Journal of Computational and Applied Mathematics 327 (2018) 372–387

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

Fast search local extremum for maximal information coefficient (MIC)

Shuliang Wang [a,*], Yiping Zhao [b], Yue Shu [c], Hanning Yuan [a,*], Jing Geng [a], Shaopeng Wang [a]

[a] School of Software, Beijing Institute of Technology, Beijing, China
[b] Software Center, Bank of China, Beijing, China
[c] Tecent Technology Company Limited, Beijing, China

ARTICLE INFO

Article history:
Received 2 September 2016
Received in revised form 16 March 2017

Keywords:
Local extremum
Fast search
Gradient descent
Intelligent maximal information coefficient (iMIC)
Big data

ABSTRACT

Maximal information coefficient (MIC) is an indicator to explore the correlation between pairwise variables in large data sets, and the accuracy of MIC has an impact on the measure of dependence for each pair. To improve the equitability in an acceptable run-time, in this paper, an intelligent MIC (iMIC) is proposed for optimizing the partition on the y-axis to approximate the MIC with good accuracy. It is an iterative algorithm on quadratic optimization to generate a better characteristic matrix. During the process, the iMIC can quickly find out the local optimal value while using a lower number of iterations. It produces results that are close to the true MIC values by searching just n times, rather than n^2 computations required for the previous method. In the compared experiments of 169 indexes about 202 countries from World Health Organization (WHO) data set, the proposed algorithm offers a better solution coupled with a reasonable run-time for MIC, and good performance search for the extreme values in fewer iterations. The iMIC develops the equitability keeping the satisfied accuracy with fast computational speed, potentially benefitting the relationship exploration in big data.

Contents

* Corresponding authors.
E-mail addresses: slwang2011@bit.edu.cn (S. Wang), yhn6@bit.edu.cn (H. Yuan).

http://dx.doi.org/10.1016/j.cam.2017.05.038
0377-0427/

1. Introduction

Data have been continuously created and accumulated in various fields [1–5]. New characteristics appear in the volume, variety, velocity and veracity of big data. To extract the value from the big data, it is eager to detect the useful correlations between random variables [6,7]. However, the relationships in the real world are complex and the obtained data are further often dirty, which makes it difficult to find out the interested relation behind the data [8]. Reshef et al. [9] proposed the maximal information coefficient (MIC) that is a statistic measure for identifying novel associations between pairwise variables in large data sets. The range of MIC is between 0 and 1. If MIC is closer to 1, then the two variables are closer relationship; if MIC is closer to 0, then the two variables are more likely two independent variables. MIC has two properties: generality and equitability, which makes it to discover more interesting relationships,without inclinations to specific relation types. Compared with other methods, including mutual information estimation, principal curve-based methods, the Spearman correlation coefficient, distance correlation, and maximal correlation [10–12], the MIC showed its equitability through simulations.

However, the accuracy of MIC has a great impact on each pair of dependence, which directly affects similarity judgement. The increasing accuracy of MIC leads to the changing rank of pairwise variables. The approximation with low accuracy will result in the deficiency in the process of feature engineering, thus reducing the accuracy of the system. As the closely related pairs of variables are more likely overlooked due to the low accuracy of MIC, more valuable associations can be detected when MIC comes closer to its true value. The MIC should be greatly enhanced by exploring its properties because computing MIC in the standard approximation algorithm leads to some deviations from the equitability property. The problem of losing some important pairs can be avoided by using MIC with high accuracy in data mining. Meanwhile, compared to other distance correlation methods [10], the statistical power of MIC is lower. In many important relationships [11,13] MIC may lose its advantage due to the power drawback.

In a follow-up study, Reshef et al. [14] believed that the use of the approximation algorithm affected the equitability through the comparison to a lower efficiency algorithm that provided a more intensive search for optimal grids, and suggested that the essence of the MIC did not lead to the deviations from equitability of the MIC values. In particular, they had been expecting the approximation algorithms to have better accuracy-time tradeoffs.

An unavoidable problem in MIC is to find out the global optimum value. An exhaustive search method attempted to get a better approximation [14]. However, the mesh partition is limited due to the extensive computation, furthermore, MIC is limited between two vectors. The fact is that the relationships can be generated between multi-vectors, which means that the vectors seem have no relationship, but may do have strong relationship. Furthermore, it was found that the power of MIC to represent the effect of the algorithm on statistics was low, thus greatly reducing the reliability of MIC.

The use of the approximation algorithm that affects the equitability [9,14] motivates us to better the approximation algorithm to generate a better characteristic matrix, making it possible to obtain higher MIC values to improve the equitability in an acceptable run-time. Also, the experiments with the exhaustive algorithm [14] have inspired us to search for a more optimal y-axis partition than the y-axis equipartition. Based on the original algorithm [9,14], Wang et al. developed the equitability of MIC [15].

Therefore, in this paper, in order to further refine the algorithm and make it more generic, we propose an intelligent MIC (iMIC) for approximation algorithm, being verified with much more experiments. The rest of this paper is organized as follows. Section 2 summarizes the work that has been done previously in this field. Section 3 presents the principles of the iMIC, i.e. algorithms,the mathematical improvement offered by the iMIC, and iMIC compared with the MIC on algorithms [9,14] in terms of their effectiveness and accuracy. In Section 4, iMIC is applied to a real-world data set to show its performance. The final conclusion is drawn in Section 5.

2. Related work

To identify the arbitrary correlations in massive data sets, many algorithms are provided to discover the relationship between pairwise variables [3,16,17]. On the basis of mutual information [18], Reshef et al. [9] proposed the MIC. It is an important indicator to take full advantage of big data. For example, MIC can be used to measure the dependence of data mining system or general recommending system.

Now it has been applied in many fields, and further help to improve their applications [19–23]. Nature Biotechnology [12] showed great interest about it and invited eight experts to discuss its usefulness. Speed [24] stated that the MIC detects nonlinear correlations in data sets equitably, which had reached the crest of the field called mutual information. An exploratory landscape analysis measure called maximum entropic epistasis was given to quantify the level of variable interactions in a continuous optimization problem by using MIC [25]. In airborne laser scanning, to predict forest attributes, an optimization method based on the MIC was proposed [26]. MIC was applied for mapping nonlinear relationships between the variables [27]. In a case of monthly streamflow prediction in the Xiangxi River, China, the nonlinearity between the hydro-meteorological variables and the streamflow measurements was characterized through the measure of MIC [28]. Employing MIC, a graph model where nodes denoted influencing factors of railway accidents and edges represented dependence of the two linked factors was proposed for preventing railway accidents which avoided complex mathematical computation [29]. In greedy equivalent search, a draft of the real network was designed based on conditional independence tests and MIC, which helped in finding more correct dependent relationship between variables [30]. To overcome the prerequisites that the

biological annotations do not co-occur with each other in particular, MIC was suggested to discover the hidden regularities between biological annotations [31].

However, there are also stimulated questions in MIC. Simon and Tibshirani [13] pointed out that the MIC would cause false positives in data analysis due to the low power. Gorfine et al. [32] noted that the equitability of MIC was less practical than a new test, HHG [33], for data exploration by power comparisons. A mathematical proof was also provided by Kinney and Atwal [34] to support mutual information rather than the MIC.

Then, Reshef et al. [14] provided a more intensive search for optimal grids. They modified the original algorithm [9] to produce an approximation algorithm that was less efficient but had greater accuracy. This more exhaustive algorithm had better equitability than the original algorithm [9], which meant that the deviations from the equitability of the original MIC values were a result of the accuracy of the approximation algorithm, rather than the nature of the MIC. Based on the original algorithm [9,14], the equitability of MIC was developed [15].

This paper proceeds from the following three aspects. Firstly, by analyzing the algorithm flow, the time complexity of the algorithm is obtained. It is proved in this paper that the time complexity of iMIC algorithm is higher than that of MIC original algorithm, but it is much lower than the exhaustive search algorithm. The time cost is acceptable. Secondly, according to the definition of mutual information, it is proved that the accuracy of the result of iMIC algorithm is higher than that of MIC, and it is not lower than the algorithm proposed in [9,14]. Thirdly, prove that the equalization of the iMIC algorithm is higher than that of the MIC original algorithm.

In terms of experiment, firstly, the difference of characteristic matrix between the original algorithm and iMIC is analyzed by using data sets which contain Gaussian noise or vertical noise. The data sets are generated by functions which are used by Reshef et al. [9] and other functions. Then, we use five data sets, which is generated by five different functions, to show the results of the characteristic matrix analysis. In the same way, we analyze the equitability by the data sets. And, we select the data sets created by using four functions to demonstrate the experiment results.

3. Principles of iMIC

Inspired by gradient descent to find local extremum, an intelligent MIC (iMIC) is proposed to find a better partition on the y-axis than the simple equipartition by quadratic optimization as an alternative to the exhaustive searching method. It helps to create a better characteristic matrix, which makes it possible to obtain higher MIC values to improve the equitability in an acceptable run-time.

