Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Correspondence:

Causality fields in nonlinear causal effect analysis*

Aiguo WANG[1], Li LIU‡[2], Jiaoyun YANG‡[3], Lian LI[3]

[1]*School of Electronic Information Engineering, Foshan University, Foshan 528225, China*

[2]*School of Big Data & Software Engineering, Chongqing University, Chongqing 400044, China*

[3]*School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230009, China*

E-mail: wangaiguo2546@163.com; dcsliuli@cqu.edu.cn; jiaoyun@hfut.edu.cn; llian@hfut.edu.cn

Received Apr. 21, 2022; Revision accepted May 5, 2022; Crosschecked May 24, 2022

https://doi.org/10.1631/FITEE.2200165

Compared with linear causality, nonlinear causality has more complex characteristics and content. In this paper, we discuss certain issues related to nonlinear causality with an emphasis on the concept of causality field. Based on widely used computation models and methods, we present some viewpoints and opinions on the analysis and computation of nonlinear causality and the identification problem of causality fields. We also reveal the importance and practical significance of nonlinear causality in handling complex causal inference problems via several specific examples.

1 Introduction

According to the three-level causal hierarchy proposed by Judea Pearl, having the knowledge of causality rather than correlation endows us with the ability to better answer the critical questions related not only to association, but also intervention and counterfactuals (Pearl, 2019). In principle, causality studies the cause–effect relationship, and thus helps reveal the data generation procedure. However, statistical correlation does not necessarily indicate causation. For example, a positive correlation between yellow fingers and cough is falsely generated by smoking. In this case, smoking is the common cause (i.e., confounder) of both yellow fingers and cough, and we cannot say that yellow fingers causally lead to cough. The Simpson paradox, in which uncontrolled and even unobserved variables could reverse or eliminate an association between two variables, is a typical phenomenon in probability and statistics. For example, the COVID-19 case fatality rate in Italy is higher than that in China overall, but for age groups, the case fatality rates in Italy are lower than those in China (von Kügelgen et al., 2021). One strategy to treat this phenomenon is to use causality. In addition, causal inference contributes to the development of machine learning in better solving open problems (e.g., explainability, robustness, adaptability, and transfer learning) and further paves the way for next-generation artificial intelligence with the abilities such as reflection and reusable mechanisms (Yue et al., 2020; Schölkopf et al., 2021). Therefore, understanding and analyzing cause–effect relationships is of great value and of primary interest in a variety of fields, e.g., economics, education, genomics, epidemiology, and medical science (Stavroglou et al., 2019).

Although significant progress has been achieved in causal inference, to our knowledge, most of existing studies deal with linear causal inference, while

‡ Corresponding authors

* Project supported by the National Major Science and Technology Projects of China (No. 2018AAA0100703), the National Natural Science Foundation of China (No. 61977012), the China Scholarship Council (No. 201906995003), and the Fundamental Research Funds for the Central Universities, China (No. 2021CDJYGRH011)

ORCID: Aiguo WANG, https://orcid.org/0000-0001-6150-8068; Li LIU, https://orcid.org/0000-0002-4776-5292

© Zhejiang University Press 2022

little attention has been devoted to nonlinear causal inference (Guo et al., 2021; Yao et al., 2021). To this end, we present some viewpoints and opinions on nonlinear causal inference, including its definition, characteristics, and its differences to linear causal inference, especially the causality field problem.

2 Causality fields

We first present related notations before illustrating the nonlinear causal effect. Suppose that D is a set of data points and $u\in D$ is defined by

$$u=\begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix}, \tag{1}$$

where x_i (i=1, 2, ⋯, n) denotes the i^{th} variable (feature) and a_i indicates its corresponding value. Specifically, a_i can be binary, multi-valued, or continuous, and u=(a_1, a_2, ⋯, a_n) when there is no ambiguity. Although the binary case of a_i is the simplest and perhaps the most commonly used one in practice, we are often required to handle multi-valued, discrete, and even continuous values in the nonlinear causality problem.

Given the cause variable X, effect variable Y, variable Z that influences X and Y (often called the covariable in the potential outcome framework, or the backdoor variable in the causal structural model), and exogenous variable U, we can obtain Eq. (2), known as the structural causal equation (SCE), via the data-fitting or regression methods (Rubin, 2005; Pearl, 2009).

$$Y=f(X,Z,U). \tag{2}$$

The average causal effect (ACE) for discrete variables is denoted by Eq. (3):

$$\begin{aligned}&\mathrm{ACE}(X\to Y)\\&=\left\{E(Y|\mathrm{do}(X=x))-E(Y|\mathrm{do}(X=x'))\right\}/(x-x')\\&=E\left\{E_Z\left[f(x,Z,U)-f(x',Z,U)\right]\right\}/(x-x').\end{aligned} \tag{3}$$

It is known that ACE($X\to Y$) indicates the causal effect of X on Y by measuring the change in Y for a one-unit change in X. For a continuous case, it is expressed as

$$\begin{aligned}&\mathrm{ACE}(X\to Y)\big|_{X=x}\\&=\lim_{\Delta x\to 0}E\left\{E_Z[f(x+\Delta x,Z,U)-f(x,Z,U)]\right\}/\Delta x\\&=E\left\{E_Z\frac{\partial}{\partial X}f(x,Z,U)\right\},\end{aligned} \tag{4}$$

which essentially estimates the ACE at X=x.

