Journal of Manufacturing Systems xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

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

Deep learning for smart manufacturing: Methods and applications

Jinjiang Wang [a,*], Yulin Ma [a], Laibin Zhang [a], Robert X. Gao [b], Dazhong Wu [c]

[a] School of Mechanical and Transportation Engineering, China University of Petroleum, Beijing 102249, China
[b] Department of Mechanical and Aerospace Engineering, Case Western Reserve University, Cleveland, OH 44106, USA
[c] Department of Mechanical and Aerospace Engineering, University of Central Florida, Orlando, FL 32816, USA

ARTICLE INFO

Article history:
Received 13 November 2017
Received in revised form 1 January 2018
Accepted 2 January 2018
Available online xxx

Keywords:
Smart manufacturing
Deep learning
Computational intelligence
Data analytics

ABSTRACT

Smart manufacturing refers to using advanced data analytics to complement physical science for improving system performance and decision making. With the widespread deployment of sensors and Internet of Things, there is an increasing need of handling big manufacturing data characterized by high volume, high velocity, and high variety. Deep learning provides advanced analytics tools for processing and analysing big manufacturing data. This paper presents a comprehensive survey of commonly used deep learning algorithms and discusses their applications toward making manufacturing "smart". The evolvement of deep learning technologies and their advantages over traditional machine learning are firstly discussed. Subsequently, computational methods based on deep learning are presented specially aim to improve system performance in manufacturing. Several representative deep learning models are comparably discussed. Finally, emerging topics of research on deep learning are highlighted, and future trends and challenges associated with deep learning for smart manufacturing are summarized.

1. Introduction

Over the past century, the manufacturing industry has undergone a number of paradigm shifts, from the Ford assembly line (1900s) to Toyota production system (1960s), flexible manufacturing (1980s), reconfigurable manufacturing (1990s), agent-based manufacturing (2000s), cloud manufacturing (2010s) [1,2]. Various countries have developed strategic roadmaps to transform manufacturing to take advantage of the emerging infrastructure, as presented by Internet of Things (IoTs) and data science. As an example, Germany introduced the framework of Industry 4.0 in 2010, which has been evolved into a collaborative effort among member countries in the European Union. Similarly, in 2011 the Smart Manufacturing Leadership Coalition (SMLC) in the U.S. created a systematic framework for implementing smart manufacturing. The plan "China Manufacturing 2025", introduced in China in 2015, aims to promote advanced manufacturing. As manufacturing machines are increasingly equipped with sensors and communication capabilities, there is significant potential to further improve the condition-awareness of manufacturing machines and processes, reduce operational downtime, improve the level of automation and product quality and response more timely to dynamically changing customer demands [3–8]. Statistics shows that 82% of the companies using smart manufacturing technologies have experienced increased efficiency and 45% of the companies of the companies experienced increased customer satisfaction [9].

Smart manufacturing refers to a new manufacturing paradigm where manufacturing machines are fully connected through wireless networks, monitored by sensors, and controlled by advanced computational intelligence to improve product quality, system productivity, and sustainability while reducing costs. Recent advancement of Internet of Things (IoTs), Cloud Computing, Cyber Physical System (CPS) provides key supporting technologies to advance modern manufacturing [10–13]. By leveraging these new technologies in manufacturing, data at different stages of a product's life, ranging from raw materials, machines' operations, facility logistics, and even human operators, is collected and processed [12]. With the proliferation of manufacturing data, data driven intelligence with advanced analytics transforms unprecedented volumes of data into actionable and insightful information for smart manufacturing as illustrated in Fig. 1. Data driven intelligence models the complex multivariate nonlinear relationships among data, with no in-depth understanding of system physical behaviours required.

Data driven intelligence has attracted extensive research effort for manufacturing data distilling and decision making. In [14],

* Corresponding author.
E-mail address: jwang@cup.edu.edu (J. Wang).

https://doi.org/10.1016/j.jmsy.2018.01.003
0278-6125/

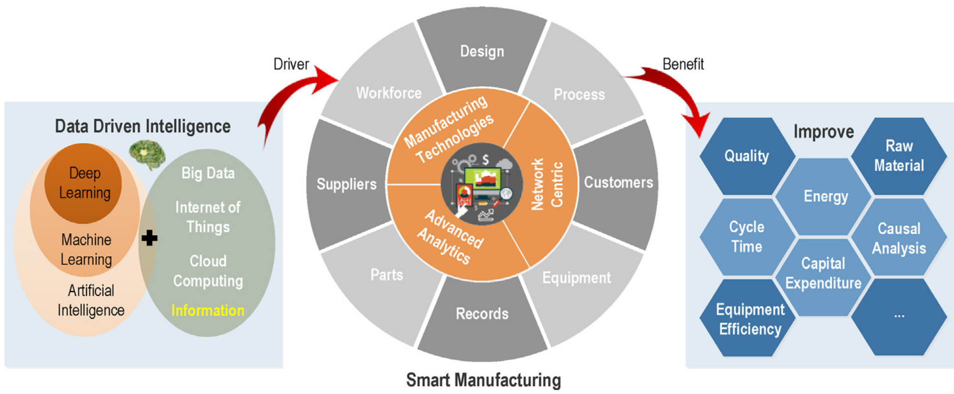

Fig. 1. The role of data driven intelligence in smart manufacturing.

data mining techniques are classified into five categories, including characterization and description, association, classification, prediction, clustering and evolution analysis. The barriers to data-driven decision making in manufacturing are also identified. Typical machine learning techniques are reviewed in [15,16] for intelligent manufacturing, and their strengths and weaknesses are also discussed in a wide range of manufacturing applications. A comparative study of machine learning algorithms including Artificial Neural Network, Support Vector Machine, and Random Forest is performed for machining tool wear prediction. The schemes, techniques and paradigm of developing decision making support systems are reviewed for the monitoring of machining operations, and these techniques include neural networks, fuzzy logic, genetic algorithms, and hybrid systems [17,18]. The potential benefit and successful application examples of typical machining learning techniques including Bayesian Networks, instance-based learning, Artificial Neural Network, and ensemble methods are discussed in [19]. Cloud enabled prognosis techniques including data driven approach, physics based as well as model-based techniques are reviewed in [20], with the benefits from both advanced computing capability and information sharing for intelligent decision making. Traditional machine learning is usually designed with shallow structures, such as Artificial Neural Network, Support Vector Machine, and logistic regression, etc. By coping with limited handcrafted features, it achieves decent performance in a variety of applications. However, the massive data in smart manufacturing imposes a variety of challenges [18,19], such as the proliferation of multimodal data, high dimensionality of feature space, and multicollinearity among data measurements. These challenges render traditional algorithms struggling and thus greatly impede their performance.

As a breakthrough in artificial intelligence, deep learning demonstrates outstanding performance in various applications of speech recognition, image recondition, natural language processing (e.g. translation, understanding, test questions & answers), multimodal image-text, and games (e.g. Alphago). Deep learning allows automatically processing of data towards highly nonlinear and complex feature abstraction via a cascade of multiple layers, instead of handcrafting the optimum feature representation of data with domain knowledge. With automatic feature learning and high-volume modelling capabilities, deep learning provides an advanced analytics tool for smart manufacturing in the big data era. It uses a cascade of layers of nonlinear processing to learn the representations of data corresponding to different levels of abstraction. The hidden patterns underneath each other are then identified and predicted through end-to-end optimization. Deep learning offers great potential to boost data-driven manufacturing applications, especially in the big data era [17,21].

In light of the above challenges, this paper aims to provide a state-of-the-art review of deep learning techniques and their applications in smart manufacturing. Specifically, the deep learning enabled advanced analytics framework is proposed to meet the opportunistic need of smart manufacturing. The typical deep learning models are briefly introduced, and their applications to manufacturing are outlined to highlight the latest advancement in relevant areas. The challenges and future trends of deep learning are discussed in the end.

The rest of paper is constructed as follows. Firstly, data-driven artificial intelligence techniques are reviewed in Section 2, with the superiority of deep learning techniques outlined. Next, the challenge and opportunistic need of deep learning in smart manufacturing are presented, and typical deep learning models are briefly discussed in Section 3. Then, the latest applications of deep learning techniques in the context of smart manufacturing are summarized in Section 4. Finally, the challenges as well as future trends of deep learning in smart manufacturing are discussed.

2. Overview of data driven intelligence

2.1. The evolution of data-driven artificial intelligence

Artificial intelligence is considered as a fundamental way to possess intelligence, and listed as the first place in Gartner's Top 10 strategic technology trends in 2017 [22]. Artificial intelligence has experienced several lifecycles, from the infancy period (1940s), through the first upsurge period (1960s) and the second upsurge period (1980s), and the present third boom period (after 2000s). The development trend and typical artificial intelligence models are summarized in Table 1.

The origin of Artificial Neural Network started back in 1940s, when MP model [23] and Hebb rule [24] were proposed to discuss how neurons worked in human brain. At the workshops in Dartmouth College, significant artificial intelligence capabilities like playing chess games and solving simple logic problems were developed [24]. The pioneering work brought artificial intelligence to the first upsurge period (1960s). In 1956, a mathematical model named Perceptron [25] was proposed to simulate the nervous system of human learning with linear optimization. Next, a network model called Adaptive Linear Unit [26] was developed in 1959 and had been successfully used in practical applications such as communication and weather forecasting. The limitation of early artificial intelligence was also criticized due to the difficulty in handling non-linear problems, such as XOR (or XNOR) classification [27].

