A Naive Bayes Based Tree Classification System

Ying Yu Chen and Xun Lu*

School of Computer Science and Technology

Soochow University, Suzhou 215006, Jiangsu, China

1127404017@suda.edu.cn,luxun@suda.edu.cn

Keywords: Machine Learning; Naive Bayes; Classification; Intelligent System

Abstraction. This paper presents the design and implementation of an automatic classification system based on Bayes tree. Analyzed the characteristics of the tree, the system research potential issues that may affect the classification results, factors that should be considered when the trees classification modeling is proposed and then induced by these factors, and complete the characteristics of six main aspects of. This article from the high abstract point of view, the most relevant content and information classification, extraction set trees classification model, and finally implement a Bias classification model of automatic classification system based on tree Bias. This paper designs and implements the self learning function parameters in tree automatic classification system, in order to obtain the best classification results of actual data through self-study and certain adjustment of parameters. The experimental results show that the proposed method can better solve tree automatic classification problem.

Introduction

Tree classification is a process of assigning predefined categories according to their biological function classification tree. Leaves deciduous trees in the winter off, leaving them only the trunk, branch and other sparse organism. These trees are called tree in this state, and their classification is called trees classification.

No leaves or fruit, it is difficult to determine the tree. At present, most of the work is classified by gardeners use their experience or through literature, is of low efficiency, high error rate. Sometimes too much heavy people work is returned by the incorrect results. Therefore, the design and be imperative to realize automatic trees classification system.

The 1 tree classification

Automatic tree classification relates to many different fields and biology, botany, statistics and computer science and other disciplines. It is the combination of many aspects, and has its own process and framework. In this paper we will focus on how the application of computational techniques to trees classification.

In the tree, the tree is divided into a set of predefined categories by its function. That is to say, to separate directory tree for a given set of tree and a given set of types, tree classification is the true or false value is assigned to each of the tasks. If you belong to, the allocation of true, otherwise the assigned false.

Tree classification is, by using the classification function, find the approximate function of an unknown, to best represent the given tree category.

Tree classification generally consists of three parts: training, testing and classification. Training is obtained after the tree classifier, can be used to determine the stage and classifications of the detection of the trees. The test is similar to the classification, in addition to test example of target classification known. The calculator will be used to evaluate the test results of the classifier, and given scores.

Tree Classifier

Tree widely distributed in the world. In the winter, tree has no leaves but the bud, leaf scars, leaf trace, branching and shape are useful classification. Here we Engler classification systems, used in Botany, tree classification system used can set the classification logic.

Leaves, flowers and fruits are the most important factors, in the classification of the trees, but they are not available in winter green deciduous trees. After a lot of research, we found that can be used for other important factors   tree. Winter buds of these factors include the position and shape of the arrangement, the number of bud scales, shapes and leaf scars and leaf trace, shape and medulla brunch structure and so on, which are all available, though the winter.

According to Bias's theory, we assume that all features are independent. Therefore, the probability of a tree belonging to a predefined class is weighted characteristic equation 1 as a combination.

Where is the weight function in a tree, the likelihood also belongs to the class of functions. Here, are obtained from the training data.

We performed with   trees classification Bayesian theory. We first consider the corresponding similarity, its shape, branches, bark, winter bud, leaf scars, leaf trace, and then combine their weight to keep the overall feature similarity.

Adaptive Features Of Xiang Quanzhong

In the natural world, tree from each other a lot of changes in their morphology. Different kinds of trees have distinctive features a variety of their own. Even for a given function, it may have different weights in different kinds of trees. For example, some tree tree bark is more obvious, so it should be given greater weight, while the other trees could more easily identify the leaf traces, and should therefore be highly weighted features of leaf trace.

Obviously, to set all the tree it is not a wise choice overall weight coefficient. Therefore, the ability to automatically learn the classifier is very necessary, adjust the feature weight, improve the precision of classification.

And parameter learning of Bayesian networks, we can get all the functions, we can get the best parameters of the best classification results of classifier. To realize the adaptive self learning, we first need to raise and the training set, and then adjust the weight coefficient, the last test, until you reach the satisfactory classification results. The initial classifier parameters are given a random set of test cases, and then returns the results, which will be compared with the actual type. In the automatic adjustment of parameters of each iteration, until no improvement is available.