3.1. From MIC to iMIC

For computation of the MIC, the original algorithm is a simple equipartition of the y-axis into 2 or 3 rows [9], and the modified original algorithm [14] is an exhaustive quadratic optimization approach that exhaustively searches an equipartition of the y-axis for up to 20 rows to find the best subpartition into 2 or 3 rows for all grids with 2 or 3 rows, respectively. [14] had less efficient but had greater accuracy than [9]. That is, it is time consuming but offers high accuracy.

We conduct a search of the optimal grid by a second optimization on the y-axis for each element of the characteristic matrix. To be specific, we compute the largest mutual information for each grid partition $((x, y_1), x \in \{2, 3, \ldots, \frac{Q}{y_1}\}; Q = m^{0.6}$, m is the number of data points) using the algorithm [9] with y-axis partition of size y_1. Then we fix every kind $(x \in \{2, 3, \ldots, \frac{Q}{y_1}\})$ to repartition the y-axis to correspond to the x-axis partitions. Well, we just fix the x-axis partition in the one-step iMIC, then search for the largest value in the exhaustive algorithm results for a given y row by repartitioning the y-axis once only. As the two algorithms above for computation of the MIC are equivalent, so we can search for a local extremum with fewer iterations with iMIC.

One-step iMIC uses quadratic optimization on the y-axis just once to search for a more optimal y-axis partition, rather than the equipartition of [9,14]. For each row of the characteristic matrix generated by the algorithm [9], we attain a better solution based on the largest value of this row. Here is an example from MIC to iMIC shown in Fig. 1(a)–(e). Actually, by using the original algorithm [9], from the y-axis equipartition of Fig. 1(a), it is obvious that the grid partition displayed as Fig. 1(b) achieves the largest mutual information. Simultaneously, to search for the optimal grid, the rest of the steps (Fig. 1(c)–(e)) are added in iMIC. It indicates that the equipartition is a subset of the quadratic optimization on the y-axis with the same partition on the x-axis; the quadratic optimization on the y-axis achieves a better maximal normalized mutual information score than all the scores in one row of the original characteristic matrix. iMIC improves the maximum score of the original characteristic matrix for each row, while the original characteristic matrix of MIC has the largest score. As a result, the iMIC can achieve greater accuracy. The degree of improvement will be proven in Section 3.3.

3.2. iMIC algorithm

Deduction 1. *Local extremum of MIC can be reached by calculating one-step iMIC with different axes fixed alternatively, we call it iMIC algorithm.*

iMIC algorithm can be described as follows. At each step, we fix one axis, and enact one-step iMIC to calculate the largest information coefficient. Then, the other axis is fixed, and calculate the one-step iMIC again, to reach a larger information coefficient. We can do these two steps alternatively, until the information coefficient converges.

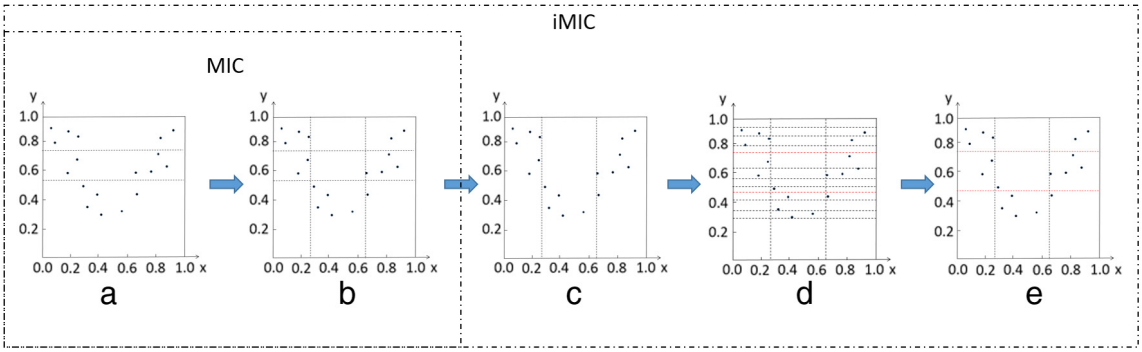


Fig. 1. From MIC to iMIC.

Just like stochastic gradient descent algorithm used in BP neural networks [35] , iMIC can be used as an efficient tool to calculate the local extremum of MIC.

Specifically, given a specific number of rows y_1, the logical path of quadratic optimization is described as follows:

(a) First, equipartition the y-axis into y_1 rows and optimize the x-axis in the manner of Reshef et al. [9].
(b) Second, select the grid partition of integers (x_1, y_1) with the largest normalized mutual information and get a detailed partition of size x_1.
(c) Third, fix the partition of size x_1 on the x-axis and optimize the y-axis into y_1 rows, and then calculate the largest normalized mutual information under the new partition of integers (x_1, y_1) to replace the original value given in the characteristic matrix.
(d) At last, multiple iterate the above steps until the convergence.

After generation of the whole characteristic matrix, we select the maximum score in the characteristic matrix as the MIC. To present the iMIC clearly, the detailed steps used to create the characteristic matrix, $M(x, y)$, each element of which is the largest normalized mutual information achieved by any x-by-y grid, are concluded as a flow chart in Fig. 2.

Let $t = \frac{Q}{2}$, and t is the upper limit of y values. We make k_x be the number of clumps on the x-axis, then the time complexity of the original algorithm [9] was $O(tk_x^2xy)$. Similarly, we make k_y be the number of clumps on the y-axis. So the iMIC will involve two parts: one is to find the optimal solution on the x-axis and the other is the second optimal process of the y-axis. Based on this, the run time is concluded to be $O(tk_x^2xy + tk_y^2yx)$.

Now, we see that how the one-step iMIC reduces the time complexity. As we described above, every time the largest possible mutual information under a specific partition size is computed in [9,14], which represents an element in the characteristic matrix. Thus, in [9,14], n^2 computations must be performed. Using one-step iMIC, the largest possible mutual information represents the maximum value of one row in the characteristic matrix. Therefore, we just perform n computations in total.

3.3. iMIC's mathematical improvement on MIC

In the first two steps, the iMIC is the same as the MIC. For example, in MIC, first, conduct a partition on y-axis, second, compute the largest mutual information, then obtain the results of the optimal grid (Fig. 1).

In iMIC, the optimal grid is searched by a second optimization on the y-axis , and the better results will be used to replace the values in the original characteristic matrix. Thus, the results must be equal to or better than that of the original MIC.

If we get B_1 after the equalization of y-axis and the algorithm in [9,14] and iMIC get the same best two steps search on y-axis noted as B'_1. In MIC, based on the same partition on the y-axis in iMIC, assume that the search result on the x-axis is A and the other partition is A'. Since each point (x, y) in the matrix represents only one optimal partition.

(1) If the partition on x-axis used in [9,14] is equal to that of iMIC, the results of MIC are the same as that of iMIC before the second optimization on the y-axis. Also, the applied second optimization is the original dynamic planning. Thus, iMIC $=$ MIC.

(2) If the partition on x-axis of MIC is not equal to that of iMIC, as the current partition on y-axis of iMIC is the best results, thus the partition set S_1 on x-axis of MIC is a subset of the partition set S_2 on x-axis of iMIC by the definition of MIC [9].

Assume that the partition on y-axis is the same when the partition set S on x-axis becomes larger.

Set $F(A, A') = (H(A) - H(A, B'_1)) - (H(A') - H(A', B'_1))$, when $|A| > |A'|$, proving $F(A, A') > 0$.

