



A kind of artificial intelligence model based on nature without statistic

Jun Meng^{1,2} · Jie-zhou Yu¹ · Xiao Chen³

Received: 24 June 2019 / Accepted: 26 June 2020 / Published online: 9 July 2020
© The Author(s) 2020

Abstract

Simple rules can generate complexity and human can also learn fast with rules. Can machines learn in a similar way? Can artificial intelligence be independent of statistics? Machine learning is growing rapidly but models are poorly interpretable and depend on statistics. We propose a method by iteration based on causality which is the real one exists in the system. It is composed of fixed goals and basic rules called DNA rules. These DNA rules can be obtained from the definition and is not statistical rules. The causality in rules promises the process to be precise, because the potential attractor of the system is deterministic, because it is subjected to the rules although the system is complex especially under uncertain interference. Such a model not only works well in the traditional deterministic systems like the stable point and limited circle but also can work in some seemingly random and systems which are considered to be stochastic systems. The model is taken to play a game and it makes the machine learns fast and adaptively, and it is also interpretable with the causality and independent from the amount of data for it is based on causal iteration. It learns and even predicts the seemingly random interference in the game. We found such a model is adaptable, and it works well even in out-of-sample situations. The model is compared with an LSTM network in prediction a seemingly random sequence, the result shows the causality-based model also works well. We think that it may help to solve some problems hard for the traditional statistical method and become an enrichment for the current models.

Keywords Statistics · Causality · DNA rules · Prediction · Adaptability · Intelligence

Introduction

Alan Turing outlined his ideas and proposed equations supporting them in his original paper ‘The chemical basis of morphogenesis’ in 1952 [18]. In his paper, a complex pattern resembling an animal’s fur pattern is generated by a

mathematical formula, and such patterns are related to the Turing structure. Actually, simple rules can generate complex phenomena that are even similar to life in nature. As is shown in Fig. 1, turbulence can be generated by convection [9], and the pattern similar to the shell can be generated by rule 30th with the cellular automaton [20]. Similar to complexity, the human may learn to play some games fast without many trials after knowing the rules. Therefore, we wonder if artificial intelligence can be generated based on simple rules? Can such intelligence adapt to out-of-sample conditions? (Fig. 1).

Recently, statistic learning methods boast notable capabilities. However, these models need much data tab with complete information reflecting the system’s characteristics [4, 10, 17]. In addition, the models are hard to interpret, their function is single and they cannot predict in the out-of-sample conditions [8, 12]. For example, the model is hard to interpret, because it is a ‘black box’ [12, 14, 16]. The structure and working process inside the ‘black box’ are unknown. The structure is important for the network [6], because data with different characteristics need networks

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40747-020-00174-z>) contains supplementary material, which is available to authorized users.

✉ Jun Meng
junmeng@zju.edu.cn

¹ Center for Data Mining and Systems Biology, College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China

² Research & Educational Center for the Control Engineering of Translational Precision Medicine (R-ECCE-TPM), School of Biomedical Engineering, Dalian University of Technology, Dalian 116024, China

³ HangZhou Hongsheng Intelligent Technology Co., Ltd, Hangzhou 310027, China

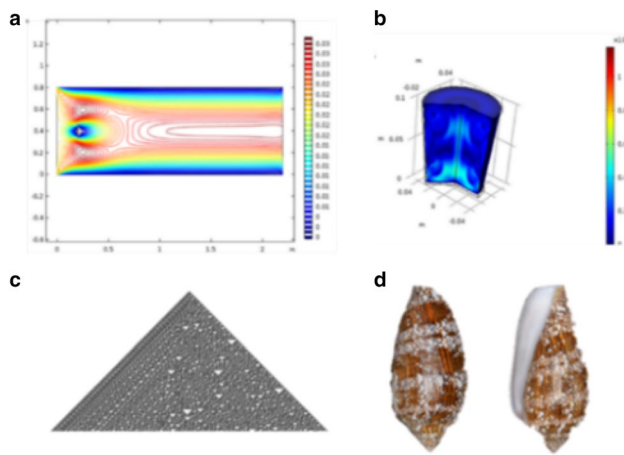


Fig. 1 Examples of the complex phenomenon and simple rules. Turbulence seems to be very chaotic, it can be generated based on simple rules. We simulated the process and generated it by convection between cold and heat. **a** and **b** are the cases of turbulence in a plane and a cup. The pattern generated by the iteration of the simple rule 30th in the cellular automaton in **c** is similar to the pattern on the shell in **d**. Sparse coding and IFS(iterating function system) also show that the basic abstract features that can constitute complex phenomena

with different structures, CNN is good at processing image and RNN is good at time sequence. It is difficult to learn data with different features well by a common ‘black box’ without knowing the inside of the box.

In contrast, the rule-based models have better interpretability. However, rules directly learned from data by statistical learning may lack mathematical logic [13, 21]. Our method is based on ‘DNA rules’ which are different from the rules learned by statistics, because they are real causalities that exist in the system. In addition we call them ‘DNA rules’ for they are like DNA on some aspect. Such a model can be interpreted by the causality. A biological mutation experiment showed that mutations considered to be random have the same mutation results under the same conditions in several experiments [2], which imply that there are certainty and causality even in complicated phenomena. The method we propose is based on a similar view, we believe there are causality and certainty between the causes and results and the causality can be mined, then we can take the causality to deductive and generate the evolution process of the system. This process is logically strict with certainty. The model can be interpreted by causality. In addition, it can generate a complete attractor corresponding to all the possible states in the system’s evolution. Therefore, comparing with statistical models, the model can also work well in out-of-sample conditions. With the attractor, the model is more complete and has better adaptability to changes. *In addition, the causality here not only works well in the traditional deterministic systems like the stable point and limited circle but also can*

work in some seemingly random and systems which are considered to be stochastic systems in the traditional statistical method. The method shows that these seemingly random systems can be mined, learned, and predicted in the short term. The method can be applied in some control and predicting in some complex and nonlinear evolutionary systems and could be an enrichment to the current model.

