IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 28, 2022, accepted April 3, 2022, date of publication April 11, 2022, date of current version April 18, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3166172

A Novel Fault Identification Method Driven by Knowledge and Data

QIHAO WAN AND HEMING ZHANG
Department of Automation, Tsinghua University, Beijing 100084, China
Corresponding author: Heming Zhang (hmz@mail.tsinghua.edu.cn)

This work was supported by the National Key Research and Development Program of China under Grant 2020YFB1707800.

ABSTRACT In the field of intelligent manufacturing, fault identification is an effective way to improve product service by identifying the cause of failures. For addressing it, the Generalized Bayesian Network (GBN) model is extended based on the traditional Bayesian Network in this paper, which redefines the directed edges and probability parameters among nodes. Compared with Bayesian Network, the GBN model has the ability to simultaneously define causality and correlation of variables. In addition, the structure of network is not only based on statistical data but also driven by expert knowledge. In order to achieve the collaboration of data and knowledge while maintaining the consistency, a hierarchical collaborative framework is designed including the data layer and knowledge layer. Furthermore, a hierarchical multi-objective optimization algorithm, namely Hierarchical Non-dominated Sorting Genetic Algorithm II (HNSGA-II), is advanced to solve the proposed model. Finally, an industrial case study for fault cause identification targeting the product service helps illustrate all details.

INDEX TERMS Fault identification, generalized bayesian network (GBN), data and knowledge, hierarchical collaborative framework, hierarchical non-dominated sorting genetic algorithm II (HNSGA-II).

I. INTRODUCTION

The manufacturing industry is one of the most important components for national economic development. Many countries have successively carried out manufacturing revolutions to innovate technology and increase productivity [1]. With the development of information science and internet technology, lots of advanced theoretical research has been proposed, such as Artificial Intelligence, Big Data, Internet of Things and so on, which extends the development space in the manufacturing [2]–[4]. In addition, in order to meet the demands of market and improve competitiveness, enterprises have to introduce new technologies to improve production efficiency and reduce production costs while ensuring production quality [5]. In this context, intelligent manufacturing has been proposed and has received extensive attention. Fault identification is one of the main research topics.

Fault identification refers to identifying observed variables most relevant to faults according to the phenomenon of failures during product service, and detecting the cause of faults [6]. Fault identification focuses more on exploring the cause of faults. The ability to timely find the cause and troubleshoot faults can enormously reduce the loss in production process and ensure the production efficiency, which greatly improves the competitiveness of enterprises in the market [7].

In the manufacturing field, expert knowledge generated by employees and data recorded by machines contain a mass of valuable information for tapping the production potential [8]. Prior to the development of computer technology, production models can be constructed by expert knowledge solely. In that case, expert system and knowledge engineering had been extensively studied. However, knowledge of enterprises may not only be explicit, that is, formalized in the documentation, but also be tacit in the environment and available in the minds of employees [9], [10]. Moreover, the manufacturing system is too complicated to construct the internal models fully, therefore many problems cannot be solved based on knowledge engineering or expert system solely. Fortunately, with the mature of information technology, data-driven statistical science has broken through these problems. Big data technology based on statistical analysis has the ability to use the data-driven simulation to construct models without clearing the internal mechanism. While the data-driven theoretical research is applied in production process, other problems have arisen. In actual production, lots of data cannot

The associate editor coordinating the review of this manuscript and approving it for publication was Nagarajan Raghavan.

be available or analyzed in production process. Hence, the practical applications cannot satisfy the basic assumption of theoretical research, that is, the production data is available and sufficient. In other words, the data source may be sparse or incomplete.

With the development of Artificial Intelligent technology, the human-machine collaboration has attracted more and more attention in many fields [11]. In that case, the integration between expert knowledge and statistical data is developed, which can effectively solve the problems caused by insufficient data and incomplete expert knowledge. Furthermore, Bayesian Network (BN) is known as a probabilistic graphical model, which can describe the causal relationship through conditional probability of uncertain variables [12]. And it is a well-established network framework for reasoning the causality of uncertain variables. Compared with rule-based reasoning methods, BN has the ability to deal with complex, fuzzy and uncertain models based on the graph theory and probability theory [13]. Therefore, BN can be an excellent tool to incorporate statistical data and expert knowledge in the manufacturing field. However, BN also has some serious limitations [14]. In BN structure, edges of nodes are unidirectional, which can express that one node causes the other one, but cannot reflect the interaction between them [15]. In other words, although BN can describe the causality of some variables, it cannot express the correlation of other variables. Moreover, the acyclic structure of BN results that it fails to define the circular dependency of variables.

Now solutions to these problems are discussed widely. In this paper, the Generalized Bayesian Network(GBN) based on traditional BN is designed, which redefines the directed edges and probability parameters of nodes. It can simultaneously take into account of the causality and correlation of variables. And a novel hierarchical collaborative optimization framework driven by statistical data and expert knowledge is extended to construct the GBN. The basic idea of this framework comes from the previous work [16], where it integrated various modular standards and traded off the contradiction of different standards so as to optimize the modularization of product platform with multiple hierarchical objectives. In this study, this framework is extended to integrate statistical data and expert knowledge while maintaining the consistency of information and complementing the inadequacy of each other. For solutions to the framework, a hierarchical multi-objective optimization algorithm in previous works, the Hierarchical Non-dominated Sorting Genetic Algorithm II(HNSGA-II), is updated. Further, the problem and model are formulated to solve the networks. For illustrating the model and method proposed, an application about fault identification in products service system is reported as an example, where the results of experiments are compared and discussed.

The rest of this paper is organized as follows: Section II introduces the related work about the study. Section III proposes the generalized Bayesian network model and hierarchical collaborative optimization framework. Furthermore, for solutions to it, the hierarchical multi-objective optimization algorithm, HNSGA-II, is also given in Section III. The formulation of problems for fault identification and the formulated design of models are proposed in Section IV. Section V introduces a case study in the field of industry and gives solutions and results of experiments to demonstrate the models and methods. In addition, applications and possible extensions of proposed models are discussed in Section V. Finally, the conclusion of this study is reported in Section VI.

II. RELATED WORK

With the maturity of information technology, data-statistics science has been applied in the field of manufacturing, and problems related to faults are widely studied [17]. Ricardo *et al.* proposed to measure the reliability of the reconstruction procedure and determine the principal components analysis (PCA) model for best reconstruction. Based on the fault subspace, fault magnitude, and the squared prediction error, necessary and sufficient conditions were provided to determine if the faults are detectable, reconstructable, and isolatable [18]. A reconstruction-based fault identification approach using a combined index for multidimensional fault reconstruction and identification was proposed to reconstruct the fault along a given fault direction [6]. The functional-failure identification and propagation framework was introduced as a novel approach for evaluating and assessing functional-failure risk of physical systems during conceptual design, which was based on combining hierarchical system models of functionality and configuration, with behavioral simulation and qualitative reasoning [19]. Santos *et al.* present a transient-based algorithm for high-impedance fault identification on distribution networks, which used the discrete wavelet transform to monitor high-frequency and low-frequency voltage components at several points of the power system, being able to indicate the most likely area within which the disturbance has occurred, without requiring data synchronization nor the knowledge of feeder or load parameters [20]. Brian S *et al.* investigated a fault diagnosis framework based on detection with feature extraction methods and identification based on data-driven process topology methods [21]. Hojjat A *et al.* developed a new data-driven fault tolerant model predictive control, which does not need the post-fault model. The model identification and control (re)calculation were combined together and were performed simultaneously to efficiently use the critical post-fault/failure time [22]. In addition, the researchers at Centre for Risk, Integrity and Safety Engineering (C-RISE) in Memorial University of Newfoundland have made outstanding contributions on the research of fault detection. For example, Alauddin *et al.* present a new data-driven fault detection model using an artificial neural network (ANN) and variable mosquito flying optimization (V-MFO) technique, and the model parameters had been tuned using the V-MFO algorithm for maximizing the fault detection rate (FDR) while minimizing the FAR [23]. Amin *et al.* present a novel methodology for dynamic risk analysis, integrating the multivariate data-based process monitoring and logical dynamic failure prediction model and it also generated a multivariate probability for a fault class in each time-step, which was used for dynamic failure prognosis by different paths a fault can lead a process to failure [24]. The current research on fault identification is mainly driven by data information. The technology research driven by knowledge and data is a new development direction for fault identification, which has the ability to break through the limitations of data information and improve the identification effect.

Recently, it is argued that collaboration between expert knowledge and data-driven science can result in significant model improvements [25]. A learning function was incorporated to combine the conditional probability distributions in terms of weighted mean, which were modeled by a Bayesian Network that was traversed to return a probabilistic solution according to the symptoms given by the user [26]. Jing Li *et al.* proposed a causal modeling approach to improve an existing causal discovery algorithm by integrating manufacturing domain knowledge with the algorithm, which was demonstrated by discovering the causal relationships among the product quality and process variables in a rolling process [27]. Qin Zhang *et al.* proposed to use i-mode, e-mode, and h-mode of the dynamic uncertain causality graph to model such complex cases and then transform them into either the standard i-mode or the standard e-mode for knowledge representation and reasoning in use of statistical data and domain knowledge [28]. Constantinou *et al.* focused on modeling the impact of some additional expert variable, where a method was proposed for eliciting expert judgment that ensured the expected values of a data variable were preserved under all the known conditions [29]. According to the review on the collaboration between expert knowledge and data-driven science, the current research driven by knowledge and data lacks practical application in the field of industry.