I. If B'_1 goes through some point sets that are partitioned by A, we then choose the simplest state to establish the mathematical derivation. Because $A' \subseteq A$, every clump that is divided by A' is equal to some adjacent clumps that are divided by A. Assume that clump D'_k is generated by A' and that D_l and D_{l+1} are generated by A. Let $D'_k = D_l + D_{l+1}$, and B'_1 just cross D_l, separating it into D_{l1} and D_{l2}. Suppose that D_{l1}, D_{l2}, and D_{l+1} contain n_1, n_2, and n_3 data points,

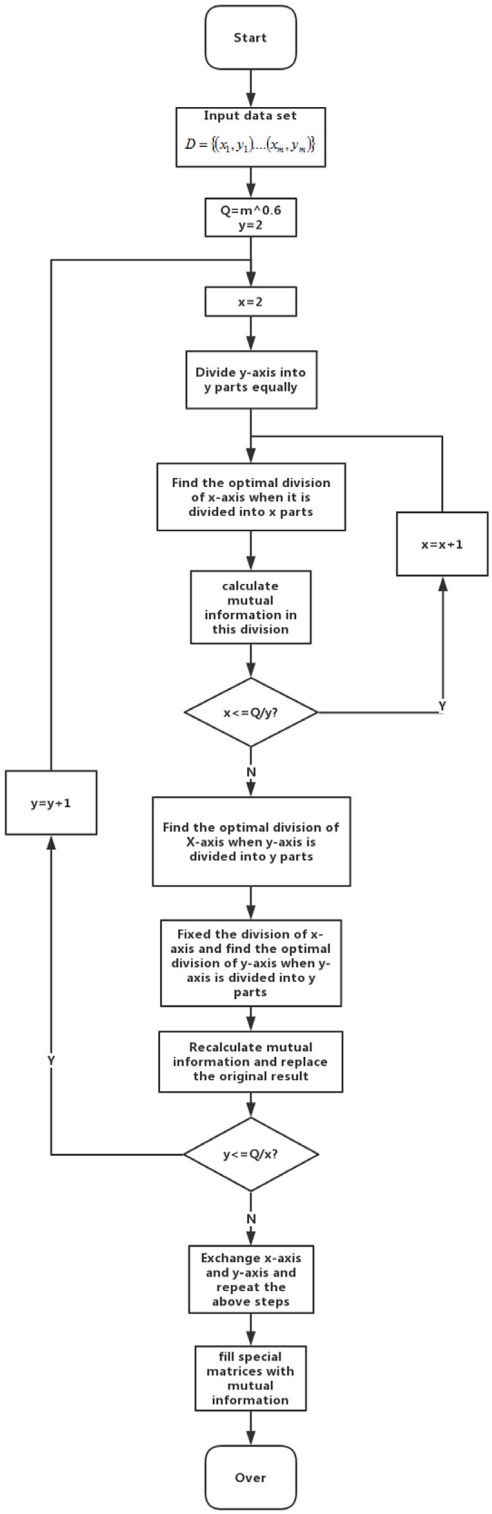


Fig. 2. iMIC flow chart.

respectively; $\sharp_{ij}$ represents the number of data points in the grid of row i and column j; then from the SOM of [9], we learned Eqs. (2) and (3).

$$F(A, A') = I(A; B'_1) - I(A'; B'_1) \tag{1}$$

$$I(A; B'_1) = H(A) - H(A, B'_1) = \sum_{j=1}^{|A|} \frac{\sharp_{*,j}}{m} \log \frac{m}{\sharp_{*,j}} - \sum_{j=1}^{|A|} \sum_{i=1}^{|B'_1|} \frac{\sharp_{i,j}}{m} \log \frac{m}{\sharp_{i,j}} = \sum_{j=1}^{|A|} \sum_{i=1}^{|B'_1|} \frac{\sharp_{i,j}}{m} \log \frac{\sharp_{i,j}}{\sharp_{*,j}} \tag{2}$$

$$I(A'; B'_1) = H(A') - H(A', B'_1) = \sum_{j=1}^{|A'|} \frac{\sharp'_{*,j}}{m} \log \frac{m}{\sharp'_{*,j}} - \sum_{j=1}^{|A'|} \sum_{i=1}^{|B'_1|} \frac{\sharp'_{i,j}}{m} \log \frac{m}{\sharp'_{i,j}} = \sum_{j=1}^{|A'|} \sum_{i=1}^{|B'_1|} \frac{\sharp'_{i,j}}{m} \log \frac{\sharp'_{i,j}}{\sharp'_{*,j}}. \tag{3}$$

Therefore, the counterpart's residual between A and A' is presented below.

$$\begin{aligned}
F(A, A') &= \sum_{j=l}^{l+1} \sum_{i=1}^{|B'_1|} \frac{\sharp_{i,j}}{m} \log \frac{\sharp_{i,j}}{\sharp_{*,j}} - \sum_{j=k}^{k} \sum_{i=1}^{|B'_1|} \frac{\sharp'_{i,j}}{m} \log \frac{\sharp'_{i,j}}{\sharp'_{*,j}} \\
&= \left(\sum_{i=1}^{|B_1|} \frac{\sharp_{i,j}}{m} \log \frac{\sharp_{i,j}}{\sharp_{*,j}} + \sum_{i=1}^{|B'_1|} \frac{\sharp_{i,l+1}}{m} \log \frac{\sharp_{i,l+1}}{\sharp_{*,l+1}} \right) - \sum_{i=1}^{|B_1|} \frac{\sharp'_{i,k}}{m} \log \frac{\sharp'_{i,k}}{\sharp'_{*,k}} \\
&= \left(\frac{n_1}{m} \log \frac{n_1}{n_1 + n_2} + \frac{n_2}{m} \log \frac{n_2}{n_1 + n_2} + \frac{n_3}{m} \log \frac{n_3}{n_3} \right) \\
&\quad - \left(\frac{n_1 + n_3}{m} \log \frac{n_1 + n_3}{n_1 + n_2 + n_3} + \frac{n_2}{m} \log \frac{n_2}{n_1 + n_2 + n_3} \right) \\
&= \frac{n_1}{m} \left(\log \frac{n_1}{n_1 + n_2} - \log \frac{n_1 + n_3}{n_1 + n_2 + n_3} \right) + \frac{n_2}{m} \left(\log \frac{n_2}{n_1 + n_2} - \log \frac{n_2}{n_1 + n_2 + n_3} \right) \\
&\quad + \frac{n_3}{m} \left(\log \frac{n_3}{n_3} - \log \frac{n_1 + n_3}{n_1 + n_2 + n_3} \right) \\
&= \frac{1}{m} \left[n_1 \log \frac{n_1}{n_1 + n_3} + (n_1 + n_2) \log \frac{n_1 + n_2 + n_3}{n_1 + n_2} + n_3 \log \frac{n_1 + n_2 + n_3}{n_1 + n_3} \right].
\end{aligned} \tag{4}$$

Take n_2 as an independent variable, find the first derivative of $F(A, A')$

$$\begin{aligned}
\frac{dF(A, A')}{dn_2} &= \frac{1}{m} \left\{ 0 + \log \frac{n_1 + n_2 + n_3}{n_1 + n_2} + (n_1 + n_2) \frac{n_1 + n_2}{(n_1 + n_2 + n_3) \times \ln 2} \right. \\
&\quad \left. \times \left[1 - \frac{n_1 + n_2 + n_3}{(n_1 + n_2)^2} \right] + \frac{n_3}{(n_1 + n_2 + n_3) \ln 2} \right\} \\
&\quad \times \frac{1}{m} \left\{ \log \frac{n_1 + n_2 + n_3}{n_1 + n_2} + \frac{(n_1 + n_2)^2}{(n_1 + n_2 + n_3) \times (\ln 2)} \right. \\
&\quad \left. \times \left[1 - \frac{n_1 + n_2 + n_3}{(n_1 + n_2)^2} \right] + \frac{n_3}{(n_1 + n_2)^2 \times (\ln 2)} \right\} \\
&= \frac{1}{m} \left\{ \log \frac{n_1 + n_2 + n_3}{n_1 + n_2} + \left[\frac{(n_1 + n_2)^2 + n_3}{n_1 + n_2 + n_3} - 1 \right] \div \ln 2 \right\}
\end{aligned} \tag{5}$$

$\because$ n_1, n_2, n_3 are positive integers.
$\therefore$ $n_1, n_2, n_3 \geq 1$.
$\therefore$ $\frac{(n_1+n_2)^2+n_3}{n_1+n_2+n_3}$ is an increasing function when $n_2 \in [1, +\infty)$.
$\therefore$ $\frac{(n_1+n_2)^2+n_3}{n_1+n_2+n_3} \geq (n_1 + 1)^2 + n_3 n_1 + 1 + n_3 > 0$.
$\because$ $\log \frac{n_1+n_2+n_3}{n_1+n_2} = \log(1 + \frac{n_3}{n_2+n_2}) > 0$.
$\therefore$ $\frac{dF(A,A')}{dn_2} > 0$ and $F(A, A')$ is an increasing function when $n_2 \in (0, +\infty)$.
$\therefore$ $F(A, A') = \frac{1}{m}(n_1 \log \frac{n_1}{n_1+n_3} + n_1 \log \frac{n_1+n_3}{n_1} + n_3 \log \frac{n_1+n_3}{n_1}) = \frac{n_3}{m} \log(1 + \frac{n_3}{n_1}) > 0$.

II. If B'_1 crosses more than one point set, just as Reshef et al. indicated in their SOM [9], the optimal solution can be achieved by drawing the x-axis partition on the edges of clump. So, gathering points by partitioning more clumps is the key for maximizing mutual information. Here, the region obtained from A is a kind of subdivision of that obtained

Table 1
The function list.

No.	Function name	Description ($x \in [0, 1]$)
1	Linear	$y = 2x$
2	Parabolic	$y = 8(x - \frac{1}{2})^2$
3	Sinusoidal	$y = \sin(5\pi x(2 + x))$
4	Categorical	200 points chosen from the set: {(1, 0.975), (2, 0.604), (3, 0.161), (4, 0.560), (5, 0.402)}
5	Circle	$\{(\cos t; \sin t) : t \in [0, 2\pi]\}$

Table 2
The function list.

