Engineering 5 (2019) 1010–1016

Contents lists available at ScienceDirect

Engineering

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

Research
Smart Process Manufacturing: Deep Integration of AI and Process Manufacturing—Perspective

Data Analytics and Machine Learning for Smart Process Manufacturing: Recent Advances and Perspectives in the Big Data Era

Chao Shang [a], Fengqi You [b,*]

[a] Department of Automation, Tsinghua University, Beijing 100084, China
[b] Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

Article history:
Received 6 November 2018
Revised 12 January 2019
Accepted 28 January 2019
Available online 18 October 2019

Keywords:
Big data
Machine learning
Smart manufacturing
Process systems engineering

ABSTRACT

Safe, efficient, and sustainable operations and control are primary objectives in industrial manufacturing processes. State-of-the-art technologies heavily rely on human intervention, thereby showing apparent limitations in practice. The burgeoning era of big data is influencing the process industries tremendously, providing unprecedented opportunities to achieve smart manufacturing. This kind of manufacturing requires machines to not only be capable of relieving humans from intensive physical work, but also be effective in taking on intellectual labor and even producing innovations on their own. To attain this goal, data analytics and machine learning are indispensable. In this paper, we review recent advances in data analytics and machine learning applied to the monitoring, control, and optimization of industrial processes, paying particular attention to the interpretability and functionality of machine learning models. By analyzing the gap between practical requirements and the current research status, promising future research directions are identified.

© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Process industries are playing a dominating role in promoting the growth of the global economy and safeguarding social benefits. This is demonstrated by the fact that the list of the world's top 500 enterprises includes many process industrial companies such as Sinopec, Shell, and ExxonMobil, to name just a few. With the development of chemical engineering, equipment manufacturing, and information technology, the spatial scale and functional complexity of production processes in modern process industries are increasing rapidly. This trend also gives rise to significant challenges for optimal and safe operations at different levels. At the lower level of control, due to dense connections between various plants and processes, multi-loop and multiscale coupling phenomena commonly exist, presenting direct obstacles to the efficient design of plant-wide control strategies. Furthermore, because processes tend to be exposed to disturbances and fault sources, which are difficult to take into account in the design phase, the risk of abnormal events increases enormously. At the higher level of scheduling and planning, decisions must be made in a real-time and flexible manner in response to varying factors in the external environment; such decisions are imperative to save operational costs and improve economic profits under increasing global competition.

To meet stringent requirements on safety, efficiency, and sustainability in modern process industries, cutting-edge technologies and innovations for smart manufacturing are urgently needed. These needs concurrently present both challenges and opportunities for the so-called Fourth Industrial Revolution, which is also known as Industry 4.0. The Third Industrial Revolution, which is now approaching its end, arose from the development of information technology, whereas the prosperity of process industries in the past 30 years is largely due to the wide application of automatic control strategies. In the presently occurring Fourth Industrial Revolution, it has been commonly recognized that machines should not only be capable of relieving humans from intensive physical work—which was a key focus of preceding industrial revolutions—but also be effective in taking on intellectual labor and even producing innovations on their own. In process industries, all manufacturing devices and processes should be "smart" so that, as a whole, they can intelligently sense the environment, discover new knowledge, and make rational decisions. Furthermore, machine intelligence could be classified into lower-level

* Corresponding author.
E-mail address: fengqi.you@cornell.edu (F. You).

https://doi.org/10.1016/j.eng.2019.01.019
2095-8099/© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

intelligence and higher-level intelligence, where lower-level intelligence would be able to functionally resemble humans, while higher-level intelligence would go well beyond the human level, as an ultimate goal to continually pursue in future.

A salient feature of the Fourth Industrial Revolution is the explosive availability of data, which has penetrated nearly all disciplines and motivated renewed inspections of traditional methodologies for problem-solving. As useful tools to model, interpret, process, and utilize data, and eventually achieve machine intelligence, data analytics and machine learning have been well established in the past decades, and have also played a central role in continually pushing beyond traditional boundaries in process systems engineering. The earliest success can be traced back to the late 1980s, arising from the gigantic upsurge of neural networks and back-propagation algorithms [1,2]. Later, statistical learning approaches, including principal component analysis (PCA), partial least squares (PLSs), and support vector machines (SVMs) have received increasing attention thanks to their clear statistical interpretations, ease of model training, and desirable capabilities in handling small sample problems. These have been primarily applied to descriptive modeling tasks, including multivariate statistical process monitoring (MSPM) and soft sensing. Due to the increasing power of machine learning, data information can be leveraged effectively, resulting in significant improvements based upon generic system identification techniques [3].

The current era of big data has witnessed a much broader spectrum of the application of data analytics and machine learning in process industries. As depicted in Fig. 1, these methods penetrate into various hierarchies in process industries, in terms of both passive applications in low-level control loops such as process monitoring and soft sensing, and active applications such as optimal control and high-level decision-making [4]. The former aims to assist industrial practitioners to better observe and manipulate the process and identify variations of importance, without straightforwardly influencing processes. In contrast, decisions obtained through active applications exert a direct effect on industrial processes.

In this work, we seek to revisit recent advances, point to relevant literature, and offer our views on future research directions. We do not intend to provide a systematic and thorough literature review of versatile methodologies; rather, we concentrate on two issues. For passive applications, the literature review provided here is streamlined by a focus on the interpretability of data analytics and machine learning models—that is, the physical meanings behind models and their correspondence with our task-dependent understanding of processes. As for active applications, a focal point is the new functionality of data analytics and machine learning, which refers to the relationships or phenomena that machine learning models aim to describe. By analyzing bottlenecks in current research progress, we point out some promising directions for future investigations.

