IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 13, 2021, accepted March 28, 2021, date of publication April 5, 2021, date of current version April 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3070925

Detecting Associations Based on the Multi-Variable Maximum Information Coefficient

TAOYONG GU[1], JIANSHENG GUO[1], ZHENGXIN LI[1,2], AND SHENG MAO[1]

[1]Equipment Management and UAV Engineering College, Air Force Engineering University, Xi'an 710051, China
[2]School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China

Corresponding author: Taoyong Gu (gutaoyong@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61502521.

ABSTRACT The maximum information coefficient (MIC) is a novel and widely-using measure of association detection in large datasets. The most outstanding feature of MIC is that it has both generality and equability. However, MIC can only deal with two variables and cannot precisely estimate coupling associations of multiple variables. In this paper, we propose an extension of MIC to deal with multi-variable datasets, called the multi-variable maximum information coefficient (MMIC). Some inherited and novel properties of MMIC are proved, including generality, equability, monotonicity, and subadditivity. We design an algorithm based on greedy stepwise strategy and upper confidence bound (UCB) for an approximate calculation of MMIC. The tests of MMIC on generated datasets and examples on real datasets are carried out to detect known and novel associations.

INDEX TERMS Data mining, association detection, information entropy, maximum information coefficient, upper confidence bound.

I. INTRODUCTION

In recent decades, data science has been rapidly developed. Large datasets which hold a lot of information have driven a novel idea "follow the data" [1]. Association detection is a classic, common, meaningful but still challenging work in the field of data analysis. It is not feasible or profitable to detect associations manually in many large and complex datasets. A relatively reasonable measure is needed to evaluate whether the variables are associated or not, and to what extent they are associated. In many cases, we do not know the type of associations contained in the datasets. The associations may be non-linear, non-monotonic, and cannot be expressed by a mathematical function.

In order to identify these complex associations, the maximum information coefficient (MIC) is proposed [2]. MIC is a measure of association based on mutual information entropy which has both generality and equability. Generality guarantees that MIC can capture varied types of associations. Equability guarantees that MIC can give similar scores to different types of associations with equal noise [3]–[6].

A large number of researchers have done a lot of research on MIC. The research can be classified in three categories: theoretical verification, calculation, and application.

In the research of theoretical verification, MIC is proved that it has both generality and equality [3]–[6]. Moreover, MIC is confirmed to Rényi's axioms [7] and three new axioms when variables are continuous [8].

In the research of calculation, the main efforts are concentrated on improving the calculation accuracy and the reducing the time complexity. The intelligent MIC (iMIC) [9] is proposed for optimizing the partition on the y-axis to get approximate scores of MIC with acceptable accuracy. SuperMIC [10] uses MapReduce framework to improve the calculating efficiency of MIC. The improved algorithm for approximation of MIC (IAMIC) [11], [12] improves the accuracy of MIC by searching more optimal partitions on the y-axis than equipartition. The most commonly used toolkit for MIC is the Maximal Information-based Nonparametric Exploration (MINE) [2] which is compatible with C++, Python, and R. There are also many data independence testing tools based on MIC, including testforDEP [13] for R, MICtools [14] for Python. RapidMic [15] is a cross-platform tool for the rapid calculation of MIC based on parallel technology.

In practical applications, MIC has advantages dealing with complex associations which cannot be measured precisely by classic methods. MIC plays a role in association detection and data driven feature selection in many fields (see Table. 1), including computer science [16]–[19], biology [20]–[22], environmental science [23]–[25], medicine [26], [27], and

The associate editor coordinating the review of this manuscript and approving it for publication was Paul D. Yoo.

TABLE 1. Number of related papers of MIC in the different fields (Web of Science, 2011-2020).

Application field	Number of papers (Web of Science, 2011-2020)
computer science	24
biology	19
environmental science	12
medicine	11
genetics	10
engineering	14
energy	9
others	14

genetics [28]. Datasets in these fields have a similar feature: there are some associations in the datasets, but the associations are difficult to be expressed by accurate models or mathematical formulas. In the fields of computer science, MIC is used for feature selection [16], [17], model evaluation [19] combined with machine learning methods. Applications in environmental science mainly focus on the analysis of hydrological activities [23], [24] and geology activities [25]. Biology mainly uses MIC on the analysis of neuroscience [20]–[22]. Medicine and genetics explore the associations of genes [28] and diseases [26], [27] based on MIC. MIC is mainly used in feature extraction [29], prediction [30], and estimation [31] based on energy datasets. MIC is also used in the fault diagnosis [32], [33] and prediction [34] of complex systems and equipment. Besides, MIC plays a role in engineering [35], astronomy [36], chemistry [37], sociology [38], optics [39], and agronomy [40]. Overall, MIC is an effective association detection method for feature selection [41], classification [18], and prediction [42].

Although MIC has made great achievements in theory, calculation, and application, there is still a fundamental problem worth expanding research: how to evaluate multi-variable associations based on MIC. The most obvious defect of MIC is that it can only deal with two variables. In many application scenarios, we not only want to detect two-variable associations, but also want to detect associations of multiple variables. Furthermore, multi-variable associations cannot be derived precisely by pairwise combinations of two-variable associations based on MIC. Therefore, an extension of MIC which enables it to deal with multi-variable associations is meaningful.

The contributions of this paper can be summarized as follows:

- We design a measure based on MIC, and name it as the multi-variable maximum information coefficient (MMIC). Then we prove generality, equability, and several intuitive multi-variable properties of MMIC.
- We propose a calculation algorithm of MMIC based on greedy stepwise strategy and upper confidence bound (UCB) [43], and analyze the time complexity of the algorithm. In order to verify the accuracy and the performance of the algorithm, we test MMIC on a series of generated datasets with different types of association, relative noises, and dimensions.
- We apply MMIC on real datasets to verify the feasibility and effectiveness of MMIC.

II. RELATED WORKS

Based on MIC, there has been some similar work of multi-variable extension, including the three-variable maximum information coefficient (3MIC) [44] and the bisecting k-means clustering maximum information coefficient (BKM-MIC) [45]. We will brief the ideas of them and analyze their advantages and disadvantages.

A. THE MAXIMUM INFORMATION COEFFICIENT

For a finite dataset $D \subset R^2$ containing two variables, $\mathbf{x}$ and $\mathbf{y}$ partitioned into dx and dy blocks. If the grid size dx and dy are fixed, we can get the maximum mutual information entropy of the dataset D with a fixed grid G (denoted as $I\left(D|_G\right)$) [2].

$$I^*\left(D, dx, dy\right) = \max I\left(D|_G\right) \tag{1}$$

The mutual information entropy for each grid size is normalized to the range of $[0, 1]$ and put into the counting matrix M.

$$M\left(D\right)_{dx,dy} = {I^*\left(D, dx, dy\right)}\Big/{\log \min\left\{dx, dy\right\}} \tag{2}$$

Based on the traversal of all feasible grid sizes dx and dy, we can get the maximum element in M as MIC. There is an upper limit of grid size for preventing overfitting which is denoted by $B\left(n\right)$. The data volume is denoted by n.

$$MIC\left(D\right) = \max_{dx*dy<B(n)} \left\{M\left(D\right)_{dx,dy}\right\} \tag{3}$$

B. THE THREE VARIABLE MAXIMUM INFORMATION COEFFICIENT

3MIC adopts a different calculating method of information coefficient. For a finite dataset $D \subset R^3$ with data volume n, each element of counting matrix M with grid size dx, dy, dz is calculated based on mutual and conditional information entropy.

$$\begin{aligned} I\left(D, dx, dy, dz\right) &= I(\mathbf{x}; \mathbf{y}; \mathbf{z}) = H(\mathbf{x}) - H(\mathbf{x} \mid \mathbf{z}) \\ &\quad + H(\mathbf{z} \mid \mathbf{x}, \mathbf{y}) - H(\mathbf{z} \mid \mathbf{y}) \end{aligned} \tag{4}$$

Except the calculation of $I\left(D, dx, dy, dz\right)$, the rest of 3MIC's definition and calculation are the same as MIC.

$$I^*\left(D, dx, dy, dz\right) = \max I\left(D|_G\right) \tag{5}$$

$$M\left(D\right)_{dx,dy,dz} = {I^*\left(D, dx, dy, dz\right)}\Big/{\log \min\left\{dx, dy, dz\right\}} \tag{6}$$

$$3MIC(D) = \max_{dx*dy*dy<B(n)} \left\{M\left(D\right)_{dx,dy,dz}\right\} \tag{7}$$

The advantage of 3MIC is that it can be calculated under a relatively low time complexity. However, 3MIC has some defects of mathematical properties.

