International Journal of Science and Engineering Applications
Volume 6 Issue 01, 2017, ISSN-2319-7560 (Online)

A Survey on Different Classification Techniques In Data Mining

Meghana.L[1], N.Deepika[2]
[1]New Horizon College of Engg. Bangalore, India
[2] New Horizon College of Engg. Bangalore, India

Abstract
Data mining is a procedure which finds valuable patterns from large amount of data. Data mining is the examination step of the "knowledge discovery in databases" process, or KDD. The general goal of the data mining process is to retrieve information from a data set and convert it into an explicable structure for further use. There are various data mining methods like classification, frequent patterns, association, clustering. The present paper provides a survey on different classification techniques involved in data mining. The idea of Classification analysis is the organization of data in given classes according to some constraints. The goal of this survey is to provide a comprehensive review on the advantages and disadvantages of the classification techniques.
Keywords: *data mining, data mining techniques, classification, Bayesian, Decision tree, Rule based.*

1. Introduction

Terabytes or petabytes of data are stored in various storage devices. This explosive growth of data in large volume has led to the introduction of data mining. The data mining technique is the process of retrieve the formerly unknown, useable patterns and interesting information from large data set. The extracted information is transformed into an understandable structure for further use. Data mining process use software techniques tools for finding patterns and regularities in sets of data.

2. Data mining techniques

Data mining is one of the tasks in the process of knowledge discovery in database (KDD). Data mining approaches are used to produce the kind of patterns to be created in data mining tasks. The different data mining techniques are as follows:

i) Classification:

Classification analysis is the forming of data in given classes. It is also known as supervised classification. The classification uses given class labels to direct the objects in the data collection. Classification procedure usually use a training set where all objects are already related with known class labels.

ii) Characterization:

Data characterization is a summarization of overall features of objects in a target class, and provides characteristic rules. The data applicable to a user-specified class are normally regained by a database query and run through a summarization module to extract the principle of the data at various levels of abstractions.

iii) Discrimination:

Data discrimination yields the discriminant rules and the technique is fundamentally the evaluation of the general structures of objects between two classes which are, the target class and the contrasting class. The techniques used for data discrimination are very similar to the techniques used for data characterization with the exception that data discrimination results include comparative measures.

iv) Association Analysis:

Association analysis is the detection of association rules. It studies the frequency of items taking place together in transactional databases, and identifies the frequent item sets.

v) Clustering:

Similar to classification, clustering is the formation of data in classes. In clustering, class labels are unidentified and it is a job of the clustering algorithm to discover suitable classes. Clustering technique can also be called as unsupervised classification.

vi) Outlier Analysis:

Outliers are some data rudiments that cannot be gathered in a given class or cluster. They are either known as exceptions or surprises. The outliers are considered noise and are rejected in some domains, but they can disclose important information in other domains. This method can be very substantial and their analysis can be valuable. [8]

3. Classification methods

Classification is a data mining method that allocates items to the target categories or classes in a given collection. The goal of classification is to precisely forecast the target class for each case in the data. Data classification is of two-step process, which involves a) learning step, where a classification model is constructed and b) a classification step, where the constructed model is used to guess the class labels for given data setThe different classification methods are as follows:

i. Decision tree induction
ii. Bayes classification method
iii. Rule-Based classification
iv. Classification by Backpropagation
v. Support Vector Machines
vi. k nearest neighbors method [3]

3.1 Decision tree induction

A decision tree is a tree in which each branch node signifies a choice between a number of substitutes, and each leaf node represents a particular decision. Decision tree are normally used for gaining the information with the purpose of making the decision.
Algorithm:
Generate a decision tree from the training tuples of data Partition D.
Input:
• Data partition, D, which is a set of training tuples and their associated class labels.
• Attribute list, the set of candidate attributes.
• Attribute selection method, a procedure to determine the splitting criterion that "best" partitions the data tuples into individual classes
Output: A decision tree.
Method:
1. Create a node N.
2. If tuples in D are all of the same class, C then
a. Return N as a leaf node labelled with the class C.
3. If attribute list is empty then
4. Return N as a leaf node labelled with the majority class in D.
5. Apply Attribute selection method (D, attribute list) to find the "best" splitting criterion.
6. Label node N with splitting criterion.
7. If splitting attribute is discrete-valued and multi-way splits allowed then
8. Attribute list ß attribute list - splitting attribute.
9. For each outcome j of splitting criterion
a. Let Dj be the set of data tuples in D satisfying outcome j;
b. If Dj is empty then
c. Attach a leaf labelled with the majority class in D to node N;
d. Else attach the node returned by Generate decision tree (Dj, attribute list) to node N;
10. End for
11. Return N;

Attribute selection measure:
Three popular selection measure:
a. Information gain: The information gain measure is used to choose the test attribute at each node in the created tree. The attribute which has the highest information gain is selected as the test attribute for the current node.

$$\text{Info}(D) = -\sum_{i=1}^{n} p_i \log_2 (p_i) \quad (1)$$

Where, Pi is the probability that an indiscriminate tuple in D belongs to Ci and is valued by |Ci, D|/ |D|
$\text{Info}_A(D)$ is the anticipated information essential to organize a tuple from D base on partitioning by A.

$$\text{Info}_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \text{Info}(Dj) \quad (2)$$

Gain is given by $\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D)$
b) Gain ratio: This overwhelms the injustice of Information gain. It gains the information using a split information value by applying a normalization method. The split information value denotes the possible information that is created by splitting the training data set D into v partitions, equivalent to v outcomes on attribute A.

$$\text{SplitInfo}_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2 \frac{|D_j|}{|D|} \quad (3)$$

The gain ratio is defined as:

$$\text{GainRatio}(A) = \text{Gain}(A) / \text{Splitinfo}_A(D) \quad (4)$$

The attribute with the maximum gain ratio is selected as the splitting attribute.

c) Gini index: The Gini Index measures the contamination of a data partition D.

$$Gini(D) = 1 - \sum_{j=1}^{m} pi^2 \quad (5)$$

Where m is the number of classes, pi is the probability that a tuple in D belongs to class Ci.
This is a weighted sum of the impurity of each partition.

$$Gini_A(D) = \frac{|D1|}{|D|} Gini(D1) + \frac{|D2|}{|D|} Gini(D2) \quad (6)$$

The attribute that maximizes the reduction in impurity is chosen as the splitting attribute.[1]

3.2 Bayes classification method

A Bayesian classifier is a classification technique which is used to define or decide if the given tuple belongs to a particular class or not. The classification is based on Bayes' theorem.
Bayes theorem is given by the following:

$$P(H/D) = \frac{P(D/H)\,P(H)}{P(H)} \quad (7)$$

Where, P(H) : Prior probability of hypothesis h
P(D) : Prior probability of training data D
P(H/D) : Probability of H given D
P(D/H) : Probability of D given H

Algorithm:
Method:
1. Let D be set of training tuples
2. Each Tuple is an 'n' dimensional attribute vector like
X : (x1,x2,x3,…. xn) are attributes
3. Let there be 'm' Classes: C1,C2,C3…Cm
4. Naïve Bayes classifier predicts X belongs to Class Ci iff
P (Ci/X) > P(Cj/X) for $1 \le j \le m$, $j \ne i$
5. Compute Maximum Posteriori Hypothesis
P(Ci/X) = P(X/Ci) P(Ci) / P(X)
6. Maximize P(X/Ci) P(Ci) as P(X) is constant
7. With many attributes, it is computationally expensive to evaluate P(X/Ci).To reduce that expensive Naïve Assumption of "class conditional independence" is made.

$$P(X/Ci) = \prod_{k=1}^{n} P(x_k / Ci)$$

$$P(X/Ci) = P(x1/Ci) * P(x2/Ci) *\ldots * P(xn/ Ci)$$

We can easily estimate the probabilities $P(x1/Ci/)$, $P(x2/Ci)$, …. , $P(xn/Ci)$ from the training tuples. And here xk refers to the value of attribute Ak for tuple $\boldsymbol{X}$.
8. To predict the class label of $\boldsymbol{X}$, the equation $P(\boldsymbol{X}/Ci)P(Ci)$ is evaluated for each class Ci .
Therefore Naive Bayes goes out to be excellent in certain applications. Text classification is one area where it really works better. [4]

3.3 Rule Based Classification

A rule-based classifier makes usage of a set of IF-THEN rules for the purpose of classification. These rules are created either using a decision tree or directly from the training data which uses an algorithm called sequential covering algorithm.
An IF-THEN rule is an expression of the form:

IF *condition* THEN *conclusion*

Where, Condition (or LHS) IF part is rule antecedent/precondition and Conclusion (or RHS) THEN part is rule consequent.
Let's consider an example of rule $R1$,
$R1$: IF *age = youth* AND *student = yes* THEN *buys_computer = yes*.

$R1$ can also be written as
R1: ((age = youth) ∧ (student = yes)) → (buys_computer = yes)

If the condition or all the attribute tests in a rule antecedent holds true for a given tuple, then we can say that the rule antecedent is satisfied and that the rule covers the tuple.

A rule R can be measured by coverage and accuracy.
Coverage of a rule: The percentage of instances that satisfy the antecedent of a Rule.

$$Coverage(R) = \frac{n_{covers}}{|D|} \quad (8)$$

Accuracy of a rule: The percentage of instances that satisfy both the antecedent and consequent of a rule.

$$Accuracy(R) = \frac{n_{correct}}{n_{covers}} \quad (9)$$

3.3.1 Rule Induction Using a Sequential Covering Algorithm

Algorithm for sequential covering

Input: *D*, a data set of class-labeled tuples;

Att_vals, the set of all attributes and their possible values.
Output**:** A set of IF-THEN rules.
Method:
1. Rule set = { } // an initial set of rules is empty
2. for each class *c* do
3. repeat
4. Rule = Learn_One_Rule(*D*, *Att_vals*, *c*);
5. remove tuples covered by Rule from *D*;
6. Rule_set = Rule_set + Rule; // add new rule to rule set
until terminating condition;
7. endfor
8. return Rule_set ;[7]

3.4 Classification by Backpropagation

Backpropagation is a neural network learning algorithm. Neural network learning is also referred to as connectionist learning due to the connections between units.

Neural Network

Backpropagation pick up by iteratively giving out a data set of training tuples, relating the network's expectation for each tuple with the actual known goal value. The goal value may be the identified class label of the training tuple (for classification problems) or a continuous value (for numeric prediction). For each training tuple, the heaviness are different so as to minimize the mean-squared error between the network's prediction and the actual goal value. These modifications are made in the "backwards" direction (i.e., from the output layer) through each hidden layer down to the first hidden layer.

Algorithm:Backpropagation Neural network learning for classification or numeric prediction, using the backpropagation algorithm.
Input:
D, a data set consisting of the training tuples and their associated target values;
l, the learning rate; *network*, a multilayer feed-forward network.
Output: A trained neural network.
Method:
1. Initialize all weights and biases in *network*;
2. while terminating condition is not satisfied {
3. for each training tuple *X* in *D* {
// Propagate the inputs forward:
4. for each input layer unit *j* {
5. $O_j = I_j$; // output of an input unit is its actual input value
6. for each hidden or output layer unit *j* {
7. $I_j = \sum_i w_{ij} O_i + \theta_j$; //compute the net input of unit *j* with respect to the previous layer, *i*
8. $O_j = 1/ (1 + e^{-ij})$; } // compute the output of each unit *j*
// Backpropagate the errors:
9. for each unit *j* in the output layer
10. $Err_j = O_j(1-O_j)(T_j - O_j)$; // compute the error
11. for each unit *j* in the hidden layers, from the last to the first hidden layer
12. $Err_j = O_j(1-O_j)\sum_k Err_k w_{jk}$; // compute the error with respect to the next higher layer, *k*
13. for each weight w_{ij} in *network* {
14. $\Delta w_{ij} = (l) Err_j O_i$; // weight increment
15. $w_{ij} = w_{ij} + \Delta w_{ij}$; } // weight update
16. for each bias θ_j in *network* {
17. $\Delta \theta_j = (l) Err_j$; // bias increment
18. $\theta_j = \theta_j + \Delta \theta_j$; } // bias update
19. }} [6]

3.5 Support Vector Machines

SVM was first introduced by Vapnik and has been very effective method for regression, classification and general pattern recognition. It is well thought-out a good classifier because of its high broad view performance without the need to add a priori knowledge, even when the aspect of the input space is very high.

Method:
1. In this algorithm we plot each data item as a point in n-dimensional space with the value of a particular coordinate.
2. Then we perform classification by finding the hyperplane that differentiate the two classes.
3. Maximize the distances between nearest data points and hyperplane. This distance is called margin.
4. Select a hyperplane that has high margin.

