This is an Accepted Manuscript of an article published by Elsevier in Journal of Manufacturing Systems on 5 MAR 2018, available online: http://doi.org/10.1016/j.jmsy.2018.02.004.

A Survey of the Advancing Use and Development of Machine Learning in Smart Manufacturing

Michael Sharp[a], Ronay Ak[a], Thomas Hedberg, Jr.[a,*]

[a] *Engineering Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Stop 8260, Gaithersburg, MD 20899 USA*

Abstract

Machine learning (ML) (a subset of artificial intelligence that focuses on autonomous computer knowledge gain) is actively being used across many domains, such as entertainment, commerce, and increasingly in industrial settings. The wide applicability and low barriers for development of these algorithms are allowing for innovations, once thought unattainable, to be realized in an ever more digital world. As these innovations continue across industries, the manufacturing industry has also begun to gain benefits. With the current push for Smart Manufacturing and Industrie 4.0, ML for manufacturing is experiencing unprecedented levels of interest; but how much is industry actually using these highly-publicized techniques? This paper sorts through a decade of manufacturing publications to quantify the amount of effort being put towards advancing ML in manufacturing. This work identifies both prominent areas of ML use, and popular algorithms. This also allows us to highlight any gaps, or areas where ML could play a vital role. To maximize the search space utilization of this investigation, ML based Natural Language Processing (NLP) techniques were employed to rapidly sort through a vast corpus of engineering documents to identify key areas of research and application, as well as uncover documents most pertinent to this survey. The salient outcome of this research is the presentation of current focus areas and gaps in ML applications to the manufacturing industry, with particular emphasis on cross domain knowledge utilization. A full detailing of methods and findings is presented.

*Corresponding Author. Tel: +1 3019754247
Email address: `thomas.hedberg@nist.gov` (Thomas Hedberg, Jr.)

April 10, 2018

Keywords: Machine Learning, Industrie 4.0, Smart Manufacturing

1. Introduction

Machine learning (ML) has seen increased usage in manufacturing over the past 20 years. Two surges in the use of ML occurred in manufacturing; the first in the 1980s, with the second currently underway. While ML saw significant attention in the 1980s, industrial adoption was not high because the methods were difficult to implement and ahead of the technology available at the time [1, 2]. Many companies and researchers in industry are revisiting past work, focusing primarily on domain-specific models. We postulate there has been very little focus on cross-domain models for connecting information across the product life cycle. ML has remained "siloed" in each phase of the product life cycle: conception, design, manufacture, quality, and sustainment. With increased adoption of the Industrial Internet of Things (IIoT), Industrie 4.0, and Smart Manufacturing, even more data is being generated. Therefore, how does one effectively and efficiently take advantage of all that data?

Applications such as Total Design theory [3], Design for Six Sigma [4], and Design for Manufacturing [5, 6] require knowledge of the various phases of the product life cycle. In a sampling of 35 defense-acquisition programs [7], development-cost growth averaged 57 percent and procurement-cost growth averaged 75 percent. Decision making dominated both types of cost growth. It follows that mitigating the negative effects of decisions earlier in the lifecycle could be advantageous to both the cost and the quality of a production program. Such mitigation requires knowledge of the full lifecycle and an understanding of how a decision in one phase of the lifecycle affects other phases of the lifecycle.

How does one gain such knowledge? Hedberg Jr et al. [8] proposed three research directions to enable using knowledge earlier in the product life cycle: (1) dynamically generate knowledge bases, (2) determine minimum information requirements, and (3) data-interoperability support. ML are poised to greatly assist with the first two of these, dynamic knowledge generation and minimum information requirements. Synthesizing the work of Hedberg Jr et al. [8] with other literature [9, 10, 11, 12, 13] identifies a need for automated methods to collect, transmit, analyze, and act on the most appropriate data. This sets the goal of using ML tools that can "observe" data, apply

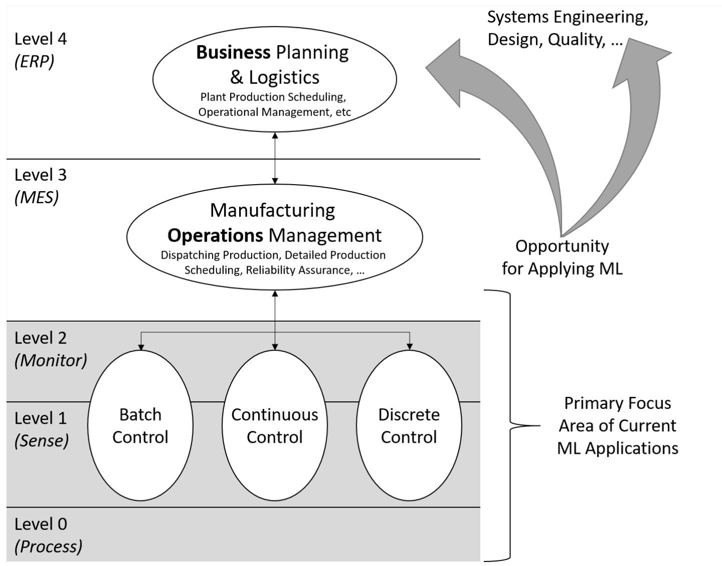

Figure 1: Hypothesized current application areas and opportunities for applying machine learning in manufacturing and beyond (adapted from [19])

context, and generate knowledge – these tools must be cross-domain (cross-phase) observatories.

This paper provides a literature survey on the application of ML to mutlidisciplinary, cross-domain focus areas that make up the product life cycle using manufacturing data in support of developing a life-cycle-wide "data observatory" [14]. The motivation of this work is to survey and enable the integration of previous domain-specific works, such as those described by Jennings et al. [15], Li et al. [16], Wang et al. [17], into the systems and enterprise level of the life cycle. For example, Figure 1 presents the scope and hierarchy of the ISA-95 [18] framework and identified hypothesized current application areas and opportunities for applying ML in manufacturing and beyond. The survey was conducted with the hypothesis that most applications of machine learning are applied to low-level manufacturing problems (ISA-95 Level 0, 1, 2) and little to no application of machine learning has been applied to systems level (ISA-95 Level 3), enterprise level (ISA-95 Level 4), and other phases / domains of the product lifecycle (e.g., systems engineering, design, quality).

To accomplish extending the application of ML to cross-domain focus areas, the gaps (e.g., what questions remain) must be identified so that they may be closed through research and development. Also, to ensure successful adoption of ML solutions, the real-life applications that exist and their benefits must be determined. To accomplish this, the literature survey was conducted with two aims. i) investigate the current state-of-the-art for ML methods; and ii) investigate any cross-domain applications of ML in the product lifecycle.

Three survey questions were asked:

- What types of algorithms are used and with what frequency?
- Are certain applications of machine learning frequently occurring? If so, which applications and at what level of the manufacturing systems?
- Is further research needed to capture the opportunities of applying machine learning to cross-domain focus areas of the product life cycle? If so, where and what?

The scope of the survey was limited to the integration of the design, fabrication, and quality domains / functions of the product life cycle. Sustainment, or customer and product support, was considered out of scope for this survey. While sustainment is important, the initial focus was on knowledge development to support design and manufacturing decisions.

