

Semi-Supervised Learning Machine Based on Multi-View Twin Support Vector Machine

Rui Yao

College of Mathematics and Systems Science, Xinjiang University, Urumqi Xinjiang
Email: 1437912205@qq.com

Received: May 9th, 2019; accepted: May 23rd, 2019; published: May 30th, 2019

Abstract

In many practical problems of machine learning, the data has multi-views; multi-views complement each other; and the classification effect is better. This paper mainly studies the semi-supervised method of multi-view dual support vector machine, and divides the data according to different characteristics. Multi-views are used to find two non-parallel hyperplanes for each view, and the model is constructed to solve the dual problem. Experimental results show that the proposed algorithm can reduce the dimension of the data and has good classification accuracy. The support vector machine algorithm shortens the running time, reduces the computational complexity, and predicts better performance.

Keywords

Semi-Supervised Learning, Classification, Support Vector Machines, Multi-View

基于多视图双支持向量机半监督学习机

姚 瑞

新疆大学, 数学与系统科学学院, 新疆 乌鲁木齐
Email: 1437912205@qq.com

收稿日期: 2019年5月9日; 录用日期: 2019年5月23日; 发布日期: 2019年5月30日

摘要

在机器学习的诸多实际问题中, 数据有多个视图, 多个视图优势互补, 分类效果更好。本文主要针对多视图双支持向量机的半监督方法进行了研究, 根据不同的特征将数据分成多个视图, 分别给每个视图找两个非平行超平面, 构建模型, 求解对偶问题。实验结果表明, 本文的算法可以对数据进行降维, 有很

好的分类精确度，该算法较双支持向量机算法缩短了运行时间，减少了计算复杂度，预测性能较好。

关键词

半监督学习，分类问题，支持向量机，多视图

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

多视图学习(MVL)可用于表示多个不同特征的数据集[1]，不同的特征经常互相补充信息，与传统的单视图学习相反，MVL 分别为每个特征视图构造一个学习函数，然后通过利用相同输入数据的冗余视图共同优化学习函数。到目前为止，已经提出了许多 MVL 算法，可以分为两组：协同训练[2]和协同正则化算法[3]。协同训练算法通过迭代，最大化多个不同视图上的学习函数，以确保在相同数据的一致性。在协同正则化算法中，需要最小化不同视图学习目标函数中的正则项，典型的方法包括稀疏多视图 SVMs [4]，multiview Laplacian SVMs [5]，多视图矢量值多重正则化方法[6]。

半监督学习是一种综合学习方法，它使用少量标记数据来训练初始分类器，并使用大量未标记数据来进一步提高初始分类器的性能，最终实现准确学习。在学习过程中，学习机在训练过程中使用的样本是一个混合样本集，标签样本较少，未标记样本较多。它不仅利用标记样本的优点来精确描述单个样本，而且还可以发现未标记样本在整个样本集中的重要作用[7]。

支持向量机(SVM)是一种机器学习方法，主要针对小样本分类问题，遵守结构风险极小化原则，获取全局最优解。1982 年 Vapnik 首次提出支持向量机[8]，他将支持向量机归结为一个二次型方程求解问题。其学习的基本想法是求解能够正确划分训练数据集并且使得几何间隔最大的分离超平面。在近几年的研究中，支持向量机在人脸检测、语音识别、文字识别、手写识别、图像处理等研究方面取得了很多研究成果并受到了许多领域学者的关注。因此，研究也从最初的简单模式发展到各种互补应用。研究人员通过添加函数、变量或系数来对公式进行变形，以生成具有独特优势或一系列应用的新算法，例如：B. Schlkopf 等人提出的 γ -SVM 算法[9]，主要通过 γ 参数控制支持向量的个数和误差；Lee 等人通过引入光滑函数来近似正类样本标签从而提出一种光滑支持向量机(Smooth SVM, SSVM) [10]，Mangasarian 等人提出的一种拉格朗日支持向量机(Lagrangian SVM, LSVM) [11]，Fung 等人提出的近似支持向量机(Proximal SVM, PSVM) [12]，以及 Lee 等人提出的约简支持向量机(Reduced SVM, RSVM)等[13]。

半监督支持向量机[14]是一种能够同时兼顾标签与无标签样本的学习方法，因而半监督支持向量机的学习方法在处理大规模数据识别与分类的过程中处理良好。半监督支持向量机最初应用于文本分类[15]，研究人员之后提出了用于解决半监督支持向量机目标函数非凸问题的一系列方法，主要的研究方法有：梯度下降法(Gradient descent) [16]；凹凸法(Convex-Concave Procedure) [17]；确定性退火方法(Deterministic annealing) [18]等。

本文提出了一种基于多视图的双支持向量机的半监督学习方法，此方法将双支持向量机[19]与半监督多视图[20]学习相结合。给定训练集，将训练集分成两个视图，分别给视图一和视图二找两个非平行的超平面，构造两个二次规划模型，求解相对应对偶问题，每个视图分别得到一个超平面，并可以求出决策函

数。事实上，基于多视图双支持向量机半监督学习方法是对数据集进行降维。数据表明，与双支持向量机相比，多视图双支持向量机半监督学习方法训练的耗时更少，缩短了运行时间，减少了计算复杂度，有较高的准确率。

本文结构如下：第2节介绍了支持向量机及半监督多视图学习的基本思想；第3节介绍了基于多视图双支持向量机半监督学习的优化模型推导及算法；第4节对基于多视图双支持向量机半监督学习方法进行人工数据和UCI数据的数值实验，并与其它的分类算法进行比较；第5节给出一些结论及展望。

2. 相关工作

2.1. 支持向量机

对于训练集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ ，其中， $x_i \in R^n$ 是输入的样本点， $y_i \in Y = \{-1, 1\}$ 是输出的标签。线性的标准支持向量机要寻找一个超平面 $f(x)$ 将两类点分开。

$$f(x) = w^T x + b = 0, \quad (1)$$

其中， $w \in R^n$ ， $b \in R$ 分别是超平面的法向量和截距。引入松弛变量 ξ ，标准支持向量机的矩阵形式可以表示为：

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} (\|Aw + e_1 b\|^2 + ce_2^T \xi, \\ \text{s.t.} \quad & -(Bw + e_2 b) \geq e_2 - \xi, \\ & \xi \geq 0, \end{aligned} \quad (2)$$