Geometrically, the margin match up to the shortest distance between the closest data points to a point on the hyperplane. Having this geometric definition agree us to travel around how to maximize the margin, so that even though there are an infinite

number of hyperplanes, only a a small number of qualify as the solution to SVM. To make certain that the maximum boundary hyperplanes are actually set up, an SVM classifier efforts to maximize the following function with respect to w and b

$$L_p = \frac{1}{2}\|\vec{W}\| - \sum_{i=1}^{t} \alpha_i y_i (\vec{W} \cdot \vec{X}_i + b) + \sum_{i=1}^{t} \alpha_i \qquad (10)$$

where t is the number of training examples, and αi , i = 1, . . . , t, are positive numbers such that the derivatives of LP with respect to αi are zero. αi are the Lagrange multipliers and LP is called the Lagrangian.

In this equation, the vectors and constant b define the hyperplane. A book learning machine, such as the SVM, can be displayed as a function class based on some parameters α. Different function classes can have various capacity in learning, which is represented by a parameter h known as the VC dimension.

3.6 k-Nearest Neighbor Algorithm for Classification

K-NN classifiers are lazy learners. It does not build models explicitly.

Assume each sample in our data set has n attributes which forms an n-dimensional vector:

x = (x1, x2, . . . , xn).

These n attributes are the liberated variables. Every single sample also has another attribute, denoted by y, whose value be governed by on the other n attributes x. We assume that y is a categoric variable, and there is a scalar function, f, which assigns a class, y = f(x) to every such vectors. Suppose the set of T is organized with its classes like,

x(i), y(i) for i = 1, 2, . . . , T

This set is the training set. The idea in k-Nearest Neighbor methods is to identify k samples in the training set whose independent variables x are similar to u, and to use these k samples to classify this new sample into a class, v.

Method:

1. Compute distance between two points using the below formula.s

Euclidean distance.

$$Dist(\boldsymbol{x_1}, \boldsymbol{x_2}) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2} \qquad (11)$$

2. Determine the class from nearest neighbor list

3. Take the majority vote of class labels among the k-nearest neighbors

4. Weigh the vote according to distance

weight factor, $w = 1/Dist^2$

5. Choose the value of k with respect to the below two aspects.

a) If k is too small, sensitive to noise points

b) If k is too large, neighborhood may include points from other classes

6. Attributes are to be scaled to prevent distance measures from being dominated by one of the attributes. [8]

4. Advantages and Disadvantages of classification methods

Table 1: advantages and disadvantages

Methods	Advantages	Disadvantages
1. Decision tree induction	Rules can be generated and they are easy to interpret and understand. It is scalable for large database	It does not handle continuous data. Handling missing data is difficult.
2. Naive Bays Algorithm	It improves performance by removing the irrelevant Features. Good performance. It has short computational time	It requires a very large number of records to obtain good results.
3. Rule based classifiers	Easy to interpret. Can classify new instances rapidly Can easily handle missing values and numeric attributes	It is inefficiency Computational cost is high It involves complex domains
4.Backpropogation	Involves complex	Difficult to understand the

classification	relationship between input and output	model
5. Support vector machines	Popular in text classification problems	Computationally expensive, thus runs slow.
6. K-Nearest Neighbor Algorithm	Easy to understand and Training is very fast.	It has memory limitation

5. Conclusion

The data mining process provides the useful information from large data set by extraction. This process involves many techniques like classification, clustering, outlier analysis, frequent patterns, association. The present paper provides the explanation of different classification methods. There is a detailed description of each method. The goal of this paper to provide the advantages and disadvantages of each classification methods is also figured out in detail.

References

[1] Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, 2ndedition.
[2] Baik, S. Bala, J. (2004), A Decision Tree Algorithm For Distributed Data Mining.
[3] Classification Algorithms for Data Mining: A Survey: International Journal of Innovations in Engineering and Technology (IJIET) Vol. 1 Issue 2 August 2012, ISSN: 2319 – 1058:
[4] Survey on Classification Techniques Used in Data Mining and their Recent Advancements: International Journal of Science, Engineering and Technology Research, Volume 3, Issue 9, September 2014, ISSN: 2278 – 7798
[5] Classification algorithm in Data mining: An Overview: International Journal of P2P Network Trends and Technology (IJPTT) –
Volume 4 Issue 8- Sep 2013 ISSN:2249-2615 http://www.ijpttjournal.org
[6] Survey on Classification Techniques in Data Mining: International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 2 Issue: 7 1983 – 1986
[7]Ruled based classification method: https://www.tutorialspoint.com/data_mining/dm_rbc.html
[8] Data mining techniques and applications: Bharati M. Ramageri / Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305 ISSN : 0976-5166
[9] Review on Classification Algorithms in Data Mining: International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 5, Issue 1, January 2015)

First Author
Meghana L received B.E degree in Computer science and Engineering from Vivekananda Institute of Technology in 2016. Currently pursuing M. Tech degree in Computer science and Engineering. Her area of interests includes Data Mining, Information and Network Security.

Second Author
N. Deepika Sr. Asst. Professor having 15 years of experience in academics as pursued her M.Tech from JNTU Hyderabad and B.Tech from SVU. She is currently working in NHCE, Dept. of CSE, Bangalore. She has guided many UG and PG students for their projects. Her research areas include clustering technique, data mining, Web mining, and Big Data analysis.

International Journal of Science and Engineering Applications
Volume 6 Issue 01, 2017, ISSN-2319-7560 (Online)

Impact of Cleaner Production and Environmental Management on Sustainable Product Innovation and Performance: A study of Manufacturing Industry of Iran

Mahmood Zohoori
Master of Environmental Management
Putra University of Malaysia
Birjand, Iran

Ali Ghani
Master of Business Administration (MBA)
Industrial Management Institute
Birjand, Iran

Hojat Zakeri
Master of Environmental Pollution
Islamic Azad University,
Science and Research Branch,
Birjand, Iran

Abstract: By review of previous research, cleaner production (CL) and environmental management (EM) can improve sustainable product innovation (SPI) and thereby company's performance (financial/non-financial). This study is trying to measure the impact of cleaner production and environmental management on sustainable product innovation and company's performance while performance includes four perspectives according to the balanced scorecard (BSC) model. The proposed perspectives of BSC consist of financial performance, customer satisfaction, internal process, and learning and growth. 255 data were gathered from 10 large manufacturing companies of Iran. The outcome of regression analysis showed that CP has significant impact on customer satisfaction, financial performance, and learning growth. Besides, EM only affects customer satisfaction and internal process significantly and positively. Except financial performance all perspectives of performance can be affected by SPI.

Keywords: Performance, Balanced Scorecard, Cleaner Production, Environmental Management, Sustainable Product Innovation, Iran

1. INTRODUCTION

According to the conducted research by scholars of current study which is published in this issue, cleaner production and environmental management have high potential which impacts sustainable product innovation and organizational performance. According to BSC's model (Kaplan and Norton, 1997), there are four perspectives to evaluate the performance of a company including, financial, customer satisfaction, internal process, and learning and growth (L&G). Recent scholars (e.g. Hojabri et al., 2013; Manafi and Subramaniam, 2015) also believe that these four perspectives can measure company performance properly.

The Environmental Management (EM) practices and also Cleaner Production (CP) techniques are proper tools which aim for efficiency in process of production, using the inputs and also generating industrial wastes. Such tools could remarkably help to Sustainable Product Innovation (SPI) because of rational utilization of natural resources and also reduction of generated waste. Also, sustainable product innovation comes as an opportunity in order to launch new products on market which can meet the imposed pressures by global society as well as legislation (Severo et al., 2017).

Severo et al. (2017) believe that cleaner production and environmental management can affect SPI and thereby financial performance. However, non-financial performance also can be affected by these factors when non-financial consists of customer satisfaction, learning and growth, and internal process (as it is presented in the research of Kaplan and Norton (1997), Hojabri et al. (2013), and Manafi and Subramaniam (2015)). The reason is that cleaner production and EM include knowledge and experience which can affect learning and growth and also internal process. Obviously, customer satisfaction will be affected by learning and growth and internal process. So, this study aims to examine how cleaner production and EM affect SPI and different perspectives of performance.

2. LITERATURE REVIEW

Innovation could be as some changes in services and products and also the way they are supposed to be produced or being offered to market. Hence, innovation is described as developing and implementing new ideas that would be a collective outcome (Van De Ven, 1986; Van De Ven et al., 1989; Gracia and Calantone, 2002). Here, innovating the sustainable products could be identified as presenting new or remarkably improved goods or services, with considering its specific characteristics or intended usages from previously produced products by companies

(Balachandra and Friar, 1997). It should focus on environmental requirements in order to minimize exploitation of natural resources, energy, water, material, production processes improvements and also those environmental practices which reduce effects on environment and minimizes production of pollutants and waste (Placet et al., 2005; Potts, 2010).

The Cleaner Production (CP) technology means using and developing new techniques, methods, energy sources and materials which reduce production waste and within the life cycle of the product (Dunn and Bush, 2001; Tseng et al., 2009; Yonga et al., 2016). The CP methodologies have been developed by the Industrial Development Organization and also Environment Programmes of the United Nations, hence organizations systematically can minimize their emissions and waste (Kliopova and Staniskis, 2006; Hicks and Dietmar, 2007; Lukena et al., 2016).

The environmental sustainability quest is currently transforming competitive landscape, demands to business modifications in products development, technologies, processes and also business models as well as initiating rational utilization of natural resources, treating them and providing the generated waste properly in order to preserve environment (Severo, 2013).

In context of an organization, the related environmental practices to processes and product development, help to sustainable development, offering organizations more economic stability and environmental sustainability. Thus, companies are able to reduce environmental effects through utilizing the environmental management technologies, including: Effluent Treatment Plant (WWTP), Waste Management System (SGR), Life Cycle Analysis (LCA), Environmental Management System (EMS) ISO 14001, Cleaner Production (CP), Industrial Symbiosis (IS), Industrial Ecology (IE), recycling, etc. (Anton et al., 2004; Chertow, 2007; Darnall et al., 2008).

3. Research Methodology and Developed Hypothesis

According to Severo et al. (2017) and Przychodzen (2015), generally eco-innovations were considered with more return on the equity and assets. Moreover, organizations which present eco-innovations also were remarkably showing more performant levels and encountered with less financial risk and have more available cash flow compared to the conventional organizations. On the other hand, they agree that EM and CP both can impact the financial performance. Additionally, Cheng et al. (2014) mentioned that sustainable innovation has strong impact on business performance in Taiwanese firms.

As mentioned above, CP and EM have high potential to influence SPI and non-financial performance as well. Consequently, we will have following hypotheses:

H1: CP affects SPI significantly and positively

H2: EM affects SPI significantly and positively

H3: CP affects financial performance significantly and positively

H4: EM affects financial performance significantly and positively

H5: CP affects customer satisfaction significantly and positively

H6: EM affects customer satisfaction significantly and positively

H7: CP affects learning and growth significantly and positively

H8: EM affects learning and growth significantly and positively

H9: CP affects internal process significantly and positively

H10: EM affects internal process significantly and positively

H11: SPI affects financial performance significantly and positively

H12: SPI affects customer satisfaction significantly and positively

H13: SPI affects learning and growth significantly and positively

H14: SPI affects internal process significantly and positively

In order to test above hypotheses, 10 large manufacturing companies in Iran were chosen. 300 questionnaires were distributed among managers (top/middle) and expert engineers and employee of R&D departments. At the end 255 usable questionnaires were received. This survey

was carried out on 2016. The next part will show the results of data analyzing by SPSS 22.

4. Data analysis and Discussion

Table 1 demonstrates the demographic information of 255 respondents of this study. According to this table, the frequency of male (141) is greater than frequency of female (114).

Table 1: Demographics

	Frequency	Percentage
Gender		
Male	141	55.3
Female	114	44.7
Age		
Less than 18	0.00	0.00%
19-25	27	10.86%
26-40	122	47.8%
Greater than 45	106	41.5%
Education		
Diploma or below	3	1.1%
Bachelor	88	34.5%
Master or higher	164	64.3%
Salary		
Less than 500 $	0.00	0.00%
501-1000 $	22	8.62%
More than 1001$	233	91.3%

The highest frequency (122) of age refers to 26-45 while the lowest frequency (0.00) is for less than 18. Out of 255 respondents, 164 respondents are working with master or doctorate degree while the frequency of diploma and bachelors totally is equal to 91. Regarding the salaries, 233 of respondents receive more than $1001 monthly while others' salaries are less than $1001.

