Front. Eng. Manag. 2021, 8(2): 157–171
https://doi.org/10.1007/s42524-020-0126-0

RESEARCH ARTICLE

Lixin TANG, Ying MENG

Data analytics and optimization for smart industry

Abstract Industrial intelligence is a core technology in the upgrading of the production processes and management modes of traditional industries. Motivated by the major development strategies and needs of industrial intellectualization in China, this study presents an innovative fusion structure that encompasses the theoretical foundation and technological innovation of data analytics and optimization, as well as their application to smart industrial engineering. First, this study describes a general methodology for the fusion of data analytics and optimization. Then, it identifies some data analytics and system optimization technologies to handle key issues in smart manufacturing. Finally, it provides a four-level framework for smart industry based on the theoretical and technological research on the fusion of data analytics and optimization. The framework uses data analytics to perceive and analyze industrial production and logistics processes. It also demonstrates the intelligent capability of planning, scheduling, operation optimization, and optimal control. Data analytics and system optimization technologies are employed in the four-level framework to overcome some critical issues commonly faced by manufacturing, resources and materials, energy, and logistics systems, such as high energy consumption, high costs, low energy efficiency, low resource utilization, and serious environmental pollution. The fusion of data analytics and optimization allows enterprises to enhance the prediction and control of unknown areas and discover hidden knowledge to improve decision-making efficiency. Therefore, industrial intelligence has great importance in China's industrial upgrading and transformation into a true industrial power.

Keywords data analytics, system optimization, smart industry

Received July 28, 2019; accepted April 4, 2020

Lixin TANG (✉), Ying MENG
Key Laboratory of Data Analytics and Optimization for Smart Industry, Ministry of Education, Northeastern University, Shenyang 110819, China
E-mail: lixintang@mail.neu.edu.cn

This work is supported by the Major International Joint Research Project of the National Natural Science Foundation of China (Grant No. 71520107004), the Major Program of National Natural Science Foundation of China (Grant No. 71790614), the Fund for Innovative Research Groups of the National Natural Science Foundation of China (Grant No. 71621061), and the 111 Project (Grant No. B16009).

1 Background

Industrial intelligence is a core technology that enhances the global competitiveness of China's industrial sector and brings about significant changes in the production processes and management modes of traditional industries. The increased integration of manufacturing processes and operation management systems transforms and upgrades traditional industries, making them highly efficient, refined, and environmentally friendly. Therefore, industrial intelligence has great importance in China's industrial upgrading and transformation into a true industrial power.

The steel industry is a typical example of highly polluting industries that consume a large amount of resources and energy. The steel industry in China, a major steel producer, faces dual pressure. First, traditional steel enterprises need to transform and upgrade for strategic development. Second, new steel enterprises need to pursue a sustainable development path. Therefore, the industry must conserve energy and reduce emissions through the intelligent transformation and upgrading of production modes to achieve environment-friendly manufacturing. Reducing energy consumption, improving product quality, and enhancing competitiveness are the most practical ways to achieve this goal. Given the physical essence of the dynamic operation processes in manufacturing and the demands of the physical system, intelligent steel plants should be constructed by integrating the structural optimization of manufacturing processes and digital information systems (Yin, 2017). Yin (2016) explored the physical essence of the dynamic operations in metallurgical processes, proposed a theoretical framework for establishing new-generation steel manufacturing processes and described the concept, theory, and method of a dynamic and precise design for steel plants. Smart industry involves the integration of machines, resources, products, and humans by maximizing the use of information, communication, and optimization techniques based on the analytics of industrial big data (Shao, 2017).

The application of production management in the steel industry has attracted global attention in recent years. The main focus has been on the research and development of integrated production management systems. Many researchers have conducted in-depth studies on the theory and practice of production management in the steel industry and achieved remarkable results. Tang et al. (2001) reviewed key planning and scheduling problems and solution methods for integrated steel production. Tang et al. (2002b; 2012a; 2014a; 2014b; 2016a) defined two modes of production, namely, serial-batch production (processing jobs on the machine in sequence) and parallel-batch production (processing jobs simultaneously on the same machine). Moreover, Tang and Zhao (2008) proposed a semi-continuous batch-scheduling model for typical batch production scheduling problems in the industry.

Tang et al. (2014a) studied an integrated charge batching and casting width selection problem in the steelmaking and continuous casting production process and employed optimization technologies to make batching decisions that improve the production efficiency of steelmaking. Tang et al. (2014b) focused on the dynamic scheduling problem of steelmaking and continuous casting production and proposed an improved differential evolution algorithm with an incremental mechanism to solve it. Tang et al. (2002b) studied scheduling problems in steelmaking to ensure the continuous, smooth, and timely execution of production processes. The problem was successfully solved by a hybrid algorithm based on Lagrange relaxation and dynamic programming.

Yasuda et al. (1984) studied planning and scheduling in the hot-rolling production process to determine coil sequence and assignment to orders. They adopted a two-stage method to solve the planning and scheduling problem. In the first stage, a rough schedule was generated according to the change in characteristics (e.g., thickness, width, and heating temperature) between adjacent rolls. In the second stage, the coils were allocated to orders based on a rough scheduling solution. Lopez et al. (1998) and Fang and Tsai (1998) solved the hot-rolling scheduling problem using a tabu search algorithm and a genetic algorithm, respectively.

Regarding the cold-rolling stage, Tang et al. (2016a) studied the coil batching problem of batch annealing operations considering energy utilization. A branch-and-price-and-cut algorithm and a tabu search algorithm were proposed to obtain the optimal and near-optimal solutions, respectively. Sahay and Kapur (2007) proposed a mathematical programming model for continuous annealing process scheduling based on the heat transfer and annealing dynamics of the production process. An optimization algorithm was used to solve the model and determine a schedule to improve the production efficiency of the continuous annealing furnace. Sahay and Krishnan (2007) established a mathematical model based on the characteristics of continuous annealing, which predicted the changing trends in the temperature and hardness of the steel coil in an annealing furnace, and proposed an algorithm to increase productivity. Valls Verdejo et al. (2009) developed a mathematical programming model for production scheduling problems in a continuous galvanizing production line and designed a tabu search algorithm to identify feasible solutions. Tang et al. (2012a) studied a coil sequencing problem in steel color-coating production. A tabu search-based algorithm with composite neighborhoods was proposed to obtain quickly the near-optimal solution.

Regarding the logistics scheduling, Tang et al. (2002a) studied the slab shuffling problem when steel-rolling schedules were implemented. The problem was formulated as an integer programming model. Then, a modified generic algorithm with tailored generic operators was proposed to solve the problem. Tang et al. (2012b) studied item shuffling problems that arise in the logistics system of steel production. They formulated the problem as a linear integer programming model with additional sets of valid inequalities, and proposed polynomial time algorithms for special cases and a greedy heuristic for general cases. Tang et al. (2015b) studied the integrated scheduling of loading and transportation in steelmaking, which is characterized by separated tractors and semitrailers. Tang et al. (2015c) formulated the stowage problem as a mixed integer programming model. They derived five valid inequalities for the model and developed a tabu search algorithm. Tang et al. (2019) studied the integrated production and delivery scheduling problem. Some effective algorithms were proposed for online and offline problems.

