Decoding the Sentiment Dynamics of Online Retailing Customers: Time Series Analysis of Social Media

Noor Farizah Ibrahim[1]* and Xiaojun Wang[2]

1. School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia
2. School of Economics, Finance and Management, University of Bristol, Bristol, UK

*Corresponding author's email: noor.farizah@gmail.com

Abstract

The Twittersphere often offers valuable information about current events. However, despite the enormous quantity of tweets regarding online retailing, we know little about customers' perceptions regarding the products and services offered by online retail brands. Therefore, this study focuses on analysing brand-related tweets associated with five leading UK online retailers during the most important sales period of the year, covering Black Friday, Christmas and the New Year's sales events. We explore trends in customer tweets by utilising a combination of data analytics approaches including time series analysis, sentiment analysis and topic modelling to analyse the trends of tweet volume and sentiment and to understand the reasons underlying changes in sentiment. Through the sentiment and time series analyses, we identify several critical time points that lead to significant deviations in sentiment trends. We then use a topic modelling approach to examine the tweets in the period leading up to and following these critical moments to understand what exactly drives these changes in sentiment. The study provides a deeper understanding of online retailing customer behaviour and derives significant managerial insights that are useful for improving online retailing service provision.

Keywords: Online retailing; service provision; time series; social media; big data analytics

1 Introduction

Social media is no longer a passing fad for online retail brands. Users post millions of messages expressing their opinions regarding online retail brands on social media platforms every day, and this has gradually altered both how online retail brands communicate with customers and how they work. The question is, how can online retail brands keep up with events in the world of Twitter? One way to do so is to ensure that they establish an online presence to gain access to the information available on Twitter. As of December 2015, 35 million posts had been posted by brands across Facebook, Twitter and Instagram, prompting a total of 65 billion customer engagements (Yuki 2015) – that is, 65 billion instances of online retail brands reaching out to customers through their corporate Twitter accounts, Facebook pages, and official Instagram accounts. Through these channels, online retail brands can learn what drives business outcomes and discover what strategies work best for them (Hewett, Rand, Rust, and van Heerde 2016). Exploring trending brand-related content has become essential for businesses to better understand their customers and helps generate more advanced exposure for them to gain marketplace intelligence, new business deals and partnerships. In fact, the timely analysis of customer messages on these platforms can help a brand to keep abreast of customers' views and opinions of them. In past years, brands have dutifully followed the trend of using social media platforms to engage and communicate with customers. One of the great opportunities for firms lies in the fact that customer opinions on Twitter can convey meaningful messages, which is valuable to businesses as a way to create good relationships with customers.

Despite the rapid shift from traditional to social media, companies still face difficulties and challenges in understanding customers' needs and concerns thoroughly. As a substantial number of customers join social media every day, online retail brands should be more actively learning about their customers' behaviour through analysing their messages on social media and pay attention to those areas that affect customers' perceptions of products or services. Despite the mass of data available on social media, companies still need to accurately analyse customer opinions expressed on social media (Ghiassi, Zimbra, and Lee 2017). A major challenge is to understand customers' messages in a timely manner in order to offer appropriate and quick responses and management (Bendle and Wang 2016;

Ibrahim, Wang, and Bourne 2017). In addition, an overwhelmingly large volume of data (Netzer, Feldman, Goldenberg, and Fresko 2012) and a lack of practical tools with which to analyse unstructured data (Archak, Ghose, and Ipeirotis 2011) also complicate such analysis. Since Twitter content is produced abundantly every day, changes in customer sentiment and the relationship between those changes and emerging issues on Twitter are difficult to capture. The root cause of such changes may lie in a range of factors, including external events, accidents and disasters, and users' daily lives.

Therefore, to understand what exactly drives changes in customer sentiment, this study attempts to exploit the content of these real-time data to understand trends in customer tweets. Thus, the research questions are as follows:

- What are the volume and sentiment trends of online retailing brand–related tweets on Twitter?
- What are the main reasons behind the changes in the volume and sentiment trends of online retail brand–related tweets?
- How can online retailers learn from social media users to improve their online retailing service provision?

To answer these questions, we analyse changes in customer sentiment and the underlying customer concerns that evolve through the time series. We examine those tweets or comments posted by customers that are directly related to five leading UK retailers namely Amazon, Argos, Asda, John Lewis and Tesco during the period spanning Black Friday, Christmas, Boxing Day and New Year's Day in the United Kingdom. Using a machine learning method, we extract the primary concerns of customers expressed through their tweets to theorise about the causes behind changes in sentiment. This method has been used in previous studies that applied text analytical techniques to obtain richer insights (Daniel, Neves, and Horta 2017; Ghose, Ipeirotis, and Li 2012; Kim, Dwivedi, Zhang, and Jeong 2016; Lee and BradLow 2011). Our analysis incorporates sentiment analysis and topic modelling into time series analysis to gain an understanding of changes in sentiment and the related causal factors. We also offer different event detection techniques to better verify the critical time points identified in the time series.

This study offers several contributions to the online retailing and social media literature. First, through a combination of multiple social media analytics approaches, we look into the underlying reasons behind the change in customer sentiments. Therefore, this study complements to the existing literature by demonstrating an effective approach of providing more accurate analysis in examining customers' opinions on Twitter (Ghiassi et al. 2017; Li, Li, and Zhu 2016; Naaman, Becker, and Gravano 2011; Vergeer and Franses 2015). Second, building from prior studies that emphasized the difficulties in analysing social media data which are unstructured, informal and massive (Gandomi and Haider 2015; McAfee, Brynjolfsson, and Davenport 2012), this study analyses Twitter data in online retailing context and highlights the main online retailing service areas that drive the shift in customer sentiments. This work extends the online retailing literature by identifying priority service areas which companies need to manage, improve, and conduct their responses. Moreover, by examining the dynamics of the topics in the time series, we present a solution for companies to monitor their performance and prioritise which area that requires extra attention from them. By doing this, we make a valuable practical contribution to the online retailing field by providing a guidance for the company on how to improve their online retailing service and pinpoint their focus towards the critical area that urgently need their attention.

The remainder of the paper is organised as follows. After reviewing several related studies making use of time series analysis of sentiments and topics in online service research, the research methodology is outlined. This is followed by the presentation of the analysis and the results thereof. In the last section, we discuss the practical and theoretical implications of this study.

2 Related Work

2.1 Twitter for Online Retail Brands

Social media applications such as Twitter have been acting as platforms on which customers can communicate and share information about brands. These platforms also provide data that can be analysed and facilitate knowledge discovery, including personal daily activities, picture sharing, political opinions, product reviews, and expressions of sentiment (Bian et al. 2016). With the increasing use of social media by the public, its applications in the business context are also increasing as a vast number of companies resort to social media as an economical and efficient way to reach a large number of customers (Kaplan and Haenlein 2010). For instance, Hennig-Thurau et al. (2010) demonstrated that digital media such as Twitter serve as a guide for companies and have affected their understanding of customer behaviour and of how to efficiently manage customer interactions. Furthermore, Hajikhani, Porras, and Melkas (2017) analysed user content in the US smartphone market using crowd evaluation to assess the impact of different profile types on the social media strategies of companies.

2.2 Time Series Analysis

Understanding time series is critical to uncovering the hidden trends and insights in a set of longitudinal data. In its application, time series analysis uses continuous data as an input to identify systematic patterns, overall trends or turning points in a time series data set over a particular time interval. Time series analysis has been effectively applied in a variety of settings, including political events (Lansdall-Welfare, Dzogang, and Cristianini 2016), disasters (Naaman et al. 2011), stock market predictions (Bollen, Mao, and Zeng 2011; Daniel et al. 2017), economic forecasting (Arias, Arratia, and Xuriguera 2013), and sales forecasting (Archak et al. 2011).

In many research areas, time series analysis allows researchers to analyse phenomena that change over time using event detection or spike detection methods. Figueiredo, Almeida, Gonçalves, and Benevenuto (2014) used YouTube videos to analyse how the popularity of videos evolves over time. They applied time series analysis to discover the popularity growth patterns of videos according to the number of times the videos were viewed and then correlated them with referrers characteristics such as viral, external, internal, search, mobile, featured, and social. In another study, Lansdall-Welfare et al. (2016) examined changes in public sentiment in the UK pertaining to the Brexit referendum using multivariate time series analysis. The findings of their study identified three change points in public sentiment in the period leading to the referendum by increases in negative affect. A significant correlation between the GBP/EUR exchange rate and change in public sentiment was detected in the hours following the Brexit vote.

For market prediction and forecasting, Naaman et al. (2011) extracted 50 Twitter trends and divided them into exogenous (e.g., news events and national holidays) and endogenous (e.g., memes, RTs) trend types using both quantitative and qualitative approaches. Archak et al. (2011) evaluated Amazon sales data and product reviews for digital cameras and camcorders using time series analysis. The study revealed key product attributes and the results were used to estimate their respective values and predict future Amazon sales.

2.3 Sentiment and Time Series Analysis of Twitter Data

Despite the unstructured nature of its text, tweets have become an interesting and popular topic of research in social media analysis over the past few years and have been widely studied by companies, marketers and even political analysts (Giachanou and Crestani 2016; Wu, Song, and Huang 2016).

A growing body of literature studies the influence of sentiment on specific time series to establish, for instance, what precipitates changes in sentiment over time. For example, O'Connor, Balasubramanyan, Routledge, and Smith (2010) explored the relationship between public opinion in presidential polls and sentiment expressed on Twitter and discovered a strong correlation between the sentiment scores of Twitter data and poll results over time. In the same vein, Bollen et al. (2011) applied time series analysis to sentiment to explore the impact of real events on user sentiment on Twitter. Based on public tweets posted between August and December 2008, the study demonstrated that changes in public sentiment were associated with social, cultural, political and economic events.

Other studies have focused on the application of sentiment analysis to messages posted by users and have used that sentiment for forecasting or future prediction. Zhang, Fuehres, and Gloor (2011) collected six months' worth of tweets to predict the movement of the Dow Jones, NASDAQ and S&P 500 indices, and found that emotional tweets were negatively correlated with stock market prices. Similarly, Ranco, Aleksovski, Caldarelli, Grcar, and Mozetic (2015) explored the relationship between sentiment and stock price returns using Twitter and NASDAQ data for forecasting and revealed that user sentiment affects stock market returns. Finally, Arias et al. (2013) examined the influence of public sentiment on stock market and movie box office revenues. Using a data set of more than 300,000 tweets, they compared several self-constructed forecasting models and tested their respective predictive power. Their findings revealed that, unlike linear models, nonlinear models have greater predictive capacity over time series.

Most of these studies explored and identified correlations between changes in sentiment and stock market prices, but without explaining the factors causing such sentiment change, and therefore offered limited insight into the factors that may link sentiment to market prices. A closer examination of such factors, particularly those that lead to a decrease in customer sentiment, would prove of greater use to companies and enable them to take the actions necessary to address any underlying issues. Moreover, the reviewed studies focused more on market prediction than identifying the root causes of inadequate service provision that might upset the market. This gap in the existing literature provides the basis for the work undertaken in this study.