The outline of this article is as follows. In Sections 2 and 3, we revisit representative data-driven methods with passive applications (process monitoring and soft sensing) and active applications (optimal control and high-level decision-making), respectively. In Section 4, an outlook on future research directions is provided, followed by concluding remarks in the last section.

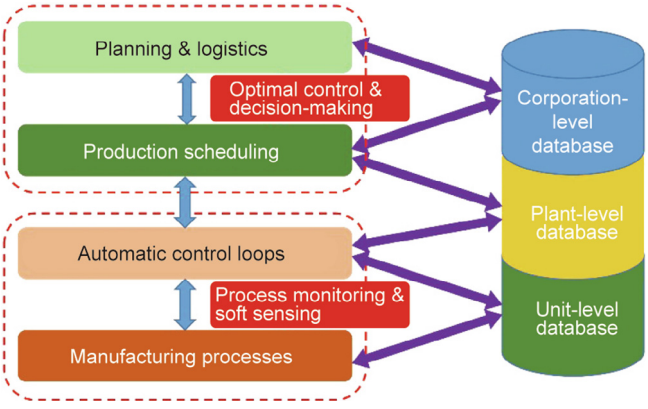

Fig. 1. Hierarchical applications of data analytics and machine learning in process industries.

2. Passive applications: Multivariate statistical process monitoring and soft sensing

2.1. Representation learning: A new road toward data analytics and machine learning

It has been commonly accepted that modeling tasks can generally be classified into unsupervised learning and supervised learning. In unsupervised learning, descriptive models are built to characterize the underlying structure within input data; these are primarily used in monitoring to describe the distribution of process data. In supervised learning, including both regression and classification, a functional mapping between input and output is established, with the prediction accuracy of the output being of special concern. This is most used in the soft sensing of crucial quality variables in industrial processes using fast-rate process variables. Recently, more research attention has been focused on representation learning or feature learning [5], in which the incorporation of domain-specific knowledge when building the model is particularly important. As such, the interpretability of models can be significantly enhanced, which further improves model performance. An example of representation learning is the application of neural networks with piecewise linear units in computer vision. Since abstract features in a figure possess local invariance, which can be described as piecewise linearity, using piecewise linear units as a domain-specific knowledge can be helpful in improving the model performance [6].

Representation learning provides a unified viewpoint of unsupervised learning and supervised learning. Unsupervised learning methods can be regarded as "feature detectors" that are used to extract interpretable underlying features from input data. These features are then used as inputs for a classifier or regressor, thereby significantly enhancing the performance of supervised learning. This is the exact enabling technology used in the deep learning technique [7]. In other words, unsupervised learning and supervised learning are not isolated from each other; rather, the former can greatly benefit the latter.

All in all, a desirable model, no matter how complicated it is, should be endowed with clear physical interpretations. The natural question then follows: What prior knowledge can well fit the characteristics of process data? In fact, this is a common yet implicit focal point of many MSPM studies. To clarify this issue, we review some recent advances in MSPM, and then turn to soft sensing methodologies.

2.2. Feature learning-based MSPM

The effect of a tiny malfunction could be drastically enlarged due to the large size and strong coupling of modern process industries. Therefore, continually monitoring the operation status and taking necessary maintenance actions are crucial to ensure the safety of manufacturing processes, although doing so entails a heavy manual workload [8]. Since the 1980s, MSPM has

established itself as an answer to this conundrum, and a great variety of classic machine learning algorithms have been applied, which well exemplify the intelligence of industrial manufacturing. Some review articles have already provided good summaries of this topic [9,10].

Recent attempts have aimed to utilize prior knowledge that is tailored to continuous manufacturing processes in order to build effective MSPM models. Because the settling time of manufacturing processes is typically long, the entire system to be monitored shows some inertia characteristics. This can be described as the underlying states of processes tending to have slow variability. Therefore, slowness has been suggested as a meaningful attribute to induce underlying features, and thus to appropriately capture process dynamics and yield better descriptions [11]. This perspective motivates the usage of slow feature analysis (SFA) to model process data and to achieve effective process monitoring and diagnosis [12,13]. In comparison with classic MSPM methods such as PCA, independent component analysis (ICA), and canonical variate analysis (CVA), SFA has its unique properties in that it enables separate descriptions of the steady states and temporal behaviors of industrial processes [12]. As such, by designing monitoring statistics tailored to process dynamics anomalies, more meaningful information can be provided; hence, nominal operating point switches can be clearly distinguished from real faults that incur dynamic anomalies. It is shown in the Tennessee Eastman benchmark process that such a strategy can reduce false alarm rates by one order of magnitude [12]. Motivated by the slowness principle, a number of monitoring approaches have been proposed, including recursive SFA for adaptive monitoring [14] and probabilistic monitoring [15]. Successful applications have also emerged in managed pressure-drilling processes [16] and batch production processes [17–19].

