International Journal of Future Generation Communication and Networking
Vol. 14, No. 1, (2021), pp. 1127-1135

# A Comparative Study on Estimating Project Combined with Different Clustering Techniques

Dr. M. Padmaja[1], Dr. D. Haritha[2],Dr.Susmitha Valli.Gogula[3]

[1]*Assistant Professor, Department of CSE*
[1]*GIT, GITAM (Deemed to be University)*
[1]*Visakhapatnam, 530045*
[2]*Professor and HOD, Department of CSE*
[2]*University College of Engineering, JNTUK*
[2]*Kakinada*
[3]*Professor, Department of CSE*
*MLRIT,*
*Hyderabad-500049*
[1]*e-mail: padmaja.madugula@gmail.com*
[2]*e-mail: harithadasari9@yahoo.com*
[3]*e-mail: susmitagv@gmail.com*

***Abstract***

*The challenge of the manager is accurately estimating the projects. The objective of the developer is to provide the product within specified type and reach to customer expectations. In order to this, management or development teams properly estimate the project in the areas of effort, cost and schedule. To achieve this, they are used suitable or an efficient method. Here, propose Grey Relational Analysis (GRA) and also combined this model with different clustering techniques. The interesting task of the developer is the selection of equal project from historical projects for the present project. From the historical data, we can choose nearer the project of our current project, but it can be difficult because it consists of many projects. So, the inventor will apply the clustering method to divide the data set into 'n' number of clusters, it can help to choose the closest project of the current project and also helps to estimate the effort in less time. Finally, projected method can be evaluated using some metrics.*

***Keywords:*** *Estimate the effort, GRA, COCOMO, MMRE, PRED, K-Means Clustering, EM Clustering*

## 1. Introduction

In the bidding process, to develop a project, clients can invite many companies to accept their constraints. In this, many competitors attended and check the customer's requirements or constraints and their capabilities. Which organization or company bid a minimum quote than the others, then that one will receive a project? Actually, here two things happened. First one, the company quotes a bid more than the other companies, it was loose the deal. The second one, company quotes a bid more than the client's budget; it was loss the profit [1]. In order to this, challenge a competition pressure to receive a project and develop or implement a project, deliver in-time and satisfy the customer constraints based on accurate estimation.

In the bidding process, project estimation and effort estimation of a project are more important tasks. Based on customer requirements, management can decide about which parameters are required to estimate the project. May be clients has a basic idea of a project, means they don't

have complete knowledge of the performance of a product. So the customer will provide relevant or irrelevant data. From this data, the development team can take the decision which data or parameters are required to develop a project. It is the challenging task for management team or development team.

In order to choose the nearer project, the analogy approach [2] can be applied. Mainly, the developer takes the decision with large data sets. It means, deciding on the most relevant project from the historical data to our relevant current project from larger dataset is to take more time. For this, the developer uses a clustering technique to know the most influential projects efficiently.

In this study, two types of clustering methods are chosen, such as K-Means clustering and Expectation-Maximization clustering [3-4]. These clustering techniques applied on large dataset to find a relevant clustering category in our current project with less time. These methods separately applied to the same data set (COCOMO 81) to choose required projects to develop a current project. After finding the relevant projects to the current project, then use suggested method (Grey Relational Analysis) for estimate the effort. Finally, compare the results in between these two clustering techniques and evaluate the method with help of metrics like MMRE, MdMRE and PRED.

Not only apply clustering methods, but also use the correlation method to choose influenced parameters from the data provided by the customer. Whatever data provided by the customer, that data is relevant or not to develop the product; it can be identified by the developer finally. Sometimes customer not aware of the working knowledge of a project with relevant features. So, the developer's task is to choose required features from all the data provided by the customer [5]. In order to this, use some methods for selection of necessary data. In this paper, use two methods, namely analysis of variance and Pearson correlation coefficient to selection of necessary data, new dataset was prepared this is known as reduced dataset.

