Towards Making Unlabeled Data Never Hurt

制作无标签数据永远不会受到伤害

Yu-Feng Li LIYF@LAMDA.NJU.EDU.CN

Zhi-Hua Zhou ZHOUZH@LAMDA.NJU.EDU.CN

National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China

李玉峰LIYF@LAMDA.NJU.EDU.CN

周志华ZHOUZH@LAMDA.NJU.EDU.CN

南京大学新型软件技术国家重点实验室，南京210093

# Abstract

It is usually expected that, when labeled data are limited, the learning performance can be improved by exploiting unlabeled data. In many cases, however, the performances of current semi-supervised learning approaches may be even worse than purely using the limited labeled data. It is desired to have *safe* semisupervised learning approaches which never degenerate learning performance by using unlabeled data. In this paper, we focus on semisupervised support vector machines (S3VMs) and propose S4VMs, i.e., safe S3VMs. Unlike S3VMs which typically aim at approaching an optimal low-density separator, S4VMs try to exploit the candidate low-density separators simultaneously to reduce the risk of identifying a poor separator with unlabeled data. We describe two implementations of S4VMs, and our comprehensive experiments show that the overall performance of S4VMs are highly competitive to S3VMs, while in contrast to S3VMs which degenerate performance in many cases, S4VMs are never significantly inferior to inductive SVMs.

抽象

通常预期，当标记数据有限时，可以通过利用未标记数据来提高学习性能。然而，在许多情况下，当前半监督学习方法的性能可能甚至比纯粹使用有限标记数据更差。期望具有安全的半监督学习方法，其通过使用未标记的数据从不退化学习性能。在本文中，我们关注半监督支持向量机（S3VM）并提出S4VM，即安全的S3VM。与通常旨在接近最佳低密度分离器的S3VM不同，S4VM尝试同时利用候选低密度分离器，以降低使用未标记数据识别不良分离器的风险。我们描述了S4VM的两种实现，我们的综合实验表明，S4VM的整体性能与S3VM相比具有很强的竞争力，而与在许多情况下退化性能的S3VM相比，S4VM从未明显逊于归纳SVM。

# Introduction

During the past decade, many effective semi-supervised learning approaches have been developed (Chapelle et al., 2006b; Zhu, 2006; Zhou & Li, 2010). It is expected that, when labeled data are limited, the use of unlabeled data will help improve the performance. However, it has been found that the performances of current semi-supervised learning approaches may be even worse than purely using labeled data in many cases (Nigam et al., 2000; Cozman et al., 2003; Grandvalet & Bengio, 2005). It is very desired to have *safe* semi-supervised learning approaches which never degenerate performance by using unlabeled data.

1.简介

在过去十年中，已经开发了许多有效的半监督学习方法（Chapelle等，2006b; Zhu，2006; Zhou＆Li，2010）。 当标记数据有限时，预计使用未标记数据将有助于提高性能。 然而，已经发现，在许多情况下，当前半监督学习方法的性能甚至可能比纯粹使用标记数据更差（Nigam等，2000; Cozman等，2003; Grandvalet＆Bengio，2005）。 非常希望有安全的半监督学习方法，它们不会通过使用未标记的数据来降低性能。

|  |
| --- |
| Appearing in *Proceedings of the 28𝑡ℎ International Conference on Machine Learning*, Bellevue, WA, USA, 2011. Copyright 2011 by the author(s)/owner(s). |

Among popular semi-supervised learning approaches,

S3VMs (Vapnik, 1998; Bennett & Demiriz, 1999; Joachims, 1999) are based on the low-density assumption and try to learn a low-density separator which favors the decision boundary going across low-density regions in the feature space (Chapelle & Zien, 2005). These approaches have already been applied to diverse applications such as text classification (Joachims, 1999), image retrieval (Wang et al., 2003), bioinformatics (Kasabov & Pang, 2004), natural language processing (Goutte et al., 2002), etc. Similar to other semi-supervise approaches, however, it has been found that S3VMs may degenerate the performance by using unlabeled data (Zhang & Oles, 2000; Wang et al., 2003; Chapelle et al., 2006b; 2008).

在流行的半监督学习方法中，

S3VMs（Vapnik，1998; Bennett＆Demiriz，1999; Joachims，1999）基于低密度假设并尝试学习低密度分离器，这有利于决策边界穿越特征空间中的低密度区域（Chapelle） ＆Zien，2005）。 这些方法已经应用于各种应用，如文本分类（Joachims，1999），图像检索（Wang等，2003），生物信息学（Kasabov＆Pang，2004），自然语言处理（Goutte等，2002）。 然而，与其他半监督方法类似，已经发现S3VM可以通过使用未标记的数据来降低性能（Zhang＆Oles，2000; Wang等，2003; Chapelle等，2006b; 2008）。。

To address this problem, in this paper we present the S4VMs (safe S3VMs). In contrast to common S3VMs which typically focus on approaching one optimal lowdensity separator, S4VMs try to exploit multiple candidate low-density separators. Our motivation lies in the observation that, given a few labeled data and abundant unlabeled data, there usually exist more than one large-margin lowdensity separators (see Figure 1), while it is hard to decide which one is the best based on the limited labeled data. Though these low-density separators all coincide with the limited labeled data well, they are often diverse and therefore, a wrong selection may cause a large loss and result in a degenerated performance. Furthermore, the optimal objective value may deviate from the ground-truth because of the limited training data. Thus, selecting one optimal lowdensity separator according to the objective value may not be really optimal, and instead, we will try to consider all the candidate low-density separators.

为了解决这个问题，在本文中我们介绍了S4VM（安全S3VM）。与通常专注于接近一个最佳低密度分离器的常见S3VM相比，S4VM尝试利用多个候选低密度分离器。我们的动机在于观察到，鉴于一些标记数据和丰富的未标记数据，通常存在多个大边缘低密度分离器（见图1），而很难根据有限的数据确定哪一个是最好的。标记数据。尽管这些低密度分离器都与有限的标记数据很好地吻合，但它们通常是多样的，因此，错误的选择可能导致大的损失并导致性能退化。此外，由于训练数据有限，最佳目标值可能偏离地面实况。因此，根据目标值选择一个最佳低密度分离器可能不是真正的最佳，相反，我们将尝试考虑所有候选低密度分离器。

Specifically, focusing on transductive setting, we construct S4VMs by optimizing the label assignment for unlabeled instances in the worse case. Theoretical analysis discloses that if the ground-truth label assignment can be realized by a low-density separator, as assumed by current S3VMs, our S4VMs will never degenerate performance. We present two implementations of S4VMs; one tries to find diverse large-margin low-density separators based on global simulated annealing search, while the other is based on a sim-

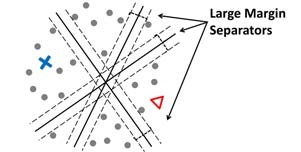


Figure1.There are usually multiple large-margin low-density separators coincide well with labeled data (cross and triangle)

图1.通常有多个大边距低密度分离器与标记数据（交叉和三角形）重合得很好

pler and efficient sampling strategy. Comprehensive experiments show that the overall performance of S4VMs are highly competitive with S3VMs, while contrasting to S3VMs which degenerate performances in many cases, our S4VMs are never significantly worse than inductive SVMs (i.e., SVMs considering only labeled data).

具体而言，关注转换设置，我们通过在最坏的情况下优化未标记实例的标签分配来构建S4VM。 理论分析表明，如果地面实况标签分配可以通过低密度分离器实现，如当前S3VM所假设的那样，我们的S4VM将永远不会降低性能。 我们提出了两种S4VM实现方式; 人们试图找到基于全局模拟退火搜索的多种大边缘低密度分离器，而另一种则基于更简单有效的采样策略。 综合实验表明，S4VM的整体性能与S3VM相比具有很强的竞争力，而与在许多情况下退化性能的S3VM相比，我们的S4VM从未明显比感应式SVM差（即仅考虑标记数据的SVM）。

The rest of this paper is organized as follows. We briefly introduce S3VMs in Section 2, and then present our S4VMs in Section 3. Experimental results are reported in Section 4, and finally we conclude the paper in Section 5.

