Computers in Industry 100 (2018) 227–243

Contents lists available at ScienceDirect

Computers in Industry

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

Review

Deep learning for big data applications in CAD and PLM – Research review, opportunities and case study

Jonathan Dekhtiar[a], Alexandre Durupt[a,*], Matthieu Bricogne[a], Benoit Eynard[a], Harvey Rowson[b], Dimitris Kiritsis[c]

[a] Université de Technologie de Compiègne, Department of Mechanical Engineering, UMR UTC/CNRS 7337 Roberval, CS 60319, 60203 Compiègne Cedex, France
[b] DeltaCAD, 795 Rue des Longues Raies, 60610 La Croix-Saint-Ouen, France
[c] Swiss Federal Institute of Technology at Lausanne (EPFL), STI-IPR-LICP, CH-1015 Lausanne, Switzerland

ARTICLE INFO

Keywords:
Deep learning
Machine learning
Computer vision
Product Lifecycle Management
Digital mock-up
Shape retrieval

ABSTRACT

With the increasing amount of available data, computing power and network speed for a decreasing cost, the manufacturing industry is facing an unprecedented amount of data to process, understand and exploit. Phenomena such as Big Data, the Internet-of-Things, Closed-Loop Product Lifecycle Management, and the advances of Smart Factories tend to produce humanly unmanageable quantities of data.

The paper approaches the aforesaid context by assuming that any data processing automation is not only desirable but rather necessary in order to prevent prohibitive data analytics costs. This study focuses on highlighting the major specificities of engineering data and the data-processing difficulties which are inherent to data coming from the manufacturing industry. The artificial intelligence field of research is able to provide methods and tools to address some of the identified issues. A special attention was paid to provide a literature review of the most recent (in 2017) applications, that could present a high potential for the manufacturing industry, in the fields of machine learning and deep learning.

In order to illustrate the proposed work, a case study was conducted on the challenging research question of object recognition in heterogeneous formats (3D models, photos and videos) with deep learning techniques. The DICE project – DMU Imagery Comparison Engine – is presented and has been completely open-sourced in order to encourage reuse and improvements of the proposed case-study. This project also leads to the development of an open-source research dataset of 2000 CAD Models, called DMU-Net available at: https://www.dmu-net.org.

1. Introduction

With the advent of phenomena such as *Cloud Computing*, *Extended Enterprise*, or *Smart Products and Manufacturing*, the manufacturing industry is facing an unprecedented increase in accessible and available data [1]. New opportunities, usages and needs arise from this observation such as Knowledge Discovery, Process Automation, Automated or Intelligent Control, Decision Support Systems, Recommendation Engine, Key Performance Indicators (KPI) Forecasting.

However, due to massive data scale, these opportunities all require to be, at least partly, automated. Recurring issues such as *massive amounts of data, high data-dimensionality* [2], *heterogeneous data aspects, and low data-quality* [3] greatly reduce, not only data integration and consumption, but also automation possibilities for statistical analysis.

Recent technological advances in terms of network throughput, data availability, computation capabilities and storage capacities allow companies and researchers to design algorithms and self-learning systems achieving industrial-grade performances. These recent advances in machine learning and deep learning offer the manufacturing industry tools and techniques reaching, at this date, some of the best performances in the current state-of-the-art to tackle the aforesaid recurring challenges.

This paper aims to:

- explore the specificities of the *engineering data* and the major challenges one may encounter when applying machine learning or deep learning in the manufacturing industry.
- relate a literature review of the existing applications in the fields of machine learning and deep learning that could be adapted to the manufacturing industry to address engineering issues.
- release and open to the research community the dataset (and possible future recurring research challenge) DMU-Net in order to insure results repeatability and statistical robustness: https://www.dmu-net.org. This dataset has also been designed to facilitate

* Corresponding author.
E-mail address: alexandre.durupt@utc.fr (A. Durupt).

https://doi.org/10.1016/j.compind.2018.04.005
Received 8 September 2017; Received in revised form 2 March 2018; Accepted 10 April 2018

research collaboration between researchers in Computer Science and Manufacturing. It should also help to disseminate achieved results and allow a clear understanding of the state-of-the-art by providing an unique scale to compare methods and models.
- detail a method illustrated with a case-study addressed with a deep learning software product focusing imagery data processing (from 3D to 2D data), highly contextualised in engineering, able to outperform the existing state-of-the-art tools and techniques in the manufacturing industry. The DMU-Net dataset constitutes the training dataset of this study. In order to facilitate the understanding of the proposed approach and fits into the open-science guidelines, every aspects of this study has been released as open source software.

