ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.17, No.12, December 2006, pp.2438−2449
DOI: 10.1360/jos172438

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

失效区域紧致性对适应性随机测试的性能影响*

陈宗岳，郭斐菁，孙昌爱+

(Faculty of Information and Communication Technologies, Swinburne University of Technology, Melbourne, VIC 3122, Australia)

Impact of the Compactness of Failure Regions on the Performance of Adaptive Random Testing

CHEN Tsong-Yueh, KUO Fei-Ching, SUN Chang-Ai+

(Faculty of Information and Communication Technologies, Swinburne University of Technology, Melbourne, VIC 3122, Australia)

+ Corresponding author: Phn: +61-3-92148589, Fax: +61-3-98190823, E-mail: changai_sun2002@hotmail.com, http://www.swin.edu.au/ict/

Chen TY, Kuo FC, Sun CA. Impact of the compactness of failure regions on the performance of adaptive random testing. *Journal of Software*, 2006,17(12):2438−2449. http://www.jos.org.cn/1000-9825/17/2438.htm

Abstract: Adaptive random testing (ART) is an enhanced version of random testing (RT). It has been observed that the compactness of failure regions is one of the factors that affect the performance of ART. However, this relationship has only been verified with rectangular failure regions. This paper further investigates the relationship between the compactness of failure regions and the performance of ART by conducting simulation experiments, where various regular and irregular failure regions are studied. The experimental results have shown that ART's performance improves as the compactness of failure regions increases. This study has provided further insights into the conditions where ART outperforms RT.

Key words: software testing; random testing; adaptive random testing; failure pattern

摘 要: 适应性随机测试是一种增强的随机测试方法.已有的研究发现:失效区域的紧致程度是影响适应性随机测试性能的几个基本因素之一,并仅在失效区域为长方形的情形下验证了上述猜想.采用仿真实验的方法进一步研究失效区域的紧致程度与适应性随机测试的性能之间的精确关系.研究了几种基本规则形状的和不规则形状的失效区域.实验结果表明:适应性随机测试方法的性能随着失效区域的紧致程度的增强而提高.该研究进一步地揭示了适应性随机测试优于随机测试的基本条件.

关键词: 软件测试;随机测试;适应性随机测试;失效模式
中图法分类号: TP311 文献标识码: A

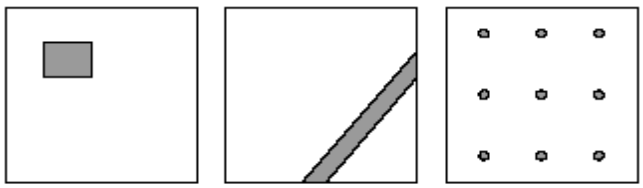
1 Introduction

Random Testing (RT)[1] is a widely used software testing method in practice. In this method, testers just randomly select test cases to test their programs, without any assumption on the program under test. RT is very

* Supported by an Australia Research Council Discovery under Grant No.DP0557246
Received 2006-01-19; Accepted 2006-04-03

simple to perform, and can be used to deliver reliability predictions. However, RT may be inefficient since it does not make use of any information about the likely characteristic of the program under test.

When the failure features of program under test are carefully examined, failure-causing inputs tend to cluster to form certain patterns[2−5]. Chan, *et al*.[6] observed some typical patterns of failure regions within the input domain, including the *block* pattern, the *strip* pattern, and the *point* pattern, as illustrated in Fig.1. The sample programs for these failure patterns are further provided in [6] and it is pointed out that the *block* and *strip* pattern are the most common ones.



(a) Block (b) Strip (c) Points

Fig.1 Failure patterns

Based on the above observation, Chen, *et al*.[7] have developed *Adaptive Random Testing* (ART) to enhance the effectiveness of RT. In ART, test cases are randomly selected and evenly spread. Intuitively speaking, a more widespread set of random test cases have a better chance to detect a failure when the failure-causing inputs are clustered into one or more failure regions[8]. It has been observed that ART can detect the first failure using about 60% test cases used by RT[7−12].

ART can be implemented through different approaches, such as by distance[7,8], by restriction[9], or by lattice[11], etc. Some of these approaches have been combined to improve ART's cost-effectiveness, that is, fewer test cases to detect the first failure and less computations[12]. The concept of mirroring[10] has recently been introduced to improve the efficiency of some ART implementations.

A series of simulation experiments have been conducted to identify the fundamental factors that affect the fault-detection capabilities of ART [13]. These experiments have observed that ART's performance depends on the failure rates, the number of failure regions, the existence of predominant failure regions, and the compactness of failure regions.

The previous work has preliminarily reported the impact of the compactness of failure regions on the performance of ART, where the failure region was set to be rectangular[13]. However, in real-life programs, the failure region may be of various shapes. In this study, we further investigate the impact of the compactness of a failure region on the performance of ART, where both regular and irregular shapes of failure regions are studied. Our experimental results show that there exists a strong relation between the compactness of failure regions and the performance of ART.

