DEVELOPMENT OF CONCEPTUAL MODEL FOR SOCIAL COMMERCE RESEARCH THROUGH INTEGRATION WITH BIG DATA ANALYSIS

Xuemei Tian  
*Swinburne University of Technology*, stian@swin.edu.au

Libo Liu  
*Swinburne University of Technology*, liboliu@swin.edu.au

Kristijan Mirkovski  
*Victoria University of Wellington*, kmirkovski@swin.edu.au

Mengxiang Li  
*University of Wollongong*, mli@uow.edu.au

Follow this and additional works at: [http://aisel.aisnet.org/pacis2016](http://aisel.aisnet.org/pacis2016)
Abstract

Information systems designers face great opportunities and challenges in developing a holistic big data research approach for the new analytics savvy generation. In addition, business intelligence is largely utilized in the business community and thus can leverage the opportunities from the abundant data and domain-specific analytics in many critical areas. The aim of this paper is to assess the relevance of these trends in the current business context through evidence-based documentation of current and emerging applications as well as their wider business implications. In this paper, we use BigML to examine how the two social information channels (i.e., friends-based opinion leaders-based social information) influence consumer purchase decisions on social commerce sites. We undertake an empirical study in which we integrate a framework and a theoretical model for big data analysis. We conduct an empirical study to demonstrate that big data analytics can be successfully combined with a theoretical model to produce more robust and effective consumer purchase decisions. The results offer important and interesting insights into IS research and practice.

Keywords: Big Data Analysis, Social Commerce, Business Intelligence.
1 INTRODUCTION

Since the 1960s, the interplay between hardware, software, and communications has led to previously unforeseen advances in information systems. These developments have been accompanied by shifts in the relationship between ‘carrier and content’, and between manifestations of the latter as various data, information, knowledge, and indeed wisdom (Buchholtz et al., 2014). During this time, concerns over data and information overload increased, along with those focused on related organizational and managerial challenges (Reich and Benbasat, 2000).

One response to these concerns saw the emergence of sub-disciplines, such as competitive and business intelligence, and data, information, and knowledge management (Biere, 2003; Williams and Williams, 2010; Wixom et al., 2011). Others came from management and organizational theory, notably the resource-based theory, which emphasized the importance of internal capabilities and competencies (Wernerfelt, 1984; Barney, 2001), and later, knowledge-based views (Eisenhardt and Santos, 2002; Spender and Grant, 1996), which focused on the heterogeneous bundles of intangible resources (i.e., imperfectly mobile, imperfectly imitable and non-substitutable) as a basis for competitive advantage and organizational success.

As understanding of the nexus between information technology, information systems, and business has increased, another core lesson has been learned – that technology on its own is unlikely to provide the answer to the questions continually posed by the management (Reynolds and Yetton, 2015). Whether it is relational databases, data mining, or enterprise systems, the latest ‘big thing’ usually turns out to be much less of a transformative force when introduced to the market. This is as likely to be true for the latest such breakthrough, namely Big Data (Lohr, 2012). How to better utilize big data assets, in addition to business assets, and human capital, to create value has become a fertile ground for corporate competitive advantage. As big data analysis becomes the next frontier for advancement of knowledge, innovation, and enhanced decision-making process, the significance of its impact on society as a whole should not be underestimated (McAfee et al., 2012).

The aim of this paper is to assess the relevance of these trends in the current business context through evidence-based documentation of current and emerging applications as well as their wider business implications. In this paper, we use the BigML software package to investigate how two social information channels (i.e., friends-based and opinion leaders-based social information) influence consumer purchase decisions on social commerce sites. We undertake an empirical study by combining a theoretical model with the analysis of big data using BigML. This methodology demonstrates that big data analytics can be successfully combined with a theoretical model to produce more robust and effective consumer purchase decisions. However, for this to happen, big data technologies must be further developed to cope with the proposed big data analysis framework.

