BUSINESS   INTELLIGENCE AND ANALYTICS IN BIG DATA

M.Reddy prasanna,M.Swapna,Dr k.venkataramana

Abstract:

Business intelligence and analytics (BI&A) has emerged as an important area of study for both practitioner sand 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. The opportunities associated with data and analysis in different organizations have helped generate significant interest in BI&A, which is often referred to as the techniques, technologies ,systems, practices, methodologies, and applications that analyze critical business data to help an enterprise better understand its business and market and make timely business decisions. In addition to the underlying data processing and analytical technologies, BI&A includes business-centric practices and methodologies that can be applied to various high-impact applications such as e-commerce, market intelligence, e-government, healthcare, and security.
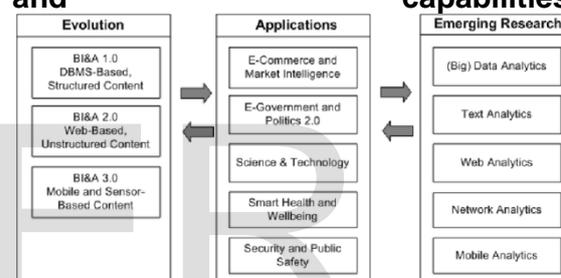
Keywords: Business intelligence and analytics, big data analytics, Web 2.0

Introduction

To make effective decisions. Hal Varian, Chief Economist at Google and emeritus professor at the University of California, Berkeley, commented on the emerging opportunities for IT professionals and students in data analysis as follows:So what's getting ubiquitous and cheap? Data. And what is complementary to data? Analysis. So my recommendation is to take lots of courses about how to manipulate and analyze data: databases, machine learning, econometrics, statistics, visualization, and so on.

The opportunities associated with data and analysis in different organizations have helped generate significant interest in BI&A, which is often referred to as the techniques, technologies ,systems, practices, methodologies Business intelligence and analytics (BI&A) and the related field of big data analytics have become increasingly important in both the academic and the business communities over the past two decades. Industry studies have highlighted this significant development. For example, based on a survey of over 4,000 information technology (IT) professionals from 93 countries and 25 industries, the IBM Tech Trends Report (2011) identified business analytics as one of the four major technology trends in the 2010s.

BI&A Evolution: Key Characteristics and   capabilities

The term intelligence has been used by researchers in artificial intelligence since the 1950s. Business intelligence became a popular term in the business and IT communities only in the 1990s. In the late 2000s, business analytics was introduced to represent the key analytical component in BI (Davenport 2006). More recently big data and big data analytics have been used to describe the data sets and analytical techniques in applications that are so large (from terabytes to ex   a bytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis, and visualization technologies.

- *M..reddy prasanna   is currently pursuing 2nd MCA in KMMIPS, Tirupathi, PH-01123456789. E-mail: reddiprasanna3@gmail.com*
- *M.swapna   is currently pursuing 2nd MCA in KMMIPS, tirupathi , PH-01123456789.*
- *Dr k.venkataramana head of the department of computer applications in KMMIPS tirupathi PH-01123456789 Email:ramanakv4@gmail.com*

In this article we use business intelligence and analytics (BI&A) as a unified term and treat big data analytics as a related field that offers new directions for BI&A research.

BI&A 1.0

As a data-centric approach, BI&A has its roots in the longstanding database management field. It relies heavily on various data collection, extraction, and analysis technologies . The BI&A technologies and applications currently adopted in industry can be considered as BI&A 1.0, where data are mostly structured, collected by companies through various legacy systems, and often stored in commercial relational database management systems (RDBMS). The analytical techniques commonly used in these systems ,popularized in the 1990s, are grounded mainly in statistical methods developed in the 1970s and data mining techniques developed in the 1980s.