No.	Function name	Function description ($x \in [0, 1]$)
1	Linear + Periodic, High Freq	$y = \frac{1}{10}\sin(12(2x - 1)) + \frac{1}{2}(2x - 1)$
2	Cosine, High Freq	$y = \cos(20\pi x)$
3	Line	$y = 2x$
4	Sine, High Freq	$y = \sin(18\pi x)$

from A'; in the same y partition B'_1, A can effectively gather the points together, resulting in larger mutual information. Thus, $F(A, A') \geq 0$.

In conclusion, $F(A, A') \geq 0$. Therefore, $I(A'; B'_1) \leq I(A; B'_1)$. This means that the one-step iMIC leads to the same MIC value as the algorithm in [9,14].

3.4. iMIC's improvements of characteristic matrix and equitability

As described above, the iMIC can provide higher accuracy with an acceptable time performance. In this section, we present the performance of iMIC from two aspects: the characteristic matrix and the equitability. The experimental results show that the iMIC achieves a good performance.

I. Improvement of the characteristic matrix. The residual matrices of relationships, as parabola function, sinusoidal function etc., represent that some values in characteristic matrices calculated by the iMIC algorithm are greater than that of original method [9,14]. It means that the proposed method reaches higher precision of MIC since MIC value is the maximum of the characteristic matrix, which results in more interesting relationships being discovered. For each function in Table 1, we first generate one data set of 2300 points spaced evenly along the curve described by the function, each of which is displayed in Fig. 3(a0)–(e0); then, we create another data set with the same sample size by adding uniform vertical noise, as shown in Fig. 4(f0)–(j0), in which R^2 (the coefficient of determination) equals 0.64. We then compute the characteristic matrix of the two data sets using both the original algorithm MIC and the proposed algorithm iMIC. In this work, we set the value of the exponent α as 0.7 in the function $B(n) = n^{\alpha}$, and we only show the results for the range where $1 < x \leq 15$ and $1 < y \leq 15$. On the relationships shown in Fig. 3(a0)–(e0) and Fig. 4(f0)–(j0), Fig. 4(f1)–(j1) correspond to the visualizations of the residual matrices produced by $10 \times (x_2 - x_1)(x \in \{a, b, \ldots, e\})$, respectively.

II. Improvement of the equitability. In data exploration, we would like different relationships with equal noise levels to have similar MIC value, which means exact equitability in this paper. Equitability is a very important indicator of correlation analysis that avoids unfair preference of specific relationship, hence the improvements of equitability perfectly show better performance of computing algorithm. Select 4 different functional relationships which are shown in Table 2. To analyze the equitability clearly, we produce a noiseless data series with a sample size $n = 1000$, and 249 additional noise data series with the same sample size. The above steps are iterated 100 times. The results of approximating the MIC by adding incremental uniform vertical noise are shown in Fig. 5. To perform the same equitability analysis performed in Fig. 5, we also add Gaussian noise in Fig. 6.

4. Practical analysis of WHO data sets

In this section, we select 169 comprehensive evaluation indexes about 202 countries from the data sets, which is published by World Health Organization (WHO) from 1960 to 2005. We compute MIC value with two different algorithms on these global indicators that generated 28 392 pairwise variables and show the rank 100 relationships identified by iMIC in Table 3.

Then, the paper analyzes the difference of MIC value (*d-value*) calculated by two different accuracy algorithms. Table 4 shows the statistic indicators of *d-value*, which means iMIC improves 79% of MIC value, the maximal MIC value improvement reaches 0.46306 and average improvement is 0.046472.

In this paper, we emphasize 11 relationships (A–K) of *d-value* that are more than 0.1, which are shown in Table 5.

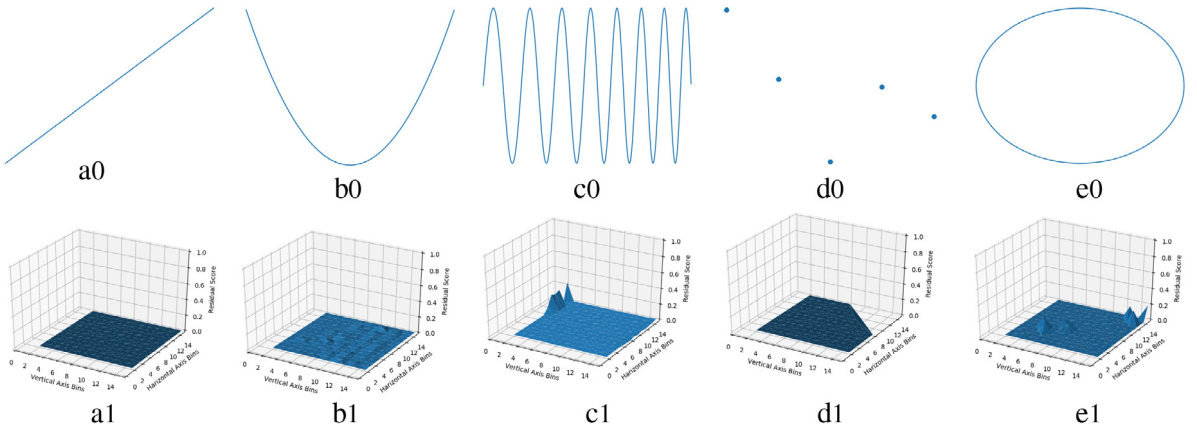


Fig. 3. Characteristic matrices and residual matrices of noiseless relationships.

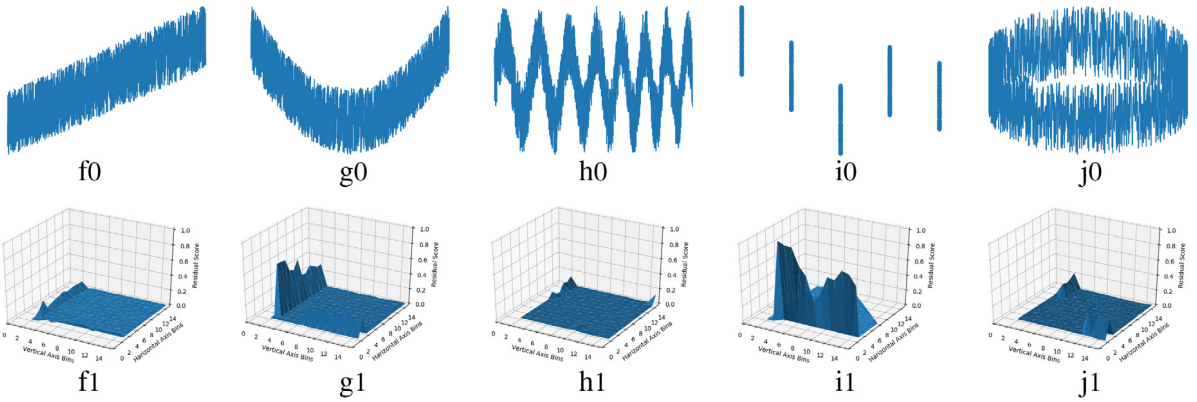


Fig. 4. Characteristic matrices and residual matrices of noisy relationships.

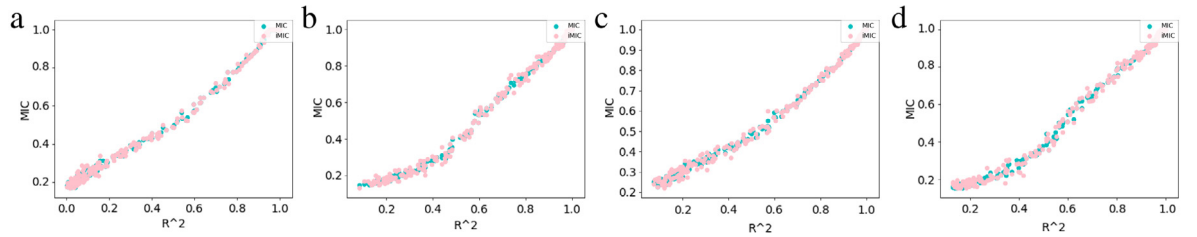


Fig. 5. Comparison of the equitability with incremental uniform vertical noise.

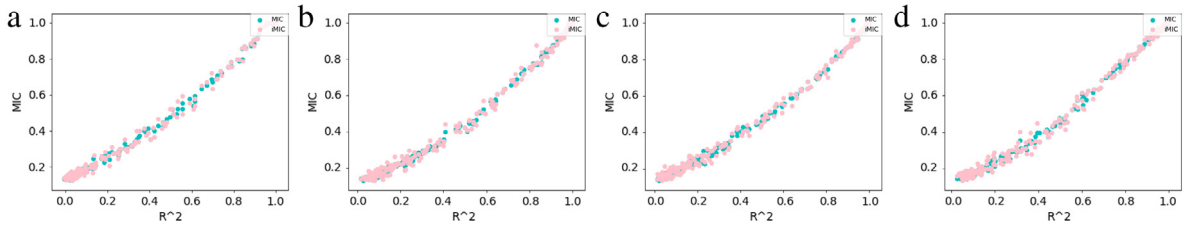


Fig. 6. Comparison of the equitability with Gaussian noise.

