Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach

Abstract

This study investigates if online reviews (e.g. valence and volume), online promotional strategies (e.g. free delivery and discounts) and sentiments of online reviews can predict product demands. Our data were extracted from Amazon.com, and neural network was employed to examine the relationships between the predictors and product sales. We designed a big data architecture and deployed Node.JS agents for scraping the Amazon.com pages using asynchronous Input/Output calls. The completed web crawling and scraping datasets were then preprocessed for sentimental and neural Network analysis. Our results showed that variables from both online reviews and promotional marketing strategies are important predictors of product demands. Variables in online reviews in general were better predictors as compared to online marketing promotional variables. Our findings led to an understanding of what attributes of online reviews and online promotional marketing can contribute to the prediction of product sales. Our results also showed how online sentiments can moderate the relationships between review valence and review volume with product sales. This study also demonstrated how Big Data architecture, in combination with sentimental and neural network analysis, can facilitate future business researches for predicting product sales in an online environment.

Keywords: Product demands; online reviews; promotional marketing; online marketplace; big data; neural network

1.0 Introduction

One of the key challenges faced by organizations today is the dynamic, global and unpredictable business environment which they operate in. With greater customer expectations on price and quality, manufacturers today can no longer rely solely on cost advantage that they have over their rivals (Chong et al., 2009). Instead, a key strategy for manufacturers today include managing their supply chain efficiently and understanding customer demands better (Chong and Zhou, 2014). Information technologies have helped manufacturers to improve their management of supply chain significantly over the years. Applications such as Enterprise Resource Planning (ERP), B2B websites, and Radio Frequency Identify (RFID) are all able to help organizations manage their supply chain better, and also reduce traditional supply chain challenge such as the Bullwhip Effect (Croom, 2005).

An important aspect of managing supply chain efficiently is to have better forecast on product sales such that a manufacturer will not over or under product products. An emerging area in forecasting the sales of products is in big data and user-generated contents (McKinsey, 2013). Recent marketing reports have all shown the impacts of user-generated contents on the sales of products. Nielsen in a report in 2012 found that online consumer reviews are the second most trusted form of reviews with 70% of consumers surveyed claimed that they trust the platform. Cone Inc in 2011 found that 4 out of 5 consumers reverse purchase decisions as a result of negative online reviews. Luca (2011) also found that a one star increase in popular review site Yelp will lead to a 5-9% increase in revenue. Such user-generated contents have an impact on e-commerce, with social commerce now being one of the most popular e-commerce business model. Companies such as Amazon.com and Alibaba's Taobao.com are examples of

The following publication Chong, A.Y.L., Li, B., Ngai, E.W.T., Ch'ng, E. and Lee, F. (2016), "Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach", International Journal of Operations & Production Management, Vol. 36 No. 4, pp. 358-383 is available at https://doi.org/10.1108/IJOPM-03-2015-0151.

successful social commerce companies that allow potential customers to see recommendations from others before making their purchasing decisions.
Given that user-generated contents play an important role in influencing the purchasing decisions of consumers, their role in helping organizations to understand and forecast product demand should be further investigated. In a social commerce environment, consumers are presented with many information which can influence their purchasing decisions. In Amazon.com for example, these information include product information such as the price of the product, promotional offers related to discounts and free deliveries, and online reviews information such as its valence, volume, and the sentiments of the reviews. As E-commerce storefront and marketplace are now becoming one of the main purchasing channel by consumers, predicting potential customer's purchasing decision and sales performance is now vital to a company's management of its supply chain. Previous demand forecasting techniques include methods such as historical sales data or data from test markets. However, with customers making real time decisions using various information online in the social commerce environment, it is now possible to be predict product sales and customer demands based on data captured online (Duan and Whinston, 2008). These data include product information such as price and descriptions, promotional marketing information such as discounts, and online reviews.

The main objective of this study is to investigate if product demands (or sales) can be predicted through comparative influence of promotional marketing strategies such as discounts and the provision of free delivery options, user generated contents such as volume and valence of online reviews, and the sentiments of the online reviews. Although previous studies have examined the roles of review rating valence and volume of review (Ye et al., 2011), studies examining the roles of sentiments in online reviews remain sparse (Hu et al., 2014). In this study, we designed a series of Big Data algorithms which sit on a Big Data architecture used for Web data and social media analytics (Ch'ng, 2014). The algorithms use asynchronous I/O (input/output) to request, extract and preprocess data in real-time from Amazon.com. After getting the data, the texts of reviews are then processed using an online classifier via text-processing Application Programmer Interface (API). The resulted sentiment is labelled as positive, negative or neutral for further analysis. This study will then use neural network to predict how these variables can collectively be used to predict product sales, as well as the predictive of the interactions effects of the online sentiments on promotional strategies and online reviews. Thus in this study, besides examining the predictive roles of promotional strategies, online reviews and review sentiments, the interplay between these variables will also be examined.

This research has several important contributions. Firstly, although forecasting product demand is an important strategy to manage a company's supply chain, there are limited studies which examine whether social commerce data can be used to predict product demand. Secondly, this study examines the differential and interplay of predictors of product sales such as marketing promotional strategies, online reviews and sentiments. Lastly, we demonstrate how Big Data technologies and architecture are applied to extract data online, in combination with neural networks and sentiment analysis to predict product sales.

2.0 Literature Review

2.1 Online promotional marketing

With advances in the Information System and Technology, new information channels are being increasingly utilized by individuals to access information (Tang, Fang, and Wang, 2014). This easier access to information has several implications especially apparent within the electronic industry. First is the increase in the amount of product information available to consumers' exposure. This means that there are more variables affecting consumers' purchasing decision due to the amount of information on products that are made available to consumers (Floyd et al., 2014). Second, companies are becoming increasingly pressurized to secure sales on their products within a briefer period. This is illustrated by the trend of new product development in the aeronautical industry, automotive industry, as well as electronic products, the one to be used in this study (Tyagi and Sawhney, 2010). A means to achieve this is by turning to an online platform for promotional marketing purposes. One of the most established marketing strategy implemented by vendors are price discounts offerings, which is also applicable in an online environment.

2.1.1 Discount Value

According to Lichtenstein, Netemeyer, and Burton (1990), based on the transaction posit utility theory, products offering with a higher discount rate lead to a higher rate of consumers' demand. This is because discount offerings are perceived by consumers as a bargain, causing consumers to believe that they have received a good deal on that product (ibid.). A study by Gendall et al. (2006) has shown that the popularity of price discounts offering is rooted in the fact that it enables short term, immediate increase in product sales. Furthermore, the quantitatively measurable nature of the price discounts effect enables vendors to guarantee the availability of their products to their consumers (ibid.). However, even though many studies have examined the influence of discount on product demands (Ehrenberg, Hammond, and Goodhardt, 1994; McNeill, 2013), Drozdenko and Jensen (2005) have highlighted contradictories and inconsistencies on the effect of price discounts on product demands. A study by Gupta and Cooper (1992) for example, has suggested that a price discount threshold level of 15% is enough to encourage purchases of the product in the US. They (ibid.) also noted that the saturation point on rate of discount is generally stood at 20% to 30%. This means that products with a price reduction higher than 30% would not necessarily translates to an increase in sales of the product in the US (ibid.). Similar study conducted by Marshall and Leng (2002) have also concluded that while a price reduction of 10% to 50% of the product correlates with higher sales of the product, a further increase to 60% to 70% on the other hand have failed to stimulate further increase in sales. Furthermore, exposure to frequent discount offerings might leads to differing discount rate expectation among consumers from differing regions. Marshall and Leng (2002) have found that Singaporean customers are more sensitive to price reductions compared to their US counterparts, with a lower price discount threshold level of 10% for Singaporean customers compared to 15% for US customers.

Furthermore, price of a product is also used by customers to determine many other variables. Suri, Manchanda, and Kohli (2000) and Marshall and Leng (2002) have noted that price information of a product may be used as a cue to determine its inherent perceived quality by customers. Attribution theory posits that consumers rely on price and other attributional information when processing information to come with a final evaluation of the product's quality (Drozdenko and Jensen 2005). This attributional information on the product may result in customers' final evaluation of the product's discount as being perceived to be a monetary gain to them. Attributional information in a form of external price references for a product has shown that there is also a difference of the effect of that external price references on products offered physically or through online stores (Jensen et al. 2003). These contradictory findings however, have yet to be verified in an online setting. Price reduction by itself is still useful to be utilized, especially for vendors trying to offer substitutes of a main product that might be low on stock. However, due to the contradictory results of the literatures on this subject, there needs to be other accompanying variables to explain and better predict sales of products.