In addition, Bayesian network, as a highly advantageous probabilistic graph model, is widely applied in the field of computer science, manufacturing, bioinformatics and so on [30]. Amit Sata *et al.* present a Bayesian inference-based methodology for analysis and reduction of casting defects, where the values of posterior probability of each input parameter were computed using Bayesian inference to identify the most influencing parameters and the avoidable range of their values [31]. Codetta-Raiteri *et al.* described a fault detection, identification, and recovery cycle composed of the tasks of diagnosis, prognosis and recovery, which was characterized through a dynamic Bayesian network model for autonomous spacecrafts [32]. Turkoz *et al.* developed a data-driven Bayesian approach for fault identification that addressed the limitations posed by the normality assumption, which was computationally efficient for high-dimensional data compared with existing approaches [33]. The current research on the application of BN cannot consider the causation and correlation of variables and keep the consistency of information from multiple sources. Considering the wide application of Bayesian network, the structure identification of it also become a very active research area. Silander *et al.* proposed a straightforward method to find the globally optimal Bayesian networks and demonstrated its feasibility [34]. Friedman *et al.* introduced an algorithm that achieved faster learning by restricting the search space, which restricted the parents of each variable to belong to a small subset of candidates [35]. The current structure learning of Bayesian networks is mainly based on data, and there are few methods driven by the collaboration of data and knowledge.

Overall, there has been growing interest in research driven by data and knowledge. However, little research on the integration of knowledge and data is applied in the field of manufacturing. Moreover, it should be noted that current research is mainly limited in that expert knowledge is only incorporated with the data-driven studies in the form of constraints. In other words, there is no real integration of expert knowledge and statistical data. In addition, the main difficulty of incorporating knowledge and data is that expert knowledge must be consistently integrated with data, that means, expert constraints should be coherent with the conditional independency found in data [36]. On the other hand, although Bayesian Network has an advantage in describing the causality of variables, it cannot define the circular dependency and cannot completely reflect the interaction of variables.

III. MODEL AND METHOD

A. GENERALIZED BAYESIAN NETWORK

In essence, Bayesian Network is a Directed Acyclic Graph (DAG) model, which consists of nodes representing random variables and directed edges describing the relationships of variables. Based on BN, this study proposes the Generalized Bayesian Network(GBN) model, which can simultaneously define the causality and correlation of each variable.

Likewise, the GBN is a combination of nodes and directed edges, shown in Fig1. Different from BN, in the GBN, each variable node is constructed by two sub-nodes, the condition node and output node, which are the red one and the blue one in Fig1-a. When the variable acts as a parent node to represent conditions, the condition node is used for constructing networks. In contrast, when the variable describes outputs as a child node, the output node is used. Actually, in terms of the causality, the condition node expresses reasons and the output node defines results, which means that, condition nodes cause output nodes. In other words, a directed edge always points from the condition node of one variable to the output node of another one. As stressed, there may be three forms for edges between two nodes, including no edge (Fig1-b), one-way edge (Fig1-c) and two-way edge (Fig1-d). There is no edge between two nodes indicating that the two variables are independent for each other. The one-way edge describes causality of two nodes, that is one node causes another one with a certain probability. And the two-way edge expresses correlation of nodes that means the two nodes can influence each other with different probabilities.

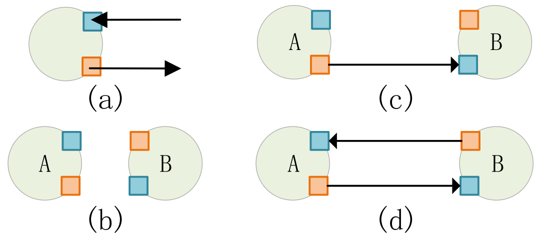

FIGURE 1. GBN: (a) node model (b) no edge model (c) one-way edge model (d) two-way edge model.

TABLE 1. The conditional probability table.

	$x_j = 0$	$x_j = 1$
$x_i = 0$	Not consider	$1 - P(x_i \mid x_j)$
$x_i = 1$	Not consider	$P(x_i \mid x_j)$

Different from BN, parameters of GBN are defined in the form of conditional probability based on edges instead of joint probability based on nodes. In detail, the parameter for a directed edge from A to B refers to the probability of B in condition of A. The parameters are defined based on edges between A and B, which have nothing to do with other nodes. In mathematics, GBN is also considered as a pair $\{G, \theta\}$ on a variable set X, where

1) $G = (I, E)$ indicates the structure graph of networks, in which E means the directed edges set of nodes set I;
2) θ defines the set of conditional probability parameters between random variables, which quantitatively describe the influence of parent nodes on child nodes.

$$\theta(X) = P\left(x_i | x_j\right) \tag{1}$$

In (1), there is a directed edge from the node x_j to x_i, and the parameter is defined based on the directed edge.

In order to better explain the definition of parameters, the conditional probability table is shown in Table1. The table describes the parameters $P\left(x_i | x_j\right)$ when the model has a directed edge from the node x_j to x_i. The parameters are defined based on the edges, and they only describe the effect on other variables when one variable occurs. When one point does not occur, the impact on other nodes is not considered.

It is worth noting that the variables may be discrete or continuous. In this paper, discrete variables are discussed as examples. In detail, the schematic diagram of GBN is shown in Fig2 as an example. In the model, the nodes set I includes variates A, B, C, D and E. It is the two-way edge model among nodes A, B and C, which describes the correlation of variates based on conditional probabilities. And it is the one-way model between A and D, in which there is a directed edge from A to D with the conditional probability parameter P(D|A). It defines the causality of A and D, that is the variate A causes the variate D with conditional probability P(D|A). There is no edge between A and E, which indicates that the variate A and E are independent for each other.

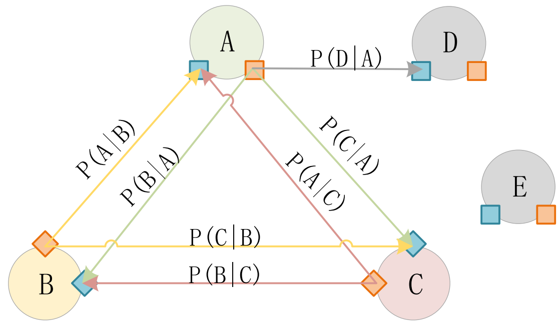

FIGURE 2. Generalized Bayesian network model.

Compared to BN, GBN redefines edges and parameters of networks. It may be a multidirectional ringed graph structure with conditional probabilities instead of directed acyclic graph. Essentially, GBN can be considered as a multi-layer network nested by multiple Bayesian networks with the same variables set. As shown, the nodes A, B and C construct the GBN, which can be considered as three Bayesian networks shown by the different colors. Therefore, the developed GBN model has the ability to express the independence, the causality and the correlation of variables on the same networks.

B. HIERARCHICAL COLLABORATIVE OPTIMIZATION FRAMEWORK

As stressed, data-driven collaborative engineering with expert knowledge can make up for the problems of insufficient data on data-driven engineering and the limitations of tacit knowledge for production process. Realizing the advantages of collaboration between expert knowledge and statistical data, this study develops a hierarchical collaborative optimization framework proposed in previous works to construct the GBN based on the collaboration of knowledge and data [16]. However, the key challenge is how to maintain the consistency of knowledge and data. Moreover, there is no accurate method to quantitatively measure the influence of expert knowledge for production process in the field of manufacturing.

For addressing these difficulties, in this study, the hierarchical collaborative optimization framework based on multi-objective optimization is developed, shown in Fig3. The framework divides expert knowledge and statistical data into different layers so as to optimize the structure of GBN separately. Meanwhile, the collaboration parameters generated by one layer will be used for optimization of another layer to collaborate different layers. Actually, each optimized layer is considered as a multi-objective optimization problem, and it collaborates with other optimized layers to optimize the results.

As shown, the framework developed is divided into two layers, including Knowledge Layer (KL) and Data Layer (DL), which respectively correspond to two multi-objective optimization problems. The KL optimizes the causality objective and correlation objective of variables based on

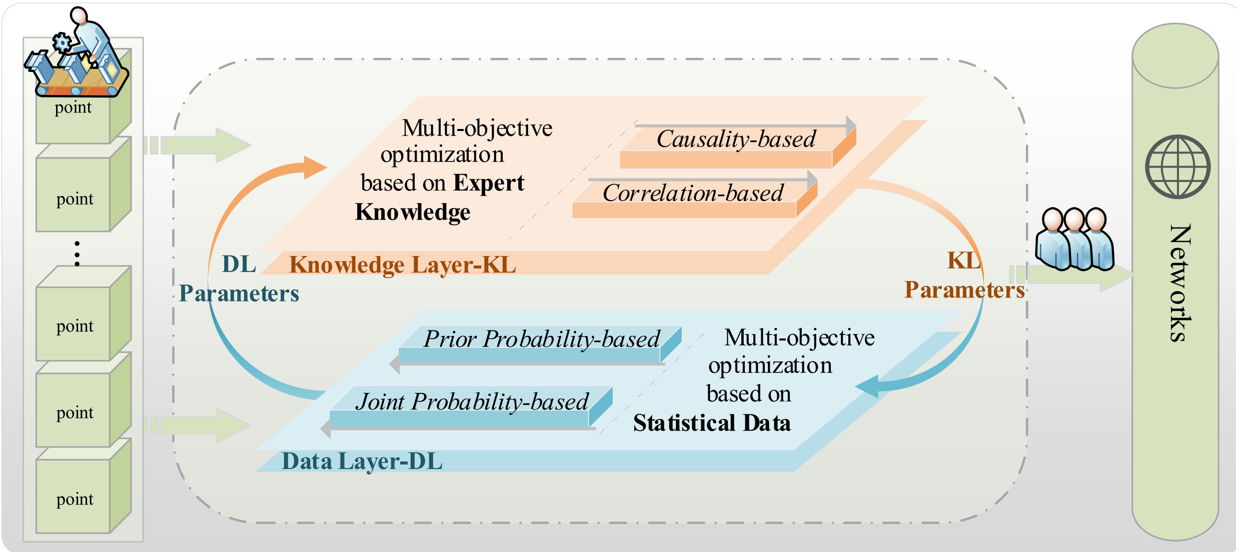

FIGURE 3. Hierarchical collaborative optimization framework.