Researchers have studied similar management problems for other industries. Brunaud and Grossmann (2017) studied multilevel decision-making problems in the process industry. Tang et al. (2015a; 2016b) investigated the reshuffling and stacking problems for a terminal yard in a logistics system. Tang and Che (2013) investigated the generation scheduling under a CO_2 emission reduction policy in the energy industry.

The management issues in steel production are usually large-scale combinatorial optimization problems. The steel production process consists of multiple stages, each containing multiple parallel production lines. Moreover, consecutive production stages are logistically linked. Therefore, the whole production process forms a large-scale crossover network. In previous research, problems were normally solved through modeling and optimization methods under specific assumptions. These assumptions, together with subjective settings of the model parameters, inaccurate relationships between inputs and outputs, and a lack of feedback in the whole process, result in deviations from the practical production situation. Data analytics can further explore the potential optimization capacity for production management in steel enterprises. Introducing data analytics can transform production management issues into dynamic system models with feedback through the whole optimization process, thus improving the efficiency and effectiveness of production management.

Following two industrial revolutions, modern industry realized mechanical and electrical automation. In recent years, the rise of computers and the Internet, as well as the development of operations research and management, has promoted operations management optimization. Moreover, scholars and industrial practitioners have begun an in-depth exploration of data and intellectual resources with the development of artificial intelligence (e.g., big data, data analytics, and machine learning). Therefore, applying operations research and artificial intelligence to traditionally dominant industries, such as manufacturing, resources and materials, energy, and logistics systems, is critical in implementing smart and environment-friendly strategies for industry development in China.

2 Innovative fusion structure of data analytics and optimization in smart industry

Operations research is the application of scientific and especially mathematical methods to optimize the decision-making process in the system. Artificial intelligence endows the system with intellectual analytical abilities, such as reasoning, discovering rules, and learning from experience. System optimization is employed in operations research and data analytics in artificial intelligence, mainly in the industrial process. In this study, an innovative fusion structure of data analytics and optimization for smart industry is proposed. In the structure, data analytics and optimization are combined to improve further the intelligent capability of industries. On the one hand, the fusion of data analytics and optimization demonstrates the utilization of existing optimization methods to open further the "black box" in industries and improve the prediction and control of unknown areas. On the other hand, it reveals the potential of data analytics methods to discover hidden knowledge and thus improve decision-making efficiency.

Figure 1 illustrates the fusion structure of data analytics and optimization for smart industry adopted in this study. The central circle demonstrates the fusion of data analytics and optimization (DAO), which is the core of the whole research. The first ring around the central circle presents the general fusion methodology of data analytics and optimization. The second ring presents key technologies based on the general fusion methodology, including intelligent perception (e.g., understanding and description of the industrial process), intelligent discovery (e.g., production condition diagnosis and production quality prediction), optimal execution (e.g., process optimization and optimal control), and optimal decision-making (e.g., whole-process production and inventory planning and production/logistics batching and scheduling). The last ring indicates the application of the methodology and technology to manufacturing, resources and materials, energy, and logistics systems to improve the industrial intellectualization level and refine environment-friendly manufacturing.

3 General fusion methodology of data analytics and optimization in smart industry

This section describes the methodology for the fusion of data analytics and optimization for smart industry. Specifically, system optimization is combined with data analytics along the following lines. 1) A fusion modeling method is proposed to describe the complicated industrial system. System optimization is used to model mathematically the identifiable and quantifiable parts of the industrial production process. Meanwhile, data analytics supplements the mathematical model by constructing the parts that are difficult to model and forming the parameters of the model. 2) An analytics-based efficient system optimization method is proposed to make optimal decisions for engineering management and execution issues. In the analytics-based method, the landscape of the solution space can be obtained by dynamically analyzing the search process of the system optimization method. The landscape of the solution space can be used to guide search direction and route, significantly improving the optimization efficiency. 3) An optimization-based high-precision data analytics method is proposed to discover the rules implied in the complicated industrial systems and lay the foundation for effective decision-making to address large-scale engineering optimization problems. Accordingly, the system optimization method is adopted to improve the accuracy of data analytics. In summary, the fusion methodology strongly supports the accurate description of industrial systems, precise analytics for complicated processes, and efficient optimization for decision-making. Some examples of the integration of system optimization methods and data analytics methods are illustrated in Fig. 2.

3.1 Fusion modeling method

This study proposes the fusion of data analytics and optimization to overcome the limitations of traditional mathematical modeling. Decision-makers can scientifically measure, diagnose, and forecast anticipated conditions and objects by applying data analytics to resources, energy, logistics, equipment, and quality data of

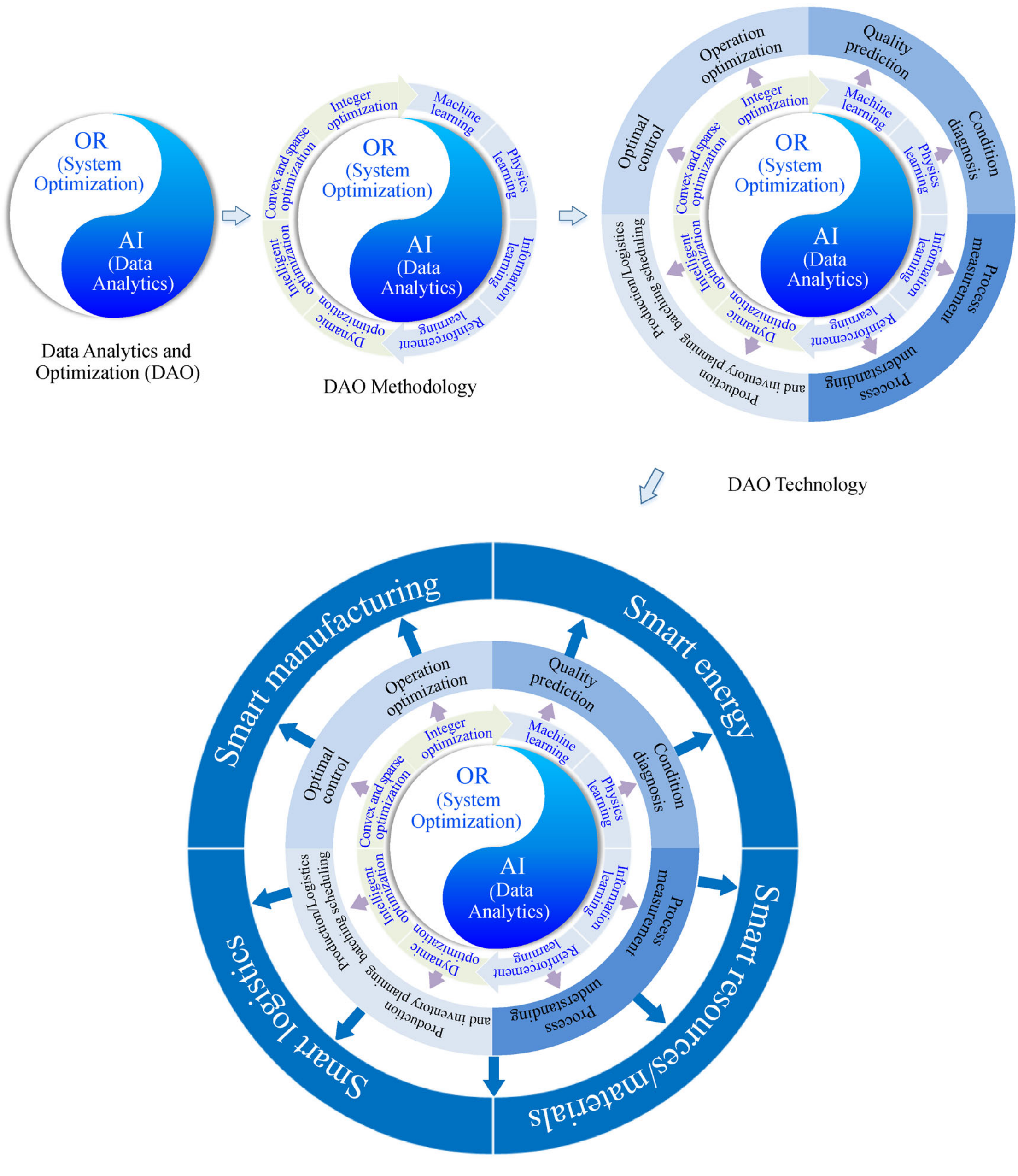

Fig. 1 Fusion structure of data analytics and optimization for smart industry.

