Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 104 (2021) 975–980

www.elsevier.com/locate/procedia

54th CIRP Conference on Manufacturing Systems

Knowledge Discovery in Heterogeneous and Unstructured Data of Industry 4.0 Systems: Challenges and Approaches

Simon Kamm[a,*], Nasser Jazdi[a], Michael Weyrich[a]

[a]*Institute of Industrial Automation and Software Engineering, University of Stuttgart, Pfaffenwaldring 47, 70569 Stuttgart, Germany*

* Corresponding author. Tel.: +49 711 685 67293; fax: +49 711 685 67302. *E-mail address:* simon.kamm@ias.uni-stuttgart.de

Abstract

With the rise of the Internet of Things and Industry 4.0, the number of digital devices and their produced data increases tremendously. Due to the heterogeneity of devices, the generated data is mostly heterogeneous and unstructured. This challenges established approaches for knowledge discovery, which typically consume structured data from one source. The paper first describes aspects of data heterogeneity and their relevance for Industry 4.0 systems. Following, the upcoming challenges for different steps inside the knowledge discovery process for Industry 4.0 systems, such as for data integration and data mining, are discussed. Additionally, it mentions approaches to tackle them.

Keywords: : Heterogeneous Data, Industry 4.0, Internet of Things, Knowledge Discovery, Machine Learning, Semantic Data Integration, Unstructured Data

1. Introduction

The Internet of Things (IoT) and Industry 4.0 (I4.0) are already well-suited terms in the field of industrial automation. Many concepts are already transferred from academia to industrial usage. One big evolution coming with these concepts is connected and digital devices, e.g., in the form of cyber-physical production systems (CPPS) [2], which can produce a tremendous amount of data. Therefore, the “Big Data” era arises, which brought many new opportunities and business models, as well as new challenges for data exchange and management [3]. One open challenge is to ensure a standardized information exchange in IoT and I4.0 applications [4]. On the one hand, interoperability between different devices is required. Information about, e.g. a production process shall be shared between different machines or even different companies along a process chain. On the other hand, this data and information shall be preserved in a data storage (e.g., database or data warehouse). This stored data and information shall be used for data mining and knowledge generation based on the available data and information, following the DIKW-pyramid (**D**ata **I**nformation **K**nowledge **W**isdom) [1] as shown in Fig. 1. Artificial Intelligence (AI) is considered a key technology for data processing and shows a great deal of potential for benefits [5]. Throughout the domain of AI, Machine Learning (ML) methods and techniques are one of the main drivers; especially Deep Learning-based approaches show outstanding results for specific tasks, e.g. in image recognition [6]. These approaches need a huge amount of well-structured and homogeneous data for training a

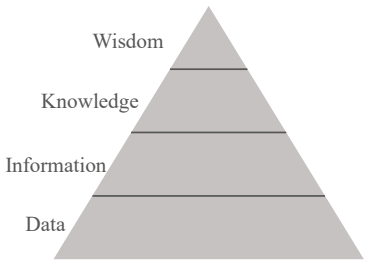

Fig. 1. DIKW pyramid [1]

10.1016/j.procir.2021.11.164

model. However, in industrial automation applications, a huge amount of data is often unstructured and heterogeneous coming from multiple sources [7]. Unstructured data has no concrete data schema for data storage, e.g., a distributed data storage via folder systems. Heterogeneity from the data has several reasons, which will be discussed in this paper. Although the data is unstructured and heterogeneous, much data is available and shall be used to generate knowledge with the help of data mining methods such as machine learning. To achieve this, the data has to be integrated in a best-structured way, to access this data easily. This paper aims at summarizing the challenges for the knowledge discovery in heterogeneous and unstructured data coming from the literature as well as showing existing approaches to tackle those since the challenges are increasing with more data being generated.

*Objectives and paper outline: C*hapter two shows the basics of Knowledge Discovery for the field of industrial automation and the occurring data with its properties are discussed with a focus on the variety. The fields of semantic data integration and machine learning with related work are shortly introduced. In chapter three, we derivate challenges for the data flow along the knowledge discovery process. We discuss different approaches for discovering knowledge in heterogeneous and unstructured data in chapter four. In the final chapter five, future possible research directions are highlighted.

