TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 6276 - 6288

Enrichment of Fault Features by Forming Ml Hypothesis

P. Patchaiammal
Research Scholar,
Bharath University, Chennai

R. Thirumalaiselvi
Research Supervisor,
Bharath University, Chennai,
Assistant Professor,
Computer Science Department,
Govt. Arts College (Men), Nandanam, Chennai.

Article Info
Volume 82
Page Number: 6276 – 6288
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 30 January 2020*

Abstract:
Modeling and optimization of applications of engineering sciences and technology advances in fault detection and diagnostics. By utilizing past data one is to promote environmentally safe modeling approaches. The software developers and users have found difficult to learn the software fault because softwares are developed using most of the learning algorithms. So, the developers needed some learning technique in order to prevent and identify the fault in pre-development.This will leads to the introduction of green engineering in software development. This paper examines and forms the hypothesis space for fault features classification in post release so as to form the learning technique to identify it in the development stage itself to reduce rework. This paper also checks the classification of input features that are to be relevant to the outcome to be predictedare not by using different hypothesis testing. Our result signifies the hypothesis space using machine learning for finding feature set of fault prediction feature set. Eight NASA PROMISE Repositories are used in this paper for the hypothesis testing. This paper used to identify the best Hypothesis Testing for solving the feature selection problem in machine learning Hypothesis Space. Several performance measures are calculated and results of the experiment revealed that choosing chi-square hypothesis testing produces more relevant result for fault prediction feature set formation.

Keywords: *Machine Learning(ML), Null Hypothesis(H_0),Alternative Hypothesis(H_1),Green Engineering,F-test, t-test, z-test, chi-square test, p-value, Genetic Hypothetical Goal Question Metrics (GHGQM).*

1. INTRODUCTION

Advances in software modeling include optimization in fault detection product modeling. This interoperability simulates the tools for sustainable environment. The goal of this paper is to promote environmentally safe engineering by utilizing the past approaches between software faults. To develop a proper fault prediction feature set a meaningful hypothetical question is to be formed. It is done by Genetic Hypothetical Goal Questionary Metrics (GHGQM).This method used to create a null hypothesis H_0 and seek to prove it wrong. This will help to omit the unnecessary data that is noisy in random set.Modeling and optimization of this type

January - February 2020
ISSN: 0193 - 4120 Page No. 6276 - 6288

of feature set creates environmentally safe software. Genetic Hypothetical Goal Question Metrics is based on machine learning classifier technique used to learn the model from the training data. Training set is used to model target function of learning algorithm. This method need step by step processing in order to approximate the target function using training dataset. Hypothesis has set of estimation which will happen in future as the values which is to be modified and additional experiments are made to predict. It compare actual value with observed evidence. Data is not fulfilled without a hypothesis to evaluate. The hypothesis is the 'model of reality' if the experiment may be repeated by others to see if they get the same results. If results are not the same, then the hypothesis is rejected. If results are same, then the hypothesis is accepted. A Classifier is a special case of hypothesis. It is a discrete valued function which is used to assign class labels to particular data points.

If 'H' is a hypothesis space H_s. By using ML Algorithm classifier function is formed. Then that function is used to categorize the attributes (labels) in dataset. In fault report dataset, the classifier of hypothesis is either faulty or non-faulty.Hypothesis test is used to determine whether the input attributes that is the input features are relevant to the outcome to be predicted. Hypothesis Test is mainly used for feature selection. Hypothesis Test is very useful in classification problem in which the input are to be categorized mainly.

2. MACHINE LEARNING

Learning focuses on the growth of model by using algorithm to expose new prediction based on available data history. Machine learning is a subset of Artificial Intelligence. Machine learning means empowering the computer system with the ability to "learn". Basic ML models become progressively better but still need adjustments to make accurate prediction. ML forms a training set using an already available data set. Then use that training set for prediction of result. This ML process needs expert algorithms. The term Machine learning means "giving the ability to the computer to learn without fully programmed".

ML is used to formulate complex models by using algorithms which lend themselves for prediction. Machine Learning has four types of learning methods. The methods are supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning.

Supervised Learning has labels for training data set. It helps to train the machines to learn the relationships between input and the result. In this the label is the known description given to the objects in the data set. These labels train the machines and also provide the structure of the algorithm outputs, as the final result must be from any one of the outputs. In this learning machines learn by using the patterns to classify the data and then apply those patterns to classify new data. In this learning labeled data is loaded into the system. It is done by a human trainer. Then the model is trained. The inputs are connected with the outputs. As new data is introduced, the algorithm is applied and finally the output is categorized form of data. Supervised algorithms are off two categories known as classification (unordered limited values) and regression (label is real number).

Unsupervised learning have no idea or knowledge about the resultant label. In this type, machines find patterns in the data by its own. There is no specific predicted outcome. In this learning raw unlabeled data is loaded into the system. Algorithm is used to find the pattern by its own. Similar patterns are identified to provide output result. Unsupervised algorithms are off two categories known as clustering (segment the data into several groups) and dimension reduction (reduce the variable number so as to find the exact solution).

Semi-supervised learning includes the combination of supervised and unsupervised learning types together. In this learning a part of the data is labeled and other part is raw unlabeled data. In this type,

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 6276 - 6288

initial training set is loaded with labeled data. The model is trained on those data. Then new raw unlabeled data is added, algorithm is used on new raw unlabeled data for classification of pattern. The process is repeated until resultant pattern is found.

Reinforcement learning includes a set of actions, parameters, and end values. It will train the machine by trial and error method. It will learn from past efforts to achieve the best possible result.

ML is used to improve the decision so as to increase the quality and productivity. ML also used for prediction analysis. Even though there is growth of technology there is a need for better tools to understand the current dataset. To achieve this the development of smart learning machines.