本文的其余部分安排如下。 我们在第2节简要介绍S3VM，然后在第3节介绍我们的S4VM。实验结果在第4节中报告，最后我们在第5节中总结了论文。

# S3VMs

Inspired by the success of large margin principle, S3VMs are extensions of supervised SVMs to semi-supervised learning by simultaneously learning the optimal hyperplane and the labels for unlabeled instances. It was disclosed that S3VMs realize the low-density assumption (Chapelle & Zien, 2005) by favoring the decision boundary going across low-density regions.

2. S3VMs

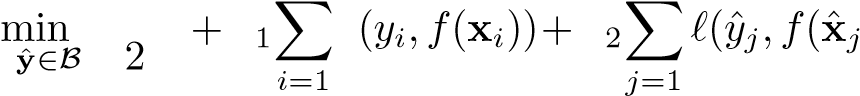
受大边缘原理成功的启发，S3VM通过同时学习最佳超平面和未标记实例的标签，将监督SVM扩展到半监督学习。 据透露，S3VM通过支持跨越低密度区域的决策边界来实现低密度假设（Chapelle＆Zien，2005）。

Formally, considering binary classification, we are given a set of labeled data {**x***𝑖,𝑦𝑖*}*𝑙𝑖*=1 and a set of unlabeled data

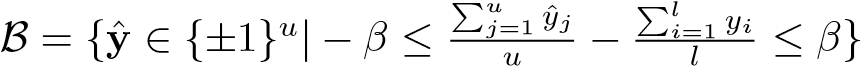
, where **x***,***x**ˆ ∈ 𝒳, *𝑦* ∈ {±1}, *𝑙* and *𝑢* are the number of labeled and unlabeled instances, respectively. The goal is to find a function *𝑓* : 𝒳 → {±1} and **y**ˆ ∈ {±1}*𝑢* such that the following functional is minimized:

形式上，考虑二进制分类，我们给出一组标记数据{x𝑖，𝑦𝑖}𝑙𝑖= 1和一组未标记数据

  ，其中x，x∈𝒳，𝑦∈{±1}，𝑙和respectively分别是标记和未标记实例的数量。 目标是找到一个函数𝑓：𝒳→{±1}和y∈{±1}𝑢，以便最小化以下函数：

 ∥*𝑓*∥ℋ *𝐶 𝑙 ℓ*  ))*,* (1)

*𝑓,*

where

is induced by the balance constraint to avoid trivial solution (Joachims, 1999; Chapelle et al., 2008), ℋ is the Reducing Kernel Hilbert Space (RKHS) induced by a kernel *𝑘* and *ℓ*(*𝑦,𝑓*(**x**)) = max{0*,𝑦𝑓*(**x**) − 1} is the hinge loss, *𝐶*1 and *𝐶*2 are regularization parameters trading off the complexity and the empirical error on label and unlabeled data, respectively. As can be seen from Eq. 1, S3VMs enforce the decision boundary to lie in low-density regions, and otherwise a large loss will occur with respective to the objective function (Chapelle & Zien, 2005).

∥𝑓∥ℋ𝐶𝑙ℓ）），（1）

𝑓，

哪里

由平衡约束诱导以避免微不足道的解（Joachims，1999; Chapelle等，2008），ℋ是由核𝑘和l（𝑦，𝑓（x））= max诱导的还原核希尔伯特空间（RKHS） {0，𝑦𝑓（x） - 1}是铰链损失，𝐶1和𝐶2分别是标准化和未标记数据的复杂性和经验误差的正则化参数。 从方程式可以看出。 如图1所示，S3VM强制决策边界位于低密度区域，否则将与目标函数相关地发生大量损失（Chapelle＆Zien，2005）。

Unlike inductive SVMs with convex formulation, the formulation of S3VMs (i.e., Eq. 1) is non-convex and the optimal solution is intractable in general. Great efforts have been devoted to avoiding S3VMs getting stuck in poor local minima. Roughly, there are three categories of approaches. The first kind of approaches are based on global combinatorial optimization (e.g., branch-andbound search), and achieve promising performances on small data sets (Bennett & Demiriz, 1999; Chapelle et al., 2007). The second kind of approaches are based on global heuristic search, which gradually increases the difficulty of solving the non-convex part in Eq. 1. Examples include the TSVM (Joachims, 1999) which gradually increases the value of *𝐶*2, the deterministic annealing approach (Sindhwani et al., 2006) which gradually increases the temperature of an entropy function, the continuation method (Chapelle et al., 2006a) which gradually decreases the smoothing of a surrogate function, etc. The third kind of approaches are based on convex relaxation, which transforms Eq. 1 into a relaxed convex problem. Examples include the semi-definite programming (SDP) relaxation (De Bie & Cristianini, 2004; Xu & Schuurmans, 2005), the minimax relaxation (Li et al., 2009b;a), etc.

与具有凸形公式的归纳SVM不同，S3VM（即，方程1）的公式是非凸的，并且最佳解决方案通常是难以处理的。我们已经付出了巨大的努力来避免S3VM陷入糟糕的本地最低点。粗略地说，有三类方法。第一种方法基于全局组合优化（例如，分支和边界搜索），并且在小数据集上实现有希望的性能（Bennett＆Demiriz，1999; Chapelle等，2007）。第二种方法是基于全局启发式搜索，这逐渐增加了求解方程中非凸部分的难度。 1.例子包括TSVM（Joachims，1999），它逐渐增加𝐶2的值，确定性退火方法（Sindhwani等，2006）逐渐增加熵函数的温度，延续方法（Chapelle等， 2006a）逐渐减少代理函数的平滑等。第三种方法基于凸松弛，它改变了方程。 1成松弛凸问题。例子包括半定规划（SDP）放松（De Bie＆Cristianini，2004; Xu＆Schuurmans，2005），极小极大松弛（Li et al。，2009b; a）等。

Avoiding inappropriate local minima when approaching the optimal solution of Eq. 1 can be regarded as a strategy towards safe S3VMs; however, this is quite challenging. To the best of our knowledge, there was no proposal of safe S3VMs in literature.

在接近方程的最优解时，避免不适当的局部最小值。 1可以被视为安全S3VM的策略; 然而，这非常具有挑战性。 据我们所知，文献中没有关于安全S3VM的提议。

# S4VMs

As mentioned, given limited labeled data and abundant unlabeled data, there usually exist multiple large-margin low-density separators coincide well with the labeled data. Without further prior information for distinguishing these separators, it might be risky to select any one of them. So, we suggest to consider all these candidate separators.

3. S4VMs

如上所述，鉴于有限的标记数据和丰富的未标记数据，通常存在多个大边距低密度分离器与标记数据一致。 如果没有进一步的先验信息来区分这些分隔符，选择其中任何一个可能存在风险。 因此，我们建议考虑所有这些候选分隔符。

In the following, we first introduce how to construct S4VMs given a number of diverse large-margin separators, by optimizing the label assignment for unlabeled instances such that the worst-case performance improvement over inductive SVM is maximized; then, we present two S4VM implementations which search for diverse largemargin separators by a global simulated annealing search and an efficient sampling strategy, respectively.

在下文中，我们首先介绍如何通过优化未标记实例的标签分配来给定多个大边距分隔符来构造S4VM，以便最大限度地提高对感应式SVM的性能改进; 然后，我们提出了两个S4VM实现，分别通过全局模拟退火搜索和有效采样策略搜索各种大型分离器。

## Constructing S4VMs

Given the predictors of multiple low-density separators

, suppose that **y**∗ is the ground-truth label assignment and let **y***𝑠𝑣𝑚* denote the predictions of the inductive SVM on unlabeled data. For any label assignment **y** ∈

{±1}*𝑢*, denote 

as the increased and decreased accuracies compared to the inductive SVM, respectively. Our goal is to learn **y** such that the improved performance over the inductive SVM is maximized; this can be cast as the optimization problem:

3.1。 构建S4VM

鉴于多个低密度分离器的预测因素

  ，假设y \*是地面实况标签分配，让y𝑠𝑣𝑚表示感应SVM对未标记数据的预测。 对于任何标签分配y∈

{±1}𝑢，表示

与感应SVM相比，分别增加和减少了准确度。 我们的目标是学习y，使得感应式SVM的性能得到最大化; 这可以作为优化问题：

max *𝑢 𝑒𝑎𝑟𝑛*(**y***,***y**∗*,***y***𝑠𝑣𝑚*) − *𝜆 𝑙𝑜𝑠𝑒*(**y***,***y**∗*,***y***𝑠𝑣𝑚*)*,* (2)

