Evaluating Feature Selection Techniques in Simulated Soccer Multi Agents System

Fahimeh Farahnakian and Nasser Mozayani
School of Computer Engineering
Iran University of Science &Technology
Tehran, Iran
Farahnakian@comp.iust.ac.ir, Mozayani@iust.ac.ir

Abstract

Since the quality of data affects the success rate of data mining and learning algorithms, it is always attempted to identify and remove the irrelevant and redundant information in a dataset.
Robotic soccer is a multi-agent system in which agents play in real-time, dynamic, complex and noisy environment. Many parameters affect the result of shooting toward the goal and agents must response to variations in soccer field rapidly. Therefore it is impossible to use all features in scoring behavior.
This paper selects dataset for effective features of scoring behavior simulated soccer agents, then compares the size of the trees and accuracy produced by each feature selection scheme against the size of the trees and accuracy produced by C4.5 with no feature selection method. Experimental results have shown that dimensionality reductions lead to operate faster and more effective learning algorithm in real-time simulated soccer agent.

1. Introduction

Feature selection generally involves a combination of search and feature utility estimation plus evaluation with respect to specific learning schemes [1].
In many practical situations there are far too many features for learning algorithm to handle. Consequently, the data must be pre-processed to select a subset of the features to use in learning. Of course, learning methods themselves try to select features appropriately and ignore irrelevant or redundant ones, but in practice their performance can frequently be improved by pre selection [2]. Feature selection is a preprocessing technique commonly used on high dimensional data [3]. Reducing the dimensionality of data reduces the size of the solution space and allows algorithms to operate faster and more effectively [4]. Thus the generic purpose pursued is the improvement of the inductive learner, either in terms of learning speed, generalization capacity or simplicity of the representation [5].

Feature selection can be found in many areas of data mining such as classification, clustering, association rules, and regression. For example, feature selection is called subset or variable selection in Statistics [6].Also, in soccer robots that time learning is effective in team performance, it is necessary to find the bests and the least features for learning algorithms.

Figure 1 shows feature selection methods how used to reduce the dimensionality of both original training data and testing data. Both reduced datasets may then be passed to a machine learning algorithm for training and testing [1, 7].
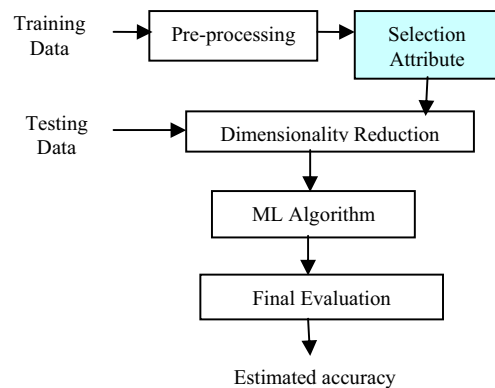

Figure1: Usage of attribute selection methods for learning machine algorithm