3. HYPOTHESIS STUDY

Hypothesis is some guess of education which needs some evaluation. A proper hypothesis is always testable which is either true (or) false. Hypothesis is formal before the test outcome is known. A good hypothesis is used to find the evidence to make predictions about sample. It is also used to explain the presence of relationship of survey. In problem domain hypothesis is used for function approximation. Hypothesis approximation target function which maps inputs to outputs in domain knowledge. This type of model approximation of learning target function is known as hypothesis in Machine Learning. In problem domain hypothesis is used for function approximation which advances in prediction modeling including optimization and fault diagnostics. Target function which map inputs and outputs in domain knowledge. This type of model approximation of learning target function is known as hypothesis in machine learning. Learning is used to search the space of hypothesis for training set. Learning for a Machine Learning used to navigate the chosen space of hypothesis towards the best to target function approximation.

3.1 Hypothesis definition

Let 'x' denotes the input and 'y' denotes the output then y=f(x) is the target function. In the problem solving technique 'f' is an unknown function then h(x) is the hypothesis function to approximate the unknown function 'f'. The set of all h(x) that is h_1 (x), h_2 (x), h_3 (x)… h_n (x) forms the hypotheses space 'H'.

3.2 Notations used in Hypothesis

- h – Single hypothesis.
- h_1 (x), h_2 (x), h_3 (x) ….h_n (x) – Hypotheses set
- H – Hypothesis set (or) space.

3.3 Parameter in Hypothesis Testing

Using historical data, hypothesis testing creates statistical decision. It is used for creating population from dataset. Always hypothesis testing is essential for machine learning. It is used to determine a statement about population set. Hypothesis testing is best for sample dataset evaluation. Hypothesis testing is based on standard normalization.

3.3.1 Null Hypothesis

It is a basic general statement based on population domain. It is represented by H_0. Always Null Hypothesis is tested to check the likelihood of this statement being true in order to make decisions to accept (or) reject our alternative hypothesis. It include equal, less than equal (or) greater than equal.

3.3.2 Alternative Hypothesis

It is opposite to null hypothesis. It is taken from real observation results with little change if necessary.

3.3.3 Significance Level

Generally, 95% label is set as significance level and 5% is error rate. Testing is applied to check error rate if it is less than 5% then one has to reject the null hypothesis and prepare the actual one.

January - February 2020
ISSN: 0193 - 4120 Page No. 6276 - 6288

3.3.4 Target function

It is of the form f(x) = y. It is used to define the predictive modeling. It is also used to modeling a process by approximating a particular function. In our fault prediction feature set, the target function is used to distinguish defective and non-defective software data during the software development.

3.3.5 Hypothesis Mapping

It is the mapping data set into the hypothesis space. Data can be specified. This Mapping space consists of all hypothesis that can explain the parts of the observed data. Hypothesis at the hypothesis space will explain some part of the data. In domain of ML the hypothesis may refers to specific methods now working properly. In software development fault occur refers to the hypothesis space. The mapping of data space to the hypothesis space is very complex. The complexity arises that path of data space and part of hypothesis space, mapping between M $\rightarrow$N and also there is interactivity feature. It means H_3 is present because of H_2 present H_4 is present because H_3 is present and so on. If data space 'D'= {D_1, D_2, D_N} then hypothesis space 'H'= {H_1,H_2,……H_N}. The mapping helps to deal with only abstract data for a specific field not to deal with raw data.

3.3.6 P-value

It is used to determine the significance of the hypothesis result. If p-value is $\leq$ 0.05 then the result is the strong evidence of rejection of the null hypothesis. If p-value is $\geq$ 0.05 means null hypothesis is accepted. If the p-value is = 0.05 means the null hypothesis is either accepted or rejected which is said to be marginal.

4. MAJOR CLASSIFICATION FORMULAS USED IN HYPOTHESIS TESTING LEARNING PROCESS

Hypothesis testing is about population parameter assumption which is used for making decision using dataset. It is basic for machine learning problem creation. It is classified into four types. Those are F-test, t-test, z-test and chi-square test.

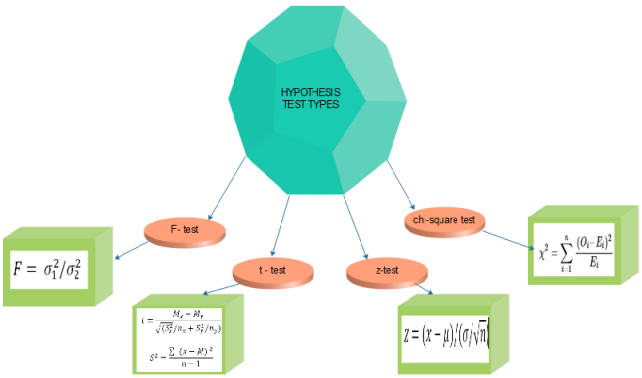

Diagram 1 Basic Structure of Hypothesis testing classification

4.1. T – Test

It determines the significance the difference between two groups of certain features.

$$t = \frac{M_x - M_y}{\sqrt{(S_x^2/n_x + S_y^2/n_y)}} \quad (1)$$

$$S^2 = \frac{\Sigma\ (x-M)^{\ 2}}{n-1} \quad (2)$$

Where, M_x and M_y are the mean values of the two samples from the population. $n_x$and$n_y$are the sample space of the two samples. S is the standard deviation. n is the total number values. x and y are the individual value in each samples.If thecalculated t-test value is greater than the critical value then null hypothesis is rejected. It means there is no significance difference between the populations taken.If the value is lesser then null hypothesis is accepted it means there is significance difference between the populations taken.

4.2. Z – Test

It is used in large data size and also in randomly selected population in which each attribute has equal chance of being selected. It takes the population parameter such as mean and standard deviation in order to check the hypothesis.

$$z = (x - \mu)/(\sigma/\sqrt{n}) \quad (3)$$

Where, x is a sample mean. μ is a population mean. $\sigma/\sqrt{n}$is the population standard deviation. If the

January - February 2020
ISSN: 0193 - 4120 Page No. 6276 - 6288

result of z is lesser than the critical value thennull hypothesis is accepted otherwise the null hypothesis is rejected.

4.3. F-Test (ANOVA - Analysis of variance)

This test is used to compare more than two groups of data at same time.

$$F = \sigma_1^2/\sigma_2^2 \qquad (4)$$

Where $\sigma_1^2$is variance 1 and σ_2^2 is variance 2.F test used to measure the significance difference between all pairs of attributes. If the result is lesser then we accept the null hypothesis and conclude that there is a significance difference between the attributes otherwise we reject the null hypothesis.

