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

Purpose: This study aims to investigate if online reviews (e.g. valence and volume), online promotional strategies (e.g. free delivery and discounts) and sentiments of user reviews can help predict product sales.

Design/methodology/approach: 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. The neural network was employed to examine which variables in our study are important predictors of product sales.

Findings: This study found that although online reviews, online promotional strategies and sentiments can all predict product sales, some variables are more important predictors than others. We found that the interplay effects of these variables become more important variables than individual variables. For example, online volume's interactions with sentiments and discounts are more important predictors than discounts, sentiments or online volume individually.

Originality/value: This study designed a big data architecture, in combination with sentimental and neural network analysis, can facilitate future business researches for predicting product sales in an online environment. This study also employed a predictive analytic approach (e.g. neural network) to examine the variables in this study, and this approach is useful for future data analysis in a big data environment where prediction can have more practical implications than significance testing. This study also examined the interplays between online reviews, sentiments and promotional strategies which up to now have mostly been examined individually in previous studies.

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

Introduction

One of the key challenges faced by organizations today is the dynamic, global and unpredictable business environment in which they operate in. With growing customer expectations for 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 includes 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 Identification (RFID) are all able to help organizations manage their supply chain better, and also reduce traditional supply chain challenges such as the Bullwhip Effect (Croom, 2005; Huang and Handfield, 2015).

An important aspect of managing supply chain efficiently is to have a better forecast on product sales such that a manufacturer will not over or under purchase production products. An

emerging area in forecasting the sales of products is in big data and user-generated contents. Recent marketing reports have all shown the impacts of user-generated content 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 claiming 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 content has an impact on e-commerce, with e-commerce now being one of the most popular e-commerce business model (Wen et al., 2014). Companies such as Amazon.com and Alibaba's Taobao.com are examples of successful e-commerce companies that allow potential customers to see recommendations from others before making their purchasing decisions.

Given that user-generated content plays an important role in influencing the purchasing decisions of consumers, its role in helping organizations to understand and forecast product demand should be further investigated. In an e-commerce environment, consumers are presented with readily available information that can influence their purchasing decisions. In Amazon.com for example, this information includes product information such as the price of the product, promotional offers related to discounts and free deliveries, and online review information regarding the valence, volume, and the sentiments of the reviews. As E-commerce storefronts and marketplaces are now becoming one of the main purchasing channels by consumers, predicting a 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 e-commerce environment, it is now possible to predict product sales and customer demands based on data captured online (Duan et al., 2008).

The main objective of this study is to investigate if product demands (or sales) can be predicted through the 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 particular, the roles of valence and ratings are not independent of sentiments given that users are exposed to all these information on a website, and therefore the role of review sentiments' interactions with reviews ratings valance and volume need to be further investigated. Previous studies on online reviews also often used data collected via online survey and use experimental approaches to understand users' behavior towards online reviews (e.g. Sparks and Browning, 2011; Vermeulen and Seegers, 2009). This research extends these studies by capturing real data from Amazon.com, as well as actual sentiments via their online comments. Lastly, as big data focuses on the three main Vs: Volume, Velocity and Variety, our approach of capturing the data online has taken into considerations of these three characteristics of big data as we our big data algorithm captured real time data, and all data available for the product we studied, as well as information available to the users who view the product page. In this study, we designed a series of big data algorithms that 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 labeled as positive, negative or neutral for further analysis. This study will then use a neural network to predict

how these variables can collectively be used to predict product sales, as well as the predictability 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 that examine whether online data such as reviews, sentiments and promotional strategies can be used to predict product demand. Secondly, this study examines the differentials 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.

This paper is organized as follows. In Section 2, we provide a literature review on predictors derived from online promotional marketing techniques and online reviews. This is followed by a discussion on neural network. Section 4 presents the research methodology which includes details on our big data architecture design, sentimental analyses and findings from neural network. Lastly, our paper presents the discussion, the conclusions and implications, and the limitations of the study and future research directions.

2.0 Literature Review

2.1 Online promotional marketing

With advances in information systems and technology, new information channels are being increasingly utilized by individuals to access information (Tang et al., 2014). This easier access to information has several important implications especially apparent within the electronics industry. First is the increase in the amount of product information easily available to consumers. This means that there are more variables affecting consumers' purchasing decisions due to the amount of information on products that are made available to consumers (Floyd et al., 2014). Secondly, companies are becoming increasingly pressured to secure sales on their products within a shorter time period. This is illustrated by the trend of new product development in the aeronautical industry and the automotive industry, as well as electronic products; the latter one being 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 strategies implemented by vendors are price discounts offerings, which is also applicable in an online environment.