2. Basics and Related Work

2.1. Knowledge Discovery in Databases

Knowledge Discovery in Databases (KDD) is a nontrivial process to find knowledge in existing data. In [8], the process was first defined to give a better understanding of the different approaches in the field of Knowledge Discovery and how those fit together in this multidisciplinary field. The KDD process model tries to model all relevant steps, from accessing and selecting the data up to data mining. Data mining is often used interchangeably with KDD, but it is just a single step within the overall KDD process. There are different models for knowledge discovery in databases, such as the KDD model [8] or the CRISP-DM model (Cross Industry Standard Process for Data Mining) [9]. In this work, we refer to the KDD model from Fayyad et al. [8] as it is widely used in practice. The model with its different steps is shown in Fig. 2. To discover knowledge, data first has to be selected, pre-processed and transformed, before data mining can be performed. The extracted patterns can then be interpreted and evaluated to finally discover new knowledge.

2.2. Heterogeneous and Unstructured Data

Big Data has been referred to as a revolution, which transforms many industries. The main purpose is to extract knowledge and value out of the big amount of data and enable better decision-making. Big Data itself was earlier defined by three Vs (volume, velocity, and variety) [7, 10]. However, a more commonly accepted definition nowadays is using four Vs (volume, velocity, variety, and veracity) [11, 12]. Each V itself brings unique challenges to the different steps of the knowledge discovery process. In this paper, we will focus on challenges for heterogeneous and unstructured data, which fits the Big Data property variety. Variety is seen to be more relevant than the pure volume of the data [7]. In the context of Big Data, variety describes two kinds of variations. One is the structural or syntactical variation of a dataset and of the concrete data types, which occurs when two data sources are not expressed equally. The other one is the variation of how the data is represented or the semantic variation, and thus how the data has to be semantically interpreted; this is sometimes called conceptual heterogeneity [11, 13]. In literature, more types of data heterogeneity can be found, which will be introduced in the following.

Heterogeneity from a statistical point of view means different statistical properties across a dataset. A common data mining assumption is that statistical properties are similar or equal across a dataset. The statistical heterogeneity even enlarges with a growing amount of data, as given in Big Data [12]. Further, data can be terminological heterogeneous, when names for the same entities vary from different sources (e.g., sensors integrated into different PLC (Programmable Logic Controller) programs). In addition, data and entities which are different interpreted by people are called semiotic heterogeneous. In some sources, this property is named pragmatic heterogeneity [11, 13]. In summary, the paper introduces five kinds of data heterogeneity. In the following, these aspects of heterogeneity are named, reviewed and rated based on their relevance for applications in the domain of industrial automation.

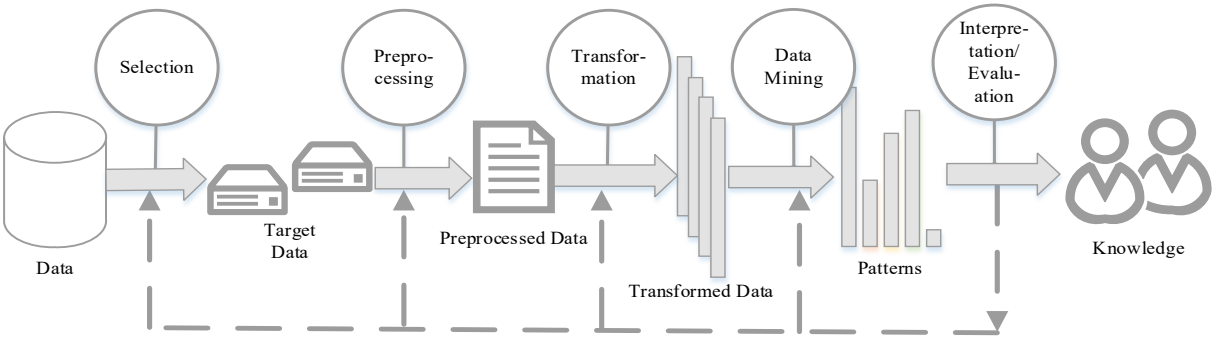

Fig. 2. Knowledge Discovery in Databases (KDD) Process [9]

