International Journal of Computer Integrated Manufacturing

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tcim20

An integrated CRN-SVR approach for the quality consistency improvement in a diesel engine assembly process

Yan-Ning Sun, Qun-Long Chen, Jin-Hua Hu, Hong-Wei Xu, Wei Qin, Xiao-Xiao Shen & Zi-Long Zhuang

To cite this article: Yan-Ning Sun, Qun-Long Chen, Jin-Hua Hu, Hong-Wei Xu, Wei Qin, Xiao-Xiao Shen & Zi-Long Zhuang (20 Apr 2023): An integrated CRN-SVR approach for the quality consistency improvement in a diesel engine assembly process, International Journal of Computer Integrated Manufacturing, DOI: 10.1080/0951192X.2023.2204469

To link to this article: https://doi.org/10.1080/0951192X.2023.2204469

Published online: 20 Apr 2023.

INTERNATIONAL JOURNAL OF COMPUTER INTEGRATED MANUFACTURING
https://doi.org/10.1080/0951192X.2023.2204469

An integrated CRN-SVR approach for the quality consistency improvement in a diesel engine assembly process

Yan-Ning Sun[a], Qun-Long Chen[b], Jin-Hua Hu[b], Hong-Wei Xu[b], Wei Qin[b], Xiao-Xiao Shen[b] and Zi-Long Zhuang[b]

[a]School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China; [b]School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China

ABSTRACT

As the last production link, the diesel engine assembly process (DEAP) significantly impacts the quality consistency of diesel engine products. Therefore, the quality consistency improvement of DEAP has become an urgent problem for academia and industry. The challenge is how to mine the causal relationship in DEAP and establish a reliable quality prediction model. This paper attempts to describe DEAP using a causal relationship network (CRN) and to provide an effective data-based scheme for improving quality consistency by integrating CRN with support vector regression (SVR). First, a two-step CRN learning method is proposed for describing the DEAP. In the first step, the association relationship network is developed by a hybrid direct association detection method of the maximal information coefficient and network deconvolution, which can accurately measure the data relations. In the second step, the information geometric causal inference is employed to determine the direction of the edges in the association relationship network, thus forming the CRN of DEAP. Then, an integrated CRN-SVR approach is proposed to realize the predictive modeling of the critical quality indicators in DEAP, which integrated SVR into CRN. At the same time, it also provides a feasible idea for the interpretability of existing machine learning techniques. Finally, the proposed approach is tested and verified in a real-world DEAP and the obtained RMSE is only 0.033. The results of this study provide theoretical support and technical guarantee for quality consistency improvement in DEAP.

ARTICLE HISTORY
Received 26 June 2022
Accepted 16 April 2023

KEYWORDS
Intelligent manufacturing; big data analytics; machine intelligence; diesel engine; complex assembly process; quality consistency improvement

1. Introduction

The diesel engine has the advantages of ample torque, high thermal efficiency, and strong environmental adaptability, so it is widely used in heavy trucks, large buses, construction machinery, ships, power generation equipment, and other vital fields (Papakostas et al. 2015; Pattanaik and Jena 2018; Jing et al. 2019; Du et al. 2021; Sun et al. 2022). In 2020, China sold 6.341 million diesel engines, an increase of 17.62% over the same period last year, the highest in nearly 6 years. In this context, the safe and efficient production of diesel engines is facing significant challenges. According to the statistics (Qin et al. 2022), taking a diesel engine manufacturer as an example, the proportion of first-class diesel engines in a batch is only 24.1%, and the proportion of third-class diesel engines is only 68.7%. That is, 31.3% of the diesel engines in a batch still have excessive quality deviation. As the last production link, the diesel engine assembly process (DEAP) significantly impacts product quality consistency. Therefore, the quality consistency improvement (QCI) of DEAP has become an urgent problem.

With the popularization of information technology, a large amount of data in the manufacturing process is collected and stored. Countries around the world began to explore new industrial development models that belong to their own, such as German Industry 4.0, made in China 2025, and so on. Relying on the rapid development of Industry 4.0, the concept of industrial big data came into being. To realize big data analysis in the DEAP, it is necessary to deeply analyze all links in the production line and manufacturing system, and have a comprehensive three-dimensional analysis of the diesel engine. The field engineers usually try the production repeatedly according to their experience and constantly adjust the process parameters until the product quality reaches the standard. However,

CONTACT Yan-Ning Sun sunyanning@shu.edu.cn School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China; Wei Qin wqin@sjtu.edu.cn School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

this method based on expert experience is limited by different levels of experts, and the cost is high. To weaken the limitation of expert experience in QCI, some scholars have done a lot of research (Lu et al. 2016; Solanki, Sonigra, and Vajpayee 2023). Statistical process control (SPC) is one of the most effective tools for QCI. It judges whether the process is under control according to the manufacturing process changes described in the control chart, and eliminates potential problems according to the abnormal conditions presented in the control chart as much as possible (Hwarng and Hubele 1993). Rashidi et al. (Rashidi, Singh, and Zhao 2018) measured the strength of the causal relationship between the state variables and the process deviation through the transfer entropy and used it for real-time online diagnosis of the root cause of anomalies. Chen and Ge (Chen and Ge 2020) improved the traditional multivariable SPC based on a hierarchical Bayesian network, which is used for monitoring and decision-making of large-scale industrial processes. However, such methods are suitable for continuous industrial production processes (Zhu et al. 2018; Ma, Dong, and Peng 2020; Zhou et al. 2021), and it is difficult to distinguish the changes in different stages of DEAP.

Intelligent optimization is another important QCI method, which usually includes three steps: data acquisition, model construction, and algorithm solution (Guo et al. 2019). Siltepave et al. (Siltepavet, Sinthupinyo, and Chongstitvatana 2012) used a decision tree model to mine the control variables of hard disk quality for the hard disk manufacturing process and adopted a variety of intelligent algorithms to optimize the solution. Lughofer et al. (Lughofer et al. 2019) proposed a dynamic adaptive quality prediction model for the early identification and decision-making of product quality potential problems in a multi-stage manufacturing process and then used a multi-objective evolutionary algorithm to optimize the solution, and achieved good decision-making results. However, this kind of method depends on the quality prediction model. Once the model is inconsistent with the actual manufacturing system or has a large difference, the effect of the optimization decision will also be greatly reduced or even invalid.

In recent years, diesel engine manufacturers are increasingly using sensors and wireless technologies to capture massive data at critical operation processes, including material properties, the temperatures, and vibrations of equipment, and even the logistics of supply chains and customer details (Kusiak 2017; Gao, Shen, and Li 2019; Wang et al. 2019; Tao et al. 2019; Roy et al. 2020, 2020; Roy and Samui 2021; Roy et al. 2022). Nevertheless, the process operation mechanism is hidden behind these massive data set but has not been fully excavated. DEAP is a typical complex system in which the interaction of upstream and downstream operation processes through material and information flows. The execution of each operation process may directly or indirectly impact subsequent processes and the final quality of diesel engines. A number of error sources also exist in the assembly lines, which may come from a fixture, workbench, or spindle of CNC (Li, Wang, and Wang 2020). When assembling a diesel engine, these errors will gather together laterally and result in the flatness error, the roundness error, and so on. Meanwhile, because the post-processing is based on pre-processing data, there is a coupling relationship between these operational processes. Therefore, data errors from pre-processing will be transferred into post-processes, which dramatically challenges the QCI of DEAP (Guo et al. 2020; Hui et al. 2020; Xu et al. 2020). To tackle these challenges, it is first necessary to reveal the law of quality deviation transfer in diesel engine assembly by analyzing the causal relationships from massive process data.

In addition, causal relationships can provide direct guidance for the QCI of DEAP and obtain the critical process parameters and critical state variables that affect the quality indicators of diesel engines. Taking these critical data as input features, the prediction relationship of quality indicators can be established for optimizing and adjusting the DEAP. Based on these motives, some scholars have conducted a series of research on the QCI of DEAP. For example, to interpret assembly process data for QCI, Zha et al. (Zha et al. 2017) constructed the mixed copula function using the weighted linear model to describe the asymmetric tail behavior of diesel engine performance testing data. Diesel engine assembly quality-related critical features are selected using mutual information (MI) and the network deconvolution (ND) algorithm (Qin, Zha, and Zhang 2018). Meanwhile, the selected critical features from the point of the structure and operation mechanism of the diesel engine are discussed. Sun, Qin, and Zhuang

(Sun, Qin, and Zhuang 2021) proposed an integrated nonparametric-copula-entropy (NCE) and network deconvolution (ND) method to reveal causal relationships between process parameters and quality indicators in the diesel engine assembly process. Furthermore, Sun et al. (Sun et al. 2021) proposed an abnormal root-cause analysis method based on the information geometric causal inference (IGCI), which can help field operators to take corrective measures in time to resume the normal process. In the following study, Qin et al. (Qin et al. 2022) established an effective quality-level prediction of diesel engines based on machine learning techniques, which is of great significance for improving production quality consistency.