With the development of Hopfield network circuit [28], artificial intelligence stepped forward to the second upsurge (1980s). Back Propagation (BP) algorithm was proposed to solve non-linear problems in complex neural network in 1974 [29]. A random mech-

Table 1
List of typical artificial intelligence models.

Timeline	Proposed models	Reference
Infancy period (1940s)	MP model	[23]
	Hebb rule	[24]
First upsurge period (1960s)	Perceptron	[25]
	Adaptive Linear Unit	[26]
Second upsurge period (1980s)	Hopfield network circuit	[28]
	Back Propagation	[29]
	Boltzmann Machine	[30]
	Support Vector Machine	[31]
	Restricted Boltzmann Machine	[32]
	Auto Encoder	[33]
Third boom period (after 2000s)	Recurrent Neural Network	[34]
	Long short-term Memory	[35]
	Convolutional Neural Network	[36]
	Deep Belief Network	[37,38]
	Deep Auto Encoder	[39]
	Sparse Auto Encoder	[40,41]
	Deep Boltzmann Machine	[42]
	Denosing Auto Encoder	[43]
	Deep Convolutional Neural Network	[44]
	Generative Adversarial Network	[45]
	Attention-based LSTM	[46]

anism was introduced into Hopfield network and put forward the Boltzmann Machine (BM) in 1985 [30]. With the development of statistical learning, Support Vector Machine (SVM) was developed with kernel functions transformation in 1997, and showed decent performance on classification and regression [31]. However, these traditional machine learning techniques require human expertise for feature extraction to reduce the dimension of input, and thus their performance highly relies on the engineered features.

The birth of deep learning benefits not only from the rich accumulation of traditional machine learning techniques, but also the inspiration of statistical learning. Deep learning uses data representation learning rather than explicit engineered features to perform tasks. It transforms data into abstract representations that enable the features to be learnt. In 1986, Restricted Boltzmann Machine (RBM) was developed by obtaining the probability distribution of Boltzmann Machine [32], and the hidden layers were used as feature vectors to characterize the input data. Meanwhile, Auto Encoder (AE) was proposed using the layer-by-layer Greedy learning algorithm to minimize the loss function [33]. In 1995, a neural network with directed topology connections between neurons, called Recurrent Neural Network (RNN), was proposed for feature learning from sequence data [34]. In 1997, an improved version of recurrent neural network, named Long short-term Memory (LSTM), was proposed to tackle the vanishing gradient problem and deal with complex time sequence data [35]. In 1998, Convolutional Neural Network (CNN) was put forward to handle two dimensional inputs (e.g. image), in which features learning were achieved by stacking convolutional layers and pooling layers [36].

As the hierarchical structures of deep learning models getting deeper, model training and parameter optimization become more difficult and time consuming, even leading to overfitting or local optimization problems. Many attempts were made to develop deep learning models, but no satisfactory performance was reported before 2006. Deep Belief Network (DBN) was developed and achieved success in 2006 [37,38]. It allowed bidirectional connections in top layer only instead of stacking RBMs directly to reduce computational complexity, and the parameters were successfully learned through layer-wise pre-training and fine tuning. Meanwhile, Deep Auto Encoder was proposed by adding more hidden layers to deal with high nonlinear input [39]. The model parameters were firstly pre-trained using a greedy layer-by-layer unsupervised learning algorithm and then fine-tuned using BP algorithm. One year later, Sparse Auto Encoder (SAE) was put forward to reduce dimensionality and learn sparse representations [40,41].

Deep learning gained increasing popularity. In 2009, Deep Boltzmann Machine with a bidirectional structure was proposed to learn ambiguous input data robustly, and the model parameters were optimized using layer-wise pre-training [42]. In 2010, Denoising Auto Encoder was presented to reconstruct the stochastically corrupted input data, and force the hidden layer to discover more robust features [43]. Deep Convolutional Neural Network (DCNN) was introduced with deep structure of Convolutional Neural Network in 2012 [44], and it showed superior performance in image recognition. Generative Adversarial Network (GAN) was proposed in 2014 [45], and it contained two independent models acting as adversaries. The generative model was designed to generate random samples similar to real samples while the discriminative model was used for training and classification with both real and generated random samples. In 2016, an attention-based LSTM model was proposed by integrating attention mechanism with LSTM [46]. Nowadays, more and more new models are being developed even per week.

2.2. Comparison between deep learning and traditional machine learning

Both deep learning and traditional machine learning are data-driven artificial intelligence techniques to model the complex relationship between input and output as shown in Fig. 2. In addition to the high hierarchical structure, deep learning also has distinctive attributes over traditional machine learning in terms of feature learning, model construction, and model training.

Deep learning integrates feature learning and model construction in one model by selecting different kernels or tuning the parameters via end to end optimization. Its deep architecture of neural nets with many hidden layers is essentially multi-level non-linear operations. It transfers each layer's representation (or features) from original input into more abstracted representation in the higher layers to find the complicated inherent structures. For example, the features such as edge, corner, contour, and object parts, are abstracted layer-by-layer from an image. These abstracted feature representations are then input to the classifier layer to perform classification and regression tasks. Overall, deep learning is an end-to-end learning structure with the minimum human inference, and the parameters of deep learning model are trained jointly.

On the contrary, traditional machine learning performs feature extraction and model construction in a separated manner, and each module is constructed step-by-step. The handcrafted features are firstly extracted by transforming raw data into a different domain (e.g., statistical, frequency, and time-frequency domain) to take the representative information requiring expert domain knowledge. Next, feature selection is performed to improve the relevancy and reduce the spurious redundancy among features before feeding into the machine learning model. Traditional machine learning techniques usually has shallow structures with at most three layers (e.g. input, output, and one hidden layer). Thus, the performance of the constructed model not only relies on the optimization of adopted algorithms (e.g. BP Neural Network, Support Vector Machine, and logistic regression), but also is heavily affected by the handcrafted features. Generally, the feature extraction and selection are time-consuming, and highly depend on domain knowledge.

Therefore, deep learning has distinctive difference with traditional machine learning techniques as illustrated in Table 2. The high level abstract representation in feature learning makes deep learning more flexible and adaptable to data variety. Because the data are abstracted, the diverse data types and sources do not have

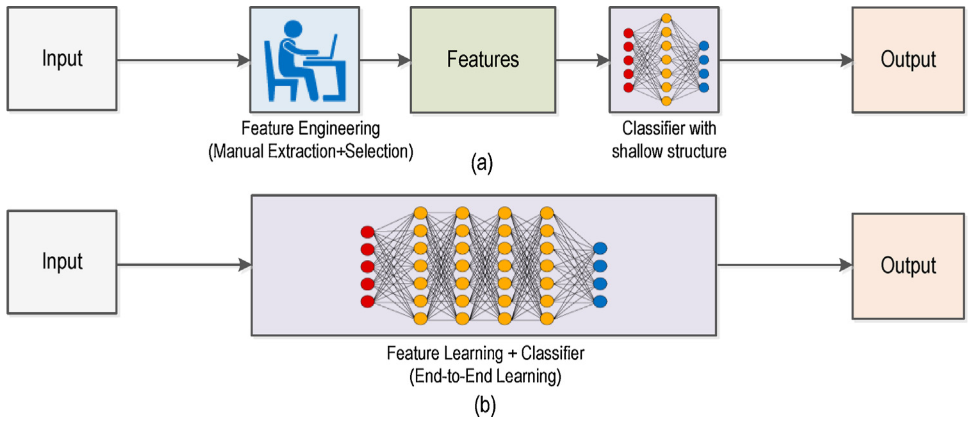

Fig. 2. Comparison between two techniques: a) traditional machine learning, b) deep learning.

Table 2
Comparison between traditional machine learning and deep learning.

	Feature learning	Model construction	Model training
Traditional machine learning	Explicit engineered features extracted with expert domain knowledge.	Use extracted features to construct data-driven model, usually with shallow structures.	Each module is trained step-by-step.
Deep learning	Features are learned by transforming data into abstract representations.	An end-to-end high hierarchical model structure with nonlinear combination of multi-layers.	Parameters are trained jointly.

strong influence on the analysis results. On the other hand, the deep hierarchical structure in deep learning is easier to model the nonlinear relationship using compositional function comparing with the shallow structure which is regarded as a generic function in traditional machine learning. The superiority of deep network had been proven mathematically in [47]. As the size and variety of dataset grow in the big data context, it becomes more difficult to create new, highly relevant features. In the context of big data era in smart manufacturing, the ability to avoid feature engineering is regarded as a great advantage due to the challenges associated with this process.

3. Deep learning for smart manufacturing

With new technologies (e.g. IoT, big data) embraced in smart manufacturing, smart facilities focus on creating manufacturing intelligence that can have a positive impact across the entire organization. The manufacturing today is experiencing an unprecedented increase in available sensory data comprised of different formats, semantics, and structures. Sensory data was collected from different aspects across the manufacturing enterprise, including product line, manufacturing equipment, manufacturing process, labour activity, and environmental conditions. Data modelling and analysis are the essential part of smart manufacturing to handling increased high volume data, as well as supporting real-time data processing [48].

