一种新的面向迁移学习的 L_2 核分类器

许 敏*①② 王士同① 史荧中①②
①(江南大学数字媒体学院 无锡 214122)
②(无锡职业技术学院物联网技术学院 无锡 214121)

摘 要：基于密度差(Difference Of Density, DOD)思想，L_2 核分类器算法具有良好的分类性能及稀疏性，然而其训练域与测试域独立同分布的假设限制了其应用范围。针对此不足，该文提出一种新的面向迁移学习的 L_2 核分类器(Transfer Learning-L_2 Kernel Classification, TL-L_2KC)，该方法既保持了 L_2 核分类器算法良好的分类性能，又能处理数据集缓慢变化及训练集在特定约束条件下获得导致训练集和未来测试集分布不一致的问题。基于人造数据集和 UCI 真实数据集的实验表明，该文提出的 TL-L_2KC 算法较之于经典的迁移学习分类方法，具有相当的、甚至更好的性能。

关键词：支持向量机；迁移学习；密度差；L_2 核分类器
中图分类号：TP181 **文献标识码**：A **文章编号**：1009-5896(2013)09-2059-07
DOI: 10.3724/SP.J.1146.2012.01647

A Novel Transfer-learning-oriented L_2 Kernel Classifier

Xu Min①② Wang Shi-tong① Shi Ying-zhong①②
①(*School of Digital Media*, *Jiangnan University*, *Wuxi* 214122, *China*)
②(*School of Internet of Things Engineering*, *Wuxi Institute of Technology*, *Wuxi* 214121, *China*)

Abstract: Based on the concept of Difference Of Density (DOD), L_2 Kernel Classifier(L_2KC) exhibits its good performance. However, the assumption that the training domain and testing domain are independent and identically distributed severely constrains its usefulness. In order to overcome this shortcoming, a novel classifier named Transfer Learnging-L_2 Kernel Classification (TL-L_2KC) is proposed for the changing environment. The proposed classifier can not only inherit the advantage of L_2KC, but also deal with the problem that the distribution inconsistency between the training and testing sets which is caused by the slow change of the datasets or the training set obtained with specific constraints. As demonstrated by extensive experiments in simulation datasets and UCI benchmark datasets, the proposed classifier TL-L_2KC shows the performance which is comparable to or better than that of the classical algorithms on the transfer learning classification problems.

Key words: Support Vector Machine (SVM); Transfer learning; Difference Of Density (DOD); L_2 Kernel Classification (L_2KC)

1 引言

分类是机器学习和模式识别领域非常重要的内容之一。著名的有支持向量机(Support Vector Machine, SVM)[1]，该算法的最终决策函数取决于支持向量数目，而不是样本维数，避免了“维数灾难”。Kim 等人[2−4]提出 L_2 核分类器(L_2 Kernel Classifier, L_2KC)，该算法与 SVM 一样是二次规划问题，其解具有稀疏性，便于快速决策，其性能也与 SVM 相当。若退化为单类数据集，可应用于概率密度估计。

上述分类方法的前提是训练数据(源域)与测试数据(目标域)独立且同分布，但在实际应用场景中，测试数据域和训练数据域的分布可能会不同。例如，医院临床研究中，常见做法是挑选符合条件的志愿者作为训练数据建立模型，然后将模型应用到整个群体。因为这些志愿者通常不是随机选择，故训练集和测试集可能会有不同的分布。如果直接使用训练集建立模型对测试集进行分类，因测试集可能包含新信息，导致与训练集分布不同，分类精度会降低；若完全丢弃训练数据，重新构建新的训练集代价昂贵，且源域训练集对测试集的分类具有一定的指导作用。

针对上述问题，迁移学习(Transfer Learning, TL)[5−9]得以提出。迁移学习是为了解决训练数据与测试数据分布不一致，如何更好利用相似领域已知标签数据而提出的研究新方向，主要有归纳式、直

2012-12-18 收到，2013-03-15 改回
国家自然科学基金(61272210, 61170122)和江苏省研究生创新工程项目(CXZZ12-0759)资助课题
*通信作者：许敏 xum@wxit.edu.cn

推式和无监督式 3 种形式[10,11]。目前，已有学者对基于 SVM 的迁移学习进行研究，提出了 TrSVM 算法[12]，本文主要研究面向迁移学习的 L_2 核分类器。

L_2 核分类器本身不适合动态数据的分类，不能利用已有知识进行迁移学习。Kim 等人[2−4]证明了 L_2 核分类器二次规划形式对应的优化问题是一特殊 SVM 问题，在数学模型上与 SVM 分类器等价。故本文在 L_2 核分类器基础上，利用与其对偶形式等价的 SVM，提出具有迁移学习能力的 L_2 核分类器(Transfer Learning-L_2 Kernel Classfier,TL-L_2KC)，其特点是已知少量目标域(target domain)测试样本类标签，需要结合其它领域(source domain)训练集知识，找到适合目标域的最佳分类器。下面首先介绍 L_2 核分类器。

2 L_2 核分类器

2.1 L_2 核分类器的基本思想

设两分类问题样本及所属类别为 $(\boldsymbol{x}_1,y_1),\cdots,(\boldsymbol{x}_n,y_n)$，其中 $\boldsymbol{x}_i\in R^d$ 表示一个 d 维样本，$y_i\in\{1,-1\}$ 是该样本的类标签。设 $f_+(\boldsymbol{x})$ 和 $f_-(\boldsymbol{x})$ 分别表示两类已知标签样本集的条件概率密度。根据决策理论，最优分类器形式如式(1)：

$$g^*(\boldsymbol{x})=\mathrm{sign}\{f_+(\boldsymbol{x})-\gamma f_-(\boldsymbol{x})\} \tag{1}$$

