to the movies' release date, the F-Measure is at around 67 per cent. Figure 10 provides a summary of models' performance based on feature values from the release date and all days prior to that. In this case the Linear SVM correctly classified 64 out of 86 movies correctly. Specifically 17 of the 36 hits, and 47 of the 50 busts were correctly classified.

Consistent with Asur and Huberman (2010), Jain (2013) and Najafi and Miller (2017), the results presented here demonstrate that tweet rate and sentiment are associated with movie box office performance. Similar to Najafi and Miller (2017) and Divakaran *et al.* (2017)

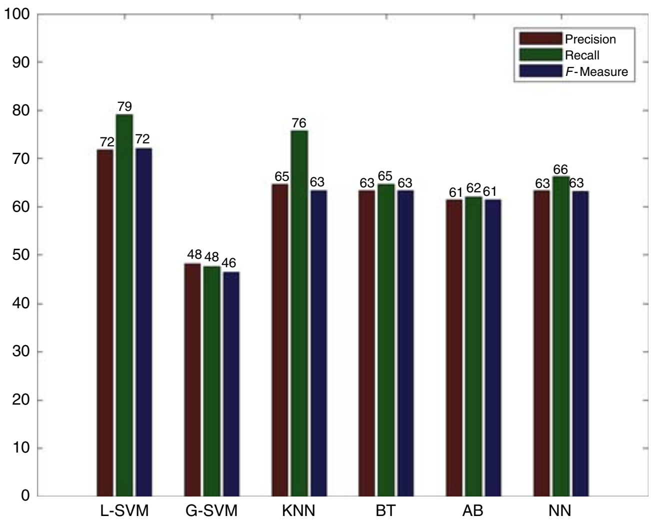

Figure 10.
Performance of the classification models using training data form release date and all dates prior to that

the method presented here utilises features representative of awareness, eWOM, expectation and behavioural intention. The main contribution of this study is the use of pre-release temporal tweet patterns to predict release date and post-release date movie box office success, using a wide range of machine learning models. In an effort to compare the results of the proposed research with exiting methods, we looked at how short-term temporal tweet patterns impact prediction performance. For the purpose of this comparison, we focused on the best performing classifier (i.e. Linear SVM) using F-measure. Figure 11 shows the results of this experiment. The "Temporal" model is trained on a vector of data from days in the sampling period, while the "Non-Temporal" model, is only train on data from that day. Note that the proposed temporal-based approach has superior overall performance as compared to the existing methods.

Conclusion

It would be of great advantage to motion picture studios to accurately predict opening night box office receipts prior to the release of the movie. More important to motion picture executives would be the ability to influence a movie's success. The results of this study show that use of temporal patterns of social media content improves performance of existing methods and can be used to develop predictive models that could effectively forecast success of a movie days in advance of its release.

Results also demonstrate that discriminant patterns in social media feature such as number of daily tweets, average tweet length and tweet sentiments develop days in advance of a movie's release. Movies studios can engineer measures to include these patterns and potentially influence a movie's success. This is significant, since analysis and manipulation of social media can be done more quickly and at lower costs than traditional methods. In addition, once an approach is designed and validated, it could be implemented quickly for future movies and at a small marginal cost.

Methods designed to gauge sentiments, especially the likelihood of behaviours, from the analysis of social media content are in their infancy. Also, mechanisms to influence volume and quality of social media content are in their early years. However, it is highly likely that

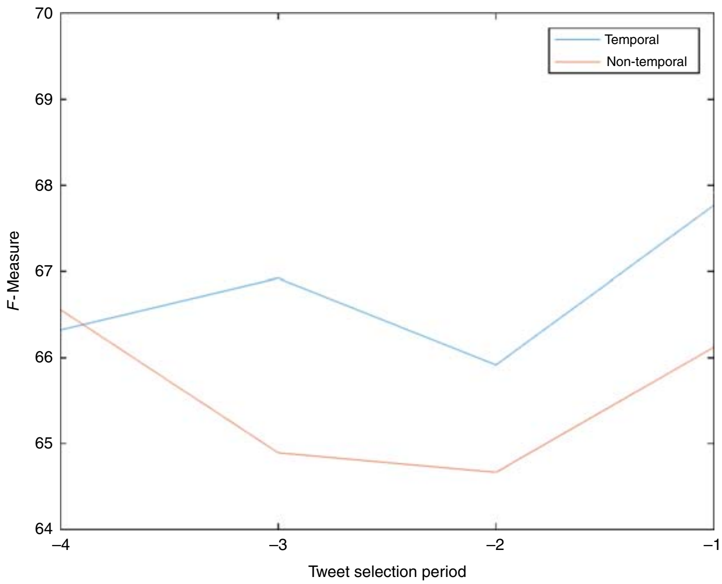

Figure 11.
Performance of the Linear SVM models on temporal and non-temporal data

these methods will one day provide increasingly precise predictions of consumer behaviour and its related financial effects. These methods could be applied to forecasting demand for other entertainment services such as sporting events, concerts, live theatre, streaming media, and mobile apps. Companies are beginning to use these techniques to track satisfaction. It seems reasonable that they could use them to anticipate the need for voluntary recalls. Also, many companies have launched social media campaigns as part of their marketing communications campaigns but with mixed success (Holt, 2016). As these analytical techniques improve it seems conceivable that they could completely replace traditional methods for reaching many marketing research objectives.

