International Journal of Engineering & Technology, 7 (4) (2018) 2552-2557

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.14881
Research paper

Class level software fault prediction using step wise linear regression

Sumangala Patil [1] *, A. Nagaraja Rao [2], C. Shoba Bindu [3]

[1] Department of computer Science and Engineering, Research Scholar, JNTUA, Ananthapuramu, Andhra Pradesh, India
[2] Department of Computer Science and Engineering, Head of the Department & Professor, SJT, Vellore Institute of Technology, University, And Tamil Nadu, India
[3] Department of Computer Science and Engineering, Professor, JNTUA, Anantapuramu, Andhra Pradesh, India
**Corresponding author E-mail: Sumangalaapatil@gmail.com*

Abstract

Programming measurements was utilized for foreseeing issue in modules of programming ventures. Convenient forecast of flaws enhances programming quality and subsequently its dependability. In this paper, a framework towards subspace grouping of large data set was pro-posed at class level to minimize the error. We composed an iterative calculation for grouping of high dimensional datasets for improvement of a target work. At that point the bunched data sets were examined utilizing Step-Wise Linear Regression to investigate the relationship among a structure variable and the autonomous factors in order to anticipate of damaged and non-faulty classes. To evaluate the supportive-ness of the model, we drove a practical learning on the Attitude Survey Data. The proposed strategy specifically managed blunder variables and consequently gave precise fault prediction least standard error (0.003) when contrasted with the current technique (4.687). Root mean square error which measures the distinction between the assessed error and the real error was (0.8) in the proposed technique. The results demonstrated that the forecast models based on subspace clustering were essentially predominant to the current techniques.

Keywords: *Software Metrics Model; Fault Prediction; Subspace Clustering; Stepwise Linear Regression; Standard Error; Root Mean Square Error*.

1. Introduction

T Fault Prediction is reasonably another examination range of programming quality affirmation. Points of the undertaking group are to make a decent fine item with zero or few deformities. Nature of item is connected with the scope of deficiencies notwithstanding time and cost.

This causes fault inclined items conveyed to clients, and first issues in framework steadfastness and wellbeing. Cost is the most essential factor that advances this bother to get to the programming framework in which flaws will probably be identified to perceive the quality affirmation engineers. Therefore the testing and quality affirmation assets might be conveyed more effectively providing the blemish forecast impacts.

Here, this work for better fault expectation process alluded a portion of the current works [1], [2], which demonstrated that the technique to compute faults in a class was to secure from classes that include same qualities inside it rather than learning from programming units with homogeneous attributes, was predominant than the information from entire frameworks.

Reference [3] shows that class subordinate models are much prevalent than bundle based expectation models. Also, a reasonable look on late open source programming [4] demonstrated that semantic mistakes, which remain application-particular and return inconsistence with a similar outline necessities or the plan reason have turned into the vital root reasons.

In some different works [5-6] proposed to forsee their deficiencies by making routine with regards to bundles. Generally segments to bunch programming measurements have the probabilities of disappointment of a unit. Generally bundles for the most part are not free on its concern range [5].

In addition, the results of a sensable report on open source programming [7] set forward that the misstep partaking in the modules particular than different classes and may depend upon the issue range of the subject approach. Unfortunately, the design decay of programming frameworks was not continually expressed. To anticipate faults in a class, in future this work consolidates these two methodologies with a suspicion that the finest technique acquires from the classes are identified with it.

In this proposed work we focus on fault expectation models, subspace grouping calculation for taking care of high dimensional information concentrating on numerical measurement and we composed an iterative calculation for bunching high clear cut datasets in light of the enhancement of a target work for bunching. The proposed straight overlaps programming flaw forecast with subspace bunching approach is changed from previous fault expectation works that were basically relied upon setting together information and classes. The fundamental target of this exertion is the change of a computerized method for passing on mistake free classes to the parts of the frameworks and to dispose of the subjective expert feeling.