**y**∈{±1}

where *𝜆* is a parameter for trading-off how much risk the user would like to undertake. For simplification of notation, we denote *𝐽*(**y***,***y**ˆ*,***y***𝑠𝑣𝑚*) as *𝑒𝑎𝑟𝑛*(**y***,***y**ˆ*,***y***𝑠𝑣𝑚*) − *𝜆 𝑙𝑜𝑠𝑒*(**y***,***y**ˆ*,***y***𝑠𝑣𝑚*) in the sequel.

max𝑢𝑢（y，y \*，y𝑠𝑣𝑚） - λ𝑙𝑜𝑠𝑒（y，y \*，y𝑠𝑣𝑚），（2）

y∈{±1}

其中λ是用于权衡用户想要承担多少风险的参数。 为了简化表示法，我们在后续中将𝐽（y，y，y𝑠𝑣𝑚）表示为𝑒𝑎𝑟𝑛（y，y，y𝑠𝑣𝑚） - λ𝑙𝑜𝑠𝑒（y，y，y𝑠𝑣𝑚）。

Note that the difficulty for solving Eq. 2 lies in the fact that the ground-truth **y**∗ is unknown; otherwise it is trivial to get the solution. Similar to existing S3VMs, here we assume that the ground-truth boundary **y**∗ can be realized by a lowdensity separator in, i.e., **y**. We consider optimizing the worst-case improvement over the inductive SVM, that is,

注意解决Eq的难度。 2事实上，地面真相y \*是未知的; 否则获得解决方案是微不足道的。 与现有的S3VM类似，这里我们假设地面实况边界y \*可以通过低密度分离器实现，即y。 我们考虑优化归纳SVM的最坏情况改进，即

**y**¯ = argmax min *𝐽*(**y***,***y**ˆ*,***y***𝑠𝑣𝑚*)*.* (3) **y**∈{±1}*𝑢* **y**ˆ∈ℳ

Theorem 1. *If* **y***, the accuracy of* **y**¯ *is never worse than that of* **y***𝑠𝑣𝑚.*

*Proof.* **y**¯ is the optimal solution and ∀**y**ˆ, *𝐽*(**y***𝑠𝑣𝑚,***y**ˆ*,***y***𝑠𝑣𝑚*) is always zero, and thus, we have min *𝐽*(**y**¯*,***y**ˆ*,***y***𝑠𝑣𝑚*) ≥ min *𝐽*(**y***𝑠𝑣𝑚,***y**ˆ*,***y***𝑠𝑣𝑚*) = 0*.* (4) **y**ˆ∈ℳ **y**ˆ∈ℳ

Since **y**∗ ∈ ℳ, we have

*𝐽*(**y**¯*,***y**∗*,***y***𝑠𝑣𝑚*) ≥ min *𝐽*(**y**¯*,***y**ˆ*,***y***𝑠𝑣𝑚*)*.* (5) **y**ˆ∈ℳ

According to Eqs. 4 and 5, we have 

0, i.e., 

call that *𝜆* ≥ 1, we have *𝑒𝑎𝑟𝑛*(**y**¯*,***y**∗*,***y***𝑠𝑣𝑚*) ≥ , and the theorem is proved.

y¯= argmax min𝐽（y，y，y𝑠𝑣𝑚）。 （3）y∈{±1}𝑢y∈ℳ

定理1.如果y，y的精度永远不会比y𝑠𝑣𝑚的精确度差。

证明。 y¯是最优解，∀y，𝐽（y𝑠𝑣𝑚，y，y𝑠𝑣𝑚）总是为零，因此，我们得到min（y，y，y𝑠𝑣𝑚）≥min𝐽（y𝑠𝑣𝑚，y，y𝑠𝑣𝑚）= 0。 （4）y∈ℳy∈ℳ

由于y \*∈ℳ，我们有

𝐽（y¯，y \*，y𝑠𝑣𝑚）≥min𝐽（y¯，y，y𝑠𝑣𝑚）。 （5）y∈ℳ

根据Eqs。 4和5，我们有

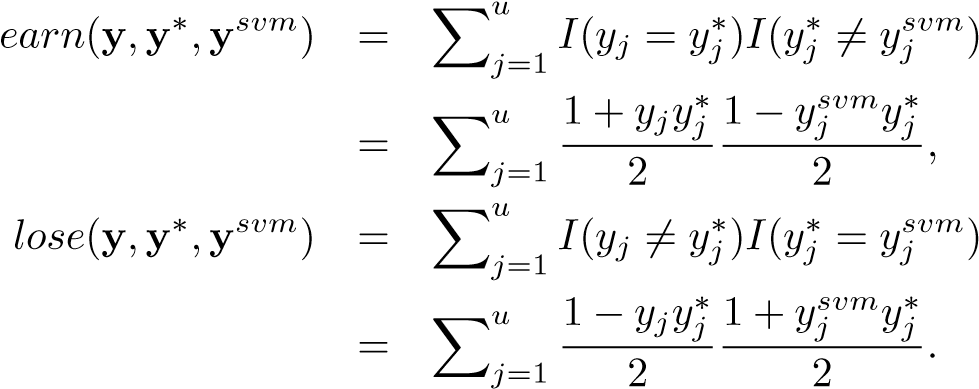
0，即

称λ≥1，我们得到y（y¯，y \*，y𝑠𝑣𝑚）≥，证明了该定理。

Theorem 1 shows that the S4VM is never worse than the inductive SVM. It is easy to get the following proposition:

Proposition 1. *If* *, the accuracy of any* **y** *satisfying* min**y**ˆ∈ℳ *𝐽*(**y***,***y**ˆ*,***y***𝑠𝑣𝑚*) ≥ 0 *is never worse than that of* **y***𝑠𝑣𝑚.*

To solve Eq. 3, note that the following are linear functions of **y**:



Without lose of generality, let *𝐽*(**y***,***y**ˆ*𝑡,***y***𝑠𝑣𝑚*) = **c**′*𝑡***y** + *𝑑𝑡*.

Eq. 3 can be cast as max *𝑢 𝜃* s.t. *𝜃* ≤ **c**′*𝑡***y** + *𝑑𝑡,*∀*𝑡* = 1*,...,𝑇.* (6)

定理1表明S4VM永远不会比归纳SVM差。 很容易得到以下命题：

命题1.如果，任何满足miny∈ℳℳ（y，y，y𝑠𝑣𝑚）≥0的y的精度永远不会比y𝑠𝑣𝑚的精度差。

解决Eq。 3，注意以下是y的线性函数：

不失一般性，让𝐽（y，y𝑡，y𝑠𝑣𝑚）=c'𝑡y+𝑑𝑡。

式。 3可以铸成最大𝑢θs.t. θ≤c'𝑡y+𝑑𝑡，∀𝑡= 1，...，𝑇。（6）

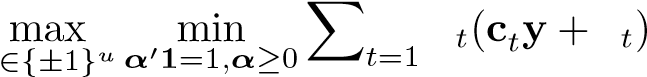
**y**∈{±1}

Though Eq. 6 is an integer linear programming, according to Proposition 1, we do not need to get the optimal solution for achieving our goal, and thus we employ a simple heuristic method to solve Eq. 6. Specifically, we first solve a convex linear programming by relaxing the integer constraint of **y** in Eq. 6 to [−1*,*1]*𝑢* and then project it back to integer solution with minimum distance. If the function value of the resulting integer solution is smaller than that of **y***𝑠𝑣𝑚*, **y***𝑠𝑣𝑚* is output as the final solution instead. It is evident that our final solution satisfies Proposition 1.

Note that prior knowledge on low-density separators can be easily incorporated into our framework. Specifically, by introducing dual variables 𝜶 for constraints in Eq. 6, according to KKT condition, Eq. 6 can be reformulated as

虽然Eq。 6是整数线性规划，根据命题1，我们不需要得到实现目标的最优解，因此我们采用简单的启发式方法来求解方程。 具体来说，我们首先通过放宽方程式中y的整数约束来求解凸线性规划。 6到[-1,1]𝑢然后将其投影回最小距离的整数解。 如果得到的整数解的函数值小于y𝑠𝑣𝑚的函数值，则输出y𝑠𝑣𝑚作为最终解。 很明显，我们的最终解决方案满足命题1。

请注意，有关低密度分离器的先验知识可以很容易地纳入我们的框架中。 具体而言，通过引入方程式中的约束的双变量α。 6，根据KKT条件，Eq。 6可以重新表述为

*𝑇*

*𝛼* ′ *𝑑 .* (7)

**y**

Here *𝛼𝑡* can be interpreted as a probability that **y**ˆ*𝑡* discloses the ground-truth solution. Hence, if prior knowledge about the probabilities 𝜶 is available, one can learn the optimal **y** with respect to the target in Eq. 7 using the known 𝜶.

