Abstract

Parametric software estimation models rely on the availability of historical project databases from which estimation models are derived. In this case, a single mathematical model cannot properly capture the diverse nature of the projects under consideration. The use of a single mathematical model offers poor quality of adjustment, due to several factors, one of which consists on heterogeneity of data used to elaborate the model.

It is necessary to deal with this problem. The idea of splitting the project database in projects groups is the main motivation of this work. The members of these groups show a more homogeneous relationship between them. This task is automatically done using Artificial Intelligence techniques such us clustering algorithms, that divides data into segments of related projets.

A new estimation model is presented in this work, called the segmented parametric software estimation model, which produces a set of clusters made up each of them by many different projects, after the use of a clustering algorithm over the entire project database. The projects clusters obtained in this manner present more homogeneous characteristics than others not clusterized.

A mathematical model is given for each cluster get in this manner. This mathematical model consist of a parametric equation obtained by means of regression analysis. The quality of adjustment of this multi model is better than the single parametric model in the evaluations carried out.

The task can be carried out recursively if considered appropriate, getting even more

homogeneous subclusters in consecutive steps.

The straightforward clustering process over all the projects does not use previous expert knowledge that they could get over them. Using this principle a new proposal has been made, this consists on creating a partition of projets before the clustering process can be done. This partition process divide projects, using expert knowledge, in groups of similar characteristics based on the influence that some cost drivers have over the effort estimation. In this manner, the clustering process will be improved due to a more suitable framework.

The MMRE (*Mean Magnitude of Relative Error*) and PRED(l [2]) (*Prediction level*) measurements have been used to compare the two models, the segmented parametric model versus the not segmented parametric, to evaluate the accuracy and quality of adjustment of the technique proposed in this work. The segmented parametric software estimation model has provided better results than the not segmented model.

Keywords: clustering, segmentation, recursive segmentation, EM algorithm, software effort estimation, segmented parametric software estimation model, Software Engineering

---

[2] *Conte et. al (Conte et al. 1986) establish $l = 0,25$ or $l = 0,3$ as good values.*