Obviously, if $f(X, Z, U)=aX+bZ+cU$, which is a linear equation, Eqs. (3) and (4) would return ACE($X\to Y$)=a; that is, the causal effect equals the coefficient a of X irrespective of the values of Z and U, which is a characteristic of linear causality. However, it raises complex issues for the nonlinear case. Then, we present the definitions and characteristics related to nonlinear causal analysis.

Definition 1 (Nonlinear causality) If the functional relationship $f(X, Z, U)$ is nonlinear, the causal relation between X and Y is called nonlinear causality.

For nonlinear causality, the causal effect of X on Y depends not only on X but also on Z and U, and different values of X, Z, and U lead to different ACEs. Consequently, this makes nonlinear causality different from linear causality to a large extent and complicates the nonlinear causal effect. Specifically, we can divide the nonlinear causal effect into strong nonlinear and quasi-linear types.

Strong nonlinear: if X appears in the derivative $\frac{\partial}{\partial X}f(X, Z, U)$, the causal relation is strong nonlinear. That is, the causal effect depends on the values of X, Z, and U, and their different values generally lead to different causal effects.

Quasi-linear: if X does not appear in the derivative $\frac{\partial}{\partial X}f(X, Z, U)$, the causal relation is quasi-linear. For this case, the causal effect is closely related to Z and U.

This indicates that the expectation of ACE($X\to Y$) in quasi-linear causality may vary only with the values of Z and U, and that the expectation of ACE($X\to Y$) in strong nonlinear causality depends also on the value of X besides Z and U. For the purpose of explanation, Table 1 gives an illustrative example, where X=1 indicates the treatment group (drug taker), X=0 is the control group (drug nontaker), Y has three outcomes, including recover (Y=1), worsen (Y=−1), and unchanged (Y=0), and Z={1, 0} denotes daytime and night respectively. Z influences both treatment X and effect Y.

Suppose that the results of drug test effectiveness are as shown in Table 1. The SCE is

$$Y = ZX - (1 - Z)X. \tag{5}$$

Table 1 The effect of a drug considering the time

Time	With or without drug	Effect
Daytime (Z=1)	With (X=1)	Recover (Y=1)
	Without (X=0)	Unchanged (Y=0)
Night (Z=0)	With (X=1)	Worsen (Y=−1)
	Without (X=0)	Unchanged (Y=0)

This case is obviously quasi-linear. According to Eq. (3), we can obtain

$$\begin{aligned}&\mathrm{ACE}(X \to Y)\\ &\quad = E(Y|\mathrm{do}(X=1)) - E(Y|\mathrm{do}(X=0)) = 0.\end{aligned} \tag{6}$$

However, we have $\mathrm{ACE}|_{Z=1}=E(Y|\mathrm{do}(X=1), Z=1)-E(Y|\mathrm{do}(X=0), Z=1)=1$ for Z=1, and $\mathrm{ACE}|_{Z=0}=E(Y|\mathrm{do}(X=1), Z=0)-E(Y|\mathrm{do}(X=0), Z=0)=-1$ for Z=0. Obviously, ACEs are different for different Z values. That is, the causal effect of X on Y is fluctuant and related to Z, where there is a positive effect during daytime (i.e., taking the drug helps recovery) and a negative effect at night (i.e., the condition worsens upon taking the drug). Thus, if we conclude from ACE=0 that there is no causal effect of X on Y, it does not conform to the facts or our common sense. This indicates that, in contrast to the linear causal effect, the expectation of Z probably counteracts the positive and negative effects in quasi-linear causality, and subsequently yields fluctuating interdependencies of X on Y existing in the nonlinear causal effect. This example prompts us to consider the problem of nonlinearity in causality analysis. Obviously, one naive solution is to calculate the causal effects with respect to each value of Z.

To better illustrate the characteristics of nonlinear causality, we introduce the concept of causality field.

Definition 2 (Causality field) This consists of the following three types of causal effect of X on Y:

Positive causality: X causes a homodromous change in Y;

Negative causality: X causes an antagonistic change in Y;

Null causality: X causes no change in Y.

The three fields are noted as

$$\begin{cases}\Omega^{+} = \left\{(x,z,u)|\dfrac{\partial}{\partial X}f(X,Z,U) > 0\right\},\\ \Omega^{-} = \left\{(x,z,u)|\dfrac{\partial}{\partial X}f(X,Z,U) < 0\right\},\\ \Omega^{0} = \left\{(x,z,u)|\dfrac{\partial}{\partial X}f(X,Z,U) = 0\right\},\end{cases} \tag{7}$$

called the causality fields of X on Y.

For linear causality, the causal effect is constant and is independent of X, Z, and U. However, the causal effect is fluctuant for nonlinear causality, which requires us to identify different causality fields for solving a specific problem.