Syntactical or structural heterogeneity exists in nearly every use case, where different data sources are present (e.g., multiple sensors from different companies). This heterogeneity is in some cases unavoidable and desired when e.g., different sensors as an optical or acoustic sensor are used to get different views on the same entity since they deliver their data most likely in a different syntax [12]. Nevertheless, it also can be undesired, if for example, no unified communication standard between similar devices (such as the communication protocol of a CAN-Bus) exists, the syntactical and structural representation of the data differ [4]. This leads to different data types, although the same kind of data is recorded. Data fields can e.g. have different ranges for the same kind of data (e.g., a sensor measures a distance in meter, while another one measures it in miles). By knowing the data and its sources, the data can and has to be translated to a unified data type, unit, and range in this case. This is an essential part of every data integration and analysis task, and therefore is relevant in the context of industrial automation, too.

Semantical or conceptual heterogeneity is the difference in representing the same domain of interest [11, 13]. This difference can further be separated. Relevant for the domain of industrial automation are differences in coverage, which occurs when two data sources describe different regions of the real world at the same level of detail and from a unique perspective. For example, equal sensors are mounted on different places of a machine with a different coverage region. Another relevant heterogeneity factor is the difference in perspective (or difference in scope). Different data sources have different perspectives while modeling the same region of the real world at the same level of detail, e.g., a pressure and a temperature sensor, which cover the same region of a machine and measure in the same time interval. Lastly, semantic heterogeneous data can have a difference in meaning and interpretation of data values. This often occurs, when different and independent parties developed a data schema for the same domain. These aspects of differences have to be considered for industrial automation systems to give the available data the correct meaning in the further process of knowledge discovery.

Statistical heterogeneity is a challenge for all data processing tasks in real-world applications, as real datasets usually are not perfectly and equally distributed as expected in theory. Therefore, this is a general challenge, which has to be considered in every knowledge discovery approach. For data integration and information exchange, this heterogeneity aspect is not seen to be relevant. However, for data mining techniques such as machine learning, this heterogeneity type can be the most relevant aspect, as model training is directly affected by this [12]. An example is a dataset for a classification module, where "good" data without any failure can easily be recorded during run time. Since products with failures are necessary for "bad" data, this data is expensive to produce. As a result, often statistical heterogeneous and imbalanced datasets exist.

Terminological heterogeneity stands for the variation of names in different data sources when the same entities are referred [11, 13]. This can exist in reality for industrial automation systems, but may easily be resolved by renaming and consistency checking of the names when the same entities are modeled in different systems. Due to this resolvability, this aspect is not considered further in the paper, although it probably occurs frequently.

Semiotic or pragmatic heterogeneity occurs when different interpretations of an entity by people exist. It depends on the interpretation of humans and is hard to detect or resolve for a computer [11, 13].

Based on the more detailed definition and classification of the heterogeneity aspects above, we argue that for industrial automation applications, syntactical and semantical data heterogeneity are the main challenges to resolve. For these two aspects, the concrete challenges for data integration and data mining need to be derived further.

2.3. Semantic Data Integration

For every data analytics task, the relevant data first has to be integrated and made available for further processing. Following the KDD process, several steps have to be performed before new patterns can be detected in the data. These tasks are the selection, preprocessing, and transformation of data (see Fig. 2). They often take a huge amount of the overall work in the KDD process. Following [14], solely the preparation step can require up to 45% of the total effort in a KDD process. When having multiple data sources with, e.g., syntactical and semantical heterogeneous data, the process is becoming even more challenging, since the selection, preprocessing and transformation have often to be handled separately. Nevertheless, we argue that the goal is to perform one data mining step based on all available data coming from different sources, when this data affects the observed behavior (e.g., failure of a machine). One approach to handle data integration for heterogeneous data is semantic data integration, which aims to combine data from different sources and consolidate the available data into meaningful and valuable information by using semantic technologies [15]. To finally perform data mining on heterogeneous and unstructured data, first, the before-mentioned parts of the KDD process have to be considered and processed in applications. Semantic data integration shall enable the import and transformation of heterogeneous data from multiple sources. Different approaches in the context of industrial automation exist that try to resolve semantical and syntactical conflict with the help of ontologies. In the following, some of them are shortly introduced. These approaches try to tackle the topic of semantic integration for the industrial automation domain in a general way and are viewed as a good starting point for future works:

Semantic Sensor Network (SSN) Ontology was defined and developed by the W3C (World Wide Web Consortium), which targets to describe the properties of sensors and their capabilities as well as the resulting observations [16]. The data shall be available in a machine-readable and interpretable form to allow autonomous or semi-autonomous data collection, processing, and acting on sensors and their observations. This ontology enables sensing data and resolving semantic heterogeneity already in the data acquisition phase.