2.4 Time Series Analysis and Topic Modelling of Twitter Data

Messages posted by customers on social media have been receiving much attention as brands see them as an important business opportunity. These messages can be processed and turned into meaningful knowledge using various approaches. Among them, topic modelling has been widely applied to discover the hidden themes contained in a large data set. Over the years, topic modelling has drawn the attention of researchers in various academic and industrial research areas (Fan, Zhao, and Xu 2015; Wilkinson and Thelwall 2012).

There has been a significant amount of research on the link between the various topics extracted from Twitter data and time series analysis using various methods such as topic modelling, word frequency and content analysis. In regard to natural disasters, Lampos and Cristianini (2012) inferred levels of rainfall by analysing posts from Twitter to reveal actual levels of rainfall and when it had happened. Dueñas-Fernández, Velásquez, and L'Huillier (2014) carried out an experiment to explore the effectiveness of topics extracted from weblogs incorporating opinion mining, topic modelling and time series analysis. The study detected trends in the themes and modelling over time, and the authors later proposed their own trend-detection methodology. More recently, Robertson and Yee (2016) studied the use of T. witter for surveillance during an influenza outbreak in North America using outbreak detection based on a linear time series model and latent Dirichlet allocation (LDA) topic model for the specific event. The study demonstrated that peak detection is associated with real-world events and identified a number of topics relevant to outbreaks.

Other studies have focused on topic trends to understand how topics on Twitter change over time. One study explored news-related tweets from seven different locations including the UK, the US, Australia and New Zealand spanning a nine-month period (Wilkinson and Thelwall 2012). The study applied time series analysis to establish the differences and similarities between top trending topics between countries. Similarly, Heverin and Zach (2012) analysed the trends and topics of tweets related to three campus shootings in the US using a time series method. They found that opinion sharing increased over time, and the various themes identified in the conversation threads during the crises contributed to a better understanding of information behaviours and sense-making. In the same line of study, Vergeer and Franses (2015) explored the relations among issue salience in televised debates and issue salience in Twitter debates using content and time series analysis. Surprisingly, the influence of issue salience in TV debates on Twitter activities was found to be limited or close to non-existent. Lai et al. (2016) explored topic histograms and the emergence of topics over time to identify topics with an abnormal structure and recovered hidden interactions between the topics. Finally, to understand the influence of traditional media sources on tweets, Kim, Gonzenbach, Vargo, and Kim (2016) analysed how the issues and agendas of three media sources, TV advertisements, newspapers and Twitter, were correlated using autoregressive integrated moving average (ARIMA) time series.

2.5 Event Detection on Social Media

Event detection can be defined as a process of detecting novel events from text corpus (Zhang et al., 2016). Several studies have demonstrated effectively event detection analysis in variety of settings including political events (Lansdall-Welfare et al., 2016), disasters predictions (Naaman, Becker, & Gravano, 2011), epidemic disease outbreaks (Odlum & Yoon, 2015), or market predictions (Daniel et al., 2017). Often, high frequency phenomena, which exhibit uncommon trending and draw people attention, are regarded as events (Kaneko & Yanai, 2016). It normally provides substantial information describing the related scenarios during the events or crises.

The detection of events using news, blogs and recently social media has been widely discussed by researchers. Social media is becoming an important platform to discover events in real time with the increasing rate and volume of data streams. For instance, a comprehensive study of 200 million posts by Fan et al. (2015) investigated topic dynamics on Weibo from the perspectives of time, demographics, emotion, geography, correlation and retweets. The study revealed interesting patterns in and insights into user profiling, event detection, and content recommendation on social media. Meanwhile, Sakaki, Okazaki & Matsuo (2010) conducted an experiment using social media data to detect earthquake activities across Japan. The research findings revealed that by using social media data as social sensor, they were able to detect a high probability of earthquakes at a faster rate compared to many other warning systems. In a recent study, Daniel et al. (2017) evaluated the influence of tweets published in financial community on financial market. In their study, the specific type of event was found to be correlated with future fluctuations in stock markets.

There also appears to be strong interest in understanding and discovering the emerging of an event within diversified data sources and high volume of data. For instance, Kim, Lee and Kyeong (2013) collected 18,720,902 tweets to study the social hot topics by normalising the high frequency words within a longitudinal Twitter stream and revealed the terms related to events such as major holidays appeared as hot topics. In another study, Odlum and Yoon (2015) explored tweets posted during the 2014 West African Ebola virus epidemic to uncover the correlation between tweets and real-world events. Using time-series and content analysis, the study revealed the intersection of social media and public health outbreak surveillance and provided insights that led to recommendations for future precautions. More recently, Pohl, Bouchachia and Hellwagner (2016) studied social media data related to Hurricane Sandy 2012 and revealed that dynamic indexing and online clustering of crisis-related data are needed for emergency management to identify events in real-world emergencies.

Undeniably, there have been studies on event detection and the potential value of the technique, but most of these have focused on different settings such as natural disasters, politics or randomly analysed public tweets rather than focusing on online retailing (Robertson & Yee, 2016; Wilkinson & Thelwall, 2012). Taking this as departure point, this present study aims to to detect significant changes in customer tweet volume and sentiment trends in online retailing context and explain the underlying causes leading to those changes.

2.6 Summary

Table 1 provides a brief overview of some important literature featuring time series analysis of social media in different settings and focus areas. As can be seen from the table, time series analysis and topic modelling have been used in various applications such as the environment, the economy, health, and politics. However, the paucity of such research in the online retailing context must be addressed. Among the studies reviewed above, the work of Dueñas-Fernández et al. (2014), in which topics were extracted from weblogs using sentiment analysis and topic modelling, is most closely related to the present study. However, in the context of this research, we are interested in decoding changes in the sentiment of tweets using the topic modelling approach, and our work differs from the aforementioned studies in several respects.

First, we use social media data to explore and decode the sentiment dynamics of Twitter users regarding online retailing brands. Along with the potential opportunities, it also brings the challenges due to the nature of the social media data that is unstructured, large in volume, informal and noisy (Archak, Ghose, & Ipeirotis, 2011; Netzer, Feldman, Goldenberg, & Fresko, 2012; Stieglitz, Mirbabaie, Ross, & Neuberger, 2018). Social media research is a relatively new research field. A large amount of unstructured data is often beyond what traditional methods can manage (Lansdall-Welfare, Dzogang, & Cristianini, 2016) and a daunting task for many organisations to harvest (Sheng, Amankwah-Amoah, & Wang, 2017). Hence, this demands more research to offer solutions that address these challenges and fulfil the potentials. Second, most of the previous studies on social media research (Kim, Dwivedi, et al., 2016; Li & Liu, 2017) have focused mainly on one or two techniques. By integrating sentiment analysis, topic modelling, and time series analysis in decoding the sentiment dynamics of online retailing customers, this research introduces distinct approach and contributes to the need for more accurate analysis of customer opinions on the social media (Ghiassi et al., 2017). For instance, the use of time series and sentiment analysis does not only detect significant shifts in customer sentiment, but also allow to determine the main reasons for the sentiment changes. We also make use of an event detection technique to determine the time intervals for topic modelling in our experiment. The multi-pronged approach of event-detection techniques utilised in this study provides opportunities to triangulate data to strengthen and validate the detection of significant dates. Hence, this research does not just collect and report the metrics of social media data but gain a better understanding of the data from a semantic point of view by examining the reasons behind dynamic changes in customer sentiment. Unlike any analysis of aggregated data, time series analysis of customer sentiment offers online retailers time specific analysis of the strengths and weaknesses of their services, enabling them to take appropriate action.

Table 1 Overview of Literature on Time Series Analysis of Social Media

Focus	Area	Data	Method	Source
Pattern/ Trend	Environment, Multiple topics	Weibo, Twitter, YouTube, Weblog	Time series, Trend features, Burst, Referrer analysis, NB, SVM, Time series, Spiking terms, Bayesian classifier, NMF, EMD, TS histograms, Opinion mining, Topic modelling, Time series	Li and Liu (2017), Naaman et al. (2011), , Wilkinson and Thelwall (2012), Fan et al. (2015), Lai et al. (2016), Dueñas-Fernández et al. (2014); (Figueiredo et al. 2014)
Event outbreak	Health,	Twitter	Bolasso (PMTK), Regression, Linear time series, LDA topic model, Time series, K-means	Lampos, Bie, and Cristianini (2010), Robertson and Yee (2016), Odlum and Yoon (2015)
Forecasting/ Prediction	Environment, Technology, Economy, Politics	Twitter, Blogs, NASDAQ, DJIA, Exchange rate, Amazon sales, Reviews	Semantic analysis, SVM, Baseline Method, OpinionFinder, GPOMS, Linear models, Time series, Clustering, AMT, Event detection, My Sentiment API, TextBlob, SentiStrength, Affin, MVA, AMT, ITA, C-SPAN metadata, Granger-causal analysis, Classifiers, Word frequency, Correlation, LASSO, Regression, n-gram	Bollen et al. (2011), Arias et al. (2013), Ranco et al. (2015), Lansdall-Welfare et al. (2016), Archak et al. (2011), Daniel et al. (2017), O'Connor et al. (2010), , , Zhang et al. (2011), Lampos and Cristianini (2012)
Relationship	Politics, Environment	Twitter, TV, News	Time series, Word frequency, Content analysis, ARIMA	Vergeer and Franses (2015), Heverin and Zach (2012), Kim, Gonzenbach, et al. (2016)

3 Method

3.1 Model Framework

Figure 1 illustrates model framework for analysing Tweets, which includes data source, method & analysis, and finding. First, data sources are the tweets that mentioned the five leading online retail brands (according their online retailing market share in the UK) covering a crucial sales period. Among these five lead brands, Amazon UK conducts its retailing operation online exclusively whereas other four brands have both online and offline retailing operations. Similar to other social media research, the implementation of method and analysis starts with pre-process of the tweets in order to filter out and remove the noise of the social media data. After cleaning the data source, sentiment analysis is first performed to explore trends in tweet sentiment. It is followed with time series analysis to identify critical time points and changes in sentiment. Topic modelling is then applied to the tweets in the time series where significant sentiment shifts were identified and key topics discussed and shared among the Twitter users are extracted to pinpoint the underlining causes that lead to these sentiment changes. More detail about each stage of the analysis and applied methods are further elaborated in the following sub-sections. The findings from the analysis such as trends in tweet volume and sentiment, critical time points, changes in sentiment, and topic dynamics can be used as reference point of service improvement for the online retailers.