2.1.1 Discount Value

According to Lichtenstein et al. (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 good value 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 a 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 et al., 1994; McNeill, 2013), Drozdenko and Jensen (2005) have highlighted contradictions 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 has generally stood at 20% to 30%. This means that products with a price reduction higher than 30% would not necessarily translate to an increase in sales of the product in the US (ibid.). Similar studies 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 lead to differing discount rate expectations 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, the price of a product is also used by customers to discern many other variables. Suri et al. (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 attributive information when processing information to come up with a final evaluation of the product's quality (Drozdenko and Jensen, 2005). This attributive information on the product may result in the customers' final evaluations of the product's discount as being perceived to be a monetary gain to them. Attributive information in a form of external price references for a product has shown that there is also a difference of the effect of the 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 (Ramanathan, 2012), especially for vendors trying to offer substitutes of a main product that might be low on stock. The monetary value of price deduction as well as the actual price of product after such deduction, if any, is thus both listed as potential predictors of product sales. 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 is used in research on this subject. Chen et al. (1998) have argued that consumers' absolute and relative perceptions of price may influence their perception to price reductions. Based on the psychophysics-of-price heuristics theory, the 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 predict 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 free delivery services as well as price reductions 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) in which 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. This was due to the fact that once these companies removed their free delivery offerings it became increasingly difficult for them to retain their customer base. This is to be expected as electronic commerce becomes more mature, consumers utilizing this technology would increase their standard expectation. This means that while free delivery used to be something perceived as an additional incentives for customers, is now becoming increasingly standardized, leading to changes in the way consumers perceive promotions offered by electronic vendors. Free delivery has become a service that is expected to be offered by electronic vendors, and a failure to offer this service might translate to a loss 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 play 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 such as blogs, product reviews, wikis and Twitter (Tirunillai and Tellis. 2012). This particular information could be defined as User Generated Content (UGC). UGC "refers to media content created by users to share information and/or opinions with other users" (Tang et al., 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 et al., 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 et al., 2008). Furthermore e-UGC enables customers to obtain more elaborate information on a product (Floyd et al., 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 (Chevalier and Mayzlin 2006; Clemons et al., 2006; Duan et al., 2008; Ghose and Ipeirotis 2006). However, in past literature 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 literature, this paper will also utilize the valence of online review as one of the variables measured.

2.2.1 Online review valence

Online review valence refers to the nature of e-UGC content, which is usually presented in the form of an evaluation score which can be positive, negative or mixed (Yang et al., 2012). Although previous research has found 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 et al. (2004) have also found a strong predictive 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 correlation with the aggregation of weekly box office revenue. Similarly, Davis and Khazanchi (2008) found that online review valance could not 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 et al. (2008) found that online review valence has an indirect impact on box office revenue by significantly influencing 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) found that the effect of online reviews on sales could be diluted by the use of 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 influencing sales. 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 et al., 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 for example in the case of consumer electronics) 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 previous literature, 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 literature has shown that online review volume has a quantitatively measurable impact on product sales (Davis and Khazanchi, 2008; Duan et al., 2008; Liu, 2006; Lu et al., 2013). It is believed that the increased consumer product awareness caused by online review volume leads to higher sales (Liu 2006; 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 et al., 2008). However, Cui et al. (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 an even more important role in predicting sales than online review volume (ibid.). Based on previous literature, it is clear that online review volume is an important predictor of 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 tends to boost purchase behavior of consumers while negative UGC is likely to discourage purchases (East et al., 2008). However, studies have shown that not only negative online e-UGC spreads at a faster pace compared to positive e-UGC (Cui et al., 2012), but it may also have a more pronounced impact on customer~~s~~'s decision to purchase a product (Cheung and Thadani, 2012; Lee et al., 2008). Cui et al. (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 et al., 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 et al., 2008)

A case study on how movies are evaluated based upon the type of e-UGC has shown that positive e-UGC has considerably less influence on end evaluations of the movies compared to negative e-UGC (Chakravarty et al., 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 program usage has also shown that negative e-UGC are more influential in decreasing consumption of software compared to the mitigating effect of positive e-UGC in consumption (Zhang et al., 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 the 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 has 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 increase the trust and sales in e-commerce context (Ng, 2013; Ou et al., 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 a numeric rating, sentiment is a qualitative feature of the text. Although the numeric rating of a review is often used to reflect the valence of that review (e.g. Ye et al., 2011; Chevalier 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 the text sentiment of online reviews is consistent with and thus may be overlooked by a 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 et al., 2006). Such bimodality makes the review rating difficult to reflect the true product quality, and may undermine its importance as a determinant to the 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 types of cognitive appraisals from previous customers, and such appraisals provide a useful information set for a 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 a strong influence in product sales.

Apart from influencing purchase decisions through a 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). The traditional point of view is that such a 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 the actual purchasing of a product (Tsai, 2001).

Based on previous literature, this study seeks to distinguish online review text sentiment from online review ratings, and considers text sentiment as a predictor of product sales. Specifically, Amazon.com offers the peer evaluation for '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 other reviews. First, such helpfulness ratings are perceived as 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). Secondly, the most helpful reviews are listed on the first page, and have a 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)

The Internet is becoming increasingly integrated into everyday life. There is more information that individuals are able to access today than in any previous time in history. This access however, has had an unexpected consequence of information overload. The amount of online information that is present today poses a challenge for customers in the decision making process on purchasing decisions (Chong and Ngai, 2013). This can also be seen by the fact that increasingly, vendors are likely to adopt multi-dimensional approaches to market their products in an online setting, as opposed to using a single-dimensional marketing strategy approach that is 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. Does UGC in the form of online reviews better predict the potential sales count for a product? Or is a one-way marketing effort by a company 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 in the same way as they affect, for example, 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 design, how could older products still be relevant to customers? Also of a considerable importance is the fact that previous findings on discounting has offered contradictory results on how a product sells, as discounts in certain instances might instead be perceived as a negative quality of that product (products with high discounts may be perceived to be of lower quality) (Marshall and Leng, 2002; Suri et al., 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 findings, 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 rates 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 e-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 et al., 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 products, 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 evaluation cues based on a multitude of information such as product attributes, functions, and performance 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 underexplored. In our study, the most helpful reviews sentiments are being used instead of sentiments of all reviews. This is because on the product page, the most helpful reviews are being shown prominently, and most users will go through the most helpful reviews first instead of all the product reviews (Hu et al., 2014). Thus, although it is possible that a product has a high rating, the most helpful reviews' sentiments, either positive or negative, can alter the review rating's influence on product sales.

