Volume 8, No. 5, May-June 2017
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

ISSN No. 0976-5697

Improved Time Performance of Adaptive Random Partition Software Testing by Applying Clustering Algorithms

K. Devika Rani Dhivya
Assistant Professor / Research Scholar,
Bharathiar University /
Department of BCA & MSc SS,
Sri Krishna Arts and Science College,
Coimbatore-8, Tamilnadu, India

Dr. V.S. Meenakshi
Assistant Professor/ Research Supervisor, Bharathiar University /
Department of computer science,
Chikkanna Government Arts College,
Tirupur-2, Tamilnadu, India

Abstract: Software testing is generally accepted technique for evaluating and taming software quality. Random testing (RT) is a major software testing strategy and a basic testing technique randomly generates test cases from the set of all possible program inputs. The simplicity of this testing makes it likely the most efficient testing strategy with respect to the time required for test case selection. Though very simple, RT is still considered as one of the state-of-the-art testing techniques, along with other more complicated and systematic testing methods. Its efficacy is notified to be less while considering its capacity of defect detection. This has been proven to pertinently overcome by Adaptive Testing (AT), on the other hand the methodology of AT is comprised of intricate complexity and high computational cost as its main constituents. Adaptive random testing (ART) is one major approach for enhancing RT. Another category of testing techniques is partition testing, which involves dividing the input domain up into a fixed number of disjoint partitions, and choosing test cases from within each partition. Partition testing has powerful, intuitive appeal, and analytical results show that even simple partitioning schemes may be more effective in fault detection than random testing. The existing hybrid approach is a combination of AT and RPT which is called as ARPT strategy, which enhances the AT. The objective of this proposed research is to improve random partition in ARPT strategies by utilizing clustering algorithms like Expectation Maximization (EM) algorithm and Nonnegative Matrix Factorization (NMF) clustering algorithms and the Self-organizing map (SOM) which can be efficiently utilized for partition. In this way random partitioning is improved to reduce the time conception and complexity in ARPT testing strategies.

Keywords: Software Testing, Random Testing, Adaptive Testing, Adaptive random Testing, ARPT, Clustering Algorithms, EM Algorithms, NMF& SOM.

1. INTRODUCTION

Software testing is generally accepted technique for assessing and improving software quality. One basic testing technique is randomly generating test cases from the set of all possible program inputs. Though very simple, random testing (RT) is still considered as one of the state-of-the-art testing techniques, along with other more complicated and systematic testing methods. In RT the generation and selection of test case are in one process and the selection of the test case is simply selecting test cases from an entire domain randomly and independently. [17][15][6][9]

Adaptive random testing (ART) is one major approach to enhancing RT.[8] ART is based on various empirical observations showing that many program faults result in failures in contiguous areas of the input domain, known as failure patterns. Another category of testing techniques is partition testing which involves dividing the input domain up into a fixed number of disjoint partitions, and choosing test cases from within each partition. In Random testing the test case are selected from whole test cases where as Random Partition techniques the test cases are randomly selected from the different partition, at that moment whole test cases are divided into several partition. In the traditional partitioning strategies, it is possible that the test cases in any two partitions are very close with each other.[4][5]

By combining AT and RPT which is called as ARPT strategy reduces the computational complexity of AT and to improve defect detection effectiveness.[8][21] Two variants for ARPT are ARPT-1 and ARPT-2, ARPT-1 exhibits better performance for different subject programs, whereas ARPT-2 requires considerable knowledge of ATs performance to achieve an acceptable overall performance. Therefore, ARPT-1 is recommended over ARPT-2 due to the higher and more robust performance of ARPT-1[1][2][3].

The main objective of this research is improving random partition in ARPT strategies by utilizing clustering algorithms[18][11].

2. BACKGROUNDSTUDY

Chen et al.,[2] mentioned that the RT is a poor method as it does not make use of any information to guide the generation of test cases although it is a commonly used testing technique for practitioners. It makes minimal use of the information from the specification or program code.

T.Y. Chen *et al.,* [1] projected two algorithms on adaptive random partitioning that offers more performance advantages than simple random testing, with considerably lower overhead than other ART algorithms. One is ART by random partitioning and the other is ART by bisection. The first algorithm exhibits failure measure 25-30% less than the random testing in block patterns, and 5% less for strip patterns. Negligible amount of failure measure for point patterns is found.