The rest of this paper has been organized into 4 sections. Section “DNA rules” presents the concepts of DNA rules. Section “Our method” shows our method and related concepts in detail. Section “Causality can be obtained” shows experiments based on the causality-based method. Lastly, Section “Experiment” discusses and outlines the advantages of the model. Conclusions and some inferences are also proposed. The method is proved to be theoretically feasible in the supplement and the properties of the model are proved.

DNA rules

DNA rules are the basic rules. They are the causal relationship in the system rather than the rules learned by statistical methods. Statistical rules may be wrong [12], because the data may mislead the model, but the causality of the system itself must be right. The DNA rules correspond to the system’s complete hypothesis space of the solution (Theorem 3 in Supplement) and can generate the attractor of the system which corresponding to the possible evolution process of the system. In contrast, the statistical rules may not be the basic rules, and their hypothesis space is not complete because of the constraints of the rules. Taking building blocks to compare these two rules. DNA rules are the basic building blocks, and the traditional statistical rules are blocks that have been shaped and fixed. It is clear the basic building blocks can make more shapes. Table 1 lists the main differences between DNA rules and traditional rules. The model based on DNA rules not only has the advantage of being interpretable but also has better adaptability to changes than the data-driven model [8].

Our method

The model we propose is based on the belief that the evolution of the systems is based on causality. The causality is composed of goals and DNA rules. The goal is a fixed and general one that applies regardless of the state of the system. For example, the principle of generalized energy minimization that the system moves in a manner of “minimum energy consumption” in generalized motion, like following the gradient direction, getting more awards. In addition, the DNA rules are the basic causal rules of the system. Since the goal and DNA rules are fixed, the

Table 1 Main differences between DNA rules and traditional rules

Differences	DNA rules	Rules in traditional concepts (for example, in expert system)
Causality based	Accurate descriptions of real causalities in system	System’s statistical associations mined by statistical learning
Basic	Most basic rules, cannot or need not be subdivided into more basic rules	Not the most basic rules, and are composed by DNA rules
Invariable	Invariable in different conditions	Variable in different conditions, new rules should be proposed for new conditions
Iterable	Can generate system’s complete attractors by iteration. The iteration process corresponds to the evolution of the system	Cannot iterate to generate system’s complete attractors, because they are not causalities
Adaptable	Can be applied to different conditions. In addition, the attractor they generate also make the system to be adaptable	Not adaptable, new rules should be proposed in new conditions. Therefore, the adaptability of such rules is worse
Interpretable	Can be interpreted by causality	Statistical associations in black box. May not be real and interpretable

The DNA rules are the basic causalities. They are Invariable in different conditions. In addition, iteration based on DNA rules can generated the attractor of the system. The traditional rules are not the most basic rules, and the statistical rules may be wrong. For example, the statistical conclusion that ‘the swan is white’. Such rules will constrain the model from adapting to the new condition

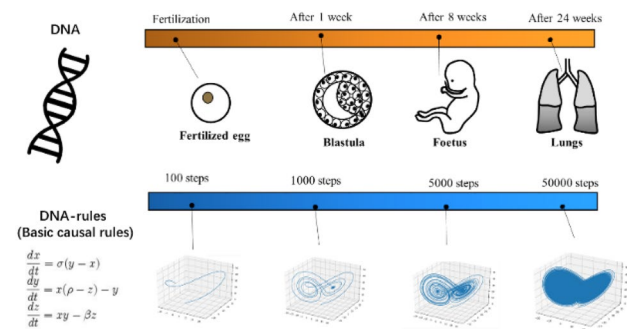


Fig. 2 DNA rules and the DNA. The DNA defines life, and the DNA rules in figure define the Lorenz system. With the DNA rules, the evolution process of the Lorenz system can be accurately generated in the form of an attractor. Such a rule is different from the rules in the expert system. It is a rule with an interpretable causal relationship. At the same time, systems based on DNA rules are highly adaptable. Just as twins grow up in different environments behave differently because of adaptation

causality is fixed, then the process of system evolution is logically strict and deterministic. From the perspective of system evolution, such a causality-based evolution process for the above system can be expressed by Formula (1), where X is the states of the system, A is the deterministic laws in the system, B is the external input from the environment. In the formula, the deterministic laws are composed of the DNA rules and the general fixed goal, they are the laws of nature that exist in the system. A is deterministic and B is uncertain:

$$\dot{X} = AX+B. \tag{1}$$

From the phase-space point of view, the method can generate attractors based on the iterative process with the causality [5, 15, 18], the Lorenz attractor in Fig. 2 is an

example. The attractor is a set of numerical values toward which a system tends to evolve, for a wide variety of starting conditions of the system [1]. It reflects the system’s evolving process and shows a range, where the system is more stable. According to Theorem 1, the attractor can be classified by Lyapunov exponent.