Past studies on the relationships between the volume of reviews and product sales do not have consistent results (Chevalier and Mayzlin, 2006; Dellarocas et al., 2004). 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 a scenario, it is possible that when review volume is 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 examines 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 the neural network (Chiang et al., 2006). In a neural network model, the networks are presented as systems of interconnected neurons that can compute various values from input information, and can learn the intrinsic nature of patterns or processes from data sets (Sim et al., 2014). In recent studies, neural networks have shown to be an effective alternative to traditional statistical techniques (Chong, 2013). In particular, a neural network can offer better prediction than traditional regression approaches, and is suitable to be employed to test large scale data with a relatively large number of input variables such as in a big data environment (George et al., 2014). In addition, neural network is suitable to solve complicated problems by learning from available situations and making predictions in other situations. For example, Ayat and Pour (2014) used different artificial neural network algorithms to predict the student performance with input factors related to learning and educational progress. Because neural networks can model complex relationships between input and output variables, considering the complexity in predicting online sales with a list of inputs such as the price, discount, delivery and customer reviews, neural network is suitable to use in our study.

Previous studies have shown that neural network techniques such as the multilayer perceptron (MLP) can be trained to approximate the smoothest, measureable functions (Gardner and Dorling, 1998). Furthermore, the MLP is not parametric and thus does not have restricted assumption about the variables. Also, MLP can include both non-linear and linear functions

into its model without pre-definition of the relationships, and can be trained to accurately generalize new and unforeseen data (ibid.). In our study, the variables we consider, such as price and number of customer reviews, are likely to have non-linear relationships with sales. Therefore, MLP is superior to traditional linear and parametric statistical methods in this situation (Gevrey et al. 2003).

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 layer is between the input and output layers, and it contains unobservable nodes. The value of each hidden unit is some function of the input units. In MLP, two hidden layers are 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 an appropriate supervised learning algorithm, the MLP will analyze 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.

Figure 1 provides an example of MLP.

<<Figure 1 about here>>

Apart from neural network, there are also other approaches, such as tabu search and genetic algorithm, which are commonly used to analyze complex situations (Baker and Ayechew, 2003). For example, tabu search is used in Lopez et al. (1998)'s research to address the Hot Strip Mill Production Scheduling Problem, while in Bolduc et al. (2009)'s paper, tabu search is used for split delivery Vehicle Routing Problem with Production and Demand Calendars (VRPPDC). Moreover, genetic algorithm is applied by Li et al. (1998) to solve the problem of Earliness and Tardiness Production Scheduling and Planning (ETPSP), and a spanning-tree based genetic algorithm is used to achieve an efficient system for production, distribution and inventory management in Gen and Syarif (2005)'s research. However, the aims of tabu search and genetic algorithm are the optimization of a stable problem (Zhao and Wu, 2000), and they are often used to find the optimal parameters that people can control to get a best result, such as optimal control of fruit-storage process (Erenturk and Erenturk, 2007). However, in our situation, although optimizing sales would be ideal, it is practically inappropriate to try to find optimized parameters that lead to sales, because the parameters in online marketplace are difficult to control. Even if all the parameters can be controlled to the optimized value as identified by these approaches, customers behaviors will change corresponding to that, and the results will still be different.

Based on above discussion, MLP is applied to predict the factors influencing product sales in our study. Although studies in the past applied explanatory statistical techniques to examine their research models, there is an increase in recent studies to apply predictive analytic approaches such as MLP in information systems research (Lu et al., 2013). A key advantage of MLP is that it can offer a useful and practical model which can help researchers to develop new theories. It can also overcome the challenges faced by traditional statistical analysis

relying on p-value which may not be effective in an environment with large data sets (e.g. false correlations) (George et al., 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 cameras, televisions, Hi-Fi sound systems, computers, 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 a 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 of 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 for the product (i.e. product sales) is measured using sales rank. This is consistent with the approaches by previous studies by 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 such needs arise. Our project aims to access, manage and process tens of thousands of web page content, including cleaning data, in real-time. As opposed to a conventional desktop and network connection, our specific requirements suggest that scalable technology is required. Big data technology plays an important role 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 2. 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. Our algorithms are coded server side in JavaScript via Node.js, an event driven, non-blocking I/O model built on Google's V8 JavaScript engine. With asynchronous I/O approach, our algorithms can manage real-time, data-intensive applications, and efficiently store our data within Mongo DB, a scalable cross platform NoSQL database, as the data comes in. The data is then stored into a CSV (Comma Separated Value) file so that it can be used for neural network analysis. 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 utilized as the data was not sufficiently large to require multicore processing.

The process of our technical methodology is shown in Figure 2:

- Developmental workstation where our Node.JS agents are deployed for scraping the web using asynchronous I/O calls.

- Physical server hosting Ubuntu 64bit virtual machines, and where data is stored and horizontally scaled.

- 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 of 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 are deployed to scrape, in real-time, the Amazon websites using JavaScript's DOM (Document Object Model) and processes the scattered HTML tags where our target information is embedded into structured key-value pair datasets. 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 was implemented so that we could control *n* requests per set. It takes on an average 1.1392 seconds to call a HTTP request, obtain a HTML response and scrap the page of all required data. We extracted all the existing electronics data available on the Amazon pages, and we could have continued data scraping with such a 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 15,433. Our sample includes 468 Audio and Hi-Fi devices, 7,917 Camera and Photo related devices, 6,549 Computers and Accessories, 416 Electronic Accessories and miscellaneous electronic devices, and 92 outdoors and sports related electronic devices.

Figure 3 and Table 1 provides the summary of variables used for this study and their descriptions.

