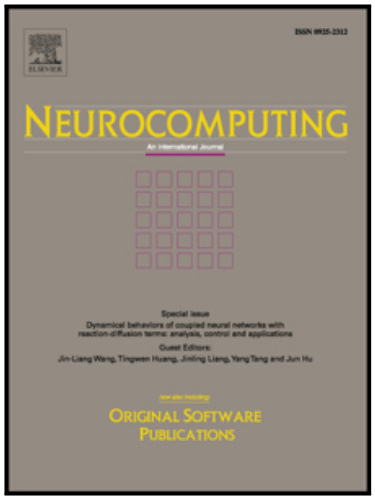
Communicated by Dr. Haijun Zhang

Accepted Manuscript



Efficient Tree Classifiers for Large Scale Datasets

Fei Wang, Quan Wang, Feiping Nie, Weizhong Yu, Rong Wang

PII: S0925-2312(18)30032-8
DOI: 10.1016/j.neucom.2017.12.061
Reference: NEUCOM 19215

To appear in: *Neurocomputing*

Received date: 3 August 2017
Revised date: 29 November 2017
Accepted date: 25 December 2017

Please cite this article as: Fei Wang, Quan Wang, Feiping Nie, Weizhong Yu, Rong Wang, Efficient Tree Classifiers for Large Scale Datasets, *Neurocomputing* (2018), doi: 10.1016/j.neucom.2017.12.061

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Efficient Tree Classifiers for Large Scale Datasets

Fei Wang[a], Quan Wang[a], Feiping Nie[b,*], Weizhong Yu[a,b], Rong Wang[b]

[a]*School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China*
[b]*Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China*

Abstract

Classification plays a significant role in production activities and lives. In this era of big data, it is especially important to design efficient classifiers with high classification accuracy for large scale datasets. In this paper, we propose a randomly partitioned and a Principal Component Analysis (PCA)-partitioned multivariate decision tree classifiers, of which the training time is quite short and the classification accuracy is quite high. Approximately balanced trees are created in the form of a full binary tree based on two simple ways of generating multivariate combination weights and a median-based method to select the divide value having ensured the efficiency and effectiveness of the proposed algorithms. Extensive experiments conducted on a series of large datasets have demonstrated that the proposed methods are superior to other classifiers in most cases.

Keywords: big data, classification, multivariate decision tree

1. Introduction

As it is one of the many important techniques in data mining, classification plays a significant role in helping humans understand things and comprehend nature. Simply because of this, classification has drawn a large interest among the various academic communities, such as, in statistics, machine learning, pattern recognition, data mining, and so on. Studies on classification have been applied in many different domains,

*Corresponding author
Email addresses: wfx@mail.xjtu.edu.cn (Fei Wang), bhxspring@stu.xjtu.edu.cn (Quan Wang), feipingnie@gmail.com (Feiping Nie), yuwz05@xjtu.edu.cn (Weizhong Yu), wangrong07@tsinghua.org.cn (Rong Wang)

Download English Version:

https://daneshyari.com/en/article/6864489

Download Persian Version:

https://daneshyari.com/article/6864489

Daneshyari.com