Recent Development in Big Data Analytics for Business Operations and Security Management

T. Parvathavardhini ,C. Dharani
Annai college of Engineering Technology,
Kumbakonam

Abstract:- **"Big data" is an emerging topic and has attracted the attention of many researchers and practitioners in industrial systems engineering and cybernetics. Big data analytics would definitely lead to valuable knowledge for many organizations. Business operations and risk management can be a benefi-ciary as there are many data collection channels in the related industrial systems (e.g., wireless sensor networks, Internet-based systems, etc.). Big data research, however, is still in its infancy. Its focus is rather unclear and related studies are not well amalgamated. This paper aims to present the challenges and opportunities of big data analytics in this unique application domain. Technological development and advances for industrial-based business systems, reliability and security of industrial systems, and their operational risk management are examined. Important areas for future research are also discussed and revealed.**

Index Terms— Big data analytics, business intelligence (BI), operational risk analysis, operations management, systems reliability and security.

I. INTRODUCTION

NFORMATION technology (IT) not only introduces con-**I**venience, but creates many new improvement opportunities which were impossible in the past. For example, advances of business intelligence (BI) methods [19] and data min-ing techniques have brought huge improvements to modern business operations [27]. Nowadays, in the "big data era," a massive amount of data is available for all kinds of indus-trial applications [32], [34]–[36], [45]–[48]. For example, the cloud service can be considered as a data warehouse which provides a useful source of data [27], [44]. Wireless sen-sor networks [e.g., radio frequency identification (RFID), near field communications] can be used to collect use-ful data ubiquitously [3], [37], [38], [41]. An evolving topic on the Internet of things (IoTs), which consists of devices

Manuscript received December 17, 2014; revised April 6, 2015 and October 12, 2015; accepted December 5, 2015. The work of T.-M. Choi was supported by The Hong Kong Polytechnic University under Grant G-UA1Q. This paper was recommended by Associate Editor D. D. Wu. *(Corresponding author: Tsan-Ming Choi.)*

T.-M. Choi is with The Hong Kong Polytechnic University, Hong Kong (e-mail: jason.choi@polyu.edu.hk).

H. K. Chan is with Nottingham University Business School, Ningbo 315100, China.

X. Yue is with the University of Wisconsin–Milwaukee, Milwaukee, WI 53201 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2015.2507599

capable of communicating via the Internet environment, also provides a platform for gathering an enormous amount of data [27], [40]. In other words, it is now easier to collect data than ever before. That being said, extracting and utilizing useful information from such huge and dynamic databases for "big data" is far from easy [128]. Since these data are linked to real-time events, they can be employed, if prop-erly (e.g., via BI schemes), for rescheduling or replanning activities in business applications which finally reduce the level of risk and improve profitability and efficiency of the operations. This undoubtedly can supplement traditional opti-mization techniques, which are *a priori* in nature. For instance, Zhang *et al.* [122] considered a dynamic workload schedul-ing problem with the help of big data stored in distributed cloud services. They developed an evolutionary optimization algorithm and simulated the performance under different sce-narios. In another study, Zhang *et al.* [123] analyzed the cost minimization issue of moving data around geographically dispersed data. Such data migration problem is very important yet challenging as the volume of big data is growing quickly. Dou *et al.* [124] developed a service optimization model for handling big data stored in cloud systems when privacy is a critical concern (e.g., the medical data). Service quality may be compromised if a cloud server refuses to provide the data due to the privacy issue. Such optimization model can maximize the service quality and is verified by a simulation study. Another application of big data is on smart grids [118]. Simmhan *et al.* [118] predicted the demand of a cloud-based smart grid system and derived the optimal pricing strategy, based on the big data on real-time consumption. The approach is possible due to the data mining algorithm the authors developed. The relationship between cloud systems and big data models will be further discussed in Section II.

Owing to the importance of big data analytics for business applications, this paper is developed. With respect to the core topic on big data analytics for business operations and risk management, we organize this paper into three big sec-tions, namely: 1) BI and data mining; 2) industrial systems reliability and security; and 3) business operational risk man-agement (ORM). Each of these sections: 1) examines some carefully selected papers; 2) outlines the related research chal-lenges; and 3) proposes the future research directions. To the best of our knowledge, this is the first paper in the literature which focuses on how big data analytics can be employed for reducing systems risk and enhancing efficiency in business operations.

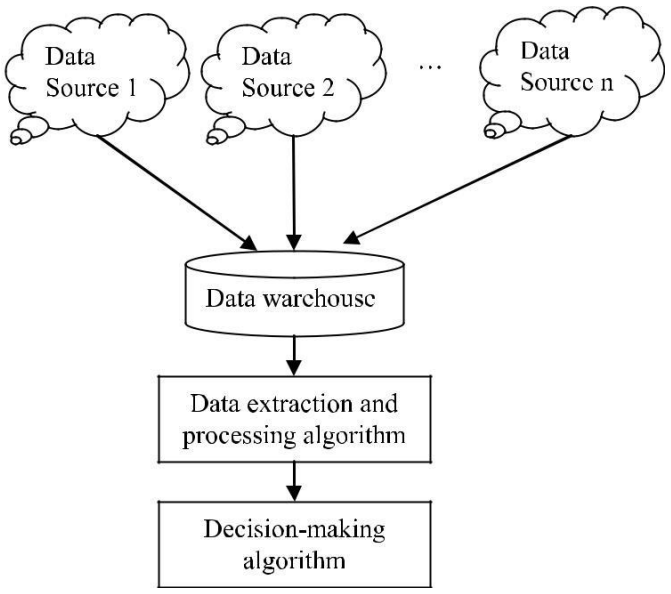

Fig. 1. Simplified BI system.