1.1. Related study

Reference [8] shows that at two changed models which fuse conduct towards work into the expectation procedure and assess the prescient capacity of such plans. Together two models accomplished apparently better. While the estimation measure has mulled over endeavor these fault expectation. Point of the forecast

models is distinguishing blunder inclined models of a product framework to great component affirmation exercises, for example, test or code surveys. Already some numbers of research specialists have been effectively looked into for such models more than decade. However for every framework the vast majority of the proposed frameworks cost is most imperative factor for quality affirmation exercises. These frameworks make utilization of irrelevantly classifier requesting documents just by their size. In any event when exertion was disregarded or aimed to the estimation of a classifier is performed shockingly well.

Reference [9] shows analyzed the connection between product measures and the blunder inclined modules by considering relations between the two. To foresee programming absconds have been broadly utilized by gauge of the product items. However these measures bolster grow great order models, ponders to set forward for connection between programming measures and blunder, basically it should be evaluated and perceived by the basic arrangement of the product measures that are intensely connected with mistakes crosswise over five datasets of PROMISE vault.

Reference [10] proposed double new measurements, in particular, PIEDG and PIMDG for anticipating shortcomings based on reliance between the document level segments of the code. The proposed measurements utilized the possibility of interest of hubs in the product reliance diagram to distinguish the parts of the code which have shortcomings. This exploration likewise analyzed the consequence of self-image chart on the product reliance worldwide diagram. The possibility of the product fault indicator which was based on the proposed measurements was assessed utilizing the Eclipse bug dataset and got promising results in desire of the product issues.

Reference [11] proposed utilization of hyper-quad tree made k-means calculation to anticipate their flaws. This show how we can apply k-strategy without giving scope of group and the focuses of bunch and ascertain programming with blunder degree by changing the scope of Δ.But it is important to evacuate clamor before applying this calculation which is delicate to boisterous occasions.

2. Problem statement

A This research study points problem of selecting a class level fault prediction using software clustering from back ground domain datasets which permits to elaborates a prediction model connected latest software clustering project. In comparable of past datasets has negative collide to modern performance. Evaluation shows accuracy for the selection of statistical techniques along with multivariate liner regression gives best performance when linear regression model is selected.

The following method was evaluated by using software clustering as class level fault perdition.

- Fuzzy C- Mean(FCM)
- Sparse Subspace clustering (SSC)

Among the many clustering methods, we selected two best clustering methods to represents the difference .were selected to represent the difference between clustering methods and selecting the best method. The choosing clustering method for numerical data and took the information from past and used to their research work done by [5] classes are connected to cluster and small, the smaller one easy to handle rather than thee entering system.

The difficult in FCM algorithm [12], each point has a degree associated with a specific cluster that point does not fit in a cluster as much as weak or strong association to the clusters, another important point is that unable to identify some natural cluster with lower value of termination criterion it get good results but if more number iteration is done it becomes expensive.

The above problems overcome by sparse subspace clustering algorithm [13] it help in selecting data points that exists in multiple possible overlapping subspaces and integrate feature by evaluation and clustering in order to find cluster in different subspace it can also handle both linear and affine subspace explicitly.

2.1. Method of evaluation

Prediction validation

Towards the estimation of the excellent prediction of system received by the Step wise Linear Regression, we conducted the k-method. The k-method is broadly applied to evaluate how the outcomes of a statistical analysis may be generalized to a nondependent information group. [5],[14] Specifically, when the aim is the prediction, the k-method is used to evaluate efficaciously how a predictive model will carry out in exercise. The k-method is going thru k series. Every round of the validation includes the diving of basic statistics groups into N distinct subsets of training and check groups with at least one review for every test. It can be viable that Step wise linear regression flops to construct fault prediction fashions. This occurs even classes made from the instructions in a group do not have any faults or only a few faults. Step wise linear regression flops to construct prediction fashions additionally whilst the variety of classes is small. This could appear most effective on joined clusters. When prediction models aren't constructed, the predicted faults inside the take a look at set are acquired by using thinking about the mean of the faults inside the training unit. Towards assess the best of fault prediction accomplished by SWLR inside the SSC system we calculated: Sum (S), Median (MD), Mean (M), and Standard Deviation (SD) of Absolute Residuals (AR). Specified the prediction also Actual Values x and y, then the Absolute Residual is same to $|x - y|$. That is a broadly used overall performance degree. The lesser the range, the higher prediction of faults is. To associates the proposed technique and the baseline in terms of fault prediction accuracy; we computed the usage of the subsequent method:

$$\text{error} = \frac{\text{MAR(clustering)} - \text{MAR(Flat)}}{\text{StDAR(Flat)}} \quad (1)$$

