Computers & Industrial Engineering 112 (2017) 459–465

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

BIG DATA and Data-Driven Intelligent Predictive Algorithms to support creativity in Industrial Engineering

Francis Rousseaux

Reims Champagne Ardenne University, France
Institut de Recherche et de Coordination Acoustique/Musique, France

ARTICLE INFO

Article history:
Available online 4 November 2016

Keywords:
Knowledge management
Predictive analysis
Inductive data-driven algorithms
Big data
Figural collections
Creativity in Industrial Engineering

ABSTRACT

Computer scientists, as far as they aim at modeling knowledge, are faced with the so-called "curse of dimensionality", which is nothing but the common name of the "combinatorial explosion" pitfall.

This situation is directly linked to the type of knowledge model they are looking after, fitted to traditional computation, namely formal and hierarchical, based on sub-categorization: better modeling quality is supposed to follow better accuracy, indefinitely, leading to costly combinatorial explosions.

What if we consider situated and/or embodied knowledge theories, where each piece of knowledge can make different sense depending on its local variations, including its time and space configuration? Is that even worst for computer scientists? Most of current computer science researchers answer positively: because the traditional "knowledge modeling" paradigm is not questioned enough.

In this paper, we propose a solution to go beyond that difficulty, using Big Data and Data-Driven Intelligent Predictive Algorithms to support creativity in "knowledge collection making", which aims at electing meaningful knowledge spatiotemporal configurations. Thanks to that innovation, accuracy is not anymore the only parameter to play with for targeting knowledge improvement, but also its relative disposition.

1. Introduction

The classical processes of decision making in Information Technology are traditionally based on deductive knowledge modeling and simulation, so that explanation and theory—based on validations or refutations (Pierce, 1903)—usually constitute the kernel of industrial engineering knowledge management systems.

Things turn different within Data-Driven Intelligent Predictive Algorithms and Big Data, because practitioners do not need anymore domain theories and explanation-based processes to succeed in providing useful results (Hazen, Skipper, Ezell, & Boone, 2016; Li, Ch'ng, & Yee-Loong, 2016; Zhong, Newman, Huang, & Lan, 2016).

2. Predictive analysis through Big Data: a Tsunami in IT decision systems

2.1. What changes make Inductive Data-Driven Algorithms and Big Data in the data ecosystem?

In the world of computing, we are currently living through a period in which the epistemological nature of data is being thoroughly changed, as Big Data seems to be revolutionizing most of the methods and approaches of digital decision-making tools and processes (O'Neil, 2016). Indeed it is as if the Curse of Dimensionality and the pitfall of the Combinatorial Explosion that limited computer science in the 20th century have been left behind us. Natural allies of Data Mining and Machine Learning, the tendency is now towards Big Data and Data-Driven Intelligent Predictive Systems (DDIPS).

These systems are called Data-Driven because they mobilize Big Data to make connections or analogies, aiming to create certain configurations or to anticipate situations that might cause delays or predictive challenges. They are called Predictive because, unlike 20th century computer systems, their performances are measured more by their ability to predict or to discover rather than to understand, explain or theorize.

They are called Intelligent because they frequently make use of input from supervised or unsupervised automatic learning techniques, data mining, and even Deep Learning based on Convolutional Neural Networks, which contribute to performances that far outstrip the preceding generation's.

Because of this epistemological turnaround, and with the help of DDIPS, a flood of predictive data are being added to the usual data, although this new type of data has never been directly input

E-mail address: francis.rousseaux@univ-reims.fr

http://dx.doi.org/10.1016/j.cie.2016.11.005
0360-8352/

or calculated by determinist or classically deductive methods (Gkoulalas-Divanis, Saygin, & Verykios, 2011; Slavkovic & Smith, 2012). Thus Predictive-Data – to be distinguished henceforward from other types of data – are considered an additional raw material to be exploited, even if their reliability must be closely scrutinized.

2.2. What about the current positions and statements made by main actors?

IT professional and industrial companies, but also start-ups working in that domain, are putting forward on promising business and marketing opportunities, furnishing and putting tool and technical solutions on market places.[1]

But in parallel observers and researchers, often coming from social sciences fields, forewarn potential users from a new calculus order that could mean their end of autonomy. To mention only a few of them:

"The end of Theory"

Following George Box saying: "*All models are wrong, and increasingly you can succeed without them*", Anderson (2012), after he has been the editor-in-chief of WIRED magazine, puts forward that: "The Data Deluge Makes the Scientific Method Obsolete".

