2020 IEEE 13th International Conference on Software Testing, Validation and Verification (ICST)

Poster: Is Euclidean Distance the best Distance Measurement for Adaptive Random Testing?

Rubing Huang, Chenhui Cui, Weifeng Sun
School of Computer Science and Communication Engineering
Jiangsu University
Zhenjiang, China
rbhuang@ujs.edu.cn

Dave Towey
School of Computer Science
University of Nottingham Ningbo China
Ningbo, China
dave.towey@nottingham.edu.cn

***Abstract*—Adaptive random testing (ART) aims at enhancing the testing effectiveness of random testing (RT) by more evenly spreading test cases over the input domain. Many ART methods have been proposed, based on various, different notions. For example, *distance-based ART* (DART) makes use of the concept of distance to implement ART, attempting to generate new test cases that are far away from previously executed ones. The Euclidean distance has been a popular choice of distance metric, used in DART to evaluate the differences between test cases. However, is the Euclidean distance the most suitable choice for DART? To answer this question, we conducted a series of simulations to investigate the impact that the Euclidean distance, and its many variations, has on the testing effectiveness of DART. The results show that when the dimensionality of the input domain is low, the Euclidean distance may indeed be a good choice. However, when the dimensionality is high, it appears to be less suitable.**

***Index Terms*—adaptive random testing, Euclidean distance, dissimilarity metrics**

I. Introduction

Software testing is an important technique for evaluating and improving software quality. One basic approach involves randomly generating test cases from the *input domain* (i.e., the set of all possible program inputs): This is called *random testing* (RT) [1]. Although RT has been widely applied to many applications such as database systems [2], it is still the focus of many questions over its testing effectiveness [3].

Many studies have investigated on how to enhance the effectiveness of RT, including in terms its ability to detect failures. *Adaptive random testing* (ART) [4], for example, aims to enhance RT's testing effectiveness by more evenly spreading test cases over the input domain, thus generating test cases with greater diversity [5]. Based on various different notions and intuitions, a number of approaches have been proposed to implement the principles of ART. *Distance-based ART* (DART) [6]–[8], for example, is an ART implementation based on the concept of distance. DART selects an element from among several candidates as the next test case such that it is the farthest away from the previously executed test cases.

The Euclidean distance is a special case of Minkowski p-distance (where p is equal to 2) [9], and is a commonly used metric to evaluate the difference between two points. The Euclidean distance has also been extensively used for the evaluation of test cases in software testing, including for DART. Although both simple and popular, to date, no studies have specifically examined whether or not the Euclidean distance is the most suitable for DART.

To answer this question, in this paper we report on a series of simulation studies conducted to investigate the impact of different versions of Minkowski p-distance (including Euclidean distance) on the testing effectiveness of DART. The simulation results show that when the input domain is of low dimensionality, then different p values for Minkowski p-distance have minimal differences in terms of fault detection capability, indicating that the Euclidean distance may be a good choice for DART. However, for higher dimension input domains, DART with lower p values provides better performance, which means that the Euclidean distance is not suitable.

The rest of this paper is organized as follows: Section II introduces some background information. Section III presents the simulation studies, including discussion of the results. Section IV addresses the threats to validity, and, finally, Section V concludes the paper.

II. Background

In this section, we present some background information about ART and the Minkowski p-distance [9].

A. Adaptive Random Testing

Adaptive random testing (ART) [4] enhances RT by attempting to ensure a more even spread of test cases over the input domain. This is motivated by observations of the contiguity of failure regions — the set of all inputs that cause a software failure, also known as the *failure-causing inputs*. If failure regions are contiguous, then non-failure regions should also be contiguous [10]. Therefore, if a test input t_i is a non-failure-causing input, then its neighbours are also unlikely to cause a failure. Thus, candidate test inputs that are far away from t_i may have a higher probability of causing failure than t_i's neighbours. Because there are many ways to achieve an even spread of test cases, according to various notions, many different ART algorithms have been proposed [4]. Using the notion of *distance*, for example, *Distance-based ART* (DART) [6]–[8] has been introduced. In this paper, we focus on one implementation of DART, the *fixed-size-candidate-set* version of ART (FSCS-ART) [6].