Section 2 presents the relevant work, Section 3 describes about the methodology to estimate the effort and efficient methodology, Section 4 shows about how to select relevant features using ANOVA and PCC methods, Section 5 shows experimental outcomes, Section 6 presents the conclusion and also further work.

## 2. Literature Review

Effort estimation is the challenging task of the software environment to accept the project. Once the project can handle, that project is delivering the product within specified the constraints by the customer and that project can meet the customer expectations. Finally, the customer is satisfied that product and provides higher quality.

Kowsalya et al [6] has proposed an analogy method to pick the relevant projects from the existing database. Geeta Nagpal et al [7] has chosen regression technique with clustering on different datasets to estimate the effort.

Sun-Jen Huang et al [8] was applied grey relational method and it is also combined with genetic algorithm on different data sets for estimating the effort. Bindu Madhuri et al. [9-11] had used the grey method clustering to evaluate the website and compare the results with AHP.

Deng [12-13] introduced the grey system theory. It was applied to uncertain problems in various environments. K. H. Hsia and J. H. Wu [14] examines the history of grey relational analysis. Lin. Yi and Liu. Sifeng [15] describes the complete history of grey relational analysis from its background.

Qinbao Song et al [16] was applied the grey relational method with a few projects for effort estimation. Chao-Jung et al [17] have applied the grey relational analysis method on various weighted methods and evaluate the results.

M. Padmaja and Dr D. Haritha [18-19] have examined the working of the GRA method on few existing projects and compare the results with different methods and also combined this method with clustering applied on the available features and selected features.

M. Padmaja and Dr D. Haritha [20-22] combined the grey relational method with Taguchi to find which degree of variation is sufficient for the effort estimation of a project. And also compare this method with optimization technique, such as Meta heuristic technique with and without clustering methods.

## 3. Methodology

The GRA algorithm was applied to determine the relevant projects of the current project. This paper aims at producing the good estimation of our selected project. The projected method is applied to the COCOMO81 dataset as presented in Figure 1.

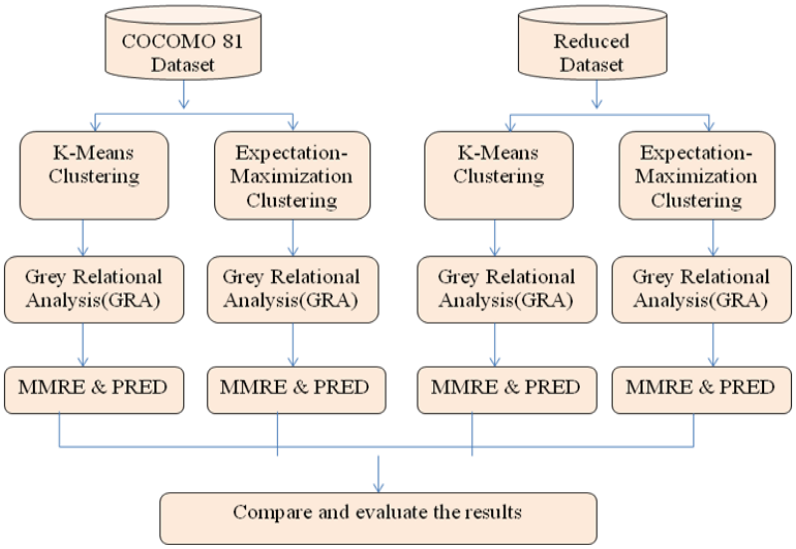


**Fig 1. The structural design of the projected technique**

Here, the projected technique is applied on COCOMO81 dataset to predict the effort. This dataset has a number of parameters, so we choose only required features based on our current project. For this, choose two methods to find maximum influential parameters, such as ANOVA and PCC. With these methods found only eight parameters are required to estimate the effort. With these eight parameters, this is known as reduced data set. The existing database has all parameters provided by the customer.