expert knowledge. Similarly, the DL is based on statistical data to optimize the prior probability objective and joint probability objective. For realizing the collaboration between layers, the KL generates the optimized parameters during optimizing KL objectives, KL parameters. It becomes one of the parameters for calculating the fitness functions of DL-layer optimization. So the KL parameters can be considered as the collaborative parameters for optimization of DL. Likewise, the DL parameters generated by optimizing DL objectives are considered as collaborative parameters to participate the optimization of KL. Further, DL and KL are circularly iterated to output results. The optimized result is the structure of GBN considering fault points as variables. In other words, all uncertain variables, including potential fault points and fault phenomena, are constructed to GBN through the collaborative circulation optimization of KL and DL. Finally, under interventions of experts, the GBN is applied for fault identification in practical production process to reason the possible fault causes and give the best troubleshooting order for potential fault points when certain factors have failed.

To sum up, this framework optimizes the GBN structure based on expert knowledge and statistical data. It is worth mentioning that the hierarchical collaborative idea, as the key point, integrates knowledge in the field of expert systems and data of statistic science to optimize the same objectives. That means the collaborative optimization method combines the advantages and balances the disadvantages of knowledge and data. The method has the ability to hold the consistency of knowledge and data and solve problems on the collaboration of knowledge and data. In detail, it can avoid the potential contradiction of information through collaboration between layers and it also can change the collaborative weight to keep the unification on magnitude of information. Furthermore, the method not only considers the causality and correlation between variables, but also considers the independence of variables in terms of joint probability and prior probability.

C. HIERARCHICAL MULTI-OBJECTIVE OPTIMIZATION ALGORITHM

Essentially, the hierarchical collaborative optimization framework proposed needs to solve a hierarchical multi-objective optimization problem in the field of manufacturing [37], [38]. For multi-objective optimal problems, many scholars are committed to relevant research and build many superior algorithms, of which NSGA-II is the most common one [39]. As well known, the Non-dominated Sorting Genetic Algorithm-II (NSGA-II) is one of the most popular optimization algorithms, which has not only the ability of optimizing multiple goals with several decision variables and constraint conditions simultaneously but also the ability of finding global optimum solutions [40]. However, it cannot deal with the multi-objective collaborative optimization for multidiscipline intersection well, such as the expert system and probability statistics proposed in this study. For overcoming this difficulty, in previous works [16], a novel hierarchical optimization algorithm based on NSGA for matching the method designed, Hierarchical Non-dominated Sorting Genetic Algorithm (H-NSGA), was considered to solve the problem. In the past work, considering that the number of Pareto population is not vast for components of product platform, the proposed algorithm is advanced based on NSGA, instead of NSGA-II. Whereas, in this paper, the hierarchical collaborative multi-objective optimization algorithm, Hierarchical Non-dominated Sorting Genetic Algorithm II (HNSGA-II), is developed based on NSGA-II for fault identification in product service.

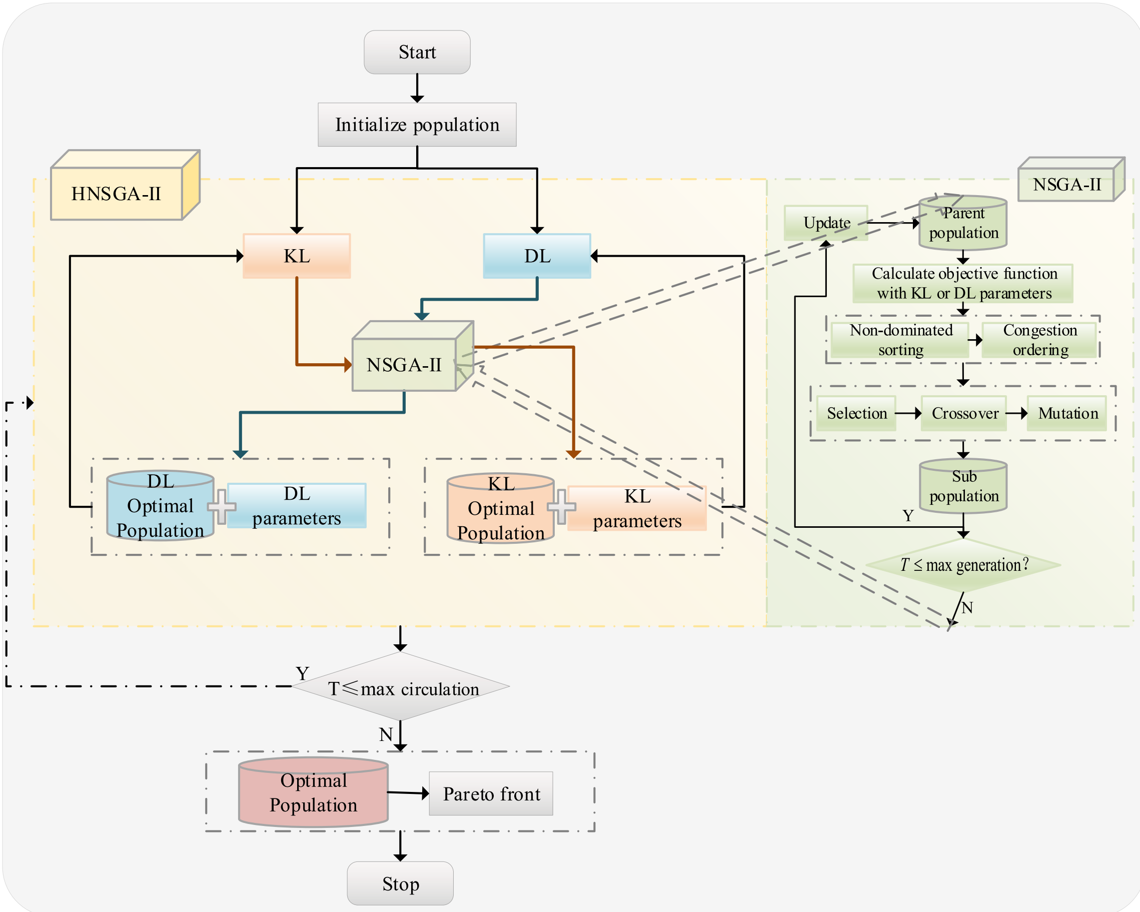

FIGURE 4. **HNSGA-II.**

The main part of this algorithm is the hierarchical cycle optimization based on the hierarchical collaborative optimization framework proposed. Through the multi-objective optimization of each layer based on NSGA-II, the HNSGA-II can achieve optimal results balancing objectives of different layers. The specific flow is shown in Fig4.

In HNSGA-II, NSGA-II is designed as the basic algorithm unit for hierarchical cycle optimization [41]. Firstly, initialized population is optimized by Knowledge Layer (KL) objectives with initialized DL parameters based on the model mentioned above to obtain the KL optimal population and KL parameters for collaboration. Next the KL optimal population is optimized by Data Layer (DL) objectives in the model with KL parameters, which can achieve the DL optimal population and DL parameters as the new parent population of KL and collaborative parameters for KL optimization. That means that the one layer gives the optimized population and collaborative parameter as new initialized population and optimal constraint parameters. Like this, KL and DL objectives are cyclically optimized with collaborative parameters until end of iteration. It is worth mentioning that the collaborative parameters are the important parts of the algorithm, which control the improvement of objectives on non-optimized layer so as to realize the collaboration between layers. Following the two layers' cycle optimization, the collaborative optimized solutions can be achieved. The specific steps of HNSGA-II are as follows:

- *Step 1 (Initialize population):* Do initial settings for this algorithm, including the initialized population with N individuals, collaborative parameters and so on;
- *Step 2 (KL optimization):* Get the parent population and collaborative parameters as the initial data. Carry out the optimization of parent population with DL parameters by NSGA-II for the multiple objectives of KL. The optimization generates KL optimized population and KL parameters as the new parent population and collaborative parameters for next step;
- *Step 3 (DL optimization):* Like the Step 2, make the DL optimization of DL objectives with KL parameters by NSGA-II and achieve the DL optimized population and DL parameters as the new parent population and collaborative parameters;
- *Step 4 (Hierarchical circulation optimization):* Repeat the optimization from step 2 to step 3. Considering the integration of multiple layers, circulation optimization is carried out to optimize population so as to get

the optimized population with objectives of different layers;
- *Step 5 (Terminate condition):* If the circulation generation is up to the set value, then get the final optimized population, which is the Pareto front. If not, do the circulation optimization;
- *NSGA-II:* After getting parent population and collaborative parameters, calculate the objective functions with collaborative parameters and do the non-dominated sorting based the value of functions. According to the sorting, make genetic evolution, including selection, crossover and mutation, to obtain the new population. Repeat the above options until the generation is up to the maximum. Finally, the optimized population is given.