From sensory data to manufacturing intelligence, deep learning has attracted much attention as a breakthrough of computational intelligence. By mining knowledge from aggregated data, deep learning techniques play a key role in automatically learning from data, identifying patterns, and making decisions as shown in Fig. 3. Different levels of data analytics can be produced including descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics. Descriptive analytics aims to summarize what happens by capturing the product's conditions, environment and operational parameters. When the product performance is reduced or the equipment failure happens, diagnostic analytics examine the root cause and report the reason it happens. Predictive analytics utilizes statistical models to make predictions about the possibility of future production or equipment degradation with available historical data. Prescriptive analytics goes beyond by recommending one or more courses of action. Measures can be identified to improve production outcomes or correct the problems, showing the likely outcome of each decision.

With the advanced analytics provided by deep learning, manufacturing is transformed into highly optimized smart facilities. The benefits include reducing operating costs, keeping up with changing consumer demand, improving productivity and reducing downtime, gaining better visibility and extracting more value from the operations for globally competitiveness.

Up to date, various deep learning architectures have been developed and the relevant research topics are fast-growing. To facilitate the investigation of manufacturing intelligence, several typical deep learning architectures are discussed including Convolutional Neural Network, Restricted Boltzmann Machine, Auto Encoder, and Recurrent Neural Network and their variants. The feature learning capability and model construction mechanism were emphasized since these models are the building blocks to construct comprehensive and complex deep learning techniques.

3.1. Convolutional neural network

Convolutional Neural Network (CNN) is a multi-layer feed-forward artificial neural network which is firstly proposed for two-dimensional image processing [36]. It has also been investigated for one-dimensional sequential data analysis including natural language processing and speech recognition recently [49]. In CNN, the feature learning is achieved by alternating and stacking convolutional layers and pooling operations. The convolutional layers convolve with raw input data using multiple local kernel filters and generate invariant local features. The subsequent pooling layers extract the most significant features with a fixed-length

Please cite this article in press as: Wang J, et al. Deep learning for smart manufacturing: Methods and applications. J Manuf Syst (2018), https://doi.org/10.1016/j.jmsy.2018.01.003

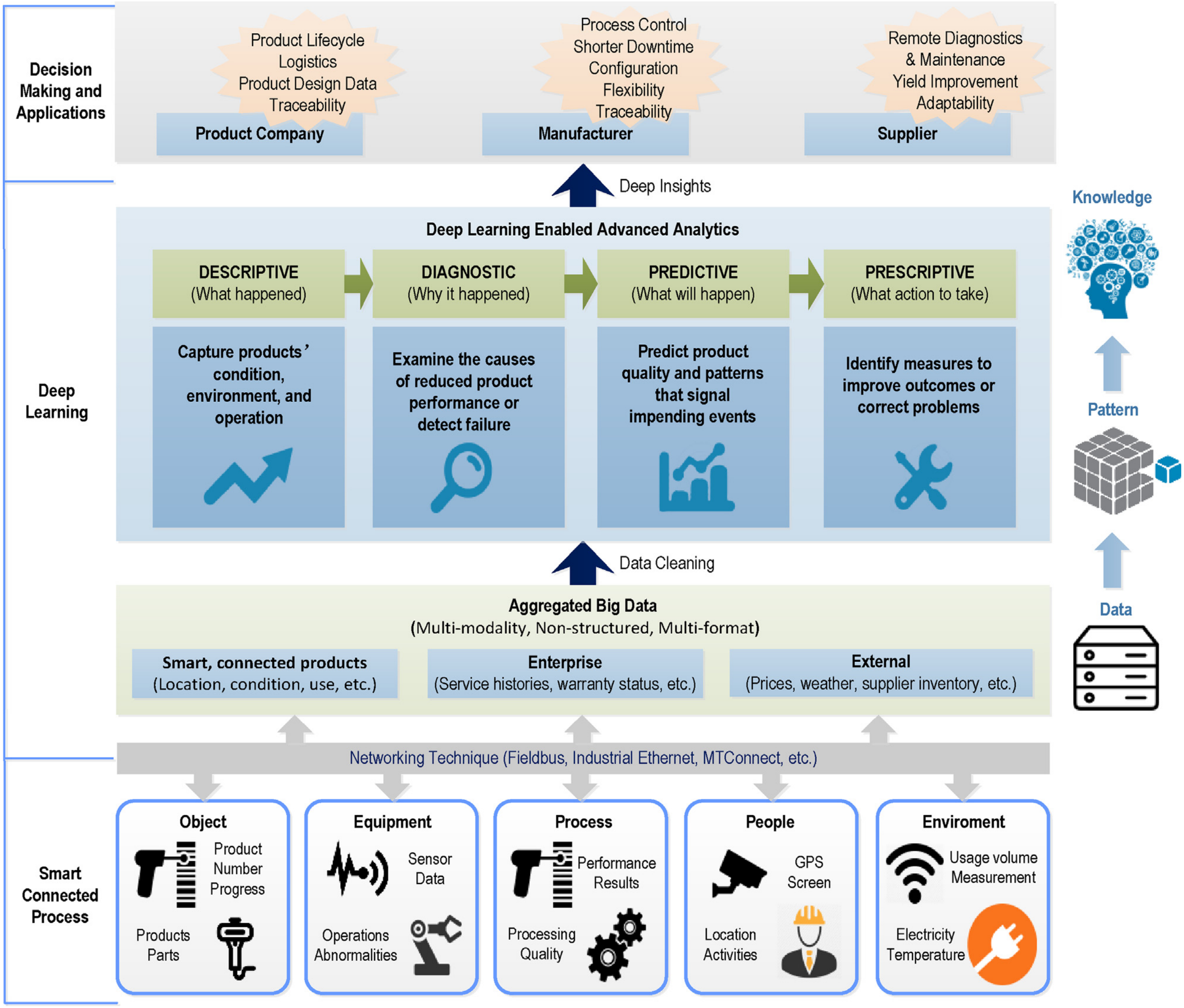

Fig. 3. Deep learning enabled advanced analytics for smart manufacturing.

over sliding windows of the raw input data by pooling operations such as max pooling and average pooling. Max pooling selects the maximum value of one region of the feature map as the most significant feature. Average pooling calculates the mean value of one region and takes it as the pooling value of this region. Max pooling is well suited to extract sparse features, while pooling operation on all samples may not be optimal.

After multi-layer feature learning, fully-connected layers convert a two- dimensional feature map into a one dimensional vector and then feed it into a softmax function for model construction. By stacking convolutional layers, pooling layers, and fully-connected layers, a typical CNN is constructed as shown in Fig. 4. Gradient based backpropagation is usually used to train convolutional neural network by minimizing the minimum mean squared error or cross-entropy loss function. CNN has the advantageous properties including sparse interactions with local connectivity, parameter sharing with reduced numbers, and equivariant representation which is invariant to object locations.

3.2. Restricted Boltzmann machine and its variant

Restricted Boltzmann Machine (RBM) is a two-layer neural network consisting of visible and hidden layer. There exists a symmetric connection between visible and hidden units, but there are no connections between each neuron within the same layer. It is an energy based model in which the visible layer is used to input data while the hidden layer is used to extract features. All hidden nodes are assumed conditionally independent. The weights and offsets of these two layers are tuned over iterations in order to make the output of the visible layer as the approximation of the original input. Finally, the hidden layers are regarded as different representations of the visible layer.

The parameters in hidden layers are treated as the features to characterize the input data to realize data coding and dimension reduction. Then, supervised learning methods such as logistic regression, Naïve Bayes, BP Neural Network, and Support Vector Machine, etc. can be used to implement data classification and regression. RBM takes the advantages of extracting required features from training datasets automatically, which avoids the local minimum value and thus has received a growing number of attentions. Utilizing RBM as the basic learning module, different variant models have been developed [32].

Deep Belief Network (DBN): DBN is constructed by stacking multiple RBMs, where the output of the l^{th} layer in hidden units is used as the input of the $(l+1)^{th}$ layer in visible units. For DBN training, a fast greedy algorithm is usually used to initialize the network and

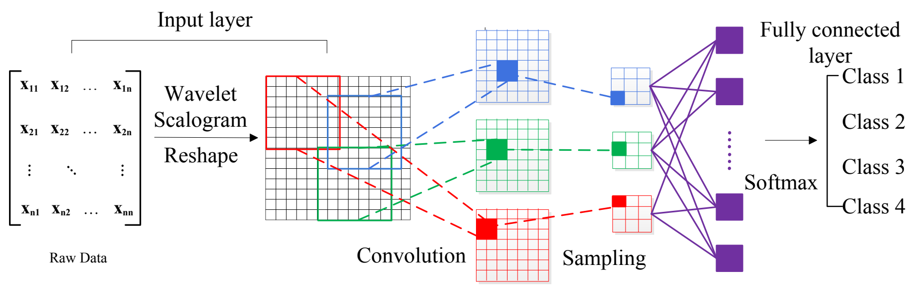

Fig. 4. Architecture of convolutional neural network model.

