Mining Social Network Content of Online Retail Brands: A Machine Learning Approach

Noor Farizah Ibrahim and Xiaojun Wang
School of Economics, Finance and Management, University of Bristol, Bristol, United Kingdom
ni14406@bristol.ac.uk
xiaojun.wang@bristol.ac.uk

Abstract: The growing popularity of social media networks such as Twitter, Facebook, Instagram and so forth is taking the Internet sphere to a higher level, creating a huge volume of social network-generated data, including tweets. As a result, all these data and information can be easily or exclusively found on the Internet by the public. However, little is known yet about how to turn this data into useful knowledge and the collection of the data has not been sufficiently researched, especially in the online retail industry. In an effort to help companies in the online retail industry to utilise the data, this paper explores social network-generated contents on Twitter sites through topics identification that can subsequently facilitate companies to improve their services and performances. This paper describes a study based on data gathered on the Twitter sites of the leading UK retailers: Amazon, Argos and Tesco, during the periods of Black Friday, Boxing Day and Christmas in the UK. We chose Twitter as a platform because, whilst other previous works addressed online reviews and website contents in their studies, microblogs are just starting to be explored due to their characters' limitations and informal content making it harder to interpret. In this paper, we infer topics from short-text tweets posted by customers mentioning brand names by employing a machine learning approach. We then analyse the tweets to determine the primary issues or topics focused by customers regarding these online retail brands by interpreting the themes of the specific keywords set represented. The topics which emerge based on the collection of tweet posts contribute to various aspects of companies' operations, such as delivery, customer service and product performance. We believe that this study derives some implications and insights for online retail brand companies in improving their service provisions, especially in customer service management. Insights on such topics can be beneficial in numerous ways, such as marketing, innovation and public image.

Keywords: data mining, social media, online retailing, topic model, machine learning, Twitter

1. Introduction

Over recent years, social media has received abundant attention and is becoming an important communication tool across the globe. Tagging along is the content generated by social media that has become increasingly popular by providing information for the public. In a business context especially, social media such as Facebook and Twitter have been extensively used by online retail brands to communicate with customers and facilitate business operations. It acts as a medium of online sources created by customers to educate each other about brands, products, services, personalities and issues (Blackshaw & Nazzaro, 2006). Twitter for instance, offers various information, purposes and functions with a wide diversity of content and coverage. From Twitter, it is crucial for companies to accurately measure people's opinion from what customers write and post online so as to better tailor and adapt their policies and products (Fong et al, 2013). Social media also facilitates in managing relationship with customers to improve sales and customer service (Safko, 2010). Some businesses even allocate an investment for social media channels to create and propagate their brand through viral content, contests and other engagement efforts. In fact, the traditional campaign methods are believed to be gradually changing from a one-way advertising message into a two-way dialogue with customers (Kumar & Mirchandani, 2012).

Under the Twitterverse, brands and customers engage in unstructured dialogue, including sharing and delivering information and updates. Nair (2011) argued that engagement between customers and brands involving the combination of sociology and technology perspectives cannot be underestimated, since it has impact to the brand of an organisation. According to a survey data from Twitter 2015, customers prefer to choose and discuss customer service queries with brand companies on Twitter (see Figure 1) rather than Facebook. This provides prospect for the brands to narrow down their focus platform and explore opinions from customers on the Twitter microblogging platform.

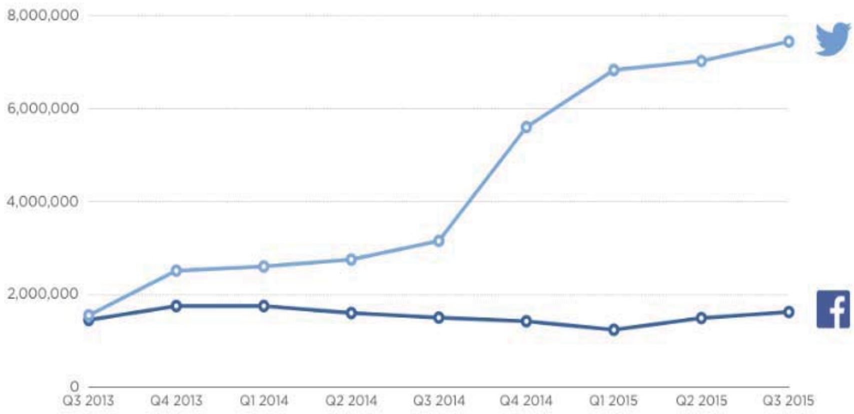

Figure 1: Customer service queries by platform (Picazo, 2016)