- 3MIC does not conform to Rényi's axioms [7]. The scores of 3MIC are not always in [0,1]. For example,

when $\mathbf{x} = \mathbf{y}$, $\mathbf{x}$ and $\mathbf{z}$ are independent at the meanwhile, $I(\mathbf{x}; \mathbf{y}; \mathbf{z}) \approx I(\mathbf{z}) - I(\mathbf{x}) - I(\mathbf{y})$. In this case, 3MIC can be negative.
- 3MIC is not an accurate multi-variable MIC, but a roughly estimating algorithm based on MIC of each two variables. Besides, 3MIC is not proved to have generality and equality.

C. THE BISECTING K-MEANS MAXIMUM INFORMATION COEFFICIENT

BKM-MIC is a multi-variable MIC calculating method based on a kind of dimensional greedy stepwise strategy. The core idea of BKM-MIC can be concluded as: (1) calculating BKM-MIC for n variables, (2) fixing the grid of BKM-MIC, (3) adding one new variable and calculating BKM-MIC for $n + 1$ variables.

For example, if we want to calculate $BKM\text{-}MIC\ ([\mathbf{x_1}, \mathbf{x_2}], \mathbf{y})$, we can calculate $MIC\ (\mathbf{x_1}, \mathbf{x_2})$, fix the grid G_X, then calculate $BKM\text{-}MIC\ ([\mathbf{x_1}, \mathbf{x_2}], \mathbf{y})$.

$$I^*(D, G_X, dy) = \max I\left(D|_{G_X, G_y}\right) \tag{8}$$

$$M(D)_{dy} = I^*(D, G_X, dy)\big/\log\min\{dX, dy\} \tag{9}$$

In the equations, dX is a fixed value decided by G_X.

$$BKM\text{-}MIC(D) = \max_{dX * dy < B(n)} \left\{M(D)_{dy}\right\} \tag{10}$$

BKM-MIC is compatible with MIC and ranges from 0 to 1. However, BKM-MIC still has some flaws to be improved:

- The newly added variables cannot directly affect the already fixed grid. The coupling associations may not be fully detected.
- Premature fixation of grid size makes it difficult to capture data feature adequately.

III. THE MULTI-VARIABLE MAXIMUM INFORMATION COEFFICIENT

To guarantee the integrity and rigor of MIC for multiple variables and improve the rationality and feasibility, we propose MMIC. In this section, we will discuss MMIC from three aspects: definition, properties, and calculation.

A. DEFINITION OF THE MULTI-VARIABLE MAXIMUM INFORMATION COEFFICIENT

The definition of MMIC can be extended from MIC fundamentally. Given a finite dataset $D \subset R^m$, variables in D are divided into two groups, $\mathbf{X}$ and $\mathbf{Y}$. MMIC is the information coefficient of the "best" grid. The information coefficient is the normalized mutual information entropy decided by the count of samples in each box. The "best" means that the grid is corresponding to the maximum normalized mutual information entropy. The number of columns in $\mathbf{X}$ and $\mathbf{Y}$ can be arbitrary, denoted as m_x and m_y, $m = m_x + m_y$. When $m_x = m_y = 1$, MMIC can give the same score as MIC to keep compatibility with MIC.

The i-th $\mathbf{X}$ or $\mathbf{Y}$ variable is partitioned to dx_i or dy_i parts, and the size of grid G is denoted as $dG = \left[dx_1, \ldots, dx_{m_x}, dy_1, \ldots, dy_{m_y}\right]$. For each grid size dG, we can get the grid which is corresponding to the maximum mutual information entropy of $\mathbf{X}$ and $\mathbf{Y}$ as $M\left(D|dG\right) = \max I\left(D|_G\right)/\log\min\{dX, dY\}$. $dX = \prod_{i=1,2,\ldots m_x} dx_i$ and $dY = \prod_{i=1,2,\ldots m_y} dy_i$ are the products of grid size of $\mathbf{X}$ and $\mathbf{Y}$. $I\left(D|_G\right)$ is the mutual information entropy of the grid. Each element in the counting matrix M is the result of the "best" grid of different size. MMIC is the maximum element of the matrix M.

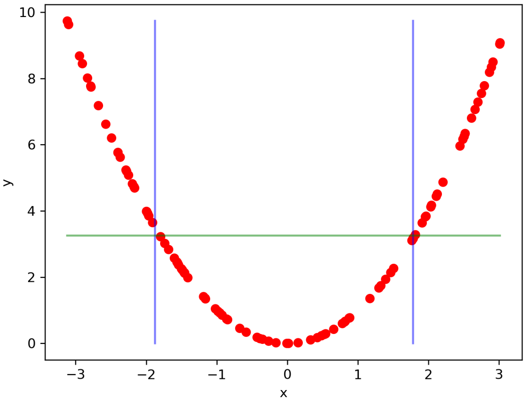


FIGURE 1. Grid of two-variable maximum information coefficient (noiseless quadratic function, calculated by MMIC of two variables, MMIC = 1.00).

The definition of MMIC can be summarized as follows.

$$I^*(D, dX, dY) = \max I\left(D|_G\right) \tag{11}$$

The mutual information entropy $I^*(D, dX, dY)$ is normalized and stored in counting matrix M.

$$M(D)_{dX,dY} = I^*(D, dX, dY)\big/\log\min\{dX, dY\} \tag{12}$$

$$MMIC\ (D) = \max_{dX * dY < B(n)} \left\{M(D)_{dX,dY}\right\} \tag{13}$$

For example, MIC can be regarded as the "best" grid of two-dimensional space like Fig. 1, and MMIC of three variables can be regarded the "best" grid of three-dimensional space like Fig. 2.

B. PROPERTIES OF THE MULTI-VARIABLE MAXIMUM INFORMATION COEFFICIENT

1) PROPERTIES INHERITED FROM THE MAXIMUM INFORMATION COEFFICIENT

MMIC reserves all good mathematical properties of MIC, including symmetry, normalized range, generality, and equability.

- Symmetry: $\text{MMIC}\,(\mathbf{X}, \mathbf{Y}) = \text{MMIC}\,(\mathbf{Y}, \mathbf{X})$ and $MMIC\ ([X_1, X_2], Y) = MMIC\ ([X_2, X_1], Y)$.
Since MMIC is based on mutual information, the symmetry of mutual information entropy guarantees the symmetry of MMIC. Therefore, the score of MMIC is not affected by the order of variables.
- Normalized range: The scores of MMIC are always in [0, 1].