It is worth mentioning that, by considering all candidate large-margin low-density separators, S4VMs are relevant to ensemble methods (Zhou, 2009), and the spirit may also be extended to other semi-supervised learning approaches.

𝑇

α'𝑑。（7）

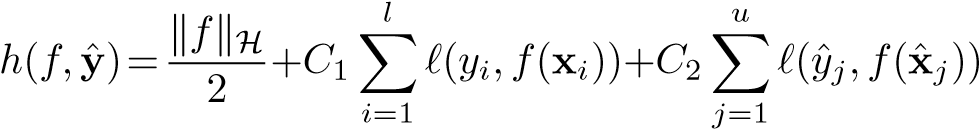
ÿ

这里α𝑡可以解释为y𝑡揭示地面实况解的概率。 因此，如果关于概率α的先验知识可用，则可以在等式1中学习关于目标的最优y。 7使用已知的α。

值得一提的是，通过考虑所有候选大边缘低密度分离器，S4VM与集合方法相关（Zhou，2009），并且精神也可以扩展到其他半监督学习方法。

## Two Implementations

Now we consider how to find diverse large-margin lowdensity separators. Let *ℎ*(*𝑓,***y**ˆ) denote the functional to be minimized by the objective function of S3VMs (i.e., Eq. 1):

*.*

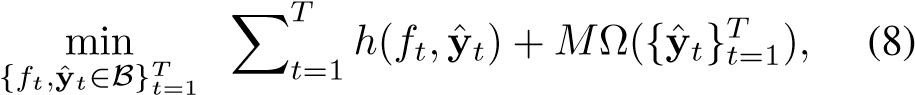
Our goal is to find multiple large-margin low-density separators and the corresponding label assignments  such that the following functional is minimized:

3.2。 两个实现

现在我们考虑如何找到各种大边缘低密度分离器。 设ℎ（𝑓，y）表示通过S3VM的目标函数（即方程1）最小化的函数：

 。

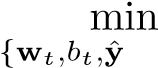
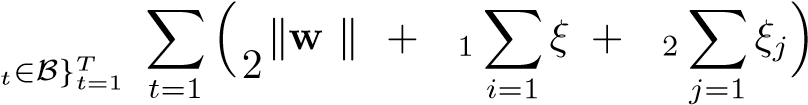
我们的目标是找到多个大边距低密度分离器和相应的标签分配，以便最大限度地减少以下功能：



where *𝑇* is the number of separators, Ω is a quantity of penalty about the diversity of separators, and *𝑀* is a large constant (e.g., 105 in our experiments) enforcing large diversity. It is evident that minimizing Eq. 8 favors the separators with large-margin as well as large diversity.

In this paper, we consider as sum of pairwise

terms, i.e., where **I** is the identity function and *𝜖* ∈ [0*,*1] is a constant, but note that other penalty quantities are also applicable.

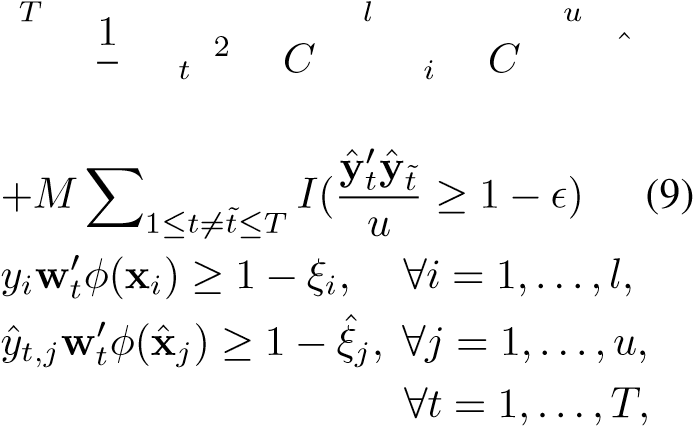
Without loss of generality, suppose that *𝑓* is a linear model, i.e., *𝑓*(**x**) = **w**′*𝜙*(**x**) + *𝑏* where *𝜙*(**x**) is a feature mapping induced by the kernel *𝑘*. Thus, Eq. 8 is cast as:

其中𝑇是分离器的数量，Ω是关于分离器多样性的惩罚量，而𝑀是一个大的常数（例如，在我们的实验中为105），强制实现大的多样性。 很明显，最小化Eq。 8有利于具有大利润和多样性的分离器。

在本文中，我们将成对视为总和

术语，即I是身份函数，ε∈[0,1]是常数，但请注意其他惩罚量也适用。

在不失一般性的情况下，假设𝑓是线性模型，即𝑓（x）=w'φ（x）+𝑏其中φ（x）是由核induced引起的特征映射。 因此，Eq。 8被视为：

s.t. 

where *𝑦*ˆ*𝑡,𝑗* refers to the *𝑗*th entry of **y**ˆ*𝑡*. Eq. 9 is non-convex and in the following we will present two solutions. It is evident that this can also be implemented by other solutions, especially those based on efficient S3VMs.

其中𝑦𝑡，𝑗是指y𝑡的第entry条。式。 图9是非凸的，下面我们将提出两种解决方案。 很明显，这也可以通过其他解决方案实现，尤其是那些基于高效S3VM的解决方案。

3.2.1. GLOBAL SIMULATED ANNEALING SEARCH

Our first implementation is based on global search, e.g., simulated annealing (SA) search (Kirkpatrick, 1984; Cernˇ y, 1985). SA is a probabilistic method for approach-` ing global solutions of objective functions suffering from multiple local minima. Specifically, at each step, SA replaces current solution by a random nearby solution with a probability depending on the value difference between their

全球模拟退火搜索

我们的第一个实现基于全局搜索，例如模拟退火（SA）搜索（Kirkpatrick，1984; Cern y，1985）。 SA是一种概率方法，用于处理受多个局部最小值影响的目标函数的全局解。 具体而言，在每个步骤中，SA通过随机附近解决方案替换当前解决方案，其概率取决于它们之间的值差异

corresponding function targets as well as a global parameter, i.e., the temperature *𝑃*, which gradually decreases during the process. When *𝑃* is large, current solution almost changes randomly; while as *𝑃* goes to zero, the changes are increasingly “downhill”. In theory, according to the convergence analysis of Markov Process, the prob ability that SA converges to the global solution approaches to 1 as SA procedure is extended (Laarhoven & Aarts, 1987).

相应的功能目标以及全局参数，即温度𝑃，在该过程中逐渐减小。 当𝑃很大时，当前的解决方案几乎随机变化; 当𝑃变为零时，变化越来越“下坡”。 理论上，根据马尔可夫过程的收敛性分析，随着SA程序的扩展，SA收敛于全局解的概率接近于1（Laarhoven＆Aarts，1987）。

To alleviate the low convergence rate of pure SA, inspired by (Sindhwani et al., 2006), a deterministic local search scheme is used. Specifically, once  are fixed,  are solved via multiple individual SVM subroutines; once are fixed, are updated based on local binary search, iteratively until convergence.

Algorithm 1 presents the pseudo-code of simulated annealing approach for Eq. 9, where the local search subroutine is given in Algorithm 2.