4.4. Chi-square Test

It is applied two-categorical variables of same population. It is used for determining whether there is significant association between two variables.This test is applied to find the goodness of the fit test. This test also determines the matching of samples with the population. This test also used to check the independence of the variables in the table for fitting the data.

$$\chi^2 = \sum_{i=1}^{n} (O_i - E_i)^2/E_i \qquad (5)$$

Where O_i is the observed data,E_i is the expected data and n is the total number of instances.If the value of chi-square is small then the data is fit in to the population taken. If the value of chi-square is large then the data is doesn't fit in to the population. It means the variables are independent.

In this paper we check the machine learning Hypothesis test to form the hypothesis space for the fault prediction feature set formation. This is done by setting the null hypothesis using Genetic Hypothetical Goal Questionary Metrics (GHGQM). All four major hypothesis are tested to select the best hypothesis test for hypothesis space feature (attribute) set formation

5. GENETIC HYPOTHETICAL GOAL QUESTION METRICS (GHGQM) FRAMEWORK

The general fault classification is to find the occurrence of fault in software development phase. The GHGQM framework helps to experience on teaching software engineer to make environmentally safe fault diagnostics. The GHGQM is formed by using historical data form NASA PROMISE repository which creates green engineering in software development.

a. Is stable solution necessary?
   From the sample data understand and improve the performance by several analysis.
b. Do you need the input variables factorization?
   This is done by constructing the sum of feature clustering.
c. Is individual feature understanding is necessary?
   Understanding of system by ranking in large system is needed to get baseline result.
d. Is prediction model necessary?
   This is used to check the need of new prediction model by using the available features.

6. NULL AND ALTERNATE HYPOTHESIS FOR FAULT PREDICTION FEATURE SET

Hypothesis Space formation is the starting point for designing a learning experiment. Hypothesis must be a reasonable conjecture. Hypothesis would be “educated guess”. Hypothesis provide the path way to understanding. An experimentable answer to a scientific question is known as hypothesis. By doing experimentable testing, one can determine whether the hypothesis is right or wrong. By making one or more predictions, hypothesis is tested. A hypothesis may have multiple prediction. To prove scientifical question, one or two predictions is enough.By using hypothesis, a set of possible approximations of mapping function of 'F' from training sample 'X' formed to make hypothesis space.Hypothesis test is used to check a claim is correct or not. In hypothesis

January - February 2020
ISSN: 0193 - 4120 Page No. 6276 - 6288

we have null hypothesis H_0 and alternative hypothesis H_1. A claim is accepted in null hypothesis then the alternative hypothesis is rejected. If null hypothesis is rejected based on evidence an alternative hypothesis needs to be accepted always the assumption starts with the null hypothesis is true.

Null hypothesis: It refers to the "result" that is forbidden by the hypothesis under consideration.

H_0: Software Development Metrics Report along with organizational post release fault dataset does not provide better fault prediction feature set.

H_1: Software Development Metrics Report along with organizational post release fault dataset does not provide better fault prediction feature set.

6.1 Attributes Used in Fault Classification

It is used to identify the known and unknown fault so that as to detect the exact phase of a fault. It is also used to form fault report which may use for perfective maintenance to reduce research. If one use only the fault occurred in the implementation phase as to prepare by report then this type of predictions used to form only a simple implementation fault classification. If a fault classified in one phase is misplaced in fault report then it creates bias in classification. So there is a need for all software development fault prediction. In this paper, we did classification of fault from Genetic Hypothetical Goal Question Metrics (GHGQM) Framework. By using public extent projects from Apache, Mozilla and also some hacking system reports of various organization.

Function approximation is necessary for the learning of mapping function from input to output prediction. The bug predicting model may be discrete or continuous. Classification is used for predicting discrete output is used for predicting continuous output. The prediction model using classification is done by using historical data. The conversion of predictive model to mathematical problem using mapping function (f) from training input feature (x) to output variable (y).

6.2. Dataset collection

Public dataset is the one which can be freely available in the Promise repositories. In the experiment we taken 29 class level metrics out of 64 from the PROMISE repository which is classified as faulty and non-faulty.

The following list describes 29 attribute classes which are responsible for the fault software modules. These attributes was applied in fault prediction machine learning algorithm. This paper used to identify whether the 29 attributes are necessary for feature selection by omitting other noisy in the dataset. Various hypothesis testing are used to check the attributes against the null hypothesis H_0 for best fault prediction feature set. The list of 29 attributes are listed in table 1.

Table 1 List of feature attributes for hypothesis testing

S.NO	ATTRIBUTES SELECTED	HYPOTHESIS SET
1	BRANCH_COUNT: It includes the number of branches in the code.	h_1
2	CONDITION_COUNT: It includes number of conditions in the code.	h_2
3	CYCLOMATIC_COMPLEXITY: It is the cyclomatic complexity values counted by metrics.	h_3
4	CYCLOMATIC_DENSITY: It includes the destiny value described by cyclomatic metrics.	h_4
5	DECISION_COUNT: It includes number of decisions in the code.	h_5
6	DECISION_DENSITY: It includes the destiny value of decisions in the code.	h_6
7	DESIGN_COMPLEXITY: It includes the complexity value related to design phase.	h_7
8	DESIGN_DENSITY: It includes the destiny value related to design phase.	h_8
9	ESSENTIAL_COMPLEXITY: It means metrics essential complexity.	h_9