Now, on these two types of data sets (original and reduced datasets) apply K-Means and EM techniques to choose a referential project of the current project. Here, we apply two clustering techniques such as K-Means and Expectation-Maximization techniques to divide into clusters chosen by the developer. These dividing categories help to identify a similar project to our current project to estimate the project. Then decide which clustering method is to find a similar project of the current project. So, in this paper compare the results with these clustering methods using measures.

After finding out the effort estimation of a project and then evaluate the efficiency of the proposed method using metrics. Here, these metrics are chosen like MMRE, MdMRE and PRED.

## 4. Selection of required features

In this work, the COCOMO81 dataset was selected it consists of several parameters and many (63) projects. The estimation of the project is mainly depends on the selection of a nearer project from the background projects. Based on this, on the available data, choose required parameters or features by using one-way Analysis of analysis and Pearson correlation coefficient methods. With these required features helpful to predict the effort accurately. The influential parameters are found [20], the required parameters are presented in the below table.

**Table 1. Most influential features using ANOVA and PCC**

| S.No | Features |
|---|---|
| 1 | Rely |
| 2 | Data |
| 3 | Turn |
| 4 | Acap |
| 5 | Aexp |
| 6 | Pcap |
| 7 | Lexp |
| 8 | Modp |

## 5. Experimental Results

In this paper, the GRA method was applied to COCOMO81 dataset and reduced dataset with clustering methods (K-Means and EM). These methods are applied on that data is divided into clusters. In this process, 'k' value is chosen '3', so data is divided into three clusters. Cluster details are presented in Table 2.

**Table 2. Clusters are obtained using K-Means**

| | COCOMO 81 (Original Data) | COCOMO81 (Reduced Data) |
|---|---|---|
| Cluster No (C.No) | Number of projects | Number of projects |
| C0 | 6 | 26 |
| C1 | 40 | 30 |
| C2 | 17 | 7 |
| Total projects | 63 | 63 |

The expectation maximization method is applied to two types of datasets, that data is divided into three clusters as shown in below Table 3.

**Table 3. Clusters are obtained using Expectation-Maximization**

| Data Set | COCOMO81 | COCOMO81 |
|---|---|---|

| | (Original Data) | (Reduced Data) |
|---|---|---|
| **Cluster No (C.No)** | **Number of projects** | **Number of projects** |
| C0 | 29 | 12 |
| C1 | 17 | 26 |
| C2 | 17 | 25 |
| Total projects | 63 | 63 |

On the clustered datasets of both methods, to forecast the effort, the GRA method was applied. After finding it, the efficiency of the method is calculated using the mentioned metrics.

**Table 4. Assessment results on original data with k-means**

| | Number of projects | MMRE | MdMRE | PRED |
|---|---|---|---|---|
| C0 | 6 | 0.3032 | 0.2791 | 0.3333 |
| C1 | 40 | 0.2713 | 0.2124 | 0.575 |
| C2 | 17 | 0.3036 | 0.2137 | 0.5882 |
| Average | Total (63) | 0.2927 | 0.2351 | 0.4988 |

**Table 5. Assessment results on reduced data with k-means**

| | Number of projects | MMRE | MdMRE | PRED |
|---|---|---|---|---|
| C0 | 26 | 0.1313 | 0.1179 | 0.9230 |
| C1 | 30 | 0.3148 | 0.2073 | 0.5333 |
| C2 | 7 | 0.2516 | 0.2378 | 0.5714 |
| Average | Total (63) | 0.2325 | 0.1877 | 0.6759 |

Here, the proposed method applied with k-means clustering on both data types, the results are as shown in Tables 4 and 5.