II. BUSINESS INTELLIGENCE AND DATA MINING

A. *Business Intelligence and the Enabling Technologies*

The term BI has been in existence for a long time. The oldest relevant study of BI is probably [19], which defines BI systems as, in simple terms, automatic data retrieving and processing systems that can help make intelligent deci-sions based on various data sources. Waston and Wixom [20] provided another even simpler definition of BI systems: "get-ting data in and getting data out." Such intelligent systems are highly related to the later development on decision sup-port systems (DSSs) in the 1970s [21]. Companies can make a careful use of such systems to enhance operations decisions making [22]. The concept is briefly illustrated in Fig. 1.

Some famous industrial applications of BI can be found in the airline industry for revenue management [23]; in the auto-mobile industry by standardizing the processes which led to reduction in various costs [24]; routing problems in transporta-tion networks [25]; minimizing the impact of uncertainties in supply chain systems [2]. All these applications assume data are accessible on a real-time basis, which is a techno-logical challenge in BI system designs. This limitation on real-time data collection has been improved by advances in IT, particularly the wireless sensor technology such as RFID [3]. This enables traditional BI systems to migrate to pervasive BI systems [20]. RFID is considered as the "the most exciting and fastest-growing technology in terms of scope of applica-tion in the next generation of BI" [37]. On the one hand, the RFID technology is a superb channel for coordination in industrial systems especially at item-level [3]. Therefore, operational aspects such as risk assessment [1], [103] and inventory management [3] are potential improvement areas with the collected data. On the other hand, the item-level applications are in fact technologically constrained so the contribution of this type of technology is limited from the BI systems' point of view. Nevertheless, this and similar sensor networks can form a vital part of the overall BI system [38].

Recent development on the IoTs also facilitates handling of an immense size of real-time dataset [26]. With a proper middleware, objects with RFID devices attached can be con-sidered as IoTs [41]. The major distinction between IoTs and other traditional sensor networks (including the RFID network) is that they are Internet-enabled which means the objects are configured in an Internet environment. In other words, such interactions would improve the ability to control activities of the objects by adding "intelligence" via better communi-cation. Despite the fact that it is still an infant technology, applications can already be found in environmental monitor-ing, inventory management, food supply chains, transportation, and so on [39]. IoT-based and also RFID-based networks, however, are subject to an intrinsic constraint, which is cre-ated by the heterogeneous nature of sensors [40], especially when the networks need to handle large datasets. It is there-fore not surprising that some relevant studies linked to the agent technology have been proposed in order to resolve this issue [17], [32]. With a proper system design, IoTs can be integrated to BI systems for many industrial applications. With the real-time information being collected, risk of man-agement, planning, and control activities can definitely be reduced. Another highly related development is the cloud service (implicitly depicted in Fig. 1). This is also an emerg-ing technology which can be used to store data in remote locations [27]. It is not a co-incidence that both big data and cloud technology emerge almost concurrently [117], as they do share some similar characteristics [121]. The cloud service provides a channel to store and process a lot of datasets [120], [123], which are originated from various loca-tions and then data analytics such as data mining, clustering, and so on, can take place somewhere else without the phys-ical connection with the data collection sites. In addition, a cloud service could even provide a platform to reduce computation effects for the above data analytics subject to the cloud infrastructure. The Apache Hadoop ecosystem is a nice example [119] that many companies have adopted in their cloud systems [117]. In other words, the cloud can act as more than a data warehouse to improve business agility via the Hadoop architecture [120], [121]. Successful deploy-ment of the technology can be found in the telecommunication industry [44] and smart grid [118], for example.

The above development on cloud services is again a con-sequence of the evolution of the Internet so subsequent discussion will focus on higher level Web-based BI sys-tems than just cloud services. Web-based BI systems have their obvious advantages: cheaper cost, shorter time to

collect data, and a single platform for data collection and a master database may be achievable. Nevertheless, such service adds concerns on security issues [28], [43]. Although nowadays vir-tually no information systems or technologies can exempt from risk and vulnerability, what is worrying here is that cloud-specific vulnerability is not clearly known and thus not well managed [43]. In addition, mastering one piece of data always has its challenges [29]. In other words, attention should be paid in designing Web-based BI systems. Table I summarizes

TABLE I
SUMMARY OF DIFFERENT BI ENABLING TECHNOLOGIES
(*X* = NOT SUPPORTED IF STANDALONE AND *O* = SUPPORT)

	Internet Enabled	Intelligent	Pervasive
RFID	X	X	O
IoT	O	X	O
DSS	X	O	X
Cloud Service	O	X	O
Web-BI	O	O	X

the key enabling technologies to BI. They are, however, never the synonyms to BI, and can only be part of the data collec-tion platforms in any BI systems. In addition, the development is limited to a high level system design, which on its own is not a real limitation but the implication is that we are still far from real-life mass-scale BI applications, not mentioning if that is achievable or not. This also brings out an impor-tant issue: shall we develop the data mining algorithms first (see the next section) or the systems first? It is really a chicken and egg question! Aforementioned technologies such as IoTs and cloud services are tightly coupled to the issue to be dis-cussed in the next section: data mining [30], which is the core engine to deliver any recommendation from the BI systems, be it intelligent or not.

B. Data Mining for Business Applications

The data extraction and processing algorithm in Fig. 1 has led to the later development of data mining, which is essen-tially a machine learning method [47]. This is the process to discover or identify useful relationship, patterns, among themes, factors, or similar items in a dataset [45], [47]. Therefore, many data mining approaches are coupled with statistical tests [27]. Insurance and associated risk analysis is a typical data mining application domain [46]. Data min-ing is also useful in medical research to identify influential factors to a disease [47]. To facilitate such mining process, on-line analytical processing (OLAP) has been the core of many BI systems for business applications [31]. OLAP was first developed in the 1970s together with the wave of DSSs development. Nowadays, many data mining studies are still linked to OLAP (see [32]). These algorithms normally include some learning approaches in order to identify patterns, behav-iors, or specific relationships the researchers would like to find out [45]. In fact, data mining algorithms can generally be categorized into two groups from the machine learning point of view, namely supervised learning and unsupervised learning [33]. The former relies heavily on known knowledge to train the mining systems in order to make reliable predic-tions; whereas the latter attempts to correlate (e.g., using the cluster analysis) items, themes, or factors from the data in order to extract their (unknown) relationship. In recent years, many data mining studies concern the latter due to a lack of suitable training sets and the nature of the data (see below). Successful applications can definitely generate benefits, e.g., maximizing revenues for online shopping [34], better traffic control [35], etc.

