Engineering Applications of Artificial Intelligence 97 (2021) 104089

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Causal artificial neural network and its applications in engineering design

Di Wu *, G. Gary Wang

Product Design and Optimization Laboratory, Mechatronic System Engineering Department, Simon Fraser University, Surrey, BC V3T 0A3, Canada

ARTICLE INFO

Keywords:
Artificial Neural Network
Causal graph
Bayesian Network
Engineering design

ABSTRACT

To reduce the computational cost in engineering design, expensive high-fidelity simulation models are approximated by mathematical models, named as metamodels. Typical metamodeling methods assume that expensive simulation models are black-box functions. In this paper, in order to improve the accuracy of metamodels and reduce the cost of building metamodels, knowledge about engineering design problems is employed to help develop a novel metamodel, named as causal artificial neural network (causal-ANN). Cause–effect relations intrinsic to the design problem are employed to decompose an ANN into sub-networks and values of intermediate variables are utilized to train these sub-networks. By involving knowledge of the design problem, the accuracy of causal-ANN is higher than the traditional metamodeling methods that assume black-box functions. Additionally, one can identify attractive subspaces from the causal-ANN by leveraging the structure of the causal-ANN and the theory of Bayesian Networks. The impacts of fidelity of causal graphs and design variable correlations are also discussed in the paper. The engineering case studies demonstrate that the causal-ANN can be accurately constructed with a small number of expensive simulations, and attractive design subspaces can be identified directly from the causal-ANN.

1. Introduction

More and more high-fidelity expensive simulation models, such as Computational Fluid Dynamics (CFD) and Finite Element Analysis (FEA) are applied in engineering design to improve the accuracy of the simulation. The high computational cost of simulation varies depending on the complexity of the model and computational power, but in general simulation models are much more expensive than mathematical models. To balance computational efficiency and modeling accuracy in engineering design, metamodels are developed to approximate the time-consuming simulation models. Wang and Shan (2007) listed popular models, such as kriging models (Joseph et al., 2008), radial basis functions (RBF) models (Fang and Horstemeyer, 2006), response surface models (RSM) (Hill and Hunter, 1966), and support vector machines (SVM) (Collobert and Bengio, 2001). Recently, more variables tend to be considered in design processes to improve flexibility, which means the new design can be adapted to different tasks. With the increase of dimensionality, the accuracy of metamodels will decrease even if increasing the number of samples, where accuracy is taken as the ability of the metamodel to provide predicted responses close to the actual values, measured by the R^2 value as defined here by Eq. (5). As a result, High-Dimensional Model Representation (HDMR) (Sobol', 1990) is applied to approximate high-dimensional simulation models.

To simplify the process of metamodel construction and develop general methodologies, simulation is usually treated as a black-box function when constructing a metamodel, which means only the input and output data are considered in the process. An issue of treating simulation models as black-box functions is that users are assumed not to know about information such as functional form, importance of variables, (non)linearity of the objective/constraint function with respect to each variable, and variable correlations. The shortcoming of the black-box assumption is that more computational cost is involved since the properties of the problem are unknown. A totally unknown design space implies that more sample points are needed to establish enough information to construct an accurate metamodel in the design space. In real-world engineering design, however, practitioners usually have some knowledge of the problem such as the variables involved, input–output relations, values of some intermediate variables, and so on. Such information is not commonly used in approximating simulation-based engineering models. By employing appropriate information in the right situations, the efficiency of the engineering design process can be improved. The potential applications of knowledge for engineering optimizations are discussed in Wu and Wang (2020), where multiple research directions are identified in knowledge assisted optimization (KAO). In Wu et al. (2019), knowledge, specifically cause–effect relations, is employed to reduce the dimensionality and decompose high-dimensional engineering design problems. In this work, a method of incorporating knowledge in metamodeling is developed to increase the computational efficiency by reducing the number of samples used to capture the behavior of the design functions.

* Corresponding author.
E-mail addresses: dwa88@sfu.ca (D. Wu), gary_wang@sfu.ca (G.G. Wang).

https://doi.org/10.1016/j.engappai.2020.104089
Received 5 August 2019; Received in revised form 8 October 2020; Accepted 6 November 2020
Available online 24 November 2020
0952-1976/

How to systematically incorporate knowledge and information of the design problem into metamodeling for engineering design? In this paper, cause–effect relations combining with artificial neural networks (ANN) are employed to help in metamodeling. ANNs are metamodeling methods for nonlinear problems (Bishop, 2006). The number of nodes, layers, and training methods can affect the accuracy of an ANN model. Thus, how to determine the topology of a neural network is one of the issues for ANN approximation. On the other hand, a network graph with nodes and links, namely a causal graph, is usually used to represent the cause–effect relations between variables and objectives. Thus, with a similar network structure, the causal graph is employed as a guide in generating the ANN structure in this work. Additionally, even in a black-box function model, actual values of some intermediate variables can be obtained by simulation. After employing the causal graph to determine the structure of an ANN, values of intermediate variables are used to improve the approximation accuracy of ANN. Thus, the designer is assumed to know the design problem just enough to be able to derive high-level cause–effect relationships. With this premise, a causal-ANN is defined and developed that involves the cause–effect knowledge and values of intermediate variables to improve the accuracy of metamodels, especially for high dimensional engineering design problems.

Apart from predicting responses close to their actual values, the causal-ANN can be used in different applications in engineering design. Considering the structural representation of a causal-ANN, not only the objective values, but also values of intermediate variables can be predicted from the causal-ANN. Therefore, combining with the theory of Bayesian Networks (Ben-Gal et al., 2007), distribution of the variables and objectives can be estimated through the causal-ANN. By analyzing the distribution, attractive design subspaces can be identified, which means the subspace where the optimum solution may locate. In this paper, the application of causal-ANN in identifying attractive design spaces is also developed.

The paper is organized as follows. ANN architecture and the definition of Bayesian network and causal graph are briefly reviewed in Section 2. The proposed causal-ANN and its application in attractive sub-space identification are introduced in Section 3. To illustrate the performance of the proposed method, two case studies are tested in Section 4. Section 5 is the conclusion.

2. Related works

2.1. Artificial neural network architecture

ANNs have been widely used in different fields for real-world problem approximation and prediction (Ebrahimi et al., 2016; Fonseca et al., 2003; Hippert et al., 2001), and feedforward ANN is one of the most popular types. Building a proper ANN, however, is still a nontrivial task due to difficulties of determining the architecture of the networks, which affects the prediction accuracy (Zhang et al., 1998). The architecture of an ANN includes the number of hidden layers, number of hidden nodes, and connections between nodes. An improper architecture of ANN may lead to overfitting, which will significantly reduce the accuracy of the metamodel. In general, the number of layers and number of hidden nodes are determined based on experience. An ANN with two hidden layers usually provides more benefits for different types of nonlinear problems compared with the network with one hidden layer (Cheng and Titterington, 1994). On the other hand, different guidelines for the number of hidden nodes have been developed, including "$2n + 1$" (Lippmann, 1987), "$2n$" (Wong, 1991), "$n/2$" (Kang, 1992), and so on, where n is the number of input nodes, but none of them outperforms the others when considering all kinds of problems. A fully connected layer structure, i.e., all nodes are connected with each other, is usually used in ANN.

The main issue of determining the ANN architecture by experience is that the guidelines may not perform well in every situation. Research has been conducted to develop a more intelligent architecture determination method for different approximation tasks. Akaike's Information Criterion (AIC) was used to determine the number of hidden nodes in ANN (Kang, 1992), where statistic properties of the training set were considered to generate the network structure. Another architecture determination method is based on the performance of the network. Different structures are tested and the most accurate selected as the structure. Srivastava et al. (2014) used the dropout method to find the appropriate structure of ANN to avoid overfitting. Nodes in the ANN are randomly dropped out during the training process to find the most accurate structure. Optimization is also employed to search for the structure with the highest accuracy. A layer-wise structure-learning method based on multi-objective optimization is developed to construct a deep neural network (Liu et al., 2018). By employing the structure-learning method, the network is no longer fully connected, and some of the connections will be deleted based on approximation accuracy. Moreover, some researchers focus on breaking the layer-wise structure of the neural network, which means there exists links connecting nodes not in adjacent layers. For instance, genetic evolution methods are employed to find out the optimum topology of the network (Maniezzo, 1994). In these aforementioned methods, the architecture of the network is determined purely from data. To find a more accurate structure of ANN, a large amount of computation is usually required.