Table 3
Rank 100 relationships identified by iMIC.

Rank iMIC	Var-A	Var-B	iMIC	MIC
1	Women who have had mammography (%)	Population living below the poverty line (% living on US$1 per day)	1	1
2	Women who have had PAP smear (%)	Births by caesarean section (%)	1	1
3	Births attended by skilled health personnel ratio highest–lowest educational level of mother	Births attended by skilled health personnel (%) lowest educational level of mother	1	1
4	Births attended by skilled health personnel ratio highest-lowest wealth quintile	Births attended by skilled health personnel (%) lowest wealth quintile	1	1
5	Private expenditure on health as percentage of total expenditure on health	General government expenditure on health as percentage of total expenditure on health	0.99997	0.99997
6	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) female	Infant mortality rate (per 1000 live births) female	0.99982	0.98961
7	Women who have had mammography (%)	Births by caesarean section (%)	0.99952	0.94824
8	Per capita total expenditure on health at average exchange rate (US$)	Per capita government expenditure on health at average exchange rate (US$)	0.99951	0.98961
9	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) male	Infant mortality rate (per 1000 live births) male	0.99951	0.99951
10	Out-of-pocket expenditure as percentage of private expenditure on health	Environment and public health workers density (per 10 000 population)	0.99836	0.99836
11	Women who have had PAP smear (%)	Women who have had mammography (%)	0.99777	0.92659
12	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) both sexes	Infant mortality rate (per 1000 live births) both sexes	0.99773	0.99299
13	Infant mortality rate (per 1000 live births) both sexes	Births by caesarean section (%)	0.99695	0.90612
14	Infant mortality rate (per 1000 live births) female	Births by caesarean section (%)	0.99695	0.90612
15	Measles immunization coverage among one-year-olds ratio urban–rural	Measles immunization coverage among one-year-olds difference urban–rural	0.99679	0.99679
16	Prevalence of tuberculosis (per 100 000 population)	Incidence of tuberculosis (per 100 000 population per year)	0.99672	0.97925
17	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) rural	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) lowest educational level of mother	0.99514	0.92851
18	Population with sustainable access to improved drinking water sources (%) total	Births by caesarean section (%)	0.99335	0.89027
19	Life expectancy at birth (years) male	Adult mortality rate (probability of dying between 15 to 60 years per 1000 population) both sexes	0.99144	0.95852
20	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) both sexes	Infant mortality rate (per 1000 live births) male	0.98973	0.97091
21	Incidence of tuberculosis (per 100 000 population per year)	Deaths due to tuberculosis among HIV-negative people (per 100 000 population)	0.98961	0.98961
22	Prevalence of tuberculosis (per 100 000 population)	Deaths due to tuberculosis among HIV-negative people (per 100 000 population)	0.98961	0.98961
23	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) female	Infant mortality rate (per 1000 live births) both sexes	0.98933	0.966
24	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) rural	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) lowest wealth quintile	0.98789	0.92851
25	Infant mortality rate (per 1000 live births) male	Infant mortality rate (per 1000 live births) both sexes	0.98786	0.9857
26	Incidence of tuberculosis (per 100 000 population per year)	Laboratory health workers density (per 10 000 population)	0.98678	0.98678
27	Prevalence of tuberculosis (per 100 000 population)	Laboratory health workers density (per 10 000 population)	0.98678	0.98678
28	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) female	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) both sexes	0.98559	0.97016
29	Measles immunization coverage among one-year-olds ratio highest-lowest wealth quintile	Measles immunization coverage among one-year-olds difference highest-lowest wealth quintile	0.98522	0.94988
30	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) male	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) both sexes	0.98453	0.9841
31	Infant mortality rate (per 1000 live births) both sexes	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) urban	0.98424	0.90144
32	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) urban	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) rural	0.98423	0.895
33	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) male	Infant mortality rate (per 1000 live births) both sexes	0.98395	0.97756
34	Number of physicians	Number of dentistry personnel	0.98028	0.97436
35	Infant mortality rate (per 1000 live births) female	Infant mortality rate (per 1000 live births) both sexes	0.98023	0.95917
36	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) both sexes	Infant mortality rate (per 1000 live births) female	0.97939	0.95812

(*continued on next page*)

Table 3 (*continued*)

37	Per capita total expenditure on health at average exchange rate (US$)	Gross national income per capita (PPP international $)	0.97826	0.95724
38	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) highest wealth quintile	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) highest educational level of mother	0.97687	0.89054
39	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) lowest wealth quintile	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) lowest educational level of mother	0.97687	0.9133
40	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) urban	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) lowest educational level of mother	0.97641	0.895
41	Adult mortality rate (probability of dying between 15 to 60 years per 1000 population) male	Adultmortality rate (probability of dying between 15 to 60 years per 1000 population) both sexes	0.97227	0.9078
42	Per capita government expenditure on health at average exchange rate (US$)	Gross national income per capita (PPP international $)	0.9719	0.95244
43	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) male	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) female	0.96953	0.9507
44	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) urban	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) highest educational level of mother	0.96698	0.85348
45	Per capita total expenditure on health at average exchange rate (US$)	Per capita total expenditure on health (PPP int. $)	0.96607	0.9191
46	Per capita government expenditure on health at average exchange rate (US$)	Per capita government expenditure on health (PPP int. $)	0.96511	0.95007
47	Per capita government expenditure on health (PPP int. $)	Births by caesarean section (%)	0.96454	0.83766
48	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) both sexes	Births by caesarean section (%)	0.96312	0.93682
49	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) male	Births by caesarean section (%)	0.96312	0.91055
50	Number of nursing and midwifery personnel	Number of laboratory health workers	0.96223	0.92393
51	Per capita total expenditure on health (PPP int. $)	Per capita government expenditure on health (PPP int. $)	0.959	0.92738
52	Adult mortality rate (probability of dying between 15 to 60 years per 1000 population) female	Adult mortality rate (probability of dying between 15 to 60 years per 1000 population) both sexes	0.95672	0.93691
53	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) female	Infant mortality rate (per 1000 live births) male	0.95647	0.91055
54	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) lowest wealth quintile	Total fertility rate (per woman)	0.95598	0.83394
55	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) urban	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) highest wealth quintile	0.95598	0.93554
56	Infant mortality rate (per 1000 live births) both sexes	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) highest wealth quintile	0.95598	0.84113
57	Measles immunization coverage among one-year-olds (%) rural	Measles immunization coverage among one-year-olds (%) lowest wealth quintile	0.95443	0.9133
58	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) male	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) urban	0.95145	0.8934
59	Measles immunization coverage among one-year-olds ratio highest-lowest educational level of mother	Measles immunization coverage among one-year-olds difference highest-lowest educational level of mother	0.95087	0.95087
60	Measles immunization coverage among one-year-olds (%) lowest wealth quintile	Measles immunization coverage among one-year-olds (%) lowest educational level of mother	0.9508	0.9508
61	Measles immunization coverage among one-year-olds ratio highest-lowest wealth quintile	Measles immunization coverage among one-year-olds (%) lowest wealth quintile	0.94988	0.94988
62	Population proportion under 15 (%)	Population median age (years)	0.9491	0.93352
63	Births attended by skilled health personnel ratio urban–rural	Births attended by skilled health personnel (%) rural	0.94807	0.94807
64	Years of life lost to non-communicable diseases (%)	Years of life lost to communicable diseases (%)	0.94796	0.93907
65	Prevalence of HIV among adults aged >=15 years (per 100 000 population)	Deaths due to HIV/AIDS (per 100 000 population per year)	0.9468	0.8938
66	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) rural	Total fertility rate (per woman)	0.9466	0.85232
67	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) both sexes	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) urban	0.94505	0.8934
68	Infant mortality rate (per 1000 live births) male	Births by caesarean section (%)	0.945	0.91426
69	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) male	Infant mortality rate (per 1000 live births) female	0.94346	0.91962
70	Life expectancy at birth (years) both sexes	Adult mortality rate (probability of dying between 15 to 60 years per 1000 population) both sexes	0.94294	0.94294

(*continued on next page*)

Table 3 (*continued*)