In relation to these categories, one downside of data col-lection and hence data processing in BI systems for risk and operations management must be discussed. This is the nature of the data, which are unstructured on many occasions. This is also one of the reasons why unsupervised learning is popularly adopted. It is partly because of the fact that the data come from a variety of sources. This multidimensional-ity implies that the solutions are not always generalizable. No matter which method(s) we choose, the ultimate objective of any data miming algorithms is to yield a high quality solu-tion. Haug *et al.* [36] discussed 12 barriers to achieve quality data and many of these barriers are related to IT systems. Therefore, data mining algorithms and system designs cannot be separated completely.

In recent years, the data mining community also encoun-ters other challenges, which include the volume of data and the velocity of data change. Together with aforemen-tioned characteristics which can be labeled as variety, these volume, velocity, and variety features characterize the term “big data” [48]. One key challenge in managing big data is to clean up the dataset in order to reduce the “noise” presented in the datasets [45], [47]. Presence of this noise will affect the quality of the recommendations from the BI systems.

Data mining, particularly in the big data era, is not with-out concerns. Concerns arise when the data collection pro-cesses link to surveillance (the “big brother” issue in other words) [49]. This is particularly alarming when the datasets contain public sector information, i.e., “governmental” big data. However, this concern has now extended to private cor-porations as well since big data have penetrated to every corner in our daily life. Technologically, mining big data is uneasy owing to its fast changing pace. For example, a learning algo-rithm suitable to today’s need can be outdated very quickly. In addition, many data mining applications are customized to dif-ferent needs which implies that an off-the-shelf solution cannot be found easily. In that sense, organizations need to devote their effort to developing such algorithms, which make the true return on investment uncertain. In other words, data min-ing can help reduce the risk of the process of concern, but its development is another type of risk to leverage! Despite

these issues, proper applications of data mining would create additional business value effectively.

BI systems and data mining always go hand-in-hand. Since its first discussion in the end of the 1950s, the attention of BI systems has been diverted by other developments, such as DSSs. In recent years, there are many driving forces that bring BI back to the attention of many researchers and practitioners. This can be supported by Fig. 2 in which the respective num-bers of publications in IEEE TRANSACTIONS over the last decade are listed. In addition, the growth of different topics actually coupled with each other, which makes the trend very obvious. In other words, one paper published in 2013 or so may cover all four topics.

Regarding the driving forces, the big data mentioned above is one of them. Maturity of machine learning and data min-ing development, such as soft computing techniques (fuzzy logic, genetic algorithms, etc.) [45], [105], [106], is also on the list. However, all these factors actually relate to basic data collection. With the recent development in IT and the Internet, data collection is much easier than in the past. With

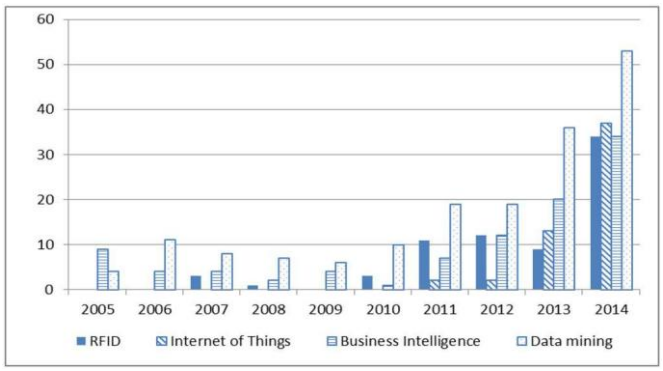

Fig. 2. Summary of recent publications in IEEE TRANSACTIONS on RFID, IoTs, BI, and Data Mining.

such valuable datasets, the need to put them into scrutiny for useful applications is perfectly understandable. Security and risk management to be discussed in the next two sections are potential improvement areas with big data analytics.

III. INDUSTRIAL SYSTEMS RELIABILITY AND SECURITY

A. *Use of Information for Systems Reliability*

Reliability is a critical element of virtually any system. It represents the probability that the system will perform its required function under certain given conditions for a time interval. A reliable system can operate continuously and safely [4], [50]. In terms of reliability, published research studies include many areas, such as product lifecycle man-agement (PLM), evaluation of dissemination and reliability of data, efficient exchange, and transmission of data. In industrial systems, information is widely used in PLM to ensure reliable production. The efficient tracking and tracing method for prod-uct lifecycle data is essential for avoiding any delays and errors affecting accuracy and completeness of enterprise applications, especially in the closed-loop PLM where several partners and organizations are involved. To this end, an overall framework for tracking and tracing product lifecycle data is developed in the closed-loop PLM [51]. It contains a schema to manage a huge amount of event data of product embedded informa-tion devices, and a processing mechanism to transform the big data into meaningful information. Ondemir and Gupta [52] proposed mathematical models that utilize product life-cycle data and remaining life estimations in order to fulfill the demands for used components and products that have a cer-tain remaining life. Information retrieved from each end-of-life product (EOLP) is stored in a database along with the unique identification number. This piece of information can also be shared among the parties of a manufacturing alliance in an industrial system [53].

