Abstract

High-Performance Computing (HPC) proved notable performance enhancements especially on fields where data processing is exceedingly time consuming. Such data raise the curse of dimensionality problem in which several methods followed to maintain the number of features describing that data. Feature Selection is one of the known procedures applied to overcome the drawback caused by the data size. In this work, a feature selection model designed and tested. Genetic Algorithm (GA) is the search algorithm involved, Linear Discriminant Analysis (LDA) used as a classifier, and both form the feature selection model. GA estimates an optimal solution that saves the enormous amount of time might be consumed by a brute force search, and LDA performs as its fitness object. HPC techniques implemented since the computational power was one of the leading obstacle causing an extensive processing time. The developed feature selection model saves 89% of the original time consumed while using common computing facilities. It also maintains an accuracy rate of almost 86% selecting 37% of the original number of features.
References


Index Terms

Computer Science

Information Sciences

Keywords

Genetic Algorithm, Linear Discernment Analysis, Islands Model, Message Passing Interface, Boost.