It can be seen that some meaningful progress has been made in its research on this topic. However, the following two key questions have not been systematically answered: 1) How to learn the causal relationship between critical process parameters, critical state variables, and critical quality indicators from the massive data to form a profound insight into DEAP? 2) How to integrate the causal relationship to establish a quantitative prediction model of critical quality indicators to guide the optimization and adjustment of DEAP?

This paper extends the work (Sun, Qin, and Zhuang 2020) initially presented at the 30th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM 2021). To answer the above two key questions systematically, the main contributions of this paper can be summarized as follows:

- A two-step causal relationship network (CRN) learning method is proposed for DEAP. In the first step, the association relationship network is developed by a hybrid direct associations detection method of the maximal information coefficient (MIC) and ND, which can accurately measure the data relations. In the second step, IGCI is employed to determine the direction of the edges in the association relationship network, thus forming the CRN of DEAP.
- An integrated CRN-SVR approach is proposed to realize the predictive modeling of the critical quality indicators in DEAP, which integrated support vector regression (SVR) into CRN. At the same time, it also provides a feasible idea for the interpretability of existing machine learning techniques. Finally, the proposed approach is tested and verified in a real-world DEAP.

The rest of this paper is organized as follows: Section 2 introduces the works related to this paper, Section 3 gives the theoretical basis, and Section 4 describes the integrated CRN-SVR approach. After that, the application validation on a real-world DEAP is carried out in Section 5, and Section 6 summarizes concluding remarks and some outlooks.

2. Related works

This section reviews the works related to the topic of this paper, which is divided into two sub-sections, 'Causal relationship analysis' and 'Quality indicators prediction', and some critical discussions are also given.

2.1. Causal relationship analysis

DEAP is a typical multivariable complex system, in which the fluctuations of process variables are coupled and superimposed, eventually affecting the quality indicators of engine products. However, not all variables will have a significant impact (Wang, Zheng, and Zhang 2020). Blind QCI will not only cause a waste of human resources, material, and financial resources but also weaken the control of critical links. Therefore, the first task is to measure and analyze meaningful relationships from massive process data, including associations and causality (Wang et al. 2021). At present, the widely used association analysis methods are the Pearson correlation coefficient (Frigieri, Ynoguti, and Paiva 2019), MI (Shi et al. 2020; Xu et al. 2022), and copula entropy (Alpettiyil, Ganapathy, and Sankaran 2019; Sun et al. 2022), which can help us identify the critical influence factors of quality consistency in DEAP. Moreover, causality strictly distinguishes between cause and effect variables, irreplaceable in revealing process mechanisms and guiding optimization decisions. There are three main methods to find causal relationships from massive process data: Granger causality analysis, Bayesian network (BN) structure learning, and function causality analysis.

2.1.1. Granger causality analysis

Yuan and Qin (Yuan and Qin 2014) diagnosed the vibration sources and propagation paths that cause oscillations in closed-loop control systems without the need to establish a system mechanism model. By combining Granger causality and topology networks, the propagation analysis method is proposed for oscillatory disturbances in an industrial sheet machine (Landman et al. 2014). Keskin and Aste (Keskin and Aste 2020) confirmed that the Granger causality analysis method is only suitable for Gaussian distribution data and can only analyze the linear causality between variables. Thus, they proposed a nonlinear transfer entropy method and proved that Granger causality and transfer entropy are equivalent to Gaussian distribution data. Nevertheless, the transfer entropy is sensitive to the selection of parameters. The algorithm has a large number of calculations, so it is not easy to directly apply it to the highly complex diesel engine assembly process.

2.1.2. BN structure learning

Nannapaneni, Mahadevan, and Rachuri (Nannapaneni, Mahadevan, and Rachuri 2016) proposed an uncertainty quantification method for performance prediction of manufacturing systems based on BN and took the injection molding process and friction welding process as examples to construct BN structure. A score search-based BN structure learning method is presented to identify quality-related fault propagation paths (Ma, Dong, and Peng 2018). However, it involves many graph search processes and has high time complexity. Independence test-based BN structure learning method is usually divided into two steps: causal skeleton graph learning and direction learning, that is, an undirected graph is constructed based on the measurement of statistics or information theory, and then the causal direction is inferred from conditional independence (Sun, Qin, and Zhuang 2021). Based on this idea and considering the transfer coupling effect of assembly deviation, Qin, Zha, and Zhang (2018) proposed an MI-ND approach to construct the causal skeleton graph, which can identify critical features in DEAP.

2.1.3. Function causality analysis

These methods mainly study the asymmetry of input and output data caused by system transfer function and noise and analyze the causal relationship and direction between the data. Shimizu et al. (Shimizu et al. 2006) proposed a linear non-Gaussian acyclic model, which assumes that the noise obeys the non-Gaussian distribution. In the correct causal direction, the noise and variables should be independent, while in the opposite direction, the noise and variables should not be independent. Nevertheless, this method is ineffective when the noise obeys the Gaussian distribution. Janzing et al. (Janzing et al. 2012) proposed an IGCI-based causal inference algorithm from the perspective of information geometry, which uses the independence of causal variable distributions and causal function mechanisms to judge the causal relationship between variables.

In summary, aiming at the QCI of DEAP, the current causal analysis methods have strict assumptions and adaptive conditions. They have not yet formed perfect research framework and verification methods.

2.2. Quality indicators prediction

Quality indicators prediction is of great significance for the QCI of DEAP. Establishing a prediction model can rely not only on the assembly process's first principle but also on the massive process data to establish a data-driven model based on statistical learning or deep learning methods.

2.2.1. First principle methods

As for small-scale and local manufacturing processes, they can effectively explain the source of errors and how they affect quality indicators. Mao, Chen, and Zhang (Mao, Chen, and Zhang 2015) proposed a mechanical assembly accuracy prediction method based on the state-space equation. This method improves the assembly accuracy of small six-axis precision CNC machine tools. Ma et al. (Ma, He, and Wu 2012) established a mechanism model between product qualified rate and manufacturing system reliability based on the Weibull analysis. They proved that it is feasible to evaluate the reliability of manufacturing processes based on product yield. However, to establish the first principle model of the production process, it is necessary to deeply study the physical or chemical mechanism (Shi and Zhou 2009; Sun et al. 2022), which is difficult for DEAP.

2.2.2. Statistical learning methods

Combining a multi-classification support vector machine method with genetic algorithm optimization, Hui et al. (2022) estimated the assembly quality of the linear axis in a three-axis vertical machining center. Jiang et al. (Jiang et al. 2020) proposed a quality-related modeling and monitoring method based on the optimized sparse partial least squares. Sun et al. (Sun et al. 2021) developed a novel framework of multivariate quality prediction for the injection process using copula entropy and multi-output SVR. It was further tested by the experiment on a real-world injection molding process dataset. Although the above methods have made some progress, the shallow statistical learning methods are relatively inadequate in exploring complex nonlinear relationships and need further improvement.

2.2.3. Deep learning methods

They are to improve prediction accuracy by constructing a statistical learning model with many hidden layers to learn more valuable features from massive training data. Ren et al. (Ren et al. 2020) proposed an improved wide-deep-sequence (WDS) model for highly redundant industrial processing data to extract quality features from non-time series data. At the same time, quality features are also extracted from time-series data based on long-term and short-term memory (LSTM) networks. Their quality prediction results are output through multi-layer perceptron (MLP). Zhou et al. (2017) developed a data-driven robust model of a blast furnace ironmaking process using an improved random vector functional-link network (RVFLN) with a Cauchy-weighted M-estimator. However, the deep learning method has the problem of poor interpretability, and the diesel engine's dynamic and changeable assembly environment limits the practical application of these methods.

In summary, the research methods and contents of quality indicators prediction are rich, from first principle models to data-driven models, from shallow statistical learning to deep learning. The research object extends from simple small-scale systems to large-scale complex systems. DEAP is a typical large-scale complex system that requires strong nonlinear fitting ability. We must fully consider various uncertainty factors to improve the interpretability and generalization performance of the algorithm.

3. Theoretical basis

3.1. Network deconvolution

Network deconvolution (ND) is a general method to distinguish direct dependencies in networks (Qin, Zha, and Zhang 2018; Sun, Qin, and Zhuang 2021; Sun et al. 2021). It assumes that the observed data are the sum of both direct and indirect effects, and by using the Eigen decomposition principle, the observed dependency matrix G_{obs} can be written as follows:

$$\begin{aligned} G_{obs} &= G_{dir} + G_{indir} = U\sum_{dir}U^{-1} + U\sum_{dir}^{2}U^{-1} + \cdots = U\left(\sum\nolimits_{dir} + \sum_{dir}^{2} + \cdots\right)U^{-1} \\ &= U\begin{pmatrix} \sum_{i\geq 1}\left(\lambda_{dir}^{i}\right)^{i} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{i\geq 1}\left(\lambda_{dir}^{n}\right)^{i} \end{pmatrix}U^{-1} = U\begin{pmatrix} \frac{\lambda_{dir}^{1}}{1-\lambda_{dir}^{1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\lambda_{dir}^{n}}{1-\lambda_{dir}^{n}} \end{pmatrix}U^{-1} \end{aligned} \tag{1}$$

where $n \longrightarrow +\infty$, G_{dir} is the direct association matrix, G_{indir} is the indirect association matrix, U and Σ_{dir} represent eigenvectors and the diagonal matrix of eigenvalues of G_{dir}, and$\lambda_{dir}^{i}$is the i-th diagonal component of matrix Σ_{dir}.