He claimed that: "*Petabytes allow us to say: Correlation is enough. We can stop looking for models. We can analyse the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot*".

This opinion is so radical that we propose to consider it like a prophecy, more than an assertion.

"*Technological Solutionism*"

In one of his recent bestseller books Morozov (2014), the Byelorussian expert analyst Evgeny Morozov explained: "*There is something, to me, that is very worrying about the idea of replacing causality with correlation, because if you do want to engage in reform, you do need to understand the causal factors that you will be reforming. If you just focus on correlations, all you'll be doing is basically adjusting the behaviour of the system without understanding the root causes that are driving it*".

And also: "*The proliferation of big data and the ability to track things that we do is good only if we can actually understand why we engage in those behaviours. The ability to understand why I think is fundamental to understanding what it is that needs to be changed*".

This opinion is also radical, because « replacing causality with correlation » is not the real matter. "Can correlation lead to causality discovery?" would have been a more productive question.

"*Algorithmic Governementality*"

The European researcher Rouvroy (2012, chap. 5) claims that (Rouvroy, 2014): "*The implicit belief accompanying the growth of 'big data' is that, provided one has access to massive amounts of raw data (and the world is actually submersed by astronomical amounts of digital data), one might become able to anticipate most phenomena (including human behaviours) of the physical and the digital worlds, thanks to relatively simple algorithms allowing, on a purely inductive statistic basis, to build models of behaviours or patterns, without having to consider either causes or intentions. I will call 'data behaviourism' this new way of producing knowledge about future preferences attitudes, behaviours or events without considering the subject's psychological motivations, speeches or narratives, but rather relying on data. The 'real time operationality' of devices functioning on such algorithmic logic spares human actors the burden and responsibility to transcribe, interpret and evaluate the events of world. It spares them the meaning-making processes of transcription or representation, institutionalization, convention and symbolization*".

Her positions are close to Catherine O'Neil's ones (O'Neil, 2016). For those researchers, some use of certain big data tools could lead to unexpected side effects they are trying to characterize. The questions then are: "How to control and limit those dangerous side effects?" and: "How to control and generate some interesting side effects?".

2.3. Why not to take better advantages of the specificities of Big Data technologies?

This epistemological breakthrough in the IT classical environments allows our digital societies to revisit an old human utopia: the management of our overabundant and oversized collections.

Many of our modern computerized activities, may they be personal, professional or even artistic, involve searching, classifying and browsing large numbers of digital objects.

Until recently, the usual tools we had at hand, however, were poorly adapted as they were often too formal, because the current models for information search often assume that the function and variables defining the categorization are known in advance.

In practice, however, when searching for information, experimentation plays a good part in the activity, not due to technological limits, but because the searcher does not know all the parameters of the class he wants to create. He has hints, but these evolve as he sees the results of his search. The procedure is dynamic, but not totally random.

The collectors always experiment ways to carry out objects situated in space/time. Here, the intension of the future category has an extensive figure in space/time. The system of extension gives as many ideas as it does constraints. What is remarkable is that when we collect something, we always have the choice between two systems of constraints, irreducible one to the other. This artificial 'u ndifferentiation' for similarity/contiguity is the only possible kind of freedom allowing us to categorize by experimentation.

Nowadays, our software design could strongly become backed up by both artistic and psychological knowledge concerning the ancient human activity of collecting, which can be described as a metaphor for categorization in which two irreducible cognitive modes are at play: aspectual similarity and spatiotemporal proximity.

Inductive Data-Driven Algorithms and Big Data could help to allow the creation of a new operational space, in between formal classes/categories and radical singularity.

3. The need for computer-aided collections management tools

3.1. An illustrative example

Let us illustrate this situation with a daily example. Looking for new material and classifying are two important processes involved in collecting. Indeed, when someone decides to start building a collection he usually already possesses a few items. Then, to extend this collection, new items must be added. In order to do so, the collector goes into the world and looks for these new items. As the collection builds up, the need to arrange the items into categories will become clearer, as the collection cannot simply remain a messy stack of unordered items (Pomian, 1987).

Let us describe a particular example: the music collector (Pachet, Aucouturier, La Burthe, Zils, & Beurive, 2006; Pachet, 2003). This collector will surely possess some initial items; these

[1] See for example http://www.bigdataparis.com/guide/BD14-15_Guide_BD_14136_2.pdf