



Business intelligence and analytics: From big data to big impact

N Abhishek, M Rakesh Chowdary

Assistant Professor, Department of CSE, Sree Dattha Group of Institutions, Hyderabad, Telangana, India

Abstract

Business intelligence and analytics (BI&A) has emerged as an important area of study for both practitioners and researchers, reflecting the magnitude and impact of data-related problems to be solved in contemporary business organizations. This introduction to the MIS Quarterly Special Issue on Business Intelligence Research first provides a framework that identifies the evolution, applications, and emerging research areas of BI&A. BI&A 1.0, BI&A 2.0, and BI&A 3.0 are defined and described in terms of their key characteristics and capabilities. Current research in BI&A is analyzed and challenges and opportunities associated with BI&A research and education are identified. We also report a bibliometric study of critical BI&A publications, researchers, and research topics based on more than a decade of related academic and industry publications. Finally, the six articles that comprise this special issue are introduced and characterized in terms of the proposed BI&A research framework.

Keywords: business intelligence and analytics, big data analytics, Web 2.0

Introduction

Business knowledge and examination (BI&A) and the related field of enormous information investigation have turned out to be progressively imperative in both the scholastic and the business groups in the course of recent decades. Industry thinks about have featured this noteworthy improvement. For instance, in view of a review of more than 4,000 data innovation (IT) experts from 93 nations and 25 ventures, the IBM Tech Patterns Report (2011) [15] distinguished business investigation as one of the four noteworthy innovation inclines in the 2010s. In a review of the condition of business examination by Bloomberg Businessweek (2011), 97 percent of organizations with incomes surpassing \$100 million were found to utilize some type of business examination. A report by the McKinsey Worldwide Establishment (Manyika *et al.* 2011) [27] pre-dicted that by 2018, the Unified States alone will confront a short-age of 140,000 to 190,000 individuals with profound investigative aptitudes, and also a deficiency of 1.5 million information keen supervisors with the know-how to examine huge information to settle on compelling choices.

BI&A evolution: key characteristics and competencies

The time period intelligence has been utilized by researchers in artificial intelligence for the reason that Nineteen Fifties. Enterprise intelligence have become a famous term in the business and it communities simplest within the 1990s. In the late 2000s, business analytics changed into delivered to symbolize the important thing analytical issue in bi (davenport 2006) [14]. Extra currently large data and big statistics analytics have been used to describe the records sets and ana- lytical techniques in programs which might be so big (from terabytes to exabytes) and complicated (from sensor to social media information) that they require advanced and precise records1“hal varian solutions your questions,” february 25, 2008 ([http:// www.Freakonomics.Com/2008/02/25/hal-varian-answers-your-questions/](http://www.Freakonomics.Com/2008/02/25/hal-varian-answers-your-questions/)).Garage, control, analysis, and visualization technol- ogies. In this newsletter we use business intelligence and ana- lytics (bi&a) as a unified time period and deal with big records analytics as a associated area that offers new directions for bi&a studies.

