Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery

Nedim Dedić (🖂) and Clare Stanier

Faculty of Computing, Engineering & Sciences,
Staffordshire University, College Road, Stoke-on-Trent, ST4 2DE, United Kingdom

nedim.dedic@research.staffs.ac.uk, c.stanier@staffs.ac.uk

**Abstract.** Confusion, ambiguity and misunderstanding of the concepts and terminology regarding different approaches concerned with analysing massive data sets (Business Intelligence, Big Data, Data Analytics and Knowledge Discovery) was identified as a significant issue faced by many academics, fellow researchers, industry professionals and domain experts. In that context, a need to critically evaluate these concept and approaches focusing on their similarities, differences and relationships was identified as useful for further research and industry professionals. In this position paper, we critically review these four approaches and produce a framework, which provides visual representation of the relationship between them to effectively support their identification and easier differentiation.

**Keywords**: Business Intelligence, Big Data, Data Analytics, Knowledge Discovery.

Introduction

During our academic and industry based work, we identified an issue faced by many academics, fellow researchers and domain experts – namely confusion, ambiguity and misunderstanding of the concepts and terminology regarding the different approaches that are concerned with analysing massive data sets. For example, during the development and validation of a framework for Business Intelligence (BI), we were asked why we omitted concepts, such as Big Data (BD), Data Analytics (DA) and Knowledge Discovery (KD) from the proposed framework. Such questions are justified as there is no generally accepted unified standard or framework encompassing the fields of BI, BD, DA and KD, which we collectively define as a *cluster of concepts concerned with analysing massive data*. Although these and similar questions may be outside the scope of a particular research project, it can require additional time and effort to answer them, especially in a research context as discussion has to be based on analysis of the existing literature.

The literature evaluated in this paper, typically deals thoroughly with each of elements from the *cluster of concepts concerned with analysing massive data* but approaches the topic from specific perspectives*.* If treated at all in the literature, issues such as possible confusion, ambiguity and misunderstanding of concepts and the interrelations between approaches, are treated only superficially, and are mostly focused on differentiation between BD and BI concepts from data structure perspective.

Thus, we identified a need to critically evaluate the *cluster of concepts concerned with analysing massive data.*  The aims of this position paper are: (i) to critically examine and review BI, DA, BD and KD to identify disparities, similarities and relationships each to another, (ii) to provide “a tool” in the form of graphical representation to be used by other researchers to quickly respond to the issues, questions and problems based on misunderstanding of the concepts concerned in this paper, (iii) to create a base for, and to provoke further discussion, between researchers and scientists, and (iv) to propose an initial conceptual framework, which will support further research and enable easy identification and differentiation of the concepts and approaches discussed in this paper.

The rest of this paper is structured as follows: in section II, we review BI, in section III we review DA and BD and Big Data Analytics (BDA) and in section IV we discuss KD. We identify the similarities, differences and relationships between concepts and approaches. In section V, we present the framework shown in Figure 1 and discuss the validation approach.

Business Intelligence

BI can be understood as a philosophy, which includes the strategies, processes, applications, data, products, technologies and technical architectures used to support the collection, analysis, presentation and dissemination of business information [1, 34]. It helps companies to out-think the competition through better understanding of the customer base [2], which could lead to creating a closer and stronger relationship with customers and enhanced revenue [3]. It plays a critical role for business in terms of organizational development by providing competitive advantage in the context of achieving positive information asymmetry [1, 4, 5], and contributes to optimising business processes and resources, maximizing profits and improving proactive [6], and strategic decision-making [7]. Besides its strategic and tactical advantages, Business Intelligence is also used at operational level.

BI could enable various types of users to spot emerging trends, make faster decisions, take actions and cope with the organizational problems as soon as they arise [8]. Its purpose is to help stakeholders to better understand their organization’s operations, make wiser, more informed business decisions, and manage operational performance [9].

We can use BI to extract meaningful information and hidden knowledge from operational data produced on a daily basis, which would help business stakeholders in variety of predictions, calculations and analysis [10]. Conventional BI focused on activities such as ETL, data warehousing and reporting, thus covering research areas of data manipulation, propagation and visualisation [34]. However, the new generation of BI has an additional research focus on areas such as data exploration and visualisation [11, 12, 34]. There is also evidence of shifting from static reports to interactive visualisations, which extends research issues from metrics overview to discovering causes and effects of the phenomena the metrics express [12]. Additionally, the competition pressure of business causes new trends in BI and related research, such as near real-time BI, data mining and text analytics [13], self-service BI and BI in cloud [11].