Next, we study a complex example: the typical competitive–symbiotic relationship between two biological populations, rabbits and foxes. Such a relationship is given by the following differential equation, the so called Lotka–Volterra equation (Takeuchi et al., 2006; Sugihara et al., 2012):

$$\begin{cases}\dfrac{\mathrm{d}N_{\mathrm{R}}}{\mathrm{d}t} = N_{\mathrm{R}}\left(\alpha - \beta N_{\mathrm{F}}\right),\\ \dfrac{\mathrm{d}N_{\mathrm{F}}}{\mathrm{d}t} = -N_{\mathrm{F}}\left(\gamma - \delta N_{\mathrm{R}}\right),\end{cases} \tag{8}$$

where N_{R} denotes the number of rabbits, N_{F} denotes the number of foxes, t denotes time, $\mathrm{d}N_{\mathrm{R}}/\mathrm{d}t$ and $\mathrm{d}N_{\mathrm{F}}/\mathrm{d}t$ represent the instantaneous growth rates of rabbits and foxes respectively, and α, β, γ, and δ are positive real parameters.

Fig. 1 presents the illustration plot, where the arrows denote the population trends between rabbits and foxes. We can observe that rabbits and foxes have competitive and symbiotic relationships. For example, under different conditions, an increase in the population size of rabbits can lead to the increase in the population size of foxes (i.e., area A, ecological mutualism model), the decrease in the population size of foxes (i.e., area B, ecological competition model), or the decrease in both the rabbit and the fox population (i.e., area C, ecological total-loss model). Obviously, the population of rabbits affects that of foxes under a certain causality existing between them; however, this causality fluctuates between positive and

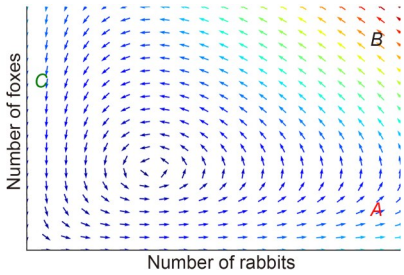

Fig. 1 Relationship between two biological populations, rabbits and foxes

The two species have a different number of individuals and their interdependency changes periodically among Ω^+, Ω^-, and Ω^0

negative values, and has different patterns in different fields. Particularly, null causality exists at the boundary between the positive and negative regions, although its probability measure may be zero. Therefore, the use of ACE probably distorts the true causality and fails to reflect the quantitative relationships between two species. In this situation, the causality field is an important concept to study these relations.

3 Computation for nonlinear causal relation

In this section, we discuss how to estimate nonlinear causality and focus on how to analyze the causality field. First, we introduce the following three assumptions to enable causal effect estimation, which play a crucial role in the calculation and analysis of causal effects (Rubin, 2005):

Assumption 1 (Stable unit treatment value assumption, SUTVA) SUTVA states that the potential outcome of a unit (or individual) i is independent of the treatment applied to other units, and that two units having the same treatment value would receive the same treatment. SUTVA basically involves no interference and the well-defined treatment levels. It indicates that there is no difference between the assignments of the treatment group and control group; that is, the experimental results would not change no matter the units are assigned to the treatment or control group. SUTVA is essentially equivalent to the independent causal mechanism (ICM) principle, stating that the causal generative process of variables is composed of autonomous modules that do not inform or influence each other.

Assumption 2 (Ignorability) The potential outcome Z is independent of the treatment assignment X conditioned on the covariable Z, i.e., $(Y^{X=1}, Y^{X=0}) \perp X|Z$, where $Y^{X=x}=E_Z(Y^{X=x}(z))$. This is also called the unconfounder assumption, since it posits the nonexistence of unobserved confounders. However, the means to verify this is a nontrivial task. This assumption enables us to calculate the potential outcome:

$$P\left(Y^{X=x'}|X=x,Z=z\right)=P\left(Y|X=x',Z=z\right), \quad (9)$$

which transforms the counterfactual problem into the statistical analysis of observational data. Judea Pearl has proven that ignorability is equivalent to the backdoor criterion (Yao et al., 2021).

Assumption 3 (Positivity) This means the choosability treatment assignment, i.e., $P(X=x|Z=z)>0$, $\forall x, z$. This assumption makes the potential outcomes meaningful; otherwise, it is pointless to discuss and calculate $P(Y|X=0, Z)$ if there are no data points in any case with $X=0$ conditioned on Z, i.e., $P(X=0|z)=0$.

Furthermore, we often call ignorability and positivity together as strong ignorability. With the above assumptions, we can calculate the ACE. The potential outcome framework is commonly used to infer the causal effect with regard to the treatment–outcome pair $(X, Y|Z)$, where Y is the outcome of treatment X applied to a population, and Z indicates the covariable. Then, we define the treatment effect as the difference between the outcomes of different treatments. Specifically, we use $Y_z^{X=x}=E(Y|X=x, Z=z)$ to denote the potential outcome with $X=x$ conditioned on $Z=z$ and define the z-specific causal effect (SCE_z) as

$$\mathrm{SCE}_z=E\left(Y_Z^{X=1}-Y_Z^{X=0}\right), \quad (10)$$

where $X=0$ ($X=1$) corresponds to the control (treatment) trial. We can also define the ACE over a population:

$$\begin{aligned}\mathrm{ACE}&=E\left(E_Z\left(\mathrm{SCE}_z\right)\right)\\&=E\left[E_Z\left(Y_Z^{X=1}\right)-E_Z\left(Y_Z^{X=0}\right)\right]\\&=E\left[E_Z(Y|X=1,Z=z)-E_Z(Y|X=0,Z=z)\right],\end{aligned} \quad (11)$$

where E_Z is the expectation with respect to Z. Thus, ACE equals the expectation of SCE_z associated with Z. Note that Z should be the set of backdoor variables.

