

Synergy of multi-label hierarchical ensembles, data fusion, and cost-sensitive methods for gene functional inference

Nicolò Cesa-Bianchi · Matteo Re · Giorgio Valentini

Received: 30 September 2010 / Accepted: 7 November 2011 / Published online: 17 December 2011
© The Author(s) 2011

Abstract Gene function prediction is a complex multilabel classification problem with several distinctive features: the hierarchical relationships between functional classes, the presence of multiple sources of biomolecular data, the unbalance between positive and negative examples for each class, the complexity of the whole-ontology and genome-wide dimensions. Unlike previous works, which mostly looked at each one of these issues in isolation, we explore the interaction and potential synergy of hierarchical multilabel methods, data fusion methods, and cost-sensitive approaches on whole-ontology and genome-wide gene function prediction. Besides classical top-down hierarchical multilabel ensemble methods, in our experiments we consider two recently proposed multilabel methods: one based on the approximation of the Bayesian optimal classifier with respect to the hierarchical loss, and one based on a heuristic approach inspired by the true path rule for the biological functional ontologies. Our experiments show that key factors for the success of hierarchical ensemble methods are the integration and synergy among multilabel hierarchical, data fusion, and cost-sensitive approaches, as well as the strategy of selecting negative examples.

Keywords Hierarchical multilabel classification · Data integration · Cost-sensitive classification · Ensemble methods · Gene function prediction

1 Introduction

Multilabel learning (see, e.g., Tsoumakas and Katakis 2007 for a review) is an emerging thread in machine learning research, as witnessed by the number of recent papers and work-

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

N. Cesa-Bianchi · M. Re · G. Valentini (✉)

DSI, Dipartimento di Scienze dell'Informazione, Università degli Studi di Milano, Milan, Italy
e-mail: valentini@dsi.unimi.it

N. Cesa-Bianchi

e-mail: cesa-bianchi@dsi.unimi.it

M. Re

e-mail: re@dsi.unimi.it