其中， $A \in (m \times n)$ ， $c \geq 0$ 是惩罚参数， e_1 和 e_2 是相应的维数为 1 的向量。我们通过求解(2)式的对偶问题可以得到使得 w 和 b ，从而可以得到两个平行的支持超平面 $w^T x + b = 1$ 和 $w^T x + b = -1$ 。

2.2. 线性可分的双支持向量机

对于训练集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ ，其中 $x_i \in R^n$ 是输入的样本点， $y_i \in Y = \{-1, 1\}$ 是输出的标签。

双支持向量机[19]寻找两个非平行的超平面 $f_1(x)$ 和 $f_2(x)$ 使得其中的一个超平面离一类样本点尽可能近，而离另一类样本点尽可能远，反之亦然。

$$f_1(x) = w_1^T x + b_1 = 0, f_2(x) = w_2^T x + b_2 = 0, \quad (3)$$

其中， $w_1, w_2 \in R^n$ 和 $b_1, b_2 \in R$ 分别是两个非平行超平面的法向量和截距。引入松弛变量 ξ ，线性双支持向量机的模型为：

$$\begin{aligned} \text{TWSVMM1:} \min_{w_1, b_1, \xi} \quad & \frac{1}{2} (\|Aw_1 + e_1 b_1\|^2 + c_1 e_2^T \xi, \\ \text{s.t.} \quad & -(Bw_1 + e_2 b_1) \geq e_2 - \xi \\ & \xi \geq 0, \end{aligned} \quad (4)$$

和

$$\begin{aligned} \text{TWSVMM2:} \min_{w_2, b_2, \xi} \quad & \frac{1}{2} (\|Bw_2 + e_2 b_2\|^2 + c_2 e_1^T \xi, \\ \text{s.t.} \quad & Aw_2 + e_1 b_2 \geq e_1 - \xi, \\ & \xi \geq 0, \end{aligned} \quad (5)$$

其中， $A \in (m_1 \times n)$, $B \in (m_2 \times n)$, $c_1 > 0, c_2 > 0$ 是惩罚参数， e_1 和 $e_2 > 0$ 是相应维数的单位向量。

简单的说，双支持向量机是一对二次规划问题，对每个小的二次规划问题构造拉格朗日函数，求解对偶问题，可以得到 w_1, w_2, b_1, b_2 ，进而可得到双支持向量机的决策函数：

$$f(x) = \arg \min_{k=1,2} \frac{|w_k^T x + b_k|}{\|w_k\|}, \quad (6)$$

其中， $k = 1, 2$ 分别对应正负类标签 $\{+1, -1\}$ 。

2.3. 线性不可分的双支持向量机

显然，并不是所有问题是线性可分的，对于非线性分划问题，我们可以选择一个适当的内核 K ，用曲面 $g_1(x)$ 和 $g_2(x)$ 代替平面 $f_1(x)$ 和 $f_2(x)$ 。

$$g_1(x) = K(x^T, C^T)u_1 + b_1 = 0, g_2(x) = K(x^T, C^T)u_2 + b_2 = 0, \quad (7)$$

其中 $C^T = [A \ B]^T$ ，定义 $w_1 = C^T u_1$ 和 $w_2 = C^T u_2$ ，根据 2.2 节可构造优化问题如下：

$$\begin{aligned} \text{KTWSVM1: } & \min_{u_1, b_1, \xi} \frac{1}{2} \left(\|K(A, C^T)u_1 + e_1 b_1\|^2 \right)^2 + c_1 e_2^T \xi, \\ \text{s.t. } & -[K(B, C^T)u_1 + e_2 b_1] \geq e_2 - \xi, \\ & \xi \geq 0, \end{aligned} \quad (8)$$

和

$$\begin{aligned} \text{TWSVM2: } & \min_{u_2, b_2, \xi} \frac{1}{2} \left(\|K(B, C^T)u_2 + e_2 b_2\|^2 \right)^2 + c_2 e_1^T \xi, \\ \text{s.t. } & K(A, C^T)u_2 + e_1 b_2 \geq e_1 - \xi, \\ & \xi \geq 0, \end{aligned} \quad (9)$$

其中 $c_1 > 0, c_2 > 0$ 是惩罚参数， e_1 和 $e_2 > 0$ 是相应维数的单位向量。构造拉格朗日函数求解对偶问题，便可得到其决策函数。

2.4. 半监督多视图学习

多视图用于表示数据的不同特征集，本文研究的均为双视图。在多视图学习算法中，给定有标签数据集 $\{(x_i, y_i)\}_{i=1}^l$ 和无标签数据 $\{x_i\}_{i=l+1}^{l+u}$ ，任意一个数据点 x 我们都可以分成两个视图 $x^{(1)}$ 和 $x^{(2)}$ ，即 $x = (x^{(1)}, x^{(2)})$ ，其中 $x_i \in R^n$, $y_i \in Y = \{-1, 1\}$ ，则目标函数可以表示为 $f = (f^{(1)}, f^{(2)})$ ， $f^{(1)}$ 和 $f^{(2)}$ 分别为两个视图下的目标函数。半监督多视图学习方法中，在协同正则化最小二乘方法下的目标函数为：

$$\begin{aligned} f = \arg \min & \sum_{i=1}^l \left[y_i - f^{(1)}(x_i^{(1)}) \right]^2 + \mu \sum_{i=1}^l \left[y_i - f^{(2)}(x_i^{(2)}) \right]^2 + \gamma_1 \|f^{(1)}\|^2 \\ & + \gamma_2 \|f^{(2)}\|^2 + \frac{\gamma C}{l+u} \sum_{i=1}^{l+u} \left[f^{(1)}(x_i^{(1)}) - f^{(2)}(x_i^{(2)}) \right]^2, \end{aligned} \quad (10)$$