Theorem 1 *In mathematics, the Lyapunov exponent of a dynamical system is a quantity that characterizes the rate of separation of close orbits [1]. The orbits are the trajectories of the system’s evolution in phase space.*

For the discrete system, the average Lyapunov exponent is given by Formula (2) [3], where the λ is the average Lyapunov exponent in function (2), f represents the iteration process and n represents the times of iteration:

$$\lambda(x_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \ln |f'(x_i)|. \tag{2}$$

The direction of the multiple initial separation vectors of the system is different, and the separation rate is also different. Among them, the maximal Lyapunov exponent (MLE) determines the predictability of the system [19]. When the Lyapunov exponent in all direction is negative, the attractor is a stable attractor, such a system is easy to predict for the system converges to the stable point. When all the Lyapunov exponent is greater than zero, the system is divergent and cannot be further predicted. When the MLE is greater than zero, and the system is not infinitely divergent, which means the system also has the Lyapunov exponent less than zero or equal to zero, the system has a strange attractor. The classification is shown in Table 2.

Because the MLE is positive, the orbits in the attractor of the system separate. Therefore, the system will leave from

Table 2 Classify systems according to the Lyapunov index

MLE (λ)	System attractor type	Traditional method	Method in this paper
$\lambda < 0$	Stable point	Statistics (easy)	Causality (easy)
$\lambda = 0$	Limit cycle	Statistics (easy)	Causality (easy)
$0 < \lambda < \infty$	Strange attractor (main focus of this paper)	Statistics (random and not deterministic)	Causality (deterministic)
$\lambda = \infty$	Divergent system	Cannot solve	Cannot solve

The system are classified by the MLE(X) and there are 4 types. When $\lambda < 0$ or $\lambda = 0$, both the tradition statistical method and causality-based method can easily mine the its inner law. When $0 < \lambda < \infty$, the system has a strange attractor, and it becomes more complex. The tradition statistical method treats the system as a stochastic system. However, statistical method may not work well in some scenarios. The method in this paper proposed a new method mainly for such a system. The method does not treat the system as a stochastic system anymore for there are still causalities in the system although the system looks very random and complex. The attractor and the orbits in attractor in deterministic. In addition, we can learn the system and predict it in short term according to its causality

the current orbit when there are external interferences, which means the attractor is sensitive to the changes and orbits are unstable. In addition, the orbits also converge to a certain range in the phase space. Therefore, the attractor has dense orbits in that certain range. These features cause the evolution of the system to seem to be random and uncertain especially when there is an interference, for there seem to be many possibilities for the orbital migration.

However, there are certainties in this seemingly random process. The DNA rules and goals are fixed, so, the causality is fixed. With the constraints of causality, the attractor of the system and its orbits are determined. Even there is interference, the process of system evolution is still subjected to causality. System attractors can be obtained based on causality or data reconstruction methods. The process of attractor generation also depicts the process that the system changes the unstable orbits because of external interferences. In addition, the attractor is coupling with the interference by the fixed goal. The more time of iteration, the more the method can get the approximate shape and orbits of the attractor. It needs emphasis that we do not need to traverse all the points of the attractor, an approximate shape of the attractor and some of its inner orbits are enough. System values that get close enough to the attractor values remain close even if slightly disturbed. In addition, because the causalities are fixed, there are fractal features in the attractor [1], we can also complement the attractors with fractal features. Therefore, the method can also work for large size problems.

System attractor is a prerequisite for complex decision making. Complex decision making is learning, adapting, and predicting complex behaviors. The process of generating attractors is the process of learning the system. In addition, there are two constraints and a fixed goal that promise the model can learn the system successfully and effectively predict and adapt. The first constraint is that the Lyapunov exponent is greater than zero but not infinity. The Lyapunov exponent reflects the rate of separation of orbits is limited, which means system values that get close enough to

the attractor values remain close even if slightly disturbed. The second constraint is we know the current state of the system. In addition, the goal is fixed. In a system evolving under interference, the fixed goal the law of the passive response of the system to external interference, and it drives the system to adapt during the iterative evolution. For example, the goal can be following the gradient direction, getting more awards, they are essentially following the lowest-energy principle. Because the goal is fixed and the attractor is sensitive to changes, the process of adapting to changes based on the goal can be done spontaneously. Compared to other models, the external interferences may be fatal for statistical models; however, the instability of the attractor makes the model adaptable in such a method. In addition, according to the fixed goal, the responses of the system to interferences are recorded, which is the orbital migration in the generated attractor. Furthermore, the local manifold can be constructed based on the system's current position in the attractor. Short-term prediction can be realized on the local attractor manifold for the rate of separation of orbits is not infinity (Theorems 1, 5 and 6 in Supplement).

The method of this paper is main according to the above theory, and it mainly focuses on the above nonlinear evolutionary system which may seem random with a strange attractor. The DNA rules can be expressed in two forms: constraint rules and affirmative rules. These two forms can be converted to each other. The hypothesis space of the attractor of the system is subjected to DNA rules (Theorem 3 in Supplement). The goal works based on the system's perception of external interferences and then drives the system to adapt to external conditions. The water may be a great example, although the shape of the water is complex and diverse, it flows following the gradient direction. As a result, in different environments, it can adapt and generate ever-changing shapes. We believe that causality is fixed in the process of system evolution although the external conditions may always change. The changes in the conditions, the system's state, and evolution are coupled by the causality.

Therefore, external changes will cause and drive the system to change and evolve. Although the results may look very complex and random in terms of timing, this problem may be solved on the phase-space. The attractor converges to a certain range with certainty, its inner orbits are also fixed although dense. Even in the case of external interferences and the system may leave from the current orbit of the attractor to a new one because of the sensitivity, short-term prediction can still be achieved by attractor manifolds. Therefore, the method is to construct a system to generate the strange attractors and to construct an intelligent model with properties of attractors.