978-1-7281-5778-8/20/$31.00 ©2020 IEEE
DOI 10.1109/ICST46399.2020.00049

FSCS-ART makes use of two sets of test cases: the *candidate set* (denoted C); and the *executed set* (denoted E). C contains k test inputs that are randomly generated from the input domain. E contains those previously executed test cases that did not reveal any failures. The test case selection criterion of FSCS-ART is defined as follows: A candidate $c \in C$, is chosen as the next test case such that it satisfies, $\forall c' \in C$, $\forall e \in E$,

$$\min_{e \in E} dist(c, e) \geq \min_{e \in E} dist(c', e) \tag{1}$$

where $dist(c, e)$ measures the distance (difference) between c and e.

B. Minkowski p-Distance

The Minkowski p-distance between N-dimensional points $x = (x_1, x_2, \cdots, x_N)$ and $y = (y_1, y_2, \cdots, y_N)$ is defined as:

$$dist_p(x, y) = \left(\sum_{i=1}^{N} |x_i - y_i|^p \right)^{1/p} \tag{2}$$

Obviously, p is a constant that must be fixed in advance. When $p = 2$, for example, Minkowski p-distance becomes the Euclidean distance; when $p = 1$, it becomes the Manhattan distance. In this paper, we examine different p values as different variations of Minkowski p-distance (including the Euclidean distance).

III. Simulations

In this section, we report on a series of simulation studies to investigate how different p values of Minkowski p-distance influence ART, and thus answer the research question of whether or not the Euclidean distance is most suitable.

A. Simulations Setup

The studies involved simulations of faulty programs under different scenarios. Generally speaking, a faulty program has two basic features: the *failure rate* and the *failure pattern*. The failure rate, denoted θ, can be defined as the ratio of the number of failure-causing inputs to the number of all possible inputs. The failure pattern refers to the shapes of the failure regions together with their distributions over the input domain. Both the failure rate and pattern are fixed after coding, but unknown before testing. Therefore, the failure rate and pattern were predefined in the simulations, but the failure regions were located randomly within the input domain. When a test case was chosen from inside a failure region, then a failure was considered to have been detected.

Without loss of generality, we considered a d-dimensional unit input domain $\mathcal{D}$ (of size of 1.0, $\mathcal{D} = [0, 1.0)^d$) in our simulations: $\mathcal{D} = \{(x_1, x_2, \cdots, x_d) | 0 \leq x_i < 1.0, i = 1, 2, \cdots, d\}$. Following previous studies [11], we considered three different failure patterns: *block*; *strip*; and *point*. The block pattern simulations involved a single hypercube failure region (such as a square in a two-dimensional input domain). For the strip pattern, we randomly selected d adjacent edges of the input domain, randomly choosing a point on each edge. After that, we translated each of d points to generate another point on the adjacent, and then connected these $2d$ points to construct the strip pattern with the desired area/volume. The point patterns involved 25 equally-sized, non-overlapping hypercubes. The values of each parameter were assigned as follows:

- Dimension d: 1, 2, 3, 4, 5, 7, and 10.
- Failure rate θ: 0.01, 0.005, 0.001, and 0.0005.
- Failure pattern: block, strip, and point patterns.
- Parameter p for Minkowski p-distance: 0.1, 0.3, 0.5, 0.8, 1.0, 2.0, 4.0, 7.0, and 10.0.

Because of the randomness in ART, we ran each simulation 3000 times, and report the average result.

B. Evaluation Metrics

The *F-measure* is a popular metric for evaluating ART. It is defined as the expected number of test case executions required to detect the first failure [6]. We used the F-measure as the evaluation metric in this study, with F_{RT} and F_{ART} denoting the F-measure of RT and ART, respectively; and *ART F-ratio* denoting the ratio of F_{ART} to F_{RT} (which shows the improvement of ART over RT). The smaller the ART the F-ratio is, the better the ART performance is.