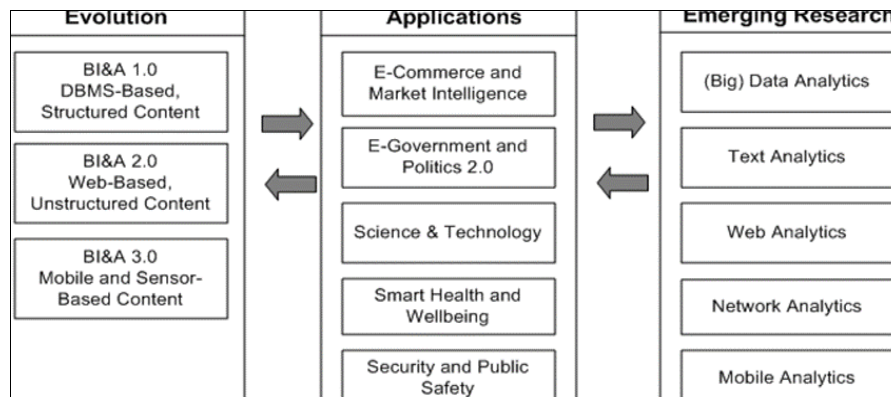


Fig 1: BI&A Overview: Evolution, Applications, and Emerging Research

BI&A Applications: From Big Data to Big Impact

Several global business and IT trends have helped shape past and present BI&A research directions. International travel, high-speed network connections, global supply-chain, and outsourcing have created a tremendous opportunity for IT advancement, as predicted by Thomas Freeman in his seminal book, *The World is Flat* (2005) [16]. In addition to ultra-fast global IT connections, the development and deployment of business-related data standards, electronic data interchange (EDI) formats, and business databases and information systems have greatly facilitated business data creation and utilization. The development of the Internet in the 1970s and

the subsequent large-scale adoption of the World Wide Web since the 1990s have increased business data generation and collection speeds exponentially. Recently, the Big Data era has quietly descended on many communities, from governments and e-commerce to health organizations. With an overwhelming amount of web-based, mobile, and sensor-generated data arriving at a terabyte and even exabyte scale (The Economist 2010a, 2010b), new science, discovery, and insights can be obtained from the highly detailed, contextualized, and rich contents of relevance to any business or organization.

Table 1: BI&A Evolution: Key Characteristics and Capabilities

	Key Characteristics	Gartner BI Platforms Core Capabilities	Gartner Hype Cycle
BI&A 1.0	DBMS-based, structured content <ul style="list-style-type: none"> ▪ RDBMS & data warehousing ▪ ETL & OLAP ▪ Dashboards & scorecards ▪ Data mining & statistical analysis 	<ul style="list-style-type: none"> ▪ <i>Ad hoc</i> query & search-based BI ▪ Reporting, dashboards & scorecards ▪ OLAP ▪ Interactive visualization ▪ Predictive modeling & data mining 	<ul style="list-style-type: none"> ▪ Column-based DBMS ▪ In-memory DBMS ▪ Real-time decision ▪ Data mining workbenches
BI&A 2.0	Web-based, unstructured content <ul style="list-style-type: none"> ▪ Information retrieval and extraction ▪ Opinion mining ▪ Question answering ▪ Web analytics and web intelligence ▪ Social media analytics ▪ Social network analysis ▪ Spatial-temporal analysis 		<ul style="list-style-type: none"> ▪ Information semantic services ▪ Natural language question answering ▪ Content & text analytics
BI&A 3.0	Mobile and sensor-based content <ul style="list-style-type: none"> ▪ Location-aware analysis ▪ Person-centered analysis ▪ Context-relevant analysis ▪ Mobile visualization & HCI 		<ul style="list-style-type: none"> ▪ Mobile BI

BI&A Research Framework: Foundational Technologies and Emerging Research in Analytics

The opportunities with the abovementioned emerging and high-impact applications have generated a great deal of excitement within both the BI&A industry and the research community. Whereas industry focuses on scalable and integrated systems and implementations for applications in different organizations, the academic community needs to continue to advance the key technologies in analytics. Emerging analytics research opportunities can be classified into five critical technical areas—(big) data analytics, text analytics, web analytics, network analytics, and mobile analytics—all of which can contribute to BI&A 1.0, 2.0, and 3.0. The classification of these five topic areas is intended

(Big) Data Analytics

Data analytics refers to the BI&A technologies that are grounded mostly in data mining and statistical analysis. As mentioned previously, most of these techniques rely on the mature commercial technologies of relational DBMS, data warehousing, ETL, OLAP, and BPM (Chaudhuri *et al.* 2011) [5]. Since the late 1980s, various data mining algorithms have been developed by researchers from the artificial intelligence, algorithm, and database communities. In the IEEE 2006 International Conference on Data Mining (ICDM), the 10 most influential data mining algorithms were identified based on expert nominations, citation counts, and a community survey. In ranked order, they are C4.5, k-means, SVM

(support vector machine), Apriori, EM (expectation maximization), PageRank, AdaBoost, kNN (k-nearest neighbors), Naïve Bayes, and CART (Wu *et al.* 2007). These algorithms cover classification, clustering, regression, association analysis, and network analysis. Most of these popular data mining algorithms have been incorporated in commercial and open source data mining systems (Witten *et al.* 2011) advances such as neural networks for classification/prediction and clustering and genetic algorithms for optimization and machine learning have all contributed to the success of data mining in different applications.