January - February 2020
ISSN: 0193 - 4120 Page No. 6276 - 6288

10	ESSENTIAL_DENSITY: It means metrics destiny value related to essential metrics.	h_{10}
11	LOC_EXECUTABLE: It includes the number of executable LOC values.	h_{11}
12	PARAMETER_COUNT: It includes the number of parameters in the code.	h_{12}
13	HALSTEAD_CONTENT: It includes the content of Halstead metrics related to the code.	h_{13}
14	HALSTEAD_DIFFICULTY: It includes the detail of difficulties occurred in the Halstead metrics of the code.	h_{14}
15	HALSTEAD_EFFORT: It includes the value of the effort related to Halstead metrics used in the code.	h_{15}
16	HALSTEAD_ERROR_EST: It describes the essential error list related to Halstead metrics used in the code.	h_{16}
17	HALSTEAD_LENGTH: It includes the detail of length of Halstead metrics.	h_{17}
18	HALSTEAD_LEVEL: It includes the level value of the Halstead metrics used in the code.	h_{18}
19	HALSTEAD_PROG_TIME: It describes the time value taken by Halstead metrics used in the code.	h_{19}
20	HALSTEAD_VOLUME: It includes the total volume of the Halstead metrics used in the code.	h_{20}
21	MAINTENANCE_SEVERITY: It describes the severity details about maintenance phase.	h_{21}
22	MODIFIED_CONDITION_COUNT: It includes the number of conditions modified in the code.	h_{22}
23	MULTIPLE_CONDITION_COUNT: It includes the multiple number of conditions used in the code.	h_{23}
24	NODE_COUNT: It describes the number of nodes in the code.	h_{24}
25	NORMALIZED_CYLOMATIC_COMPLEXITY: It includes the normalized value of cyclomatic complexity related to the code.	h_{25}
26	NUM_OPERANDS: It includes the number of measure of operands in the code.	h_{26}
27	NUM_OPERATORS: It includes the number of measure of operators in the code.	h_{27}
28	NUM_UNIQUE_OPERANDS: It includes the unique number of the measure of operands in the code.	h_{28}
29	NUM_UNIQUE_OPERATORS: It includes the unique number of the measure of the operators in the code.	h_{29}

7. IMPLEMENTATION STEPS FOR PROPOSED STUDY

7.1 Accumulating the Attributes

The dataset available consist of both numerical and non-numerical data. The training set accept only numerical values. Therefore, we need to convert the non-numerical data by using encoding technique.

7.2 Experiment Results

In this section, the results of datasets are analyzed and performances are summarized in this work four experiments has been done. In the first experiment, F-test is applied against the 29 fault dataset attributes and the corresponding p-values are found. The experiment result shows that 21 attributes rejecting the null hypothesis and 8 attributes has accepting the null hypothesis. So, contradiction occurs in the null hypothesis. It is described in Table 2 and the corresponding performance is shown in Diagram 2. From Table 2, the yellow color value indicates the null hypothesisacceptance under F-test.

January - February 2020
ISSN: 0193 - 4120 Page No. 6276 - 6288

Table 2 Performance summary of 'F' Hypothesis test

S.No.	Hypothesis Test Type	attribute selected	pval	Null Hypothesis Accept/Reject
1	f test	BRANCH_COUNT	0.003275594	reject null hypothesis
2	f test	CONDITION_COUNT	0.002756104	reject null hypothesis
3	f test	CYCLOMATIC_COMPLEXITY	0.004262935	reject null hypothesis
4	f test	CYCLOMATIC_DENSITY	0.000289853	reject null hypothesis
5	f test	DECISION_COUNT	0.002970684	reject null hypothesis
6	f test	DECISION_DENSITY	0.597298366	accept null hypothesis
7	f test	DESIGN_COMPLEXITY	0.000812361	reject null hypothesis
8	f test	DESIGN_DENSITY	0.061815566	accept null hypothesis
9	f test	ESSENTIAL_COMPLEXITY	0.067830552	accept null hypothesis
10	f test	ESSENTIAL_DENSITY	0.825082021	accept null hypothesis
11	f test	LOC_EXECUTABLE	1.45E-05	reject null hypothesis
12	f test	PARAMETER_COUNT	0.287056425	accept null hypothesis
13	f test	HALSTEAD_CONTENT	1.95E-06	reject null hypothesis
14	f test	HALSTEAD_DIFFICULTY	0.002731623	reject null hypothesis
15	f test	HALSTEAD_EFFORT	0.081439885	accept null hypothesis
16	f test	HALSTEAD_ERROR_EST	0.000124302	reject null hypothesis
17	f test	HALSTEAD_LENGTH	9.99E-05	reject null hypothesis
18	f test	HALSTEAD_LEVEL	0.020885844	reject null hypothesis
19	f test	HALSTEAD_PROG_TIME	0.081439719	accept null hypothesis
20	f test	HALSTEAD_VOLUME	0.000178819	reject null hypothesis
21	f test	MAINTENANCE_SEVERITY	0.110685859	accept null hypothesis
22	f test	MODIFIED_CONDITION_COUNT	0.002705763	reject null hypothesis
23	f test	MULTIPLE_CONDITION_COUNT	0.002997194	reject null hypothesis
24	f test	NODE_COUNT	0.003240586	reject null hypothesis
25	f test	NORMALIZED_CYLOMATIC_COMPLEXITY	0.014227053	reject null hypothesis
26	f test	NUM_OPERANDS	0.00017244	reject null hypothesis
27	f test	NUM_OPERATORS	7.69E-05	reject null hypothesis
28	f test	NUM_UNIQUE_OPERANDS	2.61E-06	reject null hypothesis
29	f test	NUM_UNIQUE_OPERATORS	7.46E-06	reject null hypothesis

