

On Scientific Discovery – A Decision-Machine Approach

Lizhi Xin,^{1,*} Kevin Xin,^{1,†}

¹XINVISIONQ, INC., Chicago, Illinois 60607, USA

ABSTRACT. Throughout the history of scientific discovery, the question of could a machine be able to find the laws of nature directly from observed data without relying on any prior information has been unimaginable until the emergence of modern-day computing and Artificial Intelligence. We develop a framework as an operator operation, evaluation, and optimization for a decision-machine to conduct scientific discovery: both nature’s “behavior” and the decision-machine’s “actions” are modeled with a formalized system under Hilbert Space; three inductive rules are utilized to evaluate the decision-machine’s performance; and the evolutionary algorithm is applied to optimize the best way to reconstruct the historical data and effectively predict its future. A simulated random dataset is used to show that the decision-machine is able to reasonably reconstruct the experimental data and effectively predict the future. Crucially, our developed framework is a versatile and experimentally feasible tool for conducting scientific discovery by machine that has broad implications for forecasting, AI for science, and fundamental scientific discovery of the natural and social sciences.

I. INTRODUCTION.

The essence of scientific discovery is to seek out the laws that govern nature as effectively as possible; in which there are four main parts that comprise of all generic scientific discovery – nature, observer, measuring instrument, and the observed data. Nature is what is being studied, its infinite possibilities are the main objective of scientific discovery. The observers are who that’s conducting the studying, their proposed hypotheses are the subjective of scientific discovery. The measuring instrument is the tool of scientific discovery, and is the apparatus that directly interacts with nature to record data. Observed data is the medium of scientific discovery, the data that’s recorded by the apparatus is the main target for the observer when studying nature [1-3].

Under the current established scientific method, the goal of the observer is to find the laws of nature by proposing a hypothesis, formulating experiments, and validating the original hypothesis [4-6]. For these hypothesis’ to be considered sufficient enough to become a theory, they have to meet two simple criteria: 1) reconstruct the experimental data, and 2) make effective forecasts of future data.

There are three main parts for the observer when conducting scientific discovery: observed data X , proposed hypotheses H , and validation criteria V .

- 1) $X: \{x_1, x_2, \dots, x_n\}$: The observer’s medium of scientific discovery.
- 2) $H: \{h_1, h_2, \dots, h_n\}$: The proposed hypotheses of the observer.
- 3) $V: \{v_1, v_2, \dots, v_n\}$: The metric used to validate the proposed hypothesis.

With the rise of computing power and artificial intelligence, the question of whether a machine could discover the laws of nature has become how – in which we ask:

Could we develop a machine to find the laws of nature directly from observed data by means of data-driven machine learning without relying on any prior knowledge?

In this Letter, we propose a framework to develop a decision-machine to conduct scientific discovery in three distinct parts.

First, a formalized symbol framework to describe both nature’s “behavior” and the decision-machine’s “actions” is defined; second, follow three predetermined inductive rules to evaluate the decision-machine’s performance; and third, continually optimize the scientific discovery process with an evolutionary algorithm.

A simulated random-walk dataset is generated for the decision-machine to “study”. We show that the decision-machine is able to reconstruct its historical data with a MAE of 4.26 and produce 3 forecasts with MAE of 8.74, 6.9, and 6.43; and MAPE of 9.66, 7.06, and 6.82. Remarkably the decision-machine is able to produce a better MAE than traditional methods in reconstructing the historical data, and is able to produce a better MAE and MAPE on average when forecasting future data.

II. SETUP

We consider a generated one dimensional randomly fluctuating dataset; which is generated by a predefined program following two simple rules:

*Contact author: lizhi@xinvisionq.com

†Corresponding author: kevin@xinvisionq.com

- 1) Randomly generate 1 or -1; if 1 then the succeeding point will be greater than the preceding one; if -1 then the succeeding point will be lesser than the preceding one.
- 2) Randomly generate a number from $\{1 \sim 9\}$; this number will be the distance between the two subsequent (consecutive) data points.

The program generates 35 data points with the starting point x_0 of 100 as in (1).

$$x_0 = 100 \quad (1a)$$

$$x_{k+1} = x_k + \text{Rand}(-1,1) * \text{Rand}(\{m_i\}) \quad (1b)$$

Where $\text{Rand}(-1,1)$ represents the algorithm either randomly generates 1 or -1; $\{m_i\} = \{1,2,\dots,9\}$; $\text{Rand}(\{m_i\})$ represents randomly select a number from $\{1,2,\dots,9\}$.