2.1.2 Discount Rate

Ratio of discount in comparison to the actual price of the product is another variable that are used in researches on this subject. Chen, Monroe, and Lou (1998) have argued that consumers' absolute and relative perception of price may influence their perception to price reductions. Based on the psychophysics-of-price-heuristics theory, psychological utility of consumers derived from saving a fixed amount of money is inversely related to a product's price (ibid.). As of now we have yet to find a previous study comparing an absolute discount and a relative discount. Does a 35$ savings appeal to more consumers compared to a 10% price reduction offers on a 350$ coat? Which one of these offers may better predicts product's demand by customers?

2.1.3 Free delivery

Another important variable likely to predict online product purchases is the availability of a free delivery option. Yip and Law (2002) have noted that online stores that offer, other than special price reduction regime, also free delivery services tend to attract more online users. Furthermore, a study conducted by Doern and Fey (2006) on electronic commerce in Russia has confirmed that free delivery service offerings have a positive relationship with customers' trust and loyalty. It is important to note however, that this offering might not be beneficial in some instances, as shown by a study conducted by Smith and Rupp (2003). They have found that online companies such as kozmo.com and urbanfetch.com have ended with failures due to their move to offer free delivery incentives to their customer base. Because of the fact that once these companies removed their free delivery offerings, it was becoming increasingly difficult to retain their customer base. This is to be expected as electronic commerce is becoming more mature, consumers utilizing this technology would increase their standard expectation. This means that free delivery that is used to be something perceived as an additional incentives for customers, is becoming increasingly normalized, leading to changes in the way consumers perceive promotions offered by electronic vendors. Free delivery becomes a service that is to be expected to be offered by electronic vendors, and a failure to offer this service might translates to a lost in customer base. In accordance with these findings, free delivery is utilized as one of the variables used to predict in consumers' product demands. One of particular interest is on how this variable will plays out when combined with other variables chosen for this study.

2.2. Online Reviews

Because of the continuous growth of online media, users are seen to be more active and regular in giving out and sharing opinions on products and services with each other through several platforms like blogs, product reviews, wikis and twitter (Tirunillai and Tellis 2012). This particular information could be defined as Users Generated Content (UGC). UGC "refers to media content created by users to share information and/or opinions with other users" (Tang, Fang, and Wang, 2014). In comparison with the traditional methods of advertisement like television and newspapers advertisements, electronic UGC is perceived by potential customers to be more reliable, balanced, and neutral than those provided through private channels (one-way company advertisements for example) (Davis and Khazanchi 2008; Lee, Park, and Han 2008; Mudambi and Schuff 2010; Senecal and Nantel 2004). For example, Lu et al. (2013) have noted that in e-UGC both favorable and unfavorable information are segregated and made available for information seekers to read. Furthermore, these e-UGC are more readily available throughout the internet (Bakos and Dellarocas, 2011; Davis and Khazanchi 2008; Duan, Gu, and Whinston, 2008). Furthermore e-UGC enables customers to obtain more elaborate information on a product (Floyd, 2014). Many B2C websites offer video and photo upload features on their review sections.

Previous studies have examined the role of e-UGC in influencing the sales of products and services (Chen, Dhanasobhon, and Smith 2008; Cui, Lui, and Guo 2012; Chevalier and Mayzlin 2006; Clemons, Gao, and Hitt 2006; Duan, Gu, and Whinston 2008; Ghose and Ipeirotis 2006). However, among past literatures, the most commonly examined variables of e-UGC are usually the valence, volume, and rate of dispersion of reviews (Lu et al. 2013). Chevalier and Mayzlin (2006) for example, have noted that the volume and valence of online reviews may be able to influence sales rank of certain books in amazon.com and barnesandnoble.com. Lu et al. (2013) have noted that based on the data that they have collected on online reviews in a restaurant review website; they concluded that there is a statistically measurable and significant relationship between the valence and volume of the reviews and sales of (restaurant) products sold. Basing on previous literatures, this paper will also utilize the valence of online review as one of the variable measured.

2.2.1 Online review valence (average rating)

Online review valence refers to the nature of e-UGC content, which is usually presented in the form of evaluation score, which can be positive, negative or mixed (Yang et al., 2012). Although previous research have find a persuasive effect of online review valance on consumer purchase decision making (Cheung and Thadani, 2012), there is no consensus about the effect of online review valence on sales. The studies of Lu et al. (2013) and Chevalier and Mayzlin (2006) both support the impacting power of online review valance on product sales. Dellarocas, Awad, and Zhang (2004) have also found a strong predicting role of online review valence on product sales. On the other hand, Liu (2006) found that compared to volume, online review valance shows a weaker explanation power of the aggregation of weekly box office revenue. Similarly, Davis and Khazanchi (2008) found that online review valance cannot impact product sales in their study of multi-layered product categories online. However, Chevalier and Mayzlin (2006) show a mixed result of the effects of online valence on online bookstore sales. There is significant online review valance effect on sales for Amazon.com but little effect for barnesandnoble.com. Duan, Gu, and Whinston (2008) found that online review valence has indirect impact on box office revenue by significantly influence online review volume.

There are conflicting explanations for the low explanatory power of online review valance on sales. By analyzing the heterogeneous characteristics of products in the movie market, Yang et al. (2012) find that the effect of online review on sales can be diluted by the more diverse marketing communication channels and higher marketing budgets. Zhu and Zhang (2010) found that the impact of online review valance on sales in moderated by product and consumer characteristics, which is supported by psychology literature that highlights the moderating role of context and environment on the effectiveness of an influencer. Online review valance also has different predicting power on the sales of experience products, such as books and movies than search products like electronics (Cui, Lui, and Guo 2012). Unlike experience products where extrinsic attributed related cues are crucial in decision-making processes of potential customers (ibid.), potential customers tend to utilize different variables (technical aspects, specifications, and performance in the case of consumer electronics for example) when assessing search products. Furthermore, in an online environment, such information is readily available and could be easily accessed by anyone (ibid.) Thus, extrapolating from the previous literatures, online review valence is considered as a predictor to consumer electronic products sales in this paper.

2.2.2 Online review volume

Online review volume refers to the number of reviews for either a product or a service (Lu et al., 2013). Previous literatures have shown that online review volume has a quantitatively measurable impact on product sales (Liu, 2006; Duan, Gu and Whinston 2008; Davis and Khazanchi 2008; Lu et al. 2013). It is believed that the increased consumer product awareness caused by online review volume leads to higher sales (Anderson and Salisbury 2003; Bowman and Narayandas 2001; Chen, Wu, and Yoon 2004; Godes and Mayzlin 2004; Liu 2006; Van den Bulte and Lilien 2001; Yang et al., 2012). For example, Lu et al. (2013) found that online review volume has led to an increase in restaurant sales. Both theoretical and practical evidence supports the predicting power of online review volume on sales (Cheung and Thadani, 2012; Duan, Gu and Whinston 2008). However, Cui, Lui and Guo (2012) found that online review volume is more effective in predicting the sales of experience products than search products. For experience goods, online review volume has a strong prediction power because it can reflect the extrinsic cues such as product popularity (ibid.). However, for electronic products where consumers can experience the actual product's attributes, online review valance plays more important role in predicting sales than online review volume (ibid.). Based on previous literatures, it is clear that online review volume is an important predictor to product sales. Thus we decided to include volume (along with valence) of online reviews as variables in predicting sales of electronic products.

2.2.3 Percentage of negative online reviews

It is generally agreed that positive UGC tend to boost purchase behavior of consumers while negative UGC is likely to discourage purchases (East, Hammond, and Lomax, 2008). However, studies have shown that not only negative online e-UGC is spread at a faster pace compared to positive e-UGC (Cui, Lui, and Guo, 2012), but it may also have a more pronounced impact on customers' decision to purchase a product (Cheung and Thadani, 2012; Lee, Park, and Han 2008). Cui, Lui, and Guo (2012) have further noted that while a positive e-UGC generally reflects a product's positive attributes such as good quality and brand image, a negative e-UGC generally reflects consumers' lack of confidence on the attributes (quality and brand image for example) of the product. This negative e-UGC in particular may negatively affect sales of a product (Sonnier, McAlister, and Rutz, 2011). This is why the proportion of positive and negative e-UGC in the form of product reviews may also influence consumers' decision to purchase a product, along with other variables such as product rating. Furthermore, Ito et al.

(1998) have posited that positive information may not have a significant impact on evaluations in comparison to negative information. This is why in the decision-making process; negative information ultimately trumps positive information in terms of influence (Lee, Park, and Han, 2008).

A case study on how movies are evaluated based upon the type of e-UGC has shown that positive e-UGC has a considerably less influence on end evaluations of the movies compared to negative e-UGC (Chakravarty, Liu, and Mazumdar, 2010). Individuals, particularly those that are not frequent moviegoers are still affected by negative e-UGC even in the presence of positive reviews given by movie critics (ibid.). Another case study examining software programs usage have also shown that negative e-UGC are more successful in preventing consumption of a software compared to the mitigating effect of positive e-UGC in consumption prevention (Zhang, Craciun, and Shun, 2010). Chevalier and Mayzlin (2006) have also found that negative reviews have a more pronounced impact on sales of books in amazon.com and barnesandnoble.com compared to positive reviews. Thus, extrapolating from these previous studies percentage of positive reviews will be examined in this study to establish a comparative analysis and to examine whether negative e-UGC in the context of electronic products also have a more pronounced impact in the sales of products compared to positive e-UGC.

