International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

A Review on Artificial intelligence Approach on Prediction of Software Defects

Poornima Gupta
Computer Science & Engineering,
B.B.D.U., Lucknow, India
purnima50gupta@gmail.com

Priya Sahai
Assistant Professor, Computer Science & Engineering,
B.B.D.U., Lucknow, India
priyasahai25101989@gmail.com

***Abstract*- This paper presents a literature review of the use of various techniques for artificial intelligence techniques in the field of software defect prediction. This literature review is very useful, since it brings a better understanding of the field of study, and this is an important contribution of this paper. Faults in software systems continue to be a major problem. A software fault is a defect that causes software failure in an executable product. Knowing the causes of possible defects as well as identifying general software process areas that may need attention from the initialization of a project could save money, time and work. The possibility of early estimating the potential faultiness of software could help on planning, controlling and executing software development activities. This paper surveys literature review of articles for the past many years in order to explore how various prediction methodologies have been developed during this period in order to take care of the issues related to software defect.**

Keywords- Software defect, Artificial intelligence, Neural Network, Fuzzy logic, Data prediction.

1. Introduction:

These days, numerous programming improvement organizations have their own particular advancement store, which ordinarily incorporates a rendition control framework and a bug following framework. This has most likely demonstrated valuable for programming deformities expectation. The forecast of the imperfection inclination of a product framework diminish the venture disappointments and reduce the aggregate expense amid the advancement and support stages. As the span of programming tasks gets to be bigger, software defect forecast methods will assume a vital part to strengthen engineers and in addition to accelerate time to advertise with more dependable programming items. Defect prediction strategies are proposed either to group the product modules as blemished or non-imperfect or speculation the quantity of imperfections in the product classes or modules. Imperfection expectation due to defects will give one more opportunity to the advancement group to retest the modules or records for which the inadequacy likelihood is high.

Machine learning calculations are connected adequately in issue areas with changing qualities and regularities. Programming quality assessment is a decent sample of the above explanation. The product metric information of modules or programming joined with defect information frames the structure of the machine learning calculations. These calculations execute a likelihood conveyance and investigate blunders. The methods prominently utilized for programming imperfection forecast issues are selection trees, neural system and Bayesian conviction system.

There are various properties or metric attributes found and utilized for programming defects prediction as a part of the writing. Rather than managing such a variety of property, it would be useful and simple in the event that we could decide the arrangement of metrics that are most vital and utilize them more to foresee the defects. In this paper we have performed review of various techniques like ANN, Fuzzy logic, Fuzzy Art map etc. to decide the compelling connections among the software metric attributes and defect prediction. In software defect prediction there are many supervised machine learning techniques like (neural network, support vector machine, Naïve Bayes etc.) exist which can reduce the time span and cost of the project. For Predicting the defect in Software there is also one new supervised machine learning technique i.e. Fuzzy ARTMAP. The metric attributes are utilized in the considered review papers generally has used software datasets which made openly accessible with a specific end goal to empower repeatable, obvious, refutable, and/or improvable prescient models of software engineering. In most of the cases there is NASA MDP dataset which are accessible to anticipate the imperfections in the software products. The point of proposed study is to investigate whether metric based assessment in the early lifecycle (i.e. necessity measurements), metric based assessment in the late lifecycle (i.e. code measurements) and metric based assessment in the early lifecycle (i.e. necessity measurements) joined with metric based assessment in the late lifecycle (i.e. code measurements) can be utilized to recognize defectiveness issue inclined modules utilizing different supervised and unsupervised prediction and estimation techniques.

2. Related Work:

In this section we have given a summarized overview of several studies that are performed in the field of software defect prediction in last couple of decades. We have described about various techniques their benefits and the concluding results which have contributed a lot in the

All Rights Reserved © 2016 IJRDASE

development of highly reliable stage of software defect prediction mechanism.