This generated dataset is then used as the raw data the decision-machine will study, reconstruct, and subsequently produce effective forecasts of. The generated dataset is split into two respective portions – training and verify. Of the 35 points, 1-29 is used as training and 30-35 is used as verify data.

III. MODEL

For the data generated, and for any one-dimensional sequential data in general, there are two main characteristics – 1) the succeeding point either moves up or down relative to the preceding point; and 2) there is a certain distance (change of movement) between the successive points. The main problem for the decision-machine conducting scientific discovery is how to successfully find both the trend movement and the distance movement of nature; in which nature does not tend to laminate all the information regarding these two characteristics.

For human observers, traditional scientific discovery has strictly followed an entity – properties – structure approach, one which has led to the formulation of concepts of force, energy, and wave-function, etc.; this approach has widely used differential equations to formulate a function for describing natural phenomena. However, nature's "behavior" tends to be "chaotic", and the uncertainty of the real world makes it difficult to directly find a function that can explain all observed phenomena.

From an information perspective, the decision-machine will conduct scientific discovery abiding by an event – information – operation approach; events

*Contact author: lizhi@xinvisionq.com

†Corresponding author: kevin@xinvisionq.com

present the decision-machine with information not property. With event being what is studied, information being the message "passed along" between nature and the decision-machine, and operation being the series of "actions" performed by the decision-machine to learn more about nature.

Instead of directly attempting to formulate a function, we can transform the observed dataset into an event series, with each subsequent data point and the distance between successive points representing an event "performed" by nature as in (2).

$$\{e_1, e_2, \dots, e_n\} \quad (2a)$$

$$e_k = (q_1|q_2, \Delta x_k) \quad (2b)$$

Where $\{e_1, e_2, \dots, e_n\}$ are the corresponding event series of the observed data points $\{x_1, x_2, \dots, x_n\}$; e_k represents a given point in the event series; q_1 represents the succeeding point is greater than the preceding point (up), q_2 represent the succeeding point is less than the preceding point (down), and Δx_k is the distance between the two successive points.

For the decision-machine to study the dataset and simulate nature, there are also two main characteristics – 1) "guess" whether the data will trend upwards or downwards, and 2) calculate the specific distance between the successive points as best as possible; the decision-machine's "actions" to simulate nature's movement are represented as an action sequence in (3).

$$\{a_1, a_2, \dots, a_n\} \quad (3a)$$

$$a_k = (g_1|g_2, \Delta x'_k) \quad (3b)$$

Where $\{a_1, a_2, \dots, a_n\}$ is the corresponding action sequence that can be taken by the decision-machine; a_k represents a given point in the action sequence, g_1 represents the decision-machine "guesses" that the data trends up, g_2 represents the decision-machine "guesses" that the data trends down, and $\Delta x'_k$ is the distance between two successive points calculated by the decision-machine.

To model both nature's event and the decision-machine's "action" under one formalized framework, the first three axioms of quantum mechanics are utilized [7-10].

The two orthogonal possible states of an event e_k (up or down) and an action a_k taken by the decision-machine ("guess" up or down) can be represented by a superposed state in Hilbert Space.

Nature's event e_k becomes (4).

$$|e_k\rangle = c_1|q_1\rangle + c_2|q_2\rangle \quad (4)$$

Where q_1 represents upwards; q_2 represents downwards. $\omega_1 = |c_1|^2$ is the objective frequency that the event moves upwards; $\omega_2 = |c_2|^2$ is the objective frequency that the event moves downwards.

The decision-machine's action a_k becomes (5).

$$|a_k\rangle = d_1|g_1\rangle + d_2|g_2\rangle \quad (5)$$

Where g_1 represents the decision-machine "guesses" upwards; g_2 represents the decision-machine "guesses" downwards. $p_1 = |d_1|^2$ are the decision-machine's "degree of beliefs" that the event moves upwards; $p_2 = |d_2|^2$ are the decision-machine's "degree of beliefs" that the event moves downwards.

For the decision-machine to reconstruct the observed dataset, corresponding density operators must be defined.