Data Analytics & Big Data Analytics

Data Analytics is the process of supporting effective decision-making through analysis of the existing data sets by using computer systems [14]. Ridge provides a broader definition and defines DA as any activity that involves applying analytical processes to data for the purpose of deriving insight from data [15].

It is an interdisciplinary field that includes many other scientific disciplines, such as computational intelligence, statistics, machine learning, signal theory, pattern recognition, machine learning, operations research [14], predictive analytics, data mining, artificial intelligence, natural language processing [16], business intelligence, prescriptive analytics and descriptive analytics. As such, research areas and issues relevant for DA concepts include research areas and issues such as visualisation, cloud computing or data exploration, already identified as research areas in BI.

The mission of DA is to access and analyse data, and to gain insight into significant trends or patterns in organizations [17]. It provides managers with access to timely information and supports decision makers highlighting useful information [18]. As it supports advanced continuous monitoring and auditing [18], we can use it to examine various data sets to support operations in different industries [19]. It is also offers the opportunity, for example, to discover new customer segments, identify associated products, understand seasonal trends, or identify the quality of suppliers’ [16].

DA is used by various industries, such as governmental organizations [18], healthcare, medicine [20], security [21], business, engineering, finance, operation management [22] and biomedical research [23].

Big Data

BD is concerned with large-volume, complex, ever growing data sets coming from various often autonomous sources [24], such as environmental and body sensors, mobile devices, administrative claims data, social media, emails, laboratory studies, electronic medical records, internet, business transactions, geospatial devices or sensors [22, 23, 25]. The rate of BD generation is extremely fast and BD may be generated by machines or humans [19].

Similar to Wu et al [24] and Barton [23], many BD definitions and explanations are focused on the volume of data. However, BD is not only related to massive data [19], and there are other characteristics of BD, which are important and must be considered [16]. Traditional definitions of BD, include variety and velocity in addition to volume as basic constituent elements [16, 26]. Those three attributes are conventionally known as “three Vs”.

Some definitions of BD go even further and include even more dimensions such as Veracity [25, 27, 28], Validity, Value, Variability, Venue, Vocabulary, and Vagueness [29] although the relevance of some of these elements is not clear. BD is everywhere around us [22]; in education [30], health care [23], engineering, operations managements, genomics [22], biomedical research [31] any many other fields.

However, massive and ever increasing data is useful only if it can be analysed [19]. Seen in the past as a technical problem, BD is today seen as a business opportunity [18], which can provide new opportunities based on the analysis of data [22]. The basic challenge of BD is to explore large data for the purpose of extracting useful information and competitive knowledge [32]. Unlocking the value of BD in complex and rapidly changing markets can bring competitive advantage and enable better response by businesses [25].

The definition of BD encompasses variety of data, which can be unstructured, semi-structured, and even structured. However, BD is most often concerned with unstructured and unorganized data [19, 23, 25, 27, 31].

On the other side, BI is mostly related to structured data, thus, we consider BD as a parallel philosophy to BI. However, this does not mean that they or their components are mutually exclusive. For example, DW as a core component of BI that works with structured data, can be used as additional part of the BDA process [25]. Notably, Chan [25] identified synergy between DW and Hadoop type BD architecture.

We identified quality and data exploration as the main research areas of BD and visualization as an important research area. Because of the size of the data considered, we see great potential in the research area of BD in the cloud.

Big Data Analytics

As traditional DA is not able to handle very large quantities of data [29], and because Big Data is too large and complex to be manipulated or managed by using standard tools and methods [26], we are witnessing a new trend - namely Big Data Analytics (BDA).

BDA is defined as large-scale analysis and processing of information [22], encompassing data sets that go beyond the capacity of conventional databases [33]. It is advanced analytics operating with big data sets [16]. BDA is a rapidly expanding field [20]. We see BDA as similar to DA since BDA includes inspecting, cleaning, transforming, and modelling data to discover and communicate useful information and patterns, suggest conclusions, and support decision making, however, by using BD data sets [26].