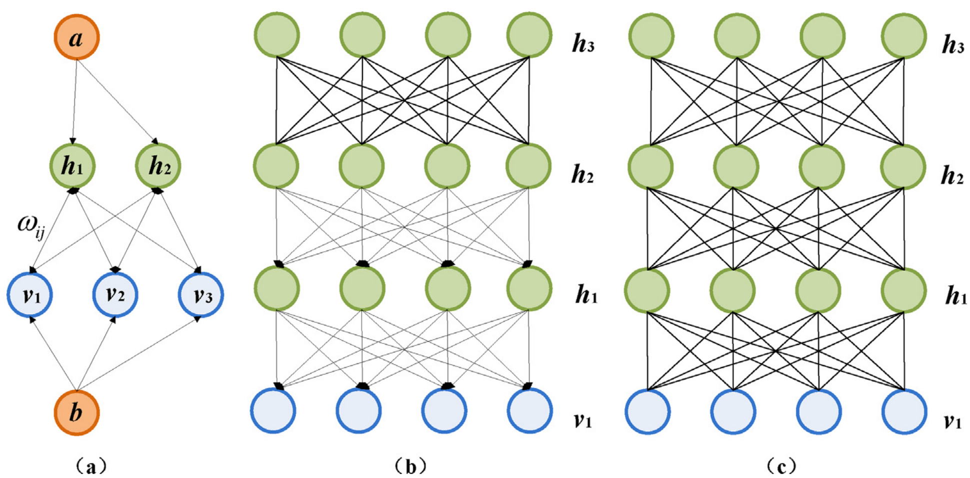

Fig. 5. Architecture of (a) RBM, (b) DBN, and (c) DBM.

the parameters of this deep architecture are then fine-tuned by a contractive wake-sleep algorithm [37]. Bayesian Belief Network is applied to the area which is close to the visible layers, and RBMs are used to the area far from the visible layers. That is to say, the highest two layers are undirected and the other lower layers are directed, as shown in Fig. 5.

Deep Boltzmann Machine (DBM): DBM can be regarded as a deep structured RBMs where hidden units are grouped into a hierarchy of layers. The full connections between two adjacent layers are enabled, but no connection is allowed within a layer or between non-neighbouring layers as shown in Fig. 5. By stacking multi-RBMs, DBM can learn complex structures and construct high-level representation of input data [42]. Compared to DBN, DBM is a fully undirected graphical model while DBN is a mixed directed/undirected one. Accordingly, the DBM model is trained jointly and more computationally expensive. On the contrary, DBN can be trained layer-wisely to be more efficiently.

3.3. Auto encoder and its variants

Auto Encoder (AE) is an unsupervised learning algorithm extracting features from input data without label information needed. It mainly consists of two parts including encoder and decoder as shown in Fig. 6. The encoder can perform data compression especially in dealing input of high dimensionality by mapping input to a hidden layer [33]. The decoder can reconstruct the approximation of input. Suppose the activation function is a linear function and we have less hidden layers than the dimensionality of input data, then the linear Auto Encoder is similar to principle component analysis (PCA). If the input data is highly nonlinear, more hidden layers are required to construct the deep Auto Encoder. Stochastic gradient descent (SGD) is often investigated to calculate parameters and build auto-encoder by minimizing the objective loss function in terms of the least square loss or cross-entropy loss.

Several variants of AE have been developed and listed as follows:

1) Denoising Auto Encoder (DAE): DAE is an extension version of the basic Auto Encoder, which is trained to reconstruct the stochastically corrupted input data by adding isotropic Gaussian noise to x and forcing the hidden layer to discover more robust features [43].
2) Sparse Auto Encoder (SAE): SAE makes the most of the hidden unit's activations close to zero by imposing sparsity constraints on the hidden units, even the number of hidden units is large [40,41].
3) Contractive Auto Encoder (CAE): In order to force the model resistant to small perturbations, CAE encourages learning more robust representations of the input x [50].

3.4. Recurrent neural network and its variants

Compared with traditional neural networks, Recurrent Neural Network (RNN) has unique characteristic of topology connections between the neurons formed directed cycles for sequence data as shown in Fig. 7. Thus, RNN is suitable for feature learning from sequence data. It allows information persists in hidden layers and captures previous states of a few time steps ago. An updated rule is applied in RNN to calculate the hidden states at different time steps. Take the sequential input as a vector, the current hidden state can be calculated by two parts through a same activation function (e.g. *sigmoid* or *tanh* function). The first part is calculated with the input while the second part is obtained from the hidden state at the previous time step. Then, the target output can be calculated with the current hidden state through a softmax function. After processing the whole sequence, the hidden state is the learned representation

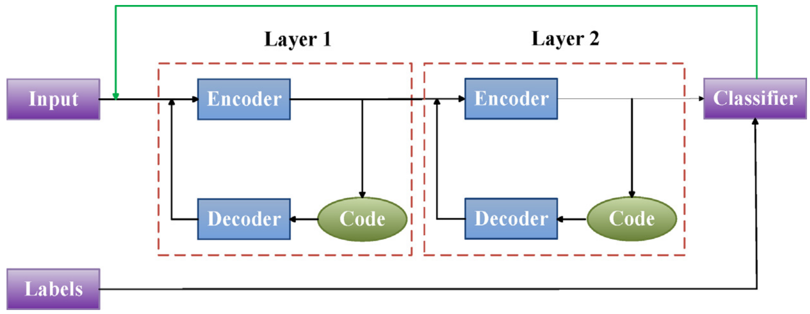

Fig. 6. The architecture of AE.

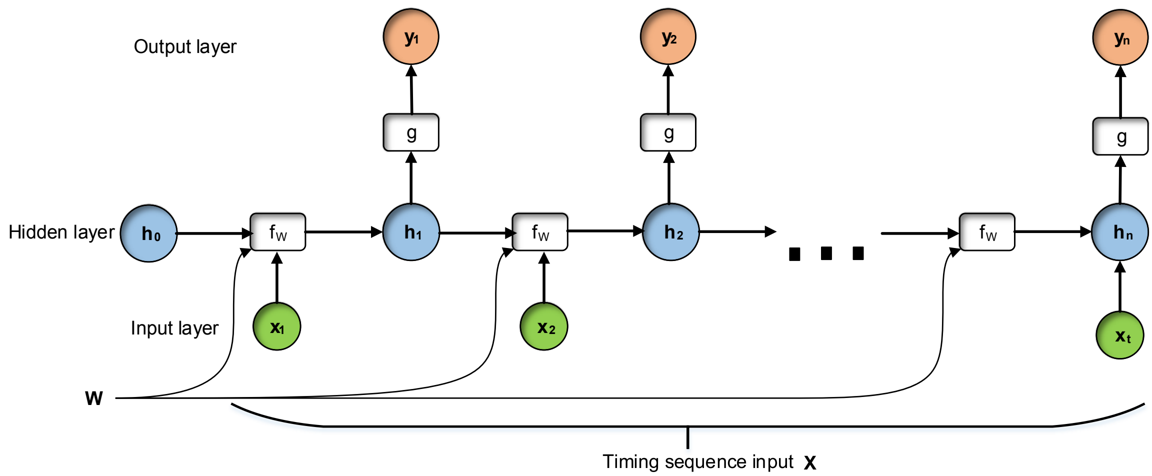

Fig. 7. Architecture of recurrent neural network model.

of the input sequential data and a conventional multilayer perceptron (MLP) is added on top to map the obtained representation to targets.

Different from traditional neural networks, the model training in RNN is performed by Backpropagation Through Time (BPTT). RNN is firstly unrolled according to time and each unrolled time step is considered as an additional layer. Then backpropagation algorithm is applied to calculate gradients. Due to the vanishing/exploding gradient problem using BPTT for model training, RNN cannot capture long-term dependencies. In other words, RNN has difficulty in dealing with long-term sequence data.

A variety of enhancements are proposed to solve these problems, among which long short-term memory (LSTM) is widely investigated for its effectiveness [35]. The most important idea of LSTM is cell state, which allows information flow down with linear interactions. Comparing with single recurrent structure in RNN, the gates including forget gate layer, input gate layer and output gate layer, are used in LSTM to control the cell state. It enables each recurrent unit to adaptively capture long-term dependencies of different time scales.

3.5. Model comparison

With the above illustration, it can be found that CNN and RNN provide complex composition mechanism to learn representation and model construction. The RBM and AE can be used for layer-by-layer pretraining of deep neural network to characterize input data. In these deep learning models, the top layers normally represent the targets. For classification where targets are discrete values, softmax layers are applied. For prediction with continuous targets, linear regression layers are added. According to the dependence on labelled data, DBN, AE and their variants are unsupervised learning or semi supervised learning. CNN, RNN and their variants are supervised learning. The Pros and Cons of these typical deep learning models are presented in Table 3.

Fortunately, a number of typical deep learning packages including open source or commercial software are available to public as summarized in Table 4. They facilitate the investigation of deep learning techniques in different manufacturing scenarios.

4. Applications to smart manufacturing

Computational intelligence is an essential part of smart manufacturing to enable accurate insights for better decision making. Machine learning has been widely investigated in different stages of manufacturing lifecycle covering concept, design [60], evaluation, production, operation, and sustainment [61] as shown in Fig. 8. The applications of data mining in manufacturing engineering are reviewed in [62], covering different categories of production processes, operations, fault detection, maintenance, decision support, and product quality improvement. The evolution and future of manufacturing are reviewed in [63,64], emphasizing the importance of data modelling and analysis in manufacturing intelligence. The application schemes of machine learning in manufacturing are identified as summarized in [65,66]. Smart manufacturing also requires prognostics and health management (PHM) capabilities to meet the current and future needs for efficient and reconfigurable production [67].