Fenton and Neil (1991) contend that despite the fact that there are such a large number of studies in the writing, software defect forecast issue is a long way for getting perfect results. There are some wrong presumptions about how defects are characterized or watched and this has brought on misdirecting results. Their case can be seen better when we see that a few papers characterize defects as watched inadequacies while a few others characterize them as remaining ones. They can gauge the software attributes by anticipating the size and multifaceted nature, testing, process quality information, multivariate methodology and so on [1]. R. Chidamber and Kemerer (1994) paper provided new groups of software metrics that is to be followed for the object oriented design. By assessing these metrics they observed that these metrics tend to various properties and propose a few methods in which object oriented approach might differ from conventional methodology. They execute six-diverse sets of metrics known as WMC, RFC, NOC, DIT, CBO, and LOCM [2]. Venkata U.B. Challagulla (2005) proposed a distinctive machine learning models for recognizing flawed real-time software modules utilizing diverse set of NASA dataset like KC1,PC1,CM1,JM1 is to be taken to anticipate the s/w product defects. When we ascertain the mean absolute error of various available software predictions it is to be found that KC1 dataset is best to anticipate the defect [3]. A.Gunes koru (2005) takes a few machines learning algorithms to foresee programming defects in software module in five NASA datasets i.e. CM1, JM1, KC1, KC2, and PC1. They performed defect prediction utilizing class-level information for KC1 instead of method level information. For this situation, the utilization of class-level information brought about enhanced forecast execution against the utilizing method level information [4]. Anuradha Chug et al. (2010) propose different grouping and bunching strategies with a target to anticipate programming imperfection. The execution of three information mining classifier calculations named J48, Random Forest, and Naive Bayesian Classifier (NBC) are assessed in view of different criteria such as ROC, Precision, MAE, RAE and so forth. Grouping method is then connected on the information set utilizing k-implies, Hierarchical Clustering and Make Density Based Clustering calculation [5]. Shanthini et al.(2012) covered the work with the primary objective to break down the execution of different classifiers on defect expectation based on open area NASA information set KC1; they broke down the execution of the classifiers utilizing customary measures, for example, exactness, review and F-measure. This study affirms that development of the SVM models is acceptable, versatile to OO frameworks, and valuable in anticipating shortcoming inclined classes for more elevated amount of measurements (class) [6]. C. Akalya Devi et al. (2012) proposed a hybrid component determination technique which gives a superior defect estimation than the conventional routines. NASA's open dataset KC1 accessible at promise software data storage is utilized. To assess the execution performance of the product defects forecast models accuracy, mean absolute errors (MAE), Root mean squared errors (RMSE) qualities measurements are utilized [7]. Ahmet Okutan et.al. (2012) proposed a novel system utilizing Bayesian systems to investigate the connections among programming measurements and imperfection inclination. They utilize nine information sets from Promise software dataset and demonstrate that RFC, LOC, and LOCQ are more viable on defect prediction expectation [8]. Martin Shepperd et.al.(2012), investigates the degree to which distributed examinations taking into account the NASA software defect datasets sets are significant and furthermore prescribed that the provenance of the information sets they utilize .This report also describes that pre-processing of dataset in adequate point of interest is useful to empower important replication and it also put exertion in comprehension the information dataset prior to apply in applying in machine learning [9]. V.Jayaraj et.al. (2013) proposed the exactness of the defect prediction of Boosting strategies for s/w defects expectation taking into account the KC1 dataset is examined. They utilized 21 system level metrics to anticipate the defects in the information data by utilizing three machine learning calculation of boosting method [10]. Malkit Singh (2013) worked on a neural system strategy and Levenberg Marquardt (LM) mechanism is created and the exactness of proposed framework is better that polynomial relation based neural system functions. In this paper Levenberg-Marquardt (LM) calculation based neural system learning method is utilized for the defect estimation due to programming imperfections at an early phase of the product advancement life cycle [11].

Martin Shepperd (2014) defect forecast analysts focused on the direct use of blind analysis it enhance reporting conventions and lead more intergroup studies keeping in mind the end goal to found simpler expertise issues. Ultimately, research is required to figure out if this inclination is common in different applications area. In this paper they take a wide range of Dataset for imperfection forecast where the exactness of Nasa MDP is superior to anything other dataset [12]. Pushpavathi T.P et al. (2014) reports an investigation for anticipating defect prediction in s/w modules utilizing coordinated methodology of genetic algorithm based fuzzy c-means clustering with random forest calculation. This technique was produced utilizing Genetic Algorithm based Fuzzy C-implies bunching with Random Forest grouping connected on exact information set and investigation was performed. At long last results were approved with the use of five NASA open space software defect information sets [13]. Sunida Ratanothayanon et al. (2012)has been proposed two classifier for the software defect prediction i.e. Back Propagation Neural Network and Radial Basis Functions with Gaussian kernels as classifiers and produce results on NASA dataset are demonstrated and investigated on the basis of mean square error and percent of accuracy [14].

3 Methodology Involved in S/W defect prediction:

We have reviewed several articles to analyse the different techniques that are involved in this decade for perfect software defect prediction. It has been observed that artificial intelligence techniques are widely used for

All Rights Reserved © 2016 IJRDASE

estimating the abnormalities in software by utilizing the various kinds of software metrics.

Table 1.1: Summarized Description of Technology used in s/w Defect Prediction in Year Wise Manner.