The observable (density operator) of nature's event is in (6).

$$\rho_{\text{event}} = |e_k\rangle\langle e_k| = \omega_1|q_1\rangle\langle q_1| + \omega_2|q_2\rangle\langle q_2| + c_1c_2^*|q_1\rangle\langle q_2| + c_1^*c_2|q_2\rangle\langle q_1| \quad (6)$$

Where the first term denoting the event moves upwards; the second term denoting the event moves downwards. The "quantum interference term" (third and fourth terms) shows that for future events that have not happened yet the movement of the event can be superposed in a state of "both" upwards and downwards; however, this "superposed state" is not a physical representation that the event is "simultaneously" moving upwards and downwards, it only represents that the movement of the event is just undeterminable.

The observable (density operator) of the decision-machine's actions is in (7).

$$\rho_{\text{operation}} = |a_k\rangle\langle a_k| = p_1|g_1\rangle\langle g_1| + p_2|g_2\rangle\langle g_2| + d_1d_2^*|g_1\rangle\langle g_2| + d_1^*d_2|g_2\rangle\langle g_1| \quad (7)$$

Where the first term denoting the decision-machine "guesses" that the event will move upwards; the second term denoting the decision-machine "guesses" that the event moves downwards. The third and fourth terms shows that the decision-machine cannot

"determine" whether the event moves upwards or downwards, in which it has not "decided" to "take" a corresponding "action" yet.

ρ_{event} and $\rho_{\text{operation}}$ signifies the undetermined superposition state. Once it's "measured", an event actually takes place and the decision-machine takes an "action", a projection happens as (8) and (9).

$$\rho_{\text{event}} \rightarrow \rho'_{\text{event}} = \omega_1|q_1\rangle\langle q_1| + \omega_2|q_2\rangle\langle q_2| \quad (8)$$

$$\rho_{\text{operation}} \rightarrow \rho'_{\text{operation}} = p_1|g_1\rangle\langle g_1| + p_2|g_2\rangle\langle g_2| \quad (9)$$

The data series $\{x_1, x_2, \dots, x_n\}$ is reconstructed as in (10) with ρ_{event} that simulates nature's "behavior".

$$x_{k+1} = x_k + \text{Sign}(\rho_{\text{event}}) * |x_k - x_{k-1}| \quad (10a)$$

$$\text{Sign}(\rho_{\text{event}}) = \begin{cases} 1, & \rho_{\text{event}} = 0 \\ -1, & \rho_{\text{event}} = 1 \end{cases} \quad (10b)$$

Where $\rho_{\text{event}} = 0$ represents that the event moves upwards; and $\rho_{\text{event}} = 1$ represents that the event moves downwards.

$\rho_{\text{operation}}$ is utilized by the decision-machine to calculate nature's position $\{x'_1, x'_2, \dots, x'_n\}$ as in (11).

$$x'_{k+1} = x'_k + \text{Sign}(\rho_{\text{operation}}) * \overline{\Delta x} + \sum_{k=2}^N \text{Rand}(-1,1) \overline{\Delta^k x} / k \quad (11a)$$

$$\text{Sign}(\rho_{\text{operation}}) = \begin{cases} 1, & \rho_{\text{operation}} = 0 \\ -1, & \rho_{\text{operation}} = 1 \end{cases} \quad (11b)$$

Where $\overline{\Delta x}$ represents the average of the first-order distance of the event's position; $\text{Rand}(-1,1)$ represents the decision-machine either randomly generates -1 or 1; $\overline{\Delta^k x}$ represents average of the kth-order distance of the event's position; $\rho_{\text{operation}} = 0$ represents the decision-machine "guesses" that the event will move upward; $\rho_{\text{operation}} = 1$ represents the decision-machine "guesses" that the event will move downward.

The decision-machine applies $\rho_{\text{operation}}$ to perform an "action" that simulates nature's behavior by "deciding" whether the event will move upwards or downwards. The main goal of $\rho_{\text{operation}}$ is to "guess" as correct as possible compared to nature's ρ_{event} with the greatest confidence. For each event of nature e_k , the decision-machine's corresponding a_k should

*Contact author: lizhi@xinvisionq.com

†Corresponding author: kevin@xinvisionq.com

aim to be correct with the greatest confidence ($p_1 = p_2 = 1$). If this is achieved, then the decision-machine is able to simulate nature's events series perfectly every single time.