Overall, HNSGA-II is designed as a hierarchical collaborative multi-objective optimization algorithm to solve the optimization problem between layers. It can achieve the optimized population with objectives of different layers as the solutions of model proposed, and it has the ability of collaborating expert knowledge and statistical data to construct GBN.

IV. FORMULATION OF METHODS

A. PROBLEM FORMULATION

The proposed methods and models are applied in the research of fault identification on the product service system. However, the data collected may be sparse, incomplete, or even with errors. And the knowledge may be one-sided and cannot describe the internal mechanism of models fully. Hence, the data and knowledge may contain overlapping or different information, which may be contradictory. The key point is how to solve problems driven by production data and expert knowledge collaboratively, meanwhile keeping the consistence of data and knowledge. In consequence, the problems framework is given in Fig5, which describes how to solve fault identification based on data and knowledge in this paper.

The possible fault points as variables are achieved from the work flows. First of all, the hierarchical collaborative optimization is run based on fault points. The expert knowledge and statistical data are recorded as basic information in practical production. The GBN is constructed driven by knowledge and data to achieve the basic Pareto Networks. In the case of expert intervention, the optimal set of GBN is given. When a fault occurs, the corresponding network is selected based on prior information to find the cause of fault and achieve the maintenance order so as to resume work. Meanwhile, the latest case will update database, knowledge base and fault points. And the optimized GBN will also be advanced based on the updated information.

B. FORMULATED DESIGN OF MODEL

For solutions to problems, possible fault points are considered as random variables, that are the nodes of GBN. While the relationships of nodes are defined by conditional probability, the types of directed edges in GBN are decided. The GBN are constructed to describe the causality and correlation of potential fault points in product service platform. For optimizing the structure of GBN, this paper designs the optimized objectives and collaborative parameters based on the proposed hierarchical collaborative framework, including KL and DL. Mathematical notations used in formulas are listed in Table2.

TABLE 2. The definition of notations.

Notation	Definition
$Diff$	The difference between optimized results and information source.
P_0	The probability of information source.
P	The probability of output.
P_{CP}	The probability of collaborative information.
P_r	The probability of optimized results.
x, y	The decision variables.
ω	The weight coefficient.
F	The number of possible fault points.
C	The probability of correlation.
N_r	The number of Pareto optimal individual.
f, l, i	The function probability coefficient, location probability coefficient and influence probability coefficient.
CP	The collaborative coefficient.
n, m	The serial number of variables.

The hierarchical collaborative framework aims to optimize the conditional probability of variables to build GBN. In particular, the GBN is optimized based on two layers, including expert knowledge and statistical data, where the different optimized objectives are compared based on different information source. Hence, the basic principles of different objectives for optimization are all defined as minimizing differences between optimized results and information source. Detailedly, in KL, the structure is optimized based on expert knowledge, and optimal objectives consist of causality and correlation of variables. Firstly, the difference based on causality is defined by (2), in which ω_{nm} is defined to describe the weight of relationship of point n and m in product service platform. In addition, $Diff_{CP}$ is the collaborative difference based on collaborative parameters, and it will be introduced later. Considering the information source may be incomplete, $x_{KL1_{nm}}$ is defined as a decision variable. When the expert knowledge $P_0(n|m)$ is available, $x_{KL1_{nm}}$ is set as 1, otherwise 0.

$$Diff_{KL1} = Diff_{CP} + \frac{\sum_{n=1}^{F}\sum_{m=1}^{F} \omega_{nm} \cdot |P(n|m) - P_0(n|m)| \cdot x_{KL1_{nm}}}{\sum_{n=1}^{F}\sum_{m=1}^{F} x_{KL1_{nm}}} \tag{2}$$

Before defining the difference of correlation between variables, the correlation needs to be designed to meet the problem of fault identification for product service. According to the practical engineering, the correlation is concerned with functions, locations and influence of factors. Thus, the correlation is defined in (3) based on function probability coefficient, location probability coefficient and influence probability coefficient, in which ω_{nm} is designed as the weight of coefficients.

$$C = \omega_f \cdot f + \omega_l \cdot l + \omega_i \cdot i \tag{3}$$

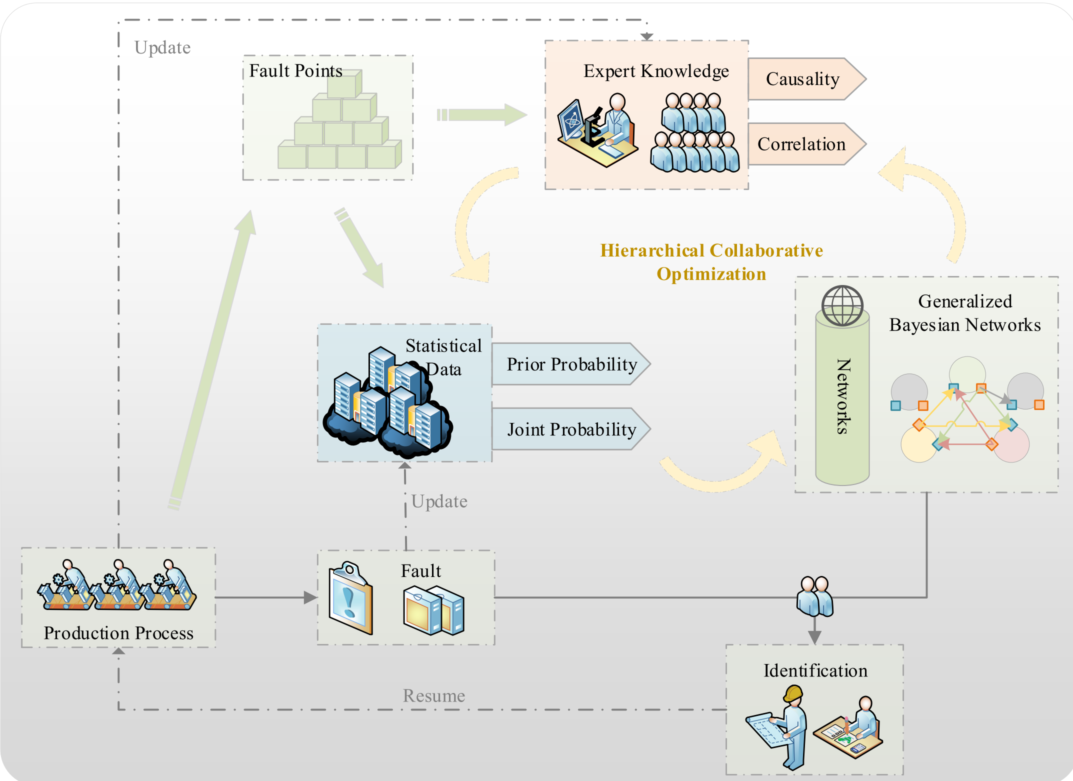

FIGURE 5. **The problem framework.**

Actually, the correlation defined is equivalent to probability which describes the influence between variables. In the same fashion, the difference based on correlation between variables is defined by (4).

$$Diff_{KL2} = Diff_{CP} + \frac{\sum_{n=1}^{F}\sum_{m=1}^{F}\omega_{nm}\cdot|P(n|m)-C(n|m)|\cdot x_{KL2_{nm}}}{\sum_{n=1}^{F}\sum_{m=1}^{F}x_{KL2_{nm}}} \tag{4}$$

In contrast, the DL optimizes the structure of GBN based on statistical data, of which optimal objectives are the difference of prior probability and joint probability. The difference between results and prior probability is designed by (5) based on Bayesian formula, $P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}$.

$$Diff_{DL1} = Diff_{CP} + \frac{\sum_{n=1}^{F}\sum_{m=1}^{F}\omega_{nm}\cdot\left|\frac{P_0(n)}{P_0(m)}-\frac{P(n|m)}{P(m|n)}\right|\cdot x_{DL1_{nm}}}{\sum_{n=1}^{F}\sum_{m=1}^{F}(x_{DL1_{nm}}+y_{DL1_{nm}})} + \frac{\sum_{n=1}^{F}\sum_{m=1}^{F}\omega_{nm}\cdot|P_0(m)-P(m|n)|\cdot y_{DL1_{nm}}}{\sum_{n=1}^{F}\sum_{m=1}^{F}(x_{DL1_{nm}}+y_{DL1_{nm}})} \tag{5}$$

In practical application, it is concerned that the data may be incomplete, that means the variables P_0 or P may be 0. So the decision parameters $x_{DL1_{nm}}$ and $y_{DL1_{nm}}$ are designed as follows:

$$x_{DL1_{nm}} = \begin{cases}1, & P_0(m)\neq 0\,\&\ \ P(m|n)\neq 0\\ 0, & otherwise\end{cases} \tag{6}$$

$$y_{DL1_{nm}} = \begin{cases}1, & P_0(n)\neq 0\,\&\ \ P_0(m)=0\,\&\ \ P(n|m)=0\\ 0, & otherwise\end{cases} \tag{7}$$

Similarly, according to joint probability formula, $P(A\cdot B) = P(A)\cdot P(B|A)$, the difference of joint probability is given by (8), in which $x_{DL2_{nm}}$ is considered as the decision variable. Only when $P_0(n)\neq 0$, $x = 1$, otherwise, $x = 0$.