Deep learning, as an emerging technique, has been investigated for a wide range of manufacturing systems recently. To give an overview, the applications of state-of-the-art deep learning techniques in manufacturing are discussed in this study, especially in

Please cite this article in press as: Wang J, et al. Deep learning for smart manufacturing: Methods and applications. J Manuf Syst (2018), https://doi.org/10.1016/j.jmsy.2018.01.003

Table 3
Comparison between different deep learning models.

Model	Principle	Pros.	Cons.
CNN	Abstracted features are learned by stacked convolutional and sampling layers.	Reduced parameter number, invariance of shift, scale and distortion	High computational complexity for high hierarchical model training
RBM	Hidden layer describes variable dependencies and connections between input or output layers as representative features.	Robust to ambiguous input and training label is not required in pre-training stage	Time-consuming for joint parameter optimization
AE	Unsupervised feature learning and data dimensionality reduction are achieved through encoding	Irrelevance in the input is eliminated, and meaningful information is preserved	Error propagation layer-by-layer and sparse representations are not guaranteed
RNN	Temporal pattern stored in the recurrent neuros connection and distributed hidden states for time-series data.	Short-term information is retained and temporal correlations are captured in sequence data.	Difficult to train the model and save the long-term dependence

Table 4
A list of deep learning tools.

Tools	Type	Description
Caffe/Caffe2 [51]	Open-source	Feedforward network, and suitable for image processing.
Theano [52]	Open-source	Consist of the most of state-of-the-art neural networks, originated at the University of Montreal Numerical in 2007.
TensorFlow [53]	Open-source	Open source software library for deep neural networks using data flow graphs, developed by Google Brain Team.
Pytorch [54]	Open-source	Deep learning framework widely used by Facebook and Twitter, originally developed at New York University in 2002. Excellent for convnets and rich set of RNNs.
CNTK [55]	Open-source	Microsoft Cognitive Toolkit, and well known in the speech community.
Google Cloud machine learning platform [56]	Commercial	Allow users to build and train machine learning models by using TensorFlow in Google Cloud Platform.
Amazon machine learning [57]	Commercial	Cloud-based service for users to use machine learning technology.
Microsoft Azure [58]	Commercial	Machine learning library
IBM Watson analytics [59]	Commercial	Cloud-based machine learning platform for data exploration, visualization and predictive analytics.

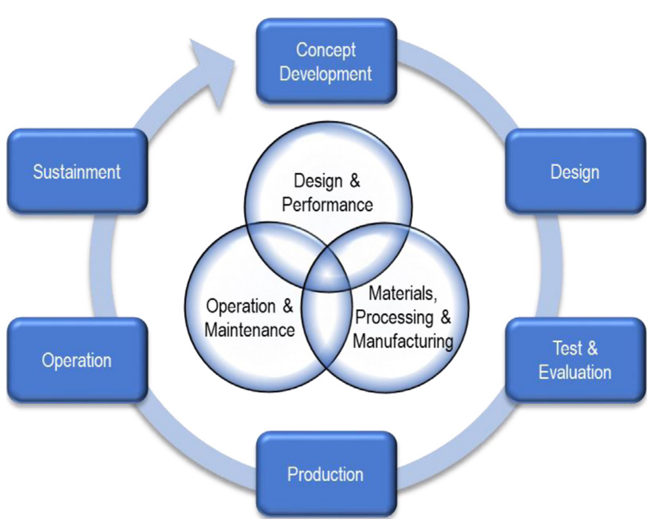

Fig. 8. Typical application scenarios of machine learning in smart manufacturing.

the areas of product quality inspection, fault diagnosis, and defect prognosis, as highlighted in Table 5.

Table 5
A list of deep learning models with applications.

Deep learning model	Application Scenarios	Reference
CNN	Surface integration inspection	[72–75]
	Machinery fault diagnosis	[77–84]
DBN	Machinery fault diagnosis	[85–92]
	Predictive analytics & defect prognosis	[109–112]
AE	Machinery fault diagnosis	[93–103]
RNNs	Predictive analytics & defect prognosis	[104–108]

4.1. Descriptive analytics for product quality inspection

Surface integration inspection is usually inspected employing machine vision and image processing techniques to detect surface defect for enhanced product quality in manufacturing [68]. Traditional machine learning has made remarkable progress and yields reliable results in many cases [69], but different pre-processing approaches including structural-based, statistical-based, filter-based, and model based techniques are needed to extract representative features with expert knowledge [70]. However, flexible configuration in modern manufacturing system could shift production from one product to another quickly. The feature representation may need redesign from scratch for traditional machine learning. Additionally, a new product may present complicated texture patterns or intensity variations, and the surface defects could be in an arbitrary size, orientation and shape. Therefore, manually designed features in traditional machine learning technique may lead to insufficient or unsatisfactory inspection performance in complex surface scenarios or dynamic changing process [71]. To address these challenges, deep learning has been investigated to learn high-level generic features and applied to a wide range of textures or difficult-to-detect defects cases.

Convolutional Neural Network, originally designed for image analysis, is well fit for automated defect identification in surface integration inspection. In [72], a Deep Convolutional Neural Network architecture is designed and the hyper-parameters are optimized based on backpropagation and stochastic gradient descent algorithms. A max-pooling Convolutional Neural Network is presented in [73] to perform feature extraction directly from the pixel representation of steel defect images and shows lower error rates comparing with multi-layer perceptron and support vector machine. The image analysis is studied with convolutional neural network in [74] to automatically inspect dirties, scratches, burrs, and wears on surface parts. The experimental results show that CNN works properly with different types of defects on textured or non-textured surfaces. A generic approach based on CNN is proposed in [75] to extract patch feature and predict defect area via thresholding and segmenting. The results show the pretrained CNN model works well on small dataset with improved accuracy for automated surface inspection system.

4.2. Diagnostic analytics for fault assessment

Manufacturing systems are usually subject to failures caused by degradation or abnormal operating conditions, leading to excessive load, defection, fracture, overheating, corrosion, and wear. The failure may incur higher operating costs, lower productivity, more disqualified part waste, and even unexpected downtime. In order to implement smart manufacturing, it is crucial for a smart factory to monitor machinery conditions, identify the incipient defects, diagnose the root cause of failures, and then incorporate the information into manufacturing production and control [75].

With aggregated data from smart sensory and automation systems, more and more deep learning techniques have been widely investigated for machinery fault diagnosis and classification [76]. Convolutional Neural Network integrates feature learning and defect diagnosis in one model, and has been used in many aspects, such as bearing [77–80], gearbox [81,82], wind generator [83], and rotor [84], etc. Since CNN was originally developed for image analysis, different approaches are investigated to construct two dimensional input from time series data. The frequency spectrum of multi-channel vibration data is also investigated in [77] to fit the model requirement. In [75], permutation is performed by transforming time series data into a matrix, and then normalized as image. In [84], time frequency spectrum of vibration signal by wavelet transform is used as image input of a CNN model. Deep Belief Network has fast inference as well as the advantage of encoding high order network structures by stacking multiple Restricted Boltzmann Machines. It has been investigated for fault diagnosis of aircraft engine [85], chemical process [86], reciprocating compressor [87], rolling element bearing [88,89], high speed train [90,91], and wind turbine [92]. The input of a DBM model is usually the preprocessed features by Teager-Kaiser energy operator or wavelet transform rather than raw data. Then the DBM model is constructed following a supervised layer-by-layer learning process.

Auto Encoder has been investigated for unsupervised feature learning, and the learned features are then fed into a traditional machine learning model for model training and classification. In [93], a five-layer deep neural network is presented using the features learned by an Auto Encoder for planetary gearbox diagnosis under various operating conditions. Different variants are also investigated including sparse Auto Encoder [94], stacked denoising Auto Encoder [95], and Contractive Auto Encoder [96], etc. Sparse Auto Encoder is investigated in [97] to learn the features of motor current signal, and partial corruption is performed on the input to improve the robustness of feature representation. A multi-layered extreme learning machine based Auto Encoder is investigated in [98] to learn feature representations for wind turbine fault classification. In [99], a continuous sparse Auto Encoder is presented by adding Gaussian stochastic unit into an activation function to extract nonlinear features of the input data. To improve diagnostic analytics, the comprehensive deep learning models are also developed. In [100], a sparse filtering based deep neural network model is investigated for unsupervised features learning. Synthesized deep learning models are discussed in [101,102] for signal denoising and fused feature extraction. Different deep learning models including Deep Boltzmann Machine, Deep Belief Network, and Stacked Auto Encoder with different preprocessing schemes are comparatively studied in [103] for rolling bearing fault diagnosis. The results show that stacked Auto Encoder performs the best. From the above literatures, it is concluded that the deep learning models outperform traditional machine learning techniques with engineered features such as support vector machine, and BP Neural Network in terms of classification accuracy.

4.3. Predictive analytics for defect prognosis

In order to increase manufacturing productivity while reducing maintenance cost, it is crucial to develop and implement an intelligent maintenance strategy that allows manufacturers to determine the condition of in-service systems in order to predict when maintenance should be performed. The temporal behaviour in the historical data is important for prediction, and deep recurrent neural network has demonstrated its capability to model temporal pattern. Recently, a general recurrent neural network, named long short term memory, has been investigated to predict defect propagation and estimate remaining useful life (RUL) of mechanical systems or components. In [104], a competitive learning-based RNN has been proposed for long-term prognosis of rolling bearing health status. In [105], a new local feature-based gated recurrent unit network has been proposed to learn the representation of the sequence of local features and the proposed method is verified on three real machine health monitoring tasks. In [106], an integrated approach of CNN and bi-directional LSTM is presented for machining tool wear prediction, in which CNN is used to extract local features from sequential signals and bi-directional LSTM to capture long-term dependence for prediction. Vanilla LSTM is investigated in [107] to estimate the remaining useful life of an aircraft turbofan engine under complex operating conditions and strong background noise, and the experimental results confirm that Vanilla LSTM provides good prediction accuracy. A stacked LSTM network enables the learning of higher level temporal features, and has been presented for anomaly prediction of space shuttle and engine [108].

