Information Sciences 314 (2015) 152–168

Contents lists available at ScienceDirect

Information Sciences

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

Laplacian unit-hyperplane learning from positive and unlabeled examples

Yuan-Hai Shao [a,*], Wei-Jie Chen [a], Li-Ming Liu [b], Nai-Yang Deng [c,*]

[a] Zhijiang College, Zhejiang University of Technology, Hangzhou 310024, PR China
[b] School of Statistics, Capital University of Economics and Business, Beijing 100070, PR China
[c] College of Science, China Agricultural University, Beijing 100083, PR China

ARTICLE INFO

Article history:
Received 10 July 2014
Received in revised form 24 March 2015
Accepted 29 March 2015
Available online 3 April 2015

Keywords:
PU learning
Biased support vector machine
One-class support vector machine
Regularization

ABSTRACT

In machine learning and data mining, learning from positive and unlabeled examples (PU learning) has attracted a great deal of attention, and the corresponding classifiers are required because of its applications in many practical areas. For PU learning, we propose a novel classifier called Laplacian unit-hyperplane classifier (LUHC), which determines a decision unit-hyperplane by solving a quadratic programming problem (QPP). The advantages of our LUHC are as follows: (1) Both geometrical and discriminant properties of the examples are exploited, resulting in better classification performance. (2) The size of QPP to be solved is small since it depends only on the number of the positive examples, resulting in faster training speed. (3) A meaningful parameter ν is introduced to control the upper bounds on the fractions of positive examples with margin errors. Preliminary experiments on both synthetic and real data sets show high level of agreement with aforementioned hypothesis, suggesting that our LUHC is superior to biased support vector machine, spy-expectation maximization, and naive Bayes in both classification ability and computation efficiency.

1. Introduction

It has been observed that only positively labeled examples and unlabeled examples are available in many applications such as text classifications and information retrieval [36,44,16,15]. As a consequence, learning from positive and unlabeled examples (PU learning) has been received a great deal of attention since both supervised and semi-supervised classification algorithms are inapplicable [29,32,31,14,8,4,30].

For PU learning, several approaches have been proposed [24,18,17,7,47,45,43]. One family of these methods is related to one-class classification, which estimates the distribution of positive class from only positive examples, such as one-class support vector machine (OSVM) [34,49]. In order to train a classifier, only labeled positive examples are required. Since there is no information about the distribution of the negative examples, the number of positive examples should be sufficiently large so that the boundary of the positive class can be induced precisely. In fact, it has been proved that these methods are effective only for the case where the number of positive examples is large enough to capture the characteristics of the positive class, and their performance would be rather poor when this number is very small [25,48].

* Corresponding authors. Tel./fax: +86 057187313551.
E-mail addresses: shaoyuanhai21@163.com (Y.-H. Shao), dengnaiyang@cau.edu.cn (N.-Y. Deng).

http://dx.doi.org/10.1016/j.ins.2015.03.066
0020-0255/

Another family of these methods wishes to seek some examples from unlabeled examples that are reliable to be negative, and then apply supervised leaning to find more and more negative examples iteratively, such as expectation maximization (EM) [15,25,13,6] and positive examples-based learning (PEBL) [46,9]. The EM estimates the parameters in the negative model by a subset of unlabeled examples that are highly reliable to be negative and performs the EM operations to infer the labels of the other unlabeled examples. A drawback of EM is that it works well only when the number of positive examples is large enough because it splits the set of positive examples into a set of positive training examples. The PEBL starts from training a support vector machine (SVM) classifier with the positive examples and an initial set of negative examples. It then uses the obtained model to find more negative examples. The new negative examples are then used to train a new SVM classifier. During each iteration, new negative examples are identified from the unlabeled data set. In a word, the performance of the methods in this family depends heavily on the reliability of the extracted negative examples. However, their initial negative examples are usually identified by heuristical rules or weak classifiers. Therefore, these methods are unreliable in the general cases and their application is limited in practice.