Compared with the statistical methods, this method is not to model the data but to reconstruct the system's causality and attractor. So that model can make logically rigorous and precise predictions. It may work in some conditions that traditional ML models cannot work well [8], because it is based on causality and has the complete attractor corresponding to the hypothesis space of the whole system. In contrast, statistical learning methods assume the model parameters to be fixed and mine the statistical relationship in data. Such a model may not be sensitive to small probability problems. In addition, the rule that the model learned from data is not DNA rules, and it is fixed. In fact, the parameters may vary in time. In addition, the statistical correlation obtained by statistical learning may not exist after the change.

As is the architecture of the method shown in Fig. 3, the DNA rules and the goal can be obtained to compose the causality. In addition, the causality can also be learned and mined from the data. Then, we can take the network, decision tree, formula, or other models to learn the mined causality, and then taking advantage of the fast iterative speed of the computer to generate the attractor based on the causality. In addition, such a system will be adaptable to changes and can be predicted in the short term. In addition, it needs to emphasize that the network in this model only learns the mined basic causal rules and causality instead of mining the statistical correlation between input and output. Although the inside of the network is a 'black box', the whole system is still interpreted for the relationship between input and output is the causality. The comparison of proposed method and the traditional statistical method is shown in Fig. 4.

Causality can be obtained

Causality is composed of goals and DNA rules. The DNA-rules are different from the rules learned in a statistical way, there are logical with strict causal relationships. The DNA rules can be obtained from the definition (proved in Supplement by Theorems 2) for the definition is the precise constraint of things. Especially, the definition of games is composed of DNA rules, as is shown in Table 3.

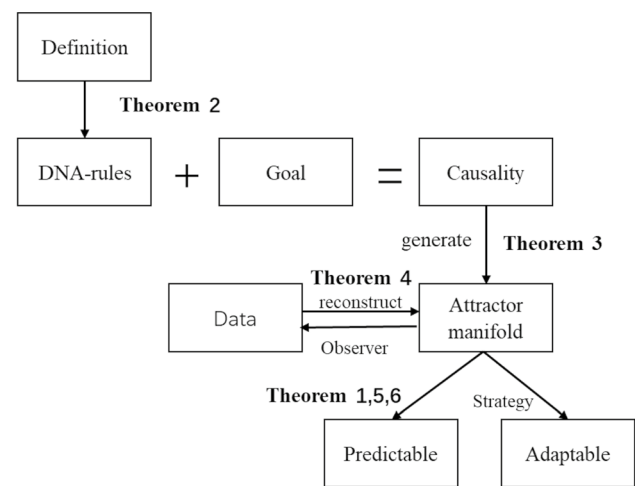


Fig. 3 System architecture and viewpoint of the proposed method. The theorem used and proposed in the paper is also listed. DNA rules are the basic evolution causalities of the system. According to Theorem 2, for some system, like the go game, DNA rules can be obtained from the definition. The causalities are composed of the basic evolution rules and the fixed general evolution goal, like win the game. According to Theorem 3, based on the causalities, the attractor manifold and evolution of the system can be generated by iteration. The attractor is a complete set of orbits that corresponding to the evolution of the system in the phase space. From this view, the time-series data of the system is a part of the attractor observed from a certain dimension. The main causalities of the system can be mined and learned date, according to the Theorem 4, the attractor manifold can also be reconstructed from the time-series data when we cannot directly get the DNA rules to generate attractor of the system. Then according to the attractor, the complex decisions can be made. According to Theorem 1, 5, 6, a system based on the attractor is predictable, so the model constructed by this method can be used to make predictions. In addition, According to Theorem 1, the model based on attractor can learn complex behaviors and adapt to change, which makes the model based on the attractor adaptable to changing conditions

The goal is also independent of statistics, for example, getting more awards, following the lowest-energy principle, and gradient. It can be artificially defined, or it can be the response of the system itself. For example, in the jump and jump game below, there is an artificial goal to adapt the parameters to win more awards. In addition, in the process of predicting the time series, the target is the law of the passive response of the system to external interference, which is decided by the system itself (Table 3).

Causal can also be mined by data reconstruction [7]. Equivalent projections of the attractors can be reconstructed using the time sequence of the system as shown in Fig. 5, and the causality can be further explored. The phase-space reconstruction method can be used to extend the one-dimensional data to the m-dimensional, and then the projection of the original system can be obtained. The phase-space reconstruction method is shown in Formula (3), where the x is the time-series data, and X is