Table 2 shows the central indicators of mean analysis for each variable.

Table 2: Descriptive Statistics

	N	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
CP	255	3.1310	1.04359	.333	.153	-1.019	.304
EM	255	2.9506	1.02638	.196	.153	-1.169	.304
Financial Performance	255	3.0910	1.03203	.202	.153	-1.262	.304
Learning & Growth	255	3.0541	1.04470	.218	.153	-1.039	.304
SPI	255	3.1710	1.06962	.317	.153	-1.046	.304
Internal Process	255	3.1624	1.06800	.329	.153	-1.029	.304
Customer satisfaction	255	3.1647	1.07316	.324	.153	-1.052	.304
Valid N (listwise)	255						

Table 3: Correlations

		CP	EM	SPI	Financial Performance	Customer satisfactions	L&G	Internal Process
CP	Pearson Correlation	1	.388**	.766**	.362**	.518**	.503**	.200**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.001
EM	Pearson Correlation	.388**	1	.380**	.341**	.227**	.221*	.189*
	Sig. (2-tailed)	.000		.000	.000	.000	.021	.044
SPI	Pearson Correlation	.766**	.380**	1	.405**	.627**	.672**	.230**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000
Financial Performance	Pearson Correlation	.362**	.341**	.405**	1	.139*	.125*	.224**
	Sig. (2-tailed)	.000	.000	.000		.027	.046	.000
Customer satisfaction	Pearson Correlation	.518**	.227**	.627**	.139*	1	.560**	-.092
	Sig. (2-tailed)	.000	.000	.000	.027		.000	.144
L&G	Pearson Correlation	.503**	.221*	.672**	.125*	.560**	1	.050
	Sig. (2-tailed)	.000	.021	.000	.046	.000		.424
Internal Process	Pearson Correlation	.200**	.189*	.230**	.224**	-.092	.050	1
	Sig. (2-tailed)	.001	.044	.000	.000	.144	.424	

**. Correlation is significant at the 0.01 level (2-tailed).

Table 3: Correlations

		CP	EM	SPI	Financial Performance	Customer satisfaction	L&G	Internal Process
CP	Pearson Correlation	1	.388**	.766**	.362**	.518**	.505**	.200**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.001
EM	Pearson Correlation	.388**	1	.380**	.341**	.227**	.221*	.189*
	Sig. (2-tailed)	.000		.000	.000	.000	.021	.044
SPI	Pearson Correlation	.766**	.380**	1	.405**	.627**	.672**	.230**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000
Financial Performance	Pearson Correlation	.362**	.341**	.405**	1	.139*	.125*	.224**
	Sig. (2-tailed)	.000	.000	.000		.027	.046	.000
Customer satisfaction	Pearson Correlation	.518**	.227**	.627**	.139*	1	.360**	-.092
	Sig. (2-tailed)	.000	.000	.000	.027		.000	.144
L&G	Pearson Correlation	.505**	.221*	.672**	.125*	.360**	1	.050
	Sig. (2-tailed)	.000	.021	.000	.046	.000		.424
Internal Process	Pearson Correlation	.200**	.189*	.230**	.224**	-.092	.050	1
	Sig. (2-tailed)	.001	.044	.000	.000	.144	.424	

** Correlation is significant at the 0.01 level (2-tailed).

Referring to the table 3, CP and EM have significant relationships with SPI and all performance's perspectives. The highest estimated correlation value refers to CP and SPI (.766) while the lowest value is (.189) referred to relationship between EM and internal process. Besides, SPI has significant relationship with all variables.

Table 4: Regression Analyses Results

Effects	Hypotheses	Coefficients	R-Square	p-value	Supported
CP→SPI EM→SPI	H1: CP affects SPI significantly and positively H2: EM affects SPI significantly and positively	.212	.655	0.00	Yes
		.184		.001	Yes
CP→FP EM→FP	H3: CP affects financial performance significantly and positively H4: EM affects financial performance significantly and positively	.235	.564	.005	Yes
		.091		.069	No
CP→CS EM→CS	H5: CP affects customer satisfaction significantly and positively H6: EM affects customer satisfaction significantly and positively	.331	.424	0.00	Yes
		.071		.091	No
CP→L&G EM→L&G	H7: CP affects learning and growth significantly and positively H8: EM affects learning and growth significantly and positively	.277	.522	.010	Yes
		.321		.007	Yes
CP→IP EM→IP	H9: CP affects internal process significantly and positively H10: EM affects internal process significantly and positively	.056	.405	.088	No
		.301		0.00	Yes
SPI→FP	H11: SPI affects financial performance significantly and positively	.079	.111	.079	No
SPI→CS	H12: SPI affects customer satisfaction significantly and positively	.461	.604	0.00	Yes
SPI→L&G	H13: SPI affects learning and growth significantly and positively	.275	.701	.006	Yes
SPI→IP	H14: SPI affects internal process significantly and positively	.261	.668	.003	Yes

Based on the developed hypotheses of this study, nine regression analyses were done. Table4 shows the results of hypothesis testing. Except H4, H6, H9, H11, other hypotheses are supported by this study.

As shown in Table 4, cleaner production has significant impacts on customer satisfaction, financial performance and learning and growth while it does not have significant impact on internal process. This result can be considered as a contribution of this study.

As another contribution of this study, EM has significant impacts on learning and growth and internal process.

The outcome of regression analysis showed that except financial performance other perspectives of performance can be affected by SPI significantly and positively. This result is inconsistent with the research conducted by Severo et al. (2017).

5. Conclusion

By review of previous research, cleaner production and environmental management can improve sustainable product innovation and thereby company's performance (financial/non-financial). This study attempted to measure the impact of cleaner production and environmental management on sustainable product innovation and company's performance while performance includes four perspectives according to the balanced scorecard model. The proposed perspectives of BSC consist of financial performance, customer satisfaction, internal process, and learning and growth. 255 data were gathered from 10 large manufacturing companies of Iran. The outcome of regression analysis showed that CP has significant impact on customer satisfaction, financial performance, and learning growth. Besides, EM only affects customer satisfaction and internal process significantly and positively. Except financial performance all perspectives of performance can be affected by SPI.

6. References

Anton, W.R.Q., Deltas, G., Khanna, M., 2004. Incentives for environmental selfregulation and implications for environmental performance. J. Environ. Econ. Manag. 48 (1), 632e654.

Balachandra, R., Friar, J.H., 1997. Factors for Success in R&D Projects and New Product Innovation: a Contextual Framework. IEEE Trans. Eng. Manage. N. J. 44 (3), 276-287.

Cheng, C.C.J., Yang, C-l., Sheu, C., 2014. The link between eco-innovation and business performance: a Taiwanese industry context. J. Clean. Prod. 64, 81-90.

Chertow, M.R., 2007. Unconvering industrial symbiosis. J. Industrial Ecol. New Haven 1 (1), 11-30.

Darnall, N., Jolley, G.J., Handfield, R., 2008. Environmental management systems and green supply chain management: complements for sustainability? Bus. Strategy Environ. 17 (1), 30-45.

Dunn, R.F., Bush, G.E., 2001. Using process integration for cleaner production. J. Clean. Prod. 9, 1-13.

Elkington, J., 1999. Cannibals with Forks. New Society, Canada.

Garcia, R., Calantone, R., 2002. A critical look at technological innovation typology and innovativeness terminology: a literature review. J. Prod. Innovation Manag. 19 (2), 110-132.

Hicks, C., Dietmar, R., 2007. Improving cleaner production through the application of environmental management tools in China. J. Clean. Prod. 15, 395-408.

Hojabri, R., Manafi, M., Eftekhar, F., Ghassemzadeh, H., Sharifi, M. and Kaliannan, M., 2013. Effective methods for health care organizations: An evaluation of excellence models. African Journal of Business Management, 7(27), p.2665.

Kaplan, R.S. and Norton, D.P., 1997. Balanced Scorecard-Strategien erfolgreich umsetzen.

Kliopova, I., Staniskis, J.K., 2006. The evaluation of cleaner production performance in Lithuanian industries. J. Clean. Prod. 14, 1561-1575.

Lukena, R.A., Van Berkelb, R., Leuenbergerc, H., Schwagerb, P., 2016. A 20-year retrospective of the national cleaner production centres programme. J. Clean. Prod. 112, 1165-1174.

Manafi, M. and Subramaniam, I.D., 2015. Balancing performance by human resource management practices. Asian Social Science, 11(10), p.386.

Placet, M., Anderson, R., Fowler, K.M., 2005. Strategies for sustainability: innovation and customization are critical, studies for the cement industry and state of Arizona reveal. ResearchTechnologyManagement 48 (5), 32-41.

Przychodzen, J., Przychodzen, W., 2015. Relationships between eco-innovation and financial performance: evidence from publicly traded companies in Poland and Hungary. J. Clean. Prod. 90, 253-263.

Potts, T., 2010. The natural advantage of regions: linking sustainability, innovation, and regional development in Australia. J. Clean. Prod. 18 (8), 713-725.

Severo, E.A., 2013. Innovation and Environmental Sustainability in the Automotive Companies of the Metal Mechanic Cluster Located in the Serra Gaúcha. Thesis (Doctorate in Business Administration). Postgraduate Programme Doctorate in Business Administration, Pontificate Catholic University of Rio Grande do Sul/University of Caxias do Sul (accessed in April 2016). https://repositorio.ucs.br/jspui/bitstream/11338/676/1/Tese%20Eliana%20Andrea%20Severo.pdf.

Severo, E. A., de Guimarães, J. C. F., & Dorion, E. C. H., 2017. Cleaner production and environmental management as sustainable product innovation antecedents: A survey in Brazilian industries. *Journal of Cleaner Production*, *142*, 87-97.

Tseng, M.L., Lin, Y.H., Chiu, A.S.F., 2009. Fuzzy AHP-based study of cleaner production implementation in Taiwan PWB manufacturer. J. Clean. Prod. 17 (14), 1249-1256.

Van De Ven, A.H., 1986. Central problems in the management of innovation. Manag. Sci. 32 (5), 590-607.

Van De Ven, A.H., Angle, H.L., Poole, M.S., 1989. Research on the Management of Innovation: the Minnesota Studies. Harper & Row, New York.

Van Hoff, B., Lyon, T.P., 2013. Cleaner production in small firms taking part in Mexico's Sustainable Supplier Program. J. Clean. Prod. 41, 270-282.

Van Hoof, B., 2014. Organizational learning in cleaner production among Mexican supply networks. J. Clean. Prod. 64, 115-124.

Yonga, J.Y., Klemesa, J.J., Varbanova, P.S., Huisinghb, D., 2016. Cleaner energy for cleaner production: modelling, simulation, optimisation and waste management. J. Clean. Prod. 111, 1-16.

International Journal of Science and Engineering Applications
Volume 6 Issue 01, 2017, ISSN-2319-7560 (Online)

Relationship between Cleaner Production, Environmental Management, and Sustainable Product on Performance

Mahmood Zohoori
Master of Environmental Management
Putra University of Malaysia
Birjand, Iran

Ali Ghani
Master of Business Administration (MBA)
Industrial Management Institute
Birjand, Iran

Hojat Zakeri
Master of Environmental Pollution
Islamic Azad University,
Science and Research Branch
Birjand,

Abstract: The current study attempts to link cleaner production and environmental management to sustainable product innovation and performance. Extant research only focused on financial performance while this study applied balanced scorecard (BSC) model for this purpose. As shown in the proposed framework of this study, cleaner production and environmental management can affect sustainable product innovation and thereby performance (financial, customer satisfaction, learning and growth, and internal process). It should be mentioned that the focal scope of this study is Iran.

Keywords: Cleaner Production, Environmental Management, Sustainable Product Innovation, Financial Performance, Customer Satisfaction, Learning and Growth, Internal Process, Iran)

1. INTRODUCTION

The global industrialization's intensification, explosion of population, new products development as well as excessive consumption and high population all contribute the economy to develop, however, led to serious ecosystems degradation environmentally. In this regard, the corporate environmentalism development that is a strategic business part, is considered as one of the most remarkable changes that started to take place in various markets in 21th century.

Such actions within environmental context were proactive and began to be considered as innovations inherent in organizational competitive strategy that needed research and development of enduring products (Gelbmann and Hammerl, 2015; Severo et al., 2017).

There are many environmental practices which could be employed as the study of eco-design and product life cycle that has been a globally growing trend in various fields such as architecture, engineering as well as design. It has clue for developing new services, systems and products and also minimizing usage of non-renewable resources to reduce their impacts environmentally (Chou, 2014; Kuo et al., 2016).

