Int J Comput Vis (2013) 104:343–361
DOI 10.1007/s11263-012-0561-4

Optimization of Robust Loss Functions for Weakly-Labeled Image Taxonomies

Julian J. McAuley · Arnau Ramisa · Tibério S. Caetano

Received: 30 November 2011 / Accepted: 13 August 2012 / Published online: 1 September 2012

Abstract The recently proposed *ImageNet* dataset consists of several million images, each annotated with a single object category. These annotations may be imperfect, in the sense that many images contain *multiple* objects belonging to the label vocabulary. In other words, we have a multi-label problem but the annotations include only a single label (which is not necessarily the most prominent). Such a setting motivates the use of a *robust* evaluation measure, which allows for a limited number of labels to be predicted and, so long as one of the predicted labels is correct, the overall prediction should be considered correct. This is indeed the type of evaluation measure used to assess algorithm performance in a recent competition on ImageNet data. Optimizing such types of performance measures presents several hurdles even with existing structured output learning methods. Indeed, many of the current state-of-the-art methods optimize the prediction of only a single output label, ignoring this 'structure' altogether. In this paper, we show how to directly optimize continuous surrogates of such performance measures using structured output learning techniques with latent variables. We use the output of existing binary classifiers as input features in a new learning stage which optimizes the structured loss corresponding to the robust performance measure. We present empirical evidence that this allows us to 'boost' the performance of binary classification on a variety of weakly-supervised labeling problems defined on image taxonomies.

Keywords Image labeling · Image tagging · Image taxonomies · Structured learning

J.J. McAuley (✉)
InfoLab, Stanford University, Gates Building 4A, Stanford, CA 94305-9040, USA
e-mail: jmcauley@cs.stanford.edu

A. Ramisa
Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Parc Tecnologic de Barcelona, c/ Llorens i Artigas 4–6, 08028 Barcelona, Spain

T.S. Caetano
Principal Researcher, Machine Learning Group, NICTA, Locked Bag 9013, Alexandria NSW 1435, Australia

1 Introduction

The recently proposed *ImageNet* project consists of building a growing dataset of images, organized into a taxonomy based on the WordNet hierarchy (Deng et al. 2009). Each node in this taxonomy includes a large set of images (in the hundreds or thousands). From an object recognition point of view, this dataset is interesting because it naturally suggests the possibility of leveraging the image taxonomy in order to improve recognition beyond what can be achieved independently for each image. Indeed this question has been the subject of much interest recently, culminating in a competition in this context using ImageNet data (Berg et al. 2010; Lin et al. 2011; Sánchez and Perronnin 2011).

Each image in ImageNet may contain several objects from the label vocabulary, however the annotation includes only a single label per image, and this label is not necessarily the most prominent. This 'imperfect' annotation suggests that a meaningful performance measure in this dataset should somehow not penalize predictions that contain legitimate objects that are missing from the annotation. One way to deal with this issue is to use a *robust* performance measure based on the following idea: an algorithm is allowed to predict more than one label per image (up to a maximum of K labels, so that the solution is not degenerate), and so