<<Figure 3>>

<<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' sentiments and opinions embedded in online reviews and UGC (Archak et al., 2011; Pang and Lee, 2008)

In our study, we conducted sentiment analysis on the most helpful reviews that are shown on the first pages 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 a 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 labelling. Based on the probabilities, sentiment was labelled 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 from several data sets by Pang and Lee (2008). This process of sentiment analysis is shown in Figure 4.

<< Figure 4 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) was developed for this study. According to Chiang et al., (2006), within the field of e-commerce, back-propagation neural networks are the most commonly used networks. Drawing from previous studies, the data analysis 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 (product sales) are then provided for the neural network.

The difference between the actual output (e.g. product sales) 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 et al., 2006).

4.5 Validation of Neural Networks

We applied a 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 any 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 nodes 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 15 predictors, four 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. 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. As shown in Table 2, relative errors computed by SPSS 19.0 was computed to compare data from the training and testing sets to ensure that there was not much difference between the two tables.

<<Table 2 about here>>

From Table 2, the average relative error for the training model is 0.812 while the testing model is 0.827. We conducted T-Test to compare to relative errors of both the training and testing tests, and found no significance between errors in both data sets. 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 15 predictors are found to be relevant to all ten networks. The average result showed that the two main and most important predictors are Sentiments*Volume and Percentage of Negative Reviews. Thus when there are more reviews, together with sentiments of these reviews will predict the sales of products very well. The result showed that in general, many of the online reviews, online marketing strategies, and sentiments do not individually predict the sales of products. Instead, we found that the interactions between these factors become strong predictors of product sales. For example, discount strategies are not important predictors of product sales. However, together with online volume or discount rates, they become important predictors. Another observation is that sentiments although is a predictor of product sales, it is not an important predictor compared to the others. However, when there is high volume of reviews, sentiments together with volume become important. Surprisingly, we found that valance, which is a commonly studied online review attributes, is not a strong

predictor of product sales compared to other variables used in this study. Its joint effects with the sentiments of reviews however, can predict product sales very well.

5.0 Discussions

The result of our analysis has confirmed all variables and predictors of product sales. In general we found that the interplays between online reviews, sentiments and online marketing promotional strategies are important predictors of product sales. Thus unlike existing studies which examined many of these variables separately, we found that the interactions between these variables can help us forecast the sales of products much better. Interaction between sentiment and volume and discount rate and volume for example, are more important predictors of product sales than other variables used in this study. 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 a 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 products compared to Volume alone. This means that an increase in sales performance 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 better sales.

Chevalier and Mayzlin (2006) have noted that based on their review-length data on book sales, the actual content of the online reviews (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 a more important predictor to product sales compared to text sentiment. These comparisons have shown that there are differing sales predictor variables for search products and experience products. However it is also important to note that while sentiment alone is not a good predictor to product sales, its effect is tremendously amplified when it's combined with the volume and valence of the reviews.

Previous literature offers inconclusive results on the differing importance between volume and valence in predicting product sales (Davis and Khazanchi, 2008). Although both have been confirmed to be 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 a more important predictor to product sales than valence. Other than volume and valence, number of online questions 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 all predictors are found to be important in our neural network analyses, there are several variables that are less important as a predictor of product sales. We found out that free delivery and sentiments have only less prediction importance on product sales 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 directly to vendors in regards to their products) to further increased their product sales.

6.0 Conclusions and Implications

The rich information embedded in online ecommerce websites has attracted increasing attention. In this study, we employed big data architecture, sentiment analysis and three-layered neural network modeling to examine predictors of sales of product (consumer electronics) in e-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 as a stand-alone variable in sales rating prediction (only a minor predictor and insignificant when compared to valence as opposed to Chevalier and Mayzlin (2006)'s study. Some predictors seem to play a more important role compared to others. Of particular interest 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 could conclude several important implications.

Firstly, 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 data enabled us to more accurately predict the online products' consumer demand, due to the greater generalization associated with larger sample sizes. Although we have only extracted samples from one website (Amazon.com), this utilization of big data architecture can be extended to a larger scale, such as extracting connected but dispersed social media data. The capability of extracting large amounts 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, et al., 2014). Instead, we make use of an 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 (ibid.)

Secondly, we have demonstrated and discussed the influence of various marketing promotional strategies and e-UGC on product sales and our finding provides significant guidance on e-commerce management. Our research results confirmed the importance of e-UGC-based social interactions on an e-commerce website, and suggest that online vendors should pay attention to online interactivity; not only to customer reviews but also to 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 sellers 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 the level of interactivity is launch apply an instant messaging service so that customers can more easily interact with customers and online sellers. This is an extension of previous studies by Lu et al. (2013) where the authors examined the roles of online promotional strategies and online reviews, but did not include sentiments.

Sentiments of reviews play and important part of consumers' decisions in purchasing (Hu et al., 2012), and therefore our proposed model is a new contribution to existing theories on online reviews or sentiments by examining these information concurrently and using real data captured from Amazon.com.

Third, our research showed high importance of negative and positive review percentage, suggesting that customers perceive polarised review with high value. The interaction between review and discount rate also show their influence on sales. This suggests that the polarised reviews and the separated rating number should also be displayed in easily noticeable part of the webpage, and discount can be offered wisely to moderate the effect of reviews.

Fourthly, the key characteristics of big data is the velocity, variety and volume of the data (Gandomi and Haider). In the context of e-commerce studies and online reviews, most studies have not address these three characteristics as data are often collected from surveys (Sparks and Browning, 2011). However, this study examined previous research models on online reviews by developing a big data architecture which was able to capture real time data, in large quantity, and different data that can influence users' decisions (e.g. sentiments, online ratings, discount etc.).