Data management and warehousing is considered the foundation of BI&A 1.0

BI&A 2.0

Since the early 2000s, the Internet and the Web began to offer unique data collection and analytical research and development opportunities. The HTTP-based Web 1.0 systems, characterized by Web search engines such as Google and Yahoo and e-commerce businesses such as Amazon and eBay, allow organizations to present their businesses online and interact with their customers directly. In addition to porting their traditional RDBMS-based product information and business contents online, detailed and IP-specific user search and interaction logs that are collected seamlessly through cookies and server logs have become a new gold mine for understanding customers' needs and identifying new business opportunities. Web intelligence, web analytics, and the user-generated content collected through Web 2.0-based social and crowd-sourcing systems have ushered in a new and exciting era of BI&A 2.0 research in the 2000s, centered on text and web analytics for unstructured web contents.

BI&A 3.0

Whereas web-based BI&A 2.0 has attracted active research from academia and industry, a new research opportunity in BI&A 3.0 is emerging. As reported Prominently in an October 2011 article in The Economist (2011), the number of Mobile phones and tablets (about 480 million units) surpassed the number of laptops and PCs (about 380 million units) for the first

time in 2011. Although the number of PCs in use surpassed 1 billion in 2008, the same article projected that the number of mobile connected devices would reach 10 billion in 2020. Mobile devices such as the iPad , iPhone, and other smart phones and their complete ecosystems of downloadable applications, from travel advisories to multi-player games, are transforming different facets of society, from education to healthcare and from entertainment to governments. Other sensor-based Internet-enabled devices equipped with RFID ,barcodes, and radio tags (the "Internet of Things") are opening up exciting new steams of innovative applications.

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). 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.

Key Characteristics
Gartner BI Platforms Core
Capabilities Gartner Hype Cycle

BI&A 1.0 DBMS-based, structured content
- RDBMS & data warehousing
- ETL & OLAP
- Dashboards & scorecards
- Data mining & statistical analysis
- Ad hoc query & search-based BI
- Reporting, dashboards & scorecards
- OLAP
- Interactive visualization
- Predictive modelling & data mining
- Column-based DBMS
- In-memory DBMS
- Real-time decision
- Location-aware analysis
- Person-centred analysis
- Context-relevant analysis

- Mobile visualization & HCI
- Mobile BI

E-Commerce and Market Intelligence

The excitement surrounding BI&A and Big Data has arguably been generated primarily from the web and e-commerce communities. Significant market transformation has been accomplished by leading e-commerce vendors such Amazon

and eBay through their innovative and highly scalable ecommerce platforms and product recommender systems.

Major Internet firms such as Google, Amazon, and Face book continue to lead the development of web analytics, cloud computing, and social media platforms.

.Science and Technology

Many areas of science and technology (S&T) are reaping the benefits of high-throughput sensors and instruments, from astrophysics and oceanography, to genomics and environmental research. To facilitate information sharing and data analytics, the National Science Foundation (NSF) recently mandated that every project is required to provide a data management plan. Cyber-infrastructure, in particular, has become critical for supporting such data-sharing initiatives.

Smart Health and Wellbeing

Much like the big data opportunities facing the e-commerce and S&T communities, the health community is facing a tsunami of health- and healthcare-related content generated from numerous patient care points of contact, sophisticated medical instruments, and web-based health communities. Two main sources of health big data are genomics-driven big data (genotyping, gene expression, sequencing data) and payer–provider big data (electronic health records, insurance records, pharmacy prescription, patient feedback and responses) (Miller 2012a). The expected raw sequencing data from each person is approximately four terabytes. Partially funded by the National Institutes of Health (NIH), the NSF BIGDATA program solicitation includes common interests in big data across NSF and NIH.

Security and Public Safety

Since the tragic events of September 11, 2001, security research has gained much attention, especially given the increasing dependency of business and our global society on digital enablement. Researchers in computational science, information systems, social sciences, engineering, medicine, and many other fields have been called upon to help enhance our ability to fight violence, terrorism, cyber crimes, and other cyber security concerns. Critical mission areas have been identified where information technology can contribute, as suggested in the U.S. Office of Homeland Security's report "National Strategy for Homeland Security," released in 2002, including intelligence and warning, border and transportation security, domestic counter-terrorism, protecting critical infrastructure (including cyberspace), defending against catastrophic terrorism, and emergency preparedness and response.

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.