2. Motivation and Background

The goal of this paper is to estimate the level of interest and actual effort being put towards the incorporation of ML technologies to the modern manufacturing industry by quantifying the presence of these concepts in the current literature. Further, this work seeks to ascertain prominent areas of the use of these technologies with both general and specific examples of applications in the literature by isolating sub populations of coordinated literature as well as targeting specific works on the subject. Last, any relevant gaps in the current level of deployment or development will be identified and presented as areas of future research.

With the availability of digital publications, it is now possible via automated techniques [20, 21] to search a wider breadth and depth of literature within an area than is feasible with manual methods. Search engines and online repositories of technical documents can quickly provide a host of information based on queries of a few simple phrases. However, these searches are mostly word matching techniques and do not match the underlying concept or contextual content to a document. Most often in the English language, a collection of words or phrases presented in a particular order is required to convey a concept or idea instead of any single word. Thus, analyzing collections of words is the basis of many forms of linguistic analysis, and as related to this methodology, is what partially drove the motivation to move beyond simple key word matching as a basis for document comparison. Towards that end, key NLP techniques were identified and applied to a large corpus of technical publication abstracts in addition to simple word matching analytics. These techniques included Bag of Words/Features, and Latent Semantic Analysis (LSA) to develop a measure of 'similarity' between the documents and concepts to identify key trends [22]. This places more emphasis on not just simple word matching as with traditional searches, but core concept matching.

In this work, a large collection of digital technical abstracts is mined via ML and NLP techniques to better understand emerging trends within both industry and academia. This base corpus is created using the word matching techniques native to many online repositories, and will be the space that all the automated information mining techniques will be applied to. Beyond automated techniques, several key papers were manually identified for a more in-depth review based on the concept searching criteria developed in this paper. For this more in-depth review, both knowledge assimilation and dissemination relies on traditional human efforts. A more complete description of the development and use of the base corpus is presented in the following section. The results from both the automated and manual information gathering efforts are presented below.

3. Methodology

A corpus of technical documents was reviewed using computer aided searching and NLP methodologies to assess AI and ML applications in manufacturing. A list of over 4000 unique articles pulled from a variety of digital resources (Primarily Engineering Village and Google Scholar) formed the search space and basis for the automated techniques used in this survey. The publication date of articles within is limited between 2005 and 2017. This section outlines the digitally assisted methodology for gathering these articles, identifying those most focused to the interests of this survey, and the general assessment of key concepts throughout the survey search space.

Table 1: Initial keyword search terms

Contains	Does NOT Contain
Manufacture	Food
Machine Learning	Bio
Artificial Intelligence	Social Media
Quality Control	
Inspection	
Manufacturing Design	

3.1. Initial Construction / Key Word Search

The initial creation of the article list began with the largely common place key term search capability familiar to many document repositories. Various combinations of the terms shown in Table 1 identified potential articles that could be of interest. Only those terms listed in the table are used to populate the base corpus. Note that some words were used as exclusion parameters, listed under the 'Does NOT Contain' section of the table. These terms were selected to create a broad scope of papers that encompasses the primary focus of this paper.

This resulted in over 4000 articles that met the key word qualifications. It would be unpractical to manually sort through each article and rate it on pertinence to the subject matter of interest. This effort would be particularly wasteful noting that a tagged document may simply contain one or more of the searched phrases, but not actually deal with the target subject matter. To overcome such difficulties, an automated NLP driven approach was taken to characterize the corpus and expedite identification of those article that are of the most interest to ML in manufacturing.

3.2. Corpus Characterization

In this survey, document characterization is performed in two broad categories: total search space characterization, and relative similarity to a focus subject document. Both analyses take the consolidated vector of semantic features and their associated values from the LSA algorithm and compare them to other established vectors using the cosign distance metric. In the case of the search space characterization, this comparison is made between each distinct article, whereas for the focused search the comparison is made to a single predefined key or “prime vector”. A semantic feature of a document can be thought of as a combined magnitude of related words and phrases.

To prepare the corpus for analysis, each document was cleaned by removing punctuation and trivial “stop words” (e.g., ’also’, ’just’, ’that’, etc.) [23]. Token phrases were constructed to have between one and three words, using the N gram approach, parsing each into a document/frequency matrix. These were then condensed into a lower dimensional feature space with Term Frequency – Inverse Document Frequency (TF-IDF) weighting and LSA. This helps to ensure that words and phrases used with similar context contribute similarly to the “semantic location” or position in the feature space. Combining elemental ideas or similar word concepts is a convenient way to think of this process in regard to this work.

Once the document feature set is established for each article, it is a simple matter to calculate the angle between two documents to determine their similarity. It is important to note that when calculating the similarity between documents for the total search space characterization, feature phrases that appear in one document but not another are assumed as zeros for comparative calculations. However, in the case of the prime document comparison (a directed subject search), any feature phrases not in the prime document are simply omitted from the cosign similarity calculation. This is to help place more focus on the prime documents topics, and mitigate exclusion of searched documents that may contain additional topics other than the topics of interest.

Using these methods, a value of similarity between any two documents within the search space, or the prime document can be made. However, simple document to document similarity is not the most informative in terms of characterizing the total search space. By identifying similar patterns within documents and grouping them, trends within the overall search space can be characterized; from these, conclusions about the state of the art in both application and research can be made.

3.3. Document Grouping

As a method of further characterizing both the overall search space and the broad scale trends relating to the prime document, a form of fuzzy k-means clustering was chosen to help group the documents. This additional layer of analysis allows for the informed selection of a broader spectrum of documents to deeply analyze for the purposes of the survey. By selecting papers that are the most like the prime document within each group, the survey will better be able to characterize a broad scope of the state of the art without over-focusing on any one sub-section artificially or miss a prominent area by accident.

To characterize each group, the most frequently appearing words and phrases are identified to help identify the central ideas and topics of the group. The topic analysis is performed both with the absolute overall most frequent phrases and those that are exclusive of the top phrases across all groups. Lastly, each group is characterized by its average similar similarity to the prime document to quantify how relevant it is to the central topic of the survey. The results of the grouping as well as the similarity and characterization analysis are presented in the results section of this survey.

4. Results

4.1. Search Space Characterization

A pool of nearly 4000 papers was created and used to evaluate of the state of the art of manufacturing cross-domain (design-manufacture-quality) use of ML. To help confirm that the constructed corpus of articles is centered in the field of manufacturing, Figure 2 shows that words such as "Manufacturing," "Process," and "Design" comprise the most frequent words.

The analysis of density-based clusters within the corpus feature space provides further characterization of the total search space. The top 10 clusters are analyzed, each one reflecting a different region of density within the data. As shown in Figure 3, the clusters are labeled from 1 to 10 based on the total number of documents held within them. The upper chart compares the number of documents contained in each group, while the lower chart shows the proximity of each group center within the search space. Groups represented in the lower left of the chart are both close to the center of the search space and close to other groups, those in the upper right are far from center and have no near neighbors. Those in the far right are considered near the edges of the search space.