SAREF-Ontology (Smart Appliances Reference) Ontology [17] is a model standardized by the ETSI (European Telecommunications Standards Institute) committee Smart Machine-to-Machine communications (SmartM2M) to interconnect data and enable communication across different protocols and standards for IoT devices in the area of smart applications. The focus lies there on home and building sensors. To enroll this standard further for other application domains, the SAREF ontology is extended. SAREF4INMA is the resulting extension for the industry and manufacturing domain [18] to align with related initiatives in the domain, such as the Reference Architecture Model for Industry 4.0 (RAMI). SAREF4INMA shall enable interoperability between various types of production equipment and between organizations along the value chain.

The **Optique Project** [19] worked on ontology-based data access (OBDA) technology to build up a semantic end-to-end connection between different data sources and a user. In this project, the main target was to integrate heterogeneous and distributed data sources and make the vast amount of Big Data accessible for a user.

Modoni et al. [20] developed a methodological approach for supporting new semantic model development with the focus on reusing existing data models (e.g. as introduced above) and the semantic conversion of legacy models.

The above-mentioned approaches aim to resolve the semantic heterogeneity of the data with semantic technologies. To resolve the syntactical heterogeneities, different approaches exist, such as SensorML [21].

With approaches as listed above, the heterogeneous data coming from multiple data sources of industrial automation systems can be accessed, preprocessed, transformed, and stored according to syntactical and semantic rules. Based on this, data mining techniques can explore and detect patterns in this heterogeneous data.

2.4. Machine Learning

Machine learning, one of the central sub-areas of AI, was stated as one of the main drivers for innovation [22] and is one of the most often used data mining techniques. Machine learning in general is a learning-based or data-driven approach to design a processing rule. This processing rule is learned based on examples [23]. For conventional ML classifiers, features are extracted based on defined rules, e.g., from a domain expert. With the extracted features, the following desired task (e.g., classification) is performed by statistical methods. In Deep Learning methods, just one network (a Deep Neural Network, short DNN) exists for the task of feature extraction and, e.g., classification. The DNN consists of multiple layers building a "deep" model structure. The model learns and adapts the network parameters by a proper loss function while the training phase. Deep Learning is a special type of ML [24]. The domain of ML is traditionally divided into three parts: Supervised Learning, Unsupervised Learning, and Reinforcement Learning [22]. In some sources, semi-supervised learning, which is a mixture of supervised and unsupervised learning, is added as the fourth part of it. Depending on the data and label quality, availability, and application, a suitable method has to be chosen. For supervised learning, every data needs a corresponding label, as given for classification tasks or regression tasks (e.g., wear-out prognosis). In unsupervised learning, no labels are given. Typical use cases are clustering or dimensionality reduction. In reinforcement learning, a reinforcement learning agent interacts with the environment and performs actions. Based on the feedback from the environment (rewards) the agent is reinforced to a specific behavior. As a result, the agent learns a task by trial and error. Robotic navigation and skill acquisition use methods from this domain. This field of ML has drawn a lot of attention in the manufacturing field in the last years, e.g. to develop smart control agents for control processes, such as for the adaptive control of a laser welding process [25].

The performance of an ML algorithm depends heavily on the representation of the data they consume [24]. The models are usually domain-specific and trained in a pre-defined environment. One major challenge in ML is how to deal with heterogeneous data [12], which will be discussed in the next chapter in more detail. In this case, often separate and even manual processing of data is performed on each type of data, thus losing the benefit of data variety, which can enable a better overall performance, when the data is used in a holistic approach.

3. Challenges for Knowledge Discovery in heterogeneous data for Industrial Automation Systems

When facing heterogeneous and unstructured data, the existing knowledge discovery approaches are limited and often cannot handle the variety of the data suitably to gain the expected benefit of it. Therefore, different approaches are necessary, which shall be able to handle the challenges in a KDD process for heterogeneous data, especially for industrial automation systems. The below-discussed challenges are not complete but shall give an overview of the most challenging questions when facing heterogeneous and unstructured data.