$$Diff_{DL2} = Diff_{CP} + \frac{\sum_{n=1}^{F}\sum_{m=1}^{F}\omega_{nm}\cdot|P_0(n\cdot m)-P_0(n)\cdot P(m|n)|\cdot x_{DL2_{nm}}}{\sum_{n=1}^{F}\sum_{m=1}^{F}x_{DL2_{nm}}} \tag{8}$$

Especially, the difference of collaboration based on collaborative parameters, $Diff_{CP}$, exists in all formulas of optimized objectives. The $Diff_{CP}$ is a key parameter, which shows the collaboration of hierarchical optimization between layers, defined by (9), in which CP is defined as the collaborative coefficient to adjust the collaborative influence between layers. The larger CP, the greater influence.

$$Diff_{CP} = \frac{\sum_{n=1}^{F}\sum_{m=1}^{F}\omega_{nm}\cdot|P_{CP}(n|m)-P(n|m)|}{F^2}\cdot CP \tag{9}$$

In particular, P_{CP} is designed as the collaborative standard value by (10). P_{CP} of KL is calculated based on the optimal result P_r of DL. Likewise, the optimal result of KL will be used to compute the collaborative standard value P_{CP} of DL. Essentially, the probability of collaborative parameters is defined based on the expected value of all Pareto optimal individuals. It describes the average distribution of the optimized population.

$$P_{CP}(n|m) = \frac{\sum_{i=1}^{N_r} P_{r_i}(n|m)}{N_r} \tag{10}$$

V. CASE STUDY

A. BACKGROUND OF PROBLEM

The SAC, Guodian Nanjing Automation Co., Ltd., focuses on R&D and product services for automation and information technology, meanwhile, it hopes to develop a series of automation businesses, including smart grid automation, power plant and industrial automation, rail transit automation, information and security technology, power electronics and so on. With the centennial opportunity of smart grid as the leading force of low-carbon economy globally, the SAC actively pushes forward its transformation of development mode and structural optimization. As a result, the SAC wants to develop the product service platform for achieving intelligent maintenance of products, in which fault identification is one of functions. When a certain factor has failed, the enterprise needs to quickly identify the cause of failures and relevant potential fault points, next to find the best maintenance sequence to eliminate failures and recover operation of equipment as soon as possible. According to this plan, a case study of fault cause identification in the field of smart grid for SAC is reported to demonstrate the proposed models and algorithm.

The HERE platform of SAC is a comprehensive platform that serves operations of smart grid daily, in which the problem record and engineering service record of products are two functions of product service. From 2017 to 2019, the platform records nearly 35,000 pieces of information, which will be considered as the basic information for this study. In order to better explain the information form, the Table3 lists several pieces of information about the fault description and causes of faults as the brief examples for information sources.

For the purpose of demonstrating research, the case study takes some transformer substations as examples to research, and it only considers the product service processes of which the first classification is Installation and the second classification is Hardware. In detail, the study chooses 10 potential fault points, simplified named from A to J for protecting the information of enterprises. It represents the 10 components of products which may fail. According to the types and the number of faults, the prior probability and joint probability data of fault points can be computed, listed in Table4. In the table, the value of Pr is the prior probability, and other data is the joint probability. In addition, the causality and correlation coefficient including function probability coefficient, location probability coefficient and influence probability coefficient are calculated based on the knowledge expression from technical staff and experts of SAC. And the coefficients are expressed based on 0-1 scoring system, in which 0 is defined as no relationships and 1 means the equal relationship. The coefficients are between 0 and 1. The results of knowledge expression are shown as causality knowledge and correlation knowledge in Table5 and Table6. Specially, in the tables, x expresses that the information is lacking. Because the expert information for internal models is limited and inadequate in practical engineering, the causality knowledge and correlation knowledge are sparse.

B. SOLUTIONS AND RESULTS

Based on the information sources, other variables including the collaborative coefficient CP and all weight parameters are given by experts of SAC based on expert knowledge. Furthermore, other parameters of HNSGA-II are set based on the optimal algorithm for experiments. The number of initial population is 20. The maximal generation and hierarchical circulation are 50 and 5. And the crossover and mutation rate are 0.8 and 0.1. The algorithm is run by Matlab 2016a on the system, macOS High Sierra 10.13.1, 1.6GHz Inter Core i5, 8GB 1600MHz DDR3. In the current environment, the algorithm only takes a few seconds to run. Considering DL optimization as the main optimal process, which means the hierarchical cycle optimization jumps out from DL to output results, the optimized results are shown in Fig6 and Fig7.

Fig6 shows the Pareto optimization population considering the collaborative difference, which takes the DL as the main process (DL-optimization). That means the prior probability (DL1) and joint probability (DL2) objectives are the main optimal objectives, and the causality (KL1) and correlation (KL2) based on KL are considered as the influence factors. The picture gives the value of $Diff_{DL1}$ and $Diff_{DL2}$ for initial population and Pareto population. It can be seen that the model and algorithm are successful to optimize objectives and achieve Pareto front. However, it is worth noting that some points are out of Pareto front. That is because the optimized objectives contain the collaborative difference $Diff_{CP}$, and the reported results have been added to the difference. In addition, the optimal results are based on the collaborative optimization, which means the algorithm has to consider the objectives of KL while optimizing the objectives of DL.

When calculating objective functions of Pareto population without the collaborative difference, the results are shown in the left picture of Fig7. The left one shows $Diff_{DL1}$ and $Diff_{DL2}$ of the optimized population and initial population without the collaborative difference $Diff_{CP}$. According to it, the difference of optimal population is smaller than initial population for DL objectives. Moreover, the right one shows that the difference of KL, $Diff_{KL1}$ and $Diff_{KL2}$. Compared to initial population, the difference of optimal population is smaller than initial population. According to these results, the objectives of KL are improved in the hierarchical optimized process with setting the DL as the main optimized objectives.

TABLE 3. Information source.

Transformer Substation	First Classification	Second Classification	Fault Description	Cause Identification
Xiaodong	Installation	Hardware	Measurement and control digital samplings have been interrupted.	Replace the device template
Sibei	Installation	Hardware	The optical fiber has been unstable.	Replace the CPU
Zhenjiang	Installation	Software	The data of electric current has been displayed as 0.	Upgrade measurement and control system
Tianjing	Design	Engineering	The terminal equipment has shown an error.	Wiring error

TABLE 4. Prior probability and joint probability of statistical data.

	A	B	C	D	E	F	G	H	I	J
Pr	0.250	0.058	0.066	0.185	0.310	0	0.130	0.360	0.370	0.150
A										
B	0.023									
C	0	0.021								
D	0.160	0	0							
E	0.026	0	0.002	0.122						
F	0	0	0	0	0					
G	0.110	0.050	0.009	0	0.031	0				
H	0.240	0.014	0	0.162	0.025	0	0.005			
I	0.170	0	0	0	0.256	0	0.015	0.350		
J	0	0.022	0.058	0	0	0	0.005	0	0.085	

TABLE 5. Causality knowledge.

	A	B	C	D	E	F	G	H	I	J
A	x	0	0.8	0	0	x	x	0.1	0.7	0
B	0.75	x	x	x	0	0.82	x	x	0	x
C	0	x	x	0	x	x	0.32	0.54	x	0
D	0.3	x	0.95	x	0	0.31	x	x	x	0
E	0.6	0.5	x	0.85	x	x	0	x	0	x
F	x	0	x	0	x	x	0	0	0	x
G	x	x	0	x	0.98	0.77	x	x	x	x
H	0	x	0	x	x	0.96	x	x	0	x
I	0	0.62	x	x	0.57	0.2	x	0.6	x	0.88
J	0.9	x	0.65	0.78	x	x	x	x	0	x

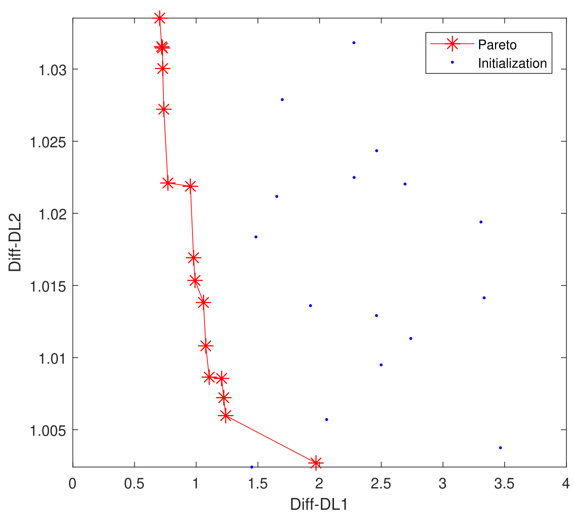

FIGURE 6. The Pareto of DL-optimization with the collaborative difference.

Compared with initial population, the main optimized objectives (DL1 and DL2) are improved to achieve the Pareto front, and the other objectives (KL1 and KL2) are also optimized. This is the effect of the hierarchical collaboration idea in the algorithm. The overall results indicate the hierarchical

TABLE 6. Correlation knowledge.