Two other data analytics approaches commonly taught in business school are also critical for BI&A. Grounded in statistical theories and models, multivariate statistical analysis covers analytical techniques such as regression, factor analysis, clustering, and discriminant analysis that have been used successfully in various business applications. Developed in the management science community, optimization techniques and heuristic search are also suitable for selected BI&A problems such as database feature selection and web crawling/spidering. Most of these techniques can be found in business school curricula. Due to the success achieved collectively by the data mining and statistical analysis community, data analytics continues to be an active area of research. Statistical machine learning, often based on well-grounded mathematical models and powerful algorithms, techniques such as Bayesian networks, Hidden Markov models, support vector machine, reinforcement learning, and ensemble models, have been

applied to data, text, and web analytics applications. Other new data analytics techniques explore and leverage unique data characteristics, from sequential/temporal mining and spatial mining, to data mining for high-speed data streams and sensor data. Increased privacy concerns in various e-commerce, e-government, and healthcare applications have caused privacy-preserving data mining to become an emerging area of research. Many of these methods are data-driven, relying on various anonymization techniques, while others are process-driven, defining how data can be accessed and used (Gelfand 2011/ 2012) [17]. Over the past decade, process mining has also emerged as a new research field that focuses on the analysis of processes using event data. Process mining has become possible due to the availability of event logs in various industries (e.g., healthcare, supply chains) and new process discovery and conformance checking techniques (van der Aalst 2012). Furthermore, network data and web content have helped generate exciting research in network analytics and web analytics, which are presented below.

Text Analytics

A significant portion of the unstructured content collected by an organization is in textual format, from e-mail communication and corporate documents to web pages and social media content. Text analytics has its academic roots in information retrieval and computational linguistics. In information retrieval, document representation and query processing are the foundations for developing the vector-space model, Boolean retrieval model, and probabilistic retrieval model, which in turn, became the basis for the modern digital libraries, search engines, and enterprise search systems (Salton 1989). In computational linguistics, statistical natural language processing (NLP) techniques for lexical acquisition, word sense disambiguation, part-of-speech-tagging (POST), and probabilistic context-free grammars have also become important for representing text (Manning and Schütze 1999). In addition to document and query representations, user models and relevance feedback are also important in enhancing search performance. In addition to the above research directions, text analytics also offers significant research opportunities and challenges in several more focused areas, including web stylometric analysis for authorship attribution, multilingual analysis for web documents, and large-scale text visualization. Multi-media information retrieval and mobile information retrieval are two other related areas that require support of text analytics techniques, in

addition to the core multimedia and mobile technologies. Similar to big data analytics, text analytics using MapReduce, Hadoop, and cloud services will continue to foster active research directions in both academia and industry.

Web Analytics

Over the past decade, web analytics has emerged as an active field of research within BI&A. Building on the data mining and statistical analysis foundations of data analytics and on the information retrieval and NLP models in text analytics, web analytics offers unique analytical challenges and opportunities. HTTP/HTML-based hyperlinked web sites and associated web search engines and directory systems for locating web content have helped develop unique Internet-based technologies for web site crawling/spidering, web page updating, web site ranking, and search log analysis. Web log analysis based on customer transactions has subsequently turned into active research in recommender systems. However, web analytics has become even more exciting with the maturity and popularity of web services and Web 2.0 systems in the mid-2000s (O'Reilly 2005).

