

University of Macau

Abstract

USING GENETIC ALGORITHMS AND BOOSTING FOR
DATA PREPROCESSING

by Celestino Lei

Thesis Supervisor: Prof. Otakar Babka
Master of Science in Software Engineering

Conventional learning methods such as connectionist methods, decision trees and nearest neighbor methods perform reasonably well on small to medium-sized training datasets. When facing with larger datasets, the learning and classification times of these learning methods might increase noticeably. This effect can be lightened with the use of data preprocessing techniques.

Genetic Algorithms and Boosting are methods developed recently in the field of machine learning. Genetic Algorithms are conventionally used for optimization problems which have no straightforward solutions. Boosting is used as a method of creating a committee of classifiers in order to provide a higher classification accuracy than using a single classifier. In this thesis, genetic algorithms and boosting methods are used as data preprocessing techniques for reducing the size of a training dataset.

In order to show the adequacy of the learning methods and data preprocessing techniques in real-life applications, two practical applications are presented, and the methods are experimented on these applications. One of the applications is related to the recognition of handwritten digits and another to the choice of the type of construction structure to be used in a construction site. Specific software systems are developed to conduct the comprehensive experiments that tend to show the feasibility of using Genetic Algorithms and Boosting in real-life applications.