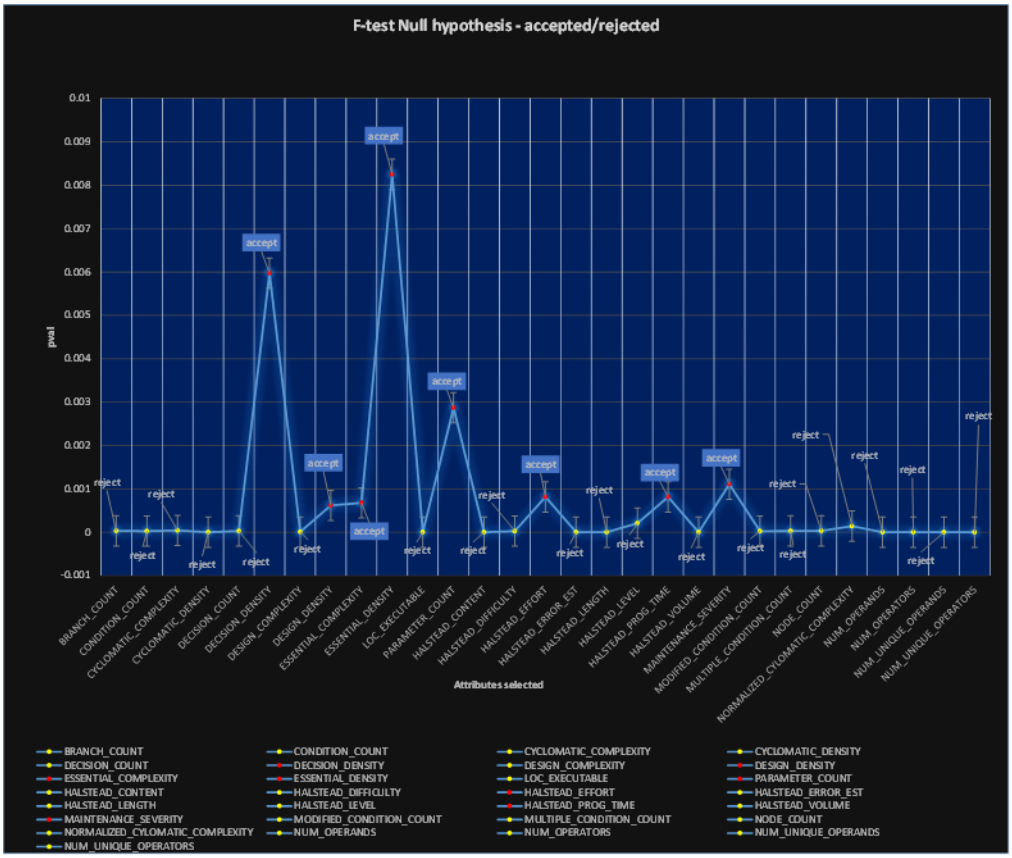

Diagram 2 A Plot of 'F' Hypothesis test Performance

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 6276 - 6288

The performance summary of z-test has been demonstrated in Table 3. This table includes the 29 fault dataset attributes and the corresponding p-values. By analyzing the result, it is found that 2 attributes known as CYCLOMATIC_COMPLEXITY and DESIGN_DENSITY are accepting the null hypothesis. This result shows the contradiction in rejecting the null hypothesis. Table 3 shows the performance summary and Diagram 3 shows the corresponding performance strategy.

Table 3 Performance summary of 'z' hypothesis test

S.No.	Hypothesis	attribute selected	pval	Null Hypothesis Accept/Reject
1	ztest	BRANCH_COUNT	4.3480418161251600E-10	reject null hypothesis
2	ztest	CONDITION_COUNT	8.4198502678969800E-20	reject null hypothesis
3	ztest	CYCLOMATIC_COMPLEXITY	8.6621215974075700E-01	accept null hypothesis
4	ztest	CYCLOMATIC_DENSITY	0.0000000000000000E+00	reject null hypothesis
5	ztest	DECISION_COUNT	1.8422472521801100E-04	reject null hypothesis
6	ztest	DECISION_DENSITY	0.0000000000000000E+00	reject null hypothesis
7	ztest	DESIGN_COMPLEXITY	9.4587210005578100E-34	reject null hypothesis
8	ztest	DESIGN_DENSITY	3.9270615668239600E-01	accept null hypothesis
9	ztest	ESSENTIAL_COMPLEXITY	3.1568293121841400E-71	reject null hypothesis
10	ztest	ESSENTIAL_DENSITY	0.0000000000000000E+00	reject null hypothesis
11	ztest	LOC_EXECUTABLE	7.2238355231517600E-36	reject null hypothesis
12	ztest	PARAMETER_COUNT	0.0000000000000000E+00	reject null hypothesis
13	ztest	HALSTEAD_CONTENT	2.3496688884900300E-77	reject null hypothesis
14	ztest	HALSTEAD_DIFFICULTY	6.3856497898069400E-55	reject null hypothesis
15	ztest	HALSTEAD_EFFORT	2.9125677297985900E-08	reject null hypothesis
16	ztest	HALSTEAD_ERROR_EST	4.0018172220755200E-17	reject null hypothesis
17	ztest	HALSTEAD_LENGTH	2.8965671469587300E-43	reject null hypothesis
18	ztest	HALSTEAD_LEVEL	0.0000000000000000E+00	reject null hypothesis
19	ztest	HALSTEAD_PROG_TIME	3.1564231976661800E-08	reject null hypothesis
20	ztest	HALSTEAD_VOLUME	1.9788787965690400E-31	reject null hypothesis
21	ztest	MAINTENANCE_SEVERITY	0.0000000000000000E+00	reject null hypothesis
22	ztest	MODIFIED_CONDITION_COUNT	1.5291896073498400E-05	reject null hypothesis
23	ztest	MULTIPLE_CONDITION_COUNT	1.7412307440454000E-05	reject null hypothesis
24	ztest	NODE_COUNT	1.0006698725934600E-29	reject null hypothesis
25	ztest	NORMALIZED_CYLOMATIC_COMPLEXITY	0.0000000000000000E+00	reject null hypothesis
26	ztest	NUM_OPERANDS	2.2146851240483600E-37	reject null hypothesis
27	ztest	NUM_OPERATORS	1.7160487706731800E-41	reject null hypothesis
28	ztest	NUM_UNIQUE_OPERANDS	1.8928485771469900E-40	reject null hypothesis
29	ztest	NUM_UNIQUE_OPERATORS	2.4141610984079800E-120	reject null hypothesis