production and logistics operations. Data analytics is then combined with system optimization to establish the fusion model. The fusion modeling method should consider the following three cases.

3.1.1 Fusion modeling method for the parts that are hard to model

Considering the complexity, dynamics, and uncertainty of industrial production processes, expressing the management objectives or production restraints by mathematical formulations is often difficult. Data analytics is employed to handle the parts that are hard to model. The functional relationship between management objectives and decision variables and among production restraints can be obtained from historical production data. The fusion of data analytics and optimization can effectively reduce the gap between the established model and the objective system and thus lay the foundation for the optimization of complex industrial production processes.

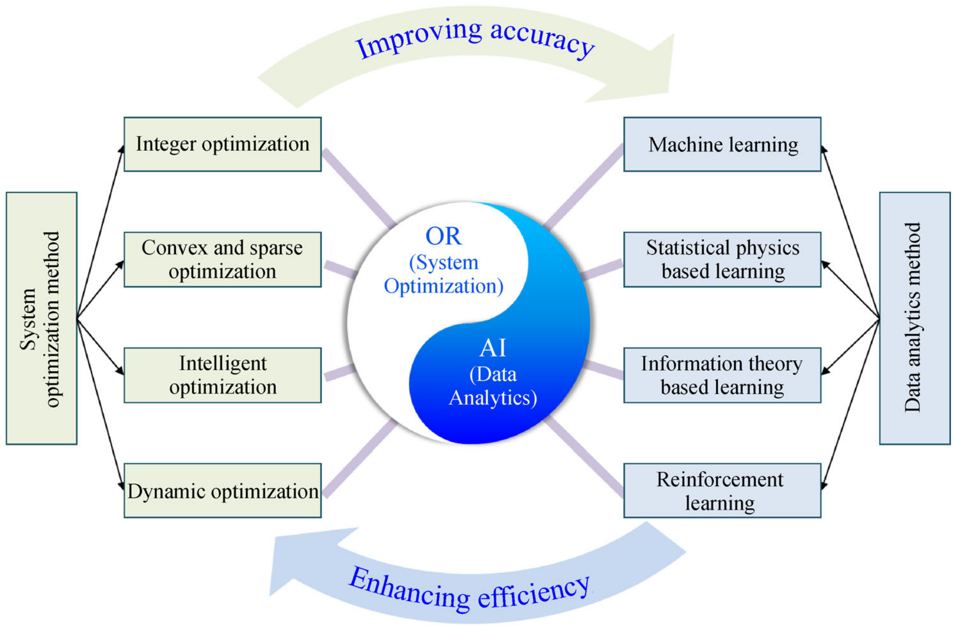

Fig. 2 Fusion methodology of data analytics and optimization.

3.1.2 Fusion modeling method for the parameters that are hard to set

When the industrial system is expressed as a mathematical model, the relationships among multi-objectives, process arguments, and constraint coefficients are represented by parameters. Parameters are usually configured by personal experience, simulations, and statistical methods. However, in most cases, the regular methods can hardly reflect the objective reality accurately because of the complexity of the industrial system. Therefore, data analytics is used to determine the model parameters. The input and output relationships in the mathematical model are used as the input for the machine learning method. Then, parameter values for the mathematical model are obtained, and the self-learning of model parameters is realized for complex industry systems through data analytics.

3.1.3 Fusion modeling method for dynamic models

Complex industrial systems are often characterized by dynamics and uncertainty. Process constraints and management requirements change with the production conditions, leading to changes in parameters and optimization objectives. Existing modeling methods are static and need to be manually adjusted offline when the production conditions change. Thus, they cannot satisfy the real-time requirements of industrial systems. Therefore, system optimization is combined with data analytics according to time-varying production conditions. Through learning, prediction, feedback, and adjustment, an online adjustment method with a closed-loop feedback structure is proposed. The model parameters are adaptively modified, and the optimization schemes are re-adjusted and re-optimized.

3.2 Analytics-based efficient system optimization method

Considering that industrial production is a large-scale, multi-objective, dynamic, and nonlinear process, the efficient system optimization methods combined with data analytics are proposed. Through data analytics methods, such as machine learning and reinforcement learning, the search process of the system optimization method is dynamically analyzed, and the structural features of the optimization space and the influence of the search direction on the objective are revealed. Then, the landscape of optimization space can be outlined, which contributes to reducing the solution space and learning the search direction. Introducing data analytics into system optimization methods can significantly improve optimization efficiency and thus provide an effective optimization solution for large-scale, practical, and complex engineering management problems. Four types of analytics-based system optimization methods are listed in the following.

3.2.1 Integer optimization method based on data analytics

Traditional integer optimization methods can generate optimal solutions for small- and middle-sized problems. However, the solution process of the integer optimization method can be seen as a “black box”. Problem characteristics and objective rules in the solution process cannot be identified, resulting in large computational costs, long solution time, and a limited scale of solvable instances. Thus, integer optimization cannot be applied to solve large-scale industrial problems. Considering the above limitations, data analytics is introduced into the integer optimization method. By intelligently analyzing the iterative process of the algorithm, the structural features of the decision space and the distribution rules of the optimal solution are learned. Accordingly, iterative parameters and the optimization direction of the algorithm can be adaptively selected to significantly increase the optimization capabilities of the algorithm. The integer optimization method combined with data analytics can solve practical industrial problems efficiently.

3.2.2 Convex and sparse optimization method based on data analytics

Given the mechanism complexity of industrial production processes, accurately establishing or solving the convex optimization model is often difficult. Therefore, common convex and sparse optimization methods are hardly employed in industrial practice. By introducing data analytics into convex optimization, feedback analytics is carried out on the actual data in the production process to update and improve the convex optimization model dynamically. Moreover, the historical data in the optimization solution process is fully utilized to obtain the rules and properties of the problem or algorithm, thus speeding up the solution process.