The rest of paper is organized as follows: Section 2 describes an implementation of ART; Section 3 discusses the underlying experiment design, including the performance measurement, compactness measurement and experimental settings; Section 4 reports the experiment result and provides further analysis; Section 5 concludes the paper.

2 Adaptive Random Testing (ART)

As mentioned above, there are various ways to implement ART. In our study, ART is implemented using FSCS (fixed size candidate set) ART algorithm[7]. In this method, two sets of test cases are maintained, namely the *executed set* and the *candidate set*. ART generates the candidate set using RT strategy, and selects a test case from

the candidate set, which is farthest away from all the elements in the executed set. With this approach, test cases in ART are both randomly selected and widely spread. The detailed FSCS ART algorithm is shown as follows:

1) The executed set *TS* is initially set as empty. RT strategy is employed to generate the first test case t_i which is then executed and appended to *TS*. If no failure is detected, then go to Step 2); otherwise the process ends;

2) RT strategy is employed to generate a candidate set $CS=\{c_1,c_2,\ldots,c_n\}$. A test case c_k ($1\le i\le n$) is selected from the candidate set *CS* such that $\underset{j=1..m}{Min}(dist(t_j,c_k))=\underset{i=1..n}{Max}(\underset{j=1..m}{Min}(dist(t_j,c_i)))$, where *m* is the size of *TS*, $t_j\in TS$, $c_i\in CS$ and $dist(t_j,c_i)$ denotes the distance between two test cases t_j and c_i[10];

3) The chosen test case c_k is tested and then added to the executed set *TS*. If a failure is detected, the process stops; otherwise, Step 2) is repeated until a failure is detected, or the number of test cases in the resulting *TS* exceeds a predefined value.

3 Design of Simulation Experiment

In this section, we discuss how to design and conduct the simulation experiments for different shapes of failure regions. We take the same performance measure and compactness measure as used in the previous simulation experiments[13].

3.1 Performance measurement

There are several common performance measures for test case selection strategies. One of them is *F-measure* which is defined as the number of test cases used to detect the first failure[8]. In this study, we use the *ART F-Ratio* which is defined as F_{art}/F_{rt} to measure the performance of ART against RT, where F_{art} and F_{rt} denote the *F-measure* of ART and RT, respectively. In our experiments, F_{art} is achieved by simulation while F_{rt} is calculated by the predefined failure rate *r* ($F_{rt}=1/r$). Obviously, the smaller the *ART F-Ratio* is, the better the performance of ART is.

3.2 Compactness measurement

There exist many compactness measurements[14]. The most intuitive one which has been used in Ref.[13] is defined as the ratio of the area of a given shape to a circular shape, assuming both shapes have the same perimeters.

In this study, we use this compactness measurement. For the two-dimensional shapes, the measurement is $\frac{4\pi A}{P^2}$,where *A* and *P* represent the area and perimeter of a geometric shape, respectively.

3.3 Settings

Prior to seeding a failure region in the input domain, we need to determine its compactness, size and location. It can be done as follows:

1) In this study, we focus on some irregular shapes and some two-dimensional regular primitive shapes, including square, rectangle, circle, ellipse and isosceles triangle.

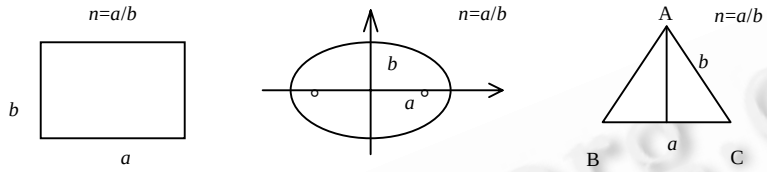
For the regular shapes of failure regions illustrated in Fig.2, their compactness is uniquely decided by the corresponding parameter *n*. For the rectangular case, *n* is defined as the ratio of the longer edge to the shorter edges ($n\ge1$); for ellipses, *n* is defined as the ratio of the semi-major to the semi-minor ($n\ge1$); for isosceles triangle, *n* is defined as the ratio of the bottom edge to the side edge ($0<n<2$ since $b+b>a$). For rectangles and ellipses, we set the parameter *n* with 1, 2, 4, 5, 7, 10, 20, 30, 40 and 50; for isosceles triangles, we set the parameter $1/n$ with 5/8, 3/4, 7/8, 1, 2, 4, 5, 7, 10, 20, 30, 40 and 50. These parameter values hereby provide various degrees of compactness.

For the irregular shapes of failure regions, we enumerate a variety of shapes, which are constructed with 3 to 5 connected square units.