为了减轻纯SA的低收敛速度，受（Sindhwani等人，2006）的启发，使用确定性局部搜索方案。 具体来说，一旦修复，就可以通过多个单独的SVM子程序来解决; 一旦被修复，基于本地二进制搜索更新，迭代直到收敛。

算法1给出了方程式的模拟退火方法的伪代码。 在图9中，在算法2中给出了局部搜索子程序。

3.2.2. REPRESENTATIVE SAMPLING

To further alleviate the computational complexity, our second implementation is based on heuristic sampling search. Recall that the goal of Eq. 8 can be realized by finding mul-

tiple large-margin low-density separators and then keeping only representative ones with large diversity; this motivates Algorithm 1 Solving Eq. 9 by Simulated Annealing Search

{

**x**

*𝑖*

*,𝑦*

*𝑖*

}

*𝑙*

*𝑖*

=1

,

{

ˆ

**x**

*𝑗*

}

*𝑢*

*𝑗*

=1

,

*𝑇*

{

ˆ

**y**

*𝑏𝑒𝑠𝑡*

*𝑡*

}

*𝑇*

*𝑡*

=1

*𝑃*

←

1

,

*𝑒*

←

1

,

{

ˆ

**y**

*𝑡*

}

*𝑇*

*𝑡*

=1

(

{

ˆ

**y**

*𝑡*

}

*𝑇*

*𝑡*

=1

*,𝑜*

)

←

Localsearch

(

{

ˆ

**y**

*𝑡*

}

*𝑇*

*𝑡*

=1

)

3:

ˆ

**y**

*𝑏𝑒𝑠𝑡*

*𝑡*

←

ˆ

**y**

*𝑡*

,

∀

*𝑡*

=1

*,...,𝑇*

Input:

Output:

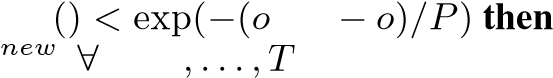
1: Initialize 2:

4: while *𝑃 > 𝑚𝑖𝑛𝑃* do

5: {**y**ˆ*𝑡𝑛𝑒𝑤*}*𝑇𝑡*=1 ← neighbour({**y**ˆ*𝑡*}*𝑇𝑡*=1)

6: Localsearch

7: if *𝑜 < 𝑜* then

8: *𝑜* ← *𝑜𝑛𝑒𝑤*; **y**ˆ*𝑡𝑏𝑒𝑠𝑡* ←**y**ˆ*𝑡* ←**y**ˆ*𝑡𝑛𝑒𝑤*, ∀*𝑡* = 1*,...,𝑇*

9: else if random *𝑛𝑒𝑤*

10: **y**ˆ*𝑡* ←**y**ˆ*𝑡* , *𝑡* = 1

11: else

12: *𝑒* ← *𝑒* + 1

13: end if

14: if *𝑒* = *𝑒𝑚𝑎𝑥* then

15: *𝑃* ← cooling(*𝑃*); *𝑒* ← 1

16: end if

17: end while

代表性抽样

为了进一步减轻计算复杂性，我们的第二个实现基于启发式采样搜索。 回想一下Eq的目标。 8可以通过找到mul-来实现

尖端的大边缘低密度分离器，然后只保留具有大量多样性的代表性分离器; 这激发了算法1求解方程。 9通过模拟退火搜索

输入：

输出：

1：初始化2：

4：而𝑃>𝑚𝑖𝑛𝑃做

5：{y𝑡𝑛𝑒𝑤}𝑇𝑡= 1←neighbor（{y𝑡}𝑇𝑡= 1）

6：Localsearch

7：如果𝑜<𝑜那么

8：𝑜←𝑜𝑛𝑒𝑤; y𝑡𝑏𝑒𝑠𝑡←y𝑡←y𝑡𝑛𝑒𝑤，∀𝑡= 1，...，𝑇

9：否则如果随机𝑛𝑒𝑤

10：y𝑡←y𝑡，𝑡= 1

11：其他

12：𝑒←𝑒+ 1

13：结束如果

14：如果𝑒=𝑒𝑚𝑎𝑥那么

15：𝑃←冷却（𝑃）; 𝑒←1

16：结束如果

17：结束

Algorithm

2

Localsearch 本地搜索

Input: {**y**ˆ*𝑡*}*𝑇𝑡*=1; (Denote [*𝑚*] = {1*,...,𝑚*})

Output: (

1: while not converged do

2: Fix, solve via multiple SVMs

3: while not converged do

4: cyclically random pick *𝑗* ∈ [*𝑢*], *𝑡* ∈ [*𝑇*]

5: optimize *𝑦*ˆ*𝑡,𝑗* ∈{±1} according to Eq. 9

6: end while

7: end while

8: Output and corresponding objective value *𝑜𝑏𝑗*

输入：{y𝑡}𝑇𝑡= 1; （表示[𝑚] = {1，...，𝑚}）

产量:(

1：虽然没有收敛

2：修复，通过多个SVM解决

3：虽然没有融合

4：循环随机选择𝑗∈[𝑢]，𝑡∈[𝑇]

5：根据方程式优化𝑦𝑡，𝑗∈{±1}。9

6：结束

7：结束

8：输出和相应的目标值𝑜𝑏𝑗

us to have a two-stage method, by searching for multiple large-margin low-density separators at first and then selecting the representative separators.

Algorithm 3 shows the pseudo-code of our second implementation. As can be seen, multiple candidate large-margin low-density separators are first obtained via local search similar to that of Algorithm 2. A clustering algorithm is then applied to identify the representative separators. This approach is simple, and experiments in Section 4 show that it is efficient and effective.

我们采用两阶段方法，首先搜索多个大边距低密度分离器，然后选择代表性分离器。

算法3显示了我们的第二个实现的伪代码。 可以看出，首先通过类似于算法2的局部搜索获得多个候选大边缘低密度分离器。然后应用聚类算法来识别代表性分离器。 这种方法很简单，第4节中的实验表明它是高效和有效的。

# Experiments

We evaluate S4VMs on a broad range of tasks including seven SSL benchmark data sets[[1]](#footnote-1), i.e., *digit1, USPS, BCI, g241c, g241n, COIL, Text*, and sixteen UCI data sets[[2]](#footnote-2). Table 1 summarizes the statistics of the data sets.

Both linear and RBF kernels are used in our experiments. As for benchmark data sets, the archival includes two sets of twelve data splits, one with 10 while the other with 100 Algorithm 3 Solving Eq. 9 by Representative Sampling

{

**x**

*𝑖*

*,𝑦*

*𝑖*

}

*𝑙*

*𝑖*

=1

,

{

ˆ

**x**

*𝑗*

}

*𝑢*

*𝑗*

=1

,

*𝑇*

{

ˆ

**y**

*𝑏𝑒𝑠𝑡*

*𝑡*

}

*𝑇*

*𝑡*

=1

𝒮

=

{

ˆ

**y**

*𝑛*

}

*𝑁*

*𝑛*

=1

Input: ;

Output:

1: Randomly sampling *𝑁* number of **y**ˆ’s, i.e.,

2: for *𝑛* = 1 : *𝑁* do

3: while not converged do

4: Fix **y**ˆ*𝑛*, solve {**w***𝑛,𝑏𝑛*} via SVM solver

5: Fix {**w***𝑛,𝑏𝑛*}, update **y**ˆ*𝑛* w.r.t S3VM’s objective function via sorting (Zhang et al., 2007)

6: end while

7: end for

8: Perform clustering (e.g., *𝑘*-means) for 𝒮 where *𝑘* = *𝑇*

9: Output **y**ˆ’s with the minimum objective value within each cluster

4.实验

我们在广泛的任务上评估S4VM，包括七个SSL基准数据集，即digit1，USPS，BCI，g241c，g241n，COIL，Text和16个UCI数据集。 表1总结了数据集的统计数据。

线性和RBF内核都用于我们的实验中。 对于基准数据集，存档包括两组十二个数据分割，一个用于10个，而另一个用100个算法3求解方程。 9代表采样

输入：;

输出：

1：随机抽样𝑁的数量，即

2：对于𝑛= 1：𝑁做

3：虽然没有融合

4：修复y𝑛，通过SVM求解器求解{w𝑛，𝑏𝑛}

5：修复{w𝑛，𝑏𝑛}，通过排序更新y𝑛w.r.tS3VM的目标函数（Zhang et al。，2007）

6：结束

7：结束

8：对𝒮执行聚类（例如，𝑘-means），其中𝑘=𝑇

9：在每个簇内输出具有最小目标值的y

Table1. Experimental data sets.

ID

Data

# Inst

# Feat

ID

Data

# Inst

# Feat

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | BCI | 400 | 117 | 13 | vehicle | 435 | 16 |
| 2 | g241c | 1500 | 241 | 14 | house-votes | 435 | 16 |
| 3 | g241d | 1500 | 241 | 15 | clean1 | 476 | 166 |
| 4 | COIL | 1500 | 241 | 16 | wdbc | 569 | 14 |
| 5 | Digit1 | 1500 | 241 | 17 | isolet | 600 | 51 |
| 6 | USPS | 1500 | 241 | 18 | breastw | 683 | 9 |
| 7 | Text | 1500 | 11960 | 19 | austra | 690 | 42 |
| 8 | house | 232 | 16 | 20 | australian | 690 | 15 |
| 9 | heart | 270 | 9 | 21 | diabetes | 768 | 8 |
| 10 | haberman | 306 | 14 | 22 | german | 1000 | 59 |
| 11 | liverDisorders 345 | | 6 | 23 | optdigits | 1143 | 42 |
| 12 | ionosphere 351 | | 33 |  |  | |  |

labeled examples. As for UCI data sets, we randomly select 10 and 100 examples to be used as labeled examples, and use the remaining data as unlabeled data. The experiments are repeated for 30 times and the average accuracy and standard deviations are recorded.