其中， μ 是平衡两个视图的参数， γ_1, γ_2 是正则化参数， γC 是耦合参数，用于调节无标签数据的兼容性。

根据表示定理(10)式可表示为：

$$f = \min \left[\sum_{i=1}^{l+u} \alpha_i K^{(1)}(x_i^{(1)}, x_i^{(1)}) + \sum_{i=1}^{l+u} \beta_i K^{(2)}(x_i^{(2)}, x_i^{(2)}) \right], i = 1, 2, \dots, l+u \quad (11)$$

其中, $K^{(1)}$ 和 $K^{(2)}$ 是核函数。 α 和 β 可以通过下列线性方程组求出:

$$\begin{aligned} \left[\frac{1}{l} JK_1 + \gamma_1 I + \frac{\gamma C}{l+u} K_1 \right] \alpha - \frac{\gamma C}{l+u} K_2 \beta &= \frac{1}{l} Y, \\ \left[\frac{u}{l} JK_2 + \gamma_2 I + \frac{\gamma C}{l+u} K_2 \right] \beta - \frac{\gamma C}{l+u} K_1 \alpha &= \frac{u}{l} Y, \end{aligned} \quad (12)$$

其中, J 和 Y 构成的对角矩阵, K_1 和 K_2 是核函数 $K^{(1)}$ 和 $K^{(2)}$ 的格拉姆矩阵。

3. 基于多视图双支持向量机的半监督学习方法

针对分类问题, 本文提出了基于多视图双支持向量机的半监督学习方法, 此方法将双支持向量机应用到多视图半监督学习中。给定训练集 $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \cup \{x_{l+1}, \dots, x_{l+u}\}$, 其中 $x_i \in R^n, i=1, \dots, l, l+1, l+u, y_i \in Y = \{-1, 1\}$, 需要找到与 $\{x_i\}_{i=l+1}^{l+u}$ 对应的输出值 $y_u = (y_{l+1}, \dots, y_{l+u})^T$ 和在多视图双支持向量机中的决策函数:

$$\begin{aligned} f(x) &= \min \{f_1(x), f_2(x)\} \\ &= \min \left\{ \min \left| w^{(1i)^T} x + b^{(1i)} \right|, \min \left| w^{(2i)^T} x + b^{(2i)} \right| \right\}, \\ i &= 1, 2, \end{aligned}$$

本文提出的多视图双支持向量机半监督学习方法是将(2.2)节中的矩阵 A, B 分别从两个视图进行描述, 得到线性可分的多视图双支持向量机的半监督学习(MV-TWSVM)的优化模型为:

$$\begin{aligned} \text{MV-TWSVM1: } \min \quad & \frac{1}{2} \left(\|A_1 w^{(11)} + e_1 b^{(11)}\|^2 + \|A_2 w^{(12)} + e_1 b^{(12)}\|^2 \right) + c_{11} e_2^T \xi^{(11)} + c_{12} e_2^T \xi^{(12)} \\ & + \frac{1}{2} \gamma_1 c_1 \left\| (U_1 w^{(11)} + e_3 b^{(11)}) - (U_2 w^{(12)} + e_3 b^{(12)}) \right\|^2, \\ \text{s.t. } \quad & -(B_1 w^{(11)} + e_2 b^{(11)}) \geq e_2 - \xi^{(11)}, \\ & -(B_2 w^{(12)} + e_2 b^{(12)}) \geq e_2 - \xi^{(12)}, \\ & \xi^{(11)} \geq 0, \xi^{(12)} \geq 0, \end{aligned} \quad (13)$$

和

$$\begin{aligned} \text{MV-TWSVM2: } \min \quad & \frac{1}{2} \left(\|B_1 w^{(21)} + e_2 b^{(21)}\|^2 + \|B_2 w^{(22)} + e_2 b^{(22)}\|^2 \right) + c_{21} e_1^T \xi^{(21)} + c_{22} e_1^T \xi^{(22)} \\ & + \frac{1}{2} \gamma_2 c_2 \left\| (U_1 w^{(21)} + e_3 b^{(21)}) - (U_2 w^{(22)} + e_3 b^{(22)}) \right\|^2, \\ \text{s.t. } \quad & (A_1 w^{(21)} + e_1 b^{(21)}) \geq e_1 - \xi^{(21)}, \\ & (A_2 w^{(22)} + e_1 b^{(22)}) \geq e_1 - \xi^{(22)}, \\ & \xi^{(21)} \geq 0, \xi^{(22)} \geq 0, \end{aligned} \quad (14)$$

其中, $A_v \in (m_1 \times n)$ 和 $B_v \in (m_2 \times n)$, m_1 和 m_2 表示数据点的数量, v 表示视图的个数, $\gamma_1, \gamma_2, c_1, c_2$ 是参数, e_1, e_2, e_3 表示相应维数的单位向量 $\xi^{(1v)}, \xi^{(2v)}$ 是损失项。

对(3.2)式引入拉格朗日乘子并构造拉格朗日函数为:

$$\begin{aligned}
& L(w^{(11)}, w^{(12)}, b^{(11)}, b^{(12)}, \alpha_{11}, \alpha_{12}, \beta_{11}, \beta_{12}) \\
&= \frac{1}{2} (A_1 w^{(11)} + e_1 b^{(11)})^T (A_1 w^{(11)} + e_1 b^{(11)}) + \frac{1}{2} (A_2 w^{(12)} + e_1 b^{(12)})^T (A_2 w^{(12)} + e_1 b^{(12)}) \\
&+ \frac{1}{2} \left[(U_1 w^{(11)} + e_3 b^{(11)}) - (U_2 w^{(12)} + e_3 b^{(12)}) \right]^T \left[(U_1 w^{(11)} + e_3 b^{(11)}) - (U_2 w^{(12)} + e_3 b^{(12)}) \right] \\
&+ \alpha_{11}^T [B_1 w^{(11)} + e_2 b^{(11)} + e_2 - \xi^{(11)}] + \alpha_{12}^T [B_2 w^{(12)} + e_2 b^{(12)} + e_2 - \xi^{(12)}] \\
&+ c_{11} e_2^T \xi^{(11)} + c_{12} e_2^T \xi^{(12)} - \beta_{11}^T \xi^{(11)} - \beta_{12}^T \xi^{(12)},
\end{aligned} \tag{15}$$