Eq. (11) also embodies the causal effect Eq. (12) proposed by Judea Pearl, such as

$$E\left(E_Z\left(Y^{X=x}\right)\right)=E\left(Y|\mathrm{do}\left(X=x\right)\right). \tag{12}$$

When X, Z, and Y all take binary values, there is no difference between linear and nonlinear causality. However, when they are multi-valued, ACE and SCE probably change following the values of Z, X, or both Z and X. Specifically, it is more challenging to handle continuous cases. According to the above discussion, if X is multi-valued,

$$\mathrm{ACE}=E\left[E_Z\left(Y^{X=x}\right)-E_Z\left(Y^{X=x'}\right)\right]/\left(x-x'\right), \tag{13}$$

where $Y^{X=x}$ is the observational outcome of the present treatment X=x, and $Y^{X=x'}$ indicates the counterfactual outcome of the imagined treatment $X=x'$. The ACE can then be explained as the difference between taking X=x and X=x'. If X, Z, and Y are continuous variables, we obtain the so-called SCE, as show in Eq. (4).

We can observe that ACE depends on the values of X, Z, and U, which leads to the problem of causality field. In this situation, ACE=0 does not mean that there is no causal effect of X on Y, since their causal effects vary with different causality fields and the ACE could be zero.

However, there are cases where the assumptions do not hold. For example, the interdependency among individuals is clear and unavoidable in the situation of rumor dissemination in social networks, often characterized by the fact that a rumor believed by more people usually generates more people that believe it, which violates the STUVA assumption.

The most popular modeling paradigms might be the causal graphical model and the causal equation model. Machine learning can be seen as a special case of the causal equation model, which is currently popular among researchers. If the graph or the causal equation is given, we can perform a Markov (causal) factorization of the joint distribution (or the term disentangle in machine learning) (Schölkopf et al., 2021):

$$P\left(X_1,X_2,\cdots,X_n\right)=\prod_{i=1}^{n}P\left(X_i|\mathrm{PA}_i\right), \tag{14}$$

where PA_i represents the parents of the node (variable) X_i. To learn the causal relations, current approaches typically fall into two categories: constraint-based approaches that use conditional independence tests to determine direct causal relations between observed variables, and score-based approaches that quantitatively score possible explanations of the causal relations (Spirtes and Zhang, 2016). Unfortunately, checking the Markov factorization consistency of the relations from observations and evaluating all possible structures become infeasible with the growth of the modeling size, especially in the case of high-dimensional datasets. This is mainly because it is hard to determine the directionality between two variables that are statistically dependent and can generate mutual information. Most importantly, some datasets are often finite and not causally sufficient (i.e., unobserved common causes exist to confound two observed variables).

Compared to the graphical model that describes the observational distribution, SCE is more intuitive for machine learning researchers who are more accustomed to thinking while considering estimating functions rather than probability distributions (Pearl, 2019).

$$X_i=f_i\left(\mathrm{PA}_i,U_i\right), \tag{15}$$

where f_i denotes a function and U_i represents an exogenous variable. To satisfy causal sufficiency, it is required that U_1, U_2, $\cdots$, U_n be independent according to the common cause principle. Given a set of variables X_i (i=1, 2, $\cdots$, n), if i<j, without loss of generality, $X_j\notin\mathrm{PA}_i$, which means X_j will not be on the right side of the equation of X_i. If Z is the covariable, and the cause variable X and effect variable Y are determined, then X and Z must be the ancestor nodes of Y. Accordingly, we can derive Y=f(X, Z, U) by combining the associated equations and solve it by applying machine learning models or fitting a model to the raw data. However, we need to know the parent nodes of X and Y to obtain the fitting equation. We should select Z and U that influence X and Y but exclude the variables that are influenced by X or Y according to the associations in the dataset. This is the key step to derive the fitting equations. Although there are methods currently available, their performance is still unsatisfactory. Typically, for machine learning methods, assumptions on function

classes should be given due to the arbitrarily slow convergence. An open research question in the frontier of machine learning is how to derive a model of nonlinear causal relations for a large number of variables, under which the causal inference framework can best exploit the power of machine learning technology. Moreover, additional challenges exist in the era of big data in learning causal relations from high-dimensional data and large-scale mixed data. Recently, a series of causal inference algorithms have been proposed to handle these challenges by employing advanced machine learning technologies, such as deep learning networks, to identify nonlinear relations. Score-based methods use a scoring function to measure the quality of fitting a causal graph to the data, and a lower score indicates the existence of incorrect conditional independence. These methods consist of two components: structural equation and scoring function, where the former generally uses a parameter θ to decide whether or not one fitting equation is adopted, and the latter can translate a candidate causal graph into a parameterized score of the structural equation. Propensity score analysis is another commonly used method that aims to create a scoring function $e(x) \in \mathbb{R}$. This method stratifies the data according to the values of $e(x)$ and obtains a balanced matching for covariate distribution in the same stratum, as long as $X \perp Z|e(z)$. Thus, the propensity score method projects the n-dimensional data into a reduced dimension, which greatly reduces the computational cost. One disadvantage of this method is that it requires prior knowledge to design the scoring function. Note that causal models, either the causal graph or SCE, can explicitly model interventions and generalize under certain distribution shifts in the light of the ICM principle, and thus can recognize the causality field by computing.