标记的例子。 对于UCI数据集，我们随机选择10和100个示例作为标记示例，并将剩余数据用作未标记数据。 将实验重复30次并记录平均准确度和标准偏差。

Inductive SVM[[3]](#footnote-3) and TSVM[[4]](#footnote-4) (Joachims, 1999) are evaluated as baselines. Linear programming is conducted by linprog function in MATLAB. The regularization parameters *𝐶*1, *𝐶*2 and *𝛽* in balance constraint are fixed as 100, 0*.*1 and 0*.*1 for all S3VMs. We call our S4VM which uses simulated annealing as S4VM*𝑎*, and the one which uses sampling as S4VM*𝑠*. For S4VM*𝑎*, *𝜖* and *𝑇* are simply fixed as 0*.*05 and 3, respectively. For S4VM*𝑠*, the sampling size *𝑁* and the number of separators *𝑇* are simply fixed as 100 and 10, respectively. *𝜆* is fixed as 3 for S4VMs. For 10 labeled examples, the width of RBF kernel is set as *𝛿*, i.e., the average distance between instances; for 100 labeled examples, the width of RBF kernel is selected by 5-fold cross validation from the set of {0*.*25*𝛿,*0*.*5*𝛿,𝛿,*2*𝛿,*4*𝛿*}.

归纳SVM和TSVM（Joachims，1999）被评估为基线。 线性编程由MATLAB中的linprog函数进行。 对于所有S3VM，平衡约束中的正则化参数𝐶1，𝐶2和β固定为100,0.1和0.1。 我们称我们的S4VM使用模拟退火作为S4VM𝑎，而使用采样作为S4VM𝑠。 对于S4VM𝑎，ε和𝑇分别固定为0.05和3。 对于S4VM𝑠，采样大小𝑁和分隔符数𝑇分别固定为100和10。 对于S4VM，λ固定为3。 对于10个标记的例子，RBF内核的宽度设置为δ，即实例之间的平均距离; 对于100个标记的实例，通过来自{0.25δ，0.5δ，δ，2δ，4δ}的集合的5倍交叉验证来选择RBF核的宽度。

## Results of S4VM*𝑎*

In addition to inductive SVM and TSVM, we also compare

S4VM*𝑎* with three variants using multiple low-density separators. S3VM*𝑏𝑒𝑠𝑡𝑎* presents the best performance among the multiple candidate separators (note that this method is impractical), S3VM*𝑚𝑖𝑛𝑎* selects the low-density separator with minimum objective value, and S3VM*𝑐𝑜𝑚𝑎* combines the candidate separators using uniform weights. Though simulated annealing was used to improve the efficiency of S3VMs (Sindhwani et al., 2006), note that it is still with high computational load, and therefore Table 2 only reports the performances on UCI data sets with RBF kernels.

4.1。 S4VM的结果

除了归纳SVM和TSVM，我们还进行了比较

S4VM𝑎有三种型号，使用多个低密度分离器。 S3VM𝑏𝑒𝑠𝑡𝑎在多个候选分隔符中表现最佳（注意此方法不切实际），S3VM𝑚𝑖𝑛𝑎选择具有最小目标值的低密度分隔符，S3VM𝑐𝑜𝑚𝑎使用统一权重组合候选分隔符。 尽管使用模拟退火来提高S3VM的效率（Sindhwani等，2006），但是注意它仍然具有高计算负荷，因此表2仅报告了具有RBF内核的UCI数据集的性能。

Table 2 shows that the overall performance of S4VM*𝑎* is highly competitive with TSVM. In terms of pairwise accuracy comparison, S4VM*𝑎* is better than TSVM on 6 out

of 12 data sets for 10 labeled examples, while this number rises to 11 out of 12 data sets for 100 labeled examples. In terms of average accuracy, S4VM*𝑎* is slightly worse (better) than TSVM for 10 (100) labeled examples. S3VM*min𝑎* and S3VM*com𝑎* do not perform as well as S4VM*𝑎*.

表2显示S4VM𝑎的整体性能与TSVM竞争激烈。 在成对精度比较方面，S4VM𝑎优于TSVM6

对于10个标记的示例，12个数据集的数量增加到12个数据集中的11个，用于100个标记的示例。 就平均准确度而言，对于10（100）个标记的示例，S4VM𝑎比TSVM略差（更好）。 S3VMmin𝑎和S3VMcom𝑎的性能不如S4VM𝑎。

More importantly, unlike TSVM which is significantly worse than inductive SVM on 4 out of 12 data sets for 10 labeled data, and 7 out of 12 for 100 labeled data, S4VM*𝑎* never degenerates the performance significantly. Both S3VM*min𝑎* and S3VM*com𝑎* are capable to reduce the chance of significantly degenerating performance compared with TSVM, however, they still degenerate performance significantly in many cases.

更重要的是，与12个数据集中的4个中的感应SVM相比，对于10个标记数据，并且对于100个标记数据中的12个中的7个，显着差于TSVM，S4VM𝑎从未显着降低性能。 与TSVM相比，S3VMmin𝑎和S3VMcom𝑎都能够降低性能显着退化的可能性，但在许多情况下，它们仍然会显着降低性能。

Though the condition of Theorem 1 is relaxed than traditional assumption of S3VMs, the theorem does not always hold owing to many factors, e.g., the ground-truth is not among the low-density separators. Even in such cases, however, S4VM may still work. Note that Theorem 1 presents a sufficient rather than necessary condition for S4VMs, and the relevance to ensemble methods provides an explanation to S4VMs’ superiority to single separators.

尽管定理1的条件比传统的S3VMs假设更宽松，但由于许多因素，定理并不总是成立，例如，地面实况不属于低密度分离器。 但即使在这种情况下，S4VM仍然可以工作。 注意，定理1为S4VM提供了充分而非必要的条件，并且与集合方法的相关性解释了S4VM对单个分隔符的优越性。

## Results of S4VM*𝑠*

Similar to S4VM*𝑎*, three variants, i.e., S3VM*best𝑠* , S3VM*min𝑠* and S3VM*com𝑠* are compared with S4VM*𝑠* in Table 3, in addition to inductive SVM and TSVM.

4.2 S4VM的结果

与S4VM𝑎类似，除了归纳SVM和TSVM之外，还将三种变体（即S3VMbest𝑠，S3VMmin𝑠和S3VMcom𝑠）与表3中的S4VM𝑠进行了比较。

Table 3 shows that the overall performance of S4VM*𝑠* is highly competitive with TSVM, though the ground-truth is seldom realized by a low-density separator (see the performance of S3VM*best𝑠* ). In terms of pairwise accuracy comparison, S4VM*𝑠* outperforms TSVM on 15/13 and 13/18 out of the 23 data sets with linear/RBF kernels for 10 and 100 labeled examples, respectively. In terms of average accuracy, S4VM*𝑠* is slightly worse (better) than TSVM for 10 (100) labeled examples. Except for the case of S3VM*min𝑠* on 100 label examples, S3VM*min𝑠* and S3VM*com𝑠* do not perform as well as S4VM*𝑠*.