Finally, another implication of particular importance in this study is that when sentiment interacts with volume and valence of e-UGC, it becomes a more important predictor of product sales compared to variables studied in previous studies such as online valance. This is a unique finding since literature to this date has yet to examine the interaction effect of sentiment on other predictors for online products' sales ratings. While Chevalier and Mayzlin (2006)'s study has found that sentiment better predicts sales when compared to e-UGC valence, and Yu et al. (2012)'s study has confirmed that sentiment could be a single predictor to product sales, this paper has extended and confirmed that sentiment could further enhance sales if its combined with the number of online reviews and average weighed ratings of products. This finding implicates the importance of business in e-sectors to pay particular attention when analysing their text sentiment not as a singular predictor, but rather to also weigh in volume and valence of online reviews to understand customer opinion of their products better, and also to better predict their products' sales.

7.0 Limitations and future directions

Despite the conclusions that this paper has contributed, there are several limitations worth noting. Firstly this study only examines electronic products. One major 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 products. Thus, future research in this study could further contribute by incorporating more products and examining 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 a 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 and a very diverse and varied amount of products availability (Garcia et al. 2011), thus offering us a suitable sample source where big data architecture can be implemented. It is hoped that future research could incorporate other similar websites and utilize this study as a foundation to examine further and confirm our findings.

Acknowledgement

"The authors acknowledge 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

Archak, N., Ghose, A. and Ipeirotis, P.G. (2011), "Deriving the Pricing Power of Product Features by Mining Consumer Reviews", *Management Science,* INFORMS, Vol. 57 No. 8, pp. 1485–1509.

Ayat, S. and Pour, Z. A. (2014), "Comparison between artificial neural network learning algorithms for prediction of student average considering effective factors in learning and educational progress", *Journal of Mathematics and Computer Science,* Vol. 8, pp. 215-225.

Baker, B. M. and Ayechew, M. A. (2003), "A genetic algorithm for the vehicle routing problem", *Computers & Operations Research,* Vol. 30 Iss. 5, pp. 787-800.

Bolduc, M. C., Laporte, G., Renaud, J. and Boctor, F. F. (2010), "A tabu search heuristic for the split delivery vehicle routing problem with production and demand calendars", *European Journal of Operational Research,* Vol. 202, pp. 122-130.

Chakravarty, A., Liu, Y. and Mazumdar, T. (2010), "The Differential Effects of Online Word-of-Mouth and Critics' Reviews on Pre-release Movie Evaluation", *Journal of Interactive Marketing,* Vol. 24 No. 3, pp. 185–197.

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

Chevalier, J. a and Mayzlin, D. (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews", *Journal of Marketing Research,* Vol. 43 No. August, pp. 345–354.

Chong, A.Y.-L., Ooi, K.-B. and Sohal, A. (2009), “The relationship between supply chain factors and adoption of e-collaboration tools: an empirical examination”, *International Journal of Production Economics*, Elsevier, Vol. 122 No. 1, pp. 150–160.

Chong, A.Y.-L. and Zhou, L. (2014), “Demand chain management: Relationships between external antecedents, web-based integration and service innovation performance”, *International Journal of Production Economics*, Elsevier, Vol. 154, pp. 48–58.

Content, U., Performance, S., Tirunillai, S. and Tellis, G.J. (2012), “Does Chatter Really Matter ? Dynamics of”, Vol. 31 No. 2, pp. 198–215.

Croom, S.R. (2005), “The impact of e-business on supply chain management: An empirical study of key developments”, *International Journal of Operations & Production Management*.

Cui, G., Lui, H. and Guo, X. (2012a), “The effect of online consumer reviews on new product sales”, *International Journal of Electronic …*, available at: http://www.tandfonline.com/doi/abs/10.2753/JEC1086-4415170102 (accessed 10 March 2015).

Cui, G., Lui, H.-K. and Guo, X. (2012b), “The Effect of Online Consumer Reviews on New Product Sales”, *International Journal of Electronic Commerce*, Vol. 17 No. 1, pp. 39–58.

Davis, A. and Khazanchi, D. (2008), "An empirical study of online word of mouth as a predictor for multi‐product category e‐commerce sales", *Electronic Markets*, Vol. 18 No. 2, pp. 130-141.

Dellarocas, C.N., Awad, N. and Zhang, X. (Michael). (2004), “Using Online Reviews as a Proxy of Word-of-Mouth for Motion Picture Revenue Forecasting”, *SSRN Electronic Journal*, available at: http://www.ssrn.com/abstract=620821.

Drozdenko, R. and Jensen, M. (2005), “Risk and maximum acceptable discount levels”, *Journal of Product & Brand Management*, Vol. 14 No. 4, pp. 264–270.

Duan, W., Gu, B. and Whinston, A. (2008), “The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry”, *Journal of retailing*, available at: http://www.sciencedirect.com/science/article/pii/S0022435908000171 (accessed 10 March 2015).

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

Erenturk, S. and Erenturk, K. (2007), “Comparison of genetic algorithm and neural network approaches for the drying process of carrot”, *Journal of Food Engineering*, Vol. 78 No. 3, pp. 905-912

Floyd, K., Freling, R., Alhoqail, S., Cho, H.Y. and Freling, T. (2014), “How online product reviews affect retail sales: A meta-analysis”, *Journal of Retailing*, Vol. 90 No. 2, pp. 217–232.

Gandomi, A. and Haider, M. (2015), "Beyond the hype: Big data concepts, methods, and analytics", *International Journal of Information Management*, Vol. 35 No. 2, pp. 137-144.