References

Asur, S. and Huberman, B.A. (2010), "Predicting the future with social media", *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on IEEE*, Vol. 1, August, pp. 492-499.

Baek, H., Oh, S., Yang, H.D. and Ahn, J. (2017), "Electronic word-of-mouth, box office revenue and social media", *Electronic Commerce Research and Applications*, Vol. 22, March-April, pp. 13-23.

Barberá, P. and Rivero, G. (2015), "Understanding the political representativeness of Twitter users", *Social Science Computer Review*, Vol. 33 No. 6, pp. 712-729.

Bearden, W.O. and Teel, J.E. (1983), "Selected determinants of consumer satisfaction and complaint reports", *Journal of Marketing Research*, Vol. 20 No. 1, pp. 21-28.

Bermingham, A. and Smeaton, A.F. (2011), "On using Twitter to monitor political sentiment and predict election results", paper presented at Sentiment Analysis where AI meets Psychology (SAAIP) Workshop at the International Joint Conference for Natural Language Processing (IJCNLP), Chiang Mai.

Bifet, A. and Frank, E. (2010), "Sentiment knowledge discovery in twitter streaming data", *International Conference on Discovery Science*, Springer, Berlin and Heidelberg, pp. 1-15.

Bollen, J., Mao, H. and Zeng, X. (2011), "Twitter mood predicts the stock market", *Journal of Computational Science*, Vol. 2 No. 1, pp. 1-8.

Brown, J., Broderick, A.J. and Lee, N. (2007), "Word of mouth communication within online communities: conceptualizing the online social network", *Journal of Interactive Marketing*, Vol. 21 No. 3, pp. 2-20.

Chen, H., De, P. and Hu, Y.J. (2015), "IT-enabled broadcasting in social media: an empirical study of artists' activities and music sales", *Information Systems Research*, Vol. 26 No. 3, pp. 513-531.

Cheung, C.M. and Thadani, D.R. (2012), "The impact of electronic word-of-mouth communication: a literature analysis and integrative model", *Decision Support Systems*, Vol. 54 No. 1, pp. 461-470.

Christiansen, T. and Tax, S.S. (2000), "Measuring word of mouth: the questions of who and when?", *Journal of Marketing Communications*, Vol. 6 No. 3, pp. 185-199, available at: www.comingsoon.net/movies

comScore (2017), "Box Office Essentials", available at: www.comscore.com/Insights/Blog/Record-2016-at-North-American-Box-Office-Proves-Continued-Enthusiasm-for-Movie-Going-Experience (accessed February, 2016).

Culotta, A. (2010), "Towards detecting influenza epidemics by analyzing Twitter messages", *Proceedings of the first Workshop on Social Media Analytics*, ACM, July, pp. 115-122.

Davidov, D., Tsur, O. and Rappoport, A. (2010), "Enhanced sentiment learning using twitter hashtags and smileys", *Proceedings of the 23rd International Conference on Computational Linguistics: Posters, Association for Computational Linguistics*, pp. 241-249.

Dhanawade, P., Patil, R., Chougule, R. and Adwant, K. (2016), "Prediction of movies box office performance using social media", *International Engineering Research Journal*, Vol. 2 No. 6, pp. 2026-2029.

DiGrazia, J., McKelvey, K., Bollen, J. and Rojas, F. (2013), "More tweets, more votes: social media as a quantitative indicator of political behavior", *PloS One*, Vol. 8 No. 11, p. e79449.

Divakaran, P.K.P., Palmer, A., Søndergaard, H.A. and Matkovskyy, R. (2017), "Pre-launch prediction of market performance for short lifecycle products using online community data", *Journal of Interactive Marketing*, Vol. 38, pp. 12-28.

East, R., Hammond, K. and Lomax, W. (2008), "Measuring the impact of positive and negative word of mouth on brand purchase probability", *International Journal of Research in Marketing*, Vol. 25 No. 3, pp. 215-224.

Eliashberg, J., Elberse, A. and Leenders, M.A. (2006), "The motion picture industry: critical issues in practice, current research, and new research directions", *Marketing Science*, Vol. 25 No. 6, pp. 638-661.