Another structure determination method is based on knowledge and a knowledge based neural network was developed for microwave design problems (Wang and Zhang, 1997). The existing knowledge, such as empirical formulations, is involved to construct the knowledge layer in the network. In Morris et al. (2017), the intermediate variables in a Bayesian network are used as the hidden nodes to construct an ANN. However, the Bayesian network can only represent the input–output relations between variables, and mathematical relations cannot be captured from the Bayesian network. Therefore, in this paper, the Bayesian network is employed to guide the modeling of the structure of the causal-ANN rather than using the Bayesian network directly. Also, mathematical relations involved in the Bayesian network and causal-ANN will be used to construct a more accurate metamodel by considering the values of intermediate variables in Bayesian network.

2.2. Bayesian network and causal graph

Bayesian networks (BNs), also known as belief networks, belong to the family of probabilistic graphical models (GMs) (Ben-Gal et al., 2007). These graphical structures can be used to represent knowledge about an uncertain domain. BNs can also be regarded as a directed acyclic graph (DAG), which means that there is no circle or loop in the graph (Pearl, 2014). A more formal definition of a BN is given as follows: a Bayesian network is an annotated acyclic graph that represents a joint probability distribution over a set of random variables (Friedman et al., 1997). Hence, in a BN, there are two main members, the variables and the conditional probability distribution of each variable.

A causal graph is one of the variances of BN, representing the cause–effect relations embedded in human thinking. Compared to the original BN, the edges in a causal graph contain directions, which express the judgment that certain events or actions will lead to particular outcomes. Causal graphs have been used in the decision-making field to represent relationships between different factors. Besides, the causal graph is also a useful tool in representing structures of engineering systems. Based on causal graphs, the Dimensional Analysis Concept Modeling (DACM) framework was developed to gather and organize information associated with an engineering problem during the concept design phase (Coatanea et al., 2016).

Constructing causal graphs often depends on expert knowledge, which may lead to errors in some parts of the graph. The proposed approach does not demand a detailed causal graph in order to alleviate or at least minimize such risk. In addition, Bayesian networks can represent the conditional and joint probability distributions of the variables in the network. Those distributions can give useful information about the objective function and interesting design subspaces.

3. Causal ANN and application in attractive sub-space identification

In this section, the utilization of cause–effect relations to help neural network construction is presented. According to the causal graph, the entire network is divided into multiple sub-networks. Intermediate variables are used together with the design variables and objective to train each sub-network. The constructed causal-ANN can be used to identify the attractive sub-spaces where the optimum design may locate. The likelihood of design variables can be estimated from the causal-ANN and the attractive sub-spaces selected through the likelihood distribution. In this section, the process of constructing a causal-ANN is described and its application in identifying attractive sub-spaces is represented. Case studies are given in Section 4 for more detailed description of each step.

3.1. Causal artificial neural network

The main purpose of the causal-ANN is to overcome the high computational cost challenge of ANNs caused by the unknown structure of the network. We note that engineers typically have certain understanding of the problem at hand, and furthermore, during engineering simulation, values of some intermediate variables can be obtained through one simulation along with the objective value. But those values are not employed in constructing metamodels. Causal relations are used to form the structure of the ANN and values of the intermediate variables are used in training the ANN. The process of constructing the ANN is as follows.

Step 1. Generate the causal relations of the design problem. A high-level causal relations map (i.e., simplified causal graph) is needed before constructing a causal-ANN. The simplified causal graph needs inputs, output, and key intermediate variables. Such intermediate variables can be the coupling variables, the outputs from each sub-system, or variables whose values can be obtained from simulations as by-products. Usually, key intermediate variables can be selected according to the problem simulation process and experience of the designers.

There are two ways to generate a high-level causal graph. One method is to simplify an existing causal graph. A causal graph of an engineering problem usually contains all variables involved in the problem, however not all of the variables in the complete causal graph are important to the design. By keeping the key variables and removing others, a complete causal graph can be simplified into a high-level causal graph. If the causal graph does not exist, knowledge of the design problem, such as flow charts, can be used to generate the high-level causal relations. By connecting inputs and output with the selected key intermediate variables, a high-level causal graph can be generated. Case studies in Section 4 will show examples and some guidelines of high-level causal relation map construction will be discussed.

Step 2. Generate sub-networks according to the causal relations. The high-level causal graph is divided into multiple sub-graphs, which include only two layers, inputs and outputs. For example, for the causal graph in Fig. 1, three sub-graphs can be generated, [A, B] to D, [A, B, C] to E and [D, E] to F, where A, B, and C are design variables, D and E are value-known intermediate variables, and F is the objective.

Step 3. Construct neural network according to the sub-graphs and values of the intermediate variables. In the paper, an ANN with two hidden layers is employed to generate the metamodel. Since the intermediate values are defined as the by-product variables, the values of the intermediate variables and the objective can be obtained through one simulation. The sampling data of design variables, intermediate variables and objective are used to train each sub-network respectively. If the sub-ANN is between intermediate variables and the objective (e.g., [D, E] to F in Fig. 1), the actual values of the intermediate variables D and E are used as the inputs of the ANN. Note that when using the system causal-ANN in prediction after the model is completely

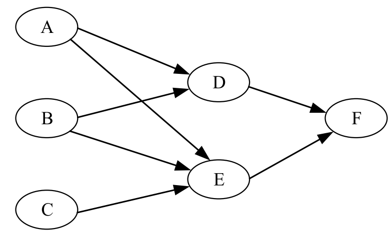

Fig. 1. An example of high-level causal graph.

constructed, values of the intermediate variables are estimated from previous sub-networks.

The purpose of the causal-ANN constructing method is to employ knowledge in constructing more accurate metamodels using a limited number of samples. It is often difficult to train an ANN to approximate large-scale nonlinear problems. In causal-ANN, the existing knowledge, i.e., cause–effect relations, is used to decompose the complex ANN network into small sub-networks. Then, the sub-ANNs are trained based on the values of intermediate variables. By achieving high accuracy of each sub-network, the accuracy of the entire metamodel can be improved. The main advantage of causal-ANN is thus the reduction of the complexity of ANN. It is also important to note that there is no extra simulation involved in constructing a causal-ANN. Compared to the time spent in simulation, the cost of constructing a high-level causal graph is thus negligible.

The key step in causal-ANN is the generation of the high-level causal graph. As aforementioned, a high-level causal graph can be generated from simplifying an existing causal graph or from known input–output relations. Since it is difficult to construct a complete and correct causal graph for a complex system, this work only requires a high-level causal graph with key variables. The cause–effect links with doubts can be removed from the high-level causal graph. If the removed links are not important, the causal-ANN can remain high accuracy. On the other hand, if the removed links are important, it can be detected by checking the accuracy of the sub-networks. The high-level causal graph construction and failure tolerance of the causal graph are discussed in Section 4.3 with the case studies.

3.2. Attractive sub-space identification method

Once an accurate causal-ANN is constructed, it can be used in multiple scenarios. In this section, an attractive sub-space identification method using causal-ANN is proposed, where the attractive sub-space is defined as the small space where the best design may be located. Because the attractive sub-space is always a smaller design space as compared to the original space, the effort used to search the attractive sub-space is much smaller than searching the original space. In the Mode-Pursuing Sample (MPS) method (Wang et al., 2004), a large number of samples are generated by evaluating the metamodel and an attractive sub-space is estimated through those cheap samples and their responses. In this work, causal-ANN combined with Bayesian theorem is applied in attractive sub-space identification. By generating sample points from the causal-ANN, Bayesian probability inference can be performed with lower computational cost than on the actual simulation.

