TRANSFORMATIONAL ISSUES OF BIG DATA AND ANALYTICS IN NETWORKED BUSINESS

Baesens B, Bapna R, Marsden JR, Vanthienen J, Zhao JL (2016)

September 21, 2023

Summery

Introduction

Big data has been defined by the 4 V's:

- volume
- velocity
- variety
- veracity

- The 4V definition of big data provides perimeters, but lacks guidance for operations within.
- Proposes a fifth "V" – *Value* – to the 4V framework for a business perspective.
- Extracting value from big data requires:
  - Understanding the four major sources of big data.
  - Recognizing prediction and causality as primary methodological paradigms.
- Big data can be classified by its origin:
  - Transactions vs. non-transactions.
  - Internal vs. external sources.
- Emphasizes the increasing relevance of data from outside corporate boundaries, like social networks and IoT.

- Identifies five major sources of big data in today's networked economies:
  - **Large-scale enterprise systems:** Systems like ERP((enterprise resource planning), CRM(customer relationship management), and SCM (supply chain management) that have been in use for almost two decades.
  - **Online social graphs:** Social networks like Facebook, Twitter, Weibo, and WeChat where users' interactions leave a trackable digital trail.
  - **Mobile devices:** Almost 5 billion handsets globally, where every user action can be tracked and possibly geotagged.
  - **Internet-of-things:** Sensor-enabled ecosystem connecting objects and humans, generating big data.
  - **Open data/public data:** Increasingly available data on weather, traffic, maps, and more.

- **Business Analytics vs Business Intelligence:**
  - Analytics is more than advanced reporting; it links variables to business outcomes.
  - Focuses on predictive or causal inferences.
  - Encompasses techniques from statistics, econometrics, machine learning, and computing.
- **Difference from Data Science:**
  - Business analytics focuses on organizational data-driven decision-making.
- **Industry Response:**
  - Businesses are integrating data specialists.
  - Academic institutions are developing programs in data analytics.
  - McKinsey report highlighted a shortage of analytical experts in the U.S.

- **Big Data's Transformational Impacts:**
  - *For businesses:*
    - Affects business processes.
    - Rebalances power dynamics in decision-making.
    - Changes optimization challenges.
  - *For academics:*
    - IT research integration with other fields.
    - Opportunity for data triangulation.
    - Challenges to develop new methodologies.
- **Power in the Age of Big Data:**
  - Power = Information (data) + Trust.
  - Focus has been on data, often neglecting trust.
- **Research Gap:**
  - Although data quality concerns are not new, the transformational impacts and growing role of big data add significance to the importance of data quality

Data Quality

Data Quality and Its Significance

- The GIGO (garbage in, garbage out) principle: Data quality is crucial for analytical models.
- Quality varies with application: sales forecasting, risk management, medical diagnostics.
- Data quality is multidimensional: accuracy, completeness, latency, security, interpretability, traceability (Moges 2014).
- Despite its importance, investments in data quality are often overlooked for perceived high costs.

The Trust and Performance Equation

- Trust in data and analytics is critical for decision-making.
- Performance boosts trust: Effective models enhance decision-making.
- Even with quality data, without trust, analytics won't be adopted.
- Improving data quality improves analytical performance leading to profit gains or cost savings.

Data Triangulation and Corporate Governance

- Triangulation validates data using (1)direct market observations, (2)indirect empirical data, and (3)survey data.
- Consistency across various data sources enhances quality.
- Few firms have dedicated roles ensuring data quality throughout.
- Quality data demands governance and oversight: (1) initial collection, (2) storage and updating, (3) retrieval, and (4) processing and preparation for analysis.
- Senior management's trust and understanding of data's importance is crucial for investing in quality data.

Methodological Paradigms and Challenges of Big Data

Big Data: Methodological Paradigms and Opportunities

- Big data combines techniques from machine learning, statistics, econometrics, and experimental design.
- Global networked economy acts as a large-scale laboratory for real-world experiments.
- Opportunities:
  - (1) effects of peer influence, (2) impacts of the influence of dynamic ties, (3) impacts of anonymity on online relationships, (4) results from alternative pricing strategies for digital media, (5) the sway of carefully designed next-generation recommender systems, and (6) the changing preference structures of Generation Y and Z consumers.
  - Generate new causal insights, test age-old norms, and build and test new theories.

- Recent experiments by Facebook and OkCupid raise ethical questions about online social experimentation.
- Benefits: Avoid bad decisions, better understand human interactions, and optimize resource allocation.
- Precedence of experimentation in advancing scientific knowledge.
- 33% of recent marriages in the U.S. start online, making digital platforms pivotal.

