NeuroImage 55 (2011) 1109–1119

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Semi-supervised pattern classification of medical images: Application to mild cognitive impairment (MCI)

Roman Filipovych *, Christos Davatzikos
and For the Alzheimer's Disease Neuroimaging Initiative [1]

Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, 3600 Market St., Suite 380, Philadelphia, PA 19104, USA

A R T I C L E I N F O

Article history:
Received 16 July 2010
Revised 9 December 2010
Accepted 24 December 2010
Available online 31 December 2010

Keywords:
Semi-supervised classification
Semi-supervised SVM
Alzheimer's
MCI

A B S T R A C T

Many progressive disorders are characterized by unclear or transient diagnoses for specific subgroups of patients. Commonly used supervised pattern recognition methodology may not be the most suitable approach to deriving image-based biomarkers in such cases, as it relies on the availability of categorically labeled data (e.g., patients and controls). In this paper, we explore the potential of semi-supervised pattern classification to provide image-based biomarkers in the absence of precise diagnostic information for some individuals. We employ semi-supervised support vector machines (SVM) and apply them to the problem of classifying MR brain images of patients with uncertain diagnoses. We examine patterns in serial scans of ADNI participants with mild cognitive impairment (MCI), and propose that in the absence of sufficient follow-up evaluations of individuals with MCI, semi-supervised strategy is potentially more appropriate than the fully-supervised paradigm employed up to date.

Introduction

High-dimensional pattern classification has gained significant attention in recent years, and has been found to be a promising technique for capturing complex spatial patterns of pathological brain changes (Davatzikos et al., 2009; Fan et al., 2008c; Vemuri et al., 2009; McEvoy et al., 2009; Hinrichs et al., 2009; Duchesne et al., 2010; Kloppel et al., 2008). Importantly, pattern classification methods have begun to provide tests of high sensitivity and specificity on an individual patient basis, in addition to characterizing group differences. As the result, these methods can potentially be used as diagnostic and prognostic tools. Pattern classification approaches were shown to work particularly well in the task of classifying patient populations from normal cohort in various clinical studies (e.g., Alzheimer's (Duchesne et al., 2010; Fan et al., 2008a; Kloppel et al., 2008; Misra et al., 2009), autism (Ecker et al., 2010), schizophrenia (Fan et al., 2008b), etc.).

The state-of-the-art brain image classification methods work by learning a classification function from a set of labeled training examples, and then apply the learned classifier to predict labels of the test data. These methods belong to the family of supervised classification approaches and assume that the labels for all training data are available. Depending on the machine learning method applied, there are many different classification functions that separate a given pair of classes. Support vector machines (SVM) have been shown to provide high classification accuracy, and are among the most widely used classification algorithms in the brain MRI classification literature (Fan et al., 2008a; Kloppel et al., 2008; Misra et al., 2009; Ecker et al., 2010; Fan et al., 2008b). However, many disorders, especially progressive ones, are characterized by uncertain or transient diagnoses for specific subgroups of patients. For example, one might be interested in classifying subjects with mild cognitive impairment (MCI) into classes that either exhibit or do not exhibit future convergence to Alzheimer's disease (AD). Unfortunately, many subjects are likely to have insufficient follow-up studies to be called converters or non-converters with high confidence. Training a supervised classifier in the scenarios where diagnoses (i.e., labels) are uncertain or unavailable may not be appropriate. Semi-supervised classification approaches are specifically designed to handle cases where only part of the data is labeled. These approaches simultaneously use both labeled and unlabeled data to infer a classifier that provides *good* classification of the unlabeled data into the two classes.

Semi-supervised SVM (Vapnik, 1998) extend the theory of traditional SVM to the case of partially labeled datasets, and offer both the accuracy of traditional SVM, and the ability to use unlabeled data to learn more reliable classification functions. Additionally, semi-supervised SVM have been shown to be more efficient than the traditional SVM in problems with a small number of labeled examples (Joachims, 1999). One of the reasons why semi-supervised SVM learning can benefit from unlabeled data is that unlabeled data can

* Corresponding author. Fax: +1 (215) 614 0266.
E-mail address: roman.filipovych@uphs.upenn.edu (R. Filipovych).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://loni.ucla.edu//ADNI//Collaboration//ADNI_Authorship_list.pdf.

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.12.066

help the classifier better learn the structure of the manifold on which image samples lie. A schematic example of one of the benefits that consideration of unlabeled data provides is depicted in Fig. 1. While a fully-supervised classifier can be constructed to separate labeled points as in Fig. 1(a), it fails to generalize well if the actual distribution is more complex than the distribution of the labeled instances (i.e., Fig. 1(b)). In contrast, semi-supervised SVM considers both labeled and unlabeled data, and may be more appropriate in the scenario where the labeled population does not entirely reflect the structure of the data.

The application focus of this paper is on Alzheimer's Disease (AD). The incidence of Alzheimer's Disease (AD) doubles every 5 years after the age of 65, rendering the disease the major cause for dementia, as well as a very important health and socioecomic issue, particularly in view of increasing life expectancy (Bain et al., 2008; Hebert et al., 2001). Although most currently approved treatments are symptomatic and do not directly slow AD pathology progression, it is anticipated that new disease modifying treatments will be available in the near future. It is also expected that treatment decisions will greatly benefit from diagnostic and prognostic tools that identify individuals likely to progress to dementia sooner. This is especially important in individuals with mild cognitive impairment (MCI), who present a conversion rate of approximately 15% per year.

