

基于共同决策方向矢量的多源迁移 及其快速学习方法

张景祥^{1,2}, 王士同¹

(1. 江南大学数字媒体学院, 江苏无锡 214122; 2. 江南大学理学院, 江苏无锡 214122)

摘 要: 多源迁移学习提取了多个相似领域之间有用信息, 提高了学习效率, 但存在计算核矩阵的空间和时间复杂度较高的问题. 提出了一种多源迁移学习方法, 该方法基于结构风险最小框架理论, 以共同决策方向矢量为基准, 将多个相似领域的决策方向矢量嵌入到支持向量机的训练过程中, 提高了目标领域分类器的分类性能. 并结合核心向量机理论提出了共同决策方向矢量核心向量机, 实现对大样本数据集的快速分类学习. 模拟和真实数据集实验表明了所提算法的有效性.

关键词: 共同决策矢量; 多源迁移学习; 分类; 核心集向量机

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2015)07-1349-07

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2015.07.015

Common-Decision-Vector Based Multiple Source Transfer Learning Classification and Its Fast Learning Method

ZHANG Jing-xiang^{1,2}, WANG Shi-tong¹

(1. School of Digital Media, Jiangnan University, Wuxi, Jiangsu 214122, China;

2. School of Science, Jiangnan University, Wuxi, Jiangsu 214122, China)

Abstract: Multiple source transfer learning (MSTL) has been obtaining more and more applications especially from several related source domains to help the learning task on target domain. However, multiple source transfer learning algorithms often deal with the corresponding quadratic programming problems which may suffer a big computational burden caused by the kernel matrix computation. In this paper, a novel common-decision-vector based multiple source transfer classification learning (CDV-MSTL) is proposed which doesn't depend on the intrinsic structure of data. This algorithm is based on the structural risk minimization principle and the SVM like framework, so it has good adaptability and better accuracy. Based on the theory of CVM, CDV-MSTL is extended to its CVM based version which can realize fast training for large scale data. Extensive experiments on synthetic and real-world datasets demonstrate the significant improvement in classification performance obtained by the proposed algorithm over existing MSTL algorithm.

Key words: common decision vector; multiple source transfer learning; classification; core vector machine

1 引言

多源迁移学习是在迁移学习的基础上提出的一种适用性更广的学习框架, 其强调在多个相似但存在一定差别的领域、任务和分布之间进行知识迁移, 从而实现提高目标领域分类器泛化性能的目的. 多源迁移学习继承了迁移学习的灵活性, 同时兼顾了多任务学习的适应性, 已经成为近几年机器学习领域的研究热点.

很多学者从不同的应用场景提出了多源迁移学习的算法, 大致可分为两类: 一类是目标领域有少量数据标签已知, 通过估计或减小小领域间的差异实现多领域间的适应学习. 基于流形结构的 MMDE 算法^[1,2]; 基于香农熵的一致性^[3]; 多源多视角 Adaboost 迁移学习算法^[4,5]; 另一类是目标领域数据标签完全未知, 借助源领域的数据或分类器实现迁移学习. 基于 LSSVM 光滑假设^[6]; 基于异构一致性学习框架^[7]; ML 型模糊迁移系统^[8].

上述方法存在核矩阵空间复杂度为 $\mathcal{O}(N^2)$ 以及时间复杂度为 $\mathcal{O}(N^3)$ 高的问题, 对于大样本数据集参数寻优过程的时间复杂度则更高, 因而对算法的适应性影响较大. 针对大样本, 虽有算法^[9-11], 但多源迁移学习大样本问题还没有提出有效的解决方法.

针对上述问题, 本文提出了一种基于共同决策方向矢量的多源迁移学习方法 (Common-Decision-Vector Based Multiple Source Transfer Learning, CDV-MSTL). 该方法以目标领域的决策矢量为共同决策方向矢量, 兼顾到各源领域上一个较小的特有偏移决策方向矢量, 基于结构风险最小框架理论将全部领域的决策方向矢量嵌入到特定核空间中的支持向量机训练过程中, 通过优化可以较好的实现多领域间知识迁移学习; 其次, 针对大样本的问题, 结合中心约束最小包含球 (Center-Constrained Minimal Enclosing Ball, CCMEB) 理论^[12,13], 提出 CDV-MSTL 快速求解版本 CDV-CVM 算法.