2.2.4 Number of Customer Questions Answered

In Amazon.com, there are Customer Questions and Answers that allow customers to ask and answer each other's questions. This extends the interaction between customers beyond merely posting and reading of reviews. Previous research found that high social interactions will increase the trust and sales in social commerce context (Ng, 2013; Ou, Pavlou, and Davison, 2014). Therefore, we consider the number of answered questions as an indicator of the interactivity level, and include it as a predictor of sales.

2.2.5 Text sentiment of online reviews

Review sentiment is also an indicator of the valence in online reviews. However, different from the numeric rating, sentiment is a qualitative feature of the text. Although the numeric rating of review is often used to reflect the valence of review (e.g. Ye, Law, Gu and Chen, 2011; Chevlier and Mayzlin, 2006), in e-UGC communication the positivity and negativity of a UGC is fundamentally determined by the its actual content. This highlights the importance to study the text sentiment of an online review.

It is commonly assumed that text sentiment of online reviews is consistent with and thus replaceable by numeric review rating (Hu et al, 2014). However, a large proportion of reviews tend to have either extremely high numeric ratings or extremely low ones (Hu, Pavlou and Zhang, 2006). Such bimodality makes the review rating difficult to reflect the true product quality, and may undermine its importance as a determinant to purchasing decision of a prospective buyer (Ghose and Ipeirotis, 2011). Moreover, customers tend to read not only review ratings but also review texts before making purchase decisions (Chevalier and Mayzlin, 2006). Such text sentiment can be seen as unique type of cognitive appraisals from previous customers, and such appraisals provide a useful information set for prospective consumer's cognitive processing (Hu, et al. 2014; Yin, et al. 2014). Therefore, apart from the online review rating, text sentiment of an online review is also an important factor that may influence purchase decisions. The review with positive sentiment calculated from its text can be seen as a piece of positive e-UGC, while the one with negative sentiment as a negative e-UGC. Such valence of UGC has strong influence in product sales.

Apart from influencing purchase decisions through cognitive process, review sentiment may also influence purchase behavior via emotional contagion. Emotion contagion can be understood as a phenomenon in which ‘exposure to an individual expressing positive or negative emotion can produce a corresponding change in the emotional state of the observer’ (Pugh, 2001, pp.1020). Traditional point of view is that such contagion effect is more obvious in face-to-face communication (e.g. Hatfield and Cacioppo, 1994). However, recent research indicates that emotional contagion does occur in computer-mediated communication via text (Kramer, et al. 2014; Hancock, et al. 2008). Marketing studies suggest that emotions of customers significantly affect their purchasing behavior, with positive emotions associated with actual purchasing of a product (Tsai, 2001).

Based on previous literatures, this study seeks to distinguish online review text sentiment from online review rating, and consider text sentiment as a predictor of product sales. Specifically, Amazon.com offers the peer evaluation of review helpfulness for each review, and we choose the most helpful reviews that are listed on the product information page as the object of sentiment analysis (Hu, et al., 2014). Those reviews are targeted because they tend to have more influence on customers than others reviews. First, such helpfulness rating is perceived as a criteria to filter high quality reviews, and customers are more likely to actively select the most helpful reviews to refer to (Mudambi and Schuff, 2010). Second, the most helpful reviews are listed on the first page, and is has higher exposure to customers. Therefore, if taking the sentiment of all reviews into consideration, the power of sentiment might be diluted. Thus, we only conduct sentiment analysis to those most helpful reviews.

2.3 Interactions between online reviews and online promotional marketing (discount rate)

Internet is becoming increasingly integrated into everyday life. There is more information that individuals are able to access to today, compared to on any point of time in history. This however, has an unexpected consequence of information overload. The amount of online information that are present today pose a challenge for customers in the decision making process on purchasing decision (Chong and Ngai, 2013). This can also be seen by the fact that increasingly, vendors are likely to adopt multi-dimensional approach to market their products in an online setting, as opposed to using a single-dimensional marketing strategy approach that are more common in the brick and mortar store (Lu et al., 2013). This however, leads to the question of efficacy between differing marketing approaches to predict product sales. Is UGC in the form of online reviews better predicts potential sales count for a product? Or does a one-way marketing effort by a company be a better predictor?

Lu et al. (2013) have examined the role of online UGC in moderating the online marketing promotions effect. It was however, focused on products that have inherent differences in terms of characteristics (products offered at restaurants as opposed to electronic products studied in this paper). Thus it is still unclear as whether price reductions and online UGC may affect sales outcomes of electronic products the same as they affect restaurant sales. One difference that is of particular significance is the shorter product life cycle of electronics. Consumers are prone to look for the latest release of a specific electronic product, as opposed to other products. It is also still not well known whether new product linings would directly affect sales of older product linings. In consumer electronics specifically, as newer products tend to offer more features in terms of performance, specifications, and even design, how would older products still be relevant to customers? Also of a considerable importance is the fact that previous findings on discount has offered contradictory results on how product sales, as discounts in certain instances might instead be perceived as a negative quality of that product (products with high discounts are perceived to be of lower quality) (Marshall and Leng, 2002; Suri, Manchanda, and Kohli, 2000). Furthermore findings on restaurant sales (Lu et al., 2013) have shown that when online UGC volumes are high, coupon promotions become redundant and ineffective in predicting product sales. Extrapolating from these previous finding, we are trying to examine how specific attributes of online reviews (volume, valence, rate of positive comments) interacts with other cues such as price reductions, discount rate, and free delivery. Can these constructs better predict sales of electronic products?

2.4 Interactions between sentiments and valence and volume

Previous studies have shown that online reviews (Duan et al., 2008) and sentiments (Yu et al., 2012) can individually influence the sales of products online. In social commerce sites such as Amazon.com, both online review valence and volume are measured by the ratings of the products and the number of reviews. Studies in the past have examined the impact of review valence and volume on product sales (Liu, 2006; Duan and Gu, 2008). Besides reviews, an important influence on product sales is the review sentiments. However, there has been limited study on the roles of review sentiments on product sales due to the difficulty in doing so (Hu et al., 2014).

In Hu et al. (2014)'s study, they found that online reviews do not have a direct influence on product sales, but instead, have an indirect impact through review sentiments. However, it should be noted that Hu et al. (2014)'s study is based on experience product, namely books. Cui et al. (2012) in their study on the impact of online reviews on product sales, found that valence of reviews is actually stronger in search products such as electronics. As found by Cui et al. (2012), search goods can be evaluated by instrumental evaluate cues based on multitude of information such as product attributes, functions, and performances on the online e-commerce website. Search goods usually also have their ratings prominently displayed. Thus how the role of online reviews, such as valence and volume can be influenced by online sentiments in predicting search goods sales remain under explored. In our study, most helpful reviews sentiments are being used instead of sentiments of all reviews. This is because on the product page, most helpful reviews are being shown, and most users will go through the most helpful reviews instead of all the product reviews (Hu et al., 2014). Thus, although it is possible that a product has high rating, the most helpful reviews sentiments which are positive and negative can alter the review rating's influence on product sales.

Past studies on the relationships between volume of reviews and product sales do not have consistent results (Dellarocas et al., 2007). However, consumers may not entirely trust the online reviews when the numbers of reviews are too few for them to check for consistency and to draw conclusions. In such scenario, it is possible that when review volume are low, their impacts can be strengthened or weakened by the sentiments of the reviews. However, to the best of our knowledge, there is so far no study that examine how the interactions between review sentiments and online review volumes can predict sales of products.

3.0 Neural network

A popular machine learning technique that is inspired by the human brain is neural network (Chiang, Zhang, and Zhou 2006). In a neural network model, the networks are presented as systems of interconnected neurons which can compute values from input information. A neural network can learn the intrinsic nature of patterns or processes from data sets (Sim et al., 2014). In recent studies, neural network have shown to be an effective alternative to traditional statistical techniques (Chong, 2013). In particular, neural network can offers better prediction than traditional regression approaches, and is suitable to be employed to test large scale data such as in a big data environment (George et al., 2014). Previous studies have shown that neural network technique such as the multilayer perceptron (MLP) can be trained to approximate most smooth, measureable function (Gardner and Dorling, 1998). Furthermore, the MLP can also model non-linear functions, and can be trained to accurately generalize new and unforeseen data (Gardner and Dorling, 1998).