2) In our simulation experiments, the input domain is assumed to be of size d, and the failure rate r is the ratio of the size of failure region f to d (that is $r=f/d$). Our simulation experiments investigate the failure rate r of 0.0005, 0.001 and 0.005 (the expected F_{rt} is 2000, 1000, and 200, respectively).

3) Finally, we randomly choose the location of the failure regions in the input domain.

To achieve a reliable simulation result, say with the confidence level of 95% and the accuracy of ±5, we need to collect a large sample of data, the size of which can be decided by means of the *central limit theorem*[7].



(a) Rectangle (b) Ellipse (c) Isosceles triangle

Fig.2 Three regular shapes

4 Result and Analysis

In this section, we report the impact of the compactness of failure regions on the performance of ART under the circumstances that a failure region is assumed to be three kinds of regular shapes and some irregular shapes, and provide further analysis across different shapes of failure regions.

4.1 Rectangle/Square

This experiment is first reported in Ref.[13]. We briefly describe the result for the sake of completeness. The compactness of a failure region can be expressed in terms of the ratio n of length to width. Recall the definition of compactness measure in Section 3. For the rectangle illustrated in Fig.2(a),

$$compactness = \frac{4\pi A}{P^2} = \frac{4\pi ab}{(2a+2b)^2} = \frac{\pi}{n+2+\frac{1}{n}},\ n \geq 1.$$

Obviously, *compactness* is a decreasing function in the domain of $n \geq 1$, and has the maximum value when $n=1$. The result of our experiments is reported in Fig.3 (refer to Table 3 for the raw data) which shows that the larger n, the larger the value of the *ART F-Ratio*. ART performs best when n is 1, that is, the rectangle is a square. Based on the above compactness analysis and simulation result, we can conclude that when the failure region is a rectangle/square, the larger the compactness is, the better the performance of ART is.

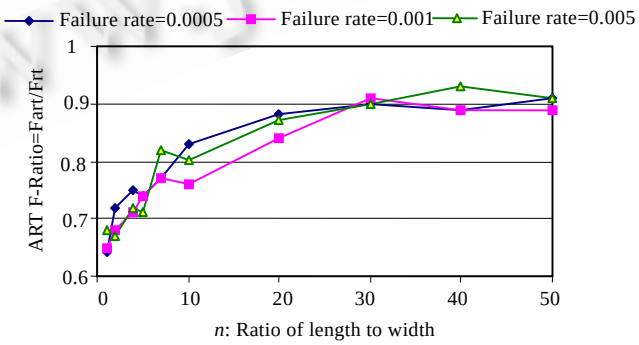


Fig.3 The impact of n (the ratio of length to width for rectangles) on the performance of ART

4.2 Ellipse/Circle

In this experiment, the compactness of a failure region is expressed in terms of the ratio n of semi-major to semi-minor. The perimeter of an ellipse, as illustrated in Fig.2(b), can be calculated using the infinite serial formula. In practice, the perimeter p of an ellipse can be approximately calculated using the following formula [15]:

$$P = \pi(3/2(a+b) - \sqrt{ab}).$$

Then

$$compactness = \frac{4\pi A}{P^2} = \frac{4\pi^2 ab}{\pi^2(3/2(a+b)-\sqrt{ab})^2}.$$

Since $n = a/b \geq 1$, we have

$$compactness = \left(\frac{2}{3/2(\frac{1}{\sqrt{n}}+\sqrt{n})-1}\right)^2.$$

Let $x = \sqrt{n}$, we have

$$compactness = \left(\frac{2}{3/2(x+\frac{1}{x})-1}\right)^2.$$

Let $f(x) = 3/2\left(x+\frac{1}{x}\right)-1$, $\frac{df(x)}{dx} = 3/2\left(1-\frac{1}{x^2}\right)$. Obviously, x=1 is a minimum point. When $x \geq 1$ (i.e. $\sqrt{n} = \sqrt{a/b} \geq 1$), $\frac{df(x)}{dx} = 3/2\left(1-\frac{1}{x^2}\right) \geq 0$. Therefore $f(x)$ is an increasing function, and hence *compactness* is a decreasing function. Thus, the circle (that is n=1) has the maximum compactness.

The results of our simulation is reported in Fig.4 (refer to Table 4 for the raw data) which shows that the larger n, the larger the value of the *ART F-Ratio*. Fig.4 also shows that ART performs best when n is 1, or equivalently the failure region is a circle.

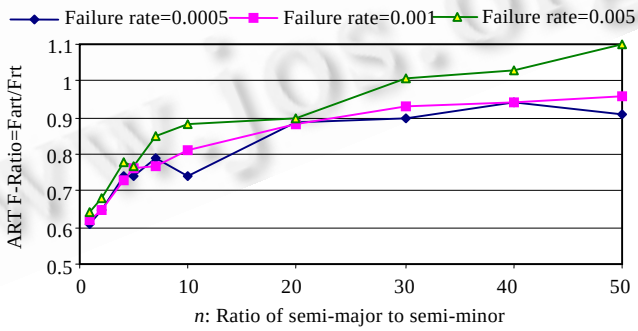


Fig.4 The impact of n (the ratio of semi-major to semi-minor for ellipses) on the performance of ART

Based on the above compactness analysis and simulation result, we can further conclude that when the failure region is a circle/ellipse, the larger the compactness is, the better the performance of ART is.

