Abstract

Software effort estimation requires high accuracy, but accurate estimations are difficult to achieve. Increasingly, data mining is used to improve an organization's software process quality, e.g., the accuracy of effort estimations. There are a large number of different method combinations for software effort estimation, selecting the most suitable combination becomes the subject of research in this paper. In this study, three simple preprocessors are taken (none, norm, log) and effort is measured using COCOMO model. Then results obtained from different preprocessors are compared and norm preprocessor proves to be more accurate as compared to other preprocessors.

References

- Sehra, S. K., Brar, Y. S., Kaur, N., "Soft Computing Techniques for Software project Effort estimation," International Journal of Advanced Computer and Mathematical Sciences, Vol. 2, Issue 3, November 2011, ISSN 2230-9624, PP. 160-167.
Effect of Data Preprocessing on Software Effort Estimation

- Srivastava, D. K., Chauhan, D. S., & Singh, R. (2012, January), "VRS Model: A Model for Estimation of Efforts and Time Duration in Development of IVR Software System". International Journal of Software Engineering (IJSE), Vol. 5 No. 1, pp. 27-46.
- Keung, J., Kocaguneli, E., & Menzies, T. (2011, February), "A Ranking Stability Indicator for Selecting the Best Effort Estimator in Software Cost Estimation". Springer Science+Business Media, pp. 1 - 19.
- Saini, K. R., & Singh, P. (2011). "Data Mining Techniques and Tools for Knowledge Discovery in Agricultural Datasets". (A. Arora, P. K. Malhotra, S. Marwah, & A. Bhardwaj, Eds.) Division of Computer Applications Indian Agricultural Statistics Research Institute (ICAR).
- http://en.wikipedia.org/wiki/COCOMO. (n. d.).
- Dejaeger, K., Verbeke, W., Martens, D., & Baesens, B. (2012, March-April), "Data Mining Techniques for Software Effort Estimation: A Comparative Study". IEEE Transactions on Software Engineering, Volume 38, no. 2, pp. 375-397.
- Ji, S. H., et. al, "Data preprocessing method for cost estimation of building projects". 27th International Symposium on Automation and Robotics in Construction (ISARC 2010)
- Dongare Y. V, et al, "RDBNorma: A semi-automated tool for relational database schema normalization up to third normal form". International Journal of Database Management Systems (IJDMS), Vol. 3, No. 1, February 2011.
- Murtaza, N. And Sattar, A. R. et al. (2010), "Enhancing the Software Effort Estimation using Outlier Elimination Methods". Pak. j. life soc. Sci. (2010), 8(1): 54-5.

- Nagpal, G. and Uddin, M. et al. (2012), "A Comparative Study of Estimation by Analogy using Data Mining Techniques". J Inf Process Systg, Vol. 8, No. 4
  - MATLAB®Documentation, http://www.mathworks.com/help/techdoc/
  - Kaur, N. and Kaur, J. (2012) "Efficient k-means Clustering Algorithm Using Ranking Method in Data Mining". ISSN: 2278 – 1323 International Journal of Advanced Research in Computer Engineering & Technology Volume 1, Issue 3.
  - M. V. Deshpande and S. G. Bhirud, "Analysis of Combining Software Estimation Techniques". International Journal of Computer Applications (0975 – 8887) Volume 5 – No. 3
  - Y. Singh, K. K. Aggarwal. Software Engineering Third edition, New Age International Publisher Limited New Delhi.
  - Ziauddin, Khan, K. Z., Tipu, S. K., & Zia, S. (2012). "An Intelligent Software Effort Estimation System". Journal of Expert Systems (JES), 1(4), 91 - 98.

Index Terms

Computer Science
Software Engineering
Effect of Data Preprocessing on Software Effort Estimation

Keywords
Software effort estimation  Data preprocessing  COCOMO Model  Kilo Line of Code (KLOC)