To better understand some of core the concepts that each group centers around, phrases that represent strong contributors to that point in the feature space are extracted and reported below. This can provide a broad understanding of the overall focus as well as areas of concentration of the corpus. The most frequent phrases overall in the search space have already been noted in Figure 2, therefore it is more informative to look at the most

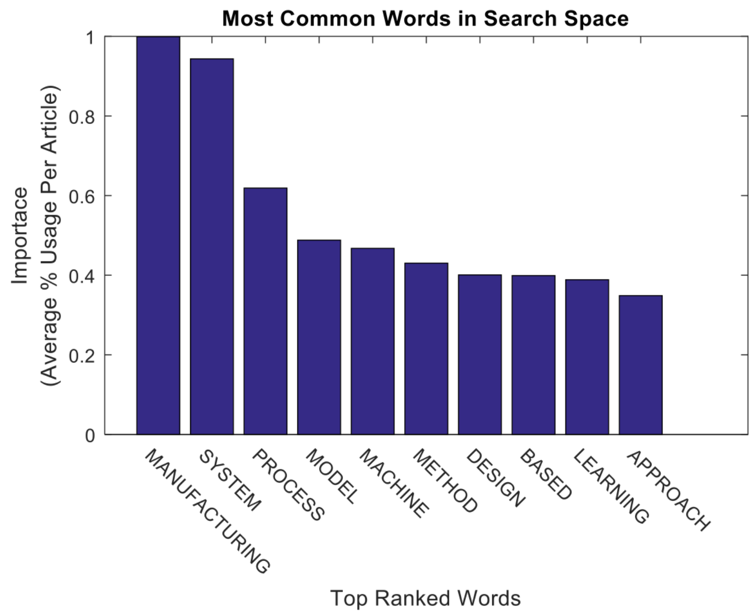

Figure 2: Search space characterization

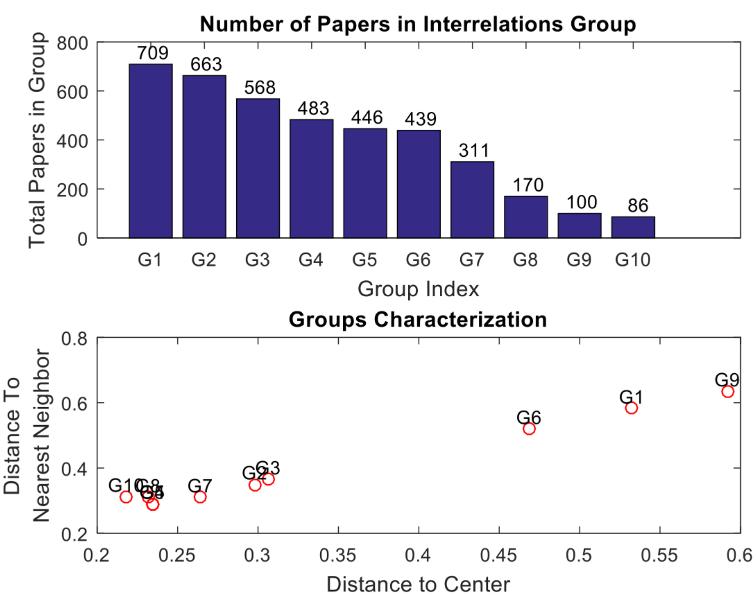

Figure 3: Cluster characterization

frequent phrases of each group exclusive of those common to all groups (See Table 2).

From this information, we can conclude that there is a large portion (approximately 28%) of the search space focused on the development of new algorithms, particularly inclined towards ML in manufacturing. Similarly, we see that approximately 34% is interested in production control and management. Even in this loose characterization of the search space, the interest in smart manufacturing and data utilization is apparent.

Following from this characterization, two explicit ML algorithms that have dedicated areas of density towards them are Neural Networks (NNs) and Support Vector Machines (SVM), implying a growing popularity of them within manufacturing. With approximately 11% and 4% of the corpus focused around these generically applicable ML algorithms respectively, indications are strong that they are being applied to a broad range of problems or applications in the manufacturing domain. This is most likely due to the low barrier of entry for use of the algorithms, both in understanding to create and physical resources required. Open source, pre-made implementations of these tools are widely available on the internet, further promoting adoption. This is not meant to imply that these are the only algorithms in use in the manufacturing domain, only that there is a prominence of both NNs, SVMs and related algorithms.

Note that, evidenced by the key phrases of Group 10, approximately 2% of the articles found seem to be related to the manufacturing of elements and isotopes through radioactive production methods. This is not a focus set of our survey, and can be marked as a trivial matching; another weakness of a simplistic key word search algorithm which created the overall corpus. As this is a suitably low percentage of the overall search space, these articles are not removed from the analysis. However, these and other outlying trivial, or incidental matches exist within the search space and are expected to have contributions to the low end of the similarity distribution for any manufacturing industry related search.

After establishing the search space as an acceptable representative sampling of recent publications regarding the manufacturing industry, two key documents were used to help infer and characterize the overall thrust and trends of focus areas as pertaining to the digital thread and process design in these publications. The individual similarity of each publication was evaluated and recorded for each key document separately. By knowing the focus and intent of each of the key documents, a sense of how prevalent these top-

Table 2: Most frequent phrases for each group

	Select Characterizing Phrases	Search Space Coverage
Group 1	CONTROL	17.80%
	MANUFACTURING SYSTEM	
	PRODUCTION	
	INTELLIGENT	
	MLTA	
Group 2	ENERGY	16.60%
	INDUSTRY	
	DEVELOPMENT	
	MANAGEMENT	
Group 3	SMART	14.20%
	ENERGY	
	GRID	
	APPLICATION	
Group 4	MLTA	12.20%
	MACHINE LEARNING	
	ALGORITHM	
	TECHNIQUE	
Group 5	ALGORITHM	11.20%
	PROPOSED	
	NEURAL NETWORK	
	RESULT	
Group 6	MANUFACTURING PROCESS	11%
	CONTROL	
	PRODUCTION	
	MLTA	
	QUALITY	
Group 7	STUDENT	7.80%
	ENGINEERING	
	PROJECT	
	COURSE	
Group 8	SUPPORT VECTOR MACHINE	4.30%
	ALGORITHM	
	MLTA	
	RESULT	
Group 9	SUPPLY CHAIN	2.50%
	PRODUCT	
	PERFORMANCE	
	NETWORK	
	AGILITY	
Group 10	PHOTOCATALYTIC	2.20%
	VISIBLE LIGHT	
	ACTIVITY	
	DEGRAMLTION	
	IRRADIATION	

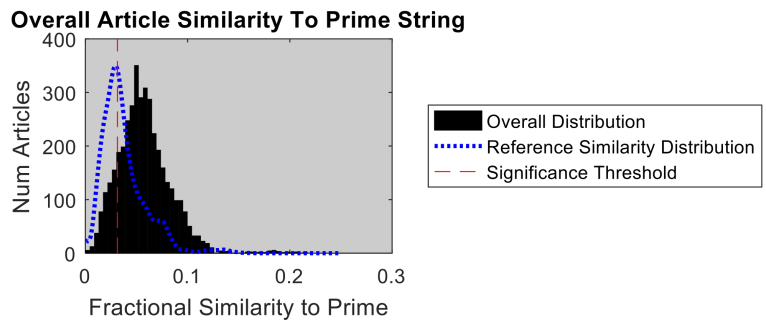

Figure 4: Search space similarity to the introduction of this paper

ics are within the search space can be gained. The first key document is the introduction section of this paper, the second is a directed set of select phrases designed to highlight important concepts.

Additionally, to establish a baseline, a short article unrelated to the field of manufacturing is also used to create a key document. In this case, the reference document selected is the abstract of an article related to peach farming that contains many of the key phases used to construct the total survey search space, but is expected to only have incidental similarity at best. This serves to aid in distinguishing random matching within the similarity algorithm.