E-Commerce and Market Intelligence E-Government and Politics 2.0 Science &Technology Smart Health and

Wellbeing Security and Public Safety Applications

- Recommender systems
- Social media monitoring and analysis
- Crowd-sourcing systems
- Social and virtual games
- Ubiquitous government services
- Hypothesis testing
- Knowledge discovery
- Human and plant genomics
- Healthcare decision support
- Patient community analysis
- Crime analysis
- Computational criminology
- Terrorism informatics
- Open-source intelligence
- Cyber security

Data • Search and user logs
- Customer transaction records
- Customer generated content
- Government information and services
- Rules and regulations
- Citizen feedback and comments
- Sensor and network content

- Genomics and sequence data
- Electronic health records (EHR)
- Health and patient social media
- Criminal records
- Crime maps
- Criminal networks
- News and web contents
- Terrorism incident databases
- Viruses, cyber attacks, and botnets

Analytics

- Association rule mining
- Database segmentation and clustering
- Anomaly detection
- Graph mining
- Social network analysis
- Text and web analytics
- Sentiment and affect analysis
- Information integration
- Content and text analytics
- Government information semantic services and onto logies
- Social media monitoring and analysis
- Social network analysis
- Sentiment and affect analysis
- Health social media monitoring and analysis
- Health text analytics
- Health onto logies

Impact

Long-tail marketing, targeted and personalized recommendation, increased sale and customer satisfaction Transforming governments, empowering citizens, improving transparency, participation, and equality S&T advances, scientific impact Improved healthcare quality, improved long-term care, patient empowerment Improved public safety and security.

(Big) Data Analytics Text Analytics Web Analytics Network Analytics Mobile Analytics Foundational Technologies

- RDBMS
- data warehousing
- ETL
- OLAP
- PM
- Data mining
- clustering
- Regression
- Classification
- Association analysis
- Anomaly detection
- Neural networks
- Genetic algorithms

- Multivariate statistical analysis
- Optimization
- Heuristic search
- Information retrieval
- Document representation
- query processing
- Relevance feedback

Emerging Research

- Statistical machine learning
- Sequential and temporal mining
- Spatial mining
- Mining high-speed data streams and sensor data
- Process mining
- Privacy-preserving data mining
- Network mining
- Web mining
- Column-based DBMS
- In-memory DBMS
- parallel DBMS
- cloud computing
- Hadoop
- Map Reduce

(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). 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), Page Rank, AdaBoost, kNN (k-nearest neighbours), 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

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 addition to document and query representations, user models and relevance feedback are also important in enhancing search performance.

Since the early 1990s, search engines have evolved into mature commercial systems, consisting of fast, distributed crawling; efficient inverted indexing; in link-based page ranking; and search logs analytics. Many of these foundational text processing and indexing techniques have been deployed in text-based enterprise search and document management systems in BI&A 1.0. A major emerging component in web analytics research is the development of cloud computing platforms and services, which include applications, system software, and hardware delivered as services over the Internet.
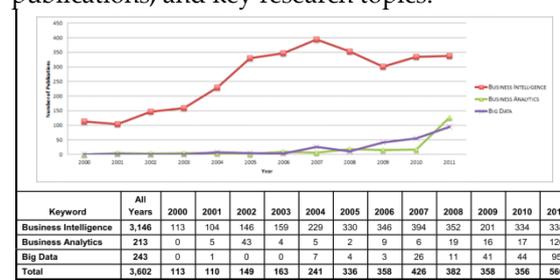
Network Analytics

Network analytics is a nascent research area that has evolved from the earlier citation-based bibliometric analysis to include new computational models for online community and social network analysis. Grounded in bibliometric analysis, citation networks and co authorship networks have long been adopted to examine scientific impact and knowledge diffusion. The h-index is a good example of a citation metric that aims to measure the productivity and impact of the published work of a scientist or scholar (Hirsch 2005). Since the early 2000s, network science has begun to advance rapidly with contributions from sociologists, mathematicians, and computer scientists.