An alternative dynamic process data analytics approach called dynamic-inner PCA (DiPCA) has been put forward in Refs. [20,21], in which principal time series as latent variables are sequentially extracted based on predictability. In the work of these scholars, auto-regressive (AR) models are used to build regression models, based on which the predictability of different principal time series is defined. In our opinion, DiPCA is similar to SFA, in that both methods maximize the dynamic contents of latent variables. Roughly speaking, predictability can be seen as a special case of slowness, because time-series data that can be well described by AR models tend to have slow variations. For processes with non-negligible dynamics, the aforementioned methods can provide better descriptive models than traditional dynamic statistical models such as dynamic PCA (DPCA) [22] and dynamic ICA (DICA) [23].

Classic machine learning models are commonly designed to be unimodal. For large-scale industrial processes, multiple operating conditions are present and switches between different modes frequently occur. Therefore, multi-modality should be conceptualized as a domain-specific feature in devising machine learning models for MSPM. The simplest model for multi-mode process monitoring is the Gaussian mixture model (GMM) [24]. Unfortunately, the GMM does not provide information about transition probabilities between different modes, which further motivates the use of the hidden Markov model (HMM) in multi-model monitoring [25]. In both the GMM and HMM, distributions of process data at a single mode are assumed to be Gaussian, although this assumption is quite restrictive in practice. Therefore, more general models are designed to alleviate this assumption [26].

In Ref. [27], a different approach to process monitoring is suggested, which is also in line with the idea of feature learning. This approach regards generic process monitoring charts, such as T^2 and squared prediction error (SPE) charts, as types of low-level features, which then serve as inputs for a high-level process monitoring model (e.g., PCA). In this way, effective fusion of information from multiple process monitoring models can be achieved, thereby systematically accounting for different characteristics of process data. Because the extracted features tend to be Gaussian distributed in a statistical sense, it is rational to use PCA as the high-level process monitoring model.

2.3. Feature learning-based soft sensing

The history of soft sensing can be traced back to the inferential control strategy proposed by Brosilow and Tong [28] in 1978. As an intelligent-sensing technology, soft sensing uses easy-to-measure process variables to furnish online estimates of hard-to-measure but important indices, such as product quality and other environment indices. It is worth mentioning that the key performance indicator (KPI) forecast is another rising application of soft sensors [29]. Some important performance indices must be evaluated based on time-consuming experimental tests, and predictive soft sensors can be developed to provide real-time estimations of these indices, which are useful in assisting operators' decision-making. In principle, the development of soft sensors can be regarded as a regression problem, so various supervised machine learning algorithms have been applied, as is comprehensively documented in Ref. [30].

The promise of utilizing representation learning in building soft sensors was first pointed out in Ref. [11], where probabilistic SFA (PSFA) is employed to induce slowly varying features, based on which a simple least squares regression model is built. Because the slowly varying features well represent underlying variations of the process, some of them tend to be highly correlated to quality indices. Compared with traditional dynamic PLS (DPLS), this approach shows much better dynamic prediction accuracies. Furthermore, it enables a desirable synthesis of fast-rate process data and irregularly sampled quality data, thereby providing a semi-supervised learning scheme. A number of extensions were developed later on. Ref. [31] proposes a Bayesian learning approach to extract dynamic features with shifting dynamics, where the slowness of slow features is assumed to have gradual alterations. In Ref. [32], another layer of flexibility is introduced to tackle the varying number of useful slow features. A modified regularized SFA is later proposed for the quality prediction of industrial terephthalic acid hydropurification processes [33].

The notable deep learning technique itself well exemplifies the spirit of representation learning. The first attempt to apply the deep learning technique to soft sensing was made in Ref. [34], in which soft sensors are built with a deep neural network (DNN) to predict the cut-point temperature of heavy diesel in a crude distillation unit. The training procedure of a DNN involves two distinct steps: an unsupervised learning step to initialize the weights of the DNN, and a supervised learning step to fine-tune the weights based on input–output data. Therefore, the unsupervised learning step can be deemed to be extracting nonlinear underlying features that induce nonlinear correlations between process variables, which further benefit the development of the regression model. Along this line, the deep learning technique has been further applied to crude oil classification [35] and carbon dioxide (CO_2)-capture process modeling [36], both of which demonstrate the advantages of using the deep learning technique in modeling "big process data." Based on inherent features extracted in the unsupervised learning step, DNNs can also be applied to process monitoring and fault diagnosis [37,38].

Needless to say, soft sensing models can also be built based on other features, such as correlations within a low-dimensional subspace. The earliest method in this regard is principal component regression (PCR), in which feature extraction is performed based on simple PCA. The use of low-dimensional latent variable models in soft sensing has been comprehensively summarized in Ref. [39].

In Ref. [40], neighborhood preserving embedding is used to first learn the intrinsic nonlinear structure of data, based on which predictors are established.

3. Active applications: Optimal control and high-level decision-making

3.1. Data analytics and machine learning for optimal control

In advanced control of industrial processes, model predictive control (MPC) is a notable and well-established approach, which is based on a precisely known mathematical model to describe system behaviors and to plan optimal control sequences in the near future [41]. However, the underlying assumption of MPC may be ideal in practice, and unknowns such as model mismatch, unmeasured disturbances, and random noises commonly exist. In these cases, a promising approach is to integrate mechanistic models with data analytics and machine learning, which show great potential to deal with the unknowns [42]. According to functionalities, their applications can be classified into two categories.