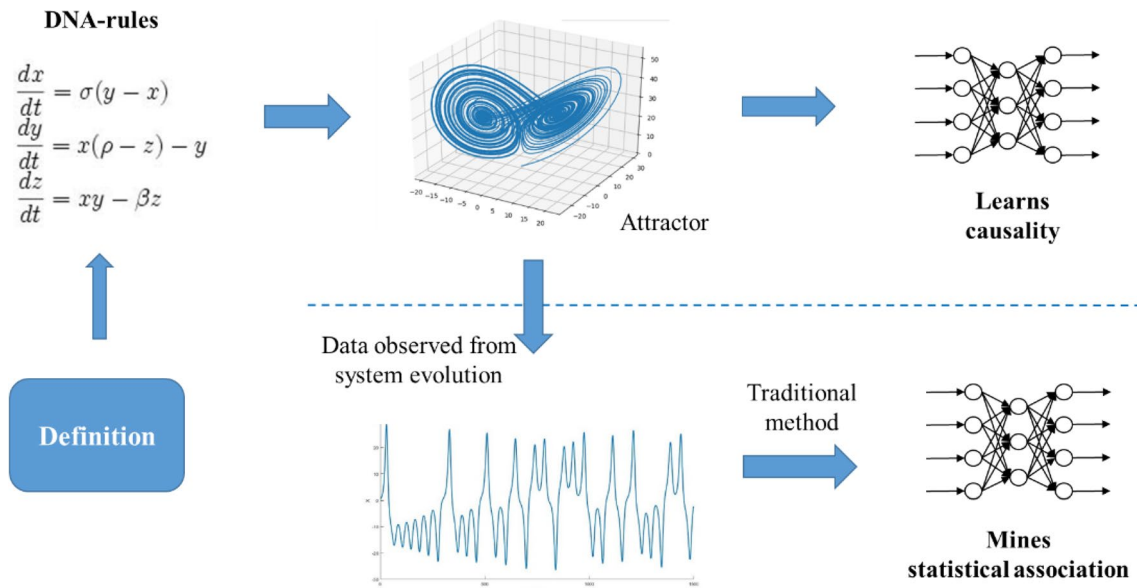


Fig. 4 Comparison of methods. Based on the viewpoint of this article, the evolution of the system is a process in which the system iterates based on DNA rules and generates attractors. The time-series data of system evolution is the observation result of the system attractor in certain dimensions. The statistical method is to use neural networks to mine possible statistical associations directly based

on time-series data of system evolution. Our method is to generate attractors through the iterative evolution of DNA rules or construct the equivalent attractor projection of the system through time-series data. In addition, for some systems, DNA rules can be obtained from the definition of the system

reconstructed m -dimensional data. In the formula, parameters τ and m can be obtained by the mutual information method and the Cao method [8]:

$$X_{m \times N_m} = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_m \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_{N_m} \\ x_{1+\tau} & x_{2+\tau} & \dots & x_{N_m+\tau} \\ \dots & \dots & \dots & \dots \\ x_{1+(M-1)\tau} & x_{2+(M-1)\tau} & \dots & x_N \end{bmatrix}. \quad (3)$$

Experiment

The jump and jump game is easy to set up and everyone can reproduce the experiment. In the game, the player operates the doll to jump to the center of the next box by pressing the screen, and the distance the doll jumps is proportional to the duration of pressing time. The distance between the different boxes is different. The more boxes the doll jumps, the higher the score the player gets. A set of experimental devices was used to play the game, as shown in Fig. 6.

The experimental device is not a linear system because of the interference. For example, the interference in the motor is always changing and affected by many factors. Therefore, as shown in Fig. 7, the interference is changing and seems to be random. In addition, the conversion ratio r of distance and time also changes according to the

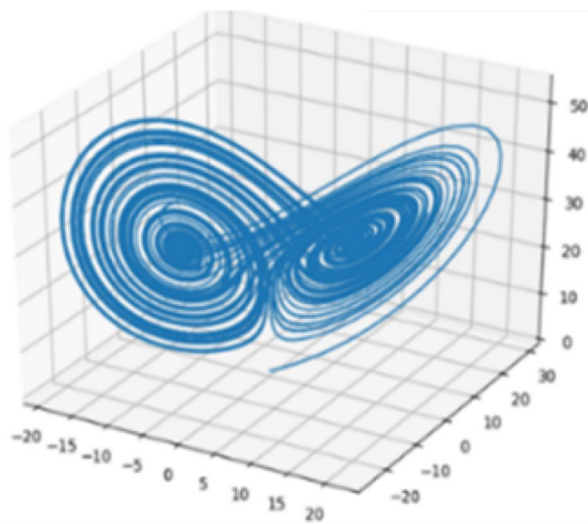
display pixel on different phones. If we set a fixed ratio r to play the game, it will fail because of the changing interference. Therefore, the ratio here should also be a variable value rather than an invariable value. It need not only convert the distance to pressing time but also consider the interference. That means the system is not to simply find a fixed optimal ratio r , it should become adaptable and even predict the interference.

In the above game, the 4 DNA rules shown in Table 2 can be obtained. In addition, the iteration process is based on these rules. In the process of playing, the control error because of the interference is given by $t_{e(k)}$, the control time is $t_{c(k)}$ and the actual time to press the screen is $t_{a(k)}$. The distance that the doll to the center of the next box is $l_{d(k)}$, the distance the doll actually jumps is $l_{a(k)}$ and the error is $l_{e(k)}$. The s is the state after each jump: 1 means jump onto the box and 0 means not on the box. f_r means the iteration process based on four DNA rules. There are relationships as follows:

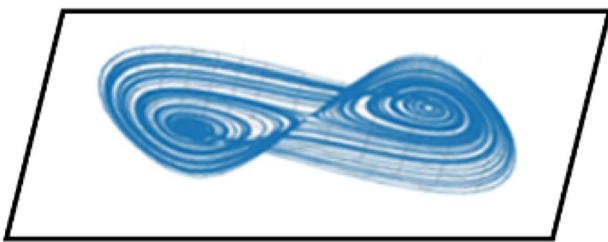
$$t_{a(k)} = t_{c(k)} + t_{e(k)}. \quad (4)$$

$$l_{a(k)} = l_{d(k)} + l_{e(k)}. \quad (5)$$

$$r_{(k+1)} = f_r(r_{(k)}, l_{a(k)}, l_{d(k)}, s_{(k)}) = \begin{cases} \frac{2l_{d(k)} - l_{a(k)}}{l_{d(k)}} r_{(k)}, S_{(k)} = 1 \\ \left(1 + 0.2 \frac{l_{d(k)} - l_{a(k)}}{|l_{d(k)} - l_{a(k)}|}\right) r_{(k)}, S_{(k)} = 0 \end{cases}. \quad (6)$$



a The Lorenz attractor



b The reconstructed attractor

Fig. 5 Data-based causal mining and reconstruction. **a** Lorenz attractor, and **b** reconstructed attractor obtained by the phase-space reconstruction method. It is a projection of the original attractor on the observed data dimension

Therefore, the $r_{(k)}$ is changing to adapt to the interference by iteration process based on DNA rules. In addition, the iteration process is shown by Algorithm 1.