Bayesian networks (BNs) are one kind of belief graphic modeling method that gives the joint distribution of each variable. By constructing a Bayesian network of the engineering design problem, the distribution of the objective $p(f|\boldsymbol{x}, D, G)$ can be found, where f is the objective, $\boldsymbol{x}$ is the design variable vector, D is the data and G is the graph structure. After obtaining the distribution of the objective, the likelihood of the objective $p(\boldsymbol{x}|f, D, G)$ can be calculated via the Bayesian theorem as shown in the following equation

$$p(\boldsymbol{x}|f, D, G) = \frac{p(f|\boldsymbol{x}, D, G)p(\boldsymbol{x})}{p(f)} \tag{1}$$

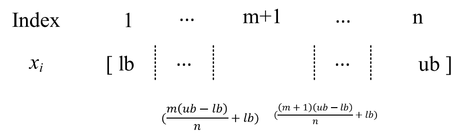

Fig. 2. Variable discretization.

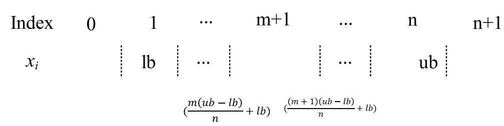

Fig. 3. Discretization for the variable without fixed bounds.

where, $p(\boldsymbol{x})$ and $p(f)$ are the distribution of the design variables and the objective respectively, which can be estimated through analyzing the sample data. In general, for an individual engineering design problem, designers or the decision makers often have an expected objective value or range. The likelihood of the objective gives the information about what area (or range) of the design variables has a higher probability to generate expected designs. Details of the method are shown as follows.

Step 1. Generate sample points. Sample points are generated following the uniform distribution (or other variable distributions if known). The causal-ANN model is evaluated to calculate the responses of the sample points. Note that the responses include the objectives and also the intermediate variables.

Step 2. Discretize all the variables and objective. Most of BNs only deal with discrete variables, while the variables in design problems are usually continuous. One method to deal with the problem is to discretize variables and the objective.

At the beginning, all the variables including inputs, intermediate variables and outputs are assumed to follow the uniform distribution. Then, the range of each variable is divided into n intervals with certain indices, as shown in Fig. 2.

lb and ub are lower and upper bounds of the variables, respectively. If the sample falls between $(\frac{m(ub-lb)}{n}+lb)$ and $(\frac{(m+1)(ub-lb)}{n}+lb)$, $m = 0,\ldots,n$, the index of the sample is $m+1$. Note that when the variable does not have a fixed lower bound or upper bound, a rough bound can be determined and then two additional sections, which are smaller than the lower bound and larger than the upper bound are added, as shown in Fig. 3.

Step 3. Calculate the joint probability of the objective, $p(f|\boldsymbol{x},D,G)$. The approximate inferencing method is employed to generate the conditional distribution of the variables, $p(xi|Pxi,D,G)$, where, xi is the intermediate variables, Pxi is the parents of xi. The conditional distribution can be calculated as follows

$$p\left(xi=a|Pxi=b,D,G\right)=\frac{N_{xi=a,Pxi=b}}{N_{Pxi=b}} \tag{2}$$

where, $N_{Pxi=b}$ is the number of samples that $Pxi=b$, and $N_{xi=a,Pxi=b}$ is the number of samples that $xi=a$ as well $Pxi=b$. Because the design variables are generated following the uniform distribution, the prior probability of the design variable can be calculated as $p\left(x=a\right)=1/n$. Then, the joint probability of objective can be calculated as follows

$$\begin{aligned} &p\left(f=a|x,D,G\right)\\ &=\sum_{i_1=1}^{n_1}\cdots\sum_{i_k=1}^{n_k}\sum_{i_x=1}^{n_x}\left(p\left(f=a|Pxi_1\right)p\left(Pxi_1|Pxi_2\right)\cdots p\left(Pxi_k|x\right)p\left(x\right)\right)\\ &=\sum_{i_1=1}^{n_1}(p(f=a|Pxi_1)\cdots\sum_{i_k=1}^{n_k}p\left(Pxi_k|x\right)\sum_{i_x=1}^{n_x}p\left(x\right)) \end{aligned} \tag{3}$$

where, n_k gives the discrete number of each parent variable (i.e., intermediate variables), and n_x represents the number of discrete sections of design variables. By counting the data and analyzing the Bayesian network, the joint probability of objective can be estimated.

Step 4. Estimate the likelihood of the design variables and find the interesting area of each variable. The likelihood is estimated according to the Bayesian theorem. $p(f)$ is estimated through the function, $p\left(f=a\right)=\frac{N_{f=a}}{N}$, where N is the number of samples and $N_{f=a}$ is the number of samples where the objective value falling in the section a. Finally, the likelihood of the design variable is estimated via (3). The interval with the largest likelihood of the design variables is selected as the interesting area.

Note that when there are multiple parents for one variable, the correlations of those parents should be considered. However, to estimate the joint distribution considering the correlations of multiple parents, a huge amount of samples are needed to cover all the possible combinations of the multiple parents. One of the methods is assuming the probability distribution when given each parent is independent. For example, if A and B are the parents of C, the distribution $p\left(C|A\right)$ and $p(C|B)$ are calculated independently. However, ignoring the correlations between parents may lead to wrong likelihood estimation when the correlations between parents are very strong. Therefore, a method named "Noisy-or" is employed to estimate the probability distribution. In Noisy-or method, the joint distribution given multiple parents can be calculated as follows.

$$P\left(f=a|x_1,x_2,\ldots,x_n\right)=1-\prod_{i=1}^{n}P(f\neq a|x_i) \tag{4}$$

In the Noisy-or method, the probability distribution considering correlation can be estimated by the probability distribution given each parent, which can reduce the number of samples significantly.

By comparing the likelihood of each interval, the interesting subspace can be determined. However, the number of samples used in likelihood estimation is usually very large. In this work, we use causal-ANN to generate the samples and thus the computational cost of identifying attractive sub-spaces is negligible.

4. Case studies

Two cases are employed to illustrate the performance of the proposed method, the power converter design problem and the aircraft design problem. First the Causal-ANN is constructed for both problems and then the attractive sub-space in each problem is identified. To introduce the method clearly, the power converter problem is presented step-by-step to represent the process of the proposed method.

4.1. Power converter design problem

A power converter design problem (Padula et al., 1996; Wang et al., 2007) is used to test the performance of the proposed dimension reduction methodology. The design problem has six design variables, as shown in Table 1.

Table 1
Design variables in power converter design.

Variables	Name	Description	Lower bound	Upper bound
x_1	C_w	Core center leg width (m)	0.001	0.1
x_2	$Turns$	Inductor turns	1.0	10
x_3	A_{cp}	Copper size (m^2)	7.29e^{-8}	1.0e^{-5}
x_4	$L_f/PINDUC$	Inductance (H)	1.0e^{-6}	1.0e^{-5}
x_5	C_f	Capacitance (F)	1.0e^{-5}	0.01
x_6	w_w	Core window width (m)	0.001	0.01

The upper and lower bounds defined in Wang et al. (2007) are used in this paper. The objective of the problem is to minimize the weight of the power converter. The formulation of the problem is defined as follows and all constant values are taken from Wang et al. (2007).

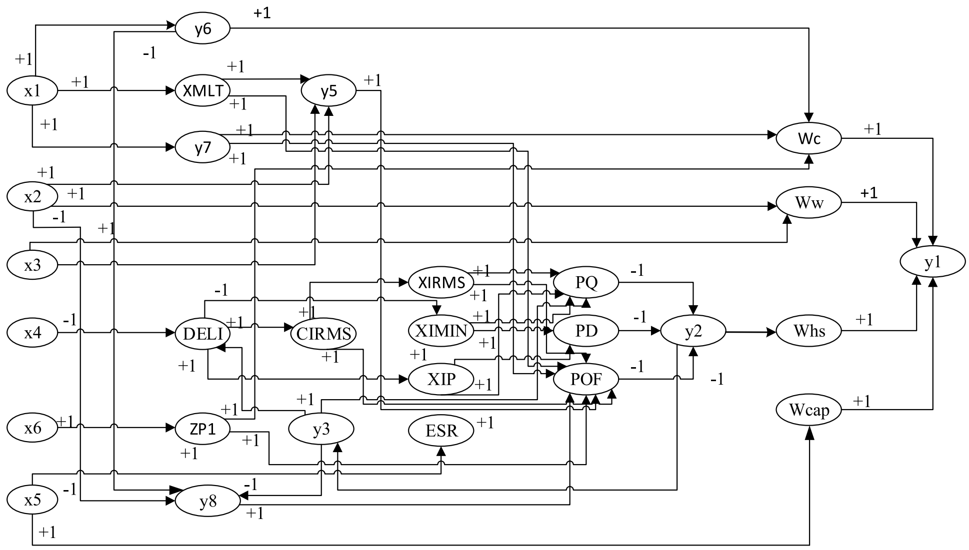

Fig. 4. Causal graph of the power converter problem.