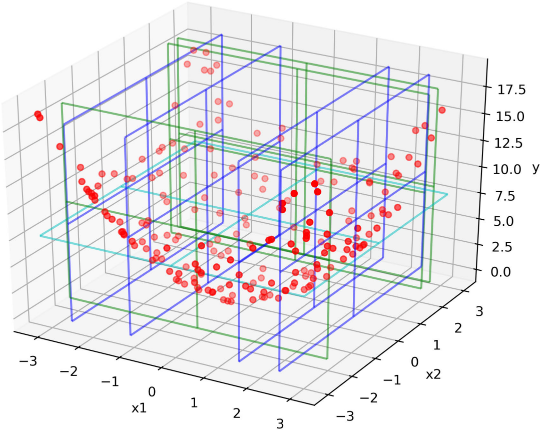


FIGURE 2. **Grid of three-variable maximum information coefficient (noiseless quadratic function, calculated by MMIC of three variables, MMIC = 1.00).**

MMIC tends to be 1 when at least one variable $\mathbf{y}$ in $\mathbf{Y}$ keeps a noiseless association of $\mathbf{X}$, and at least one variable $\mathbf{x}$ in $\mathbf{X}$ keeps a noiseless association of $\mathbf{Y}$. MMIC tends to 0 when each variable $\mathbf{y}$ in $\mathbf{Y}$ and each variable $\mathbf{x}$ in $\mathbf{X}$ are independent with sufficient samples. The premise, sufficient samples, is set to prevent random associations due to the effect of small sample. This premise is more prominent when the dimension of data is higher. Higher dimension brings more boxes partitioned by the grid and fewer points in each box. Therefore, the data samples are diluted by dimension, and MMIC is more likely to capture some random associations.

- Generality and equability.
  We will prove that MMIC can give similar scores for different types of association with equal noises by experimental verification. Besides, we will discuss the relationship between MMIC and R^2. MMIC is roughly equal to R^2 when the associations can be expressed by mathematical functions (see experiments and results in Section IV).

2) PROPERTIES FOR MULTIPLE VARIABLES

Besides the mathematical properties inherited from MIC listed above, MMIC also has some intuitive properties for multiple variables, including monotonicity and subadditivity. We will prove these two properties as follows.

Theorem 1: Monotonicity of MMIC

$$\mathrm{MMIC}\left(\mathbf{X_1},\mathbf{Y}\right) \leq \mathrm{MMIC}\left(\left[\mathbf{X_1},\mathbf{X_2}\right],\mathbf{Y}\right)$$

Proof: More information usually leads to better choices. Monotonicity can be deducted based on the definition of MMIC.

Denote the partitioning grid of MMIC $(\mathbf{X_1},\mathbf{Y})$ by $G_1 = \left[G_{\mathbf{X_1}}, G_{\mathbf{Y}}\right]$. $G_{\mathbf{X_1}}$ is the grid of $\mathbf{X_1}$, and $G_{\mathbf{Y}}$ is the grid of $\mathbf{Y}$. Presume that $\mathbf{X_2}$ is not partitioned. Therefore, $G_{\mathbf{X_2}}$ is empty. Since the grid $G = \left[G_{\mathbf{X_1}}, G_{\mathbf{X_2}}, G_{\mathbf{Y}}\right]$ is corresponding to MMIC $([\mathbf{X_1},\mathbf{X_2}],\mathbf{Y})$, the scores of all other grids are smaller than it, including the grid G_1. Therefore, MMIC $(\mathbf{X_1},\mathbf{Y}) \leq$ MMIC $([\mathbf{X_1},\mathbf{X_2}],\mathbf{Y})$. ■

Theorem 2: Subadditivity of MMIC

$$\mathrm{MMIC}\left(\mathbf{X_1},\mathbf{Y}\right) + \mathrm{MMIC}\left(\mathbf{X_2},\mathbf{Y}\right) \geq \mathrm{MMIC}\left(\left[\mathbf{X_1},\mathbf{X_2}\right],\mathbf{Y}\right)$$

Proof: MMIC can be regarded as an information extracting method based on mutual information entropy. Mutual information entropy conforms to the Jensen's Inequation $\mathrm{H}\left(\mathbf{X_1},\mathbf{Y}\right)+\mathrm{H}\left(\mathbf{X_2},\mathbf{Y}\right) \geq \mathrm{H}\left(\left[\mathbf{X_1},\mathbf{X_2}\right],\mathbf{Y}\right)$. Therefore, MMIC keeps subadditivity of mutual information entropy. The equal sign is satisfied only if $\mathbf{X_1}$ and $\mathbf{X_2}$ are dependent. ■

C. CALCULATION OF MULTI-VARIABLE MAXIMUM INFORMATION COEFFICIENT

The most challenging work of MMIC is not the theoretical derivation, but the calculation. The calculation of MIC is time-consuming when dealing with complex associations and large amounts of data samples. Complex associations lead to more different partitioning grids which needed to be calculated and compared. Large amounts of data samples lead to more time cost of information entropy calculation of each grid. MMIC introduces the third factor of time complexity: the dimension. Dimensional explosion heavily exacerbates the disaster of time complexity.

1) CALCULATION STRATEGIES

Almost all calculating methods of MIC are approximate algorithms balancing accuracy and time complexity. Like MIC, calculating the exact value of MMIC is not feasible in most cases. In order to weigh accuracy and time complexity, we design and implement three calculation strategies: limitation of grid, greedy stepwise strategy and upper confidence bound (UCB).

a: LIMITATION OF Grid

The first strategy is setting reasonable limitations of the maximum grid size and the potential partitioning positions. The maximum grid size can be interpreted as the model capacity of association. In theory, MMIC can detect associations of any complexity. However, in practice, it can only accurately measure associations below the model capacity. This limitation is not only designed to reduce the time complexity of calculation, but also designed to prevent recognizing associations caused by coincidence of small sample. Different from the limitation of the maximum grid size, the limitation of potential partitioning positions is merely intended for simplicity of calculation. In tests, we find that the impact on calculation accuracy caused by the limitation of the grid can be ignored, and most of the interesting and common associations can be measured with tolerable error.

b: GREEDY STEPWISE STRATEGY

The second strategy is a kind of greedy stepwise strategy including two steps: initialization and optimization. In initialization, a grid is generated based on an initial strategy. In most cases, grids based on uniform partition and two-variable MIC can be expedient choices. After obtaining the initial grid, the grid is adjusted to get the mutual information entropy maximum by a kind of neighborhood searching strategy: finding the maximum positive gain $I(D|G^*) - I(D|G)$ and replacing G with G^*.

c: UPPER CONFIDENCE BOUND

The third strategy is UCB. UCB is a searching strategy which can balance exploration and exploitation [43]. The calculation of MMIC is a classic problem of solution searching: selecting optimal combinations from candidate items under certain conditions. In the calculation of MMIC, we not only need to select the partitioning position, but also need to select the partitioning variable. Therefore, the selection based on UCB maintains an average gain for each variable as:

$$\hat{\mu}_i = \text{average}\left(\text{gain}\left(V_i\right)\right) = \frac{\sum \text{gain}\left(V_i\right)}{T_i}. \tag{14}$$

In variable selection, the average gain $\hat{\mu}_i$ and the explorations times T_i are weighted by UCB.

$$UCB_i = \hat{\mu}_i + \eta \ln \sqrt{\frac{T}{T_i}} \tag{15}$$

According to our test, $\eta = 0.1$ is appropriate in most cases. For the calculation of MMIC, UCB has two advantages:

- UCB can reduce the randomness of the calculation, and enhance the robustness in the grid searching.
- UCB can not only give the value, but also give the confidence.

2) ALGORITHM AND TIME COMPLEXITY

Based on the strategies listed above, to balance calculation accuracy and time complexity of MMIC, there are four adjustable parameters according to data dimension and volume in Algorithm. 1: α, β, γ and θ. α is a parameter inherited from MIC [2] to control the maximum grid size. β is designed for controlling the potential partitioning position. Besides, γ and θ are designed for controlling the number of iterations corresponding to max_iter(global) and max_iter(local) in pseudo code respectively.

