Wrapper Feature Selection based on Genetic Algorithm for Recognizing Objects from Satellite Imagery

Nabil M. Hewahi, Department of Computer Science, University of Bahrain, Zallaq, Bahrain
Eyad A. Alashqar, Department of Computer Science, Islamic University of Gaza, Gaza City, Palestinian Territories, Israel

ABSTRACT

Object recognition is a research area that aims to associate objects to categories or classes. The recognition of object specific geospatial features, such as roads, buildings and rivers, from high-resolution satellite imagery is a time consuming and expensive problem in the maintenance cycle of a Geographic Information System (GIS). Feature selection is the task of selecting a small subset from original features that can achieve maximum classification accuracy and reduce data dimensionality. This subset of features has some very important benefits like, it reduces computational complexity of learning algorithms, saves time, improve accuracy and the selected features can be insightful for the people involved in problem domain. This makes feature selection as an indispensable task in classification task. In this work, the authors propose a new approach that combines Genetic Algorithms (GA) with Correlation Ranking Filter (CRF) wrapper to eliminate unimportant features and obtain better features set that can show better results with various classifiers such as Neural Networks (NN), K-nearest neighbor (KNN), and Decision trees. The approach is based on GA as an optimization algorithm to search the space of all possible subsets related to object geospatial features set for the purpose of recognition. GA is wrapped with three different classifier algorithms namely neural network, k-nearest neighbor and decision tree J48 as subset evaluating mechanism. The GA-ANN, GA-KNN and GA-J48 methods are implemented using the WEKA software on dataset that contains 38 extracted features from satellite images using ENVI software. The proposed wrapper approach incorporated the Correlation Ranking Filter (CRF) for spatial features to remove unimportant features. Results suggest that GA based neural classifiers and using CRF for spatial features are robust and effective in finding optimal subsets of features from large data sets.

Keywords: Feature Extraction, Feature Selection, Genetic Algorithm, Satellite Imagery, Wrapper Approach

DOI: 10.4018/JITR.2015070101
1. INTRODUCTION

Object recognition is one of research areas that aims to classify objects. Usually recognition of objects such as building, tree, mountains, roads, and rivers is a very complex task in terms of time and cost. Traditional recognizing methods, such as hand-digitizing, are slow, tedious and the availability of numerous spectral, spatial, and textural features renders the selection of optimal features a time consuming.

The goal of the Feature Selection (FS) is to detect irrelevant and/or redundant features as they harm the learning algorithm performance (Lee & Moore, 2014). A FS algorithm can effectively remove irrelevant and redundant features and take into account feature correlation. This not only leads up to an insight understanding of the data, but also improves the performance of a learner by enhancing the generalization capacity and the interpretability of the learning model (Wang et al., 2014). In other words, no new feature is created, the features that are considered irrelevant or redundant are discarded, and we ideally would end up with the best possible feature subset, that is, the subset with minimum size and which leads to the minimum classification error rate.

Feature selection with subset evaluation requires defining how to search the space of feature subsets (search method) and what measure to use when evaluating a feature subset (evaluation criterion) as well as the initial feature set and a termination condition.

Feature Selection methods fall into two broad categories: Wrapper and Filter (Kohavi & John, 1997; Molina, Belanche, & Nebot, 2002). The Wrapper approach uses the accuracy of the classification algorithm as the evaluation function to measure a feature subset as you shown in Figure 1, while the evaluation function of the Filter approach is independent of the classification algorithm. The accuracy of the Wrapper approach is usually high; however, the generality of the result is limited, and the computational complexity is high. In comparison, Filter approach is of generality, and the computational complexity is low. Because the Wrapper approach is computationally expensive (Vafaei & De Jong, 1992), the Filter approach is usually a good choice when the number of features is very large. Thus, we focus on the Wrapper method in our experiment, because we have only 38 features.

Figure 1. Feature selection based on wrapper method (Kohavi & John, 1997)
From Pilot to Practice: Streamlining Procurement and Engineering at Lawrence Livermore National Laboratory
Judith Gebauer and Frank Farber (2000). Organizational Achievement and Failure in Information Technology Management (pp. 77-90).
www.igi-global.com/chapter/pilot-practice-streamlining-procurement-engineering/27855?camid=4v1a