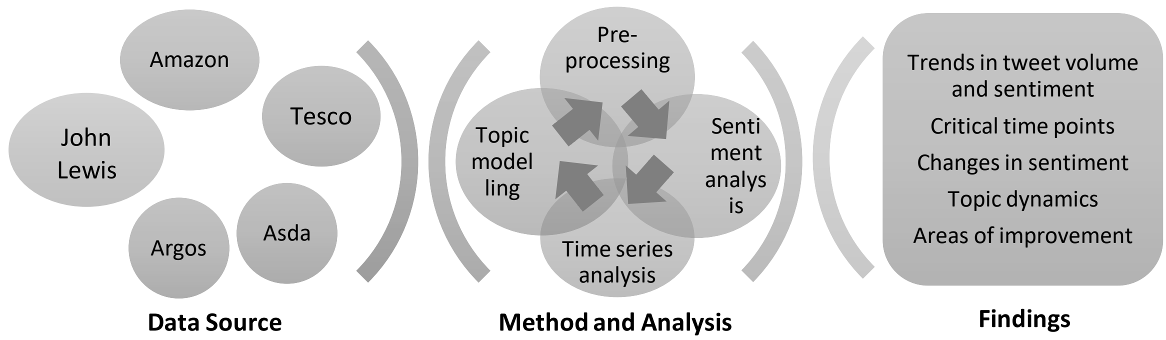

Figure 1 Model Framework for Analysing Tweets

3.2 Data Collection and Pre-processing

In this study, 386,379 tweets referring to five leading online retail brands, namely Amazon, Tesco, John Lewis, Argos and Asda, were extracted using NodeXL Pro. These online retailers are the leading online UK retailers and among the largest retailers – both online and offline – in the UK, as listed on the Top 500 retailers website (News 2017; Retailing 2016). Furthermore, they are considered very active on Twitter and have large Twitter followings, indicating that they are influential. This ensures that our findings will be practical and relevant to other researchers and practitioners. The tweets mentioning these five brands were collected from 20 November 2016 to 20 January 2017, the period covering the Black Friday, Christmas, Boxing Day and New Year's sales events in the UK. The tweets were extracted using a consistent keyword: [brand name] lang:en since:[start date] until:[end date]. During the pre-processing stage, we filtered non-English tweets and cleaned the text by removing punctuation and converting all text to lowercase. We then removed stop words and omitted terms that appeared with low frequency. Finally, we filtered out all meaningless words.

3.3 Analysing Tweet Sentiment

A lexicon-based approach was used to analyse tweet sentiment. Customer tweets were measured and divided into three categories – positive, negative and neutral – using SentiStrength. SentiStrength is a lexicon-based sentiment analysis tool that classifies tweets on an 11-point scale ranging from -5 (negative) to 0 (neutral) and 5 (positive). The SentiStrength algorithm applies linguistic rules involving negations, booster words and emoticons in computing negative and positive outputs (Thelwall, Buckley, and Paltoglou 2012). This method was considered relevant because it was designed for a wide range of texts, including unstructured social media data. The lexicon also contains word lists that measure different dimensions of psychological and behavioural characteristics in the text. In this study, we classified the sentiment strength of each collected tweet as either positive, negative or neutral. Since each classified text returns both a positive and negative score, we then calculated its polarity to obtain a final score. This algorithm has been tested for accuracy and works well with text of the same nature as our Twitter data, that is, short, informal text (Thelwall, Buckley, Paltoglou, Cai, and Kappas 2010).

3.4 Analysing Time Series

The study period was divided into smaller time ranges using a time series analysis of daily Twitter volume. Time series analysis was used to identify the peaks of Twitter activities to understand the underlying pattern at work at a particular time (Ranco et al. 2015). Such analysis detects changes in the situational information in relation to a subject over time using continuous data as input. Time series analysis describes events in real-time and has been introduced in various applications, such as economics, the environment, science and medicine. Using several techniques including frequency method, difference in mean and moving average, we observed whether changes had occurred and identified the times of occurrence of any such changes. We highlighted all abrupt changes in relative volume and sentiment as events to break out into discrete time series.

3.5 Extracting Topics

To extract key points from our data set, we employed the topic modelling approach, a popular machine learning method, to identify emerging topics in a particular sub-period of our study. LDA is a well-known topic extracting tool that facilitates the analysis of large collections of unstructured data. This approach has been demonstrated effectively in previous studies in a variety of settings such as Twitter topic coverage (Kim, Jeong, Kim, Kang, and Song 2016), Wikipedia topic identification (Yildirim, Üsküdarlı, and Özgür 2016) and microblog user discovery (Li, Yan, Weihong, and Ding 2014). We extracted 20 topics for each period using the MALLET LDA toolkit. In this experiment, we confined our extraction to 20 topics because a large number of topics can return unstable results and contain junk topics which are difficult to interpret (Nikolenko, Koltcov, and Koltsova 2016).

Conversely, fewer than 10 topics may show small granularity and be insufficient to create a representative analysis (Lugmayr and Grueblbauer 2016). Moreover, most studies on microblogs have predefined the number of topics in their studies, for example, by using 10 or 20 topics (Goswami and Kumar 2016; Kim, Jeong, et al. 2016). For instance, Lo, Chiong, and Cornforth (2016) argued that 20-topic LDA models perform better than 10 and 30-topic models. On this basis, we selected 20 as the optimal number of topics to be extracted in this study. This is supported by a number of other scholars who decided to evaluate only the best 20 topics in their studies (Ma, Sun, and Cong 2013; Zhang, Wang, Cao, Wang, and Xu 2016; Zhou, Wan, and Xiao 2016). After extracting the topics and keywords from our Twitter corpus, the topics were then manually labelled into meaningful themes based on the combined judgement of the researchers with reference to different sets of successful factors and constructs of online retailing (Collier and Bienstock 2006; Francis 2009; Luo, Ba, and Zhang 2012). We labelled those topics that could best explain the combination of extracted keywords.

4 Analysis and Results

In this section, we present the results of our time series analysis. First, we obtained the volume and sentiment scores of tweets. This was followed by the performance of event detection analysis to detect shifts in volume and sentiment and identify the times of any significant changes. We focused on the critical points at which sharp declines in customer sentiment were apparent in the spike detection graph, and applied topic modelling to tweets showing negative sentiment to understand the underlying issues that led to customers dissatisfaction.

4.1 Trends in Tweet Volume and Sentiment

Customers express their opinions in a variety of ways, ranging from outrage to satisfaction and happiness. To explore these reactions in full, we examined tweets and explored the trends displayed by those tweets from two different perspectives: volume and sentiment. We first observed the emerging patterns in the daily volume of tweets. Figure 2 shows the trends of tweets associated with our five chosen online retail brands during the study period. It shows a pronounced spike on 25 November, a day before Black Friday, suggesting that at this time, customers were most likely actively shopping in anticipation of the festive season. We observed another spike on 12 December, with 10,522 tweets, when one of the brands under examination organised a Twitter contest for its customers. The high volume of participation indicates that using Twitter for marketing is a smart and effective strategy. This also suggests that customers use Twitter to talk and express their opinions about brands, since it is easy to access and the majority of online retail brands nowadays own social media accounts to interact and engage with customers.

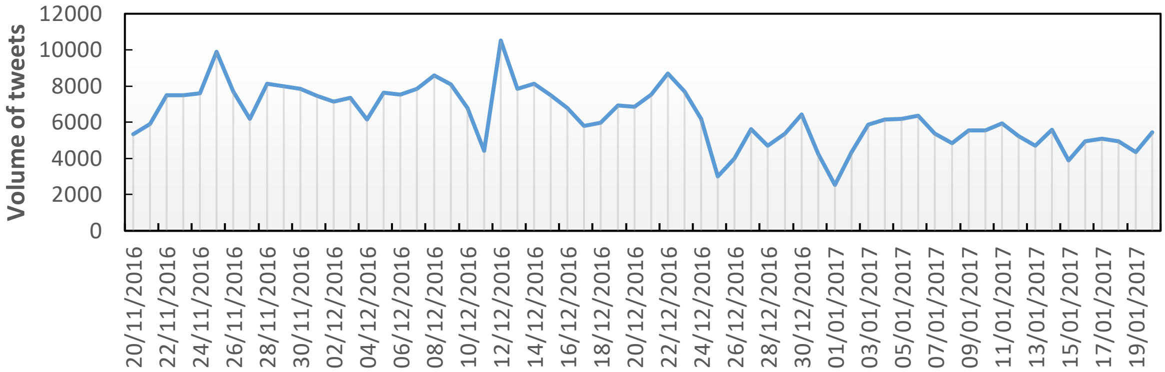

Figure 2 Twitter Trends of the Online Retail Brand's Search Term

SentiStrength was used to extract the overall positive, negative and neutral sentiments from customer tweets in relation to the brands under examination. Figure 3 shows the frequency distribution of each score. The distribution line of positive and negative tweets was seen to have a similar proportion, whilst neutral sentiments remained higher. This indicates that neutral tweets govern the whole data set. This is explained by the value of standard deviation, shown in Figure 4; the small standard deviation shows that the distribution of the sentiments in our study was concentrated and not widely spread out.

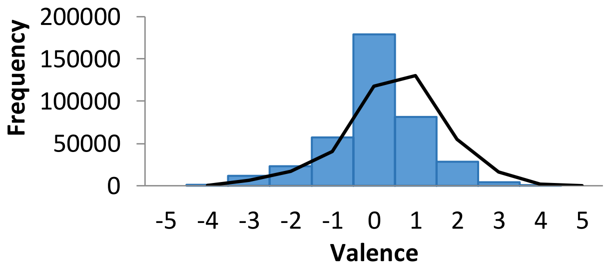

Figure 3 Frequency Distribution of the Overall Tweets

We then narrowed down the focus by observing only the fluctuation of positive and negative sentiment throughout the study period, as shown in Figure 4. Although the volume patterns for both positive and negative tweets are in line with the overall volume trend observed in Figure 3, tweets expressing positive and negative sentiment were found to spike at different points in time. For instance, the highest negative spike occurred on 23 December, a few days before Christmas and Boxing Day, while 9 December showed the most positive sentiment. However, subsequent to these events, the spikes dropped in magnitude, lasting only a couple of days. This is perhaps due to the decrease in the volume of tweets posted, as the holiday and shopping seasons were coming to an end. Moreover, the standard deviation (σ) trendline was found to be consistent throughout the period, whilst the mean (μ) decreased due to a lower number of tweets towards the end of the festive season, when people were busy celebrating the events.

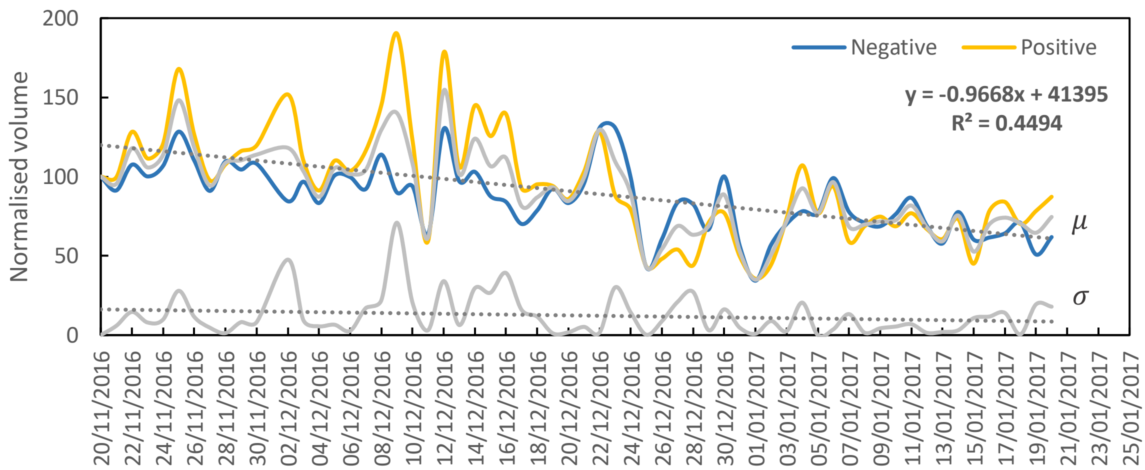

Figure 4 Volume of Negative and Positive Tweets

To help clarify the trend followed by the sentiments in the time series, Figure 5 shows the distribution of the daily average sentiment score. We observed the movement of this score over time, and discerned a number of significant spikes in the graphs, including 2 December (12%), 9 December (13%) and 16 December (11%) for positive sentiments in Figure 5(a), and 20 November (13%), 21 November (11%), 23 December (12%), 28 December (12%), 6 January (11%) and 15 January (11%) for negative sentiments in Figure 5(b). Most of the spikes were detected in the build-up to the sales events, indicating a pre-Christmas shopping frenzy. During this time, retailers tend to offer one-day-only deals on all stock and sometimes extend their discounts during the week. As shown in the graph, the mean score for negative sentiment has a bigger range gap compared to the trend of positive sentiments, which implies that, on some days, customer tweets can range from negative to extremely negative.