2 基于共同决策方向矢量的多源迁移学习

2.1 相关概念及问题描述

为了便于讨论相关的迁移学习算法, 首先对一些相关概念和符号进行统一说明.

定义 1 领域 (Domain): 领域常用 $D = \{(\mathbf{x}_i, \mathbf{y}_i) \in \mathbf{X} \times \mathbf{Y}\}$ 来表示. 其中对于上述目标函数, 说明如下:

$\mathbf{X} \in \mathbf{R}^d$ 表示领域样本的特征空间, \mathbf{Y} 表示样本的标签向量, $(\mathbf{x}_i, \mathbf{y}_i)$ 表示领域 D 的一个样本.

由定义 1 知, 多源迁移学习中存在多个源领域可表示为: $\mathbf{SD} = \{\mathbf{SD}_1, \dots, \mathbf{SD}_n\}$; 目标领域表示为: $\mathbf{TD} = \{(\mathbf{x}_i^1, \mathbf{y}_i^1), \dots, (\mathbf{x}_i^{N_i}, \mathbf{y}_i^{N_i})\}$.

2.2 CDV-MSTL 算法

根据多源迁移学习的策略, 本文引入共同决策方向矢量作为各源领域和目标领域之间的桥梁, 提出多源迁移学习模型, 其目标函数为:

$$\min_{\mathbf{w}_t, \Delta \mathbf{w}_j, \Delta b_j^S, \xi_j^S} \frac{1}{2} \|\mathbf{w}_t\|^2 + \frac{1}{2} b_t^2 + \frac{\gamma}{2n} \sum_{j=1}^n (\lambda_j \Delta \mathbf{w}_j)^2 + C \sum_{j=1}^n \sum_{i=1}^{N_j^S} \xi_{ij}^S + \frac{\gamma}{n} \sum_{j=1}^n \Delta b_j^S \quad (1)$$

$$\text{s.t. } y_{ij}^S ((\mathbf{w}_t + \lambda_j \Delta \mathbf{w}_j) \varphi(\mathbf{x}_{ij}^S) + (b_t + \lambda_j \Delta b_j^S)) \geq 1 - \xi_{ij}^S$$

$$\xi_{ij}^S \geq 0, i = 1, 2, \dots, N_j^S; j = 1, 2, \dots, n$$

这里通过引入常数向量 $\boldsymbol{\lambda} = (0, 1, 1, \dots, 1)_{1 \times n}$ 可实现将 n 个领域划分为目标领域和 $n-1$ 个源领域. \mathbf{x}_i^t 表示目标领域的第 i 个样本 ($i = 1, 2, \dots, N_t$, N_t 表示目标域样本个数); \mathbf{x}_{ij}^S 表示第 j 个源领域的第 i 个样本 ($i = 1, 2, \dots, N_j^S$, N_j^S 表示第 j 个源领域的样本个数, $j = 1, 2, \dots, n-1$); \mathbf{w}_t 表示目标领域数据在 RKHS 内待求解的分类器的决策方向矢量, $\Delta \mathbf{w}_j$ 表示第 j 个源领域在 RKHS 内

待求决策方向矢量与共同决策方向矢量的偏差矢量; b_t 表示目标领域的待求解的分类器偏置项; Δb_j^S 表示源领域待求解分类器偏置差异量; 参数 C 是正则化参数; ξ_{ij}^S 和 ξ_i^t 分别表示各源领域和目标领域分类器的松弛变量.