其中 γ 是人为设置用来反映先验概率的固定参数。

重新定义类标签 $y_i\in\{-\gamma,1\}$，并定义 $I_+=\{i\,|\,y_i=+1\}$，$I_-=\{i\,|\,y_i=-\gamma\}$，根据核密度估计模型，正负两类样本的密度函数可用权重因子 $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_n)$ 分别表示：

$$\left.\begin{aligned}\hat{f}_+(\boldsymbol{x};\boldsymbol{\alpha})&=\sum_{i\in I_+}\alpha_i k_\sigma(\boldsymbol{x},\boldsymbol{x}_i)\\ \hat{f}_-(\boldsymbol{x};\boldsymbol{\alpha})&=\sum_{i\in I_-}\alpha_i k_\sigma(\boldsymbol{x},\boldsymbol{x}_i)\end{aligned}\right\} \tag{2}$$

式 (2) 满足约束 $A=\left\{\alpha\left|\sum_{i\in I_+}\alpha_i=\sum_{i\in I_-}\alpha_i=1,\alpha_i\ge 0,\forall i\right.\right\}$ 及高斯核 $k_\sigma(\boldsymbol{x},\boldsymbol{x}_i)=(2\pi\sigma^2)^{-d/2}\cdot\exp\left\{-\dfrac{\|\boldsymbol{x}-\boldsymbol{x}_i\|^2}{2\sigma^2}\right\}$，窗宽 $\sigma>0$。

L_2 核分类器的核心思想为：使用 $\hat{d}_\gamma(\boldsymbol{x};\boldsymbol{\alpha})=\hat{f}_+(\boldsymbol{x})-\gamma\hat{f}_-(\boldsymbol{x})$ 估计真实的 $d_\gamma(\boldsymbol{x})$，并提出用最小化 $\hat{d}_\gamma(\boldsymbol{x})$ 和 $d_\gamma(\boldsymbol{x})$ 的 L_2 距离(Integrated Squared Error, ISE)来估计参数 $\boldsymbol{\alpha}$：

$$\begin{aligned}\mathrm{ISE}(\boldsymbol{\alpha})&=\left\|\hat{d}_\gamma(\boldsymbol{x};\boldsymbol{\alpha})-d_\gamma(\boldsymbol{x})\right\|_{L_2}^2\\&=\int(\hat{d}_\gamma(\boldsymbol{x};\boldsymbol{\alpha})-d_\gamma(\boldsymbol{x}))^2\mathrm{d}\boldsymbol{x}\\&=\int\hat{d}_\gamma^2(\boldsymbol{x};\boldsymbol{\alpha})\mathrm{d}\boldsymbol{x}-2\int\hat{d}_\gamma(\boldsymbol{x};\boldsymbol{\alpha})d_\gamma(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\\&\quad+\int d_\gamma^2(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\end{aligned}$$

因 $d_\gamma(\boldsymbol{x})$ 中不包含参数 $\boldsymbol{\alpha}$，故第 3 项可省略，又因 $d_\gamma(\boldsymbol{x})$ 值不可知，故需要估计第 2 项的值。设

$$H(\boldsymbol{\alpha})=\int\hat{d}_\gamma(\boldsymbol{x};\boldsymbol{\alpha})d_\gamma(\boldsymbol{x})\mathrm{d}\boldsymbol{x}=\sum_{i=1}^n\alpha_i y_i h_i$$

运用留一法给出 $H(\boldsymbol{\alpha})$ 的一个无偏估计 $H_n(\boldsymbol{\alpha})=\sum_{i=1}^n\alpha_i y_i\hat{h}_i$，其中

$$\hat{h}_i=\begin{cases}\dfrac{1}{N_+-1}\displaystyle\sum_{j\in I_+,j\ne i}k_\sigma(\boldsymbol{x}_j,\boldsymbol{x}_i)-\dfrac{1}{N_-}\sum_{j\in I_-}k_\sigma(\boldsymbol{x}_j,\boldsymbol{x}_i),\\ \qquad i\in I_+\\ \dfrac{1}{N_+}\displaystyle\sum_{j\in I_+}k_\sigma(\boldsymbol{x}_j,\boldsymbol{x}_i)-\dfrac{1}{N_--1}\sum_{j\in I_-,j\ne i}k_\sigma(\boldsymbol{x}_j,\boldsymbol{x}_i),\\ \qquad i\in I_-\end{cases} \tag{3}$$

其中 $N_+=|I_+|$，$N_-=|I_-|$。

令 $c_i=y_i\hat{h}_i$，最终求解 $\boldsymbol{\alpha}$ 的二次规划形式为

$$\boldsymbol{\alpha}=\min_{\boldsymbol{\alpha}\in A}\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n\alpha_i\alpha_j y_i y_j k_{\sqrt{2}\sigma}(\boldsymbol{x}_i,\boldsymbol{x}_j)-\sum_{i=1}^n c_i\alpha_i \tag{4}$$

2.2 L_2 核分类器的几何解释

式(4)在核空间中与式(5)相应优化问题的对偶问题一致：

$$\left.\begin{aligned}&\min_{\boldsymbol{w},\xi_+,\xi_-}\frac{1}{2}\boldsymbol{w}^2+\xi_++\xi_-\\&\text{s.t.}\quad y_i\cdot\left\langle\boldsymbol{w},\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_i)\right\rangle\ge c_i-\xi_+,\ i\in I_+\\&\text{s.t.}\quad y_i\cdot\left\langle\boldsymbol{w},\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_i)\right\rangle\ge c_i-\xi_-,\ i\in I_-\end{aligned}\right\} \tag{5}$$