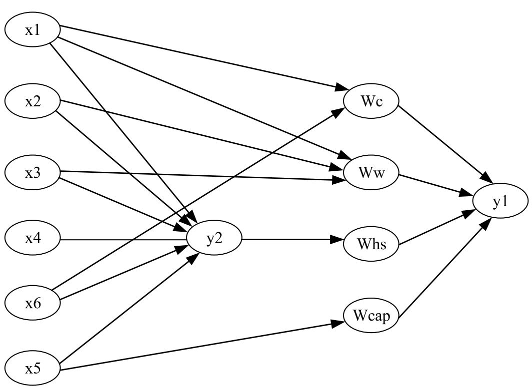

Fig. 5. High-level causal graph of the power converter design problem.

The problem is mainly dominated by the coupling between the circuit efficiency (y_2) and the duty circle (y_3). To show the cause–effect relations in the power converter design problem clearly, the complete causal graph is presented in Fig. 4. The guidelines of how to construct a high-level causal graph are discussed in Section 4.3. The six variables at the left side are the six design variables and the one at the right side is the objective. There are 21 intermediate variables involved in the problem, which are shown between the design variables and the objective in Fig. 4. The definitions of those variables can be found in Kott and Gabriele (1993). As shown in Fig. 4, y_2 is influenced by y_3 through different routes; while y_2 influences y_3 directly. All the design variables are involved in loops through different links and then finally influence the objective.

4.1.1. Constructing causal-ANN

The causal graph can be simplified to generate the high-level causal graph. Since the objective of the problem is to minimize the total mass of the power converter, the mass of the four components, i.e., W_c, W_w, W_{cap}, and W_{hs} can be outputs from the simulation. Additionally, the circuit efficiency y_2 as one coupled variable can also be an output from the simulation. Thus, the simplified causal graph can be formed in Fig. 5. Please note the simplified causal graph, or the high-level causal graph, does not have to be generated from the detailed causal graph, and can be created directly from a designer's knowledge, which we assume is a more likely situation.

Table 2
Comparison of R^2 values among three metamodels.

	Causal-ANN	ANN	RBF
R^2	0.967	0.691	0.732

According to the high-level causal graph, six sub-networks are divided as shown in Fig. 6. For each sub-network, an ANN network with two hidden layers and four hidden nodes in each layer is constructed. Note that the objective is a sum of the mass of each component. Thus, the sixth sub-graph is represented by the summation of each component and the other five graphs are used to construct ANN. For constructing the fourth ANN, the actual values of y_2 are used as the input of this ANN.

In this case, 200 sample points are generated for training. The Matlab neural network toolbox is employed to construct the ANN. In this work, the R^2 value (Torrie, 1960) shown in (5), which measures the extrapolation error on new test points (2000 in this work), is used to calculate the approximation accuracy of the causal-ANN.

$$R^2 = 1 - \frac{\sum_i (f_i - \hat{f}_i)^2}{\sum_i (\hat{f}_i - \bar{f})^2} \tag{5}$$

where f is the actual output value; $\hat{f}$ is the predicted value, and $\bar{f}$ is the average value of the actual output. R^2 value is always smaller than one. The closer the R^2 value is to one, the more accurate is the metamodel. Additionally, an RBF model and an ANN with two hidden layers between the design variables and the objective are constructed on the same 200 training samples. The R^2 value is also calculated with the same 2000 test points and compared in Table 2.

As shown in Table 2, the R^2 value of the causal-ANN is the highest among the three metamodeling methods, which means the causal-ANN is the most accurate metamodel. The lower value of R^2 of ANN and RBF is caused by the high non-linearity of the problem, especially for ANN. To further illustrate the performance of the causal-ANN, the R^2 value of each sub-network is shown in Table 3. It can be found that all sub-networks are accurate. Note that the third network is between all the six design variables and y_2, which has the same number of inputs as the entire design problem. The accuracy of this sub-network (0.994) is very high compared with the accuracy of the entire model (0.967). By dividing the entire network to sub-networks to reduce the complexity of each one, the accuracy of each sub-network can be improved.

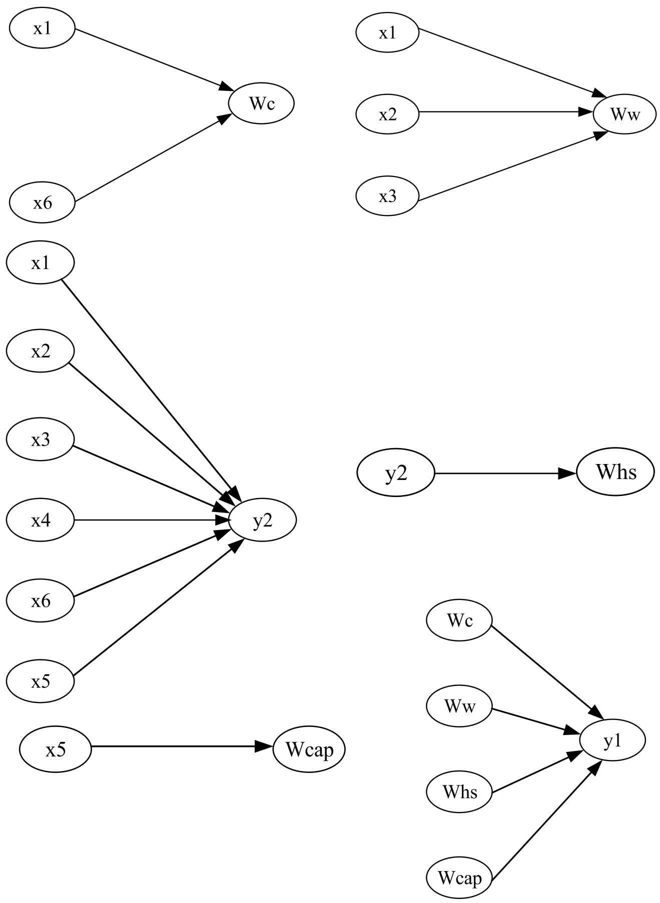

Fig. 6. Six sub-networks for the power converter design problem.

Table 3
R^2 value of each sub-network.

	y_2	W_c	W_w	W_{hs}	W_{cap}
R^2	0.994	1.000	0.974	0.934	1.000

4.1.2. Attractive sub-space identification

After constructing the causal-ANN, the probability distribution of the objective values and likelihood of the design variables can be estimated on the samples generated from the causal-ANN. At the beginning, the design variables, intermediate variables, and objective are discretized. For this case, the upper and lower bounds are used to determine the interval of design variables. While for the intermediate variables and objective, the minima and maxima are selected to determine the boundary of the intervals. Thus, all the variables and objective are divided into five intervals based on their own bounds.

The objective of the power converter problem is to minimize the mass, which means a smaller objective value is desired. Therefore, the first interval of the objective, i.e., $y = 1$, is selected and the conditional probability $P(y = 1|\boldsymbol{x})$ and likelihood $P(\boldsymbol{x}|y = 1)$ are estimated in this problem. Considering the correlations among the six design variables, the Noisy-or method is employed and the probability distribution of each design variable $P\left(y \neq 1|x_i\right), i = 1, 2, \ldots, 6$ is calculated. To estimate the probability distribution and the likelihood, 10,000 samples are generated from both the actual model via expensive simulation and the causal-ANN. The probability distributions estimated from the actual model and causal-ANN, $P\left(y \neq 1|x_i\right)$ and $P_{prediction}\left(y \neq 1|x_i\right)$, $i = 1, 2, \ldots, 6$, are shown in Tables 4 and 5, where $x_i = 1$ means the sample locates in the first interval of x_i.