As the technique for document comparison is designed to capture similarities beyond simple word matching, the reference key document is expected to be better constructed from natural language. Towards that end, the introduction section of this paper is used (prior to minor editorial changes) as a key document (Prime String) to compare the search space similarity to the core concepts of this survey. Figure 4 shows the distribution of the search space similarity to the introduction section of this paper.

In comparison to the baseline chance similarity developed from the corpus similarity to the peach farming article (dotted line in Figure 4), there is a significant increase in the overall similarity distribution (approx. 0.02 mean shift). This clear visual and statistical shift strongly implies that the core concepts described within the introduction of this paper (Section 1) are of overall interest in manufacturing. However, the fact that the peak significance seems mostly normal without a strong secondary peak suggests that few (if any) of the articles address all the concepts within this survey. Were a significant number the articles to address all or most of the themes in the

Table 3: Machine learning base phrases

Including	Excluding
Machine Learning	Social
Neural Networks	Food
Reinforcement Learning	Oil
Smart Manufacturing	Gas
Intelligent Manufacturing	Smart Grid
Flexible Manufacturing	Finance
Agile Manufacturing	Young Firms
Reconfigurable Manufacturing	Radioactive Decay
Data Driven	Irradiation
Product Life Cycle	
Cross-Domain	
Data Analytics	

survey, those articles would be expected to produce a second, significant peak centered at a value higher than the bulk similarity of the corpus. As is, there may be some articles addressing a large portion of the themes, but not a significant enough number of these articles to fall outside the normal corpus similarity distribution. These facts together lend credence to the suggestion that this an area in need of further development.

To further corroborate this hypothesis, a vector of weighted and directed key words was also analyzed as a key document for comparison with the search space of manufacturing related articles. This vector was constructed to capture many key concepts within the areas of machine learning to help evaluate the extent of research on novel utilizations of automated and rapid analysis methods in the manufacturing search space. As a base line reference for this vector, a similar sized vector was constructed from the top occurring phrases in the Peach Farming abstract mentioned before. This new set baseline is necessary due to the increased chance of randomly matching a vector with lower total phrases, either in part or total. The base phrases used to construct the machine learning query vector of phrases are provided in Table 3. All subsequent substrings of phrases are also added to the query vector to aid in capturing core concept similarity.

Figure 5 shows the similarity distributions for both the directed search vector of phrases and the reference vector constructed from the random doc-

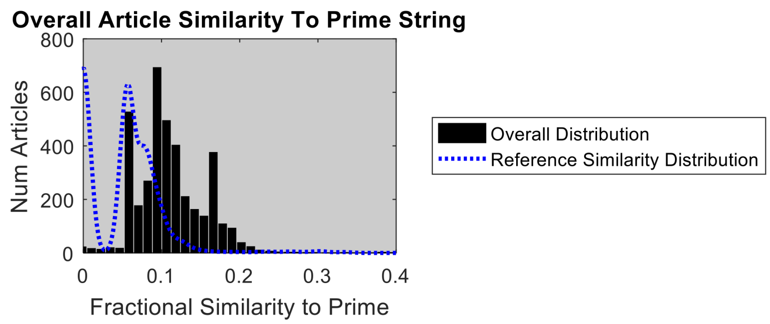

Figure 5: Directed vector similarity distribution

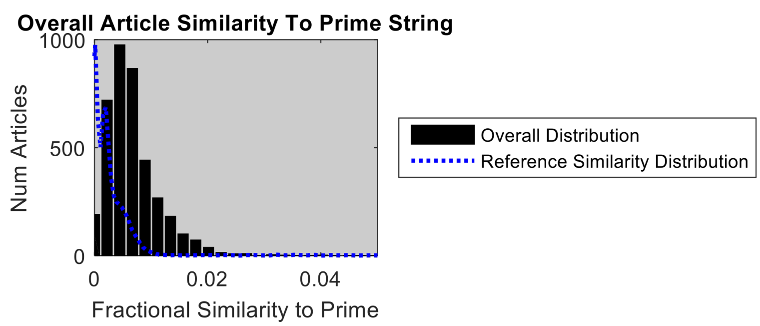

Figure 6: Convolution of similarity distributions

ument. From this, it is evident that the targeted search vector again has a significant overall increase in similarity to the reference distribution; confirming interest in ML throughout the search space. As an aside, the seeming multi-modal nature of the Reference distribution is due to the comparatively lower number of possible feature phrases to match. This means that the jump between near zero matches and at least one match becomes much more significant. Additionally, another intuitive reason for this jump is that if one match exists, the probability of multiple matches increases as well.

Further insight into the state of the art can be gained by looking at the convolution of similarity distributions (Figure 6) for both the prime document and the directed search vector. This provides a sense of how many articles in the search space are targeting smart manufacturing processes with an emphasis on ML algorithms.

Once again, overall the similarity distribution is low, implying few or zero exact matches exist in the search space, but the marked increase over

Table 4: Average similarity summary

	Reference	Query	% Difference
Prime Document	0.0373	0.0575	54.15
ML Targeted Vector	0.0498	0.1076	116.06
Convolution	0.0027	0.007	159.25

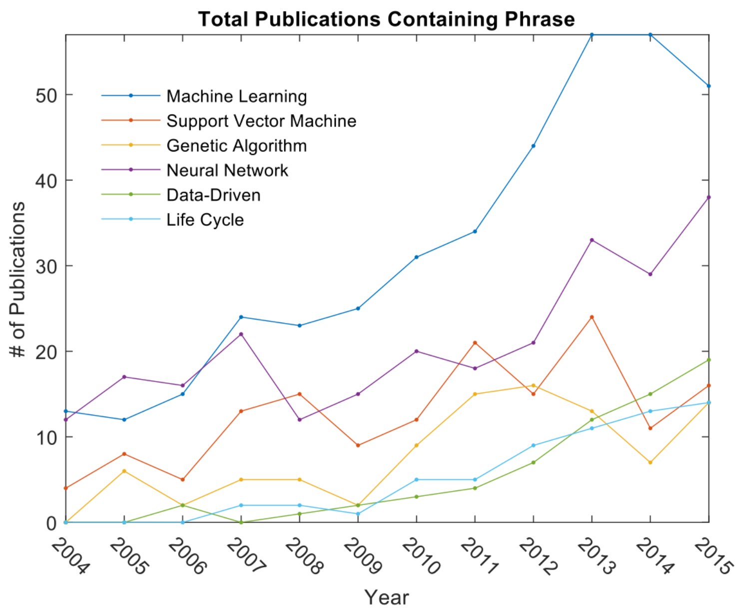

Figure 7: Visualization of interest in ML by year

the reference distribution implies strong interest in the general concepts. A summary of the relative increases for each similarity query is provided in Table 4 below.

It is convenient and intuitive to use these percentages of relative similarity as metrics indicating interest in the concepts they represent within the search space. This would seem to imply a marked interest in ML within the field of manufacturing with regards to product life-cycle management (PLM). A final testimony for the supposition that the field of manufacturing is becoming increasingly interested in ML and associated techniques can be seen in the visualization below. Figure 7 shows how many publications relate key phrases by year, indicating a growing interest over time.

Notice that many common concepts and algorithms within machine learning are increasingly being the subject of publication. Similarly, interest in life-cycle management can be seen to have an increasing presence. With this information in mind, the remainder of this document reviews a selection of articles within the workspace that are selected to both be like the prime document as well as be a representative selection from across the different regions of density within the search space. Documents for review were chosen by the authors both for their calculated similarity to the prime document and for special interest to the survey themes as judged by the authors.