C. Results

Tables I to III show the ART F-ratio simulation results for block, strip, and point patterns, respectively. In the tables, the third column shows the ART F-ratio values, and the bold typeface indicates the minimum value on each row. From these tables, we have the following observations:

(1) Different p values may lead to different performances for ART, and there is no p value that is always best, regardless of dimensions, failure rates, and failure patterns.

(2) For the block patterns, when the dimension d is small (such as $d = 1, 2, 3, 4$), then the differences among the ART F-ratios for different p values seem small, regardless of failure rates — the maximum difference is about 0.15. However, this changes when d is greater than or equal to 5: As the p values increase, FSCS-ART generally has worse performance, with the differences becoming larger. FSCS-ART with $p = 0.1$ has the best performance overall, and the maximum difference reaches more than 4.50.

(3) For the strip patterns, the differences among the ART F-ratios are very small for different p values, regardless of failure rates, dimensions, and p values. When $d = 1$, the ART F-ratio values are about 0.55. For other values of d, however, the ART F-ratio values are around 1.0, which means that FSCS-ART has similar performance to RT.

(4) The observations for the point patterns are similar to those for block patterns, i.e., when d is small, the ART F-ratio differences are small. However, when d becomes larger, the differences also become larger. Similarly, FSCS-ART with $p = 0.1$ performs best overall.

In answer to the research question, although it is evident that FSCS-ART with Euclidean distance (i.e., $p = 2.0$) may sometimes have the best ART F-ratio values, overall, it generally does not provide the best performance.

TABLE I
ART F-RATIO RESULTS FOR *Block Patterns*

Failure Rate (θ)	**Dimension (d)**	**Parameter Value (p) for Minkowski p-distance**									***Maximum Difference***
		$p=0.1$	$p=0.3$	$p=0.5$	$p=0.8$	$p=1.0$	$p=2.0$	$p=4.0$	$p=7.0$	$p=10.0$	
$\theta=0.01$	$d=1$	**0.5594**	0.5647	0.5631	0.5611	0.5668	0.5722	0.5769	0.5703	0.5611	*0.0175*
	$d=2$	0.7298	0.7008	0.6945	0.6801	0.6733	0.6747	**0.6552**	0.6696	0.6850	*0.0746*
