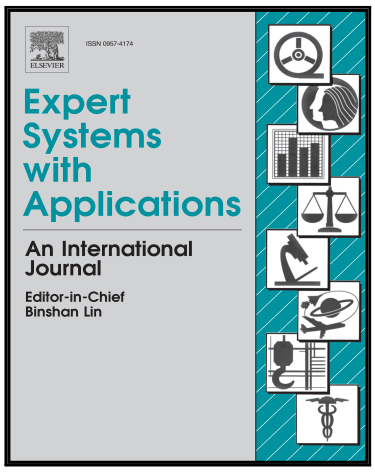
Accepted Manuscript



Region compatibility based stability assessment for decision trees

Lihong Wang, Qiang Li, Yanwei Yu, Jinglei Liu

PII: S0957-4174(18)30181-7
DOI: 10.1016/j.eswa.2018.03.036
Reference: ESWA 11883

To appear in: *Expert Systems With Applications*

Received date: 31 May 2017
Revised date: 19 March 2018
Accepted date: 20 March 2018

Please cite this article as: Lihong Wang, Qiang Li, Yanwei Yu, Jinglei Liu, Region compatibility based stability assessment for decision trees, *Expert Systems With Applications* (2018), doi: 10.1016/j.eswa.2018.03.036

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

Region compatibility based stability assessment for decision trees

Lihong Wang[1], Qiang Li[2], Yanwei Yu[1,3,*], Jinglei Liu[1]

[1]*School of Computer and Control Engineering, Yantai University*
Yantai, Shandong 264005, China
[2]*School of Economics and Management, Yantai University*
Yantai, Shandong 264005, China
[3]*College of Information Science and Technology, Pennsylvania State University*
University Park, PA 16802, USA
∗Corresponding author. Email: wanglh_000@163.com (Lihong Wang), lq130@163.com (Qiang Li), yuyanwei@ytu.edu.cn (Yanwei Yu), jinglei_liu@sina.com (Jinglei Liu).

Abstract

Decision tree learning algorithms are known to be unstable, because small changes in the training data can result in highly different decision trees. An important issue is how to quantify decision tree stability. Two types of stability are defined in the literature: structural and semantic stability. However, existing structural stability measures are meaningless when applied to apparently different decision trees, and semantic stability only focuses on prediction accuracy without considering structural information. This paper proposes a region compatibility based structural stability measure for decision trees that considers the structural distribution of leaves from the view of basic probability assignments in evidence theory. To the best of our knowledge, we are the first to use basic probability assignments to quantify decision tree stability. We prove convergence for region compatibility, and show that apparently different decision trees have some inherent similarity from the view of region compatibility. We also clarify the meaning of region compatibility for measuring decision tree stability, and derive a method to select a relatively stable learning algorithm for a given dataset. Experimental results validate that region compatibility is effective to quantify the stability

Download English Version:

https://daneshyari.com/en/article/6854967

Download Persian Version:

https://daneshyari.com/article/6854967

Daneshyari.com