3.2.3 Multi-objective intelligent optimization method based on data analytics

The optimization of complex industrial production processes, which is essentially a nonlinear multi-objective optimization problem, usually includes multiple conflicting objectives. To solve this problem, most studies on multi-objective evolutionary algorithms have focused on the design of evolutionary strategies and evolutionary operators and the adaptive selection of parameters while ignoring the analytics and utilization of intermediate information generated during the iterative process of evolutionary algorithms. To exploit and utilize such information fully, a multi-objective optimization algorithm based on data analytics is proposed. The algorithm first dynamically estimates and constructs the shape of the Pareto front of the optimization problem by analyzing the intermediate solutions obtained in the evolution process using data analytics. The decomposition technology is integrated on this basis. The scatter vector is dynamically adjusted according to the constructed shape of the Pareto front to be able to evenly distribute the decomposition vector across the front. Finally, the evenly distributed decomposition vector improves the search dispersion and efficiency of the multi-objective evolutionary algorithm, ensuring that the distribution of new population is close to the real Pareto front. In summary, the method aims to outline the landscape of the solution space through data analytics and uses the obtained information about the solution space to guide the optimization process. The proposed method can be used to obtain high-quality solutions for complex multi-objective industrial problems.

3.2.4 Dynamic programming method based on data analytics

Dynamic programming (DP) transforms the multi-stage decision process into a series of sub-problems and uses the sequential relationships between such small problems to obtain the optimal solution to the original problem. With the increase in problem size, the number of state variables and the computational complexity of DP increase exponentially, resulting in “the curse of dimensionality”, which limits the capability of DP. To solve this problem, expectations can be introduced in each stage to approximate the utility function and improved with robust optimization or data-driven methods. Another alternative is to design an approximate function to estimate the future impact of current decisions or control strategies in the DP equations for them to meet the Bellman optimality principle. The final optimal decision is obtained via iterative approaching. This method is called approximate dynamic programming (ADP). The design of the approximate and evaluation functions directly affects the performance of the ADP algorithm. Describing the future impact (reward) of different decisions is equivalent to a “black box” or random problem. Data analytics methods can be used to learn problem mechanisms and characteristics via a large amount of historical data. The obtained relationship between reward and decision forms closed-loop feedback for algorithm improvement.

3.3 Optimization-based high-precision analytics method

For complex industrial systems, data analytics methods, such as machine learning, statistical physics-based learning, information theory (IT)-based learning, and reinforcement learning, are used to analyze the historical data and reveal hidden production rules. However, traditional data analytics methods are usually based on a fixed modeling framework, making little use of problem characteristics and system optimization technologies. Thus, over-fitting and low generalization often occur in practical applications. To solve this problem, system optimization methods need to be used to transform the theory and methods of traditional machine learning, thus improving modeling accuracy and generalization ability. The system optimization method is an important tool for extracting data features, detecting outliers, tuning multiple parameters, and constructing deep adversarial networks to improve the accuracy of the machine learning model. Some examples of optimization-based analytics methods are introduced in the following.

3.3.1 Machine learning method based on multi-objective optimization

Ensemble learning is an important research field in machine learning. Traditional methods include AdaBoost, Bagging, and Random Forest. The limitation of these methods is that the learning framework is fixed, which may lead to over-fitting in practical applications. Therefore, machine learning based on multi-objective optimization targets at two conflicting objectives, namely, precision and generalization ability, and uses evolutionary optimization to construct ensemble learning machines. It can achieve the multi-objective optimization of the ensemble learning construction process and the self-adaptive evolution of the ensemble architecture. The combination of machine learning with multiple-objective optimization can effectively overcome the limitation of traditional ensemble learning and provide a new ensemble learning modeling method that satisfies the precision requirements of the industrial production process.

3.3.2 Machine learning method based on statistical physics

Data analytics is essentially a physics discovery problem. Therefore, based on statistical physics, quantum physics, and thermodynamics theory, and combined with statistics, convex optimization, and intelligent optimization methods, we study new learning methods based on physical theories including the following. 1) Based on the theory of statistical physics, the correlation is established between the movement of a large number of microscopic particles and the characteristics of macroscopic behaviors, corresponding microscopic particles to data, and macroscopic behaviors to knowledge, and an explainable learning model with parameters having physical meaning is constructed. 2) Based on the ground electronic state and energy in quantum physics theory, quantum potential energy, electron spin, space position, angular momentum, and other factors are considered, and quantum space topology is used to construct an ultra-micro learning model and characterize the desired macroscopic properties index. 3) Based on the entropy concept of the measurement system orderliness in thermodynamics, a probability learning model based on entropy and enthalpy is established, and the accuracy of the learning model is improved by analyzing the correlation and evolution characteristics of micro data and macro thermal phenomena. In the above learning models based on different physical theories, the effectiveness and robustness of learning are further improved by using convex optimization and intelligent optimization techniques.

3.3.3 Machine learning method based on IT

IT involves discovering effective methods for application issues in communication systems using basic probability and mathematical statistics tools. IT provides quite robust measurements for data sets in the sense of probability distributions. Such measurements can be used as underlying components for new machine learning methods, showing that an IT-based method helps provide reasonable learning models. Moreover, knowledge of the "black box" of optimization algorithms can be precisely measured according to the IT principle, which is used to reveal the complex emergence behavior of the search process. IT-based optimization algorithms can be used to enhance the solution quality of optimization problems that stem from the IT-based machine learning method. Therefore, the accuracy of the method can be significantly improved.

3.3.4 Reinforcement learning-method based on intelligent optimization

Reinforcement learning is used to optimize system performance through iterative interactions and the evaluation of environmental feedback. The mapping strategy is continuously modified from state to action. The key is to adjust the optimal action under different states through learning. Traditional reinforcement learning often causes problems, such as sensitive learning parameters, long learning processes, and poor convergence. Given the advantages of self-learning, self-adaptability, and parallel computing, intelligent optimization is used to optimize the parameters and determine the search direction for reinforcement learning to speed up the convergence and improve the stability of reinforcement learning. Combining reinforcement learning with intelligent optimization, the large-scale stochastic dynamic issues in the industrial production process can be dealt with effectively.

4 Technologies for data analytics and optimization in smart industry

With the above methodology, several key technologies are identified to handle analytics and optimization issues in smart manufacturing, as shown in Fig. 3. In terms of analytics, certain technologies are employed to understand the complicated process and sequentially discover unknown correlations and hidden production rules to forecast trends and perform backtracking diagnosis. On this basis, optimization is carried out to improve process operations and production management to achieve smart manufacturing capability.

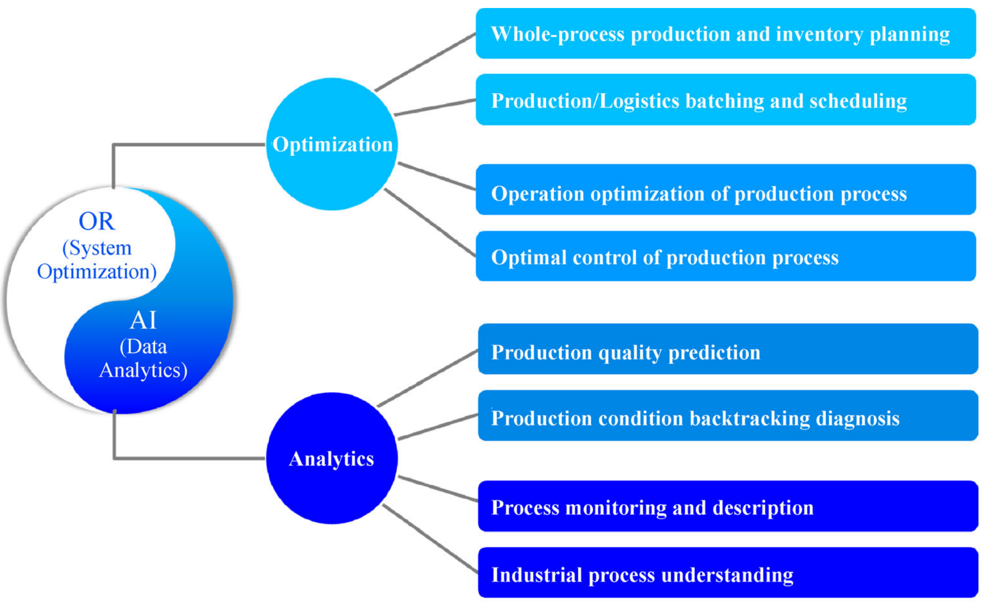

Fig. 3 Data analytics and optimization are used to handle key issues in smart industry.