Another application is the evaluation of dissemination and reliability of data, which is of significance to aeronautical engi-neering, power systems, traffic field, etc. Woochul *et al.* [54] presented a novel middleware architecture called the real-time data distribution service (RDDS) and demonstrate the viabil-ity of the proposed approach by implementing a prototype of RDDS. They show that, compared to baseline approaches, RDDS achieves highly efficient and reliable sensor data dis-semination as well as robustness against unpredictable work-loads. Shah *et al.* [55] proposed a distributed control algorithm for data delivery performance evaluation in a simulation envi-ronment. They show that the framework can achieve the required quality of service. Li and Meeker [56] provided an introduction to the basic ideas of using Bayesian methods for a reliability data analysis and illustrate the methods with four basic kinds of reliability data. Real-time information becomes the key factor for reliable delivery of power from the gen-erating units to the end-users [57]. The impact of equipment failures, capacity constraints, and natural emergency, can be largely avoided by online power system condition monitoring, diagnostics, and protection.

There are also many studies investigating how to guar-antee a real-time exchange and reliable transmission of data and information in industrial systems. For example, Gungor *et al.* [58] conducted a survey on smart grid potential applications and communication requirements. They posited that information and communication technologies represent a significant element in the growth and performance of smart grids. A sophisticated, reliable, and fast communication infras-tructure is necessary for the connection among the huge amount of distributed elements, such as energy storage sys-tems, users, and generators, which enable a real-time exchange of data and information. All these can enhance the efficiency, and reliability of all the elements involved in the smart grid. Dietrich *et al.* [59]

investigated the communication and com-putation in buildings. They proposed that the critical focal functional areas of the home energy management system are energy efficiency, data measurement, and transmission. The real-time consumption data collected from appliances are mea-sured and transferred to a data concentrator. Statistical analysis and intelligent advice generation based on the consumption data can enable in-home displays to inform consumers about their consumption behavior. In short, in the big data era, we can collect more information related to the reliability of sys-tems from more sources, which include real-time data. We also have more powerful computational ability to process a higher volume of information on systems reliability. In this area, the main challenges include how to construct the evaluation index systems based on the systems reliability data provided by big data technologies and how to establish the forecast-ing and warning mechanism to process the information of the systems reliability. Table II summarizes the major findings, applications, and the weaknesses of the models examined in this section. A quick finding is that most papers only explore the use of information for systems reliability of the respective functional area but not the whole big system in the enterprise. Plus, in order to achieve the desirable results in enhancing systems reliability, the data requirements are much higher than other traditional works in the area.

B. Data-Driven Industrial Systems Reliability

Industrial processing plants are usually heavily instrumented with a number of sensors to deliver data for process monitoring and control to guarantee the systems reliability and stability.

TABLE II
MAJOR FINDINGS AND APPLICATIONS OF THE REVIEWED MODELS WITH THE USE OF INFORMATION FOR SYSTEMS RELIABILITY

Models	Major Findings and Applications	Weaknesses	Probable Solutions
K,J&X[51]	Build a framework to track and trace product lifecycle data	Data loss may happen during the data processing step	Enhance data processing
O&G[52]	Develop a mathematical model to fulfill demand for used components	Data on product lifecycle and remaining life must be available	Have a tool to ensure the real time data on product lifecycle and remaining life are available
W,K&S[54]	Derive a real time middleware architecture for sensor data dissemination	The system is not yet linked up with other applications in the enterprise systems	Use an ERP system
S,G&A[55]	Propose a control algorithm for data delivery in smart grid applications	Relatively complex for industrial implementation	Examine a simpler version of the control algorithm
L&M[56]	Develop a Bayesian approach for reliability data analysis	Assessment of prior distribution and collection of unbiased data are both critical	Increase the weight of the observed information if the prior distribution is difficult to estimate
G,B&H[57]	Discuss the applications of wireless sensor networks in smart grid	The performance relies heavily on the availability of real time and reliable data	Ensure that real time data are reliable and available
G et al. [58]	Examine the future applications of smart grid for systems reliability	Does not consider the systems optimization issues	Revise the model formulation to achieve systems optimization.

Wireless-networked sensors will form a new "Web" [60]. The soft sensors can be applied in many industrial operations. For instance, Kadlec and Gabrys [62] proposed an ensemble approach for soft sensor development based on multilayer perceptrons. They solved the problem of optimal network complexity selection in the context of ensemble methods. They applied the soft sensor to an industrial drier process. Another application is process monitoring and process fault detection. The systems can be trained to analyze the nor-mal operating state or to recognize possible process faults. Kämpjärvi *et al.* [63] gave a complex soft sensor for the detection and isolation of process faults, which can be devel-oped into an ethylene cracking process. The authors pointed out that an improved accuracy of the system is obtained after including calculated variables (which are generated using pro-cess knowledge). The knowledge-based approach for fault diagnosis assesses online monitored data according to a set of rules which are learned from the experience of human experts. This approach automates human intelligence for pro-cess supervision [64]. Rongsheng *et al.* [65] presented an efficient method to discover knowledge for classification prob-lems through data summarization. It is applied on the welding fault diagnosis in manufacturing. It discretizes continuous fea-tures and then summarizes the data using a contingency table. The inconsistency rate for different subsets of features can then be easily calculated from the contingency table. After the best feature subset is found, knowledge can be intuitively derived from the data summary. Besides, knowledge can be updated quickly whenever new datasets are obtained.

Observe that data-driven techniques are widely used in sup-ply chain management, and many problems are present. For example, the supply chain system may not perform well in responding to market demands and supplier conditions in real time. The big data collected through sensors and equipments have the potential to solve these problems. In fact, sensors and RFID tags have been integrated to be applied in supply chain management. This integration introduces the concept of com-bined intelligent products and enables the storage of static and dynamic data on the product, which leads to promising new opportunities in EOLP management [66].

