POMS PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT POMS

Vol. 27, No. 10, October 2018, pp. 1745–1748
ISSN 1059-1478 | EISSN 1937-5956 | 18 | 2710 | 1745

DOI 10.1111/poms.12892

Big Data in Supply Chain Management

Nada R. Sanders
D'Amore-McKim School of Business, Northeastern University, Boston, Massachusetts 02115, USA, n.sanders@northeastern.edu

Ram Ganeshan*
Raymond A. Mason School of Business, The College of William and Mary, Williamsburg, Virginia 23185, USA,
ram.ganeshan@mason.wm.edu

"Big data" has become ubiquitous. It is impacting every aspect of how companies organize and manage their supply chains. Supply chains are evolving into digital networks connected via devices and sensors revolutionizing how data is generated, shared, and communicated. It is also unlocking new research streams. In this study, we introduce papers in this issue and showcase Big Data research trends across supply chain management.

Key words: big data; supply chain management
History: Received: September 2017; Accepted: March 2018 by Kalyan Singhal, after 3 revisions.

1. Introduction

In addition to the data generated by traditional transaction-based enterprise systems (e.g., POS, RFID, ERP), supply chains and enterprises now generate vast amounts of data from unstructured data sources. Today's supply chains are heavily instrumented – sensors, tags, trackers, and other smart devices are collecting data in real time on a wide variety of business processes. These also include Internet data (e.g., Google Trends, online reviews, social media postings, blog/wiki entries, and forum discussions), digital clickstreams, camera and surveillance footage, imagery, and environmental data (e.g., weather, energy usage). Advances in computing architecture such as cluster computing and cloud computing have enabled the storage, retrieval, analysis, sharing, and distribution of data and insights easier and cheaper. As one reads through this issue, one will find examples of innovative ways to generate, collect, and analyze data across various supply chain contexts.

What does this deluge of data mean for supply chain research? It includes use of new features and variables to derive new insights, formulating models to describe new phenomenon, and integrating existing methodologies or developing new methodologies to deal with large datasets.

Through this Special Issue, we intend to provide a snapshot of the current research using big data in supply chain management. We have received an enthusiastic response to our call with forty-five papers submitted. We showcase eight of those papers in this issue, each demonstrating some innovative approaches and/or interesting problem contexts.

Also, as preparation for this Special Issue, we spent a considerable amount of time over the last two years communicating with data science leaders, academic colleagues, and policy makers. We synthesize these discussions on current big data research trends and overlay papers in this issue, described in section 2. In section 3, we offer our concluding thoughts.

2. Research in This Issue

A majority of the articles bring unique applications in specific industries (e.g., online apparel retail, medical device, online specialty food store, emergency physician network), while the rest focus specific issues in broader contexts (e.g., implications of analytical capability, techniques of dealing with Big Data). We will briefly summarize each article and point out some of the common threads.

Four articles in this issue use **predictive analysis**. In "The Operational Value of Social Media Information," Cui, Gallion, Moreno and Zhang propose a comprehensive framework, involving feature extractions and statistical machine learning, to predict daily aggregate sales for an online exclusive retailer of men's clothing. They use Facebook posting data, which records the company's interactions with customers, to supplement the operational data to generate the sales forecast. Using the proposed framework, they are able to quantify the forecast accuracy improvement through utilizing social media inputs. In a similar spirit, Boone, Ganeshan, Hicks and Sanders show in their article titled "Can Google Trends Improve Your Sales Forecast" that keywords search data from Google Trends leads to improved sales forecast for multiple product categories for an online specialty food and cookware retailer. Lau, Zhang and Xue describe a sentiment-based sales forecast method in their article titled "Parallel Aspect-Oriented Sentiment Analysis for Sales Forecasting with Big Data." They use Web crawler data from major online sources (e.g., Amazon, Newegg, JD.com, Dianping) to learn not just product aspect taxonomy, but also the association between product aspects and user sentiments for each product category. The learning generates weekly sales forecast for specific products. In the article "RFID Tag Performance: Linking the Laboratory to the Field through Unsupervised Learning," Ellis, Rao, Raju, and Goldsby address the issue that firms fail to reap the predicted benefit of implementing RFID due to the tag reading issue in the field. By linking the laboratory testing data and the scanned field data, they develop an approach to predict the field reading performance of RFID tags, which is shown to significantly outperform the prediction by deploying traditional logistics regression.