4.3 Isosceles triangle/equilateral triangle

In this experiment, the compactness of an isosceles triangle is expressed in terms of the ratio n of the bottom

edge to the side edge. The compactness of an isosceles triangle, as illustrated in Fig.2(c), can be defined as follows.

$$compactness = \frac{4\pi A}{p^2} = \frac{2a\pi\sqrt{b^2 - \frac{1}{4}a^2}}{(a+2b)^2} = \frac{n\pi\sqrt{4-n^2}}{(n+2)^2}.$$

Let $x=n+2$, then

$$compactness = \pi\sqrt{-1+\frac{8}{x}-\frac{20}{x^2}+\frac{16}{x^3}}.$$

Let $f(x) = -1+\frac{8}{x}-\frac{20}{x^2}+\frac{16}{x^3}$, $\frac{df(x)}{dx} = \left(-\frac{8}{x^2}+\frac{40}{x^3}-\frac{48}{x^4}\right)$. When $x=3$ or $x=2$, $\frac{df(x)}{dx}=0$. Note that $2<x<4$ (as $0<n<2$), thus the unique solution of $\frac{df(x)}{dx}=0$ is x=3. When $2<x<3$ (that is $0<n<1$) and $\frac{df(x)}{dx}>0$, $f(x)$ is an increasing function and hence *compactness* is an increasing function. When $3<x<4$ (that is $1<n<2$) and $\frac{df(x)}{dx}<0$, $f(x)$ is a decreasing function and hence *compactness* is a decreasing function. Thus, *compactness* has the maximum value when $x=3$(that is n=1, the shape is an equilateral triangle). In other words, an equilateral triangle has the maximum compactness among all isosceles triangle shapes.

The result of our experiments is reported in Fig.5 (refer to Table 5 for the raw data), and shows that:

1) When *n* is equal to or larger than 1, the larger *n* is, the larger the value of the *ART F-Ratio* is;

2) When *n* is smaller than 1, the smaller *n* is, the larger the value of the *ART F-Ratio* is.

For all failure rates considered, ART performs best when the ratio *n* is 1. It is noted that when *n* is 1, an isosceles triangle is an equilateral triangle.

Therefore, we can further conclude that when the failure region is an isosceles triangle, the larger the compactness of the failure region is, the better the performance of ART is.

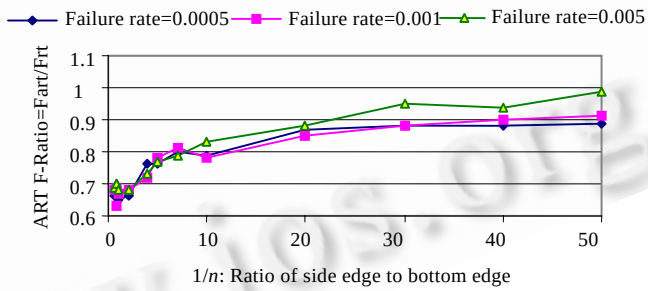


Fig.5 The impact of *n* (the ratio of bottom edge to side edge for isosceles triangles) on the performance

4.4 Irregular shapes

In this experiment, we have only considered 28 irregular shapes of failure regions, as illustrated in Fig.6. Each irregular shape consists of 3 to 5 connected same-size square units.

The first column in Table 1 indicates the index of each irregular shape, as illustrated in Fig.6. The *compactness* of each failure region shape can be calculated using the compactness measurement $\frac{4\pi A}{P^2}$, described in Section 3.2. For the majority of data reported in Table 1 (under the confidence level of 95% and accuracy of ±5 (irregular shapes)), at the same failure rate, the performance of ART improves with the increasing of compactness of failure regions. For example, for the shapes “5D” and “5I”, the *ART F-Ratio* is 0.65 and 0.81 when the failure rate is 0.0005; while the compactness is 0.628 and 0.157, respectively. For failure regions which have the same

compactness but different shapes, their ART performances are very similar.

