ABSTRACT

One of the most exceptionally important tasks in software development life cycle is to produce precise effort estimations and schedule. Developers may use effort estimate as input to project planning, budgeting, pricing policies and investment analyses. Missing or Incomplete data are exasperation to a certain extent everywhere but especially in effort estimation it is a greater hindrance. Over estimation of the software effort can result in losing the chance to win a bid and underestimation can lead to detrimental effect on the quality of software or monetary loss. Accuracy in effort estimation is decisive for developers and customers. For a developer it is required to do preplanning, budgeting, risk analysis and productivity assessment and customers requires estimation for taking place in contract negotiation process. The importance of effort estimation is seen on different phases of SDLC; in the initial phase it provides some insights on whether to proceed forward or not. Rough validation and progress of a project are done in the intermediate phase. At the completion phase cost estimates helps to assess productivity.

The major challenge in Software estimation is mainly concerned about quality of data. High degree of inaccuracy is mainly because of missing values. The common measure usually adopted in missing values is to ignore the whole project data or its feature. Accuracy of such data is termed as the
characteristics of the dataset. A great deal of uncertainty exists in such dataset. This makes the estimation imprecise and vague.

All the popular empirical estimation models require accurate inputs for estimation of effort. In other words, the accuracy of the output of these models is depends upon the accuracy in the input parameters. It is very difficult to obtain such accurate information at the early development stage of a software project. Controlling the huge uncertainty in existing estimation model is hard since it could not be modeled purely based on mathematics. These issues led to the introduction of variety of techniques to solve the relevant issues. Among the variety of techniques, imputation based models yield much attraction and focus of the researchers due to their enhanced performance.

Recently many machine learning algorithms were proposed and used as the imputation techniques for estimating the effort. In majority of the cases we find that the imputation models posses’ significance compared to other models in terms of performance. Among those models, Genetic Algorithm (GA) is stochastic especially in domains where a direct search method cannot reveal better outputs and of course in the large search spaces. They have an advantage to operate with incomplete data to extract significant rules. In this series, this research proposes a hybrid model of imputation technique, which holds better significance than other recent estimation models.
The objective of this proposed research is to formulate a comprehensive effort estimation model using Graphical and Genetic Approaches. This proposed hybrid effort estimation model is an extension to the well-known software effort estimation model, which produces accurate outputs even with incomplete and obscure data. The proposed model has two folds: the graphical segment offers a complete unique approach to devise a mere estimation scheme for effort estimation; later, the mere estimation scheme is being optimized by using genetic algorithm. The hybrid version of the proposed model gives better performance compared to existing models.

The performance of the proposed approach is validated in terms of standard performance factors and demonstrated the efficiency of the same. The ability of the COCOMO II and proposed estimation models are examined using standard performance validation measures like Magnitude of Relative Error (MRE), Magnitude of Absolute Error (MAE), Correlation, Root Mean Square (RMS), Mean Magnitude of Relative Error (MMRE), Relative Root Mean Square (RMSE) and PRED. The results attained offer the confidence over the proposed model and of course proffer the lead for different directions in the relevant field of research.