The second algorithm is effective as 25% more than random testing in block patterns, 5-8% more effective when strip patterns are considered. Marginal effectiveness in case of point patterns is shown. The main disadvantage of this work is that the point pattern doesn't exhibit any higher effectiveness.

TsongYueh Chen *et al.*, [7] proposed some ART algorithms by offsetting the edge preference, and offer a new family of ART algorithms. A series of simulations are conducted and it proved that these new algorithms not only select test cases more effectively, but also have better failure detection capabilities.

This paper investigated the edge preferences of FSCS-ART and RRT, and a new family of algorithms, namely ART with Partitioning by Edge and Centre (ECP-ART) was offered.

There exists one particular ART algorithm, namely ART by bisection (ART-B) that does not have any preference in the test case selection. The main disadvantage is that FSCS-ART and RRT only outperform ART-B when the failure rate is small.

The below table 1 shows Comparison of different Techniques used by RT, AT and ART. [20]

TABLE I. Comparison of different Techniques used by RT, AT and ART [20]

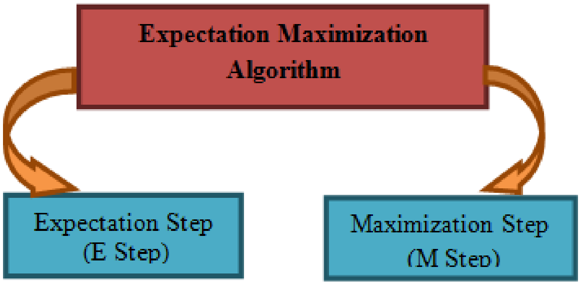
S.No	**Year**	**Authors**	**Technique**	**Computational Overhead**	**Code / Test Case Coverage**	**Fault Detection**	**Time Taken**
1.	**2004**	T.Y. Chen et al	ART by random partitioning and the other is ART by bisection	Reduced	Increased	Improved	-
2.	**2005**	J. Mayer	Bisection with Restriction.	Reduced	Not improved	Improved	Reduced
3.	**2006**	T. Y. Chen et al	Iterative Partition	Reduced	-	-	-
4.	**2008**	TsongYueh Chen et al	FSCS-ART and RRT	Reduced	-	Improved	Increased
5.	**2009**	Andrew F. Tappenden et al	Evolutionary algorithm like eAR and FSCS	Reduced	-	Improved	Reduced
6.	**2009**	Zhiquan Zhou et al	A Dynamic partitioning strategy	Reduced	-	Not improved	Reduced
7.	**2010**	TsongYueh Chen et al	ART	Reduced	-	-	Reduced
8.	**2012**	J. Mayer[10]	Adaptive Random Testing by Bisection with Restriction	-	Increased	Improved	Reduced
9.	**2013**	Ali Shahbazi et al	RT- RBCVT, ART-RBCVT and QRT-RBCVT	-	Increased	Improved	Reduced
10.	**2013**	Cliff Chow et al	Divide and Conquer method in ART technique	Reduced	Remains Same	Improved	Reduced
11.	**2014**	JunpengLv, Hai Hu et al.,	ARPT-1, ARPT-2	Reduced	Increased	Improved	Reduced
12.	**2015**	BoJiang et al.,	Adaptive-Randomized Techniques (APFD) Average Percentage of Fault Detection	Reduced	Increased	Improved	Reduced

III. METHODOLOGY

A. EM Algorithm for ARPT

The EM algorithm is utilized to find (locally) most extreme probability parameters of a factual model and it can be utilized as a part of situations where the issue can not be fathomed straightforwardly by the conditions. It incorporates expansion to obscure parameters and known information perceptions with inert variables.Typically Finding a most extreme probability arrangement needs enamoring the subsidiaries of the probability work with regard to all the obscure esteems the parameters and the inactive factors and simultaneously settling the subsequent equations.[12]

The EM algorithm seeks to find the MLE of the marginal likelihood by iteratively applying the following two steps:



E-step: Calculate the expected value of the log likelihood function, with respect to the conditional distribution of given under the current estimate of the parameters :

$$q(t+1) = arg_Q \max F(q, \theta^{(t)}) \quad \rightarrow \quad 1$$

M-step: Find the parameter that maximizes this quantity:[19]

$$\theta^{(t+1)} = arg_\theta \max F\left(q^{(t+1)}, \theta\right) \rightarrow 2$$