To get rid of the influence of data type and programming language, we define that the complexity of calculating information coefficient of a fixed grid is 1. When the amount of data samples is n and the dimension of data samples is m, based on the maximum grid size [2] and the potential partitioning position, the number of feasible grid size is approximately $m^\alpha \cdot 2^n$. The amounts of searching space for each global iteration and local iteration are $\beta \log_2 n$ and $\theta \gamma m$ respectively. Therefore, for each feasible grid size, there are about $\beta\gamma\theta m \log_2 n$ grids being compared. Overall, the theoretical time complexity of Algorithm 1 is approximately $O\left(m^\alpha \cdot 2^n \cdot \beta\gamma\theta m \log_2 n\right)$ (see Fig. 3).

Algorithm 1 Calculation of MMIC Based on Greedy Stepwise Strategy and UCB

```
Input: dataset $D$, dividing to variables $\mathbf{X}$ and $\mathbf{Y}$
Output: MMIC, best grid
  for feasible grid sizes $dG$ do
      Initialize uniform grid $G$ with equal density
      for iter(global) from 1 to max_iter(global) do
          Select variable $V_i$ from $\mathbf{X}$ and $\mathbf{Y}$ by UCB (see (14) and (15))
          for iter(local) from 1 to max_iter(local) do
              for all $G'$ from $G$ and neighborhood($G$) do
                  $G^* \leftarrow \arg\max_{G'} I\left(D|_{G'}\right)$
                  gain $\leftarrow I(G^*) - I(G)$
              end for
              if gain $> 0$ then
                  $G \leftarrow G^*$
                  $T_i \leftarrow T_i + 1$
                  $\hat{\mu}_i = \left(\hat{\mu}_i T_i + \text{gain}\right) / (T_i + 1)$
              end if
          end for
      end for
      $M(D)_{dX,dY} \leftarrow \max I_G / \log \min\{dX, dY\}$
  end for
  return $MMIC(D) \leftarrow \max M(D)$, $best\ grid \leftarrow G^*$
```

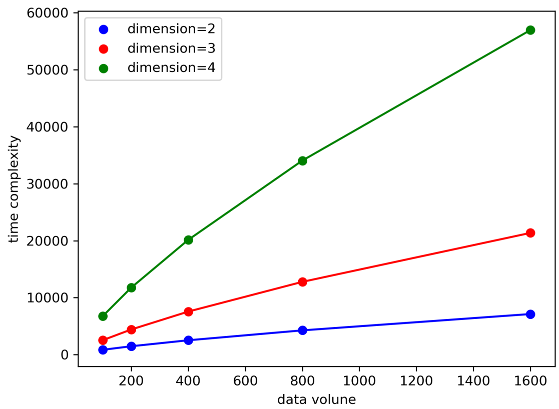


FIGURE 3. Time complexity of MMIC of two, three, and four variables.

IV. EXPERIMENTS AND RESULTS

A. TEST ON GENERATED DATASETS

To prove that MMIC is as effective as MIC, we generate datasets based on combining function with different dimensions, data volumes, and relative noises. The datasets contain non-linear, non-monotonic associations (see Table. 2). All the programs are running on a laptop which contains an Intel Core i7-8750H CPU and 16GB memory. The algorithm is coded in pure Python and posted on https://github.com/GuTaoyong/The-Multi-Variable-Maximum-Information-Coefficient. The default values of controlling parameters are $\alpha = 0.6$, $\beta = 2$, $\gamma = 1$, and $\theta = 1$.

TABLE 2. **Types of function in tests.**

Function type	Function expression	Dimension
sum	$y = \sum_{i\in\{1,\ldots,m\}} x_i$	$m \in \{1,2,3\}$
sum of power	$y = \sum_{i\in\{1,\ldots,m\}} x_i^2$	$m \in \{1,2,3\}$
sum of exponent	$y = \sum_{i\in\{1,\ldots m\}} 2^{x_i}$	$m \in \{1,2,3\}$
product	$y = \prod_{i\in\{1,\ldots m\}} x_i$	$m \in \{1,2,3\}$
product of power	$y = \prod_{i\in\{1,\ldots m\}} x_i^2$	$m \in \{1,2,3\}$
product of exponent	$y = \prod_{i\in\{1,\ldots m\}} 2^{x_i}$	$m \in \{1,2,3\}$
sine of sum	$y = \sin\left(\sum_{i\in\{1,\ldots m\}} x_i\right)$	$m \in \{1,2,3\}$

TABLE 3. **MIC and MMIC of noiseless function associations.**

Function type	MIC [2]	MMIC (two variables)	MMIC (three variables)
sum	1.00	1.00	1.00
sum of power	1.00	1.00	1.00
sum of exponent	1.00	1.00	1.00
product	1.00	1.00	1.00
product of power	1.00	1.00	1.00
product of exponent	1.00	1.00	0.98
sine of sum	1.00	0.92	0.95
random	0.18	0.15	0.16

In the tests, the calculation accuracy of MMIC is satisfactory. The scores of MMIC of noiseless functions are near 1, very close to that of MIC. The worst result is 0.92 which is acceptable with the limited computational resources (see Table. 3). More detailed results can be seen in Table. 7 and 8 in Section V. The association detection ability of MMIC can be described as equivalent to MIC.

In terms of calculating efficiency, when the data volume is 200, MMIC of two-variable, three-variable, and four-variable data can be calculated in less than 20 seconds, one minute, and five minutes respectively. It is proved that even on a lightweight computing device, the calculation of MMIC is still feasible.

In addition, we compare the results of MMIC with the goodness of fit (R^2). Some conclusion can be drawn from the tests:

- MMIC is roughly equal to R^2, and the changing trends of MMIC and R^2 are consistent (see Fig. 4). This means that MMIC can do roughly well as R^2 without the preset of association types.
- In most two-variable cases, MMIC is less than R^2. It is consistent with the conclusion that R^2 is the approximate theoretical upper bound of MIC [2]. However, in some three and four-variable, or highly noisy two-variable cases, this relationship may change (see Fig. 4). This is an overfitting and underfitting dilemma in nonparametric association detection, including MIC and MMIC. Unlike R^2, we do not preset the type of associations in the calculation of MMIC. Therefore, MMIC may detect random associations when the grids are too dense, and miss associations when the grids are too sparse.

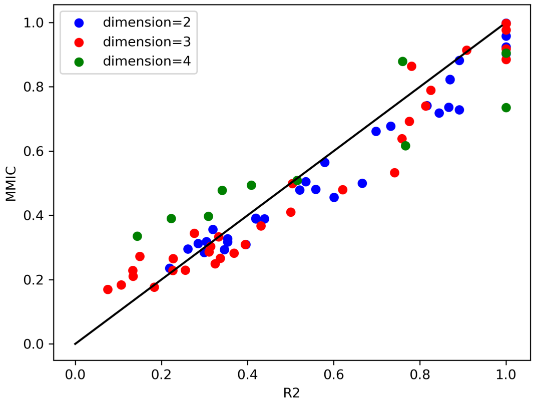


FIGURE 4. **MMIC and R^2 of two, three, and four variables in generated datasets.**

These two conclusions of MMIC are also consistent with MIC. Therefore, the inheritance between MMIC and MIC is further verified. More detailed experimental results are presented in section V.

B. EXAMPLES ON REAL DATASETS

Besides the tests on generating datasets, we also use MMIC to detect known and unknown associations based on real datasets, including datasets from the World Health Organization (WHO) [46] and the National Climatic Data Center (NCDC) [47].

1) EXAMPLES ON WHO DATASET

The WHO dataset contains 355 indicators of 202 countries and regions. The purpose of the examples on the WHO dataset is to get the most relevant indicator sets of a certain indicator. Besides the scores of MMIC, we can also get the partitioning grids of MMIC which may have some reference value.