Therefore, from Equation (1), if

$$\frac{\lambda_{dir}^{i}}{1-\lambda_{dir}^{i}} = \lambda_{obs}^{i}, \textit{for all } 1 \leq i \leq n \tag{2}$$

Rewritten Equation (2):

$$\frac{\lambda_{obs}^{i}}{1+\lambda_{obs}^{i}} = \lambda_{dir}^{i}, \textit{for all } 1 \leq i \leq n \tag{3}$$

Thus, the direct association matrix G_{dir} can be written as the deconvolution formula:

$$G_{dir} = U\begin{pmatrix} \lambda_{dir}^{1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{dir}^{n} \end{pmatrix}U^{-1} \tag{4}$$

To guarantee the infinite series of Equation (1) converge, a tuning parameter was introduced to linearly scale the $\lambda_{obs}^{i}$to cover more regions:

$$\begin{cases} \lambda_{dir}^{i} = \frac{\alpha\lambda_{obs}^{i}}{1+\alpha\lambda_{obs}^{i}} \\ \alpha \leq \min\left(\frac{\beta}{\left(1-\beta\lambda_{obs}^{+(\max)}\right)}, \frac{-\beta}{\left(1+\beta\lambda_{obs}^{-(\min)}\right)}\right) \end{cases} \tag{5}$$

where $\lambda_{obs}^{+(\max)}$and$\lambda_{obs}^{-(\min)}$are the most significant positive and smallest negative eigenvalues of G_{obs}. The β parameter is network dependent and $\beta = 0.95$ in this paper.

Algorithm 1 gives the pseudocode of the ND algorithm. Because of its theoretical impact as a foundational graph theoretic tool, ND is widely applicable for computing direct dependencies in network science across diverse disciplines.

Algorithm 1: Network deconvolution

Input: Observed dependency matrix G_{obs}
Output: Direct association matrix G_{dir}
1: **Linear Scaling Step**: The observed dependency matrix G_{obs} is scaled linearly so that all eigenvalues of the direct association matrix G_{dir} are between −1 and 1
2: **Decomposition Step**: The observed dependency matrix G_{obs} is decomposed to its eigenvalues and eigenvectors such that $G_{obs}=U\Sigma_{obs}U^{-1}$
3: **Deconvolution step**: A diagonal eigenvalue matrix Σ_{dir} is formed whose i-th component is shown in Equation (5)
4: **Return** $G_{dir}= U\Sigma_{dir}U^{-1}$

3.2. Information geometric causal inference

If X is the cause of Y ($X \rightarrow Y$), the distribution of X and the function f mapping X to Y are independent since they correspond to independent mechanisms of nature (Janzing et al. 2012). Based on this idea, let ζ_X and ζ_Y define exponential families of 'smooth' reference distributions for X and Y, respectively. Let u denote the projection of p_X onto ζ_X and u_f its image under f. If $X \rightarrow Y$, then:

$$D(p_Y||\zeta_X) = D(p_X||\zeta_X) + D(u_f||\zeta_Y) \tag{6}$$

where $D(\cdot||\cdot)$ denotes the relative entropy distance or the Kullback–Leibler divergence between two probability densities. Its definition is as follows:

$$D(p||q) = {}^{p}(x)\log\frac{p(x)}{q(x)}\mathrm{d}x \tag{7}$$

where p and q denote probability densities. The following causal inference method named information geometric causal inference (IGCI) is employed in this paper: given $C_{X\rightarrow Y}$ and $C_{Y\rightarrow X}$, infer that X causes Y if $C_{X\rightarrow Y}<0$, or that Y causes X if $C_{Y\rightarrow X}<0$. $C_{X\rightarrow Y}$ and $C_{Y\rightarrow X}$ can be denoted as:

$$C_{X\rightarrow Y} = -C_{Y\rightarrow X} = D(p_X||\zeta_X) - D(p_Y||\zeta_Y) \tag{8}$$

As a common approach, the IGCI method can infer deterministic causal relations between variables with various domains. In this paper, the following simple estimator is used:

$$\hat{C}_{X\rightarrow Y} = \frac{1}{N-1}\sum_{j=1}^{N-1}\log\frac{|y_{j+1}-y_j|}{|x_{j+1}-x_j|} \tag{9}$$

where $x_1<x_2< \ldots <x_N$ are the observed values of X (arrange in ascending order). Algorithm 2 gives the pseudocode of the IGCI algorithm.

Algorithm 2: Information geometric causal inference

Input: Observed data of the two variables X and Y
Output: Causal direction between X and Y
1: Estimate $C_{X\rightarrow Y}$ and $C_{Y\rightarrow X}$ using Equation (9)
2: **If** $C_{X\rightarrow Y} \leq C_{Y\rightarrow X}$
3: **Then** X causes Y
4: **Else** Y causes X
5: **Return**: $X\rightarrow Y$ or $X\rightarrow Y$

3.3. Bayesian networks

Bayesian networks (BN) are a class of graphical models (also called CRN) that allow an intuitive representation of the probabilistic structure of multivariate data using graphs. Given a probability distribution P on a set of variables $X=(X_1, X_2, \ldots, X_n)$, a directed acyclic graph (DAG) $G=(X, A)$ is called a CRN, and A denotes the set of the directed arcs represent direct probabilistic dependencies (Nannapaneni, Mahadevan, and Rachuri 2016). The mathematical symbolic expression of CRN can be expressed as:

$$p(X) = \prod_{i=1}^{n} p(X_i|\mathrm{pa}(X_i)) \tag{10}$$

where $\mathrm{pa}(X_i)$ represents the set of parent nodes of X_i and $p(X_i|\mathrm{pa}(X_i))$ represents the conditional probability distribution of X_i, given its parent nodes. If X_i has no parent nodes, then $p(X_i|\mathrm{pa}(X_i))$ represents the marginal probability distribution of X_i.

In the field of BNs, model selection and estimation are collectively known as learning, a name borrowed from artificial intelligence and machine learning. BN learning is usually performed as a two-step process:

- Structure learning: learning the structure of the DAG;
- Parameter learning: learning the local distributions implied by the structure of the DAG learned in the previous step.

Both steps can be performed either as unsupervised learning, using the information provided by a data set, or as supervised learning, by interviewing experts in the fields relevant to the modeled phenomenon.

4. Integrated CRN-SVR approach

This section introduces the integrated CRN-SVR approach. A two-step CRN structure learning method is presented to describe the DEAP, and then the SVR model is integrated into the developed CRN structure.

4.1. A two-step CRN structure learning method

Inspired by Zhang et al. (Zhang, Zhang, and Xie 2013), this section introduces a two-stage CRN structure construction scheme for root cause analysis and propagation path identification of quality deviation (see Figure 1). Compared with Zhang et al. (Zhang, Zhang, and Xie 2013), this study uses the IGCI method instead of the greedy equivalent search algorithm to determine the causal relationship between variables, significantly improving the calculation efficiency.

4.1.1. Step 1: undirected graph construction

In this step, a hybrid method of the maximal information coefficient (MIC) and the network deconvolution (ND) is proposed to detect direct associations between variables of the manufacturing process more accurately.

Step 1.1: Compute MIC (Reshef et al. 2011) for all nodes with respect to all other nodes in the dataset. Traverse all variables, the corresponding observed dependency matrix G_{obs} can be written as:

$$G_{obs} = \begin{pmatrix} 1 & \mathrm{MIC}_{12} & \cdots & \mathrm{MIC}_{1N} \\ \mathrm{MIC}_{21} & 1 & \cdots & \mathrm{MIC}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{MIC}_{N1} & \mathrm{MIC}_{N2} & \cdots & 1 \end{pmatrix} \tag{11}$$

Step 1.2: Perform the ND algorithm on G_{obs} for filtering the transitive noises and get the direct dependencies G_{dir} based on Eqs. (4) and (5).

$$G_{dir} = \begin{pmatrix} 0 & \mathrm{MIC}_{12}^{ND} & \cdots & \mathrm{MIC}_{1N}^{ND} \\ \mathrm{MIC}_{21}^{ND} & 0 & \cdots & \mathrm{MIC}_{2N}^{ND} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{MIC}_{N1}^{ND} & \mathrm{MIC}_{N2}^{ND} & \cdots & 0 \end{pmatrix} \tag{12}$$

Step 1.3: Get the maximum MIC^{ND} (MMIC) for each node and set a threshold value $\gamma = 0.9$ of the MMIC for each node. If either of the following Equation (13) is satisfied, those two variables will establish an undirected edge. After this phase, a preliminary undirected graph is established, written as G_{pu}.

$$\mathrm{MIC}_{ij}^{ND} \geq \gamma \mathrm{MMIC}_i \mathrm{or} \mathrm{MIC}_{ji}^{ND} \geq \gamma \mathrm{MMIC}_j \tag{13}$$