Algorithm 1 Process of changing the ratio r to adapt

Input: the distance of the objection $l_{d(k)}$ and actually jump $l_{a(k)}$ at the time of k .

conversion ratio at the time of k : $r_{(k)}$

the state of the doll at the time of k : $S_{(k)}$

Output: appropriate conversion ratio r

1: **repeat**

2: $r_{(k+1)} \leftarrow$ update according to rules

$f_r(r_{(k)}, l_{a(k)}, l_{d(k)}, S_{(k)})$

The iteration based on DNA rules is actually reconstructing the nonlinear system which includes the changing interference, it generates the attractor of the whole system. Then it can adapt and predict the system within the Lyapunov time based on the attractor. Furthermore, the iterative process continually corrects r , so the errors will not accumulate and lead to failure. Therefore, it is possible to jump to the center of the next box in the following game and get a very high score as shown in Fig. 8. And the iterative process in the game is shown in Fig. 9.

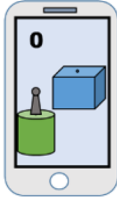

We also compare our method with the statistical method by predicting the disturbance sequence in the above system. The interference sequence is shown in Fig. 7. The data is generated by pressing the screen at the same time interval, the real interval between each pressing is changing. The interference is a part of the whole experiment device, and the causality of the interference is unknown, so we tried to mining the causality from the data and reconstruct it.

In the experiment, a low-pass filter is firstly used to filter out the glitch while maintaining the characteristics of the data. Then the phase-space reconstruction method is taken to reconstruct the attractor and mine the causality. A simple BP (Back Propagation) Network learns the reconstructed data and causality and makes further predictions. The LSTM (Long–Short-Term Memory) Network is used to compare the effectiveness of the model for its network structure is suitable for sequence data. The results are shown in Fig. 10.

The results show the trend of the LSTM predicted curve and the causality-based method curve is similar, they both fit the main trend of the time series. It implies this seemingly random interference has inherent rules and they can be mined and learned. Therefore, this sequence is not actually random. Compare with the LSTM model, the causality-based model requires only a simple BP network rather than a complex network. The model only needs to learn the mined causality. This makes ordinary networks more versatile. Further from the details, the reconstructed network predicts better performance than LSTM in some details. In addition, the mean square error of these two models is similar. The LSTM is 6.8 and the BPNN is about 6.9, which means the BPNN also works well. In addition, the results also show both models have shortcomings in detail fitting. On the one hand, the disturbance is very short, it is easily affected by other secondary factors. On the other hand, there is an error in the process of collecting data, because the accuracy of the timer is limited.

In the jump and jump game, an attractor is generated by iteration based on causality, and the adaptability of the attractor is utilized to realize the adaptation to the constantly changing interference in the system, so the system can get a very high score. The second experiment realized the prediction by mining and learning the causality. Learning the causality is equivalent to learning the system’s attractor,

Table 3 DNA rules and the definition

System	Definition	Rules
<p>Jump and jump game</p> 	<ul style="list-style-type: none"> • A game for one player. • Operate the doll to jump onto the next box • There is a linear relationship between the distance of the jump and the duration of pressing the screen • more boxes jumped, higher score can get. 	<ul style="list-style-type: none"> • Rule1: When the doll jumps onto the box and exceeds the center of it, adjust r slightly smaller. • Rule2: When the doll jumps onto the box and does not reach the center of it, adjust r slightly bigger. • Rule3: When the doll does not jump to the box and exceeds it, adjust r much smaller • Rule4: When the doll does not jump to the box and does not reach it, adjust r much bigger. <p>(The r is a ratio of distance and time)</p>
<p>A shooting game</p> 	<ul style="list-style-type: none"> • A game for one player. • Shot the slowly moving target • Try to get more score in certain times 	<ul style="list-style-type: none"> • Move the direction with the target • Rule1: Move the gun to the left if last time shoot on the right of target • Rule2: Move the gun to the right if last time shoot on the left of target • Rule3: Move the gun to the up if last time shoot on the below of target • Rule4: Move the gun to the down if last time shoot on the top of target

We can obtain the DNA rules according to the definitions of the game rather than by statistic. The rules are equivalently obtained from the definition of the game is exactly composed of the DNA rules. Take the shooting game as an example. According to the definition, the player should try to shoot the center of the slowly moving target, so it will adjust the orientation to shoot. Then on the basis of considering the movement of the target, four DNA rules for four situations. If you let the machine to play the game, the machine can also use four DNA rules to play. Because we know that the orientation influent the trajectory of the bullet. The causality does not need to learn from amounts of data. That is the advantage of the causal model. In addition, the jump and jump game is similar. In addition, 4 DNA rules can be obtained to play the game and make the ratio constantly adapts to the changing disturbance

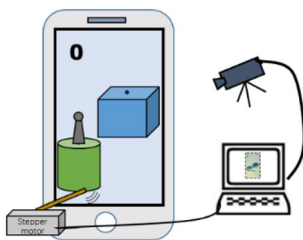


Fig. 6 Experiment devices. The game runs on the phone and the real-time images are pushed to the PC. The PC analyses images and controls the stepper motor to press the screen with a capacitive pen. There is a changing interference in the device and we should constantly change the ratio of time and distance to adapt to it and even predict it

because attractor can be generated by causality. According to the experiments, we believe that the data appearing to be chaotic and random in time sequence is predictable from the perspective of the attractor except for the system of infinite divergence. The reason for the uncertainty in the time sequence data may because the evolutionary iteration of the system in phase space may seem to be random when observing from a single dimension, or the uncertainty may also be caused by the external interference which will drive

