Abstract

Data mining is an interdisciplinary field of computer science and is referred to extracting or mining knowledge from large amounts of data. Classification is one of the data mining techniques that maps the data into the predefined classes and groups. It is used to predict group membership for data instances. There are many areas that adapt Data mining techniques such as medical, marketing, telecommunications, and stock, health care and so on. The C4. 5 can be referred as the statistic Classifier. This algorithm uses gain radio for feature selection and to construct the decision tree. It handles both continuous and discrete features. C4. 5 algorithm is widely used because of its quick classification and high precision. This paper proposed a C4. 5 classifier based on the various entropies (Shannon Entropy, Havrda and Charvĕ entropy, Quadratic entropy) instance of Shannon entropy for classification. Experiment results show that the various entropy based approach is effective in achieving a high classification rate.

References
Classification Through Machine Learning Technique: C4.5 Algorithm based on Various Entropies

- Agarwal, S., Pandey, G. N., & Tiwari, M. D. Data Mining in Education: Data Classification and Decision Tree Approach.
Mathur, N., Kumar, S., Kumar, S., & Jindal, R. The Base Strategy for ID3 Algorithm of Data Mining Using Havrda and Charvat Entropy Based on Decision Tree.


Index Terms

Computer Science

Artificial Intelligence
Keywords

Data Mining  Classification technique  Machine learning  Decision tree technique
C4. 5 algorithm.