How to integrate, store and describe data is a challenge naturally arising when heterogeneous data sources exist. An easy and naive approach would be to assume that standardized data formats and data descriptions can be used. This would naturally resolve the most problems of semantical and syntactical heterogeneity. Although most parties in a domain probably agree with this, the development of a common and standardized data and communication format typically takes a long time [4], so this approach will not resolve the currently faced challenges. Due to the lack of one common data language, multiple data protocols and semantics exist, e.g., different companies use different data formats, communication protocols or have a different semantic meaning for the same variable name [12]. This leads in the end to a big variety of data protocols, which require flexible interfaces instead of predefined and fix database interfaces. Integrating data coming from multiple sources and adapt to new sources, is a requirement [26]. Once the data is integrated, the data has to be stored in an appropriate way, which allows fast data access, extendibility, and changeability of the storage for new features (e.g., when a new sensor is installed or

changed). In addition, the relations within the data shall be preserved. Classical relational database systems are challenged by the increasing heterogeneity and can often not handle the big amount of relations in the data [27]. As a result, data from one or multiple similar sources are often held in separate so-called data silos. Each silo typically has a (slightly) different semantic. When data from another data silo is needed, ETL (Extract, Transform, Load) processes are mandatory to translate the data from the data schema and semantic of one silo to another silo. Naturally, it is desired to have a common description and storage of the data, which makes the data available, easily accessible, and interchangeable.

After the data is integrated and stored, the task is to discover patterns in the data, which in the end deliver additional value. For this purpose, the available unstructured and heterogeneous data shall be consumable for the data mining methods (e.g., DNNs). For typical ML methods, this sort of data is not usable directly [28]. Without any data normalization or transformation, the models will most probably not converge. With multiple heterogeneous data sources, the effort for data normalization and transformation is increasing tremendously, when e.g., every source has to be normalized separately or even manually by analyzing the data and its properties. The resulting heterogeneous features coming from the diverse data sources shall be used in combination to benefit from the full data variety. In a classical ML approach, different models are used for different data sources and types (e.g., for image data, time-series data) and the extracted knowledge can be used further. In this approach, the different data sources would not be combined by the data-driven model [29]. Following this, classical ML methods will not use the advantage of the Big Data variety. The challenge is to combine the heterogeneous data in a machine-readable form to enable "end-to-end" learning for a data-driven learning approach instead of multi-step learning. As the last challenge for industrial automation systems, we see the availability and usage of expert knowledge and analytical models. Most processes exist already for a long time and experts for them are present in the companies. Moreover, in some scenarios, additional analytical models are available, which can describe the system behavior at least up to a specific accuracy. This knowledge shall be used for knowledge discovery in heterogeneous and unstructured data, too. However, it is an open challenge how expert knowledge and analytical models can be integrated efficiently into the data-driven data mining step.

Overall, different challenges exist in the knowledge discovery process for heterogeneous and unstructured data. In this chapter we introduced four challenges, formulated as open questions:

1. How to integrate the heterogeneous data?
2. How to describe and store the data uniformly and standardized?
3. How to combine heterogeneous data to enable "end-to-end" learning?
4. How to integrate available expert knowledge and analytical models?

In the following chapter, some existing approaches that try to answer the questions are named and discussed.

4. Approaches and Discussion

As already introduced previously, there exist ***semantic approaches***, which handle data integration and modeling through semantic data integration. Data integration and transformation can be performed with the help of *ontologies* [15, 30, 31]. An approach for additionally tackling the data storage challenges can be *Knowledge Graphs*. As in [32], they use semantic knowledge bases to resolve semantic heterogeneity and model the data with the help of ontologies. Following [28], the Knowledge Graph in combination with new algorithms, such as Graph Analytics or Graph Convolutional Neural Networks [33], can enable "end-to-end" learning of a model from the available heterogeneous data. This would also tackle the third mentioned challenge.