Population growth rate is an issue of widespread concern and frequent discussion. Therefore, we use MMIC to analyze the associations of population growth rate and other indicators based on the dataset.

Firstly, we calculate MMIC of each two variables to determine the range of MMIC and roughly judge whether the variables are associated or not. Population annual growth rate is selected as **Y**. According to the monotonicity and subadditivity of MMIC, variables with larger MMIC are given preference in the selection of **X**. The calculation of two-variable MMIC shows that, besides the other population indicators (population growth and urban population growth), the most relevant indicators are population living below the poverty line, children per woman, contraceptive prevalence, and under-5 mortality rate (see Table. 4).

TABLE 4. Most relevant indicators of population annual growth rate evaluated by two-variable MMIC.

Order	Column name in dataset	MMIC
1	Population_growth	0.89
2	Urban_population_growth	0.74
3	Population living below the poverty line (% living on <US$1 per day)	0.73
4	Children_per_woman	0.70
5	Contraceptive prevalence (%)	0.70
6	Under-5 mortality rate (Probability of dying aged <5 years per 1 000 live births)	0.65

TABLE 5. Most relevant indicators of population annual growth rate evaluated by three-variable MMIC.

Order of variables in Table 4	MMIC	R^2	Pearson	Spearman	Kendall
$[3, 4]$	0.89	0.75	0.87	0.83	0.78
$[4, 5]$	0.88	0.39	0.62	0.69	0.49
$[3, 5]$	0.87	0.25	0.50	0.49	0.37
$[3, 6]$	0.80	0.78	0.78	0.81	0.76
$[4, 6]$	0.79	0.67	0.71	0.75	0.68
$[5, 6]$	0.77	0.45	0.67	0.72	0.52

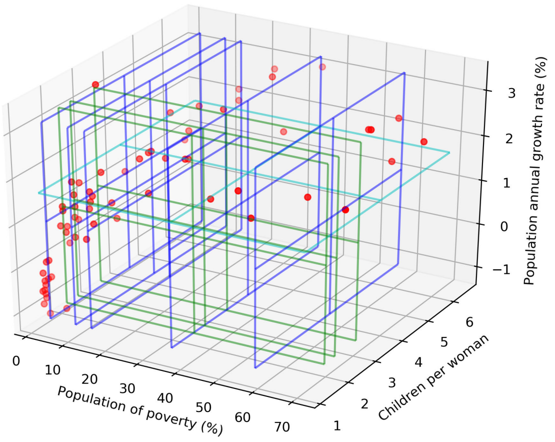
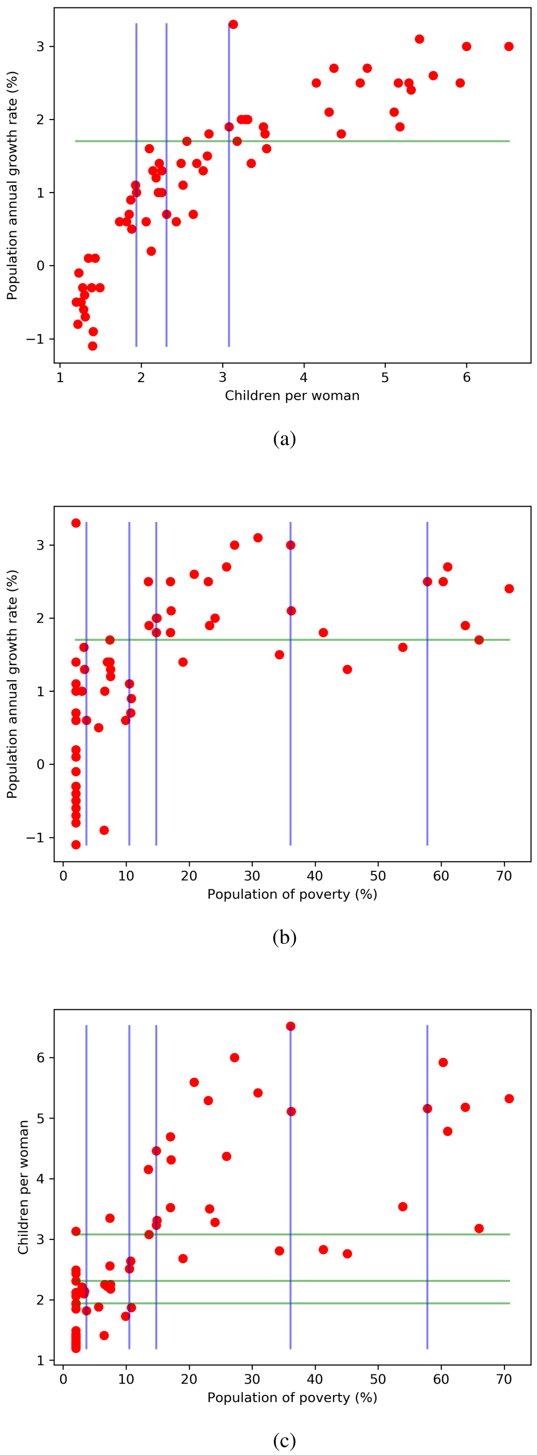


FIGURE 5. Grid of population of poverty, children per woman, population annual growth rate. The 'population of poverty' in the figure is the 'population living below the poverty line (% living on < US$1 per day)' in the dataset. The 'Children per woman' in the figure is the 'Children_per_woman' in the dataset. MMIC = 0.89, R^2 = 0.75.

Then we select $\mathbf{X}$ from these four variables in pairs, and set population annual growth rate as $\mathbf{Y}$ to calculate MMIC. The results are shown in Table. 5. Compared with R^2, Pearson, Spearman, and Kendall correlation coefficient of linear regression, MMIC can give higher scores for non-linear and non-monotonic associations.

For example, the most relevant two-variable combination is population living below the poverty line and children per woman. In the grid of MMIC of these indicators, the partitioning points are 3.7, 10.5, 14.8, 36.1, and 57.8 for population living below the poverty, 1.94, 2.31, and 3.08 for children per woman, 1.7 for population annual growth rate. As shown in Fig. 5 and 6, the partitioning points of the grid are located in the position where the trends of the indicators change. MMIC of these three indicators is near 0.9, exceeding MMIC of two indicators in them (0.73 of population annual growth rate and population living below the poverty, 0.70 of population annual growth rate and children per woman) and R^2 of linear regression of these three indicators(0.75).



(a)

(b)

(c)

FIGURE 6. Grids of population of poverty, children per woman, and population annual growth rate project to each two-dimensional space.

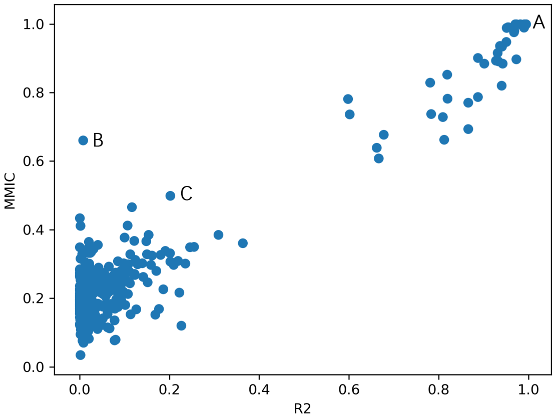


FIGURE 7. MMIC and R^2 of meteorological variables.

TABLE 6. Theorical time complexity of the calculation of MMIC (Algorithm. 1, $\alpha = 0.6$, $\beta = 2$, $\gamma = 1$, $\theta = 1$).