Esuli, A. and Sebastiani, F. (2007), "Sentiwordnet: a high-coverage lexical resource for opinion mining", *Evaluation*, pp. 1-26.

Godes, D. and Mayzlin, D. (2004), "Using online conversations to study word-of-mouth communication", *Marketing Science*, Vol. 23 No. 4, pp. 545-560.

Goh, K.Y., Heng, C.S. and Lin, Z. (2013), "Social media brand community and consumer behavior: quantifying the relative impact of user-and marketer-generated content", *Information Systems Research*, Vol. 24 No. 1, pp. 88-107.

Harrison-Walker, L.J. (2012), "The measurement of word-of-mouth communication and an investigation of service quality and customer commitment as potential antecedents", *Journal of Service Research*, Vol. 4 No. 1, pp. 70-65.

Hatzivassiloglou, V. and McKeown, K.R. (1997), "Predicting the semantic orientation of adjectives", *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, July*, pp. 174-181.

Holt, D. (2016), "Branding in the age of social media", *Harvard Business Review*, Vol. 94 No. 3, pp. 41-45, 50.

Howard, P.N., Duffy, A., Freelon, D., Hussain, M.M., Mari, W. and Maziad, M. (2011), "Opening closed regimes: what was the role of social media during the Arab spring?", available at: https://ssrn.com/abstract=2595096 or http://dx.doi.org/10.2139/ssrn.2595096

Jain, V. (2013), "Prediction of movie success using sentiment analysis of tweets", *The International Journal of Soft Computing and Software Engineering*, Vol. 3 No. 3, pp. 308-313.

Jangid, B.M., Jadhav, C.K., Dhokate, S.M., Jadhav, G.M. and Bhandari, G.M. (2016), "Survey on movies popularity prediction system using social media feature", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 4 No. 9.

Jansen, B.J., Zhang, M., Sobel, K. and Chowdury, A. (2009), "Twitter power: tweets as electronic word of mouth", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 11, pp. 2169-2188.

Kim, S.M. and Hovy, E. (2004), "Determining the sentiment of opinions", *Proceedings of the 20th International Conference on Computational Linguistics. Association for Computational Linguistics, August*, p. 1367.

Kouloumpis, E., Wilson, T. and Moore, J.D. (2011), "Twitter sentiment analysis: the good the bad and the omg!", *LCWSM*, Vol. 11 Nos 538/541, p. 164.

Kumar, S. (2014), "Stock market prediction using twitter mood", *International Journal of Scientific & Engineering Research*, Vol. 5 No. 5, pp. 44-47.

Lamberton, C. and Stephen, A.T. (2016), "A thematic exploration of digital, social media, and mobile marketing: research evolution from 2000 to 2015 and an agenda for future inquiry", *Journal of Marketing*, Vol. 80 No. 6, pp. 146-172.

Loughlin, C. and Harnisch, E. (2014), "The viability of StockTwits and google trends to predict the stock market", available at: http://stocktwits. com/research/Viability-of-StockTwits-and-Google-Trends-Loughlin_Harnisch.pdf (accessed February, 2017).

Money, R.B., Gilly, M.C. and Graham, J.L. (1998), "Explorations of national culture and word-of-mouth referral behavior in the purchase of industrial services in the United States and Japan", *The Journal of Marketing*, Vol. 62 No. 2, pp. 76-87.

Mulay, S.A., Joshi, S.J., Shaha, M.R., Vibhute, H.V. and Panaskar, M.P. (2016), "Sentiment analysis and opinion mining with social networking for predicting box office collection of movie", *International Journal of Emerging Research in Management & Technology*, Vol. 5 No. 1, pp. 74-79.

Najafi, H. and Miller, D.W. (2015), "Comparing analysis of social media content with traditional survey methods of predicting opening night box-office revenues for motion pictures", *Journal of Digital & Social Media Marketing*, Vol. 3 No. 3, pp. 262-278.

Najafi, H.L. and Miller, D.W. (2016), "Predicting long-term box office performance of movies using social media and its comparison with traditional survey method", *International Journal of Research and Reviews in Applied Sciences*, Vol. 26 No. 3, pp. 120-135.

Najafi, H.L. and Miller, D.W. (2017), "Assessing alternative methods for predicting box office success using tweets", *Journal of Advancements in Applied Business Research*, Vol. 6 No. 1, pp. 26-38.

Nielsen, F.A. (2011), "A new ANEW: evaluation of a word list for sentiment analysis in microblogs", *Proceedings of the ESWC2011 Workshop on "Making Sense of Microposts": Big things come in small packages, in Heraklion, Crete*, pp. 93-98.

Oghina, A., Breuss, M., Tsagkias, M. and de Rijke, M. (2012), "Predicting imdb movie ratings using social media", *European Conference on Information Retrieval. Springer, Berlin and Heidelberg*, pp. 503-507.