If the position series $\{x'_1, x'_2, \dots, x'_n\}$ calculated by the decision-machine perfectly overlaps with nature's position series $\{x_1, x_2, \dots, x_n\}$, then the decision-machine "discovers" the "kinetic laws" of nature.

IV. EVALUATION

We term the density operator $\rho_{\text{operation}}$ as the operation operator; $\rho_{\text{operation}}$ is utilized by the decision-machine to simulate nature's event ρ_{event} ; $\rho_{\text{operation}}$ can be constructed with the data set and operation set as in (12). For more details see Xin, et al [11-15].

$$\text{Data set: } \left\{ \begin{array}{l} H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ S = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}, D = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \\ T = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{bmatrix}, I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{array} \right\} \quad (12a)$$

$$\text{Operation set: } \{+, *, V\} \quad (12b)$$

The operation operator takes the final form of a logic tree. $\rho_{\text{operation}}$ traverses through each observed event by guessing whether the event will trend upwards or downwards with a certain degree of beliefs. In order to evaluate how well the decision-machine performs, three inductive rules (IR) are defined for it to follow:

- 1) Similarity degree (SD): how well the decision-machine's predicted result coincides with the actual event that happened.
- 2) Effective operational level (EOL): how confident and accurate the decision-machine predictions are.
- 3) Consistency: how consistent the decision-machine is when making predictions to assure relative effectiveness and objectiveness.

1. Similarity Degree

After the decision-machine traverses through the entire dataset and makes a respective "guess" at each point, the similarity degree is the ratio of all the correct predictions made by the decision-machine and all of the total predictions made as in (13).

$$SD = N_{\text{correct}} / N_{\text{total}} \quad (13)$$

*Contact author: lizhi@xinvisionq.com

†Corresponding author: kevin@xinvisionq.com

The similarity degree reflects how accurate the operation operator "guesses" of the event's movement direction is. If $SD = 1$ then the decision-machine can simulate nature completely; if $SD = 0$ then the decision-machine knows nothing about nature; most times SD will be in between 0 and 1.

2. Effective Operational Level

There needs to be an "incentive" for the decision-machine to "guess" the correct corresponding action of nature's "events"; in which a reward and deficit system can be introduced based on the four possible outcomes as in (14); ρ' is the complex system of nature's events and the decision-machine's actions.

$$\rho' = \rho'_{\text{event}} \otimes \rho'_{\text{operation}} = \omega_1 p_1 |q_1\rangle \langle q_1| \otimes |g_1\rangle \langle g_1| + \omega_1 p_2 |q_1\rangle \langle q_1| \otimes |g_2\rangle \langle g_2| + \omega_2 p_1 |q_2\rangle \langle q_2| \otimes |g_1\rangle \langle g_1| + \omega_2 p_2 |q_2\rangle \langle q_2| \otimes |g_2\rangle \langle g_2| \quad (14)$$

Where the first term denotes the decision-machine predicts correctly that the event is going upwards; the second term denotes the decision-machine predicts incorrectly as the event is going upwards; the third term denotes the decision-machine predicts incorrectly as the event is going downwards; the fourth term denotes the decision-machine predicts correctly that the event is going downwards.

At any given point, when the operation operator has "guessed" and compared it with the event that actually happened; in which there is only four possible outcomes as in (15).

$$r_k = \begin{cases} \omega_1 p_1 |x_k - x_{k-1}| \\ -\omega_1 p_2 |x_k - x_{k-1}| \\ -\omega_2 p_1 |x_k - x_{k-1}| \\ \omega_2 p_2 |x_k - x_{k-1}| \end{cases} \quad (15)$$

Where the first and fourth terms represent that the operation operator "guessed" correctly; and the second and third terms represent that the operation operator "guessed" wrong.

To ensure that the decision-machine is able to accumulate experience for every prediction made, it has to strictly follow the predefined system of reward and deficit; if correct then it is rewarded, otherwise a deficit is incurred [16-17].

The total expected returns by the decision-machine then becomes as in (16).

$$R_{\text{returns}} = \sum_{k=1}^N r_k \quad (16)$$

The maximum expected returns becomes when the decision-machine always “guesses” correctly at any given point with the greatest confidence as in (17).

$$\max R_{\text{returns}} = \sum_{\text{up}} \omega_1 |x_k - x_{k-1}| + \sum_{\text{down}} \omega_2 |x_k - x_{k-1}| \quad (17)$$

Where the first term represents that the decision-machine accumulates rewards when the event trends upwards; the second term represents that the decision-machine accumulates rewards when the event trends downwards.