其中 $\phi_\sigma(\boldsymbol{x})$ 是高斯核希尔伯特空间的核映射。该优化形式等价于变形的 SVM。与标准 SVM 有如下不同：首先，右边的约束项用 c_i 代替 1，意味着若 c_i 足够大，变形的 SVM 更强调 $\boldsymbol{x}_i$ 分类的正确性；其次，每一类只有一个平滑因子 ξ_+，ξ_-，且不需要非负。

L_2 核分类器提出 ISE 准则下最小化 $\hat{d}_\gamma(\boldsymbol{x})$ 和 $d_\gamma(\boldsymbol{x})$ 间的差异，用于分类，其分类问题与 SVM 一样是二次规划问题，且具有稀疏性，加快分类决策效率。但 L_2 核分类器只能解决训练域与测试域同分布的场合，若数据域存在缓慢变化或训练域和测试域分布近似但不同，则分类效率会降低。基于此问题，本文提出了具有迁移学习能力的 L_2 核分类器。

3 面向迁移学习的 L_2 核分类器

3.1 问题陈述

源域和目标域有相同的属性空间和标签集合 Y ($Y=\{1,-\gamma\}$)，γ 定义见 2.1 节，两域分布近似但不同。训练集 T 包含两个样本集：目标域训练集 T_t、源域训练集 T_s，两训练集中所有样本均有类标签。此外，还有来自目标域的无标签测试集 E。

3 个样本集满足如下关系：T_t 和 E 来自目标域，

T_s 和 E 来自不同域，即它们分布不同。T_s 样本数相对较大，但因与测试集 E 分布不同，影响分类精度；T_t 样本数较少，不能体现目标域分布特点，故虽与测试集 E 分布相同，但很难获得较高的分类性能。对 T_s, T_t 和 E 定义如下：

定义 1　目标域中的训练集：$T_t=\{(\boldsymbol{x}_1^t,y_1^t),(\boldsymbol{x}_2^t,y_2^t),\cdots,(\boldsymbol{x}_n^t,y_n^t)\}$，其中，$\boldsymbol{x}_i^t\in X(i=1,2,\cdots,n)$ 是第 i 个样本，$y_i^t\in Y$ 是相应的类标签，n 为训练集规模。定义 $I_+^t=\{i\mid y_i^t=+1\}$, $I_-^t=\{i\mid y_i^t=-\gamma\}$。

定义 2　源域中的训练集：$T_s=\{(\boldsymbol{x}_1^s,y_1^s),(\boldsymbol{x}_2^s,y_2^s),\cdots,(\boldsymbol{x}_m^s,y_m^s)\}$，其中，$\boldsymbol{x}_i^s\in X(i=1,2,\cdots,m)$ 是第 i 个样本，$y_i^s\in Y$ 是相应的类标签，m 为训练集规模。定义 $I_+^s=\{i\mid y_i^s=+1\}$, $I_-^s=\{i\mid y_i^s=-\gamma\}$。

定义 3　目标域中的测试集：$E=\{\boldsymbol{x}_1^e,\boldsymbol{x}_2^e,\cdots,\boldsymbol{x}_r^e\}$，其中，$\boldsymbol{x}_i^e\in X(i=1,2,\cdots,r)$ 是第 i 个样本，$y_i^s\in Y$ 是相应的类标签，r 为测试集规模。

本文所提出的 TL-L_2KC 算法的目标是通过学习 T_s 和 T_t，找到一个分类器 L，使得 L 能对测试集 E 进行精确分类。其中，E 集合在训练阶段未知，不能在训练过程中使用。

3.2 TL-L_2KC 理论依据

若目标域与源域同分布，则可直接利用源域 L_2 核分类器求解；若目标域与源域分布近似但不同，则直接利用源域 L_2 核分类器求解会导致分类精度下降。若目标域已知标签样本足够多，则可直接利用目标域 L_2 核分类器求解；若目标域已知标签样本过少，则所得 L_2 核分类器可能会导致泛化误差过大。为避免上述源域与目标域分布不一致且目标域样本不足以建立分类器的问题，需进行源域训练集与目标域训练集间的迁移学习。L_2 核分类器本身不具备迁移学习条件，我们利用 L_2 核分类器等价的变形 SVM 形式进行迁移学习。优化问题如式(6)：

$$\left.\begin{aligned}&\min_{\boldsymbol{w}_t,\boldsymbol{w}_s,\xi_+^t,\xi_-^t,\xi_+^s,\xi_-^s}\frac{1}{2}\|\boldsymbol{w}_t\|^2+\xi_+^t+\xi_-^t+\frac{1}{2}\|\boldsymbol{w}_s\|^2\\&\qquad+\xi_+^s+\xi_-^s+\mu\|\boldsymbol{w}_t-\boldsymbol{w}_s\|^2\\&\text{s.t.}\quad y_i^t\cdot\left\langle\boldsymbol{w}_t,\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_i^t)\right\rangle\geq c_i^t-\xi_+^t,\quad i\in I_+^t\\&\text{s.t.}\quad y_i^t\cdot\left\langle\boldsymbol{w}_t,\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_i^t)\right\rangle\geq c_i^t-\xi_-^t,\quad i\in I_-^t\\&\text{s.t.}\quad y_i^s\cdot\left\langle\boldsymbol{w}_s,\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_i^s)\right\rangle\geq c_i^s-\xi_+^s,\quad i\in I_+^s\\&\text{s.t.}\quad y_i^s\cdot\left\langle\boldsymbol{w}_s,\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_i^s)\right\rangle\geq c_i^s-\xi_-^s,\quad i\in I_-^s\end{aligned}\right\}\tag{6}$$