**Table 6. Evaluation criteria on original data with EM clustering**

| | Number of projects | MMRE | MdMRE | PRED |
|---|---|---|---|---|
| C0 | 29 | 0.24665 | 0.20954 | 0.45262 |
| C1 | 17 | 0.32872 | 0.32108 | 0.75864 |
| C2 | 17 | 0.22806 | 0.18025 | 0.42672 |
| Average | Total (63) | 0.26781 | 0.23695 | 0.54599 |

**Table 7. Evaluation criteria on reduced data with EM clustering**

| | Number of projects | MMRE | MdMRE | PRED |
|---|---|---|---|---|
| C0 | 12 | 0.1745 | 0.1426 | 0.9546 |

| | | | | |
|---|---|---|---|---|
| C1 | 26 | 0.1893 | 0.1032 | 0.6547 |
| C2 | 25 | 0.2717 | 0.1784 | 0.5986 |
| Average | Total (63) | 0.21185 | 0.14140 | 0.73597 |

And also apply another integration method, such as GRA and the EM method on both data sets. The results are obtained and shown in above Tables 6 and 7.

Finally, compare the results with metrics to decide which clustering method is suitable with and without influential parameters. From all results, to observe with metrics is shown in below Tables and shown in below Figures.

**Table 8. Compare the error rate (MMRE) on clustering techniques**

| | **MMRE (Original Data)** | **MMRE (Reduced Data)** |
|---|---|---|
| K-Means | 0.2927 | 0.2325 |
| Expectation Maximization (EM) | 0.2678 | 0.2118 |

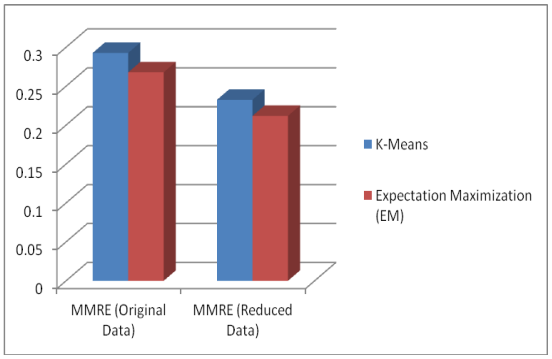


**Fig 2. Compare the results with MMRE**

In order to evaluate the proposed method with two clustering techniques on both types of datasets observe with MMRE metric. The results are shown in above Table 8 and as depicted in Figure 2.

**Table 9. Compare the results with PRED on clustering techniques**

| | **PRED (Original Data)** | **PRED (Reduced Data)** |
|---|---|---|
| K-Means | 0.4988 | 0.6759 |
| Expectation Maximization (EM) | 0.5234 | 0.7012 |

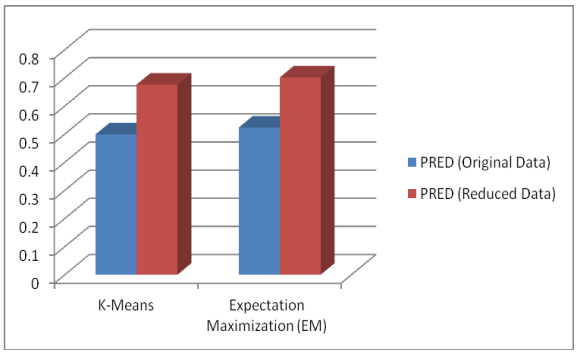


**Fig 3. Compare the results with PRED**

In order to evaluate the proposed method with two clustering techniques on both types of datasets observe with PRED metric. The results are shown in above Table 9, and as depicted in Figure 3.