The effective operational level is then defined as the ratio of the actual expected returns and the maximum expected return as in (18).

$$\text{EOL} = R_{\text{returns}} / \max R_{\text{returns}} \quad (18)$$

Effective Operational Level is used in conjunction with Similarity Degree to ensure that the operation operator $\rho_{\text{operation}}$ is able to “guess” correctly with the greatest degree of confidence.

If the decision-machine can simulate nature correctly with the greatest confidence, then EOL is 1; if the decision-machine is very confident in its “guesses” even though it simulates nature completely wrong, then EOL is -1. When EOL is 0, then neither a reward or deficit is incurred to the decision-machine; and when EOL is greater than 0, then that means the decision-machine is rewarded and subsequently found valuable information.

3. Consistency

A comparison to traditional statistical methods is made with the Mean Average Error (MAE). The MAE of traditional methods are defined in as (19).

$$\text{MAE}_{\text{mean}} = \sum_{k=1}^N |\bar{x} - x_k| / N \quad (19)$$

Where \bar{x} is the actual mean value of $\{x_k\}$ of the observed time series data.

The MAE of the decision-machine is defined as in (20).

*Contact author: lizhi@xinvisionq.com

†Corresponding author: kevin@xinvisionq.com

$$\text{MAE}_{\text{DM}} = \sum_{k=1}^N |x'_k - x_k| / N \quad (20)$$

Where x_k is the actual observed position, x'_k is the position calculated by the decision-machine.

In order to evaluate the consistent effectiveness of the decision-machine’s predictions, predetermined threshold levels of SD, EOL, and MAE are set. While these levels can be adjusted, in order to obtain the effective operation operator, we’ve set them as follows:

- 1) The similarity degree needs to be greater than the threshold value (SD>0.8).
- 2) The effective operational level needs to be greater than 0 (EOL>0.3).
- 3) The MAE_{DM} should be less than the MAE_{mean} .

V. OPTIMIZATION

In order for the decision-machine to conduct effective scientific discovery, the question that must be asked is:

How can an effective $\rho_{\text{operation}}$ be obtained where $\text{SD} > 0.8$, $\text{EOL} > 0.3$, and $\text{MAE}_{\text{DM}} < \text{MAE}_{\text{mean}}$?

We cannot hope for the decision-machine to be able to randomly generate a $\rho_{\text{operation}}$ and fulfill the predetermined levels of SD, EOL, and MAE by “pure luck”; just as Feynman said he didn’t believe a machine could find theories on its own [18].

The evolutionary algorithm [19-21] is used to optimize the process of finding the most effective operation operator that fulfills the consistency criteria. Without the continual optimization and refining done by the evolutionary algorithm, the decision-machine would just be a “blind guesser”.

The evolutionary algorithm first generates a population of 300 operation operators and calculates the fitness of each one. After the initial structural change of each operation operator through crossover and mutation, the population is then ordered by levels of “fitness”; the “fitter” ones are selected and “parented”, subsequently producing offspring; the next generation is then ordered again by levels of “fitness”, and the process of crossover, mutation, and selection is reinitiated. This leads to an iterative generation after generation of evolution, that hypothetically should eventually produce the “fittest” operation operator.

The evolutionary algorithm is suited for solving the optimization problems of complex and uncertain environments – the more uncertain it is the more of an advantage the evolutionary algorithm has. The evolutionary algorithm balances exploration and utilization very well, i.e. it's able to keep the existent most optimal solution (stored in the entire population as whole instead of in a single individual) as well as utilize crossover and mutation to explore even more optimal solutions; or in other words to utilize the already existent valuable information as well as build up on new unknown available information.

Throughout the process of obtaining the most effective operation, the decision-machine first starts off knowing nothing about nature, but throughout continuous evolution, the “fittest” operation naturally arises, which should be the one that fulfills the predetermined levels of SD, EOL, and MAE_{DM} .

The most effective operation operator $\rho_{operation}$ evolved by the evolutionary algorithm in this study as in (21).

$$\rho_{operation} = \left(\left(X \vee \left(((T + S) \vee D) * ((Z + (X + Z)) * (((Y \vee Y) \vee I) \vee H)) \right) \right) + H \right) \quad (21)$$

This effective operation operator outputted becomes the one that aids the decision-machine in reconstructing the historical data and producing effective forecasts.

