Mach Learn
DOI 10.1007/s10994-013-5380-5

Discriminative extended canonical correlation analysis for pattern set matching

Ognjen Arandjelović

Received: 29 November 2011 / Accepted: 2 May 2013

Abstract In this paper we address the problem of matching sets of vectors embedded in the same input space. We propose an approach which is motivated by *canonical correlation analysis* (CCA), a statistical technique which has proven successful in a wide variety of pattern recognition problems. Like CCA when applied to the matching of sets, our *extended canonical correlation analysis* (E-CCA) aims to extract the most similar modes of variability within two sets. Our first major contribution is the formulation of a principled framework for robust inference of such modes from data in the presence of uncertainty associated with noise and sampling randomness. E-CCA retains the efficiency and closed form computability of CCA, but unlike it, does not possess free parameters which cannot be inferred directly from data (inherent data dimensionality, and the number of canonical correlations used for set similarity computation). Our second major contribution is to show that in contrast to CCA, E-CCA is readily adapted to match sets in a discriminative learning scheme which we call *discriminative extended canonical correlation analysis* (DE-CCA). Theoretical contributions of this paper are followed by an empirical evaluation of its premises on the task of face recognition from sets of rasterized appearance images. The results demonstrate that our approach, E-CCA, already outperforms both CCA and its quasi-discriminative counterpart constrained CCA (C-CCA), *for all values of their free parameters*. An even greater improvement is achieved with the discriminative variant, DE-CCA.

Keywords Set · Matching · Vectors · Principal · Angles

1 Introduction

Central to any applied problem of pattern recognition is the issue of how the entities of interest should be represented. A numerical description based on readily measurable quantities is sought, one which (as much as possible) minimizes variability due to confounding factors

Editor: Jean-Philippe Vert.

O. Arandjelović (✉)
Centre for Pattern Recognition and Data Analytics, Deakin University, Geelong, VIC 3216, Australia
e-mail: ognjen.arandjelovic@gmail.com

Published online: 14 August 2013