- X Observed variables Z Latent (unobserved) variables
- θ(t) The estimate of the parameters at iteration t.
- `(θ) The marginal log-likelihood logp(x|θ)log p(x,z|θ) The complete log-likelihood, i.e., when we know the value of Z.
- q(z|x,θ) Averaging distribution, a free distribution that EM gets to vary.
- Q(θ|θ(t)) The expected complete log-likelihood
- Pzq(z|x,θ)logp(x,z|θ) H(q) Entropy of the distribution q(z|x,θ). [14]

Note that in typical models to which EM is applied:

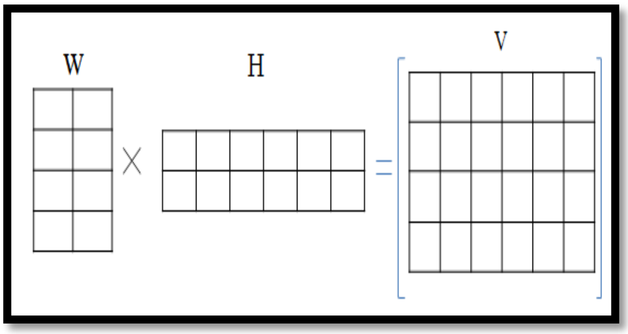
1. The watched information focuses might be discrete (taking esteems in a limited or countably boundless set) or nonstop (taking esteems in an uncountably unbounded set). There may in actuality be a vector of perceptions related with every information point.

2. The missing values(aka latent variables) are discrete, drawn from a settled number of qualities, and there is one idle variable for each watched information point.

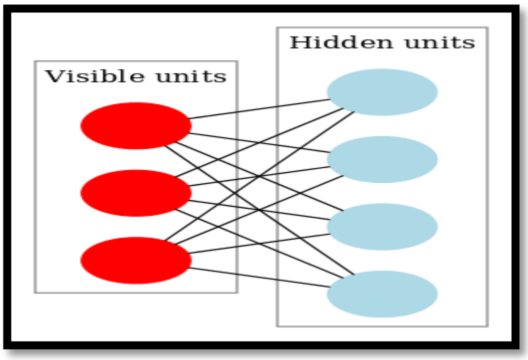
3. The parameters are ceaseless, and are of two sorts: Parameters that are related with all information focuses, and parameters related with a specific estimation of a dormant variable.

B. Non-Negative Matrix Factorization (NMF) for ARPT

Non-negative matrix factorization (NMF), also non-negative matrix approximation is a group of algorithms in multivariate analysis and linear algebra where a matrix V isfactorized into (usually) two matrices W and H, with the property that all three matrices have no negative elements.This non-negativity makes the resulting matrices easier to inspect. Also, in applications such as processing of audio spectrograms non-negativity is inherent to the data being considered. Since the problem is not exactly solvable in general, it is commonly approximated numerically.The approximate non-negative matrix factorization: the matrix V is represented by the two smaller matrices W and H, which, when multiplied, approximately reconstruct V.[21]



NMF as a probabilistic graphical model: visible units (V) are connected to hidden units (H) through weights W, so that V is generatedfrom a probability distribution with mean.



C. A Self-Organizing Map for ARPT

Self-organizing map (SOM) is characterized as a neural strategy for grouping. They are not the same as other simulated neural systems as they apply focused learning instead of mistake remedy learning, and as in they utilize an area capacity to protect the properties of the information space. It is showing the two spatial spaces of connection among bunches. SOM has possessed the capacity to display the information focuses that are in one or three-dimensional space, that given by SOM abilities. Additionally, due to the simple of representation and the exchange off between data content two dimensional spaces have been utilized all the more frequently [16][13].

The SOM algorithm is presented in detail as follows:[16]
1. Select two parameters from the log record which are Session ID and asset address, to put in the cluster.
2. Information parameter esteems are standardized and given as a numeric lattice.
3. The lattice is as contribution of the SOM calculation.
4. Set the learning rate and neighborhood separate with cycle number for deciding the groups of experiments by running the SOM calculation. For this situation analyzed and
5. Utilizing the separation capacity to check the closeness degree between experiments. Where x is info test and w is the weight vector of i'th hub.
6.The victor of the opposition between hubs in a system hub with the base separation is chosen.
7. The weights to all hubs inside a topological separation refreshed by rehash step 6 for all passages of the grid.
8. The yield is given test suite closeness test cases in similar gatherings.
After the test cases are trained through repeated presentations of all test, present unit input vectors of every test case to the trained case and assign the winning test case

for related application. Update the number by labeling the node as the number of test cases allocated.