The rest of the paper is organized as follows. In section two, we provide a review of the literature related to big data, business intelligence, social information processing theory and online consumer reviews. In section three, we identify the research gap based on the literature review. In section four, we conduct an empirical study to develop and explain our methodology and compare the results from different models. In section five, we provide robust checks for our final model and conclude by discussing the wider implications of our study.

2 LITERATURE REVIEW

2.1 Big Data and Business Intelligence

Big data has been labelled as the fifth wave in the technology revolution, after the mainframe, the PC, the Internet, and Web 1.0 eras, and more recently, mobile and Web 2.0 eras (Freeland, 2012). It combines an architectural paradigm shift in data movement in which instead of bringing data for centralized computation, the aim is to push the computation to distributed locations (Wixom et al.,
Big data analytics have been used to describe data sets and analytical techniques in applications that are so large (from terabytes to exabytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis, and visualization technologies (McAfee et al., 2012). Considering the seemingly limitless spectrum, big data can utilized for social networks, web server logs, traffic flow sensors, satellite imagery, broadcast audio streams, banking transactions, MP3s of rock music, the content of web pages, scans of government documents, GPS trails, telemetry from automobiles, financial markets and so on (Dumbill, 2012). Dumbill (2012) explains that although any data can be regarded as ‘Big’ when size is too large to be handled by conventional systems, it is also the case that the dramatically increased volume (there is too much of it), velocity (it is moving too fast), and variety (it is not structured in a useable way) do not longer fit in the structures of database architectures (Dumbill, 2012). More recently, three more characteristics have been mentioned by other researchers, such as Daniel (2014): (1) value (a source of competitive advantage); (2) veracity (the biases, noise and abnormality in data), and (3) verification (refers to data verification and security). Daniel (2014) argues that these three attributes also represent Big Data’s fundamental properties as they are linked to data accuracy, a concept associated with the longevity of data and their relevance to analysis outcomes, as well as the length required to store data in a useful form for appropriate value-added analysis.

Furthermore, while there is plenty of enthusiasm in business circles for big data, there are those who argue that there is little special about it, seeing it as merely a magnification of the features found in knowledge management, and others predicting that its popularity will eventually wane owing to its complexity and a shortage of qualified workers (Boulton, 2013). It is also important to balance matters of data size against the context of its application. A survey conducted jointly by IBM and Oxford University Business School revealed that 30 per cent of respondents did not know what ‘big’ meant for their organizations (Schroeck et al., 2012). Elsewhere it has been observed that considerations of adequacy and relevance might require ‘not so big data’ (Zhang, 2013), something that might apply to smaller businesses, who are also expected to adopt these technologies, for example through SaaS or as a component of ERP systems (Economist, 2010).

An important component of the ongoing case for big data is calls for education and training of a new breed of statisticians and analysts for the proper use of BI applications (Lund, 2013), big data scientists operating as multifunctional problem solvers communicating between different departments (Davenport, 2006). Likewise, the professional development and career progression of in-house analysts – already familiar with the organization’s unique business processes and challenges – has also been identified as a top priority for business executives (Schroeck et al, 2012). This might well extend to the operational personnel, to reflect their needs for analytical and decision making competencies consequent upon ongoing systems integration and the breaking down of traditional organizational hierarchies (Berner et al., 2014).

As with other such concepts, the meaning and interpretation of Business Intelligence (BI) has shifted over the years. At its simplest, BI refers to the ability to use information to gain a competitive advantage. This includes data on education, skills, and the past performance of employees to help businesses identify the critical talent within their organizations and ensure its development and retention (Davenport, 2006). BI is often referred to as the techniques, technologies, systems, practices, methodologies, and applications that analyze critical business data to help an organization better understand its business and respond to the market through timely decisions (Chen et al., 2012). Moving beyond BI to the development of more advanced apps and an enhanced ability to drill down and answer the questions asked by BI engines (Bennett, 2014), BI and Big Data initiatives must advance together, rather than separately on parallel tracks. The real value lies in integrating the existing BI and analytics capabilities with the new big data technologies and techniques to focus on how these new capabilities can augment and extend the existing environment (Bennett, 2014).