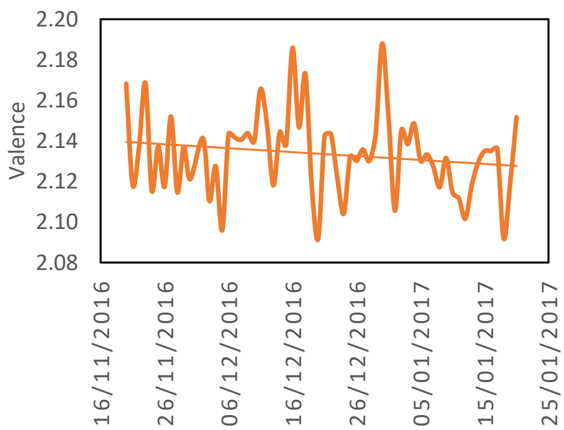

Figure 5(a) Trend of Positive Score

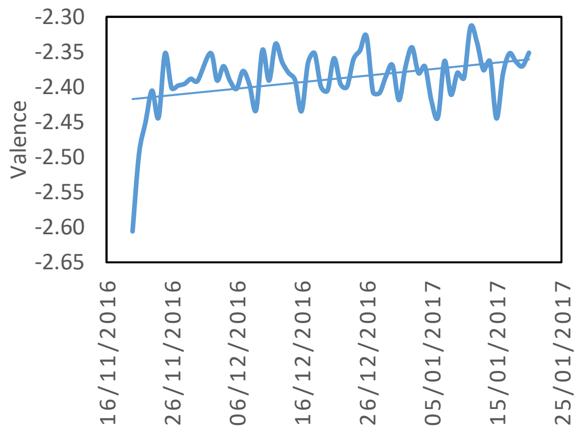

Figure 5(b) Trend of Negative Score

4.2 Event Detection in Longitudinal Time Series Data

In this section, we employed event detection to understand the changes in sentiment patterns and reasons for the occurrence of such changes at particular times. Event or spike detection is used to identify times at which changes occur in the time series. This captures any change in the mean and frequency of correlation of the process. To capture the critical time points, we deployed three approaches: frequency, the difference in mean, and moving average.

4.2.1 The Frequency Method

The frequency method is utilised to detect the magnitude of shifts in frequency. This popular method is in common use by public online search tools such as Google Trends and Google Books Ngram Viewer. Kulkarni, Al-Rfou, Perozzi, and Skiena (2015) argued that changes in frequency are the quickest way to detect sequences of events. As it is also simple to deploy, we began with the frequency method.

Figure 6 illustrates the largest drops and increases in sentiment in the time series. During this stage, the frequency method technique was applied to each sentiment in our time series. First, Figure 6(a) demonstrates the detected time interval on negative sentiment. The change in negative tweet sentiments appearing over time was tracked to capture the dates leading up to and following the change. For instance, major frequency changes in negative sentiment occurred on 12 December, 25 December and 31 December. These changes in sentiment could be attributed to Boxing Day and Christmas, known as the ‘sales season’. For instance, on 25 December, a sharp drop could be observed by the dotted line showing the degree of the peak. Second, Figure 6(b) shows the time series of positive sentiment with the detected time intervals. As can be seen from the graph, the detected changes in frequency occurred on 25 November and between 10 and 13 December. These dates were not found to be significant, as they did not accord with the event dates in the study period, hence the results extracted from later analysis would not have been practical or relevant. Finally, Figure 6(c) shows the combination of positive and negative sentiment frequency. This figure displays a superior result compared to Figure 6(b) in terms of time interval because the detected dates are more scattered. This captured some important dates such as Christmas (25 December) and New Year (31 December), but the remainder of the dates (11 December–13 December) were not significant and did not provide much value.

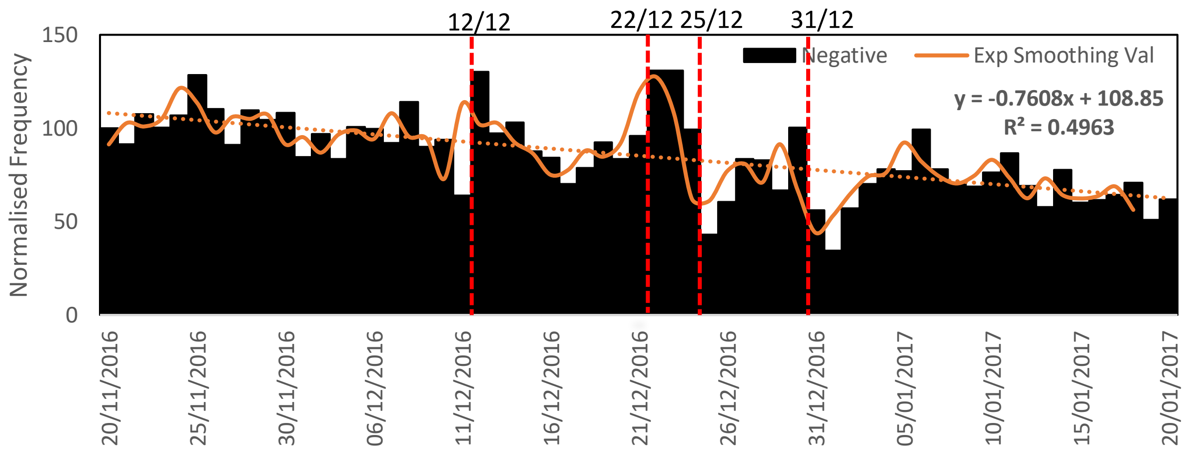

Figure 6(a) Frequency of Negative Sentiment with Time Interval

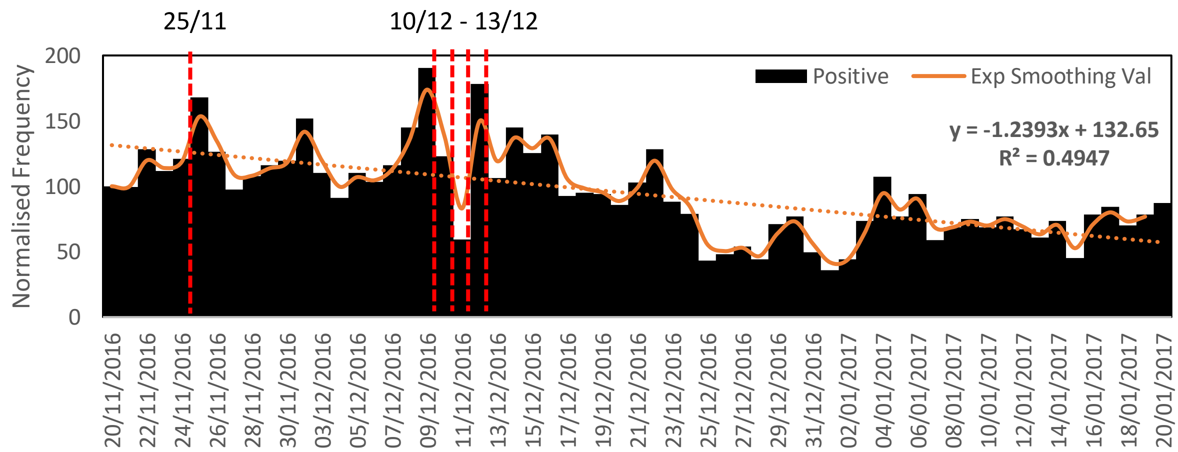

Figure 6(b) Frequency of Positive Sentiment with Time Interval

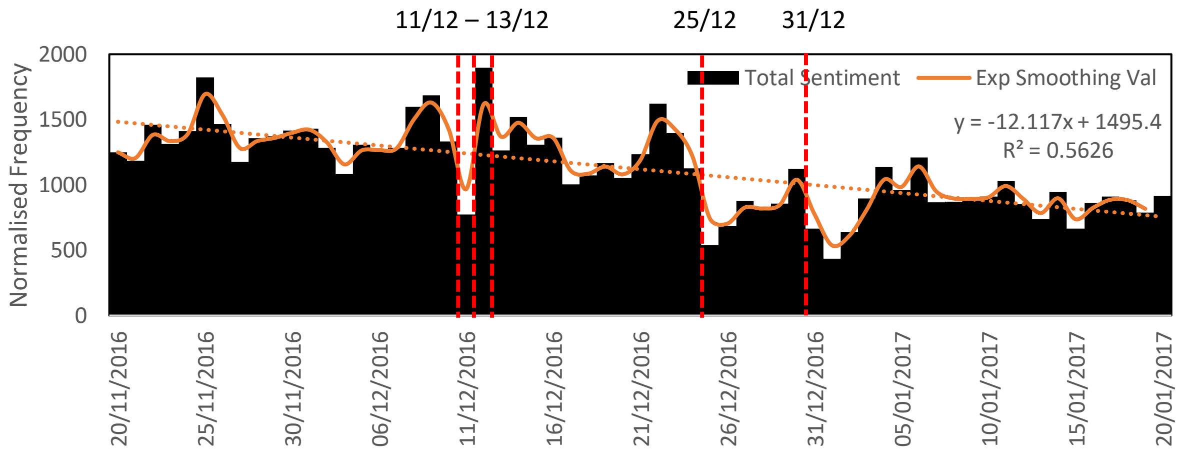

Figure 6(c) Frequency of Positive and Negative Sentiment with Time Interval

Compared to the results from Figures 6(b) and (c), the analysis of the negative-sentiment tweets in Figure 6(a) detected some critical points that coincided with key events over the study period, including Black Friday, Christmas Day, and New Year's Eve. To verify the results, two other methods were deployed for further analysis of the negative-sentiment tweets in the next phase.

4.2.2 The Difference in Mean

The next phase involved establishing the mean difference (Lansdall-Welfare, Lampos, and Cristianini 2012) to identify significant changes in customer sentiment and the associated dates. This involved tracking the mean difference in the data set. Key dates were detected by applying a shift point to the time series to identify sudden changes in the mean. We calculated the mean difference, d, by measuring the absolute difference between the mean value for each day:

$$d = \mu_1 - \mu_2 \qquad [1]$$

Figure 7 shows the daily mean difference, where peaks can be interpreted as days which displayed the most abrupt or largest shifts. As can be seen, there was a large spike in negative sentiment on 25 November (Black Friday), followed by a similar spike between 9 and 10 December. The same pattern also can be seen on 26 and 27 December due to the Boxing Day sales events held by most UK online retailers. The remaining spikes occurred after the New Year, on 7 and 15 January.