4.2. Survey Results

A brief survey was performed consisting of papers targeted by the NLP methodology described in previous section to further expand upon the ideas characterizing the current state of knowledge-based usage and ML across all stages of the manufacturing process. This also serves to further validate the assertions and conclusions formed from the NLP algorithms while also adding a degree of context to those results. The following section collects and summarizes notable uses or potential uses of machine learning within industry, seeking to answer questions such as:

- How is machine learning currently being applied in manufacturing?
- How is machine learning enabling better information management and decision support?
- What digital knowledge bases exist and how are they being utilized and maintained?
- What are the current approaches to total Product Life-Cycle Management (PLM)?
- What gaps exist in these areas that must be filled to provide a fully realized digital production environment that maximally utilizes available resources across all stages of production?

4.3. Current Application of ML in Manufacturing

The first of these questions is possibly the most generic, yet with the most potential impact upon the others. Machine learning is broadly defined by the concept of having a computer update a model or response based upon new data or experiences through its learning lifetime. Self-updating algorithms have the potential to greatly benefit the manufacturing industry at all stages and levels of operations management.

4.3.1. Decision Support

Perhaps the most intuitive uses of machine learning occur at some of the more basic and fundamental levels of the manufacturing process. Even before the first part is made, manufacturers need to have an accurate estimation of the production cost to not only set pricing, but on a more fundamental level, to determine if the proposed product is financially viable or feasible for the company. Many factors go into the determination of cost, such as materials, process and tool costs, batch sizes, scheduling, outsourcing, etc. There is often heavy interplay within these variables, thus making it difficult for any single person to make fully informed and accurate estimations. This is particularly true in instances where the product is largely novel or dissimilar to the companys typical products. Deng and Yeh [24] propose a solution involving a Least Squares approach utilizing Support Vector Machines (SVM) to characterize the cost space and find a unique optimal cost estimation for a product. The use of SVM lends itself to rapid incorporation of new or updated cost estimations that can easily added to the model after the actual cost of a job is determined, either strengthening or, where needed, superseding areas of the model with better information. This work was later extended upon comparing two machine learning methods – back-propagation neural networks (BPNs) and least squares support vector machines (LS-SVMs) to the life-cycle cost estimation problem [25]. Authors combine LS-SVMs with a data transformation – the log-sigmoid transfer function, and conclude that such data preparations play an important role in obtaining more accurate, available, and generalizable cost estimation model can be provided by this novel combining mechanism. Data preprocessing where practitioners clean, filter and transform data is a common and important step to solve outliers, missing data or scaling issues. This step also includes data reduction, which is selection and extraction of both features (input variables) and samples in a database [26].

Another prime example of machine learning aiding in manufacturing decisions is proposed by Woodward and Gindy [27]. They briefly describe a decision support system that utilizes Genetic Algorithms (GAs) to explore and optimize some collected set of elemental heuristics as guided by some predefined set of hyper-heuristics for finding solutions to difficult questions with often “softly defined” criteria. Within the paper, the application is proposed to aid in determination of decisions affecting ecological impact of the process, but the idea can easily be adapted to other important decisions where many factors interplay and have far reaching effect that may not be instantly apparent. The authors of this paper go on to assert that the library of elemental heuristics should be drawn from a diverse set of experts, each from different relevant areas of the issue in question. This library should be able to grow and update over time, making it as collaborative and integrated as possible, creating a framework for collecting valuable experience and knowledge from both people and machines. This symbiotic and archival approach leverages the ability of humans to create highly accurate approximate rules without the distortion often associated with “tribal knowledge”, or information passed from person to person, while also allowing the computer to rapidly manage and apply more of that information than any single person could reasonably be expected to. Beyond simple solution optimization, bootstrapping and clustering can be applied to obtain and present the end user with groups of similar solutions from which a decision maker can choose one that best suits the companys needs.

Managing scheduling of machine time for operations on a production line is rarely a trivial task with flexible manufacturing systems (FMS). Like many other production line decisions, optimal scheduling is a function of many interplaying variables that are difficult to completely manage and totally account for by a human decision maker. Yusof et al. [28] addresses the flexible manufacturing systems (FMS) and considering a machine-loading problem in FMS environment. The machine loading problem can be defined as the allocation of part operations and required tools to the machines, to optimize some objective(s) subject to some technological constraints. The problem covers many objectives such as maximization of the utilization of resources, minimization of processing and tooling costs and maximization of throughput. The authors point out that purely heuristics-based algorithms will often oversimplify a problem or be too specific to generalize across multiple projects. Conversely, many more mathematically rigorous methods can become bogged down with expanded complexity or dimensionality of many problem formulations, in terms of both computation and configuration time. They propose the idea of utilizing harmony search (HS) algorithm based on the musical harmony [29], to aid in creating an optimal schedule. The algorithm suggested by the authors seeks to strike a balance between those two extremes by using rough heuristics to determine the feasibility of a solution, then applying machine learning algorithms to optimize the problem within the, now, more limited scope. Methods such as this could greatly reduce the time needed to design and implement process work flow plans, while still providing strong justification and a high level of confidence in the management decisions that were made.

Maintenance scheduling is another area often performed sub-optimally (if at all) in many smaller and sometimes even larger industrial companies. Reactive maintenance, repairing an item or process only after a failure has been identified, is largely considered the least optimal method of maintaining a system, incurring large amounts of unplanned downtime and often allowing cascading failures that need not have occurred if preventative maintenance had been properly performed. Most preventative maintenance can broadly be classified into three categories; cycle based, current condition based, and predicted condition based. Cycle-based plans schedule maintenance after a set number of operational/calendar hours or duty cycles, such as miles driven on a tire. Current condition-based plans assess the current state of the system and perform maintenance when it is within a predefined limit or when a triggering event occurs, such as the growth of a fault frequency beyond a set threshold. The last category looks at the current and past states of the system as well as its expected future operational load to predict when a fault will likely occur, or how much Remaining Useful Life (RUL) a system has. Wu et al. [30] uses ML and NNs to predict the percentage of remaining useful life in rotating equipment. The authors propose an intelligent decision support system based on this technology to promote optimal maintenance strategies. Further, by reducing unnecessary maintenance, a production plant can maximize equipment utilization and availability while also reducing costs of repairs.

4.3.2. Digital Knowledge Management: Plant and Operations Health Management

For a fully developed maintenance regime, diagnostic and prognostic information exists for a multitude of individual units that interact within a manufacturing process production line. Although information regarding individual units can be incredibly useful for identifying and scheduling maintenance plans, there is much more information to be gained from the analysis of the total system. As an example, this form of analysis could help in identifying the interplay between linked units, each possessing non-terminal or even seemingly trivial amounts of degradation, causing an overall cumulative deleterious effect on the product. Choo et al. [31] introduces a hierarchical Markov Decision Process known as Adaptive Multi-Scale Prognostics and Health Management (AM-PHM) to help manage produced diagnostic and prognostic knowledge at all levels of the manufacturing system. This approach helps to overcome the problems of exponential complexity as information is aggregated up the manufacturing system starting at individual components, and moving to work cell and assembly line levels. After information is pulled to and managed at the high level, decisions made can then be translated back down to the lower levels, informing specific tasks to undertake to ensure reliability of the production line as well as the quality and integrity of the product produced.