	A	B	C	D	E	F	G	H	I	J
A	x	0.84	0	0	0.56	x	0.67	0.49	0.78	x
B	x	x	x	x	x	0.47	0.23	x	x	0.32
C	0.96	0.62	x	0.52	0.20	x	0.94	0.51	0.25	x
D	0.25	x	0.32	x	x	0.34	x	0	x	0.14
E	x	x	x	x	x	0.94	0.01	0	0.91	0.31
F	0.41	0	x	x	0	x	x	0.27	x	x
G	0.98	0.66	0.91	0.70	0.87	0.69	x	0.83	x	0.58
H	x	x	x	x	0.14	x	0.59	x	0.08	0.17
I	0	0.83	0	0.62	x	0.25	x	0.84	x	0.45
J	x	x	0.07	x	0.93	0.77	0.37	x	0.03	x

collaborative optimization framework can optimize the GBN based on the collaboration of knowledge and data effectively.

Likewise, Fig8 and Fig9 report the optimization process which makes KL as the main process (KL-optimization). Fig8 shows the optimized results with collaborative difference and gives the difference of causality (KL1) and correlation (KL2) for initial population and Pareto population. Furthermore, in Fig9, the left figure displays the KL1 and KL2 difference functions of initial population and optimized population without the collaborative difference, and the other one indicates that the DL1 and DL2 difference functions are also optimized based on the hierarchical collaborative optimization. Similarly, the network is successfully optimized based on the collaboration of knowledge and data. On the whole, the case effectively optimizes the initial population and achieves the Pareto front, whether the DL or KL is considered as the main process. Moreover, the other objectives are also improved collaboratively. The current results can illustrate the hierarchical collaborative algorithm HNSGA-II is effective for the proposed model. Besides, it is worth noting that there are lots of vacancy information and 0 value in the basic information source, but they have no effects on results due to the decision parameters x and y. The decision parameters are designed to eliminate the impact of inadequate information so as to guarantee robustness of the algorithm. On the other hand, it makes the algorithm more scalable and has the ability to accommodate various types of information.

It should be mentioned that the model is designed to solve problems of fault identification. Hence, this case considers DL as the main process to explain how it works for fault identification of SAC. Now, it assumes that the possible fault point A has been faulty. The proposed model will give the cause of this fault and maintenance order based on the results of experiments considering DL as the main objectives. According to results, we have achieved the Pareto front with

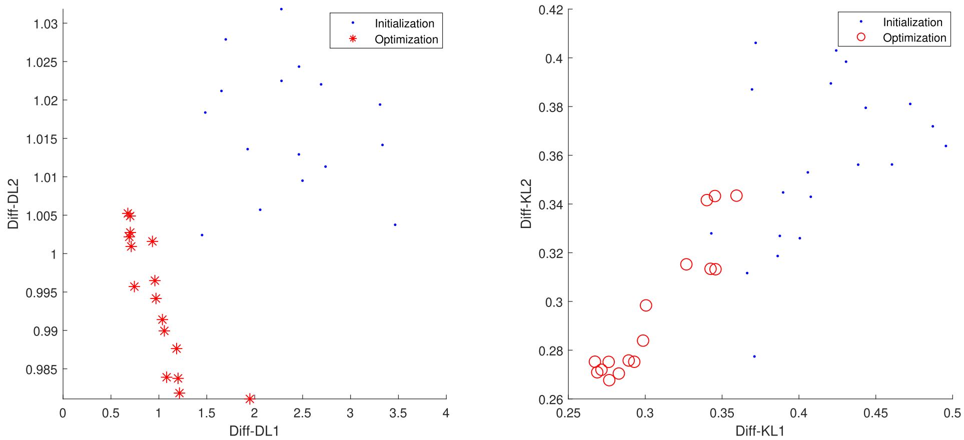

FIGURE 7. **The DL-optimization functions without the collaborative difference.**

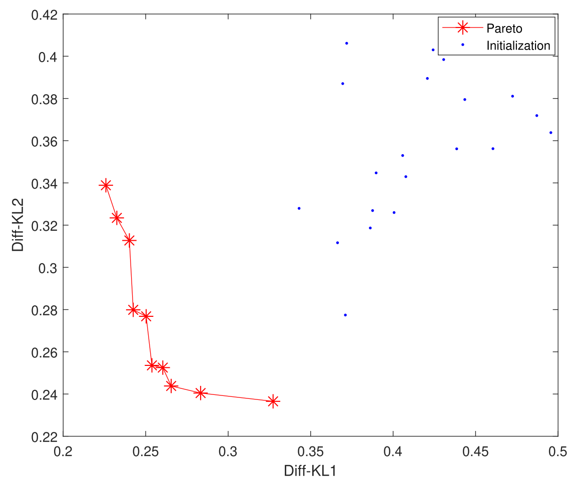

FIGURE 8. **The Pareto of KL-optimization with the collaborative difference.**

16 individuals. Considering the actual characteristics of component A, experts select the corresponding individual based on expert knowledge as the compositions of GBN by (11), where the weights of objectives for Pareto front are decided by experts in SAC.

$$Diff = \omega_{KL1} \cdot Diff_{KL1} + \omega_{KL2} \cdot Diff_{KL2} + \omega_{DL1} \cdot Diff_{DL1} + \omega_{DL2} \cdot Diff_{DL2} \quad (11)$$

Under the intervention of experts, the 1st Pareto individual has the minimized difference, which will be as the GBN parameters to find the cause of the fault and maintenance order for component A. The 1st Pareto individual is shown in Table7, but it is not the final GBN. Experts will adjust the results based on expert knowledge to give the final GBN, as shown in Fig10.

As shown, the GBN gives the possible causes and maintenance order for the fault of A. The most possible cause and the first order to test will be $G \rightarrow F \rightarrow B \rightarrow C$. And the next order will be decided by the results of last test until the work can be resumed. If G is failed in test, next order $F \rightarrow H$ will be inspected. In general, the GBN gives all possible causes and probability for the fault of A. So the experiments confirm that the hierarchical collaborative model can deal with fault identification problems driven by statistical data and expert knowledge.

C. COMPARATIVE EXPERIMENT

Compared with the initial population, the above optimal results display the effectiveness of optimization based on the optimization of main functions and shows the ability of collaborative optimization based on the optimization of assistant functions. In order to demonstrate the effect of hierarchical collaboration well, the comparative experiments are designed in the same conditions and parameters, which optimize the population with the single multi-objective algorithm, NSGA-II, instead of the hierarchical optimization algorithm. By comparing the secondary objectives between the single optimization and hierarchical collaborative method, the effects of hierarchical collaboration can be revealed.

The comparative results are shown in Fig11 and Table8. Fig11 shows the KL objective functions ($Diff_{KL1}$ and $Diff_{KL2}$) value of optimal population. In detail, the single optimization takes DL as optimized objectives based on NSGA-II. And the hierarchical collaborative optimization takes DL as the main process to optimize objectives based on HNSGA-II. Similarly, Table8 gives the DL functions ($Diff_{DL1}$ and $Diff_{DL2}$) value of optimal population taking KL as the main optimized objectives for single optimization (S-DL1 and S-DL2) and hierarchical collaborative optimization (H-DL1 and H-DL2).

According to the results shown, the KL values of hierarchical method are more optimized than the single model in Fig11, and the DL value of hierarchical method is better improvement than single model in Table8. Compared to single optimization, the optimal population of hierarchical collaborative optimization has the better performance on collaborative objectives. The results illustrate the effect of hierarchical collaboration. In other words, the hierarchical collaborative method greatly improves the performance of collaborative functions while optimizing main objectives.

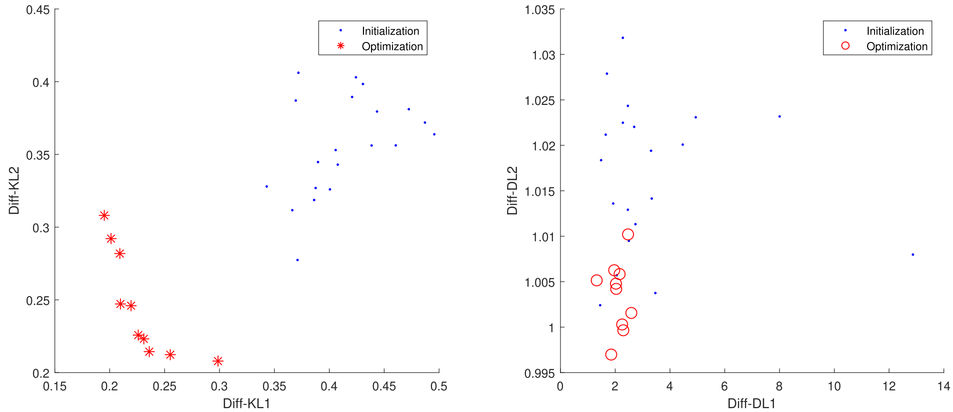

FIGURE 9. The KL-optimization functions without the collaborative difference.

TABLE 7. The GBN parameters for the case.

	A	B	C	D	E	F	G	H	I	J
A	1	0.946	0.903	0.760	0.380	0.948	0.968	0.100	0.716	0.379
B	0.692	1	0.806	0.109	0.013	0.705	0.305	0.069	0.120	0.169
C	0.201	0.363	1	0.255	0.418	0.580	0.593	0.057	0.161	0.200
D	0.456	0.334	0.929	1	0.985	0.837	0.409	0.564	0.706	0.717
E	0.387	0.039	0.683	0.982	1	0.298	0.614	0.638	0.614	0.630
F	0.034	0.164	0.455	0.199	0.504	1	0.567	0.037	0.534	0.444
G	0.712	0.449	0.624	0.529	0.273	0.687	1	0.759	0.219	0.395
H	0.252	0.469	0.221	0.937	0.607	0.704	0.514	1	0.892	0.751
I	0.405	0.360	0.249	0.931	0.540	0.289	0.441	0.918	1	0.378
J	0.267	0.349	0.379	0.361	0.459	0.201	0.446	0.559	0.372	1

TABLE 8. The comparative results of KL-optimization.