To learn the causal effects from a nonlinear model, there are two types of methods with respect to data observation. If the values of all the variables X, Y, Z, and U are observable, the approaches such as regression adjustment, propensity score, covariate balancing, and machine learning based models can be leveraged to learn the casual effect. On the other hand, when there are unobserved variables or confounders, we can calculate the estimated casual effects by adopting other special variables, such as the instrumental variable (IV), mediator, and running variable, or using tools such as the front-door criterion, to avoid the collection of useless unobserved variables.

In the nonlinear causal model, there are some other assumptions for relaxing the complexity of causal effect computation. One common assumption is that Z and X are independent if Z is unobservable, and Y is dependent on Z (usually called unobserved heterogeneity). Another assumption, monotonicity, expresses that a change from X=false to X=true cannot make Y change from true to false under any circumstance. In this regard, we refer interested readers to an excellent book (Wooldridge, 2010) that summarizes a series of approaches for learning causal effects from different observable situations of data under various assumptions.

Another interesting topic is causal representation learning in machine learning. Since real-world observations, such as objects in an image that are usually not structured into the units or variables that can be directly described as a function, such functions should be learned from the high-dimensional and low-level data (e.g., pixels). Accordingly, Schölkopf et al. (2021) discussed three issues of modern machine learning, including disentangled representation learning, transferable mechanism learning, and interventional model learning and reasoning. The first one has been widely studied in machine learning, aiming to learn a latent disentangled representation $W=\{w_1, w_2, \cdots, w_k\}$ ($k \ll n$) from the observation data by a feature transformation function g:

$$g: \{x_1, x_2, \cdots, x_n\} \rightarrow \{w_1, w_2, \cdots, w_k\}, \quad (16)$$

where $\{x_1, x_2, \cdots, x_n\}$ represents the original features. The g function can be modeled using neural networks (e.g., the encoder–decoder framework). Unlike the disentanglement approach, the latter two approaches are currently in their infancy but are essential in machine learning. Since the size of training data in each domain is limited and large-scale manual labeling is burdensome, it is not surprising that future artificial intelligence models possess the capacity of reusing the fundamental modules or structural knowledge across domains. This coincides with the ICM principle. Therefore, new machine learning models should be designed to identify the causal modules (i.e., graphs or equations) that do not inform or influence each other, and

such modules should be able to characterize inherent and invariant relations beyond the representation of statistical dependence structures and be reused across different environments.

Overall, using machine learning models or fitting equation methods to calculate causal relations involves two major steps. The first is to learn optimal feature representation via machine learning models to disentangle for covariables or backdoor variables, and the second is to obtain a fitting function $f(X, Z, U)$ by taking machine learning models as the generator of function. The greatest challenge in using machine learning to calculate the causal effect and to recognize causality fields is choosing appropriate data coding, known as the feature representation problem. The raw observational data typically contain natural features (e.g., pixels in the image, primary physiological and pathological indicators of a patient), and it is difficult to directly use them to do causal inference via potential outcome or the backdoor criterion. Thus, performing feature space optimization (e.g., feature transformation and feature compression) on the natural features is expected to better uncover the latent causality and to help identify covariables or backdoor variables. Accordingly, researchers have proposed a wealth of data-driven causality analysis and computing models, including the advanced machine learning models (e.g., deep learning and variational autoencoder), toward better data representation. Please refer to the literature (Guo et al., 2021; Schölkopf et al., 2021; Yao et al., 2021) for details. Most importantly, for all of the causal inference algorithms, a crucial issue is how to identify the causality field to reveal the true causal effect among variables. However, this topic still lacks in-depth discussion. Note that ACE is often used to measure the causal effect, which could be inappropriate in nonlinear causality and can distort the true causality.

4 Examples of the causality field

In this section, we apply the SCE model on several examples to illustrate the differences between the causality field and ACE.

For this task, we obtain three datasets from Kaggle (https://www.kaggle.com/), including the alcohol consumption dataset, life expectancy (WHO) dataset, and diabetes dataset. These contain 213, 2938, and 768 samples, respectively. We then analyze the causality field of the alcohol consumption (X) on the suicide rate (Y) with income per person (Z) as the confounding variable, body mass index (BMI) (X) on life expectancy (Y) with gross domestic product (GDP) (Z) as the confounding variable, and insulin (X) on diabetes (Y) with age (Z) as the confounding variable.

To calculate the causality field, we first need to obtain SCE, as denoted by Eq. (2). As this is a nonlinear equation, we apply machine learning models to learn the equation to achieve a good fitting effect. Considering that there are only two variables as input, we apply support vector regression (SVR) to fit the model for the first two datasets and support vector machine (SVM) with the Gaussian kernel function for the diabetes dataset. A five-fold cross validation is conducted to optimize the hyper-parameters. Assuming that M denotes the trained model, $y=M(x, z)$ represents the causal equation of Y.