Although Twitter is known as a rich source with potentially useful information, its content has not been well studied (Lin & He, 2009). The large volume of textual data available on Twitter is beyond what traditional methods can handle and the investigation of social network content is particularly crucial and timely in current research. Hence, we conducted this study to examine the pattern of contents generated on Twitter based on three leading UK online retailers: Amazon, Argos and Tesco. The Latent Dirichlet Allocation (LDA) topic modelling approach was conducted to identify the patterns hidden in these unstructured textual data. This unsupervised topic modelling approach was employed on our dataset as it is also one of the most widely used algorithms in the computer science and engineering field (Guo et al, 2016). In this research, we extracted the tweets topically for summarising and analysing the Twitter content. The following are the research questions generated in this study:

RQ1: What are the most common concerns and interest of customers that are discussed and shared among Twitter users regarding online retail brands?

RQ2: How do these identified topics differ from previous works?

RQ3: How can companies learn from Twitter users to improve their online retail service provision?

To fit our objectives, we conducted this research to effectively identify and browse online retail brands-related topics on Twitter. Mallet LDA algorithm was proposed to extract keywords to summarise the microblog content of the selected brands. However, there are also challenges in using this dataset for current topic modelling, including: (1) tweets are much shorter that online reviews or traditional articles; (2) various unstructured styles of writing; (3) diversity of topics due to a wide platform; (4) informal and slang text; (5) often ungrammatical and having spelling errors; (6) texts are noisy. In the following sections, we discuss the methods followed by the experiments and results. Then, the implications of this study are explained under the discussion section and our conclusions regarding the objective of the research and possible areas of future studies are outlined.

2. Related works

The popularity of social media has started to draw researchers' attentions over the past years (Zhao et al, 2011). It is due to the nature of microblogging that provides researchers with large amount of texts and may potentially contain useful information that can hardly be found in available traditional information sources. Recent social media studies examined Twitter from variety of perspectives including trending topics (Madani, Boussaid, & Zegour, 2015; Yıldırım, Üsküdarlı, & Özgür, 2016), movie prediction (Lin et al, 2009), breaking news and political events (Guo et al, 2016; Kim et al, 2016) and community learning (Beykikhoshk et al, 2015; Bian et al, 2016).

There are numerous studies related to social network content in a business environment. Valuable information and knowledge discovered from this network content is useful for business intelligence at various managerial levels in organisation and contribute to the success of a business (Abbasi et al, 2008). Companies have progressively started to include social media in their new business strategy (Ngai, Tao, & Moon, 2015) and part of brand building (Cawsey & Rowley, 2016). Cawsey et al (2016) argued that social media is a strategic approach for brand building and highlighted that brand image and brand awareness enhancement were the most common in adopting social media. According to Mangold and Faulds (2009), numerous researchers have recommended companies to deploy mixed marketing strategy including blogs handling, social media tools and promotional tools to engage with customers.

Furthermore, a topic modelling approach is becoming a standard tool in extracting topics from social networks and facilitates researchers to understand large collections of unstructured data. A number of studies has demonstrated that the approach has been effectively applied in variety of tasks including sentiment analysis (Titov & McDonald, 2008), multi-document summarisation (Haghighi & Vanderwende, 2009) and image labelling (Feng & Lapata, 2010). Kim et al (2016) conducted the LDA topic modelling technique and found that topic coverage on Twitter was more precise and entities such as person, organization and location can be extracted from the tweets. Li et al (2014) in Sina microblogs study found that LDA topic model provided insights in discovering user interests in microblogs. Similarly, Yıldırım et al (2016) identified topics in microblogs in Wikipedia using LDA approach.

