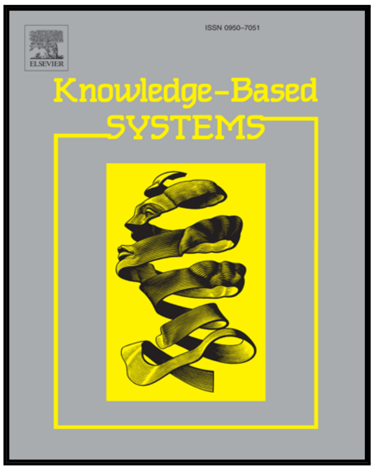
Accepted Manuscript



Sparse Feature Space Representation: A Unified Framework for Semi-Supervised and Domain Adaptation Learning

Long Liu , Lechao Yang , Bin Zhu

PII: S0950-7051(18)30240-5
DOI: 10.1016/j.knosys.2018.05.011
Reference: KNOSYS 4331

To appear in: *Knowledge-Based Systems*

Received date: 5 January 2018
Revised date: 8 May 2018
Accepted date: 9 May 2018

Please cite this article as: Long Liu , Lechao Yang , Bin Zhu , Sparse Feature Space Representation: A Unified Framework for Semi-Supervised and Domain Adaptation Learning, *Knowledge-Based Systems* (2018), doi: 10.1016/j.knosys.2018.05.011

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Sparse Feature Space Representation: A Unified Framework for Semi-Supervised and Domain Adaptation Learning

Long Liu [a], Lechao Yang [a], Bin Zhu [b*]

[a] (School of Automation and Information Engineering, Xi'an University of Technology Xi'an 710048, China)

[b] (College of Communication Engineering, Chongqing University, Chongqing 400040, China)

Abstract: In a semi-supervised domain adaptation (DA) task, one has access to only few labeled target examples. In this case, the success of DA needs the effective utilization of a large number of unlabeled target data to extract more discriminative information that is useful for generalization. To this end, we exploit in this paper the feature space embeddings of the target data as well as multi-source prior models to augment the discrimination space for the target function learning. Therefore, we propose a novel multi-source adaptation learning framework based on Sparse Feature Space Representation (SFSR), or called SFSR-MSAL for short. Specifically, the SFSR algorithm is first presented for the further construction of robust graph, on which the discriminative information can be smoothly propagated into the unlabeled target data by additionally incorporating the geometric structure of the target data. Considering the robustness in the semi-supervised DA, we replace the traditional l_2-norm based least squares regression with the $l_{2,1}$-norm sparse regression, and then construct the SFSR-graph based semi-supervised DA framework with multi-source adaptation constraints. Our framework is universal and can be easily degraded into semi-supervised learning by just tuning the regularization parameter. Moreover, to select the discriminative SFSR-graph Laplacians, we also introduce the ensemble SFSR-graph Laplacians regularization into SFSR-MSAL, thus further improving the performance of SFSR-MSAL. The validity of our methods including semi-supervised and DA learning are examined by several visual recognition tasks on some benchmark datasets, which demonstrate the superiority of our methods in comparison with other related state-of-the-art algorithms.

Keywords: Domain adaptation; sparse representation; Laplacian regularization; feature space embedding

1. Introduction

Traditional learning tasks usually require large amounts of training data to establish a classifier with satisfactory generalization capability. However, the acquisition of labeled training data is usually nontrivial for one learning task in that it needs to manually annotate input data with ground-truth labels by experts, which is often difficult, expensive, and time-consuming [1]. In the past decades, semi-supervised learning (SSL) has become a classical paradigm by exploiting a large number of unlabeled data [2]. The success of SSL is mainly attributed to certain assumptions that holds for the data distribution (also referred to as a cluster or a manifold assumption [3]), which considers both local and structure smoothness of the data distribution [4]. In particular, the manifold assumption has been applied for regularization where the geometric structure behind labeled and unlabeled data is explored with a graph-based representation [3]. In such a representation, examples are expressed as the vertices and the pairwise similarity between data is described as a weighted edge. Thus, graph-based SSL (GSSL) algorithms make good use of the manifold structure to propagate the known label information over the graph into the unlabeled data [5, 8, 11, 12].

Although most of existing GSSL algorithms have shown promising achievements in their specific applications, there still exist several limitations in them, e.g., performance sensitivities to model parameters and noise data [8]. Besides, while exploiting the vast amount of unlabeled data directly in the GSSL paradigm is valuable in its own right, it is still beneficial to leverage a plenty of labeled data of relevant categories across domains. For example, it is increasingly popular to enrich our limited collection of training data with those from the Internet. One problem with this strategy, however, arises from the possible misalignment between the target domain of interest and the auxiliary (or source) domain that provides some prior information. This makes it harmful to directly incorporate data from the source domain into the target domain. Addressing this issue has inspired recent research efforts into the domain adaptation (DA) problems [14, 15] in computer vision and machine learning [16]. Recently, many cross-domain learning techniques have been

* Corresponding author. E-mail address: zhubin@cqu.edu.cn.

ACCEPTED MANUSCRIPT

Download English Version:

https://daneshyari.com/en/article/6861311

Download Persian Version:

https://daneshyari.com/article/6861311

Daneshyari.com