Multilabel classification with meta-level features in a learning-to-rank framework

Yiming Yang · Siddharth Gopal

Received: 1 October 2010 / Accepted: 2 November 2011 / Published online: 15 December 2011 © The Author(s) 2011

Abstract Effective learning in multi-label classification (MLC) requires an appropriate level of abstraction for representing the relationship between each instance and multiple categories. Current MLC methods have focused on learning-to-map from instances to categories in a relatively low-level feature space, such as individual words. The fine-grained features in such a space may not be sufficiently expressive for learning to rank categories, which is essential in multi-label classification. This paper presents an alternative solution by transforming the conventional representation of instances and categories into meta-level features, and by leveraging successful learning-to-rank retrieval algorithms over this feature space. Controlled experiments on six benchmark datasets using eight evaluation metrics show strong evidence for the effectiveness of the proposed approach, which significantly outperformed other state-of-the-art methods such as Rank-SVM, ML-kNN (Multi-label kNN), IBLR-ML (Instance-based logistic regression for multi-label classification) on most of the datasets. Thorough analyses are also provided for separating the factors responsible for the improved performance.

Keywords Multilabel classification · Learning to rank

Y. Yang

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

The work is supported, in part, by the National Science Foundation (NSF) under grant IIS_0704689. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

Language Technologies Institute & Machine Learning Department, Carnegie Mellon University, Pittsburgh, USA e-mail: yiming@cs.cmu.edu