Overall, social media research varies from other forms of online research. Social media serves good benefits for businesses includes increasing profitability (Kaplan & Haenlein, 2010), improving customer service, promotion and product innovation (Berthon et al, 2012). Topic modelling approach on the other side has been widely used in extracting topics from unstructured data like microblogs (Guo et al, 2016). Many existing topic extraction methods exist in diverse settings. The difference is that, our work concentrates on three different online brand communities and we extract the textual data posted by customers on Twitter that have their own challenges.

3. Methods

3.1 Model framework

In this study, the workflow analysis consists of steps as depicted in Figure 2. We also visualised the extracted topics using topic cloud to provide a clear image and easy-to-understand graphical presentations to substantiate our findings.

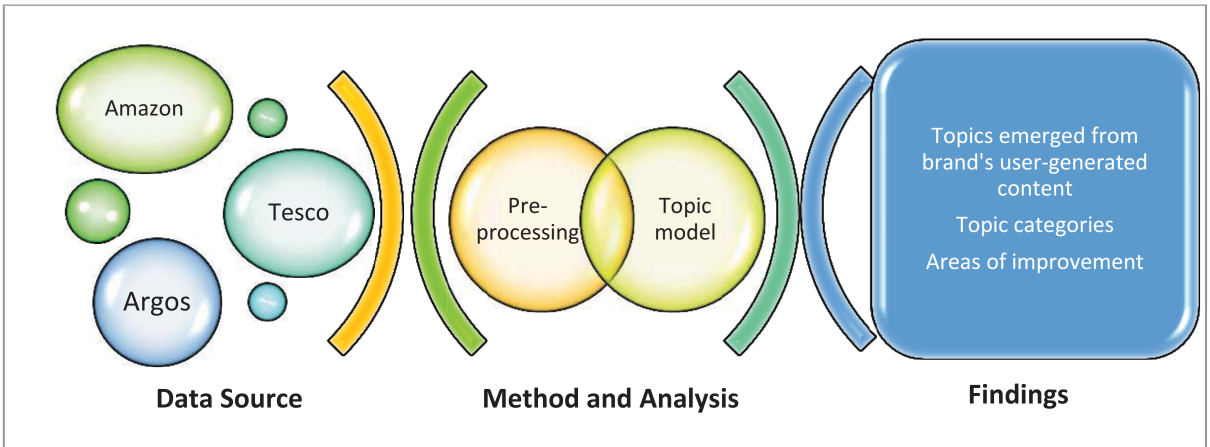

Figure 2: The model framework for analysing tweets

3.2 Tweets pre-processing

Tweets that mentioned a brand name were collected and pre-processed for the analysis. Three leading brands comprises Amazon, Argos and Tesco were selected from the UK top 20 online retailing site. The dataset represents the tweets surrounding the brands from 20 November 2016 to 20 January 2017, the periods of Black Friday, Boxing Day, Christmas and New Year sales events in the UK. We used keyword searching for brand names with the only English tweets were selected. The dataset included 300,249 qualified tweets as shown in Table 1.

Table 1: Dataset

Start Time (GMT)	End Time (GMT)	Amazon # of tweets	Argos # of tweets	Tesco # of tweets
Nov 20, 2016 00:00	Jan 20, 2017 23:59	186,885	31,509	81,855

As seen in Figure 3, we first filtered non-English tweets and cleaned the text by removing numbers, punctuation and converted all text to lowercase. We then tokenised the words by breaking up the texts into discrete words. Further actions were removing all the stop words and reduced the words to their stems. Then we applied $tf - idf$ to omit terms that have low frequency and filtered out the meaningless words from the texts. Mallet LDA was executed to find which topic each tweet has been assigned into.

