Computers in Industry 94 (2018) 52–61

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

An argumentation reasoning approach for data processing

Erisa Karafili[a,*], Konstantina Spanaki[b], Emil C. Lupu[a]

[a] Department of Computing, Imperial College London, 180 Queen's Gate, SW7 2AZ London, UK
[b] School of Business and Economics, Loughborough University, LE11 3TU Leicestershire, UK

ARTICLE INFO

Article history:
Received 31 May 2017
Accepted 8 September 2017
Available online xxx

Keywords:
Data processing
Data quality
Usage control
Argumentation reasoning
Data manufacturing
Case scenarios

ABSTRACT

Data-intensive environments enable us to capture information and knowledge about the physical surroundings, to optimise our resources, enjoy personalised services and gain unprecedented insights into our lives. However, to obtain these endeavours extracted from the data, this data should be generated, collected and the insight should be exploited. Following an argumentation reasoning approach for data processing and building on the theoretical background of data management, we highlight the importance of data sharing agreements (DSAs) and quality attributes for the proposed data processing mechanism. The proposed approach is taking into account the DSAs and usage policies as well as the quality attributes of the data, which were previously neglected compared to existing methods in the data processing and management field. Previous research provided techniques towards this direction; however, a more intensive research approach for processing techniques should be introduced for the future to enhance the value creation from the data and new strategies should be formed around this data generated daily from various devices and sources.

1. Introduction

Big data paradigm as a shifting phenomenon [14,33,7,40] provides access to a large pool of data resources; coupled with new storage, management and analytical techniques. These unprecedented opportunities arising from 'Big Data' are stressing firms to position themselves within this highly competitive data-intensive scheme, altering the way they generate, collect and transform data to actionable knowledge [10]. Gaining a competitive posture requires more than analytical techniques [36]; making sense [68] of this data is the challenge of organisations, forming a new way of data-based decision-making, disrupting the business landscape while moving from the world of "making things" to a "world of outcomes" [44]. The industrial world was disruptively changed during the last century; while the manufacturing process has gone through multiple revolutions (agricultural, industrial, digital etc.) and respectively multiple alterations in the way products and services are formed. A critical view of the literature associated with the manufacturing context [41,62] of the last decades, reveals that this evolution around data has influenced radically also the production, distribution and supply chain world.

The data evolution (also often referred as "Big Data" evolution) has also accelerated the popularity of service industries around data (for storing, analysing, processing etc.). In the context of these data services; myriads of data are shared as well as stored, used and transformed, not only by users but also by companies, organisations, and governments. Security concerns regarding shared data, e.g., privacy, confidentiality, integrity, and the necessity of protecting them are becoming serious issues. Towards this direction, the regulatory and legislative environment is continuously updated for meeting citizens' rights, protecting their data privacy, anonymity, and security. The regulatory and legislative implications of security issues are becoming rather more severe when we extend the scope across the immediate organisational, governmental or even country borders; where multiple legislative domains are applied, and frequently they lead to conflicting behaviours and interests. Exchanging data implies that all parties agree on the associated rules to be enforced; this is often referred as a data sharing agreement [59] (DSA). The DSA can be seen as a contract, between two or more parties, and the different rules are the terms of the contract. The terms express how and who is permitted or denied to access, delete, use, and share the data, along with the different constraints that should be respected.

* Corresponding author.
E-mail addresses: e.karafili@imperial.ac.uk (E. Karafili), k.spanaki@lboro.ac.uk (K. Spanaki), e.c.lupu@imperial.ac.uk (E.C. Lupu).

http://dx.doi.org/10.1016/j.compind.2017.09.002

The enforcement is done when the data are used, and it evaluates the requests and usage of the data against the set agreements.

Constructing and representing DSAs is not trivial, as they should incorporate data access and usage rules for the security and privacy of the data, as well as users preferences, business and legislative rules applicable to that case. All the above rules are applied to the same bunch of data; therefore there can be conflicting behaviours between rules, and differentiated legislative, regulatory rules and domain contexts that bring inconsistencies inside the DSAs rules. Different techniques [42,63,70] were introduced for data services towards various security solutions, with a major focus on data access permissions and an efficient usage control [17,69], suitable for stored or integrated data from multiple parties and sources. Access and usage control is a well-studied research area [39,53], however the existing solutions do not permit a fine grained representation of the different types of rules and their constraints, as well as the associated conflict detection and resolution.