S-DL1	4.7687	4.8213	8.6076	2.8114	3.8131
	3.8404	2.6125	2.2939	1.6649	2.1109
	2.1167	2.1779	1.8558	2.1046	2.3665
S-DL2	1.0129	1.0130	1.0074	1.0138	1.0152
	1.0149	1.0141	1.0125	1.0064	1.0057
	1.0060	1.0071	1.0059	1.0078	1.0045
H-DL1	2.5884	2.0359	1.9629	2.4686	2.1610
	1.8536	2.2946	2.2530	2.0264	1.3238
H-DL2	1.0016	1.0042	1.0063	1.0102	1.0058
	0.9970	0.9997	1.0003	1.0048	1.0051

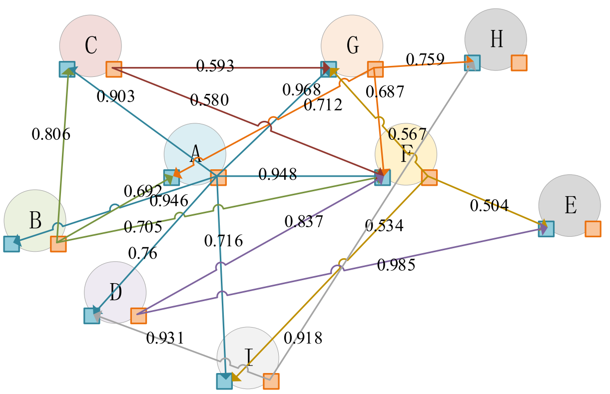

FIGURE 10. GBN for the case.

In particular, for demonstrating the effects of circular optimization, the other experiment is designed, of which results are shown in Fig12. In this case, the number of circulation generation is set as 5. The comparative experiment outputs the total difference of Pareto population by (11) for each circulation, where labels of different colors represent the current generation number. According to results, with the increase of cyclic generation, the performance of total difference is improved. The results of this experiment illustrate that the circular optimization improves the performance of objectives.

Overall, the model and algorithm proposed not only can optimize the main objective functions, but also improve the collaborative objectives. It can effectively solve the hierarchical collaborative multi-objective optimization problems for GBN. Moreover, compared to related works, the novel method has the ability to collaborate knowledge and data so as to achieve the results of identification in manufacturing. The comparative experiment explains that the results of hierarchical collaborative optimization are better than the traditional methods. In addition, in the actual production information, the mistake and lacking of information are very common. There exists great uncertainty in the basic information. The collaborative of two information sources has the ability to complement each other to find the causes of faults based on the uncertain information.

D. APPLICATIONS AND POSSIBLE EXTENSIONS

This paper is studied based on practical engineering programs of SAC for fault identification in product service platform [42]. Finally, the research will be used for building the product service system of SAC and be applied in fault identification service. The system serves the product service process and focuses on maintenance, overhaul, identification and test in production workflow [43], which is one of the indispensable steps for enterprises to achieve intelligent manufacturing [44]. The intelligent services of products on fault identification can intelligently troubleshoot the cause of failures and reduce the loss caused by failures of factors [45].

However, there are still some limitations in this study. In the model of fault identification, lots of expert knowledge

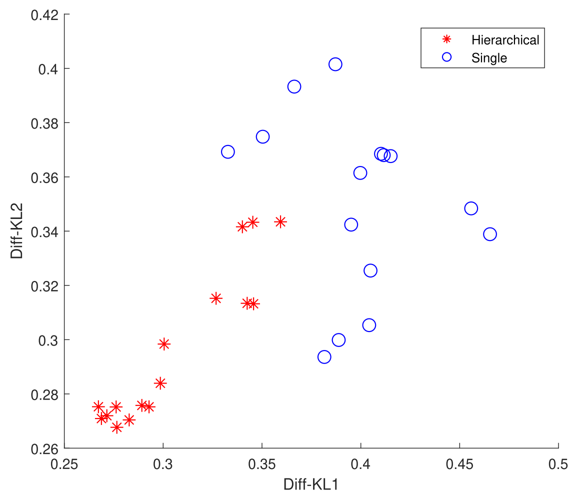

FIGURE 11. The comparative results of DL-optimization.

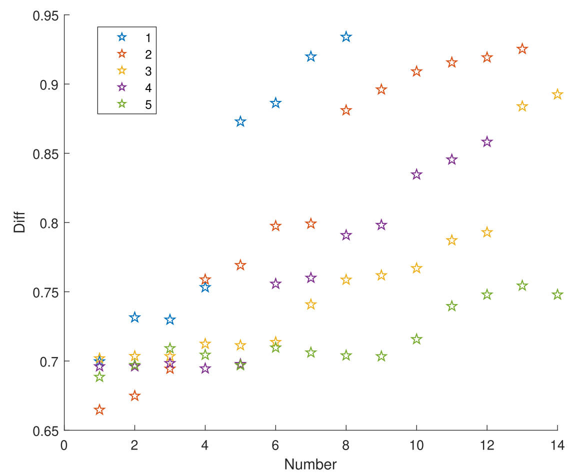

FIGURE 12. The comparative results of circulation.

need to be recorded and many important parameters, such as weights and collaborative parameters, are decided by technical staff. They are influenced by human factors and have great subjectivity. For solutions to it, a feedback mechanism needs to be designed in future works to optimize the technical parameters based on results of models so as to ensure objectivity and accuracy of methods. On the other hand, the proposed GBN has only addressed the questions of adjacent variables. This paper only introduces the definition and application of GBN, where the conditional probability between adjacent variables is used for analysing the causality and correlation of variables. In the future, more theoretical research on generalized Bayesian Network will be studied, for example, how to achieve the probabilistic reasoning between variables based on the proposed GBN.

VI. CONCLUSION

This study focuses on fault identification for product service in the field of industry and proposes a new sight to solve it, that is, to construct the generalized Bayesian Network driven by statistical data and expert knowledge. The GBN has the ability to describe all relationships of variables, including the independence, the causality and the correlation. Furthermore, the features of GBN enable it to be constructed by both information sources, including data and knowledge. For addressing it, this paper develops a hierarchical collaborative framework, which consists of the data layer and the knowledge layer. Each layer is considered as an independent multi-objective optimization process, while the collaborative parameters generated by the optimized results of one layer will participate in the optimization of another one. Hence, the framework can realize the collaboration between data and knowledge while maintaining their consistency. Further, a hierarchical multi-objective optimization algorithm, HNSGA-II, is developed to solve it.

A case study of fault identification for product service has been given. The results of experiments show the optimized Pareto front and give the causes and maintenance order in the case when one component has failed. In addition, compared the single multi-objective optimization with hierarchical collaborative optimization, the model and algorithm proposed can not only optimize the main objective functions, but also improve other assistant functions. And according to the outputs of each circulation, it is illustrated that circular optimization improves the performance of objectives. Everything considered, with statistical analysis and knowledge inference, the method based on the GBN is a powerful tool for fault identification in practice.

ACKNOWLEDGMENT

The authors would like to thank the SAC in Nanjing for their supports for data and knowledge about the product service.

REFERENCES

[1] T. DebRoy, H. L. Wei, T. Mukherjee, J. W. Elmer, J. O. Milewski, A. M. Beese, A. Wilson-Heid, A. De, and W. Zhang, "Additive manufacturing of metallic components–process, structure and properties," *Prog. Mater. Sci.*, vol. 92, pp. 112–224, Feb. 2018.

[2] R. Y. Zhong, X. Xu, E. Klotz, and S. T. Newman, "Intelligent manufacturing in the context of industry 4.0: A review," *Engineering*, vol. 3, no. 5, pp. 616–630, 2017.

[3] T. Pan, J. Chen, J. Xie, Y. Chang, and Z. Zhou, "Intelligent fault identification for industrial automation system via multi-scale convolutional generative adversarial network with partially labeled samples," *ISA Trans.*, vol. 101, pp. 379–389, Jun. 2020.

[4] X. Yan, Y. Liu, and M. Jia, "Multiscale cascading deep belief network for fault identification of rotating machinery under various working conditions," *Knowl.-Based Syst.*, vol. 193, Apr. 2020, Art. no. 105484.

[5] R. H. C. Palácios, I. N. da Silva, A. Goedtel, and W. F. Godoy, "A comprehensive evaluation of intelligent classifiers for fault identification in three-phase induction motors," *Electr. Power Syst. Res.*, vol. 127, pp. 249–258, Oct. 2015.

[6] H. H. Yue and S. J. Qin, "Reconstruction-based fault identification using a combined index," *Ind. Eng. Chem. Res.*, vol. 40, no. 20, pp. 4403–4414, 2001.

[7] A. T. Bouloutas, S. Calo, and A. Finkel, "Alarm correlation and fault identification in communication networks," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 523–533, Feb. 1994.

[8] A. Giret, E. Garcia, and V. Botti, "An engineering framework for service-oriented intelligent manufacturing systems," *Comput. Ind.*, vol. 81, pp. 116–127, Sep. 2016.