71	Measles immunization coverage among one-year-olds difference urban–rural	Children aged 6–59 months who received vitamin A supplementation (%)	0.94268	0.47962
72	Population using solid fuels (%) urban	Population using solid fuels (%) rural	0.9363	0.88081
73	Population with sustainable access to improved sanitation (%) total	Population with sustainable access to improved sanitation (%) rural	0.93574	0.93574
74	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) female	Births by caesarean section (%)	0.9352	0.87911
75	Deaths among children under five years of age due to other causes (%)	Women who have had PAP smear (%)	0.93355	0.80876
76	Infant mortality rate (per 1000 live births) both sexes	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) rural	0.93125	0.85533
77	Infant mortality rate (per 1000 live births) male	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) rural	0.93125	0.85533
78	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) female	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) urban	0.92957	0.85358
79	Life expectancy at birth (years) female	Adult mortality rate (probability of dying between 15 to 60 years per 1000 population) female	0.92924	0.85989
80	Healthy life expectancy (HALE) at birth (years) female	Births by caesarean section (%)	0.92837	0.86829
81	Infant mortality rate (per 1000 live births) both sexes	Healthy life expectancy (HALE) at birth (years) both sexes	0.92816	0.82657
82	Infant mortality rate (per 1000 live births) male	Infant mortality rate (per 1000 live births) female	0.92654	0.91846
83	Per capita total expenditure on health at average exchange rate (US$)	Per capita government expenditure on health (PPP int. $)	0.92593	0.92593
84	Births attended by skilled health personnel (%) rural	Births attended by skilled health personnel (%) lowest wealth quintile	0.92413	0.87684
85	Births attended by skilled health personnel ratio urban-rural	Births attended by skilled health personnel (%) lowest wealth quintile	0.92413	0.85161
86	Years of life lost to communicable diseases (%)	Births by caesarean section (%)	0.92098	0.87911
87	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) female	Healthy life expectancy (HALE) at birth (years) both sexes	0.92026	0.82695
88	Deaths due to HIV/AIDS (per 100 000 population per year)	Laboratory health workers density (per 10 000 population)	0.9188	0.9188
89	Measles immunization coverage among one-year-olds (%) urban	Measles immunization coverage among one-year-olds (%) lowest wealth quintile	0.91871	0.73843
90	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) male	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) highest educational level of mother	0.91866	0.77368
91	Births attended by skilled health personnel ratio urban-rural	Births attended by skilled health personnel difference urban-rural	0.91829	0.8386
92	Infant mortality rate (per 1000 live births) male	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) highest wealth quintile	0.91774	0.86589
93	Deaths among children under five years of age due to diarrheal diseases (%)	Births by caesarean section (%)	0.91766	0.83028
94	Healthy life expectancy (HALE) at birth (years) male	Healthy life expectancy (HALE) at birth (years) both sexes	0.91763	0.88844
95	Infant mortality rate (per 1000 live births) both sexes	Healthy life expectancy (HALE) at birth (years) female	0.91645	0.81057
96	Physicians density (per 10 000 population)	Women who have had mammography (%)	0.91543	0.86767
97	Life expectancy at birth (years) both sexes	Adult mortality rate (probability of dying between 15 to 60 years per 1000 population) female	0.91394	0.89679
98	Healthy life expectancy (HALE) at birth (years) female	Healthy life expectancy (HALE) at birth (years) both sexes	0.91259	0.89098
99	Deaths among children under five years of age due to malaria (%)	Under-5 mortality rate (Probability of dying aged 5 years per 1000 live births) lowest wealth quintile	0.91238	0.91238
100	Life expectancy at birth (years) both sexes	Healthy life expectancy (HALE) at birth (years) male	0.91137	0.88843

Table 4
The statistic values of difference between iMIC and MIC.

$MAX_{d\text{-}value}$	$MIN_{d\text{-}value}$	$AVG_{d\text{-}value}$	$d\text{-}value > 0$
0.46306	0	0.046472	79%

Table 6 is the rank comparison between iMIC and MIC of 11 relationships in Table 5. From Table 6, the relationship A ranks at 2763 when using original algorithm MIC, but increased to 71 after using the proposed algorithm iMIC. The rank of other relationships also changed to a large extent. Table 6 indicates that iMIC algorithm could detect some important relationships that may be ignored by original algorithm.

The scatter plots of these 11 pairwise variables are described in Fig. 7.

We know from Fig. 7, that B, C, D, E, F, G, H, J are positive correlation, while I and K are negative correlation obviously.

Relationship B says with measles immunization coverage among one-year-olds urban improves, measles immunization coverage among one-year-olds lowest wealth quintile also raised, this suggests that the measles immunization rate of

Table 5
The relationships of *d-values* that are more than 0.1.

Relationship	Var-A	Var-B	iMIC	MIC
A	Children aged 6–59 months who received vitamin A supplementation (%)	Measles immunization coverage among one-year-olds difference urban-rural	0.94268	0.47962
B	Measles immunization coverage among one-year-olds (%) urban	Measles immunization coverage among one-year-olds (%) lowest wealth quintile	0.91871	0.73843
C	Under-5 mortality rate (probability of dying by age 5 per 1000 live births) male	Under-5 mortality rate (Probability of dying aged at 5 years per 1000 live births) highest educational level of mother	0.91866	0.77368
D	Per capita government expenditure on health (PPP int. $)	Births by caesarean section (%)	0.96454	0.83766
E	Deaths among children under five years of age due to other causes (%)	Women who have had PAP smear (%)	0.93355	0.80876
F	Under-5 mortality rate (Probability of dying aged at 5 years per 1000 live births) lowest wealth quintile	Total fertility rate (per woman)	0.95598	0.83394
G	Infant mortality rate (per 1000 live births) both sexes	Under-5 mortality rate (Probability of dying aged at 5 years per 1000 live births) highest wealth quintile	0.95598	0.84113
H	Under-5 mortality rate (Probability of dying aged at 5 years per 1000 live births) urban	Under-5 mortality rate (Probability of dying aged at 5 years per 1000 live births) highest educational level of mother	0.96698	0.85348
I	Infant mortality rate (per 1000 live births) both sexes	Healthy life expectancy (HALE) at birth (years) female	0.91645	0.81057
J	Population with sustainable access to improved drinking water sources (%) total	Births by caesarean section (%)	0.99335	0.89027
K	Infant mortality rate (per 1000 live births) both sexes	Healthy life expectancy (HALE) at birth (years) both sexes	0.92816	0.82657

Table 6
The comparison of rank scores and time costs.

Relationship	*d-value*	Rank iMIC	Rank MIC	MIC_time (ms)	iMIC_time (ms)
A	0.46306	71	2963	9	10
B	0.18028	89	459	6	8
C	0.14498	90	321	7	7
D	0.12688	47	155	7	8
E	0.12479	75	210	5	5
F	0.12204	54	165	4	4
G	0.11485	56	147	2	3
H	0.1135	44	126	2	3
I	0.10588	95	207	8	10
J	0.10308	18	86	2	2
K	0.10159	81	183	7	9

urban children reflects the attitude of whole country to measles vaccine. When the nation pays great attention on measles vaccination, the children in poor area can also get measles immunization. We can learn from Relationship C that the higher mortality rate of male under-5, the higher under-5 mortality rate with highest educational level of mother, which suggests that male children died under the age of five due to bad social conditions. Therefore regardless of the extent to which the education of mother, the overall social environment will affect the survival rate of children.

Relationship D tells us that the more per capita government expenditure on health, the more births by caesarean section. Because per capita government expenditure on health always means the stage of social development. Women in developed country could choose birth by caesarean section, which is expensive, to avoid difficult labor.

Relationship E means weak-positive correlation between woman who have had PAP smear rate and death rate of children under five years of age due to other causes.

Relationship F shows that the total fertility rate per woman increases, under-5 mortality rate lowest wealth quintile also increases. Woman's fertility rate is high, which means that she may not have enough care of every child, even possibly to affect the entire population of the country, which leads to insufficient per capita resource allocation, social development level lags behind and then unable to provide basic guarantee for the children.

Relationship G means positive correlation between infant mortality rate and under-5 mortality rate from highest wealth is quintile. That is because high infant mortality rate usually represents the poor medical level of the society. Even rich families are also restricted by the backward social conditions. With the same trend of relationship C, obviously relation H is reasonable. Relationship J shows that higher population with sustainable access to improved drinking water source total always presents the stage of social development, as shown in relationship D, the higher births by caesarean section.