Based on XML and Internet protocols (HTTP, SMTP), web services offer a new way of reusing and integrating third party or legacy systems. New types of web services and their associated APIs (application programming interface) allow developers to easily integrate diverse content from different web-enabled system, for example, REST (representational state transfer) for invoking remote services, RSS (really simple syndication) for news “pushing,” JSON (JavaScript object notation) for lightweight data-interchange, and AJAX (asynchronous JavaScript + XML) for data interchange and dynamic display. Such lightweight programming models support data syndication and notification and “mashups” of multimedia content (e.g., Flickr, Youtube, Google Maps) from different web sources—a process somewhat similar to ETL (extraction, transformation, and load) in BI&A 1.0. Most of the e-commerce vendors have provided mature APIs for accessing their product and customer content (Schonfeld 2005). For example, through Amazon Web Services, developers can access product catalog, customer reviews, site ranking, historical pricing, and the Amazon Elastic Compute Cloud (EC2) for computing capacity. Similarly, Google web APIs support AJAX search, Map API, GData API (for Calendar, Gmail, etc.), Google Translate, and Google App Engine for cloud computing resources. Web services and APIs continue to provide an exciting stream of new data

Table 2: Summary of Special Issue Papers within the BI&A Research Framework

Authors and Titles	Evolutions	Applications	Data	Analytics/ Research	Impacts
Chau and Xu, “Business Intelligence in Blogs: Understanding Consumer Interactions and Communities”	BI&A 2.0 on social media & network analytics	Market intelligence on consumers and communities	User-generated content extracted from blogs	<ul style="list-style-type: none"> ▪ Text and network analytics ▪ Community detection ▪ Network visualization 	Increased sales and customer satisfaction
Park <i>et al.</i> , “A Social Network-Based Inference Model for Validating Customer Profile Data”	BI&A 1.0 & 2.0 on social network analysis and statistical analysis	Market intelligence in predicting customers’ profiles	Self-reported user profiles and mobile call records	<ul style="list-style-type: none"> ▪ Network analytics ▪ Anomaly detection ▪ Predictive analytics 	Personalized recommendation and increased customer satisfaction
Lau <i>et al.</i> , “Web 2.0 Environmental Scanning and Adaptive Decision Support for Business Mergers and Acquisitions”	BI&A 1.0 and 2.0 on scorecards and web analytics	Market intelligence on environmental scanning	Business information extracted from Internet and proprietary financial information	<ul style="list-style-type: none"> ▪ Text and web analytics ▪ Sentiment and affect analysis ▪ Relation mining 	Strategic decision making in mergers and acquisitions

Hu <i>et al.</i> , “Network-Based Modeling and Analysis of Systemic Risk in Banking Systems”	BI&A 1.0 on statistical analysis	Systemic risk analysis and management in banking systems	U.S. banking information extracted from FDIC and Federal Reserve Wire Network	<ul style="list-style-type: none"> ▪ Network and data analytics ▪ Descriptive and predictive modeling ▪ Discrete event simulation 	Monitoring and mitigating of contagious bank failures
Abbasi <i>et al.</i> , “MetaFraud: A Meta-Learning Framework for Detecting Financial Fraud”	BI&A 1.0 on data mining and meta-learning	Fraud detection	Financial ratios, and organizational and industrial-level context features	<ul style="list-style-type: none"> ▪ Data analytics ▪ Classification & generalization ▪ Adaptive learning 	Financial fraud detection
Sahoo <i>et al.</i> , “A Hidden Markov Model for Collaborative Filtering”	BI&A 1.0 on statistical analysis	Recommender systems with changing user preferences	Blog reading data, Netflix prize data set, and Last.fm data	<ul style="list-style-type: none"> ▪ Data and web analytics ▪ Statistical dynamic model ▪ Collaborative filtering 	Personalized recommendations

BI&A Knowledge and Skills

BI&A education should be interdisciplinary and cover critical analytical and IT skills, business and domain knowledge, and communication skills required in a complex data-centric business environment. Analytical and IT skills include a variety of evolving topics. They are drawn from disciplines such as statistics and computer science for managing and analyzing both structured data and unstructured text. Coverage of these topics ranges from BI&A 1.0 to BI&A 3.0. The academic programs intended to produce BI&A professionals should consider these analytical and IT skills as suggested in Table 3 of our research framework.