The Environmental Management (EM) and Cleaner Production (CP) practices are techniques which result in production process efficiency, employing the inputs and generating industrial waste. Such tools remarkably can help to Sustainable Product Innovation (SPI), because of using the national resources rationally and reducing the generated waste. Thus, sustainable product innovation is assumed as an opportunity for launching new products on market which can meet the imposed pressure by global society and legislation (Severo et al., 2017).

In Iran, during recent years, many new policies have been dispensed regarding the importance of green environment. However, they are not sufficient and there should be more investigation on it. More studies in this field will contribute Iran to grow as same as other developed countries. In addition, different companies attempt to comply themselves with global standards.

The recent research conducted by Sevro et al. (2017) states that cleaner production and environmental management affect sustainable product innovation and thereby financial performance. The current study believes that besides of financial performance, nonfinancial performance can be affected by cleaner production, environmental management and sustainable product innovation. The reason is

that balanced scorecard (BSC) model of Kaplan and Norton (1997). According to this model performance has four perspectives including, financial, customer, learning and growth, and internal process. Besides, there are interactions between theses dimensions. Hence, this study aims to show that how cleaner production and environmental management affect sustainable product innovation and thereby company performance (financial and non-financial).

2. LITERATURE REVIEW

From environmental sustainability perspective, many surveys have been performed to understand the organizational and technical variables related to developing a sustainable product innovation (Elkington, 1999; Horbach, 2008; Lin et al., 2012; Marchi, 2012; Khalili and Duecker, 2013; Boons et al., 2013; Silverstre and Silve Neto, 2014).

Innovation could be as some alterations in services or products and how they are being offered or produced. Therefore, innovation is considered as implementation and development of new ideas that is a collective outcome (Van De Ven, 1986; Van De Ven et al., 1989; Garcia and Calantone, 2002). In this regard, innovating the sustainable products could be identified as introducing new or remarkably improved services and products by considering its characteristics or intended application of products produced by company previously (Balachandra and Friar, 1997).

It has to focus on environmental requirements in order to minimize employing the water, natural resources, materials, energy, improvements in environmental practices and production processes which reduce impacts on environment and minimizes production of pollutants and waste (Placet et al., 2005; Potts, 2010).

It should be considered that companies are able to optimize new product development and production process by using the environmental practices for example Cleaner Production methods (CP) that was presented by United Nations Industrial Development Organizations as well as United Nations Environment Programmes, hence organizations systematically can decrease their emissions and waste (Klipova and Staniskis, 2006; Hicks and Dietmar, 2007; Lukena et al., 2016).

The CP concept means those actions which enable an organization to qualify itself as an effective energy and raw material user within the production process, attempting to improve productivity which leads to improve in competitiveness and finally organizational performance (Severo et al., 2015)/ Here, CP technologies means employing and developing new techniques, methods, energy sources and material which reduce production waste and throughout life cycle of product (Dunn and Bush, 2001; Tseng et al., 2009; Yonga et al., 2016).

One of the integral programs of a company is Cleaner Production which presents interfering working teams with process of production which develops certain relationships with the suppliers and environmentally seek proper use of energy and natural resources, assuming productive requirements and reducing the environmental effects of the business activity. But, it would be a long-run strategy which can result in having competitive advantage over other rivals (Tseng et al., 2009; Bhupendraa and Sangleb, 2016; Lukena et al., 2016).

The CP approach has been employed in various segments and different sizes of organizations for embracing the production process efficiency that could help to having sustainable innovation (Boons et al., 2013; Silvestre and Silva Neto, 2014). CP should impact development process of new products via minimizing the usage of resources as well waste emission (Boons et al., 2013; Sivestre and Silva Neto, 2014; Kuo et al., 2016).

In a research conducted by Geng et al. (2010) it was revealed that increasing environmental issues in Chin made the regional Chinese governments to look for and promote CP such as coordination and measures of different stakeholders through offering financial

support and through suggesting proper policies and performing training programs.

A study conducted by Van Hoff and Lyon (2013) on small and medium sized (SMEs) of Mexico, asserted that using the CP methods in emerging countries could help to having remarkable environmental and economic benefits across supply chains globally.

In addition, Massote and Santi (2013) explained that CP implementation is considered as an effective approach for obtaining eco-efficiency. Here, CP methods present viable alternatives for the companies via the implementation in process of production that contributes to reduction of waste, atmospheric emissions and wastewater as well as offering proper application of energy and raw materials and water consumption rationalization, so providing businesses environmental and economic benefits (Zeng et al., 2010; Ortolano et al., 2014; Van Hoof, 2014).

In 2013, Severo explained that metal-mechanic cluster firms in Southern Brazil already have presented many product innovations in market by means of CP techniques. From total 333 studied companied, 25.8% of them applied CP techniques and occasioned benefits by CP might prevail for sustainable development of new services and products.

2.1. Environmental management and sustainable product innovation

The competitive landscape is transforming due to environmental sustainability quest as well as demand to changes in business to develop processes, products, business models and technologies and also rationally using the natural resources, treating and providing appropriately the generated waste in order to preserve the surrounding environment (Severo, 2013).

Within the organizational context, related environmental practices to processes and product development help to the sustainable development, offering the organizations more economic stability and environmental sustainability. Hence, companies can reduce environmental impacts through using the environmental management methods including: Effluent Treatment Plant (WWTP), Waste Management System (SGR), Environmental Management System (EMS) ISO 14001, Cleaner Production (CP), Life Cycle Analysis (LCA), Industrial Symbiosis (IS), Industrial Ecology (EI), Recycling, etc. (Anton et al., 2004; Sharma and Henriques, 2005; Marshall et al., 2005; Chertow, 2007; Darnall et al., 2008).

Regarding the related aspects to management, Marshall et al. (2005) asserted that Environmental Management process employs regulations and rules which impact proactive environmentalism.

In an organization, environmental management requires more effort in order to manage supplier and internal processes and also big initial investment due to it leads to improved performance financially (Darnall et al., 2008; Fenga et al., 2016).

In addition, previous studies demonstrated that organizations understand the necessity of environmental management not just for reducing the environmental impacts but also to achieve competitive advantage (York et al., 2003; Da Rosa et al., 2015). The methodologies for both organizational changes and environmental management increase environmental innovation. They form a type of innovation which helps to obtain sustainable development (Severo et al., 2017). In 2013, Youn et al studied 141 South Korean organizations that contextualize employing the environmental practices in order to promote improvement of services and products globally.

The other research performed by Jabbour et al., (2012) on total 75 Brazilian automotive industry firms, asserted that environmental management has positive impact on business operations performance, emphasizing on auto parts segment and also automotive components. A sustainable and innovative

company simultaneously attempts to be effective economically, considering the environmental support capacity and that would not change resources availability in its ecosystems (Agyeman and Evans, 2004; Anton et al., 2004).

Thus, the probability to have environmental concerns in business innovation via proper practices environmentally could become important to keep future businesses in future. So, innovations attempt to improve production procedures, lower energy consumptions and lower costs will be critical for keeping the environmental sustainability (Nidumolu et al., 2009).

2.2. Sustainable product innovation and financial performance

Innovating sustainable products as well as financial performance is the main focus of many studies, specifically in contemporary organizations. As noted by Paladino (2007), financial performance concept emphasizes on profitability and quality of products and services to provide a certain return on investment (ROI), and also minimizing the operation costs and generating total performance and competitiveness. According to Przychodzen and Przychodzen (2015), generally, eco-innovations were known as "higher" return on quality and assets. Moreover, those firms which present eco-innovations also were remarkably more performant and probably encounter with less financial risks and will have more "available" cash flow compared to conventional organizations.

An investigation has been conducted by Marchi (2012) in 6047 Spanish organizations and it was revealed that introducing the product innovations positively impact environment, through creating environmental innovations and through cooperation of research and development. But, related to environmental challenges, product innovation should be planned appropriately and should consider environmental issues for creating competitiveness. However, probability of failure in those projects which involve innovation could reach unfavorable proportions (Barbieri et al., 2010).

In addition, Cheng et al. (2014) asserted that sustainable innovation has strong impact on business performance in Taiwanese organizations. But, the findings demonstrated that managers should know about the positive relationship between sustainable innovation and the organizational performance.

3. PROPOSED FRAMEWORK

All above discussion there are reasonable relationships between cleaner production and sustainable product innovation and thereby financial performance. It is also true for environmental management. However, it seems that cleaner production and environmental management can affect non-financial performance. By considering BSC's perspectives we can improve the framework of Severo et al. (2017).

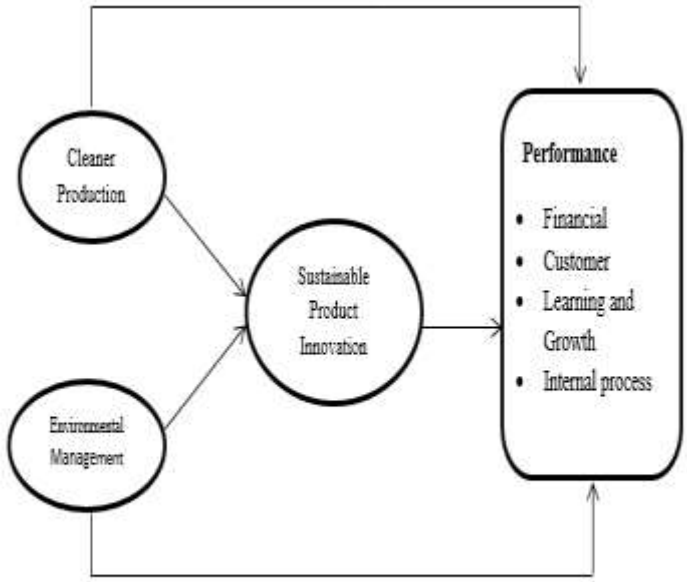

Figure 1: Proposed Framework of this study

As shown in Figure 1, cleaner production and environmental management can affect various perspectives of performance directly. Furthermore, current study believes that cleaner production and environmental management should be connected to each dimension of performance. There are several research conducted by Manafi and

subramaniam (2015 a, b) which are concentrated on the vital role of knowledge in increasing innovation. They asserted that innovation can increase economic growth. As a result, the environmental management and also cleaner production include knowledge can be applied for innovation and thereby higher performance.

4. CONCLUSION

Extant research showed that cleaner production and environmental management have potential to affect sustainable product innovation. Besides, the impact of sustainable product innovation on financial performance was concentrated by Severo et al. (2017). Since according to BSC model there are four perspectives for performance, we could develop the proposed framework of Severo et al. (2017).

As a suggestion for future study, the proposed framework of this study (Figure 1) can be tested in different industries of Iran to highlight which dimensions of performance are being affected by the independent variable of this study. The results will be useful for policy makers in government as well. Besides, the proposed framework of this study has high potential to be applied for various industries in various countries.

5. REFERENCES

Agyeman, J., Evans, B., 2004. Just sustainability: the emerging discourse of environmental justice in Britain? Geogr. J. 170 (2), 155-164.

Anton, W.R.Q., Deltas, G., Khanna, M., 2004. Incentives for environmental selfregulation and implications for environmental performance. J. Environ. Econ. Manag. 48 (1), 632e654.

Balachandra, R., Friar, J.H., 1997. Factors for Success in R&D Projects and New Product Innovation: a Contextual Framework. IEEE Trans. Eng. Manage. N. J. 44 (3), 276-287.

Barbieri, J.C., Vasconcelos, I.F.G., Andreassi, T., Vasconcelos, F.V., 2010. Innovation and sustainability: new models and propositions. RAE 50 (2), 146-154.

Bhupendraa, K.V., Sangleb, S., 2016. Strategy to derive benefits of radical cleaner production, products and technologies: a study of Indian firms. J. Clean. Prod.126, 236-247.

Boons, F., Montalvo, C., Quist, J., Wagner, M., 2013. Sustainable innovation, business models and economic performance: an overview. J. Clean. Prod. 45, 1-8.

Cheng, C.C.J., Yang, C-l., Sheu, C., 2014. The link between eco-innovation and business performance: a Taiwanese industry context. J. Clean. Prod. 64, 81-90.

Chertow, M.R., 2007. Unconvering industrial symbiosis. J. Industrial Ecol. New Haven 1 (1), 11-30.

Chou, J.-R., 2014. An ARIZ-based life cycle engineering model for eco-design. J. Clean. Prod. 66, 210-223.

Darnall, N., Jolley, G.J., Handfield, R., 2008. Environmental management systems and green supply chain management: complements for sustainability? Bus. Strategy Environ. 17 (1), 30-45.