MAR (clustering) is the Median of the Mean values are absolutely residuals calculated for the systems constructed at cluster degree. MAR (Flat) and StDAR (Flat) also remain the same old deviation values of absolute residuals received after the fault prediction model constructed on the complete system, considerably. The variable fault can take the range in between -1 and +1. Specified software program launches, a positive ranges shows that the baseline is well, whilst a negative ranges show that our method out plays the baseline. In comparison with the base line, the absolute value of errors suggests that the magnitude of how good or bad our system is. If two error are zero then the techniques remain the same and the prediction errors will be equivalent. To affirm whether the SSC method is considerably higher than the baseline, we examined the subsequent null hypothesis (HnO):

Null hypothesis

With the baseline technique, mistake made by our technique are notably more. This assumption is one-sided because we expect that our approach is more correct within the prediction of faults. To test Hn0, we deliberate to apply the unpaired t-test on the two distributions of +ve and -ve errors values. The use of this test is possible, if the +ve and –ve error values and poor mistakes values are usually disbursed. This assumption is validated using the Shapiro-Wilk test.

Threats to validity

There are numerous threats that might affect the validity of the effects. In order to mitigate those threats we took the subsequent steps; the pre-processing of the dataset has not been carried out. This allows different researchers to replicate our-examine without problems. We exploited a public data set from the mind-set of Survey statistics, which have been employed for fault prediction in other empirical research. Within the SSC study we used all of the systems in that dataset for which we are able to find supply code. Outside validity threats are continuously exist while using facts from a specific environment. We justifying this threat with the aid of considering structures from distinct domains and replicating with other systems to verify or contradict the consequences

If you want to draw an accurate conclusion from the evaluation of the combined records, we selected a set of statistical assessments to make certain that our observations are not taking place by way of any chance.

Accuracy (A)

Accuracy is described as the proportion of predicted faults inclined modules which can be inspected by all modules.

$$A = \frac{TN+TP}{TN+FN+FP+TP} \quad (2)$$

Where, if the predictions class value and the actual class value is real i.e. known as Real Positive (TP). In case, the actual class value is untrue and the prediction class value is true means Untrue Positive (FP). If, the actual class value is real and the prediction class value is untrue means Untrue Negative (FN). If, both the actual and the prediction class value is untrue means Real Negative (TN). Figure1 depicts the prediction accuracy of the SSC and the FCM methods for fault prediction of classes in system for Attitude Survey Data. From the graph 1, it is clear that the SSC regression method of fault prediction performed with high accuracy. This is mainly because of the fact that the SSC method directly dealt with error factors and minimized the inaccuracy components. Thus the standard errors are also minimum while employing the SSC fault detection method.

3. Techniques used

3.1. A fuzzy C-means clustering (FCM)

A clustering process is dived into groups based on a closeness method. The each dataset may belong to more than one group; on the base of the degree of the membership each dataset gives probability distribution over the cluster.

The FCM algorithm attempts to partition a finite collection of n elements X={x1….xn} into a collection of c Fuzzy clusters with respect to some given criterion. Given a finite set of data, the algorithm returns a list of c clusters centers C= {c1…, cn) and a partition matrix.

i) Iteration

Each iteration of FCM algorithm the following objective function is optimized

$$\arg_c \min \sum_{i=1}^{n} \sum_{j=1}^{c} w_{ij}^{m} \left\| x_i - c_j \right\|^2 \quad (1)$$

ii) Centroid

The arrangement of the coefficients will give the level of the k^{th} group $w_k(x)$ at any given point x. c-means is the mean of all focuses of the centroid of a group, weighted by their level of having a place with the group:

$$C_k = \frac{\sum_z w_k(x)^m x}{\sum_x w_k(x)^m} \quad (2)$$

Where degree of membership for a given data point X_i the degree of its membership to cluster j is calculated as follows

$$W_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}}} \quad (3)$$

iii) Fuzziness Coefficient

In the equation (2) and (3) the fuzziness coefficient m where $1 < m < \infty$ measure the tolerance of the required clustering this value determine how much the cluster can overlap with one another. The higher the fuzziness a larger number of data point will fall inside the fuzzy bound.