对 L 的分量求导并令其为 0 得：

$$\nabla_{w^{(11)}} L : A_1^T (A_1 w^{(11)} + e_1 b^{(11)}) + \gamma_1 c_1 U_1^T \left[(U_1 w^{(11)} + e_3 b^{(11)}) - (U_2 w^{(12)} + e_3 b^{(12)}) \right] + B_1^T \alpha_{11} = 0, \tag{16}$$

$$\nabla_{w^{(12)}} L : A_2^T (A_2 w^{(12)} + e_1 b^{(11)}) + \gamma_1 c_1 U_2^T \left[(U_1 w^{(11)} + e_3 b^{(11)}) - (U_2 w^{(12)} + e_3 b^{(12)}) \right] + B_2^T \alpha_{12} = 0, \tag{17}$$

$$\nabla_{b^{(11)}} L : e_1^T (A_1 w^{(11)} + e_1 b^{(11)}) + \gamma_1 c_1 e_3^T \left[(U_1 w^{(11)} + e_3 b^{(11)}) - (U_2 w^{(12)} + e_3 b^{(12)}) \right] + e_2^T \alpha_{11} = 0, \tag{18}$$

$$\nabla_{b^{(12)}} L : e_1^T (A_2 w^{(12)} + e_1 b^{(12)}) - \gamma_1 c_1 e_3^T \left[(U_1 w^{(11)} + e_3 b^{(11)}) - (U_2 w^{(12)} + e_3 b^{(12)}) \right] + e_2^T \alpha_{12} = 0, \tag{19}$$

$$\nabla_{\xi^{(11)}} L : c_{11} e_2^T - \alpha_{11}^T - \beta_{11} = 0, \tag{20}$$

$$\nabla_{\xi^{(12)}} L : c_{12} e_2^T - \alpha_{12}^T - \beta_{12} = 0, \tag{21}$$

$$\nabla_{\alpha_{11}} L : \alpha_{11}^T (B_1 w^{(11)} + e_2 b^{(11)} - \xi^{(11)}) = 0, \tag{22}$$

$$\nabla_{\alpha_{12}} L : \alpha_{12}^T (B_2 w^{(12)} + e_2 b^{(12)} - \xi^{(12)}) = 0, \tag{23}$$

$$\nabla_{\beta_{11}} L : \beta_{11}^T \xi^{(11)} = 0, \tag{24}$$

$$\nabla_{\beta_{12}} L : \beta_{12}^T \xi^{(12)} = 0, \tag{25}$$

$$-(B_1 w^{(11)} + e_2 b^{(11)}) + \xi^{(11)} \geq e_2, \tag{26}$$

$$-(B_2 w^{(12)} + e_2 b^{(12)}) + \xi^{(12)} \geq e_2, \tag{27}$$

由(16)和(18)得：

$$\left[\begin{bmatrix} A_1^T \\ e_1^T \end{bmatrix} [A_1 \quad e_1] + c_1 \gamma_1 \begin{bmatrix} u_1^T \\ e_3^T \end{bmatrix} [u_1 \quad e_3] \right] \begin{bmatrix} w^{(11)} \\ b^{(11)} \end{bmatrix} - c_1 \gamma_1 \begin{bmatrix} u_1^T \\ e_3^T \end{bmatrix} [u_2 \quad e_3] \begin{bmatrix} w^{(12)} \\ b^{(12)} \end{bmatrix} + \alpha_{11} \begin{bmatrix} B_1^T \\ e_2^T \end{bmatrix} \alpha_{12} = 0, \tag{28}$$

由(17)和(19)得：

$$-c_1 \gamma_1 \begin{bmatrix} U_2^T \\ e_3^T \end{bmatrix} [U_1 \quad e_3] \begin{bmatrix} w^{(11)} \\ b^{(11)} \end{bmatrix} + \left(\begin{bmatrix} A_2^T \\ e_1^T \end{bmatrix} [A_2 \quad e_1] + c_1 \gamma_1 \begin{bmatrix} U_2^T \\ e_3^T \end{bmatrix} [U_2 \quad e_3] \right) \begin{bmatrix} w^{(12)} \\ b^{(12)} \end{bmatrix} + \begin{bmatrix} B_2^T \\ e_2^T \end{bmatrix} \alpha_{12} = 0, \tag{29}$$

定义

$$H_1 = [A_1 \quad e_1], H_2 = [A_2 \quad e_1],$$

$$P_1 = [U_1 \quad e_3], P_2 = [U_2 \quad e_3],$$

$$G_1 = [B_1 \quad e_2]^T, G_2 = [B_2 \quad e_2]^T,$$

令

$$\begin{aligned} H &= H_1^T H_1 + c_1 \gamma_1 P_1^T P_1, F = -c_1 \gamma_1 P_1^T P_2, \\ P &= H_2^T H_2 + c_1 \gamma_1 P_2^T P_2, Q = F^{-1} H - P^{-1} F^T, \\ u_1 &= \begin{bmatrix} w^{(11)} & b^{(11)} \end{bmatrix}^T, u_2 = \begin{bmatrix} w^{(12)} & b^{(12)} \end{bmatrix}^T, \end{aligned} \quad (30)$$