比较式(6)与式(5), TL-L_2KC 算法与 L_2KC 算法的不同之处在于训练目标域分类器的同时，训练源域分类器，且通过 $\min\limits_{\boldsymbol{w}_t,\boldsymbol{w}_s}\mu\|\boldsymbol{w}_t-\boldsymbol{w}_s\|^2$ 建立两分类器间的联系，两分类器互相学习、互相影响，影响程度由参数 μ 决定。通过这种方式，使源域分类器能将源域样本分类正确，目标域分类器能将目标域已知标签样本分类正确的同时，目标域分类器与源域分类器互相吸引、逼近。

式(6)优化问题对应的拉格朗日形式为

$$\begin{aligned}&L(\boldsymbol{w}_s,\boldsymbol{w}_t,\xi_+^s,\xi_-^s,\xi_+^t,\xi_-^t,\boldsymbol{\alpha},\boldsymbol{\beta})\\&=\frac{1}{2}\|\boldsymbol{w}_t\|^2+\xi_+^t+\xi_-^t+\frac{1}{2}\|\boldsymbol{w}_s\|^2+\xi_+^s+\xi_-^s\\&\quad+\mu\|\boldsymbol{w}_t-\boldsymbol{w}_s\|^2\\&\quad+\sum_{i=1}^m\alpha_i\left(c_i^s-\xi_i^s-y_i^s\left\langle\boldsymbol{w}_s,\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_i^s)\right\rangle\right)\\&\quad+\sum_{j=1}^n\beta_j\left(c_j^t-\xi_j^t-y_j^t\left\langle\boldsymbol{w}_t,\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_j^t)\right\rangle\right)\end{aligned}\tag{7}$$

其中，若源域 $\boldsymbol{x}_i^s\in I_+^s$, $\xi_i^s=\xi_+^s$，若 $\boldsymbol{x}_i^s\in I_-^s$, $\xi_i^s=\xi_-^s$，目标域 ξ_i^t 设置同理。$\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_m)^{\mathrm{T}}$，$\boldsymbol{\beta}=(\beta_1,\beta_2,\cdots,\beta_n)^{\mathrm{T}}$ 为拉格朗日乘子列向量。分别置方程 $L(\boldsymbol{w}_s,\boldsymbol{w}_t,\xi_+^s,\xi_-^s,\xi_+^t,\xi_-^t,\boldsymbol{\alpha},\boldsymbol{\beta})$ 对原始变量 $\boldsymbol{w}_s$, $\boldsymbol{w}_t$, ξ_+^s, ξ_-^s, ξ_+^t 和 ξ_-^t 的偏导数为 0，可得

$$\left.\begin{aligned}&\boldsymbol{w}_s=\frac{2\mu\sum\limits_{j=1}^n\beta_jy_j^t\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_j^t)+(2\mu+1)\sum\limits_{i=1}^m\alpha_iy_i^s\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_i^s)}{4\mu+1}\\&\boldsymbol{w}_t=\frac{2\mu\sum\limits_{i=1}^m\alpha_iy_i^s\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_i^s)+(2\mu+1)\sum\limits_{j=1}^n\beta_jy_j^t\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_j^t)}{4\mu+1}\\&\sum_{i\in I_+^s}\alpha_i=1,\ \sum_{i\in I_-^s}\alpha_i=1,\ \sum_{j\in I_+^t}\beta_j=1,\ \sum_{j\in I_-^t}\beta_j=1\end{aligned}\right\}\tag{8}$$

将式(8)代入式(7)，并设 $\boldsymbol{A}=[a_i]=[Y_i^s\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_i^s)]$, $\boldsymbol{B}=[b_j]=[Y_j^t\phi_{\sqrt{2}\sigma}(\boldsymbol{x}_j^t)]$，$\boldsymbol{cc}=[c_1^s,c_2^s,\cdots,c_m^s,c_1^t,c_2^t,\cdots,c_n^t]^{\mathrm{T}}$，则上述优化问题对应的二次规划形式为

$$\left.\begin{aligned}&\min_{\boldsymbol{\alpha},\boldsymbol{\beta}}[\boldsymbol{\alpha}^{\mathrm{T}}\quad\boldsymbol{\beta}^{\mathrm{T}}]\frac{\begin{bmatrix}(2\mu+1)\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}&2\mu\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}\\2\mu\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}&(2\mu+1)\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\end{bmatrix}}{2(4\mu+1)}\\&\qquad\cdot\begin{bmatrix}\boldsymbol{\alpha}\\\boldsymbol{\beta}\end{bmatrix}-\begin{bmatrix}\boldsymbol{\alpha}^{\mathrm{T}}&\boldsymbol{\beta}^{\mathrm{T}}\end{bmatrix}\cdot\boldsymbol{cc}\\&\text{s.t.}\ \sum_{i\in I_+^s}\alpha_i=\sum_{i\in I_-^s}\alpha_i=1,\ \sum_{j\in I_+^t}\beta_j=\sum_{j\in I_-^t}\beta_j=1\end{aligned}\right\}\tag{9}$$