	$d=3$	0.8439	0.8210	0.8304	**0.8173**	0.8477	0.8194	0.8700	0.8449	0.8628	*0.0527*
	$d=4$	**0.9743**	1.0123	1.0093	1.0162	1.0474	1.0753	1.1148	1.1121	1.1365	*0.1622*
	$d=5$	**1.0182**	1.1512	1.1986	1.2384	1.2729	1.3221	1.3790	1.4325	1.4407	*0.4225*
	$d=7$	**1.0103**	1.4817	1.6441	1.7014	1.8171	2.0206	2.2130	2.3450	2.5064	*1.4961*
	$d=10$	**1.0020**	1.9997	2.2764	2.6761	2.9121	3.8455	4.8611	5.1685	5.5783	*4.5763*
$\theta=0.005$	$d=1$	0.5658	0.5711	0.5629	0.5646	0.5639	0.5641	0.5586	**0.5557**	0.5563	*0.0154*
	$d=2$	0.7020	0.6818	0.6630	0.6659	0.6624	0.6547	**0.6532**	0.6710	0.6548	*0.0488*
	$d=3$	0.8538	0.8051	0.8133	0.8111	**0.8040**	0.8077	0.8133	0.8378	0.8286	*0.0498*
	$d=4$	0.9896	0.9797	0.9590	0.9830	**0.9531**	1.0043	1.0394	1.0460	1.0468	*0.0937*
	$d=5$	**0.9946**	1.1178	1.1358	1.1741	1.1746	1.2438	1.3289	1.3225	1.3326	*0.3380*
	$d=7$	**1.0142**	1.4449	1.4864	1.6470	1.7013	1.8852	2.0327	2.2013	2.2167	*1.2025*
	$d=10$	**1.0086**	1.9591	2.2498	2.5468	2.8178	3.5806	4.3353	4.8331	4.8433	*3.8347*
$\theta=0.001$	$d=1$	0.5671	0.5625	0.5581	0.5664	0.5586	**0.5511**	0.5627	0.5681	0.5580	*0.0170*
	$d=2$	0.7135	0.6607	0.6523	0.6502	**0.6278**	0.6357	0.6400	0.6438	0.6395	*0.0857*
	$d=3$	0.8370	0.7620	0.7538	0.7500	0.7569	**0.7462**	0.7764	0.7625	0.7645	*0.0908*
	$d=4$	0.9666	0.8920	**0.8698**	0.8734	0.8948	0.8963	0.9156	0.9143	0.9456	*0.0968*
	$d=5$	**0.9924**	1.0575	1.0641	1.0358	1.0573	1.0844	1.1284	1.1627	1.1857	*0.1933*
	$d=7$	**0.9887**	1.3094	1.4226	1.4586	1.4698	1.6005	1.7031	1.8143	1.8237	*0.8350*
	$d=10$	**1.0057**	1.8615	2.0482	2.3533	2.4071	2.8741	3.1375	3.5314	3.5589	*2.5532*
$\theta=0.0005$	$d=1$	0.5679	0.5550	0.5532	0.5599	**0.5496**	0.5540	0.5565	0.5635	0.5513	*0.0183*
	$d=2$	0.7040	0.6479	0.6430	0.6354	0.6397	**0.6313**	0.6335	0.6360	0.6352	*0.0727*
	$d=3$	0.8336	0.7731	0.7490	**0.7427**	0.7434	0.7495	0.7646	0.7458	0.7779	*0.0909*
	$d=4$	0.9570	0.8969	0.8882	**0.8518**	0.8647	0.8924	0.8685	0.8872	0.8976	*0.1052*
	$d=5$	**0.9980**	1.0189	1.0382	1.0124	1.0219	1.0663	1.0710	1.1304	1.1192	*0.1324*
	$d=7$	**1.0335**	1.2940	1.2941	1.3784	1.4238	1.5039	1.6939	1.6559	1.7420	*0.7085*
	$d=10$	**0.9725**	1.8123	1.9529	2.1555	2.2908	2.7049	3.0199	3.1290	3.2155	*2.2430*

