Neural Network based Approach for Module Sequence Generation for Software Implementation

Manjot Kalsi¹, Janpreet Sigh²
¹M.Tech Research Scholar, Department of Computer Science & Engineering
²Assistant Professor, Department of Computer Science & Engineering
¹, ²Lovely Professional University, Phagwara, (Punjab), India
¹kalsimanjot@gmail.com, ²janpreet.singh@lpu.co.in

Abstract: For the detection of fault prone modules various predictive models can be used based on source code metrics as input for the classifiers [13]. The identification of metric subset for the enhancement of performance for predictive objective would not only improve the model but also provides strength to the structural properties of the modules. Unbalanced datasets also itself is a kind of difficulty for building predictive modeling. Unbalanced datasets are common in empirical software engineering as a majority of the modules are not likely to be faulty [8, 28]. We propose a method of applying search based metric selection and oversampling of NASA dataset [3]. Selection approach uses the weights of Neural Network to identify the sequence of implementation of software modules.

Keywords: Metric subset, Predictive models, unbalanced datasets, NASA dataset, Neural Network.

I. Introduction

The main aim of organization is to produce high quality products in a timely manner as well as cost effective manner [1]. To increase the utilization of available resources, when taking mitigating actions, such as code inspection, refactoring etc., the ability to identify potentially, referenced components would assist. Predictive models have been a focus of research in empirical studies of software systems [20].

Number of researchers work on this area of quantification of implementation and structural design of systems. Their study observes that feature selection is the process of identifying a subset of features that improves a model’s specific performance [5]. Many researchers suggest to use search based optimization with testing release date planning and cost estimation [31]. In this study we propose a Neural Network based algorithm as a search based feature selection strategy in order to find a subset of source code metrics that will generate an implementation sequence that enhances and simplifies a predictive model of software quality. To built predictive models for generating efficient sequence of modules for implementation, fit databases are required which contains instances of faulty and non-faulty modules. Preparation of balanced fit dataset is also always not possible when using industrial systems. Many modules in the fit dataset are actually imbalanced i.e. these exists a large difference between the number of fault prone modules and non faulty modules. This causes performance degradation of fault-proneness modules.

In conventional systems, the predictive accuracy of minority class is degraded since the prediction accuracy of the majority class is dominant in satisfying the objective function of model [25, 9]. In this paper we propose sampling method to balance the fit dataset in order to increase classifier performance as a classifier we took cohesion and coupling parameters of different modules under consideration.

II. Proposed Objectives

We propose the following steps as the objectives of the system to produce implementation sequence of different modules of software by predictive modeling.

- Selection of modules under consideration from KC1 dataset.
- To find the maximum value of cohesion and coupling parameters of selected modules to find the percentage of cohesion and coupling for each selected module.
- To find the collective percentage of cohesion and coupling for a single module.
To assign weights to the nodes of Neural Network by the values of collective percentage calculated for each module.

To find sequence of implementation of modules on the basis of weights assigned to the Neural Network as the final output of the system.

To find the performance measures of the modules sequence as SNR, RMSE, CPU time and accuracy.

### III. Methodology

The following diagram describes the flow of methodology used in the proposed system.

#### Fig1: Proposed Methodology

Some of the features of the software are selected to find the sequence of implementation of modules. In the proposed system, we use coupling and cohesion as parameters to take decision about the generation of module sequence. Neural Network helps to identify the exact module to be implemented as we are considering the percentage of different parameters to take decision about which module to implement next. The judgment is done on the basis of weights assigned to the nodes of the Neural Network.

#### IV. NASA Dataset

In this study we used dataset form NASA project available from the PROMISE repository. Project KC1 is used for storage management and processing of ground data. It was developed in C++ [33, 24]. A total of twenty one metrics calculated per method for KC1, are presented in the following table:

#### Table1: KC1 source code metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>loc blank</td>
<td>branch count</td>
</tr>
<tr>
<td>loc code and comments</td>
<td>loc comments</td>
</tr>
<tr>
<td>cyclomatic complexity</td>
<td>design complexity</td>
</tr>
<tr>
<td>essential complexity</td>
<td>loc executable</td>
</tr>
<tr>
<td>halstead content</td>
<td>halstead difficulty</td>
</tr>
<tr>
<td>halstead effort</td>
<td>halstead error est</td>
</tr>
<tr>
<td>halstead level</td>
<td>halstead length</td>
</tr>
<tr>
<td>halstead prog time</td>
<td>halstead volume</td>
</tr>
<tr>
<td>num operands</td>
<td>num operators</td>
</tr>
<tr>
<td>num unique operands</td>
<td>num unique operators</td>
</tr>
<tr>
<td>loc total</td>
<td></td>
</tr>
</tbody>
</table>

#### V. Conclusion

From the literature review, it has been observed that, after
deciding the architectural representation during the design phase while we are following some Model of software engineering, the choice of selecting a module from the architecture for coding and development is also very much important and we have to decide the sequence of selection of modules to go with the development. Any mistake during this selection process may leads to software crisis as if, after the implementation of some modules it is found that the section process was not good and we are not able to proceed further. So we have to start from the beginning. This may affect the overall cost of development, Time deadlines of the developers etc. So the selection sequence of development of modules is very important and somehow this aspect of software systems is not considered in the software implementation models. To perform the perfect selection of sequence of modules for development, we proposed a new technique with the help of Neural Networks, which helps to generate this sequence. The selection process depends on the properties of modules specified during design phase of software development cycle. These properties are coupling between the modules, Lack of Cohesion, etc. These properties will helps to decide the weights of Neural Network and then with the help of Neural Network generation of sequence of execution of modules is completed. For the performance evaluation of the proposed system, I will use KC1 data set of NASA. This data set consist 22 different properties related to 2000 different modules. SNR, RMSE, CPU time and Accuracy are the parameters under consideration.

References


[20] Naeem Seliya, Member, IEEE, and Taghi M. Khoshgoftaar, Member, IEEE software Quality Analysis of Unlabeled Program Modules With Semisupervised Clustering