VI. DATASET 1 – TRAINING

Datapoints 1-29 of the generated dataset by (1) are the training data that the decision-machine “studies”.

Figure 1 shows the graphical results produced; Table I shows the numerical values of the fitting results as produced by the decision-machine.

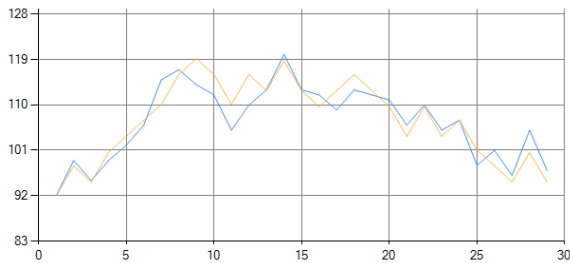


FIG 1. Fitting results graph produced by the decision-machine.

*Contact author: lizhi@xinvisionq.com

†Corresponding author: kevin@xinvisionq.com

Label	Q	X	A	P	X'
1		92			92
2	0	99	0	0.57	98
3	1	95	1	0.5	95
4	0	99	0	0.5	101
5	0	102	0	0.5	104
6	0	106	0	0.5	107
7	0	115	0	0.5	110
8	0	117	0	0.5	116
9	1	114	0	0.57	119
10	1	112	1	0.5	116
11	1	105	1	0.5	110
12	0	110	0	0.73	116
13	0	113	1	0.5	113
14	0	120	0	0.5	119
15	1	113	1	0.5	113
16	1	112	1	0.5	110
17	1	109	0	0.5	113
18	0	113	0	0.5	116
19	1	112	1	0.5	113
20	1	111	1	0.5	110
21	1	106	1	0.43	104
22	0	110	0	0.73	110
23	1	105	1	0.5	104
24	0	107	0	0.5	107
25	1	98	1	0.5	101
26	0	101	1	0.5	98
27	1	96	1	0.5	95
28	0	105	0	0.5	101
29	1	97	1	0.5	95

TABLE I. Numerical fitting results produced by the decision-machine.

In Figure 1, the blue line is the raw data; the yellow line is the reconstructed fitted data by the decision-machine.

In Table I, the columns are: column 1 are the labels; column 2 is the direction of the raw data (0 trend upwards, 1 trend downwards); column 3 are the positions of the raw data; column 4 are the decision-machine's “decisions” (0 “decide” upwards, 1 “decide” downwards); column 5 are the decision-machine's degree of beliefs; column 6 are the calculated positions by the decision-machine; and the numbers highlighted in red are the incorrect decisions made by the decision-machine.

VII. DATASET 2 – VERIFY

Datapoints 30-35 of the generated dataset by (1) are used as the verify dataset and are subsequently forecasted by the decision-machine. This dataset is termed verify because the decision-machine produces three forecasts of datapoints 30-35 without actually knowing what the exact values are; after the forecasts are made the predicted values are then compared to the actual generated points 30-35.

Figure 2 shows the graphical results of the three forecasts; Table II shows the numerical results of the three forecasts as produced by the decision-machine.

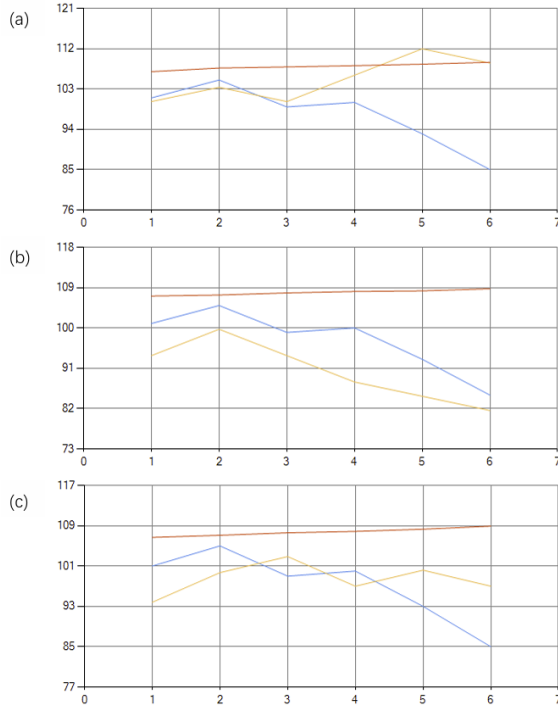


FIG 2. The three predictions (a)-(c) produced by the decision-machine; (a) Forecast 1; (b) Forecast 2; (c) Forecast 3.

