Knowledge-Based Systems 66 (2014) 82–91

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

# Modelling contextual constraints in probabilistic relaxation for multi-class semi-supervised learning

Adolfo Martínez-Usó *, Filiberto Pla, José M. Sotoca

*Institute of New Imaging Technologies – Dept. of Computer Languages and Systems, Universitat Jaume I, Av. Sos Baynat s/n, 12071 Castellón, Spain*

A R T I C L E I N F O

*Article history:*
Received 26 December 2013
Received in revised form 14 April 2014
Accepted 15 April 2014
Available online 30 April 2014

*Keywords:*
Semi-supervised
Probabilistic relaxation
Database classification
Hyperspectral images
Multi-class

A B S T R A C T

This paper proposes a semi-supervised approach based on probabilistic relaxation theory. The algorithm performs a consistent multi-class assignment of labels according to the contextual information constraints. We start from a fully connected graph where each initial sample of the input data is a node of the graph and where only a few nodes have been labelled. A local propagation process is then performed by means of a support function where a new compatibility measure has been proposed. Contributions also include a comparative study of a wide variety of data sets with recent and well-known state-of-the-art algorithms for semi-supervised learning. The results have been provided by an analysis of their statistical significance. Our methodology has demonstrated a noticeably better performance in multi-class classification tasks. Experiments will also show that the proposed technique could be especially useful for applications such as hyperspectral image classification.



## 1. Introduction

There are many tasks which are too specialised to allow the use of unsupervised techniques and, at the same time, too time-consuming if every single case is to be solved manually by an expert. These tasks are especially common in remote sensing and medical imaging applications, images and videos in the web, speech signals, handwritten digits/letters/words recognition, etc., where a huge amount of data is available but only a few are labelled. Therefore, there is an increasing demand in many applications fields for a process that is able to annotate such data in an accurate way without too much participation of an expert [1].

In order to meet these demands, semi-supervised learning (SSL) has received increasing attention in recent years. Graph in Fig. 1 illustrates how the interest in SSL has been growing. This tendency in the literature is shown by means of the records collected from [2,3] over the last decade. Semi-supervised approaches arise from the idea of using a large amount of unlabelled data together, which is often cheap and easy, and a small amount of labelled data, which is hard to obtain since it requires costly human-expert time or special devices. The important point here is to obtain a better solution than that which is derived from the unlabelled data alone [4], thus reducing the annotation effort.

There are actually two different SSL settings depending on the goal of the semi-supervised classification. *Inductive semi-supervised learning* predicts the labels on future test data whereas *transductive semi-supervised learning* predicts the labels on the unlabelled instances in the training data. Transductive semi-supervised approaches are especially useful in applications where all the data is available and not much supervised information can be achieved. Despite the works that extend transduction to induction algorithms [5], there is still a debate in the machine learning community on these learning philosophies. However, it is well recognized that transductive learning provides an important insight into the exploitation of unlabelled data [6]. For instance, in remote sensing, a comprehensive analysis of the contents of all the bands and classes that form a hyperspectral image could be very time-consuming. However, if a small amount of data is provided by an expert, the whole data can be labelled with acceptable accuracy. The work here presented is focused on solving the transductive problem in multi-class classification and annotation scenarios.

Semi-supervised learning algorithms can be categorized into several paradigms [4], including generative parametric models, semi-supervised support vector machines, graph-based approaches and, more recently, the disagreement-based semi-supervised learning [6], especially the co-training algorithm [7]. The methodology proposed here belongs to the graph-based SSL methods, which is probably the most important sub-family within the SSL techniques. Graph-based methodologies assume that labelled and unlabelled data may be reasonably represented by a graph, where nodes

* Corresponding author. Tel.: +34 964728359.
*E-mail address:* auso@uji.es (A. Martínez-Usó).