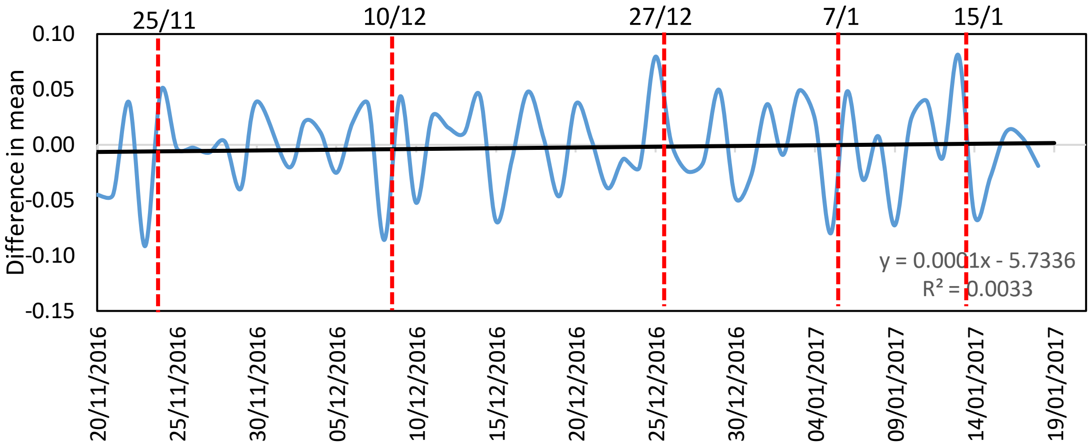

Figure 7 Dates of Significant Negative Sentiment Change

4.2.3 The Moving Average (MA)

To further validate the results, we computed the moving average of the negative sentiment score to identify the change points or the changes in our Twitter corpora. Moving averages are useful when working with real-time data that rapidly rises and falls each day and helps smooth data and filter out the noise (O'Connor et al., 2010). In this study, we smoothed out the sentiment score using a basic exponential approach to identify any rise or fall in sentiment.

$$MA_t = \frac{1}{k}\left(x_{t-k+1} + x_{t-k+2} + \cdots + x_t\right) \qquad [2]$$

Figure 8 shows the moving average (MA4) of negative sentiment using the unsmoothed version under different time windows (k). It is clear that the significant shifts detected in the time series reflect changed customer sentiment. In agreement with the second approach, this approach also identified 25 November, 10 and 27 December, and 7 and 15 January as the key dates on which the largest changes in the mean of the sentiment score occurred.

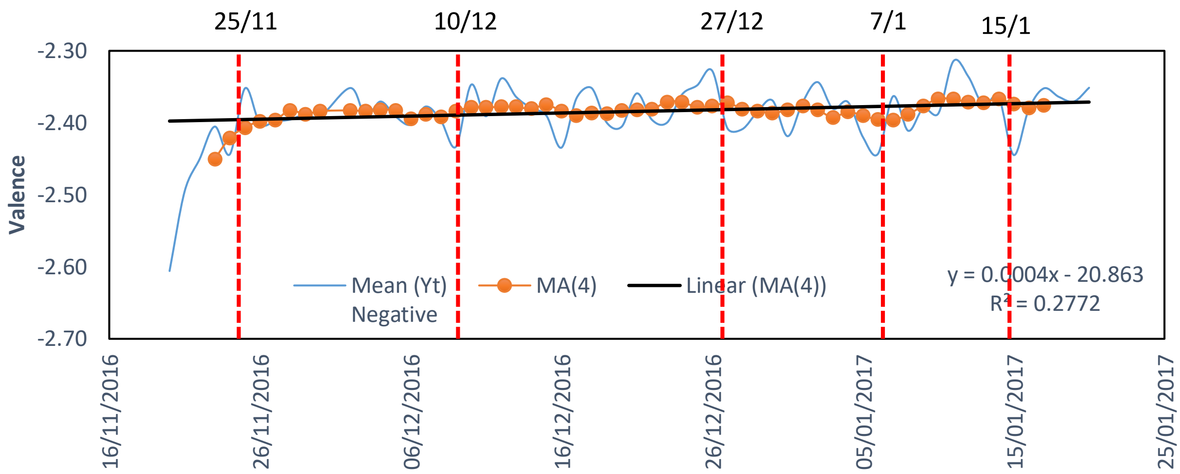

Figure 8 Moving Average of Negative Sentiment

Due to the nature of Twitter data, the application of event detection techniques to Twitter streams has been challenging (Daniel et al. 2017). The analysis above identified several key dates which warranted further investigation. Both the mean difference and moving average methods identified the same key dates that fit and partitioned our time series. To understand customers' perceptions of online retailing service, we focused on these chosen dates to establish the underlying reasons contributing to changes in customer sentiment. For instance, one of segmented dates is 25 November, which coincides with Black Friday sales in the UK. This event appeared to have precipitated lasting changes in customer sentiment. In both the UK or US, Black Friday is becoming one of the most important days in the annual retailing calendar.

As for 25 and 26 December, these dates represent the holiday season in the UK. The Christmas and Boxing Day sales events explains the increase in negative sentiment before the festive week. The results of the spike detection approach demonstrated a build-up of negative tweets from customers in the lead-up to these events. These events precipitated the highest frequency of negative tweets. As the aim of this study is the exploration of those areas in which online retail brands need to improve, we kept our focus on the spikes indicating changes in the frequency of negative tweets.

4.3 Changes in Customer Sentiment

We constructed the time series computed in the previous section to quantify the significance of sentiment changes throughout the time period. Figure 9 presents the patterns of sentiment that changed over time in our data set. As seen, the number of negative sentiments was higher than that of positive sentiments throughout the time series, due to Black Friday, Christmas and Boxing Day. These peaks of negative tweets corresponded with higher sales volumes that most likely led to a greater number of customers later visiting Twitter to complain about their dissatisfaction with their shopping experiences. The influx of complaints was expected as these are the season's biggest shopping days, leading to considerable sales volumes in the UK. Moreover, negative sentiments tend to be more numerous because customers spread the word about bad service if they do not receive quick responses from the companies involved (Grégoire, Salle, and Tripp 2015). Furthermore, we observed a rise in negative sentiment before Christmas and Boxing Day until after these events (P3 to P4) from 9 per cent to 10 per cent, with sentiment dropping by about 1 per cent after the New Year (P4 to P5). While we had not expected negative tweets to decrease after the events, there were several reasons underlying the drop in negative sentiment. Moreover, negative sentiment remained at 9 per cent after Black Friday and rose in the lead-up to Christmas (P2 and P3). Thelwall, Buckley, and Paltoglou (2011) utilised a similar volume change method to investigate the changes in mood associated with changes in common words.

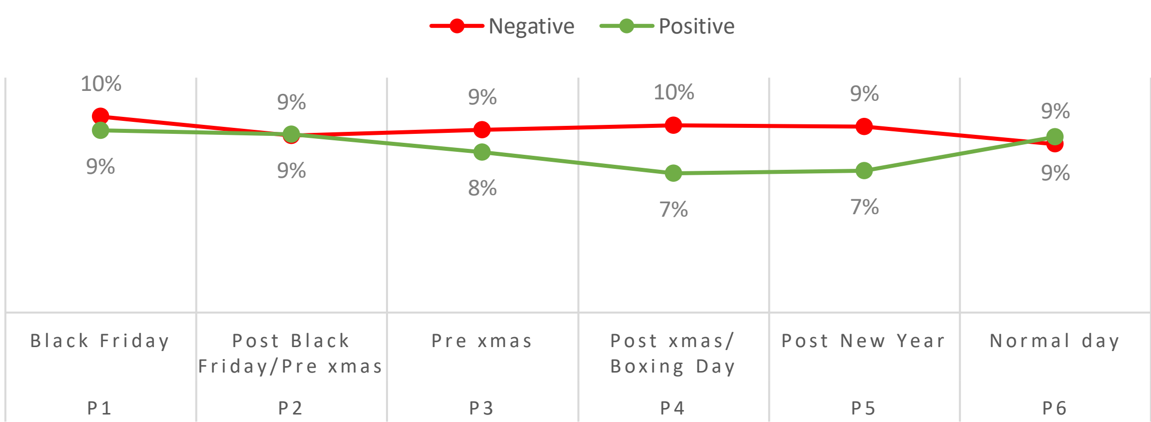

Figure 9 Sentiments Tweets by Events

From the previous analysis in Figure 5(b), we noted that the trends of the negative sentiments changed dynamically, yet this figure does not tell us why. Hence, using this information, we narrowed down the focus by examining trends in a small timeframe leading up to and following the key events to observe the changes in sentiment by using a seven-day moving window. Adopting Lansdall-Welfare et al. (2012) approach, we computed the mean difference in the sentiment score for each day before and after the event. This approach was also used by Giachanou and Crestani (2016), who tracked sentiment over time. Figure 10 shows the volume and mean difference for the two significant sales events, Black Friday and Boxing Day, in the seven-day window frame. This measure showed abrupt peak changes before the events and drastic drops immediately thereafter. There was a corresponding drop in the volume of tweets. Significantly, the volume decrease implies that customers were relatively more active on Twitter in the build-up to the events. Generally, these kinds of events are considered important in the UK and therefore much-discussed among customers. Argos, for example, records its highest sales on Black Friday, when they offer substantial discounts on products such as toys and household appliances. However, the decrease of sentiment indicates low customer satisfaction during this peak sales season, possibly due to bad shopping experiences and poor customer service, or as a result of other factors. Building on this foundation, we employed topic modelling in our next analysis to establish the exact reasons for these changes.

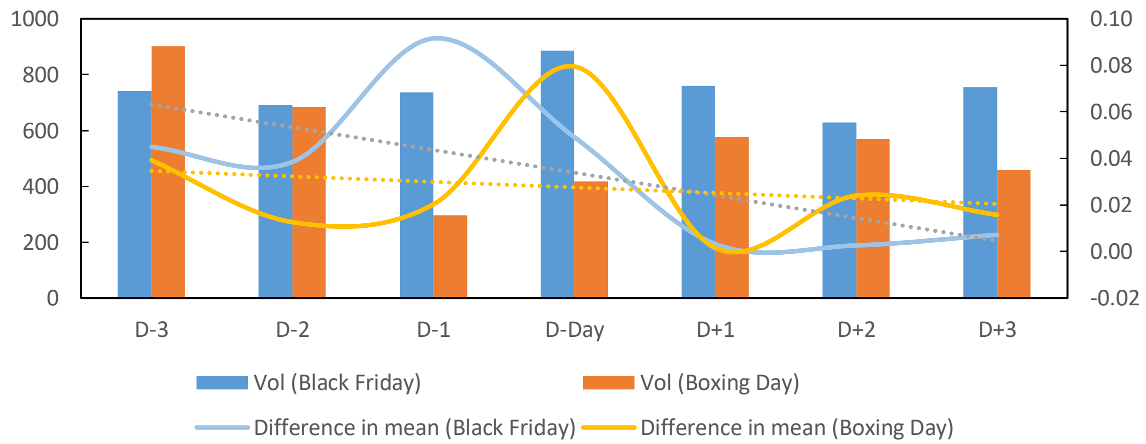

Figure 10 Longitudinal Analysis for Seven-day Moving Window

4.4 Topic Dynamics

The analysis in section 4.2 demonstrated that our methods are able to detect sentiment shifts in customer tweets. To discover the reasons underlying those changes in sentiment and the key topics causing customers concern, we applied the topic modelling approach to each time period leading up to the critical points at which significant shifts in sentiment occurred. As argued by Fan et al. (2015) and Daud (2012), studying the different topics provides us with information on the spectrum of customer interests. Figure 11 reveals the topic modelling results using the data set containing all negative sentiments expressed during the study period. As seen here, the time series was divided into six shorter time periods based on the event detection results. For each period, only the top 10 topics were presented according to their weights, that is, the importance of the extracted topics. To identify the top ten topics, we merged duplicate topics and measured their accumulated weights. We then labelled the topics according to the method set out in section 3.5. Through topic modelling, we aimed to decode the sentiment dynamics of online retailing customers.