Step 1.4: Connectivity detection and recovery to obtain a complete undirected graph fulfilling the connectivity. First, the connected component of G_{pu} is obtained by using the depth-first search (DFS) algorithm (Tarjan 1972). For any two connected components, their variables set are V_k and V_t, and the number of variables is m and n, respectively. The MIC between the two connected components is defined as MIC*:

$$\begin{gathered} \mathrm{MIC}^*(V_k, V_t) = \max\{\mathrm{MIC}_{ij}^{ND}\} \\ \mathrm{Variable} X_i \in V_k, \mathrm{Variable} X_j \in V_t \end{gathered} \tag{14}$$

This study calculates the MIC* of any two connected components and set an undirected edge between

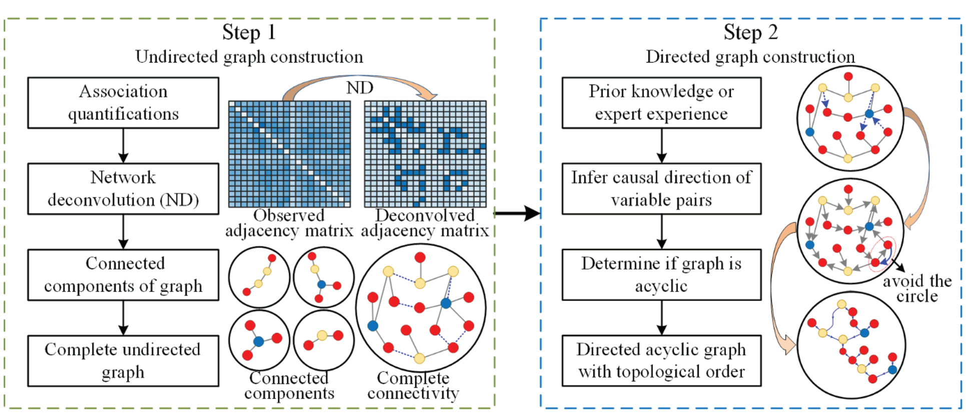

Figure 1. A two-step CRN structure learning method.

corresponding variables to recover the graph's connectivity.

4.1.2. Step 2: directed graph construction

This step is to infer the direction of the edges of the undirected graph obtained from the last step. There are also four main substeps:

Step 2.1: The partial digraph can be obtained by using prior or expert knowledge to determine the direction of some edges.

Step 2.2: Infer the causal direction of variable pairs for all undirected edges by using the IGCI method.

Step 2.3: Use the DFS algorithm again to determine whether there is a loop in the digraph. If there is a loop, break the edge between two variables with a smaller $\mathrm{MIC}^{\mathrm{ND}}$.

Step 2.4: Perform the topological order of the directed acyclic graph to get DAG with topological order (Inoue and Minato 2014).

4.2. Integrated SVR method to CRN structure

Learning the input-output quantitative relationship of DEAP from the actual production data requires that the modeling method has high interpretability. In a very intuitive way, the structure of CRN provides an interpretable insight for establishing a quality indicators prediction model. This study improves the interpretability of the prediction model by integrating the causality of DEAP. The edges and directions in CRN represent the transmission path of the information flow of the critical process parameters, critical state variables, and quality indicators of DEAP. As shown in Figure 2, the SVR model is integrated into the CRN. It is further trained with critical process parameters and critical state variables as input and quality indicators as output features.

SVR is an efficient and easy-to-use regression model in machine learning, which constructs a hyperplane or set of hyper-planes in a high or infinite-dimensional space. At present, there are mainly ε-SVR and v-SVR. In this paper, ε-SVR is employed, and its basic principle is as follows.

Given the input sample dataset $\mathbf{x} \in \mathbb{R}^{k\times l}$ and the output sample dataset $\mathbf{y} \in \mathbb{R}^{k}$, the input sample data $\boldsymbol{x}$ is first mapped into m-dimensional feature space through a nonlinear transformation $\phi(\boldsymbol{x})$. Then, a linear model in this m-dimensional feature space is established to estimate the regression function as follows:

$$f(\boldsymbol{x}, \boldsymbol{w}) = \boldsymbol{w} \cdot \phi(\boldsymbol{x}) + b \tag{15}$$

where w is the weight vector, and b is the offset.

After $\phi(\boldsymbol{x})$ transformation, the nonlinear regression problem in the original space can be solved as a linear regression problem in the high-dimensional space. The ε insensitive loss function is as follows:

$$L_{\varepsilon}(x_i, y_i, f) = \max\{0, |y_i - f(x_i)| - \varepsilon\} \tag{16}$$

The SVR model is to find a suitable function $f(x)$ to fit these training data samples such that the error between the real value y_i and the predicted value $f(x)$ is minimized, and the error can be represented by ε insensitive loss function. As shown in Figure 2, when a training data sample is within the two blue dotted lines, the error of the training data sample is considered to be 0. Further, ε-SVR solves the following optimization problem:

$$\begin{aligned} &\min_{w,b,\zeta,\zeta^*} \tfrac{1}{2} w^T w + C \sum_{i=1}^{n} (\zeta_i + \zeta_i^*) \\ &\text{subject to} \quad y_i - w^T \phi(x_i) - b \le \varepsilon + \zeta_i, \\ &\qquad\qquad\quad w^T \phi(x_i) + b - y_i \le \varepsilon + \zeta_i^*, \\ &\qquad\qquad\quad \zeta_i, \zeta_i^* \ge 0, i = 1, \ldots, n \end{aligned} \tag{17}$$

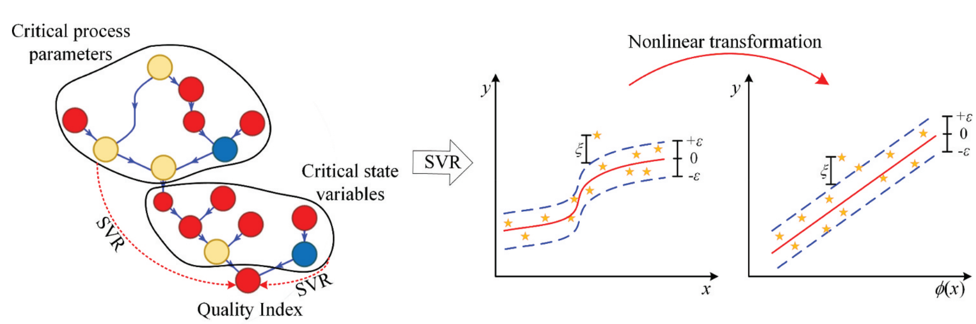

Figure 2. Schematic diagram of the integrated CRN-SVR approach.

where this study penalizes samples whose prediction is at least ε away from their actual target. These samples penalize the objective by ζ_i or ζ_i, depending on whether their predictions lie above or below the ε tube. Equation (17) is optimized by introducing the Lagrange function, and the solution of Equation (15) is obtained by solving the dual problem:

$$f(\boldsymbol{x}) = \sum_{i=1}^{l_{sv}} (a_i + a_i^*) K(\boldsymbol{x}_i, \boldsymbol{x}) + b \tag{18}$$

where a_i, a_i^* $(i = 1, 2, \ldots, n)$ are Lagrange multipliers, and their corresponding samples are support vectors; l_{sv} is the number of support vectors; $K(\boldsymbol{x}_i, \boldsymbol{x})$ is the kernel function. The commonly used kernel functions include polynomial kernel function, Gaussian radial basis kernel function, sigmoid kernel function, Fourier series kernel function, etc.

It can be seen from the structure of the SVR model that it does not require a specific function form. A trained SVR model can capture the complex input–output relationship between nonlinear output variables (the quality indicators of DEAP) and input variables (critical process parameters and critical state variables). Therefore, it is very effective to use SVR to predict the quality indicators of DEAP. In addition, the SVR model is a convex quadratic optimization problem, which ensures that the extremum found is the global optimal solution, unlike other nonlinear optimization methods, which easily fall into a local minimum.

5. Application validation

5.1. Dataset description

This study uses the diesel engine production line of Guangxi Yuchai (a diesel engine manufacturer from China) to verify the proposed approach in this paper. The diesel engine production line consists of four parts: the main assembly line, sub-assembly lines 1–5, performance test line, and package line (see Figure 3). The diesel engine block and cooler, crankshaft, oil pump, camshaft, piston connecting rod unit, and cylinder are assembled by sub-assembly lines 1–5, respectively, and delivered to the main assembly line. Moreover, the main assembly line carries on the overall assembly of the diesel engine. The performance indicators of diesel engines are tested in the performance test lines, such as power and fuel consumption rate.