3.2. Sparse subspace clustering (SSC)

The algorithm is based on the sparse representation of the data usually a, high dimensional data resides in multiple low dimensional subspace Even if the representation might not be unique but the core idea behind the algorithm is to represent every data point as linear combination of other points from its own subspace. This is a look for cluster in subspace of unknown dimensions. Recall at this instant, where the records points considered from fixed linear or affine subspaces is contaminated with noise.

Particularly assume that $\bar{y}_i = y_i + \zeta_i$ ith recorded point corrupted with noise ζ_i bounded by using$\|\zeta_i\|_2 \leq \epsilon$. ‖.With the intention to improve the sparse substitution of $\bar{y}_i$, we can examine for the sparse answer of $\bar{y}_i = y_i C_i$ with error of $\|y_i C_i - \bar{y}_i\|_2 \leq \epsilon$. it can find this type of sparse representation via fixing the subsequent hassle;

$$\min \|C_i\|_1 \text{Subject to } \|y_i C_i - \bar{y}_i\|_2 \leq \epsilon \quad (4)$$

However, in many conditions we do not recognize the noise degree ϵ earlier. In such instances, the Lasso optimization set of rules [8] may be used to get better solution from

$$\min \|C_i\|_1 + r\|y_i C_i - \bar{y}_i\|_2 \quad (5)$$

In this ϒ >0 and is an unchanged value. In case the facts considered from number of affine subspaces are contaminated with noise, the sparse substitution may be received by means of fixing the problem and adopting the modified Lasso optimization algorithm.

$$\min \|C_i\|_1 + r\|y_i C_i - \bar{y}_i\|_2$$

$$\text{Subject to} C_i^T 1 = 1 \quad (6)$$

Partition of the records into exclusive subspaces follows with the aid of making use of spectral clustering to the Laplacian of $\widetilde{G}$.

3.3. Stepwise linear regression (SWLR)

Within the proposed research, we took the (SWLR) approach, which permits computing linear regression in levels and to evaluate model used for fault prediction with suitable consequences. To improve the accuracy of class level fault prediction, we chose linear regression and basic relationship between source code classes. Step wise linear explores the connection among a dependent variable and one or more independent variables, presenting a version defined with the aid of a linear equation:

$$y = b_1 x_1 + b_2 x_2 + \cdots + b_n x_n + c \quad (7)$$

In which y is the dependent variable,x1,x2,…,xn are the independent variables, b_i is the coefficient that represents the quantity variable y adjustments whilst variable x_i changes by means of one unit, and c is the intercept. Step wise Linear Regression provides computing the above expression in steps in which the selection of the nondependent variable is conducted through an automatic system. The variables may be selected making use of three techniques: forward, backward, or a merging of both. The forward selection begins with evolved without a variable in the version, testing the variables one after the other and adding each variable in the system if they may be substantially connected with the established variable. The backward technique begins with entire variables and testing the variables one after the other. If variables is not substantially connected with the established variable are detached from the system. SSC system uses aggregation of both techniques. In each step, this aggregation method adds or detaches variables separately with in the system if they're or are not notably connected with the dependent variable, correspondingly.

4. Aspherical design

4.1. Dataset

In this study paper, we took a real-time data set, named, Attitude Survey Data [15]. This data set encloses questions related to employee satisfaction with their supervisors in a big financial agency. The version changed and designed to measure the general performance of a supervisor, in addition to questions that associated with particular events regarding interaction among manager and worker. On this observation, we attempted to provide an explanation for the specific supervisor characteristics and common pride with the supervisor as perceived by using the personnel. Table1 offers the outline of the variables used inside the attitude Survey statistics. Here we have a look for the two large classes of variables. The variables X1, X2, and X5 relate to direct interpersonal relationships between employee and manager, while variables X3 and X4 are of a much less private nature and relate to the process. The variable X6 isn't always an immediate assessment of the supervisor but serves extra as a standard degree of the way the worker perceives his or her personal development inside the business enterprise. The information used within the analyses is given in table 1 which was received by using aggregating the responses obtained from the different employee. Y is dependent or response variable and p predictor or explanatory variables, X1, X2...XP are the resulting data of 30 observations. The Y and X1, Xtwo ..., Xp are related by linear model as shown in table1.