Volume	Dimension	Grid size	Position	Time complexity
100	2	40	16	842
200	2	46	24	1469
400	2	52	36	2518
800	2	58	55	4258
1600	2	64	84	7123
100	3	60	16	2527
200	3	69	24	4407
400	3	78	36	7554
800	3	87	55	12774
1600	3	96	84	21369
100	4	80	16	6739
200	4	92	24	11752
400	4	104	36	20143
800	4	116	55	34063
1600	4	128	84	56984

2) EXAMPLES ON METEOROLOGICAL DATASETS

The meteorological datasets are collected from 18 observation points. Our analysis focuses on the associations of temperature, wind, sky coverage, and liquid precipitation. The purpose of the examples of MMIC on meteorological datasets is to evaluate the strength and non-linearity of the associations of the indicators.

We use MMIC $- R^2$ to evaluate the non-linearity of associations. R^2 denotes the goodness of fit of linear regression. This measure corresponds to the non-linearity measure of MIC: MIC $- \rho^2$, where ρ denotes the Pearson correlation coefficient [2]. MMIC $- R^2$ is near 0 for linear associations and large for non-linear associations with high scores of MMIC. After examining all associations, we find there are 26.2% (88/335) of them are non-linear (MMIC $- R^2 > 0.2$).

We select three typical points in Fig. 7 for illustration. Point A, B, and C represent strong linear (dew point temperature, wind direction, and air temperature), non-linear (sky condition, air temperature, and liquid precipitation), and weak non-linear association (sea level pressure, air temperature, and liquid precipitation) respectively (see Fig. 8).

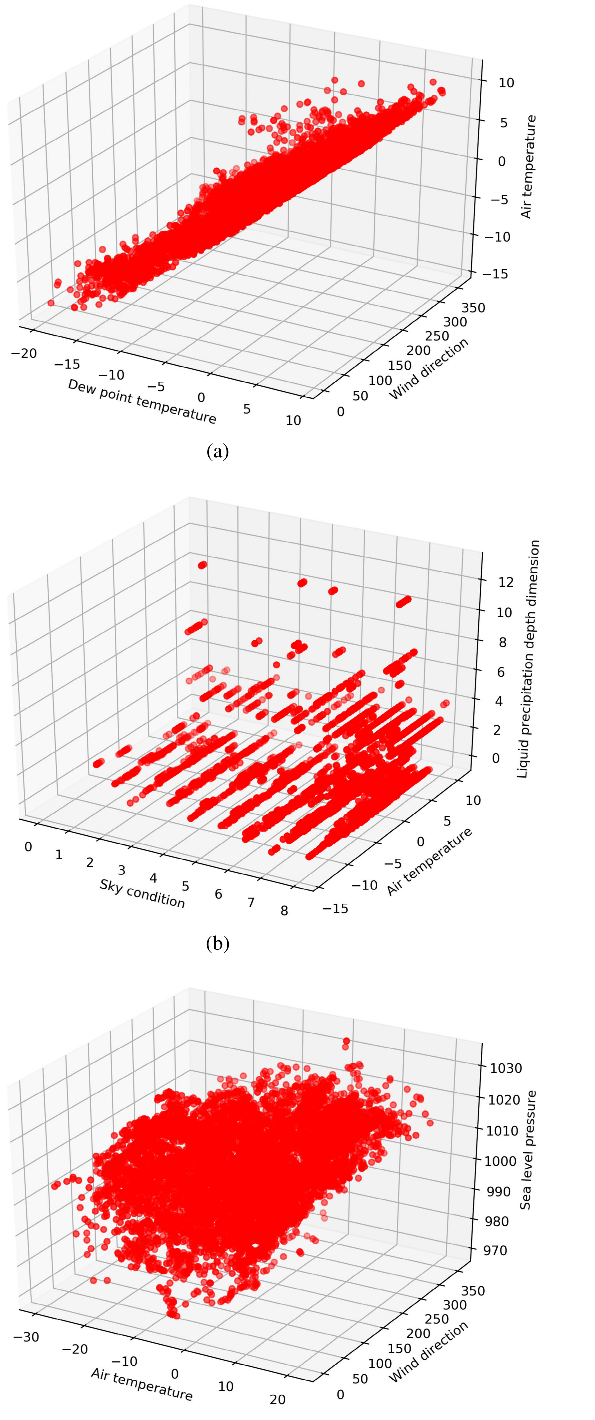


FIGURE 8. Point distributions of three variables in meteorological datasets (a)Strong linear association of dew point temperature (degrees Celsius, the temperature to which a given parcel of air must be cooled at constant pressure and water vapor content in order for saturation to occur) [47], wind direction (angular degrees), and air temperature (degrees Celsius). Point A in Fig. 7. MMIC = 0.99, $R^2 = 0.97$. (b)Non-linear association of sky condition total coverage code (0-19), air temperature (degrees Celsius), and liquid precipitation depth dimension (millimeters, six hour duration). Point B in Fig. 7. MMIC = 0.66, $R^2 = 0.01$. (c)Weak non-linear association of wind direction (angular degrees), air temperature (degrees Celsius), and sea level pressure (hectopascals). Point C in Fig. 7. MMIC = 0.41, $R^2 = 0.10$.

TABLE 7. Relative noise, MMIC, and R^2 of two variables.

Function type	Relative noise	MMIC	R^2
sum	0	1.00	1.00
sum	0.2	0.74	0.87
sum	0.4	0.48	0.52
sum	0.6	0.31	0.40
sum	0.8	0.24	0.22
sum of power	0	1.00	1.00
sum of power	0.2	0.88	0.89
sum of power	0.4	0.56	0.58
sum of power	0.6	0.32	0.35
sum of power	0.8	0.33	0.35
sum of exponent	0	1.00	1.00
sum of exponent	0.2	0.72	0.84
sum of exponent	0.4	0.46	0.60
sum of exponent	0.6	0.29	0.35
sum of exponent	0.8	0.32	0.30
product	0	1.00	1.00
product	0.2	0.82	0.87
product	0.4	0.66	0.70
product	0.6	0.39	0.42
product	0.8	0.36	0.32
product of power	0	1.00	1.00
product of power	0.2	0.73	0.89
product of power	0.4	0.50	0.67
product of power	0.6	0.39	0.44
product of power	0.8	0.28	0.30
product of exponent	0	1.00	1.00
product of exponent	0.2	0.74	0.82
product of exponent	0.4	0.51	0.54
product of exponent	0.6	0.30	0.26
product of exponent	0.8	0.31	0.29
sine of sum	0	0.92	1.00
sine of sum	0.2	0.82	0.87
sine of sum	0.4	0.68	0.73
sine of sum	0.6	0.48	0.56
sine of sum	0.8	0.39	0.42
random	-	0.15	-

TABLE 8. Relative noise, MMIC, and R^2 of three variables.

