The Role of Key Elements in Software Development Using Cluster Analysis

Dr.Seetharam .K.
Professor , Dept of CSE/IT
MeRITS, Udayagiri 524226
Research Scholar Satyabama
Chennai

Chandrakanth Pujari
Asst.Prof, Dept of MCA
Dr.AIT, Bangalore -64
Research Scholar Satyabama
Chennai

ABSTRACT

The reliability is one very important parameter of applications software[1,6,9].The most straight restriction in most software reliability models is the assumption of statistical independence among successive software factors considered. Any measurement requires the set of elements associated with the associated process. Here the seven factors considered are size, effort, duration, S1 (customer participation), S2 (staff availability), S3(standards use) and S4(methods use)[4,13,14] Here the paper discusses Qualitative/quantitative measurement of software using cluster analysis. In this paper four different cases are carried out. First analysis with size as predominant factor, second analysis with effort as predominant factor.[3,10] third analysis with duration as predominant factor, finally including all the three associated in the list of seven factors with software reliability performance.

KEYWORDS:

Software reliability, cluster, fuzzy logic, size, efforts, duration

1.INTRODUCTION:

Cluster analysis is used to explain the Cophenetic correlation coefficient for the hierarchical cluster tree representation. Here three different groups are considered with each group of five element variables4. The variables are related to software development. Here 50 validated projects data are used. Some of the outer values are excluded and remaining data are used for software performance measurement analysis[17].

2. ANALYSIS:

2.1 From the software projects validated data, Three different groups are considered. The predictor variables are
1) Application size, S1,S2, S3 & S4
2) Duration,. S1, S2, S3 & S4
3) Effort, S1, S2, S3 & S4
4) Size, effort, duration, S1, S2, S3 & S4 where
S1=Customer participation
S2=Staff availability
S3= Standard Use
S4=Methods use

Five different levels are identified (see appendix) for each variables considered above separately by means of fuzzy logic[2,8]. This is necessary to make it a measurable variable. Four analysis are carried out. The variables are standardized using normal distribution principles. Following five different Pairwise distances between observations are considered.[15,16]
Euclidean distance (Euclidean),
Standard.Euclidean distance (seuclidean),
City Block distance (cityblock),
Mahalanobis distance (mahalanobis),
Minkowski – distance (minkowski).
Linkage to create a hierarchical cluster tree using five algorithm chosen.
'single' --- nearest distance,
'complete' --- furthest distance,
'average' --- group average distance,
'weighted' --- weighted average distance,
'ward' --- inner squared distance
Cophenet: Cophenetic correlation coefficient (CCC) is used for measuring hierarchical cluster tree correlation coefficient[7,11,12]

2.2 Size, S1, S2, S3 & S4:

Taking Size, S1, S2, S3 and S4 as arguments euclidean YS1, seuclidean YS2, cityblock YS3, mahalanobis YS4, minkowski YS5 are computed. Five clusters z11, z12, z13, z14 and z15 are generated taking Euclidean distance YS1 as base with five different methods. The Cophenetic Correlation coefficient (CCC) of these five clusters with the base YS1 is computed. Typical five clusters are shown in figure. Similar process is repeated taking YS2, YS3, and YS4 as base distances and CCC with their respective five set clusters is computed. This is shown in Size_CCC table.

2.3 Duration, S1, S2, S3, S4:

Taking Duration ,S1,S2,S3 and S4 as arguments YD1,YD2,YD3,YD4, YD5 are calculated. Five clusters are generated taking YD1 as base with five different methods. CCC of base YD1 with the five clusters is computed. Similar process is repeated taking YD2,YD3 and YD4 as base distance. Duration_CCC table is shown.

2.4 Effort, S1, S2, S3, S4:

Taking Effort, S1, S2, S3 and S4 as arguments YE1, YE2,YE3, and YE4 are calculated. Five clusters are generated taking YE1 as base with five different methods. CCC of base YE1 with the five clusters is computed. Similar process is repeated taking YE2,YE3,YE4 and YE5 as base distance. Effort_CCC table is shown.

3.0 CONCLUSION

Three different group cases are in detail,

- Cluster analysis is done using a covariance matrix,
- Five different Pairwise distances between observations,
- Hierarchical cluster tree used by five algorithms,
- Cophenetic correlation coefficient table is calculated.