4.1 Key technologies for data analytics in smart industry

Some key analytics-related technologies that play an important role in smart industry are briefly introduced in the following.

4.1.1 Industrial process understanding technology

Industrial process understanding mainly includes the recognition of industrial image and video, comprehension of sound and speech, and visualization of industrial processes.

Image and video recognition plays an important role in detecting and monitoring production processes. It is usually carried out by experienced operators through the actual observation of images. However, large-scale image data and complex industrial production environment make manual image recognition labor-intensive and less accurate. Therefore, research on image and video recognition based on image processing and deep learning, combined with industrial practice, is the key to improve the intellectualization level of quality detection, monitoring, and fault diagnosis in industrial production processes.

The comprehension of sound and speech helps reveal and recognize the real-time characteristics of production lines and equipment by an analysis of sound signals from the production process. Regressing production data provide a mathematical expression that is consistent with the propagation wave of sound signals. Therefore, the comprehension of sound and speech technology should involve sound-to-text recognition and production mechanism modeling. First, digital data are converted into sound data for multi-dimensional monitoring. Then, the sound data, combined with the collected sound signals, are analyzed, and the status of the equipment and production lines is obtained. Finally, the adjustment scheme based on the resulting status is fed back to the equipment and production lines in the form of voice data for the audibility of the production process.

The visualization of the industrial process restores the dynamic production process to the greatest extent. The whole production process model, integrated with a three-dimensional simulation method, is established based on virtual reality. Then, the established model is performed by virtual reality devices for process visualization, which is considered a “black box” (e.g., ironmaking). Moreover, the process model can be combined with relevant production parameters collected through monitoring the equipment status, operator, and environment. Thus, the real-time monitoring of the production process is allowed through the linkage analysis of the process model and production parameters.

4.1.2 Process monitoring and description technology

The monitoring and description of complex industrial production processes are important to ensure safe production, save energy consumption, and reduce emission. Monitoring and description are used to measure the production process (e.g., energy and resource consumption at each production stage). For example, concerning energy consumption, measurement problems can be divided into three dimensions, namely, production process, product, and medium, according to different measurement objects. The primary task of energy measurement is to analyze and filter out abnormal data and supplement missing data. In the process dimension, the consumption and recovery of resources and energy in each production process are statistically calculated, and the unit consumption and recovery of resources and energy media in each process are obtained. The product-oriented energy measurement problem should be studied considering the various product types in each process, differences in processes and production lines, and intersection among products. The consumption and recovery of resources and energy media per unit output of each product play a vital role in improving product quality and optimizing resource allocation, which is determined by the statistical allocation of resources and energy consumption and the recovery of each process to products. The production indicators of different dimensions should be identified to monitor energy consumption effectively. In addition to process- and product-oriented energy and resources measurement, estimating energy and resource consumption based on the media dimension and generating consumption statistics for each medium are also important.

4.1.3 Production condition backtracking diagnosis technology

Given that steel manufacturing processes are complex and changeable, enterprises often encounter problems in diagnosing production conditions, which have direct effect on production efficiency and product quality. Therefore, the analyses of complex production conditions have practical significance in improving the safety and reliability of the production process. Although the production process cannot be accurately modeled because of high temperatures and pressures, a condition diagnosis model based on machine learning and statistical learning can be established. The model construction requires a collection of substantial production process data and the use of statistical methods for data cleaning, feature clustering, and correlation analysis. Thus, current and future production conditions can be determined, and operators can be guided toward the optimal scenario.

4.1.4 Product quality prediction technology

Analyzing quality data in the production process involves recognizing variations in product quality from the mathematical perspective. It combines the analysis and induction of the relationship among various parameters in the production process, environmental indicators, and product performance indicators. Quality analytics allows the enterprises to interpret effectively the relationship between the current state of the production process (considering the input) and the product performance index (considering the output). Quality analytics applied to the production process mainly aims to identify the influence of product quality parameters, product performance indicators, and enterprise benefits in the production process. The correlation among the production process analytics from the physical, mathematical, data, and economic perspectives can be identified. The quality analytical model of the production process is based on physical principles and the mathematical relationship among parameters. Given that actual industrial production processes are complex and changeable, analytically describing the material and energy transfer relations among different processes is difficult. Therefore, the mechanism is not clear, and the input–output relationship of the black-box problem cannot be determined. On the premise of collecting and analyzing big data, a quality analytics model with a suitable input–output relationship can be established to identify the influence of the related factors on product quality and help managers to test product quality during the production process. Such a model also enables staff to optimize the follow-up production process and achieve the goals of reducing enterprise costs and energy consumption and improving product quality.

4.2 Key technologies for system optimization in smart industry

In this section, some key technologies that are used to handle optimization issues in a steel industry are briefly introduced.

4.2.1 Whole-process production and inventory planning technology

The steel industry is characterized by long production processes, including ironmaking, steelmaking, hot rolling, and cold rolling, with complex logistics networks, high temperatures, high energy consumption, high setup costs, high changeover costs, long production cycles, and high inventory levels (Fig. 4). The business includes multiple production and logistics stages. Plant-wide inventory planning aims to determine stock flows from one production operation to another, inventory levels, changes in each item of stock held, and the logistics of parallel production lines of raw materials, in-process products, and finished products in multiple production stages. A scientific and rational determination of plant-wide inventory planning can reduce production and logistics costs, inventory costs, and energy consumption and improve resource utilization to optimize the enterprise's equipment usage and maximize the overall benefits.

The complex and delicate production process, changeable production conditions, and unpredictable market demands in the steel industry make accurately expressing practical problems using existing mechanisms difficult. However, with the progress in science and technology, precise onsite data can easily be acquired and stored. Data analytics can effectively extract critical data from a large volume of incomplete and noisy practical data, thus obtaining objective information and knowledge. Data analytics-based plant-wide inventory planning in the steel industry focuses on plant-wide production and inventory control, considering all processes (i.e., raw materials, ironmaking, steelmaking, continuous casting, hot rolling, cold rolling, and sales) as a whole. Combining the mechanism model with data analytics helps to establish an appropriate optimal control model and construct a reasonable production and inventory control strategy using convex optimization.

4.2.2 Production/Logistics batching and scheduling technology

Considering the huge equipment and high set-up costs in the steel production process, the products must be produced on equipment in batches to reduce production costs. However, the products required by customers are of high variety and low volume. The contradiction between large-variety requirements and mass production mode has brought great challenges to production management (Fig. 5). How to reduce production costs and increase profits by combining dispersive customer requirements into batches is the key management problem faced in the steelmaking, hot rolling, and cold rolling stages in the steel industry.