MapReduce is one of the new advances of "support-ing systems." It is a framework for distributed computing which uses the "divide approach" to decompose complex big data problems into small units of work and process them in parallel [67]. It is efficient and fault tolerant for analyzing a large set of data, which makes it a popular tool for pro-cessing large scale data [68]. Chen *et al.* [69] conducted an empirical analysis of MapReduce traces. They found a char-acterization of new MapReduce workloads which are driven in part by an interactive analysis. Herodotou *et al.* [70] intro-duced starfish, a self-tuning system for timely, cost-effective and reliable analytics. Cohen *et al.* [71] studied the emerg-ing practice of magnetic, agile, deep data analysis as a radical departure from the traditional data warehousing and BI.

Big data technologies enable us to more comprehensively monitor and control the operations of data-driven industrial systems. The design, manufacturing, application, and man-agement of industrial systems have many new characteristics. Expectedly, the evaluation, analysis, diagnosis, and early warn-ing of data-driven industrial systems reliability will encounter many "big data challenges" [60], [61], [66], [70].

Table III summarizes the major findings and weaknesses of several models reviewed in Section III-B. From Table III, we can see that there are various studies on the use of big data for industrial systems reliability analysis. However, most of them are exploratory in nature. Thus, deeper theoretical research should be conducted in the future.

C. Security Breaches

There are many big data applications to solve problems associated with security breaches. For example, big data tech-niques are used to conduct security assessment, and solve security problems in industry systems. It can also be used in the public sector. Xu *et al.* [72] used a data mining tech-nique for fast dynamic security assessment. They examined the method and proved that it has better efficiency and accuracy in the transient stability assessment. Ding *et al.* [73] pro-posed a data-based global operations approach to minimize the effect on the production performance caused by unex-pected variations to guarantee security in the operations of a mineral processing plant. They extracted relevant rules based on the operational data of the mineral processing plant and justified the effectiveness of the approach. Demand response. 6

TABLE III
MAJOR FINDINGS AND APPLICATIONS OF THE REVIEWED MODELS WITH THE DATA DRIVEN INDUSTRIAL SYSTEMS RELIABILITY

Models	Major Findings and Applications	Weaknesses	Probable Solutions
C,B&B [61]	Develop a robust algorithm to find the optimal tradeoff between energy spent in transmission and data compression	Only focus on having a longer network lifetime, without considering the corresponding potential volatility of lifetime	Lifetime is a random variable and hence its volatility affects the systems performance. To overcome this issue, one can incorporate the consideration of lifetime volatility into the optimization model. For example, the mean-variance analysis can be conducted [103]
K &G[62]	Derive an analytical model to solve the optimal network complexity selection problem	The artificial neural network is slow	Examine the use of faster AI methods such as extreme learning machines to help
K et al. [63]	Generate process knowledge to improve accuracy in the detection of process faults for ethylene cracking process	The method includes several calculated variables which rely heavily on the quality of the process knowledge data	Improve the quality of the process knowledge data
R et al. [65]	Build an efficient method to reveal knowledge for classification problems through data summarization and is applicable in fault diagnosis in manufacturing process	Data will be lost during the required discretization process	Enhance input quantity and quality so that the impact brought by data loss would be minimized
C et al. [69]	Examine the interactive analytical processing in big data systems	The finding is based on an empirical study which is exploratory in nature	Extend the analysis to analytical modeling
C et al. [71]	Discuss new practices with big data analysis	No detailed analytical comparison is made	Conduct comprehensive studies on analytical comparisons

has become a key feature of the future smart grid. Yong [74] presented a comprehensive framework of defense architecture for power system security and stability, which includes the security assurance system and the stability control system. He found that it can effectively guarantee the security of power systems. Wang *et al.* [75] proposed an event-driven emergency demand response scheme to enhance the efficiency and secu-rity of a power system. Big data can also be used to ensure the security of buildings and physical infrastructure, which is the basis of home video surveillance and security systems. It also offers solutions across industry sectors for monitor-ing and protecting sensitive infrastructures [76]. Big data can be used in the public sector to improve decision making with automated systems to enhance security and innovate [77]. Samuelson *et al.* [78] provided techniques for strengthen-ing the security of automatic dependent surveillance-broadcast systems. These include a message authentication code algo-rithm that provides message authentication and an encryption scheme that safeguards message content. It is widely used in the air traffic control system. Finke *et al.* [79] evaluated the limitations of legacy systems used in air traffic control and studied the feasibility of employing format-preserving encryption.

With respect to security models and methods, big data driven security models have two common characteris-tics related to security: 1) advanced monitoring systems that continuously analyze systems, resources, and make consid-erations based on the respective behaviors and risk models and 2) integration of security and risk management tools to facilitate detailed investigations of potential problems [80]. Krishnan *et al.* [81] presented an efficient sampling strat-egy that maximizes the database information content while minimizing computing requirements. The proposed approach is used to derive operations rules against voltage security issues. Zhou and Chao [82] designed a media-aware security framework, which enables various multimedia applications in IoTs. They first presented a novel multimedia traffic clas-sification and analysis method for dealing with the hetero-geneity of diverse applications. Then a media-aware traffic security architecture is proposed based on the given traf-fic classification. Furthermore, they provided a design rule and strategy, which can achieve a good balance between the system's flexibility and efficiency. Their study creates a media-aware security architecture by integrating the characteristics of multimedia traffic, security service, and IoTs.

The big data technology enhances our understanding of security breaches in industrial systems. It also helps iden-tify the signs of security breaches in real time and make more effective reactions. Research challenges on how to:
3) build a more perfect security evaluation system; 2) set up more reasonable security breach diagnosis rules; and take legal and effective measures to deal with the security breaches [73], [76], [77], [81], [82] are critically important.