Another application towards extending the life of manufacturing equipment, specifically robotic arms, focuses on incorporating Linear Temporal Logic (LTL) in to their monitoring and control scheme [32]. This technology adds intelligent autonomous diagnostic systems that can connect continuous and discrete prognostics. Having monitoring systems tolerant of connected systems is imperative and invaluable in the interconnected world of modern manufacturing work cells. The authors envision that this LTL-monitor system could be extended to selectively guide robotic motion sets towards those that produce the most even wear on joints to even further extend the life of these systems.

4.3.3. Digital Knowledge Management: Data Management

The volume of data produced by manufacturing systems is rapidly growing beyond the capabilities of traditional algorithms, especially for users who want the most useful information from their data. High sample volume as well as huge numbers of dissimilar data sources are creating a need for both information consolidation and isolation algorithms that can be implemented in a distributed parallel fashion to meet the computational speed requirements necessary for prompt knowledge utilization. Collected information that is not able to be correctly interpreted or made useful in a timely manner is rarely even so much as marginally better than having not collected the data. Kumar et al. [33] propose utilizing tools such as the Hadoop framework and cloud computing to help overcome this problem. Working with map-reduce algorithms, the authors go on to propose a method for dealing with data imbalance issues (having a large discrepancy between the number of exemplar cases for different categories of data). Specifically, they highlight an algorithm for overcoming data imbalance for the goal of fault detection or identification through traditional machine learning classification algorithms. Automating large scale diagnostic data alarming could produce much more informed decisions about maintenance scheduling, workload, and demand cycles to maximize out and reliability of the system.

The concept of cloud computing, the delivery of computing services over the internet (the cloud), was born to address the administration and storage of big data, and the scalability of services challenges, and to increase efficiency. Having the potential to be the next major driver of business innovation, cloud services could be part of a business strategy for manufacturing companies. One benefit of moving towards cloud manufacturing is the ability to store large amounts of critical data in the cloud, and access to resources in real-time. A detailed work on "cloud manufacturing", and what parts of a company can easily and quickly adopt cloud-based solutions has been prepared by Xu [34]. In this work, Xu analyzes the benefits of integrating cloud technology into a typical manufacturing company after discussing the essential requirements of a cloud computing system. IIoT and cloud services are two key paradigms for the construction of virtual manufacturing. In this context, the author describes MTconnect [35], STEP (Standard for the Exchange of Product Model Data) [36] and STEP-NC (STEP-Compliant Data Interface for Numeric Controls) [37]. MTconnect is a manufacturing communication protocol used for data integration, and STEP provides a way to share product data over the entire life cycle of a product. STEP aims to accurately capture product definition and provide data interoperability between native systems, such as: Computer-Aided Design (CAD), Computer-Aided Manufacturing (CAM), Analysis (CAE), and Inspection (CMM) software. Healthy adoption of cloud solutions must include effective integration of the existing data-exchange standards and/or protocols.

Brodsky et al. [38] have developed a system for managing a repository and conducting analysis and optimization on manufacturing models in Brodsky et al. [39] and Brodsky et al. [38], respectively. The former work proposes an architectural design and framework for fast development of software solutions for descriptive, diagnostic, predictive, and prescriptive analytics of dynamic production processes. The uniqueness and novelty of the proposed architectural design and framework is its middleware layer, which is based on a reusable, modular, and extensible Knowledge Base (KB) of process performance models. However, this related effort lacked a systematic design of the unit manufacturing process (UMP) repository and possible ecosystems around the repository, as well as a specific architecture for such a repository.

Furthermore, it did not address an implementation of a reusable repository and support for populating it with dynamic production processes. The authors address these gaps in their following work. They first propose the concept of a reusable KB of manufacturing process models, its functionality and high-level system architecture capable of supporting future ecosystems around it. Then, they implement an initial collection of performance models for milling and drilling as well as a composite performance model for machining. They also develop a system for managing a repository and conducting analysis and optimization on manufacturing models where the initial scope of the system includes (1) an Integrated Development Environment (IDE) and its interface through the use of Atom Studio [40], (2) simulation and deterministic optimization of performance models through the use of Unity Decision Guidance Management System (DGMS), and (3) model management and version control through the use of the standard interface of GitLab [41].

Manufacturing standards provide the means for industries to effectively and consistently deploy the necessary measurement science to assess process performance. These assessments ultimately set the stage for controlling the manufacturing systems and processes and enabling continuous improvement within the enterprise. Several evolving manufacturing-related standards lay foundations for modeling and integrating manufacturing systems and related services. Bloomfield et al. [42] proposed a framework to standardize the data exchange between manufacturing applications throughout the product life cycle. By implementing the Core Manufacturing Simulation Data Information Model (CMSDIM) developed by researchers at NIST [43], and chartered by the Simulation Interoperability Standards Organization (SISO), they aim to enhance interoperability between manufacturing applications at multiple stages of the product life cycle.The Core Manufacturing Simulation Data (CMSD) standard specifies the information entities common to manufacturing simulations to facilitate simulation model construction and data exchange between simulation and other manufacturing applications within a shop floor. Authors [42] discuss information gaps between the lean design engineering software and discrete event simulation. With their developed software, called "UA translator", the authors report that they decreased the time to develop manufacturing applications, could eliminate of human error and introduce of process time variation. Another data interoperability standard for manufacturing quality data is Quality Information Framework (QIF) [44]. QIF is an XML-based standard that was created and managed by the Dimensional Metrology Standards Consortium (DMSC). It supports Digital Thread concepts in engineering applications ranging from product design through manufacturing to quality inspection [44].

4.3.4. Lifecycle Management

To facilitate the total integration a manufacturing system such that it can fully utilize the volumes of information being produced about it, there needs to be clear system of communication. The concept of the "Internet of Things (IoT) is exactly this, with both components and controllers directly communicating with each other as well as system coordinators and decision makers. An interesting extension of this is presented in the work of Aruväli et al. [45], detailing the notion of Digital Object Memory (DOMe). With this notion, information relating to each unit on a production line such as g-code, diagnostics, quality information, and even a complete list of machine interactions could follow a product through its entire lifetime from initiation of production to consumer purchase. As an example, the authors suggest that the manufacturing machines could stream real time information regarding surface roughness to the product item giving it the ability to self-assess its quality after production. Unfortunately, as is also explained in their work, this goal is currently not practical due to the difficulties of developing communications with largely dissimilar pieces of equipment, components, parts, and etc. The DOMe implementation could be both hardware and software driven to ensure high fidelity and storage of production information, but would need some open source communications standard that universally applies to all the relevant constituents in manufacturing systems.

Building upon the idea of cyber physical systems, Barthelmey et al. [46] describe a system to use both hardware and software to track changes in a manufacturing facility automatically. The goal of this work is to create an up to date set of documentation detailing the capabilities and status of a facility. This "self-organized creation of technical documents" could create a general cost savings by eliminating or reducing the costly upkeep of technical documentation as it quick goes out of date due to the ever-changing manufacturing systems. Much of this change tracking can, and should (as explained by the authors) be automated; keyed to some initiating event. Events that could trigger autonomous updates of the documentation include some physical change of the system as reported by various sensors and monitoring equipment, or the passage of some preset amount of duty cycles or calendar hours. Additionally, a prompt for document updates could occur after any soft change to the system, such as maintenance. This logging scheme could promote not only well-maintained documentation in a context sensitive database, but also well-informed decisions regarding process planning.