Considering the production characteristics of the steel industry, production/logistics batching and scheduling refers to the assignment of jobs with identical or similar characteristics to batches of reasonable size, which are considered as production objects. The main task is to determine the composition and size of batches, the composition and length of campaigns, and the assignment and schedule of batches (or campaigns) to machines and devices to ensure product quality, shorten production cycle, and reduce inventory, production costs, and material and energy consumption. Most previous production and logistics batch scheduling problems were solved using deterministic parameters. For uncertain parameters, the most common solution method is stochastic optimization. Given that process parameters are generally difficult to determine, data analysis is used to estimate the parameters

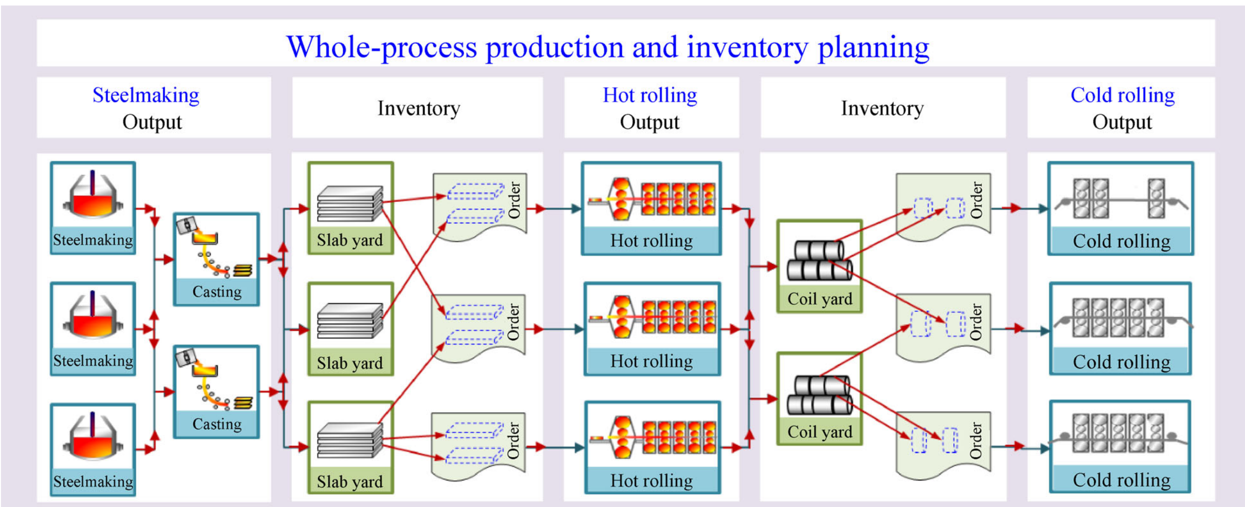

Fig. 4 Whole-process production and inventory planning of steel production.

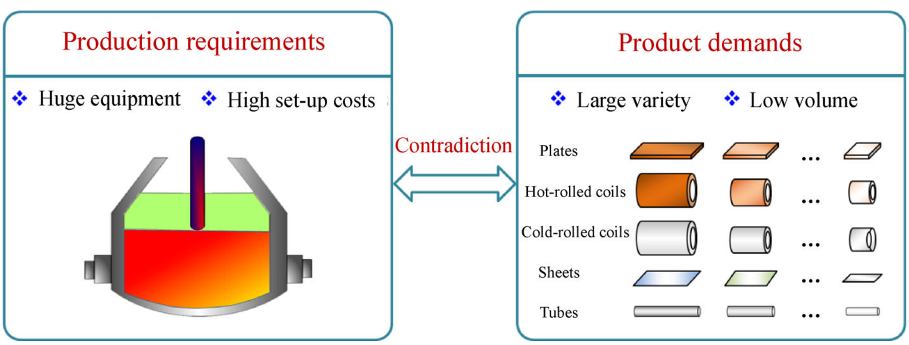

Fig. 5 Main contradiction between production requirements and product demands.

in actual production and logistics operations. Combining data analytics with an optimization method can help solve production and logistics batch scheduling problems.

4.2.3 Operation optimization technology for the production process

According to the production process information, operation optimization for the production process scientifically sets the values of production process parameters, such as flow rate, pressure, and temperature. These values are set without changing process flow and increasing equipment to enable the quality, output, and cost of products to satisfy the expected requirements (Fig. 6). Operation optimization for production processes can be divided into stable-state and dynamic optimization problems. For the stable-state problem, the related information is available, and the process parameters are relatively stable within a period. However, for the dynamic problem, the related information is dynamically available along with the production process. The parameters of the production process must be dynamically adjusted to adapt to frequent changes in production conditions. Operating the industrial production process is usually an optimization problem with complex constraints, as well as large-scale, multi-objective, and dynamic features. Therefore, establishing a precise mechanism model is difficult. These challenges motivate researchers to study the modeling and solution method for operation optimization problems by using data analytics.

A dynamic model related to the time dimension is established for the industrial production processes. The state and operational variables in the model are time-dependent. Mechanism models are developed for explicit processes, whereas data analysis models are established for processes with complex and uncertain factors through the collection and analysis of massive amount of data. Dynamic mechanism models have a large number of differential-algebraic equations. A suitable discretization method must be chosen for highly accurate solutions. Dynamic operation optimization identifies a set of time-dependent operation curves to optimize the performance index. For example, with the blast furnace, the distribution process is analyzed from the physics perspective, then the relationship between the distribution matrix and the radial distribution of the ore-to-coke ratio is modeled. The radial distribution of the ore-to-coke ratio is set by optimizing the distribution matrix through an intelligent optimization algorithm. Research on the optimization of the production process is of great importance for product quality improvement and the reduction of resource and energy consumption as well as production and operation costs.

4.2.4 Optimal control technology for the production process

Optimal control produces control policies with optimal evaluation indexes (energy or costs) based on the production parameters obtained through operation optimization. It allows the dynamic production system to be maintained in the expected state. The basic framework of optimal control is shown in Fig. 7. Typical research cases are described in the following.

Steelmaking, for example, is a complex dynamic batch process for which the mechanism model is difficult to establish. When converter steelmaking is studied in practice, the furnace temperature and carbon content in the steelmaking process are predicted dynamically. On the premise of sufficient data, the least-squares support vector machine (LSSVM) is used to establish a dynamic prediction model, and soft sensing technology is employed

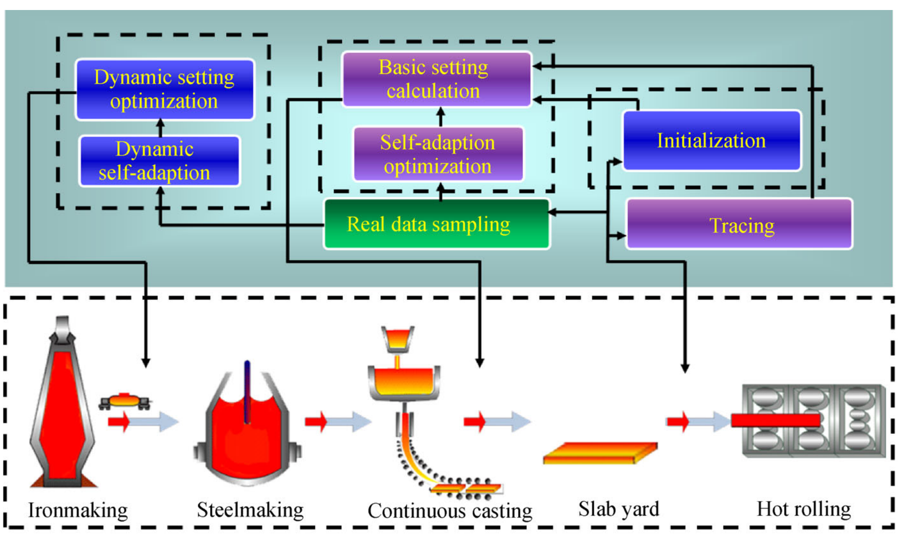

Fig. 6 Operation optimization in the steel industry.

to predict the above parameters. To achieve the required error accuracy, the distribution estimation algorithm optimizes uncertain model parameters, ensuring that the final predicted value satisfies the production process requirements. In the dynamic optimization process, an analytical optimization model is established for the relevant control variables based on the LSSVM. The distribution estimation algorithm helps optimize the model and select the appropriate control strategy.