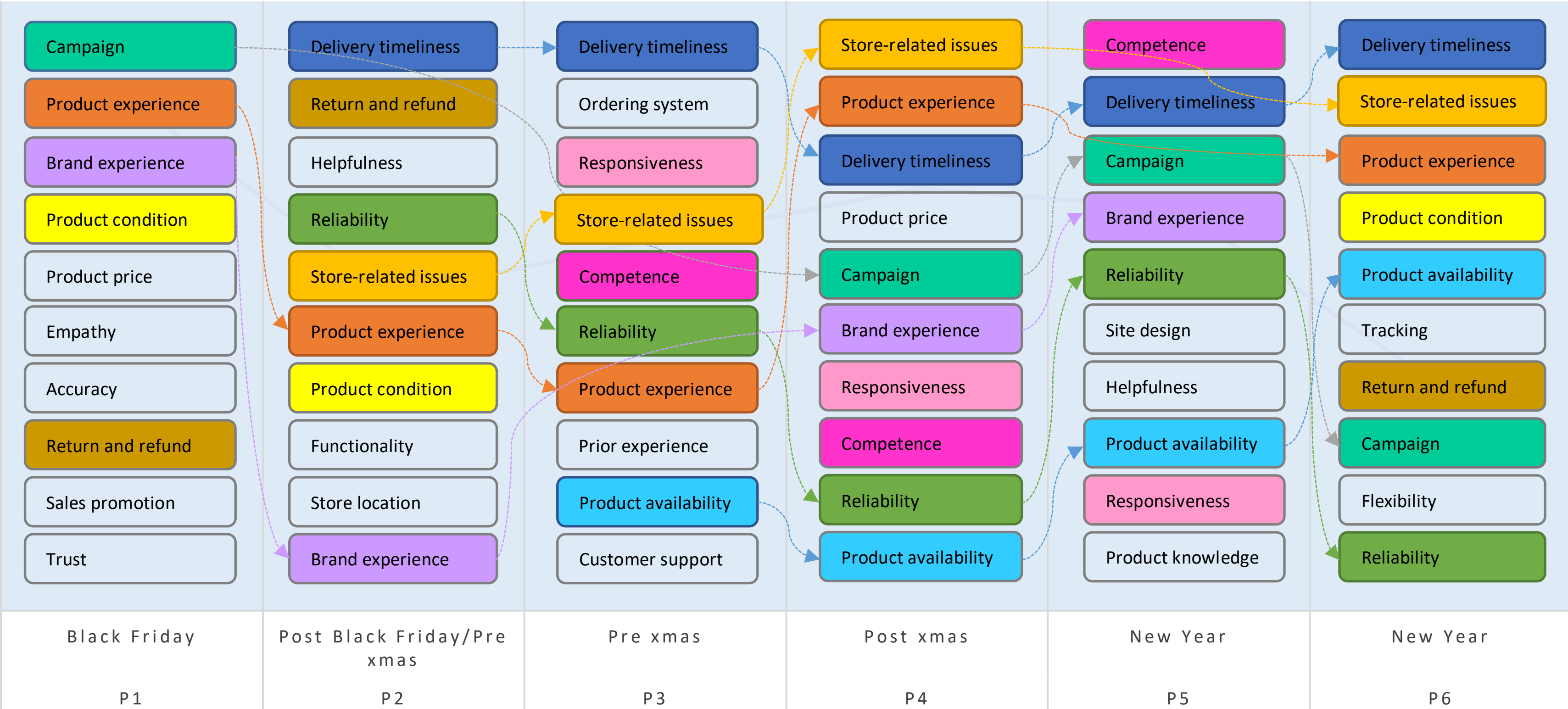

Figure 11 The Evolution of Topics Over Time

Note: Only three to five connections were highlighted for each recurring topic

Overall, a number of recurring topics emerged from the time series. As shown in Figure 11, delivery timeliness, product experience and reliability appeared consistently in five out of the six time periods, indicating that these areas greatly affected customers throughout the time series. On the other hand, in the case of Black Friday, only four topics appeared repeatedly both before and after the event, whereas the rest were different. In contrast, for Boxing Day and Christmas, customers' concerns were the same both before and after the events, except for a few topics such as ordering systems, customer support and product prices that made their first appearances. As illustrated by Figure 11, six main factors – delivery timeliness, product experience, campaign, brand experience, reliability, product availability, and store-related issues – emerged as the principal topics concerning customers as demonstrated by their cumulative weight and their recurrence in the time series.

As can be seen from Figure 11, delivery timeliness was the issue most concerning customers, repeatedly appearing uppermost in the time series. In online retailing, delivery timeliness reflects the speed with which orders are distributed; it is essential for brands to take the least amount of time possible to keep customers satisfied. This is particular challenging during the Black Friday and Christmas sales period due to high transaction volumes and extreme winter weather. Courier companies may not be able to meet promised delivery times for various reasons. This is corroborated by other studies (Clemes, Gan, and Zhang 2014; Lin, Wang, and Chang 2011), which have noted that brands need to focus on timeliness and expanding their staff, particularly in the delivery department, to avoid delays and ensure that customers are satisfied with the service they have paid for. The prevalence of this issue is illustrated by two recent news articles that reported that more than half of customers had either not received their Christmas deliveries when expected, or not received them at all (Drewett 2017; Skynews 2017).

Surprisingly, campaign made a few appearances among the key topics in the time series (see P1, P4, P5 and P6). This was due to the impact of some highly charged external issues that had led to campaigns participated in by Twitter users. Campaigns refer to any promotional activities (e.g., #BlackFriday) or coordinated efforts to reinforce social issues on Twitter, and are particularly noticeable when they go viral and attract the attention of other Twitter users. The example emerging from our results was the #stopfundinghate hashtag. This concerned an issue involving bigotry that trended on Twitter in the UK for a number of hours during the time series. We found that few brands in our study sample were involved in this activism, which raised the ire of some of their customers, who tweeted at them in protest. This corroborates the argument put forth by Kim, Gonzenbach, et al. (2016), who confirmed that traditional media such as newspapers tend to influence issues on Twitter. In contrast, a sales campaign launched by one of the brands also ranked high in the top ten topics. To capitalise on the Black Friday sales event, multiple brands used the #BlackFriday hashtag to connect with customers and raise interest in their sales. Thus, the tweets appearing in the time series comprised a mix of Black Friday sales promotion and activism.

Furthermore, store-related issues also emerged as one of customers' principal concerns that require further attention from online retailing brands. Such issues include a combination of online and offfline store problems such as customers frustrated with online service being compelled to visit a physical store instead, issues with click-and-collect schemes in stores, loss of trust in online retailing service, rude cashiers, too few staff and long queues at pay points. Store-related issues were commonly raised by customers who prefer to visit physical stores for their Christmas shopping. This is reasonable since, out of the five brands in our study, only Amazon has operated fully online since its founding, whilst the others have physical stores alongside an online presence. These differences in customers' preferences imply that the information on Twitter is an integrated hybrid of online and offline store information. This supports the view of (Melis, Campo, Breugelmans, and Lamey 2015) on how customer perceptions of online retail service have influenced offline service. Store-related issues with multichannel retailers potentially affect customer satisfaction, highlighting the importance of a swift response from both online customer service officers and store managers to ensure a speedy resolution to customer problems.

Although the fraction of tweets regarding returns and refunds was small, the issue ranked second in importance after Black Friday (P2) and emerged again after the Christmas and New Year sales when customers started returning their unwanted Christmas gifts. This illustrates that the provision of timely refunds is important, especially after sales events (Griffis, Rao, Goldsby, and Niranjan 2012). Thus, streamlining returns and refunds can help retailers gain customer trust and improved customer loyalty.

As regards products, product experience and brand experience dominated customer concerns on Twitter. Complaints regarding products and services were made to online customer service desks when goods were found to be faulty, defective, damaged, or not functioning as described. Again, these are understandable phenomena because of the volume of sales at this time of the year, when customers are more likely to shop and spend money to capitalise on good deals and in the hunt for the perfect Christmas gift. At this time, customer experiences with purchased products dominated the conversation. There was also much talk about product availability – many customers were frustrated by being unable to obtain desired products because they were out of stock despite stores having been asked to replenish them. Some in-store staff even disregarded customers' requests to restock shelves. During the sales season, it is critical to maintain sufficient levels of stock; the unavailability of products causes great dissatisfaction, resulting in unhappy customers (Ramanathan, Subramanian, and Parrott 2017).

Further observation demonstrated that the emerging issues of trust, sales promotion, ordering systems and site design received relatively low levels of attention. These topics were found to be unconnected and the patterns trivial. One explanation for this is that people face few issues with ordering systems and site design, since companies nowadays focus on providing the best available website technology to aid their buyers. Nevertheless, ordering system–related tweets increased in frequency during the pre-Christmas period when some brands faced difficulties managing high order volumes in the lead-up to Christmas. This scenario must be avoided, since customers are unlikely to return if the problem persists.

4.5 Efficiency of the Approach

The scalability of the method is a data-driven way of the approach from the traditional data analysis method. Although a few studies have analysed Twitter data in online retailing, the analysis of such data is complicated by its nature: it appears in large volumes, is unstructured, noisy and informal, and takes a short-text format. This, in conjunction with the lack of practical tools, has made the analysis of social media data difficult and challenging for researchers (Archak et al., 2011; Netzer et al., 2012; Stieglitz et al., 2018). In addition, several studies also note that traditional analytical techniques are not suited to the analysis of vast amounts of data on a daily basis (He et al., 2017; Lansdall-Welfare et al., 2016).

This research differs from other studies in that it uses Twitter as its data source and integrates different social media analytics techniques to offer possible solutions to these challenges and enhance the literature. Thus, the value of Twitter data is clear: online retail brands can improve their online service and brand image based on the insights derived from the analysis of customers' tweets.