Deep Belief Network, as the feature learning approach in regression models, has also been investigated for predictive analytics. In [109], Deep Belief Network is investigated to model the complex relationship between material removal rate and chemical mechanical polishing process parameters in semiconductor manufacturing. An integrative approach of Deep Belief Network and particle filter is presented in [110] for the RUL prediction of a ceramic bearing. By aggregating the output of ensemble DBNs, Support Vector Regression model is investigated to predict electricity load demand [111]. To predict the resource request in cloud computing, DBN is proposed in [112] to optimize job schedule and balance the computational load.

5. Discussions and outlook

As the evolution of smart manufacturing, more and more machineries are equipped with smart sensors and meshed with Internet of Things. Currently, most companies do not know what to do with the data they have, and they lack software and modelling to interpret and analyse them. On the other hand, manufacturers need practical guidance to improve their processes and products, while the academics develop up-to-date artificial intelligence models without considering how they will be applied in practice. As manufacturing process becomes more complex, more difficulty comes along to clear the data and formulate the right problems to model. Five gaps are identified in smart manufacturing innovation including adopted strategies, improved data collection, use and sharing, predictive model design, generalized predictive models, and connected factories and control processes [48].

To meet the high demand of advanced analytics in smart manufacturing, deep learning with feature learning and deep network offers great potential and shows advantageous properties. To handle overwhelming data characterized by high-volume, high-velocity and high-variety, there are still some challenges associated with manufacturing industry to adopt, implement, and deploy deep learning for real-world applications. To address the challenges, the future development trends of deep learning for smart manufacturing are discussed in terms of data matter, model selection, model visualization, generic model, and incremental learning.

5.1. Data matter

A common presumption in machine learning is that algorithms can learn better with more data, and thus the performance of deep learning model heavily depends on the scale and quality of datasets. So far deep learning shows the effectiveness when it is applied to limited types of data (e.g. images, speech, and vibration, etc.) and well-defined tasks. Multi-sensory has been instrumented to capture data at all stages of a product's life. Deep learning algorithm may be infeasible to directly handle such high dimensional, multi-modality, and non-structured data, and even susceptible to the curse of dimensionality. Extracting the relevant data to reduce the size and applying appropriating task-specific regularization term may improve the performance of deep learning. On the other hand, the class imbalance problem is another challenge. The class follows a highly-skewed distribution in real life, representing most data samples belong to few categories. For example, the dataset of surface defects is normally too small and costly to collect. The ratio of good to bad parts is highly imbalanced ranging from 9:1 to even less than one million. Thus, it is difficult to apply standard classification techniques to differentiating good parts from scraps. Appropriate measures such as class resampling, cost-sensitive training, and integration of boot strapping may be necessary for deep learning model to address class imbalance issues [113].

5.2. Model selection

Different problems are tacked with specialized models. A number of deep learning models are available at present. Considering the complexity in manufacturing process, model selection has become a major challenge. Generally, some empirical criteria can be adopted when applying deep learning models. Firstly, supervised or unsupervised deep learning algorithms can be chosen depending on the available data set in hand. Supervised deep learning algorithms are appropriate to dealing with data rich but knowledge sparse problems, namely labelled data are available. If there is no expert knowledge to solve the problem, unsupervised deep learning algorithms might be suitable. Secondly, since one algorithm has its strength and weakness, the general applicability of the selected algorithm should be considered [16,19].

5.3. Model visualization

The analytics solutions of deep learning need to be understood by manufacturing engineers. Otherwise, the generated recommendations and decisions may be ignored. Due to the model complexity behind, deep neural network is usually regarded as a black-box model. It is hard to explain the internal computation mechanism or interpret the abstract feature representation physically. Visualization of the learned features and model architecture may offer some insights, and thus facilitate the construction and configuration of deep neural network models for complex problems. On the other hand, the engineered features by domain expertise have demonstrated its effectiveness. Complementing the abstract features with engineered features by visualization and fusion may contribute a more effective model. Some visualization techniques have been proposed including t-SNE model [114] for high dimensional data visualization, and visualization of activations produced by each layer of deep neural network via regularized optimization [115].

5.4. Generic model

The application of deep learning is not bonded to specific machines, thus deep learning models can be a general solution to address manufacturing intelligence problems. Although many improvements including dropout and activation function have been investigated to handle large datasets, it is still challenging to build a high hierarchical model with multi-layers for complex problems. Both the architecture design and hyper-parameter optimizations have significant impacts on the performance of deep learning models. One way to improve the architecture is to increase its width or depth. The determination of the optimal hyper-parameters relies on appropriate optimization algorithms in a computationally efficient way [116]. On the other hand, parallel implementation of deep learning algorithms can be applied to large scale and real time analytics using parallel computing, graphic processing unit (GPU), and Hadoop technology. On the top, setting up the correct problem to be optimized, and choosing appropriate models should be the basis of developing generic models.

5.5. Incremental learning

The deep learning algorithms are not fundamentally built to learn incrementally and are therefore susceptible to the data velocity issues. For a new problem setup, deep learning may need to rebuild the model from scratch and the existing knowledge may be difficult to utilize. Additionally, the data in the new scenarios is also an issue. It is necessary to enable deep learning with incremental learning capabilities. Transfer learning aims to extract the knowledge from one source task and then applies the learned knowledge to a different but related task [117]. It could employ the pre-trained deep learning model from a relevant task for model initialization and fine-tuning to enable knowledge reuse and updating as transferred deep learning. Some previous works focusing on transferred feature extraction/dimensionality reduction have been done. A maximum mean discrepancy (MMD) measure evaluating the discrepancy between source and target domains is added into the target function of deep neural networks [118]. Thus, transferred deep learning is meaningful and promising for smart manufacturing to enable knowledge updating and intelligence upgrading.

6. Conclusions

Deep learning provides advanced analytics and offers great potentials to smart manufacturing in the age of big data. By unlocking the unprecedented amount of data into actionable and insightful information, deep learning gives decision-makers new visibility into their operations, as well as real-time performance measures and costs. To facilitate advanced analytics, a comprehensive overview of deep learning techniques is presented with the applications to smart manufacturing. Four typical deep learning models including Convolutional Neural Network, Restricted Boltzmann Machine, Auto Encoder, and Recurrent Neural Network are discussed in detail. The emerging research effort of deep learning in applications of manufacturing is also summarized. Despite of the promising results reported so far, there are still some limitations and significant challenges for further exploration.

As the evolution of computing resources (e.g., cloud computing [119–124], fog computing [125,126], etc.), computational intelligence including deep learning may be push into cloud, enabling more convenient and on-demand computing services for smart manufacturing.

Acknowledgments

This research acknowledges the financial support provided by National Key Research and Development Program of China (No. 2016YFC0802103), National Science foundation of China (No. 51504274), and Science Foundation of China University of Petroleum, Beijing (No. 2462014YJRC039).