[9] J. Patalas-Maliszewska and M. Śliwa, "An approach to tacit knowledge classification in a manufacturing company," *Tehnički Vjesnik*, vol. 26, no. 3, pp. 650–657, 2019.

[10] S. Shin, S. Hyun, Y.-J. Shin, J. Song, and D.-H. Bae, "Uncertainty based fault type identification for fault knowledge base generation in system of systems," in *Proc. 16th Int. Conf. Syst. Syst. Eng. (SoSE)*, Jun. 2021, pp. 216–221.
[11] P. Smart, E. Simperl, and N. Shadbolt, "A taxonomic framework for social machines," in *Social Collective Intelligence*. Berlin, Germany: Springer, 2014, pp. 51–85.
[12] J. Pearl, "Causality: Models," in *Reasoning Inference*, vol. 29. Berlin, Germany: Springer, 2000.
[13] A. C. Constantinou, N. Fenton, W. Marsh, and L. Radlinski, "From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support," *Artif. Intell. Med.*, vol. 67, pp. 75–93, Feb. 2016.
[14] S. Hosseini and K. Barker, "A Bayesian network model for resilience-based supplier selection," *Int. J. Prod. Econ.*, vol. 180, pp. 68–87, Oct. 2016.
[15] I. Chaturvedi, E. Ragusa, P. Gastaldo, R. Zunino, and E. Cambria, "Bayesian network based extreme learning machine for subjectivity detection," *J. Franklin Inst.*, vol. 355, no. 4, pp. 1780–1797, Mar. 2018.
[16] Q. Wan, J. Zhao, J. Xue, C. Zhang, B. He, and H. Zhang, "A novel hierarchical collaborative method based on multi-objective optimization for modularization of product platform," in *Proc. IEEE 23rd Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, May 2019, pp. 215–220.
[17] S. Kiranyaz, A. Gastli, L. Ben-Brahim, N. Al-Emadi, and M. Gabbouj, "Real-time fault detection and identification for MMC using 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8760–8771, Nov. 2019.
[18] R. Dunia and S. J. Qin, "Subspace approach to multidimensional fault identification and reconstruction," *AIChE J.*, vol. 44, no. 8, pp. 1813–1831, 1998.
[19] T. Kurtoglu and I. Y. Tumer, "A graph-based fault identification and propagation framework for functional design of complex systems," *J. Mech. Des.*, vol. 130, no. 5, May 2008, Art. no. 051401.
[20] W. C. Santos, F. V. Lopes, N. S. D. Brito, and B. D. Souza, "High-impedance fault identification on distribution networks," *IEEE Trans. Power Del.*, vol. 32, no. 1, pp. 23–32, Feb. 2017.
[21] B. S. Lindner and L. Auret, "Data-driven fault detection with process topology for fault identification," *IFAC Proc. Volumes*, vol. 47, no. 3, pp. 8903–8908, 2014.
[22] H. A. Izadi, B. W. Gordon, and Y. Zhang, "A data-driven fault tolerant model predictive control with fault identification," in *Proc. Conf. Control Fault-Tolerant Syst. (SysTol)*, Oct. 2010, pp. 732–737.
[23] M. Alauddin, F. Khan, S. Imtiaz, and S. Ahmed, "A variable mosquito flying optimization-based hybrid artificial neural network model for the alarm tuning of process fault detection systems," *Process Saf. Prog.*, vol. 39, Mar. 2020, Art. no. e12122.
[24] M. T. Amin, F. Khan, S. Ahmed, and S. Imtiaz, "A novel data-driven methodology for fault detection and dynamic risk assessment," *Can. J. Chem. Eng.*, vol. 98, no. 11, pp. 2397–2416, 2020.
[25] Y. Zhou, N. Fenton, and M. Neil, "Bayesian network approach to multinomial parameter learning using data and expert judgments," *Int. J. Approx. Reasoning*, vol. 55, no. 5, pp. 1252–1268, Jul. 2014.
[26] M. A. Helal and M. Mouhoub, "A probabilistic model for automobile diagnosis system: Combining Bayesian estimator and expert knowledge," in *Proc. Int. Conf. Ind., Eng. Other Appl. Appl. Intell. Syst.* Berlin, Germany: Springer, 2018, pp. 257–264.
[27] J. Li and J. Shi, "Knowledge discovery from observational data for process control using causal Bayesian networks," *IIE Trans.*, vol. 39, no. 6, pp. 681–690, 2007.
[28] Q. Zhang and Q. Yao, "Dynamic uncertain causality graph for knowledge representation and reasoning: Utilization of statistical data and domain knowledge in complex cases," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1637–1651, May 2018.
[29] A. C. Constantinou, N. Fenton, and M. Neil, "Integrating expert knowledge with data in Bayesian networks: Preserving data-driven expectations when the expert variables remain unobserved," *Expert Syst. Appl.*, vol. 56, pp. 197–208, Sep. 2016.
[30] A. N. Terent'yev and P. I. Bidyuk, "Method of probabilistic inference from learning data in Bayesian networks," *Cybern. Syst. Anal.*, vol. 43, no. 3, pp. 391–396, May 2007.
[31] A. Sata and B. Ravi, "Bayesian inference-based investment-casting defect analysis system for industrial application," *Int. J. Adv. Manuf. Technol.*, vol. 90, nos. 9–12, pp. 3301–3315, Jun. 2017.
[32] D. Codetta-Raiteri and L. Portinale, "Dynamic Bayesian networks for fault detection, identification, and recovery in autonomous spacecraft," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 1, pp. 13–24, Jan. 2015.
[33] M. Turkoz, S. Kim, Y.-S. Jeong, M. K. Jeong, E. A. Elsayed, K. N. Al-Khalifa, and A. M. Hamouda, "Bayesian framework for fault variable identification," *J. Qual. Technol.*, vol. 51, no. 4, pp. 375–391, 2018.
[34] T. Silander and P. Myllymaki, "A simple approach for finding the globally optimal Bayesian network structure," 2012, *arXiv:1206.6875*.
[35] N. Friedman, I. Nachman, and D. Pe'er, "Learning Bayesian network structure from massive datasets: The 'sparse candidate' algorithm," 2013, *arXiv:1301.6696*.
[36] H. S. Sousa, F. Prieto-Castrillo, J. C. Matos, J. M. Branco, and P. B. Lourenço, "Combination of expert decision and learned based Bayesian networks for multi-scale mechanical analysis of timber elements," *Expert Syst. Appl.*, vol. 93, pp. 156–168, Mar. 2018.
[37] Z. Qiao and X. Shu, "Coupled neurons with multi-objective optimization benefit incipient fault identification of machinery," *Chaos, Solitons Fractals*, vol. 145, Apr. 2021, Art. no. 110813.
[38] C. Yang and M. Jia, "Hierarchical multiscale permutation entropy-based feature extraction and fuzzy support tensor machine with pinball loss for bearing fault identification," *Mech. Syst. Signal Process.*, vol. 149, Feb. 2021, Art. no. 107182.
[39] Q. Wan, G. Peng, J. Zhao, and H. Zhang, "A multi-objective collaborative optimization algorithm based on cycle subsystem model with features for product family design," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2018, pp. 3614–3619.
[40] J. Gong, X. Zhu, Q. Hu, Y. Hu, and B. Li, "A fast optimized algorithm based on the NSGA—II for microwave Windows," in *Proc. 18th Int. Vac. Electron. Conf. (IVEC)*, Apr. 2017, pp. 1–2.
[41] A. Kamjoo, A. Maheri, A. M. Dizqah, and G. A. Putrus, "Multi-objective design under uncertainties of hybrid renewable energy system using NSGA-II and chance constrained programming," *Int. J. Elect. Power Energy Syst.*, vol. 74, pp. 187–194, Jan. 2016.
[42] M. Matsas, G. Pintzos, A. Kapnia, and D. Mourtzis, "An integrated collaborative platform for managing product-service across their life cycle," *Proc. CIRP*, vol. 59, pp. 220–226, Jan. 2017.
[43] H. Cai, L. D. Xu, B. Xu, C. Xie, S. Qin, and L. Jiang, "IoT-based configurable information service platform for product lifecycle management," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1558–1567, May 2014.
[44] W. Shen, Q. Hao, H. J. Yoon, and D. H. Norrie, "Applications of agent-based systems in intelligent manufacturing: An updated review," *Adv. Eng. Informat.*, vol. 20, no. 4, pp. 415–431, Oct. 2006.
[45] N. Nishino, S. Wang, N. Tsuji, K. Kageyama, and K. Ueda, "Categorization and mechanism of platform-type product-service systems in manufacturing," *CIRP Ann.*, vol. 61, no. 1, pp. 391–394, 2012.

QIHAO WAN received the bachelor's degree from the School of Automation Science and Electrical Engineering, Beihang University, in 2016. He is currently pursuing the Ph.D. degree with the National Engineering Research Centre of Computer Integrated Manufacturing System (CIMS-ERC), Tsinghua University, Beijing, China.

HEMING ZHANG received the Ph.D. degree in mechanical engineering from Zhejiang University, China, in 1995. He joined as a Faculty Member at the Department of Automation, Tsinghua University, after completing a two-year Postdoctoral Fellowship in National Engineering Research Center of Computer Integrated Manufacturing System (CIMS-ERC), Tsinghua University, China. He is currently a Professor at CIMS-ERC. His research interests include multidisciplinary modeling and collaborative simulation, service-oriented modeling and simulation, concurrent, and collaborative design.