5 Engineering implementation of data analytics and optimization technology in smart industry

Dorf and Bishop (2011) pointed out that engineering is concerned with understanding and controlling the materials and forces of nature for the benefit of humankind. Engineering practice depends on science research and technology innovation. Science research, technology innovation, and engineering practice generally present a V-shaped structure in the time dimension, as shown in Fig. 8. Yesterday's technology innovation is based on science research of the day before yesterday and implemented in today's engineering practice. In other words, today's science research lays the foundation for tomorrow's technology innovation and the engineering practice of the day after tomorrow. Thus, research on the fusion methods of data analytics and optimization technology provides a strong basis for future engineering practice.

A four-level framework for smart industry based on theoretical and technological research on the fusion of data analytics and optimization is proposed, as presented in Sections 3 and 4. Under the four-level framework, data analytics and optimization technology are employed to overcome critical issues commonly faced by manufacturing, resources and materials, energy, and logistics systems, such as high energy consumption, high production costs, poor production technology, low labor productivity, low utilization rate of resources, and heavy environmental pollution.

Figure 9 shows the four levels of the framework, namely, perception, discovery, execution, and decision-making. The collected sensory data (i.e., image, speech, and text) are understood and described at the perception level. Then, the production process is accurately diagnosed, and product quality is predicted at the discovery

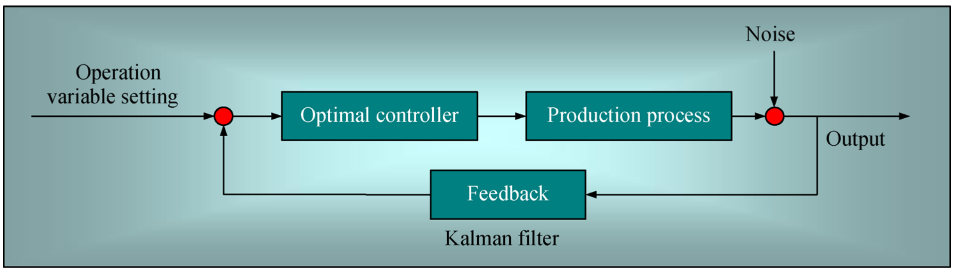

Fig. 7 Optimal control in the steel industry.

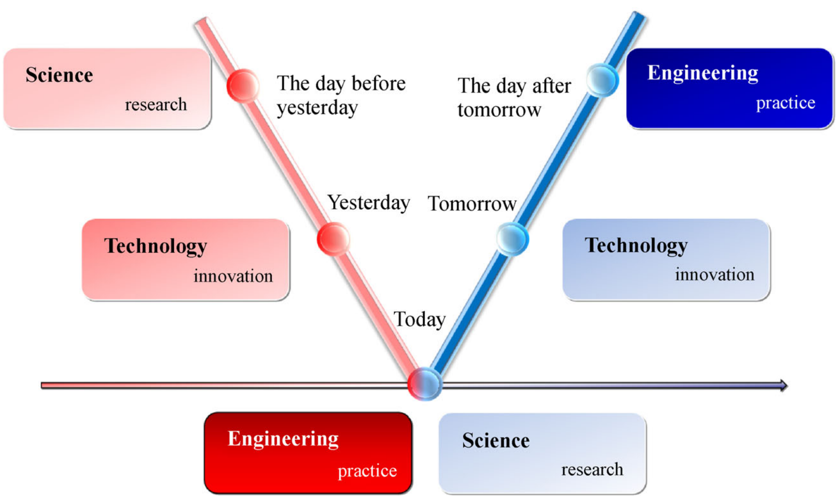

Fig. 8 V-shaped structure of the relationship among science research, technology innovation, and engineering practice.

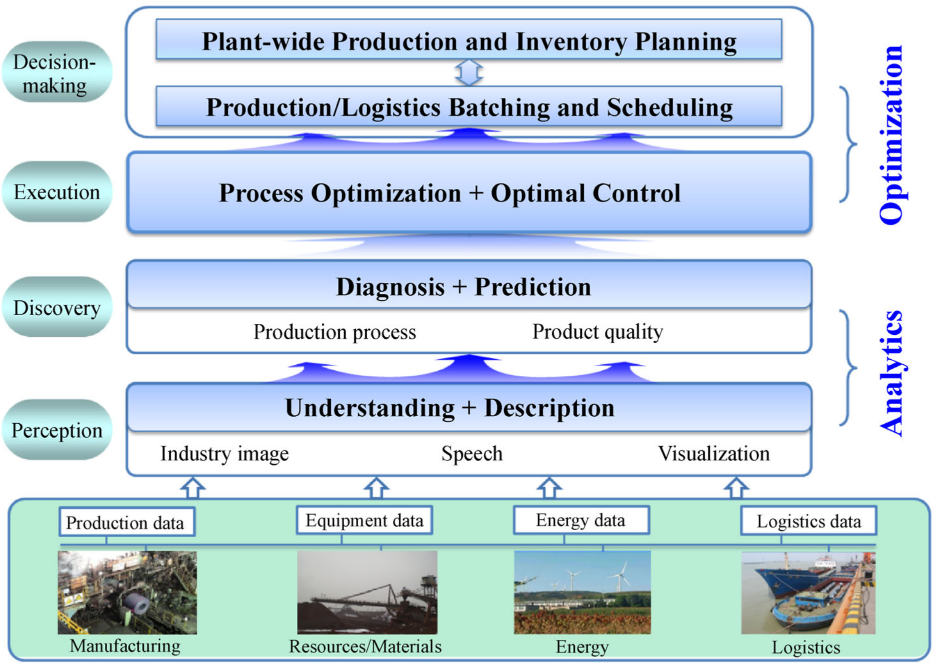

Fig. 9 Four-level framework for data analytics and optimization application in smart industry.

level. Operation optimization and optimal control are executed accordingly. Finally, the decision-making of production planning and scheduling is optimized to gain the capability of smart industry. Perception and discovery are key to data analytics, whereas execution and decision-making are key to system optimization. Under this framework, data analytics supports system optimization, and system optimization is dynamically adjusted according to feedback from data analytics. The four levels of perception, discovery, execution, and decision-making constitute a closed-loop feedback structure.