We additionally propose two approaches based on known ***ML methods***, which we call *feature extraction approach* and *ensemble learning*. In the *feature extraction approach*, different algorithms are trained in the first stage to extract meaningful features and reduce the dimensionality of the data. The training of these feature extractors is performed separately per data type (e.g., one extractor for images, one for time-series data, etc.). Thereafter, the extracted features are fed into a second stage ML model, which produces the final output (e.g., class label or regression output). For this approach, a two-staged learning process is required [29]. In [34], a *feature extraction approach* is realized with two image sources for a quality diagnosis platform for laser-based manufacturing processes. When one feature extraction algorithm has to be changed, the stage-2 algorithm has to be trained again. This can further be improved by Transfer Learning paradigms. The stage-2 algorithm shall solely adapt to the changed input data instead of retraining on all data, which is a typical scenario for Transfer Learning [35]. The second approach is *ensemble learning*, where different models are separately trained to perform the desired task (e.g., classification or regression). These outputs are then combined to the final output (e.g., class label) [36]. This combination can be a simple majority voting or some sort of weighting. Nevertheless, none of these three approaches (*Knowledge Graph, Feature Extraction, and Ensemble Learning*) integrates available knowledge or analytical models. Thus, hybrid ML models are required, which combine analytical with data-driven models [37], but they do not fulfill further challenges. In the end, we see currently no approach, which can fulfill all mentioned challenges. Therefore, we argue that a new concept is mandatory, which shall contain the following functionalities:

- Semantic data integration and modeling
- Data storage while preserving the relation of data and the underlying data model with fast data access
- Performing data mining using the available data variety and enabling "end-to-end" learning
- Possibility to include and use available expert knowledge and analytical models

5. Conclusion and Future Work

Overall, the amount of unstructured and heterogeneous data is increasing drastically over the past year. This growth will

probably continue. To use this Big Data and especially the variety coming with this data in the context of industrial automation systems, different challenges have to be resolved. This paper shows that the widely known KDD process and the methods used in it need adaptions to these challenges. The heterogeneous data existing in the industrial automation domain has different aspects, as semantic and syntactical heterogeneity. The two aspects are detected as the most relevant ones in this field. These kinds of heterogeneity have to be resolved to enable the data mining step and discovering patterns in the underlying data. Therefore, we argue that semantic approaches are necessary to describe the data in a common language and integrate the data coming from different sources. Machine learning as a frequently used technique for data mining has to be adapted to the specific characteristic of the data. Four concrete challenges were derived for the knowledge discovery process for heterogeneous and unstructured data with a focus on industrial automation systems. We show three different approaches (Knowledge Graphs with Graph Analytics, Feature Extraction, and Ensemble Learning) as proposals for heterogeneous data mining. Anyway, none of the approaches can resolve all challenges on his own. We, therefore, introduced necessary functionalities for a new concept to resolve them. A further concretization of these functionalities and a detailed comparison of existing approaches will be part of the upcoming research. The definition of a first concrete concept and its components is planned as future work, too. The focus of the future work lies in industrial automation systems with specific environmental conditions, such as timing requirements or distributed partial processes.

References

[1] Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science 2007; 2:163–80.

[2] Müller T et al. Cyber-Physical Production Systems: enhancement with a self-organized reconfiguration management. Procedia CIRP 2020.

[3] Jagadish H et al. Big data and its technical challenges. Commun. ACM 2014; 7:86–94.

[4] Plattform Industrie 4.0. Details of the Asset Administration Shell. Part 1 – The exchange of information between partners in the value chain of Industrie 4.0. [Online] Available: https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/2018-verwaltungsschale-im-detail.html. Accessed on: Dec. 10 2020.

[5] Lindemann B, Jazdi N, Weyrich M. Adaptive Quality Control for discrete large-scale Manufacturing Systems subjected to Disturbances. 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)2020; 379–86.

[6] Maschler B et al. Distributed Cooperative Deep Transfer Learning for Industrial Image Recognition. Procedia CIRP 2020; 437–42.

[7] Marek Obitko, Václav Jirkovský, and Jan Bezdíček. Big Data Challenges in Industrial Automation. In Industrial Applications of Holonic and Multi-Agent Systems. 305–16.

[8] Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth. From Data Mining to Knowledge Discovery in Databases. AI Magazine 1996; 3:37–54.

[9] Wirth R, Hipp J. CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. 29–39.

[10] McAfee A et al. BigData: The Management Revolution. Harvard business review 2012; 60–8.