There are more than 100 workstations (WS) on the diesel engine assembly line. A total of 172

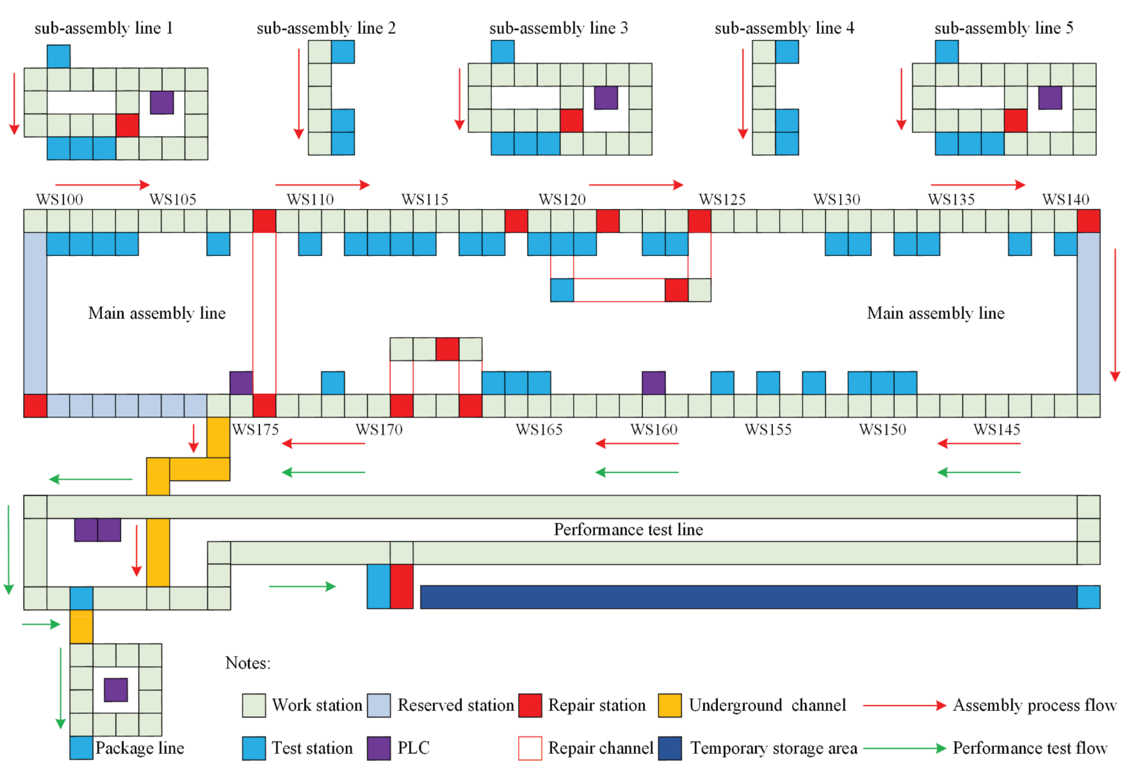

Figure 3. Diagram of diesel engine manufacturing system.

Table 1. Selected manufacturing process parameters (numbers 1–16) and performance test parameters (numbers 17–41).

Numbers	Symbols	Descriptions	Numbers	Symbols	Descriptions
1	CLPH1	Cylinder liner protrusion height 1	22	OP	Oil pressure
2	CLPH2	Cylinder liner protrusion height 2	23	IAT	Intake air temperature
3	CLPH3	Cylinder liner protrusion height 3	24	IWT	Inlet water temperature
4	CLPH4	Cylinder liner protrusion height 4	25	CFC	Cumulative fuel (gas) consumption
5	CLPH5	Cylinder liner protrusion height 5	26	T	Torque
6	CLPH6	Cylinder liner protrusion height 6	27	ET	Exhaust temperature
7	ST	Start torque	28	FT	Fuel temperature
8	RT	Run torque	29	FC	Fuel consumption
9	AC	Axial clearance	30	FCR	Fuel consumption rate
10	CTM	Crankshaft turning moment	31	WG	Water gate
11	PPH1	Piston protrusion height 1	32	SI	Smoke Intensity
12	PPH2	Piston protrusion height 2	33	Th	Throttle
13	PPH3	Piston protrusion height 3	34	RT	Running time
14	PPH4	Piston protrusion height 4	35	IOT	Intercooler outlet temperature
15	PPH5	Piston protrusion height 5	36	IOP	Intercooler outlet pressure
16	PPH6	Piston protrusion height 6	37	IIT	Intercooler inlet temperature
17	OWT	Outlet water temperature	38	IIP	Intercooler inlet pressure
18	AH	Ambient humidity	39	RS	Rotational speed
19	AP	Atmospheric pressure	40	P	Power
20	PL	Piston leakage	41	Pd	Power deviation
21	OT	Oil temperature			

assembly process parameters are tested. The diesel engine enters the stage of performance test after assembly. This study first analyzed the correlation between assembly process parameters and performance test parameters. Then, the assembly process parameters that correlate significantly with the performance indicators are selected as the candidate set of factors. As is shown in Table 1, 16 assembly process parameters (numbers 1–16) and 25 performance test parameters (numbers 17–41) are selected to construct the dataset of this section. The rated power of produced diesel engines is 254 kW. According to the deviation between the actual power and the rated power, it is stipulated that if the actual power falls within the range of 254 ± 3% kW (246.38 kW-261.62 kW), then the diesel engines are qualified products. If the power deviation exceeds ±3%, the diesel engines are unqualified. The scatter diagram of power deviation data of a batch of diesel engines is plotted in Figure 4, showing a serious low power consistency problem. Furthermore, the histogram of power data of a batch of diesel engines is shown in Figure 5. It can be seen from the figure that the rated power of 254 kW deviates from the average value of the data, which shows that in addition to stochastic factors, there are technical reasons leading to lower power consistency. For example, some process parameters are not controlled in a reasonable range.

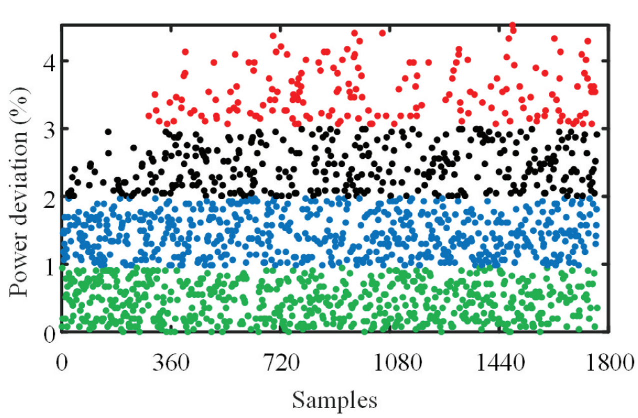

Figure 4. Scatter diagram of power deviation data of a batch of diesel engines.

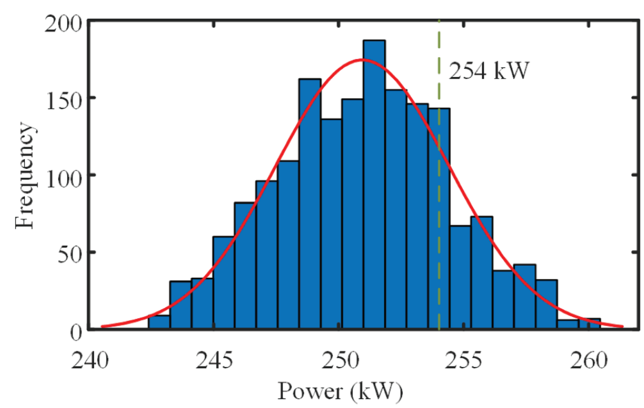

Figure 5. Histogram of power data of a batch of diesel engines.

5.2. Result and discussion

This subsection analyzes the root cause and propagation path of diesel engine production's power

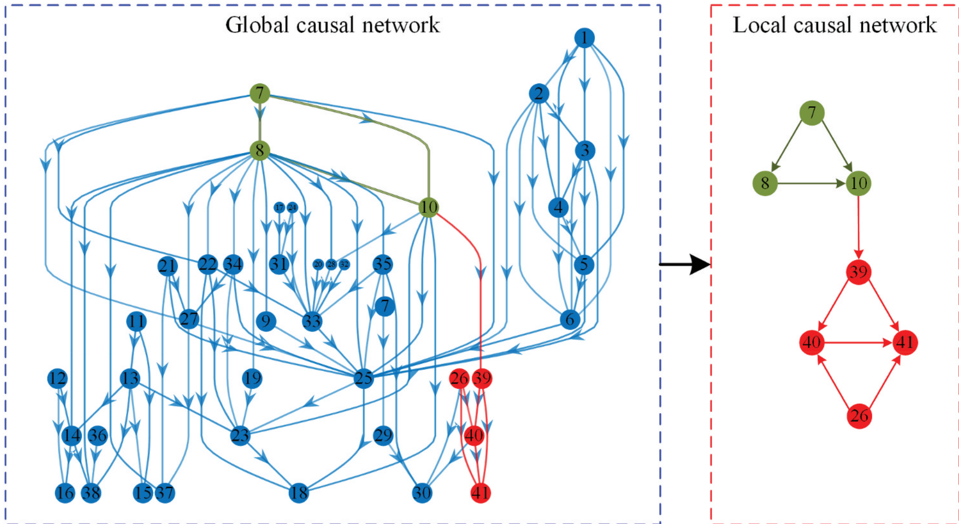

Figure 6. Causal Bayesian network of the DEAP.

consistency problem. All experiments and related codes are carried out in Python 3.7. The computer hardware configuration that the authors used is Intel (R) Core (TM) i7–8700 CPU @ 3.20 GHz 32.00 G RAM. A training dataset consisting of 1763 samples generated from DEAP is used to learn the BN structure, as shown in Figure 6. In this figure, the descriptions of numbers 1–41 are listed in Table 1. It can be seen that the parameters associated with power deviation are 7, 8, 10, 26, 39, and 40, where the first three are assembly process parameters (marked in green) and the last three are engine operating state parameters (marked in red).