Oh, C., Roumani, Y., Nwankpa, J.K. and Hu, H.F. (2017), "Beyond likes and tweets: consumer engagement behavior and movie box office in social media", *Information & Management*, Vol. 54 No. 1, pp. 25-37.

Oliver, R.L., Rust, R.T. and Varki, S. (1997), "Customer delight: foundations, findings, and managerial insight", *Journal of Retailing*, Vol. 73 No. 3, pp. 311-336.

Pak, A. and Paroubek, P. (2010), "Twitter as a corpus for sentiment analysis and opinion mining", *LREC*, Vol. 10 No. 2010, pp. 1320-1326.

Pang, B. and Lee, L. (2008), "Opinion mining and sentiment analysis", *Foundations and Trends® in Information Retrieval*, Vol. 2 Nos 1/2, pp. 1-135.

Paul, M.J. and Dredze, M. (2011), "You are what you tweet: analyzing twitter for public health", *Icwsm*, Vol. 20, pp. 265-272.

Rui, H., Liu, Y. and Whinston, A. (2013), "Whose and what chatter matters? The effect of tweets on movie sales", *Decision Support Systems*, Vol. 55 No. 4, pp. 863-870.

Sakaki, T., Okazaki, M. and Matsuo, Y. (2010), "Earthquake shakes twitter users: real-time event detection by social sensors", *Proceedings of the 19th International Conference on World Wide Web, ACM, April*, pp. 851-860.

Sawhney, M.S. and Eliashberg, J. (1996), "A parsimonious model for forecasting gross box-office revenues of motion pictures", *Marketing Science*, Vol. 15 No. 2, pp. 113-131.

Shi, L., Agarwal, N., Agrawal, A., Garg, R. and Spoelstra, J. (2012), "Predicting US primary elections with Twitter", available at: http://snap. stanford. edu/social2012/papers/shi. pdf (accessed January, 2017).

Si, J., Mukherjee, A., Liu, B., Pan, S.J., Li, Q. and Li, H. (2014), "Exploiting social relations and sentiment for stock prediction", *EMNLP*, Vol. 14, pp. 1139-1145.

Singh, V.K., Piryani, R., Uddin, A. and Waila, P. (2013), "Sentiment analysis of movie reviews: a new feature-based heuristic for aspect-level sentiment classification", *2013 IEEE International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*, pp. 712-717.

Socher, R., Pennington, J., Huang, E.H., Ng, A.Y. and Manning, C.D. (2011), "Semi-supervised recursive autoencoders for predicting sentiment distributions", *Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics*, pp. 151-161.

Spann, M., Ernst, H., Skiera, B. and Soll, J.H. (2009), "Identification of lead users for consumer products via virtual stock markets", *Journal of Product Innovation Management*, Vol. 26 No. 3, pp. 322-335.

St Louis, C. and Zorlu, G. (2012), "Can twitter predict disease outbreaks?", *BMJ*, Vol. 344, p. e2353.

Stephen, A.T. and Galak, J. (2012), "The effects of traditional and social earned media on sales: a study of a microlending marketplace", *Journal of Marketing Research*, Vol. 49 No. 5, pp. 624-639.

Twitter (2015), "Twitter usage: company facts", available at: https://about.Twitter.com/company (accessed January 2015).

Yu, H. and Hatzivassiloglou, V. (2003), "Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences", *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, July*, pp. 129-136.

Zeithaml, V.A., Berry, L.L. and Parasuraman, A. (1993), "The nature and determinants of customer expectations of service", *Journal of the Academy of Marketing Science*, Vol. 21 No. 1, pp. 1-12.

Further reading

Howard, P.N. (2010), *The Digital Origins of Dictatorship and Democracy: Information Technology and Political Islam*, Oxford University Press.

About the authors

Najafi Hossein is a Professor of Computer Science at the University of Wisconsin – River Falls. He received his PhD in Electrical Engineering from the University of Minnesota. His research interests include machine learning, neural networks, fuzzy logic and their application to real-world problems. His recent research has been published in *Journal of Digital & Social Media Marketing*, the *International Journal of Intelligent Computing and Cybernetics*, *Journal of Applied Intelligence*, and *Journal of the Academy of Finance*. Najafi Hossein is the corresponding author and can be contacted at: hossein.najafi@uwrf.edu

Darryl W. Miller is a Professor of Marketing at the College of Business and Economics, University of Wisconsin – River Falls. He received his PhD in Business Administration from Kent State University. His research interests include advertising, branding strategy, and services marketing. His recent research has appeared in *Journal of Marketing Communications*, *Services Marketing Quarterly*, *Journal of Historical Research in Marketing*, and *Journal of Digital & Social Media Marketing*.