As shown in Tables 4 & 5, the probability distributions estimated from the predicted model is the same as the distribution calculated from the actual model, which means that the causal-ANN is accurate estimating the predicted distribution. In some cases, the probability distribution is equal to zero, for example when $x_3 = 2$, which means that if the third coordinate of the sample is located in the second interval, all the objective values will be located in its first interval. This is caused by the distribution of the objective values. By setting the upper bound of the objective at the maximum value, over 95% objective values will locate at the first interval. The ill-defined boundary of the objective may render the likelihood estimation useless because the likelihood of some intervals may reach 100% according to Eq. (5). Therefore, the upper bound of the objective should be reduced to avoid 0% appearing in the probability distribution. In this case, 11 is selected as the upper bound according to the distribution of the objective values and then the objective is discretized into six intervals. The first interval of the objective is still the desired space. Then, the probability distributions estimated on the actual model and the causal-ANN is shown in Tables 6 and 7.

As shown in Tables 6 and 7, the probability distributions estimated from the predicted model are close to the values estimated from the actual model. Note that only 200 expensive simulation points are used to construct the causal-ANN, and probability estimation is performed on the causal-ANN, whose computation cost is negligible. No extra simulation is required to perform the probability estimation.

Table 4
Probability distribution $P\left(y \neq 1|x_i\right), i = 1, 2, \ldots, 6$ on actual model.

	x_1	x_2	x_3	x_4	x_5	x_6
$x_i = 1$	0	0.0005	0.038	0.004	0.007	0.007
$x_i = 2$	0.002	0.002	0	0.0095	0.009	0.006
$x_i = 3$	0.0055	0.0075	0	0.009	0.012	0.009
$x_i = 4$	0.0105	0.014	0	0.011	0.006	0.007
$x_i = 5$	0.02	0.014	0	0.0045	0.004	0.009

Table 5
Probability distribution $P_{prediction}\left(y \neq 1|x_i\right)$, $i = 1, 2, \ldots, 6$ on causal-ANN.

	x_1	x_2	x_3	x_4	x_5	x_6
$x_i = 1$	0	0.0005	0.038	0.004	0.007	0.007
$x_i = 2$	0.002	0.002	0	0.0095	0.009	0.006
$x_i = 3$	0.0055	0.0075	0	0.009	0.012	0.009
$x_i = 4$	0.0105	0.014	0	0.011	0.006	0.007
$x_i = 5$	0.02	0.014	0	0.0045	0.004	0.009

Table 6
Probability distribution $P\left(y \neq 1|x_i\right), i = 1, 2, \ldots, 6$ with new upper bound.

	x_1	x_2	x_3	x_4	x_5	x_6
$x_i = 1$	0.0505	0.5775	0.7915	0.6455	0.6290	0.6035
$x_i = 2$	0.2245	0.6075	0.6315	0.6470	0.6345	0.6160
$x_i = 3$	0.9270	0.6580	0.6215	0.6360	0.6295	0.6570
$x_i = 4$	1	0.6695	0.5780	0.6425	0.6455	0.6435
$x_i = 5$	1	0.6895	0.5795	0.6310	0.6635	0.6820

Table 7
Probability distribution $P_{prediction}\left(y \neq 1|x_i\right)$, $i = 1, 2, \ldots, 6$ with new upper bound.

	x_1	x_2	x_3	x_4	x_5	x_6
$x_i = 1$	0.0495	0.5820	0.7915	0.6445	0.6290	0.6035
$x_i = 2$	0.2215	0.6060	0.6295	0.6465	0.6355	0.6145
$x_i = 3$	0.9265	0.6540	0.6200	0.6340	0.6290	0.6530
$x_i = 4$	1	0.6670	0.5765	0.6420	0.6420	0.6435
$x_i = 5$	1	0.6885	0.5800	0.6305	0.6620	0.6830

Table 8
Interesting variable intervals with the largest likelihood.

	x_1	x_2	x_3	x_4	x_5	x_6
Actual model	1	1	5	5	1	1
Predicted model	1	1	5	5	1	1
Optimal solution	1	2	5	5	1	1

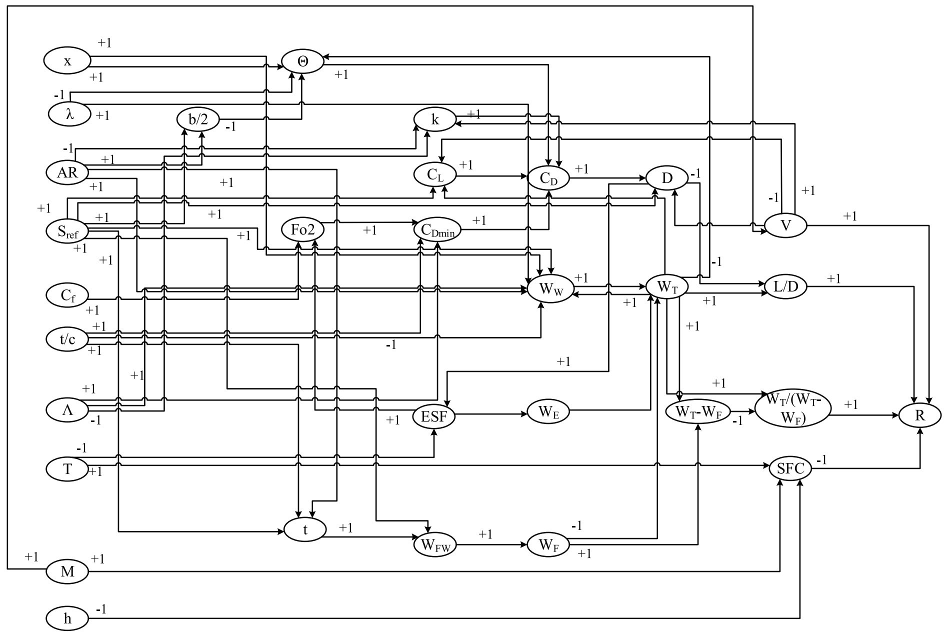

Fig. 7. Causal graph of the aircraft concept design problem.

Table 9
Design variables in aircraft concept design.

	Variables	Description	Lower bound	Upper bound
1	M	Mach number	1.4	1.8
2	T	Throttle setting	0.1	1.0
3	S_{REF}	Wing surface area (ft^2)	500	1500
4	AR	Aspect ratio	2.5	8.5
5	t/c	Thickness/chord ratio	0.01	0.09
6	λ	Wing taper ratio	0.1	0.4
7	Λ	Wing sweep (deg)	40	70
8	x	Wing box x-section area (ft^2)	0.9	1.25
9	C_f	Skin friction coefficient	0.75	1.25

Then, by employing the Noisy-or method and the Bayes theory, the interval of the design variable with the largest likelihood can be determined, which is shown in Table 8. In the table, the number of each design variable represents the interval of each variable. As with the above comparison, the likelihood is estimated on both the actual and predicted models. Additionally, the interval that the optimum is located in is also represented in the table. As shown in the table, the interesting interval generated from the prediction model is the same as the result from the actual model, which is almost the same as the interval where the actual optimum locates, except for x_2. This is because that the second design variable of the optimum point is located near the boundary of the first and the second interval and the likelihood distribution cannot capture it accurately.

4.2. Aircraft concept design problem

The aircraft concept design problem (Sobieszczanski -Sobieski et al., 2000) is used to test the performance of the proposed method. There are nine design variables (listed in Table 9 and three coupled disciplines (structure, aerodynamics and propulsion). The objective of the problem is to maximize the range computed by the Breguet equation (Sobieszczanski -Sobieski et al., 2000). The causal graph is shown in Fig. 7, which is given only for reference as it is not needed to construct a high-level causal graph to build the causal-ANN.

Two coupled variables, total weight of the aircraft, W_T and the drag D are selected as two intermediate variables. Also, the weight of the fuel (W_F) from the structural discipline and the specific fuel consumption (SFC) from the propulsion discipline are selected as the other two intermediate variables. Thus, the high-level causal graph is shown in Fig. 8.

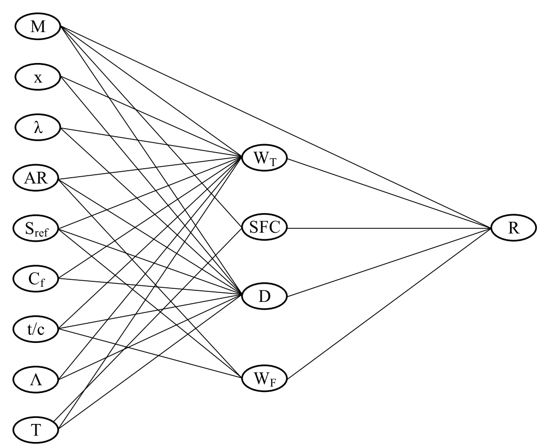

Fig. 8. High-level causal graph for aircraft concept design.