表3显示S4VM𝑠的整体性能与TSVM竞争激烈，尽管低密度分离器很少实现真实性（参见S3VMbest𝑠的性能）。 在成对精度比较方面，S4VM𝑠在23个数据集中的15/13和13/18上优于TSVM，线性/ RBF内核分别用于10和100个标记示例。 就平均准确度而言，对于10（100）个标记的示例，S4VM𝑠比TSVM略差（更好）。 除了100个标签示例中的S3VMmin𝑠外，S3VMmin𝑠和S3VMcom𝑠的性能不如S4VM𝑠。

More importantly, unlike TSVM which degenerates performance on 12 and 17 cases for 10 and 100 labeled ex-

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table2.Accuracy (mean±std.) of S4VM*𝑎* and compared methods. ‘SVM’ denotes inductive SVM. For semi-supervised methods (TSVM, S3VM*best𝑎* , S3VM*com𝑎* and S4VM*𝑎*), if the performance is significantly better/worse than SVM (paired *𝑡*-tests at 95% significance level), the corresponding entries are bolded/underlined. The win/tie/loss counts are summarized in the last row, and the method with the smallest number of losses against SVM is bolded.  表2.S4VM𝑎的准确度（平均值±标准值）和比较方法。  'SVM'表示归纳SVM。 对于半监督方法（TSVM，S3VMbest𝑎，S3VMcom𝑎和S4VM𝑎），  如果性能明显优于/差于SVM（配对𝑡-测试为95％显着性水平），则相应的条目用粗体/下划线表示。  胜利/平局/损失计数总结在最后一行，并且针对SVM的损失数最小的方法是粗体。   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | # Labeled 标记 | Data | SVM | TSVM | S3VM*best𝑎* | S3VM*min𝑎* | | | S3VM*com𝑎* | | S4VM*𝑎* | | 10 | | 66.695.765.665.971.888.0±±±±±±7.52.85.58.16.82.83.1  81.087.074.755.574.380.7±±±±±±14.79.15.98.17.5  75.6 | 67.894.765.362.472.589.8±±±±±±13.20.16.26.27.61.85.8  85.077.486.754.278.784.4±±±±±±8.69.94.88.76.3  76.6 | 12.2  6.0  6.9  9.7  76.2 | 67.4  64.8  71.9  73.7  75.1 | 12.5  6.3  7.5  10.0  19.8  10.9  9.8 | | 67.2  65.2  72.3  74.0  82.4  75.2  81.4  75.7 | 76.1 | | |  | | | 4/4/4 | 1/9/2 | 2/10/0 | | | 0/10/2 |  | | | # Labeled | Data | SVM | TSVM | S3VM*best𝑎* | S3VM*min𝑎* | | | S3VM*com𝑎* | | S4VM*𝑎* | | 100 | | 78.795.470.368.376.394.9±±±±±±2.91.02.12.83.41.71.7  92.591.599.266.597.793.6±±±±±±2.10.52.61.01.7  85.4 | 78.695.870.066.376.092.4±±±±±±2.80.72.12.63.43.32.4  90.990.696.466.196.092.4±±±±±±2.83.42.32.12.6  84.3 | 2.8  2.0  3.0  2.5  85.1 | 78.4  69.8  76.2  66.4  84.8 | | 2.8  3.3  2.7  1.3 | 78.7  76.0  66.6  97.6  85.2 | 85.4 | | |  | | | 7/4/1 | 6/6/0 | 8/4/0 | | | 4/8/0 |  | | |

amples, respectively, S4VM*𝑠* never degenerates the performance significantly. Both S3VM*min𝑠* and S3VM*com𝑠* are capable to reduce the chance of degenerating performance compared with TSVM, however, they still degenerate performance significantly in many cases.

更重要的是，与分别针对10和100个标记示例的12和17个案例的性能退化的TSVM不同，S4VM𝑠从未显着降低性能。 与TSVM相比，S3VMmin𝑠和S3VMcom𝑠都能够降低性能退化的可能性，但在许多情况下，它们仍然会显着降低性能。

Wilcoxon sign tests at 95% significant level disclose that S4VM*𝑠* is significantly better than inductive SVM for both 10 and 100 labeled examples. The other three semisupervised methods, however, do not obtain such a significance. These results validate the effectiveness of S4VM*𝑠*.

在95％显着水平的Wilcoxon符号测试表明，对于10和100个标记的实例，S4VM𝑠明显优于诱导SVM。 然而，其他三种半监督方法并没有获得这样的意义。 这些结果证实了S4VM的有效性。

## Running Time

Figure 2 plots the running time on 12 UCI data sets with 10 labeled examples. As can be seen, S4VM*𝑎* has the highest time cost, and S4VM*𝑠* scales slightly worse than TSVM but much better than S4VM*𝑎*. Note that S4VM*𝑠* is inherently parallel due to the consideration of multiple separators, and it can be speedup by parallel implementation or using efficient S3VM solutions.

4.3。 运行时间

图2绘制了12个UCI数据集的运行时间，其中包含10个标记示例。 可以看出，S4VM𝑎的时间成本最高，S4VM𝑠比TSVM略差，但比S4VM好得多。 请注意，由于考虑了多个分隔符，S4VM𝑠本质上是并行的，并且可以通过并行实现或使用高效的S3VM解决方案来加速。

## Parameter Influence

S4VM*𝑠* has four parameters, i.e., sampling size *𝑁*, cluster number *𝑇*, risk parameter *𝜆* and the kernel type. Figure 3 studies the influence of *𝑁*, *𝑇* and *𝜆* with linear/RBF kernels

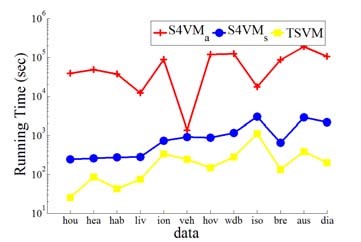


Figure2.Running time (in seconds) of S4VM*𝑎*, S4VM*𝑠* and TSVM with 10 labeled examples

图2.S4VM𝑎，S4VM𝑠和TSVM的运行时间（以秒为单位），带有10个标记示例

on five representative data sets with 10 labeled examples.

4.4。 参数影响

S4VM𝑠有四个参数，即采样大小𝑁，簇号𝑇，风险参数λ和内核类型。 图3研究了具有线性/ RBF核的𝑁，𝑇和λ对具有10个标记示例的五个代表性数据集的影响。

It can be seen that, though the number of labeled examples is small, the performance of S4VM*𝑠* is quite insensitive to the setting of the parameters. One possible reason is that, rather than simply picking one low-density separator, S4VMs optimize the label assignments in the worse case. This property makes S4VM*𝑠* even more attractive, since the performance of current S3VMs are usually sensitive to parameter settings, especially when the number of labeled examples is too few to afford a reliable model selection. Moreover, paired *𝑡*-tests at 95% significant level

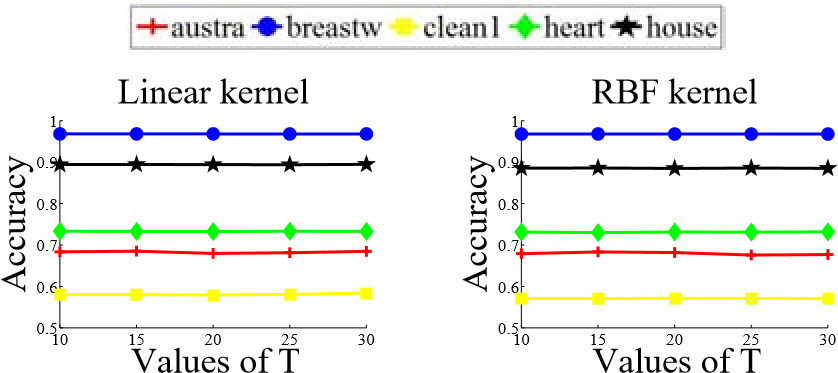
|  |
| --- |
| Table3. Accuracy (mean ± std.) of S4VM*𝑠* and compared methods (see title of Table 2 for more information). |

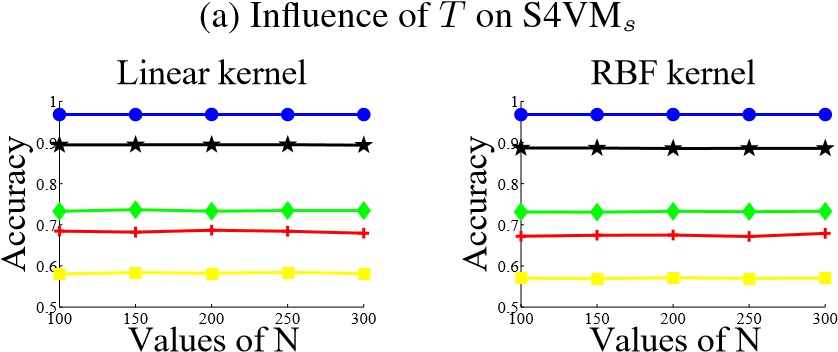
confirm that S4VM*𝑠* does not degenerate performance on all the cases in Figure 3.