Conditions in the winter deciduous Chinese, Professor Ren Xuanwei to Chinese   tree, including the description and the analysis of vivid detailed and complete data characteristics. In our experiments, a total of 1316   tree selection, of which 734   tree selection for training and close test and open test 882.

This is the similarity analysis of both meaningless two   tree, its similarity with the low, because it is in any of the two kinds of trees are similar to natural, more or less. Then we use a threshold of T to eliminate this situation. Characteristics of trees and low similarity should also be excluded similarity calculation. In our experiments, the threshold T is set to 0.01 is a very traditional values. Less than T trees similar classification is not invalid, but also bring chaos classification; therefore they should be eliminated before classification.

Table 1 in the close test and open test results

	Close test	Open test
Trees #	1868	1564
Trees #	1868	1432
Correctly classified trees #	1932	1224
Incorrectly classified trees #	0	132
Classification precision(%)	98.07	85.47

Here can not use closed test to evaluate which test cases from the training data and the test, usually with high accuracy of classifier learning ability. Close test results show that the classification is proposed in this paper has very good learning ability.

Open the test data to separate the training data set, it doesn't matter of training data. A precision of 87%, open test results show that the system can well satisfy the classification.

After the test from the winter condition data China deciduous, we test our project collected data, we will once again be the threshold is set to 0.01. Table 2 shows the test results.

Table 2 test results of our project

	result
Number of hibernal trees	2132
Number of effective hibernal trees	1782
Number of correctly classified hibernal trees	1434
Number of ineffective hibernal trees	350
Classification precision(%)	80.47

Due to incomplete data collection form of the real world, the number of tree growth is invalid. At the same time, our results than experiment in winter conditions Chinese deciduous as well. However, the precision is still to maintain more than 80%, show the efficiency of the classifier. In the two experiment, we found that the noise elimination processing is really helps to improve the accuracy of.

Conclusion and Characteristics

Tree classification is a process according to its function is divided into the trees. trees classification is the classification of tree. Small so far; the automatic classification tree to do at the same time, huge data collection in forestry and gardening very much need to classify.

We studied the characteristics of tree system, potential and may affect the classification result analysis. Through these efforts, we design and implement a tree automatic classification system based on Bayesian learning. Experimental results show that the proposed method can well solve the automatic trees classification problem.

The subject of this dissertation is an important part of the research on protection and development of landscape plants. There is a lot of work to do. For example, we are currently using the typical classification system of Bayesian learning. It wants to replace data learning network, this is our future work.

Reference

[1]. Dong H, Hui S C, He Y. Structural analysis of chat messages for topic detection[J]. Online Information Review, 2006, (5):496-516.

[2]. Abraham, Abraham, Abraham et.al. The morphologies of distant galaxies. 1: an automated

classification system[J]. Astrophysical Journal, 1994. DOI:10.1086/174550.

[3]. Pal M, M P, Pal M, et al. M5 model tree for land cover classification[J]. International Journal of Remote Sensing, 2006, 27(4):825-831.

[4]. Larsen M, M L, Larsen M, et al. Single tree species classification with a hypothetical multi-spectral satellite[J]. Remote Sensing of Environment, 2007, 110:523–532.

[5]. Zhang L, L Z, Zhang L, et al. An approach for ECG classification based on wavelet feature extraction and decision tree[C]. //Wireless Communications and Signal Processing (WCSP), 2010 International Conference on. IEEE, 2010:1 - 4.

[6]. Zimmermann A, A Z, Zimmermann A, et al. CTC - correlating tree patterns for classification[C]. //Data Mining, Fifth IEEE International Conference on. IEEE, 2005.

[7]. Yi J, J Y, Yi Jet al. Coevolutionary Based on Decision Tree Classification Algorithm[J]. Computer Knowledge and Technology(Academic Exchange), 2007.

[8]. Qinyi Z, Z Q, Qinyi Zet al. Research on Decision-tree Classification Method Based on Object Cube Structure[J]. Journal of Dali University, 2009.