Function type	Relative noise	MMIC	R^2
sum	0	1.00	1.00
sum	0.2	0.74	0.81
sum	0.4	0.66	0.59
sum	0.6	0.65	0.42
sum	0.8	0.55	0.27
sum of power	0	1.00	1.00
sum of power	0.2	0.79	0.83
sum of power	0.4	0.49	0.51
sum of power	0.6	0.32	0.22
sum of power	0.8	0.35	0.20
sum of exponent	0	1.00	1.00
sum of exponent	0.2	0.86	0.78
sum of exponent	0.4	0.66	0.51
sum of exponent	0.6	0.57	0.34
sum of exponent	0.8	0.47	0.14
product	0	1.00	1.00
product	0.2	0.64	0.76
product	0.4	0.41	0.27
product	0.6	0.42	0.28
product	0.8	0.32	0.11
product of power	0	1.00	1.00
product of power	0.2	0.69	0.78
product of power	0.4	0.41	0.29
product of power	0.6	0.32	0.16
product of power	0.8	0.35	0.23
product of exponent	0	0.98	1.00
product of exponent	0.2	0.48	0.62
product of exponent	0.4	0.52	0.32
product of exponent	0.6	0.31	0.19
product of exponent	0.8	0.23	0.09
sine of sum	0	0.95	1.00
sine of sum	0.2	0.91	0.91
sine of sum	0.4	0.68	0.71
sine of sum	0.6	0.66	0.57
sine of sum	0.8	0.56	0.43
random	-	0.16	-

V. CONCLUSION

In this paper, we propose a multi-variable extension of the maximum information coefficient, and name it as the multi-variable maximum information coefficient (MMIC). We prove that MMIC inherits all good properties of MIC including generality and equability. Besides, MMIC is also proven to have some intuitive properties of multiple variables like monotonicity and subadditivity. For the calculation of MMIC, we design an algorithm based on greedy stepwise strategy and UCB and analyze the time complexity of the algorithm. Based on generated and real datasets, we illustrate the rationality and feasibility of MMIC. MMIC can detect non-linear and non-monotonic associations of multiple variables which may be helpful for large-scale association analysis.

MMIC has some room for further research. For example, distinguishing the coincidence and real associations and mining causality through associations by MMIC are two meaningful tasks. Some heuristic searching and reinforce learning strategies can be introduced to improve the efficiency of calculation by pre-evaluating the grids. A combination of heuristic high-layer strategy and greedy or brute force lower-layer strategy may be a good choice for the calculation of MMIC.

TABLE 9. Most relevant indicators of population annual growth rate evaluated by two-variable MMIC (MMIC > 0.6).

Order	Column name in dataset	MMIC
1	Population_growth	0.89
2	Urban_population_growth	0.74
3	Population living below the poverty line (% living on <; US$1 per day)	0.73
4	Children_per_woman	0.70
5	Contraceptive prevalence (%)	0.70
6	Under-5 mortality rate (Probability of dying aged <; 5 years per 1 000 live births) lowest wealth quintile	0.65
7	Population proportion over 60 (%)	0.64
8	Children_and_elderly	0.62
9	Under-5 mortality rate (Probability of dying aged <; 5 years per 1 000 live births) rural	0.62
10	Contraceptive_use	0.62
11	Hospital beds (per 10 000 population)	0.60

APPENDIX

In this section, the detailed results are demonstrated, including the time complexity of Algorithm. 1 in Table. 6, the relation of relative noise, MMIC, and R^2 in Table. 7 and Table. 8, and the most relevant indicators of population annual growth rate in WHO dataset evaluated by two-variable MMIC in Table. 9.

More detailed codes, data, and results are posted on https://github.com/GuTaoyong/The-Multi-Variable-Maximum-Information-Coefficient.

REFERENCES

[1] A. Halevy, P. Norvig, and F. Pereira, "The unreasonable effectiveness of data," *IEEE Intell. Syst.*, vol. 24, no. 2, pp. 8–12, Mar. 2009.

[2] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518–1524, Dec. 2011.

[3] D. Reshef, Y. Reshef, M. Mitzenmacher, and P. Sabeti, "Equitability analysis of the maximal information coefficient, with comparisons," 2013, *arXiv:1301.6314*. [Online]. Available: http://arxiv.org/abs/1301.6314

[4] D. N. Reshef, Y. A. Reshef, M. Mitzenmacher, and P. C. Sabeti, "Cleaning up the record on the maximal information coefficient and equitability," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 33, pp. E3362–E3363, Aug. 2014.

[5] Y. A. Reshef, D. N. Reshef, P. C. Sabeti, and M. Mitzenmacher, "Theoretical foundations of equitability and the maximal information coefficient," 2014, *arXiv:1408.4908*. [Online]. Available: http://arxiv.org/abs/1408.4908

[6] J. B. Kinney and G. S. Atwal, "Equitability, mutual information, and the maximal information coefficient," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 9, pp. 3354–3359, Mar. 2014.

[7] A. Rényi, "On measures of dependence," *Acta Math. Academiae Scientiarum Hungarica*, vol. 10, nos. 3–4, pp. 441–451, 1959.

[8] T. F. Móri and G. J. Székely, "Four simple axioms of dependence measures," *Metrika*, vol. 82, no. 1, pp. 1–16, Jan. 2019.

[9] S. Wang, Y. Zhao, Y. Shu, H. Yuan, J. Geng, and S. Wang, "Fast search local extremum for maximal information coefficient (MIC)," *J. Comput. Appl. Math.*, vol. 327, pp. 372–387, Jan. 2018.

[10] C. Wang, D. Dai, X. Li, A. Wang, and X. Zhou, "SuperMIC: Analyzing large biological datasets in bioinformatics with maximal information coefficient," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 4, pp. 783–795, Jul. 2017.

[11] S. Wang, Y. Zhao, Y. Shu, and W. Shi, "Improved approximation algorithm for maximal information coefficient," *Int. J. Data Warehousing Mining*, vol. 13, no. 1, pp. 76–93, Jan. 2017.

[12] S. Wang and Y. Zhao, "Analysing large biological data sets with an improved algorithm for MIC," *Int. J. Data Mining Bioinf.*, vol. 13, no. 2, pp. 158–170, 2015.

[13] T. Aparicio, E. F. Pozo, and D. Saura, "An independence test based on joint recurrences," *Modern Economy*, vol. 6, no. 8, pp. 895–907, 2015.

[14] D. Albanese, S. Riccadonna, C. Donati, and P. Franceschi, "A practical tool for maximal information coefficient analysis," *GigaScience*, vol. 7, no. 4, Apr. 2018, Art. no. giy032.

[15] D. Tang, M. Wang, W. Zheng, and H. Wang, "RapidMic: Rapid computation of the maximal information coefficient," *Evol. Bioinf.*, vol. 10, Jan. 2014, Art. no. EBO.S13121.

[16] Y. Li, Z. Dai, D. Cao, F. Luo, Y. Chen, and Z. Yuan, "Chi-MIC-share: A new feature selection algorithm for quantitative structure–activity relationship models," *RSC Adv.*, vol. 10, no. 34, pp. 19852–19860, May 2020.

[17] K. Zheng, X. Wang, B. Wu, and T. Wu, "Feature subset selection combining maximal information entropy and maximal information coefficient," *Appl. Intell.*, vol. 50, no. 2, pp. 487–501, Feb. 2020.

[18] L. Xu, J. Feng, X. Li, and J. Chen, "A LiDAR data-based camera self-calibration method," *Meas. Sci. Technol.*, vol. 29, no. 7, Jul. 2018, Art. no. 075205.

[19] L. Zhao, P. Wang, B. Song, X. Wang, and H. Dong, "An efficient kriging modeling method for high-dimensional design problems based on maximal information coefficient," *Struct. Multidisciplinary Optim.*, vol. 61, no. 1, pp. 39–57, Jan. 2020.

[20] Z. Zhang, S. Sun, M. Yi, X. Wu, and Y. Ding, "MIC as an appropriate method to construct the brain functional network," *Biomed Res. Int.*, vol. 2015, Feb. 2015, Art. no. 825136.

[21] M. S. Morelli, A. Greco, G. Valenza, A. Giannoni, M. Emdin, E. P. Scilingo, and N. Vanello, "Analysis of generic coupling between EEG activity and PETCO2 in free breathing and breath-hold tasks using maximal information coefficient (MIC)," *Sci. Rep.*, vol. 8, no. 1, p. 4492, Dec. 2018.

[22] G. Valenza, A. Greco, C. Gentili, A. Lanata, L. Sebastiani, D. Menicucci, A. Gemignani, and E. P. Scilingo, "Combining electroencephalographic activity and instantaneous heart rate for assessing brain–heart dynamics during visual emotional elicitation in healthy subjects," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 374, no. 2067, May 2016, Art. no. 20150176.