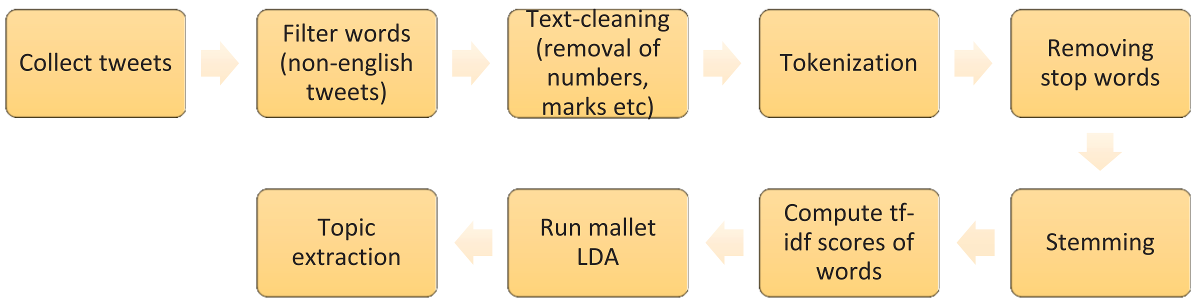

Figure 3: Overview of pre-processing and topic extraction

3.3 Processing Topic Model: LDA Analysis

LDA is a statistical unsupervised machine learning technique to identify latent topic information from a large corpora (Dokoohaki & Matskin, 2012). LDA is widely known for identifying patterns in texts that derive to the emerging topics. Topics that emerge created from words that co-occur more frequently across documents and most likely belong to the same topic (Llewellyn et al, 2015). The objective of this model is to identify and extract topics from what have been tweeted by customers.

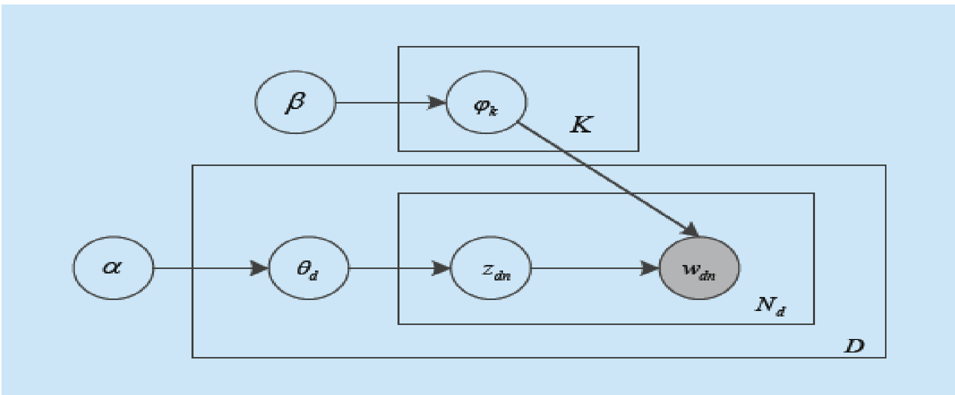

Figure 4: Graphical model representation of LDA

Figure 4 shows the graphical model presentation of LDA. K is the number of topics; D: the number of documents; Nd: the length of the $d-th$ document; θd ability distribution of the $d-th$ document; ϕk: the word probability distribution of the topic k; Wdn: the $n-th$ word in the $d-th$ document, and Zdn: the word's topic (Li et al, 2014). In this study, Mallet LDA library was utilised for modelling topic from the tweet corpora to learn more about the context of online retail brands discussed on Twitter.

4. Experiments and Results

In this section, we presented the topic model analysis to extract the relevant topics from our tweet corpora. Gibbs sampling was also deployed to compute the optimal latent topic number.

4.1 Topic modelling: LDA

We performed our experiments on the datasets retrieved from Twitter during the Black Friday, Christmas and Boxing Day events periods. We chose MALLET library to perform LDA computation because it enables the program to browse words that are frequently found together or share a common connection. Further, LDA helps identify additional and hidden terms within topics that may not be directly instinctual but relevant. We suspected that this topic model would provide less relevant topics relating to online retail brands since such topics are generally less frequently discussed among Twitter users. That could be seen through tweets that are mostly about product promotion, marketing campaigns and the tweet spam. However, our model picked up numerous relevant keywords that are important to online retail brands that will be discussed in the later section. Following the justifications of our approach, each step in Figure 5 was performed.