Da Rosa, F.S., Guesser, T., Hein, N., Pfitsche, E.D., Lunkes, R.J., 2015. Environmental impact management of Brazilian companies: analyzing factors that influence disclosure of waste, emissions, effluents, and other impacts. J. Clean. Prod. 96, 148-160.

Dunn, R.F., Bush, G.E., 2001. Using process integration for cleaner production. J. Clean. Prod. 9, 1-13.

Elkington, J., 1999. Cannibals with Forks. New Society, Canada.

Fenga, T., Caib, D.,Wangc, D., Zhanga, X., 2016. Environmental management systems and financial performance: the joint effect of switching cost and competitive intensity. J. Clean. Prod. 113, 781-791.

Garcia, R., Calantone, R., 2002. A critical look at technological innovation typology and innovativeness terminology: a literature review. J. Prod. Innovation Manag. 19 (2), 110-132.

Gelbmann, U., Hammerl, B., 2015. Integrative re-use systems as innovative business models for devising sustainable producteservice-systems. J. Clean. Prod. 97, 50-60.

Geng, Y., Xinbei, W., Qinghua, Z., Hengxin, Z., 2010. Regional initiatives on promoting cleaner production in China: a case of Liaoning. J. Clean. Prod. 18,1502-1508.

Hicks, C., Dietmar, R., 2007. Improving cleaner production through the application of environmental management tools in China. J. Clean. Prod. 15, 395-408.

Horbach, J., 2008. Determinants of environmental Innovation e new evidence from German panel data sources. Res. Policy 37, 163e173.

Kaplan, R.S. and Norton, D.P., 1997. Balanced Scorecard-Strategien erfolgreich umsetzen.

Khalili, N.R., Duecker, S., 2013. Application of multi-criteria decision analysis in design of sustainable environmental management system framework. J. Clean. Prod. 47, 188-198.

Kliopova, I., Staniskis, J.K., 2006. The evaluation of cleaner production performance in Lithuanian industries. J. Clean. Prod. 14, 1561-1575.

Kuo, T.-C., Smith, S., Smith, G.C., Huang, S.H., 2016. A predictive product attribute driven eco-design process using depth-first search. J. Clean. Prod. 112 (4), 3201-3210.

Lin, R.J., Tan, K.H., Geng, Y., 2012. Market demand, green product innovation, and firm performance: evidence from Vietnam motorcycle industry. J. Clean. Prod. 30, 1-7.

Lukena, R.A., Van Berkelb, R., Leuenbergerc, H., Schwagerb, P., 2016. A 20-year retrospective of the national cleaner production centres programme. J. Clean. Prod. 112, 1165-1174.

Manafi, M. and Subramaniam, I.D., 2015a. The Role of the Perceived Justice in the Relationship between Human Resource Management Practices and Knowledge Sharing: A Study of Malaysian Universities Lecturers. Asian Social Science, 11(12), p.131.

Manafi, M. and Subramaniam, I.D., 2015b. Relationship between Human Resources Management Practices, Transformational Leadership, and Knowledge Sharing on Innovation in Iranian Electronic Industry. Asian Social Science, 11(10), p.358.

Marchi, V., 2012. Environmental innovation and R&D cooperation: empirical evidence from Spanish manufacturing firms. Res. Policy 41, 614-623.

Marshall, R.S., Cordano, M., Silverman, M., 2005. Exploring individual and institutional drivers of proactive environmentalism in the US wine industry. Bus. Strategy Environ. 14 (2), 92-109.

Massote, C.H.R., Santi, A.M.M., 2013. Implementation of a cleaner production program in a Brazilian wooden furniture factory. J. Clean. Prod. 46, 89-97.

Nidumolu, R., Prahalad, C.K., Rangaswami, M.R., 2009. Why sustainability is now the key driver of innovation? Harv. Bus. Rev. 87 (9), 57-64.

Ortolano, L., Sanchez-Triana, E., Afzal, J., Laiq Ali, C., Rebellon, S.A., 2014. Cleaner production in Pakistan's leather and textile sectors. J. Clean. Prod. 68, 121-129.

Placet, M., Anderson, R., Fowler, K.M., 2005. Strategies for sustainability: innovation and customization are critical, studies for the cement industry and state of Arizona reveal. ResearchTechnologyManagement 48 (5), 32-41.

Paladino, A., 2007. Investigating the drivers of innovation and new product success: a comparison of strategic orientations. J. Prod. Innovation Manag. 24, 534-553.

Przychodzen, J., Przychodzen, W., 2015. Relationships between eco-innovation and financial performance: evidence from publicly traded companies in Poland and Hungary. J. Clean. Prod. 90, 253-263.

Potts, T., 2010. The natural advantage of regions: linking sustainability, innovation, and regional development in Australia. J. Clean. Prod. 18 (8), 713-725.

Severo, E.A., 2013. Innovation and Environmental Sustainability in the Automotive Companies of the Metal Mechanic Cluster Located in the Serra Gaúcha. Thesis (Doctorate in Business Administration). Postgraduate Programme Doctorate in Business Administration, Pontificate Catholic University of Rio Grande do Sul/University of Caxias do Sul (accessed in April 2016). https://repositorio.ucs.br/jspui/bitstream/11338/676/1/Tese%20Eliana%20Andrea%20Severo.pdf

Severo, E.A., Guimaraes, J.C.F., Dorion, E.C.H., Nodari, C.H., 2015. Cleaner production, environmental sustainability and organizational performance: an empirical study in the Brazilian metal-mechanic industry. J. Clean. Prod. 96, 118-125.

Severo, E. A., de Guimarães, J. C. F., & Dorion, E. C. H. (2017). Cleaner production and environmental management as sustainable product innovation antecedents: A survey in Brazilian industries. *Journal of Cleaner Production*, *142*, 87-97.

Sharma, S., Henriques, I., 2005. Stakeholder influences on sustainability practices in the Canadian forest products industry. Strategic Manag. J. 26 (2), 159-180.

Silvestre, B.S., Silva Neto, R., 2014. Are cleaner production innovations the solution for small mining operations in poor regions? the case of Padua in Brazil. J. Clean. Prod. 84, 809-817.

Tseng, M.L., Lin, Y.H., Chiu, A.S.F., 2009. Fuzzy AHP-based study of cleaner production implementation in Taiwan PWB manufacturer. J. Clean. Prod. 17 (14), 1249-1256.

Van De Ven, A.H., 1986. Central problems in the management of innovation. Manag. Sci. 32 (5), 590-607.

Van De Ven, A.H., Angle, H.L., Poole, M.S., 1989. Research on the Management of Innovation: the Minnesota Studies. Harper & Row, New York.

Van Hoff, B., Lyon, T.P., 2013. Cleaner production in small firms taking part in Mexico's Sustainable Supplier Program. J. Clean. Prod. 41, 270-282.

Van Hoof, B., 2014. Organizational learning in cleaner production among Mexican supply networks. J. Clean. Prod. 64, 115-124.

Yonga, J.Y., Klemesa, J.J., Varbanova, P.S., Huisinghb, D., 2016. Cleaner energy for cleaner production: modelling, simulation, optimisation and waste management. J. Clean. Prod. 111, 1-16.

York, R., Rosa, E.A., Dietz, T., 2003. Footprints on the earth: the environmental consequences of modernity. Am. Sociol. Rev. 68 (2), 279-300.

Zeng, S.X., Meng, X.H., Yin, H.T., Tamb, C.M., Sun, L., 2010. Impact of cleaner production on business performance. J. Clean. Prod. 18, 975-983.

International Journal of Science and Engineering Applications
Volume 6 Issue 01, 2017, ISSN-2319-7560 (Online)

Impact of brand equity on customer purchasing behavior: A study of clothing industry of Iran

Shekofeh Sadeghi
Islamic Azad University of Gheshm, Iran

Mohammad Ail Abbasi
Islamic Azad University of Abarkooh, Iran

Somaye Tabatabaei
Islamic Azad University of Firuzabad, Iran

Abstract: This study as a quantitative research attempts to postulate that brand equity has high potential to improve customer purchasing behavior in clothing industry of Iran. By literature review, brand equity has four components including, brand association, brand loyalty, perceived quality, and brand awareness. 400 data was gathered and analyzed. The results showed that the highest impact refers to perceived quality while the lowest impact refers to brand awareness.

Keywords: Brand Equity, Brand Association, Brand Loyalty, Perceived Quality, Purchasing Behavior, Clothing Industry, Iran

1. INTRODUCTION

Clothing industry is one of those industries that can grow in all of the countries across the world. Iran, also is one of those countries that has a variety of product ranges in this industry. Since, clothing is one of the high selling industries, so its growth is very critical in order to contribute country's economic growth through appropriate selling of products. It's been a while since Iranian products could not be successful to attract domestic customers. On the other hand, importing of Chinese products and also well-known western brands resulted in changes in customer's purchase behavior. Brand equity is one of the influential factors on purchase behavior (Aghaei, 2014; Kladou, S., & Kehagias, 2013; Aaker, 1991). According to literature review, brand equity constitutes of perceived quality, brand loyalty, brand awareness and brand association. The fact that how brand equity can impact purchasing behavior in clothing industry of Iran, is the main idea of current research. *Hence, this study aims to find how brand equity' components affects customer purchasing behavior in clothing industry of Iran.*

2. LITERATURE REVIEW

Many conducted studies previously, focused on brand equity notion and all of its relevant definitions. Some of the famous researches performed before, tried to discuss "brand equity" concept as a set of assets and liabilities relevant to a specific brand for example name and symbol which can impose beneficial or detrimental effects on arising values from products and services (Yasin et al., 2007; Aakar, 1991). Moreover, Keller (1999) mentioned that brand equity is able to signify many unique marketing effects on a specific brand. Thus, by considering positive brand equity features, it can occur while customers are all willing to pay higher amounts for an interesting name disclosed to products or services which has the same quality level in comparison to other products (Bello and Hobrook, 1995). However, brand equity might be damaged while it is not effectively managed. For instance, poor product quality or customer service can have negative impacts on brand image that might lead to decrease in sales volume.

In addition, Murphy (1998), presented one of the critical examples of brand as a kind of equity and provided laws to support intellectual property. His study led to some elaborate discussions and in those countries that legal system has been developed properly, the values of a brand have been considered for customers and producers both. In order to deal with piracy, many countries set up laws to protect trademarks, patents, designs and copyrights. In addition, brand is assumed as a product that is tradable and could be evaluated through financial values (Boyle, 2003).

As shown in research conducted by (Aakar, 1991), becoming familiar with a certain brand and past purchase experiences can both impact

customers' perception. Besides, Aakar (1991) asserted that brand equity can generate high margins by reduction in reliance and also premium pricing according to promotional activities. In addition, Aakar (1991) noted that by means of a positive image, customers will not focus on promotions in short-term anymore and instead assume the brand as a whole. He mentioned that brand equity is considered as a general concept and has to be subdivided later as four specific areas including brand loyalty, brand awareness, perceived quality and brand associations.

2.1. Brand Awareness

We can assume it as one of the significant attributes due to it is considered as one of the important features of brand which influences attention. There are various ideas from different scholars relevant to capability of potential customers to recall and arrange brand and making it relevant to its related product class (Aakar, 1991; Kim & Ko, 2012). The brand awareness level is at some point linked to brand recognition and also the first brand with name in highest level possible.

Regarding the potential customers, it is necessary to know about products, thus it is one of the choices to purchase. Maybe there should be a certain level of awareness for a product before penetrating a set of considerations (Homburg et al., 2010) and rising the brand awareness could be conductive with high chance of entering to the second set (Nedungadi, 1990). Here, those brands which have more levels of awareness probably would be purchased more (Kim and Ko, 2012; Yasin et al., 2007). In addition, it will explain why customers prefer to purchase the well-known brands instead of less famous brands (Macdonald and Sharp, 2000; Hoyer and Brown, 1990).

Based on a study conducted by Keller (1999), brand awareness could be increased through a lot of repeated exposures to a brand. In order to have brand awareness there are two tasks to be done including improving the identity of brand name and relating it with the class of product. Celebrity endorsement and advertisement could be considered as a helpful tool to increase brand awareness. In a lot of studies, it was identified that advertisement is an influential device (Mackenzie et al, 1986; Tsai et al, 2007; Aghaei et al, 2014).

By above discussion, the first hypothesis of this study is formulated as follow:

H1: Brand awareness affects purchasing behavior significantly and positively

2.2. Perceived Quality

A lot of previously performed studies, examined perceived quality in various fields and a lot of emphasize was on many important attributes such as perceived quality. It is known as customer perception regarding superiority of products and total service quality as well (Aakar, 1991; Keller. 1999; Yasin et al., 2007).