**Table 10. Evaluation of the results**

| | **COCOMO81 (Original Dataset)** | | **COCOMO81 (Reduced Dataset)** | |
|---|---|---|---|---|
| | K-Means Clustering | EM Clustering | K-Means Clustering | EM Clustering |
| MMRE | 0.2927 | 0.2678 | 0.2325 | 0.2118 |
| MdMRE | 0.2351 | 0.2370 | 0.1877 | 0.1414 |
| PRED | 0.4988 | 0.5234 | 0.6759 | 0.7012 |

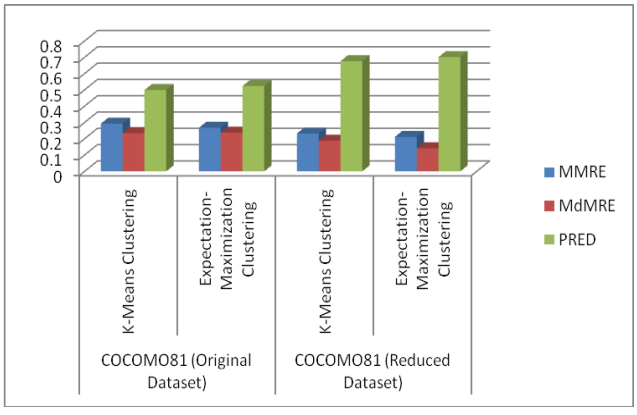


**Fig 4.Evaluation criteria over both data sets**

Finally, observe in all cases to decide the suitable method in case of forecast the effort for a particular project. For all experiments, the proposed method was suitable with most influential parameters with Expectation-Maximization than k-means clustering as shown in above Figure 4. Not only derive the method with mentioned criteria, but also compared with existing methods as also presented in below Table 11.

**Table 11. Compare and evaluate the results with other techniques**

| | MMRE | PRED |
|---|---|---|
| EM & GRA | 0.2118 | 0.7012 |
| K-Means & GRA | 0.2325 | 0.6759 |
| K-Means & PSO | 0.2303 | 0.55 |
| Original COCOMO by Human experts | 0.26 | 0.54 |
| ANN | 0.37 | 0.40 |
| FNN | 0.22 | 0.75 |
| FGRA | 0.2323 | 0.667 |

From the experimental results the proposed method also suitable when compare with other techniques in prescribed the criteria and also proven the criteria such as the combination of GRA and EM proven with a minimum error rate with required parameters only. This method was proven on the criteria of MMRE and PRED metrics are applied to the parameters to evaluate the method.

## 6. Conclusion and Future work

The development of the projects, before handling the projects, the estimation of the effort is a challenging task within the project management sector. The goal of the developer is developing the project with quality and satisfies the clients. In this paper, the projected technique is combined with EM and k-means clustering to get a reference project from the historical data and also forecast the effort. The experimental results were proven with influential parameters using Expectation – Maximization technique than k-means clustering. So, the Grey Relational Analysis method works well with EM clustering on influencing parameters and then compare with K-Means clustering. And also verify with other techniques on the same data set. Finally, the proposed method proves better results than with others. Further extensions of this work can be done by combining optimization techniques and different methods to find suitable parameters and applied to different data sets.

## References

[1] Swapna Kishore and Rajesh Naik, "Software Requirements and Estimation", Tata McGraw Hill.

[2] Martin Shepperd, Chris Schofield and Barbara Kitchenham, "Effort Estimation using Analogy", IEEE, **(2009**).

[3] Dwivayani Sentosa, Budi Susetyo, Utami Dyah Syafitri, Sutoro, "Applied Expectation Maximization (EM) Clustering for Local Variety Corn", International Journal of Scientific & Engineering Research, Volume 8, Issue 1, **(January-2017)**, pp. 1189-1192.

[4] G. Chamundeswari, Prof. G. Pardasaradhi Varma and Prof. Ch. Satyanarayana, "An Experimental Analysis of K-means Using Matlab", International Journal of Engineering Research & Technology (IJERT), Vol.1, Issue.5, **(July 2012).**

[5] Jin-Cherng Lin, Yueh-Ting Lin, Han-Yuan Tzeng and Yan-Chin Wang, "Using Computing Intelligence Techniques to Estimate Software Effort", International Journal of Software Engineering & Applications (IJSEA), Vol.4, No.1, **(January 2013).**

[6] M. Kowsalya, H. Oormila Devi and N. Shiva Kumar, "Analogy Based Software Project Effort Estimation Using Projects Clustering", International Journal of Scientific and Research Publications, Vol.7, Issue.4, ISSN.2250-3153**, (April 2017)**, pp. 320-325.