最终，源域训练集、目标域训练集的最优分类器为：

源域训练集判别函数：

$$f^s(\boldsymbol{x}^s)=\operatorname{sign}\left(\left\langle\boldsymbol{w}_s^{\mathrm{T}},\phi_{\sqrt{2}\sigma}(\boldsymbol{x}^s)\right\rangle\right)\tag{10}$$

目标域训练集判别函数：

$$f^t(\boldsymbol{x}^e)=\operatorname{sign}\left(\left\langle\boldsymbol{w}_t^{\mathrm{T}},\phi_{\sqrt{2}\sigma}(\boldsymbol{x}^e)\right\rangle\right)\tag{11}$$

3.3 TL-L_2KC 算法分析

由式(9)可知，TL-L_2KC 优化问题可转化成一个二次规划问题，式(9)中核函数 $\boldsymbol{K}(\cdot)$ 只有保证 Mercer 核时，才能保证其是二次凸规划，所求解才为全局最优解。为了验证这一问题，我们给出如下定理。

引理 1[13] 定义在 $R^n \times R^n$ 上的对称函数 $\boldsymbol{K}$ 是 Mercer 核函数的充要条件是对任意的 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_l \in R^n$，$\boldsymbol{K}$ 关于 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_l$ 的 Gram 矩阵是半正定的。

定理 1 式(9)的核函数是 Mercer 核。

证明 式(9)核矩阵为

$$\boldsymbol{K} = \begin{bmatrix} (2\mu+1)\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} & 2\mu\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B} \\ 2\mu\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A} & (2\mu+1)\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \end{bmatrix}$$

显然 $\boldsymbol{K}$ 是对称矩阵。下面证明矩阵 $\boldsymbol{K}$ 是半正定的。

设任意 $c_1, c_2, \cdots, c_m, c_{m+1}, \cdots, c_{m+n} \in R$

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{c} = \boldsymbol{c}^{\mathrm{T}}\left(2\mu\begin{bmatrix} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{B} \\ \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A} & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \end{bmatrix} + \begin{bmatrix} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} & [\boldsymbol{0}]_{m\times n} \\ [\boldsymbol{0}]_{n\times m} & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \end{bmatrix}\right)\boldsymbol{c}$$

$$= \boldsymbol{c}^{\mathrm{T}}\left(2\mu\begin{bmatrix} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{B} \\ \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A} & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \end{bmatrix}\right)\boldsymbol{c} + \boldsymbol{c}^{\mathrm{T}}\left(\begin{bmatrix} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} & [\boldsymbol{0}]_{m\times n} \\ [\boldsymbol{0}]_{n\times m} & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \end{bmatrix}\right)\boldsymbol{c}$$

证明上式第 1 部分 ≥ 0：

设 $\boldsymbol{Z} = [\boldsymbol{A} \quad \boldsymbol{B}]$

$$\boldsymbol{c}^{\mathrm{T}}\left(2\mu\begin{bmatrix} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{B} \\ \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A} & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \end{bmatrix}\right)\boldsymbol{c} = \boldsymbol{c}^{\mathrm{T}}\left(2\mu\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\right)\boldsymbol{c}$$

$$= \sum_{i=1}^{m+n}\sum_{j=1}^{m+n} c_i c_j (2\mu z_i z_j)$$

$$= 2\mu\left(\sum_{i=1}^{m+n} c_i z_i\right)\left(\sum_{j=1}^{m+n} c_j z_j\right) = 2\mu\left(\sum_{i=1}^{m+n} c_i z_i\right)^2 \geq 0$$

上式中 $\mu \geq 0$。

证明上式第 2 部分 ≥ 0：

设 $\boldsymbol{a} = [c_1, c_2, \cdots, c_m]$，$\boldsymbol{b} = [c_{m+1}, c_{m+2}, \cdots, c_{m+n}]$，则

$$\boldsymbol{c}^{\mathrm{T}}\left(\begin{bmatrix} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} & [\boldsymbol{0}]_{m\times n} \\ [\boldsymbol{0}]_{n\times m} & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \end{bmatrix}\right)\boldsymbol{c}$$

$$= \begin{bmatrix} \boldsymbol{a}^{\mathrm{T}} & \boldsymbol{b}^{\mathrm{T}} \end{bmatrix}\left(\begin{bmatrix} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} & [\boldsymbol{0}]_{m\times n} \\ [\boldsymbol{0}]_{n\times m} & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \end{bmatrix}\right)\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix}$$

$$= \boldsymbol{a}^{\mathrm{T}}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)\boldsymbol{a} + \boldsymbol{b}^{\mathrm{T}}\left(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\right)\boldsymbol{b}$$

由第 1 部分证明可知，式子第 2 部分同样 ≥ 0。故 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{c} \geq 0$，即矩阵 $\boldsymbol{K}$ 半正定为 Mercer 核函数。证毕

引理 2[13] 假设二次规划中的 Gram 矩阵为半正定矩阵，则该二次规划为凸二次规划。

引理 3[13] 假设二次规划为凸二次规划，则 KKT 条件也是充分条件，故得到的二次规划解为全局最优解。

定理 2 设 $\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}$ 是对偶问题式(9)的解，则原始问题式(6)对于 $\boldsymbol{w}_t$ 的解为全局最优解，并可表示为

$$\boldsymbol{w}_t = \frac{2\mu\sum_{i=1}^{m}\alpha_i y_i^s \phi_{\sqrt{2}\sigma}(\boldsymbol{x}_i^s) + (2\mu+1)\sum_{j=1}^{n}\beta_j y_j^t \phi_{\sqrt{2}\sigma}(\boldsymbol{x}_j^t)}{4\mu+1}$$