Relationship I is on behalf of the negative correlation between infant mortality rate on both sexes and healthy life expectancy at birth female. Just like correlation G, the high infant mortality rate always means the backward economy of

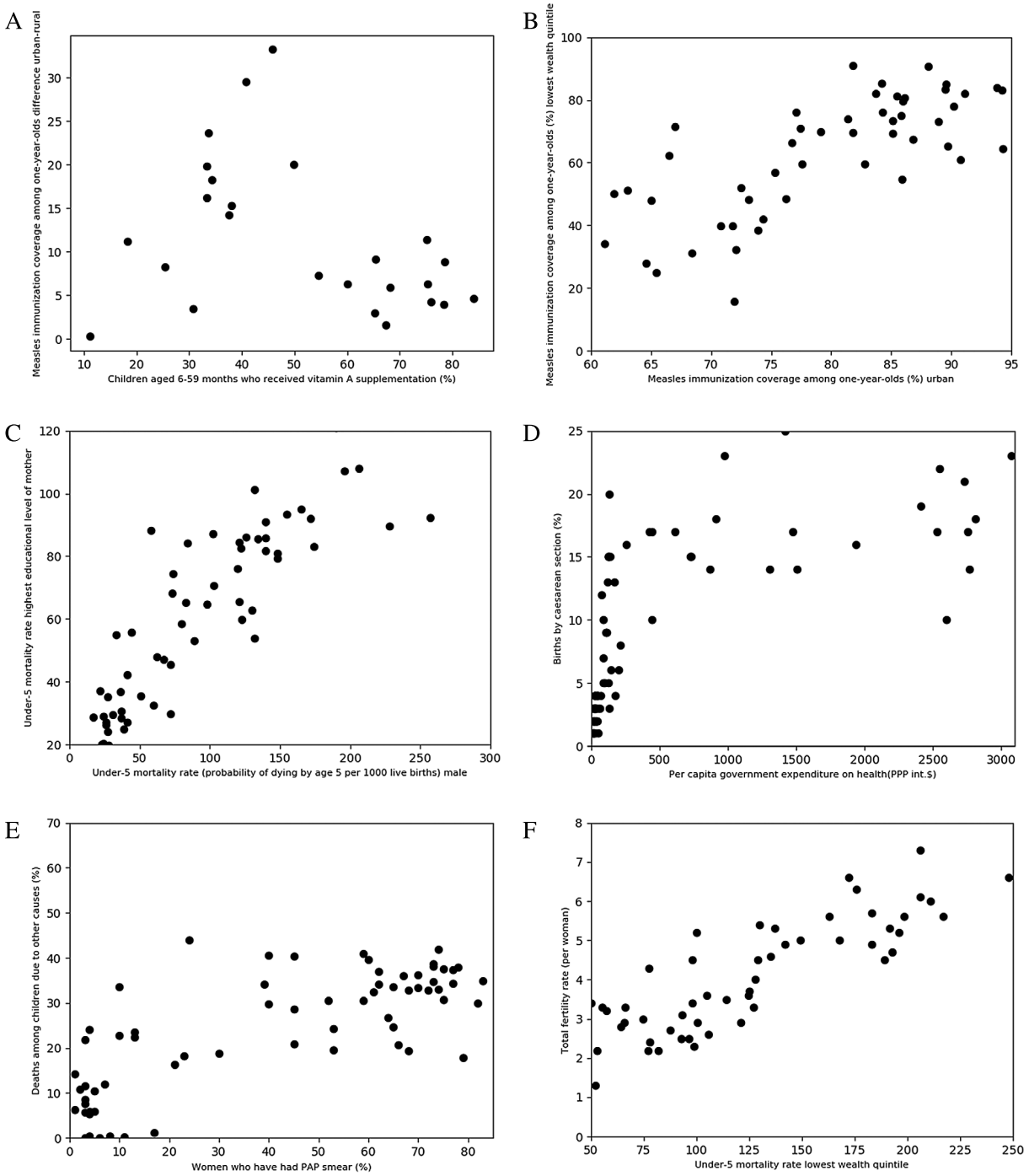


Fig. 7. Scatter plot of pairwise variables.

society. And it will lead to the undeveloped condition for healthcare, which shorten healthy life expectancy at birth female. Similar to relationship I, correlation K makes sense, infant mortality rate on both sexes is higher, is the lower health life expectancy and at both sexes.

Relationship A in Fig. 7 is special. It shows us that the accuracy of information coefficient under our iMIC algorithm doubles that under Reshef's algorithm [9]. And Fig. 8 provides the best grid partition from two algorithms. Obviously, the 3×2 grid partition constructed by iMIC aggregates more points in a few grids.

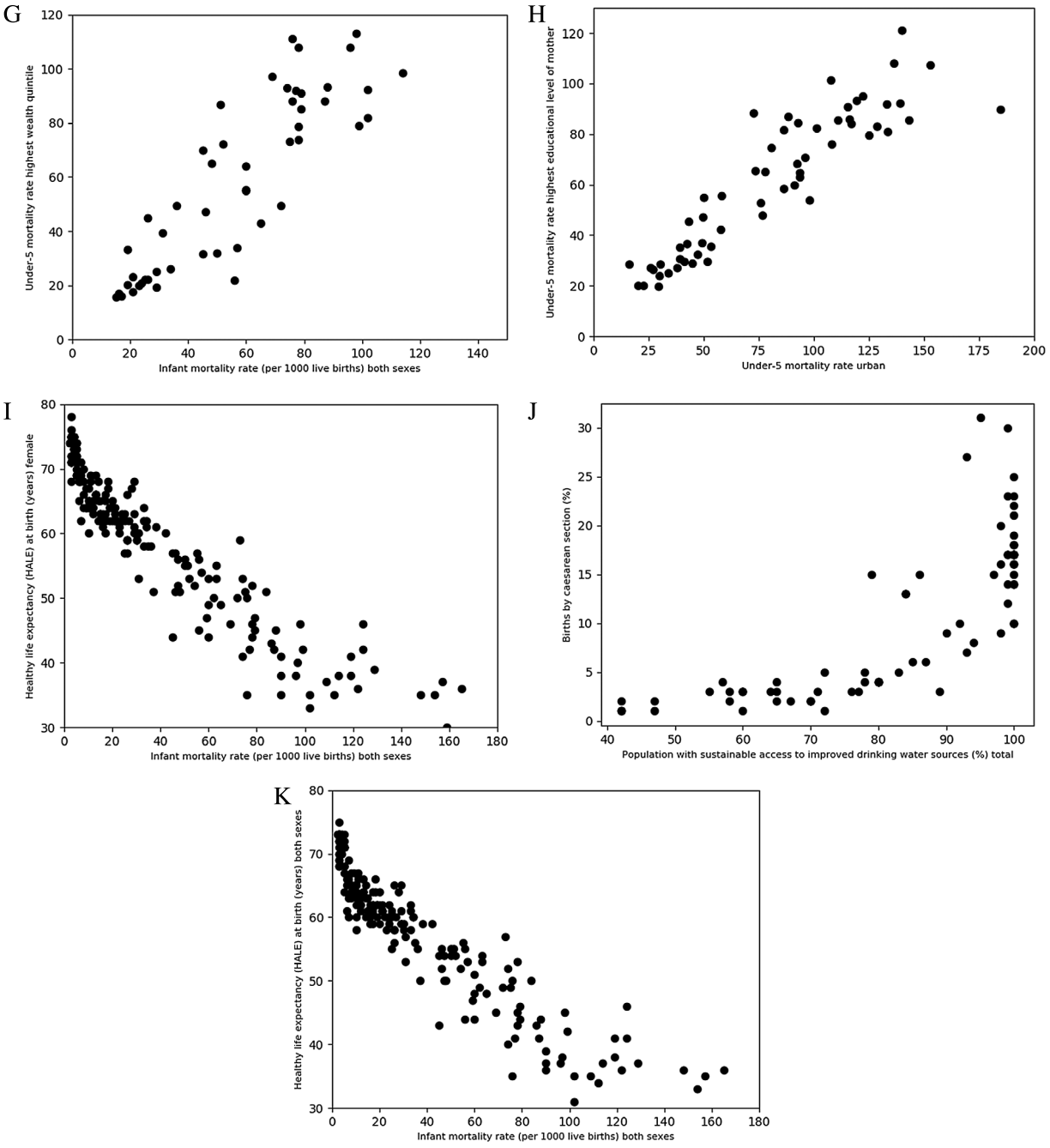
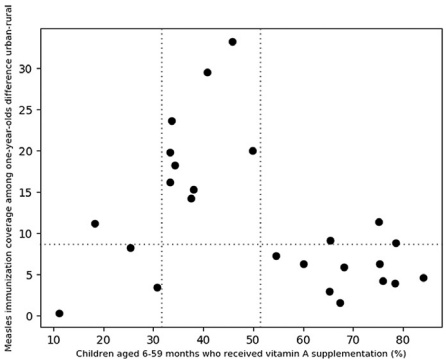


Fig. 7. (*continued*)

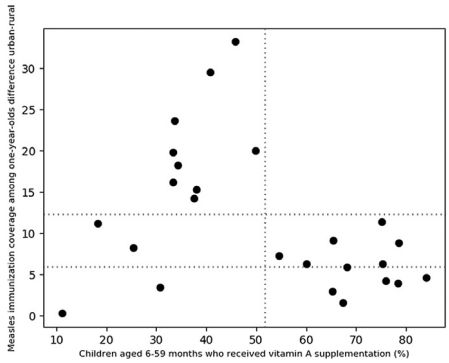
5. Conclusion

Inspired by the idea of gradient descent to find a solution to local extremum, the iMIC has been proposed to fast approximate the true value of the MIC with better accuracy in this paper. It may find out a better partition on the y-axis than the simple equipartition by quadratic optimization as an alternative to the exhaustive searching method of Reshef et al. [9,14], because the equipartition on the y-axis is a subset of the quadratic optimization on the y-axis with the same partition on the x-axis. The performance improvement of iMIC was verified experimentally.