IV. BUSINESS OPERATIONAL RISK MANAGEMENT

A. Operational Risk Management

In the industry, irrespective of the specific busi-ness nature, there are all kinds of operations and processes [111], [112]. In light of all sources of uncertainty in the real world setting, there exist all kinds of opera-tional risks [113]–[116], [126], [127]. In an influential study, Beroggi and Wallace [85] defined ORM as the "identification of an unexpected event, an assessment of its consequences and the decision to change the planned course of action." Under this definition and concept, they proposed a real-time event driven paradigm and reasoning algorithm for optimal decision making with a goal of achieving effective and imple-mentable ORM. They illustrated their argument by using two real world related industrial examples (one in transporta-tion routing, and one in emergency management). Later on, extending the individual decision maker (e.g., risk manager) based ORM frameworks, Beroggi and Wallace [86] explored the case when there are multiple decision makers involved in the process. They developed two dynamic preference aggrega-tion models which can be employed for cardinal and ordinal preference assessments. They used field data to validate their

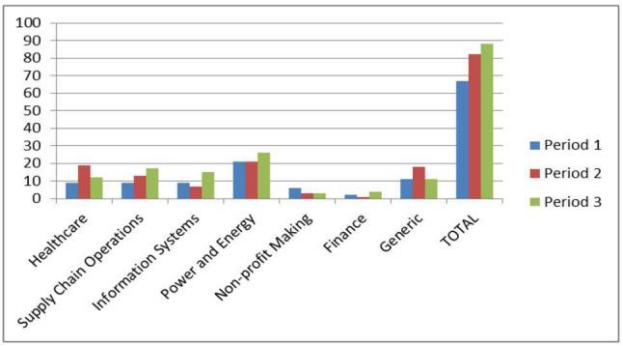

Fig. 3. Summary of recent publications related to ORM in IEEE TRANSACTIONS.

proposed models. They also conducted a sensitivity analysis to demonstrate the robustness of their models. From the perspec-tive of a practitioner, Breden [84] examined the value which can be brought by proper ORM to the companies. With the meaning very close to [85], Breden [84] defined the opera-tional risk as "the risk of loss resulting from inadequate or failed internal processes, people and systems or from exter-nal events." He argued that operational risks emerge in the industry as time evolves in the presence of changes. He then examined the role of ORM within the regulatory framework and mentions that the regulation could be treated as "an inte-gral part of the business decision making" mechanism. He also discussed the formation of the ORM framework which should ensure the operational processes are robust and resilient for evaluating and managing operational risk. Minciardi *et al.* [98] investigated the optimal resource allocation problem with the integrated preoperational and operational management frame-work for natural hazards. They extensively examined different alternatives of modeling the resources under a determinis-tic setting. Mendonca and Wallace [99] developed a cognition scientific "process-level" model in improvisation. They imple-mented this model in the domain of emergency management and proved that their model can successfully help to enhance the respective ORM. Most recently, Chronopoulos *et al.* [101] employed the real options approach to study production capac-ity planning problems. They incorporated important factors such as risk aversion of the company and the probable operational flexibility into the analytical model. They coun-terintuitively found that an increased degree of risk aversion may actually enhance investment by decreasing the optimal capacity. Fig. 3 shows the ORM related publications in IEEE TRANSACTIONS. We show the figures in three periods, namely period 1 (2005–2007), period 2 (2008–2010), and period 3 (2011–2013). The publication in 2014 is not included as we are using three years as a period, and the full set of publica-tion data in 2014 was not available at the time the dataset was collected.

From Fig. 3, we can observe several trends: 1) ORM is a popular topic and it has a steady increase in popularity over the past decade; 2) the top four most popular domains are (in a decreasing order) power and energy, healthcare, sup-ply chain operations, and information systems; and 3) only relatively few operational risk analysis related research on finance and nonprofit making organizations (outside the field of healthcare) has been published in IEEE TRANSACTIONS in the reported periods. At the first glance, it is a bit surprising to see that operational risk in the finance sec-tor is not popularly published in IEEE TRANSACTIONS. However, it is actually reasonable as there is no specialized IEEE TRANSACTIONS on "financial engineering" and there are other premier outlets in applied mathematics and busi-ness for the respective publications of top-tier research in financial ORM.

B. Operational Risk Management With Big Data

In the big data era, operational risk can come from differ-ent perspectives. First, the value of information assets is very tricky and some can be evaluated by using the value-at-risk kind of measure but not all [89]. Second, the associated cost with big data affects the real operational cost significantly. As indicated by Tallon [89], if companies tend to believe that the cost of retaining data is zero or almost zero, they will keep a lot of redundant data without any careful planning. This leads to poor data-driven decision making in real operations. In addi-tion, the real cost is projected to be much bigger than zero because

of the other associated high maintenance and soft-ware costs. Third, in the presence of the Internet, e-commerce is now very popular which helps collect big data for all kinds of business analytics. However, cultural and political risks arise from such operations which can be visualized by the debate on whether it is legal and ethical [90] for companies to keep the IP address information of customers for a long period of time (e.g., in some European countries which have strong concerns on privacy, keeping IP address information is viewed as a breach of privacy [89]). In addition, with the "new deal on data" [102], there will be a lot of legal concerns and new regulations governing the use of big data in business operations.

To achieve the big data related ORM, some measures can be taken. For instance, Esteves and Curto [97] analyzed companies' implementation of big data analytics from the per-spective of risk and proposed the planned behavior theory based model to help. Tallon [89] proposed ways to estimate the value of information assets and achieve cost control. To be specific, he argued that the financial estimate for big data storage investment can be similar to the calculation following the financial instrument such as insurance. For con-trolling cost risk in big data operations, he suggested the storage tiers method in which the most valuable (top-tier) data storage technology is used for the most important data fol-lowing the Pareto principle. Breden [84] proposed companies to set a tolerance level for each important risk related fac-tor and incorporate the regulatory bodies' guidelines into their ORM scheme. He also believed that proper ORM must be based on both internal events (e.g., operational problems and accidents) and external events (e.g., changes in the market and threat from competitors). Zou *et al.* [87] developed a Bayesian Markov chain Monte Carlo (BMCMC)-based frame-work for ORM. They noted that it is difficult to estimate the distribution parameters of the nonconjugate distribution under 8 the Bayesian framework. Thus, they proposed the BMCMC framework to help obtain the posterior distribution for the nonconjugate distribution. This BMCMC framework should be useful for ORM in the big data era because it can incorporate a massive amount of both internal and external loss data in the analysis. Recently, Zheng and Litvinov [88] developed a novel three-process ORM framework for future grid oper-ations. Their framework includes the robust systems capacity reliability process, the scenario-based commitment process, and the systems economics dispatch process with corrective measures. Cornalba *et al.* [100] explored ORM in the health-care sector. They developed a DSS which can merge prior knowledge and the available data together to help estimate the risk profile of patients in the clinic. This helps a lot with the diagnosis.