IV. RESULT AND ANALYSIS

This proposed work comprises ARPT with clustering algorithms like EM, NMF and SOM. The representativeness of the units under test is the major threat to external validity. The process of assigning ARPT with clustering algorithm for a particular test case of an application shows the effective test coverage. While comparing the existing algorithm like RT, RPT, PSS, GA, ARPT-1 & ARPT-2 algorithm with the proposed ARPT with clustering algorithm gives a best result. The enhanced ARPT with Clustering Algorithm reduces the time taken to process the unit under test.The below Table 1 shows the time coverage for each method.
Time is calculated by the time taken for each method to execute the test case of an application.

TABLE II.Time Comparison with Methods

Methodology	Time (Milli Seconds)
RT	**3.5**
RPT	**3.4**
PSS	**3.3**
GA	**3.2**
ARPT-1	**2.2**
ARPT-1(EM)	**2.1**
ARPT-1(NMF)	**2.04**
ARPT-1(SOM)	**1.8**
ARPT-2(EM)	**3.0**
ARPT-2 (NMF)	**2.9**
ARPT-2 (SOM)	**2.8**

The comparisons of time for each method are given to analyse the time taken to execute the test case. Measuring the time performance of the proposed algorithms (ARPT-1(EM), ARPT-1(NMF), ARPT-1(SOM)), by using the chart, the performance is compared between RT, RPT, PSS, GA, ARPT-1(EM), ARPT-1(NMF), ARPT-1(SOM), ARPT-2(EM), ARPT-2 (NMF), ARPT-2(SOM). Here, the performance time of all the methods calculated by Minutes. Time taken to execute test case of an applications are shown in the beloveFig 1. Thus the executing time taken by ARPT –1 with the clustering algorithms are less while comparing with the existing algorithms.

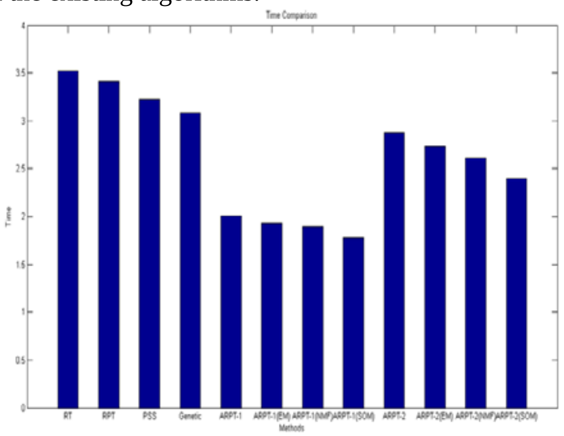


Fig 1. Comparison of the Time Vs Methods

V. CONCLUSION

ARPT with clustering algorithm automatically derives good parameter values for test application, and it is capable of achieve less time for executing a test of an application. Thus the proposed ARPT 1 with clustering algorithm solves the parameter space of problem between the target method and objective function of the test data. It achieves high test case coverage within less time and produces better accuracy. Finally, the performance of the ARPT1 with Clusterig algorithm is better and coverage of the test cases is high and less time to coverage of all test cases rather than the other methods. In future work, Optimation algorithms like Bat-inspired, Ant-colony algorithms and other advanced Meta heuristic optimization algorithm also can be used for ARPT -1 to derive an optimal solution. They may reach much better result and high test case cover with less time for an application.

VI. REFERENCE

[1] T.Y. Chen et al, “Adaptive Random Testing Through Dynamic Partitioning”, Proceedings of the Fourth International Conference on Quality Software (QSIC’04) IEEE, 0-7695-2207-6/04 .

[2] Chen *et al*, “Distributing Test Case more evenly in Adaptive Random Testing”, The Journal of Systems and Software 81 (2008) 2146–2162.

[3] Feng Ye, Chang Liu, HaiHu , “On the Computational Complexity of Parameter Estimation in Adaptive Testing Strategies”, 15th IEEE Pacific Rim International Symposium on Dependable Computing, 2009.