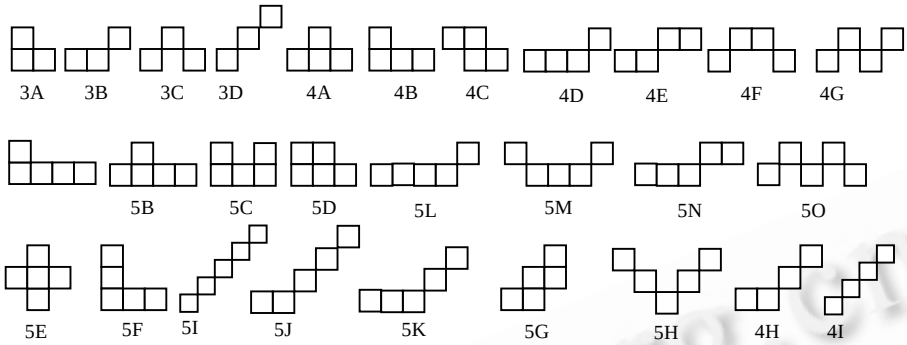


Fig.6 Irregular shapes constructed from 3 to 5 connected square units

Table 1 The impact of compactness on the *ART F-Ratio*

Index	Compactness	Failure rate		
		0.0005	0.001	0.005
5D	0.628	0.65	0.65	0.66
3A	0.589	0.68	0.70	0.71
4A	0.503	0.70	0.70	0.71
4B	0.503	0.67	0.72	0.72
4C	0.503	0.72	0.67	0.73
5A	0.436	0.74	0.73	0.77
5B	0.436	0.68	0.73	0.74
5C	0.436	0.79	0.68	0.71
5E	0.436	0.71	0.71	0.71
5F	0.436	0.70	0.73	0.75
5G	0.436	0.69	0.73	0.75
3B	0.377	0.70	0.67	0.72
4D	0.349	0.74	0.73	0.82
4E	0.349	0.77	0.75	0.78
5L	0.321	0.78	0.75	0.81
5N	0.321	0.77	0.77	0.83
3C	0.262	0.80	0.79	0.78
3D	0.262	0.80	0.81	0.77
4F	0.256	0.78	0.79	0.83
4H	0.256	0.75	0.76	0.80
5K	0.245	0.79	0.81	0.80
5M	0.245	0.79	0.78	0.80
4G	0.196	0.83	0.80	0.85
4I	0.196	0.82	0.78	0.82
5J	0.194	0.77	0.83	0.83
5H	0.157	0.79	0.84	0.84
5I	0.157	0.81	0.83	0.86
5O	0.157	0.83	0.81	0.85

4.5 Statistical analysis of the relationship between the compactness of failure region and performance of ART

For a specific shape of failure regions, the above experimental results have demonstrated that the improvement of performance of ART over RT improves with the increasing of the compactness of the failure region. It is interesting to investigate whether this trend still exists when various shapes of failure regions are taken into account. We attempt to answer this question by analyzing all experimental data reported above irrespective of the variety of failure region shapes. The resulting data are shown in Table 6.

As illustrated in Fig.7 (refer to Table 6 for the raw data), the performance of ART basically improves as the compactness of failure regions increases with a few exceptions. For example, when the failure rate is set as 0.005: (a) For a rectangle with its compactness being 0.060 (n=50), the *ART F-Ratio* is 0.91 (see Table 3); (b) For an ellipse with its compactness being 0.071 (n=30), the *ART F-Ratio* is 1.01 (see Table 4). The compactness of case (b)

is larger than case (a), but the performance of ART in case (b) is worse than that in case (a).

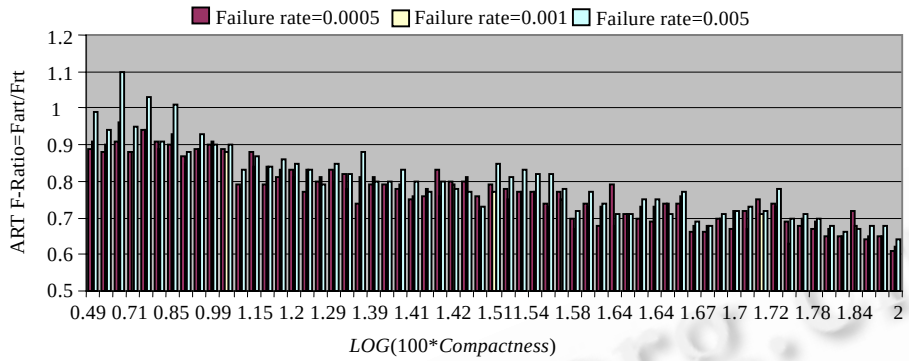


Fig.7 The impact of the compactness of failure regions on the performance of ART under different failure rates

As illustrated in Fig.7, under different failure rates, the trend of the performance of ART over compactness seems to be a decreasing function. To further evaluate the relationship between the compactness of failure regions and the performance of ART, we use the technique of *trend analysis*[16] to process the experimental data. Given a set of data, the best way to discover the trend is to find the line of best fit for the data, and correlation coefficient is the most common way of determining how well that line actually correlates with the data. We use the linear function *ART F-Ratio*=α**LOG*(100**Compactness*)+β to fit this trend, and use the technique of *least squares fitting* to determine the parameters *a* and *b* and the correlation coefficient *cor* according to the following formulas[16]

$$\alpha=\frac{\sum_{i=1}^{N}(X_i-M_x)(Y_i-M_y)}{\sum_{i=1}^{N}(X_i-M_x)^2},\quad \beta=M_y-aM_x,\quad Cor=\frac{\sum_{i=1}^{N}(X_i-M_x)(Y_i-M_y)}{\sqrt{\sum_{i=1}^{N}(X_i-M_x)^2}\sqrt{\sum_{i=1}^{N}(Y_i-M_y)^2}},$$

where M_x, M_y and N is the mean of X coordinates, the mean of Y coordinates and the number of data points, respectively.