[23] W. Li, H. Fang, G. Qin, X. Tan, Z. Huang, F. Zeng, H. Du, and S. Li, "Concentration estimation of dissolved oxygen in pearl river basin using input variable selection and machine learning techniques," *Sci. Total Environ.*, vol. 731, Aug. 2020, Art. no. 139099.

[24] K. Lin, P. Lu, C.-Y. Xu, X. Yu, T. Lan, and X. Chen, "Modeling saltwater intrusion using an integrated Bayesian model averaging method in the pearl river delta," *J. Hydroinform.*, vol. 21, no. 6, pp. 1147–1162, Nov. 2019.

[25] Y. Wang, H. Tang, T. Wen, and J. Ma, "A hybrid intelligent approach for constructing landslide displacement prediction intervals," *Appl. Soft Comput.*, vol. 81, Aug. 2019, Art. no. 105506.

[26] P. Wu, D. Zhou, Y. Wang, W. Lin, A. Sun, H. Wei, Y. Fang, X. Cong, and Y. Jiang, "Identification and validation of alternative splicing isoforms as novel biomarker candidates in hepatocellular carcinoma," *Oncol. Rep.*, vol. 41, no. 3, pp. 1929–1937, Dec. 2018.

[27] J. N. Poynter, J. R. B. M. Bestrashniy, K. A. T. Silverstein, A. J. Hooten, C. Lees, J. A. Ross, and J. Tolar, "Cross platform analysis of methylation, miRNA and stem cell gene expression data in germ cell tumors highlights characteristic differences by tumor histology," *BMC Cancer*, vol. 15, no. 1, pp. 769–779, Dec. 2015.

[28] D. Yang and H. Liu, "Maximal information coefficient applied to differentially expressed genes identification: A feasibility study," *Technol. Health Care*, vol. 27, no. S1, pp. 249–262, Jun. 2019.

[29] X. Feng, Q. Feng, S. Li, X. Hou, and S. Liu, "A deep-learning-based oil-well-testing stage interpretation model integrating multi-feature extraction methods," *Energies*, vol. 13, no. 8, p. 2042, Apr. 2020.

[30] J. Yang, L. Lin, Z. Sun, Y. Chen, and S. Jiang, "Data validation of multifunctional sensors using independent and related variables," *Sens. Actuators A, Phys.*, vol. 263, pp. 76–90, Aug. 2017.

[31] Y. Fan, S. Liu, L. Qin, H. Li, and H. Qiu, "A novel online estimation scheme for static voltage stability margin based on relationships exploration in a large data set," *IEEE Trans. Power Syst.*, vol. 30, no. 3, pp. 1380–1393, May 2015.

[32] F. Shao and K. Li, "A graph model for preventing railway accidents based on the maximal information coefficient," *Int. J. Modern Phys. B*, vol. 31, no. 3, Jan. 2017, Art. no. 1750010.

[33] X. Huang, Y.-P. Luo, and L. Xia, "An efficient wavelength selection method based on the maximal information coefficient for multivariate spectral calibration," *Chemometric Intell. Lab. Syst.*, vol. 194, Nov. 2019, Art. no. 103872.

[34] Y. Wang, N. Zhang, Q. Chen, D. S. Kirschen, P. Li, and Q. Xia, "Data-driven probabilistic net load forecasting with high penetration of behind-the-meter PV," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3255–3264, May 2018.

[35] S. Bai, M. Li, R. Kong, S. Han, H. Li, and L. Qin, "Data mining approach to construction productivity prediction for cutter suction dredgers," *Autom. Construct.*, vol. 105, no. 9, Sep. 2019, Art. no. 102833.

[36] E. Martínez-Gómez, M. T. Richards, and D. S. P. Richards, "Distance correlation methods for discovering associations in large astrophysical databases," *Astrophys. J.*, vol. 781, no. 1, pp. 39–49, Jan. 2014.

[37] B. Hemmateenejad and K. Baumann, "Screening for linearly and nonlinearly related variables in predictive cheminformatic models," *J. Chemometrics*, vol. 32, no. 4, p. e3009, Apr. 2018.

[38] G. Sagl, T. Blaschke, E. Beinat, and B. Resch, "Ubiquitous geo-sensing for context-aware analysis: Exploring relationships between environmental and human dynamics," *Sensors*, vol. 12, no. 7, pp. 9800–9822, Jul. 2012.

[39] G. Li, Z. Zhou, C. Hu, L. Chang, H. Zhang, and C. Yu, "An optimal safety assessment model for complex systems considering correlation and redundancy," *Int. J. Approx. Reasoning*, vol. 104, no. 1, pp. 38–56, Jan. 2019.

[40] Z. Chen, S. Sun, Y. Wang, Q. Wang, and X. Zhang, "Temporal convolution-network-based models for modeling maize evapotranspiration under mulched drip irrigation," *Comput. Electron. Agricult.*, vol. 169, Feb. 2020, Art. no. 105206.

[41] L. Wang, P. Xing, C. Wang, X. Zhou, and Z. Dai, "Maximal information coefficient and support vector regression based nonlinear feature selection and QSAR modeling on toxicity of alcohol compounds to tadpoles of Rana temporaria," *J. Brazilian Chem. Soc.*, vol. 30, no. 2, pp. 279–285, 2019.

[42] S. Liao, Z. Liu, B. Liu, C. Cheng, X. Jin, and Z. Zhao, "Multistep-ahead daily inflow forecasting using the ERA-interim reanalysis data set based on gradient-boosting regression trees," *Hydrol. Earth Syst. Sci.*, vol. 24, no. 5, pp. 2343–2363, May 2020.

[43] P. Auer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, no. 2, pp. 235–256, 2002.

[44] Y. Jiang and Q. Zhang, "On improved 3MIC algorithm on exploring large data sets with multi-variables and application," in *Proc. 7th Int. Conf. Intell. Hum.-Mach. Syst. Cybern.*, Aug. 2015, pp. 157–160.

[45] F. Shao and K. Li, "Detecting novel multi-variable associations in big data based on MIC," in *Proc. IEEE 5th Int. Conf. Electron. Inf. Emergency Commun.*, May 2015, pp. 39–42.

[46] WHO. (2019). *The Global Health Observatory, Explore a World of Health Data, Data/Gho Indicators [EB/OL]*. [Online]. Available: https://www.who.int/data/gho/data/indicators/

[47] NCDC. (2020). *The Integrated Surface Data (ISD), [EB/OL]*. ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-lite/2020/

TAOYONG GU was born in Zhejiang, China, in 1992. He received the B.S. degree from the Department of Computer Science and Technology, Nanjing University, in 2015, and the M.S. degree from Equipment Management and UAV Engineering College, Air Force Engineering University, Xi'an, China, in 2017, where he is currently pursuing the Ph.D. degree. His current research interests include data mining and machine learning.

JIANSHENG GUO was born in Lingbao, Henan, China, in 1965. He received the Ph.D. degree from Northwest Polytechnic University. He is currently a Professor with the Air Force Engineering University. His current research interests include equipment maintenance support, information systems, and big data technology.

ZHENGXIN LI received the Ph.D. degree in control science and engineering from Air Force Engineering University, China, in 2011. He currently holds a postdoctor position with the School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University. His research interests include time series pattern recognition, machine learning, and data mining.

SHENG MAO was born in Hubei, China, in 1993. He received the B.S. and M.S. degrees from Aeronautical and Aerospace Engineering College, Air Force Engineering University, Xi'an, China, in 2014 and 2016, respectively, where he is currently pursuing the Ph.D. degree with the Equipment Management and UAV Engineering College. His current research interests include anomaly detection, deep learning, and optimization theory.