The first application category involves the establishment of prediction models for the unknowns by fitting existing past data, such that the uncertainty can be decomposed into the deterministic part that is known in advance, and the stochastic part that represents prediction errors. For example, in the operations and control of smart grids, estimates of electricity generation from renewable energy sources, including wind energy and solar energy, can be derived based on other sources of information, such as weather forecasts and climatic factors [43]. Likewise, if product quality and other indices that cannot be measured online are involved in the optimal control problem, soft sensor models can be built to provide real-time estimates, which are indispensable for executing closed-loop control. In both cases, machine learning models such as SVM and neural networks have been extensively applied. The usefulness of developing a good prediction model lies in the opportunity to noticeably reduce the magnitude of uncertainty that is involved in the optimal control problem, which results in the attainment of better control performance. In this sense, the accuracy of prediction models is a critical issue.

The second category of applying data analytics and machine learning in MPC involves the description of the distribution of uncertainty in an unsupervised manner. In practice, the system is inevitably prone to uncertain disturbances, which may drive system states away from the nominal trajectory. To address the effect of uncertainty, robust MPC (RMPC) [44] and stochastic MPC (SMPC) [45] have been proposed and applied, where different mathematical tools are utilized to delineate uncertainty. In RMPC, uncertainty sets are responsible for representing the possible region of uncertainty realizations, while probability distributions are directly used in SMPC. A recent promising direction in RMPC and SMPC is to appropriately model the uncertainty by taking an active learning viewpoint. In RMPC, traditional norm-based sets are commonly adopted as uncertainty sets, which lack sufficient flexibility to nicely delineate the distribution of uncertainty. Therefore, data-driven uncertainty sets constructed with unsupervised learning methods can well address this issue. For example, by actively learning a compact high-density region from available data in the form of a polytope based on support vector clustering (SVC), the resulting optimal control problem can be cast as a classic robust optimization (RO) problem that is easy to solve [46]. A novel strategy has also been developed to tune the size of the polytope; this strategy provides an appropriate probabilistic guarantee for the solution of RMPC, thereby indicating that an approximate solution of SMPC will eventually be obtained [46]. The theoretical bound of the established sample size is much lower than that of the classic results in SMPC, thereby enhancing the practicability and reducing the conservatism of SMPC. This approach has been applied to irrigation control [47], which shows that the safety and closed-loop performance of systems can be considerably improved by mining meaningful information from data. In Ref. [48], a learning-based scheme is adopted in MPC to deal with repetitive control tasks in autonomous systems.

3.2. Data analytics and machine learning for high-level decision-making

Generic optimization techniques under uncertainty can be classified into stochastic programming (SP) [49], RO [50], and distributionally RO (DRO) [51]; these three methods have found extensive applications in energy system operations and supply-chain design [52,53]. Data-driven decision-making is a recently emerging paradigm that integrates model-based and data-driven systems for optimization under uncertainty. This organic integration of machine learning and mathematical programming leads to fundamentally more powerful and efficient data-driven optimization frameworks that close the loop between data analytics and decision support [54].

Scenario programs yield data-driven approximation to classic chance-constrained SPs, where scenarios collected from past experiences are directly adopted to transform chance constraints into a great number of deterministic constraints [55]. The key to ensuring the quality of scenario programs is to select a sufficient number of important scenarios. Some theoretical results have been established in this regard [56]. However, induced optimization problems always include massive constraints, which pose significant computational challenges. Because of the decomposable structure of scenario programs, many decomposition algorithms such as the L-shaped method have been developed [49,57,58]. Recent research efforts have focused on employing distributed optimization techniques [59,60], in which the original large-scale problem is first decomposed into several subproblems, and then multiple processors are used to solve the subproblems in parallel with limited communications.

In data-driven RO, uncertainty sets are typically constructed directly based on uncertainty data. From a machine learning point of view, this can be understood as an unsupervised learning task. However, not all unsupervised learning methods can be applied to this end, mainly because it is necessary to account for the tractability of induced optimization problems. On the one hand, the unsupervised learning method must be powerful enough to accurately capture the distribution of uncertainty; on the other hand, an over-complicated uncertainty set could make the optimization problem difficult to tackle or even intractable. Therefore, data-driven uncertainty sets must be meticulously devised in order to achieve a desirable balance between two conflicting objectives. Based on such motivation, a number of unsupervised learning methods have been developed recently, which are dedicated to data-driven constructions of uncertainty sets. In Ref. [61], piecewise linear kernel-based SVC is proposed as a new approach tailored to data-driven RO. By solving a quadratic program, the distributional geometry of massive uncertain data can be effectively captured as a compact convex uncertainty set, which considerably reduces the conservatism of RO problems. In addition, the fraction of data coverage of the data-driven uncertainty set can be easily selected by adjusting only one parameter, which furnishes an interpretable and pragmatic way to control conservatism and exclude outliers. In Ref. [62], data-driven uncertainty sets are established by using PCA and kernel density estimation, which can systematically handle correlations and asymmetry. In Refs. [63,64], the use of data-driven uncertainty sets in multi-stage adaptive RO (ARO) is investigated and demonstrated on

industrial-scale scheduling and planning problems in process industries. The results show that significant reduction of conservatism can be obtained in a multi-stage setting thanks to the power of machine learning. In particular, by actively learning the support of uncertain demands, more than 20% higher net present value can be achieved in a planning application, in contrast to generic RO strategies. This approach was later applied to unit commitment [65], optimal operations of industrial steam systems [66], and the resilient design of supply chains [67]. To utilize label information within multi-class uncertainty data, a data-driven stochastic RO framework is proposed in Ref. [68]. In Ref. [69], conventional robustness and minimax regret criteria are simultaneously optimized in data-driven ARO to yield rational decisions.

