Mach Learn (2013) 92:225–250
DOI 10.1007/s10994-013-5357-4

Bayesian object matching

Arto Klami

Received: 14 January 2013 / Accepted: 11 April 2013 / Published online: 30 April 2013

Abstract Matching of object refers to the problem of inferring unknown co-occurrence or alignment between observations or samples in two data sets. Given two sets of equally many samples, the task is to find for each sample a representative sample in the other set, without prior knowledge on a distance measure between the sets. Given a distance measure, the problem would correspond to a linear assignment problem, the problem of finding a permutation that re-orders samples in one set to minimize the total distance. When no such measure is available, we need to consider more complex solutions. Typical approaches maximize statistical dependency between the two sets, whereas in this work we present a Bayesian solution that builds a joint model for the two sources. We learn a Bayesian canonical correlation analysis model that includes a permutation parameter for re-ordering the samples in one of the sets. We provide both variational and sampling-based inference for approximative Bayesian analysis, and demonstrate on three data sets that the resulting methods outperform the earlier solutions.

Keywords Canonical correlation analysis · Matching · Permutation · Bayesian analysis

1 Introduction

The task in object matching is to learn correspondence of samples in two data sets. A classical example considers a set of agents and another set of jobs, and the task is to assign each job for exactly one agent. For each agent-job pair we have a specific cost, corresponding, for example, to how well they perform the job or how much it costs, and the goal is to find the assignment that minimizes (or maximizes) the total assignment or matching cost. The problem is known as the maximum weight matching in a bipartite graph, or as linear assignment problem. Efficient polynomial time algorithms exist for finding the match, such as the classical Hungarian algorithm (Kuhn 1955).

Editors: Zhi-Hua Zhou, Wee Sun Lee, Steven Hoi, Wray Buntine, and Hiroshi Motoda.

A. Klami (✉)
Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, Helsinki, Finland
e-mail: arto.klami@hiit.fi