Table IV shows a summary of different major ORM frame-works, with their strengths and weaknesses, reviewed in Sections IV-A and IV-B. From Table IV, we can see that there still lacks a whole comprehensive and holistic picture of ORM for industrial systems. In addition, most models only examine ORM of some specific functional areas but not the whole enterprise. The analysis of most studies also focuses on one optimization objective. Thus, more emphasis should be paid on employing the systems concept, considering multi-ple objectives [104], [107], and developing the enterprise risk management scheme for ORM.

In the big data era, BI is created via data mining. In the liter-ature, a number of studies are devoted to the use of data mining techniques to conduct operational risk analysis. We review them as follows. First, in the scope of financial related operational risk, Koyuncugil and Ozgulbas [91] explored how data mining can be used for developing the financial early warning system for risk detection. Hailemariam *et al.* [92] explored the data mining-based techniques and algorithms for forecasting customer loyalty and financial loan default risk. They concluded by proposing to future researchers that it is important to examine the use of multiple algorithms and big data for developing data mining-based customer loyalty and default risk prediction systems. Yu *et al.* [93] compared the performances of four commonly seen data mining tech-niques, namely the logistics regression, the decision tree, the support vector machine, and neural networks, on evaluating individual credit risk. They found that the linear regression and the support vector machine-based data mining methods yield the best classification accuracy whereas the support vec-tor machine gives the highest robustness. They further revealed that the decision tree-based data mining method is very sensi-tive to the input data and the classification result is unstable. Second, in the domain of assessing the risk of management fraud, Deshmukh and Talluru [94] discussed the application of a data mining technique which can incorporate big data into the analysis. To be specific, they employed a data mining tool to study the management fraud data collected from a big enter-prise. They compared the results with other commonly used techniques such as statistical-based methods and the neural networks. They argued that data mining can be a competitive and appealing tool for ORM. Third, in healthcare, the use of data mining for risk analysis is a timely topic. Chang *et al.* [95]

TABLE IV
SUMMARY OF DIFFERENT MAJOR ORM FRAMEWORKS
REVIEWED IN THIS PAPER

ORM Model	Strengths	Weaknesses	Probable Solutions/ Remarks
B&W[85]	Real time decision mechanism	Work for individual decision maker, not a group of decision makers	It is the pioneering systems engineering study on ORM. To overcome the weakness, extension can be made to support group decision making
B&W [86]	Flexible and can work for cardinal and ordinal preference assessments	Assuming consistency of preferences	It is an extension to [85] which supports group decision making. Further extension can be made to deal with the preference inconsistency issue
Breden [84]	Include regulation in ORM	Not analytical	From the perspective of practitioners. Further analytical study should be conducted.
Tallon [89]	Consider value, risk and cost with big data operations and provide a framework to estimate value of information assets	Not analytical	It mentions that data policies governance should strike a balance between value and risk. Further analytical study should be conducted.
F&B [90]	Consider risk, data protection and ethics with big data	Not analytical, only consider the cases in North America and the UK	Exploratory in nature. Further scientific study should be conducted.
E&C[97]	Employ planned behavior theory based model to enhance implementation of big data based ORM	The conclusion is drawn with a relatively small sample size	Exploratory in nature. Conduct scientific behavioral studies to verify the empirical observations
Z&R[87]	Overcome the challenge of estimating distribution parameters for non-conjugate distributions	No closed form analytical results	Enhance the efficiency of the algorithm
Z&L[88]	Robust and flexible	Relatively complex to use	This method is mainly for future grid operations
M,S&T[98]	Develop a two-phase framework which includes pre-operational and operational risk management of natural hazards	Assuming the decisions are taken by one decision maker who has full access to the needed information	Extend it to support group decision making
M&W[99]	Develop a scientific cognitive model for improvisation in emergency management	The model does not link up with other existing models and does not consider important factors such as risk and time constraints	Enhance linkage by using, e.g., the ERP system concept
C,B&B[100]	Develop a new system to help risk assessment based on domain knowledge and individual patient's data	The system is separated from the other systems in the clinical routine	Integrate it with the existing systems in the clinical routine
C,D&S[101]	Apply real options approach in finding the value of capacity sizing under risk aversion	The project is assumed to have an infinite lifetime	A stylized model based analysis with an increasing concave utility function for the firm

proposed the use of data mining techniques for indicating the common risk factors for multidiseases prediction. They made use of some real data from a physical examination center to