Data-driven DRO has been a popular topic in operations research in the past decade. It can be regarded as a combination of RO and SP, in that the worst-case performance on a set of probability distributions is optimized [51]. In data-driven DRO, ambiguity sets play a key role, and are typically determined based on data analytics. It is typical for the distribution of uncertainty to not be precisely known in nature; such inexactness is referred to as distributional ambiguity. To hedge against distributional ambiguity, a set of candidate probability distributions is employed. The most-used strategy to characterize the ambiguity is to extract first-order and second-order moment information from past data. The issue of determining the size of ambiguity sets is formally addressed based on hypothesis testing in Ref. [70]. In process industries, DRO has been first applied to process planning and scheduling [71], and to optimal operations of a shale gas supply chain [72].

4. An outlook on future research directions

4.1. Process monitoring

Although a variety of feature selection methods have been used to design process monitoring models, it is worth noting that the extracted features must be closely related to prior knowledge that is tailored to process characteristics. At present, although a great number of process monitoring models are designed based on combinations of different feature extraction methods, nonlinear manifolds and non-Gaussianity, which underpin many monitoring models, are essentially not unique to process data. Typically, a good process model should not only be powerful enough to describe process properties, but also allow for clear interpretations that can be easily accepted by industrial practitioners [73]. At present, however, this issue has not been given sufficient consideration. In this sense, it is worth concentrating future efforts on high-level features such as slowness, non-stationarity, and causality [3,74]. In addition, feature design could be based on prior knowledge from industrial practitioners, such as monotonicity and range information, since interpretable models are helpful for root-cause diagnosis and maintenance after the detection of potential faults [75].

Based on certain feature characteristics, transfer learning can be further adopted to synthesize data information collected under different operating conditions or from different manufacturing devices. Current research has focused on building individual models for each operating condition or each device. Despite their discrepancy, some similarities exist, so such models can be formally expressed as bearing some common information. Borrowing the idea from transfer learning [76], features should be discovered as a common knowledge that describes the fundamental principle underlying manufacturing processes, thereby potentially improving the modeling performance as well as human understanding of it.

Another useful direction is to develop user-friendly visualization strategies in data-based process monitoring in order to better assist decision-making, since visualization can make it possible to better understand high-dimensional process data [77].

4.2. Soft sensing

Due to the time-varying characteristics of industrial processes, the performance of soft sensors can easily degrade over time, which calls for a significant workload to maintain and update the model. Therefore, quality prediction is not merely a regression problem, and more research attention should be paid to the adaptive mechanism of prediction models, especially under the presence of frequent operating condition deviations [78]. Moreover, the inaccuracy of laboratory data caused by human could be considered, such as uncertain time delay, large and varying sampling interval, and sampling habits of various operators.

Traditional supervised models are commonly under a strong premise that data samples are independent and identically distributed. Nevertheless, the underlying mechanism of the process variables affecting product quality may be much more complicated. The burgeoning theory of online learning offers new solutions to modeling tasks without specific assumptions regarding data [79]. For example, online learning could systematically handle data that are generated deterministically, stochastically, or even adversarially. Therefore, attempts to use online learning techniques to address quality-control problems are worth making in the era of big data.

Meanwhile, with the rapid development of imaging technologies, more image and spectral data are being collected in industries, providing meaningful information for the establishment of high-fidelity prediction models. However, a challenge arises from the high-dimensionality and strong correlations between different dimensions. Image processing and object recognition have already been playing a leading role in fields such as remote sensing and autonomous driving [80]. Although these technologies have been applied in process industries [81,82], their development is still in its infancy. Hence, it would be a viable direction to embrace the power of advanced image-processing techniques—especially convolutional neural networks—to make full use of image and spectral data from process industries.

4.3. Data-driven optimal control

Future research efforts can be undertaken in incorporating domain-specific knowledge into devising uncertainty sets in RMPC. For example, in Ref. [47], a new concept of a conditional uncertainty set is proposed to describe the dependence of the distribution of rainfall forecast errors on forecast values. For other types of uncertainty, the question of how to devise the associated conditional uncertainty set is worth further case-by-case investigation.

Reinforcement learning (RL), which is a prevalent machine learning approach, is particularly useful for deriving action policies with no model information [83]. This approach is data-driven in nature and can adapt to time-varying environments intrinsically. Therefore, RL-based control has great potential to tackle optimal control tasks in complicated manufacturing plants, whose high-fidelity mathematical models are difficult to establish in practice [84,85].

4.4. High-level decision-making

In comparison with the other applications, high-level decision-making is the most important, since it directly affects the economic profits and environmental impacts of a process industrial company. On the one hand, high-level decisions are typically made under uncertainty according to the experiences of business leaders,

leaving much room for improvement; hence, more applications of data-driven RO and DRO for decision-making in process industries are anticipated in the future. On the other hand, there is value in future study to further improve the solution quality and computational efficiency of data-driven RO and DRO. Current DRO methodologies exploit moment information to describe the ambiguity of probability distributions. In principle, different kinds of moment information could be regarded as the outcomes of simple data analytics approaches. This situation provides motivation to use advanced unsupervised learning methods to extract high-level information such as the distribution within high-dimensional feature space, based on which ambiguity can be further considered. By utilizing the power of machine learning, the ambiguity of probability distributions can hopefully be reduced, thereby leading to less conservative solutions. For example, the ambiguity set proposed in Ref. [86] involves the probabilities of a series of nested sets; however, there are no results regarding a systematic investigation on the construction of these nested sets. In fact, kernel-based machine learning algorithms with varying regularization parameters could be utilized to derive nested sets, which capture the majority of data samples.