It is clear that much work remains to be done for the anticipated benefits of big data to be realized. The remaining issues of clarification around the specific content of these large datasets (primitive data elements, information, knowledge or meta-knowledge) and the relationship between terms like data,
information, and knowledge (Zhang, 2013). There are also risks to do with the misuse of data or with ‘false discoveries’ in trying to find meaningful needles in massive haystacks of data (Davenport, 2006). Hence, this calls for improvement in such areas as data analysis, acquisition, extraction and cleaning, integration, representation, and interpretation thorough integration of systems. More generally, there are matters of the benefits and risks of big data, both in organizational terms and as regards society as a whole, whether these relate to potential problems in the areas of privacy, security and ethics, or to emergence of a new version of the Digital Divide, a digital network divide (Buchholtz et al., 2014; McNeely et al., 2014). Critically, this ‘softer’ dimension to developments in big data extends to behavioral and cultural issues at organizational level, and would extend to a much wider set of norms and behaviors.

The promises of Big Data are real, however, there is currently a wide gap between its potential and realization. Many leading researchers, such as Agrawal, et al. (2012) and Zicari (2014), emphasize the opportunities and challenges in big data research and argue that the challenges include not just the obvious issues of scale, but also heterogeneity, lack of structure, error-handling privacy, timeliness, provenance, and visualization. These problems impede the progress at all phases in the pipeline (acquisition/recording, extraction/cleaning annotation, integration/aggregation/representation, analysis/modelling, and interpretation). Better understanding and management of these problems is the key of creating business value from data (Agrawal et al., 2012). The task of creating real value from data cannot be achieved by simply by focusing on development of new technologies; it also requires us to fundamentally rethink how we manage data analysis. The purpose of this paper is to introduce a concept on how to manage data analysis.

3 RESEARCH FRAMEWORK

In the foregoing literature review a number of commonalities emerged between sources. Firstly, researchers examined knowledge contribution in general, with a special focus on knowledge sharing. Although some studies noted that social relational activities influence knowledge sharing (Liang et al., 2011), few studies explicitly investigated social relational activities. Secondly, most studies relied on subjective data collected through surveys, case studies, and focus groups to explore how and why customers participate in online social communities. Straub et al. (1995) argued that actual usage and perceived usage are not always congruent. Instead of using self-reported contributions, we used field data to measure the actual contributions. Thirdly, we worked on the premise that extracting useful knowledge from big data requires not only scalable analytics services, but also theoretical support. Boyer et al. (2010) identified the obstacles to successful BI projects as not as much technological, as lying within the organization itself. They thus advocated strategic enterprise BI programs that would enable business users to make informed decisions based on the collaborative leveraging of enterprise-wide information. This prediction seems relevant today and to the circumstances of big data, whose analytics-driven insights must be closely linked to business strategy, easy for end users to understand, and embedded into organizational processes to enable action at the right time (Balboni and Cook, 2012). The strategic significance of such activities is emphasized by Elbashir et al. (2013) who recognize the critical importance of collaborative knowledge synergies between senior management, CIOs and IT managers in facilitating improved decision making across the organization’s value chain. For present purposes this is a reminder of the importance of the knowledge component, both in relation to current expectations of big data, and in acknowledgement of the fact that many organizations are still failing to use available data and knowledge effectively, let alone being in a position to accommodate the demands of big data.

Admittedly big data really could turn out to be different, but another lesson from experience is that the more data and information are available, the greater the need for human judgement in decision making. Even in reporting big data-related advances in cognitive augmentation and job automation, McGovern (McGovern, 2015) concedes that there are still many tasks where the combination of humans and machines produce superior results. Processing data requires not only the right technologies and
analytic tools/techniques, but also consideration of the wider organizational dimension. This will mark the transformation from big / massive data to knowledge which can be used for making business predictions and decisions. This raises questions not only as to the relative value of big data predictions and knowledge-based decisions, but also as to future relationships between big data and knowledge management. In this paper, to support the above arguments, we use two empirical studies to investigate the influence of social relational activities on customer purchase decisions. We also compare the findings from two studies, which indicate that big data analysis may not provide the best results without consideration of the organizational dimension.