The high-level causal graph can be divided into five sub-networks as shown in Fig. 9. The ANN with two hidden layers and four hidden nodes in each layer is constructed based on each sub-network. Note that for the fifth network, the actual values of W_T, W_F, SFC, and D are used as the input of the network. 200 training samples are generated according to the simulation model. The Matlab toolbox is employed to train the neural network. Additional 2000 testing points are generated and the R^2 value is calculated on the testing points to measure the estimation error of the causal-ANN. Additionally, the R^2 values of an RBF model and an ANN on the entire problem are also calculated for comparison. The R^2 values of the three metamodels are shown in Table 10. As shown in Table 10, the R^2 value of the causal-ANN is the largest among the three metamodeling methods, which means the causal-ANN outperforms other two metamodeling methods on modeling accuracy. In this case, the ANN and RBF are also accurate due to their high R^2 values. Table 11 gives the R^2 value of each sub-network. The accuracy of the sub-network between all design variables and W_T is the lowest compared with other networks, which brings down the overall accuracy

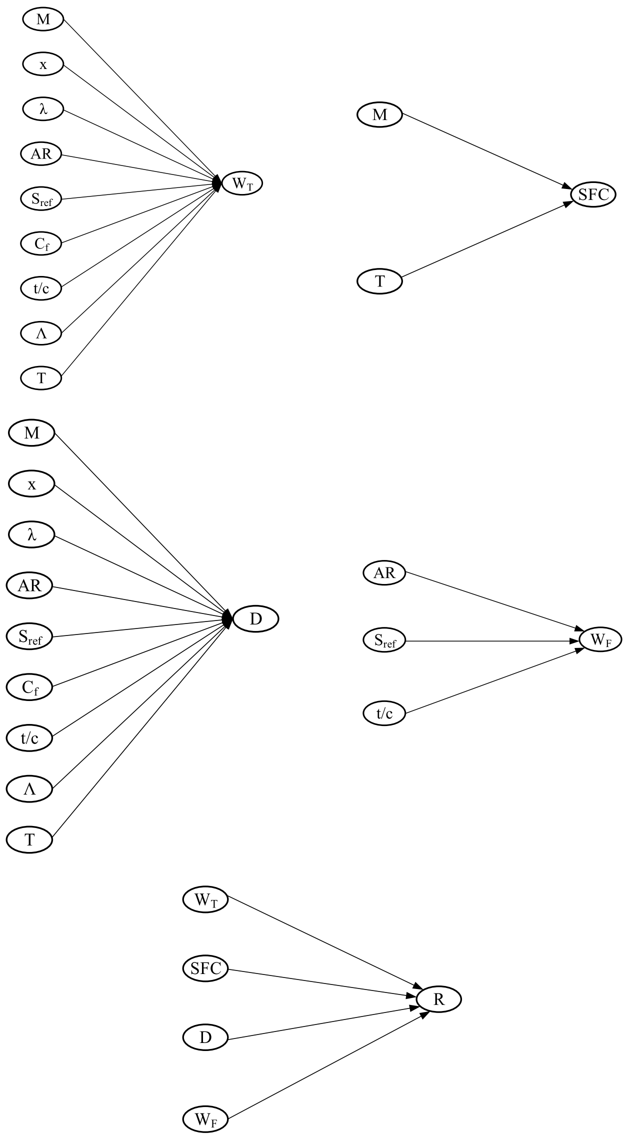

Fig. 9. Sub-networks for aircraft concept design.

Table 10
Comparison of R^2 value among three metamodels.

	Causal-ANN	ANN	RBF
R^2	0.968	0.943	0.940

Table 11
R^2 value of each sub-network.

	W_T	W_F	SFC	D
R^2	0.906	0.987	0.983	0.980

of the causal-ANN. The reason of the lower accuracy of the first sub-network is that coupling among the three disciplines is involved in this network, which increases the complexity of the sub-problem.

Once the causal-ANN is constructed, the likelihood is estimated based on the samples generated from the neural network. To illustrate the performance of the likelihood estimation on the neural network, 10,000 testing samples are generated on the actual model and the causal-ANN. The design variables, intermediate variables and objective are discretized into five intervals. As the objective is to maximize the range, the fifth interval of the objective is desired. To estimate the likelihood through the Noisy-or method, the probability distribution $P\left(y \neq 5|x_i\right), i = 1, 2, \ldots, 9$ is estimated on the actual model and the causal-ANN as shown in Tables 12 and 13. It can be found that the probability distribution estimated from causal-ANN is similar with the results from the actual model. Then, the likelihood is calculated via Bayes theory and the interval with the largest likelihood is represented in Table 14. The interval where the optimal solution locates in is also shown in the same table. It can be found that the interesting intervals generated from causal-ANN are the same as those obtained from the actual model. Additionally, these interesting variable intervals are exactly where the optimal solution locates.

4.3. Discussion

4.3.1. Generation of high-level causal graph

The causal relations are employed as the premier knowledge in constructing a causal-ANN. However, it is hard to generate an accurate and complete causal graph. In this paper, only a high-level causal graph including key intermediate variables is needed to represent the causal-effect relations in the design problem. As described in Section 3, finding the important intermediate variables is the key step in constructing a high-level causal graph. One of the criteria to select intermediate variables is if the variable value can be calculated or is an output from simulation. These intermediate variables can thus be called by-product variables. Basically, the coupling variables, outputs of each discipline, and by-product variables can be selected as key intermediate variables in the high-level causal graph. Another suggestion is to simplify the structure of an existing causal graph. Involving many variables in the causal graph may cause difficulty in constructing a causal-ANN. Thus, a causal graph with less than two intermediate layers is recommended. Additionally, for the problem with coupling loops, one variable in each coupling relation is selected to avoid coupling in the causal graph since the BN cannot deal with coupling well. Finally, the intermediate variables that have directly and prominent impact on the objective are usually selected as key variables.

In this paper, the complete causal graph exists for the two case study problems. Thus, the high-level causal relations can be generated through simplifying the causal graph. For the power converter problem, considering the objective is to minimize the total weight of the converter, the weight of each component can be selected as the key intermediate variables. Also, one of the coupling variables, circuit efficiency (y_2) is kept in the high-level causal graph. For the aircraft design problem, the total weight of the aircraft (W_T), drag (D), weight of the fuel (W_F), and specific fuel consumption (SFC), which directly influence the final objective, i.e., range, are picked as the key variables. Additionally, W_T and D are the coupled variables, while W_F and SFC are the outputs from the structure and propulsion disciplines. If a complete causal graph does not exist, high-level knowledge about the design problem can be utilized to generate the causal relations in the causal-ANN.

4.3.2. Fault tolerance studies on causal relations

Even though only high-level causal relations are required in constructing causal-ANNs, there might be errors in defining these causal relations, which may influence the accuracy of the causal-ANN. Thus, the impact of the faulty causal relations on the accuracy of the causal-ANN is discussed in this section.

