Computers & Industrial Engineering 159 (2021) 107513

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Multistage MR-CART: Multiresponse optimization in a multistage process using a classification and regression tree method

Dong-Hee Lee [a,*], So-Hee Kim [a], Kwang-Jae Kim [b]

[a] Division of Industrial Information Studies, Hanyang University, 222 Wangsimni-Ro, Seongdong-Gu, Seoul 04763, Republic of Korea
[b] Department of Industrial and Management Engineering, Pohang University of Science and Technology, 77 Cheongam-Ro, Nam-Gu, Pohang, Gyeongbuk 37673, Republic of Korea

ARTICLE INFO

Keywords:
Multistage process
Multiresponse optimization
Desirability function
Big data
Data mining
Classification and regression tree

ABSTRACT

A multistage process consists of sequential consecutive stages. In this process, each stage has multiple responses and is affected by its preceding stage, while at the same time, affecting the following stage. This complex structure makes it difficult to optimize the multistage process. Recently, it became easy to obtain a large amount of operational data from the multistage process due to development of information technologies. The proposed method employs a data mining method called a classification and regression tree for analyzing the data and desirability functions for simultaneously optimizing the multiresponse. To consider the relationship between stages, a backward optimization procedure which treats the multiresponse of the preceding stage as the input variables is proposed. The proposed method is described using a steel manufacturing process example and is compared with existing multiresponse optimization methods. The case study shows that the proposed method works well and outperforms the existing methods.

1. Introduction

A multistage process that consists of a series of sequential stages is a common structure in manufacturing lines. Most manufacturing industries require several stages to complete their final products, such as semiconductor, printed circuit board, chemical and telecommunication manufacturing processes (Pan, Li, & Wu, 2016). Fig. 1 describes a representative multistage manufacturing process consisting of K stages. The rectangles represent stages, and each stage is followed by its inspection stage shown as a circle. The raw materials enter Stage 1 and become a final product through the K stages. $\mathbf{x}_k$ and $\mathbf{y}_k$ in Fig. 1 are vectors of I_k input variables and J_k response variables, respectively, at the kth stage.

In the multistage process, each stage has multiple responses and is affected by its preceding stage, while at the same time, affecting the following stage. Several methods have been proposed to solve the multiresponse problem in the multistage process. Mukherjee and Ray (2008) employed desirability functions for optimizing multiple responses in a two-stage process. In this approach, empirical models for the multiple responses are fitted, and desirability functions are constructed by using the empirical models. Then, the optimal condition for the input variables is obtained by maximizing the desirability functions. In order to search the optimal condition, several metaheuristics such as genetic algorithm, simulated annealing, Tabu search were employed. Later, Bera and Mukherjee (2016) extended the scope of Mukherjee and Ray (2008)'s method from the two-stage process to the multistage process. Hejazi, Seyyed-Esfahani, and Mahootchi (2015) suggested a mathematical programming method and a metaheuristic algorithm using iterative seemingly unrelated regression for optimizing the multiple responses in the multistage process. Recently, Yin, He, Niu, and Li (2018) suggested a method for optimizing coal preparation production system, which is a particular multistage process. In this method, a forward iterative modeling method based on support vector regression was presented to consider the interdependency between neighboring stages. In addition, a goal-oriented and backward iterative optimization approach based on genetic algorithm was proposed to determine the globally optimal operating conditions of coal preparation system.

The above methods commonly build empirical models for the multiple responses and obtains the optimal setting for the input variables based on the empirical models. Although these methods are attractive approaches, they have a difficulty in that they require a large number of experiments for building the empirical models. In the multistage process, there are various relationships between stages and relationships between the input and responses that should be investigated for the

* Corresponding author.
E-mail address: dh@hanyang.ac.kr (D.-H. Lee).

https://doi.org/10.1016/j.cie.2021.107513
Received 6 February 2021; Received in revised form 13 May 2021; Accepted 22 June 2021
Available online 28 June 2021

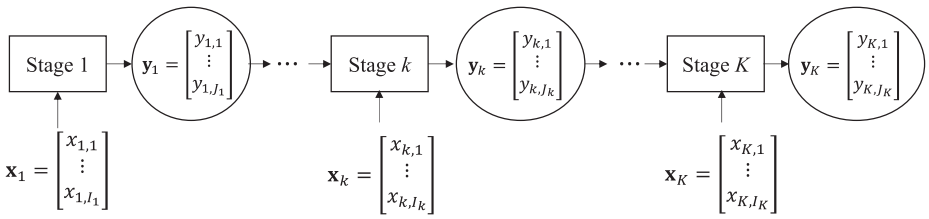

Fig. 1. Schematic of a multistage process (Reprinted from Lee DH, Yang JK, Kim SH, Kim KJ (2020). Optimizing mean and variance of multiresponse in a multistage manufacturing process using operational data. Quality Engineering, 32(4), 627–642 with permission from Taylor & Francis, https://www.tandfonline.com/doi/full/10.1080/08982112.2020.1712727).

optimization. A large number of experiments must be conducted to build empirical models that explain these relationships, which requires large amounts of resources (time, material, machine, etc.).

Alternatively, process operational data gathered from manufacturing lines can be used instead of conducting a large number of experiments. Recently, many manufacturing companies have been able to obtain a large volume and variety of operational data from the manufacturing lines due to network sensors and IoT (Internet of Things). This large and variety of operational data may contain meaningful information. Using data mining methods can be attractive when dealing with a large volume and variety of operational data. Classification and regression tree (CART) and patient rule induction method (PRIM) are representative data mining methods applicable to the process optimization. Recently, Lee, Kim, Kim, Kim, and Zhen (2021) suggested applying CART for optimizing multiple responses in a single stage process. However, none of the methods have employed CART to optimize the multistage process. The proposed method extends the scope of CART-based optimization from single stage to multistage by considering the relationship between stages. For this purpose, a backward sequential optimization procedure suggested. In this procedure, optimization is sequentially conducted from the last stage to the first stage. Additionally, the proposed method employs a desirability function method (Derringer & Suich, 1980) as the objective function of CART for simultaneous optimizing the multiple responses. The proposed method obtains the subregions in the input variables space where high desirability function value is obtained for each stage.

The rest of the paper is organized as follows. Section 2.1 provides a reviews CART which is employed in the proposed method and compares it with PRIM. Section 2.2 reviews desirability function method which is also employed in the proposed method. The proposed method is presented in Section 3 and is illustrated with a case study in Section 4. Finally, a discussion and concluding remarks are given in Section 5.

2. Literature review

2.1. Review of CART and PRIM

In this section, we review CART and PRIM, which are applicable to process optimization. CART was first introduced by Breimen et al. (1984). It is a binary recursive partitioning procedure that finds the subregion in the input variable space where the performance of the response is considerably better. When the response variable is nominal (continuous), it becomes a classification (regression) tree. CART has the advantage of being able to process various types of data (Lee, Jeong, & Lee, 2016). Above all, CART works well with a large volume of data (Tang, Zheng, Huang, & Wang, 2005).

The binary split procedure of CART starts at the root node, the origin of the decision tree, which includes all the data. The data are split into two child nodes. Here, the higher node of the child nodes is called a parent node. Each child node is in turn split into grandchildren, where the child node now becomes a parent node, and the grandchild node becomes a child node. The rule of splitting data is to divide the data into two child nodes whose data are more homogeneous than the current group with respect to the response. The procedure of splitting continues until a set of final nodes (i.e., leaf nodes) with homogeneous data is created. The path from the root node to the leaf node describes the condition of that leaf node (Steinberg, 2009).

In classification tree, the data are split into two child nodes to maximize the degree of homogeneity within each child node. Gini measure is a representative index for measuring the degree of homogeneity. In regression tree, the data are split into two child nodes to minimize mean squared error (MSE) within all of the child nodes. MSE is the average of squared difference between estimated and actual response values. The estimated response is calculated as the average response value of observations included in the leaf node.

CART has the advantage of being able to process various types of data, and it is capable of being used for classification and prediction problems, so it has been widely used in various fields such as healthcare, marketing, agriculture, financial analysis and so on (Lee et al., 2016; Haughton & Oulabi, 1997; Liu, Kanter, Messer, & Kaiser, 2013;

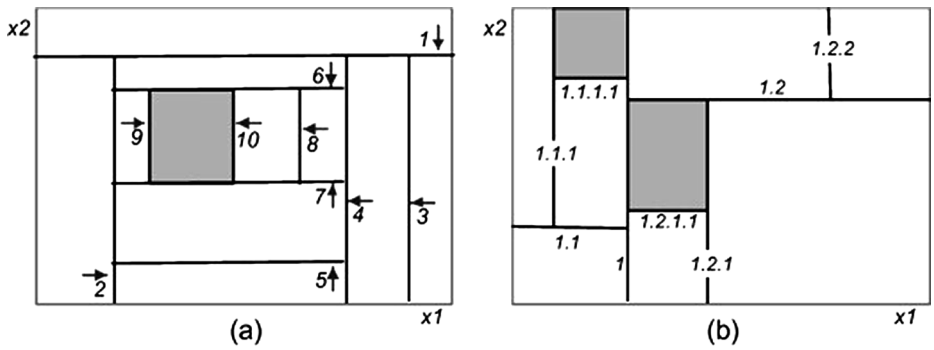

Fig. 2. Illustration of PRIM and CART (Reprinted Abu-Hanna A, Nannings B, Dougelmans D, Hasman A (2010) PRIM versus CART in subgroup discovery: When patience is harmful. Journal of biomedical informatics, 43(5), 701–708, with permission from Elsevier, https://doi.org/10.1016/j.jbi.2010.05.009).

