Mach Learn (2012) 88:69–126
DOI 10.1007/s10994-011-5276-1

Compressed labeling on distilled labelsets for multi-label learning

Tianyi Zhou · Dacheng Tao · Xindong Wu

Received: 30 September 2010 / Accepted: 12 December 2011 / Published online: 6 January 2012

Abstract Directly applying single-label classification methods to the multi-label learning problems substantially limits both the performance and speed due to the imbalance, dependence and high dimensionality of the given label matrix. Existing methods either ignore these three problems or reduce one with the price of aggravating another. In this paper, we propose a $\{0, 1\}$ label matrix compression and recovery method termed "compressed labeling (CL)" to simultaneously solve or at least reduce these three problems. CL first compresses the original label matrix to improve balance and independence by preserving the signs of its Gaussian random projections. Afterward, we directly utilize popular binary classification methods (e.g., support vector machines) for each new label. A fast recovery algorithm is developed to recover the original labels from the predicted new labels. In the recovery algorithm, a "labelset distilling method" is designed to extract distilled labelsets (DLs), i.e., the frequently appeared label subsets from the original labels via recursive clustering and subtraction. Given a distilled and an original label vector, we discover that the signs of their random projections have an explicit joint distribution that can be quickly computed from a geometric inference. Based on this observation, the original label vector is exactly determined after performing a series of Kullback-Leibler divergence based hypothesis tests on the distribution about the new labels. CL significantly improves the balance

Editors: Grigorios Tsoumakas, Min-Ling Zhang, and Zhi-Hua Zhou.

T. Zhou · D. Tao
Centre for Quantum Computation & Intelligent Systems (QCIS), Faculty of Engineering & IT, University of Technology, Sydney (UTS), 235 Jones Street, Broadway, NSW 2007, Australia

T. Zhou
e-mail: tianyi.zhou@student.uts.edu.au

D. Tao
e-mail: dacheng.tao@uts.edu.au

X. Wu (✉)
Department of Computer Science, Hefei University of Technology, Hefei, 230009, China
e-mail: xwu@cs.uvm.edu

X. Wu
Department of Computer Science, University of Vermont, 351 Votey Hall, Burlington, VT 05405, USA