Program Development

BI&A provides a unique opportunity for IS units in business schools to develop new courses, certificate programs, and degree programs charged with preparing the next generation of analytical thinkers. There are many options for delivering BI&A education. Because of the depth of knowledge required, graduate programs are the obvious choice. Viable program development options in delivering BI&A education include

- creating a Master of Science (MS) degree in BI&A
- creating a BI&A concentration in existing MS IS programs
- offering a graduate BI&A certificate program

The first option requires the effort of developing a new program. A few universities have embarked on this endeavor. A nonexhaustive list includes North Carolina State University, Saint Joseph’s University, Northwestern University, the University of Denver, Stevens Institute of Technology, and Fordham University. New York University will launch its new program in May 2013. New MS degree programs can be designed explicitly to attract analytically strong students with undergraduate degrees in areas such as mathematics, science, and computer science, and to prepare these students for careers, not only in the IS or IT groups in industry, but also in functional areas such as research and development, marketing, media, logistics, and finance.

Conclusions

Through BI&A 1.0 initiatives, businesses and organizations from all sectors began to gain critical insights from the structured data collected through various enterprise systems and analyzed by commercial relational database management systems. Over the past several years, web intelligence, web analytics, web 2.0, and the ability to mine unstructured user-

generated contents have ushered in a new and exciting era of BI&A 2.0 research, leading to unprecedented intelligence on consumer opinion, customer needs, and recognizing new business opportunities. Now, in this era of Big Data, even while BI&A 2.0 is still maturing, we find ourselves poised at the brink of BI&A 3.0, with all the attendant uncertainty that new and potentially revolutionary technologies bring. This MIS Quarterly Special Issue on Business Intelligence Research is intended to serve, in part, as a platform and conversation guide for examining how the IS discipline can better serve the needs of business decision makers in light of maturing and emerging BI&A technologies, ubiquitous Big Data, and the predicted shortages of data-savvy managers and of business professionals with deep analytical skills. How can academic IS programs continue to meet the needs of their traditional students, while also reaching the working IT professional in need of new analytical skills? A new vision for IS may be needed to address this and other questions. By highlighting several applications such as e-commerce, market intelligence, e-government, healthcare, and security, and by mapping important facets of the current BI&A knowledge landscape, we hope to contribute to future sources of knowledge and to augment current discussions on the importance of (relevant) academic research.

References

1. Borgatti SP, Everett MG, Freeman LC. UCInet for Windows: Software for Social Network Analysis, Harvard MA. Analytic Technologies, 2002.
2. Brantingham PL. Computational Criminology, Keynote Address to the European Intelligence and Security Informatics Conference, Athens, Greece, 12-14.
3. Brin S, Page L. The Anatomy of a Large-Scale Hypertextual Web Search Engine, Computer Network and ISDN Systems. 1998; (30):107-117.
4. Brumfiel G. 2911, High-Energy Physics: Down the Petabyte Highway, Nature (469) 282-283.
5. Chaudhuri S, Dayal U, Narasayya V. An Overview of Business Intelligence Technology, Communications of the ACM. 2011; 54(8):88-98.
6. Chen H. Intelligence and Security Informatics for International Security: Information Sharing and Data Mining, New York: Springer, 2006.
7. Chen H. AI, E-Government, and Politics 2.0, IEEE Intelligent Systems. 2009; 24(5):64-67.
8. Chen H. Design Science, Grand Challenges, and Societal