The other researches tried to assume it through intangibles and focused on total emotions for brand as well as discussing that it is naturally subjective and noted that product specifications could have less relevant correlation to the perceived quality. The brand's perceived quality will contribute to generate value via reasoning to buy a certain product and contributing to make some differentiation for brand positioning, offering premium price strategies and motivating the channel members to operate effectively as well as expanding the extensions into new categories of brand (Aakar, 1991). Moreover, it was identified that perceived quality is a significantly important factor to define brand loyalty and repeated purchases both (Delong et al, 2004). Besides, it was understood that obtaining a satisfactory level of the perceived quality is not easy due to there are so many fast and continuous advancements in products which can improve customer expectations about quality of a product (Park et al., 2002). However, the second hypothesis of this study is as follow:

H2: Perceived quality affects purchasing behavior significantly and positively

2.3. Brand Loyalty

The other analyzed topic in previous studies was the brand loyalty concept due to it is

assumed as a main aspect of brand equity. Based on a study conducted by Atilgan et al (2005) it was identified that brand loyalty has both positive and negative impact on brand equity regarding certain products as well as it was shown that how it can have direct and indirect impact on brand loyalty. In case of influencing the brand loyalty, a lot of studies demonstrated that customers continue to purchase a certain brand without emphasizing on superior aspects, convenience or price of competitors (Aakar, 1991).

Also (Aakar, 1991), explained that probably loyal customers will more repeat the purchase thus it has less cost for them to keep their customers compared to attracting the potential ones. Secondly, brand loyalty can increase trade leverages. Many customers who have high levels of emotion toward a specific brand will switch to shops that sell that certain brand. Thirdly, loyal customers can influence purchase behavior of other customers to purchase that brand. So, the third hypothesis of this study is developed as follow:

H3: Brand loyalty affects purchasing behavior significantly and positively

2.4. Brand Association

The brand association concept inspired scholars to perform many studies on different roles of brand association. Although there are a lot of consensus parts regarding the definition of relationship between brand and memory in general (Aakar, 1991; Keller, 1999; Yasin et al, 2007), such types of researches revealed that brand equity is dependent on association of customers to brands that will lead to building a certain brand image.

These two experts asserted that brand association is a remarkably complex concept which is related to different ideas, examples, facts and also episodes which create a network of brand knowledge (Yoo et al., 2002). Moreover, regarding the tangible products, all of the qualities which intangible such as distinctiveness and innovativeness also are considered as brand associations.

In addition, Keller (1999) classified brand associations as attitudes, benefits and attributes. The attributes mean those specific product features. They could be categorized as product non-product and product related attributes. Regarding the product related attributes, all of the service or product features should be considered. In case of non-related attributes, packaging, price information, usage imagery and user imagery have to be considered.

Brand association results in value creation for both organization and the customers.

Extant research conducted by Wong (2013), Kim and Ko (2012), and Griffin et al. (2012) emphasized on the impact of brand association on customer buying behavior. So, the fourth hypothesis will be as follow:

H4: *Brand association affects purchasing behavior significantly and positively*

3. METHODOLOGY

This research applied quantitative approach (hypothesis testing). For this purpose, 400 questionnaires were distributed among customers. This research conducted on November of 2016. The scope of this study was Iranian clothing industry, so the designed questionnaires were distributed in 10 large shopping mall located in 4 largest cities of Iran (Tehran, Shiraz, Isfahan, and Mashhad). The respondents were chosen randomly from people who had come for shopping. It should be mentioned that the questionnaire were developed based on the research of Aaker (1991) and Aghaei et al. (2014). To analyze gathered data, Pearson Correlation test, and regression analysis were carried out by SPSS 22. It should be mentioned that to measure each variable 5-point Likert scale were applied. This scale was sorted from strongly disagree (1) to strongly agree (5).

4. RESULTS AND DISCUSSION

The first analysis considers demographic factors. Table 1 demonstrates results achieved by demographic analysis.

Table 1: Demographic

	Frequency	Percentage
Gender		
Male	212	53%
Female	188	47%
Age		
Less than 18	78	19.5%
19-25	155	38.75%
26-40	101	25.25%
Greater than 45	66	16.5%
Education		
Diploma or below	148	37%
Bachelor	191	47.75%
Master or higher	61	15.25%
Salary		
Less than 500 $	203	50.75%
501-1000 $	121	30.25%
More than 1001$	76	19%

Referring to the table 1, the frequency of male is greater than female. The highest frequency of age refers to the third group of respondents (155) while the lowest frequency refers to the last group (66). The maximum frequency of education refers to the respondents who have bachelor degree. Out of 400 respondents, 203 customers have salary below 500 $ per month while 121 of respondents receive salary between 501 and 1000$ monthly. The lowest frequency of salary refers to people who receive more than 1001$ per month.

The next analysis refers to mean analysis. Table 2 demonstrates the results of mean analysis.

Table 2: Descriptive Statistics

	N	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Perceived quality	400	3.1125	1.03437	.344	.122	-.965	.243
Brand Loyalty	400	2.8905	1.01868	.159	.122	-1.262	.243
Brand Awareness	400	3.0125	1.01781	.251	.122	-1.291	.243
Brand Association	400	3.0390	1.04182	.103	.122	-1.103	.243
Purchasing behavior	400	3.1965	1.07040	.287	.122	-1.055	.243
Valid N (listwise)	400						

According to the table 2, the highest mean (3.1965) refers to purchasing behavior while the lowest estimated value for mean (2.8905) refers to brand loyalty. All of the estimated values for skewness are positive, so for each variable the pick of frequency chart is located to the left direction. The negative values of kurtosis show that all of the variables have flat distribution.

Table 3: Correlations

		PQ	BL	BA	BAss	PB
PQ	Pearson Correlation	1	.254**	.188**	.393**	.646**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	400	400	400	400	400
BL	Pearson Correlation	.254**	1	.125*	.089	.275**
	Sig. (2-tailed)	.000		.012	.076	.000
	N	400	400	400	400	400
BA	Pearson Correlation	.188**	.125*	1	.038	.249**
	Sig. (2-tailed)	.000	.012		.444	.000
	N	400	400	400	400	400
B Ass	Pearson Correlation	.393**	.089	.038	1	.441**
	Sig. (2-tailed)	.000	.076	.444		.000
	N	400	400	400	400	400
PB	Pearson Correlation	.646**	.275**	.249**	.441**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	400	400	400	400	400

**. Correlation is significant at the 0.01 level (2-tailed).

Table 3: Correlations

		PQ	BL	BA	BAss	PB
PQ	Pearson Correlation	1	.254**	.188**	.393**	.646**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	400	400	400	400	400
BL	Pearson Correlation	.254**	1	.125*	.089	.275**
	Sig. (2-tailed)	.000		.012	.076	.000
	N	400	400	400	400	400
BA	Pearson Correlation	.188**	.125*	1	.038	.249**
	Sig. (2-tailed)	.000	.012		.444	.000
	N	400	400	400	400	400
B Ass	Pearson Correlation	.393**	.089	.038	1	.441**
	Sig. (2-tailed)	.000	.076	.444		.000
	N	400	400	400	400	400
PB	Pearson Correlation	.646**	.275**	.249**	.441**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	400	400	400	400	400

*. Correlation is significant at the 0.05 level (2-tailed).

As shown in table 3, all of the independent variables have significant relationships with purchasing behavior. The highest estimated value (.646) refers to perceived quality while the lowest value (.188) refers to brand awareness.

Table 4: Regression Analysis

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.700[a]	.489	.484	.76869

a. Predictors: (Constant), BAss, BA, BL, PQ

ANOVA[b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	223.758	4	55.940	94.672	.000[a]
	Residual	233.397	395	.591		
	Total	457.155	399			

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	.108	.188		.573	.567		
	PQ	.521	.042	.503	12.333	.000	.776	1.289
	BL	.116	.039	.110	2.958	.003	.929	1.076
	BA	.139	.039	.132	3.589	.000	.957	1.045
	BAss	.235	.040	.229	5.847	.000	.844	1.184

Referring to table 4, R square is equal to .489 which means that 48.9% of variation in purchasing behavior can be accounted by brand equity's components. As shown in ANOVA table, the estimated p-value is zero, so at least one of the independent variables has significant impact on purchasing behavior. Besides, there is no multi co linearity among independent variables because all VIF are less than 5.

We are 95% confident that brand awareness has significant effect on customer purchasing behaviors in clothing industry of Iran because p-value is equal to zero. The estimated coefficient is equal to .139 which asserts that for every unit increase in brand awareness, purchasing behavior will go up .139 units. Consequently, the first hypothesis of this study (H1) is supported by the results.

We are 95% confident that perceived quality has significant effect on customer purchasing behaviors in clothing industry of Iran because p-value is equal to zero. The estimated

coefficient is equal to .521 which asserts that for every unit increase in perceived quality, purchasing behavior will go up .521 units. Consequently, the second hypothesis of this study (H2) is supported by the results.

We are 95% confident that brand loyalty has significant effect on customer purchasing behaviors in clothing industry of Iran because p-value is equal to .003 (less than .05). The estimated coefficient is equal to .116 which asserts that for every unit increase in brand loyalty, purchasing behavior will go up .116 units. Consequently, the third hypothesis of this study (H3) is supported by the results.

We are 95% confident that brand association has significant effect on customer purchasing behaviors in clothing industry of Iran because p-value is equal to zero (less than .05). The estimated coefficient is equal to .235 which asserts that for every unit increase in brand association, purchasing behavior will go up .235 units. Consequently, the fourth hypothesis of this study (H4) is supported by the results

By above results, we can formulate following regression equation:

PB= .108 + .521 (PQ)+ .235 (BAss)+ .116 (BL)+ .139 (BA)

5. CONCLUSION

By the review of literature in term of marketing management, brand equity (perceived quality, brand loyalty, brand association, and brand awareness) have high potential affect customer purchasing behavior. The findings of this research also reinforce respective theories. According to the results of this study all of the brand equity's components have significant and positive impact customer purchasing behavior in clothing industry of Iran. So, Iranian company should attempt to increase their brand equity by concentrating on its components. Since the highest impact referred to the perceived quality, Iranian company needs to improve their technologies towards higher and better quality.

The results of regression analysis shows low r-square, so future study may add another variable to improve the framework of this study.

6. REFERENCES

Aakar, D. A. (1991). Managing Brand Equity, Capitalized on the Value of a Brand Name.

Aghaei, M., Vahedi, E., Kahreh, M. S., & Pirooz, M. (2014). An Examination of the Relationship between Services Marketing Mix and Brand Equity Dimensions. *Procedia-Social and Behavioral Sciences*, *109*, 865-869.

Atilgan, E., Aksoy, S., & Akinci, S. (2005). Determinants of the brand equity: a verification approach in the beverage industry in Turkey. *Marketing intelligence & planning*, *23*(3), 237-248.

Bello, D. C., & Holbrook, M. B. (1995). Does an absence of brand equity generalize across product classes?. Journal of Business Research, 34(2), 125-131.

Boyle, E. (2003). A study of entrepreneurial brand building in the manufacturing sector in the UK. *Journal of product & brand management*, *12*(2), 79-93.

Delong, M., Bao, M., Wu, J., Chao, H., & Li, M. (2004). Perception of US branded apparel in Shanghai. *Journal of Fashion Marketing and Management*, *8*(2), 141-153.

Griffin, A., Gleason, G., Preiss, R., &Shevenaugh, D. (2012). Best practice for customer satisfaction in manufacturing firms. *Sloan Management Review*, *36*(2).

Homburg, C., Klarmann, M., & Schmitt, J. (2010). Brand awareness in business markets: when is it related to firm performance?. International Journal of Research in Marketing, 27(3), 201-212.

Hoyer, W. D., & Brown, S. P. (1990). Effects of brand awareness on choice for a common, repeat-purchase product. *Journal of consumer research*, 141-148.

Keller, L. K. (1999). Brand mantras: rationale, criteria and examples. *Journal of Marketing Management*, *15*(1-3), 43-51.

Kim, A. J., &Ko, E. (2012). Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. Journal of Business Research, 65(10), 1480-1486.

Kladou, S., &Kehagias, J. (2013).Assessing destination brand equity: An integrated approach. *Journal of Destination Marketing & Management*.

Macdonald, E. K., & Sharp, B. M. (2000). Brand Awareness Effects on Consumer Decision Making for a Common, Repeat Purchase Product:: A Replication. *Journal of business research*, *48*(1), 5-15.