BDA provides tools and methods to accumulate, manage, analyse, combine and assimilate large volumes of disparate, structured, and unstructured data [20, 25, 26]. Besides combining data, BDA sometimes requires combinations of various methods from different disciplines [26]. As its name suggests, BDA is concerned with Big Data and Analytics [16], and is a current research and application area [24]. As a concept related to BD, BDA is concerned with the same research areas as BD.

Knowledge Discovery & Data Mining

While applying appropriate tools and software [18], BDA uses various DA methods, such as clustering, classification, association rule or sequential patterns to discover new knowledge [29]. Methods and algorithms for analysing data and identifying patterns are collectively known as Data Mining (DM) [18, 22].

DM is considered a powerful approach for developing knowledge from data [35]. DM is understood as applying data analysis and discovery algorithms to produce a particular enumeration of models over existing data [36]. In this context, data exploration is the most relevant research area for DM.

Esfandiari et al [35] state, in their reference to Fayyad et al [36], that DM was originally considered as synonym for Knowledge Discovery in Databases (KDD). However, in the original text, Fayyad et al [36] regard DM as a step in the KDD process, which includes application of specific algorithms for extracting patterns from data. Chen at al. [37] regarded DM and KD in databases as synonyms.

According to Fayyad et al [36], KDD includes additional steps, which include data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge, and proper result interpretation. We consider next what is understood under the heading of Knowledge Discovery (KD)

According Cortez & Santos [38], based on the 1996 definition by Fayyad et al [36], KD is a branch of the artificial intelligence field that aims to extract high-level knowledge from complex and voluminous data, which would be useful and understandable. However, care is needed when using definitions from the period, which preceded BD. A 2004 definition by Koua & Kraak [39] defined KD as a higher-level process, which uses DM process to turn data into knowledge. More recently, 2009 [40] and 2011 [41], KD was defined not only as a branch of Artificial Intelligence, but as an interdisciplinary field with the focus on methodologies to identify valid, novel, meaningful and potentially useful patterns often from large data sets. We regard KD as a higher entity encompassing DA, which is not exclusively related only to computer-based concepts.

A Framework to differentiate Business Intelligence, Big Data, Data Analytics and Knowledge Discovery

There are many similarities between BI, DA, BD and KD, which explains the confusion concerning these concepts. However, based on our literature review and preliminary discussions with seven domain experts, we identify significant differences and argue that these concepts should be regarded as distinct approaches.

As presented in Figure 1, we see Knowledge Discovery as the highest-level concept, which in addition to other methods includes Data Analytics to discover or produce new knowledge. Within KD, we see Data Analytics as an entity, which includes various disciplines, including Big Data Analytics and Business Intelligence.



Figure 1: Visual representation of the relationships between approaches

For the purposes of our discussion, which is focused on the analysis and use of data, we regard Big Data as part of Big Data Analytics. Taking into account intention, purpose and underlying business philosophies, we see Big Data Analytics and Business Intelligence at the same level. However, taking into account technical structure, relevant software applications and data, we also see Big Data and Business Intelligence as concepts at the same level.

We see data focus as the major difference between BI and Big Data. Big Data encompasses unstructured, semi-structured and structured data, however the main focus is on unstructured data [19, 23, 15, 27, 31], while the focus of BI is on structured data. While BI requires DW and/or data marts to support reporting [42, 43], Big Data can work with DW but DW are not required [25] and there are many alternative supporting technologies such as the Hadoop platform. In reports based on traditional BI systems, there is a requirement to have structured master and transactional data. For example, to use or analyse sales transactional data we must have master data describing the properties of sales (such as store, location or product descriptions). Big Data is not subject to those requirements. For example, the analysis of the content of emails or appeals submitted to public administration institutions does not require structured data.

Validation of the framework proposed

In the academic community, validation is usually based on approaches such as formal interviews or surveys. These approaches are time consuming for participants and we were interested to see whether professional networks could be used to gather responses. The concepts proposed in this paper are intended as discussion points and for this reason we wanted to reach out to the professional practitioner community, to evaluate whether the distinctions we propose match the real world experience of those working in the field. Thus in addition to online survey and direct discussion with domain experts, we used an additional method of communication, namely special interest groups on internet-based business social networks to discuss and validate the proposed framework. We exposed the framework from Figure 1 to relevant domain experts for discussion as a post using appropriate LinkedIn groups.

