Towards Effective Adaptive Random Testing for Higher-Dimensional Input Domains

Johannes Mayer
Ulm University
Dept. of Applied Information Processing
89069 Ulm, Germany
johannes.mayer@uni-ulm.de

ABSTRACT

Adaptive Random Testing subsumes a class of algorithms that detect the first failure with less test cases than Random Testing. The present paper shows that a "reference method" in the field of Adaptive Random Testing is not effective for higher dimensional input domains and clustered failure-causing inputs. The reason for this behavior is explained, and a modified method is proposed and analyzed.

Categories and Subject Descriptors

D.2.5 [**Software Engineering**]: Testing and Debugging—*Testing tools*; G.3 [**Probability and Statistics**]: Reliability and life testing

General Terms

Algorithms, Reliability, Verification

Keywords

Adaptive Random Testing, Random Testing, Test case selection

1. INTRODUCTION

Random Testing (RT) [5], i. e. the random generation of test inputs, is a promising approach to the automation of test case generation since it is unbiased and easy to implement. Chan et al. [1] observed that failure-causing inputs form clusters within the input domain. They coarsely classified these clusters, called failure patterns, into block, strip, and point type. Based thereon, Chen et al. [4] deduced that wide-spread test cases have a better chance to detect a failure with less test cases than Random Testing and introduced Adaptive Random Testing (ART) based on this notion.

Chen et al. [4] introduced the *F-measure* which denotes the (random) number of test cases necessary to detect the first failure. The F-measure has been used to compare all ART algorithms so far. The first ART method was Distance-Based ART (D-ART) [3, 4]. This method is considered the best one (besides the similar RRT) in the ART community due to its low F-measure.

Chen et al. [2] presented simulation results for D-ART with block failure pattern and one-, two- resp. three-dimensional input domains. These results show an interesting behavior: The F-measure of D-ART gets worse the higher the dimension.

The present paper first investigates the finding by Chen et al. [2] and systematically analyzes D-ART for dimensions $d = 2, 3, \dots, 6$ with a simulation study. The not so good F-measure of D-ART in higher dimensions is explained and an improved algorithm is proposed.

After the introduction of preliminaries in Section 2, the D-ART method is explained and discussed in Section 3. An improved algorithm is presented in Section 4 accompanied by simulation results. Section 5 concludes this paper.

GECCO'06, July 8–12, 2006, Seattle, Washington, USA.
ACM 1-59593-186-4/06/0007.

2. PRELIMINARIES

An input is said to be *failure-causing* if the program execution with this input leads to a failure. The percentage of failure-causing inputs within the input domain is said the *failure rate* θ. The theoretical mean F-measure of Random Testing is simply $1/\theta$. The *relative mean F-measure* of a method is its mean F-measure related to the theoretical mean F-measure of RT.

3. DISTANCE-BASED ART

Distance-Based ART (D-ART) [3, 4] with fixed sized candidate set randomly selects the first test case. Thereafter, a set of test case candidates—each one randomly chosen—of size k is computed in each iteration. The candidate with the greatest minimal distance to all previously executed test cases is chosen as the next test case. In the following iteration, a new candidate set is chosen and the procedure is repeated until a failure is detected (or the resources for testing are exhausted). The size $k = 10$ of the candidate set has been recommended.

The relative mean F-measure of D-ART has been determined through simulation for input domains of various dimensions (d) and several failure rates (θ) each with 5000 randomly generated block failure patterns (hyper cubes with random location within the input domain). Table 1 shows the results. Each relative mean F-measure value is accompanied by its accuracy on confidence level 99% determined by the Central Limit Theorem. The simulation confirms the results by Chen et al. [2] and shows that D-ART is less effective the higher dimension and the higher the failure rate. For higher dimensions and higher failure rates, the relative mean F-measure of D-ART is even greater than 1, which means that D-ART is less effective than RT in these cases.

In order to analyze the reason for the low effectiveness of D-ART, Table 2 shows the width of a cubic block failure

Table 1: Relative mean F-measure of D-ART for the block failure pattern

	Failure Rate θ				
	0.01	0.005	0.002	0.001	0.0005
$d=2$	0.680 ($\pm$0.018)	0.645 ($\pm$0.017)	0.639 ($\pm$0.018)	0.645 ($\pm$0.018)	0.638 ($\pm$0.018)
$d=3$	0.840 ($\pm$0.024)	0.805 ($\pm$0.023)	0.768 ($\pm$0.022)	0.751 ($\pm$0.022)	0.752 ($\pm$0.023)
$d=4$	1.070 ($\pm$0.032)	1.000 ($\pm$0.030)	0.928 ($\pm$0.028)	0.893 ($\pm$0.028)	0.890 ($\pm$0.028)
$d=5$	1.315 ($\pm$0.039)	1.238 ($\pm$0.038)	1.150 ($\pm$0.036)	1.104 ($\pm$0.034)	1.041 ($\pm$0.033)
$d=6$	1.617 ($\pm$0.049)	1.537 ($\pm$0.047)	1.409 ($\pm$0.044)	1.340 ($\pm$0.042)	1.251 ($\pm$0.040)

Table 2: Relative width of a d-dimensional hyper cube for failure rate $\theta = 0.01$

d	2	3	4	5	6	7	8
w	0.100	0.215	0.316	0.398	0.464	0.518	0.562

pattern related to the width of the (d-dimensional) cubic input domain for failure rate $\theta = 0.01$. The relative width $w = \sqrt[d]{\theta}$ increases with the dimension d. For dimensions $d \geq 7$ it even contains the mid point of the input domain with probability one.

