Int J Comput Vis (2012) 99:302–318
DOI 10.1007/s11263-011-0423-5

Fast Structured Prediction Using Large Margin Sigmoid Belief Networks

Xu Miao · Rajesh P.N. Rao

Received: 23 March 2010 / Accepted: 17 January 2011 / Published online: 5 February 2011

Abstract Images usually contain multiple objects that are semantically related to one another. Mapping from low-level visual features to mutually dependent high-level semantics can be formulated as a structured prediction problem. Current statistical models for structured prediction make simplifying assumptions about the underlying output graph structure, such as assuming a low-order Markov chain, because exact inference becomes intractable as the tree-width of the underlying graph increases. Approximate inference algorithms, on the other hand, force one to trade off representational power with computational efficiency. In this paper, we present large margin sigmoid belief networks (LMSBNs) for structured prediction in images. LMSBNs allow a very fast inference algorithm for arbitrary graph structures that runs in polynomial time with high probability. This probability is data-distribution dependent and is maximized in learning. The new approach overcomes the representation-efficiency trade-off in previous models and allows fast structured prediction with complicated graph structures. We present results from applying a fully connected model to semantic image annotation, image retrieval and optical character recognition (OCR) problems, and demonstrate that the proposed approach can yield significant performance gains over current state-of-the-art methods.

Keywords Structured prediction · Probabilistic graphical models · Exact and approximate inference

X. Miao (✉) · R.P.N. Rao (✉)
Computer Science and Engineering, University of Washington, Seattle, WA 98125, USA
e-mail: xm@cs.washington.edu

R.P.N. Rao
e-mail: rao@cs.washington.edu

1 Introduction

Structured prediction is an important machine learning problem that occurs in many different fields, e.g., natural language processing, protein structure prediction and semantic image annotation. The goal is to learn a function that maps an input vector $\mathbf{X}$ to an output $\mathbf{Y}$, where $\mathbf{Y}$ is a vector representing output labels whose components take on the value $+1$ or -1 (presence or absence of the corresponding label).[1] The traditional approach to such classification problems is to train a set of binary classifiers independently. Structured prediction on the other hand also considers the relationships among the output variables $\mathbf{Y}$. For example, in the image annotation problem, an entire image or parts of an image are annotated with labels representing an object, a scene or an event involving multiple objects (Carneiro et al. 2007). These labels are usually dependent on each other, e.g., buildings and beaches occur under the sky, a truck is a type of automotive, and sunsets are more likely to co-occur with beaches, sky, and trees (Fig. 1). Such relations capture the semantics among the labels and play an important role in human cognition. A major advantage of structured prediction is that the structured representation of the output can be much more compact than an unstructured classifier, resulting in smaller sample complexity and greater generalization (Bengio et al. 2007).

Extending traditional classification techniques to structured prediction is difficult because of the potentially complicated inter-dependencies that may exist among the output variables. The problem can be modelled using a *probabilistic graphical model*, but it is well-known that exact inference over a general graph is NP-hard. Therefore, practical

[1] Structured prediction problems can involve continuous variables. In this paper, we focus only on discrete output variables.