Table 1: Description of Variables Included in Attitude Survey Data

Variable	Description
Y	Overall rating of job being done by the supervisor
X1	Handles employee complaint
X2	Does not allow special privileges
X3	Opportunity to learn new things
X4	Raised based on performance
X5	Too critical of poor performance
X6	Rate of advancing to better jobs

4.2. Empirical procedure

The model is used to predict error in class level by using linear regression to explore the relation between a dependent variable and the independent variable so as to predict defective and non-defective classes. This method directly deals with error factors. The coefficient is the metric that determines the efficiency of the regression model. The standard error is the measure of the statistical accuracy of the fault prediction which is equal to the standard deviation of the theoretical analysis. The t-value is the ratio of the departure of an estimated defect prediction parameter from its notional value and its standard error. Based on these parameters, the attributes of the datasets are evaluated for both clustering methods.

The following steps for clustering algorithms

Step1: Using hold-out method by separating training and testing sets

Step 2: Selecting features for clustering

Step 3: The clustering approach is applied on the standardized training set

Step 4: Computing a BPRM for each of the clusters

Step 5: The testing set are classified into clusters by projecting

Step 6: calculated evaluation criteria

Fuzzy C-Means algorithm tested in this experiment

1) Number of clusters $1 < c < n$ $(1 < 3 < 30)$.
2) Where c= number of clusters, n=number of observations Maximum number of iteration =5.
3) Fuzziness degree m =0.5.
4) Termination measure =0.05
5) Termination threshold =0.01

Above clustering algorithm are used to similarity measurements. Where each item may belong to more than one fuzzy (group) and the degree membership for each item is given by a probability distribution over the cluster. SSC algorithm was basically addressing the data segmentations (clustering) problem, that is output tells us which data points belongs to which group, therefore we need to deviate at the final stages of the algorithm . It is uses the eigenvectors of the graph Laplacian are to apply the K-means algorithm.

5. Results and discussion

5.1. Performance evaluation parameters

The performance evaluation parameters for Step Wise Linear Regression with subspace clustering analysis were determined based on the confusion matrix. The evaluation results on the Attitude Survey Data are given in Table 2, which show the individual values (coefficients, standard errors and t-values) of the FCM and the SSC regression models on Attitude Survey Data. The coefficient is the metric that determines the efficiency of the regression model. The standard error is the degree of the statistical accuracy of the fault prediction. The departure of an expected defect prediction parameter divided by the sum of its notional value and standard error is called t-value. Based on these parameters, the attributes of the datasets are evaluated for both FCM method and the SSC method Step Wise Linear Regression-method.

Table 2: Comparison of Results of FCM and SSC Methods

Variable	FCM Algorithm			SSC Algorithm		
	Coeffiient	Std. Error	t-value	Coeffcient	Std. Error	t-value
X1	0.1821	0.3225	0.62	0.0941	0.1777	0.53
X2	0.2521	0.2101	0.62	0.0633	0.1840	0.34
X3	0.2898	0.2501	0.24	0.0613	0.1781	0.34
X4	0.3211	0.2478	0.52	0.6920	0.4355	0.59
X5	0.3211	0.2478	0.52	0.0073	0.0503	0.15
X6	0.3521	0.2414	0.50	0.0153	0.0483	0.31

Fig. 1: Integrated Graph for FCM and SSC Methods.

Table 3: The R, P and Standard Error Values of the SSC and FCM System for Attitude Survey Data

Method	R	p-value	Standard Error
FCM Clustering	0.0023	1.7387	4.6870
SSC Clustering	0.0171	4.6831	0.00341

From the figure 1 it is clear that the SSC regression method for fault prediction performed with high accuracy .This is mainly because of the fact the SSC method directly dealt with error factor and minimized in the accuracy components thus the standard error were also minimum while employing the SSC fault detection method.

It is found from Table 3 that the SSC method performs more efficiently accurately than the FCM fault prediction.