Based on the data in Table 6 and the above formulas, parameters α and β, correlation coefficient *cor*, confidence level, M_x and M_y are computed and reported in Table 2 (between the compactness of failure regions and the performance of ART). Similarly, the correlation coefficients for various shapes of failure regions are reported in Tables 7~Tables 10 using the data from Tables 3, Tables 4, Tables 5 and Tables 1, respectively. In Tables 2 and Tables 7~Tables 10, X corresponds to *LOG*(100**compactness*), Y corresponds to the *ART F-Ratio*. Generally speaking, a value of 1 or −1 means that the data correspond perfectly with the line. Note that when |*cor*| is greater than 0.81 or equivalently confidence is greater than 66%, a strong correlation is said to exist. It is noted that all correlation coefficients in Tables 2 and Tables 7~Tables 10 indicate a strong correlation.

Table 2 The trend analysis of relationship

	Failure rate		
	0.0005	0.001	0.005
α	−0.206	−0.218	−0.254
β	1.060	1.076	1.157
Cor	−0.914	−0.942	−0.931
Confidence (%)	83.6	88.8	86.8
M_x	1.409	1.409	1.409
M_y	0.770	0.769	0.799
X=*LOG*(100**Compactness*); *Y*=*ART F-Ratio*; *N*=61			

Since α, representing the slope of the correlated line and indicating the trend, is negative for all failure rates,

and *LOG*(100**compactness*) is an increasing function over *compactness*, we can draw the conclusion that the *ART F-Ratio* will decrease with the increasing of compactness of failure regions. In other words, the performance of ART will improve as the compactness of failure regions increases.

5 Conclusion

We have reported a simulation study of the impact of the compactness of failure regions on the performance of ART. The experimental result shows that, the larger the compactness of failure regions, the better the performance of ART, for the cases of square, rectangle, circle, ellipse, isosceles triangle, equilateral triangle and some irregular failure regions. Using the regression analysis technique, we further analyzed the relationship between the compactness of failure regions and the performance of ART, regardless of the shapes of failure regions. The result shows that there exists a strong relation between the compactness of failure regions and the performance of ART.

In this simulation study, we have to assume that the failure region is of certain shapes. The validity of our simulation results is obviously restricted to those shapes studied. It should be noted that in real-life applications, failure regions may be other than those regular or irregular shapes discussed in this paper.

This study provides further insights into the conditions where ART outperforms RT. Future work may include an experimental analysis of the geometry of failure regions for real-life programs and an experimental analysis of the impact of the compactness of failure regions on the performance of ART.

References:

[1] Hamlet R. Random testing. In: Marciniak J, ed. Encyclopedia of Software Engineering. 2 ed, John Wiley & Sons, 2002.

[2] Ammann PE, Knight JC. Data diversity: An approach to software fault tolerance. IEEE Trans. on Computers, 1988,37(4):418−425.

[3] Bishop PG. The variation of software survival times for different operational input profiles. In: David P, ed. Proc. of the 23rd Int'l Symp. on Fault-Tolerant Computing (FTCS-23). California: IEEE Computer Society Press, 1993. 98−107.

[4] Finelli GB. NASA software failure characterization experiments. Reliability Engineering and System Safety, 1991,32(1−2): 155−169.

[5] White LJ, Cohen EI. A domain strategy for computer program testing. IEEE Trans. on Software Engineering, 1980,6(3):247−257.

[6] Chan FT, Chen TY, Mak IK, Yu YT. Proportional sampling strategy: Guidelines for software testing practitioners. Information and Software Technology, 1996,38(12):775−782.

[7] Chen TY, Leung H, Mak IK. Adaptive random testing. In: Maher MJ, ed. Proc. of the 9th Asian Computing Science Conf. LNCS 3321, Heidelberg: Springer-Verlag, 2004. 320−329.

[8] Chen TY, Tse TH, Yu YT. Proportional sampling strategy: A compendium and some insights. Journal of Systems and Software, 2001,58(1):65−81.

[9] Chan KP, Chen TY, Tower D. Restricted random testing. In: Kontio J, Conradi R, eds. Proc. of the 7th European Conf. on Software Quality (ECSQ 2002). LNCS 2349, Heidelburg: Springer-Verlag, 2003. 321−330.