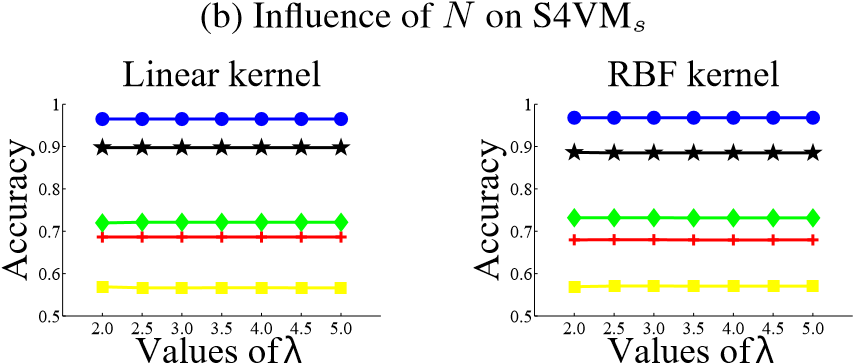
可以看出，尽管标记示例的数量很少，但S4VM𝑠的性能对参数的设置非常不敏感。 一个可能的原因是，S4VM不是简单地选择一个低密度分离器，而是在最坏的情况下优化标签分配。 此属性使S4VM𝑠更具吸引力，因为当前S3VM的性能通常对参数设置敏感，尤其是当标记示例的数量太少而无法提供可靠的模型选择时。 此外，95％显着水平的成对𝑡-测试证实S4VM𝑠不会降低图3中所有情况的性能。

# Conclusion

Semi-supervised learning tries to exploit unlabeled data to improve learning performance. Though semi-supervised learning approaches are promising, there are many cases where the performance of using unlabeled data is even worse than purely using the limited labeled data. In this paper, we focus on semi-supervised support vector machines (S3VMs) and propose the S4VMs. Unlike S3VMs which typically try to obtain one low-density separator, S4VMs attempt to exploit multiple candidate diverse largemargin low-density separators and optimize the label assignment for the worst case. We present two implementations, one uses a global simulated annealing search for the low-density separators, while the other uses a simpler and efficient sampling strategy. Comprehensive experiments validate the effectiveness of our S4VMs. It is particularly encouraging since the overall performance of S4VMs is highly competitive to TSVM, while contrasting to TSVM







(c) Influence of *𝜆* on S4VM*𝑠*

Figure3. Parameter Influence with 10 labeled examples

which often degenerates performance, S4VMs are never significantly worse than inductive SVMs.

五，结论

半监督学习试图利用未标记的数据来提高学习成绩。尽管半监督学习方法很有前途，但在许多情况下，使用未标记数据的性能甚至比纯粹使用有限标记数据更差。在本文中，我们关注半监督支持向量机（S3VMs）并提出S4VM。与通常尝试获得一个低密度分离器的S3VM不同，S4VM尝试利用多个候选多种大型低密度分离器，并针对最坏情况优化标签分配。我们提出了两个实现，一个使用全局模拟退火搜索低密度分离器，而另一个使用更简单有效的采样策略。综合实验验证了我们S4VM的有效性。这是特别令人鼓舞的，因为S4VM的整体性能与TSVM相比具有很强的竞争力，而与经常降低性能的TSVM相比，S4VM从未明显比感应式SVM差。

# Acknowledgments

We want to thank anonymous reviewers for helpful comments and Teng Zhang for help in experiments. This research was supported by the NSFC (61073097), JiangsuSF (BK2008018) and 973 Program (2010CB327903).

致谢

我们要感谢匿名评论者提供的有用评论和Teng Zhang的实验帮助。 该研究得到了NSFC（61073097），JiangsuSF（BK2008018）和973计划（2010CB327903）的支持。

# References

参考

Bennett, K. and Demiriz, A. Semi-supervised support vector machines. In *NIPS 11*, pp. 368–374. 1999.

Cernˇ y, V. Thermodynamical approach to the traveling salesman` problem: An efficient simulation algorithm. *J. Optimization Theory and Application*, 45(1):41–51, 1985.

Chapelle, O. and Zien, A. Semi-supervised learning by low density separation. In *AISTATS*, pp. 57–64, 2005.

Chapelle, O., Chi, M., and Zien, A. A continuation method for semi-supervised SVMs. In *ICML*, pp. 185–192, 2006a.

Chapelle, O., Scholkopf, B., and Zien, A. (eds.).¨ *Semi-Supervised Learning*. MIT Press, Cambridge, MA, 2006b.

Chapelle, O., Tubingen, G., Sindhwani, V., and Keerthi, S. S. Branch and bound for semi-supervised support vector machines. In *NIPS 20*, pp. 217–224. 2007.

Chapelle, O., Sindhwani, V., and Keerthi, S. S. Optimization techniques for semi-supervised support vector machines. *J. Mach. Learn. Res.*, 9:203–233, 2008.

Cozman, F., Cohen, I., and Cirelo, M. Semi-supervised learning of mixture models. In *ICML*, pp. 99–106, 2003.

De Bie, T. and Cristianini, N. Convex methods for transduction. In *NIPS 16*, pp. 73–80. 2004.

Goutte, C., Dejean, H., Gaussier, E., Cancedda, N., and Renders,´ J.M. Combining labelled and unlabelled data: A case study on fisher kernels and transductive inference for biological entity recognition. In *CoNLL*, pp. 1–7, 2002.

Grandvalet, Y. and Bengio, Y. Semi-supervised learning by entropy minimization. In *NIPS 17*, pp. 529–536. 2005.

Joachims, T. Transductive inference for text classification using support vector machines. In *ICML*, pp. 200–209, 1999.

Kasabov, N. and Pang, S. Transductive support vector machines and applications in bioinformatics for promoter recognition. In *ICNNSP*, pp. 1–6, 2004.

Kirkpatrick, S. Optimization by simulated annealing: Quantitative studies. *J. Stat. Phys.*, 34(5):975–986, 1984.

Laarhoven, P.J.M. and Aarts, E.H.L. *Simulated Annealing: Theory and Applications*. Springer, 1987.

Li, Y.-F., Kwok, J. T., and Zhou, Z.-H. Semi-supervised learning using label mean. In *ICML*, pp. 633–640, 2009a.

Li, Y.-F., Tsang, I. W., Kwok, J. T., and Zhou, Z.-H. Tighter and convex maximum margin clustering. In *AISTATS*, pp. 344–351, 2009b.

Nigam, K., McCallum, A. K., Thrun, S., and Mitchell, T. Text classification from labeled and unlabeled documents using EM. *Mach. Learn.*, 39(2-3):103–134, 2000.

Sindhwani, V., Keerthi, S.S., and Chapelle, O. Deterministic annealing for semi-supervised kernel machines. In *ICML*, pp. 841–848, 2006.

Vapnik, V. N. *Statistical Learning Theory*. Wiley, New York, 1998.

Wang, L., Chan, K., and Zhang, Z. Bootstrapping SVM active learning by incorporating unlabelled images for image retrieval. In *CVPR*, pp. 629–634, 2003.

Xu, L. and Schuurmans, D. Unsupervised and semi-supervised multi-class support vector machines. In *AAAI*, pp. 904–910, 2005.

Zhang, K., Tsang, I.W., and Kwok, J.T. Maximum margin clustering made practical. In *ICML*, pp. 1119–1126, 2007.

Zhang, T. and Oles, F. The value of unlabeled data for classification problems. In *ICML*, pp. 1191–1198, 2000.

Zhou, Z.-H. Ensemble learning. In Li, S. Z. (ed.), *Encyclopedia of Biometrics*, pp. 270–273. Springer, 2009.

Zhou, Z.-H. and Li, M. Semi-supervised learning by disagreement. *Knowl. Infor. Syst.*, 24(3):415–439, 2010.

Zhu, X. Semi-supervised learning literature survey. Technical Report 1530, Dept. Comp. Sci., Univ. Wisconsin-Madison, 2006.

1. http://www.kyb.tuebingen.mpg.de/ssl-book/ [↑](#footnote-ref-1)
2. http://archive.ics.uci.edu/ml/datasets.html [↑](#footnote-ref-2)
3. http://www.csie.ntu.edu.tw/ cjlin/libsvm/ [↑](#footnote-ref-3)
4. http://svmlight.joachims.org/ [↑](#footnote-ref-4)