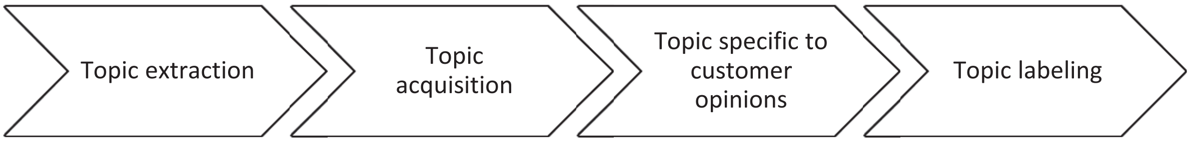

Figure 5: Steps for topic modelling

4.2 Determining topic number

Gibbs sampling was applied to determine the optimal number of topics by experimenting different values of k to check for stability. We run the sampling algorithm for 1000 iterations and tried out different values of k by adopting Griffiths and Steyvers (2004) approach and the one with the largest likelihood was selected. Using Gibbs algorithm, we tried out different values of k that fitted to the model and found the optimal value. We used each value of k to extract the log-likelihood values from each model and then computed the harmonic mean of the log-likelihood values to get the final k. Figure 6 is part of algorithm used in this study.

```
> harmonicMean <- function(logLikelihoods, precision=2000L) {
+     llMed <- median(logLikelihoods)
+     as.double(llMed - log(mean(exp(-mpfr(logLikelihoods,
+     prec = precision) + llMed))))
+ }
```

Figure 6: Finding optimal number algorithm

4.3 Topic extraction results

In this study, we extracted 62 topics that have been identified based on the probability distribution over a number of words. As LDA is an unsupervised method, topics were extracted by generating the topic model that represent the word probability for each keywords and topic. Figure 7 illustrates the topics according to its weight and due to page limitation, only the first 9 topics were shown here due to page limitation. As can be seen, many of these topics captured important aspects of retailers' operations. Each topic was visualised as topic clouds where dominant words with high probabilities are enlarged based on the proportions of probabilities in the topic.

Figure 7: Topic modelling result

From this LDA output, we observed several interesting concepts or terms that are clearly relevant to online retail brand operations such as ‘delivery’, ‘order’, ‘service’, ‘product’ etc. For instance, we can see that topic 27 is about Argos and Amazon especially on ‘order’ and ‘delivery’ as highlighted in the percentage Table 2. Additionally, it contains words related to Christmas e.g. gift and present. In Table 2, the highest probability in each row is in blue colour font and green for the second largest probabilities in cases where the probabilities are close. Based on the composition table, it implies that all brands at least have 2 topics with comparable probabilities e.g. Topic 27 and 22 for Amazon, Topic 27 and 51 for Argos and Topic 51 and 9 for Tesco respectively. It means that all brands speak or represent all those topics in proportions of the obtained probabilities. A reading of tweets in each brand proved that customers talked or conversed about the brands pertaining all those topics.

Table 2: Percentage composition breakdown of each topic for the whole corpus

Brand	Topic 27	Topic 51	Topic 44	Topic 55	Topic 22	Topic 47	Topic 9	Topic 7	Topic 8
Amazon	19.2%	2.57%	10.35%	5.17%	39.76%	12.10%	0.17%	0.68%	0.29%
Argos	24.69%	26.04%	9.4%	5.12%	1.23%	1.05%	0.26%	0%	0%
Tesco	10.5%	27.88%	11.37%	7.56%	0.34%	0.95%	20.05%	0%	0%

The topics were then manually labelled into categories based on the combination of human judgement (Guo et al, 2016) and literature on online retailing. We assigned the topics to 8 categories in which 6 (delivery, customer service, brand/product, security, website, transaction) of them are from the literature (Collier & Bienstock, 2006; Francis, 2007; Luo, Ba, & Zhang, 2012) whilst the last two (promotion and network building) were proposed as they were not mentioned in the previous work. The interpretations of the categories are detailed below.