TABLE II
ART F-RATIO RESULTS FOR *Strip Patterns*

Failure Rate (θ)	**Dimension (d)**	**Parameter Value (p) for Minkowski p-distance**									***Maximum Difference***
		$p=0.1$	$p=0.3$	$p=0.5$	$p=0.8$	$p=1.0$	$p=2.0$	$p=4.0$	$p=7.0$	$p=10.0$	
$\theta=0.01$	$d=1$	**0.5594**	0.5647	0.5631	0.5611	0.5668	0.5722	0.5769	0.5703	0.5611	*0.0175*
	$d=2$	0.9793	0.9359	0.9422	0.9495	**0.9008**	0.9392	0.9117	0.9156	0.9117	*0.0785*
	$d=3$	0.9701	0.9738	**0.9432**	0.9522	0.9655	0.9639	0.9818	0.9904	1.0005	*0.0573*
	$d=4$	0.9606	0.9870	**0.9436**	0.9531	0.9679	0.9733	1.0164	1.0094	1.0160	*0.0728*
	$d=5$	1.0224	1.0232	**0.9691**	0.9813	0.9975	1.0162	1.0051	1.0092	0.9902	*0.0541*
	$d=7$	1.0407	1.0160	1.0131	1.0153	**0.9945**	1.0185	1.0217	1.0265	1.0340	*0.0462*
	$d=10$	1.0314	1.0487	1.0035	1.0426	**1.0005**	1.0068	1.0196	1.0097	1.0159	*0.0482*
$\theta=0.005$	$d=1$	0.5658	0.5711	0.5629	0.5646	0.5639	0.5641	0.5586	**0.5557**	0.5563	*0.0154*
	$d=2$	1.0254	0.9484	0.9460	0.9604	0.9297	**0.9175**	0.9279	0.9499	0.9225	*0.1079*
	$d=3$	1.0000	0.9846	0.9550	0.9682	0.9820	**0.9404**	0.9886	0.9879	1.0019	*0.0615*
	$d=4$	0.9822	1.0309	0.9553	0.9894	**0.9280**	0.9830	0.9668	0.9687	1.0079	*0.1029*
	$d=5$	1.0227	1.0264	**0.9637**	1.0019	0.9827	1.0210	0.9934	1.0085	0.9907	*0.0627*
	$d=7$	1.0117	1.0169	1.0136	1.0069	**0.9908**	1.0395	1.0433	1.0338	1.0211	*0.0525*
	$d=10$	1.0005	1.0034	1.0124	**0.9861**	1.0155	1.0265	1.0136	1.0151	1.0275	*0.0414*
$\theta=0.001$	$d=1$	0.5671	0.5625	0.5581	0.5664	0.5586	**0.5511**	0.5627	0.5681	0.5580	*0.0170*
	$d=2$	0.9808	0.9775	1.0009	0.9682	0.9720	0.9624	0.9768	**0.9465**	0.9692	*0.0544*
	$d=3$	0.9932	1.0037	0.9653	0.9763	1.0103	**0.9514**	0.9771	0.9731	0.9981	*0.0589*
	$d=4$	1.0153	1.0125	1.0279	1.0234	**0.9605**	0.9982	0.9974	0.9891	0.9858	*0.0674*
	$d=5$	0.9965	1.0234	0.9959	0.9816	0.9819	**0.9791**	1.0249	1.0028	1.0027	*0.0458*
	$d=7$	0.9901	1.0131	**0.9808**	0.9901	1.0122	0.9910	0.9978	0.9925	0.9955	*0.0323*
	$d=10$	0.9919	**0.9797**	0.9854	0.9927	1.0307	1.0083	1.0280	0.9905	1.0233	*0.0510*
$\theta=0.0005$	$d=1$	0.5679	0.5550	0.5532	0.5599	**0.5496**	0.5540	0.5565	0.5635	0.5513	*0.0183*
	$d=2$	0.9861	0.9717	0.9725	0.9642	**0.9489**	0.9770	0.9519	0.9839	0.9764	*0.0372*
	$d=3$	1.0124	0.9757	1.0049	0.9900	0.9842	0.9978	0.9853	**0.9727**	1.0024	*0.0397*
	$d=4$	1.0162	1.0111	1.0215	0.9964	0.9877	1.0038	**0.9729**	0.9937	0.9904	*0.0486*
	$d=5$	**0.9696**	0.9981	0.9945	0.9812	1.0043	1.0236	0.9728	0.9985	0.9930	*0.0540*
	$d=7$	1.0195	1.0132	1.0014	**0.9911**	0.9971	1.0019	1.0186	1.0100	0.9921	*0.0284*
	$d=10$	0.9691	0.9780	0.9930	0.9900	**0.9663**	1.0054	1.0332	0.9783	1.0090	*0.0669*

D. Discussion

When the input domain's dimensionality d is small (i.e., $d \leq 4$), the Euclidean distance ($p = 2.0$) has comparable performance to other distance measures for FSCS-ART. Therefore, it is recommended that the Euclidean distance be used for the input domains with low dimensions. However, when d is higher (i.e., $d \geq 5$), the Euclidean distance appears less suitable for FSCS-ART, and the Minkowski p-distance with $p = 0.1$ is recommended instead. The main reason for this may be explained as follows: When the dimensionality of the input