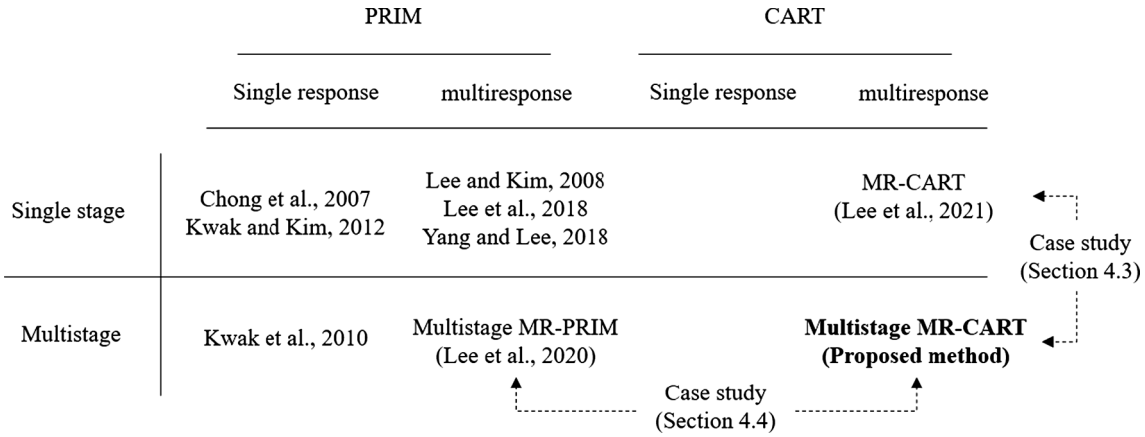

Fig. 3. Classification of data mining-based multiresponse optimization methods.

Emrouznejad & Anouze, 2010). Due to these advantages, we employ CART for optimizing multiple responses in the multistage process.

The patient rule induction method (PRIM) was first introduced by Friedman and Fisher (1999). It is a method of searching a set of subregions called boxes in the input variable space within which the performance of the response is considerably better than that of the entire input domain (Chong, Albin, & Jun 2007). PRIM starts with the entire input variable space and generates candidate boxes by removing the observations for each input variables by α% on each lower and upper side. Then, candidate boxes are generated again from the selected box. This iterative process continues until the portion of observations β becomes less than a predetermined threshold. The parameter α indicates the peeling rate, i.e., how many observations are to be removed from previous box when generating candidate boxes. The parameter β indicates the proportion of the observations remaining in the final box. It should be noted that only a small number of observations are removed to generate candidate boxes by setting a small value for α (usually $\alpha \leq 0.05$). Thus, a long sequence of boxes is generated, and each removal becomes less important in determining the final box and unfortunate removals that remove good observations can be mitigated in later steps (Chong et al., 2007). This approach is called the "patient strategy," which is the core concept of PRIM.

Both PRIM and CART are the same in that they find the optimal conditions directly from the data without constructing a specific model but are different in the way they narrow the input variable space. PRIM narrows the input variable space from the outside inwards in a 'patient' way, while CART does it through a binary splitting procedure in an aggressive way. Fig. 2 compares PRIM and CART through a two-input variable case. Fig. 2(a) illustrates how PRIM peels the two-dimensional input variable space though the iterative procedure. The numbers in Fig. 2(a) mean sequences of selected boxes and a gray rectangle indicates the final subregion. Fig. 2(b) illustrates how CART splits the two-dimensional input variable space. Each number in the Fig. 2(b) means the index of split, and gray rectangles show the final subregions with a sufficiently high response variable value. (Abu-Hanna, Nannings, Dougelmans, & Hasman, 2010).

2.2. Review of CART and PRIM-based optimization methods

Most of the existing multiresponse optimization methods for the multistage process build empirical models for the response variables and obtain the optimal setting for the input variables based on the empirical models. In the multistage process, there are various relationships between stages and relationships between the input and responses that should be investigated. A large number of experiments must be conducted to build empirical models that explain these relationships, which requires large amounts of resources. The proposed method is advantageous than this approach in that it does not require experimental data. Instead, it utilizes CART to analyze the process operational data.

In addition to the proposed method, there exist several CART and PRIM-based optimization methods which do not require the experimental data. Fig. 3 classifies the CART and PRIM-based optimization methods. For each PRIM and CART, optimization methods are classified according to the number of stages and the number of responses. In the case of PRIM, Chong et al. (2007) first proposed to use PRIM to optimize a single response in a single stage process. Then, Lee and Kim (2008), Lee, Yang, and Kim (2018), and Yang and Lee (2018) extended Chong et al. (2007)'s method from a single response optimization to a multiresponse optimization method in the single stage process by employing the desirability function method. Kwak, Kim, and Lee (2010) extended Chong et al. (2007)'s method from the single stage process to the multistage process. The extended model considers the performance of each stage in terms of the relationship between the stages. However, it assumes that there is only one response variable at each stage. Recently, Lee, Yang, Kim, and Kim (2020) extended the scope of process optimization both from single response to multiresponse, and from the mean of the multiple responses to the mean and variance of multiresponse.

In the case of CART, only Lee et al. (2021) proposed to use CART for process optimization. The proposed method is quite comparable to Lee et al. (2021) in that both methods attempt to optimize the multiple responses by employing CART. The difference between them is that Lee et al. (2021) optimized the multiple responses in the single stage process, while the proposed method does it in the multistage process. In this regard, we call Lee et al. (2021)'s method and the proposed method as MR-CART (multiresponse CART) and multistage MR-CART, respectively. The existing MR-CART can also be used in the multistage process. However, this approach neglects the relationships between the stages. Compared to the existing MR-CART, the proposed method can lead to better optimization performance in terms of the whole point of view considering relationships between the stages. This feature will be further explained with a case study in Section 4.3.

The proposed method is also quite comparable to Lee et al. (2020) in that both methods optimize multiple responses in the multistage process. The difference between them is that Lee et al. (2020) employs PRIM, while the proposed method employs CART. In this regard, we call Lee et al. (2020)'s method as multistage MR-PRIM. Both methods work well in optimizing the multiple responses in the multistage process. One critical difference between them is that the existing multistage MR-PRIM provides a single subregion as the final region while the proposed multistage MR-CART provides several subregions as the final regions, as shown in Fig. 2(b). This feature will be further explained with a case study in Section 4.4.

Compared to the existing CART and PRIM-based optimization methods, the proposed method is comprehensive in that it

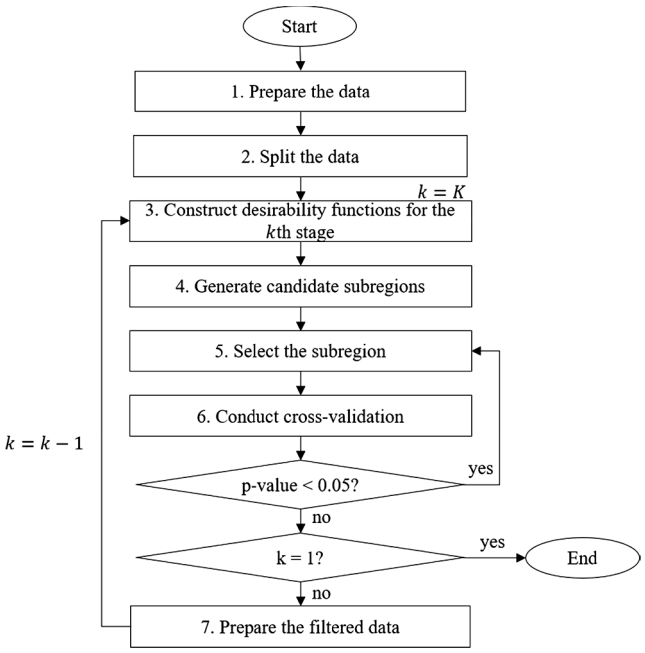

Fig. 4. Overall procedure of the proposed method.

simultaneously considers multiple responses and dependency between the stages. The existing MR-CART does not consider the dependency between the stages. The existing multistage MR-PRIM is a good alternative for optimizing the multiple responses in the multistage process. However, it only obtains a single subregion while the proposed method obtains several subregions. This is useful for optimizing the multiple responses because process engineers can compare the several subregions and better understand tradeoffs between the multiple responses.

3. Proposed multistage MR-CART

In this section, the proposed multistage MR-CART is presented. As mentioned in Section 1, it is important to obtain reliable response surface models in MRSO because the optimal solution is obtained by analyzing the response surface models. Nevertheless, it is not easy to obtain reliable response surface models, especially when dealing with the large amount of data. This is because not only the form of functional relationships between input variables and responses might not be clear, but also the singularity problem can arise when the coefficients of the response surface models are estimated by using the least squares method since the input variables are correlated (Geladi & Kowalski, 1986). In this regard, the proposed method does not construct a specific model. Instead, the proposed method finds the optimal conditions directly from the data. More specifically, the proposed method finds a subregion in the input variable space where the performance of the multiple responses in the multistage process is considerably better. The proposed multistage MR-CART considers tradeoffs between multiple responses through the desirability functions. At the same time, the relationship between the stages to optimize multistage manufacturing process is considered through a backward optimization procedure.