According to Definition 2, the causality field is achieved by calculating $\frac{\partial}{\partial X} f(X, Z, U)$. Here, the nonlinear causal equation is obtained by training machine learning models, which is out of the scope of this paper. It means that we cannot obtain the form of $M(x, z)$; therefore, we apply a numerical method to obtain the causality field through Eq. (17):

$$\begin{aligned} g(x,z) &= \frac{\partial}{\partial X} M(X,Z) \\ &= \lim_{\Delta x \to 0} \frac{M(\mathrm{do}(x+\Delta x),z) - M(\mathrm{do}(x),z)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{M(x+\Delta x,z) - M(x,z)}{\Delta x}, \end{aligned} \tag{17}$$

where $M(\mathrm{do}(x), z)$ represents the value of y when performing the intervention $X=x$, given the condition $Z=z$. Z is the unique backdoor variable, so $M(\mathrm{do}(x), z)=E_Z[M(x, z)]$.

Although X, Y, and Z may be continuous variables, their values are discrete in practical datasets. Thereby, for each x and z, we can use the established model to compute $g(x, z)$, while $\mathrm{ACE}(X \to Y)$ can be computed by $\sum_z g(x, z)P(z)$.

Fig. 2 shows the causality field and ACE of the alcohol consumption (X) on the suicide rate (Y) with income per person (Z) as the confounding variable. From ACE, we can find that with the increase of X, the causal effect first decreases and then increases. Besides, ACEs are larger than 0 regardless of the value of X, which means that X has only a positive effect on Y. This indicates that consuming alcohol can only increase the suicide rate. It seems that if the policy makers want to reduce this rate, they should strictly prohibit alcohol consumption. However, when we check the causality field, although it shows similar trends to ACE, there are still some effects that are smaller than 0 for different Z values; i.e., there are negative causality field and null causality field. This means that consuming a certain amount of alcohol can sometimes reduce the suicide rate when the values of X and Z fall into the scope of the negative causality field; hence, we should allow some alcohol consumption to reduce the suicide rate.

To obtain the particular form of the causality field, we fit a polynomial function $g(x, z)$ according to these discrete x, z values. Since the ranges of x and z considerably vary, we perform normalization on x and z before fitting the function. The mean and standard deviation of x are 11.5 and 6.7, respectively, while the mean and standard deviation of z are 2.62×10^4 and 1.51×10^4, respectively. Eq. (18) shows the function:

$$g(x,z) = 0.97 + 1.4x - 0.43z + 0.4x^2 - 0.34xz - 0.17z^2 - .21x^3 - 0.02x^2z - 0.05xz^2. \quad (18)$$

The root-mean-square error (RMSE) is 0.094. If $\{x, z\}$ satisfies $g(x, z)=0$, then we can obtain the null causality field Ω^0. If $\{x, z\}$ satisfies $g(x, z)<0$, then we can achieve the negative causality field Ω^-. We can also obtain the positive causality field Ω^+ by setting $g(x, z)>0$. For example, the values $\{-1.103, 1\}$, $\{-0.8, 1\}$, $\{0, 1\}$ of $\{X, Z\}$ fall into the scopes of null causality field, negative causality field, and positive causality field respectively. These values are normalized, and they correspond to the original values $\{4.11, 4.14\times10^4\}$, $\{6.14, 4.14\times10^4\}$, and $\{11.4, 4.14\times10^4\}$ respectively.

Figs. 3a and 3b respectively show the causality field and ACE of BMI (X) on life expectancy (Y) with GDP (Z) as the confounding variable. The normal range of BMI is 18.5 to 25. We can find that, in this normal range, ACE increases with the increase of BMI. Beyond this normal range, ACE will decrease. However, if we check the SCE_z figure, we can find that when GDP is larger than a threshold, BMI has no effect on life expectancy. This means that, for lower GDP, policy makers should pay attention to BMI to increase people's life expectancy, while for higher GDP, if they want to increase life expectancy, BMI is no longer a factor to consider. This fact could not be discovered from ACE. If polices are formed according to ACE, there will be a waste of resources. For this dataset, we do not fit functions for the unabridged form, and only estimate the causality field for some data points, as $g(x, z)$ is too complicated.

Figs. 4a and 4b respectively illustrate the causal effect and ACE of insulin (X) on diabetes (Y) with