Furthermore, the local causal network of these six parameters can be drawn (see Figure 6). As can be seen, because of the existence of the start torque (parameter 7) error, the run torque (parameter 8) and crankshaft turning moment (parameter 10) are affected. Subsequently, the error of start torque (parameter 7), run torque (parameter 8), and crankshaft turning moment (parameter 10) is accumulated. Thus, the rotational speed (parameter 39) deviates from the expected values, which affects the power (parameter 40) and its deviation (parameter 41). Based on the above analysis, the local causal network correctly reflects the actual assembly error propagation in DEAP. The start torque (parameter 7) in the network is accurately recognized as the root cause variable.

With the help of statistical analysis, the control range of start torque can be optimized, the error transmission can be reduced, and the power consistency can be improved. The original control limit of start torque is [0, 45]. The authors find the quantiles of 20%, 40%, 60%, and 80% of start torque, and the corresponding values are shown in Table 2. The five intervals defined by quantiles are defined as intervals 1–5. The probabilities of the first, second, and third rate products and unqualified products in these five intervals are shown in Figure 7. As can be seen, if the start torque is controlled in interval 2 (2.966 ~ 8.681 N·m), the unqualified product rate is the lowest. Furthermore, if the start torque is controlled in interval 4 (12.172 ~ 17.380 N·m), the first-class product rate is the highest.

Furthermore, this study evaluates prediction accuracy concerning power based on comparisons of the proposed approach (CRN-SVR), CRN-linear regression (CRN-LR), CRN-decision tree (CRN-DT), and CRN-back propagation neural network (CRN-BPNN). Experiments are also designed to verify the effectiveness of the CRN structure proposed previously. Furthermore,, the suitability of the machine learning model integrated into the CRN structure is

Table 2. Quantile list of start torque.

Quantile	0%	20%	40%	60%	80%	100%
Value	2.355	2.966	8.681	12.172	17.380	45.093

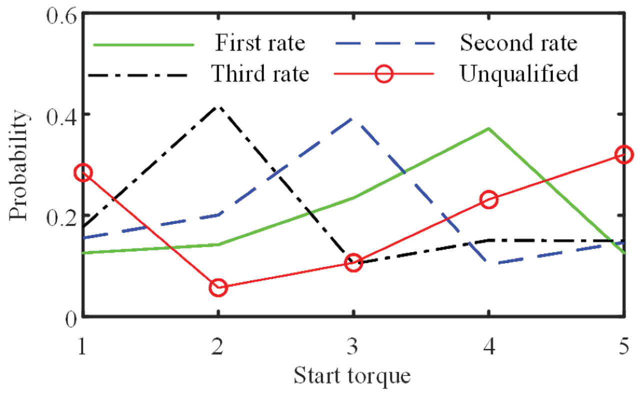

Figure 7. Variation curve of the probability of different product classes with the control range of start torque.

demonstrated by evaluating the accuracy of the proposed approach.

Five critical variables obtained from the CRN structure, i.e. torque (parameter 26), rotational speed (parameter 39), starting torque (parameter 7), run torque (parameter 8), and crankshaft turning moment (parameter 10), are selected as input features for prediction. Figure 8 shows the prediction results on the real-world industrial data of the 177 test sets by CRN-SVR, CRN-LR, CRN-DT, and CRN-BPNN. It can be seen that the errors of the four approaches are pretty minor, which can prove that the critical variables found in the CRN structure are highly effective and essential factors. Figure 9 shows the scatter diagram of the real value and the predicted value, in which the scatter points are concentrated on the diagonal attachment, indicating that the prediction error has a fairly small fluctuation bias. The probability density curve of prediction error (see Figure 10) and RMSE comparison (Table 3) were also drawn. It can be seen that the integrated SVR method to CRN structure has the best adaptability and performance.

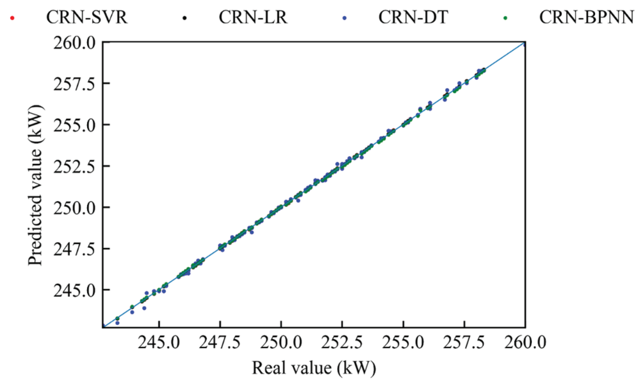

Figure 9. Scatter diagrams from different approaches.

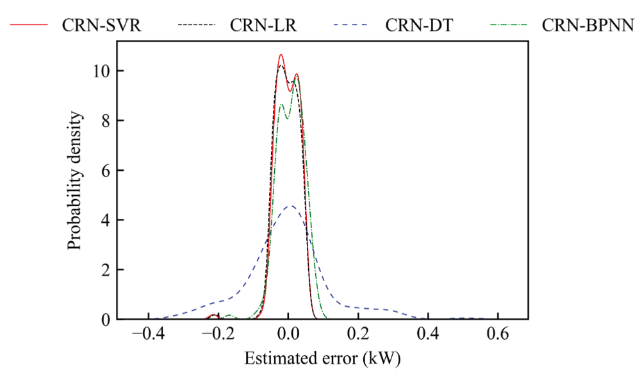

Figure 10. Probability density curves from different approaches.

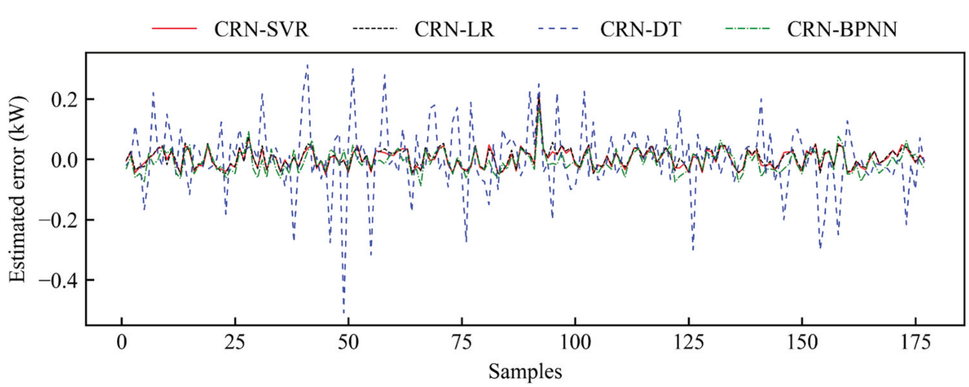

Figure 8. Quality prediction results from different approaches.

Table 3. Performance comparison of different approaches.

Metric	CRN-SVR	CRN-LR	CRN-DT	CRN-BPNN
RMSE	0.033240	0.033733	0.116340	0.037836

Table 4. RMSE comparison results.

Metric	SVR (without)	SVR (with)
RMSE	0.033240	0.033199

In addition, the CRN structure proposed before also emphasizes that the three assembly process parameters (starting torque, run torque, and crankshaft turning moment) are related to the power. To demonstrate the accuracy of this finding, the SVR model is designed to test whether the input features containing assembly process parameters are the control variable. The RMSE comparison results are listed in Table 4. It can be seen that assembly process parameters can explain the power deviation to some extent. After being added, the generalization error can be further reduced, improving the prediction accuracy of the quality indicators.

Moreover, the results of practical application indicate that the performance of the integrated CRN-SVR approach this paper proposed has superiority over the other approaches. In addition, this approach is very user-friendly and requires almost no fine-tuning of parameters to achieve excellent performance. It can overcome the blindness in the selection of predefined parameters in the conventional algorithm.

6. Conclusion

This study proposes an integrated CRN-SVR approach for quality consistency improvement in a diesel engine assembly process (DEAP). On the one hand, to learn the causal relationship between critical process parameters, critical state variables, and critical quality indicators from the massive data, a two-step causal relationship network (CRN) learning method is proposed for forming a profound insight into DEAP. On the other hand, an integrated CRN-SVR approach is proposed to realize the predictive modeling of the critical quality indicators in DEAP, which can integrate the causal relationship and guide the optimization and adjustment of DEAP. The dataset from a real-world DEAP is employed as an application study to test the proposed approach. In the future, the authors will further expand the CRN-SVR approach to form a theoretical system for the quality consistency improvement of the multi-station assembly process integrating data analysis, quality modeling, and process control.

Acknowledgements

The authors would also like to thank Foxconn Technology Group for providing us with a historical dataset of IM process and the valuable suggestions and comments of Editors and Reviewers of the International Journal of Computer Integrated Manufacturing.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The work was supported by the National Key Research and Development Program of China [2021YFB3300503].

