Classification of Documents and Images Using an Enhanced Genetic Algorithm

Nirmani Athuraliya[1], Dulshani Dasanayake[1], Hashini De Silva[1], Kavindi Fernando[1], Prasanna S. Haddela[1], and Adeepa Gunarathne[1]

[1]Department of Information Technology, Sri Lanka Institute of Information Technology, Sri Lanka.
deeshaniathuraliya2@gmail.com, dulshanisd@gmail.com, desilva.hashini27@gmail.com, kavindifernando999@gmail.com, prasanna.s@sliit.lk, dimuth.a@sliit.lk

Abstract

In 1975, John Holland proposed the Genetic Algorithm (GA). The algorithm is widely used to provide superior solutions for optimization and search problems by relying on biologically inspired operators including mutation, crossover, and selection. The fittest individuals are chosen for reproduction in this algorithm to generate the next generations offspring. Classification is a technique used in data mining to analyze the collected data and to divide them into different classes. The relationship between a known class assignment and the properties of the entity to be classed may serve as the foundation for the classification procedure. Through this research, it has mainly consider classification for documents and images using GA. In order to enhance the accuracy and to reduce the error rate of traditional models, a new approach is proposed which is based on GA. The primary benefit of using GA in conjunction with classification is the efficiency in which it can address optimization issues. The experiment results are used to verify the suggested algorithm using benchmark data sets gathered from the UCI machine learning repository.

Keywords

Genetic Algorithm, Classification, Text Documents, Medical Images