Perception level

The perception level is the foundation of smart industry. Key analytics issues at this level include industrial data understanding, process monitoring, and description. Understanding refers to recognizing industrial data (i.e., image, speech, and text) and visualizing the black-box technology by virtual reality. Description refers to obtaining and analyzing production process, resource, and energy consumption data; processing abnormal and missing data; and using data analytics to infer the resources, energy consumption, and recovery in each production stage.

Discovery level

The discovery level in smart industry is closely related to production techniques, equipment, control systems, and management. Key analytics issues are tackled for production process diagnosis, product quality prediction, and technological knowledge mining. Diagnosis refers to analyzing the data of each product in each process based on actual historical production process data; identifying the production, resource, and energy bottlenecks; and analyzing the basic causes of any failures in each process. Through the in-depth analytics of the production process, the technological knowledge behind the historical data can also be discovered, thus providing support for the levels of execution and decision-making. Prediction aims to reveal the quality of products based on the current production conditions and historical data, thus providing a scientific basis for enterprise production planning and control strategies.

Execution level

Key system optimization technologies at the execution level include operation optimization and optimal control for the production process. Operation optimization controls the production process according to the mechanism model or data analytics model, describing the quantitative relationship between the operating parameters and relevant economic indices. In other words, operations in the system are monitored, and reasonable process parameters (e.g., flow, pressure, and temperature) are set without changing the process flow and increasing the production equipment. The aim is to improve product quality, produce economic benefits, and optimize the production process. Then, control policies are optimized according to the production parameters obtained by operation optimization to ensure that the dynamic production process meets the expected requirements.

Decision-making level

Decision-making for engineering management ranks at the top in the smart industry ecosystem. Two key optimization issues (i.e., whole-process production and inventory planning and production/logistics batching and scheduling) are identified to transform the production process and improve the utilization of resources, energy, and equipment. From raw materials to semi-finished products to finished products, the whole-process production and inventory planning problem is included in optimally determining the output of each production unit, the amount flowing between two consecutive operations, and the inventory. Optimal planning can ensure a balanced production load, reasonable inventory structure, and smooth production process. Production/Logistics batching and scheduling involve grouping customer requirements into batches consistent with the whole-process production and inventory plan, allocating the batches to equipment, and sequencing and timing the processes to realize the efficient utilization and optimal allocation of resources, energy, and equipment.

6 Conclusions

This study focuses on how to upgrade the industry through data analytics and optimization. To achieve this goal, it proposes an integrated research structure encompassing the theoretical foundation and technological innovation of data analytics and optimization and their application to smart industrial engineering. For the theoretical foundation, this study proposes a fusion of data analytics and optimization to overcome the limitation of the traditional single method and improve optimization efficiency and analysis accuracy. Key data analytics and optimization technologies are then identified to realize smart manufacturing. Finally, a four-level framework for smart industry is illustrated based on the understanding of smart industry. This study may not cover all aspects of smart industry, considering the limitations of the research field. Industrial intellectualization is an ever-progressing area.

References

Brunaud B, Grossmann I E (2017). Perspectives in multilevel decision-making in the process industry. Frontiers of Engineering Management, 4(3): 256–270

Dorf R C, Bishop R H (2011). Modern Control Systems, 12th ed. San Antonio: Pearson

Fang H L, Tsai C H (1998). A genetic algorithm approach to hot strip mill rolling scheduling problems. In: Proceedings of 10th IEEE International Conference on Tools with Artificial Intelligence. Piscataway, 264–271

Lopez L, Carter M W, Gendreau M (1998). The hot strip mill production scheduling problem: A tabu search approach. European Journal of Operational Research, 106(2–3): 317–335

Sahay S S, Kapur P C (2007). Model based scheduling of a continuous annealing furnace. Ironmaking & Steelmaking, 34(3): 262–268

Sahay S S, Krishnan K (2007). Model based optimization of continuous annealing operation for bundle of packed rods. Ironmaking & Steelmaking, 34(1): 89–94

Shao A L (2017). Can industrial intelligence promote industrial transformation? Case of mining enterprises. Frontiers of Engineering Management, 4(3): 375–378

Tang L X, Che P (2013). Generation scheduling under a CO_2 emission reduction policy in the deregulated market. IEEE Transactions on Engineering Management, 60(2): 386–397

Tang L X, Jiang W, Liu J Y, Dong Y (2015a). Research into container reshuffling and stacking problems in container terminal yards. IIE Transactions, 47(7): 751–766

Tang L X, Li F, Chen Z L (2019). Integrated scheduling of production and two-stage delivery of make-to-order products: Offline and online algorithms. INFORMS Journal on Computing, 31(3): 493–514

Tang L X, Li F, Liu J Y (2015b). Integrated scheduling of loading and transportation with tractors and semitrailers separated. Naval Research Logistics, 62(5): 416–433

Tang L X, Liu J Y, Rong A Y, Yang Z H (2001). A review of planning and scheduling systems and methods for integrated steel production. European Journal of Operational Research, 133(1): 1–20

Tang L X, Liu J Y, Rong A Y, Yang Z H (2002a). Modeling and a genetic algorithm solution for the slab stack shuffling problem when implementing steel rolling schedules. International Journal of Production Research, 40(7): 1583–1595

Tang L X, Liu J Y, Yang F, Li F, Li K (2015c). Modeling and solution for the ship stowage planning problem of coils in the steel industry. Naval Research Logistics, 62(7): 564–581

Tang L X, Luh P B, Liu J Y, Fang L (2002b). Steel-making process scheduling using Lagrangian relaxation. International Journal of Production Research, 40(1): 55–70

Tang L X, Meng Y, Chen Z L, Liu J Y (2016a). Coil batching to improve productivity and energy utilization in steel production. Manufacturing & Service Operations Management, 18(2): 262–279

Tang L X, Sun D F, Liu J Y (2016b). Integrated storage space allocation and ship scheduling problem in bulk cargo terminals. IIE Transactions, 48(5): 428–439

Tang L X, Wang G S, Chen Z L (2014a). Integrated charge batching and casting width selection at Baosteel. Operations Research, 62(4): 772–787

Tang L X, Yang Y, Liu J Y (2012a). Modeling and solution for the coil sequencing problem in steel color-coating production. IEEE Transactions on Control Systems Technology, 20(6): 1409–1420

Tang L X, Zhao R, Liu J Y (2012b). Models and algorithms for shuffling problems in steel plants. Naval Research Logistics, 59(7): 502–524

Tang L X, Zhao Y, Liu J Y (2014b). An improved differential evolution algorithm for practical dynamic scheduling in steelmaking-continuous casting production. IEEE Transactions on Evolutionary Computation, 18(2): 209–225

Tang L X, Zhao Y F (2008). Scheduling a single semi-continuous batching machine. Omega, 36(6): 992–1004

Valls Verdejo V, Alarco M A P, Sorli M P L (2009). Scheduling in a continuous galvanizing line. Computers & Operations Research, 36 (1): 280–296

Yasuda H, Tokuyama H, Tarui K, Tanimoto Y, Nagano M (1984). Two-stage algorithm for production scheduling of hot strip mill. Operations Research, 32: 695–707

Yin R Y (2016). Theory and Methods of Metallurgical Process Integration. Beijing: Metallurgical Industry Press

Yin R Y (2017). A discussion on "smart" steel plant—view from physical system side. Iron and Steel, 52(6): 1–12 (in Chinese)