The task of predicting short term conversion to AD from MCI has been addressed in the past with the help of fully supervised techniques that aim at deducing a decision function from a set of labeled images (e.g., normal control, AD, MCI-Converters, etc.) (Duchesne et al., 2008; Fan et al., 2008a; Kloppel et al., 2008; Misra et al., 2009). However populations of individuals with MCI are highly heterogeneous. Previous studies suggest that some MCIs are close to AD and will convert soon, whereas some will remain stable for over a decade. Moreover, while some individuals with MCI may convert at a faster rate than others to AD, some will never develop AD and others may develop other forms of dementia. At the same time, some individuals might be labeled with relatively higher reliability. For example, AD patients are undoubtedly converters, as well as normal control subjects are non-converters. Semi-supervised SVM do not make use of uncertain labels when building a classification function, but rather attempt to separate unlabeled data into two classes in such a way that the heterogeneity of the data is disentangled, and that the classifier agrees with the reliably labeled part of the data. As the result, classification of MCI populations is likely to benefit from the semi-supervised SVM.

In this paper, we explore the potential of semi-supervised pattern classification to provide image-based biomarkers of progressive disorders in the absence of certain diagnostic information for some patients. We present a general framework that allows to detect patterns of brain pathology using a high-dimensional semi-supervised pattern classification method that is not biased by the uncertain information about the subjects' current diagnoses. We apply our approach in the ADNI study, and investigate patterns of brain atrophy that are characteristic of AD-like MCI, and which often predict conversion to AD.

Methods and materials

Methods

Semi-supervised SVM

In the two-class classification scenario, the task of classifying images into two classes (e.g., patients vs. controls) can be viewed as the task of finding a decision function that separates the two classes in a high-dimensional space. Traditional linear SVM algorithm (Vapnik, 1995) finds this decision function as the separating hyperplane with the largest margin, where the margin is the distance from the separating hyperplane to the closest training examples. Given a set of points (i.e., images) $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, and their respective labels $\{y_1, \ldots, y_n\}$, the task of finding a separating, i.e., classification, function $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$ within the framework of traditional linear SVM could be formulated as the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{w},b,\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + \beta\sum_{i=1}^{n} \xi_i \\ &\mathbf{s.t.}\ y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1 - \xi_i, \forall i = 1, \ldots, n \\ &\xi_i \geq 0, \forall i = 1, \ldots, n \end{aligned} \tag{1}$$

where the slack variables ξ_i are introduced to allow some amount of misclassification in the case of non-separable classes, and constant β implicitly controls the tolerable misclassification error. Fig. 2(a) shows a simplified example of the supervised SVM for a two-dimensional problem. Training examples that lie on the margin define

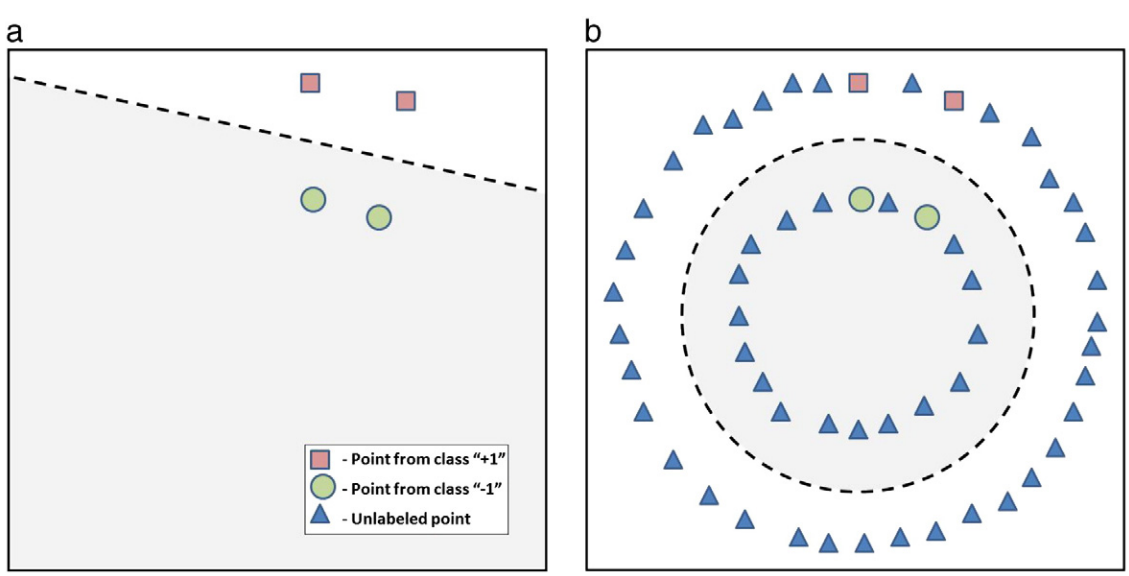

Fig. 1. Benefits of semi-supervised SVM. (a) Fully supervised classifier (represented by a dashed line) does not consider unlabeled data, and separates only the labeled points. (b) Semi-supervised classifier (represented by a dashed circle) considers unlabeled data points, and may have a better generalization ability.