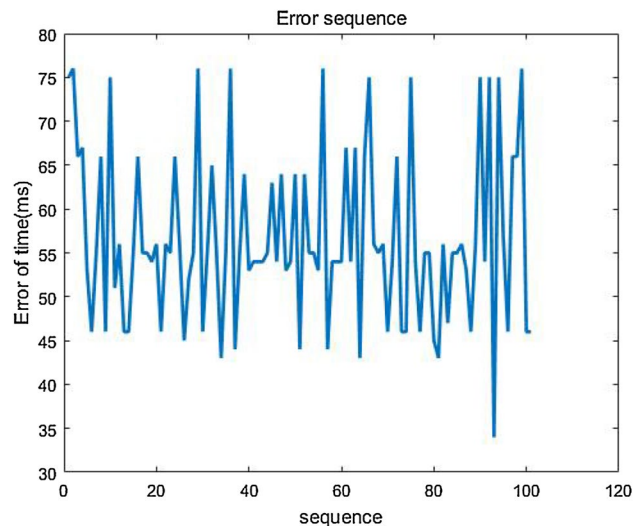


Fig. 7 Interference in the experiment devices. The data t is generated by pressing the screen at the same time interval, the real interval between each pressing is changing. This data appears to be random and difficult to predict. In addition, it cannot be predicted by simply using a network. It interferes with the system and makes it complex

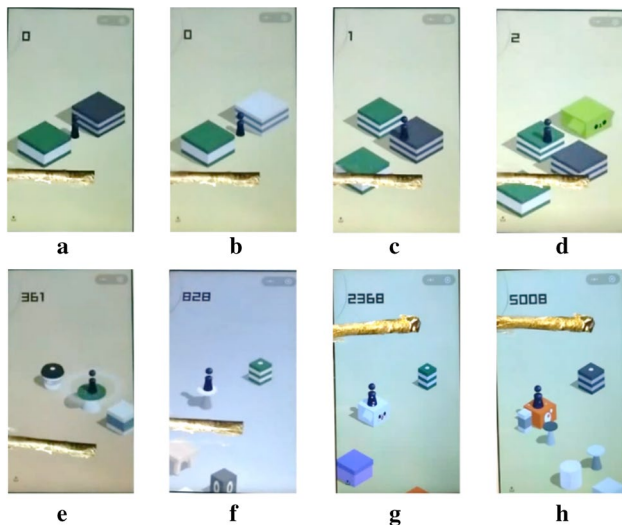


Fig. 8 Based on the method in this paper, the model can adapt to the interference and get a very high score after training. The above pictures are shot during the game. **a** Model failed for the first time just like the person plays at the beginning. **b** Model failed again, but it performed better than the first time. **c** Doll successfully jumped to the box although still inaccurate. **d** Doll jumped closer to the center of the next box. **e, f** Model plays well although there is a random interference in the game. The process from **a** to **f** shows that the model can master the game quickly. It learns, predicts, and adapts the interference as it is a part of the whole system. **g, h** Results of the model plays after adjusting the device. After adjusting, interference and ratio will also change. Fortunately, the model is highly adaptable to changes, it can quickly master the game again and get a very high score again even after changing the game device

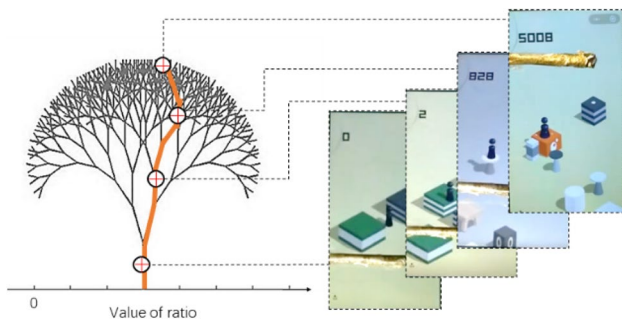
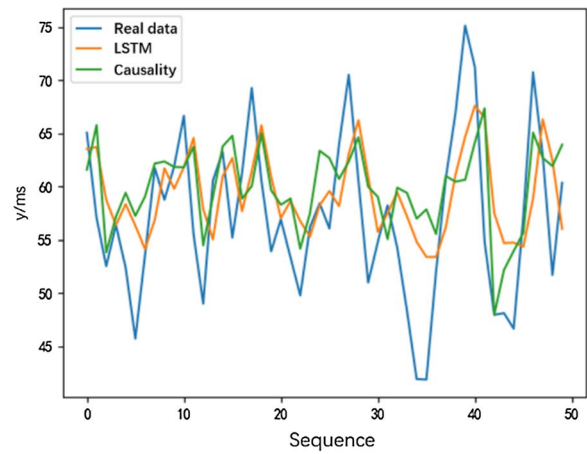