References

[1] Putnik G, Sluga A, ElMaraghy H, Teti R, Koren Y, Tolio T, et al. Scalability in manufacturing systems design and operation: state-of-the-art and future developments roadmap. CIRP Ann Manuf Technol 2013;62(2):751–74.
[2] Lee YT, Kumaraguru S, Jain S, Hatim Q, Robinson S, Helu M, et al. A classification scheme for smart manufacturing systems' performance metrics. Smart Sustain Manuf Syst 2017;1(1):52–74.
[3] Hu T, Li P, Zhang C, Liu R. Design and application of a real-time industrial Ethernet protocol under Linux using RTAI. Int J Comput Integr Manuf 2013;26(5):429–39.
[4] Ye Y, Hu T, Zhang C, Luo W. Design and development of a CNC machining process knowledge base using cloud technology. Int J Adv Manuf Technol 2016:1–13.
[5] Tao F, Qi Q. New IT driven service-oriented smart manufacturing: framework and characteristics. IEEE Trans Syst Man Cybern Syst 2017;99:1–11.
[6] Ang J, Goh C, Saldivar A, Li Y. Energy-efficient through-life smart design, manufacturing and operation of ships in an industry 4.0 environment. Energies 2017;10(5):610.
[7] Huang Z, Hu T, Peng C, Hou M, Zhang C. Research and development of industrial real-time Ethernet performance testing system used for CNC system. Int J Adv Manuf Technol 2016;83(5–8):1199–207.
[8] Lalanda P, Morand D, Chollet S. Autonomic mediation middleware for smart manufacturing. IEEE Internet Comput 2017;21(1):32–9.
[9] Smart Manufacturing Coalition. Manufacturing growth continues despite uncertain economy, according to ASQ outlook survey; 2013. https://smartmanufacturingcoalition.org/sites/default/files/12.16.13_manufacturing_outlook_survey.pdf. [Accessed 10 September 2017].
[10] Wang L, Törngren M, Onori M. Current status and advancement of cyber-physical systems in manufacturing. J Manuf Syst 2015;37:517–27.
[11] Wang P, Gao RX, Fan Z. Cloud computing for cloud manufacturing: benefits and limitations. J Manuf Sci Eng 2015;137:1–10.
[12] Lu Y, Xu X, Xu J. Development of a hybrid manufacturing cloud. J Manuf Syst 2014;33(4):551–66.
[13] Wu D, Rosen DW, Schaefer D. Cloud-based design and manufacturing: status and promise. Comput Aided Des 2015;59:1–14.
[14] Choudhary AK, Harding JA, Tiwari MK. Data mining in manufacturing: a review based on the kind of knowledge. J Intell Manuf 2009;20(5):501–21.
[15] Lade P, Ghosh R, Srinivasan S. Manufacturing analytics and industrial internet of things. IEEE Intell Syst 2017;32(3):74–9.
[16] Monostori L, Márkus A, Brussel HV, Westkämpfer E. Machine learning approaches to manufacturing. CIRP Ann Manuf Technol 1996;45(2):675–712.
[17] Teti R, Jemielniak K, O'Donnell G, Dornfeld D. Advanced monitoring of machining operations. CIRP Ann Manuf Technol 2010;59(2):717–39.
[18] Helu M, Libes D, Lubell J, Lyons K, Morris K. Enabling smart manufacturing technologies for decision-making support. Proceedings of the ASME international design engineering technical conferences and computers and information in engineering conference (IDETC/CIE) 2016:1–10.
[19] Wuest T, Weimer D, Irgens C, Klaus DT. Machine learning in manufacturing: advantages, challenges, and applications. Prod Manuf Res 2016;4(1):23–45.
[20] Gao R, Wang L, Teti R, Dornfeld D, Kumara S, Helu M, et al. Cloud-enabled prognosis for manufacturing. CIRP Ann Manuf Technol 2015;64(2):749–72.
[21] Wu D, Jennings C, Terpenny J, Gao RX, Kumara S. A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests. J Manuf Sci Eng 2017;139(7):1–10.
[22] Gartner's Top10 strategic technology trendsfor; 2017. http://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/. [Accessed 13 August 2017].
[23] Mcculloch WS, Pitts WH. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 1943;5(4):115–33.
[24] Samuel AL. Some studies in machine learning using the game of checkers II—recent progress. Annu Rev Autom Program 2010;44(1–2):206–26.
[25] Rosenblatt F. Perceptron simulation experiments. Proc IRE 1960;48(3):301–9.
[26] Widrow B, Hoff ME. Adaptive switching circuits. Cambridge: MIT Press; 1960.
[27] Minsky M, Perceptrons Papert S. Am J Psychol 1988;84(3):449–52.
[28] Tank DW, Hopfield JJ. Neural computation by concentrating information in time. Proc Natl Acad Sci USA 1987;84(7):1896.
[29] Werbos PJ. Backpropagation through time: what it does and how to do it. Proc IEEE 1990;78(10):1550–60.
[30] Sussmann HJ. Learning algorithms for Boltzmann machines. 27th IEEE conference on decision and control 1988;1:786–91.
[31] Vapnik VN. An overview of statistical learning theory. IEEE Trans Neural Netw 1998;10(5):988–99.
[32] Smolensky P. Information processing in dynamical systems: foundations of harmony theory. Parallel distributed processing: explorations in the microstructure of cognition. Cambridge: MIT Press; 1986.
[33] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986;323(6088):533–6.
[34] Hihi SE, Hc-J MQ, Bengio Y. Hierarchical recurrent neural networks for Long-Term dependencies. Adv Neural Inf Process Syst 1995;8:493–9.
[35] Hochreiter S, Schmidhuber J. Long short-Term memory. Neural Comput 1997;9(8):1735.
[36] Lécun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86(11):2278–324.
[37] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural network. Science 2006;313(5786):504–7.
[38] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Comput 2014;18(7):1527–54.
[39] Deng L, Seltzer M, Yu D, Acero A, Mohamed A, Hinton GE. Binary coding of speech spectrograms using a deep auto-encoder. Proceedings of 11th annual conference of the international speech communication association 2010;3:1692–5.
[40] Schölkopf B, Platt J, Hofmann T. Efficient learning of sparse representations with an energy-Based model. Proceedings of advances in neural information processingsystems 2006:1137–44.
[41] Ranzato MA, Boureau YL, Lecun Y. Sparse feature learning for deep belief networks. Proceedings of international conference on neural information processing systems 2007;20:1185–92.
[42] Salakhutdinov RR, Hinton GE. Deep Boltzmann machines. J Mach Learn Res 2009;5(2):1967–2006.
[43] Larochelle H, Lajoie I, Bengio Y, Manzagol PA. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 2010;11(12):3371–408.
[44] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolution neural networks. International conference on neural information processing systems 2012;25:1097–105.
[45] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. Int Conf Neural Inf Process Syst 2014;3:2672–80.
[46] Wang Y, Huang M, Zhao L, Zhu X. Attention-based LSTM for aspect-level sentiment classification. Proceedings of conference on empirical methods in natural language processing 2016:606–15.
[47] Poggio T, Smale S. The mathematics of learning: dealing with data. Not Am Math Soc 2003;50(5):537–44.
[48] Kusiak A. Smart manufacturing must embrace big data. Nature 2017;544(7648):23–5.
[49] Ince T, Kiranyaz S, Eren L, Askar M, Gabbouj M. Real-time motor fault detection by 1-D convolution neural networks. IEEE Trans Ind Electron 2016;63(11):7067–75.
[50] Hassanzadeh A, Kaarna A, Kauranne T. Unsupervised multi-manifold classification of hyperspectral remote sensing images with contractive Autoencoder. Neurocomputing 2017;257:67–78.
[51] Caffe2. https://caffe2.ai/. 2017 [Accessed 20 October 2017].
[52] Theano. http://deeplearning.net/software/theano/index.html#. 2017 [Accessed 20 October 2017].
[53] Google TensorFlow. https://www.tensorflow.org/. 2017 [Accessed 20 October 2017].
[54] Pytorch. http://pytorch.org/. 2017 [Accessed 20 October 2017].
[55] Microsoft Cognitive Toolkit. https://www.microsoft.com/en-us/cognitive-toolkit. 2017 [Accessed 20 October 2017].
[56] Google Google cloud machine learning. https://cloud.google.com/products/machine-learning/. 2017 [Accessed 20 October 2017].
[57] Amazon Web Service. Amazon AI, https://aws.amazon.com/amazon-ai/. 2017 [Accessed 20 October 2017].