The procedure associated with the proposed method is given in Fig. 4. We assumed that there are K stages to be optimized. In the multistage manufacturing process shown in Fig. 1, the raw materials enter Stage 1 and become a final product through a total of K stages. In contrast, the optimization process is conducted backwards from the final stage to the first stage. We propose this backward optimization approach is to consider that each stage is affected by its preceding stage. Steps 3–6 are iterated to optimize the multiple responses of each stage. Because Stage K is optimized first, the stage iteration counter, k, is first set as $k = K$ and decreases by one whenever the optimization is conducted. This iteration continues until the optimization reaches the first stage. The details of each step are described below.

3.1. Step 1. Prepare the data

Suppose that the entire data with K stages is prepared. For the kth stage, there are I_k input variables, denoted by $x_{k,i}$ for $i = 1,2,\cdots,I_k$, and there are J_k response variables, denoted by $y_{k,j}$ for $j = 1,2,\cdots,J_k$. We have N observations denoted by $\{\left(x_{n,k,i}, y_{n,k,j}\right)$ for $n = 1,2,\cdots,N; k = 1,2,\cdots,K; i = 1,2,\cdots,I_k; j = 1,2,\cdots,J_k\}$, where $x_{n,k,i}$ and $y_{n,k,j}$ are values of the ith input and jth response variables at the kth stage of the nth observation, respectively. Thus, every observation has a total of $\sum_{k=1}^{K} I_k + \sum_{k=1}^{K} J_k$ values.

3.2. Step 2. Split the data

To avoid over-fitting problems, cross-validation is conducted. The entire N observations are randomly divided into training and test datasets. In general, the size of the training dataset is four or five times larger than the size of the test dataset. The training dataset will be used in Steps 3–7 to implement the CART algorithm, the test dataset will be used in Step 6 to conduct cross-validation.

3.3. Step 3. Construct desirability functions for the kth stage

The stage iteration counter k is set as $k = K$ and conducts optimization with the training dataset only. Stage K (i.e., the last stage) is optimized first, and then the optimization process is conducted backwards to the first stage sequentially. Steps 3–7 consists of tree construction and selection in the stage. In this step, $y_{n,k,j}$ is converted to an individual desirability function value $d_{n,k,j}$, which has a value between 0 and 1. Three desirability functions proposed by Derringer and Suich (1980) are employed for the conversion. When $y_{n,k,j}$ is a nominal-the-best (NTB) type response, the desirability function can be expressed as follows:

$$d_{n,k,j} = \begin{cases} 0, & \text{if } y_{n,k,j} \le y_{k,j}^{\min} \text{ or } y_{n,k,j} \ge y_{k,j}^{\max} \\ \left(\dfrac{y_{n,k,j} - y_{k,j}^{\min}}{T_{k,j} - y_{k,j}^{\min}}\right)^{s_{k,j}}, & \text{if } y_{k,j}^{\min} < y_{n,k,j} < T_{k,j} \\ \left(\dfrac{y_{k,j}^{\max} - y_{n,k,j}}{y_{k,j}^{\max} - T_{k,j}}\right)^{t_{k,j}}, & \text{if } T_{k,j} \le y_{n,k,j} < y_{k,j}^{\max} \end{cases}$$

where $y_{k,j}^{\min}$ and $y_{k,j}^{\max}$ are the minimum and maximum acceptable limits for $y_{n,k,j}$, respectively. $T_{k,j}$ is a target value for $y_{n,k,j}$. $s_{k,j}$ and $t_{k,j}$ are parameters that determine the shape of $d_{n,k,j}$. When $y_{n,k,j}$ is a larger-the-better (LTB) type response, the desirability function can be expressed as follows:

$$d_{n,k,j} = \begin{cases} 0, & \text{if } y_{n,k,j} \le y_{k,j}^{\min} \\ \left(\dfrac{y_{n,k,j} - y_{k,j}^{\min}}{y_{k,j}^{\max} - y_{k,j}^{\min}}\right)^{s_{k,j}}, & \text{if } y_{k,j}^{\min} < y_{n,k,j} < y_{k,j}^{\max} \\ 1, & \text{if } y_{n,k,j} \ge y_{k,j}^{\max} \end{cases} \quad (2)$$

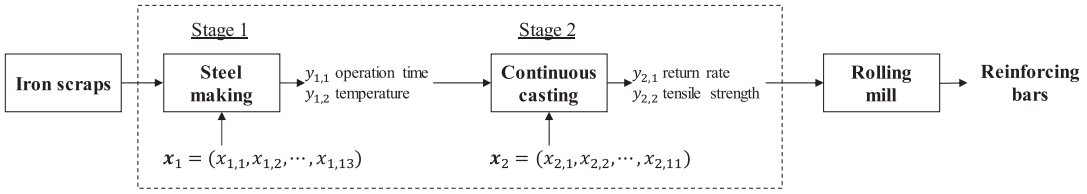

Fig. 5. Steel manufacturing process in the cast study (Lee et al., 2020). (Reprinted from Lee DH, Yang JK, Kim SH, Kim KJ (2020). Optimizing mean and variance of multiresponse in a multistage manufacturing process using operational data. Quality Engineering, 32(4), 627–642 with permission from Taylor & Francis, https://www.tandfonline.com/doi/full/10.1080/08982112.2020.1712727).

Similarly, when $y_{n,k,j}$ is a smaller-the-better (STB) type response, the desirability function can be expressed as follows:

$$d_{n,k,j} = \begin{cases} 1, & \\ \left(\dfrac{y_{k,j}^{\max} - y_{n,k,j}}{y_{k,j}^{\max} - y_{k,j}^{\min}}\right)^{t_{k,j}}, & \text{if } y_{n,k,j} \le y_{k,j}^{\min} \\ 0, & \\ & \text{if } y_{k,j}^{\min} < y_{n,k,j} < y_{k,j}^{\max} \end{cases}$$

$$\text{if } \; y_{n,k,j} \ge y_{k,j}^{\max} \tag{3}$$

Once all $y_{n,k,j}$ for $j = 1, 2, \cdots, J_k$ are converted to $d_{n,k,j}$, $d_{n,k,j}$'s are aggregated into an overall desirability function value $D_{n,k}$. More specifically, J_k individual desirability function values at the kth stage of the nth observation (i.e., $d_{n,k,1}$, $d_{n,k,2}$, $\cdots$, d_{n,k,J_k}) are aggregated into $D_{n,k}$ by Equation (4). Once every observation has $D_{n,k}$ value, $\overline{D}_k$ value, which is an average of $D_{n,k}$'s at the kth, is calculated and this value is used an objective function value of CART.

$$D_{n,k} = \left(\prod_{j=1}^{J_k} d_{n,k,j}\right)^{\frac{1}{J_k}} \tag{4}$$

3.4. Step 4. Generate candidate subregions

In this step, CART is performed to generate candidate subregions (i.e., leaf nodes) whose $\overline{D}_k$ is large. The min leaf should be determined before performing CART. Min leaf means a minimum number of observations included in the nodes of the tree; thus, it acts as a stopping criterion for splitting the nodes. The smaller the min leaf, the better the objective function value (i.e., $\overline{D}_k$, value). However, the small min leaf value makes additional tree splitting and this can result in overfitting. One important issue in performing CART is considering the relationships between stages. To reflect the relationship between Stage $k-1$ and Stage k, the response variables of Stage $k-1$ (i.e., $y_{k-1,j}$) become the input variables of Stage k. Therefore, the input variable space of Stage k is extended to a $(I_k + J_{k-1})$-dimensional space including I_k input variables of Stage k and the J_{k-1} response variables of Stage $k-1$ (i.e., $\left(y_{k-1,j}, x_{k,i}\right)$). The binary split procedure of CART starts at the root node, and the tree is constructed through the binary split procedure of CART to generate leaf nodes having large $\overline{D}_k$ values until the stopping criterion, min leaf, is satisfied. The generated leaf nodes are evaluated in Step 5.

On the other hand, when $k = K$, the root node contains all observations of the training dataset $\{(y_{n,k-1,j}, x_{n,k,i}, y_{n,k,j})$ for $n = 1, 2, \cdots, N; i = 1, 2, \cdots, I_k; j = 1, 2, \cdots, J_k\}$; thus, CART is performed by the entire data. Whenever k decreases by 1, observations are filtered in Step 7 and CART is performed by the filtered data. The filtering process will be explained in Step 7.

3.5. Step 5. Select the subregion

Among candidate subregions, a process engineer selects a subregion considering tradeoffs between the multiple responses or just simply selects the one having the largest $\overline{D}_k$ value. In addition, closeness to the existing process operation conditions and acceptability in operating the process can also be considered. That is, the selection can be flexibly conducted according to the engineer's judgement. In this context, multiple objective optimization method for decision making such as Pareto front can also be performed (Chapman, Lu, & Anderson-Cook, 2014).

