Abstract—In this study,Hierarchical Clustering approach is evaluated for Reusability Prediction of function based Software systems. Here, the metric based approach is used for prediction. Reusability value is expressed in the two linguistic values. Five Input metrics are used as Input and clusters are formed using Hierarchical Clustering, thereafter performance of the system is recorded.

Keywords—Software Reusability, Software Metrics, Function Based, Hierarchical Clustering.

I. INTRODUCTION

SOFTWARE reusability is an attribute that refers to the expected reuse potential of a software component [1]. Software reuse not only improves productivity but also has a positive impact on the quality and maintainability of software products. Software professionals have recognized reuse as a powerful means to potentially overcome the situation called as software crisis. The software development community is gradually drifting toward the promise of widespread software reuse, in which any new software system can be derived virtually from the existing systems. As a result, an increasing number of organizations are using software not just as all-inclusive applications, as in the past, but also as component parts of larger applications.

A great deal of research over the past several years has been devoted to the development of methodologies to create reusable software components and component libraries, where there is an additional cost involved to create a reusable component from scratch. That additional cost could be avoided by identifying and extracting reusable components from the already developed large inventory of existing systems. But the issue of how to identify good reusable components from existing systems has remained relatively unexplored. Our approach, for identification and evaluation of reusable software, is based on software models and metrics. As the exact relationship between the attributes of the reusability is difficult to establish so a Clustering based approach could serve as an economical, automatic tool to generate reusability ranking of software by formulating the relationship based on its training. Hence, in our study we will experiment with Hierarchical Clustering based approach for the reusability prediction of function based systems.

II. METHODOLOGY USED

Reusability evaluation System for function Based Software Components can be framed using following steps:

A. Selection and refinement of metrics

First of all, Selection and refinement of metrics targeting the quality of function based software system and perform parsing of the software system to generate the Meta information related to that Software. The metric of [2-5] are used and the metrics are as under:

The proposed five metrics for Function Oriented Paradigm are as follows [2-4]:

i) Cyclometric Complexity Using Mc Cabe’s Measure

i) Halstead Software Science Indicator

i) Regularity Metric

i) Reuse-Frequency Metric

i) Coupling Metric.

B. Calculating Metric Values

Calculate the metric values of the sampled software components.

C. Use Hierarchical Clustering Based Prediction System

There are two main methods of hierarchical clustering algorithm.

First method is agglomerative approach, where we start from the bottom where all the objects are and going up (bottom up approach) through merging of objects. We begin
with each individual objects and merge the two closest objects. The process is iterated until all objects are aggregated into a single group.

Second method is divisive approach (top down approach), where we start with assumption that all objects are group into a single group and then we split the group into two recursively until each group consists of a single object. One possible way to perform divisive approach is to first form a minimum spanning tree (e.g. using Kruskal algorithm) and then recursively (or iteratively) split the tree by the largest distance.

Step by step algorithm of agglomerative approach to compute hierarchical clustering is as follow [5]:
1. Convert object features to distance matrix.
2. Set each object as a cluster (thus if we have 6 objects, we will have 6 clusters in the beginning)
3. Iterate until number of cluster is 1
   a. Merge two closest clusters
   b. Update distance matrix

The flow chart of agglomerative hierarchical clustering algorithm is given below:

![Flowchart of Hierarchical Clustering Algorithm](image)

The given data is with five Input Attributes i.e. Coupling, Volume, Complexity, Regularity, Reuse_Frequency, and one Output attributes named as Reusability Level of the Software Component. Then hierarchical clustering based algorithm is implemented in Matlab 7.4.

The given data with five Input Attributes is applied to the hierarchical clustering algorithm and dendrogram is plotted as shown in figure 3. The dendrogram shows Indices of the components on the x-axis and distance between the components on the y-axis and the Accuracy percentage for the classification of the complements in Reusable and Non-reusable category is obtained as 85.3211%.

![Dendrogram showing the Hierarchical clustering of the Programs](image)

Deduce the results in terms of Accuracy of prediction after 10-fold cross validation.

### III. RESULTS & DISCUSSION

The function oriented dataset considered have the output attribute as Reusability value. The Reusability in the dataset is expressed in terms of two numeric labels i.e. Reusable and Non-Reusable. The Graphical representation of the count of the number of examples of certain reusability label is shown in the Figure 2.

![Bar-chart of Count of examples of the Reusability Output Attribute in the Dataset](image)

### IV. CONCLUSION

In this study, Hierarchical Clustering approach is evaluated for Reusability Prediction of function based Software systems. Here, the metric based approach is used for prediction.
Reusability value is expressed in the two linguistic values. Five Input metrics are used as Input and clusters are formed using Hierarchical Clustering, thereafter performance of the system is recorded. The proposed technique is showing Accuracy value approximately equal to 85.3211%, so it is satisfactory enough to use the Hierarchical clustering based technique for the identification of the object based reusable modules from the existing reservoir of software components.

The proposed approach is applied on the C based software modules/components and it can further be extended to the Artificial Intelligence (AI) based software components e.g. Prolog Language based software components. It can also be tried to calculate the fault-tolerance of the software components with help of the proposed metric framework.

REFERENCES