Ref No.	Publication/ Year	Methodology Applied	Accuracy
[10]	Software Defect Prediction using Boosting Techniques(2013)	Boosting with decision Stump	86.94%
		Boosting with REP tree	86.52%
		Boosting with M5	87.37%
[11]	Software Defect Prediction Tool based on Neural Network (2013)	Linear function based Neural Network	80.3%
		Quadratic Function Based Neural Network	78.8%
		Levenberg-Marquardt (LM) algorithm based neural network	88.1%
[12]	Software Fault Prediction in Object Oriented Software Systems Using Density Based Clustering Approach	Density based Clustering Approach	58.63%
[6]	Applying Machine Learning for Fault Prediction Using Software Metrics	Naïve Bayes	68.27%
		Support Vector Machine	80%
		K-STAR	72.41%
		Random Forest	70.3%
[7]	Comparative Classifier for Software Quality Assessment	Neural network	76.27%
		Radial Basis Function	69.52%
[3]	Empirical Assessment of Machine Learning based Software Defect Prediction Techniques(2005)	Support Vector Logic Regression	--
		NN for Discrete goal field	--
		Logistic Regression	--
		Naïve Bayes	--
		Instance Based Learning	--
		J48 Tree	--
		1-Rule	--

4. Conclusion:

This paper presents a literature review of the use of various techniques for artificial intelligence techniques in the field of software defect prediction. This literature review is very useful, since it brings a better understanding of the field of study, and this is an important contribution of this paper. From the literature review it can be concluded that this subject attracts a great deal of interest by researchers and a tremendous algorithms are present related to software defect prediction based coding in the area of Software Engineering. We have reviewed several articles to analyze the different techniques that are involved in this decade for perfect software defect prediction. So we conclude that we in future work there are tremendous scope of the application of Fuzzy logic and neural network based techniques for software defect prediction.

It has been observed that however various techniques like Support Vector Logic Regression, neural network, Boosting etc. techniques are analyzed for defect prediction but none of them gives the accuracy higher than 88 percent. Hence there is a large scope in the field of defect prediction. We can in future consider advanced hybrid mechanism like Fuzzy Artmap in performing more reliable defect prediction algorithm development.

References:

[1]. Norman Fenton, N.E. and Neil, M. (1999), “A Critique of Software Defect Prediction Models”, IEEE Transactions on Software Engineering, vol.25, no. 5, pp. 675-689.
[2]. R.Chidamer and Kemerer, “A Metrics Suit for Object Oriented Design” IEEE Transaction Software Engineering, vol.20, no.6, June 1994.
[3]. Venkata U.B. Challagulla, Farokh B. Bastani, I-Ling Yen, Raymond A. Paul, “Empirical Assessment of Machine Learning based Software Defect Prediction Techniques”, IEEE, 2005.
[4]. A.Gunes Koru, “An Investigation of the Effect of Module Size on Defect Prediction Using Static Measures” ACM SIGSOFT Software Engineering Notes, 2005.
[5]. Anuradha Chug, Shafali Dhall “Software Defect Prediction Using Supervised Learning Algorithm and Unsupervised Learning Algorithm”, International Journal of computer science and engineering Vol.27, no.5 2010.
[6]. Shanthini. A, Chandrasekaran.RM, “Applying Machine Learning for Fault Prediction Using Software Metrics”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6, June 2012.
[7]. C. Akalya Devi, K. E. Kannammal and B. Surendiran, “A Hybrid Feature Selection Model for Software Fault Prediction”, International Journal on Computational Sciences & Applications (IJCSA) Vo2, No.2, April 2012.
[8]. Ahmet Okutan Olcay Taner Yıldız, “Software defect prediction using Bayesian networks”, Empirical Software Eng (2014) 19:154–181 © Springer Science+Business Media, LLC, 2012.
[9]. Martin Shepperd, Qinbao Song Zhongbin Sun Carolyn Mair “Data Quality: Some Comments on the NASA Software Defect Data Sets” IEEE Transaction on Software Engineering, Vol. 28, Oct 2012.

All Rights Reserved © 2016 IJRDASE

[10]. V.Jayaraj, N.Saravana Raman "Software Defect Prediction using Boosting Techniques", International Journal of Computer Applications (0975 – 8887) Volume 65– No.13, March 2013.
[11]. Malkit Singh, Dalwinder Singh Salaria, "Software Defect Prediction Tool based on Neural Network", International Journal of Computer Applications vol.70-N0.22, May 2013.
[12]. Martin Shepperd, David Bowes, and Tracy Hall "The Use of Machine Learning in Software Defect Prediction", IEEE Transactions On Software Engineering, Vol. 40, NO. 6, June 2014.
[13]. Pushpavathi T.P, Suma V, Ramaswamy V, "Defect Prediction in Software Projects-Using Genetic Algorithm based Fuzzy C-Means Clustering and Random Forest Classifier", International Journal of Scientific & Engineering Research, Volume 5, Issue 9, September-2014.
[14]. Sunida Ratanothayanon, Chouvanee Srivisal, Sirirut Vanichayobonand, and Ladda Preechaveerakul, "Comparative Classifiers for Software Quality Assessment", IACSIT International Journal of Engineering and Technology, Vol. 4, No. 4, August 2012.
[15]. Supreet Kaur, and Dinesh Kumar, "Software Fault Prediction in Object Oriented Software Systems Using Density Based Clustering Approach", International Journal of Research in Engineering and Technology (IJRET) Vol. 1 No. 2 March, 2012.

All Rights Reserved © 2016 IJRDASE