[4] Andrew F. Tappenden, “A Novel Evolutionary Approach for Adaptive Random Testing”, IEEE transactions on reliability, vol. 58, no. 4, December, 2009.

[5] ZhiQuan Zhou, ArnaldoSinaga, Lei Zhao,”Improving Software Testing Cost-Effectiveness Through Dynamic Partitioning”, Ninth International Conference on Quality Software, 2009.

[6] Liu, H, Xie, X, Yang, J, Lu, Y and Chen, T 2010, 'Adaptive random testing by exclusion through test profile', in Proceedings of the 10th International Conference on Quality Software, QSIC 2010, USA, 14-15 July 2010, pp. 147-156.

[7] T.Y. Chen, “Adaptive random Testing: the art of test case diversity”, The Journal of Systems and Software 83 (2010) 60–66.

[8] Cliff Chow et al, “The ART of Divide and Conquer An Innovative Approach to Improving the Efficiency of Adaptive Random Testing “ Symposium on Engineering Test Harness (TSETH ’13), Proceedings of the 13th International Conference on Quality Software (QSIC ’13), IEEE Computer Society, Los Alamitos, CA (2013).

[9] Ali Shahbazi , Andrew, “ CentroidalVoronoi Tessellations— A New Approach to Random Testing”, IEEE transactions on software engineering, vol. 39, no. 2, February 2013.

[10] J. Mayer, “Adaptive Random Testing by Bisection with Restriction,” in Formal Methods and Software Engineering, 3785th ed., vol. 3785, K.-K. Lau and R. Banach, Eds. Manchester, UK: Springer Berlin Heidelberg, 2005, pp. 251–263.

[11] T. Y. Chen, D. H. Huang, and Z. Q. Zhou, “Adaptive Random Testing Through Iterative Partitioning,” in AdaEurope ’06 Proceeding of the 11th Ada-Europe International conference on Reliable Software Technologies, 2006, pp. 155–166.

[12] K. P. Chan, T. Y. Chen, and D. Towey, “Restricted Random Testing,” Proc. 7th Eur. Conf. Softw. Qual. (ECSQ 2002), vol. 2349, no. Lecture Notes in Computer Science, pp. 321–330, 2002.

[13] Kohonen, T. 2001. Self-organizing maps of massive databases. International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications, Vol.9, No.4, 179–185.

[14] Kak*et.al* ,”Expectation-Maximization Algorithm for Clustering MultidimensionalNumerical Data”,Purdue University, November 18, 2014.

[15] MojtabaRaeisiNejad Dobuneh1, Dayang N. A. Jawawi1, MojtabaVahidi Asl2, and Mohammad V. Malakooti3 “Clustering Test Cases in Web Application Regression Testing using Self-Organizing Maps”, Int. J. Advance Soft Compu. Appl, Vol. 7, No. 2, July 2015 ISSN 2074-2827.

[16] YaminiPathania, GurpreetKaur , “ Role of Test Case Prioritization based on Regression Testing using Clustering “,International Journal of Computer Applications (0975 – 8887) Volume 116 – No. 19, April 2015.

[17] JunpengLv, Hai Hu, Kai-Yuan Cai, and TsongYueh Chen, “Adaptive and Random Partition Software Testing “,EEE Transactions On Systems, Man, And Cybernetics: Systems, Vol. 44, No. 12, December 2014 PP: 1649.

[18] T. Li and C. Ding. The relationships among various nonnegative matrix factorization methods for clustering. In In Proceedings of the 2006 IEEE International Conference on Data Mining (ICDM 2006), pages 362–371, 2006.

[19] Chuong B Do &SerafimBatzoglou ,”What is the expectation maximization algorithm?”, ©2008 Nature Publishing Group, Stanford, California, USA.

[20] K. Devika Rani Dhivya, V.S. Meenakshi, Paper “A Comparative Study on Adaptive and Random Testing Technique “,International Journal of Advanced Research in Computer Science and Software Engineering Research Volume 6, Issue 2, February 2016 ISSN: 2277 128X.

[21] F.WangandP.Li.”Efficientnonnegativematrixfactorization withrandomprojections”, InProceedingsof2010SIAMDataMiningConference(SDM’10), pages 281–292, 2010.