Label	Position	Mean	F1	F2	F3	Average
30	101	106	100	94	94	106.9
31	105	106	103	100	100	107.4
32	99	106	100	94	103	107.8
33	100	106	106	88	97	108.1
34	93	106	112	85	100	108.4
35	85	106	109	82	97	108.9

TABLE II. Results of the verify dataset.

In Figure 2, (a), (b), and (c) are the three corresponding graphs of each forecast; the blue line is raw data; the

yellow line is the predicted position; and the red line is the average of 1000 forecasts by the decision-machine.

In Table II the columns are: column 1 are the labels; column 2 are the raw positions; column 3 are the traditional predictions; columns 4-6 F1-F3 are the three predictions produced by the decision-machine; column 7 is the average of 1000 forecasts.

VIII. RESULTS ANALYSIS

The SD, EOL, and MAE as calculated by the decision-machine and the comparison to traditional methods are shown in Table III; the MAE and MAPE of the traditional methods, and three forecasts produced by the decision-machine are shown in Table IV.

	SD	EOL	MAE
Decision-machine	0.86	0.42	4.26
Traditional	X	X	5.73

TABLE III. Evaluation metrics of the fitting results.

	MAE	MAPE
Traditional	8.83	9.61
Forecast 1	8.74	9.66
Forecast 2	6.9	7.06
Forecast 3	6.43	6.82

TABLE IV. MAE and MAPE of traditional methodologies and three decision-machine predictions.

In Table III, the MAE_{DM} of 4.26 is less than the traditional methods MAE_{mean} of 5.73; where the consistency conditions are met by the outputted $\rho_{operation}$ in (21); in which the SD is 0.86 and the EOL is 0.42.

In Table IV, the average MAE and MAPE of the three decision-machine's predictions are 7.35 and 7.84 respectively, both better than traditional methodology's MAE of 8.83 and MAPE of 9.61.

As seen in Table I, the four values highlighted in red reflects the four incorrect "guesses" made by the decision-machine; where the probability that the decision-machine "believes" the entity moves upwards is 54% on average; the probability that it "believes" the entity moves downwards is 49% on average; which leads to a slight increase in the average produced shown in Table II.

In Table II, the average of 1000 forecasts produced by the decision-machine as shown in the Average column is fairly close to the mean of 106 produced

*Contact author: lizhi@xinvisionq.com

†Corresponding author: kevin@xinvisionq.com

by traditional methods, which shows that the decision-machine's produced results encompass those of traditional statistical ensemble results.

When it comes to describing phenomena engaged in a random walk, traditional methods resort to using a statistical ensemble – attempting to produce a forecast for a single event is completely given up and only a probability distribution of where an entity might be is produced. By applying our methodology, the decision-machine is able to produce a forecast of the trajectory of a single event.

We'd like to emphasize that a so-called effective forecast of the future doesn't necessarily always mean it is an accurate forecast, because no one can predict what will happen in the future; an effective prediction is just means of one that complies with a consistent rule of induction, in which it performs reliable and reasonable predictions; the accuracy of a prediction can only be evaluated until after what it attempted to predict has actually happened, in which the original prediction is then compared to the actual event.

IX. CONCLUSIONS

We have developed a framework for the decision-machine to conduct scientific discovery automatically as an operator operation, evaluation, and optimization. Our proposed framework consists of three distinct parts: (1) an abstract formalism is the logic skeleton of the decision-machine to perform "deductive inference"; (2) the correlation between the abstract formalism and the observable phenomena is expressed by the three inductive rules; (3) the evolutionary algorithm is used to optimize the process of scientific discovery conducted by the decision-machine.

Using the example of a completely random-walk like one dimensional time series dataset, the decision-

machine is able to geometrically reconstruct any curve and effectively predict its future by self-learning and self-adapting with the evolutionary algorithm. Furthermore, compared to traditional methods, the decision-machine doesn't attempt to describe nature fully in terms of differential equations and don't rely on statistical ensembles, and are able to produce single-event forecasts.