3.6. Step 6. Conduct cross-validation

The cross-validation is conducted to investigate whether the selected subregion is over-fitted or not. More specifically, statistical tests such two-sample t-test or Mann Whitney test can be conducted to investigate whether the difference in the mean of D_k between training and test datasets is significant. If the p-value from the two-sample t-test is greater than 0.05, the difference between training and test datasets is regarded as not significant, thus, the selected subregion is considered as not over-fitted. Otherwise, the procedure goes back to Step 5 and selects the other candidate subregion to obtain other subregions which are not over-fitted. This repetition continues until at least one leaf node having the p-value larger 0.05 is found. If the p-value is larger than 0.05, the stage iteration counter k is investigated to go to Step 7. If k is greater than 1, k decreases by one and the procedure goes to Step 7 to filter data and to optimize its preceding stage (i.e., Stage k-1). Otherwise (i.e., $k = 1$), the procedure ends.

On the other hand, not only a single partition for cross-validation, but also k-fold cross-validation can be conducted. The k-fold cross-validation uses k-fold data partitions and uses results from each partition to evaluate the difference of performance between training and test datasets. In this paper, only a single cross-validation is performed for simplicity.

3.7. Step 7. Preparev the filtered data

In this step, filtered data are prepared to optimize Stage $k-1$. This step is conducted only when the stage iteration counter k is greater than 1. The final subregion is selected from the candidate subregions in Step 5. The final subregion is located in the extended input variables space of Stage k (i.e., $\left(y_{k-1,j}, x_{k,i}\right)$). If $y_{k-1,j}$ participate in the binary splitting procedure in generating the final subregion, the subrange of $y_{k-1,j}$ shrinks because of the binary splitting procedure. Then, the criteria for filtering is defined as $y_{n,k-1,j} \in s^*_{y_{k-1,j}}$ for $j = 1, 2, \cdots, J_{k-1}$, where $s^*_{y_{k-1,j}}$ is the subrange of $y_{k-1,j}$ in the final subregion of Stage k. Filtered data are prepared to optimize Stage $k-1$, as shown below. The number of filtered data is denoted as N_f.

$$\left\{ \left(y_{n,k-2,j}, x_{n,k-1,i}, y_{n,k-1,j}\right) | y_{n,k-1,j} \in s^*_{y_{k-1,j}} \text{for} j = 1, 2, \cdots, J_{k-1} \right\}$$

Once the filtered data are prepared, the procedure goes back to Step 3, and this iteration continues until $k = 1$. If $y_{k-1,j}$ does not participate in the binary splitting procedure generating the final subregion, the filtering process is skipped. This means that $y_{k-1,j}$ does not affect $y_{k,j}$.

3.8. Case study: Steel manufacturing process

The steel manufacturing process is a typical continuous process and a multistage process at the same time. Usually, this type of continuous multistage process includes various stages, and each stage has many input and response variables. The relationships among input and response variables are quite complex and the dependency between the stages makes the problem more complicated. The proposed method attempts to obtain optimal region of input variables by considering these complex relationships. For this purpose, the proposed method analyzes the large amount of operational data using CART and simultaneously optimizes the multiple responses using the desirability function method. To consider the relationship between stages, a backward optimization procedure which treats the multiple responses of the preceding stage as the input variables is proposed. In this section, the proposed method is illustrated by an example related to a steel manufacturing process problem.

The steel manufacturing process for producing reinforcing bars consists of four stages: iron scrap, steel making, continuous casting, and roll milling. Among the four stages, the scope of this case study is limited to the steel making and continuous casting stages (i.e., a two-stage process), and they are denoted as Stage 1 and Stage 2, respectively, as shown in the dotted box in Fig. 5.

In Stage 1, the iron scraps are melted to create molten iron. In the following Stage 2, the molten iron takes a solid shape, referred to as a billet. There are two response variables in Stage 1, which are the operation time in the steel making process ($y_{1,1}$, min) and molten iron temperature ($y_{1,2}$, ℃). $y_{1,1}$ and $y_{1,2}$ are smaller-the-better (STB) type and nominal-the-better (NTB) type variables, respectively. If $y_{1,2}$ is out of a certain range, the equipment can be damaged, or billets may harden prematurely. Therefore, it is important to meet the target of the temperature. The target value of $y_{1,2}$ is 1580℃. There are 13 input variables in Stage 1. $x_{1,1}$,$x_{1,2}$,$x_{1,3}$ are related to power consumption (kW) in the steel making process. $x_{1,4}$,$x_{1,5}$,$x_{1,6}$ are amounts of input materials (ton) injected into electric furnace in the steel making process. $x_{1,7}$,$x_{1,8}$,$x_{1,9}$,$x_{1,10}$,$x_{1,11}$ are amounts of quicklime (kg), Dolomite stone (kg), lump coal (kg), Fe-Si (kg), Si-Mn (kg), Fe-Nb (kg), respectively. Lastly, $x_{1,12}$, and $x_{1,13}$ are amounts of oil and oxygen in the steel making process, respectively.

In Stage 2, there are two response variables, which are return rate ($y_{2,1}$,%) and tensile strength ($y_{2,2}$, MPa), both of which are larger-the-better (LTB) type variables. $y_{2,1}$, an important performance measure in the steel industry, is defined as the ratio of the amount of iron scraps used in the steel-making stage to the number of billets produced in the continuous casting stage (Kwak et al., 2010). There are 11 input variables in Stage 2. $x_{2,1}$ is elapsed time (min) for transferring the molten steel in the continuous casting process. $x_{2,2}$ is the operation time (min) in the continuous casting process. $x_{2,3}$ is power consumption (kW) in the continuous casting process. $x_{2,4}$,$x_{2,5}$,$x_{2,6}$,$x_{2,7}$, and$x_{2,8}$ are amount of Dolomite stone (kg), Fe-Si (kg), Si-Mn (kg), Fe-Nb (kg), Recarburizer (kg) included in the iron, respectively. $x_{2,9}$, and $x_{2,10}$ are moving and waiting time (min) of a continuous carting machine. Lastly, $x_{2,11}$ is an elapsed time (min) for operation of the continuous casting process.

In Section 4.1, the proposed method is demonstrated by a step-by-step procedure with the steel manufacturing process example. In Section 4.2, the example is solved again to investigate the effect of the min leaf parameter. In Sections 4.3 and 4.4, the proposed method is compared with MR-CART (Lee et al., 2021) and the multistage MR-PRIM (Lee et al., 2020), respectively. We use the software package MATLAB R2018b software. Also, we conduct pruning by a technique provided MATLAB R2018b software.

Table 1
The target value and specification limits for response variables.

	Continuous casting (Stage 2)		Steel making (Stage 1)	
	$y_{2,1}$(LTB)	$y_{2,2}$(LTB)	$y_{1,1}$(STB)	$y_{1,2}$(NTB)
Lower and upper specification limit	[80, 100]	[200, 300]	[30, 50]	[1530, 1620]
Target	100	300	30	1580

4. Optimization of the steel manufacturing process problem

4.1. Step 1. Prepare the data

We have a total of 5609 observations denoted by $\{\left(x_{n,k,i}, y_{n,k,j}\right)$ for $n = 1,2,\cdots,5609; k = 1,2; i = 1,2,\cdots,I_k; j = 1,2,\cdots,J_k\}$. The numbers of input variables of Stage 1 and 2, denoted by I_1 and I_2, are 13 and 11, respectively. The numbers of response variables of Stage 1 and 2, denoted by J_1 and J_2, are two for each, as shown in Fig. 5. Thus, every observation includes 28 values (i.e., $28 = 13 + 11 + 2 + 2$).

4.2. Step 2. Split the data

The entire 5609 observations are randomly divided into training and test datasets at a ratio of 4:1 to conduct cross-validation to avoid overfitting. As a result, 4487 observations are used in Steps 3, 4, and 5 to implement the CART algorithm, and the other 1122 observations are used in Step 6 to conduct cross-validation.

4.3. Step 3. Construct desirability functions for Stage 2

Step 3 is conducted for optimizing Stage 2 with the training dataset only. For $y_{2,1}$ and $y_{2,2}$, individual desirability functions are constructed by considering the target value and specification limits. The target value and specification limits are set as shown in Table 1.

Since $y_{2,1}$ and $y_{2,2}$ are LTB type variables, Equation (2) is employed to construct $d_{2,1}$ and $d_{2,2}$. Similarly, $y_{1,1}$ and $y_{1,2}$ employs Equations (3) and (1) because they are STB and NTB type variables, respectively. We used linear functions for each individual desirability function as shown in Fig. 6. Based on the two left functions in Fig. 6, $y_{n,1,1}$ and $y_{n,1,2}$ of every observation are transformed into $d_{n,1,2}$ and $d_{n,1,2}$, respectively. Then, $d_{n,1,2}$ and $d_{n,1,2}$ of every observation is aggregated into $D_{n,2} = \sqrt{d_{n,2,1}d_{n,2,2}}$. This geometric mean means equally important for the two responses. There may be cases where one response is more important than the others. In this case, weights can be assigned to more important responses using weighted geometric mean (Derringer, 1994). Nevertheless, it is assumed in this study that the same weight is given for model simplicity.