4 EMPIRICAL STUDY

We conducted an empirical study by combining a big data tool (BigML) and a conceptual model. We used BigML to examine how the two social information channels (i.e., friends-based social information and opinion leaders-based social information) influence consumer purchase decisions on social commerce sites.

4.1 Data Collection

The data for this study was collected from a popular social commerce site, which provides a platform for consumers to share their experience, and to interact with other consumers. A consumer in the social commerce site can rate, for example, a product (of a particular brand) while sharing her or his experience about this product. The consumer can also follow other consumers whose posts or ratings are useful.

In total, the website includes 72,674 products from different brands. All of brands were used for analysis in this study. Based on the customers ID, we crawled the network data for each customer from the community. Specifically, we collected the “following” and “follower” of customers’ lists and built an egocentric network for each customer. Then we crawled friends-based social information (i.e., friends’ review valence and friends’ purchase) and opinion leaders-based social information (i.e., opinion leaders’ review valence and opinion leaders’ purchase).

4.2 Operationalization of Variables

Opinion leaders’ review valence is operationalized as the total score of ratings (on products in a particular brand) provided by individuals who are the members’ reference persons. Friends’ review valence is operationalized as the total score of ratings provided by individuals who are friend of a consumer in the social commerce site.

Opinion leaders’ purchase is operationalized as the total number of products in the buy-lists of a consumer's reference persons in the social commerce site.

Friends’ purchase is operationalized as the total number of products in the buy-lists of a consumer's friends in the social commerce site.

Consumer purchase behavior is operationalized as the number of products in a consumer's buy-list. Such operationalization was considered because consumers can add products to their buy-lists to indicate that they have already bought specific products in the social commerce site.

4.3 Data analysis tool – BigML

BigML (https://bigml.com) is an approach to machine learning. Users can set up data sources, create, visualize and share prediction models, and use models to generate predictions. It implements a set of basic data mining procedures. BigML includes several advanced procedures that support interactive analysis of big data. For example, data histograms are first rendered on a data sample and then updated when the entire data set is loaded and processed. Decision trees provide a supervised machine learning
technique that starts with the root classification node, which is then iteratively refined. Decision trees in BigML are rendered on the client application dynamically, and evolve as their nodes are induced on the computational server. More important, the process of preparing a model in BigML can be divided into several consecutive steps: 1) data preparation, 2) association model learning, 3) model selection, 4) model testing, and 5) development of model.

4.4 Phase 1: BigML analysis – Base model

To analyze our data in BigML, we set consumer purchase decisions as our objective. We first used BigML to run the selected dataset with the Base Model (1 Click Model). The results of the decision tree are shown in Figure 1. This indicates how important each factor is in the prediction of consumer purchase decisions. Friends’ purchase is the most important factor, which is followed by opinion leaders’ review and opinion leaders’ purchase. The result of the BigML analysis ran with the Default Model is not consistent with the regression analysis. This means that the Default Model does not fit the business rules well, which are artifacts of the data.

![Decision Tree](image)

Figure. 1 Results from training data with the default model

BigML’s Default Model does not fit the business rules, and therefore we should adjust the training model by integrating a conceptual framework from the existing literature.

4.5 Phase 2 BigML analysis – Integrating the Conceptual Model into the Base Model

4.5.1 Conceptual Research Model

Based on social information processing theory, we propose a conceptual research model, which should be integrated with the Default Model of BigML. Consequently we developed the following hypotheses:

H1a. Friends’ review valence will positively influence ego-consumer purchase behavior.

H1b. Friends’ purchase will positively influence ego-consumer purchase behavior.

H2a. Opinion leaders’ review valence will positively influence ego-consumer purchase behavior.

H2b. Opinion leaders’ purchase will positively influence ego-consumer purchase behavior.