To better understand the validation process, we explain which kinds of feedback are available in these forums. *Like* is the lowest level category feedback on content that a user can provide in internet-based business social networks. However, it has positive impact as it indicates that the user found the content interesting, useful or worth considering. *Comment* can be considered as a stronger category of feedback in regard to *like*. It is direct discussion about content, which can have positive or negative impact. It is also an appropriate mechanism for critically evaluating content and providing additional suggestions. *Share* can be considered as the strongest type of feedback in internet-based business social networks.It allows sharing of content on personal profiles of users called to provide feedback, in other groups, and in news feeds. By sharing respective content via internet-based business social networks, user suggests not only that respective content, concept or idea is interesting, useful or worth considering, but also worth promoting further within the community. *Share* may also have negative connotations but this was not the case in this work.

Over a period of three weeks, we received 560 feedbacks in three different forms. It was very encouraging that 384 users from business social networks found the framework good enough to be liked, while 134 decided to share the proposed framework with their professional networks as a good example explaining differences between BI, BD, DA and KD. Written comments were received from 42 users via preliminary survey and via business social networks; these comments were very positive and no negative comments were received. Suggestions for extending the framework included adding additional dimensions to the framework to categorize the type of analytics for each specific concept, such as descriptive, predictive, or prescriptive, and including concepts such as Machine Learning, Business Analytics, or Data Science. In our further work, we will review the framework and consider extending it to include additional elements.