Moving Forward: Ethics and Technology

- Ethical considerations paramount when conducting large-scale field experiments.
- Universities' ethics committees and institutional review boards set standards.
- Companies should collaborate with academia or establish ethics committees.
- Societal institutions need to adapt and catch up with technological advancements.

Innovative Big Data Applications

Online-to-Offline Commerce

- O2O commerce as a major business trend, especially in China (Lau et al. 2015).
- Goes beyond traditional "clicks and mortar" model by creating a closed loop of online and offline customer activities (Bernstein et al. 2008).
- Distinct from B2B, B2C, C2B, and C2C due to its offline component inclusion (Xiao and Dong 2015).
- New generation of eCommerce, necessitating business model changes.
- Intrinsically tied to big data: O2O leverages data from location-based services, mobile computing, and the Internet-of-things.
- Example: QB House (Japanese barbershop), with sensors in barber chairs for seat availability and haircut time analytics, integrates this with online systems and performance analysis.

Networks of Smart Vehicles

- Smart vehicles, an Internet-of-things example, use on-board diagnostics (OBD) to monitor real-time behaviors of drivers and car components (Stankovic 2014).
- The data aids in decision-making through analytics and value-driven modeling for vehicle owners, managers, and insurance firms.
- Insurance companies harness this data to offer rate incentives, promoting better driving behaviors.
- Significant impacts:
  - Formation of novel business models in vehicle and traffic management.
  - Benefits: enhanced traffic flow, fairer insurance rates, and improved fleet management.
- Usage-based insurance (UBI):
  - Leads to innovative business models for insurance companies.
  - Aids in combating insurance fraud and suboptimal schemes.
- As vehicles increasingly adopt OBD devices, insurance firms aim to develop models targeting safe drivers and bolstering profits.
- Emergence of alliances involving insurance companies, fleet-owners, and banks to structure optimal profit strategies.

Proactive Customer Care

- Location-based services provide travelers with real-time information, aiding in on-the-move purchasing decisions.
- Companies can utilize these services to issue eCoupons, attracting customers to premium products or specific stores.
- By harnessing real-time big data and analytics, firms can comprehend shifting customer needs based on their changing locations.
- Emergence of the Location Based Marketing Association showcases the business significance of these services, with members from diverse sectors.
- These services enable a proactive business approach, ensuring customers are catered to whenever and wherever a need arises.
- Shifting business paradigms:
  - Upending the traditional "location is king" principle in sales.
  - Diminishing emphasis on the "line of sight" rule for store locations, as mobile maps and location services guide consumers.
  - eCoupons coupled with location-based services can effectively steer customer traffic.
- The evolving model raises captivating research queries that challenge long-held business beliefs.

Disruptive Impacts of Big Data

Business Analysts Retooling

- The evolution of business models necessitates proficiency in big data capture, analysis, modeling, and decision-making for business analysts (Waller and Fawcett 2013).
- Current challenge:
  - Many business analysts, even recent graduates, lack adequate training in big data and analytics.
  - This knowledge gap represents a significant challenge for businesses and their existing personnel.
- There's an imperative need for developing and retooling business analysts to harness the value-added potential of big data.
- Educational Institutions:
  - Many universities aren't yet equipped to offer advanced big data courses.
  - A notable number of faculty members are unprepared to deliver efficient big data education.
- Potential Solution:
  - Forge alliances between universities and enterprises to curate academic and professional big data programs.
  - Innovative departments can model their progress on early

Integration of Data and Social Sciences

- The advent of micro-level behavioral data encourages collaboration among company researchers, social scientists, and data scientists.
- "Big data" fosters innovative projects and offers profound insights into areas like human behavior, consumer choices, and social dynamics (Lin 2015).
- Existing psychological theories, such as the regulatory focus theory or the elaboration likelihood framework, are simplistic two-dimensional models.
- Large-scale data analysis can evolve these theories into intricate and realistic models with significant behavioral implications.
- This transformation might necessitate a recalibration of tools and mindsets for social scientists.

Breakdown of Traditional Business Boundaries

- Big data analytics is challenging traditional business boundaries.
- eCommerce giants like Alibaba and Tencent have ventured into banking, causing disruptions especially in China (Financial Times, January 5, 2015).
- These companies, leveraging their extensive customer base, offer services traditional banks can't. For example, Alibaba's Yu'ErBao fund offers real-time updates and transfers, becoming China's number one fund (Yu and Shen 2015).
- The rapid expansion of eCommerce in banking is attributed to their lower operating margins and not needing physical branches.
- Disruption arises not from banks' inability to retaliate, but from being unprepared for competition outside their industry.
- Emphasis on the importance of adapting to big data's value. Those lagging risk obsolescence in the face of agile competitors (Yan et al. 2015).

