Abstract

Data mining is for new pattern to discover. Data mining is having major functionalities: classification, clustering, prediction and association. Classification is done from the root node to the leaf node of the decision tree. Decision tree can handle both continuous and categorical data. The classified output through decision tree is more under stable and accurate. In this research work, Comparison is made between ID3, C4.5 and C5.0 and after that Implementation of system is done. The new system gives more accurate and efficient output with less complexity. The system performs feature selection, cross validation, reduced error pruning and model complexity along with classification. The implemented system supports high accuracy, good speed and low memory usage. The memory used by the system, is low compare to other classifiers as the rules generated by this system is less. The major issues concerning data mining in large databases are efficiency and scalability. While in case of high dimensional data, feature selection is the technique for removing irrelevant data. It reduces the attribute space of a feature set. More reliable estimation of prediction is done by f-fold – cross - validation. The error rate of a classifier produced from all the cases is estimated as the ratio of the total number of errors on the hold-out cases to the total number of cases. By increasing the
model complexity, accuracy of the classification is increases. Overfitting is again major problem of decision tree. The system has also facility to do post pruning that is through reduced error pruning technique. Using this proposed system; Accuracy is gained and classification error rate is reduced compare to the existing system.

**References**

- TomM. Mitchel, McGrawHil, Decision Tree Learning, Lecture slides for textbook Machine Learning, , 197
- Zuleyka Díaz Martinez, José Fernández Menéndez, Mª Jesús Segovia Vargas, See5 Algorithm versus Discriminant Analysis, Spain.
- Xindong Wu • Vipin Kumar • J. Ross Quinlan • Joydeep Ghosh • Qiang Yang • Hiroshi Motoda • Geoffrey J. McLachlan • Angus Ng • Bing Liu • Philip S. Yu • Zhi-Hua Zhou • Michael Steinbach • David J. Hand • Dan Steinberg, Top 10 algorithms in data mining, © Springer-Verlag London Limited 2007
- Decision Trees for Business Intelligence and Data Mining: Using SAS Enterprise Miner &quot;Decision Trees— What Are They?&quot;
- Thair Nu Phyu, "Survey of Classification Techniques in Data Mining&quot;,
International MultiConference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, March 18 - 20, 2009, Hong Kong
- S. B. Kotsiantis, Department of Computer Science and Technology &quot;Supervised Machine Learning: A Review of Classification Techniques&quot; - , University of Peloponnese, Greece
- Classification: basic Concepts, Desision Tree, and model evalution
- Terri Oda , Data Mining Project, April 14, 2008
- Matthew N. Anyanwu manyanwu, Sajjan G. Shiva ssviva, Comparative Analysis of Serial Decision Tree Classification Algorithms
- Maria Simi , Decision tree learning
- Osmar R. Zaíane, 1999, Introduction to Data Mining, University of Alberta
- J. R. B. COCKETT, J. A. HERRERA, Decision tree reduction, University of Tennessee, Knoxville, Tennessee
- Hendrik Blockeel , Jan Struyf, Efficient Algorithms for Decision Tree Cross-validation, Department of Computer Science, Katholieke Universiteit Leuven, Belgium
- S. Rasoul Safavian and David Landgrebe, A Survey of Decision Tree Classifier Methodology, School of Electrical Engineering Purdue University, West Lafayette
- Floriana Esposito, Donato Malerba, and Giovanni Semeraro, A Comparative Analysis of Methods for Pruning Decision Trees, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 19, NO. 5, MAY 1997
- Paul E. Utgoff, Neil C. Berkman, Jeffery A. Clouse, Decision Tree Induction Based on
Efficient Tree Restructuring, Department of Computer Science, University of Massachusetts, Amherst, MA 01003
- Niks, Nikson, Decision Trees, Introduction to machine learning,
- Ron Kohavi Ross Quinlan, Decision Tree Discovery, Blue Martini Software 2600 Campus Dr. Suite 175, San Mateo, CA & Samuels Building, G08 University of New South Wales, Sydney 2052 Australia
- Paul E. Utgoff, Incremental Induction of Decision Trees, Department of Computer and Information Science, University of Massachusetts, Amherst, MA 01003
- Matti K¨a¨ari¨ainen, Tuomo Malinen, Tapio Elomaa, Selective Rademacher Penalization and Reduced Error Pruning of Decision Trees, Department of Computer Science, University of Helsinki, Institute of Software Systems, Tampere University of Technology, Tampere, Finland
- Michael Kearns, Yishay Mansour, A Fast, Bottom-Up Decision Tree Pruning Algorithm with Near-Optimal Generalization, AT&T Labs, Tel Aviv University
- Zijian Zheng, Constructing new attributes for decision tree learning, Basser Department of Computer Science, the university of Sydney, Australia
- Emily Thomas, DATA MINING: DEFINITIONS AND DECISION TREE EXAMPLES, Director of Planning and Institutional Research, State University of New York
- Kurt Hornik, The RWeka Package August 20, 2006
- Zhengping Ma, Eli Lilly and Company, Data mining in SAS® with open source software, SAS Global Forum 2011
- Simon Urbanek, Package &apos;rJava&apos;, Jan 2, 2012
- M. Govindarajan, Text Mining Technique for Data Mining Application, World Academy of Science, Engineering and Technology 35 2007
- A S Galathiya, AP Ganatra, CK Bhensdadia, An Improved decision tree induction algorithm with feature selection, cross validation, model complexity & reduced error pruning, IJSCIT march 2012

Index Terms
Computer Science Algorithms

Keywords
Rep Decision Tree Algorithm C5 Classifier C4. 5 Classifier