Three articles provide **descriptive analysis** for specific phenomena. In "Leveraging Big Data to Balance New Key Performance Indicators in Emergency Physician Management Networks" contributed by Foster, Penninti, Shang, Kekre, Hedge and Venkat, measurement indices that reflect treatment complexity, revenue potential, amount of support for the physician by assistants or nurses, average performance of a facility and patient experience are developed to identify driving forces contributing to physician performance difference. Physicians in the network are clustered based on non-performance-based attributes, and then the disparities among the physician clusters are identified. The importance of this research lies in its recognition that the same productivity level may not be attainable for all physicians at all facilities, which allows for sensible performance evaluation of physician within and across facilities. In the article "Product Recall Decisions in Medical Device Supply Chains: A Big Data Analytic Approach to Evaluating Judgment Bias," Mukherjee and Sinha examine how judgmental bias in recall decisions is related to signal characteristics extracted from unstructured user feedback and situational contexts extracted from firm characteristics. In addition, they compare different methods for their prediction accuracy on product recall probability. The predictive model allows for estimating the error rates and judgmental bias. On a different domain, the article "An Investigation of Visibility and Flexibility as Complements to Supply Chain Analytics: An Organizational Information Processing Theory Perspective" contributed by Srinivasan and Swink describes the relationship between firm's analytical capability and supply chain characteristics. The authors identify that better demand and supply visibility are observed for firms with stronger analytical capability, and the relationship between the firms' analytical capability and their operational performance becomes stronger as the supply chains become more flexible.

Different from the aforementioned studies, the article contributed by Choi, Wallace and Wang, "Big Data Analytics in Operations Management," provides a systematic discussion on various techniques and methodologies deployed in existing studies using data analytics. This study spans over a wide range of topical areas, including forecasting, inventory management, revenue management and marketing, transportation management, supply chain management, and risk analysis.

It is worth noting common themes across these studies, closely connected to the characteristics on the Big Data. Almost all of the authors recognize the complexity associated with the large "volume" of the data involved in their problem contexts. For example, Cui's et al. analysis involves 171,279 unique users interacting with the firm's Facebook page with 25,730 comments and 266,534 likes, Boone et al. hint at the potentially thousands of Google keywords to choose from, Foster et al. match over 10 million patient visits with physicians and facilities, and Mukherjee and Sinha identify 108 firms and 1358 RFID devices for unique firm-device combination. The data involved also have high "variety." For example, taxonomy aspects from user comment on a retail product (Lau et al.) or user feedback on a medical device (Mukherjee and Sinha) are highly unstructured.

Theoretically, more data and variables suggest more information and better descriptive and predictive accuracy. This notion, however, is not necessarily true unless one fully understands the contexts of the data and designs the right approach. The way to handle volume and variety is not a simple extension or expansion of an existing model or an existing method. It is of paramount importance to carefully adapt and choose the techniques to deal with specific data and problem contexts to enable predictive and descriptive power, using Big Data. Computationally, the volume and the variety associated with Big Data introduce major challenges. A careful trade-off between the amount of information used and validity of the models needs to be examined. Techniques involving parallel computing can be used to address some of these issues. Lau et al. adopt such techniques in their sampling and learning algorithms to alleviate computational bottleneck for mining the Web crawler data from major online sources. Moreover, feature extraction, which helps to reduce the number of variables, must be conducted based on specific contexts and objectives of the studies. For example, though both using unstructured data generated by users, Mukherjee and Sinha apply Latent Dirichlet allocation model for classification of text data from user feedback in their study of medical device recall, while Cui et al. adopt neural tensor network to extract sentiment features from user comments. Foster et al., in contrast, adopt indices as surrogate measures to reflect treatment complexity and revenue potential, which is a common approach in physician evaluation.

In addition, data variety often induces added challenges associated with sparsity and heterogeneity, which require proper treatment in the analysis. For example, in their study on RFID performance, Ellis et al. impute missing data by applying random forest to predict the field reading rates. Srinivasan and Swink carefully control firm size and industry, construct variables for process complexity, organizational factors, technological infrastructure, and strength of leadership in their measurements for supply visibility, demand visibility, analytical capability, flexibility, cost performance, and delivery performance.

The choice of different techniques for effective prediction and description can be problem and context specific too. In generating sales forecasts, techniques including linear regression, ARIMA models, lasso regression, support vector machines, gradient boosting model, and extended extreme learning machine are studied by Cui et al., Boone et al., and Zhang et al. In predicting RFID reading performance, Ellis et al. adopt cluster analysis and association rule mining. In predicting recall probability, Mukherjee and Sinha compare triangulation of Lasso, stepwise generalized linear mixed model, and the stepwise generalized additive model.