The third family of these methods includes logistic regression and biased support vector machines (BSVM) [17,7,12]. It considers the PU learning to be supervised learning with noise [17], where all the unlabeled examples are assumed to be negative and the noisy labels are taken into consideration, by setting different weights in loss function. That is to say, the weight is large when a labeled positive example is misclassified to be negative, while the weight is small when a negative example is misclassified as positive. However, no direct way has been provided to set up the weights. Furthermore, the performance largely depends on the number of labeled examples [7].

For PU leaning, we propose a novel Laplacian unit-hyperplane classifier (LUHC) in this paper, where both positive and unlabeled examples are used. Comparing with EM where a larger proportion of negative examples in unlabeled examples is assumed and BSVM where the PU learning is treated as supervised learning with noise [17], our LUHC try to discover and employ both geometrical and discriminant properties of the training examples directly, yielding its nice performance. In addition, a meaningful parameter ν is introduced to control the upper bounds on the fractions of positive examples with margin errors. Further, the main cost of our LUHC is to solve a quadratic programming problem (QPP) with a rather small size depending only on the number of the positive examples, resulting in its faster training speed, especially when the number of positive examples is small. Our LUHC has been compared with BSVM, EM, and naive Bayes by numerical experiments on both synthetic and real data sets. The preliminary results show that our LUHC is less sensitive to the proportion of labeled examples, and superior to others in both classification ability and computation efficiency.

The paper is organized as follows: Section 2 briefly dwells on the famous biased SVM. Section 3 proposes our Laplacian unit-hyperplane classifier and gives the corresponding theoretical analysis. Section 4 discusses the parameters selection of our classifier. Experimental results are described in Section 5, and at last, concluding remarks are given in Section 6.

2. Related works

In this paper, we consider PU learning, i.e. the problem of learning from positive and unlabeled examples: Suppose that, for every example $x \in \mathbb{R}^n$, there is a class output $y \in \{-1, 1\}$ corresponding to negative class or positive class. Now, we have a training set

$$T = \{(x_1, y_1), \ldots, (x_l, y_l)\} \cup \{x_{l+1}, \ldots, x_{l+q}\}, \tag{1}$$

where $x_i \in \mathbb{R}^n$ is its positive example and $y_i = 1$ is a positive output, $i = 1, \ldots, l; x_j \in \mathbb{R}^n$ is an unlabeled example known to belong to one of the two classes, $j = l+1, \ldots, l+q$. The task is to find a real function $f(x)$ in R^n such that the output value of y for any x can be predicted by the sign function of $f(x)$ as

$$g(x) = \text{sign}(f(x)). \tag{2}$$

Biased support vector machines (BSVM) [23,5] treats all the unlabeled examples as negative examples with noises; that is $\{y_i = 1 : i = \{1, \ldots, l\}\}$ and $\{y_j = -1 : j = \{l+1, \ldots, l+q\}\}$. Then, following the maximal margin principle, the SVM-type classifier is built. More precisely, searching for a hyperplane in the feature spaces

$$f(x) = (w \cdot x) + b = 0, \tag{3}$$

leads to the primal optimization problem

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|w\|^2 + \frac{C_+}{2}\sum_{i=1}^{l}\xi_i + \frac{C_-}{2}\sum_{i=l+1}^{l+q}\xi_i, \\ \text{s.t.} \quad & y_i((w \cdot x_i) + b) \geqslant 1 - \xi_i, \quad i = 1, \ldots, l, \ldots, l+q, \\ & \xi_i \geqslant 0, \quad i = 1, \ldots, l, \ldots, l+q, \end{aligned} \tag{4}$$

where C_+ and C_- are positive parameters. The only difference between the above problem and the one in the standard SVM is that there is two penalty parameters in the former while only one in the latter. The reason is that the examples x_j with $j = \{l+1, \ldots, l+q\}$ should be penalized slightly because they are assumed to be negative with noisy; they may be positive actually.