Chapter 66
Supervised Machine Learning for Plants Identification Based on Images of Their Leaves

Mohamed Elhadi Rahmani
GeCoDe Laboratory, Department of Computer Science, Dr. Moulay Tahar University of Saida, Saida, Algeria

Abdelmalek Amine
GeCoDe Laboratory, Department of Computer Science, Dr. Moulay Tahar University of Saida, Saida, Algeria

Reda Mohamed Hamou
GeCoDe Laboratory, Department of Computer Science, Dr. Moulay Tahar University of Saida, Saida, Algeria

ABSTRACT

Botanists study in general the characteristics of leaves to give to each plant a scientific name; such as shape, margin...etc. This paper proposes a comparison of supervised plant identification using different approaches. The identification is done according to three different features extracted from images of leaves: a fine-scale margin feature histogram, a Centroid Contour Distance Curve shape signature and an interior texture feature histogram. First represent each leaf by one feature at a time in, then represent leaves by two features, and each leaf was represented by the three features. After that, the authors classified the obtained vectors using different supervised machine learning techniques; the used techniques are Decision tree, Naïve Bayes, K-nearest neighbour, and neural network. Finally, they evaluated the classification using cross validation. The main goal of this work is studying the influence of representation of leaves' images on the identification of plants, and also studying the use of supervised machine learning algorithm for plant leaves classification.

DOI: 10.4018/978-1-7998-2460-2.ch066

1. INTRODUCTION

Plants identification is the determination of the similarities or differences between two elements. For this, it considered as a basic activity and one of the primary objectives of systematics. The comparison of an unknown plant with a named specimen and the determination that the two elements are the same also involves classification. in other words, it is the decision that an unknown plant belongs to the same group (such as species, genus, family, etc.) as a known specimen, this process allows to the information stored in classification systems to become available and applicable to the material at hand. The two processes (identification and classification) require a definition of criteria of similarities, i.e: Correct identification provides basic information about size, shape and texture of a plant and can be helpful in protecting it from various types of pests and diseases. Plant species classification can be done through various ways like flower, root, leaf, fruit etc. these two lasts allow to a botanist to compare plants and judge the final decision. According to Blackwelder (1967):" identification enables us to retrieve the appropriate facts from the system (classification) to be associated with some specimen at hand" and is "better described as the recovery side of taxonomy."

In nature, plant leaves are two dimensional containing important features that can be useful for classification of various plant species, such as shapes, colours, textures and structures of their leaf, bark, flower, seedling and morph. According to Bhardwaj et al. (2013), if the plant classification is based on only two dimensional images, it is very difficult to study the shapes of flowers, seedling and morph of plants because of their complex three dimensional structures.

The present paper proposes a comparison of the classification of different representation of plant leaves based on its margin, shape and textures; we used for each representation different classical supervised data mining algorithms. The organization of this paper is given as follows: Section 2 provides a stat of the art in which we gave a summary of machine learning and some recent works on application of machine learning in plants identification; Section 3 gives details about dataset used in our experiment, Section 4 presents used machine learning approaches, discussion of the results got by Rahmani et al. (2015) compared to the obtained results of classification of plants using multilayer neural network is shown in Section 5, and finally Section 6 gives the overall conclusion and the scope for future research.

2. VIEW OF LITERATURE

2.1. Supervised Machine Learning

In supervised learning we need classes. That means we need a qualitative attribute that can take a finite set values, so we can say that supervised learning is the inference of a function labelled training data.

2.1.1. Decision Tree

First decision tree developed by J. Ross Quinlan known as ID3 (Iterative Dichotomize), then Quinlan presented C4.5 as successor of ID3, which became a benchmark to which newer supervised learning algorithms are often com- pared. In 1984, Breiman, Friedman, Olshen, and Stone published the book Classification and Regression Trees (CART).