D-ART is designed to widely spread the test cases over the whole input domain. Therefore, this method searches in a by far too large domain. E. g. for $d \geq 7$, it is obviously sufficient to have exactly one test case located at the center of the input domain.

4. ART WITH INCREASING DOMAIN

For the reasons mentioned, we should adapt D-ART to select the test cases from a sub-domain centered around the middle of the input domain. This sub-domain has to be enlarged as the algorithm proceeds to approach the whole input domain. More formally, we assume w. l. g. that the input domain is the hyper cuboid $I := \{(0, \ldots, 0)(w_1, \ldots, w_d)\}$.[1] Then, D-ART with Increasing Domain (ID-D-ART) selects the k th test case (and all candidates used for this selection) from the hyper cuboid

$$I_k := \{(r_d(k)w_1, \ldots, r_d(k)w_d) \\ ((1 - r_d(k))w_1, \ldots, (1 - r_d(k))w_d)\},$$

instead of the whole input domain (as D-ART does), where the restriction function $r_d(k)$ is defined as

$$r_d(k) = f_0 / \sqrt[d]{k},$$

with $k \geq 1$ and $f_0 \in [0, \frac{1}{2})$. The series $I_1, I_2, \ldots$ is monotonically increasing and approaching I, with $I_1 \subset I_2 \subset \ldots \subset I$.

To evaluate the effectiveness of ID-D-ART, a simulation study analogous to the previous one (cf. Table 1) has been conducted with $f_0 = 0.49$ (as determined through a separate simulation study) and $k = 10$. Table 3 shows the results of this study. The F-measure of ID-D-ART is much better than that of D-ART and always below 1 which means that ID-D-ART is always more effective than RT. Furthermore, the effectiveness of ID-D-ART improves for higher-dimensional input domains as opposed to D-ART.

[1]We denote a hyper cuboid $\{p_{min} p_{max}\}$ by its points with minimal (p_{min}) and maximal (p_{max}) coordinates.

Table 3: Relative mean F-measure of ID-D-ART for the block failure pattern

	Failure Rate θ				
	0.01	0.005	0.002	0.001	0.0005
$d=2$	0.523 ($\pm$0.015)	0.563 ($\pm$0.016)	0.579 ($\pm$0.017)	0.599 ($\pm$0.017)	0.591 ($\pm$0.017)
$d=3$	0.405 ($\pm$0.014)	0.453 ($\pm$0.015)	0.507 ($\pm$0.017)	0.537 ($\pm$0.017)	0.572 ($\pm$0.018)
$d=4$	0.269 ($\pm$0.010)	0.315 ($\pm$0.012)	0.366 ($\pm$0.014)	0.424 ($\pm$0.015)	0.467 ($\pm$0.017)
$d=5$	0.149 ($\pm$0.007)	0.190 ($\pm$0.008)	0.252 ($\pm$0.010)	0.291 ($\pm$0.012)	0.337 ($\pm$0.013)
$d=6$	0.058 ($\pm$0.004)	0.098 ($\pm$0.005)	0.144 ($\pm$0.007)	0.183 ($\pm$0.009)	0.217 ($\pm$0.010)

5. CONCLUSION

In the present paper it has been shown that the "reference method" D-ART, is not so effective for clustered failure-causing inputs (more precisely, the block failure pattern) and higher-dimensional input domains. Therefore, its practical use is limited. However, an improved version of the algorithm has been proposed that selects the test cases from a sub-domain of the input domain. This sub-domain increases and approaches the whole input domain as the algorithm proceeds. The proposed algorithms is more effective than D-ART and also RT for the block failure pattern—even for higher-dimensional input domains.

Due to the similarity of D-ART and RRT, the proposed approach can also be applied to RRT. First simulations have shown that the results for RRT are similar.

Furthermore, a detailed investigation on the restriction function $r_d(\cdot)$ as well as on the factor f_0 should be made.

6. REFERENCES

[1] F. T. Chan, T. Y. Chen, I. K. Mak, and Y. T. Yu. Proportional sampling strategy: Guidelines for software testing practitioners. *Information and Software Technology*, 38:775–782, 1996.

[2] T. Y. Chen, F.-C. Kuo, and Z. Q. Zhou. On the relationships between the distribution of failure-causing inputs and effectiveness of adaptive random testing. In *Proceedings of the 17th International Conference on Software Engineering and Knowledge Engineering (SEKE 2005)*, Taipei, Taiwan, China, 2005.

[3] T. Y. Chen, H. Leung, and I. K. Mak. Adaptive random testing. In M. J. Maher, editor, *Proceedings of the 9th Asian Computing Science Conference (ASIAN 2004)*, volume 3321 of *Lecture Notes in Computer Science*, pages 320–329. Springer, 2004.

[4] T. Y. Chen, T. H. Tse, and Y. T. Yu. Proportional sampling strategy: A compendium and some insights. *The Journal of Systems and Software*, 58:65–81, 2001.

[5] R. Hamlet. Random testing. In *Encyclopedia of Software Engineering*, pages 970–978. Wiley, 1994.