5. Concluding remarks

In modern process industries, an increasing amount of data embodying valuable information can be collected and archived. By making use of data, data analytics and machine learning can help sense the environment, discover knowledge, and make decisions automatically and intelligently. Oriented to data-driven monitoring, prediction, control, and optimization, this paper reviews the current status of research in this field and analyzes the knowledge gaps to be filled in. In particular, we differentiate passive applications of data-driven methods, which include monitoring and soft sensing, from active applications, which include control and optimization. For the former, the interpretability of models has been suggested as a major concern, while for the latter, special attention has been paid to functionality. It is worth noting that although big data is enormously reshaping the process industries, a majority of data-driven methods have not yet been applied in practice. Data analytics and machine learning are by no means the answer to every conundrum. Most importantly, it is necessary to incorporate *a priori* knowledge of plants and processes in order to achieve successful applications, which leaves both challenges and opportunities for future research.

Acknowledgements

Chao Shang acknowledges the financial support from the National Natural Science Foundation of China (61673236, 61433001, and 61873142).

Compliance with ethics guidelines

Chao Shang and Fengqi You declare that they have no conflicts of interest or financial conflicts to disclose.

References

[1] Willis MJ, Di Massimo CD, Montague GA, Tham MT, Morris AJ. Artificial neural networks in process engineering. IEE Proc Contr TheorAppl 1991;138(3):256–66.
[2] Willis MJ, Montague GA, Di Massimo C, Tham MT, Morris AJ. Artificial neural networks in process estimation and control. Automatica 1992;28(6):1181–7.
[3] MacGregor J, Cinar A. Monitoring, fault diagnosis, fault-tolerant control and optimization: data driven methods. Comput Chem Eng 2012;47:111–20.
[4] Pillonetto G, Dinuzzo F, Chen T, De Nicolao G, Ljung L. Kernel methods in system identification, machine learning and function estimation: a survey. Automatica 2014;50(3):657–82.
[5] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013;35(8):1798–828.
[6] Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning; 2010 Jun 21–25; Haifa, Israel; 2010. p. 807–14.
[7] Bengio Y, Lamblin P, Popovici D, Larochelle H. Greedy layer-wise training of deep networks. In: Schölkopf B, Platt J, Hofmann T, editors. Advances in neural information processing systems 19: Proceedings of the 2006 Conference. Cambridge: The MIT Press; 2007. p. 153–60.
[8] Shu Y, Ming L, Cheng F, Zhang Z, Zhao J. Abnormal situation management: challenges and opportunities in the big data era. Comput Chem Eng 2016;91:104–13.
[9] Chiang L, Lu B, Castillo I. Big data analytics in chemical engineering. Annu Rev Chem Biomol Eng 2017;8:63–85.
[10] Ge Z, Song Z, Ding SX, Huang B. Data mining and analytics in the process industry: the role of machine learning. IEEE Access 2017;5:20590–616.
[11] Shang C, Huang B, Yang F, Huang D. Probabilistic slow feature analysis-based representation learning from massive process data for soft sensor modeling. AIChE J 2015;61(12):4126–39.
[12] Shang C, Yang F, Gao X, Huang X, Suykens JAK, Huang D. Concurrent monitoring of operating condition deviations and process dynamics anomalies with slow feature analysis. AIChE J 2015;61(11):3666–82.
[13] Shang C, Huang B, Yang F, Huang D. Slow feature analysis for monitoring and diagnosis of control performance. J Process Contr 2016;39:21–34.
[14] Shang C, Yang F, Huang B, Huang D. Recursive slow feature analysis for adaptive monitoring of industrial processes. IEEE Trans Ind Electron 2018;65(11):8895–905.
[15] Guo F, Shang C, Huang B, Wang K, Yang F, Huang D. Monitoring of operating point and process dynamics via probabilistic slow feature analysis. Chemom Intell Lab Syst 2016;151:115–25.
[16] Gao X, Li H, Wang Y, Chen T, Zuo X, Zhong L. Fault detection in managed pressure drilling using slow feature analysis. IEEE Access 2018;6:34262–71.
[17] Zhang H, Tian X, Deng X. Batch process monitoring based on multiway global preserving kernel slow feature analysis. IEEE Access 2017;5:2696–710.
[18] Zhang S, Zhao C. Slow-feature-analysis-based batch process monitoring with comprehensive interpretation of operation condition deviation and dynamic anomaly. IEEE Trans Ind Electron 2019;66(5):3773–83.
[19] Zhang H, Tian X, Deng X, Cao Y. Batch process fault detection and identification based on discriminant global preserving kernel slow feature analysis. ISA Trans 2018;79:108–26.
[20] Dong Y, Qin SJ. A novel dynamic PCA algorithm for dynamic data modeling and process monitoring. J Process Contr 2018;67:1–11.
[21] Dong Y, Qin SJ. Dynamic latent variable analytics for process operations and control. Comput Chem Eng 2018;114:69–80.
[22] Ku W, Storer RH, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis. Chemom Intell Lab Syst 1995;30(1):179–96.
[23] Lee JM, Yoo C, Lee IB. Statistical monitoring of dynamic processes based on dynamic independent component analysis. Chem Eng Sci 2004;59(14):2995–3006.
[24] Yu J, Qin SJ. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. AIChE J 2008;54(7):1811–29.
[25] Wang F, Tan S, Shi H. Hidden Markov model-based approach for multimode process monitoring. Chemom Intell Lab Syst 2015;148:51–9.
[26] Bai X, Lu G, Hossain MM, Szuhánszki J, Daood SS, Nimmo W, et al. Multi-mode combustion process monitoring on a pulverised fuel combustion test facility based on flame imaging and random weight network techniques. Fuel 2017;202:656–64.
[27] He QP, Wang J. Statistical process monitoring as a big data analytics tool for smart manufacturing. J Process Contr 2018;67:35–43.
[28] Brosilow C, Tong M. Inferential control of processes: part II. the structure and dynamics of inferential control systems. AIChE J 1978;24(3):492–500.
[29] Shardt YAW, Hao H, Ding SX. A new soft-sensor-based process monitoring scheme incorporating infrequent KPI measurements. IEEE Trans Ind Electron 2015;62(6):3843–51.
[30] Kadlec P, Gabrys B, Strandt S. Data-driven soft sensors in the process industry. Comput Chem Eng 2009;33(4):795–814.
[31] Ma Y, Huang B. Bayesian learning for dynamic feature extraction with application in soft sensing. IEEE Trans Ind Electron 2017;64(9):7171–80.
[32] Ma Y, Huang B. Extracting dynamic features with switching models for process data analytics and application in soft sensing. AIChE J 2018;64(6):2037–51.
[33] Zhong W, Jiang C, Peng X, Li Z, Qian F. Online quality prediction of industrial terephthalic acid hydropurification process using modified regularized slow-feature analysis. Ind Eng Chem Res 2018;57(29):9604–14.
[34] Shang C, Yang F, Huang D, Lyu W. Data-driven soft sensor development based on deep learning technique. J Process Contr 2014;24(3):223–33.
[35] Gao X, Shang C, Jiang Y, Huang D, Chen T. Refinery scheduling with varying crude: a deep belief network classification and multimodel approach. AIChE J 2014;60(7):2525–32.
[36] Li F, Zhang J, Shang C, Huang D, Oko E, Wang M. Modelling of a post-combustion CO_2 capture process using deep belief network. Appl Therm Eng 2018;130:997–1003.
[37] Zhang Z, Zhao J. A deep belief network based fault diagnosis model for complex chemical processes. Comput Chem Eng 2017;107:395–407.
[38] Wu H, Zhao J. Deep convolutional neural network model based chemical process fault diagnosis. Comput Chem Eng 2018;115:185–97.