4.4. Step 4. Generate candidate subregions

A binary splitting procedure for CART was conducted to maximize the objective function as $\overline{D}_k$ until the numbers of leaf nodes became smaller than 150. It should be noted that the response variables of Stage 1 (i.e., $y_{1,1}$ and $y_{1,2}$) were treated as the input variables of Stage 2 to reflect the relationship between Stage 1 and Stage 2. Therefore, the input variable space of Stage 2 was extended from 11-dimensional space to 13-dimensional space.

We chose a min leaf value of 150 to ensure a fair comparison between the proposed method and the multistage MR-PRIM. When the min leaf value is 150, the number of observations of the final region is similar to that of the optimal box obtained from the multistage MR-PRIM. The comparison will be explained in Section 4.4. On the other hand, a smaller min leaf produces a better the objective function value (i.e., $\overline{D}_2$), but at the same time the tree is split more finely, which can result in

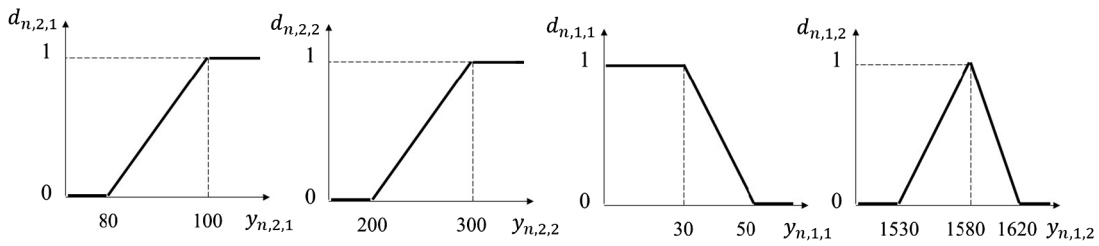

Fig. 6. Four individual desirability functions.

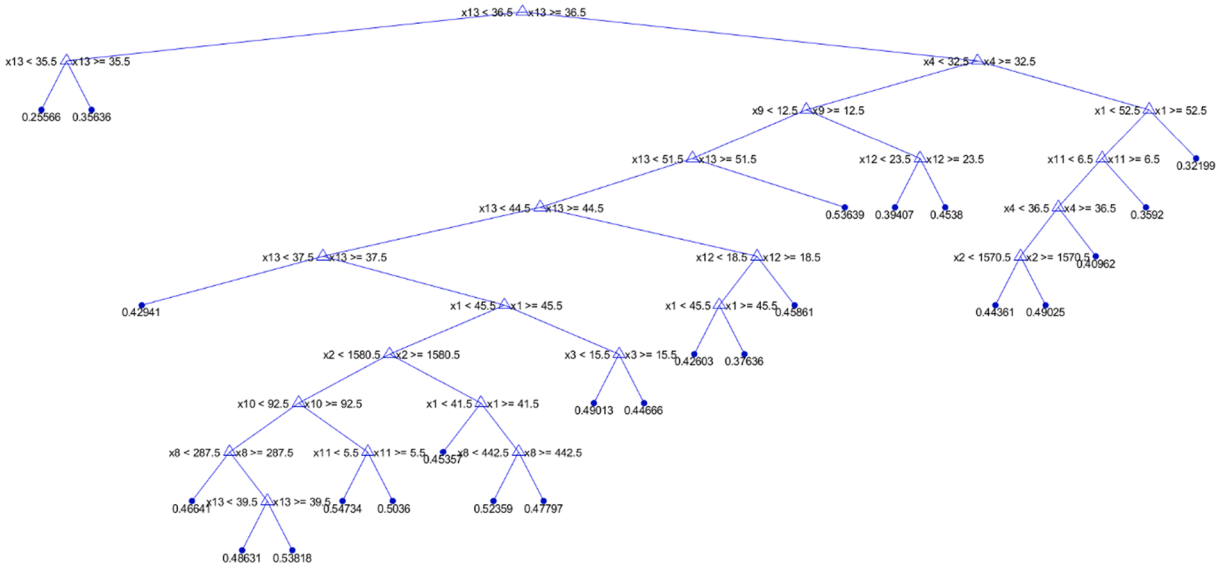

Fig. 7. Tree generated by the proposed method for the case study.

Table 2
Results of the optimization for Stage 2.

Subregion (leaf node)	Splitting criteria of input variables								The number of observations	Response at subregion		
	$y_{1,1}$	$y_{1,2}$	$x_{2,2}$	$x_{2,6}$	$x_{2,7}$	$x_{2,8}$	$x_{2,9}$	$x_{2,11}$		$\overline{D}_2$	$\overline{y}_{2,1}$	$\overline{y}_{2,2}$
1	<45.5	<1580.5	<32.5		<12.5	92.5≤	<5.5	37.5 ≤&<44.5	272	0.5473	92.45	255.50
2	<45.5	<1580.5	<32.5	287.5≤	<12.5	<92.5		39.5≤& <44.5	166	0.5382	92.49	254.68
3			<32.5		<12.5			51.5≤	240	0.5364	92.26	253.37
4	41.5≤&<45.5	1580.5≤	<32.5	<442.5	<12.5			37.5≤&<44.5	151	0.5236	92.23	250.91
5	<45.5	<1580.5	<32.5		<12.5	92.5≤	5.5≤	37.5≤ &<44.5	159	0.5036	91.91	250.21

overfitting. In this regard, we conducted a sensitivity analysis with respect to several min leaf values to investigate the effect on min leaf in Section 4.2.

Fig. 7 shows the pruned tree obtained by the proposed method. We obtained 24 leaf nodes through the binary splitting procedure. Among them, we selected top 5 leaf nodes having large $\overline{D}_2$ values. Table 2 shows the top 5 leaf nodes. Among the 13 variables, $x_{2,2}, x_{2,6}, x_{2,7}, x_{2,8}, x_{2,9}, x_{2,11}, y_{1,1}$ and $y_{1,2}$ participated in splitting the root node into the 5 leaf nodes. Splitting criteria for each leaf node are given in Table 2. Blanks in the table mean that no splitting criteria exists. The number of observations satisfying all the splitting criteria for each Subregion is shown in Table 2. For example, the 272 observations in Subregion 1 satisfy the splitting criteria $y_{1,1}$ <45.5, $y_{1,2}$ <1580.5, $x_{2,2}$ <32.5, $x_{2,7}$ <12.5, $x_{2,8}$ ≥92.5, $x_{2,9}$ <5.5, and 37.5≤ $x_{2,11}$ <44.5. The last three columns in

Table 3
Results of cross-validation for Stage 2.

Subregion (leaf node)	Training dataset			Test dataset			$\Delta\overline{D}_2$	p-value
	Number of observations	Response at subregion		Number of observations	Response at subregion			
		$\overline{D}_2$	s_D		$\overline{D}_2$	s_D		
1	272	0.5473	0.2086	49	0.5257	0.2106	0.0216	0.506

Table 4
Results of optimization for Stage 1.

Subregion (leaf node)	Splitting criteria of input variables				The number of observations	Response at subregion		
	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,10}$		$\overline{D}_1$	$\overline{y}_{1,1}$	$\overline{y}_{1,2}$
1			<8775	<155.5	220	0.5966	40.97	1570.01
2	<10135	<5825	8775≤&<9825		164	0.5771	41.87	1571.67
3		<5905	<8775	155.5≤	228	0.5591	41.73	1568.64
4	<9745	5905≤	<8775	155.5≤	175	0.5469	41.73	1567.00
5	<10135	5825≤ &<6415	8775≤ &<9825		180	0.5393	42.48	1569.48

Table 2 report $\overline{D}_2$, $\overline{y}_{2,1}$, and $\overline{y}_{2,2}$ at the 5 subregions.

4.5. Step 5. Select the subregion

Among the 5 subregions, Subregion 1 has the largest $\overline{D}_2$ value of 0.5473, so it is reasonable to select. It should be noted that selecting the subregions is not always obvious in practice. Thus, the process engineer can select the final subregion considering other things such as practicable operational input variable conditions based on their professional judgement.

4.6. Step 6. Conduct cross-validation

The cross-validation is conducted to determine whether the selected subregion is over-fitted or not. Observations from the test dataset that satisfy the splitting criteria of Subregion 1 are selected. As a result, 49 observations from the test dataset are selected. Now, we have 272 and 49 observations of training and test datasets, respectively. As shown in Table 3, $\overline{D}_2$ value from the training dataset is larger than that of the test dataset (i.e., 0.5473 >0.5257). Generally, the performance of the training dataset is better than that of the test dataset because the subregions were obtained from the training dataset. Next, we conduct a statistical test to compare the two samples. The last column of Table 3 shows the p-value obtained from the two-sample *t*-test. Because the p-value is greater than 0.05, we conclude that the difference between training dataset and test dataset is not significant which means that Subregion 1 is not-overfitted. Because $k \neq 1$, the procedure goes to Step 7.