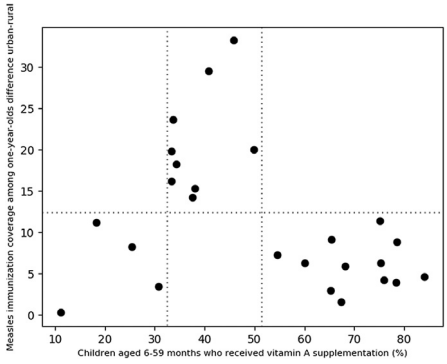
Finally, we select 169 comprehensive evaluation indexes about 202 countries from the data sets of World Health Organization (WHO), to compare the results created by two different algorithms. By practical application, we verify the ability of the algorithm. And, it is proved that the proposed algorithm is superior to original algorithm.



(a) MIC = 0.47962.



(b) MIC = 0.47962.



(c) iMIC = 0.94268.

Fig. 8. Comparison of grid partition solution about relationship A. (a) and (b) are the best grid partition computed by original algorithm, MIC=0.47962. (c) shows the best grid partition calculated by iMIC algorithm, iMIC = 0.94268.

In the future, the large amount of data is being generated in various fields. iMIC can detect the nonlinear relationships in big data. It will be the next study to calculate the global maximum value of the MIC with acceptable time complexity, which may reveal the essence of the iMIC much deeply.

Acknowledgments

Many thanks to Wenzhong Shi for his valuable comments. Also thank Ziqiang Yuan and Chuanlu Liu for their helping the paper revision. This work was supported by National Natural Science Fund of China (61472039), National Key Research and Development Plan of China (2016YFC0803000, 2016YFB0502604), and Frontier and interdisciplinary innovation program of Beijing Institute of Technology (2016CX11006).

References

[1] T. Hastie, Elements of Statistical Learning: Datamining, Inference, and Prediction, (second ed.), Springer, 2009, pp. 693–694.
[2] D. Howe, M. Costanzo, P. Fey, T. Gojobori, L. Hannick, W. Hide, D.P. Hill, R. Kania, M. Schaeffer, S.S. Pierre, The future of biocuration, Nature 455 (7209) (2008) 47–50.
[3] D.R. Li, S.L. Wang, D.Y. Li, Spatial Data Mining, Springer, Berlin, 2015.
[4] V. Mayer-Schönberger, K. Cukier, Big data: A Revolution That Will Transform How We Live, Work, and Think, Houghton Mifflin Harcourt, 2013.
[5] S. Wang, H. Yuan, Spatial data mining: a perspective of big data, Int. J. Data Warehouse Min. 10 (4) (2014) 50–70.
[6] F. Frankel, R. Reid, Big data: distilling meaning from data, Nature 455 (7209) (2008) 30.
[7] J.D. Ullman, J. Leskovec, A. Rajaraman, Mining of Massive Datasets, Cambridge University Press, 2011.
[8] M.A. Hernández, S.J. Stolfo, Real-world data is dirty: data cleansing and the merge/purge problem, Data Min. Knowl. Discov. 2 (1) (1998) 9–37.
[9] D.N. Reshef, Y.A. Reshef, H.K. Finucane, S.R. Grossman, G. McVean, P.J. Turnbaugh, E.S. Lander, M. Mitzenmacher, P.C. Sabeti, Detecting novel associations in large data sets, Science 334 (6062) (2011) 1518–1524.
[10] G.J. Szkely, M.L. Rizzo, N.K. Bakirov, Measuring and testing dependence by correlation of distances, Ann. Statist. 35 (6) (2007) 2769–2794.
[11] G.J. Székely, M.L. Rizzo, et al., Brownian distance covariance, Ann. Appl. Stat. 3 (4) (2009) 1236–1265.
[12] N. Simon, R. Tibshirani, Finding correlations in big data, Nature Biotechnol. 30 (4) (2012) 334–335.
[13] [N. Simon, R. Tibshirani, Comment on "detecting novel associations in large data sets" by Reshef et al, Science Dec 16, 2011, arXiv preprint, arXiv: 1401.7645.

[14] D. Reshef, Y. Reshef, M. Mitzenmacher, P. Sabeti, Equitability Analysis of the Maximal Information Coefficient, with Comparisons, 2013. arXiv: 1301.6314.
[15] S. Wang, Y. Zhao, Y. Shu, W. Shi, Improved Approximation Algorithm for Maximal Information Coefficient, Int. J. Data Warehouse Min. 13 (1) (2017) 76–93.
[16] D. Li, S. Wang, W. Gan, D. Li, Data field for hierarchical clustering, Int. J. Data Warehouse Min. 7 (4) (2011) 43–63.
[17] X. Wu, X. Zhu, G.-Q. Wu, W. Ding, Data mining with big data, IEEE Trans. Knowl. Data Eng. 26 (1) (2014) 97–107.
[18] E.H. Linfoot, An informational measure of correlation, Inf. Control 1 (1) (1957) 85–89.
[19] C. Lin, T. Miller, D. Dligach, R. Plenge, E. Karlson, G. Savova, Maximal information coefficient for feature selection for clinical document classification, in: ICML Workshop on Machine Learning for Clinical Data. Edingburgh, UK, 2012.
[20] J. Das, J. Mohammed, H. Yu, Genome-scale analysis of interaction dynamics reveals organization of biological networks., Bioinformatics 28 (14) (2012) 1873–1878.
[21] T.K. Anderson, W.W. Laegreid, F. Cerutti, F.A. Osorio, E.A. Nelson, J. Christopherhennings, T.L. Goldberg, Ranking viruses: measures of positional importance within networks define core viruses for rational polyvalent vaccine development, Bioinformatics 28 (12) (2012) 1624.
[22] T.V. Karpinets, B.H. Park, E.C. Uberbacher, Analyzing large biological datasets with association networks, Nucleic Acids Res. 40 (17) (2012) e131.
[23] X. Zhang, K. Liu, Z.P. Liu, B. Duval, J.M. Richer, X.M. Zhao, J.K. Hao, L. Chen, NARROMI: a noise and redundancy reduction technique improves accuracy of gene regulatory network inference, Bioinformatics 29 (1) (2013) 106–113.
[24] T. Speed, A Correlation for the 21st Century, Science 334 (6062) (2011) 1502–1503.
[25] Y. Sun, M. Kirley, S.K. Halgamuge, Quantifying variable interactions in continuous optimization problems, IEEE Trans. Evol. Comput. 21 (2) (2017) 249–264.
[26] E.B. Gorgens, R. Valbuena, L.C.E. Rodriguez, A Method for Optimizing Height Threshold When Computing Airborne Laser Scanning Metrics, Photogramm. Eng. Remote Sens. 83 (5) (2017) 343–350.
[27] C. L.T. Borges, J.A. Dias, A model to represent correlated time series in reliability evaluation by non-sequential Monte Carlo simulation, IEEE Trans. Power Syst. 32 (2) (2017) 1511–1519.
[28] Y. Fan, G. Huang, Y. Li, X. Wang, Z. Li, L. Jin, Development of PCA-based cluster quantile regression (PCA-CQR) framework for streamflow prediction: Application to the Xiangxi river watershed, China, Appl. Soft Comput. 51 (2017) 280–293.
[29] F. Shao, K. Li, A graph model for preventing railway accidents based on the maximal information coefficient, Internat. J. Modern Phys. B (2016) 1750010.
[30] Y. Zhang, W. Zhang, Y. Xie, Improved heuristic equivalent search algorithm based on maximal information coefficient for Bayesian network structure learning, Neurocomputing 117 (2013) 186–195.
[31] S. Wang, Y. Zhao, Analysing large biological data sets with an improved algorithm for MIC, Int. J. Data Min. Bioinform. 13 (2) (2014).
[32] M. Gorfine, R. Heller, Y. Heller, Comment on detecting novel associations in large data sets, 2012. (Unpublished). Available at http://emotion.technion.ac.il/gorfinm/files/science6.pdf on 11.11.12.
[33] R. Heller, Y. Heller, M. Gorfine, A consistent multivariate test of association based on ranks of distances, 2012. arXiv:1201.3522.
[34] J.B. Kinney, G.S. Atwal, Equitability, mutual information, and the maximal information coefficient, Proc. Natl. Acad. Sci. 111 (9) (2014) 3354–3359.
[35] J. Schmidhuber, Deep Learning in Neural Networks: An Overview, Neur. Netw. Off. J. Int. Neur. Netw. Soc. 61 (2014) 85–117.