TABLE III
ART F-RATIO RESULTS FOR *Point Patterns*

Failure Rate (θ)	**Dimension (d)**	**Parameter Value (p) for Minkowski p-distance**									***Maximum Difference***
		$p = 0.1$	$p = 0.3$	$p = 0.5$	$p = 0.8$	$p = 1.0$	$p = 2.0$	$p = 4.0$	$p = 7.0$	$p = 10.0$	
$\theta = 0.01$	$d = 1$	0.9324	0.9663	0.9597	0.9603	0.9942	**0.9296**	0.9495	0.9924	0.9561	*0.0646*
	$d = 2$	0.9647	0.9981	0.9858	**0.9557**	0.9702	0.9808	1.0080	1.0151	0.9926	*0.0594*
	$d = 3$	**1.0415**	1.0597	1.0757	1.0750	1.1192	1.1133	1.1334	1.1463	1.1247	*0.1048*
	$d = 4$	**1.1168**	1.1365	1.1410	1.2162	1.2912	1.2612	1.3367	1.3844	1.2959	*0.2676*
	$d = 5$	**1.2404**	1.2654	1.3258	1.3505	1.4086	1.5037	1.6161	1.6747	1.6086	*0.4343*
	$d = 7$	**1.3714**	1.5063	1.6317	1.7218	1.8910	2.0724	2.3453	2.3897	2.3848	*1.0183*
	$d = 10$	**1.4599**	1.7130	1.8225	2.0203	2.2290	2.5126	2.7860	2.8485	3.0207	*1.5608*
$\theta = 0.005$	$d = 1$	0.9249	0.9291	0.9448	0.9788	0.9379	0.9641	**0.9093**	0.9813	0.9696	*0.0720*
	$d = 2$	**0.9428**	0.9730	0.9550	0.9577	0.9698	0.9802	1.0028	0.9907	1.0055	*0.0627*
	$d = 3$	1.0219	**1.0067**	1.0268	1.0909	1.0865	1.0624	1.0855	1.026	1.0748	*0.0842*
	$d = 4$	**1.0595**	1.1073	1.1466	1.1479	1.1926	1.2410	1.2249	1.2386	1.2315	*0.1815*
	$d = 5$	**1.1955**	1.2468	1.2738	1.3111	1.3252	1.4541	1.5274	1.4947	1.5182	*0.3319*
	$d = 7$	**1.3749**	1.4661	1.5935	1.7034	1.7919	2.1001	2.2169	2.2896	2.2309	*0.9147*
	$d = 10$	**1.4586**	1.7295	1.9214	2.2186	2.3513	2.7447	3.0603	3.2049	3.1867	*1.7463*
$\theta = 0.001$	$d = 1$	0.9302	0.9696	0.9918	0.9426	0.9287	0.9497	0.9554	**0.9044**	0.9396	*0.0874*
	$d = 2$	0.9654	0.9553	0.9742	**0.9402**	0.9769	0.9879	0.9489	0.9613	0.9552	*0.0477*
	$d = 3$	1.0262	1.0105	1.0095	**0.9985**	1.0350	1.0315	1.0294	1.0245	1.0589	*0.0604*
	$d = 4$	1.0702	**1.0522**	1.0722	1.0846	1.1127	1.1151	1.1302	1.1803	1.1970	*0.1448*
	$d = 5$	**1.1413**	1.1490	1.1596	1.1853	1.2495	1.3160	1.2906	1.3632	1.3904	*0.2491*
	$d = 7$	**1.2939**	1.3739	1.4752	1.5904	1.6031	1.7366	1.7977	1.9401	2.0524	*0.7585*
	$d = 10$	**1.5568**	1.7987	1.9511	2.2569	2.3851	2.8527	3.2365	3.5006	3.5081	*1.9513*
$\theta = 0.0005$	$d = 1$	0.9577	0.9447	0.9629	**0.9049**	0.9497	0.9261	0.9934	0.9640	0.9161	*0.0885*
	$d = 2$	**0.9404**	0.9896	0.9802	0.9829	0.9877	0.9542	1.0199	0.9994	1.0120	*0.0795*
	$d = 3$	1.0175	**0.9805**	1.0080	1.0039	1.0079	1.0383	1.0015	1.0144	1.0239	*0.0578*
	$d = 4$	1.0535	1.0744	1.0721	**1.0476**	1.0716	1.0756	1.1039	1.1500	1.1231	*0.1024*
	$d = 5$	**1.1175**	1.1192	1.1584	1.2029	1.2249	1.2526	1.2806	1.2671	1.2792	*0.1631*
	$d = 7$	**1.2449**	1.3615	1.4013	1.4707	1.5443	1.6941	1.8029	1.8599	1.8777	*0.6328*
	$d = 10$	**1.5357**	1.7421	1.9525	2.2141	2.3266	2.7114	3.0086	3.3180	3.3415	*1.8058*

domain is high, the Euclidean distance cannot successfully describe the differences between test inputs.