MacKenzie, S. B., Lutz, R. J., & Belch, G. E. (1986). The role of attitude toward the ad as a mediator of advertising effectiveness: A test of competing explanations. Journal of marketing research, 130-143.

Murphy, J. (1998). What is branding?. In Brands (pp. 1-12). Palgrave Macmillan UK.

Nedungadi, P. (1990). Recall and consumer consideration sets: Influencing choice without altering

brand evaluations. Journal of consumer research, 263-276.

Park, J. W., Kim, K. H., & Kim, J. (2002). Acceptance of brand extensions: interactive influences of product category similarity, typicality of claimed benefits, and brand relationship quality. *Advances in consumer research*, *29*(1), 190-198.

Wong, I. A. (2013).Exploring customer equity and the role of service experience in the casino service encounter. International Journal of Hospitality Management, 32, 91-101.

Yasin, N. M., Noor, M. N., &Mohamad, O. (2007). Does image of country-of-origin matter to brand equity?. *Journal of Product & brand management*, *16*(1), 38-48.

Yoo, B., & Donthu, N. (2002). Testing cross-cultural invariance of the brand equity creation process. *Journal of Product & Brand Management*, *11*(6), 380-398.

International Journal of Science and Engineering Applications
Volume 6 Issue 01, 2017, ISSN-2319-7560 (Online)

Performance Enhancement of Test Case Prioritization Using Hybrid Approach

Rajanroop Walia
D.A.V. Institute of Engineering and Technology
Jalandhar, India

Harpreet K. Bajaj
D.A.V. Institute of Engineering and Technology
Jalandhar, India

Abstract: Regression Testing is an indispensable part of software testing process. It validates all the modifications that have been introduced into the system throughout the development period. Although it is an expensive process in terms of time and cost, yet it cannot be avoided. Therefore, various techniques have been introduced in the past for reducing the expenses involved in this process. Test Case Prioritization is one such technique that schedules the execution order of test cases with an aim to improve the rate of fault detection. In this paper, a hybrid approach has been presented which is a combination of two approaches, Adaptive approach and Genetic algorithm. The approach works by firstly employing an adaptive approach to prioritize the test cases according to their statement coverage. Further, the leftover test cases are prioritized using Genetic Algorithm. Finally, the results of the proposed approach are compared with those of Genetic Algorithm based on two parameters: execution time and average percentage of statement coverage (APSC) values. The results confirm that the proposed approach performs better in terms of both parameters.

Keywords: regression testing, test case prioritization, genetic algorithms, adaptive approach.

1. INTRODUCTION

Regression testing aims to verify that the software still performs in the same manner as it did before it was changed [1]. However, regression testing can be expensive and time-consuming, especially when the test suite involved in testing the software is large. This limitation triggered the efforts to truncate these expenses and thus, led to the development of three main techniques namely, test case prioritization, test case selection and test suite minimization. Test case prioritization attempts to reorder the test cases so as to improve the rate of fault detection. Test case selection selects a subset of the original test suite for execution. Finally, test suite minimization shrinks the original test suite such that it still maintains the coverage. Among these techniques, test case prioritization is considered to be most efficient since it takes into account all the test cases contained in the test suite and identifies the best execution sequence that meets a certain testing criteria. This is not so in case of other two techniques as they do not cover all the test cases of a test suite and thus increase the risk of software containing undetected errors [2].

Various prioritization techniques have been proposed in the past including genetic algorithms, ant colony optimization, particle swarm optimization, history-based approach and adaptive approach. Among these techniques, Genetic algorithms are widely used in solving test case prioritization problems, by generating results using the techniques inspired by natural selection. But they consume too much time in doing so. This is so because they carry out test suite prioritization and execution as separate phases. On the other hand, an adaptive approach which is gaining popularity nowadays, saves time by carrying out prioritization and execution of test cases simultaneously. But it only schedules the order of those test cases which have achieved some amount of statement coverage on the previous program. This means all of the test cases are not prioritized by an adaptive approach, which further implies that full statement coverage has not yet been achieved. Therefore, a hybrid approach has been proposed in this paper, which is a combination of the above two approaches. It overcomes the limitations of both the approaches by achieving almost 100% statement coverage in minimum time.

This paper is organized as follows. Section 2 describes related work. Section 3 explains some existing test case prioritization approaches. Section 4 describes the proposed work. Section 5 explains how the experiment is carried out and presents the results. Section 6 concludes the paper.

2. RELATED WORK

An in-depth analysis of regression testing was presented in order to remove the constraints associated with it. In [1] Y. Li gave a detailed description of regression testing including its definition and types. Apart from this, they also compared the *retest all* and *selective* regression testing strategies and concluded that there is tradeoff between the both. However, it was explained in [2] that as the size of test suite increases, *retest all* strategy becomes infeasible because of time and cost constraints. Thus, it revealed an increasing trend towards the different techniques to remove these constraints namely, test case prioritization, test suite minimization and test case selection. However, test case prioritization gained much popularity which is evident from the vast amount of work that has been done in this field. Y.C. Huang in [3] proposed a cost-cognizant prioritization technique that ordered test cases according to their history information by using genetic algorithm. The technique prioritized test cases on the basis of their test costs and fault severities, without analyzing the source code. Its efficiency was evaluated using a UNIX utility program and the results confirmed the usefulness of the proposed technique. In [4], a technique for identifying the test path that must be tested first in case of static testing was proposed. Test paths or scenarios were derived from source code. In order to find the path to be tested first, the approach made use of Information Flow model and Genetic Algorithm.Y. Huang in [5], proposed a method of cost-cognizant test case prioritization which was based on the use of historical records. The historical records were gathered from the latest regression testing and then a genetic algorithm was proposed to determine the most effective order. Evaluation results proved that the proposed method improved the fault detection effectiveness. In [6], the necessity of Component-Based Software testing prioritization framework was developed and proved, which uncovered more extreme bugs at an early stage and enhanced software product

deliverable quality utilizing Genetic Algorithm with java decoding technique. For this, they proposed a set of prioritization keys. An algorithm to prioritize test cases based on total coverage using a modified genetic algorithm was proposed in [7]. The performance of the proposed algorithm was compared with five other approaches and the results indicated that the proposed algorithm was better than other approaches. However, the same could not be guaranteed for bigger test suites. In [8], Y. Singh presented a regression test prioritization technique based on Ant Colony Optimization to reorder test suites in time constrained environment. On the other hand, a modified version of Ant Colony Optimization for test case prioritization was also presented in [9]. The performance in both the cases was evaluated using the Average Percentage of Faults detected (APFD) metric and the results proved the effectiveness of these techniques. Tyagi in [10] proposed a 3-step approach to perform regression testing using Multi Objective Particle Swarm Optimization. The proposed MOPSO outperformed other approaches like No Ordering, Reverse Ordering and Random Ordering as it achieved maximum fault coverage and maximum value of APFD in minimum execution time. In [11], history-based approach for prioritizing the test cases was extended to modified lines. The modified lines were prioritized first and then subsequently followed by the test cases. The results showed that the proposed approach was able to detect faults faster and with less effort as compared to previous approach. Dan Hao in [12] presented an adaptive TCP approach, which worked by determining the test case execution order simultaneously during the execution of test cases on the modified program. The results showed the proposed adaptive approach to be significantly better than the total test case prioritization approach and comparable to additional statement-coverage based test case prioritization approach. In [13], L. Mei proposed Preemptive Regression Testing (PRT), a novel strategy that rescheduled test cases based on the changes of the service under test detected in the course of each actual regression test session. Three particular PRT strategies, integrated with existing test case prioritization techniques were proposed to generate new techniques. The experimental result confirmed that one of the PRT-enriched techniques was able to test workflow-based web service. A novel family of input-based local-beam-search adaptive-randomized techniques was proposed in [14]. The results showed that these techniques achieved either higher or same mean APFD values as the existing code-coverage-based greedy or search-based prioritization techniques. A. Schwartza in [15] empirically studied the existing strategies and developed two additional Adaptive Test Prioritization (ATP) strategies using fuzzy analytical hierarchy process (AHP) and the weighted sum model (WSM). The empirical studies provided in this research showed that utilizing these strategies can improve the cost-effectiveness of regression testing.

3. EXISTING TEST CASE PRIORITIZATION APPROACHES

3.1 Genetic Algorithm

Genetic algorithm is an evolutionary method that generates solutions to optimization problems using the techniques which are based on the principles of natural selection. It works by repeatedly evolving a population of individuals represented as chromosomes, towards a better solution. During each step, it chooses individuals from the current population based on their fitness values, which are calculated in accordance to some objective function in the problem being solved. Once the best fit chromosomes are selected, they are then modified by applying the following genetic operators in order to produce the next generation:

a) Crossover: Crossover operator is used to vary the chromosomes from one generation to next in such a manner that the new chromosome formed after applying crossover is better than original chromosomes. In other words, it mimics the process of biological evolution by taking more than one chromosomes as parents and then producing a child chromosome from them. In case of one-point crossover, a random crossover point is selected in both the parent chromosomes and then their tails are swapped to get new off-springs as shown below:

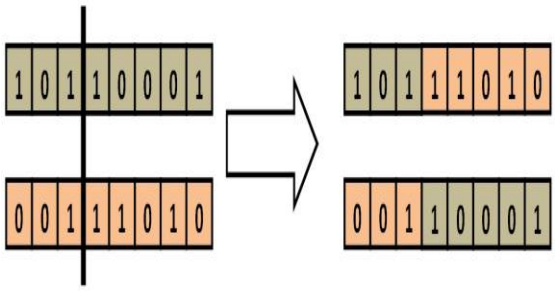

Fig 1: Crossover operation

b) Mutation: Mutation operator is applied to inject diversity in the population of chromosomes by altering one or more gene values in a chromosome. It can lead to a solution which is entirely different from the previous solution. In case of bit-flip mutation, one or more random bits are selected and flipped as shown below:

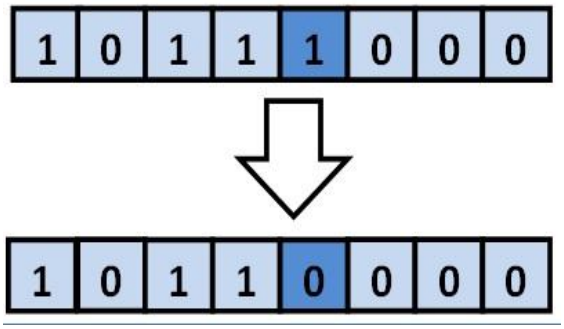

Fig. 2: Mutation operation

3.2 Adaptive Approach

An adaptive approach for solving prioritization problems has gained much popularity in the recent years. Unlike existing test case prioritization approaches that prioritize the test cases before running them on the modified program, an adaptive approach works by prioritizing the test cases simultaneously during their execution. It does so by calculating the initial fault detection capability (denoted by *Priority (t)*) of each test case according to its statement coverage on the previous program and selects a test case t_s with the largest *Priority*. This *Priority (t)* is given by the following equation:

$$Priority(t) = \sum_{S \text{ is executed by } t} Potential(s) \quad (1)$$

where *Potential (s)* denotes how likely a statement s contains faults that have not been discovered by the existing test suite where *Potential (s)* of any statement lies in the interval [0,1]. Initially, all statements have *Potential* 1. The adaptive approach then runs the test case with the largest priority and then modifies the *Potential* of each statement *s* according to whether its output is passed or failed. In other words, it modifies the *Potential* on the basis of the following equation:

$$Potential(s) = \begin{cases} Potential'(s), & s \text{ is executed by } t' \\ Potential'(s) * q, & s \text{ is executed by } t' \wedge t' \text{ is passed} \\ Potential'(s) * p, & s \text{ is executed by } t' \wedge t' \text{ is failed} \end{cases} \quad (2)$$

where *Potential'(s)* represents the probability of any statement comprising new faults before running any test case *t'*. *p* and *q* are two non-negative constants, whose value lies between 0 and1. This process is repeated until all the test cases are prioritized and executed.

4. PROPOSED WORK

Genetic Algorithms provide excellent solutions to prioritization problems but take significant amount of time to do so. This is so because firstly they schedule the order of test cases and then execute them. On the other hand, adaptive approach prioritizes the test cases on the basis of their output information. In other words, test case prioritization and execution take place simultaneously in case of adaptive approach. Since both processes occur concurrently, time expenses are reduced to a great extent. But adaptive approach prioritizes only those test cases that achieve some amount of statement coverage. The test cases which are unable to cover any statements are left non-prioritized which implies that statement coverage has not been done perfectly. Therefore, in order to prioritize all the test cases, a hybrid approach has been designed. In this approach, the test cases that cover the code statements are prioritized first using an adaptive approach. The leftover test cases that do not cover any statement are prioritized using Genetic Algorithm by applying four operations: parent selection, crossover, mutation and duplicate elimination. The benefit of this approach is that besides saving time, it achieves almost 100% statement coverage.