4.4. Identified Gaps and Needs

With the popularity of smart-phones and similar tablet devices, the practicality of implementing a standardized modular application-driven environment in industrial settings is gaining support. Gröger et al. [47] discusses the idea of a "App"-based manufacturing tool-set. With a unified platform for application development users could develop specific tools that aid in every level of production manufacturing. These tools could be task specific for a single company, or more broad reaching, such as an interactive diagnostic maintenance tool that helps a user trouble shoot equipment on the production floor. Apps linked with online information repositories could have access to and provide contextually pertinent information at times and situations where it can have the most impact. For this to be maximally utilized, some unified platform for the Apps to be built within needs to be developed, further some repository should exist where end users from diverse companies could acquire or submit applications like the Google Play Store [48]. Standards regarding input/output protocols of the applications would need to be implemented as well as security and user interfacing. While the idea of contextualized, "right place right time" digital tools has the potential to greatly speed up maintenance, production, and development tasks in manufacturing, the lack of a standard or unified platform currently prevents this from being realized.

Dekkers et al. [49] performed a survey regarding the linking of multiple stages of the product life-cycle development, where they arrived at some notable conclusions about improving production. The need for a link between product design and engineering stages of the life-cycle and the implemented manufacturing plan has been known but not properly addressed since the early 1990s, although some strides have been made. One of the major hold ups is the lack of appropriate standardized software that can help create and manage a consistent repository of the knowledge. The knowledge base should contain contextualized knowledge for all stages of the life-cycle. Such a knowledge base can aid not only in production decisions within a company, but also with concurrent engineering and design or fabrication sourcing. Although the large amounts of data associated with this may slow down the process, specialized software could aid in tracking and management of the repository. As of now, no standardized repository or method for constructing such a repository exists, but the need and growing interest in one is highly apparent and is expected to create a shift in manufacturing practices when it eventually enters the industry.

Generic solutions for applying ML to cross-domain focus areas in the product life cycle are absent in the literature. The majority of previous work has been focused on specific use cases and domains. In addition, there is little-to-no use of cross-domain data and application of ML. Hedberg Jr et al. [8] identifies several research directions for using manufacturing knowledge earlier in the product life-cycle. They suggest dynamic knowledge bases could be generated by applying ML to data from several domains / phases of the life-cycle. Additionally, industry needs guidance on the minimum information requirements for the product life-cycle because of the different data requirements for each phase of the life-cycle [8] . However, when data and information is passed between the phases, information is lost, which requires iterations of communicating to ensure all the requirements for each phase are met. Our review described in this paper supports most of the findings from Hedberg Jr et al. [8]. ML applied to cross-domain use cases is an untapped area of research that would bring significant benefit to industry.

5. Conclusions

In this work, we developed and analyzed a corpus of approximately 4000 abstracts from technical documents centered in the field of manufacturing using a series of NLP techniques. Going beyond simple key term matching, this work endeavored to provide concept matching, with a clear methodology and justification for characterizing general trends within the corpus as well as directed searches for concepts of interest in the realm of ML. From this, clear trends indicating the increasing prevalence of digital automation and ML appear throughout the manufacturing industry. Notably, generically applicable algorithms such as NNs and SVMs are gaining popularity. Algorithms such as these can produce compelling results with a low investment of time and resources to setup and maintain, making them very appealing for a wide array of problems.

During the characterization of the total corpus, several key papers were identified for a more complete, in-depth review. The results of the in-depth manual survey confirmed many of the characterizations and suppositions about the total corpus developed by the NLP information mining. Survey results show that there is a growing interest in lifecycle management, as well as in applications of ML in manufacturing. Areas of knowledge management, decision support, and lifecycle management are increasingly becoming augmented by automated technologies. Despite this, there are still significant gaps that could benefit from further development and adoption of some of these state-of-the-art technologies.

As the manufacturing industry moves toward "automated manufacturing", the role of data management and processing becomes more prominent. With the availability of data in each phase of product life-cycle, and advances in algorithms and software tools, ML is emerging as an appropriate and promising tool for more agile, lean, and energy-efficient manufacturing systems. This trend and others necessitate pushing towards the right combination of human resources, automation and data, PLM, as well as the link between ML and IIoT.

Retrieval from information silos and single-domain data re-use is the generally accepted practice currently applied with ML. A holistic view of ML applications across life-cycle is still a challenge. We recommend that academia and practitioners shift ML research and applications towards more of a life-cycle or enterprise-wide focus to take advantage of the ever growing mass of data. This would enable cross-domain data usage and could benefit industry with improved knowledge generation in each phase of the product life-cycle.

Acknowledgments

The authors wish to thank Allison Barnard Feeney (NIST), Michael Brundage (NIST), and Vijay Srinivasan (NIST), Alden Dima (NIST), and the peer-reviewers for their comments and input to this paper.

Disclaimers

 Certain commercial systems are identified in this paper. Such identification does not imply recommendation or endorsement by NIST. Nor does it imply that the products identified are necessarily the best available for the purpose.

[1] C. Piddington, M. Pegram, An ims test case - global manufacturing, in: Proceedings of the IFIP TC5/WG5.7 Fifth International Confer-

ence on Advances in Production Management Systems, North-Holland Publishing Co., 717937, 1993, pp. 11–20.

[2] K.-D. Thoben, S. A. Wiesner, T. Wuest, "industrie 4.0" and smart manufacturing a review of research issues and application examples, International Journal of Automation Technology 11 (2017) 4–16.

[3] S. Pugh, Total design : integrated methods for successful product engineering, Addison-Wesley Pub. Co., Wokingham, England ; Reading, Mass., 1991.

[4] K. Yang, B. El-Haik, Design for six sigma : a roadmap for product development, McGraw-Hill, New York, 2nd edition, 2009.

[5] C. E. S. da Silva, E. G. Salgado, C. H. P. Mello, E. da Silva Oliveira, F. Leal, Integration of computer simulation in design for manufacturing and assembly, International Journal of Production Research 52 (2014) 2851–2866.

[6] I. H. Garbie, DFSME: design for sustainable manufacturing enterprises (an economic viewpoint), International Journal of Production Research 51 (2013) 479–503.

[7] J. G. Bolten, R. S. Leonard, M. V. Arena, O. Younossi, J. M. Sollinger, Sources of Weapon System Cost Growth: Analysis of 35 Major Defense Acquisition Programs, Report, RAND Corporation, 2008.

[8] T. D. Hedberg Jr, N. W. Hartman, P. Rosche, K. Fischer, Identified research directions for using manufacturing knowledge earlier in the product life cycle, International Journal of Production Research 55 (2017) 819–827.

[9] M. Helu, T. Hedberg Jr, A. Barnard Feeney, Reference architecture to integrate heterogeneous manufacturing systems for the digital thread, CIRP Journal of Manufacturing Science and Technology (2017).

[10] Energetics Inc., Measurement Science Roadmap for Prognostics and Health Management for Smart Manufacturing System, Report, National Institute of Standards and Technology, 2015.

[11] R. Gao, L. Wang, R. Teti, D. Dornfeld, S. Kumara, M. Mori, M. Helu, Cloud-enabled prognosis for manufacturing, CIRP Annals - Manufacturing Technology 64 (2015) 749–772.