[58] Microsoft Microsoft Azure; 2017. https://azure.microsoft.com/en-us/services/machine-learning-studio/. [Accessed 20 October 2017].
[59] IBM. IBM Watson ecosystem program; 2017. http://m.ibm.com/http/www-03. ibm.com/innovation/us/watson/. [Accessed 20 October 2017].
[60] Zhang W, Jia MP, Zhu L, Yan X. Comprehensive overview on computational intelligence techniques for machinery condition monitoring and fault diagnosis. Chin J Mech Eng 2017;30(4):1–14.
[61] Lee J, Lapira E, Bagheri B, Kao H. Recent advances and trends in predictive manufacturing systems in big data environment. Manuf Lettersm 2013;1(1):38–41.
[62] Harding JA, Shahbaz M, Srinivas Kusiak A. Data mining in manufacturing: a review. J Manuf Sci Eng 2006;128:969–76.
[63] Esmaeilian B, Behdad S, Wang B. The evolution and future of manufacturing: a review. J Manuf Syst 2016;39:79–100.
[64] Kang HS, Ju YL, Choi SS, Kim H, Park JH. Smart manufacturing: past research, present findings, and future directions. Int J Precision Eng Manuf Green Technol 2016;3(1):111–28.
[65] Hazen BT, Boone CA, Ezell JD, Jones-Farmer LA. Data quality for data science, predictive analytics, and big data in supply chain management: an introduction to the problem and suggestions for research and applications. Int J Prod Econ 2014;154(4):72–80.
[66] Shin SJ, Woo J, Rachuri S. Predictive analytics model for power consumption in manufacturing. Procedia CIRP 2014;15:153–8.
[67] Vogl GW, Weiss BA, Helu M. A review of diagnostic and prognostic capabilities and best practice for manufacturing. J Intell Manuf 2016:1–17.
[68] Xie X. A review of recent advances in surface defect detection using texture analysis techniques. Elcvia Electron Lett ComputVision Image Anal 2008;7(3):1–22.
[69] Neogi N, Mohanta DK, Dutta PK. Review of vision-based steel surface inspection systems. EURASIP J Image Video Process 2014;1:1–19.
[70] Pernkopf F, O'Leary P. Visual inspection of machined metallic high-precision surfaces. EURASIP J Adv Signal Process 2002;7:667–8.
[71] Scholz-Reiter B, Weimer D, Thamer H. Automated surface inspection of cold-formed micro-parts. CIRP Ann Manuf Technol 2012;61(1):531–4.
[72] Weimer D, Scholz-Reiter B, Shpitalni M. Design of deep convolution neural network architectures for automated feature extraction in industrial inspection. CIRP Ann Manuf Technol 2016;65(1):417–20.
[73] Ren R, Hung T, Tan KC. A generic deep-learning-based approach for automated surface inspection. IEEE Trans Cybern 2017;99:1–12.
[74] Masci J, Meier U, Ciresan D, Schmidhuber J, Fricout G, Mittal A. Steel defect classification with max-pooling convolution neural networks. IEEE international joint conference on neural networks (IJCNN) 2012;20:1–6.
[75] Park JK, Kwon BK, Park JH, Kang DJ. Machine learning-based imaging system for surface defect inspection. Int J Precision Eng Manuf Green Technol 2016;3(3):303–10.
[76] Zhao R, Yan R, Chen Z, Chen Z, Mao K, Wang P, et al. Deep learning and its applications to machine health monitoring: a survey; 2016. https://arxiv.org/pdf/1612.07640.pdf. [Accessed 20 October 2017].
[77] Janssens O, Slavkovikj V, Vervisch B, Stockman K, Loccufier M, Verstockt S, et al. Convolution neural network based fault detection for rotating machinery. J Sound Vib 2016;377:331–45.
[78] Lu C, Wang Z, Zhou B. Intelligent fault diagnosis of rolling bearing using hierarchical convolution network based health state classification. Adv Eng Inf 2017;32:139–51.
[79] Guo X, Chen L, Shen C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement 2016;93:490–502.
[80] Verstraete D, Droguett E, Meruance V, Modarres M, Ferrada A. Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. Shock Vib 2017:1–29.
[81] Chen ZQ, Li C, Sanchez RV. Gearbox fault identification and classification with convolution neural networks. Shock Vib 2015;2:1–10.
[82] Wang P, Ananya Yan R, Gao RX. Virtualization and deep recognition for system fault classification. J Manuf Syst 2017;44:310–6.
[83] Dong H, Yang L, Li H. Small fault diagnosis of front-end speed controlled wind generator based on deep learning. WESEAS Trans Circuits Syst 2016;15:64–72.
[84] Wang J, Zhuang J, Duan L, Cheng W. A multi-scale convolution neural network for featureless fault diagnosis. Proceedings of 2016 international symposium on flexible automation 2016:65–70.
[85] Tamilselvan P, Wang P. Failure diagnosis using deep belief learning based health state classification. Reliab Eng Syst Saf 2013;115(7):124–35.
[86] Yu H, Khan F, Garaniya V. Nonlinear Gaussian belief network based fault diagnosis for industrial processes. J Process Control 2015;35:178–200.
[87] Tran VT, Althobiani F, Ball A. An approach to fault diagnosis of reciprocating compressor valves using teager–kaiser energy operator and deep belief networks. Expert Syst Appl 2014;41(9):4113–22.
[88] Shao H, Jiang H, Zhang X, Niu M. Rolling bearing fault diagnosis using an optimization deep belief network. Meas Sci Technol 2015;26(11):1–17.
[89] Gan M, Wang C, Zhu C. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. Mech Syst Signal Process 2016;72–73(2):92–104.
[90] Yin J, Zhao W. Fault diagnosis network design for vehicle on-board equipments of high speed railway: a deep learning approach. Eng Appl Artif Intell 2016;56:250–9.
[91] Xie H, Yang Y, Wang H, Li T, Jin W. Fault diagnosis in high-speed train running gears with improved deep belief networks. J Comput Inf Syst 2015;11(24):7723–30.
[92] Li C, Sanchez RV, Zurita G, Cerrada M, Cabrera D, Vasquez RE. Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. Mech Syst Signal Process 2016;7:6–77, 283–293.
[93] Jia F, Lei Y, Lin J, Zhou X, Lu N. Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech Syst Signal Process 2016;7:2–73, 303–315.
[94] Guo J, Xie X, Bie R, Sun L. Structural health monitoring by using a sparse coding –based deep learning algorithm with wireless sensor networks. Pers Ubiquit Comput 2014;18:1977–87.
[95] Lu C, Wang Z, Qin W, Ma J. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. Signal Process 2017;130:377–88.
[96] Shao H, Jiang H, Wang F, Zhao H. An enhancement deep feature fusion method for rotating machinery fault diagnosis. Knowl Based Syst 2017;119:200–20.
[97] Sun W, Shao S, Zhao R, Yan R, Zhang X, Chen X. A sparse auto-encoder-based deep neural network approach for induction motor faults classification. Measurement 2016;89:171–8.
[98] Yang Z, Wang X, Zhong J. Representational learning for fault diagnosis of wind turbine equipment: a multi-layered extreme learning machines approach. Energies 2016;9(379):1–17.
[99] Wang L, Zhao X, Pei J, Tang G. Transformer fault diagnosis using continuous sparse autoencoder. SpingerPlus 2016;5(448):1–13.
[100] Lei Y, Jia F, Lin J, Xing S, Ding SX. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. IEEE Trans Ind Electron 2016;63(5):3137–47.
[101] Li C, Sanchez RV, Zutita G, Cerrada M, Cabrera D, Vasquez RE. Multimodel deep support vector classification with homologous features and its application to gearbox fault diagnosis. Neurocomputing 2015;168:119–27.
[102] Guo X, Shen C, Chen L. Deep fault recognizer: an integrated model to denoise and extract features for fault diagnosis in rotating machinery. Appl Sci 2017;7(41):1–17.
[103] Chen Z, Deng S, Chen X, Li C, Sanchez RV, Qin H. Deep neural network-based rolling bearing fault diagnosis. Microelectron Reliab 2017;75:327–33.
[104] Malhi A, Yan R, Gao RX. Prognosis of defect propagation based on recurrent neural networks. IEEE Trans Instrum Meas 2011;60(3):703–11.
[105] Zhao R, Wang D, Yan R, Mao K, Shen F, Wang J. Machine health monitoring using local feature-based gated recurrent unit networks. IEEE Transa Ind Electron 2018;65(2):1539–48.
[106] Zhao R, Yan R, Wang J, Mao K. Learning to monitor machine health with convolution bi-directional LSTM networks. Sensors 2017;17(273):1–18.
[107] Wu Y, Yuan M, Dong S, Lin L, Liu Y. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. Neurocomputing 2017;226(5):853–60.
[108] Malhotra P, Vig L, Shroff G, Agarwal P. Long short term memory networks for anomaly detection in time series. In: Proceeding of European symposium on artificial neural networks, computational intelligence, and machine learning. 2015. p. 89–94.
[109] Wang P, Gao RX, Yan R. A deep learning-based approach to material removal rate prediction in polishing. CIRP Ann Manuf Technol 2017;66:429–32.
[110] Deutsch J, He M, He D. Remaining useful life prediction of hybrid ceramic bearings using an integrated deep learning and particle filter approach. Appl Sci 2017;7(649):1–17.
[111] Qiu X, Zhang L, Ren Y, Suganthan PN, Amaratunga G. Ensemble deep learning for regression and time series forecasting. IEEE symposium series on computational intelligence 2014:1–6.
[112] Zhang W, Duan P, Yang LT, Xia F, Li Z, Lu Q, et al. Resource requests prediction in the cloud computing environment with a deep belief network. Software Pract Exp 2017;47(3):473–88.
[113] Khan SH, Hayat M, Bennamoun M, Sohel FA, Tognari R. Cost-sensitive learning of deep feature representations from imbalanced data. IEEE Trans Neural Networks Learn Syst 2017;99:1–15.
[114] Maaten LVD, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9(2605):2579–605.
[115] Yu D, Yao K, Su H, Li G, Seide F. KL-divergence regularized deep neural network adaptation for improved large vocabulary speech recognition. IEEE international conference on acoustics, speech and signal processing 2013:7893–7.
[116] Vig E, Dorr M, Cox D. Large-scale optimization of hierarchical features for saliency prediction in natural images. IEEE computer vision and pattern recognition 2014:2798–805.
[117] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010;22(10):1345–59.
[118] Dziugaite GK, Roy DM, Ghahramani Z. Training generative neural networks via Maximum Mean Discrepancy optimization. Proceedings of the 31st conference on uncertainty in artificial intelligence 2015:258–67.
[119] Mell P, Grance T. The NIST definition of cloud computing. Commun ACM 2009;53(6), 50–50.
[120] Davis J, Edgar T, Porter J, Bernaden J, Sarli M. Smart manufacturing, manufacturing intelligence and demand-dynamic performance. Comput Chem Eng 2012;47(12):145–56.
[121] Lee J, Kao HA, Yang S. Service innovation and smart analytics for industry 4: 0 and big data environment. Procedia CIRP 2014;16:3–8.

Please cite this article in press as: Wang J, et al. Deep learning for smart manufacturing: Methods and applications. J Manuf Syst (2018), https://doi.org/10.1016/j.jmsy.2018.01.003

[122] Lee J, Lapira E, Bagheri B, Kao H. Recent advances and trends in predictive manufacturing systems in big data environment. Manuf Lett 2013;1(1):38–41.
[123] Chen CLP, Zhang CY. Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. Inf Sci 2014;275(11):314–47.
[124] Meziane F, Vadera S, Kobbacy K, Proudlove N. Intelligent systems in manufacturing: current developments and future prospects. Integr Manuf Syst 2000;11(4):218–38.
[125] O'Donovan P, Leahy K, Bruton K, O'Sullivan D. Big data in manufacturing: a systematic mapping study. J Big Data 2015;2(1):1–22.
[126] Wu D, Liu S, Zhang L, Terpenny J, Gao RX, Kurfess T, et al. A fog computing-based framework for process monitoring and prognosis in cyber-manufacturing. J Manuf Syst 2017;43(1):25–34.