4.7. Step 7. Prepare the filtered data

Subregion 1 is obtained by splitting 7 input variables as shown in Table 2. It should be noted that $y_{1,1}$ and $y_{1,2}$ are included in the 7 variables. The ranges of $y_{1,1}$ and $y_{1,2}$ (i.e., $y_{1,1}$<45.5, $y_{1,2}$<1580.5) are resulted from the optimization of Stage 2, therefore this subrange should be preserved to ensure the result of optimization of Stage 2 when optimizing Stage 1. Thus, the observations whose $y_{1,1}$ values were less than 45.5 and $y_{1,2}$ values were less than 1580.5 were only included in the training dataset. The number of filtered data satisfying the filtering criteria is $N_f = 2158$. The procedure goes back to Step 3 to optimize Stage 1.

4.8. Step 3. Construct desirability functions for Stage 1

Step 3 for Stage 1 is similar to those of Stage 2. The $y_{n,1,1}$ and $y_{n,1,1}$ values of the filtered observations are transformed into $d_{n,1,1}$ and $d_{n,1,2}$ based on the two right functions in Fig. 6. Like $D_{n,2}$, $D_{n,1}$ is also defined as a geometric mean of $d_{n,1,1}$ and $d_{n,1,2}$.

4.9. Step 4. Generate candidate subregions

Because the preceding stage of Stage 1 does not exist, Stage 1 is optimized only by the input variables of itself (i.e., $x_{1,i}$s) without any response variable of other stages. Table 4 reports the results of optimization for Stage 1. Among the 13 input variables, 4 of them participated in splitting the root node into the 5 leaf nodes.

4.10. Step 5. Select the subregion

Here, Subregion 1, which has the largest $\overline{D}_1$ value of 0.5966, is selected as the subregion.

4.11. Step 6. Conduct cross validation

Table 5 shows the results of cross-validation for Stage 1. The last column of Table 5 shows the p-value obtained from the two-sample *t*-test. Because the p-value is greater than 0.05, the difference between training dataset and test dataset is not significant. Thus, Subregion 1 can be considered not-overfitted. Because $k = 1$, the procedure ends.

The left part of Table 6 compares subranges of the input variables of Stage 2 before and after optimization. The subranges in "Before optimization" are obtained from the entire observations (i.e., the training dataset of 4487 observations) and those in "After optimization" are obtained from the 272 observations satisfying the splitting criteria of Subregion 1. These narrowed subranges are optimal interval where the responses in Stage 2 are optimized. The right part of Table 6 reports the subranges of input variables in Stage 1. Like Stage 2, the subranges became narrowed after optimizing Stage 1.

Fig. 8 compares $\overline{D}_2$, $\overline{y}_{2,1}$, $\overline{y}_{2,2}$, $\overline{D}_1$, $\overline{y}_{1,1}$, and $\overline{y}_{1,2}$ before and after optimization for training and test datasets. For each figure, gray and black bar represent before and after optimization, respectively. All six measures are improved compared to "before optimization," except for $\overline{y}_{1,2}$. Also, the six measures except for $\overline{y}_{1,2}$ are improved more in training dataset than test dataset as we expected. The target value of $y_{1,2}$ is 1580 as shown in Fig. 6. Before optimization, $\overline{y}_{1,2}$ was 1576 but it decreased to 1570 after optimization. This means that the proposed method improved $\overline{y}_{1,1}$ by sacrificing $\overline{y}_{1,2}$ to maximize $\overline{D}_1$. If $\overline{y}_{1,2}$ is more important than $\overline{y}_{1,1}$, weighted geometric mean suggested by Derringer (1994) can be alternative to improve $\overline{y}_{1,2}$ by assigning a large weight to $y_{1,2}$.

Table 5
Results of cross-validation for Stage 1.

Subregion (leaf node)	Training dataset			Test dataset			$\Delta\overline{D}_1$	p-value
	Number of observations	Response at subregion		Number of observations	Response at subregion			
		$\overline{D}_1$	s_D		$\overline{D}_1$	s_D		
1	220	0.5966	0.0851	56	0.5746	0.1019	0.022	0.185

Table 6
Comparison of subranges of input variables before and after optimization.

Continuous casting (Stage 2)			Steel making (Stage 1)		
Variable	Before optimization	After optimization	Variable	Before optimization	After optimization
$x_{2,1}$	[1,264]	[8,109]	$x_{1,1}$	[1050,13590]	[8170,12130]
$x_{2,2}$	[13,88]	[15,32]	$x_{1,2}$	[2330,15240]	[4470,7840]
$x_{2,3}$	[1200,12100]	[2200,4300]	$x_{1,3}$	[4970,18500]	[6650,8770]
$x_{2,4}$	[0,600]	[0,600]	$x_{1,4}$	[16,39.66]	[24.27,38.8]
$x_{2,5}$	[0,500]	[0,170]	$x_{1,5}$	[17.6,35.2]	[19.64,30.88]
$x_{2,6}$	[0,1425]	[0,835]	$x_{1,6}$	[4.36,23.86]	[6.32,20.82]
$x_{2,7}$	[0,25]	[0,12]	$x_{1,7}$	[0,2213]	[685,1910]
$x_{2,8}$	[0,270]	[95,195]	$x_{1,8}$	[0,1211]	[0,407]
$x_{2,9}$	[0,120]	[0,5]	$x_{1,9}$	[302,1416]	[507,1012]
$x_{2,10}$	[0,89]	[1,60]	$x_{1,10}$	[0,416]	[426,1052]
$x_{2,11}$	[11,99]	[38,44]	$x_{1,11}$	[399,1329]	[0,220]
$y_{1,1}$	[36,426]	[38,45]	$x_{1,12}$	[0,373]	[103,155]
$y_{1,2}$	[1518,1662]	[1541,1580]	$x_{1,13}$	[1026,3710]	[426,1052]

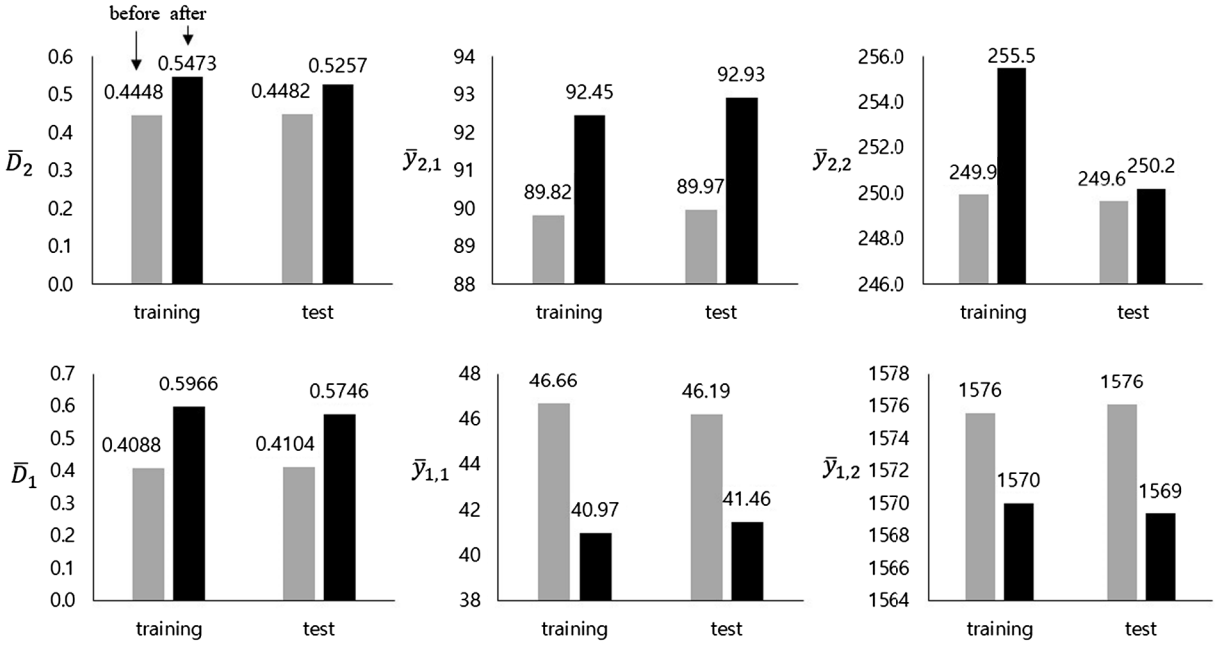

Fig. 8. Comparison of desirability function and responses before and after optimization.

Table 7
Results of a sensitivity analysis with respect to the min leaf value.

Continuous casting (Stage 2)					Steel making (Stage 1)				
Min leaf	Training $\overline{D}_2$	Test $\overline{D}_2$	$\Delta\overline{D}_2$	p-value	Min leaf	Training $\overline{D}_1$	Test $\overline{D}_1$	$\Delta\overline{D}_1$	p-value
25	0.6734	0.4291	0.2443	0.003	25	0.6545	0.6474	0.0072	0.836
50	0.6234	0.5082	0.1152	0.049	50	0.6373	0.5571	0.0802	0.001
75	0.5879	0.5248	0.0631	0.329	75	0.6163	0.5658	0.0506	0.019
100	0.5752	0.5417	0.0335	0.506	100	0.6163	0.5658	0.0506	0.019
125	0.5579	0.5663	−0.0085	0.857	125	0.5966	0.5746	0.0220	0.185
150	0.5473	0.5257	0.0216	0.506	150	0.5966	0.5746	0.0220	0.185
175	0.5439	0.5488	−0.0049	0.901	175	0.5966	0.5746	0.0220	0.185
200	0.5439	0.5488	−0.0049	0.901	200	0.5966	0.5746	0.0220	0.185

4.12. Sensitivity analysis with respect to the min leaf parameter

Min leaf reflects the minimum number of observations included in the leaf nodes of the tree. The min leaf should be set before starting the CART procedure because it acts as a stopping criterion for splitting the nodes. In Section 4.1, we set the min leaf to 150 for a fair comparison with the multistage MR-PRIM in Section 4.4. However, as mentioned in Section 3, smaller values of min leaf produce better objective function values, but at the same time the tree is split more finely, which can result in overfitting. In this regard, we conducted a sensitivity analysis with

Table 8
Comparison of MR-CART and the multistage MR-PRIM.