证明 根据引理 2 及定理 1 的证明，可知式(9)为凸二次规划，而又根据引理 3 的满足条件可知，该二次规划的解为全局最优解。

值得指出的是，上式 $\dfrac{2\mu\sum_{i=1}^{m}\alpha_i y_i^s \phi_{\sqrt{2}\sigma}(\boldsymbol{x}_i^s)}{4\mu+1}$ 是从源域学到的知识，$\dfrac{(2\mu+1)\sum_{j=1}^{n}\beta_j y_j^t \phi_{\sqrt{2}\sigma}(\boldsymbol{x}_j^t)}{4\mu+1}$ 是从当前目标域数据中学到的新知识。证毕

故式(9)TL-L_2KC 模型与其它核化方法(如 SVM, SVDD)一样，其实质都是求解 QP 问题。式(9)QP 问题的规模为源域训练集规模 m 与目标域训练集规模 n 之和，故该算法的时间复杂度为 $O((m+n)^3)$，空间复杂度为 $O((m+n)^2)$。

4 实验结果与分析

本文实验分两部分：第 1 部分通过人造数据集实验验证算法有效性及参数设置方法；第 2 部分在 UCI 真实数据集上进行实验并与 TrSVM 算法进行比较。

4.1 人造数据集实验

4.1.1 L_2KC 与 TL-L_2KC 对比实验 构造如下两类高斯分布数据各 100 个，如图 1 所示。下三角形表示源域正类训练集，上三角形表示源域负类训练集，两类数据方差均为 1，均值分别为 0 和 2.8。按相同方式生成两类高斯分布目标域测试集各 100 个，两类数据方差为 1，均值分别为 0 和 2.5。为了表示数据的缓慢变化，将两数据集向 X 轴正向移动 2 个单位，如图 2 所示，点号表示目标域正类数据，加号表示目标域负类数据，下三角形表示目标域已标签正类样本，上三角形表示目标域已标签负类样本。这样生成的两域数据既存在相关性，又不完全相同。

实验从源域样本 L_2KC 训练、目标域已知标签样本 L_2KC 训练、源域样本与目标域已知样本合并 L_2KC 训练及源域样本与目标域已标签样本 TL-L_2KC 训练 4 个方面进行研究，并分析其相关原因。

图 2 源域分类器对测试集分类精度为 0.8650。

由图可知，若直接使用源域分类器对目标域测试数据进行分类，由于目标域数据集发生了偏移，分类效果并不理想。图 2 横曲线表示训练目标域已标签样本获得的分类器，分类精度为 0.72。由图所知，由于目标域样本过少，获得信息不足以反映数据分布规律，分类精度较低。图 2 与源域分类器靠近的曲线表示将源域数据和目标域已标签数据合并训练后的分类器，分类精度为 0.8750。由图 2 可知，虽然分类精度比只有源域样本参与训练有所提高，但由于目标域已标签数据占的比重相当小，故两类分类器与仅训练源域的分类器差别不大，精度提升不明显。鉴于此，运用本文提出的 TL-L_2KC 迁移学习算法对同样的数据集进行训练。

图 3 显示了迁移学习后的两类分类器，分类精度为 0.94。可看出，TL-L_2KC 的最优分类器介于源域数据和目标域已标签样本数据分类器之间，既考虑到了目标域已标签样本本身的作用，又学习了源域数据集的分类经验，分类精度显著提升。

4.1.2 参数设置及寻优　TL-L_2KC 方法参数有 σ，μ，其中 σ 为高斯核宽，μ 为自适应参数。Kim 等人[2−4]指出，当数据集维数超过 15 维，L_2 核分类器分类性能降低。原因是二次项高斯核宽为 $\sqrt{2}\sigma$，而一次项高斯核宽为 σ，而两个核的常数项为 $(4\pi\sigma^2)^{-d/2}$ 和 $(2\pi\sigma^2)^{-d/2}$，故两常数项的比值为 $\sqrt{2}^d$，当维数高时，一次项占二次规划的主导作用。为解决此问题，引入大于 0 的参数 η，平衡一次项与二次项。二次规划形式如式(12)：

$$\left.\begin{aligned}\min_{\boldsymbol{\alpha},\boldsymbol{\beta}}[\boldsymbol{\alpha}^{\mathrm{T}}\quad\boldsymbol{\beta}^{\mathrm{T}}]&\frac{\begin{bmatrix}(2\mu+1)\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} & 2\mu\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}\\ 2\mu\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A} & (2\mu+1)\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\end{bmatrix}}{2(4\mu+1)}\\ &\cdot\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{bmatrix}-\frac{1}{\eta}\begin{bmatrix}\boldsymbol{\alpha}^{\mathrm{T}}\quad\boldsymbol{\beta}^{\mathrm{T}}\end{bmatrix}\cdot\boldsymbol{cc}\\ \text{s.t.}\ \sum_{i\in I_+^s}\alpha_i=&\sum_{i\in I_-^s}\alpha_i=1,\ \sum_{j\in I_+^t}\beta_j=\sum_{j\in I_-^t}\beta_j=1\end{aligned}\right\}\quad(12)$$

如上所述，参数设置分两种情况，若数据集维数 d 小于等于 15 维，则采用 5 重交叉验证方式获得最优的 (σ,μ) 参数对；若数据集维数 d 大于 15 维，则引入参数 η，用于平衡一次项与二次项，同样采用 5 重交叉验证方式获得最优的 (σ,μ,η) 参数组合。参数设置如表 1。其中，参数 σ,η 的设置方式如文献[4]。交叉验证规则是源域分类器对源域样本集、目标域分类器对目标域已标签样本集分类精度均值达到最优。