由(28)和(29)可到:

$$H u_1 + F u_2 = -G_1 \alpha_{11}, \quad (31)$$

$$F^T u_1 + P u_2 = -G_2 \alpha_{12}, \quad (32)$$

由(31)和(32)可解得:

$$u_1 = -Q^{-1} F^{-1} G_1 \alpha_{11} + Q^{-1} P^{-1} G_2 \alpha_{12}, \quad (33)$$

$$u_2 = -Q H^{-1} G_1 \alpha_{11} + Q(F^T)^{-1} G_2 \alpha_{12}, \quad (34)$$

由(33)式(34)式及 KKT 条件可以得到 MV-TWSVM1 的对偶问题(DMV-TWSVM1)

$$\begin{aligned} \max \quad & \frac{1}{2} \alpha_1^T M \alpha_1 + e_2^T \alpha_1, \\ \text{s.t.} \quad & 0 \leq \alpha_1 \leq c_1, \end{aligned} \quad (35)$$

其中

$$M = \begin{pmatrix} M_1 & M_2 \\ M_3 & M_4 \end{pmatrix}, \alpha_1 = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \end{pmatrix}, c_1 = \begin{pmatrix} c_{11} \\ c_{12} \end{pmatrix},$$

$$M_1 = M_{11}^T H M_{11} + M_{13}^T P M_{13} - M_{11}^T F M_{13} - M_{13}^T F^T M_{11} - 2G_1^T M_{11},$$

$$M_2 = M_{12}^T H M_{12} + M_{14}^T P M_{14} - M_{12}^T F M_{14} - M_{14}^T F^T M_{12} - 2G_2^T M_{12},$$

$$M_3 = -M_{11}^T H M_{12} - M_{13}^T P M_{14} + M_{11}^T F M_{14} + M_{13}^T F^T M_{12} - 2G_1^T M_{12},$$

$$M_4 = -M_{12}^T H M_{11} - M_{14}^T P M_{13} + M_{12}^T F M_{13} + M_{14}^T F^T M_{11} - 2G_2^T M_{13},$$

$$M_{11} = Q^{-1} F^{-1} G_1, M_{12} = Q^{-1} F^{-1} G_2,$$

$$M_{13} = Q H^{-1} G_1, M_{14} = Q(F^T)^{-1} G_2,$$

同理, 对 MV-TWSVM2 引入拉格朗日乘子并构建拉格朗日函数:

$$\left[\begin{bmatrix} B_1^T \\ e_2^T \end{bmatrix} \begin{bmatrix} B_1 & e_2 \end{bmatrix} + c_2 \gamma_2 \begin{bmatrix} U_1^T \\ e_3^T \end{bmatrix} \begin{bmatrix} U_1 & e_3 \end{bmatrix} \right] \begin{bmatrix} w^{(21)} \\ b^{(21)} \end{bmatrix} - c_2 \gamma_2 \begin{bmatrix} U_1^T \\ e_3^T \end{bmatrix} \begin{bmatrix} U_2 & e_3 \end{bmatrix} \begin{bmatrix} w^{(22)} \\ b^{(22)} \end{bmatrix} + \alpha_{21} \begin{bmatrix} A_1^T \\ e_1^T \end{bmatrix} = 0, \quad (36)$$

$$-c_2 \gamma_2 \begin{bmatrix} U_2^T \\ e_3^T \end{bmatrix} \begin{bmatrix} U_1 & e_3 \end{bmatrix} \begin{bmatrix} w^{(21)} \\ b^{(21)} \end{bmatrix} + \left[\begin{bmatrix} B_2^T \\ e_2^T \end{bmatrix} \begin{bmatrix} B_2 & e_2 \end{bmatrix} + c_2 \gamma_2 \begin{bmatrix} U_2^T \\ e_3^T \end{bmatrix} \begin{bmatrix} U_2 & e_3 \end{bmatrix} \right] \begin{bmatrix} w^{(22)} \\ b^{(22)} \end{bmatrix} + \begin{bmatrix} A_2^T \\ e_1^T \end{bmatrix} \alpha_{22} = 0 \quad (37)$$

定义

$$\begin{aligned} X &= G_1^T G_1 + c_2 \gamma_2 P_1^T P_1, Y = -c_2 \gamma_2 P_1^T P_2, \\ R &= G_2^T G_2 + c_2 \gamma_2 P_2^T P_2, S = Y^{-1} X - R^{-1} Y^T, \\ v_1 &= [w^{(21)} \ b^{(21)}]^T, v_2 = [w^{(22)} \ b^{(22)}]^T, \end{aligned} \quad (38)$$

由(36) (37)和(38)可解得:

$$v_1 = S^{-1}Y^{-1}H_1\alpha_{21} + S^{-1}R^{-1}H_2\alpha_{22}, \quad (39)$$

$$v_2 = SX^{-1}H_1\alpha_{21} + S(Y^T)^{-1}H_2\alpha_{22}, \quad (40)$$

由(39)式(40)式及 KKT 条件可以得到 MV-TWSVM2 的对偶问题(DMV-TWSVM2):

$$\begin{aligned} \max \quad & \frac{1}{2}\alpha_2^T N\alpha_2 + e_1^T \alpha_2, \\ \text{s.t.} \quad & 0 \leq \alpha_2 \leq c_2, \end{aligned} \quad (41)$$

其中

$$N = \begin{pmatrix} N_1 & N_2 \\ N_3 & N_4 \end{pmatrix}, \alpha_2 = \begin{pmatrix} \alpha_{21} \\ \alpha_{22} \end{pmatrix}, c_2 = \begin{pmatrix} c_{21} \\ c_{22} \end{pmatrix},$$

$$N_1 = M_{21}^T X M_{21} + M_{23}^T R M_{23} - M_{21}^T Y M_{23} - M_{23}^T Y^T M_{21} - 2G_1^T M_{21},$$

$$N_2 = M_{22}^T X M_{22} + M_{24}^T R M_{24} - M_{22}^T Y M_{24} - M_{24}^T Y^T M_{22} - 2G_2^T M_{22},$$

$$N_3 = -M_{21}^T X M_{22} - M_{23}^T R M_{24} + M_{21}^T Y M_{24} + M_{23}^T Y^T M_{22} - 2G_1^T M_{22},$$

$$N_4 = -M_{22}^T X M_{21} - M_{24}^T R M_{23} + M_{22}^T Y M_{23} + M_{24}^T Y^T M_{21} - 2G_2^T M_{23},$$

$$M_{21} = S^{-1}Y^{-1}G_1, M_{22} = S^{-1}Y^{-1}G_2,$$

$$M_{23} = SX^{-1}G_1, M_{24} = S(Y^T)^{-1}G_2,$$

由上述过程可导出如下算法:

算法 3.1: 多视图双支持向量机半监督学习方法:

输入: 训练集 T ;

输出: 决策函数 f ;

- 1) 输入训练集 T , $\gamma_1, \gamma_2 > 0$, 选择参数 $c_{11}, c_{12}, c_{21}, c_{22} > 0$;
- 2) 计算(35), (41)得到 α_1 和 α_2 ;
- 3) 计算(33), (34), (39), (40)得到 u_1, u_2, v_1, v_2 ;
- 4) 由(30), (38)得到 $(w^{(11)}, b^{(11)})$, $(w^{(12)}, b^{(12)})$, $(w^{(21)}, b^{(21)})$, $(w^{(22)}, b^{(22)})$;
- 5) 得到决策函数。

$$f(x) = \min \left\{ \min \left| w^{(1i)^T} x + b^{(1i)} \right|, \min \left| w^{(2i)^T} x + b^{(2i)} \right| \right\}, i = 1, 2.$$

4. 数值实验

在本节中, 将本文的方法 MV-TWSVM 与 TWSVM 进行比较。本文的数值结果是在内存 4.00 GB, 64 位操作的 Windows7 系统中用软件 MATLAB, R2014a 运行得到的。

4.1. 人工数据

本节图 1 中, 月牙型数据为视图 1, 由直线 $y = 2x + 3$ 和 $y = 2x - 3$ 生成的样本点加入白噪声得到的数据集为视图 2。图中直线是用 MV-TWSVM 得到的两个视图下的分划线。可以看出分划线很好的表示出了数据的分布, 并且有较高的准确率, 视图 1 的准确率达到 83.6759%, 视图 2 的准确率达到 99.3781%。

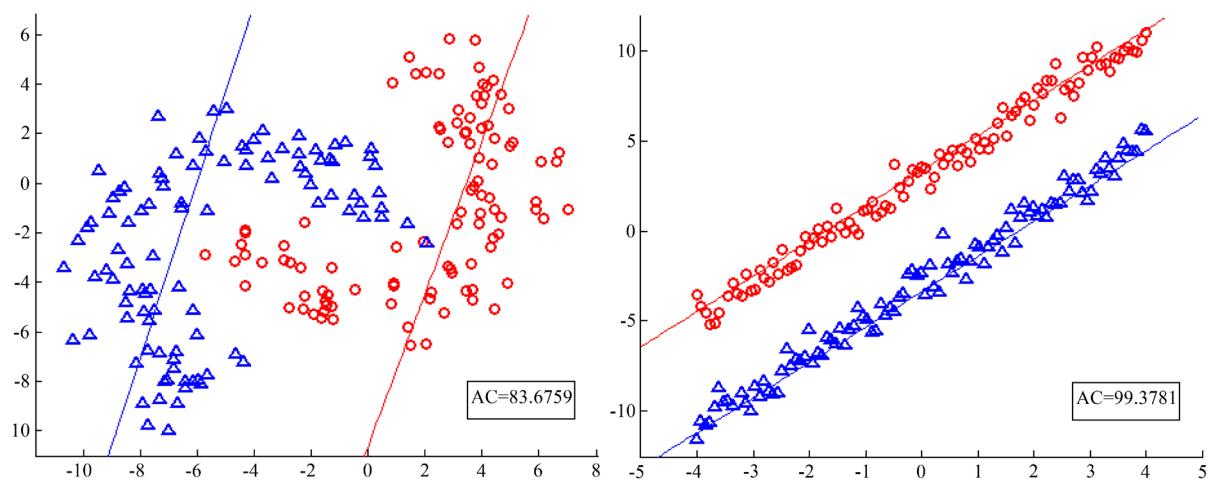


Figure 1. The left figure is view 1, and the right figure is view 2. The line in the figure is the partition hyperplane obtained by MV-TWSVM calculation of artificial data set

图 1. 左图是视图 1，右图是视图 2。图中直线是 MV-TWSVM 计算人工数据集得到的分划超平面

4.2. UCI 数据

在 UCI 数据上找了 Australian, pima, German, spectf, parkinsons, breast 六个数据集, 表 1 是数据集的详细信息, 其中数据 Australian, pima, German 选用了该数据的百分之三十进行数值实验。用本文提出的方法对六个数据集计算正确率, 并与 TWSVM 进行比较, 如表 2。表中加粗的部分正确率较高。

表 2 是线性的 TWSVM, MV-TWSVM 对所有数据集计算的正确率, 表格分为两部分, 第一部分是 TWSVM 方法, 第二部分是本文的方法, 其中第一表示 TWSVM 方法下的正确率, 第二列表示在视图 1 下的正确率, 第三列表示在视图 2 下的正确率, 第四列表示原始数据的正确率。奇数行表示将数据集前一半作为视图 1, 剩余部分作为视图二; 偶数行表示将数据集随机分成两部分, 一部分作为视图 1, 另一部分作为视图 2。表中可以看出数据 spectf, Australian, pima 在本文提出的方法下正确率较高。

当我们选择不同比例无标签的数据点时, 正确率也会不同, 如表 3。表 3 分别取了 50%, 80%, 90%, 无标签的数据点, 计算正确率和耗时。从表中可以看出, 数据集 spectf, parkinsons, breast, Pima, German, 的无标签点增多时, 在本文的方法下正确率增大。

此外, 本节还对 MV-TWSVM 与 TWSVM 进行了时间耗时比较, 如图 2, 数据集仍然是 Australian, pima, German, spectf, parkinsons, breast 6 个数据集, 从图中可以看出, MV-TWSVM 比 TWSVM 耗时少很多, 从而说明 MV-TWSVM 对数据降维事实相符。