[10] Chen TY, Kuo FC, Merkel RG, Ng SP. Mirror adaptive random testing. Information and Software Technology, 2004, 46(15):1001−1010.

[11] Mayer J. Lattice-Based adaptive random testing. In: Ireland A, ed. Proc. of the 20th IEEE/ACM Int'l Conf. on Automated Software Engineering (ASE 2005). New York: ACM Press, 2005. 333−336.

[12] Mayer J. Adaptive random testing by bisection with restriction. In: Lau KK, Banach R, eds. Proc. of the 7th Int'l Conf. on Formal Engineering Methods (ICFEM 2005). LNCS 3785, Heidelburg: Springer-Verlag, 2005. 251−263.

[13] Chen TY, Kuo FC, Zhou ZQ. On the relationships between the distribution of failure-causing inputs and effectiveness of adaptive random testing. In: Smith G, ed. Proc. of the 17th Int'l Conf. on Software Engineering and Knowledge Engineering (SEKE 2005). Skokie: UDN Digital Co. Ltd., 2005. 306−311.

[14] Maceachren AM. Compactness of geographic shape: Comparison and evaluation of measures. Geografisk Annaler, 1985,67B(1): 53−67.

[15] Bronshtein IN, Semendyayev KA. Handbook of Mathematics. 3rd English ed, New York: Van Nostrand Reinhold Co., 1985.

[16] Draper NR, Smith H. Applied Regression Analysis. 2nd ed, New York: John Wiley and Sons Inc., 1981.

Appendix I: Experimental Data

Table 3 The impact of compactness and failure rate on the *ART F-Ratio* with the confidence level of 95% and accuracy of ±5 where *n* is the ratio of length to width of a rectangle/square

n	Compactness	Failure rate		
		0.0005	0.001	0.005
1	0.785	0.64	0.65	0.68
2	0.698	0.72	0.68	0.67
4	0.503	0.75	0.71	0.72
5	0.436	0.74	0.74	0.71
7	0.344	0.77	0.77	0.82
10	0.260	0.83	0.76	0.80
20	0.142	0.88	0.84	0.87
30	0.098	0.90	0.91	0.90
40	0.075	0.89	0.89	0.93
50	0.060	0.91	0.89	0.91

Table 4 The impact of compactness and failure rate on the *ART F-Ratio* with the confidence level of 95% and accuracy of ±5 where *n* is the ratio of semi-major to semi-minor of an ellipse/circle

n	Compactness	Failure rate		
		0.0005	0.001	0.005
1	1.000	0.61	0.62	0.64
2	0.840	0.65	0.65	0.68
4	0.529	0.74	0.73	0.78
5	0.437	0.74	0.76	0.77
7	0.320	0.79	0.77	0.85
10	0.225	0.74	0.81	0.88
20	0.110	0.89	0.88	0.90
30	0.071	0.90	0.93	1.01
40	0.053	0.94	0.94	1.03
50	0.041	0.91	0.96	1.10

Table 5 The impact of compactness and failure rate on the *ART F-Ratio* with the confidence level of 95% and accuracy of ±5 where *n* is the ratio of bottom edge to side edge of an isosceles triangle/equilateral triangle

1/*n*	Compactness	Failure rate		
		0.0005	0.001	0.005
0.625	0.465	0.66	0.68	0.69
0.75	0.562	0.69	0.63	0.70
0.875	0.597	0.67	0.69	0.70
1	0.605	0.65	0.67	0.68
2	0.487	0.66	0.68	0.68
4	0.308	0.76	0.72	0.73
5	0.258	0.76	0.78	0.77
7	0.195	0.80	0.81	0.79
10	0.142	0.79	0.78	0.83
20	0.075	0.87	0.85	0.88
30	0.051	0.88	0.88	0.95
40	0.038	0.88	0.90	0.94
50	0.031	0.89	0.91	0.99

Table 6 The unified impact of the compactness on the performance of ART across the various failure region shapes