The step-by-step working of hybrid approach is shown below in the flowchart given in Figure 3:

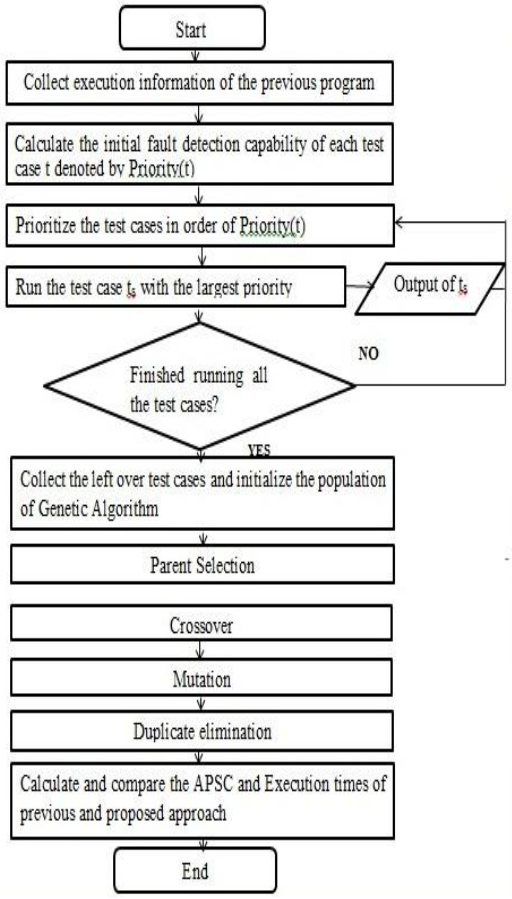

Fig. 3: Flowchart of the proposed technique

Finally, the efficiency of the proposed approach is evaluated by comparing its results with those of Genetic Algorithm on the basis of two parameters: APSC and Execution time.

5. EXPERIMENTAL EVALUATION

In order to prove the effectiveness of the proposed technique, 100 test cases along with their statement coverage have been collected from Apache Open Source by interfacing it in Eclipse and then testing it with Junit test toolkit. This dataset has been used for implementation of the proposed approach. For the purpose of comparison, Genetic Algorithm has also been implemented on the same dataset. Post implementation, the performance of both the approaches have been calculated according to two parameters: Execution Time and Average Percentage of Statement Coverage (APSC) values. APSC is defined as the degree to which a prioritized test suite covers the statements. It is calculated as shown below:

$$APSC = 1 - \frac{TS_1 + TS_2 + \cdots + TS_m}{nm} + \frac{1}{2n} \quad (3)$$

where

TS_i denotes the id of first test case that first covers the statement i in the execution sequence.

M denotes the number of statements.

N denotes the number of test cases.

The first set of each of these values for both the techniques has been acquired by altering the number of test cases in the dataset, as given by Table 1. This is done by creating four different subsets of the original dataset, containing 25, 50, 75 and 100 test cases respectively. Figures 4 and 5 show the bar graphs for APSC and Execution Time respectively. It is clearly visible that the proposed hybrid approach maximizes the statement coverage up to 5 percent and minimizes the execution time to a great extent.

Table 1: Comparison based on the number of Test Cases

No. of Test Cases	APSC values (in %)		Execution Time Values (in ms)	
	APSC (GA)	APSC (HY)	Time (GA)	Time (HY)
25	98.7	99.9	51394	14950
50	97.35	99.79	114518	61974
75	97.23	99.61	187018	115882
100	95.91	99.57	331653	292261

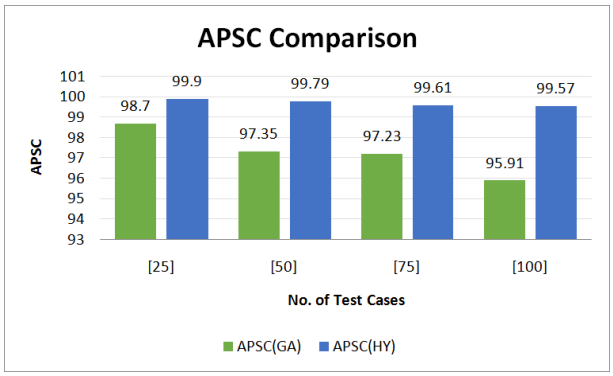

Fig. 4: Graph showing APSC values of Genetic Algorithm and Hybrid Approach corresponding to the number of test cases.

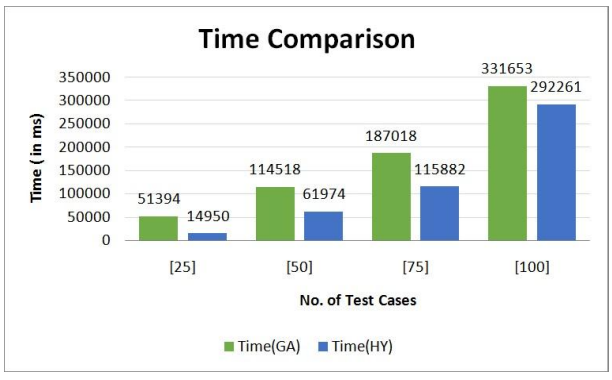

Fig. 5: Graph showing Execution Time values of Genetic Algorithm and Hybrid Approach corresponding to the number of test cases.

The next set of APSC and execution time values of both approaches has been obtained by taking into account, the number of generations. Figures 6 and 7 show the bar graphs for APSC and Execution Time respectively. Figures 8 and 9 show the line graphs for the same. From both the graphs, it can be observed that the proposed hybrid approach outperforms the genetic algorithm by 5 percent in terms of APSC values. As far as execution time is concerned, a significant difference can be observed in that also.

Table 2: Comparison based on the number of Generations

No. of Generations	APSC values (in %)		Execution Time values (in ms)	
	APSC (GA)	APSC (HY)	Time (GA)	Time (HY)
[2]	97.53	99.74	462020	212564
[3]	97.4	99.6	533850	327632
[4]	96.67	99.59	538903	338389
[5]	94.55	99.5	660924	491002

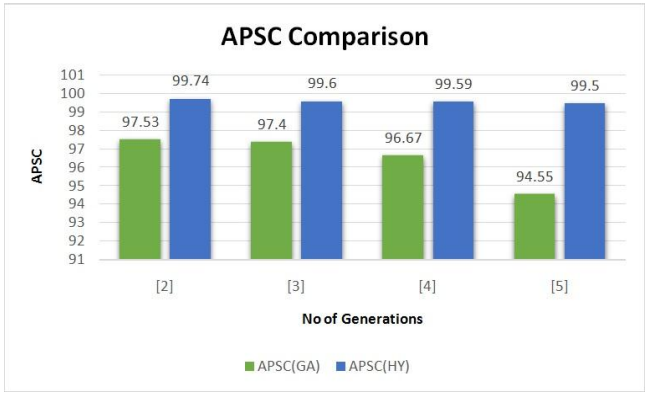

Fig. 6: Graph showing APSC values of Genetic Algorithm and Hybrid approach corresponding to number of Generations.

Fig. 7: Graph showing Execution Time values of Genetic Algorithm and Hybrid approach corresponding to number of Generations.

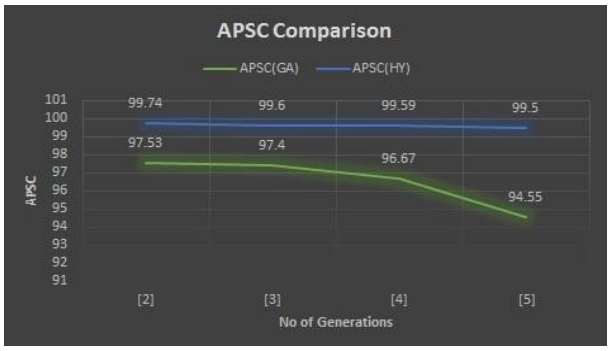

Fig. 8: Graph showing APSC values of Genetic Algorithm and Hybrid approach corresponding to number of Generations.

Fig. 9: Graph showing Execution Time values of Genetic Algorithm and Hybrid approach corresponding to number of Generations.

6. CONCLUSION

In this paper, two test case prioritization approaches, adaptive approach and genetic algorithms, have been combined to form a hybrid approach. Unlike other prioritization approaches, adaptive approach carries out prioritization and execution of test cases simultaneously. Firstly, it selects a test case according to its initial fault detection capability (priority) in the previous program. Then it executes that test case and records its output. Based on the output of first test case and the execution history of next unselected test case, it prioritizes that test case. This process continues till all the test cases which cover code statements are prioritized and executed. Further, the test cases that are unable to cover any statements are taken by Genetic algorithm and prioritized using four operations, parent selection, crossover, mutation and duplicate elimination. The performance of the hybrid approach is further compared with that of Genetic Algorithm. The experimental results show that the proposed approach outperformed the latter in terms of execution time and APSC values.

7. ACKNOWLEDGMENTS

I express my sincere gratitude and thanks to Ms. Harpreet K. Bajaj, Head of Department, Computer Science and Engineering at DAVIET Jalandhar. Under her valuable guidance and continuous support, I have been able to complete this paper.

8. REFERENCES

[1] Li Y, Wahl N J. (1999). An Overview of Regression Testing. ACM SIGSOFT Software Engineering Notes, 25(1), 69-73.

[2] Yoo S, Harman M. (2012). Regression testing minimization, selection and prioritization: a survey. Software Testing, Verification and Reliability, 22(2), 67-120.

[3] Huang Y C, Huang C Y, Chang J R. (2010). Design and Analysis of Cost-Cognizant Test Case Prioritization Using Genetic Algorithm with Test History. In: Proceedings of 34th IEEE Annual Computer Software and Applications Conference, 413-418.

[4] Sabharwal S, Sibal R, Sharma C. (2011). A Genetic Algorithm based Approach for Prioritization of Test Case Scenarios in static testing. In: Proceedings of International Conference on Computers and Communication Technology (ICCCT), 304-309.

[5] Huang CY, Peng KL, Huang YC. (2012). A history-based cost-cognizant test case prioritization technique in regression testing. Journal of Systems and Software, 85(3), 626-637.

[6] Mahajan S, Joshi S D, Khanaa V. (2015). Component-Based Software System Test Case Prioritization with Genetic Algorithm Decoding Technique Using Java Platform. In: Proceedings of IEEE International Conference on Computing Communication Control and Automation, (ICCUBEA), 847-851.

[7] Ramingwong L, Konsaard P. (2015). Total Coverage Based Regression Test Case Prioritization using Genetic Algorithm. In: Proceedings of 12th IEEE International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 1-6.

[8] Singh Y, Kaur A, Suri B. (2010). Test Case Prioritization using Ant Colony Optimization. ACM SIGSOFT Software Engineering Notes, 35(4), 1-7.

[9] Solanki K, Singh Y, Dalal S. (2015). Test Case Prioritization: An Approach Based on Modified Ant Colony Optimization (m-ACO). In: Proceedings of IEEE International Conference on Computer, Communication and Control (ICCCC), 1-6.

[10] Tyagi M, Malhotra S. (2014). Test Case Prioritization using Multi Objective Particle Swarm Optimizer. In: Proceedings of IEEE International Conference on Signal Propagation and Computer Technology (ICSPCT), 390-395.

[11] Gupta A, Mishra N, Tripathi A, Vardhan M, Kushwaha DS. (2015). An Improved History- Based Test Prioritization Technique Using Code Coverage. Advanced Computer and Communication Engineering Technology, 315, 437-448.

[12] Hao D, Zhao X, Zhang L. (2013). Adaptive Test-Case Prioritization Guided by Output Inspection. In: Proceedings of 37th IEEE Annual Computer Software and Applications Conference (COMPSAC), 169-179.

[13] Mei L, Chan W K, Tse T H, Jiang B. (2015) Preemptive Regression Testing of Workflow-based Web Services. IEEE Trans. On Services Computing, 8 (5): 740-754.

[14] Jiang B, Chan W K. (2015). Input based adaptive randomized test case prioritization: A local beam search approach. Journal of Systems and Software, 105, 91-106.

[15] Schwartza A, Do H. (2016). Cost-effective regression testing through Adaptive Test Prioritization strategies. Journal of Systems and Software, 115, 61-81.