Challenges and Research Opportunities of Big Data and Analytics

Ubiquitous Informing:

- Refers to widespread recording, storing, and sharing of data by individuals and businesses.
- Enabled by advancements in mobile computing, video streaming, social networking, and IoT.
- Challenges: Extracting value from data and leveraging analytics to deliver business insights.

Implementation Environments:

- Stack environments like Hadoop manage big data.
- Importance of using such technology judiciously to avoid underutilization and potential legacy burdens.
- Alternative: More affordable in-memory setups.

Integration Issues:

- Identifying interrelationships in big data to extract value.
- Linking various data streams can provide comprehensive customer behavior insights.
- Example: In insurance, linking data streams can uncover fraud patterns.

Value Assessment in Big Data Analytics

Analytical Techniques Challenges:

- Models derived from statistics, econometrics, ML, and AI.
- These models focus on specific accuracy criteria, often hard for end-users to grasp.
- Crucial to bridge the communication gap between data scientists and decision-makers.

Trust and Value in Analytical Models:

- Need a common language emphasizing "value" for better understanding.
- Beyond statistical metrics, real value-based criteria should dominate. Applicability depends on the business context.

Value Assessment in Big Data Analytics: Continued

Understandability and Black Box Models:

- Analytical models should be comprehensible to decision-makers.
- While complex black box models may be accurate, they can be distrusted in critical applications.
- Debate: If black-box models prove consistently accurate, should they be favored over more transparent but less accurate models?

Operational Efficiency:

- Covers model evaluation, monitoring, and updating.
- In real-time settings (e.g., fraud detection), fast model evaluation is essential.
- Adaptable models are key to a company's success, especially as conditions change.

Regulatory Guidelines:

- Increasing regulatory guidelines for analytics due to its rising strategic and societal impact.
- Basel III Capital Requirements Accord exemplifies such regulation for credit risk modeling.
- Fragmented regulatory guidelines sometimes impede global analytical standardization.

Managing Automated Decisions:

- Big data analytics lead to increased automation in decision-making.
- This necessitates effective modeling and managing of business decisions.
- Separating decisions from processes can boost business agility.
- Decision model and notation standards bridge the gap between business process design and decisions (Taylor et al. 2013).

Enhanced Compliance Checking:

- Analytic techniques facilitate better compliance analysis, e.g., process analytics.
- Can be used for compliance checking and risk management (Caron et al. 2013).

Return on Investment and Trust:

- Analytical models should be profit-driven to build trust across an organization.
- Managerial decisions prioritize economic returns over statistical significance.
- Focus on quantifying analytical model ROI and embedding economic insights into model building.

Overview of Papers in the Special Issue

Overview of the Special Issue on Big Data Analytics 82 initial submissions were received, and after thorough evaluations, 11 finalist papers were accepted, each touching upon unique facets of big data analytics:

- Yahav et al.: Data-driven tree-based method to address self-selection bias.
- Zhang et al.: Personalized social brand advertising using large-scale social media data.
- Martens et al.: Leveraging fine-grained consumer behavior data for targeted marketing.
- Ghose and Todri: Measuring the impact of display advertising on user behavior.
- Saboo et al.: Time-varying effects model for real-time resource allocation.
- Brynjolfsson et al.: Crowd-based method for search trend data selection.
- Menon and Sarkar: Addressing scalability and privacy in transactional data sharing.
- Han et al.: Modeling relationships among choice, consumption, and utility in mobile app analytics.
- Breuker et al.: Predictive modeling for business process event data.
- Shi et al.: Framework for quantifying business landscapes in a digital economy.

Conclusion and Perspectives

Conclusions on the Special Issue on Big Data and Analytics

Editing a special issue presents a myriad of emotions for all involved parties. Through careful curation, editors aim to ensure a positive experience. This special issue offers a comprehensive perspective on the evolving domain of big data and analytics in Networked Business, serving as a potential cornerstone for future research, particularly in doctoral seminars.

- The paper provided a clear blueprint on organizing and presenting complex data, especially showcasing the transformational issues in a holistic level
- It emphasized the balance between theory and practical application.
- Featured a robust discussion section that was instructive while acknowledged limitations and proposed future research directions.
- These papers are informative when conducting my own research