LOG(100**Compactness*)	Failure rate		
	0.0005	0.001	0.005
0.488	0.89	0.91	0.99
0.583	0.88	0.90	0.94
0.618	0.91	0.96	1.10
0.705	0.88	0.88	0.95
0.721	0.94	0.94	1.03
0.781	0.91	0.89	0.91
0.853	0.90	0.93	1.01
0.874	0.87	0.85	0.88
0.874	0.89	0.89	0.93
0.992	0.90	0.91	0.90
1.039	0.89	0.88	0.90
1.153	0.79	0.78	0.83
1.154	0.88	0.84	0.87
1.196	0.79	0.84	0.84
1.196	0.81	0.83	0.86
1.196	0.83	0.81	0.85
1.288	0.77	0.83	0.83
1.290	0.80	0.81	0.79
1.292	0.83	0.80	0.85
1.292	0.82	0.78	0.82
1.352	0.74	0.81	0.88
1.389	0.79	0.81	0.80
1.389	0.79	0.78	0.80
1.408	0.78	0.79	0.83
1.408	0.75	0.76	0.80
1.412	0.76	0.78	0.77
1.414	0.83	0.76	0.80
1.418	0.80	0.79	0.78
1.418	0.80	0.81	0.77
1.488	0.76	0.72	0.73
1.505	0.79	0.77	0.85
1.507	0.78	0.75	0.81
1.507	0.77	0.77	0.83
1.536	0.77	0.77	0.82
1.543	0.74	0.73	0.82
1.543	0.77	0.75	0.78
1.576	0.70	0.67	0.72
1.640	0.74	0.73	0.77
1.640	0.68	0.73	0.74
1.640	0.79	0.68	0.71
1.640	0.71	0.71	0.71
1.640	0.70	0.73	0.75
1.640	0.69	0.73	0.75
1.640	0.74	0.74	0.71
1.641	0.74	0.76	0.77
1.668	0.66	0.68	0.69
1.687	0.66	0.68	0.68
1.701	0.70	0.70	0.71
1.701	0.67	0.72	0.72
1.701	0.72	0.67	0.73
1.701	0.75	0.71	0.72
1.723	0.74	0.73	0.78
1.750	0.69	0.63	0.70
1.770	0.68	0.70	0.71
1.776	0.67	0.69	0.70
1.781	0.65	0.67	0.68
1.798	0.65	0.65	0.66
1.844	0.72	0.68	0.67
1.895	0.64	0.65	0.68
1.924	0.65	0.65	0.68
2.000	0.61	0.62	0.64

Table 7 The trend analysis of relationship between the compactness of rectangular failure regions and the performance of ART (based on Table 3)

	Failure rate		
	0.0005	0.001	0.005
α	−0.216	−0.223	−0.237
β	1.102	1.092	1.129
Cor	−0.957	−0.976	−0.966
Confidence (%)	91.6	95.3	93.3
M_x	1.383	1.383	1.383
M_y	0.803	0.784	0.801

X=*LOG*(100**Compactness*); *Y*=*ART F-Ratio*; *N*=10

Table 8 The trend analysis of relationship between the compactness of ellipse failure regions and the performance of ART (based on Table 4)

	Failure rate		
	0.0005	0.001	0.005
α	−0.219	−0.237	−0.296
β	1.084	1.122	1.260
Cor	−0.957	−0.990	−0.983
Confidence (%)	92.1	98.0	96.7
M_x	1.338	1.338	1.338
M_y	0.791	0.805	0.864

X=*LOG*(100**Compactness*); *Y*=*ART F-Ratio*; *N*=10

Table 9 The trend analysis of relationship between the compactness of isosceles triangle failure regions and the performance of ART (based on Table 5)

	Failure rate		
	0.0005	0.001	0.005
α	−0.189	−0.197	−0.234
β	1.010	1.020	1.095
Cor	−0.966	−0.971	−0.991
Confidence (%)	93.3	94.2	98.3
M_x	1.281	1.281	1.281
M_y	0.766	0.768	0.795

X=*LOG*(100**Compactness*); *Y*=*ART F-Ratio*; *N*=13

Table 10 The trend analysis of relationship between the compactness of irregular failure regions and the performance of ART (based on Table 1)

	Failure rate		
	0.0005	0.001	0.005
α	−0.253	−0.269	−0.269
β	1.132	1.156	1.181
Cor	−0.737	−0.825	−0.796
Confidence (%)	85.8	90.8	89.2
M_x	1.503	1.503	1.503
M_y	0.752	0.752	0.777

X=*LOG*(100**Compactness*); *Y*=*ART F-Ratio*; *N*=28

CHEN Tsong-Yueh was born in 1948. He is a professor of software engineering at the Faculty of Information and Communication Technologies in Swinburne University of Technology. He received his Ph.D. degree in Computer Science from the University of Melbourne in 1986; a master degree in Computer Science from Imperial College of Science and Technology in 1976; a bachelor in Mathematics and Physics from The University of Hong Kong in 1971. His current research areas include software testing, debugging, software maintenance and software design.

KUO Fei-Ching was born in 1973. She is a PhD student at the Faculty of Information and Communication Technologies in Swinburne University of Technology. She received her BSc(Honours) in Computer Science at the same university in 2002. Her current research areas include software testing, debugging, maintenance and project management.

SUN Chang-Ai was born in 1974. He is a postdoctoral research fellow at the Faculty of Information and Communication Technologies in Swinburne University of Technology. He received his Ph.D. degree in Computer Software and Theory in 2002 from Beijing University of Aeronautics and Astronautics; a bachelor degree in Computer and Its application in 1997 from University of Science and Technology Beijing. His current research areas are software testing, software maintenance and software architecture.