Root mean square error (RSME) is a frequently used measure of the differences between values predicted by a model and the values observed. In this paper, RMSE is employed to determine the difference between the estimated fault occurrence and the actual fault values. Figure [2] show the RMSE comparison of the SSC and FCM regression based fault prediction method. It clearly shows that the SSC method has prediction results with less RMSE values such that it has minimum errors.

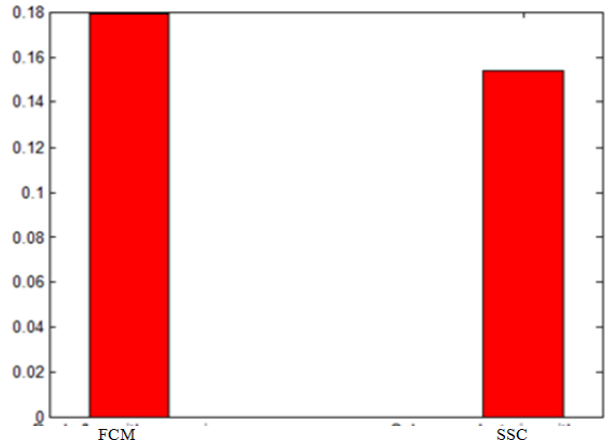

Fig. 2: RMSE Comparison of the SSC and FCM Regression Based Fault Prediction Method.

Many researchers analyzed the connection between K-means algorithm and software fault prediction. Here, we considered only those which are closely connected with an emphasis to our work, on software defect prediction process in different statistical methods, and the results of this study were compared with the existing statistical methods. In our study, the data presented in Graph (1) showed the prediction accuracy of the FCM and the SSC methods for fault prediction of classes in system for Attitude Survey Data. From the graph (1) it is clear that the SSC regression method for fault prediction performed with high accuracy. This was because of the fact that the SSC method directly dealt with error factors and minimized the inaccuracy components. Thus the standard errors were also minimum by employing the SSC fault detection method. The SSC+SWLR were statistically significant, the effect size was large and hence the fault prediction accuracy increased in the FCM+SWLR scheme. From the Graph (1), it can be found that the prediction accuracy of the FCM method varied from 0 to 0.2 (indicating 20%) while the SSC method varied from 0 to 0.7 (78%)) proving the efficiency of the SSC scheme. Table (3) showed results obtained for Step Wise Linear Regression (SWLR) with sub space clustering analysis and the FCM method in terms of R, P and Standard error values obtained for each of the attributes. The p-value is the probability to the right of our test statistic calculated using the null distribution. Further, the smaller the p-value, the stronger the evidence against the null hypothesis. Standard error is a measure of the statistical accuracy of an estimate, equal to the standard deviation of the theoretical distribution of a large population of such estimates.

It can be found that the performance measure for fault prediction in the subspace clustering with SWLR method was statistically higher than the performance measure of the FCM with SLWR system. This indicated that the local models built on a subset of the classes of the entire system showed a significantly better fit to the data compared to the global model built on all the classes together.

Graph (1) showed the standard error results for the FCM and the SSC fault prediction methods on Attitude Survey Data. The graph results indicated that the SSC method provided accurate fault prediction with minimum standard error. The FCM method had standard error in the range of 0.3 to 0.45 while the SSC method had error in the range of 0.18 to 0.43. The reason for this reduction was due to increased relevancy in the prediction of faults. Figure (1) depicted the RMSE comparison of the SSC and the FCM regression based fault prediction methods. It clearly explains that the SSC method had prediction results with less RMSE values such that it had minimum errors. The FCM scheme had RMSE of 0.18 while subspace clustering had 0.16. This reduction was due to the reduction in effort to produce false results with enhanced prediction of faults. A comparison of results based on the prediction accuracy, standard error and the RMSE values proved better performance of the subspace clustering SWLR regression based method. Thus the experimental results of the SSC and the FCM fault prediction methods proved that the subspace clustering with SWLR method provided efficient prediction of software faults with high accuracy and minimum errors. To evaluate the effectiveness of the method we conducted an empirical study on the Attitude Survey Data and experiment results revealed that the subspace clustering linear regression software fault prediction can be fully automated and effective results can be produced by using subspace clustering with software metrics. Future work of this method is to use different machine learning algorithms and regression techniques to improve the fault prediction accuracy.