Regarding the efficiency of the approach, topics modelling (LDA) requires most computational resource. For LDA, the differences in training and running time can be explained by the volume of data in the corpus, the number of topics, and the number of iterations. Smaller dataset requires shorter time to run the algorithm (Blei, 2012). In this study, since Twitter consists of short messages up to 140 characters and the topic modelling only focus on the time periods that lead to significant shift of sentiment, the execution time required was less than 7 seconds for 20 topics. However, the computation time of LDA will increase substantially when it applies to longer time series (e.g. the whole sales period) as the volume of data, the number of optimal topics and the number of iterations will all increase accordingly. The number of iterations represents the number of times that the words assignment process goes through. Note that different number of iterations also gives different run times. The model is gradually improving by correcting the assignment of words until reaching a true model based on the number of iterations. The running time of LDA models with different iterations was compared across different runs. In the experiment, the different number of iterations 700, 800 and increments of 100 up to 1000 were performed. It was decided to proceed with 1000 iterations since it produces better perplexity and has only slightly higher time difference compared to 800 iterations.

5 Discussion and Conclusion

This study aimed at understanding the causes of significant shifts in customer sentiment, particularly sharp increases in negative sentiment. To achieve this, we began with an analysis of the trends exhibited by the volume and sentiment of tweets during the time series in question. This was followed by the deployment of event detection techniques to identify critical time points that led to significant deviations in sentiment in the time series. Based on the results of the sentiment and time series analyses, topic modelling was then applied to identify specific factors that contributed to the change in customer opinions over time. The results of these analyses provide companies with insight not only into how customers feel about their products and services (the sentiment), but also why customers feel that way (the reason) and when these feelings arose (the time).

The results indicate that delivery timeliness, product experience, campaign, brand experience, reliability, product availability, and store-related issues lead the emerging topics. The observation of the time series demonstrates that delivery timeliness and product experience consistently top the emerging topics ranking, whilst other topics varied in importance over different time periods. This demonstrates that customers have consistently discussed the same issues over time, giving insight into the most important service areas for online retailing. Delivery and product are already known to be the core factors in online retailing (Cheung, Lee, and Rabjohn 2008; Luo et al. 2012); however, in this study, we also captured the topic of store-related issues, which appeared four times in the time series following Black Friday and in the periods leading up to and immediately following Christmas. Hence, this issue is also highly relevant to online retailing services nowadays. For instance, most online retailers operate both online and offline retail channels. Many customers reserve items online and pick them up in store. The perceived service customers receive from offline retail channels will affect their overall shopping experience. The list of emerging topics identified by this study allows online retailers to more fully understand customers' concerns at different time intervals and shows the need to respond to the diverse customer concerns prevalent at various times.

5.1 Research Contributions

This study offers several important contributions for researchers and practitioners in online retailing and social media literature. On a theoretical level, our study offers a better understanding of customers' opinions towards online retailing and provides insight into what customers are really thinking about by analysing their opinions as expressed on Twitter. Instead of only focusing on customer sentiment, we looked into the manner in which customer opinions evolved on Twitter over time and analysed the reasons underlying significant changes in customer sentiment towards online retailing service. We incorporated sentiment and topic modelling analyses, correlated with time series analysis, to provide an efficient approach to the analysis of information regarding online retailing on Twitter. The analyses effectively associated three factors – sentiment, topic and time – that were used to accurately examine the evolution of the data. Compare to conventional marketing approaches such as focus group, customer survey, and interviews, this study adopts a more data-driven approach that enables to derive more robust and generalizable research findings. In the era of big data, social media can act as a resource enabling companies to fully understand customers' needs, design better business strategies, and improve their online retail service provisioning.

The classification of customer opinions is a critical challenge in online retailing (Li et al. 2016). This study offers a novel contribution by not only classifying customer opinions as a whole, but also segregating them into time intervals according to the key factors that triggered them. Adding to the social media literature in online retailing, our findings examine the reasons behind dynamic changes in customer sentiment using Twitter data. The analysis of Twitter data is challenging and researchers often face difficulties due to the data being informal, unstructured and massive in size (Gandomi and Haider 2015; McAfee et al. 2012). Nevertheless, our analysis presents the manner in which the emerging topics' appearance in the time series is correlated to the rise and fall of customer sentiment. Thus, in addition to providing insights for online retailing researchers and practitioners into the reasons underlying changes in customer sentiment over time, this study also enhances the body of social media literature by offering solutions to some of the challenges faced by the field of social media analytics.

Furthermore, we extend the existing literature by presenting multiple techniques for the analysis of brand-related social media data. Most previous studies analysing social media data have focused on a single technique (Kim, Dwivedi, et al. 2016; Li and Liu 2017). More importantly, our study responds to the work of Ghiassi et al. (2017) and Lansdall-Welfare et al. (2016), both of which required the accurate analysis of big data, by presenting more accurate analytical techniques for evaluating customers' opinions, despite the vast amount of unstructured data on the social media platform. Motivated by these studies, our study combined sentiment analysis, time series analysis, and topic modelling performed on social media data, examined the patterns in the volume and sentiment of tweets, applied time series analysis to detect shifts in sentiment, and then employed topic modelling to decode the reasons behind those shifts. The consequent observations showed that customer sentiments changed due to a variety of reasons, including delivery timeliness, product experience, product availability and store-related issues, and that these factors varied in importance throughout the time series. Finally, this study complements prior studies by accentuating the importance of accurate analysis of social media big data by presenting the evolution of sentiment and customer concerns in the online retailing context.

5.2 Managerial Implications

The provision of online retailing service can be improved with regard to both performance monitoring and management using this study's identification of emerging topics during the annual UK sales season. The frequency with which topics appeared in the time series and the rank of each topic based on its weight provide insight into the areas and factors that ought to be prioritised. For instance, product experience appeared repeatedly in the time series, indicating that customers were unhappy and likely had bad experiences with certain products or services, especially following the Christmas period, when the topic ranked near the top of the chart in our LDA analysis. It stands to reason that this is common, since this is when most customers start to use and enjoy their products, but the topic's recurrence also demonstrates that product experience is a continual problem that may require extra attention from companies and ought to be prioritised. Specifically, the prominence of this factor tells companies to improve product quality for a better customer experience. From a broader perspective, the findings of this study can help online retailers learn how to set up and monitor priorities, think strategically and tactically, be responsive to customer needs, and focus on those areas of their service provision that are critical to making the business run systematically.

A major challenge faced by online retailers is understanding their customers' opinions in order to design and conduct successful marketing campaigns (Li et al. 2016). Since data analytics are increasingly important to the online retail industry, more accurate analyses of big data can significantly enhance companies' understanding of customer behaviour. In a competitive marketplace, companies need to improve the quality of their online retailing service and create distinctive marketing strategies to win the hearts of customers. The use of sentiment classification and topic extraction at multiple time points as conducted by this study gives companies a dynamic view of how customer concerns and opinions evolve over time. Such information allows companies to adjust their strategies and monitor their services in response to customer needs. It also helps clarify how firms should refine their service improvement strategies, which facilitates better operations management, inventory analysis, customer relationship management, and market prediction.

Furthermore, the findings from our study provide companies and managers with richer insight into the best areas at which to target their responses, and how to design their improvement strategies. Although responding to general issues and complaints is essential, tailoring responses to the particular problems reflected at specific points in the time series may offer better customer outcomes. The insights from this study not only allow companies to identify areas for improvement but also suggest optimal timing for the execution of specific improvement strategies. Moreover, retailers may be completely unaware of in-store issues that hold the potential to upset customers; thus, identifying such tweets may help brands in mounting a prompt response. However, as issues aired on social media have the potential to snowball, negative complaints may have a lasting impact on the stores and staff in question, and may affect the views of other existing or potential customers who are exposed to the discussion.

5.3 Future Research

This study can be extended in a number of ways. For instance, because this study focuses mainly on social media data by looking at the views of Twitter users, further research could examine other groups of online retailing customers by means of surveys, interviews or focus groups. These research extensions may provide comparative findings into how different groups of customers perceive online retailing service and the results derive from the comparative analysis can enable online retailers to design and improve service offerings that serves a wider consumer group. Furthermore, it will be valuable to incorporate other methods to improve the analysis of social media data. For instance, social network analysis can be employed to identify the relationships between various topics and observe the characteristics of the customer community, and predictive analysis can be used to predict what customers will say on social media. Finally, future studies may extend our research to other industries, disciplines or management challenges.

References

Archak, N., Ghose, A., and Ipeirotis, P. G. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science,* 57(8), 1485-1509. doi:10.1287/mnsc.1110.1370

Arias, M., Arratia, A., and Xuriguera, R. (2013). Forecasting with twitter data. *ACM Transactions on Intelligent Systems and Technology,* 5(1), 1-24. doi:10.1145/2542182.2542190

Bendle, N. T., and Wang, X. (2016). Uncovering the message from the mess of big data. *Business Horizons,* 59(1), 115-124. doi:10.1016/j.bushor.2015.10.001

Bian, J., Yoshigoe, K., Hicks, A., Yuan, J., He, Z., Xie, M., . . . Modave, F. (2016). Mining Twitter to Assess the Public Perception of the "Internet of Things". *PLoS One,* 11(7. doi:10.1371/journal.pone.0158450

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science,* 2(1), 1-8. doi:10.1016/j.jocs.2010.12.007

Cheung, C. M. K., Lee, M. K. O., and Rabjohn, N. (2008). The impact of electronic word-of-mouth. *Internet Research,* 18(3), 229-247. doi:10.1108/10662240810883290

Clemes, M. D., Gan, C., and Zhang, J. (2014). An empirical analysis of online shopping adoption in Beijing, China. *Journal of Retailing and Consumer Services,* 21(3), 364-375. doi:10.1016/j.jretconser.2013.08.003

Collier, J. E., and Bienstock, C. C. (2006). How Do Customers Judge Quality in an E-tailer? *MIT Sloan Management Review,* 48(1), 35-40

Daniel, M., Neves, R. F., and Horta, N. (2017). Company event popularity for financial markets using Twitter and sentiment analysis. *Expert Systems with Applications,* 71, 111-124. doi:10.1016/j.eswa.2016.11.022

Daud, A. (2012). Using time topic modeling for semantics-based dynamic research interest finding. *Knowledge-Based Systems,* 26, 154-163. doi:10.1016/j.knosys.2011.07.015

Drewett, Z. (2017). "This is what you should do if your Christmas deliveries are late", from http://metro.co.uk/2017/12/20/christmas-deliveries-late-7173324/

Dueñas-Fernández, R., Velásquez, J. D., and L'Huillier, G. (2014). Detecting trends on the Web: A multidisciplinary approach. *Information Fusion,* 20, 129-135. doi:10.1016/j.inffus.2014.01.006

Fan, R., Zhao, J., and Xu, K. (2015). Topic dynamics in Weibo: a comprehensive study. *Social Network Analysis and Mining,* 5(41), 1-15. doi:10.1007/s13278-015-0282-0

Figueiredo, F., Almeida, J. M., Gonçalves, M. A., and Benevenuto, F. (2014). On the dynamics of social media popularity: A YouTube case study. *ACM Transactions on Internet Technology,* 14(4. doi:10.1145/0000000.0000000

Francis, J. E. (2009). Is C-OAR-SE best for internet retailing service quality? *Managing Service Quality: An International Journal,* 19(6), 670-686. doi:10.1108/09604520911005062