表 1 参数设置

参数	TL-L_2KC 算法参数预设值
σ	0.01~10 等分成 50 点
μ	0.1,0.2,0.3,$\cdots$,1
η	1~$\sqrt{2}^d$ 等分成 20 点

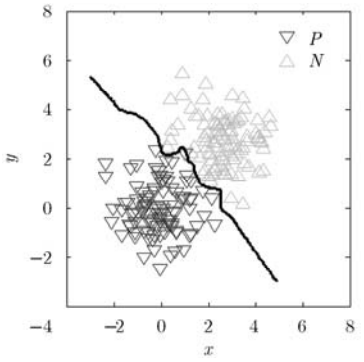

图 1 源域 L_2KC 分类效果图

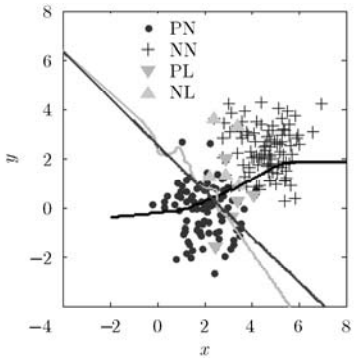

图 2 目标域已标签样本 L_2KC 分类效果图

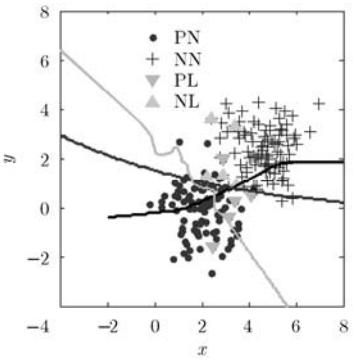

图 3 TL-L_2KC 分类效果图

表 2 数据集描述

序号	数据集	规模	属性个数	源域样本数	目标域样本数	目标域已知标签的样本数
1	Diabetes	768	8	80	688	34
2	Ionosphere	351	34	86	265	13
3	sonar	208	60	42	166	9
4	iris	150	4	20	130	4
5	wine	178	13	41	137	6

4.2 真实数据集实验

4.2.1 数据集描述 选用表 2 中 UCI 数据集进行实验，且实验按源域、目标域、源域目标域合并、源域目标域迁移学习等 4 方面进行训练与测试，并与 TrSVM 迁移算法进行比较。

按文献[14]构造源域与目标域数据集。构造源域特征偏差(feature bias)数据集，使源域、目标域分布不同。方法如下：随机选择 50%属性列，每一列属性值根据一定规则排序，规则为若是字符按字典顺序、若是数值按数值大小顺序排序，对每一属性排序结果选取排序列表最上方样本，构成源域样本集，所选取的样本规模见表 2。剩下的样本构成目标域样本集。目标域已知标签样本随机选取，不超过目标域样本总数的 5%。构造的源域训练集和目标域训练集需满足如下条件：(1)源域因特征偏差，不能代表目标域测试集所有特征，源域分类器对目标域分类精确度不高；(2)目标域已知标签样本包含目标域分布信息，但样本过少，目标域分类器分类精度也不高。

4.2.2 UCI 数据集上 TL-L_2KC 算法的实验结果及分析 表 3，表 4 列出了 5 个数据集 4 种情况下的实验结果及参数设置情况。

通过表 3，表 4 可知，因源域训练集与目标域分布存在偏差，故直接利用源域分类器对测试集进行分类，其分类精度最低；若直接使用目标域训练集构造分类器，因目标域训练集与目标域测试集同分布，故分类效果优于源域分类器；若将源域与目标域合并训练，测试集分类精度不一定提升，因为源域训练集规模大于目标域训练集，源域训练集占主导作用；若将两训练集进行迁移学习，则既能保证目标域训练集对目标域分类器的主导作用，又能学习源域训练集已有的知识，故得到较好的分类精度。

4.2.3 TL-L_2KC 与 TrSVM 实验比较 TL-L_2KC 算法的参数设置如表 1。TrSVM 算法按文献[12]设置参数 C,C_s，采用高斯核，具体见表 5。

表 6 为 TL-L_2KC 算法与 TrSVM 算法的迁移学习效果对比表。通过表 6 的实验结果发现，因 TL-L_2KC 算法同时训练源域、目标域两个分类器，可通过参数 μ 灵活控制学习源域的程度，故本文提出的 TL-L_2KC 算法在 5 个 UCI 真实数据集上的分类精确率有 3 组明显优于 TrSVM 算法，有 1 组与 TrSVM 算法持平，只有 1 组略低于 TrSVM 算法。总之，本文提出的 TL-L_2KC 算法具有较好的性能。

5 结束语

L_2 核分类器具有解的稀疏性、较高的决策效率且可用于概率密度估计等优势，但其无法进行迁移学习，不适用数据集缓慢变化，训练集、测试集分布不同的场景。为弥补此缺陷，本文从 L_2 核分类器等价的 SVM 出发，提出面向迁移学习的 L_2 核分类器，即 TL-L_2KC 算法，并通过在人造数据集和 UCI 真实数据集上进行实验，进一步验证了 TL-L_2KC 算法的有效性。