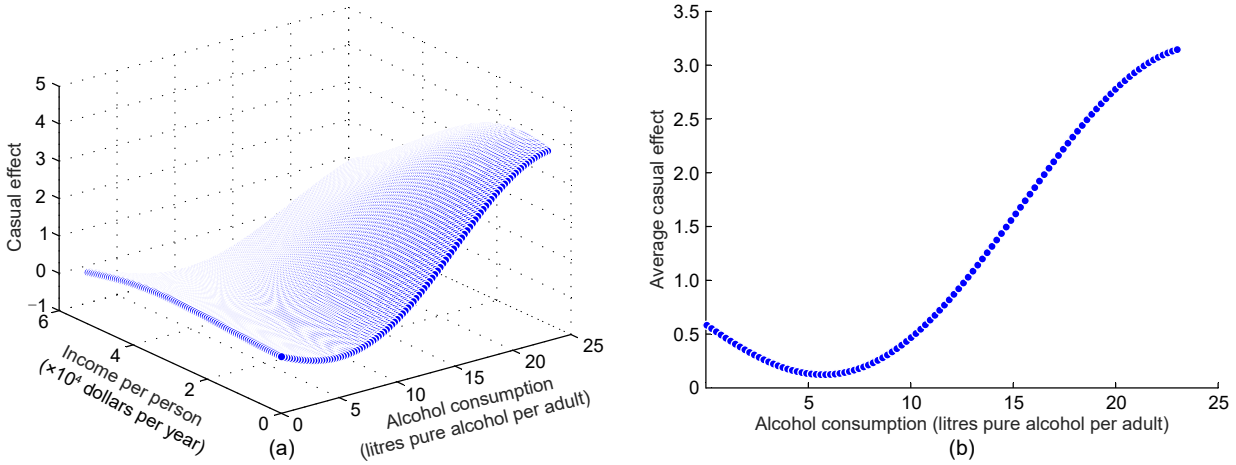

Fig. 2 Causal effect of the alcohol consumption on the suicide rate with income per person as the confounding variable: (a) causality field; (b) average causal effect

age (Z) as the confounding variable. For this dataset, Y is binary and contains only two values, namely having diabetes or not having diabetes. Therefore, we apply the softmax function to convert the output into a continuous value within the range from 0 to 1. For this dataset, the causality field and ACE show a similar trend. With the increase of insulin, the causal effect decreases regardless of age. When it is larger than a threshold, insulin has negative effect on diabetes, which means that it can prevent the onset of diabetes. When it is lower than a threshold, insulin can increase the risk of diabetes. However, there is still a small difference among different ages given the same insulin dose. It means that the threshold for insulin on diabetes varies according to age, and this variation is concealed by averaging. We can control different thresholds of insulin according to the age rather than using a fixed threshold. After fitting a polynomial function, we can obtain $g(x, z)$ as

$$g(x,z)=0.001\,467-1.28\times10^{-6}x-4.24\times10^{-6}z -7.8\times10^{-10}x^2+1.07\times10^{-9}xz-3.86\times10^{-8}z^2. \tag{19}$$

Subsequently, we can identify the causality field Ω^0, Ω^+, and Ω^- based on this equation.

From the above three examples, we can infer that the causality field differs a lot from the average causal effect. When z varies, the causal effect will change even for the same x. This inspires us that, when trying to figure out the causal effect of X on Y, we should check the confounding variable Z and

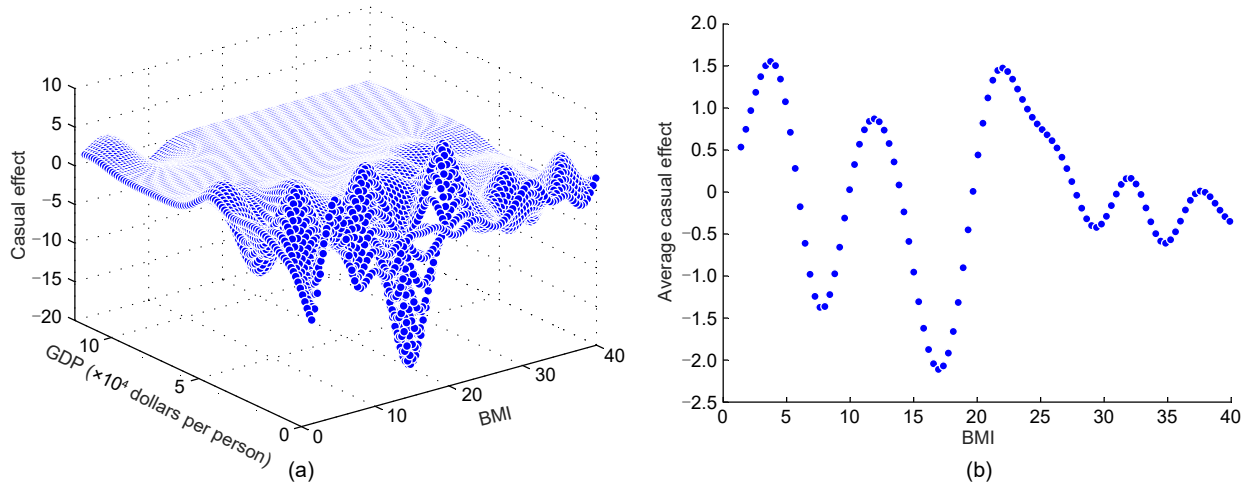

Fig. 3 Causal effect of BMI on the life expectancy with GDP as the confounding variable: (a) causality field; (b) average causal effect