[39] Ge Z. Process data analytics via probabilistic latent variable models: a tutorial review. Ind Eng Chem Res 2018;57(38):12646–112461.
[40] Yuan X, Ge Z, Ye L, Song Z. Supervised neighborhood preserving embedding for feature extraction and its application for soft sensor modeling. J Chemometr 2016;30(8):430–41.
[41] Chu Y, You F. Model-based integration of control and operations: overview, challenges, advances, and opportunities. Comput Chem Eng 2015;83:2–20.
[42] Yan Z, Wang J. Robust model predictive control of nonlinear systems with unmodeled dynamics and bounded uncertainties based on neural networks. IEEE Trans Neural Netw Learn Syst 2014;25(3):457–69.
[43] Appino RR, González Ordiano JÁ, Mikut R, Faulwasser T, Hagenmeyer V. On the use of probabilistic forecasts in scheduling of renewable energy sources coupled to storages. Appl Energy 2018;210:1207–18.
[44] Saltık MB, Özkan L, Ludlage JHA, Weiland S, Van den Hof PMJ. An outlook on robust model predictive control algorithms: reflections on performance and computational aspects. J Process Contr 2018;61:77–102.
[45] Farina M, Giulioni L, Scattolini R. Stochastic linear model predictive control with chance constraints—a review. J Process Contr 2016;44:53–67.
[46] Shang C, You F. A data-driven robust optimization approach to scenario-based stochastic model predictive control. J Process Contr 2019;75:24–39.
[47] Shang C, Chen WH, Stroock AD, You F. Robust model predictive control of irrigation systems with active uncertainty learning and data analytics. IEEE Trans Contr Syst Technol. Epub 2019 May 31.
[48] Rosolia U, Zhang X, Borrelli F. Data-driven predictive control for autonomous systems. Robot Auton Syst 2018;1:259–86.
[49] Marti K, Kall P. Stochastic programming. Berlin: Springer; 1994.
[50] Ben-Tal A, El Ghaoui L, Nemirovski A. Robust optimization. Princeton: Princeton University Press; 2009.
[51] Delage E, Ye Y. Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper Res 2010;58 (3):595–612.
[52] Gebreslassie BH, Yao Y, You F. Design under uncertainty of hydrocarbon biorefinery supply chains: multiobjective stochastic programming models, decomposition algorithm, and a comparison between CVaR and downside risk. AIChE J 2012;58(7):2155–79.
[53] Garcia DJ, You F. Supply chain design and optimization: challenges and opportunities. Comput Chem Eng 2015;81:153–70.
[54] Bertsimas D, Thiele A. Robust and data-driven optimization: modern decision making under uncertainty. In: Johnson MP, Norman B, Secomandi N, editors. Models, methods, and applications for innovative decision making. Catonsville: The Institute for Operations Research and the Management Sciences; 2006. p. 95–122.
[55] Calafiore G, Campi MC. Uncertain convex programs: randomized solutions and confidence levels. Math Program 2005;102(1):25–46.
[56] Campi MC, Garatti S. The exact feasibility of randomized solutions of uncertain convex programs. SIAM J Optim 2008;19(3):1211–30.
[57] Birge JR, Louveaux F. Introduction to stochastic programming. 2nd ed. New York: Springer Science & Business Media; 2011.
[58] You F, Grossmann IE. Multicut Benders decomposition algorithm for process supply chain planning under uncertainty. Ann Oper Res 2013;210 (1):191–211.
[59] Carlone L, Srivastava V, Bullo F, Calafiore GC. Distributed random convex programming via constraints consensus. SIAM J Contr Optim 2014;52 (1):629–62.
[60] You K, Tempo R, Xie P. Distributed algorithms for robust convex optimization via the scenario approach. IEEE Trans Automat Contr 2019;64 (3):880–95.
[61] Shang C, Huang X, You F. Data-driven robust optimization based on kernel learning. Comput Chem Eng 2017;106(2):464–79.
[62] Ning C, You F. Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods. Comput Chem Eng 2018;112:190–210.
[63] Ning C, You F. Data-driven adaptive nested robust optimization: general modeling framework and efficient computational algorithm for decision making under uncertainty. AIChE J 2017;63(9):3790–817.