The paper is structured as followed: In Section 2, a focus on the manufacturing computing context will be firstly put on. This will serve as a starting point for the proposed analysis and highlight the reasons to consider *artificial intelligence* to automate some of the engineering processes. Then a prospective study on *the current state-of-the-art on artificial intelligence* systems applicable in the manufacturing industry will be presented in Section 3. Finally, the case-study of this paper and its characteristics will be introduced in Section 4 to highlight how *deep learning* techniques could help addressing objectives of manufacturers. The applied method, commonly used in deep learning and computer vision research, will detailed in Section 5, the source code associated with the proposed case-study has been released in open-source to ease the understand and accelerate the implementation of similar software products and solutions which share common objectives.

2. Engineering and computing – massive aspects and research challenges

2.1. Available data and computer resources have a major impact on the industry

Computers took, over the years, an essential and central place in almost every engineering tasks. This trend has led the manufacturing industry to face an ever-challenging situation regarding computing issues. Data storage capacity, network traffic and computation capabilities, are all facing an exponential increase which encourages the manufacturing industry to rethink its habits and workflows manipulating engineering data at a large scale. Table 1 summarises the aforesaid situation.

Even if the accuracy of the numbers mentioned in Table 1 is questionable, there is no doubt that the industry is facing unprecedented amounts of data and network traffic all across the Information Systems (IS). It becomes increasingly challenging to handle these data on a daily basis, in such large quantities, with human capacities.

In order to deal with the present situation, process automation appears as a promising solution. This study does not only consider *automation* to be an efficient way to reduce costs and improve efficiency. *Automation* is also appraised as the only possible way to address massive scale engineering data challenges.

Thanks to recent advances in computation capabilities, it is conceivable to perform, almost in real time, complex calculations at an affordable price. Currently, artificial intelligence (AI) algorithms take full advantage of the newly available computation resources to solve complex problems that were previously out of reach.

2.2. Engineering data are mostly unstructured and heterogeneous – a challenging situation

This study focuses on any data used by domain experts and will be later referred as *Engineering Data*. These data come in very different forms and originates from various systems. This fact introduces a high degree of heterogeneity which tends to make any data usage difficult. Efficient techniques exist for number of applications using similar data, nevertheless data heterogeneity significantly impairs the possibilities for human-designed automated processes due to the broad range of engineering activities.

Two of the most common sources of data heterogeneity and their repercussions have been reviewed in the upcoming subsection:

2.2.1. Data heterogeneity – a recurrent limiting factor for big data engineering applications

2.2.1.1. Data type heterogeneity. Data type hterogeneity relates to the different types of file that one may find inside the IS.

A few examples could be: *Sketches, Drawings, Laser Scans, CAD Files, Technical Reports, etc.*

It becomes a major bottleneck in a Product Lifecycle Management (PLM) perspective when engineers need to compare the contents of files of heterogeneous (i.e. different) types but concerning the same *PLM part*.

A research question could be: *How to, effectively and automatically, compare, hundreds of files and documents that relate to a given assembly when they cannot be compared by "just" analysing the inner differences or their structure?*

2.2.1.2. Data format heterogeneity. Data format heterogeneity is inherent to the manufacturing industry. In the extended enterprise context, it is most likely that the different actors do not use the same CAD software or version. These software products mostly use proprietary formats, and this constitutes a major limitation to the automated analysis objective. Needless to say, that the *format heterogeneity* issue is more frequently problematic with engineering-specific data types. Data types, such as photo, music or video, also have inner format heterogeneity (e.g. photos: JPEG, PNG, GIF, TIFF, etc.), however they only differ in terms of performance and quality. Converting a photo from one format to another will not change the number of people present in a photograph or their relative height. Contrarily, with CAD files, it is almost certain that any conversion or translation will degrade the inner content and delete some information that might be essential for the engineers [8].

A research question could be: *How to, effectively and automatically, compare, hundreds of CAD files, in different formats, presenting different versions of the same part and extract the differences between them?*

2.2.1.3. Could STEP and JT help to address the aforesaid

Table 1
Major computing trends and their respective evolution.

Trend category	Predictor	Measured trend	Variation rate by year	Variation over 10 years
Storage		Price per GB	−22% (÷2 every 33 months)	÷12
	Kryder's law [4]	Storage capacity	+27% (2× every 36 months)	11×
Network	Cisco forecast [5]	Global IP traffic	+60% (2× every 18 months)	110×
	Nielsen's law [6]	Network speed	+50% (2× every 21 months)	57×
Computation		Price per GFLOP/s	−37% (÷2 every 18 months)	÷100
	Moore's law [7]	Computing power	+60% (2× every 18 months)	110×