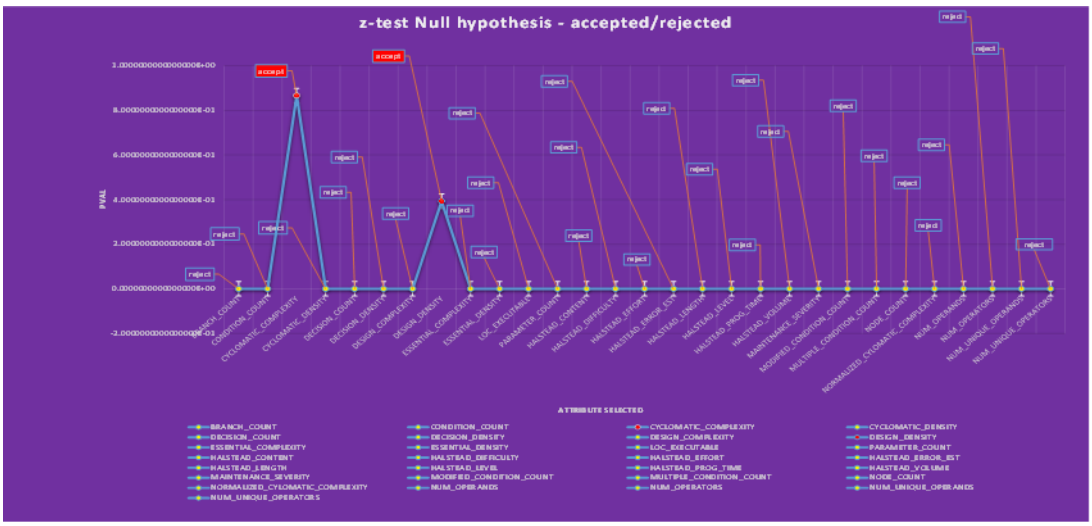

Diagram 3 A Plot of 'z' Hypothesis test Performance

The sensitivity analysis of t-test is described in Table 4. This table contains the results of 29 fault attributes selected and its specific p-values. This table has all p-values greater than the significance level but, has high variance in p-value. After the analysis of the result, it shows that there is a need for better hypothesis performance with the dataset to show the independence of the attributes. The performance plot shown in Diagram 4.

January - February 2020
ISSN: 0193 - 4120 Page No. 6276 - 6288

Table 4Performance summary of 't' Hypothesis test

S.No.	Hypothesis	attribute selected	pval	Null Hypothesis Accept/Reject
1	ttest	BRANCH_COUNT	7.5520915251098300E-36	reject null hypothesis
2	ttest	CONDITION_COUNT	2.5074003089509400E-35	reject null hypothesis
3	ttest	CYCLOMATIC_COMPLEXITY	4.3587682656615000E-36	reject null hypothesis
4	ttest	CYCLOMATIC_DENSITY	1.0827307190007300E-115	reject null hypothesis
5	ttest	DECISION_COUNT	4.9646579468535800E-36	reject null hypothesis
6	ttest	DECISION_DENSITY	3.8977561450646100E-287	reject null hypothesis
7	ttest	DESIGN_COMPLEXITY	3.5344630061308900E-36	reject null hypothesis
8	ttest	DESIGN_DENSITY	1.3858870495111100E-154	reject null hypothesis
9	ttest	ESSENTIAL_COMPLEXITY	6.6996066620370900E-33	reject null hypothesis
10	ttest	ESSENTIAL_DENSITY	9.2137818299475600E-24	reject null hypothesis
11	ttest	LOC_EXECUTABLE	3.4455619146566500E-40	reject null hypothesis
12	ttest	PARAMETER_COUNT	7.9172118197552200E-48	reject null hypothesis
13	ttest	HALSTEAD_CONTENT	1.3242655331900300E-66	reject null hypothesis
14	ttest	HALSTEAD_DIFFICULTY	1.3884764063890500E-76	reject null hypothesis
15	ttest	HALSTEAD_EFFORT	5.8023257998161100E-08	reject null hypothesis
16	ttest	HALSTEAD_ERROR_EST	2.1478598032925300E-26	reject null hypothesis
17	ttest	HALSTEAD_LENGTH	8.6586784633639600E-37	reject null hypothesis
18	ttest	HALSTEAD_LEVEL	9.3693731385618600E-70	reject null hypothesis
19	ttest	HALSTEAD_PROG_TIME	5.8028102408531800E-08	reject null hypothesis
20	ttest	HALSTEAD_VOLUME	6.0160114439521200E-27	reject null hypothesis
21	ttest	MAINTENANCE_SEVERITY	1.0939796052690900E-102	reject null hypothesis
22	ttest	MODIFIED_CONDITION_COUNT	2.2147063251594000E-34	reject null hypothesis
23	ttest	MULTIPLE_CONDITION_COUNT	2.5698576227446500E-35	reject null hypothesis
24	ttest	NODE_COUNT	8.9268623541298700E-43	reject null hypothesis
25	ttest	NORMALIZED_CYLOMATIC_COMPLEXITY	3.3082364501873900E-85	reject null hypothesis
26	ttest	NUM_OPERANDS	5.5755017160387800E-36	reject null hypothesis
27	ttest	NUM_OPERATORS	5.8004255301301100E-37	reject null hypothesis
28	ttest	NUM_UNIQUE_OPERANDS	1.8345854861385700E-46	reject null hypothesis
29	ttest	NUM_UNIQUE_OPERATORS	7.3913366086818700E-123	reject null hypothesis

