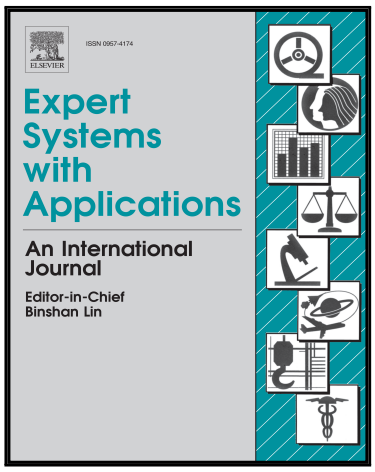
Accepted Manuscript



Learning from class-imbalanced data: Review of methods and applications

Guo Haixiang , Li Yijing , Jennifer Shang , Gu Mingyun , Huang Yuanyue , Gong Bing

PII: S0957-4174(16)30717-5
DOI: 10.1016/j.eswa.2016.12.035
Reference: ESWA 11044

To appear in: *Expert Systems With Applications*

Received date: 6 September 2016
Revised date: 23 November 2016
Accepted date: 25 December 2016

Please cite this article as: Guo Haixiang , Li Yijing , Jennifer Shang , Gu Mingyun , Huang Yuanyue , Gong Bing , Learning from class-imbalanced data: Review of methods and applications, *Expert Systems With Applications* (2016), doi: 10.1016/j.eswa.2016.12.035

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

Learning from class-imbalanced data: Review of methods and applications

Guo Haixiang[1,2,3,*], Li Yijing[1,2,*], Jennifer Shang[4,*], Gu Mingyun[1], Huang Yuanyue[1], Gong Bing[5]

1. College of Economics and Management, China University of Geosciences, Wuhan 430074, China
2. Research Center for Digital Business Management, China University of Geosciences, Wuhan 430074;
3. Mineral Resource Strategy and Policy Research Center of China University of Geosciences(WUHAN), Wuhan 43007
4. The Joseph M. Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, PA 15260, USA
5. Department of Industrial Engineering, Business Administration and Statistic, E.T.S Industrial Engineering, Universidad Politécnica de Madrid, C/José Gutiérrez Abascal, 2- 20086, Madrid, Spain

Abstract: Rare events, especially those that could potentially negatively impact society, often require humans' decision-making responses. Detecting rare events can be viewed as a prediction task in data mining and machine learning communities. As these events are rarely observed in daily life, the prediction task suffers from a lack of balanced data. In this paper, we provide an in depth review of rare event detection from an imbalanced learning perspective. Five hundred and seventeen related papers that have been published in the past decade were collected for the study. The initial statistics suggested that rare events detection and imbalanced learning are concerned across a wide range of research areas from management science to engineering. We reviewed all collected papers from both a technical and a practical point of view. Modeling methods discussed include techniques such as data preprocessing, classification algorithms and model evaluation. For applications, we first provide a comprehensive taxonomy of the existing application domains of imbalanced learning, and then we detail the applications for each category. Finally, some suggestions from the reviewed papers are incorporated with our experiences and judgments to offer further research directions for the imbalanced learning and rare event detection fields.

Keywords: rare events, imbalanced data, machine learning, data mining

1. Introduction

Rare events, unusual patterns and abnormal behavior are difficult to detect, but often require responses from various management functions in a timely manner. By definition, rare events refer

* Corresponding authors: Guo Haixiang and Li Yijing contributed equally in this paper. Tel: 86 027-67883201, Fax: 86 027-67883201
Email addresses: liyijing024@hotmail.com (Li Yijing), faterdumk0732@sina.com (Guo Haixiang), shang@katz.pitt.edu (Jennifer Shang), huangyuanyue1991@126.com (Huang Yuanyue), 550312686@qq.com (Gu Mingyun), gongbing1112@gmail.com (Gong Bing)

Download English Version:

https://daneshyari.com/en/article/4943485

Download Persian Version:

https://daneshyari.com/article/4943485

Daneshyari.com