MLP consists of a system of interconnected nodes distributed in three hierarchic layers (i.e. input, hidden and output). The input node receives the input data (i.e. predictors) while the output layer generates the final information. The output unit is the functions of the hidden units. The hidden layers are between the input and output layers, and it contains unobservable nodes. The value of each hidden unit is some functions of the input units (SPSS, 2011). In MLP, two hidden layers is allowed. The hidden layers will receive inputs from neurons in the input layer, and knowledge is then stored by the interneuron connection strengths (i.e. synaptic weights) (Haykin 1994). With appropriate supervised learning algorithm, the MLP will analyse the data sets, and the synaptic weights of the neural weight will be adjusted to attain the desired design objective (Chong et al. 2013). They are then used to store knowledge and make it available for future use (Sim et al., 2014).

Figure 1 provides an example of MLP.

<<Figure 1 about here>>

In this study, MLP is applied to predict the factors influencing the product sales. Although studies in the past applied explanatory statistical techniques to examine their research models, there is an increase in recent studies that apply predictive analytic approaches such as MLP in information systems researches (Shmueli and Koppius 2011; Lu et al. 2013). A key advantage of MLP is that it can offer useful and practical model which can help researchers to develop new theory (Shmueli and Koppius 2011). It can also overcome the challenges faced by traditional statistical analysis relying on p-value which may not be effective in environment with large data sets (e.g. false correlations) (George, Hass, and Pentland 2014).

4.0 Methodology

4.1 Research Context and data

In this study, we adopted approaches from Hu et al. (2014) and used Amazon.com as the source of our data. Search products, specifically electronic devices such as camera, television, Hi-Fi, notebook etc. are used in this study. We have chosen electronic products for our study as they have shorter product shelf lives, and it would be interesting to see how the relationships between the predictors used in our study influence product sales (Chong and Ooi 2008). Similar to another study (Lu et al. 2013), we used Amazon.com solely as we are unable to include dispersion in our study. Dispersion of eWOM is defined as the extent in which conversation on a product or service is being carried out across broad range of communities. We initially randomly selected 40,000 electronic products to gather sales and review information. Out of these 40,000 products, only 12,000 have text reviews which we could capture. We collected review information such as total number of reviews (volume), average rating of the product (valence), percentage of positive reviews, percentage of negative reviews and number of answered questions. For online promotional marketing variables, we collected free delivery, discount rate, and discount value. Sentiments were collected for the most helpful reviews. The selection of sentiments from the most helpful review is similar to previous studies (e.g. Hu et al., 2014), and potential buyers most often read such reviews instead of all the reviews available for a product. The demand of the product (i.e. product sales) is measured using sales rank. This is consistent with the approaches by previous studies by Godes and Mayzlin (2004) and Hu et al. (2014).

4.2 Big Data Technology

When selecting a suite of technologies to facilitate research, our aim is to lay a generic technical foundation prior to specialising the system for particular purposes. We first set forth the requirements in terms of the volume and velocity of data, with the potential to scale up when needs arise. Our project aims to access, manage and process tens of thousands of web page contents, including cleaning data in real-time. As opposed to a conventional desktop and network connections, our specific requirements suggest that scalable technology is required. Big Data technology has become a necessity for research in the 21st century, and we lay a foundation here for the purpose of making the data aspect of our research possible.

The system that we used is illustrated in Figure #. The system sits within our Web and Social Media Big Data client-server architecture, integrating various open-source server technologies (Ch'ng, 2014) used by large corporations (e.g., LinkedIn, Paypal, Yahoo, etc.). The system consists of 6x Linux Ubuntu 64bit Virtual Machines (VM) on 2x HP DL388p physical servers. The physical system scalable horizontally as needs arises with additional VMs. Due to the asynchronous I/O nature of our algorithms, as data comes in, they are efficiently stored within MongoDB, a cross-platform NoSQL database that is horizontally scalable prior to synchronous extraction later into CSV (Comma Separated Value) file for our Neural Network analysis. Our asynchronous algorithms are coded in server-side JavaScript via Node.js. Node.js is an event-driven, non-blocking I/O model built on Google's V8 JavaScript engine. The capability of the asynchronous I/O algorithms made it possible to manage real-time, data-intensive applications such as what we have in the present research. Our algorithms were developed and deployed on a Dell T3600 Tower Workstation with 64GB of RAM, 6 cores 12 threads, and two GPGPU cards: Quadro K4000 and Tesla K40c. The Tesla K40c was prepared for parallel processing needs, however, it was not utilised as the data was not sufficiently large to require multicore processing.

The process of our technical methodology is in Figure 2:

1. Developmental workstation where our Node.JS agents are deployed for scraping the web using asynchronous I/O calls.
2. Physical server hosting Ubuntu 64bit virtual machines, and where data is stored and horizontally scaled.
3. Completed Web crawling and scraping datasets are converted into comma separated values for Neural Network analysis.

In (1) we developed a series of asynchronous I/O algorithms, which helped us to acquire and pre-process raw Amazon.com data. The algorithms take an input file containing lines of product listings before crawling the pages by following all the paging links. After all the product pages associated with the listings are obtained, asynchronous agents hosted on our web server is deployed to scrape, in real-time, the Amazon websites using JavaScripts' DOM (Document Object Model) and processes the scattered HTML tags where our target information is embedded into structured key-value pair dataset. Regular Expressions are used for specific character data patterns such as numbers and keywords. Incoming data are immediately stored within the horizontally scalable MongoDB server (2).

Our asynchronous code is capable of sending thousands of concurrent sockets where agents requests for Amazon.com pages, however, to prevent our IP from being blocked, a recursive mechanism were implemented so that we can control *n* requests per set. It takes on average 1.1392 seconds to call a HTTP request, obtain a HTML response and scrap the page of all required data. We extracted all electronics data there are on the Amazon pages, we could have continued data scraping with our highly efficient system. Finally, a CSV file is generated when the scraper agents have completed their jobs (3).

<<Figure 2 about here>>

The total number of records in our study is 35,203. Our sample includes 813 Audio and HiFi devices, 23716 Camera and Photo related devices, 9870 Computers and Accessories, 264 Television and Home Cinema, 92 outdoor and sports related electronic devices, and 448 other electronic devices.

Table 1 provides the summary of variables used for this study and their descriptions. 15 variables, as well as 3 interactions effects (e.g. Positive Review X Discount Rate, Valence X Discount Rate and Volume X Discount Rate) are used as predictors of this study.

<<Table 1 about here>>

4.3 Sentiment Analysis

We performed sentiment analysis to identify the sentiment of online review text. Sentiment analysis is the computational treatment to classify reviews into positive and negative polarity (Pang and Lee, 2008). It is increasingly popular in e-commerce, and is often used to understand customers' sentiment and opinions embedded in online reviews and UGC (Archak, et al. 2011, Pang and Lee, 2008).

In our study, we conducted sentiment analysis to the most helpful reviews that are shown on the first page of product information. All the reviews on the same product page are analysed as one single object. This is because people tend to go through many reviews in the webpage before they construct their cognitive and affective perception about the product, and those reviews are often processed as an entirety though they were posted separately. For each object, we wrote Python script to do HTTP POST to Natural Language Processing Application Programming Interface (API) and the returned response contained the probabilities of different sentiment classes and final label. Based on the probabilities, sentiment was labelled as either positive, negative or neutral using hierarchical classification, and was coded as 1, -1 and 0 correspondingly. In the classification process, the neutrality is identified first. The probability of neutrality is ranged from 0 to 1. If the probability of labelling as neutral is greater than 0.5, the object (i.e. text of reviews) will be labelled as ‘neutral’. However, if the probability of neutrality is smaller than 0.5, the texts are classified as not neutral, and then we will determine which polarity the sentiment is. The sum of probabilities for positive label and negative is 1, and the one with higher probability will be used to label the review sentiment. NLTK trainer was used to train the classifiers, and the training data are several data sets from Bo Pang and Lillian Lee. This process of sentiment analysis is shown in Figure 3.

<< Figure 3 about here >>

4.4 Neural Network Analysis

A three-layered neural network which consists of a layer for input nodes, hidden nodes, and output nodes each (Garson, 1998) is developed for this study. According to Chiang, Zhang, and Zhou (2006), within the field of e-commerce, back-propagation neural network is the most commonly used networks. Drawing from previous studies, the data analysing on this study is also done using back-propagation neural network methods. Initial weights and biases will be given values between 0 and 1. Training data with sets of inputs (i.e. discount value, discount rate, valence of reviews, etc) and output (sales rank) are then provided for the neural network. The model in this paper is illustrated in Figure 4.

The difference between the actual output (e.g. sales rank) and the desired output will be calculated and back-propagated to the previous layers (Chong et al. 2013). The neural network applies the Delta rule to adjust the connection weight and reduces the output errors. This process is then back-propagated to the previous layer until it reaches the input layer (Chiang, Zhang, and Zhou 2006).