[64] Ning C, You F. A data-driven multistage adaptive robust optimization framework for planning and scheduling under uncertainty. AIChE J 2017;63 (10):4343–69.
[65] Ning C, You F. Data-driven adaptive robust unit commitment under wind power uncertainty: a Bayesian nonparametric approach. IEEE Trans Power Syst 2019;34(3):2409–18.
[66] Zhao L, Ning C, You F. Operational optimization of industrial steam systems under uncertainty using data-driven adaptive robust optimization. AIChE J 2019;65(7):e16500.
[67] Zhao S, You F. Resilient supply chain design and operations with decision-dependent uncertainty using a data-driven robust optimization approach. AIChE J 2019;65(3):1006–21.
[68] Ning C, You F. Data-driven stochastic robust optimization: general computational framework and algorithm leveraging machine learning for optimization under uncertainty in the big data era. Comput Chem Eng 2018;111:115–33.
[69] Ning C, You F. Adaptive robust optimization with minimax regret criterion: multiobjective optimization framework and computational algorithm for planning and scheduling under uncertainty. Comput Chem Eng 2018;108:425–47.
[70] Bertsimas D, Gupta V, Kallus N. Data-driven robust optimization. Math Program 2018;167(2):235–92.
[71] Shang C, You F. Distributionally robust optimization for planning and scheduling under uncertainty. Comput Chem Eng 2018;110:53–68.
[72] Gao J, Ning C, You F. Data-driven distributionally robust optimization for shale gas supply chains under uncertainty. AIChE J 2019;65(3):947–63.
[73] MacGregor JF, Bruwer MJ, Miletic I, Cardin M, Liu Z. Latent variable models and big data in the process industries. In: Proceedings of 9th International Symposium on Advanced Control of Chemical Processes; 2015 Jun 7–10; Whistler, BC, Canada; 2015. p. 521–5.
[74] Shu Y, Zhao J. Data driven causal inference based on a modified transfer entropy. Comput Chem Eng 2013;57:173–80.
[75] Qin SJ. Survey on data-driven industrial process monitoring and diagnosis. Annu Rev Contr 2012;36(2):220–34.
[76] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010;22(10):1345–59.
[77] Wang R, Edgar TF, Baldea M, Nixon M, Wojsznis W, Dunia R. A geometric method for batch data visualization, process monitoring and fault detection. J Process Contr 2018;67:197–205.
[78] Kadlec P, Grbić R, Gabrys B. Review of adaptation mechanisms for data-driven soft sensors. Comput Chem Eng 2011;35(1):1–24.
[79] Morariu O, Morariu C, Borangiu T, Răileanu S. Manufacturing systems at scale with big data streaming and online machine learning. In: Borangiu T, Trentesaux D, Thomas A, Cardin O, editors. Service orientation in holonic and multi-agent manufacturing. Cham: Springer; 2018. p. 253–64.
[80] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The kitti vision benchmark suite. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition; 2012 Jun 16–21; Providence, RI, USA. Washington, DC: IEEE; 2012. p. 3354–61.
[81] Duchesne C, Liu JJ, MacGregor JF. Multivariate image analysis in the process industries: a review. Chemom Intell Lab Syst 2012;117:116–28.
[82] Chen M, Khare S, Huang B. A unified recursive just-in-time approach with industrial near infrared spectroscopy application. Chemom Intell Lab Syst 2014;135:133–40.
[83] Duan Y, Chen X, Houthooft R, Schulman J, Abbeel P. Benchmarking deep reinforcement learning for continuous control. In: Proceedings of the 33rd International Conference on Machine Learning; 2016 Jun 19–24; New York, NY, USA; 2016. p. 1329–38.
[84] Lewis FL, Vrabie D, Vamvoudakis KG. Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers. IEEE Control Syst 2012;32(6):76–105.
[85] Liu D, Yang X, Wang D, Wei Q. Reinforcement-learning-based robust controller design for continuous-time uncertain nonlinear systems subject to input constraints. IEEE Trans Cybern 2015;45(7):1372–85.
[86] Wiesemann W, Kuhn D, Sim M. Distributionally robust convex optimization. Oper Res 2014;62(6):1358–76.