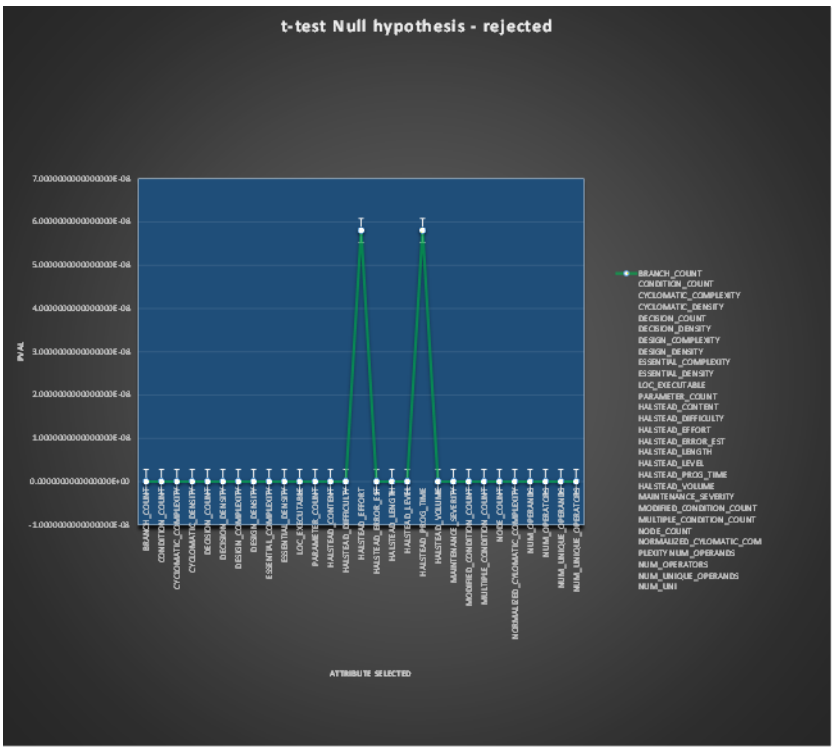

Diagram 4 A Plot of 't' Hypothesis test Performance

Table 5 shows the best result of the hypothesis test against Chi-square test. This table contains the result of p-value against the 29 fault attributes. All 29 attribute rejected the null

January - February 2020
ISSN: 0193 - 4120 Page No. 6276 - 6288

hypothesis. From the table, we analyzed there is no major variation in the p-value. The table also shows the Chi-square statistic level with high category relationship between the attributes which shows the independency of the variable. This hypothesis test performed better compared to all the other three hypothesis test. This result shows that all 29 attributes are necessary for the formation of feature set of fault prediction classification. The consequence of the table is shown in Diagram 5.

Table 5 Performance summary of 'Chi-square' Hypothesis test

S.No.	Hypothesis Test Type	attribute selected	chi-square statistic	pval	Null Hypothesis Accept/Reject
1	chitest	BRANCH_COUNT	68.59415456	1.1102230246E-16	Reject
2	chitest	CONDITION_COUNT	70.11385946	1.11E-16	Reject
3	chitest	CYCLOMATIC_COMPLEXITY	72.38389682	0.0	Reject
4	chitest	CYCLOMATIC_DENSITY	14.88745775	0.000114119	Reject
5	chitest	DECISION_COUNT	48.13172124	3.99E-12	Reject
6	chitest	DECISION_DENSITY	8.600647346	0.003360435	Reject
7	chitest	DESIGN_COMPLEXITY	79.12263725	0.0	Reject
8	chitest	DESIGN_DENSITY	34.74944733	3.75E-09	Reject
9	chitest	ESSENTIAL_COMPLEXITY	28.06969502	1.17E-07	Reject
10	chitest	ESSENTIAL_DENSITY	16.74387781	4.28E-05	Reject
11	chitest	LOC_EXECUTABLE	168.8832824	0.0	Reject
12	chitest	PARAMETER_COUNT	24.43308951	7.69E-07	Reject
13	chitest	HALSTEAD_CONTENT	168.8832824	0.0	Reject
14	chitest	HALSTEAD_DIFFICULTY	223.0268475	0.0	Reject
15	chitest	HALSTEAD_EFFORT	344.000000000	0.0	Reject
16	chitest	HALSTEAD_ERROR_EST	72.72015864	0.0	Reject
17	chitest	HALSTEAD_LENGTH	258.4791338	0.0	Reject
18	chitest	HALSTEAD_LEVEL	5.529967562	0.018693452	Reject
19	chitest	HALSTEAD_PROG_TIME	344	0.0	Reject
20	chitest	HALSTEAD_VOLUME	339.3352255	0.0	Reject
21	chitest	MAINTENANCE_SEVERITY	14.93011985	0.000111567	Reject
22	chitest	MODIFIED_CONDITION_COUNT	41.90661384	9.57E-11	Reject
23	chitest	MULTIPLE_CONDITION_COUNT	76.5597297	0.0	Reject
24	chitest	NODE_COUNT	120.9708727	0.0	Reject
25	chitest	NORMALIZED_CYLOMATIC_COMPLEXITY	6.29287219	0.012122445	Reject
26	chitest	NUM_OPERANDS	183.0874923	0.0	Reject
27	chitest	NUM_OPERATORS	235.7772312	0.0	Reject
28	chitest	NUM_UNIQUE_OPERANDS	165.6003094	0.0	Reject
29	chitest	NUM_UNIQUE_OPERATORS	64.08927855	1.22E-15	Reject

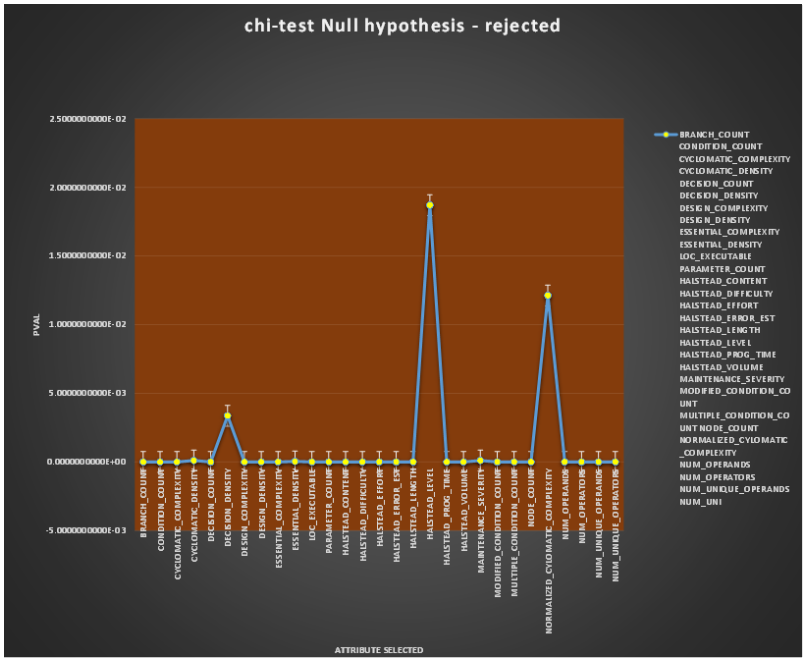

Diagram 5 A Plot of 'Chi-square' Hypothesis test Performance

January - February 2020
ISSN: 0193 - 4120 Page No. 6276 - 6288

7.3 Proposed Machine Learning Hypothesis Space

From the hypothesis testing result, 29 attributes are proved to be relevant to the fault prediction feature set. Therefore, all 29 attributes in Table 1 are considered as the feature set in genetic fault prediction hypothesis space.