ORCID

Yan-Ning Sun http://orcid.org/0000-0003-2835-7346
Wei Qin http://orcid.org/0000-0002-8527-0354

References

Alpettiyil, K., B. R. Ganapathy, and P. G. Sankaran. 2019. “Non-Parametric Estimation of Copula Based Mutual Information.” *Communications in Statistics - Theory and Methods* 49 (6): 1513–1527. doi:10.1080/03610926.2018.1563180.

Chen, G., and Z. Ge. 2020. “Hierarchical Bayesian Network Modeling Framework for Large-Scale Process Monitoring and Decision Making.” *IEEE Transactions on Control Systems Technology* 28 (2): 671–679. doi:10.1109/TCST.2018.2882562.

Du, J., S. H. Mirtalebi, M. T. Ahmadian, Y. Cao, M. Suhatril, and H. Assilzadeh. 2021. “Dynamic Control of Supported Macro/micro-Tubes Conveying Magnetic Fluid Utilizing Intelligibly Designed Axially Functionally Graded Materials.” *International Journal of Computer Integrated Manufacturing* 35 (4–5): 345–358. doi:10.1080/0951192x.2021.1879399.

Frigieri, E. P., C. A. Ynoguti, and A. P. Paiva. 2019. “Correlation Analysis Among Audible Sound Emissions and Machining Parameters in Hardened Steel Turning.” *Journal of Intelligent Manufacturing* 30 (4): 1753–1764. doi:10.1007/s10845-017-1356-9.

Gao, L., W. Shen, and X. Li. 2019. “New Trends in Intelligent Manufacturing.” *Engineering* 5 (4): 619–620. doi:10.1016/j.eng.2019.07.001.

Guo, F., X. Zhou, J. Liu, Y. Zhang, D. Li, and H. Zhou. 2019. “A Reinforcement Learning Decision Model for Online Process Parameters Optimization from Offline Data in Injection Molding.” *Applied Soft Computing* 85: 105828. doi:10.1016/j.asoc.2019.105828.

Guo, L., Z. Zhuang, Y. Sun, and W. Qin. 2020. “Data-Driven Rated Power Prediction of Diesel Engines Using Improved Multi-Class Imbalanced Learning Method.” *Procedia*

Manufacturing 51: 324–329. doi:10.1016/j.promfg.2020.10.046.

Hui, Y., X. Mei, G. Jiang, F. Zhao, Z. Ma, and T. Tao. 2022. “Assembly Quality Evaluation for Linear Axis of Machine Tool Using Data-Driven Modeling Approach.” *Journal of Intelligent Manufacturing* 1–17. doi:10.1007/s10845-020-01666-y.

Hui, Y., X. Mei, G. Jiang, F. Zhao, and P. Shen. 2020. “Assembly Consistency Improvement of Straightness Error of the Linear Axis Based on the Consistency Degree and GA-MSVM-I-KM.” *Journal of Intelligent Manufacturing* 31 (6): 1429–1441. doi:10.1007/s10845-019-01520-w.

Hwarng, H. B., and N. F. Hubele. 1993. “X Control Chart Pattern Identification Through Efficient Off-Line Neural Network Training.” *IIE Transactions* 25 (3): 27–40. doi:10.1080/07408179308964288.

Inoue, Y., and S. Minato. 2014. “An Efficient Method for Indexing All Topological Orders of a Directed Graph.” *International Symposium on Algorithms and Computation*: 103–114. doi: 10.1007/978-3-319-13075-0_9.

Janzing, D., J. Mooij, K. Zhang, J. Lemeire, J. Zscheischler, P. Daniušis, B. Steudel, and B. Schölkopf. 2012. “Information-Geometric Approach to Inferring Causal Directions.” *Artificial Intelligence* 182-183: 1–31. doi:10.1016/j.artint.2012.01.002.

Jiang, Q., X. Yan, H. Yi, and F. Gao. 2020. “Data-Driven Batch-End Quality Modeling and Monitoring Based on Optimized Sparse Partial Least Squares.” *IEEE Transactions on Industrial Electronics* 67 (5): 4098–4107. doi:10.1109/tie.2019.2922941.

Jing, X., Y. Zhu, J. Liu, H. Zhou, P. Zhao, X. Liu, G. Tian, H. Ye, and Q. Li. 2019. “Intelligent Generation Method of 3D Machining Process Based on Process Knowledge.” *International Journal of Computer Integrated Manufacturing* 33 (1): 38–61. doi:10.1080/0951192x.2019.1690687.

Keskin, Z., and T. Aste. 2020. “Information-Theoretic Measures for Nonlinear Causality Detection: Application to Social Media Sentiment and Cryptocurrency Prices.” *Royal Society Open Science* 7 (9): 200863. doi:10.1098/rsos.200863.

Kusiak, A. 2017. “Smart Manufacturing Must Embrace Big Data.” *Nature* 544 (7648): 23–25. doi:10.1038/544023a.

Landman, R., J. Kortela, Q. Sun, and S. L. Jämsä-Jounela. 2014. “Fault Propagation Analysis of Oscillations in Control Loops Using Data-Driven Causality and Plant Connectivity.” *Computers & Chemical Engineering* 71: 446–456. doi:10.1016/j.compchemeng.2014.09.017.

Li, Z., Y. Wang, and K. Wang. 2020. “A Data-Driven Method Based on Deep Belief Networks for Backlash Error Prediction in Machining Centers.” *Journal of Intelligent Manufacturing* 31 (7): 1693–1705. doi:10.1007/s10845-017-1380-9.

Lu, X., W. Chang, L. Li, and S. Zhou. 2016. “A Quality Evaluation Model for Diesel Engine Using RBF Neural Network Based on Trial Run Data.” Paper presented at the 2016 IEEE International Conference on Mechatronics and Automation, Harbin, China, August 7-10.

Lughofer, E., A. Zavoianu, R. Pollak, M. Pratama, P. Meyer-Heye, H. Zörrer, C. Eitzinger, and T. Radauer. 2019. “Autonomous Supervision and Optimization of Product Quality in a Multi-Stage Manufacturing Process Based on Self-Adaptive Prediction Models.” *Journal of Process Control* 76: 27–45. doi:10.1016/j.jprocont.2019.02.005.

Ma, L., J. Dong, and K. Peng. 2018. “A Practical Propagation Path Identification Scheme for Quality-Related Faults Based on Nonlinear Dynamic Latent Variable Model and Partitioned Bayesian Network.” *Journal of the Franklin Institute* 355 (15): 7570–7594. doi:10.1016/j.jfranklin.2018.07.035.

Ma, L., J. Dong, and K. Peng. 2020. “A Novel Hierarchical Detection and Isolation Framework for Quality-Related Multiple Faults in Large-Scale Processes.” *IEEE Transactions on Industrial Electronics* 67 (2): 1316–1327. doi:10.1109/tie.2019.2898576.

Ma, J., Y. He, and C. Wu. 2012. “Research on Reliability Estimation for Mechanical Manufacturing Process Based on Weibull Analysis Technology.” In *Proceedings of the IEEE 2012 Prognostics and System Health Management Conference (PHM-2012 Beijing)*, Beijing, China, 1–5.

Mao, J., D. Chen, and L. Zhang. 2015. “Mechanical Assembly Quality Prediction Method Based on State Space Model.” *International Journal of Advanced Manufacturing Technology* 86 (1–4): 107–116. doi:10.1007/s00170-015-8122-y.

Nannapaneni, S., S. Mahadevan, and S. Rachuri. 2016. “Performance Evaluation of a Manufacturing Process Under Uncertainty Using Bayesian Networks.” *Journal of Cleaner Production* 113: 947–959. doi:10.1016/j.jclepro.2015.12.003.

Papakostas, N., G. Pintzos, C. Giannoulis, and G. Chryssolouris. 2015. “An Agent-Based Collaborative Platform for the Design of Assembly Lines.” *International Journal of Computer Integrated Manufacturing* 29 (4): 374–385. doi:10.1080/0951192x.2015.1066862.

Pattanaik, L. N., and A. Jena. 2018. “Tri-Objective Optimisation of Mixed Model Reconfigurable Assembly System for Modular Products.” *International Journal of Computer Integrated Manufacturing* 32 (1): 72–82. doi:10.1080/0951192x.2018.1550673.

Qin, W., D. Zha, and J. Zhang. 2018. “An Effective Approach for Causal Variables Analysis in Diesel Engine Production by Using Mutual Information and Network Deconvolution.” *Journal of Intelligent Manufacturing* 31 (7): 1661–1671. doi:10.1007/s10845-018-1397-8.

Qin, W., Z. Zhuang, L. Guo, and Y. Sun. 2022. “A Hybrid Multi-Class Imbalanced Learning Method for Predicting the Quality Level of Diesel Engines.” *Journal of Manufacturing Systems* 62: 846–856. doi:10.1016/j.jmsy.2021.03.014.

Rashidi, B., D. S. Singh, and Q. Zhao. 2018. “Data-Driven Root-Cause Fault Diagnosis for Multivariate Non-Linear Processes.” *Control Engineering Practice* 70: 134–147. doi:10.1016/j.conengprac.2017.09.021.