4.5 Validations of Neural Networks

We applied multilayer perceptron training algorithm to train the neural network in this study. Similar to existing studies (Chong et al. 2013), cross validations were conducted to avoid over-fitting of the model. In order to determine the ideal number of hidden nodes, we increase the hidden nodes starting from 1, and increase the number of hidden node by one and check this against the errors in the neural network. The ideal number of hidden network is one which does not increase the neural network’s errors (Chong et al. 2013).

Networks with four hidden nodes were found to be complex enough to map the datasets without incurring additional errors to the neural network model. Our neural network therefore consists of 14 predictors, six hidden nodes, and one output variable.

The activation function for the hidden and output layers used in this study is the sigmoid function. The sigmoid function approaches the value of one for large positive numbers and 0.5 for zero and very close to zero for large negative numbers (Sim et al. 2014). As a result, it allows transition between the low and high output of the neurons. A ten-fold cross validation was performed whereby we used 90 percent of the data to train the neural network, while the remaining 10 percent was used to measure the prediction accuracy of the trained network. Root Mean Squared Error (RMSE) was computed to compare data from the training and testing sets to ensure that there are not much difference between the two tables. The RMSE of the validations are shown in Table 2.

<<Table 2 about here>>

From Table 3, the average RMSE for the training model is 0.00102 while the testing model is 0.0103. The RMSE values for the two models are relatively consistent and do not vary much, and we can therefore be confident that the network model is reliable in capturing the numeric relations between the predictors and outputs.

4.6 Sensitivity analysis

The importance for predictors in this study was calculated using sensitivity analysis. Sensitivity analysis performance was calculated by averaging the importance of the predictors over ten networks (Chong et al. 2013). The importance of the predictor variable is a measure of how much the network's model-predicted value changes for different values of the predictor variable (Chong et al. 2013). The importance was calculated by average the predictors' importance over ten networks and expressed as a percentage (Chong et al. 2013).

<<Table 3 about here>>

Table 3 shows that all 14 predictors are found to be relevant to all ten networks. The average result showed that the two main and most important predictors are volume*discount rate and number of answered questions. The result showed that in general, eWOM related variables such as positive reviews, negative reviews, valence, volume, rating of the most helpful favourable review, and rating of the most helpful critical review are better predictors of electronic product sales than online promotion strategies such as discounts and free deliveries. However, interaction effects between discount rate with volume and positive review are important predictors of electronic sales. We also conducted correlation analysis for the predictors and sales, and found that besides negative reviews which have a negative relationship with sales rank, all other predictors have positive relationship with sales rank.

5.0 Discussions

The result of our analysis has confirmed all variables that we hypothesized to be able to predict sales rank of consumer electronic products in an online setting. Some variables appear to be having a more significant influence than others. Predictors such as Percentage of Negative Reviews, Number of Questions, Volume, Current Price, and so on, have a significant impact on potential product sales. Some results that might be of particular importance however, are actually the interaction effects between two measured variables used in this study. Interaction between sentiment and volume and discount rate and volume for example, strongly influence sales rank. In Duan et al. (2008) and Davis and Khazanchi (2008)'s study, volume of e-UGC is a good predictor of product sales in the movie industry. While we also found a similar result in our study (volume being an important predictor is confirmed), we have also confirmed that additionally, volume of e-UGC could better predict the sales of search product (consumer electronics) when it interacts with Sentiment and Discount Rate. We have confirmed that interactions between Sentiment and Volume, and Discount Rate and Volume, could better predict sales of product, compared to Volume alone. This means that an increase in sales rank of consumer electronic products could be better reached if we introduce a discount offer to products that already have a high volume of online reviews, rather than relying on the volume alone. This result contradicts Lu et al. (2013)'s study on sales of restaurants whereby discount may not be able to influence sales on product with a high volume of e-UGC. It is clear that this plays differently for consumer electronic products, whereby discount is likely to be well-received by customers, which would subsequently lead to higher sales rank.

Chevalier and Mayzlin (2006) have noted that based on their review-length data on book sales, the actual content of the online review (the sentiment of the text) are able to better predict book sales when compared to the star rating (valence) of the product. However, our study has shown that this is not the case in the context of search products. Our result has shown that valence of the review is twice as strong of a predictor to product sales compared to text sentiment. These comparisons have shown that there are differing sales predictor variables for search product and experience product. However it is also important to note that while sentiment alone could not be a good predictor to product sales, its effect is tremendously amplified when it's combined with the volume and valence of the reviews.

Previous literatures offer inconclusive results on the differing importance between volume and valence in predicting product sales. Although both have been confirmed to be an important parameters in predicting sales of product, studies are divided as to which variable is deemed to be more important. Is valence more important than volume in predicting product sales (or vice versa)? However, based on our study on search product (consumer electronics), volume is able to better predict product sales. Volume is twice as strong of a predictor in sales compared to valence. Other than volume and valence, customer Q&A has also been proven to be very important in predicting sales. This might be because when customers perceive their peer assessment to be more neutral compared to the information advertised by the vendors.

Although our study has confirmed all of our previous hypotheses, there are several variables that appear to have only a weak effect on product sales. We found out that free delivery and sentiments have only a minor influence on sales rank in electronic products. Basing on the result of our study, it is probably more appropriate for business practitioners to focus their marketing efforts on strategies related to price reductions, increased online presence (through increased volume and valence of e-UGC), and providing relevant customer service (for example providing a platform by which potential customers are able to inquire on vendors in regards to their products) to further increased their sales rank of products.

6.0 Conclusions and Implications

The rich information embedded in social commerce websites has attracted increasing attention. In this study, we employed Big Data architecture, sentiment analysis and three-layered neural network modelling to examine predictors on sales rank of product (consumer electronics) in social commerce. Our analysis of the collected data has confirmed that all proposed predictors are influential, and promotional marketing strategies and social interactions such as online review and answered questions are both important for influencing sales. Many variables confirmed previous studies on their roles in predicting sales, such as review valence and volume, albeit we also have some unique findings on specific variables such as discount rate (in contrast with Lu et al (2013)'s study) and sentiment (only minor predictor and insignificant when compared to valence as opposed to Chevalier and Mayzlin (2006)'s study) as a stand-alone variable in sales rating prediction. Some predictors seem to play a more important role compared to others. Of particular interests is the role of sentiment, especially on how it interacts when combined with other variables. Our study has shown that sentiment has a significant interaction with volume and valence of online review and could significantly affect and predict product sales. Drawing from these results, we concluded several important implications.

First, this paper has demonstrated that large amount of datasets could be efficiently extracted using Big Data architecture. By employing a special set of asynchronous algorithms, we are able to extract samples and pre-process them real-time. This large amount of sample enabled us to more accurately predict online products' demand of a consumer, due to the more generality associated with larger sample sizes. Although we have only extracted samples from one website (amazon.com), this utilisation of Big Data architecture can be extended to a larger scale, such as extracting connected yet dispersed social media data. The capability of extracting large amount of real-time data makes longitudinal research possible. In addition, this paper avoids p-value in traditional statistics, considering that it works less effectively in big data context (George, Haas, and Pentland 2014). Instead, we make use of artificial neural network to examine factors’ predicting power. Neural network and other data mining and machine learning techniques are recommended when using big data to examine business theories (George, Hass and Pentland, 2014)

Second, we have demonstrated and discussed the influence of various marketing promotional strategies and e-UGC on product sales rank, and our findings provides significant guidance on e-commerce management. Our research results confirmed the importance of eUGC-based social interactions on e-commerce website, and suggest that online vendors should pay attention to online interactivity, not only customer reviews but also customer Q&A. Some sellers have adopted certain review-stimulating strategies, while little has been done on managing customer Q&A. One possible way that online seller can use to manage Q&A is to answer customer’s questions, especially when the question is too technical for other customers to answer. Another possible way to increase level of interactivity is launch apply instant messaging service so that customers can more easily interacted with customers and online sellers.

Third, our research showed high importance of negative and positive review percentage, interaction effects between sentiment and volume and valence on sales rank. This suggests that the perceived value of polarised reviews, and the separated rating number should also be displayed in obvious position. Also, sentiment of review is an important factor when interact with other review features, and thus companies should pay their attention in analysing their review text sentiment to understand their customer’s opinion, rather than merely monitoring overall numeric rating.

Fourth, our result shows that online marketing promotional strategies such as discounts should be employed together with online reviews to improve sales. Price reduction strategies such as discounts and free deliveries do not strongly influence customers' purchasing decisions. Therefore an integrated strategies taking into considerations of online reviews and discounts will be more successful in increasing product sales. Lastly, this study shows that manufacturers can predict their product demands via online marketplaces. E-commerce is now one of the main channels for selling products, and traditionally, companies focused on improving product forecasts and reducing bullwhip effects by examining data from their physical storefronts. With e-commerce increasingly becoming the main channels of product sales, manufacturers should also examine the forecast of their product demands through online marketplace. However, unlike traditional physical storefronts, online marketplace are also influenced by online reviews. As shown by this study, online reviews service as an important predictors of product sales online, and manufacturers should take account of this when planning for their productions.