Fig. 9 System iterative process during the game. The iterative process is the process of learning and adapting the game and generated the attractor of the system. The generation of attractors records the orbits of the attractors and the system’s response to interferences. When the system is iterating, ratio r iterates basing on DNA rules and the goal. This process can be vividly explained by a tree although the attractor is not a decision tree. The tree simply shows the changes in the iteration on a plane, but the iterative process is actually the attractor which is converged in a certain range in the phase space. The value in the horizontal coordinates of the tree corresponds to the value of the ratio, and each fork represents an iteration. This continuously iterated tree can reach all positive numbers by iteration. At the same time, the rules will flexibly adjust the ratio when the conditions change. The bolded portion of the orange represents the process of iterations according to the interference in a certain game and the picture on the right corresponding to the progress of the game



a The result of prediction

	Causality based method	LSTM
Interpretability	Can be interpreted by causalities	Black box
Complexity of network	Simple structure	Complex structure
Input data	Constructed by phase space reconstruction	Original time series data
Learning content	Causality	Statistical correlation

b Comparison of methods

Fig. 10 Predictions of the interference. Predictions of the interference. The prediction process is implemented by mining the causality and learning the causality by BP network. In the reconstruction, the optimal delay time of the sequence is 2, and the dimension m is 4. After reconstructing the time series, the main causality generating the interference is mined. Then the BP network can learn this causality and make predictions. **a** The simple BP network performs just like the complex LSTM method. **b** Compares these two prediction methods

the system to change orbits in the attractor. In the first case, the forecast can be achieved by constructing the attractor. In addition, for the second reason, the system can be predicted in the short term using our method basing on attractor (Theorems 1, 5 and 6 in Supplement).

Conclusion

Artificial intelligence is programming to simulate, achieve, and even eventually exceed the level of human intelligence, and prediction is one of the important issues of AI. In this paper, the method generates model by iteration based on causality and it is different from current methods based on statistical learning. It is adaptable and has the ability to predict, and it works well even in out-of-sample situations

for it has complete attractor. These promise the model can work well prediction. In addition, it is also interpretable, not dependent on the amount of data [11]. In addition, this model also robust and flexible. In addition, the attractor shows the distribution of optimal solutions is sparse and fractal. These properties are proved in the supplement. This method may be a fruitful subject for further research. In addition, conjecture is proposed.

Conjecture 1 *Definition is the best DNA rules for the iteration to generate the intelligent system.*

The model based on the definition is modeling the system, but the model based on data is different, because the data may only reflect a part of the system. The definition is the exact meaning of an object. If one exceeds the definition, then it does not belong to that category anymore. In the experiment, if the rule is more than defined, the range of the attractor will be smaller, which will result in poor adaptation and if the rule is less than the definition, the range of the attractor obtained by the iteration will become larger, resulting in making errors. We compare the solution sets of definitions and other features that describe the same thing. S_d is the solution set corresponding to the definition, and S_i is the solution set corresponding to other features. There must be the following relationship, given by Formula (7):

$$S_i \in S_d. \quad (7)$$

The solution set of S_i is a subset of the solution set of S_d . This also shows why the model based on DNA rules is more adaptable and can work even in out-of-sample conditions.

Just like the fruits that are grown without pesticides and fertilizers are more fragrant. Our model does not require complex human knowledge and statistics, so we call the model generated by an iterative process based on DNA rules corresponding to the definition as natural intelligence (NI).

Acknowledgements Project supported by Yuyao Zhejiang University Robotics Research Institute Project (K11805). We thank Kuixiang Xiong for his help in the experiment.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Boeing G (2016) Visual analysis of nonlinear dynamical systems: chaos, fractals, self-similarity and the limits of prediction. *Soc Sci Electron Pub* 4:37
- Cairns J, Overbaugh J, Miller S (1988) The origin of mutants. *Nature* 335:142–145. <https://doi.org/10.1038/335142a0>
- Cencini M, Ceconi F, Vulpiani A (2009) Chaos: from simple models to complex systems. World Scientific, Singapore
- Chen XW, Lin X (2014) Big data deep learning: challenges and perspectives. *IEEE Access* 2:514–525. <https://doi.org/10.1109/ACCESS.2014.2325029>
- Eckmann JP, Ruelle D (2008) Ergodic theory of chaos and strange attractors. *Rev Mod Phys* 57(3):617–656
- Gaier A, Ha D (2019) Weight agnostic neural networks
- George S, Robert M, Hao Y, Chih-hao H, Ethan D, Michael F, Stephan M (2012) Detecting causality in complex ecosystems. *Science* 338:496–500
- Hao Y, Richard JB, Sarah MG, Sue CHG, Chih-Hao H, Laura JR, Jon TS, George S (2015) Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proc Natl Acad Sci USA* 112:E1569
- Lorenz EN (2004) Deterministic nonperiodic flow. *J Atmos Sci* 20:130–141
- Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E (2015) Deep learning applications and challenges in big data analytics. *J Big Data* 2:1
- Nehmzow U, Walker K (2005) Quantitative description of robot–environment interaction using chaos theory. *Robot Autonom Syst* 53:177–193
- Nguyen A, Yosinski J, Clune J (2015) Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, pp. 427–436. <https://doi.org/10.1109/CVPR.2015.7298640>
- Pearl J (2000) Causality: models, reasoning, and inference. Cambridge University Press, New York
- Ritter S, Barrett DGt, Santoro A, Botvinick M (2017) Cognitive psychology for deep neural networks: a shape bias case study. international conference on machine learning
- Ruelle D (1980) Strange attractors. *Math Intell* 2:126–137
- Shwartz-Ziv R, Tishby N (2017) Opening the black box of deep neural networks via information
- Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, Schrittwieser J, Panneershelvam V, Lanctot M (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* 529:484–489
- Turing AM (1952) The chemical basis of morphogenesis. *Philos Trans R Soc B Biol Sci* 237:37–72
- Wolf A (1986) Quantifying chaos with Lyapunov exponents. *Nonlinear Science: Theory and Applications*: 273–290, Manchester Univ. Press, Manchester
- Wolfram S (1984) Cellular automata as models of complexity. *Nature* 311:419–424
- Wolfram S (2018) Logic, explainability and the future of understanding. <https://blog.stephenwolfram.com/2018/11/logic-explainability-and-the-future-of-understanding/>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.