Ren, L., Z. Meng, X. Wang, R. Lu, and L. T. Yang. 2020. “A Wide-Deep-Sequence Model-Based Quality Prediction Method in Industrial Process Analysis.” *IEEE Transactions on Neural Networks and Learning Systems* 31 (9): 3721–3731. doi:10.1109/TNNLS.2020.3001602.

Reshef, D. N., Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti. 2011. "Detecting Novel Associations in Large Data Sets." *Science* 334 (6062): 1518–1524. doi:10.1126/science.1205438.

Roy, S. S., R. Chopra, K. C. Lee, C. Spampinato, and B. Mohammadi-Ivatlood. 2020. "Random Forest, Gradient Boosted Machines and Deep Neural Network for Stock Price Forecasting: A Comparative Analysis on South Korean Companies." *International Journal of Ad Hoc and Ubiquitous Computing* 33 (1): 62–71. doi:10.1504/IJAHUC.2020.104715.

Roy, S. S., S. F. Mihalache, E. Pricop, and N. Rodrigues. 2022. "Deep Convolutional Neural Network for Environmental Sound Classification via Dilation." *Journal of Intelligent & Fuzzy Systems Preprint* 43 (2): 1–7. doi:10.3233/JIFS-219283.

Roy, S. S., and P. Samui. 2021. "Predicting Longitudinal Dispersion Coefficient in Natural Streams Using Minimax Probability Machine Regression and Multivariate Adaptive Regression Spline." *International Journal of Advanced Intelligence Paradigms* 19 (2): 119–127. doi:10.1504/IJAIP.2021.115244.

Roy, S. S., P. Samui, I. Nagtode, H. Jain, V. Shivaramakrishnan, and B. Mohammadi-Ivatloo. 2020. "Forecasting Heating and Cooling Loads of Buildings: A Comparative Performance Analysis." *Journal of Ambient Intelligence and Humanized Computing* 11 (3): 1253–1264. doi:10.1007/s12652-019-01317-y.

Shimizu, S., P. O. Hoyer, A. Hyvärinen, A. Kerminen, and M. Jordan. 2006. "A Linear Non-Gaussian Acyclic Model for Causal Discovery." *Journal of Machine Learning Research* 7: 2003–2030.

Shi, J., J. Zhao, X. Liu, L. Chen, and T. Li. 2020. "Quantifying Direct Dependencies in Biological Networks by Multiscale Association Analysis." *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 17 (2): 449–458. doi:10.1109/TCBB.2018.2846648.

Shi, J., and S. Zhou. 2009. "Quality Control and Improvement for Multistage Systems: A Survey." *IIE Transactions* 41 (9): 744–753. doi:10.1080/07408170902966344.

Siltepavet, A., S. Sinthupinyo, and P. Chongstitvatana. 2012. "Improving Quality of Products in Hard Drive Manufacturing by Decision Tree Technique." *International Journal of Computer Science (IJCSI)* 9 (3): 29–34.

Solanki, A. B., S. S. Sonigra, and V. Vajpayee. 2023. "Quality Tools Implementation Practice to Reduce Forging Defects in Crankshaft Manufacturing Industries: An Overview." In *Recent Advancements in Mechanical Engineering*, 533–555. Singapore: Springer. doi:10.1007/978-981-19-3266-3_42.

Sun, Y. N., Y. Chen, W. Y. Wang, H. W. Xu, and W. Qin. 2021. "Modelling and Prediction of Injection Molding Process Using Copula Entropy and Multi-Output SVR." In *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*, Lyon, France, 1677–1682.

Sun, Y. N., W. Qin, H. W. Xu, R. Z. Tan, Z. L. Zhang, and W. T. Shi. 2022. "A Multiphase Information Fusion Strategy for Data-Driven Quality Prediction of Industrial Batch Processes." *Information Sciences* 608: 81–95. doi:10.1016/j.ins.2022.06.057.

Sun, Y. N., W. Qin, and Z. L. Zhuang. 2020. "Quality Consistency Analysis for Complex Assembly Process Based on Bayesian Networks." *Procedia Manufacturing* 51: 577–583. doi:10.1016/j.promfg.2020.10.081.

Sun, Y. N., W. Qin, and Z. L. Zhuang. 2021. "Nonparametric-Copula-Entropy and Network Deconvolution Method for Causal Discovery in Complex Manufacturing Systems." *Journal of Intelligent Manufacturing* 33 (6): 1699–1713. doi:10.1007/s10845-021-01751-w.

Sun, Y. N., W. Qin, Z. L. Zhuang, and H. W. Xu. 2021. "An Adaptive Fault Detection and Root-Cause Analysis Scheme for Complex Industrial Processes Using Moving Window KPCA and Information Geometric Causal Inference." *Journal of Intelligent Manufacturing* 32 (7): 2007–2021. doi:10.1007/s10845-021-01752-9.

Sun, Y. N., W. Wang, H. Xu, and W. Qin. 2022. "A Nonlinear Causal Effect-Based Soft Sensing Method for Batch Processes." *Journal of Huazhong University of Science and Technology (Natural Science Edition)* 50 (06): 48–54. doi:10.13245/j.hust.220605.

Sun, Y. N., Z. L. Zhuang, H. W. Xu, W. Qin, and M. J. Feng. 2022. "Data-Driven Modeling and Analysis Based on Complex Network for Multimode Recognition of Industrial Processes." *Journal of Manufacturing Systems* 62: 915–924. doi:10.1016/j.jmsy.2021.04.001.

Tao, F., Q. Qi, L. Wang, and A. Y. C. Nee. 2019. "Digital Twins and Cyber–Physical Systems Toward Smart Manufacturing and Industry 4.0: Correlation and Comparison." *Engineering* 5 (4): 653–661. doi:10.1016/j.eng.2019.01.014.

Tarjan, R. 1972. "Depth-First Search and Linear Graph Algorithms." *SIAM Journal on Computing* 1 (2): 146–160. doi:10.1137/0201010.

Wang, J., C. Xu, J. Zhang, and R. Zhong. 2021. "Big Data Analytics for Intelligent Manufacturing Systems: A Review." *Journal of Manufacturing Systems* 62: 738–752. doi:10.1016/j.jmsy.2021.03.005.

Wang, J., P. Zheng, Y. Lv, J. Bao, and J. Zhang. 2019. "Fog-IBDIS: Industrial Big Data Integration and Sharing with Fog Computing for Manufacturing Systems." *Engineering* 5 (4): 662–670. doi:10.1016/j.eng.2018.12.013.

Wang, J., P. Zheng, and J. Zhang. 2020. "Big Data Analytics for Cycle Time Related Feature Selection in the Semiconductor Wafer Fabrication System." *Computers & Industrial Engineering* 143: 106362. doi:10.1016/j.cie.2020.106362.

Xu, H. W., W. Qin, Y. L. Lv, and J. Zhang. 2022. "Data-Driven Adaptive Virtual Metrology for Yield Prediction in Multi-Batch Wafers." *IEEE Transactions on Industrial Informatics* 18 (12): 9008–9016. doi:10.1109/TII.2022.3162268.

Xu, H., J. Zhang, Y. Lv, and P. Zheng. 2020. "Hybrid Feature Selection for Wafer Acceptance Test Parameters in Semiconductor Manufacturing." *IEEE Access* 8: 17320–17330. doi:10.1109/access.2020.2966520.

Yuan, T., and S. J. Qin. 2014. "Root Cause Diagnosis of Plant-Wide Oscillations Using Granger Causality." *Journal of Process Control* 24 (2): 450–459. doi:10.1016/j.jprocont.2013.11.009.

Zhang, Y., W. Zhang, and Y. Xie. 2013. "Improved Heuristic Equivalent Search Algorithm Based on Maximal Information Coefficient for Bayesian Network Structure Learning." *Neurocomputing* 117: 186–195. doi:10.1016/j.neucom.2013.02.015.

Zha, D., W. Qin, J. Zhang, and Z. Zhuang. 2017. "Correlation Analysis of Diesel Engine Performance Testing Data Based on Mixed-Copula Method." *Data Mining and Big Data* 10387: 48–57. doi:10.1007/978-3-319-61845-6_6.

Zhou, P., Y. Lv, H. Wang, and T. Chai. 2017. "Data-Driven Robust RVFLNs Modeling of a Blast Furnace Iron-Making Process Using Cauchy Distribution Weighted M-Estimation." *IEEE Transactions on Industrial Electronics* 64 (9): 7141–7151. doi:10.1109/tie.2017.2686369.

Zhou, P., R. Zhang, J. Xie, J. Liu, H. Wang, and T. Chai. 2021. "Data-Driven Monitoring and Diagnosing of Abnormal Furnace Conditions in Blast Furnace Ironmaking: An Integrated PCA-ICA Method." *IEEE Transactions on Industrial Electronics* 68 (1): 622–631. doi:10.1109/tie.2020.2967708.

Zhu, J., Z. Ge, Z. Song, L. Zhou, and G. Chen. 2018. "Large-Scale Plant-Wide Process Modeling and Hierarchical Monitoring: A Distributed Bayesian Network Approach." *Journal of Process Control* 65: 91–106. doi:10.1016/j.jprocont.2017.08.011.