http://dx.doi.org/10.1016/j.knosys.2014.04.023
0950-7051/

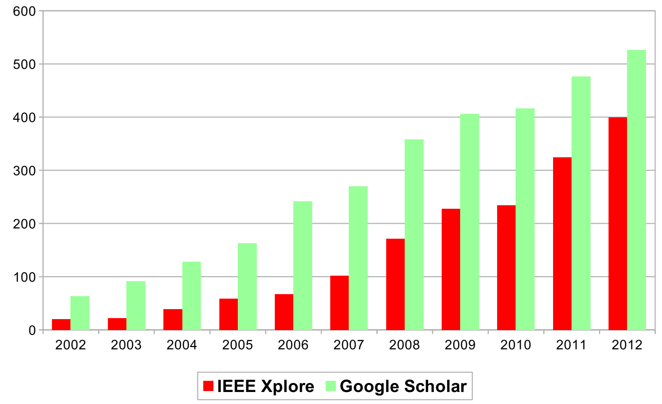


**Fig. 1.** Timeline and number of research works records on SSL from 2002 to 2012 (records collected in June 2013).

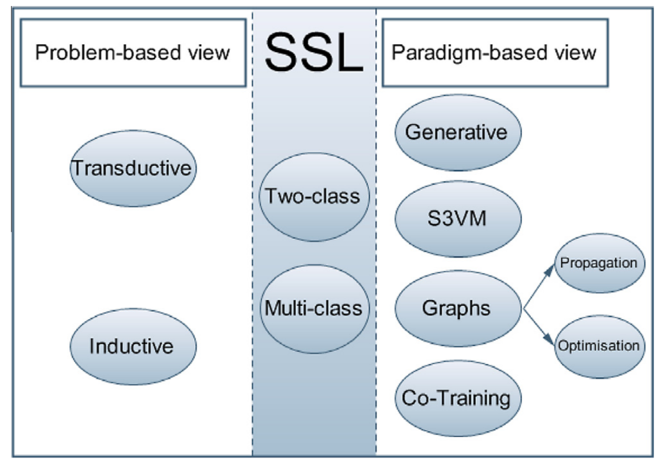


**Fig. 2.** SSL Taxonomy, overview on SSL methodologies.

represent the training data and edges represent similarities between them. Thus, a graph-based approach assumes that (i) the inherent manifold structure defined by the unlabelled data is correlated to the classification goal and (ii) closer data points tend to have similar class labels. Note that the former assumption is global, whereas the latter one is local. These assumptions lead us to the so called *Cluster-Manifold assumption*. Graph-based methods use a graph to approximate the manifold (and density) of the data so that samples that can be connected via a path through high density regions which are likely to have the same label.

Graph-based methods can be mainly divided into two categories. On the one hand, there are some methodologies that iteratively spread label information from each sample to its neighbours until a global stable state is reached [8,9]. On the other hand, there are also methodologies that optimise a loss function based on smoothness constraints derived from the graph [10–12]. In any case, the graph construction is carried out on the basis of an affinity matrix in which connections among samples are somehow described.

One of the best known approaches within graph-based SSL is [8], where authors propose a label propagation through dense unlabelled data regions. A similar label propagation algorithm is [9], where each node of the graph receives, in addition to its initial value, a normalized weighted contribution from its neighbours. This algorithm is inherently multi-class, being often used as a reference for comparison purposes. The Laplacian SVM [12] is probably the most referenced algorithm in this SSL family, due to the good performance that it generally provides. It finds class labels using a Laplacian graph for minimising inconsistencies between supervised and unlabelled data. Authors in [13] presented two algorithms, *ms3vm-iter* and *ms3vm-mkl*. These algorithms contribute to the state-of-the-art with an efficient design for the Semi-Supervised Support Vector Machines (S3VM) approaches, also achieving highly competitive or better performances than their competitors. More recently, a novel game theory approach applied to graph transduction was presented in [14]. Its main advantage consists of eliminating the restrictions on the pairwise relationships with regard to the similarity matrix. It is also an inherently multi-class algorithm that propagates the label information to the unlabelled nodes. Fig. 2 graphically overviews how the SSL methodologies have been here presented in a taxonomic view. The left side of the graphical schema divides SSL from the point of view of the problem, which is our final aim; whereas the right side shows the different paradigms within the SSL.

Relaxation methods [15] find numerical solutions for a wide range of problems in physics and engineering and, more specifically, probabilistic relaxation has been demonstrated to be very useful for pattern recognition [16]. Relaxation approaches are *iterative* processes that are mainly used in graph-based methods, being often refined by *ad hoc* or heuristic choices [17]. These methods involve *contextual information* that describes relations between single components [18], defining a neighbourhood in accordance with the properties of the system. In addition, contextual information also involves an *a priori* knowledge of the problem which is generally provided by the domain user. A general framework and the theoretical foundations for probabilistic relaxation can be found in [18,19].

This work presents a multi-class semi-supervised algorithm based on probabilistic relaxation that, while using only a few supervised instances, reaches robust solutions and is able to compete against the main state-of-the-art algorithms in terms of classification performance. The main contributions of this work are:

- A multi-class semi-supervised learning algorithm based on probabilistic relaxation, which allows context information to be introduced into the system. The algorithm proposed here works in a transductive way, that is, our main aim is labelling those samples that we already have, thus, focusing on annotation tasks.
- A comparison between the proposed technique and the main state-of-the-art algorithms. This comparison has been carried out on a wide range of databases. The analysis of results has been statistical supported.

In Section 4.6, as a matter of introducing a real annotation scenario, an application to semi-supervised graph-based hyperspectral image classification is also shown. Following previous studies in the application of semi-supervised learning techniques [20,21], the hyperspectral image of *AVIRIS Indian Pine* is used for comparing our proposal with two multi-class algorithms in a remote sensing application environment.

## 2. Probabilistic relaxation overview

Probabilistic relaxation approaches are nonlinear processes that are mainly used on graph-based methods. Graph-based methods start by constructing a graph from training data which, in SSL, involves labelled and unlabelled samples. So, let us consider a graph where each sample is a node and edges represent similarities between them. Let us also consider as contextual information both the relationship among the nodes in the graph and the initial labels provided by the user for a few nodes. In this context, a probabilistic relaxation method is an iterative process that assigns consistent