3. Closing Thoughts

Each of the research areas discussed and papers showcased represent game changers. The challenge, however, is that each of the big data applications are not evolving in isolation of one another; rather they are simultaneously being implemented across supply chain networks creating a new and unfamiliar topography. Future research may likely see new supply chain models where sophisticated postponement strategies optimize a blend of production of semi-finished products in low-cost countries with high-level personalization occurring close to customers. A new level of mass customization may occur with customer-centric plants utilizing 3-D printing and high functioning robotics that innovate products based on customer profiles gathered in real time. One area of change will certainly include product innovation with new products being directly created from captured customer needs and new design capabilities that will likely become important strategic assets. Transportation and logistics will likely experience dramatic changes with autonomous vehicles and drones. Blockchains hold the promise of transparent supply chains. The supply chain itself will likely extend as "smart products" track the customer well after the purchase and "lead" the customer to future personalized experiences.

Until now researchers have typically worked on small subsets of larger problems. Big data have the potential to change this status quo and create an opportunity for OM and supply chain management researchers to address the truly relevant aspects of large-scale challenges in a meaningful way. What will this take?

First, we argue that to continue to be relevant in this new environment, to build and maintain our reputation as a serious discipline with impactful scholars, we need to leverage these newly available datasets not only to grow the supply chain knowledge base, but by taking on the "wicked problems" of our time from human caused climate change to delivering affordable healthcare to all. To do this, we need to include "all phases of science" (Singhal and Singhal 2012) in conducting big data research.

Second, we need to make what we do interesting to a larger part of society. How can big data-based supply chain research break that barrier? Quantum physics or astronomy, for example, captivates us with abstract and arcane ideas like dark matter and Black holes – outreach activities that translate into real research dollars. Big data gives our community this opportunity. We need to be able to tell stories with the data and our models that appeal and enthrall the greater public. Doing so will not only open up funding opportunities, but get us a "seat in the table."

Third, we need to encourage interdisciplinary and multi-method research. To meet the challenges posed by big data, we envision a scenario where supply chain researchers can make their insights richer and much more impactful drawing ideas and methods from a diverse set of disciplines. Our institutions, as many already do, need to celebrate and support such efforts. Our doctoral programs need to produce graduates who are comfortable with both interdisciplinary research and methodologies appropriate for handling big data ("petabytes of data").

Lastly, the big data method-paradigm is still in its infancy. There are multiple approaches to the data – how to separate the signal from the noise; how to calibrate and filter data; which algorithms to use; and how to integrate theory with the huge amount of data. We encountered several lively debates with authors, editors, and reviewers when putting together this issue. *Why should the quantity and quality of Facebook comments or Google searchers impact forecasts? Why is calibrating RFID signals from the tags a supply chain problem? How can an algorithm capture customer sentiment? Is cluster analysis the right tool to classify doctors?* We sensed what Richard Bernstein famously called the "*Cartesian anxiety*," and suggest, as adjudicators of these research papers, that we are open to divergent methodologies and not remain dogmatic in the methods we deem acceptable.

The scientific study of operations and supply chain management may be at a tipping point. We draw this conclusion by closely following advancements in big data technology; observing its sweeping impact on organizations and society; noting the escalating complexity and scale of contemporary societal challenges; and observing publications in leading journals, as well as this special issue. Big data analytics offers unprecedented challenges to old theories, established standards, new methodologies, and possibly beliefs within our field of inquiry. At the minimum, we can say that existing theories will likely collide with this torrent of data, some may survive while others not. We can certainly conclude that we are at the cusp of something big and have to up our game as a community to address this opportunity. Is it a paradigm shift? Maybe. The revolution is certainly brewing.

Acknowledgments

We are deeply appreciative of all the authors who submitted their manuscripts, the Associate Editors and reviewers who helped improve the papers found in this special issue. We are especially thankful to Kalyan Singhal, Editor-in-Chief of the *Production and Operations Management journal*, for his leadership and support in developing this special issue. Our special thanks goes to the reviewers of this study, especially Annabelle Feng, who gave us the right words to collate the themes in the essays and the papers.

Reference

Singhal, K., J. Singhal. 2012. Imperatives of the science of operations and supply-chain management. *J. Oper. Manag*. **30**(3): 237–244.