[11] Wang L. Heterogeneous Data and Big Data Analytics. Automatic Control and Information Sciences 2017; 1:8–15.

[12] L'Heureux A et al. Machine Learning With Big Data: Challenges and Approaches. IEEE Access 2017; 7776–97.

[13] Jirkovský V, Obitko M. Semantic Heterogeneity Reduction for Big Data in Industrial Automation. In ITAT 2014.

[14] Cios K et al. Data Mining: *A Knowledge Discovery Approach*. Boston, MA: Springer Science+Business Media LLC; 2007.

[15] Giacomo G de et al. Using Ontologies for Semantic Data Integration. In *Studies in big dataA comprehensive guide through the Italian database research over the last 25 years.* Springer; 2018; 187–202.

[16] Compton M et al. The SSN ontology of the W3C semantic sensor network incubator group. Journal of Web Semantics 2012; 25–32.

[17] Daniele L, den Hartog F, Roes J. Created in Close Interaction with the Industry: The Smart Appliances REFerence (SAREF) Ontology. In Formal Ontologies Meet Industry. Springer; 2015; 100–12.

[18] Roode M de et al. SAREF4INMA: a SAREF extension for the Industry and Manufacturing domain. Semantic Web 2020; 1–16.

[19] Giese M et al. Optique: Zooming in on Big Data. Computer 2015; 3:60–7.

[20] Modoni G et al. Enhancing factory data integration through the development of an ontology: from the reference models reuse to the semantic conversion of the legacy models. International Journal of Computer Integrated Manufacturing 2017; 10:1043–59.

[21] Botts M, Robin A. OpenGIS® Sensor Model Language (SensorML) Implementation Specification2007.

[22] Bundesministerium für Wirtschaft und Energie (BMWi) Indutrie 4.0. Technology Scenario Artificial Intelligence in Industrie 4.02019.

[23] Bishop C. Pattern recognition and machine learning. New York: Springer; 2006.

[24] Goodfellow I, Bengio, Yoshua, Courville, Aaron. Deep Learning: MIT Press; 2016.

[25] Masinelli Giulio et al. Adaptive Laser Welding Control: A Reinforcement Learning Approach. IEEE Access 2020; 103803–14.

[26] Faul A, Jazdi N, Weyrich M. Approach to interconnect existing industrial automation systems with the Industrial Internet. IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)2016 - 2016;

[27] Yoon B-H, Kim S-K, Kim S-Y. Use of Graph Database for the Integration of Heterogeneous Biological Data. Genomics & informatics 2017; 1:19–27.

[28] Wilcke X, Bloem P, Boer V de. The Knowledge Graph as the Default Data Model for Machine Learning2017.

[29] Damoulas T, Girolami M. Combining feature spaces for classification. Pattern Recognition 2009; 11:2671–83.

[30] Sabou M, Ekaputra F, Biffl S. Semantic Web Technologies for Data Integration in Multi-Disciplinary Engineering. In Multi-Disciplinary Engineering for Cyber-Physical Production Systems. Springer International Publishing; 2017; 301–29.

[31] Kovalenko O, Euzenat J. Semantic Matching of Engineering Data Structures. In Semantic Web Technologies for Intelligent Engineering Applications. Cham: Springer International Publishing; 2016; 137–57.

[32] Grangel-González I et al. Knowledge Graphs for Semantically Integrating Cyber-Physical Systems. International Conference on Database and Expert Systems Applications 2018; 184–99.

[33] Zhang Z, Cui P, Zhu W. Deep Learning on Graphs: A Survey. IEEE Transactions on Knowledge and Data Engineering 2020.

[34] Stavropoulos P et al. A three-stage quality diagnosis platform for laser-based manufacturing processes. The International Journal of Advanced Manufacturing Technology 2020; 11:2991–3003.

[35] Maschler B, Weyrich M. Deep Transfer Learning for Industrial Automation: A Review and Discussion of New Techniques for Data-Driven Machine Learning2020.

[36] Sagi O, Rokach L. Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2018; 4:

[37] Rogers T et al. On a Grey Box Modelling Framework for Nonlinear System Identification. In Special Topics in Structural Dynamics, Volume 6. Springer International Publishing; 2017; 167–78.