Table 1. The details of the data set

表 1. 数据集的详细信息

Dataset	类别个数	样本个数	特征个数	类别比
Australian	2	690	14	1:1.25
pima	2	768	8	1:1.87
German	2	1000	24	1:2.33
spectf	2	267	44	1:3.85
parkinsons	2	195	22	1:3.06
breast	2	569	30	1:1.68

Table 2. The accuracy of linear TWSVM, MV-TWSVM for all data sets (mean (std)%)
表 2. 线性的 TWSVM, MV-TWSVM 对所有数据集计算的正确率

Dataset	TWSVM mean (std)%	MV-TWSVM		
		$f_1(x)(ac1)$ mean (std)%	$f_2(x)(ac2)$ mean (std)%	$f(x)$ mean (std)%
spectf	75.26(5.75)	89.31 (3.68)	48.28 (12.87)	80.73 (3.68)
		86.81 (6.15)	86.29 (7.34)	75.22 (3.68)
parkinsons	95.80(0.82)	94.89 (7.17)	71.54 (8.83)	92.39 (9.37)
		90.62 (9.27)	80.72 (10.54)	89.44 (8.08)
breast	99.60(0.82)	92.47 (1.97)	62.71 (4.96)	94.57 (4.43)
		99.26 (1.29)	68.52 (7.46)	93.59 (4.69)
Australian*	88.21(6.52)	75.48 (4.24)	68.27 (4.64)	71.34 (7.12)
		88.41 (4.24)	85.06 (3.90)	85.76 (6.49)
pima*	74.32(8.61)	74.55 (8.21)	65.47 (5.18)	72.47 (6.00)
		74.40 (8.23)	70.99 (6.54)	71.44 (7.44)
German*	75.14(6.51)	71.66 (5.10)	32.68 (4.52)	71.44 (5.69)
		71.33 (4.91)	70.51 (5.68)	66.06 (5.68)

*是指用该数据集的 33% 进行数值实验。黑体表示几个方法中最高的正确率。

Table 3. The accuracy of MV-TWSVM in computing unlabeled data sets with different ratios (mean (std)%)
表 3. MV-TWSVM 对不同比率的无标签的数据集计算的正确率

Dataset	比率	50%	80%	90%
		决策函数 mean (std)%	mean (std)%	mean (std)%
spectf	$f_1(x)$	93.35 (8.67)	93.44 (6.75)	94.17 (6.93)
	$f_2(x)$	88.08 (8.54)	83.46 (5.84)	85.42 (9.43)
	$f(x)$	71.54 (12.94)	74.76 (8.18)	68.75 (9.46)
	time	0.68	0.38	0.31
parkinsons	$f_1(x)$	97.89 (4.46)	89.88 (8.76)	88.56 (5.33)
	$f_2(x)$	72.44 (14.07)	72.54 (12.96)	76.70 (6.84)
	$f(x)$	94.78 (8.43)	87.33 (8.67)	87.48 (7.38)
	time	0.39	0.20	0.20
breast	$f_1(x)$	98.94 (1.71)	98.46 (1.79)	99.03 (1.65)
	$f_2(x)$	78.28 (7.05)	72.08 (7.65)	72.47 (6.71)
	$f(x)$	95.75 (3.20)	96.25 (7.65)	94.53 (3.03)
	time	2.19	0.45	0.34
Australian*	$f_1(x)$	86.06 (8.83)	87.32 (7.27)	85.67 (5.51)
	$f_2(x)$	80.00 (6.87)	64.85 (6.09)	69.88 (6.77)
	$f(x)$	81.47 (7.87)	79.72 (7.59)	83.89 (5.55)
	time	3.36	0.61	0.30

Continued

	$f_1(x)$	78.66 (6.58)	74.27 (3.39)	76.27 (7.03)
pima*	$f_1(x)$	68.01 (7.43)	69.54 (4.49)	65.41 (5.26)
	$f(x)$	63.81 (7.16)	70.69 (7.17)	70.76 (6.74)
	time	5.86	0.62	0.32
German*	$f_1(x)$	72.80 (5.59)	73.13 (6.16)	69.22 (6.63)
	$f_2(x)$	66.00 (7.13)	66.75 (4.80)	68.67 (5.23)
	$f(x)$	68.60 (4.66)	65.63 (5.20)	66.22 (2.98)
	time	8.44	1.09	0.56

*是指用该数据集的 33% 进行数值实验。黑体表示几个方法中最高的正确率。

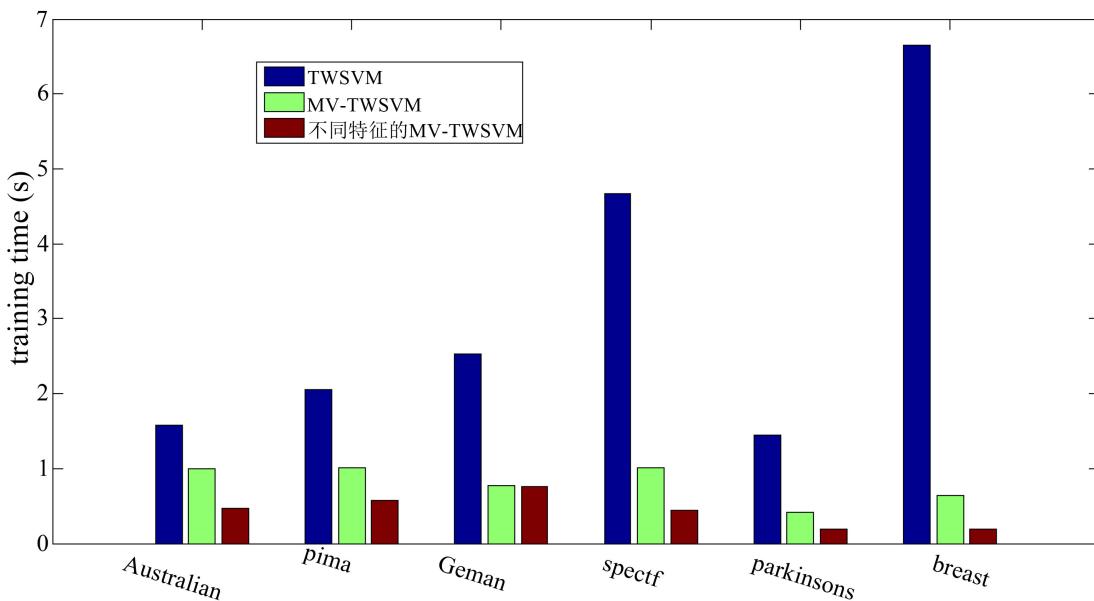


Figure 2. Comparison of training time between TWSVM and MV-TWSVM under different characteristics
图 2. 不同的特征下, TWSVM 与 MV-TWSVM 的训练时间之间的比较

