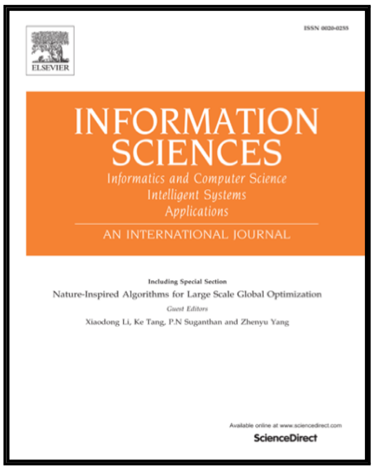
Accepted Manuscript



An efficient method for clustered multi-metric learning

Bac Nguyen, Francesc J. Ferri, Carlos Morell, Bernard De Baets

PII: S0020-0255(18)30673-X
DOI: https://doi.org/10.1016/j.ins.2018.08.055
Reference: INS 13899

To appear in: *Information Sciences*

Received date: 8 January 2018
Revised date: 8 June 2018
Accepted date: 25 August 2018

Please cite this article as: Bac Nguyen, Francesc J. Ferri, Carlos Morell, Bernard De Baets, An efficient method for clustered multi-metric learning, *Information Sciences* (2018), doi: https://doi.org/10.1016/j.ins.2018.08.055

An efficient method for clustered multi-metric learning

Bac Nguyen[a,*], Francesc J. Ferri[b], Carlos Morell[c], Bernard De Baets[a]

[a] *KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure links 653, 9000 Ghent, Belgium.*
[b] *Departament d'Informàtica, Universitat de València, 46100 Burjassot, València, Spain.*
[c] *Computer Science Department, Universidad Central "Marta Abreu" de Las Villas, Santa Clara, Cuba.*

Abstract

Distance metric learning, which aims at finding a distance metric that separates examples of one class from examples of the other classes, is the key to the success of many machine learning tasks. Although there has been an increasing interest in this field, learning a global distance metric is insufficient to obtain satisfactory results when dealing with heterogeneously distributed data. A simple solution to tackle this kind of data is based on kernel embedding methods. However, it quickly becomes computationally intractable as the number of examples increases. In this paper, we propose an efficient method that learns multiple local distance metrics instead of a single global one. More specifically, the training examples are divided into several disjoint clusters, in each of which a distance metric is trained to separate the data locally. Additionally, a global regularization is introduced to preserve some common properties of different clusters in the learned metric space. By learning multiple distance metrics jointly within a single unified optimization framework, our method consistently outperforms single distance metric learning methods, while being more efficient than other state-of-the-art multi-metric learning methods.

Keywords: Multi-metric learning, Heterogeneously distributed data, Nearest-neighbor classification, Convex optimization.

1. Introduction

The notion of similarity between objects or examples plays a key role in several machine learning tasks, including classification [28, 44, 47] and ranking [25]. There is often no obvious way of defining a (dis)similarity measure. Rather than using a default distance metric such as the Euclidean one, it is desirable to learn a distance metric that satisfies certain conditions, depending on the application domain. If some side-information is given, for instance, as provided by human

*Corresponding author
Email address: bac.nguyencong@ugent.be (Bac Nguyen)