Gardner, M. and Dorling, S. (1998), “Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences”, *Atmospheric environment*, available at: http://www.sciencedirect.com/science/article/pii/S1352231097004470 (accessed 11 March 2015).

Gen, M. and Syarif, A. (2005), “Hybrid genetic algorithm for multi-time period production/distribution planning”, *Computers & Industrial Engineering*, Vol. 48 Iss. 4, pp. 799-809.

George, G., Haas, M. and Pentland, A. (2014), “Big data and management”, *Academy of Management Journal*, available at: http://amj.aom.org/content/57/2/321.short (accessed 11 March 2015).

Gevrey, M., Dimopoulus, L. and Lek, S. (2003), “Review and comparison of methods to study the contribution variables in artificial neural network models”, Ecological Modelling, Vol.160 No. 3, pp. 249-264.

Ghose, A. and Ipeirotis, P.G. (2011), “Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics”, *Knowledge and Data Engineering, IEEE Transactions on*.

Gupta, S. and Cooper, L.G. (1992), “The Discounting of Discounts and Promotion Thresholds”, *Journal of Consumer Research*.

Hancock, J.T., Gee, K., Ciaccio, K. and Lin, J.M.-H. (2008), "I'm sad you're sad: emotional contagion in CMC", *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ACM, pp. 295–298.

Hatfield, E. and Cacioppo, J.T. (1994), *Emotional contagion*, Cambridge university press.

Hu, N., Koh, N.S. and Reddy, S.K. (2014), "Ratings lead you to the product, reviews help you clinch it? the mediating role of online review sentiments on product sales", *Decision Support Systems*, Elsevier B.V., Vol. 57, pp. 42–53.

Hu, N., Pavlou, P.A. and Zhang, J. (2006), "Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of Online word-of-mouth communication", *Proceedings of the 7th ACM conference on Electronic commerce*, ACM, pp. 324–330.

Kramer, A.D.I., Guillory, J.E. and Hancock, J.T. (2014), "Experimental evidence of massive-scale emotional contagion through social networks", *Proceedings of the National Academy of Sciences*, National Acad Sciences, Vol. 111 No. 24, pp. 8788–8790.

Lee, J., Park, D.H. and Han, I. (2008), "The effect of negative online consumer reviews on product attitude: An information processing view", *Electronic Commerce Research and Applications*, Vol. 7 No. 3, pp. 341–352.

Li, Y., Ip, W. H. and Wang, D. W. (1998), "Genetic algorithm approach to earliness and tardiness production scheduling and planning problem", *International Journal of Production Economics*, Vol. 54, pp. 65-76.

Lichtenstein, D.R., Netemeyer, R.G. and Burton, S. (1990), "Distinguishing Coupon Proneness from Value Consciousness: An Acquisition-Transaction Utility Theory Perspective", *Journal of Marketing*, Vol. 54 No. 3, pp. 54–67.

Lopez, L., Carter, M. W. and Gendreau, M. (1998), "The hot strip mill production scheduling problem: A tabu search approach", *European Journal of Operational Research*, Vol. 106, pp. 317-335.

Lu, X., Ba, S., Huang, L. and Feng, Y. (2013), "Promotional marketing or word-of-mouth? Evidence from online restaurant reviews", *Information Systems Research*, Vol. 24 No. 3, pp. 596–612.

Marshall, R. and Leng, S.B. (2002), "Price threshold and discount saturation point in Singapore", *Journal of Product & Brand Management*, Vol. 11, pp. 147–159.

McNeill, L. (2013), “Sales promotion in Asia: successful strategies for Singapore and Malaysia”, *Asia Pacific Journal of Marketing and Logistics*, Vol. 25, pp. 48–69.

Mudambi, S. and Schuff, D. (2010), “What makes a helpful review? A study of customer reviews on Amazon. com”, *MIS quarterly*, available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2175066 (accessed 10 March 2015).

Ng, C.S.-P. (2013), “Intention to purchase on social commerce websites across cultures: A cross-regional study”, *Information & Management*, Vol. 50 No. 8, pp. 609–620.

Ou Pavlou, P.A., and Davison, R.M., C.X.J. (2014), “Swift Guanxi in Online Marketplaces: The Role of Computer-Mediated Communication Technologies”, *MIS Quarterly*, Vol. 38 No. 1, pp. 209–230.

Pang, B. and Lee, L. (2008), “Opinion mining and sentiment analysis”, *Foundations and trends in information retrieval*, available at: http://dl.acm.org/citation.cfm?id=1454712 (accessed 10 March 2015).

Pugh, S.D. (2001), “Service with a smile: Emotional contagion in the service encounter”, *Academy of management journal*, Academy of Management, Vol. 44 No. 5, pp. 1018–1027.

Sonnier, G.P., McAlister, L. and Rutz, O.J. (2011), “A Dynamic Model of the Effect of Online Communications on Firm Sales”, *Marketing Science*.

Tang, T. (Ya), Fang, E. (Er) and Feng Wang. (2014), “Is Neutral Really Neutral? The Effects of Neutral User-Generated Content on Product Sales”, *Journal of Marketing*, Vol. 78 No. 71328203, pp. 41–58.

Tsai, W.-C. (2001), “Determinants and consequences of employee displayed positive emotions”, *Journal of Management* , Vol. 27 No. 4 , pp. 497–512.

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*, Vol. 27 No. 1, pp. 83–96.

Vermeulen, I. E., and Seegers, D. (2009), "Tried and tested: The impact of online hotel reviews on consumer consideration", *Tourism Management*, Vol. 30 No. 1, pp. 123-127.

Yang, J., Kim, W., Amblee, N. and Jeong, J. (2012), “The heterogeneous effect of WOM on product sales: why the effect of WOM valence is mixed?”, *European Journal of Marketing*, Vol. 46, pp. 1523–1538.