[7] Geeta Nagpal, Moin Uddin and Arvinder Kaur, "Analyzing Software Effort Estimation using k-means Clustered Regression Approach", ACM SIGSOFT Software Engineering Notes, Vol.38, No.1, **(January 2013).**

[8] Sun-Jen Huang, Nan-Hsing Chiu and Li-Wei Chen, "Integration of the grey relational analysis with genetic algorithm for software effort estimation", European Journal of Operational Research, **(2008),** pp 898–909.

[9] Bindu Madhuri Ch, Padmaja M, Srinivasa Rao T, Anand Chandulal J, "Evaluating web site based on grey clustering theory combined with AHP", International Journal of Engineering and Technology, Vol. 2, No. 2, **(2010).**

[10] Bindu Madhuri Ch, Padmaja M, "Evaluating Web Sites Based on GHAP", International Journal of Computer Science and Engineering (IJCSE), Vol. 02, No. 03, **(2010)**, pp: 674-679.

[11] Bindu Madhuri Ch, Padmaja M, "Selection of Best Web Site by Applying COPRAS-G Method", International Journal of Computer Science and Information Technologies, (IJCSIT), Vol. 1, Issue. 2, **(2010),** pp: 138-146.

[12] Deng. J, "Introduction to grey system, Journal of Grey System", Vol.1 No.1, **(1989),** pp. 1-24.

[13] Deng Julong, "Introduction to Grey System Theory", the journal of grey system, **(1989),** pp 1-24.

[14] K. H. Hsia and J. H. Wu, "A study on the data preprocessing in grey relational analysis", The Journal of Grey System, Vol. 9 (1), **(1997),** pp. 47–53.

[15] Lin. Yi, Liu. Sifeng, "A Historical Introduction to Grey Systems Theory", IEEE International Conference on Systems, Man and Cybernetics, **(2004).**

[16] Qinbao Song, Martin Shepperd and Carolyn Mair, "Using Grey Relational Analysis to Predict Software Effort with Small Data Sets", 11th IEEE International Software Metrics Symposium - METRICS **(2005).**

[17] Chao-Jung Hsu and Chin-Yu Huang, "Comparison of weighted grey relational analysis for software effort estimation", Software Qual J, Springer Science+Business Media, LLC **(2010).**

[18] M. Padmaja, Dr D. Haritha, "Software Effort Estimation using Grey Relational Analysis", MECS in International Journal of Information Technology and Computer Science, Vol. 9, No. 5, **(May 2017),** pp: 52-60.

[19] M. Padmaja, Dr D. Haritha, "Software Effort Estimation using Meta Heuristic Algorithm", International Journal of Advanced Research in Computer Science, Vol: 8, Issue: 5, **(May-June 2017),** pp. 196-201.

[20] M. Padmaja, Dr D. Haritha, "Software Effort Estimation using Grey Relational Analysis with K-Means Clustering", 4th International Conference on Information System Design and Intelligent Applications, 15th - 17th, June, 2017, Duy Tan University, 3 QuangTrung, Da Nang, VietNam, Springer, AISC Series, **(Jan 2018)**.

[21] M. Padmaja, Dr D. Haritha, "Optimization of Process Parameters Using Grey-Taguchi Method for Software Effort Estimation of Software Project", International Journal of Image, Graphics and Signal Processing, Vol: 10, No: 9, **(Sep 2018),** pp. 10-16.

[22] M. Padmaja, Dr D. Haritha, "A heuristic effort estimation method using Bat algorithm through clustering", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-8 Issue-9, **(July 2019),** pp. 1536 – 1541.