$H = \{ h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9, h_{10}, h_{11}, h_{12}, h_{13}, h_{14}, h_{15}, h_{16}, h_{17}, h_{18}, h_{19}, h_{20}, h_{21}, h_{22}, h_{23}, h_{24}, h_{25}, h_{26}, h_{27}, h_{28}, h_{29}\}$

7.4Discussion

The result of performance against the four types of hypothesis testing has been shown in Table 2, 3, 4, 5 and in Diagram 2, 3, 4, and 5. These tables and diagrams show the hypothesis test result of 29 fault attributes of PROMISE data repository. Table 2and Table 3 shows some variance in null hypothesis rejection. Table 4 and 5 shows entire rejection of the null hypothesis.Theacceptance of the null hypothesis is indicated in yellow colour in the table. The performance plots of Hypothesis testing types are shown in Diagrams 2 to 5. When we look at overall results of all the hypothesis testing performance summary, we can conclude that Chi-square hypothesis test provide better result for the formation of fault prediction feature set.

8. CONCLUSION

In this paper, we used four hypothesis testing in order to identify the attributes which are relevant to form fault prediction feature set. Eight public NASA datasets from PROMISE Repository are used in this paper. From the result, we observed that the chi-square test is the best test for the independent hypothesis space formation for genetic fault prediction feature set. In this work from the total of 69 features only 29 are used for the hypothesis testing. The result measured in the tables and the corresponding graphs are plotted. The relationship between every variables are observed. From the performance measurement, there is some acceptance of null hypothesis in F-test and Z-test. T-test shows better result, but which has more variation in the p-value. Chi-square test shows the better relationship of attributes with the population data so as to form independent features.After analyzing the results, we found that chi-squared hypothesis testing have higher statistical value and all the p-value is lesser than 0.05 significance level. This will help to choose better feature set which makes the software development as green engineering and also helps to the rejection of the null hypothesis H_0. Finally, we concluded thatSoftware Development Metrics Report along with organizational post release fault dataset provide better fault prediction feature set and chi-square test is easily used to select the features from the available data set for the formation of safe fault free software engineering. In future, this feature set can be used to design Genetic Fault Prediction Taxonomy.

REFERENCES

1. https://stattrek.com/hypothesis-test/hypothesis-testing.aspx
2. https://www.statisticssolutions.com/academic-solutions/resources/directory-of-statistical-analyses/hypothesis-testing/
3. https://en.wikipedia.org/wiki/Statistical_hypothesis_testing
4. Lehmann, E. L.; Romano, Joseph P. (2005). Testing Statistical Hypotheses (3E ed.). New York: Springer. ISBN 978-0-387-98864-1
5. Losavich, J. L.; Neyman, J.; Scott, E. L.; Wells, M. A. (1971). "Hypothetical explanations of the negative apparent effects of cloud seeding in the Whitetop Experiment". Proceedings of the National Academy of Sciences of the United States of America. 68 (11): 2643–2646. doi:10.1073/pnas.68.11.2643 PMC 389491. PMID 16591951
6. E. L. Lehmann (1997). "Testing Statistical Hypotheses: The Story of a Book". Statistical Science. 12 (1): 48–52. doi:10.1214/ss/1029963261.
7. Siegrist, Kyle. "Hypothesis Testing - Introduction". www.randomservices.org. Retrieved March 8, 2018.
8. Branch, Mark (2014). "Malignant side effects of

January - February 2020
ISSN: 0193 - 4120 Page No. 6276 - 6288

null hypothesis significance testing". Theory & Psychology. 24 (2): 256–277. doi:10.1177/0959354314525282.
9. https://sixsigmastudyguide.com/types-of-hypothesis-test/
10. Hypothesis testing, type I and type II errorsAmitav Banerjee, U. B. Chitnis, S. L. Jadhav, J. S. Bhawalkar, and S. Chaudhury Publication of the Association of Industrial Psychiatry of India 2009 Jul-Dec; 18(2): 127–131.
11. An Introduction to Genetic Algorithms, Jenna Carr, May 2014
12. Feature Selection for Image Steganalysis using Hybrid Genetic Algorithm, Zhihua Xia, Xingming Sun, Jiaohua Qin and ChangmingNiu May 2017 Science Alert
13. Genetic Algorithms and Evolutionary Computation, Adam Marczyk, April 2004, The Talk Origins Archive
14. A Combined Genetic Adaptive Search for Engineering Design, Kalyanmoy Deb and Mayank Goyal (India), Journal of Computer Science and Informatics, 1996, volume 26, page 30-45
15. Genetic Programming James McDermott and Una-May O'Reilly Evolutionary Design and Optimization Group, Computer Science and Articial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. March 2012
16. Application of Genetic Algorithm in Software Testing, Praveen RanjanSrivastava and Tai-hoon Kim, International Journal of Software Engineering and Its Applications Vol. 3, No.4, October 2009
17. A Genetic Algorithm Based Classification Approach for Finding Fault Prone Classes Parvinder S. Sandhu, Satish Kumar Dhiman, AnmolGoyal, World Academy of Science, Engineering and Technology, [18] International Journal of Computer, Electrical, Automation, Control and Information Engineering, Vol:3, No:12, 2009
18. Software Project Planning Using Genetic Algorithm, Deepak Kumar, Assistant Professor, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 1, January – February 2014, ISSN 2278-6856
19. Initial Population for Genetic Algorithms: A Metric Approach, Pedro A. Diaz-Gomez and Dean F. Hougen, School of Computer Science, University of Oklahoma, Norman, Oklahoma, USA
20. Software Testing Using Genetic Algorithms, Akshat Sharma, RishonPatani and Ashish Aggarwal, International Journal of Computer Science & Engineering Survey (IJCSES) Vol.7, No.2, April 2016
21. Test-Data Generation Using Genetic Algorithms, Roy P. Pargas, Mary Jean Harrold, Robert R. Peck, Journal of Software Testing, 1999