- Delivery - as seen in most topics (e.g. topic 27, 51, 22), most of customers talked about delivery arriving late, change in delivery time, late order status, missing package, stolen package, and unreliable shipping company. There were also problems with late refund process, tracking option, and lazy delivery guy.
- Customer service – as seen in topic 27, 44, 55, 47 for instance, customers seem to voice out their complaints about a variety of customer-related issues includes hard to reach customer service desk, horrible service personnel, dissatisfaction, insult by a cashier at brand store, and cancellation order.
- Brand/product – in topic 55, 47, 7 and 8 for instance, customers tweeting about current and past-experience with the product, product design, product review, product characteristic, global brands, and the originality of the product.
- Security and privacy – as seen in the tweets in topic 5 and 36, some customers voice out their trust towards the companies in terms of providing personal info and confidentiality of their queries.
- Website – in topic 51 and 47 as examples, customers tweeted about the need of technical support of website performance and concerns of some not working functions on the website
- Transaction – as seen in this category (e.g. topic 47), customers mentioned about ordering system error, apps error, app and website outage, cart disappear and support for system problem.
- Promotion – as expected, there are lots of tweets pertaining to product sales, product advertisement, and discounts since brands tend to post tweets when they are doing promotions or organizing an event and got retweeted by customers e.g. book sales and prime offer for boxing day and Christmas.
- Network building – as observed in topic 9, when customers talked about something trending such as 'food' that attract their attention, some users will joint conversation on the product e.g. lamb joint conversation; and this lead to customer network building.

Based on the findings, we believe that topic modelling approach is a viable approach to learn characteristics of concept based on Twitter dataset. Topics detected from this topic model analysis can be interpreted as important context derived from their top probability words associated to online retail brands. We extracted keywords to summarise microblogs content on the selected brands. According to Guo et al (2016), the easier the topics are interpreted into meaningful contexts, the better the detected topic. However, we realised that the interpreting process are rather complicated and is very subjective which depends on the comprehensive ability of an interpreter. Overall, this LDA modelling analysis discovered variety of topic of online retail brands operations.

5. Discussion

In this study, our focus is to gain a better understanding on what has been shared on Twitter platforms and concerns from customers with the purpose to improve businesses and customer services. We explored tweet contents using a topic modelling approach to find natural and hidden topics within the tweet text. This allowed us to generate and extract relevant topics to be inferred from the corpora of tweets and, from that, we gained insights on how well online retail brands are doing and in which areas they need to improve. In this study, our findings revealed that delivery, customer service and product have the most weightage and high co-occurrences in which these contents are the most talked about on Twitter in an online retail brand context. For companies, this is an area for improvement to understand and improve customer service that is purely based on the customers' words. The new identified categories, promotion and network building also yield some insights. The buzz created from brand-customer communication makes people aware about the brands, learn new products and attract potential customers. This is great as it can build a Twitter community that share the same interest and a place for building social connection and content discovery.

Furthermore, the results from this paper can be used by managers and practitioners to offer service within customers' expectations and show that they care about their customers. Untimely and late delivery might be the hardest part to improve, but showing concern in handling their queries and complaint might well lessen the resentment and wind down the customers. We believe that the issues found in our topic model analysis can be fixed and will help the brands in improving customer experience and, hence, enhanced business performance. For instance, the tweets under delivery category can explain why it has garnered so much attention from customers because it is one of the core importances in the online retail business. As such, brands absolutely need to understand that interactions with customers are needed in order to meet their expectations. By understanding customers, brands can convert the online conversations into valuable knowledge that will be useful for their businesses. We notice from the analysis that customers share diverse content on social media. Therefore, brands also can use this opportunity to share a mix of content that might interest customers and which would work much better rather than spamming customers with sales messages.

We also observed that detailed focus on consumer-generated content could help brands to rebuild their customer management efficiently. Customers sharing ideas and building network on social media platforms allow companies the ability to connect with them and keep up with what is happening on Twitter. To connect with customers, brands should show interest in the topic being talked about and engage with customers. Timely engagement in this low cost and high efficiency platform allows companies to improve customer experience (Kaplan et al, 2010). This is aligned with a study from Karimov and Brengman (2011) that demonstrated that companies need to invest in and maintain their social media channels with potential customers in order to stay competitive in the industry. This is an opportunity to satisfy customers and, at the same time, to plot strategies for a competent business.