IV. THREATS TO VALIDITY

The main threat to the validity of our study relates to the design of the simulations: We only used four failure rates, three failure patterns, seven dimensions, and nine p values for the Minkowski p-distance. Furthermore, the study only included simulations, and no experiments with real-life programs. We therefore acknowledge that additional experiments, including both simulations and real-life programs, will need to be conducted in the future.

Another threat to validity relates to the selection of DART implementation. There are many versions of DART, including *restricted random testing* (RRT) [7] and *Markov Chain Monte Carlo based ART* (MCMC-ART) [8], therefore, it will be interesting to investigate the impact of different distance measures on other DART implementations.

V. CONCLUSION

In this study, we conducted simulations to investigate the impact of the Minkowski p-distance on the testing effectiveness of ART, and hence answer whether or not the Euclidean distance is most suitable for ART. The simulation results show that Euclidean distance ($p = 2.0$) is a good choice for input domains with low numbers of dimensions, but that lower p values may be better for high dimensional input domains.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (grant nos. 61502205 and 61872167), and the China Postdoctoral Science Foundation (grant no. 2019T120396). This work is also supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (grant no. KYCX19_1614).

REFERENCES

[1] A. Orso and G. Rothermel, "Software testing: A research travelogue (2000-2014)," in *Proceedings of Future of Software Engineering (FOSE'14)*, 2014, pp. 117–132.

[2] H. Bati, L. Giakoumakis, S. Herbert, and A. Surna, "A genetic approach for random testing of database systems," in *Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB '07)*, 2007, pp. 1243–1251.

[3] G. J. Myers, *The Art of Software Testing*. John Wiley & Sons: New York, 2004.

[4] R. Huang, W. Sun, Y. Xu, H. Chen, D. Towey, and X. Xia, "A survey on adaptive random testing," *IEEE Transactions on Software Engineering*, to appear, 2019. DOI: 10.1109/TSE.2019.2942921.

[5] T. Y. Chen, F. Kuo, D. Towey, and Z. Zhou, "A revisit of three studies related to random testing," *SCIENCE CHINA Information Sciences*, vol. 58, no. 5, pp. 1–9, 2015.

[6] T. Y. Chen, F.-C. Kuo, R. G. Merkel, and T. H. Tse, "Adaptive random testing: The ART of test case diversity," *Journal of Systems and Software*, vol. 83, no. 1, pp. 60–66, 2010.

[7] K. P. Chan, T. Y. Chen, and D. Towey, "Restricted random testing: Adaptive random testing by exclusion," *International Journal of Software Engineering and Knowledge Engineering*, vol. 16, no. 4, pp. 553–584, 2006.

[8] B. Zhou, H. Okamura, and T. Dohi, "Enhancing performance of random testing through Markov Chain Monte Carlo methods," *IEEE Transactions on Computers*, vol. 62, no. 1, pp. 186–192, 2013.

[9] M. M. Deza and E. Deza, *Encyclopedia of distances*. Springer, 2009.

[10] P. E. Ammann and J. C. Knight, "Data diversity: an approach to software fault tolerance," *IEEE Transactions on Computers*, vol. 37, no. 4, pp. 418–425, 1988.

[11] C. Mao, X. Zhan, T. H. Tse, and T. Y. Chen, "KDFC-ART: a KD-tree approach to enhancing Fixed-size-Candidate-set Adaptive Random Testing," *IEEE Transactions on Reliability*, vol. 68, no. 4, pp. 1444–1469, 2019.