ABSTRACT

The present research focuses on developing certain novel optimized neural network architectures and rule mining approaches. All the developed approaches in this research work are based on variants of biologically mimicked neural network models and nature inspired evolutionary optimization algorithms to perform defect prediction if any in the software development process. The developed techniques and compounded algorithms are applied for five publicly available NASA Promise repository datasets.

Software defect prediction is always a herculean area of research and software practitioners and researchers have carried out numerous ways to predict where the fault is likely to occur in the software module and their varying degrees of success. These prediction studies results in fault prediction models and it allows software personnel to concentrate on the defect free software code, thereby resulting in software quality improvement and employing the better utility of the resources. When software quality comes into picture, then software defect prediction (SDP) plays a major role. Software is described to be of high quality when it is defect free. This research work mainly concentrates on the internal metrics of the system which includes the source code of software systems and not their functions or behaviour of the system. The prediction of software defects at an early stage will make the company professionals to deliver a quality product to the end customers, as the cost incurred for the development play a major role.

The proposed research works in this thesis are of four modules i.e., the research contributions involve developing proposed rule mining approaches, artificial neural network models and their variants with population based evolutionary optimization algorithms like Gravitational Search Algorithm, Charged System Search Algorithm and Biogeography Based Optimization for achieving optimal solution and defect identification in software development process so as to achieve the software product with high quality. Five publicly available NASA Promise repository datasets are employed for training, testing and validating the proposed software defect predictor models.

In the first module, a classification model is proposed which is a combination of relational association rules and traditional Naive Bayes method. A novel method for detecting software modules with defects, based on relational association rule mining integrating with Naive Bayes approach, called DRAR_NB (Defect Prediction using combination of Relational Association Rules and Naive Bayes Approach). The combination of association rules with Naive Bayes on Text Categorization has already demonstrated that instead of words, word relation i.e. association rules from these words derived meaningful word sets as feature set for classification. Naive Bayes is then used on derived feature set for final target class categorization. In this proposed work, the concept of association rules and Naive Bayes is extended to capture relevant relationship between object oriented metrics using binary relations and then categorizing target class 'defective or non-defective' with simple probabilistic Naive Bayes model. The proposed model achieves better performance with class imbalance data and is compared with existing machine learning based techniques – Naive Bayes, Boosting, OneR and J48.

In the second module, a novel approach based on proposed emotional ELMAN neural network (EENN) classifier and hybrid gravitational search algorithm (GSA) – charged system search algorithm (CSSA) is developed to perform effective software defect prediction for the considered NASA datasets. ELMAN neural network is a recurrent neural network model wherein the information gets transferred to the hidden layer from the recurrent layer. This enables the increased accuracy and faster convergence of the network and that is the reason of choosing ELMAN neural network in this work. Also, in this research contribution attempt is taken to introduce emotional quotients into the ELMAN neural network, based on which the error gets reduced during the convergence of the network. The merits of individual GSA and CSSA are combined together resulting in hybrid GSA – CSSA to perform effective exploration and exploitation search mechanism and achieve better weight values to train proposed EENN model. From the simulation results, it is proved that the proposed predictor model involving the merits of GSA, CSSA and EENN model has resulted in better prediction accuracy in comparison with that of the earlier existing methodologies as available in the literature. The results were simulated with respect to both the cost sensitive and non cost- sensitive case.

In the third module, software defect predictor model is developed by employing the versatile neural network algorithm – Back Propagation Neural Network (BPN) algorithm that employs gradient descent learning rule and another neural model – Kohonen Self Organizing Feature Maps (K-SOM) that employs clustering approach to perform the defect diagnosing process. The main aim of introducing GSA – CSSA combination into back propagation neural network architecture is the ability to adapt the change in environment, escape from local and global optima problem and ensure robustness of the system. In this work, the computation of optimal

connection weights is carried out and minimization of error function (mean square error – MSE) is employed as an objective function. This research work concentrates on evolution of weights using hybrid GSA – CSSA and then performing the defect prediction action using the BPN model and KSOM model. The developed hybrid GSA – CSSA based BPN and KSOM models are applied for the public datasets from the NASA Promise repository. From the simulation results, it is proved that the proposed GSA – CSSA – BPN predictor model involving the merits of GSA, CSSA and BPN model has outperformed in an effective manner with better prediction accuracy in comparison with that of the earlier existing methodologies as available in the literature.

In the fourth module, a prediction approach based on conventional Radial Basis Function Neural Network (RBFNN) and the novel Adaptive Dimensional Biogeography Based Optimization (ADBBO) model is proposed. Radial basis function neural network is a neuronal model employing Gaussian function to enable the network to attain fast convergence. In this module, cost- sensitive RBFNN is developed along with a proposed variant of Biogeography Based Optimization. The cost-sensitivity factor is added along with RBFNN to consider the effects of false-positive and false-negative costs. The results were simulated both for the non-cost-sensitive and cost-sensitive case. The cost factors were noted to possess their influence on the probability of detection, probability of false alarm and accuracy. The weights of the radial basis function neural network are tuned employing the novel ADBBO to compute optimal weights for the training process. The computed results of the proposed ADBBO – RBFNN predictor model is compared with the earlier existing algorithms in the literature and that of the other proposed models in this thesis on the five NASA datasets

and the results obtained show the performance is better for the proposed algorithm.

The proposed variants of neural network architecture models, rule mining approaches and other evolutionary computing are developed using MATLAB R2013a (Version 8.1.0.604) environment and executed in a PC with Intel Core2 Duo Processor with 2.27GHz speed and 2.00 GB RAM. Simulations are conducted with the modeled variant neural network predictor models for achieving effective and efficient software defect prediction model for the NASA datasets and the simulated computed results prove that the proposed defect predictor models are better in accuracy than that of the earlier available defect predictor models in the literature.