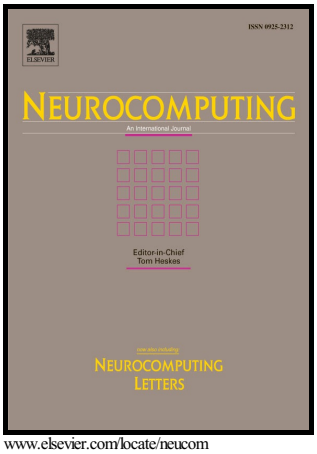
Large Cost-Sensitive Margin Distribution Machine for Imbalanced Data Classification

Fanyong Cheng, Jing Zhang, Cuihong Wen, Zhaohua Liu, Zuoyong Li



www.elsevier.com/locate/neucom

PII: S0925-2312(16)31294-2
DOI: http://dx.doi.org/10.1016/j.neucom.2016.10.053
Reference: NEUCOM17689

To appear in: *Neurocomputing*

Received date: 31 January 2016
Revised date: 11 June 2016
Accepted date: 31 October 2016

Cite this article as: Fanyong Cheng, Jing Zhang, Cuihong Wen, Zhaohua Liu and Zuoyong Li, Large Cost-Sensitive Margin Distribution Machine for Imbalanced Data Classification, *Neurocomputing* http://dx.doi.org/10.1016/j.neucom.2016.10.053

Large Cost-Sensitive Margin Distribution Machine for Imbalanced Data Classification

Fanyong Cheng[a,b,*], Jing Zhang[a], Cuihong Wen[a], Zhaohua Liu[c], Zuoyong Li[b]

Lushan South Road, Changsha

[a]*College of Electrical and Information Engineering, Hunan University, Changsha 410082, China*
[b]*Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Department of Computer Science, Minjiang University, Fuzhou 350121, China*
[c]*School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, 411201, China*

Abstract

This paper develops cost-sensitive margin distribution learning and proposes Large Cost-Sensitive margin Distribution Machine (LCSDM) to get balanced detection rate on imbalanced training data. Recently, margin theory revealed that compared with a single margin, margin distribution is more critical to the generalization performance. Large margin Distribution Machine (LDM) is designed to get superior classification performance and strong generalization performance. However, LDM generally has imbalanced margin distribution between two classes on imbalanced training data. This generally leads to the lower detection rate of the minority class, which contradicts to the needs of high detection rate of the minority class in many real applications. Therefore, cost-sensitive margin distribution learning is brought forward to obtain balanced margin distribution and detection rate between two classes. What's more, this research deduces the relation between cost-sensitive parameter and in-class detection rate, and designs LCSDM to obtain balanced detection rate. Experimental results show that LCSDM can gradually increase the margin distribution of the minority class to obtain a more balanced detection rate. As a general learning method, LCSDM is especially applicable to imbalanced data classification.

Keywords: `margin distribution`, cost-sensitive learning, imbalanced training

*Fanyong Cheng
Email address: `b12090031@hnu.edu.cn.` (Fanyong Cheng)