CHOI *et al.*: RECENT DEVELOPMENT IN BIG DATA ANALYTICS FOR BUSINESS OPERATIONS AND RISK MANAGEMENT

TABLE V
SUMMARY OF FUTURE RESEARCH AREAS

TOPICS	IMPORTANT AREAS FOR FUTURE RESEARCH
(A) Technological advances and business intelligence	1. Synergizing multiple research methodologies so that different knowledge can be dealt with by different research methods for "big-data" research. This can be done by, e.g., employing real time data to fine tune the analytical optimization model developed based on historical big data, and including behavioral factors (such as bounded rationality) into the decision making framework to support business managerial decisions. 2. Investigating mining ontologies, rather than just algorithms for analysis with big data.
(B) Systems security and reliability	1. Exploring the ICT supply chain's systems reliability and security, especially from the total systems perspective. 2. Developing theories on big-data driven industrial systems reliability based on the current exploratory studies. 3. Developing evaluation index and early warning systems for systems reliability and security. 4. Establishing the measures which can properly handle and cope with security breach. For example, companies can impose multi-layer protection schemes to enhance system security and also implement well-defined standard operating procedures to ensure the business operations would follow the rules to avoid security breach.
(C) Operational risk management	1. Looking into measures which enhance the quality of big data and formalizing these measures with the support of scientifically sound theoretical evidence. 2. Building a formal operational risk management framework specifically tailored for the big data era with special emphasis on achieving systems optimality. 3. Examining more on the under-explored sectors such as operational risk analysis for non-profit making organizations.

conduct their analysis. They found that the accuracy of the proposed data mining technique is very high and hence they believed that data mining is the right tool for this disease pre-diction application. Petrus *et al.* [96] explored the use of the decision tree method and the data mining method for studying the patient classification problem for the emergency depart-ment's operations in hospitals. They found that data mining is useful for classifying patients.

V. CONCLUSION

There is sufficient supporting evidence to conclude that data-driven approaches would be a growing research method-ology/philosophy in business operations. Countless application domains can be influenced by this big data fad. BI systems are definitely on the list as such systems highly rely on the input data to generate valuable outputs. That being said, the scope of BI systems is so wide and related research involved the multidisciplinary knowledge. Hence it is not surprising that the research focal points have been scattered around dif-ferent disciplines. Consequently, it is not easy to generalize the results from previous studies. In this connection, emerg-ing big-data-oriented research may need some adjustments. Synergizing multiple research methodologies could be one direction. Data mining is still the core engine of BI systems but previous data mining algorithms are very application-oriented. This is not a criticism but an observation. The main reason is due to the nature of the data involved. So, soft computing techniques may be more applicable in this regard. In addition, coupling with the big data era, it may be the right time to think about mining ontologies, rather than just algorithms.

There are many applications of big data in industrial sys-tems reliability [100], [109] and security. However, very few prior studies focus on the security of the "information and communication technology" (ICT) supply chain. It is critically important as it is the base of all the applications. ICT mediated supply chain is the necessary carrier of big data as it produces all the software, hardware, and information infrastructure for big data's collection, storage, and application. So in the future, we need to address the reliability and security problems that an ICT mediated supply chain faces [83]. To be specific, it is important to construct the evaluation index systems and early warning systems for systems reliability and security manage-ment. It is also critical to explore the measures which can properly deal with security breaches. In addition, as revealed by our analysis, more emphasis should be paid on achieving systems optimality (i.e., examining the whole picture from the perspective of the enterprise). Furthermore, since most of the current studies on data-driven industrial systems reliability are only exploratory in nature, deeper research on laying the the-oretical foundation on the topic should be conducted in the future.

In ORM, the existing frameworks are all mainly on traditional operations with an emphasis on risk analysis [116], [125]. In the big data era, there are new chal-lenges associated with the specific cost and value features of the data systems. There are also concerns regarding the data quality and auditing. All these are important and they affect the industrial operations significantly. As a consequence, a formal ORM framework specifically tailored for the big data era should deserve further and deeper explorations. This framework should be supported by theoretical research and can achieve the globally optimal solution for the whole enterprise. In particular, how values, costs, and risks of the big data sup-ported operations should be assessed, measured, and controlled are critical areas for future investigation. Furthermore, as revealed in our review data analysis, more ORM research should be conducted on the domain with nonprofit making organizations. We conclude this paper by presenting Table V which summarizes the future key research areas proposed in this paper.

ACKNOWLEDGMENT

The authors would like to thank the Chief Editor, the Associate Editor, and the three anonymous reviewers for their constructive and important comments.

REFERENCE

[1] N. Manwani and P. S. Sastry, "Noise tolerance under risk min-imization," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 1146–1151, Jun. 2013.
[2] H. K. Chan and F. T. S. Chan, "Early order completion contract approach to minimize the impact of demand uncertainty on supply chains," *IEEE Trans. Ind. Informat.*, vol. 2, no. 1, pp. 48–58, Feb. 2006.
[3] G. M. Gaukler, "Item-level RFID in a retail supply chain with stock-out-based substitution," *IEEE Trans. Ind. Informat.*, vol. 7, no. 2,
pp. 362–370, May 2011.
[4] K. Govindan, A. Jafarian, M. E. Azbari, and T.-M. Choi, "Optimal bi-objective redundancy allocation for systems reliability and risk management," *IEEE Trans. Cybern.*, to be published.
[5] B. Shen, T.-M. Choi, Y. Wang, and C. K. Y. Lo, "The coordination of fashion supply chains with a risk-averse supplier under the markdown money policy," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 43, no. 2,
pp. 266–276, Mar. 2013.
[6] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of*
Investment. New York, NY, USA: Wiley, 1959.
[7] D. D. Wu and D. Olson, "Enterprise risk management: A DEA VaR approach in vendor selection," *Int. J. Prod. Res.*, vol. 48, no. 16,
pp. 4919–4932, 2010.
[8] D. D. Wu and D. Olson, "Enterprise risk management: Coping with model risk in a large bank," *J. Oper. Res. Soc.*, vol. 61, no. 2,
pp. 179–190, 2010.
[9] D. L. Olson and D. D. Wu, "Risk management models for supply chain: A scenario analysis of outsourcing to China," *Supply Chain Manag. Int. J.*, vol. 16, no. 6, pp. 401–408, 2011.
[10] V. Agrawal and S. Seshadri, "Risk intermediation in supply chains," *IIE Trans.*, vol. 32, no. 9, pp. 819–831, 2000.