Future studies will include expanding our decision-machine to other real-world datasets in the natural and physical sciences.

Lastly, we are not attempting to create a fully perfect omnipotent machine to accurately predict the future, but rather a self-learning and self-adapting machine to "deal with" nature's creative evolution that can cope with the problems that even the machine designer cannot foresee.

Scientific discovery isn't just about backtracking the past and to predict the future but more importantly it's about challenging the uncertain, ever-changing, modifiable nature. We shouldn't naively believe that nature is static and there is some simple final structure "sitting" somewhere for us to find, scientific discovery is more about finding the unknown of the unknown.

ACKNOWLEDGMENTS

This work received no funding.

DATA AVAILABILITY

The authors confirm that all data generated are included in the manuscript. Data is available upon request from the corresponding author K. X. on reasonable request.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest regarding the publication of this paper.

-
- | | |
|---|--|
| <p>[1] Kuhn, T. S. (1973). <i>The structure of Scientific Revolutions</i> Thomas S. Kuhn. Univ. of Chicago Press.</p> <p>[2] Popper, K. R., & Freed, J. (1959). <i>The logic of Scientific Discovery: Karl R. popper</i>. Hutchinson.</p> <p>[3] Strevens, M. (2022). <i>The knowledge machine: How an unreasonable idea created modern science</i>. Penguin Books.</p> <p>[4] Mach, E. (2019). <i>Science of Mechanics: A Critical and historical account of its development</i>. Forgotten Books.</p> | <p>[5] Whitehead, A. N. (1957). <i>The concept of nature</i>. University of Michigan Press.</p> <p>[6] Strogatz, S. (2020). <i>Infinite powers: How calculus reveals the secrets of the universe</i>. Houghton Mifflin Harcourt.</p> <p>[7] Von Neumann, J. <i>Mathematical Foundations of Quantum Theory</i>, Princeton, NJ: Princeton University Press (1932).</p> <p>[8] Jammer, M. <i>The Philosophy of Quantum Mechanics</i>, New York, NY: John Wiley & Sons (1974).</p> |
|---|--|

*Contact author: lizhi@xinvisionq.com

†Corresponding author: kevin@xinvisionq.com

- [9] Dirac, P.A.M., the Principles of Quantum Mechanics, Oxford University Press (1958).
- [10] Heisenberg, W., the Physical Principles of the Quantum Theory, Chicago, IL: The University of Chicago Press (1930).
- [11] Xin, L., Xin, H. Decision-making under uncertainty – a quantum value operator approach. *Int J Theor Phys* 62, 48, <https://doi.org/10.1007/s10773-023-05308-w> (2023).
- [12] Xin, L.; Xin, K.; Xin, H. 2023 On Laws of Thought—A Quantum-like Machine Learning Approach. *Entropy*, 25, 1213. (doi:10.3390/e25081213).
- [13] Xin, L. Z., & Xin, K. (2025). Is the Market Truly in a Random Walk? Searching for the Efficient Market Hypothesis with an AI Assistant Economist. *Theoretical Economics Letters*, 15, 895-903. <https://doi.org/10.4236/tel.2025.154049>.
- [14] Xin, L. Z.; Xin, H. 2022. Quantum Measurement: A game between observer and nature? <https://doi.org/10.48550/arXiv.2210.16766>.
- [15] Xin, L.; Xin, K. 2025. AI Assistant Scientist: Aiding scientific discovery with quantum-like evolutionary algorithm. <https://doi.org/10.51094/jxiv.1103>.
- [16] Devlin, K. J. (2010). The unfinished game: Pascal, Fermat, and the seventeenth-century letter that made the World Modern. Basic Books.
- [17] Savage, L.J.: The foundation of Statistics. Dover Publication Inc., New York (1954).
- [18] Feynman, Richard Phillips. (1967). The character of physical law. MIT. Press.
- [19] Holland, J., *Adaptation in Natural and Artificial System*, Ann Arbor, MI: University of Michigan Press (1975).
- [20] Koza, J.R., *Genetic programming, on the programming of computers by means of natural selection*, Cambridge, MA: MIT Press (1992).
- [21] Koza, J.R., *Genetic programming II, automatic discovery of reusable programs*, Cambridge, MA: MIT Press (1994).

*Contact author: lizhi@xinvisionq.com

†Corresponding author: kevin@xinvisionq.com