7.0 Limitations and future directions

Despite contributions that this paper has contributed, there are several limitations worth noting. Firstly we only applied our hypotheses in the context of search product sales (specifically electronic products). One main reason as to why we decided to choose electronic products as our sample is due to the limited amount study that has been done in this area on search products and specifically electronic products. Our results have shown that there are several fundamental differences on how different predictors play out in electronic products, as opposed to several previous studies which focused on mainly experience product. Thus, future research in this study could further contribute by incorporating more products and examine whether our model is applicable to non-electronic product types. Another possible limitation of this study might be the size and source of our sample. As we have applied big data architecture for our data mining purposes, it is hoped that further research directions would be able to conduct studies with similar approach that also utilize a larger sample to further confirm the generality of this study. In terms of sample source, we decided to use amazon.com as our primary website by which samples are collected. Our reason for this is because amazon is the largest e-commerce website with the largest review community on the internet with a very diverse and varied amount of products that are available (Garcia et al. 2011) and thus it is suitable to be made as our sample source where big data architecture is implemented. It is hoped that future researches could incorporate other similar websites and utilize this study as a foundation to examine further and confirm our findings.

Acknowledgement

"The authors acknowledges the financial support from the National Natural Science Foundation of China (NSFC), International Doctoral Innovation Centre, Ningbo Education Bureau, Ningbo Science and Technology Bureau, China's MoST and The University of Nottingham. The project is partially supported by NSFC no. 71402076 and NBSTB Project 2012B10055"

References

Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. "Twitter mood predicts the stock market." Review of. *Journal of Computational Science* 2 (1):1-8. doi: http://dx.doi.org/10.1016/j.jocs.2010.12.007.

Bakos, Yannis and Chrysanthos Dellarocas. 2011."Cooperation without Enforcement? A Comparative Analysis of Litigation and Online Reputation as Quality Assurance Mechanisms." *Management Science* 57:1944–62.

Cao, Q., and M. J. Schniederjans. 2004. "Empirical study of the relationship between operations strategy and information systems strategic orientation in an e-commerce environment." Review of. *International Journal of Production Research* 42 (15):2915-39.

Chakravarty, Anindita, Yong Liu, and Tridib Mazumdar. 2010. "The Differential Effects of Online Word-of-Mouth and Critics' Reviews on Pre-release Movie Evaluation." Review of. *Journal of Interactive Marketing* 24 (3):185-97. doi: http://dx.doi.org/10.1016/j.intmar.2010.04.001.

Chen, Pei-yu, Samita Dhanasobhon and Michael D. Smith. 2008. All Reviews are Not Created Equal: The Disaggregate Impact ofReviews and Reviewers at Amazon.com, working paper. Carnegie Mellon University.

Chen, Mao, and Jaswinder Pal Singh. 2001. Computing and using reputations for internet ratings. Paper presented at the Proceedings of the 3rd ACM conference on Electronic Commerce.

Chen, Shih-Fen S., Kent B. Monroe, and Yung-Chien Lou. 1998. "The effects of framing price promotion messages on consumers' perceptions and purchase intentions." Review of. *Journal of Retailing* 74 (3):353-72. doi: http://dx.doi.org/10.1016/S0022-4359(99)80100-6.

Cheung, Christy M. K., and Dimple R. Thadani. 2012. "The impact of electronic word-of-mouth communication: A literature analysis and integrative model." Review of. *Decision Support Systems* 54 (1):461-70. doi: http://dx.doi.org/10.1016/j.dss.2012.06.008.

Chevalier, Judith A., and Dina Mayzlin. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews." Review of. *Journal of Marketing Research* 43 (3):345-54. doi: 10.1509/jmkr.43.3.345.

Chiang, Wei-yu Kevin, Dongsong Zhang, and Lina Zhou. 2006. "Predicting and explaining patronage behavior toward web and traditional stores using neural networks: a comparative analysis with logistic regression." Review of. *Decision Support Systems* 41 (2):514-31. doi: http://dx.doi.org/10.1016/j.dss.2004.08.016.

Ch'ng (2014) The Value of Using Big Data Technology in Computational Social Science, The 3rd ASE Big Data Science Conference, 4-7 August 2014, Tsinghua University Beijing China.

Chong, Alain Yee-Loong. 2013. "Predicting m-commerce adoption determinants: A neural network approach." Review of. *Expert Systems with Applications* 40 (2):523-30. doi: http://dx.doi.org/10.1016/j.eswa.2012.07.068.

Chong, Alain Yee-Loong, Felix TS Chan, Mark Goh, and MK Tiwari. 2013. "Do interorganisational relationships and knowledge-management practices enhance collaborative commerce adoption?" Review of. *International Journal of Production Research* 51 (7):2006-18.

Chong, Alain Yee-Loong, and Eric TW Ngai. 2013. "What Influences Travellers' Adoption of a Location-based Social Media Service for Their Travel Planning?" Review of. *PACIS 2013 Proceedings* 210.

Chong, Alain Yee-Loong, Keng-Boon Ooi, and Amrik Sohal. 2009. "The relationship between supply chain factors and adoption of e-Collaboration tools: An empirical examination." Review of. *International Journal of Production Economics* 122 (1):150-60. doi: http://dx.doi.org/10.1016/j.ijpe.2009.05.012.

Chong, Alain Yee-Loong, and Keng-Boon Ooi. 2008. "Adoption of interorganizational system standards in supply chains." Review of. *Industrial Management & Data Systems* 108 (4):529-47. doi: doi:10.1108/02635570810868371.

Clemons, Eric K., Guodong G. Gao and Lorin M. Hitt. 2006. “When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry,” *Journal of Management Information Systems*, 23:149–71.

Cui, Geng, Hon-Kwong Lui, and Xiaoning Guo. 2012. "The effect of online consumer reviews on new product sales." Review of. *International Journal of Electronic Commerce* 17 (1):39-58.

Davis, Alanah, and Deepak Khazanchi. 2008. "An Empirical Study of Online Word of Mouth as a Predictor for Multi-product Category e-Commerce Sales." Review of. *Electronic Markets* 18 (2):130-41.

Dellarocas, Chrysanthos, Neveen Awad, and Xiaoquan Zhang. 2004. "Exploring the value of online reviews to organizations: Implications for revenue forecasting and planning." Review of. *ICIS 2004 Proceedings*:30.

Dobrzykowski, David, Vafa Saboori Deilami, Paul Hong, and Seung-Chul Kim. 2014. "A structured analysis of operations and supply chain management research in healthcare (1982–2011)." Review of. *International Journal of Production Economics* 147, Part B (0):514-30. doi: http://dx.doi.org/10.1016/j.ijpe.2013.04.055.

Doern, Rachel R., and Carl F. Fey. 2006. "E-commerce developments and strategies for value creation: The case of Russia." Review of. *Journal of World Business* 41 (4):315-27. doi: http://dx.doi.org/10.1016/j.jwb.2006.08.001.

Drozdenko, Ronald, and Marlene Jensen. 2005. "Risk and maximum acceptable discount levels." Review of. *Journal of Product & Brand Management* 14 (4):264-70. doi: doi:10.1108/10610420510609285.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston. 2008. "The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry." Review of. *Journal of Retailing* 84 (2):233-42. doi: http://dx.doi.org/10.1016/j.jretai.2008.04.005.

Ehrenberg, A.S.C., Hammond, K. and Goodhardt, G.J. 1994. “The after-effects of price-related consumer promotions”, *Journal of Advertising Research*, Vol. 34 No. 4, pp. 11-21.

Flannagan, Ryan. 2014. "How important are online customer reviews?", Accessed 2014-10-30. http://nuancedmedia.com/how-important-are-online-customer-reviews/.

Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. 2014. How Online Product Reviews Affect Retail Sales: A Meta-analysis. Journal of Retailing, 90(2):217-232.

Garson, GD. 1998. *Neural networks: an introductory guide for social scientists*, London: Sage.

Gendall, Philip, Janet Hoek, Tracy Pope, and Karen Young. 2006. "Message framing effects on price discounting." Review of. *Journal of Product & Brand Management* 15 (7):458-65. doi: doi:10.1108/10610420610712847.