The aim of this study is to propose a novel data processing technique using a policy analysis language for the representation of the sharing agreements and quality attributes of the data. We extend our research scope in an industrial context where the data quality control should also be studied as it is often a neglected aspect in the context of data sharing agreements. Initially, we provide the theoretical background where this study was based, coupling two streams of literature. Previous research in data management was examined in order to explain the data manufacturing analogy and the quality problem which was also highlighted in past decades, as well as the data access and usage control background for building the concepts associated with DSAs. The following sections develop the proposed methodological approach based on a policy language for argumentation and abductive reasoning used in the data processing context, and its explanation through a use case scenario and the relevant analysis and representation. Our study ends up with a research agenda for the future implications of this approach, as well as how this can be extended and applied in industrial data-intensive environments.

2. Theoretical background

The analogy of the data processing to the manufacturing processing of physical materials is prevalent in the literature of data management. There are many similarities between the mechanisms of data processing and the manufacturing processing; however, there are some significant differences. In a manufacturing process, physical materials are input into a process, the materials are transformed, and the resulting output is a manufacturing product. In the data processing, the data represent the input into the processing mechanism, and a transformed data product is the output of the process. Data of bad quality used through the data process will remain bad until the quality is improved (until the problem is actively cleaned up or removed). Data sharing controls were not well-studied in previous decades, as the data were mostly shared within the boundaries of a company or between single databases, where the trust and security issues were solved by the individual sharing entities and the associated agreements between interested parties.

The data manufacturing framing presented in data management literature of the previous decades should be extended in a context across boundaries between individuals, firms and countries and the target should be tailored data products, as these are mostly the results of the data evolution. Data manufacturing analogy was based on the data artefact as a unit of analysis; however, data era requires novel techniques focusing on the processing of data but not solely on their processing mechanism, but also the quality and sharing attributes associated with them. This section will explore the previous research of data management and data sharing literature, with the view to provide a background for the problem area and how this is expanded in a data-intensive industrial context.

2.1. Data management and quality attributes

Plenty of examples nowadays can provide evidence that companies from multiple industries invest in data management solutions, with the view to improve and expand the production of enterprises through the use of their analytical skills, or by viewing and optimising their supply chains of their core business [62,61]. Although data can be used along with the core business focus in different industries, the recent data evolution expanded the business scope and disrupted the operating models providing opportunities to process this data, create new data and information products/services and also to resell and exchange this data. Reviews of the literature in the context of data management in an industrial context mostly focus on “Supply Chain Analytics (SCA)” [64] as a way in “developing supply chain strategies and efficiently managing supply chain operations at tactical and operational levels” [64]. Supply Chain Management (SCM) focuses mostly on how the analytics can be applied to strategic decisions related to SCM [64,62], how efficiency and effectiveness of supply chains can be improved through the use of data [61] as well as the data strategies and servitization around supply chains [46]. This direction reflects that the research focuses mostly the use of data within an industrial context; nevertheless, our research focus will be on data processing techniques, setting data manufacturing analogy from data management literature as the main theoretical background, analogous to the product/service manufacturing processes. Literature reviews and frameworks referring to data/information processing are presented and summarised in Table 1, in Appendix A. This summary reveals that data analogous to physical materials are moved through a manufacturing process which reshapes/reconfigures them in information/data products.

Data processing was initially introduced by Brodie [6] through the analogy between product manufacturing and data manufacturing when data quality was a primary concern in transforming data to valid information and knowledge [1]. Some of the most indicative studies around these areas developed the concept of data manufacturing analogy in order to find out the path for better data quality [38,1,18] and they provided frameworks to describe and track data manufacturing process [65,2,66,56]. A simple framework of input-process-output describing the similarities between the two manufacturing processes was proposed in [66] and calls for continuously defining, measuring, analysing, and improving data quality. Mostly, the data manufacturing analogy was focusing on data quality and the ways to ensure that we can trust the data we use in manufacturing processes. Recent studies following the data manufacturing path can be considered those in [19,16] where they introduce the need for continuous improvement in the SCM data production process, suggesting a framework for establishing a data quality control mechanism. These two studies are investigating the data manufacturing process via a data quality lens and expand the research focus to new topics related to the processing of data, and how to improve the quality of this data with the use of various techniques and methods.

2.2. Data access and usage control

Research around data sharing has provided data-centric solutions for protecting the used and shared data, aside from focusing on protecting the databases where they are stored [15], the network used for their transfer [28], or constructing