- Impacts, *ACM Transactions on Management Information Systems*. 2011a; 2(1)1:1-1.
9. Chen H. Smart Health and Wellbeing, *IEEE Intelligent Systems*. 2011b; 26(5):78-79.
 10. Chen H. Dark Web: Exploring and Mining the Dark Side of the Web, New York: Springer, 2012.
 11. Chen H, Brandt L, Gregg V, Traummuller R, *et al.* (Eds.). Digital Government: E-Government Research, Case Studies, and Implementation, New York: Springer, 2007.
 12. Chen H, Reid E, Sinai J, Silke A, Ganor B. (eds.). Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security, New York: Springer, 2008.
 13. Chiang RHL, Goes P, Stohr EA. Business Intelligence and Analytics Education and Program Development: A Unique Opportunity for the Information Systems Discipline, *ACM Transactions on Management Information Systems* (3:3), forthcoming, 2012.
 14. Davenport TH. Competing on Analytics, *Harvard Business Review*. 2006; 84(1):98-107. Doan A, Ramakrishnan, R., and Halevy, A. Y. 2011. Crowdsourcing Systems on the World-Wide Web, *Communications of the ACM* (54:4):86-96.
 15. Fortunato S. Community Detection in Graphs, *Physics Reports*. 2010; 486:3-5, 75-174. Frank, O., and Strauss, D. 1986. "Markov Graphs," *Journal of the American Statistical Association* (81:395), pp. 832-842.
 16. Freeman T. The World is Flat: A Brief History of the Twenty-First Century, New York: Farrar, Straus, and Giroux, 2005.
 17. Gelfand A. Privacy and Biomedical Research: Building a Trust Infrastructure-An Exploration of Data-Driven and Process-Driven Approaches to Data Privacy, *Biomedical Computation Review*, winter, 2011/2012, 23-28 (available at <http://biomedicalcomputationreview.org/content/privacy-and-biomedical-research-building-trust-infrastructure>, accessed).
 18. Hanauer DA, Rhodes DR, Chinnaiyan AM. "Exploring Clinical Associations Using '-Omics' Based Enrichment Analyses, *PLoS ONE*. 2009; 4(4):e5203.
 19. Hanauer DA, Zheng K., Ramakrishnan N, Keller BJ. Opportunities and Challenges in Association and Episode Discovery from Electronic Health Records, *IEEE Intelligent Systems*. Henschen, D. 2011. Why All the Hadoopla? *Information Week*. 2011; 26(5):83-87, 141:9-26.
 20. Hevner A, March ST, Park J, Ram S. Design Science Research in Information Systems," *MIS Quarterly*. 2004; 28(1):75-105.
 21. Hirsch JE. An Index to Quantify an Individual's Scientific Research Output, *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102:46, 16569-16572.
 22. Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M. ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Network," *Journal of Statistical Software* (24:3) (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2743438/>). IBM. 2011. "The 2011 IBM Tech Trends Report: The Clouds are Rolling In...Is Your Business Ready?," 2012-2008 <http://www.ibm.com/developerworks/techntrendsreport>; accessed.
 23. Karpf D. Blogosphere Research: A Mixed-Methods Approach to Rapidly Changing Systems, *IEEE Intelligent Systems*. 2009; 24(5):67-70.
 24. Liben-Nowell D, Kleinberg J. The Link-Prediction Problem for Social Networks, *Journal of American Society for Information Science and Technology*. 2007; 58(7):1019-1031.
 25. Lin Y, Brown RA, Yang HJ, Li S, Lu H, Chen H. Data Mining Large-Scale Electronic Health Records for Clinical Support, *IEEE Intelligent Systems*. 2011; 26(5):87-90.
 26. Lusch RF, Liu Y, Chen Y. The Phase Transition of Markets and Organizations: The New Intelligence and Entrepreneurial Frontier, *IEEE Intelligent Systems*. 2010; 25(1):71-75.
 27. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH. Big Data: The Next Frontier for Innovation, Competition, and Productivity, *McKinsey Global Institute*, 2011-2012, (http://www.mckinsey.Com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation; accessed).