

{tag}  
Techniques

{/tag}

Evolutionary Computation for Optimization  
© 2010 by IJCA Journal  
Number 2 -

Article 6  
Year of Publication: 2010

Authors:

Dipak V. Patil

R. S. Bichkar

10.5120/1537-140

{bibtex}spe140t.bib{/bibtex}

## Abstract

Missing data is one of the major issues in data mining and pattern recognition. The knowledge contains in attributes with missing data values are important in improving decision-making process of an organization. The learning process on each instance is necessary as it may contain some exceptional knowledge. There are various methods to handle missing data in decision tree learning. The proposed imputation algorithm is based on the genetic algorithm that uses domain values for that attribute as pool of solutions. Survival of the fittest is the basis of genetic algorithm. The fitness function is classification accuracy of an instance with imputed value on the decision tree. The global search technique used in genetic algorithm is expected to help to get optimal solution.

## Reference

- Little R. J. and Rubin D. B. 1987. Statistical Analysis with Missing Data. John Wiley and

Sons, New York.

- Schafer J. L. and Graham J. W. 2002. Missing data: our view of the state of the art *Psychol. Methods* 7(2), 147–177.
- Kuligowski R. J. & Barros A. P. 1998. Using artificial neural Networks to estimate missing rainfall data. *Journal AWRA* 34(6), 14.
- Brockmeier L. L., Kromrey J. D. and Hines C. V., 1998. Systematically Missing Data and Multiple Regression Analysis: An Empirical Comparison of Deletion and Imputation Techniques. *Multiple Linear Regression Viewpoints*, Vol. 25, 20-39.
- Abebe A. J., Solomatine D. P. & Venneker R. G. W. 2000. Application of adaptive fuzzy rule-based models for reconstruction of missing precipitation events. *Hydrological Sciences Journal*.45 (3), 425–436.
- Sinharay S., Stern H.S. and Russell D. 2001. The use of multiple imputations for the analysis of missing data. *Psychological Methods* Vol.4: 317–329.
- Khalil K., Panu M. and Lennox W. C. 2001. Groups and neural networks based stream flow data infilling procedures. *Journal of Hydrology*, 241, 153–176.
- Bhattacharya B., Shrestha D. L. & Solomatine D. P. 2003. Neural networks in reconstructing missing wave data in dimentation modeling. In the Proceedings of 30th IAHR Congress, Thessaloniki, Greece Congress, August 24-29 2003 Thessaloniki, Greece.
- Fessant F. & Midenet, S. 2002. Self-organizing map for data imputation and correction in surveys. *Neural Comput. Appl.* 10, 300–310.
- Musil C. M., Warner C. B., Yobas P. K. & Jones S. L. 2002. A comparison of imputation techniques for handling missing data. *Weston Journal of Nursing Research* 24(7), 815–829.
- Junninen H., Niska H., Tuppurainen K., Ruuskanen J. & Kolehmainen M. 2004. Methods for imputation of missing values in air quality data sets. *Atoms. Environ.* 38, 2895–2907.
- M. Subasi, E. Subasi and P.L. hammer, 2009. New Imputation Method for Incomplete Binary Data, *Rutcor Research Report*, August 2009.
- Amman Mohammad Kalteh & Peder Hjorth, 2009. Imputation of Missing values in precipitation-runoff process database. *Journal of Hydrology research*.40.4, pages 420—432.
- Papagelis A. and Kalles D. 2000. GAtree: Genetically Evolved Decision Trees, *Proceedings 12th International Conference on Tools with Artificial Intelligence* 13-15 November 2000 pages 203-206.
- Rajasekaran G.A., Vijayalakshmi Pai, 2004. *Neural Networks Fuzzy Logic and Genetic Algorithms Synthesis and Applications*, Prentice-Hall of India.
- Goldberg D. 1999. *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley.
- Quinlan J. R. 1993. *C4.5: Programs for machine learning*. Morgan Kaufman, San Mateo.
- Salvatore Ruggieri, 2002. Efficient C4.5, *IEEE Transaction On Knowledge and Data Engineering*, Vol. 14, No. 2 March/April.
- Endou T. and Qiangfu Zhao, 2002. Generation of comprehensible decision trees through evolution of training data, *Proceedings of the 2002 Congress on Evolutionary Computation*, 2002. Volume 2, 12-17 May.
- Quinlan J. R., 1990. Decision Trees and Decision making *IEEE Transaction On Systems, Man, And Cybernetics* vol. 20, No. 2, March/April.

- Zhiwei Fu, Fannie Mae, 2001. A Computational Study of Using Genetic Algorithms to Develop Intelligent Decision Trees, Proceedings of the 2001 IEEE congress on evolutionary Computation.
- Newman D.J. & Hettich S. & Blake, C.L. & Merz C.J. UCI Repository of machine learning databases [<http://www.ics.uci.edu/>].
- [www.cs.waikato.ac.nz/ml/weka](http://www.cs.waikato.ac.nz/ml/weka).

### Index Terms

Computation                      Computer Science                      Evolutionary

### Key words

algorithm                      missing data                      genetic  
decision tree