With the increasing use of social media, the adoption of social media tools and personnel in business strategy should be extended and widely applied. It can be a platform for customer management in handling customer service and enquiries. It can also be beneficial for brands in turning to social media because it plays an important role in effectively improving customer experience. Since companies are aware that building a great customer service and experience is a complex task, they should invest in social media management activities to improve the experience. Establishing a systematic monitoring tool of social media to manage customer queries and experience and is one of the top priorities to effectively handle customers (Ibrahim, Wang, & Bourne, 2017).

Overall, this study brings new insights to the approach in the way that topic modelling results reveal useful information for business as one of the contributions in this study. It facilitates to organise and offer insights for researchers to understand the hidden content behind the large collections of unstructured text bodies. We believe that an effective social media management enables companies to go to a place where customers meet and converse, to understand and listen to their voice. In contrast with traditional methods where companies tend to talk rather than listen (Nair, 2011), today's social media age is different; talk less, but listen more and play the game.

6. Conclusion

In this study, our objective is to analyse Twitter content to discover emergent topics associated to online retail brands. Using a topic modelling approach, we generated some insights to the practitioners from our analysis of tweets expressing their opinions and experiences towards online retail brands. Our analysis highlighted six categories of topics (delivery, customer service, product, website, security, transaction) derived from literature and two new additional categories (promotion and network building) closely associated with microblogs like Twitter were proposed. From the observation, it shows that delivery and customer service areas were discussed the most on Twitter and need a bigger improvement, especially for Amazon and Tesco. Hence, these companies need to focus on the highlighted areas and strengthen their strategies for a better service and customer experience. Our approach shows that the identified topics were somehow relevant and this knowledge can be used to reach out to customers. Based on our findings, we believe that social media is vital for companies to understand customers better and could lead to improve business strategies. Getting close to customers should be today's priority for companies and social media platforms should be fully utilised to achieve the desired purposes. However, it will be interesting to add more analysis, such as sentiment and cross validation analysis, in future work for more reliable and accurate results. Furthermore, the analysis could be extended to other brands as well to identify other new categories that might emerge. It will be interesting to explore and improve on the results.

Acknowledgement

Supported by the University of Bristol Alumni Foundation and the Government of Malaysia.

References

Abbasi, A., Chen, H., Thoms, S., and Fu, T. (2008) "Affect analysis of web forums and blogs using correlation ensembles", *IEEE Transactions on Knowledge and Data Engineering,* Vol. 20, No. 9, pp. 1168-1180.

Berthon, P. R., Pitt, L. F., Plangger, K., and Shapiro, D. (2012) "Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy", *Business Horizons,* Vol. 55, No. 3, pp. 261-271.

Beykikhoshk, A., Arandjelović, O., Phung, D., Venkatesh, S., and Caelli, T. (2015) "Using Twitter to learn about the autism community", *Social Network Analysis and Mining,* Vol. 5, No. 1.

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,* Vol. 11, No. 7.

Blackshaw, P., and Nazzaro, M. (2006) "Consumer-generated media (CGM) 101: Word-of-mouth in the age of the web-fortified consumer", *New York: Nielsen BuzzMetrics*.

Cawsey, T., and Rowley, J. (2016) "Social media brand building strategies in B2B companies", *Marketing Intelligence & Planning,* Vol. 34, No. 6, pp. 754-776.

Collier, J. E., and Bienstock, C. C. (2006) "How Do Customers Judge Quality in an E-tailer?", *MIT Sloan Management Review,* Vol. 48, No. 1, pp. 35-40.

Dokoohaki, N., and Matskin, M. (2012) "Mining Divergent Opinion Trust Networks through Latent Dirichlet Allocation".

Feng, Y., and Lapata, M. (2010) "Topic models for image annotation and text illustration", Paper read at Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics.