First, the influences of the number of layers in the causal relations are discussed. Fig. 5 illustrates a high-level causal relation including two intermediate layers. As shown in Fig. 10, one intermediate variable, y_2, is removed from the causal graph to reduce the number of intermediate layers to one. Compared with the original causal-ANN, the sub-network, $\left[x_1, \ldots, x_6\right] - y_2 - W_{hs}$ is replaced by the direct links from the design variables to W_{hs}. Thus, the total number of sub-networks to be trained is four. The causal-ANN with one intermediate layer is trained with 200 samples and the R^2 values of the objective and different intermediated variables are calculated on 2000 testing

Table 12
Probability distribution $P\left(y \neq 1|x_i\right), i = 1, 2, \ldots, 9$ on real model.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
$x_i = 1$	1	0.9955	0.9950	1	1	0.9965	0.9990	0.9985	0.9975
$x_i = 2$	1	0.9980	0.9995	1	0.9990	1	0.9990	0.9985	0.9960
$x_i = 3$	0.9990	1	0.9990	1	0.9995	0.9990	0.9965	0.9980	1
$x_i = 4$	0.9970	1	1	0.9995	0.9985	0.9995	0.999	0.9985	1
$x_i = 5$	0.9975	1	1	0.9940	0.9965	0.9985	1	1	1

Table 13
Probability distribution $P_{prediction}\left(y \neq 1|x_i\right),\ i = 1, 2, \ldots, 6$ on causal-ANN.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
$x_i = 1$	1	0.9875	0.9830	1	0.9965	0.9905	0.9920	0.9940	0.9900
$x_i = 2$	1	0.9930	0.9945	1	0.9955	0.9975	0.9950	0.9955	0.9875
$x_i = 3$	0.9985	0.9965	0.9980	1	0.9985	0.9960	0.9940	0.9940	0.9975
$x_i = 4$	0.9880	0.9990	0.9990	0.9995	0.9950	0.9950	0.9980	0.9940	1
$x_i = 5$	0.9885	0.9990	1	0.9755	0.9895	0.9960	0.9960	0.9975	1

Table 14
Interesting variable intervals with the largest likelihood.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Actual model	5	1	1	5	5	1	1	1	2
Predicted model	5	1	1	5	5	1	1	1	2
Optimal solution	5	1	1	5	5	1	1	1	2

Table 15
R^2 value of objective and intermediate variables for the causal-ANN without y_2.

	y_1	W_c	W_w	W_{hs}	W_{cap}
R^2	0.886	1.000	0.969	0.805	1.000

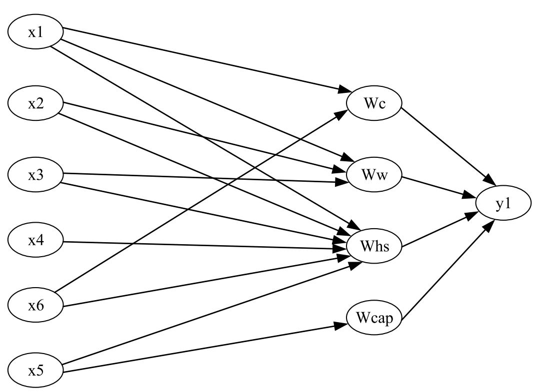

Fig. 10. Causal graph with one intermediate layer for power converter design.

samples and shown in Table 15. Note that the same training samples and test samples as in Section 4 are used in this test and the following test. Compared with the causal-ANN with y_2, the accuracy of the new causal-ANN decreases. Comparing the R^2 values of W_{hs} in Tables 3 and 15, it can be found that involving more intermediate variables in the complex networks can improve the prediction accuracy. On the other hand, compared with ANN and RBF model, the accuracy of the new causal-ANN is still better, which means a very simple high-level causal graph can still improve the accuracy of the prediction model.

Second, the influence of missing links is studied. Six causal graphs with one of the links from $\left[x_1, \ldots, x_6\right]$ to y_2 missing in each graph are created to construct six variations of the causal-ANNs and the accuracies of those causal-ANNs are calculated. The R^2 values of the objective and the intermediate variables, y_2 and W_{hs}, are listed in Table 16. It can be found that, missing the links will decrease the accuracy of the causal-ANN model. If any of the links from $\left[x_1, x_2, x_3\right]$ to y_2 is removed, the causal-ANN fails. In the causal-ANN with multiple layers, the accuracy of the previous sub-network has large impact on

Table 16
Comparison of R^2 values when missing links in causal graphs.

Missing link(s)	y_1	y_2	W_{hs}
None	0.967	0.994	0.934
$x_1 - y_2$	−60.280	0.0913	−122.265
$x_2 - y_2$	−4.120	0.6768	−9.300
$x_3 - y_2$	−62.245	−1.002	−126.222
$x_4 - y_2$	0.922	0.992	0.844
$x_5 - y_2$	0.949	0.992	0.897
$x_6 - y_2$	0.931	0.993	0.861
$[x_4, x_5, x_6] - y_2$	0.909	0.993	0.818

Table 17
ANOVA analysis results of $\left[x_1, \ldots, x_6\right]$ to y_2.

	x_1	x_2	x_3	x_4	x_5	x_6
Prob > F	0	0	0	0.111	0.165	0.565

the next sub-network and the errors will accumulate through the sub-networks. Thus, the low accuracy of y_2, when removing the links from $\left[x_1, x_2, x_3\right]$ to y_2, leads to a failed prediction of y_1 as the negative R^2 value. However, if any of the links from $\left[x_4, x_5, x_6\right]$ to y_2 is missed, the prediction accuracy will not decrease much compared with the correct causal graph. Table 17 gives the ANOVA analysis results of $\left[x_1, \ldots, x_6\right]$ to y_2, which illustrates that $\left[x_1, x_2, x_3\right]$ are important variables while $\left[x_4, x_5, x_6\right]$ are not. Therefore, missing the links of the important variables will decrease the prediction accuracy significantly while missing the links of unimportant variables will influence the accuracy slightly. In engineering design, the chance of missing less important variables is much larger than missing important variables and missing those less important variables will influence the accuracy of the causal graph slightly. On the other hand, if important variables are missed from the causal graph, the prediction of the causal-ANN will be poor or unacceptable.

4.3.3. Impact of variable correlations

In engineering problems, design variables usually correlate with each other. But considering the correlations may lead to higher simulation expenses because a large number of variable combinations should be considered and more samples should be generated. To reduce the computational cost, the multiple parents usually are assumed to be independent from each other in common probability inference. In this case, each design variable is considered independently and the interval of each design variable with the largest likelihood is determined separately and finally those intervals of variables are put together to form the interesting design subspace (Backlund et al., 2015). However, ignoring variable correlations may bring extra errors in probability inference. Thus, the impact of variable correlations is discussed in this section. To illustrate the differences between considering variable

Table 18
Interesting area detected with independent assumption in power converter design.

	x_1	x_2	x_3	x_4	x_5	x_6
Actual model	1	2	5	5	1	1
Predicted model	3	5	2	4	4	2

Table 19
Interesting area detected with independent assumption in aircraft concept design.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Actual model	5	1	1	5	5	1	1	1	2
Predicted model	5	1	1	5	4	1	1	1	2

correlations or not, the interval with the largest likelihood is estimated with the independence assumption on the same 10,000 samples, and the results are shown in Tables 18 and 19 for both the power converter design problem and aircraft design problem. By comparing the results between Tables 8 and 18 for the power converter design problem, it can be found that in the power converter design problem, ignoring the variable correlations may lead to completely wrong results. It can be explained that for a highly nonlinear problem, optimizing it along each dimension cannot find the optimal solution. When the design variables are highly correlated, the combined influence of design variables may dominate the objective value variance. On the other hand, as shown in Table 19 for the aircraft design problem, the fifth design variable tends to be in a different interval compared with the results considering correlations and the interval where the optimal solution is in from Table 14. In this case, the correlation influence of the design variables is weaker than that in the power converter problem. Thus, the interesting interval estimated independently is near the actual one. Therefore, correlations between design variables should be considered in probability inference.

5. Conclusion

To improve metamodel accuracy and efficiency, knowledge of the engineering design problem is employed in building the metamodel. The cause–effect relations are combined with ANN to develop the causal-ANN model. The entire ANN is divided into several sub-networks according to the causal graph. Values of intermediate variables are employed in constructing sub-networks. Therefore, by reducing the complexity of each sub-networks, the accuracy of the entire ANN is improved and the computational cost of building the model is reduced. The causal-ANN is employed in two engineering case studies and the results show that the prediction accuracy of the causal-ANN outperforms ANN or RBF model.

The developed causal-ANN can be used to replace the expensive simulation for design exploration and optimization. In this work, a causal-ANN based attractive space identification method is developed. Likelihood distributions of the design variables are estimated through Bayesian networks according to samples estimated from the causal-ANN. By comparing the likelihood distributions, the attractive sub-spaces can be identified, which can be used to improve the design efficiency by reducing the large design space to some small regions. In the two engineering cases, by employing the proposed method, the attractive sub-spaces can be found in both cases. Since the samples used in the distribution estimation come from the causal-ANN, there is no expensive simulation involved and thus has negligible costs. Additionally, the impacts of errors in the causal graph and variable correlations are discussed based on the testing results. For the causal graph, involving intermediate variables in the complex sub-networks can improve the prediction accuracy. Missing less important links will not influence the accuracy much but missing important links will cause the prediction to fail, which will in turn help to validate the quality of the knowledge. The variable corrections have influences on the likelihood estimation and they should be considered in attractive design space detection.