Based on the conceptual framework, we conducted the following steps to adjust the model in BigML: 1) association model learning; 2) model testing; and 3) development of model of purchase decision marking. The results from running the data set with the adjusted model are shown in Figure 2.
results indicate that the adjusted model fits the theoretical model better and therefore provides more accurate results concerning the antecedents of consumers’ purchase decisions.

![Diagram of consumer purchase decision and opinion leaders' review](image)

**Figure.** 2 BigML analysis model 2

4.6 Robust check

Using a count variable summing up all the products that are added to a buy-list represents consumer purchase decision. Negative binomial regression is often used to analyze count data. Thus we use negative binomial regression to analyze the dataset. To test the robustness of the adjusted model from BigML, we ran a negative binomial regression analysis for the model.

\[ \text{Consumer Purchase Decision} = \beta_0 + \beta_1 \text{Opinionleader Review} + \beta_2 \text{Friend Review} + \beta_3 \text{Opinionleader Purchase} + \beta_4 \text{Friend Purchase} + w \]  

(1)

The results indicate that friends’ reviews valence have a significant positive effect \((\beta = 0.012, p < 0.001)\) on consumers’ purchase decisions. The regression results suggest that opinion review also has a significant positive effect \((\beta = 0.051, p < 0.001)\) on consumers’ purchase decisions. The result reveals that opinion leaders’ review valence exerts a significant positive moderating effect on consumers’ purchase decisions \((\beta = 0.006, p < 0.001)\). The positive coefficient \((\beta = 0.035, p < 0.001)\) indicates that opinion leaders’ purchase exerts a significant positive effect on consumers’ purchase decisions. Furthermore, the results are consistent with the adjusted model of BigML. Therefore, our model developed in BigML is robust.

5 DISCUSSION AND CONCLUSION

5.1 Research Implications

Recently, the Big Data era has descended on many communities, from governments to e-commerce to health organizations. To avoid being data driven, the big data analysis process should be adopted to leverage the opportunities presented by the abundant data and domain-specific analytics. In addition, while machine learning algorithms exist, most of them produce “black box” models, which are difficult to understand. This study presents a complete implementation of a big data analysis process by integrating machine learning with a theory based on business rules from the existing literature. Instead of developing machine learning models solely on specific training datasets, researchers should integrate theory based business rules to adjust the machine learning models and use test datasets to test the adjusted machine learning models in order to attain comprehensive models that are more relevant and better reflect the reality (see Figure 3).
Figure 3 Big data analysis process

The literature of big data emphasises the application of algorithms to pattern analysis and prediction. We have created a big data processing model to emphasise that processing data requires not only the right technologies and analytic tools/techniques, but also consideration of the wider organizational dimension. This will drive the transformation from big / massive data to knowledge to business value processes (as shown in figure 5 below). This raises questions not only as to the relative value of big data predictions and knowledge-based decisions, but also as to future relationships between big data and knowledge management. Big Data is still in its infancy and is quintessentially a technology or set of still developing technologies. Experience has shown time and again that the contribution of new technologies is dependant very much on organizational and especially, human factors.

5.2 Conclusion

Making sound business decisions based on accurate and current information and knowledge requires more than simple intuition, and BI has become indispensable to organizational success in the global economy. We need not only to be skilled engineers, but also domain experts. Big data will become big knowledge through the combination of data mining of association rules and the conceptual model of business rules. Data mining and domain experts will be able to use this for extending the possibilities of the model not only for user-defined business rules, but also for models generated from association rules gained from machine learning.

This paper started from the perception that big data and business intelligence, while facing challenges in developing systems for a new data savvy generation, had the potential to leverage opportunities presented by the abundant data and domain-specific analytics needed in many critical areas.

This study assessed the relevance of these trends in the current business context through evidence-based documentation of current and emerging applications as well as their wider business implications. We used the comprehensive model after verifying that the machine learning model was inaccurate and thus unsuitable to the task in hand. The proposed framework was tested using in an empirical study the results of which suggest that big data can become big knowledge by the combination of data mining of association rules and a conceptual model of business rules. This big data analysis framework and theoretical model contributes to the existing literature and offers important and interesting insights to IS research and practice.

Reference