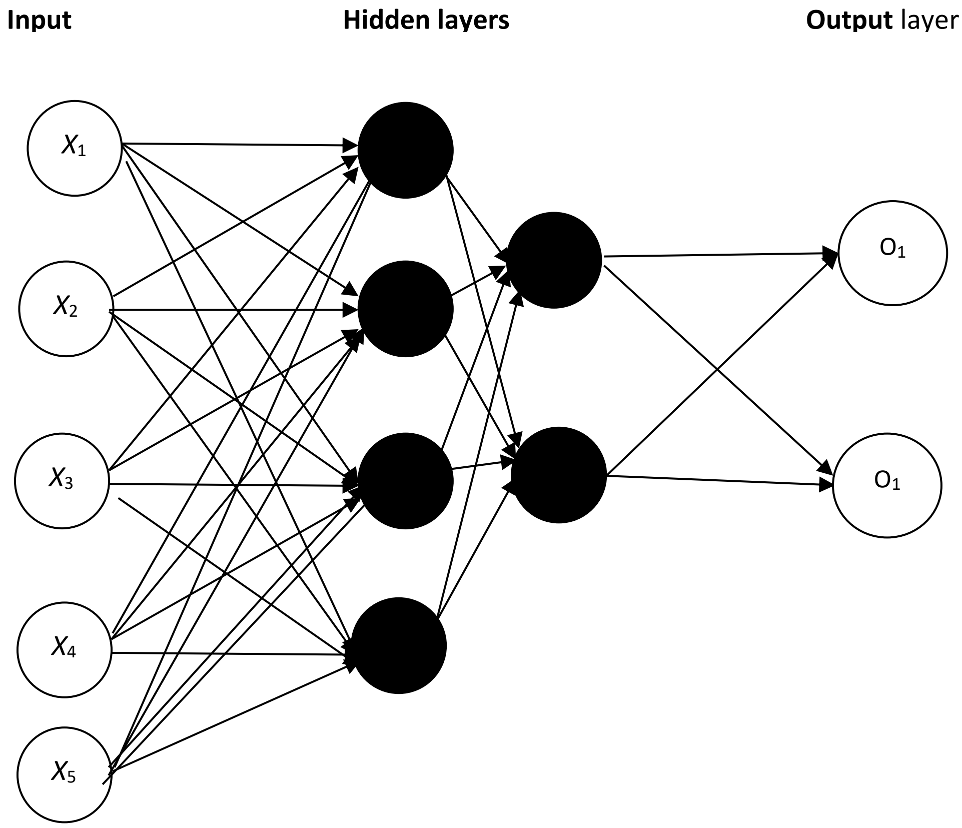
George, Gerard, Martine R Haas, and Alex Pentland. 2014. "Big Data and Management." Review of. *Academy of Management Journal* 57 (2):321-6.

Ghose, Aindya and Panagiotis G. Ipeirotis. 2006. “Towards an Understanding of the Impact of Customer Sentiment on Product Sales and Review Quality,” in *Workshop on Information Technology and Systems.*

Gupta, S. and Cooper, L.G. 1992. “The discounting of discounts and promotion thresholds”, *Journal of Consumer Research*. Vol. 19, December, pp.401-11.
Haykin, Simon. 1994. *Neural networks: A comprehensive foundation*: NY: Macmillan College Publishing.
Ito, Tiffany A, Jeff T Larsen, N Kyle Smith, and John T Cacioppo. 1998. "Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations." Review of. *Journal of personality and social psychology* 75 (4):887.
Jensen, Thomas, Jeremy Kees, Scot Burton, and Fernanda Lucarelli Turnipseed. 2003. "Advertised reference prices in an Internet environment: Effects on consumer price perceptions and channel search intentions." Review of. *Journal of Interactive Marketing* 17 (2):20-33. doi: 10.1002/dir.10052.
Lee, Jumin, Do-Hyung Park, and Ingoo Han. 2008. "The effect of negative online consumer reviews on product attitude: An information processing view." Review of. *Electronic Commerce Research and Applications* 7 (3):341-52. doi: http://dx.doi.org/10.1016/j.elerap.2007.05.004.
Lichtenstein, Donald R., Richard G. Netemeyer, and Scot Burton. 1990. "Distinguishing Coupon Proneness from Value Consciousness: An Acquisition-Transaction Utility Theory Perspective." Review of. *Journal of Marketing* 54 (3):54-67. doi: 10.2307/1251816.
Liu, Yong. 2006. "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue." Review of. *Journal of Marketing* 70 (3):74-89. doi: 10.1509/jmkg.70.3.74.
Lu, X., S. Ba, L. Huang, and Y. Feng. 2013. "Promotional Marketing or Word-of-Mouth? Evidence from Online Restaurant Reviews." Review of. *Information Systems Research* 24 (3):596-612. doi: doi:10.1287/isre.1120.0454.
Marshall, Roger, and Seow Bee Leng. 2002. "Price threshold and discount saturation point in Singapore." Review of. *Journal of Product & Brand Management* 11 (3):147-59. doi: doi:10.1108/10610420210430042.
McNeill, Lisa. 2013. "Sales promotion in Asia: successful strategies for Singapore and Malaysia", *Asia Pacific Journal of Marketing and Logistics*, Vol. 25 Iss 1 pp. 48 - 69
Min, Hye-Jin, and Jong C. Park. 2012. "Identifying helpful reviews based on customer’s mentions about experiences." Review of. *Expert Systems with Applications* 39 (15):11830-8. doi: http://dx.doi.org/10.1016/j.eswa.2012.01.116.
Mudambi, Susan M, and David Schuff. 2010. "What makes a helpful online review? A study of customer reviews on Amazon. com." Review of. *Management Information Systems Quarterly* 34 (1):11.
Ou, Carol Xiaojuan, Paul A Pavlou, and Robert M Davison. 2014. "Swift Guanxi in online marketplaces: the role of computer-mediated communication technologies." Review of. *MIS Quarterly* 38 (1):209-30.
Senecal, Sylvain and Jacques Nantel. 2004. “The Influence of Online Product Recommendations on Consumers’ Online Choices.” *Journal of Retailing* 80 (2):159–69
Shmueli, Galit, and Otto R Koppius. 2011. "Predictive analytics in information systems research." Review of. *MIS Quarterly* 35 (3):553-72.
Sim, Jia-Jia, Garry Wei-Han Tan, Jessica C. J. Wong, Keng-Boon Ooi, and Teck-Soon Hew. 2014. "Understanding and predicting the motivators of mobile music acceptance – A multi-stage MRA-artificial neural network approach." Review of. *Telematics and Informatics* 31 (4):569-84. doi: http://dx.doi.org/10.1016/j.tele.2013.11.005.
Smith, Alan D., and William T. Rupp. 2003. "Strategic online customer decision making: leveraging the transformational power of the Internet." Review of. *Online Information Review* 27 (6):418-32. doi: doi:10.1108/14684520310510055.