Ye, Q., Law, R., Gu, B. and Chen, W. (2011), “The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings”, *Computers in Human Behavior*, available at: http://www.sciencedirect.com/science/article/pii/S0747563210000907 (accessed 10 March 2015).

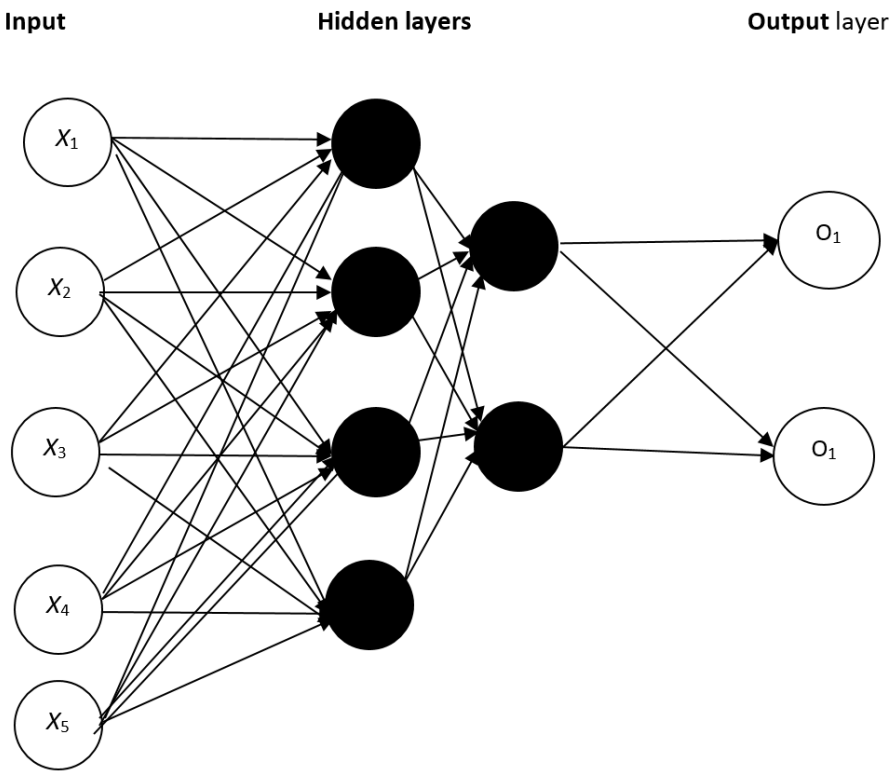
Yin, D., Bond, S. and Zhang, H. (2014), “Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews”, *Mis Quarterly*, Vol. 38 No. 2, pp. 539–560.

Zhao, C. and Wu, Z. (2000), “A genetic algorithm for manufacturing cell formation with multiple routes and multiple objectives”, International Journal of Production Research, Vol. 38 No. 2, pp. 385-395.

Zhang, J.Q., Craciun, G. and Shin, D. (2010), “When does electronic word-of-mouth matter? A study of consumer product reviews”, *Journal of Business Research*, Vol. 63 No. 12, pp. 1336–1341.

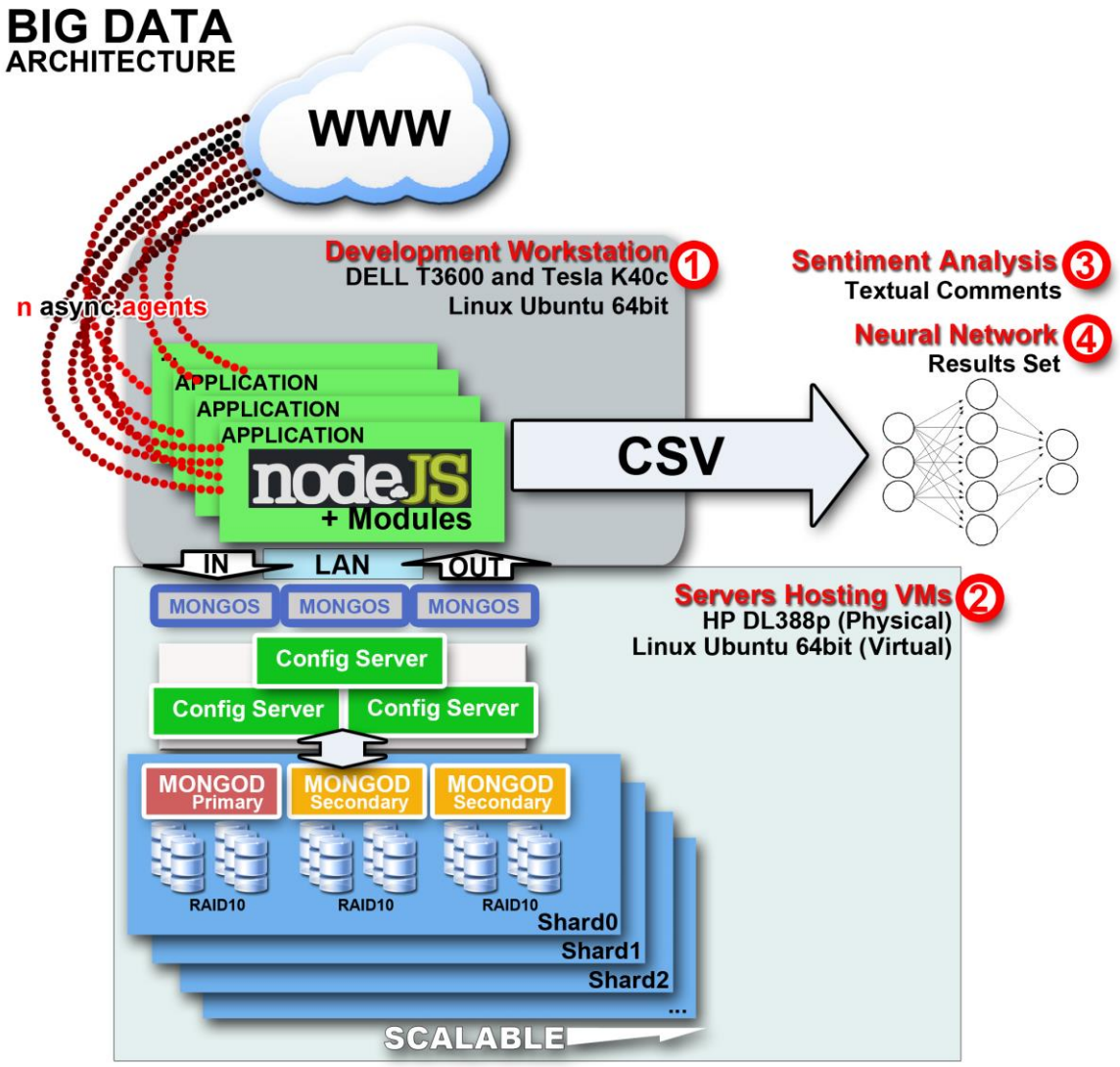
Zhu, F. and Zhang, X. (Michael). (2010), “Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics”, *Journal of Marketing*, Vol. 74 No. March, pp. 133–148.