3.1Details of variables:

S1: customer participation: how actively customer took part in development work:
1 = very low; none
2 = low; passive; client defined or approved <30% of all functions
3 = normal; client defined and approved 30-70% of all functions
4 = high; active; client defined and approved all of most important functions, and over 70% of others
5 = very high; client participated very actively, thus most functions were slightly volatile and changes had to be made.

S2: staff availability: availability of software personnel during project:
1 = very low; big problems with key personnel availability; lots of simultaneous customer and maintenance responsibilities; special know-hold required
2 = low; personnel involved in some other simultaneous projects and/or maintenance responsibilities
3 = normal; key members involve in only one other project
4 = high; project members involved almost full-time
5 = very high; qualified personnel available when needed full-time participation

S3: standards use: level and use if standards:
1 = very low; standards developed during project
2 = low; some standards, but not familiar ones; more must be developed for some tasks
3 = nominal; generally known standards applied in environment before; some tailoring needed
4 = high detailed standards applied in same environment for some time
5 = very high; stable and detailed to team; use controlled

S4: method use: level and use of methods: meeting; used by individuals
2 = low; use beginning; traditional concepts employed(stru
1 = very low;no modern design methods; mostly ctural analysis and design, top-down design etc).
3 = nominal; generally known methods used
4 = High Methods integrated in detail
and most activities are covered.
Support Existed. Used by everyone.
5 = Very high methods used during
entire life cycle.

4.0 REFERENCES:

[1] Conte, S.,Dunsmore, H.and Shen,V, *Software Engineering Metrics and models*, Benjamin Cummings, Menlo Park CA, 1986.

[2] Guanrong Chen and etal. *Introduction to fuzzy sets, fuzzy logic and fuzzy control* systems. CRC Press, 2000.

[3] Daniel T Larose *Data Mining: methods and models*. WIley interscience 2006.

[4] Fergal Mc Caffery, Philip S Taylor Gerry Coleman *Adept: A Unified Assessment method for small software companies*, Jan//Feb 2007 IEEE Software 24-31.

[5] E.M.Hall Managing risk: *Methods for software system Development* Addson-Wesley;1998.

[6] John D Musa *Software Reliability* Engineering McGraw Hill 1999.

[7] Joseph Hair, Rolph Anderson , Ronald Tatham , and William Black, *Multivariate Data Analysis*, 4th Ed, Prentice Hall , Upper Saddle River, NJ, 1995.

[8] Luigi Cerulo, Raffaelo Esposito, Maria Tortorella, Luigi Troiano RCOST –Research Centre on Software Technology Italy *Supporting software Evolution by using Fuzzy logic* 7th International IEEE workshop proceedings on Principles of software Evolution (IWPSE'04).

[9] Musa JD,*A Theory of software reliability and its application*, IEEE Trans Software Eng 1975 ;Se-1(3):312-27.

[10] Marcio Rodrigo Braz *Software Effort estimation based on use cases*, IEEE 30th International conference Proceedings COMPSAC'06 2006.

[11]Pace and Ronald Berry, sparse spatial auto regressions, *Statistics and probability Letters*, Vol 33, No 3, pp 291-297, May 5 1997Data set available from StatLib.

[12] Richard A Johnson and Dean Wichern , *Applied multivariate statistical Analysis,*
Prentice Hall, Upper Saddle River, NJ, 1998.

[13]Seetharam.k, Chandrakanth G Pujari, *Ranking of Tools use, software logical complexity, Requirement volatility, Quality requirements, Efficiency requirements in software development* IEEE Explorer Journal 2008 pp1608-1616.

[14]Seetharam.K,Chandrakanth G Pujari, *Factor analysis in the software development and refinement of reliability* Communicated.

[15] U.S.Census,Bureau, Urban area criteria for census 2000, *Federal Register*,Vol 67,No 51,March 15 2002,;

[16]Tsun Chowe, Dac-Buu Cao *A survey study of critical success factors in agile software projects* Journal of system and software 81 (2008) 961-971.

[17] Xuemei Zhang, Hoang Pham *An analysis of factors affecting software reliability* Journal of system and software 50 (2000) 43-56.