Mobile Analytics

As an effective channel for reaching many users and as a means of increasing the productivity and efficiency of an organization's workforce, mobile computing is viewed by respondents of the recent IBM technology trends survey (IBM 2011) as the second most "in demand" area for software development. Mobile BI was also considered by the Gartner BI Hype Cycle analysis as one of the new technologies that have the potential to drastically disrupt the BI market (Bitterer 2011). Mobile computing offers a means for IT professional growth as more and more organizations build applications. With its large and growing global install base, Android has been ranked as the top mobile platform since 2010.

Mapping the BI&A Knowledge Landscape: A Bibliometric Study of Academic and Industry Publications

In an effort to better understand the current state of BI&A related research and identify future sources of knowledge, we conducted a bibliometric study analyzing relevant literature, major BI&A scholars, disciplines and publications, and key research topics.

Keyword	All Years	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Business Intelligence	3,146	113	104	146	159	229	330	346	394	352	201	334	338
Business Analytics	213	0	5	43	4	5	2	9	6	19	16	17	126
Big Data	243	0	1	0	0	7	4	3	26	11	41	44	95
Total	3,602	113	110	149	163	241	336	358	426	382	358	356	560

A collection, transformation, and analytics process was followed in the study, much like a typical BI&A process adopted in other applications. To ensure data consistency and relevance across our collection, we retrieved only those publications that contained the keywords business intelligence, business analytics, or big data within their title, abstract, or subject indexing (when applicable). The choice of these three keywords was intended to focus our search and analysis on publications of direct relevance to our interest

Academic Publication Publications

Decision Support Systems 41 Communications of the AIS 19 Journal of Management Information Systems 12 Management Science 10 Information Systems Research 9 Journal of the Association for Information Systems 5 INFORMS Journal on Computing 4 Management Information Systems Quarterly 2

BI&A Education and Program Development

BI&A provides opportunities not only for the research community, but also for education and program development. In July 2012, Columbia University and New York City announced plans to invest over $80 million dollars in a new Centre for Data Science, which is expected to generate thousands of jobs and millions of dollars in tax revenues from 100 start up companies over the next 10 years (Associated Press 2012). BI&A is data science in business. Job postings seeking data scientists and business analytics specialists abound these days. There is a clear shortage of professionals with the "deep" knowledge

required to manage the three V's of big data: volume, velocity, and variety (Russom 2011).

Education Challenges

BI&A focuses on understanding, interpretation, strategizing, and taking action to further organizational interests. Several academic disciplines have contributed to BI&A, including IS, CS, Statistics, Management, and Marketing, as shown in our bibliometric study. IS programs, in particular, are uniquely positioned to train a new generation of scholars and students due to their emphasis on key data management and information technologies, business-oriented statistical analysis and management science techniques, and broad business discipline exposure (e.g., Finance, Accounting, Marketing, and Economics). Since its inception approximately 45 years ago, IS as an academic discipline has primarily focused on business needs in an era when the major challenges involved the management of internal business and transaction data. In the age of big data, these problems remain, but the emphasis in industry has shifted to data analysis and rapid business decision making based on huge volumes of information. Such time-critical decision making largely takes place outside of the IS function

Papers in this Special Issue

The idea for this special issue began in May 2009, when Delmar Straub, the Editor-in-Chief of MIS Quarterly, solicited suggestions for special issues from the editorial board members. We submitted the special issue proposal on Business Intelligence Research in August 2009, with the call-for papers approved and distributed at the 30th Annual International In the following 20 months, six of the manuscripts went through three or four review rounds and were then accepted for this issue.

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.

References

- Adomavicius, G., and Tuzhilin, A. 2005. "Toward the Next
- Generation of Recommender Systems: A Survey of the State-of the-
- Art and Possible Extensions," IEEE Transactions on
- Knowledge and Data Engineering (17:6), pp. 734-749.
- Anderson, C. 2004. "The Long Tail," WIRED Magazine (12:10)
- (http://www.wired.com/wired/archive/12.10/tail.html).
- Associated Press. 2012. "Columbia U Plans New Institute for Data
- Sciences," July 30 (http://www.cbsnews.com/8301-505245_162-
- 57482466/columbia-u-plans-new-institute-for-data-sciences/,
- accessed August 3, 2012).
- Barabási, A. 2003. Linked: How Everything Is Connected to
- Everything Else and what it Means for Business, Science, and