The most significance of the subspace clustering system is that the SSC model showed a significantly increased fit to the data compared to global models. That is, cluster rules do better than rules learned across the whole data. Again, the difference in our new approach is on the subspace clustering approach and the focused directing test effort, reducing cost, and increasing quality of software and its reliability. The subspace clustering algorithm is efficient and can handle large data points near the intersections of subspaces. Another key advantage of the algorithm with respect to the state of the art is that it can deal directly with data nuisances, such as noise, sparse outlying entries, and missing entries.

6. Conclusion

In this work we proposed a different software fault prediction technique built on clustering of classes and software metrics. This work experimented a new linear regression technique for fault prediction using subspace clustering with software metrics. The Sparse Subspace clustering technique was based on the related and similar classes. To estimate the efficiency of the technique, we carried out an experimental look on the Attitude Survey Data. The results revealed that the proposed linear regression software program fault prediction can be completely automatic and powerful consequences can be produced by means of the usage of subspace clustering with software metrics. The future work may also consider the usage of one of a kind machine gaining knowledge of algorithm and regression strategies to enhance the fault prediction accuracy

References

[1] Mizuno O. and Hata H.2010. An integrated approach to detect fault-prone modules using complexity and text feature metrics. Advances in Computer Science and Information Technology, Springer Berlin Heidelberg.457-468.

[2] Abaei G. and Selamat A.2014.Software fault prediction based on improved fuzzy clustering.Distributed Computing and Artificial Intelligence.11th International Conference. Springer International Publishing. 165-172.

[3] Catal C., Sevim U. and Diri B.Metrics-driven software quality prediction without prior fault data. Electronic Engineering and Computing Technology. Springer Netherlands. 189-199.

[4] Tan X., Peng X., Pan S. and ZhaoW...2011. Assessing software quality by program clustering and defect prediction. Reverse Engineering (WCRE).18thWorkingConferenceIEEE. 244-248.

[5] Scanniello G., Gravino C., Marcus A. and Menzies T.2010.Class level fault prediction using software clustering. Automated Software Engineering (ASE). IEEE/ACM 28th International Conference. 640-645.

[6] Oyetoyan T.D.Conradi R. and Soares Cruzes D.2013.Criticality of defects in cyclic dependent components. Source Code Analysis and Manipulation (SCAM). IEEE 13th International Working Conference. 21-30.

[7] Faragó C., Hegedűs P. and Ferenc R.2015.Code Ownership Impact on Maintainability Computational Science and its Applications ICCSA,Springer International Publishing. 3-19.

[8] Mende T. and Koschke R.2010.Effort-aware defect prediction models. Software Maintenance and Reengineering (CSMR).14th European Conference IEEE. 107-116.

[9] Zafar. H, Rana Z. and .Shamail S.M.M.2012.inding focused item sets from software defect data.Multitopic Conference (INMIC).15th International conference IEEE, 418-423.

[10] Shekofteh.,Maryam.,Keyvan Mohebbi., and Javad Kamyabi., 2015.Software defect prediction using participation of nodes in software coupling.Journal of Theoretical and Applied Information Technology .82(3): 440-446.
[11] Sashidharan R. and Sriram P.2013.Hyper-quad tree based K means algorithm for software fault prediction. Advances in intelligent system and computing, Proceding of ICC^3.246:107-118. https://doi.org/10.1007/978-81-322-1680-3_12.
[12] wang D.,Han B. and Huang M.2012.Application of fuzzy c-means clustering algorithm based on particle swarm optimization in computer forensics,International conference on applied physcis and industrial engineering.24:1186-1191.
[13] Elhamifar E. and Vidal R.2013.Sparse subspace clustering: Algorithm, theory, and applications,Pattern Analysis and Machine Intelligence. IEEE Transactions .35(11).2765-2781.
[14] Xu J.,.Ho D. and.Capretz L.F..2015. An empirical estimation models. arXiv preprint arXiv:1507.06925,2015.
[15] Chatteree S. and A.S. HadA.S.1991.Regression Analysis by Example. Second Edition, John Wiley and Sons.