Fong, S., Zhuang, Y., Li, J., and Khoury, R. (2013) "Sentiment Analysis of Online News Using MALLET", Paper read at International Symposium on Computational and Business Intelligence.

Francis, J. E. (2007) "Internet retailing quality: one size does not fit all", *Managing Service Quality: An International Journal,* Vol. 17, No. 3, pp. 341-355.

Griffiths, T. L., and Steyvers, M. (2004) "Finding scientific topics", *Proc Natl Acad Sci U S A,* Vol. 101 Suppl 1, pp. 5228-5235.

Guo, L., Vargo, C. J., Pan, Z., Ding, W., and Ishwar, P. (2016) "Big Social Data Analytics in Journalism and Mass Communication", *Journalism & Mass Communication Quarterly,* Vol. 93, No. 2, pp. 332-359.

Haghighi, A., and Vanderwende, L. (2009) "Exploring content models for multi-document summarization", Paper read at Annual Conference of the North American Chapter of the Association for Computational Linguistics.

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,* Vol. 72, pp. 321-338.

Kaplan, A. M., and Haenlein, M. (2010) "Users of the world, unite! The challenges and opportunities of Social Media", *Business Horizons,* Vol. 53, No. 1, pp. 59-68.

Karimov, F. P., and Brengman, M. (2011) "Adoption of Social Media by Online Retailers", Paper read at International Journal of E-Entrepreneurship and Innovation.

Kim, E. H.-J., 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,* Vol. 42, No. 6, pp. 763-781.

Kumar, V., and Mirchandani, R. (2012) "Increasing the ROI of social media marketing", *MIT Sloan Management Review,* Vol. 54, No. 1, pp. 55.

Li, H., Yan, J., Weihong, H., and Ding, Z. (2014) "Mining User Interest in Microblogs with a User-Topic Model", *China Communications,* Vol. 11, No. 8, pp. 131-144.

Lin, C., and He, Y. (2009) "Joint Sentiment/Topic Model for Sentiment Analysis", Paper read at 18th ACM Conference on Information and Knowledge Management.

Llewellyn, C., Grover, C., Alex, B., Oberlander, J., and Tobin, R. (2015) "Extracting a Topic Specific Dataset from a Twitter Archive", Paper read at Research and Advanced Technology for Digital Libraries: 9th International Conference on Theory and Practice of Digital Libraries.

Luo, J., Ba, S., and Zhang, H. (2012) "The Effectiveness of Online Shopping Characteristics and Well-Designed Websites on Satisfaction", *MIS Quarterly,* Vol. 36, No. 4.

Madani, A., Boussaid, O., and Zegour, D. E. (2015) "Real-time trending topics detection and description from Twitter content", *Social Network Analysis and Mining,* Vol. 5, No. 1, pp. 1-13.
Mangold, W. G., and Faulds, D. J. (2009) "Social media: The new hybrid element of the promotion mix", *Business Horizons,* Vol. 52, No. 4, pp. 357-365.
Nair, M. (2011) "Understanding and Measuring the Value of Social Media", *Journal of Corporate Accounting & Finance,* Vol. 22, No. 3, pp. 45-51.
Ngai, E. W. T., Tao, S. S. C., and Moon, K. K. L. (2015) "Social media research: Theories, constructs, and conceptual frameworks", *International Journal of Information Management,* Vol. 35, No. 1, pp. 33-44.
Picazo, S. (2016). "Customer Service on Twitter and the impact on brands", from https://blog.twitter.com/en-gb/2016/customer-service-on-twitter-and-the-impact-on-brands
Safko, L. (2010). *The social media bible: tactics, tools, and strategies for business success*: John Wiley & Sons.
Titov, I., and McDonald, R. (2008) "Modeling online reviews with multi-grain topic models", Paper read at 17th international conference on World Wide Web.
Yıldırım, A., Üsküdarlı, S., and Özgür, A. (2016) "Identifying Topics in Microblogs Using Wikipedia", *PLoS One*.
Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., and Li, X. (2011) "Comparing Twitter and Traditional Media using Topic Models", Paper read at Advances in Information Retrieval: 33rd European Conference on IR Research.