References

1. Dedić N. & Stanier C. (2016). Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting. Lecture Notes in Business Information Processing. 268. Pp 225-236.
2. Brannon, N. (2010). Business Intelligence and E-Discovery. Intellectual Property & Technology Law Journal, 22(7), 1–5.
3. Alexander, A. (2014). Case Studies: Business intelligence. Accounting Today, (June), 32.
4. Marchand, M., & Raymond, L. (2008). Researching performance measurement systems: An information systems perspective. International Journal of Operations & Production Management, 28(7), 663 – 686.
5. Thamir, A., & Poulis, E. (2015). Business Intelligence Capabilities and Implementation Strategies. International Journal of Global Business, 8(1), 34–45.
6. Olszak, C. M., & Ziemba, E. (2006). Business Intelligence Systems in the holistic infrastructure development supporting decision-making in organisations. Interdisciplinary Journal of Information, Knowledge, and Management, 1, 47–58.
7. Popovič, A., Turk, T., & Jaklič, J. (2010). Conceptual Model of Business Value of Business Intelligence Systems. Management: Journal of Contemporary Management, 15(1), 5–29.
8. Sandu, D. I. (2008). Operational and real-time Business Intelligence. Informatica Economic, XII(4), 33 – 36.
9. American Institute of CPAs. (2015). Business Intelligence. Retrieved March 27, 2015, from http://www.aicpa.org/INTERESTAREAS/INFORMATIONTECHNOLOGY/RESOURCES/BUSINESSINTELLIGENCE/Pages/default.aspx
10. Kurniawan, Y., Gunawan, A., & Kurnia, S. G. (2014). APPLICATION OF BUSINESS INTELLIGENCE TO SUPPORT MARKETING STRATEGIES: A CASE STUDY APPROACH. Journal of Theoretical and Applied Information Technology, 64(1), 214.
11. Obeidat, M. et al., 2015. Business Intelligence Technology, Applications, and Trends. International Management Review, 11(2), pp.47–56.
12. Anadiotis, G., 2013. Agile business intelligence: reshaping the landscape. , p.3.
13. Chaudhuri, S., Dayal, U. & Narasayya, V., 2011. An overview of business intelligence technology. Communications of the ACM, 55(8), p. 88–98 .
14. Runkler, T. A. (2012). Data Analytics: Models and Algorithms for Intelligent Data Analysis (1st ed.). Wiesbaden, Germany: Springer Science & Business Media.
15. Ridge, E. (2014). Guerrilla Analytics: A Practical Approach to Working with Data. Waltham, MA, USA: Morgan Kaufmann.
16. Russom, P. (2011). TDWI Best Practices Report: Big Data Analytics.
17. Lussier, R. N., & Hendon, J. R. (2016). Fundamentals of Human Resource Management: Functions, Applications, Skill Development (1st ed.). Los Angeles, US: SAGE Publications.
18. Fadairo, S.A., Williams, R. & Maggio, E., 2015. Using Data Analytics for Oversight and Efficiency. The journal of government financial management, 64(2), p.18.
19. Henry, R., & Venkatraman, S. (2015). Big Data Analytics: The next big learning opportunity. Academy of Information and Management Sciences Journal, 18(2), 17–29.
20. Belle, A., Thiagarajan, R., Soroushmehr, S. M. R., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big Data Analytics in Healthcare. BioMed Research International, 2015, 1–16. http://doi.org/10.1155/2015/370194
21. Cárdenas, A. A., Manadhata, P. K., & Rajan, S. P. (2013). Big Data Analytics for Security. IEEE Security & Privacy, 11(6), 74–76.
22. Gerard, G., Haas, M., & Pentland, A. (2014). Big Data and Management. Academy of Management Journal, 57(2), 321–326.
23. Barton, A. (2016). Big Data. Journal of Nursing Education, 55(3), 123–124. http://doi.org/10.3928/01484834-20160216-01
24. Wu, X., Zhu, X., Wu, G.-Q., & Ding, W. (2014). Data mining with big data. IEEE Transactions on Knowledge and Data Engineering, 26(1), 97–107.
25. Chan, J. O. (2013). An architecture for Big Data analytics. Communications of the IIMA, 13(2), 1.
26. Cao, M., Chychyla, R., & Stewart, T. (2015). Big data analytics in financial statement audits. Accounting Horizons, 29(2), 423. http://doi.org/10.2308/acch-51068
27. Lokhande, S., & Khare, N. (2015). An Outlook on Big Data and Big Data Analytics. International Journal of Computer Applications, 124(11), 37–41. http://doi.org/10.5120/ijca2015905658
28. IBM. (2016). The Four V’s of Big Data. Retrieved April 13, 2016, from http://www.ibmbigdatahub.com/infographic/four-vs-big-data
29. Tsai, C.-W., Lai, C.-F., Chao, H., & Vasilakos, A. (2015). Big data analytics: a survey. Journal of Big Data, 2(1), 1–32. http://doi.org/10.1186/s40537-015-0030-3
30. Metz, S. (2015). Big Data. The Science Teacher, 82(5), 6.
31. National Institutes of Health. (2016). What is Big Data? Retrieved April 13, 2016, from https://datascience.nih.gov/bd2k/about/what
32. Rajaraman, A. & Ullman, J.D., 2011. Mining of Massive Datasets 1st ed., Cambrige, UK: Cambrige University Press.
33. Nicol, D. (2013). Mobile Strategy: How Your Company Can Win by Embracing Mobile Technologies (1st ed.). Boston, MA, US: IBM Press.
34. Dedić N. & Stanier C. (2016). An Evaluation of the Challenges of Multilingualism in Data Warehouse Development. In Proceedings of the 18th International Conference on Enterprise Information Systems, 1, 196-206.
35. Esfandiari, N., Babavaliana, M. R., Amir-Masoud, E. M., & Tabarb, V. K. (2014). Knowledge discovery in medicine: Current issue and future trend. Expert Systems with Applications, 41(9), 4434–4463. http://doi.org/10.1016/j.eswa.2014.01.011
36. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (pp. 82–88). AAAI Press.
37. Chen, M.-S., Han, J., & Yu, P. . (1996). Data mining: an overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering, 8(6), 866 – 883. http://doi.org/10.1109/69.553155
38. Cortez, P., & Santos, M. F. (2013). Knowledge Discovery and Business Intelligence. Expert Systems, 30(4), 283–284.
39. Koua, E. L., & Kraak, M.-J. (2004). Geovisualization to support the exploration of large health and demographic survey data. International Journal of Health Geographics, 3(12), 13. http://doi.org/10.1186/1476-072X-3-12
40. Fred, A. (2009). “Preface.” Preface. In International Conference on Knowledge Discovery and Information Retrieval. Madeira, Portugal.
41. Aradau, C., & Munster, R. Van. (2011). Politics of Catastrophe: Genealogies of the Unknown. Chippenham, UK: Routledge. Retrieved from http://www.kdir.ic3k.org/
42. Kimball, R., Margy, R., Thornthwaite, W., Mundy, J., & Becker, B. (2008). The Data Warehouse Lifecycle Toolkit (2nd ed.). Indianapolis: John Wiley & Sons.
43. Inmon, B.W., 2005. Building the Data Warehouse 4th ed., Indianapolis: John Wiley & Sons.