[12] M. Helu, T. Hedberg, Enabling Smart Manufacturing Research and Development using a Product Lifecycle Test Bed, volume 1 of *Procedia Manufacturing*, Elsevier Science Bv, Amsterdam, pp. 86–97.

[13] M. Li, S. Gao, C. C. Wang, Real-time collaborative design with heterogeneous CAD systems based on neutral modeling commands, Journal of Computing and Information Science in Engineering 7 (2006) 113–125.

[14] T. Hedberg Jr, A. Barnard Feeney, M. Helu, J. A. Camelio, Towards a lifecycle information framework and technology in manufacturing, Journal of Computing and Information Science in Engineering 17 (2017) 021010–021010–13.

[15] C. Jennings, D. Wu, J. Terpenny, Forecasting obsolescence risk and product life cycle with machine learning, IEEE Transactions on Components, Packaging and Manufacturing Technology 6 (2016) 1428–1439.

[16] Z. Li, D. Wu, C. Hu, J. Terpenny, An ensemble learning-based prognostic approach with degradation-dependent weights for remaining useful life prediction, Reliability Engineering & System Safety (In Press).

[17] J. Wang, Y. Ma, L. Zhang, R. X. Gao, D. Wu, Deep learning for smart manufacturing: Methods and applications, Journal of Manufacturing Systems (In Press).

[18] The International Society of Automation, Enterprise-control system integration part 1: Models and terminology, 2010. ANSI/ISA-95.00.01-2010.

[19] C. Johnsson, D. Brandl, K. Unger, ISA 95 for Beginners, Report, WBF, 2006.

[20] S. Chonde, S. Kumara, J. Storer, Content mining to support design, in: 2015 Industrial and Systems Engineering Research Conference, Norcross, Nashville TN, 2015, pp. 1672–1681.

[21] S. Chonde, A methodology of machine learning in automated entity summarization, Thesis, The Pennsylvania State University, 2016.

[22] S. T. Dumais, Latent semantic analysis, Annual Review of Information Science and Technology 38 (2004) 188–230.

[23] J. Leskovec, A. Rajaraman, J. D. Ullman, Mining of massive datasets, Cambridge University Press, Cambridge, second edition, 2014.

[24] S. Deng, T.-H. Yeh, Using least squares support vector machines for the airframe structures manufacturing cost estimation, International Journal of Production Economics 131 (2011) 701–708.

[25] T.-H. Yeh, S. Deng, Application of machine learning methods to cost estimation of product life cycle, International Journal of Computer Integrated Manufacturing 25 (2012) 340–352.

[26] S. Garcia, J. Luengo, F. Herrera, Data preprocessing in data mining, Springer, New York, 2014.

[27] J. Woodward, N. Gindy, A hyper-heuristic multi-criteria decision support system for eco-efficient product life cycle, 2010.

[28] U. K. Yusof, R. Budiarto, S. Deris, Harmony search algorithm for flexible manufacturing system (fms) machine loading problem, in: 2011 3rd Conference on Data Mining and Optimization (DMO), pp. 26–31.

[29] Z. W. Geem, J. H. Kim, G. Loganathan, A new heuristic optimization algorithm: Harmony search, SIMULATION 76 (2001) 60–68.

[30] S.-j. Wu, N. Gebraeel, Y. Yih, A neural network integrated decision support system for condition-based optimal predictive maintenance policy, IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 37 (2007) 226–236.

[31] B. Y. Choo, S. C. Adams, B. A. Weiss, J. A. Marvel, P. A. Beling, Adaptive multi-scale prognostics and health management for smart manufacturing systems, International Journal of Prognostics and Health Management (IJPHM) Special Issue: PHM for Smart Manufacturing Systems 7 (2016) 014.

[32] G. Heddy, U. Huzaifa, P. Beling, Y. Haimes, J. Marvel, B. Weiss, A. LaViers, Linear temporal logic (ltl) based monitoring of smart manufacturing systems, in: Annual Conference of the Prognosticcs and Health Management Society, volume 6, PHM Society, 2015, p. 10.

[33] A. Kumar, R. Shankar, A. Choudhary, L. S. Thakur, A big data mapreduce framework for fault diagnosis in cloud-based manufacturing, International Journal of Production Research 54 (2016) 7060–7073.

[34] X. Xu, From cloud computing to cloud manufacturing, Robotics and Computer-Integrated Manufacturing 28 (2012) 75–86.

[35] MTConnect Institute, MTConnect standard, 2014.

[36] International Standards Organization, Industrial automation systems and integration – product data representation and exchange – part 242: Application protocol: Managed model-based 3D engineering, 2014. ISO 10303-242.

[37] International Standards Organization, Industrial automation systems and integration – product data representation and exchange – part 238: Application protocol: Application interpreted model for computerized numerical controllers, 2007. ISO 10303-238.

[38] A. Brodsky, M. Krishnamoorthy, W. Z. Bernstein, M. O. Nachawati, A system and architecture for reusable abstractions of manufacturing processes, in: 2016 IEEE International Conference on Big Data (Big Data), pp. 2004–2013.

[39] A. Brodsky, G. Shao, M. Krishnamoorthy, A. Narayanan, D. Menasc, R. Ak, Analysis and optimization based on reusable knowledge base of process performance models, The International Journal of Advanced Manufacturing Technology (2016) 1–21.

[40] Atom: A hackable text editor for the 21st century, 2017, Accessed: 2018-08-24. Archived by WebCite at http://www.webcitation.org/6sxVmP24h.

[41] P. Hughes, A. Nwaigwe, S. Packham, A. Dunstone Gray, About gitlab, 2017-02-09, Accessed: 2017-08-24. Archived by WebCite at http://www.webcitation.org/6sxUV6u26.

[42] R. Bloomfield, E. Mazhari, J. Hawkins, Y.-J. Son, Interoperability of manufacturing applications using the core manufacturing simulation data (cmsd) standard information model, Computers and Industrial Engineering 62 (2012) 1065–1079.

[43] S. Leong, Y. T. Lee, F. H. Riddick, A core manufacturing simulation data information model for manufacturing applications, in: Systems Interoperability Standards Organization, 2006 Fall Simulation Interoperability Workshop.

[44] Dimensional Metrology Standards Consortium, Part 1: Overview and fundamental principles in quality information framework (QIF) an integrated model for manufacturing quality information, 2014.

[45] T. Aruväli, W. Maass, T. Otto, Digital object memory based monitoring solutions in manufacturing processes, Procedia Engineering 69 (2014) 449–458.

[46] A. Barthelmey, D. Störkle, B. Kuhlenkötter, J. Deuse, Cyber physical systems for life cycle continuous technical documentation of manufacturing facilities, Procedia CIRP 17 (2014) 207–211.

[47] C. Gröger, S. Silcher, E. Westkamper, B. Mitschang, Leveraging apps in manufacturing. a framework for app technology in the enterprise, Procedia CIRP 7 (2013) 664–669.

[48] S. McIlroy, N. Ali, A. E. Hassan, Fresh apps: an empirical study of frequently-updated mobile apps in the google play store, Empirical Software Engineering 21 (2016) 1346–1370.

[49] R. Dekkers, C. M. Chang, J. Kreutzfeldt, The interface between product design and engineering and manufacturing: A review of the literature and empirical evidence, International Journal of Production Economics 144 (2013) 316–333.