	Continuous casting (Stage 2)				Steel making (Stage 1)		
	Proposedmethod	Single-stage MR-CART	Multistage MR-PRIM		Proposedmethod	Single-stage MR-CART	Multistage MR-PRIM
The number of observations	272	240	224	The number of observations	220	228	213
$\overline{D}_2$	0.5473	0.5364	0.5190	$\overline{D}_1$	0.5966	0.5591	0.4953
$\overline{y}_{2,1}$	92.45	92.26	91.68	$\overline{y}_{1,1}$	40.97	41.73	43.15
$\overline{y}_{2,2}$	255.50	253.37	254.88	$\overline{y}_{1,2}$	1570.01	1568.64	1579.73

Table 9
Comparison of criteria of the input variable with the multistage MR-PRIM.

Continuous casting (Stage 2)			Steel making (Stage 1)		
Variable	Proposed method criteria	PRIM method criteria	Variable	Proposed method criteria	PRIM method criteria
$x_{2,1}$			$x_{1,1}$		8980≤&≤10520
$x_{2,2}$	<32.5	23≤&≤32	$x_{1,2}$		4740≤&≤6950
$x_{2,3}$		2900≤&≤3900	$x_{1,3}$	<8775	7860≤&≤9830
$x_{2,4}$		290≤	$x_{1,4}$		
$x_{2,5}$			$x_{1,5}$		19.4≤&≤26.76
$x_{2,6}$		370≤&≤700	$x_{1,6}$		9.7≤&≤21.24
$x_{2,7}$	<12.5	6≤&≤20	$x_{1,7}$		935≤&≤1740
$x_{2,8}$	92.5≤	15≤&≤180	$x_{1,8}$		≤400
$x_{2,9}$	<5.5	3≤&≤10	$x_{1,9}$		410≤&≤1009
$x_{2,10}$		1≤&≤32	$x_{1,10}$	<155.5	
$x_{2,11}$		40≤&≤76	$x_{1,11}$		402≤&≤1055
$y_{1,1}$	<45.5	40≤&≤81	$x_{1,12}$		105≤&≤176
$y_{1,2}$	<1580.5		$x_{1,13}$		1971≤&≤2514

respect to several min leaf values to investigate the effect of min leaf.

Table 7 shows the results of the sensitivity analysis with respect to several min leaf values. The left part of Table 7 reports the results of optimization for Stage 2. As we expected, smaller min leaf values produced a larger $\overline{D}_2$ values in Stage 2. However, it should be noted that when min leaf is 25 or 50, p-values of the two-sample *t*-test for cross-validation are<0.05, which means that the subregions are overfitted. According to the results of the sensitivity analysis, it is recommended to set the min leaf as 75 at which $\overline{D}_2$ has the largest value except for leaf nodes that are over-fitted. The right part of Table 7 shows the sensitivity analysis for Stage 1 with respect to same min leaf values with Stage 2. $\overline{D}_1$ value increased as the min leaf value decreased as in Stage 2. Also, the difference in $\overline{D}_1$ (i.e., $\Delta\overline{D}_1$) between the training and test datasets also increased as the smaller min leaf value was applied, except when the min leaf value was set to 25.

4.13. Comparison with the MR-CART

In this section, we compare the proposed multistage MR-CART with the existing MR-CART proposed by Lee et al. (2021). We used the same data in Section 4.1 for a fair comparison. The existing MR-CART does not consider the relationships between stages; thus, it can also be called single-stage MR-CART. Specifically, the response variables of Stage $k-1$ (i.e., $y_{n,k-1,j}$) are not treated as input variables for optimizing Stage k. Thus, the dataset is not filtered as well when optimizing Stage $k-1$.

Table 8 compares the optimization results of the proposed method and the single-stage MR-CART. The second column of each stage shows the result of the optimization using single-stage MR-CART. In Stage 2, the $\overline{D}_2$ value of the proposed method was larger than that of the single-stage MR-CART. This is because the proposed method treats $y_{1,1}$ and $y_{1,2}$ as input variables for optimizing Stage 2, while the single-stage MR-CART does not. Additionally, the $\overline{D}_1$ value of proposed method was also higher than that of the single-stage MR-CART in Stage 1. This seems to be due to data filtering. The proposed method optimizes Stage 1 using a filtered dataset in the criteria for $y_{1,1}$ involved in splitting (i.e., $y_{1,1}<45.5$ and $y_{1,2}<1580.5$), as shown in Table 2. The filtered data are close to the targets $y_{1,1}$, which is advantageous in optimizing Stage 1. In contrast, the single-stage MR-CART optimizes Stage 1 using the entire dataset without any data filtering.

4.14. Comparison with the multistage MR-PRIM

In this section, we compare the proposed multistage MR-CART with the multistage MR-PRIM proposed by Lee et al. (2020). As mentioned in Section 2.3, CART and PRIM are quite comparable because both find the optimal condition directly from the data without constructing specific model but are different in the way they narrow the input variable space. PRIM narrows the input variable space in a 'patient' way by iteratively removing small amounts of data. In contrast, CART iteratively conducts binary splits to improve the objective function value, which is a rather aggressive way of narrowing.

The data removal in PRIM is conducted until the portion of observations of the remaining dataset becomes less than a predetermined threshold. We set this predetermined threshold at $\beta = 5\%$, which is a typical value in PRIM (Freidman and Fisher, 1999). In addition, we applied a data removal rate as $\alpha = 5\%$, which is also a typical value in PRIM (Freidman and Fisher, 1999). As a result, the multistage MR-PRIM obtained a final box, which included 224 observations.

Table 9 reports the criteria of the input variables obtained from the proposed method and the multistage MR-PRIM. Blanks in the table mean that no splitting or data removal occurred during the optimization. When comparing the number of input variables that participated in obtaining the final subregion (i.e., input variables that are not blank), CART has 6 input variables ($y_{1,1}$,$y_{1,2}$,$x_{2,2}$,$x_{2,7}$,$x_{2,8}$,$x_{2,9}$) for Stage 2, and 2

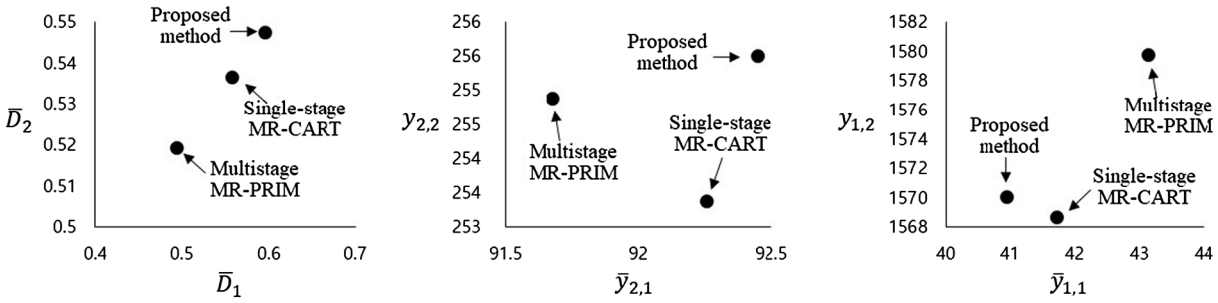

Fig. 9. Comparison of the results from the three methods.

input variables ($x_{1,3}$,$x_{1,10}$) for Stage 1, while PRIM has 10 input variables ($y_{1,1}$,$x_{2,2}$,$x_{2,3}$,$x_{2,5}$,$x_{2,6}$,$x_{2,7}$,$x_{2,8}$,$x_{2,9}$,$x_{2,10}$,$x_{2,11}$) for Stage 2, and 11 input variables ($x_{1,1}$,$x_{1,2}$,$x_{1,3}$,$x_{1,5}$,$x_{1,6}$,$x_{1,7}$,$x_{1,8}$,$x_{1,9}$,$x_{1,11}$,$x_{1,12}$,$x_{1,13}$) for Stage 1. This means that fewer input variables are required to obtain the final subregion in CART. We believe that this occurred because the two methods have different mechanisms for narrowing the input variable space. PRIM narrows the input variable space in a 'patient' way, while CART does it rather aggressively. In PRIM, only a small portion of observations are removed for each iteration, thus, many iterations are conducted by generating a long sequence of boxes. During the many iterations, input variables have a high chance to participate in data removal. In contrast, CART splits data in aggressive way, and only key variables participate in reducing the input variable space.