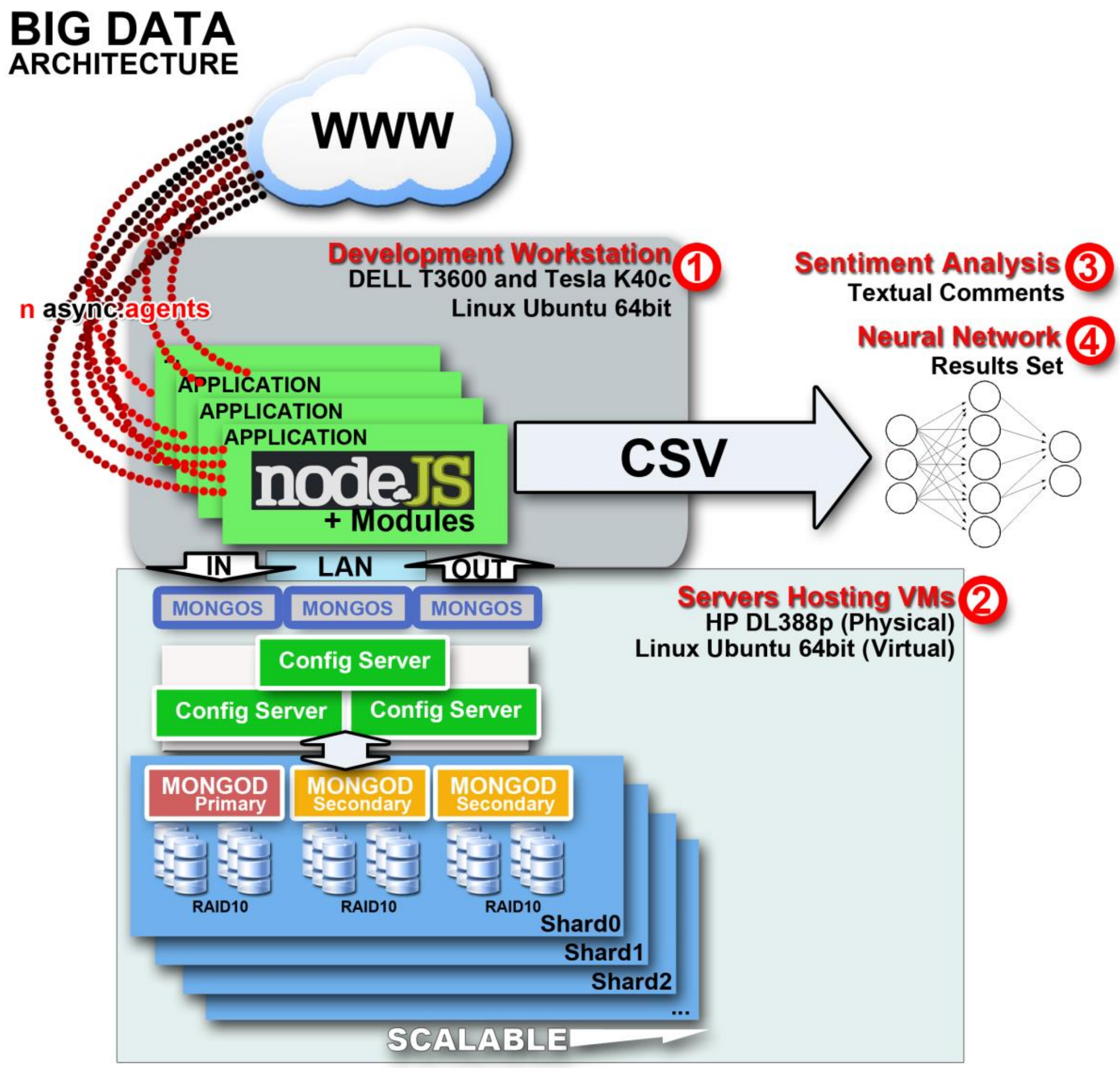
Sonnier, Garrett P., Leigh McAlister, and Oliver J. Rutz. 2011. “A Dynamic Model of the Effect of Online Communications on Firm Sales,” *Marketing Science*, 30 (4):702–716.
Suri, Rajneesh, Rajesh V. Manchanda, and Chiranjeev S. Kohli. 2000. "Brand evaluations: a comparison of fixed price and discounted price offers." Review of. *Journal of Product & Brand Management* 9 (3):193-207. doi: doi:10.1108/10610420010332458.
Tang, T., Fang, E., & Wang, F. 2014. Is neutral really neutral? The effects of neutral user-generated content on product sales. *Journal of Marketing*, 78(4):41-58.
Tirunillai, S., and G. J. Tellis. 2012. "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance." Review of. *Marketing Science* 31 (2):198-215. doi: doi:10.1287/mksc.1110.0682.
Tyagi, R.K. and Sawhney, M.S., 2010. "High-performance product management: the impact of structure, process, competencies, and role definition." *Journal of Product Innovation Management*, 27 (4), 83–96.
Westbrook, Robert A. 1987. "Product/Consumption-Based Affective Responses and Postpurchase Processes." Review of. *Journal of Marketing Research* 24 (3):258-70. doi: 10.2307/3151636.
Yip, Lily, and Rob Law. 2002. "User preferences for web site attributes: A study of the Hong Kong Disneyland." Review of. *Asia Pacific Journal of Tourism Research* 7 (1):33-44. doi: 10.1080/10941660208722108.
Zhang, Jason Q., Georgiana Craciun, and Dongwoo Shin. 2010. "When does electronic word-of-mouth matter? A study of consumer product reviews." Review of. *Journal of Business Research* 63 (12):1336-41. doi: http://dx.doi.org/10.1016/j.jbusres.2009.12.011.
Zhu, Feng, and Xiaoquan Zhang. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics." Review of. *Journal of Marketing* 74 (2):133-48. doi: 10.1509/jmkg.74.2.133.

Figure 1 Neural Network Example



X_1 to X_5 = Input vector; O_1, O_2 = Output vector

Figure 2. Big Data Architecture



Note: Numbering describes the three part process. 1) Developmental workstation where our Node.JS agents are deployed for scraping the web using asynchronous I/O calls. 2) Physical server hosting Ubuntu 64bit virtual machines, and where data is stored and horizontally scaled. 3) Completed Web crawling and scraping datasets are converted into comma separated values for Neural Network analysis.

Figure 3. The sentiment analysis process

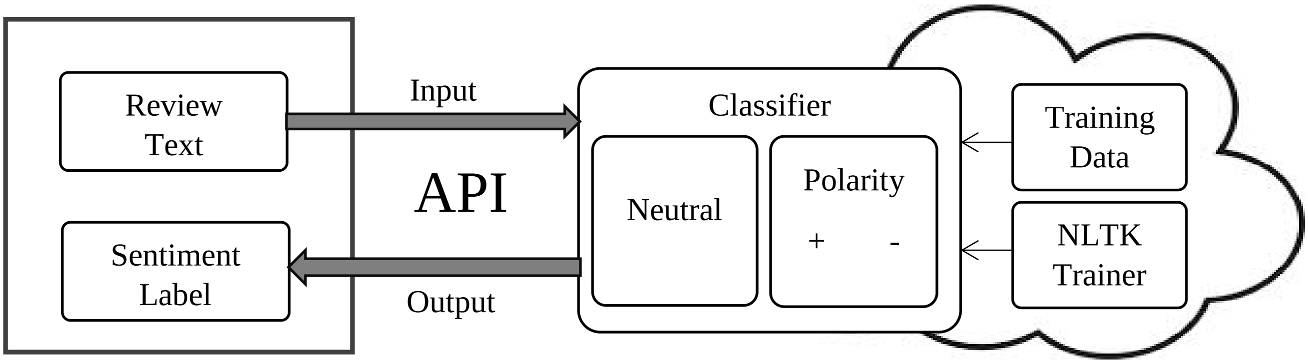


Table 1 Summary of variables used in this study

Variables	Description	Operationalization
Free Delivery	If the product offers free delivery	Binary variable: 1 means the product offers free delivery in UK as listed in Amazon.com, and 0 otherwise
Discount Value	The monetary value of the discount of the product	The fixed deductible amount of the product listed in Amazon.com, all in GBP
Discount Rate	The percentage discount of the product	The percentage of the discount on the original price of the product listed in Amazon.com, all in %
Customer Review Rating (Valence)	The valence of customer review of the product	The average of all customer review ratings for the product, listed on Amazon.com
Number of Customer Reviews (Volume)	The number of total customer reviews of the product	The listed total number of customer reviews of the product on Amazon.com
Number of Answered Questions	The number of total answered questions of the product	The listed total number of answered questions of the product on Amazon.com
Positive Reviews	The percentage of 5 and 4 stars customer review	The percentage of 5 and 4 star customer reviews (number of 5 and 4 star customer reviews divided by the total number of customer reviews) listed on Amazon.com.
Negative Reviews	The percentage of 2 and 1star customer review	The percentage of 2 and 1 star customer reviews (number of 2 and 1 star customer reviews divided by the total number of customer reviews) listed on Amazon.com. This measures the negative review.

Sales Rank	The Best Sellers Rank of the product	The ranking of the product provided by Amazon.com and this serves as a proxy of sales for this study.

Table 2 Full validation results of neural network model

Network	**Training**	**Testing**
1	.832	.810
2	.780	.784
3	.816	.825
4	.851	.861
5	.856	.872
6	.749	.794
7	.854	.864
8	.758	.805
9	.797	.834
10	.823	.819
Mean	.812	.827
Standard Deviation	0.039428	0.030554

Table 3 Sensitivity Analysis

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	Importance
Sentiment * Volume	.125	.111	.212	.085	.210	.088	.060	.096	.077	.071	11%
Percentage of Negative Reviews	.156	.143	.043	.175	.072	.069	.140	.117	.063	.090	11%
Discount Rate * Volume	.064	.050	.061	.142	.101	.097	.053	.096	.080	.156	9%
Sentiment * Valance	.057	.133	.157	.067	.079	.049	.111	.060	.089	.070	9%
Percentage of Positive Reviews * Discount Rate	.059	.115	.081	.024	.071	.091	.183	.030	.106	.056	8%
Number of Questions	.086	.077	.054	.161	.079	.096	.048	.059	.084	.066	8%
Volume	.056	.053	.063	.037	.079	.119	.043	.087	.089	.077	7%
Current Price	.074	.045	.029	.083	.079	.075	.071	.026	.041	.065	6%
Percentage of Positive Reviews	.047	.033	.049	.066	.036	.098	.044	.064	.040	.064	5%
Discount Rate	.048	.028	.066	.016	.052	.030	.019	.088	.110	.079	5%
Discount rate * Valance	.067	.094	.039	.032	.038	.023	.058	.072	.057	.047	5%
Discount Value	.035	.032	.067	.025	.037	.054	.059	.061	.025	.043	4%
Valance	.032	.032	.020	.027	.023	.049	.041	.084	.033	.064	4%
Percentage of Negative Review * Discount Rate	.035	.021	.023	.021	.007	.016	.023	.020	.050	.011	2%

Sentiments	.026	.018	.019	.019	.021	.025	.021	.024	.021	.022	2%
Free Delivery	.032	.016	.014	.019	.017	.021	.026	.015	.034	.017	2%