Gandomi, A., and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management,* 35(2), 137-144. doi:10.1016/j.ijinfomgt.2014.10.007

Ghiassi, M., Zimbra, D., and Lee, S. (2017). Targeted Twitter Sentiment Analysis for Brands Using Supervised Feature Engineering and the Dynamic Architecture for Artificial Neural Networks. *Journal of Management Information Systems,* 33(4), 1034-1058. doi:10.1080/07421222.2016.1267526

Ghose, A., Ipeirotis, P. G., and Li, B. (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science,* 31(3), 493-520. doi:10.1287/mksc.1110.0700

Giachanou, A., and Crestani, F. (2016). Like It or Not. *ACM Computing Surveys,* 49(2), 1-41. doi:10.1145/2938640

Goswami, A., and Kumar, A. (2016). A survey of event detection techniques in online social networks. *Social Network Analysis and Mining,* 6(107), 1-25. doi:10.1007/s13278-016-0414-1

Grégoire, Y., Salle, A., and Tripp, T. M. (2015). Managing social media crises with your customers: The good, the bad, and the ugly. *Business Horizons,* 58(2), 173-182. doi:10.1016/j.bushor.2014.11.001

Griffis, S. E., Rao, S., Goldsby, T. J., and Niranjan, T. T. (2012). The customer consequences of returns in online retailing: An empirical analysis. *Journal of Operations Management,* 30(4), 282-294. doi:10.1016/j.jom.2012.02.002

Hajikhani, A., Porras, J., and Melkas, H. (2017). Brand Analysis in Social Network Services: Results from Content Analysis in Twitter Regarding the US Smartphone Market. *International Journal of Innovation and Technology Management,* 13(6), 1-19. doi:10.1142/s0219877017400089

Hennig-Thurau, T., Malthouse, E. C., Friege, C., Gensler, S., Lobschat, L., Rangaswamy, A., and Skiera, B. (2010). The Impact of New Media on Customer Relationships. *Journal of Service Research,* 13(3), 311-330. doi:10.1177/1094670510375460

Heverin, T., and Zach, L. (2012). Use of Microblogging for Collective Sense-Making During Violent Crises: A Study of Three Campus Shootings. *Journal of the American Society for Information Science and Technology,* 63(1), 34-47. doi:10.1002/asi

Hewett, K., Rand, W., Rust, R. T., and van Heerde, H. J. (2016). Brand Buzz in the Echoverse. *Journal of Marketing,* 80(3), 1-24. doi:10.1509/jm.15.0033

Ibrahim, N. F., Wang, X., and Bourne, H. (2017). Exploring the effect of user engagement in online brand communities: Evidence from Twitter. *Computers in Human Behavior,* 72, 321-338. doi:10.1016/j.chb.2017.03.005

Kaplan, A. M., and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons,* 53(1), 59-68. doi:10.1016/j.bushor.2009.09.003

Kim, E. H., Jeong, Y. K., Kim, Y., Kang, K. Y., and Song, M. (2016). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science,* 42(6), 763-781. doi:10.1177/0165551515608733

Kim, Y., Dwivedi, R., Zhang, J., and Jeong, S. R. (2016). Competitive intelligence in social media Twitter: iPhone 6 vs. Galaxy S5. *Online Information Review,* 40(1), 42-61. doi:10.1108/oir-03-2015-0068

Kim, Y., Gonzenbach, W. J., Vargo, C. J., and Kim, Y. (2016). First and Second Levels of Intermedia Agenda Setting: Political Advertising, Newspapers, and Twitter During the 2012 U.S. Presidential Election. *International Journal of Communication,* 10, 4550-4569

Kulkarni, V., Al-Rfou, R., Perozzi, B., and Skiena, S. (2015). *Statistically Significant Detection of Linguistic Change.* Paper presented at the 24th International Conference on World Wide Web - WWW '15.

Lai, E. L., Moyer, D., Yuan, B., Fox, E., Hunter, B., Bertozzi, A. L., and Brantingham, P. J. (2016). Topic time series analysis of microblogs. *IMA Journal of Applied Mathematics,* 81(3), 409-431. doi:10.1093/imamat/hxw025

Lampos, V., Bie, T. D., and Cristianini, N. (2010). Flu detector-tracking epidemics on Twitter. *Machine Learning and Knowledge Discovery in Databases*, 599-602

Lampos, V., and Cristianini, N. (2012). Nowcasting Events from the Social Web with Statistical Learning. *ACM Transactions on Intelligent Systems and Technology,* 3(4), 1-22. doi:10.1145/2337542.2337557

Lansdall-Welfare, T., Dzogang, F., and Cristianini, N. (2016). *Change-Point Analysis of the Public Mood in UK Twitter during the Brexit Referendum*. Paper read at IEEE 16th International Conference on Data Mining Workshops (ICDMW 2016). 10.1109/icdmw.2016.0068

Lansdall-Welfare, T., Lampos, V., and Cristianini, N. (2012). *Effects of the Recession on Public Mood in the UK*. Paper read at International World Wide Web Conference Committee (IW3C2).

Lee, T. Y., and BradLow, E. T. (2011). Automated marketing research Using online customer reviews. *Journal of Marketing Research*, 881-894

Li, H., Yan, J., Weihong, H., and Ding, Z. (2014). Mining User Interest in Microblogs with a User-Topic Model. *China Communications,* 11(8), 131-144

Li, J., Li, X., and Zhu, B. (2016). User opinion classification in social media: A global consistency maximization approach. *Information & Management,* 53(8), 987-996. doi:10.1016/j.im.2016.06.004

Li, Q., and Liu, Y. (2017). Exploring the diversity of retweeting behavior patterns in Chinese microblogging platform. *Information Processing & Management,* 53(4), 945-962. doi:10.1016/j.ipm.2016.11.001

Lin, H. H., Wang, Y. S., and Chang, L. K. (2011). Consumer responses to online retailer's service recovery after a service failure. *Managing Service Quality: An International Journal,* 21(5), 511-534. doi:10.1108/09604521111159807

Lo, S. L., Chiong, R., and Cornforth, D. (2016). Ranking of high-value social audiences on Twitter. *Decision Support Systems,* 85, 34-48. doi:10.1016/j.dss.2016.02.010

Lugmayr, A., and Grueblbauer, J. (2016). Review of information systems research for media industry–recent advances, challenges, and introduction of information systems research in the media industry. *Electronic Markets,* 27(1), 33-47. doi:10.1007/s12525-016-0239-9

Luo, J., Ba, S., and Zhang, H. (2012). The Effectiveness of Online Shopping Characteristics and Well-Designed Websites on Satisfaction. *MIS Quarterly,* 36(4), 1131-1144

Ma, Z., Sun, A., and Cong, G. (2013). On predicting the popularity of newly emerging hashtags in Twitter. *Journal of the American Society for Information Science and Technology,* 64(7), 1399-1410. doi:10.1002/asi.22844

McAfee, A., Brynjolfsson, E., and Davenport, T. H. (2012). Big data: the management revolution. *Harvard Business Review,* 90(10), 60-68

Melis, K., Campo, K., Breugelmans, E., and Lamey, L. (2015). The Impact of the Multi-channel Retail Mix on Online Store Choice: Does Online Experience Matter? *Journal of Retailing,* 91(2), 272-288. doi:10.1016/j.jretai.2014.12.004

Naaman, M., Becker, H., and Gravano, L. (2011). Hip and trendy: Characterizing emerging trends on Twitter. *Journal of the American Society for Information Science and Technology,* 62(5), 902-918. doi:10.1002/asi.21489

Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M. (2012). Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science,* 31(3), 521-543. doi:10.1287/mksc.1120.0713

News, E. (2017). "The top 500 ecommerce retailers in the UK", from https://ecommercenews.eu/top-500-ecommerce-retailers-uk/

Nikolenko, S. I., Koltcov, S., and Koltsova, O. (2016). Topic modelling for qualitative studies. *Journal of Information Science,* 43(1), 88-102. doi:10.1177/0165551515617393

O'Connor, B., Balasubramanyan, R., Routledge, B. R., and Smith, N. A. (2010). *From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series.* Paper presented at the Fourth International AAAI Conference on Weblogs and Social Media.

Odlum, M., and Yoon, S. (2015). What can we learn about the Ebola outbreak from tweets? *Am J Infect Control,* 43(6), 563-571. doi:10.1016/j.ajic.2015.02.023

Ramanathan, U., Subramanian, N., and Parrott, G. (2017). Role of social media in retail network operations and marketing to enhance customer satisfaction. *International Journal of Operations & Production Management,* 37(1), 105-123. doi:10.1108/ijopm-03-2015-0153

Ranco, G., Aleksovski, D., Caldarelli, G., Grcar, M., and Mozetic, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. *PLoS One,* 10(9), e0138441. doi:10.1371/journal.pone.0138441

Retailing, I. (2016). "InternetRetailing UK Top500 Retailers 2016", from http://internetretailing.net/iruk/the-uks-top-500-ecommerce-retailers/

Robertson, C., and Yee, L. (2016). Avian Influenza Risk Surveillance in North America with Online Media. *PLoS One,* 11(11), e0165688. doi:10.1371/journal.pone.0165688

Skynews. (2017). "Christmas delivery chaos sees parcels delayed, damaged and left in bins", from https://news.sky.com/story/christmas-delivery-chaos-sees-parcels-delayed-damaged-and-left-in-bins-11178143

Thelwall, M., Buckley, K., and Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology,* 62(2), 406-418

Thelwall, M., Buckley, K., and Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology,* 63(1), 163-173. doi:10.1002/asi.21662

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the Association for Information Science and Technology,* 61(12), 2544-2558

Vergeer, M., and Franses, P. H. (2015). Live audience responses to live televised election debates: time series analysis of issue salience and party salience on audience behavior. *Information, Communication & Society,* 19(10), 1390-1410. doi:10.1080/1369118x.2015.1093526

Wilkinson, D., and Thelwall, M. (2012). Trending Twitter topics in English: An international comparison. *Journal of the American Society for Information Science and Technology,* 63(8), 1631-1646. doi:10.1002/asi.22713

Wu, F., Song, Y., and Huang, Y. (2016). Microblog sentiment classification with heterogeneous sentiment knowledge. *Information Sciences,* 373(1), 149-164. doi:10.1016/j.ins.2016.09.002

Yildirim, A., Üsküdarlı, S., and Özgür, A. (2016). Identifying Topics in Microblogs Using Wikipedia. *PLoS One,* 11(3), 1-20. doi:10.1371/journal

Yuki, T. (2015). What Makes Brands' Social Content Shareable on Facebook? *Journal of Advertising Research,* 55(4), 458-470. doi:10.2501/jar-2015-026

Zhang, C., Wang, H., Cao, L., Wang, W., and Xu, F. (2016). A hybrid term–term relations analysis approach for topic detection. *Knowledge-Based Systems,* 93, 109-120. doi:10.1016/j.knosys.2015.11.006

Zhang, X., Fuehres, H., and Gloor, P. A. (2011). Predicting stock market indicators through twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences,* 26, 55-62

Zhou, X., Wan, X., and Xiao, J. (2016). CMiner: Opinion Extraction and Summarization for Chinese Microblogs. *IEEE Transactions on Knowledge and Data Engineering,* 28(7), 1650-1663