表 3 源域训练与目标域训练实验结果及参数设置

数据集	源域训练			目标域训练		
	精确率	σ	η	精确率	σ	η
Diabetes	0.6265	0.4178	1	0.6701	0.2139	1
Ionosphere	0.2377	2.2527	4.1392e+004	0.6868	0.2139	1
Sonar	0.5482	0.2139	5.6513e+007	0.6807	0.2139	1
iris	0.7462	0.0100	1	0.9077	0.0100	1
wine	0.7007	0.6216	1	0.8905	0.2139	1

表 4 源域目标域合并训练及迁移学习实验结果及参数设置

数据集	源域目标域合并训练			迁移学习			
	精确率	σ	η	精确率	σ	η	μ
Diabetes	0.6788	0.2139	1	0.7093	0.0100	1	0.4
Ionosphere	0.7396	0.2139	1	0.8943	0.2139	6.8995e+003	0.2
Sonar	0.6566	0.4178	1.1303e+008	0.7892	0.2139	5.6513e+007	1.0
iris	0.9077	0.0100	1	0.9538	0.0100	1	0.1
wine	0.8029	0.6216	1	0.9489	0.0100	1	0.1

表 5 TrSVM 参数设置

参数	TrSVM 算法参数预设值
C	0.5, 1, 1.5, 2
C_S	0.5, 1, 1.5, 2
σ	$2^{-2} \sim 2^{7}$ 等分成 50 点

表 6 TL-L_2KC 算法与 TrSVM 算法的迁移学习效果对比表

数据集	TL-L_2KC	TrSVM			
	精确率	精确率	σ	C	C_s
Diabetes	**0.7093**	0.7049	73.2500	0.5	2.0
Ionosphere	**0.8943**	0.8642	8.0714	2.0	2.0
Sonar	**0.7892**	0.7530	1.0000	1.0	0.5
iris	**0.9538**	**0.9538**	0.2500	0.5	0.5
wine	0.9489	**0.9562**	0.2500	0.5	0.5

参 考 文 献

[1] 张战成, 王士同, 邓赵红, 等. 支持向量机的一种快速分类算法[J]. 电子与信息学报, 2011, 33(9): 2181–2186.
Zhang Zhan-cheng, Wang Shi-tong, Deng Zhao-hong, *et al*.. Fast decision using SVM for incoming samples[J]. *Journal of Electronics & Information Technology*, 2011, 33(9): 2181–2186.

[2] Kim J and Scott C. Kernel classification via integrated squared error[C]. Proceedings of the IEEE 14th Workshop on Statistical Signal Processing, Madison, 2007: 783–787.

[3] Kim J and Scott C. Performance analysis for L_2 kernel classification[C]. Proceedings of Advances in Neural Information Processing Systems, Vancouver, 2008: 836–843.

[4] Kim J and Scott C. L_2 kernel classification[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(10): 1822–1831.

[5] Bruzzone L and Marconcini M. Domain adaptation problems: a DASVM classification technique and a circular validation strategy[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(5): 770–787.

[6] Pan S J, Tsang I W, Kwok J T, *et al*.. Domain adaptation via transfer component analysis[J]. *IEEE Transactions on Neural Networks*, 2011, 22(2): 199–210.

[7] Zhang Hu-xiang. Transfer learning through domain adaptation[C]. Proceedings of the 8th International Symposium on Neural Networks, Guilin, 2011: 505–512.

[8] 于重重, 田蕊, 谭励, 等. 非平衡样本分类的集成迁移学习算法[J]. 电子学报, 2012, 40(7): 1358–1363.
Yu Chong-chong, Tian Rui, Tan Li, *et al*.. Integrated transfer learning algorithmic for unbalanced samples classification[J]. *Acta Electronica Sinica*, 2012, 40(7): 1358–1363.

[9] 张建军, 王士同, 王骏. 迁移学习数据分类中的 ESVM 算法[J]. 计算机工程, 2012, 38(8): 173–176.
Zhang Jian-jun, Wang Shi-tong, and Wang Jun. ESVM algorithm in transfer learning data classification[J]. *Computer Engineering*, 2012, 38(8): 173–176.

[10] Pan S J and Yang Q. A survey on transfer learning [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10): 1345–1359.

[11] Shi Y, Lan Z, Liu W, *et al*.. Extending semi-supervised learning methods for inductive transfer learning[C]. Proceedings of the 9th IEEE International Conference on Data Mining, Los Alamitos, 2009: 483–492.

[12] 洪佳明, 印鉴, 黄云, 等. TrSVM: 一种基于领域相似性的迁移学习算法[J]. 计算机研究与发展, 2011, 48(10): 1823–1830.
Hong Jia-ming, Yin Jian, Huang Yun, *et al*.. TrSVM: a transfer learning algorithm using domain similiary[J]. *Journal of Computer Research and Development*, 2011, 48(10): 1823–1830.

[13] 邓乃杨, 田英杰. 数据挖掘的新方法——支持向量机[M]. 北京: 科学出版社, 2004: 92–96.
Deng Nai-yang and Tian Ying-jie. Theory, Algorithms and Expansion-Support Vector Machines[M]. Beijing: Science Press, 2004: 92–96.

[14] Ren J, Shi X, Fan W, *et al*.. Type independent correction of sample selection bias via structural discovery and re-balancing[C]. Proceedings of the Eighth SIAM International Conference on Data Mining, Atlanta, 2008: 565–576.

许　敏：女，1980 年生，讲师，博士生，研究方向为模式识别、人工智能等.

王士同：男，1964 年生，教授，博士生导师，主要研究方向为模式识别、人工智能、数据挖掘、模糊系统等.

史荧中：男，1970 年生，讲师，博士生，研究方向为模式识别、人工智能等.