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 Variables examined in this research

Note:

I = Discount Value; **II** = Discount Rate; **III** = Current Price; **IV** = Free Delivery; **V** = Customer Review Rating (Valence); **VI** = Number of Customer Reviews (Volume); **VII** = Percentage of Negative Review; **VIII** = Percentage of Positive Review; **IX** = Numbered of Answered Questions; **X** = Sales Rank.

Figure 4. The sentiment analysis process

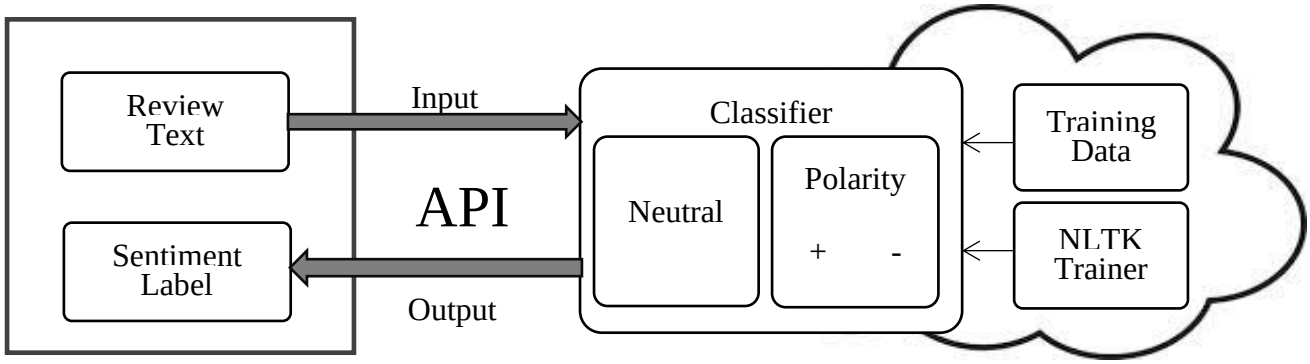


Table 1 Summary of variables used in this study

Variables	Description	Operationalization
Discount Value	The monetary value of price deduction from usual price	The value of price deduction listed as 'You Save' in Amazon.com, all in GBP
Discount Rate	The percentage of price deduction from usual price	The percentage of the price deduction listed as 'You Save' in Amazon.com, all in percentage (%)
Current Price	The price of the product	The price of the product listed on Amazon.com. If there is any discount, this is the price of the product after the discount
Free Delivery	Whether the product is delivered without delivery fee	Binary variable coded as 1 or 0. 1 denotes the product is delivered for free in UK as listed in Amazon.com, while 0 denotes there will be a fee charged for delivery
Customer Review Rating (Valence)	The accumulated average numeric rating of online review	The averaged numeric rating of all customer reviews for the product, listed on Amazon.com, ranged from 0 to 5
Number of Customer Reviews (Volume)	The number of all online reviews	The total number of all online reviews of the product, listed on Amazon.com, if there is no customer review, it is counted as 0
Percentage of Negative Review	The proportion of 1-star and 2-star reviews in total reviews	The percentage of negative reviews, calculated by dividing the number of reviews rated as 1 and 2 stars by total volume of reviews
Percentage of Positive Review	The proportion of 4-star and 5-star reviews in total reviews	The percentage of positive reviews, calculated by dividing the number of reviews rated as 4 and 5 stars by total volume of reviews
Review Text Sentiment (Sentiment)	The sentiment of most helpful reviews	Ternary variable coded as 1, -1, or 0. 1 denotes the total sentiment of most helpful reviews is positive, 0 denotes the sentiment is neutral, and -1 denotes the sentiment is negative
Number of Answered Questions	The number of answered questions in Customer Questions & Answers	The total number of answered questions of the product in Customer Questions & Answers, listed on Amazon.com
Sales Rank	The Best Sellers Rank of the product	The ranking of the product listed on Amazon.com. This rank reflects recent and historical product sales on Amazon.com and indicates the product's overall selling. In this study, it is used as a proxy of product sales situation

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%