5. 结论与展望

本文将双支持向量机应用到多视图半监督学习中, 将数据分成两个视图进行描述, 降低了数据的维数, 减少了计算的复杂度, 有效地缩短了运行时间。对于不同的数据, 多视图双支持向量机半监督学习方法取得了较好的正确率。

进一步, 我们将研究如何选择不同的核函数将线性划分推广到非线性划分。另外, 现有的多视图学习方法实际上是双视图, 如何推广到更多视图上的研究也是一个需要考虑的问题。

参考文献

- [1] Xu, J., Han, J., Nie, F. and Li, X. (2017) Re-Weighted Discriminatively Embedded K-Means for Multi-View Clustering. *IEEE Transactions on Image Processing*, **26**, 3016-3027. <https://doi.org/10.1109/TIP.2017.2665976>
- [2] Balcan, M.F., Blum, A. and Yang, K. (2004) Co-Training and Expansion: Towards Bridging Theory and Practice. *Proceedings of the Neural Information Processing Systems Conferences*, Vancouver, December 2004, 89-96.
- [3] Farquhar, J.D.R., Hardoon, D.R., Meng, H., Shawe-Taylor, J. and Szedmak, S. (2005) Two View Learning: SVM-2k,

- Theory and Practice. *Advances in Neural Information Processing Systems*, **18**, 355-362.
- [4] Sun, S. and Shawe-Taylor, J. (2010) Sparse Semi-Supervised Learning Using Conjugate Functions. *Journal of Machine Learning Research*, **11**, 2423-2455.
- [5] Sun, S. (2011) Multi-View Laplacian Support Vector Machines. *7th International Conference on Advanced Data Mining and Applications*, Beijing, 17-19 December 2011, 209-222. https://doi.org/10.1007/978-3-642-25856-5_16
- [6] Luo, Y., Tao, D., Xu, C., Liu, H. and Wen, Y. (2013) Multiview Vector Valued Manifold Regularization for Multilabel Image Classification. *IEEE Transactions on Neural Networks and Learning Systems*, **24**, 709-722. <https://doi.org/10.1109/TNNLS.2013.2238682>
- [7] 赵莹. 半监督支持向量机学习算法研究[D]: [博士学位论文]. 哈尔滨: 哈尔滨工程大学, 2010.
- [8] Vapnik, V.N. (1995) The Nature of Statistical Learning Theory. Spring, New York. <https://doi.org/10.1007/978-1-4757-2440-0>
- [9] Schlkopf, B., Smola, A., Williamson, R. and Bartlett, P.L. (2000) New Support Vector Algorithms. *Neural Computation*, **12**, 1207-1245. <https://doi.org/10.1162/089976600300015565>
- [10] Lee, Y.J. and Mangasarian, O.L. (2001) SSVM: A Smooth Support Vector Machine for Classification. *Computational Optimization and Applications*, **20**, 5-22. <https://doi.org/10.1023/A:1011215321374>
- [11] Mangasarian, O.L. and Musicant, D.R. (2001) Lagrangian Support Vector Machines. *Journal of Machine Learning Research*, **1**, 161-177.
- [12] Fung, G.M. and Mangasarian, O.L. (2005) Multicategory Proximal Support Vector Machine Classifiers. *Machine Learning*, **59**, 77-97. <https://doi.org/10.1007/s10994-005-0463-6>
- [13] Lee, Y.J. and Huang, S.Y. (2007) Reduced Support Vector Machines: A Statistical Theory. *IEEE Transactions on Neural Networks*, **18**, 1-13. <https://doi.org/10.1109/TNN.2006.883722>
- [14] Bennett, K. and Demiriz, A. (1998) Semi-Supervised Support Vector Machines. In: *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II*, MIT Press, Cambridge, 368-374.
- [15] Joachims, T. (1999) Transductive Inference for Text Classification Using Support Vector Machines. In: *Proceedings of the Sixteenth International Conference on Machine Learning*, Morgan Kaufmann Publishers Inc., San Francisco, 200-209.
- [16] Chapelle, O., Zien, A., Cowell, R., et al. (2005) Semi-Supervised Classification by Low Density Separation. *Encyclopedia of Biostatistics*, **34**, 57-64.
- [17] Collobert, R., Sinz, F., Weston, J. and Bottou, L. (2006) Large Scale Transductive SVMs. *Journal of Machine Learning Research*, **7**, 1687-1712.
- [18] Sindhwani, V., Keerthi, S. and Chapelle, O. (2006) Deterministic Annealing for Semi-Supervised Kernel Machines. *Proceedings of International Conference on Machine Learning*, Pittsburgh, June 2006, 108-116. <https://doi.org/10.1145/1143844.1143950>
- [19] Jayadeva, Khemchandani, R. and Chandra, S. (2007) Twin Support Vector Machines for Pattern Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29**, 905-910. <https://doi.org/10.1109/TPAMI.2007.1068>
- [20] Sindhwani, V. (2007) On Semi-Supervised Kernel Methods. The University of Chicago, Chicago.



知网检索的两种方式：

1. 打开知网首页 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN: 2163-1476，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>
期刊邮箱：orff@hanspub.org