Another important difference between the proposed multistage MR-CART and PRIM is that the proposed method can generate several subregions while PRIM generate only a single subregion. As shown in Fig. 7, we obtained 24 leaf nodes (i.e., subregions) by the proposed method. Among them, we considered the top 5 leaf nodes having largest objective function values. We conducted the cross validation for the 5 leaf nodes and selected the final subregion. In contrast, PRIM obtain a single subregion by gradually peeling off the initial input variable space.

The third column of in Table 8 for each stage shows the optimization results of the multistage MR-PRIM. The proposed method shows higher $\overline{D}_1$and $\overline{D}_2$values compared to those of the multistage MR-PRIM. When comparing $\overline{D}_1$and $\overline{D}_2$ values for both methods, the number of observations should be considered. This is because $\overline{D}_1$and $\overline{D}_2$ improved when the number of observations became small by applying a small min value, as shown in Section 4.2. This implies that the proposed method outperforms the multistage MR-PRIM in that it shows higher $\overline{D}_1$and $\overline{D}_2$ values with more observations.

Fig. 9 shows the optimization results of the three methods given in Table 8. In terms of $\overline{D}_1$and $\overline{D}_2$, the proposed method outperforms the single-stage MR-CART and the multistage MR-PRIM. Also, the proposed method outperforms the two methods in terms of $\overline{y}_{2,1}$and $\overline{y}_{2,2}$. In terms of $\overline{y}_{1,1}$and $\overline{y}_{1,2}$, the proposed method outperforms the single-stage MR-CART but does not outperform the multistage MR-PRIM. More specifically, the proposed method shows better $\overline{y}_{1,1}$ while the multistage MR-PRIM shows better $\overline{y}_{1,2}$. This implies that the proposed method sacrificed $\overline{y}_{1,2}$ and improved $\overline{y}_{1,1}$ to improve $\overline{D}_1$.

5. Concluding remarks

In this paper, we proposed a systematic procedure for optimizing the multiple responses in the multistage manufacturing process using CART. In the multistage process, the performance of each stage needs to be considered in the context of the relationship between the stages since each stage is influenced by its preceding stage, and it also affects the stage that follows. We consider this property by modifying the CART algorithm and employing the desirability function method, which optimizes multiple responses simultaneously. We illustrated the proposed multistage MR-CART using a steel making example and compared it with existing methods.

As a future research topic, the proposed method can be extended to optimize the multiple responses during process changes. The proposed method assumes that observations are collected from a stable manufacturing process. However, manufacturing processes often experiences small shifts due to various assignable causes such as machine deterioration over time. In such a case, recent observations are more informative than past observations to capture current status of the process. For this purpose, time-weighted methods can be employed in the proposed method. In particular, the concept of a cumulative sum control chart (which is a useful control chart in process monitoring under the small process shifts) can be employed.

Secondly, the mean and variance of multiple responses can be considered, not just the mean of multiple responses. The proposed method only considers the mean of multiple responses. However, the variance of the multiple responses is also important to maintain stable variability. In this regard, loss function approaches (Pignatiello, 1993; Vining, 1998), which can consider the variance of multiple responses can also be an alternative. The loss function approaches optimize the multiple responses by estimating the variance–covariance matrix for multiple responses and using cost matrix determined by process economics (Ko, Kim, & Jun, 2005; Wang, Ma, Ouyang, & Tu, 2016). In the proposed method, the objective function can be modified based on the loss function for simultaneous optimization of the mean and variance.

Third, the nonlinear desirability function can be considered. The proposed method assumes the linear individual desirability functions. However, nonlinear functions could be more realistic for each desirability function. For this purpose, s_j, t_j can be adjusted in Equations (1), (2), and (3). By determining the shape of d_j, it can model more flexibly than the proposed method. Besides, the other overall desirability function also can be considered. The proposed method used geometric mean as overall desirability functions. However, if it is difficult to use the geometric mean that results in unacceptable overall responses due to one unacceptable response, it can be substituted for other forms such as additive, multiplicative, and maximin functions and weighted geometric mean.

CRediT authorship contribution statement

Dong-Hee Lee: Conceptualization, Methodology, Writing – original draft. **So-Hee Kim:** Software. **Kwang-Jae Kim:** Supervision.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07049412). Also, this work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2019R1A2C1007834).

References

Abu-Hanna, A., Nannings, B., Dougelmans, D., & Hasman, A. (2010). PRIM versus CART in subgroup discovery: When patience is harmful. *Journal of biomedical informatics, 43*(5), 701–708.

Bera, S., & Mukherjee, I. (2016). A multistage and multiple response optimization approach for serial manufacturing system. *European Journal of Operational Research, 248*(2), 444–452.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Belmont, CA: Chapman and Hall/CRC.

Chapman, J. L., Lu, L., & Anderson-Cook, C. M. (2014). Process Optimization for Multiple Responses Utilizing the Pareto Front Approach. *Quality Engineering, 26*(3), 253–268.

Chong, I. G., Albin, S. L., & Jun, C. H. (2007). A data mining approach to process optimization without an explicit quality function. *IIE Transactions, 39*, 795–804.

Derringer, G. (1994). A balancing act: optimizing a product's properties. *Quality Progress, 27*(6), 51–58.

Derringer, G., & Suich, R. (1980). Simultaneous optimization of several response variables. *Journal of Quality Technology, 12*(4), 214–219.

Emrouznejad, A., & Anouze, A. L. (2010). Data envelopment analysis with classification and regression tree – a case of banking efficiency. *Expert Systems, 27*(4), 231–246.

Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing, 9*(2), 123–143.

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta, 185*, 1–17.

Haughton, D., & Oulabi, S. (1997). Direct marketing modeling with CART and CHAID. *Journal of Direct Marketing, 11*(4), 42–52.

Hejazi, T. H., Seyyed-Esfahani, M., & Mahootchi, M. (2015). Optimization of degree of conformance in multiresponse-multistage systems with a simulation-based metaheuristic. *Quality and Reliability Engineering International, 31*(4), 645–658. https://doi.org/10.1002/qre.1622

Ko, Y. H., Kim, K. J., & Jun, C. H. (2005). A new loss function-based method for multiresponse optimization. *Journal of Quality Technology, 37*(1), 50–59.

Kwak, D. S., Kim, K. J., & Lee, M. S. (2010). Multistage PRIM: Patient rule induction method for optimisation of a multistage manufacturing process. *International Journal of Production Research, 48*(12), 3461–3473.

Lee, M. S., & Kim, K. J. (2008). MR-PRIM: Patient rule induction method for multiresponse optimization. *Quality Engineering, 20*(2), 232–242.

Lee, J. C., Jeong, S. W., & Lee, W. Y. (2016). A prediction model of timely processing on medical service using classification and regression tree. *Journal of IKEEE, 20*(10), 16–25.

Lee, D. H., Yang, J. K., & Kim, K. J. (2018). Dual-response optimization using a patient rule induction method. *Quality Engineering, 30*(4), 610–620.

Lee, D. H., Yang, J. K., Kim, S. H., & Kim, K. J. (2020). Optimizing mean and variance of multiresponse in a multistage manufacturing process using operational data. *Quality Engineering, 32*(4), 627–642.

Lee DH, Kim SH, Kim ES, Kim KJ, Zhen H (2021) MR-CART: Multiresponse Optimization Using a Classification and Regression Tree Method. Quality Engineering, accepted for publication.

Liu, Z., Kanter, C. A., Messer, K. D., & Kaiser, H. M. (2013). Identifying significant characteristics of organic milk consumers: A CART analysis of an artefactual field experiment. *Applied Economics, 45*(21), 3110–3121.

Mukherjee, I., & Ray, P. K. (2008). Optimal process design of two-stage multiple responses grinding processes using desirability functions and metaheuristic technique. *Applied Soft Computing, 8*(1), 402–421.

Pan, J. N., Li, C. I., & Wu, J. J. (2016). A new approach to detecting the process changes for multistage systems. *Expert Systems with Applications, 62*, 293–301.

Pignatiello, J. J., Jr (1993). Strategies for robust multiresponse quality engineering. *IIE transactions, 25*(3), 5–15.

Steinberg, D. (2009). CART—classification and regression trees. In X. Wu (Ed.), *The Top Ten Algorithms in Data Mining* (pp. 179–201). FL: Chapman and Hall/CRC.

Tang, T. I., Zheng, G., Huang, Y., & Wang, P. (2005). A comparative study of medical data classification methods based on decision tree and system reconstruction analysis. *Industrial Engineering and Management Systems, 4*(1), 102–108.

Vining, G. G. (1998). A compromise approach to multiresponse optimization. *Journal of Quality Technology, 30*(4), 309–313.

Wang, J., Ma, Y., Ouyang, L., & Tu, Y. (2016). A new Bayesian approach to multi-response surface optimization integrating loss function with posterior probability. *European Journal of Operational Research, 249*(1), 231–237. https://doi.org/10.1016/j.ejor.2015.08.033

Yang, J. K., & Lee, D. H. (2018). Optimization of mean and standard deviation of multiple responses using patient rule induction method. *International Journal of Data Warehousing and Mining, 14*(1), 60–74.

Yin, X., He, Z., Niu, Z., & Li, Z. S. (2018). A hybrid intelligent optimization approach to improving quality for serial multistage and multi-response coal preparation production systems. *Journal of manufacturing systems, 47*, 199–216.