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


Index Terms

Computer Science  Evolutionary Computation

Key words

algorithm  decision tree

missing data  genetic