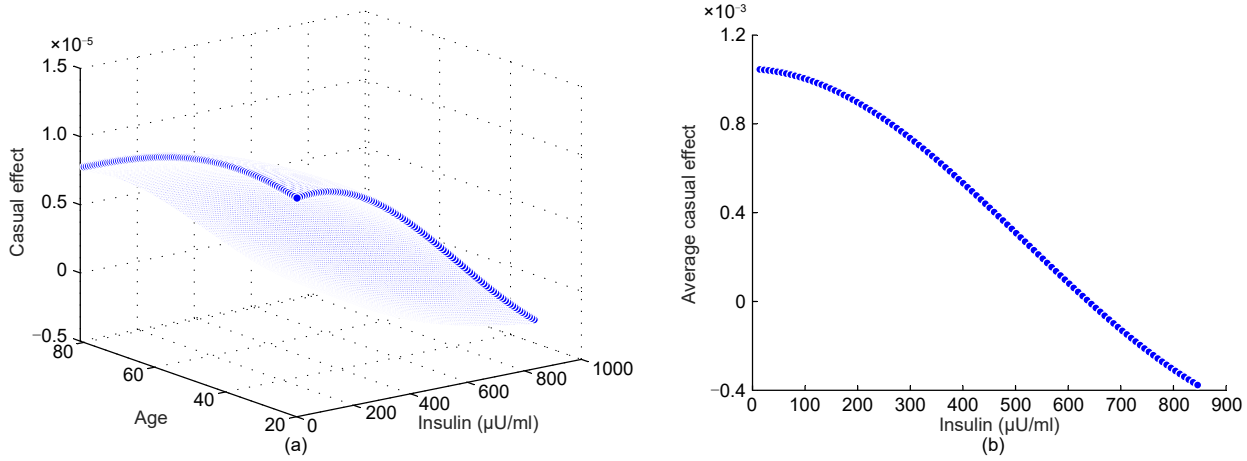

Fig. 4 Causal effect of insulin on diabetes with age as the confounding variable: (a) causality field; (b) average causal effect

choose different strategies according to its value. This can be applied for various situations, e.g., different individuals and environments. This approach could guide the individualized policy making process like precision medicine.

Notably, we have used simple confounding variables in the above three examples. In practice, it is usually difficult to determine the confounding variables. The above are just examples to illustrate the differences between the causality field and ACE. When handling complicated datasets, researchers should refer to the graph-based model, potential-outcome framework, or even deep learning based model to obtain the causality field. However, these methods are beyond the scope of this paper.

5 Conclusions

We present preliminary discussions on the problems of nonlinear causal effect calculation and the causality field, which are pivotal for solving complex practical problems. These discussions could help understand both linear and nonlinear causal relations. We expect further considerations of these problems to facilitate the research on nonlinear causal relations and their wider application.

Contributors

Aiguo WANG, Li LIU, Jiaoyun YANG, and Lian LI designed the research. Jiaoyun YANG processed the data. Aiguo WANG, Li LIU, Jiaoyun YANG, and Lian LI drafted the paper. Aiguo WANG, Li LIU, and Jiaoyun YANG revised and finalized the paper.

Compliance with ethics guidelines

Aiguo WANG, Li LIU, Jiaoyun YANG, and Lian LI declare that they have no conflict of interest.

References

Guo RC, Cheng L, Li JD, et al., 2021. A survey of learning causality with data: problems and methods. *ACM Comput Surv*, 53(4):75. https://doi.org/10.1145/3397269

Pearl J, 2009. Causality: Models, Reasoning, and Inference (2nd Ed.). Cambridge University Press, Cambridge, UK.

Pearl J, 2019. The seven tools of causal inference, with reflections on machine learning. *Commun ACM*, 62(3):54-60. https://doi.org/10.1145/3241036

Rubin DB, 2005. Causal inference using potential outcomes: design, modeling, decisions. *J Am Stat Assoc*, 100(469): 322-331. https://doi.org/10.1198/016214504000001880

Schölkopf B, Locatello F, Bauer S, et al., 2021. Toward causal representation learning. *Proc IEEE*, 109(5):612-634. https://doi.org/10.1109/JPROC.2021.3058954

Spirtes P, Zhang K, 2016. Causal discovery and inference: concepts and recent methodological advances. *Appl Inform*, 3:3. https://doi.org/10.1186/s40535-016-0018-x

Stavroglou SK, Pantelous AA, Stanley HE, et al., 2019. Hidden interactions in financial markets. *Proc Nat Acad Sci USA*, 116(22):10646-10651. https://doi.org/10.1073/pnas.1819449116

Sugihara G, May R, Ye H, et al., 2012. Detecting causality in complex ecosystems. *Science*, 338(6106):496-500. https://doi.org/10.1126/science.1227079

Takeuchi Y, Du NH, Hieu NT, et al., 2006. Evolution of predator–prey systems described by a Lotka–Volterra equation under random environment. *J Math Anal Appl*, 323(2):938-957. https://doi.org/10.1016/j.jmaa.2005.11.009

von Kügelgen J, Gresele L, Schölkopf B, 2021. Simpson's paradox in Covid-19 case fatality rates: a mediation analysis of age-related causal effects. *IEEE Trans Artif Intell*, 2(1):18-27. https://doi.org/10.1109/TAI.2021.3073088

Wooldridge JM, 2010. Econometric Analysis of Cross Section and Panel Data (2nd Ed.). MIT Press, Cambridge, USA.

Yao LY, Chu ZX, Li S, et al., 2021. A survey on causal inference. *ACM Trans Knowl Discov Data*, 15(5):74. https://doi.org/10.1145/3444944

Yue ZQ, Zhang HW, Sun QR, et al., 2020. Interventional few-shot learning. Proc 34th Conf on Neural Information Processing Systems, p.2734-2746.