This paper offers a novel view on metamodel construction, which combines knowledge about the engineering problem and data-based modeling. For future work, different applications of the causal-ANN will be researched in support of engineering design.

CRediT authorship contribution statement

Di Wu: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft. **G. Gary Wang:** Supervision, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Backlund, Peter B., Shahan, David W., Seepersad, Carolyn Conner, 2015. Classifier-guided sampling for discrete variable, discontinuous design space exploration: Convergence and computational performance. Eng. Optim. 47 (5), 579–600, Retrieved October 15, 2018 (http://www.tandfonline.com/doi/abs/10.1080/0305215X.2014.908869).
Ben-Gal, Irad, Ruggeri, F., Faltin, F., Kenett, R., 2007. Bayesian Networks. In: Encyclopedia of Statistics in Quality and Reliability.
Bishop, Christopher M., 2006. Pattern Recognition and Machine Learning. Springer.
Cheng, Bing, Titterington, D.Michael, 1994. Neural networks: A review from a statistical perspective. Statist. Sci. 2–30.
Coatanea, Eric, Roca, Ric, Mokhtarian, Hossein, Mokammel, Faisal, Ikkala, Kimmo, 2016. A conceptual modeling and simulation framework for system design. Comput. Sci. Eng. 18 (4), 42–52, Retrieved January 15, 2017 (http://ieeexplore.ieee.org/document/7499786/).
Collobert, Ronan, Bengio, Samy, 2001. SVMTorch: Support vector machines for large-scale regression problems. J. Mach. Learn. Res. 1 (Feb), 143–160.
Ebrahimi, Ebrahim, Monjezi, Masoud, Khalesi, Mohammad Reza, Armaghani, Danial Jahed, 2016. Prediction and optimization of back-break and rock fragmentation using an artificial neural network and a bee colony algorithm. Bull. Eng. Geol. Environ. 75 (1), 27–36, Retrieved November 25, 2018 (http://link.springer.com/10.1007/s10064-015-0720-2).
Fang, Hongbing, Horstemeyer, Mark F., 2006. Global response approximation with radial basis functions. Eng. Optim. 38 (4), 407–424, Retrieved November 27, 2016 (http://www.tandfonline.com/doi/abs/10.1080/03052150500422294).
Fonseca, D.J., Navaresse, D.O., Moynihan, G.P., 2003. Simulation metamodeling through artificial neural networks. Eng. Appl. Artif. Intell. 16 (3), 177–183, Retrieved November 25, 2018 (https://www.sciencedirect.com/science/article/pii/S0952197603000435).
Friedman, Nir, Geiger, Dan, Goldszmidt, Moises, 1997. Bayesian network classifiers. Mach. Learn. 29 (2–3), 131–163.
Hill, William J., Hunter, William G., 1966. A review of response surface methodology: A literature survey. Technometrics 8 (4), 571–590.
Hippert, H.S., Pedreira, C.E., Souza, R.C., 2001. Neural networks for short-term load forecasting: A review and evaluation. IEEE Trans. Power Syst. 16 (1), 44–55, Retrieved November 25, 2018 (http://ieeexplore.ieee.org/document/910780/).
Joseph, V. Roshan, Hung, Ying, Sudjianto, Agus, 2008. Blind kriging: A new method for developing metamodels. J. Mech. Des. 130 (3), 31102, Retrieved November 27, 2016 (http://mechanicaldesign.asmedigitalcollection.asme.org/article.aspx?articleid=1449652).
Kang, Suh Young, 1992. An Investigation of the Use of Feedforward Neural Networks for Forecasting. Kent State University.
Kott, G., Gabriele, G.A., 1993. Application of multidisciplinary design optimization to the power stage design of a power converter. ASME Adv. Des. Autom. 2, 359–366.
Lippmann, R., 1987. An introduction to computing with neural nets. IEEE ASSP Mag. 4 (2), 4–22, Retrieved November 25, 2018 (http://ieeexplore.ieee.org/document/1165576/).
Liu, Jia, Gong, Maoguo, Miao, Qiguang, Wang, Xiaogang, Li, Hao, 2018. Structure learning for deep neural networks based on multiobjective optimization. IEEE Trans. Neural Netw. Learn. Syst. 29 (6), 2450–2463, Retrieved November 25, 2018 (https://ieeexplore.ieee.org/document/7920404/).
Maniezzo, V., 1994. Genetic evolution of the topology and weight distribution of neural networks. IEEE Trans. Neural Netw. 5 (1), 39–53, Retrieved November 25, 2018 (http://ieeexplore.ieee.org/document/265959/).
Morris, Daniel, Antoniades, Andreas, Took, Clive Cheong, 2017. On making sense of neural networks in road analysis. In: Proceedings of 2017 International Joint Conference on Neural Networks (IJCNN), May 14-19. IEEE, Anchorage, AK, USA, pp. 4416–4421, Retrieved November 25, 2018 (http://ieeexplore.ieee.org/document/7966415/).

Padula, Sharon, Alexandrov, Natalia, Green, L., 1996. MDO test suite at NASA langley research center. In: 6th Symposium on Multidisciplinary Analysis and Optimization.

Pearl, Judea, 2014. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann.

Sobieszczanski -Sobieski, Jaroslaw, Agte, Jeremy S., Sandusky, Robert, 2000. Bi-level integrated system synthesis. AIAA J. 38 (1), 164–172.

Sobol', Il'ya Meerovich, 1990. On sensitivity estimation for nonlinear mathematical models. Mat. Model. 2 (1), 112–118.

Srivastava, Nitish, Hinton, Geoffrey, Krizhevsky, Alex, Sutskever, Ilya, Salakhutdinov, Ruslan, 2014. Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Torrie, James Hiram, 1960. Principles and Procedures of Statistics: With Special Reference to the Biological Sciences. McGraw-Hill.

Wang, G. Gary, Shan, S., 2007. Review of metamodeling techniques in support of engineering design optimization. J. Mech. Des. 129 (4), 370, Retrieved November 27, 2016 (http://mechanicaldesign.asmedigitalcollection.asme.org/article.aspx?articleid=1449318).

Wang, Liqun, Shan, Songqing, Wang, G. Gary, 2004. Mode-pursuing sampling method for global optimization on expensive black-box functions. Eng. Optim. 36 (4), 419–438.

Wang, Dapeng, Wang, Gaofeng, Naterer, Greg, 2007. Collaboration pursuing method for MDO problems. AIAA J. 45 (5), 1091–1103, Retrieved (http://arc.aiaa.org/doi/abs/10.2514/6.2005-2204).

Wang, Fang, Zhang, Qi-Jun, 1997. Knowledge-based neural models for microwave design. IEEE Trans. Microw. Theory Tech. 45 (12), 2333–2343, Retrieved November 25, 2018 (http://ieeexplore.ieee.org/document/643839/).

Wong, F.S., 1991. Time series forecasting using backpropagation neural networks. Neurocomputing 2 (4), 147–159, Retrieved November 25, 2018 (https://www.sciencedirect.com/science/article/pii/092523129190045D).

Wu, Di, Coatanea, Eric, Wang, G. Gary, 2019. Employing knowledge on causal relationship to assist multidisciplinary design optimization. J. Mech. Des. 141 (4), 41402, Retrieved February 1, 2019 (http://mechanicaldesign.asmedigitalcollection.asme.org/article.aspx?doi=10.1115/1.4042342).

Wu, Di, Wang, G. Gary, 2020. Knowledge-assisted optimization for large-scale design problems: A review and proposition. J. Mech. Des. 142 (1).

Zhang, Guoqiang, Eddy Patuwo, B., Hu, Michael Y., 1998. Forecasting with artificial neural networks:: The state of the art. Int. J. Forecast. 14 (1), 35–62, Retrieved November 25, 2018 (https://www.sciencedirect.com/science/article/pii/S0169207097000447).