定理 1 对于式(1)所示优化问题, 其对偶问题可表示为如下的二次规划形式:

$$\arg \max_{\boldsymbol{\tau}} -\frac{1}{2} \boldsymbol{\tau}^T \mathbf{H} \boldsymbol{\tau} + \mathbf{1}^T \boldsymbol{\tau} \quad (2)$$

$$\text{s.t. } \sum_{j=1}^n \tau_j \mathbf{l}_j = \mathbf{0}, 0 < \tau_j < C, j = 1, \dots, n$$

其中, $\boldsymbol{\tau} = (\underbrace{\alpha_{1,1}, \dots, \alpha_{N_s,1}}_{N_s}, \dots, \underbrace{\alpha_{1,n}, \dots, \alpha_{N_s,n}}_{N_s})^T = (\tau_1, \dots, \tau_n)^T$ 为拉格朗日乘子. $\mathbf{l} = (\mathbf{l}_1, \dots, \mathbf{l}_n)$ 为标签向

量; $\mathbf{H} = \begin{pmatrix} \tilde{\mathbf{H}}_1 + \hat{\mathbf{H}}_{1,1} & \cdots & \hat{\mathbf{H}}_{n,1} \\ \hat{\mathbf{H}}_{1,2} & \cdots & \hat{\mathbf{H}}_{n,2} \\ \vdots & \ddots & \vdots \\ \hat{\mathbf{H}}_{1,n} & \cdots & \tilde{\mathbf{H}}_n + \hat{\mathbf{H}}_{n,n} \end{pmatrix}_{n \times n}$ 为核矩阵, 其

中 $\tilde{\mathbf{H}}_k = [\hat{\mathbf{h}}_{ij}^S]_{N_i^S \times N_j^S}$, $\hat{\mathbf{h}}_{ij}^S = \frac{n}{\gamma} y_{i,k}^S y_{j,k}^S \varphi(\mathbf{x}_{j,k}^S)^T \varphi(\mathbf{x}_{i,k}^S)$, $\hat{\mathbf{H}}_{k,l} = [\hat{\mathbf{h}}_{ij}^S]_{N_i^S \times N_j^S}$, $\hat{\mathbf{h}}_{ij}^S = y_{i,l} y_{j,k} \varphi(\mathbf{x}_{j,k}^S)^T \varphi(\mathbf{x}_{i,k}^S)$; $\mathbf{1}$ 是元素为 1 的向量.

证明 构造拉格朗日函数:

$$L(\underbrace{\alpha_{1,1}, \dots, \alpha_{N_s,1}}_{N_s}, \dots, \underbrace{\alpha_{1,n}, \dots, \alpha_{N_s,n}}_{N_s}) = \frac{1}{2} \|\mathbf{w}_t\|^2 + \frac{\gamma}{2n} \sum_{j=1}^n (\lambda_j \Delta \mathbf{w}_j)^2 + C \sum_{j=1}^n \sum_{i=1}^{N_j^S} \xi_{ij}^S + \frac{\gamma}{n} \sum_{j=1}^n \Delta b_j^S - \sum_{j=1}^n \sum_{i=1}^{N_j^S} \alpha_{ij}^S \xi_{ij}^S - \sum_{j=1}^n \sum_{i=1}^{N_j^S} \alpha_{ij}^S \left[y_{ij}^S ((\mathbf{w}_t + \lambda_j \Delta \mathbf{w}_j)^T \varphi(\mathbf{x}_{ij}^S) + (b_t + \lambda_j \Delta b_j^S)) - 1 + \xi_{ij}^S \right] \quad (3)$$

对 $\mathbf{w}_t, \Delta \mathbf{w}_j, \Delta b_j^S, \xi_{ij}^S$ 求导, 由 KKT 条件可得:

$$\frac{\partial L}{\partial \mathbf{w}_t} = 0 \Rightarrow \mathbf{w}_t = \sum_{j=1}^n \sum_{i=1}^{N_j^S} \alpha_{ij}^S y_{ij}^S \varphi(\mathbf{x}_{ij}^S) \quad (4)$$

$$\frac{\partial L}{\partial (\Delta \mathbf{w}_j)} = 0 \Rightarrow \Delta \mathbf{w}_j = \frac{n}{\gamma} \sum_{i=1}^{N_j^S} \alpha_{ij}^S y_{ij}^S \lambda_j \varphi(\mathbf{x}_{ij}^S) \quad (5)$$

$$\frac{\partial L}{\partial b_t} = 0 \Rightarrow b_t = \sum_{i=1}^{N_j^S} \alpha_{ij}^S y_{ij}^S \quad (6)$$

$$\frac{\partial L}{\partial \Delta b_j^S} = 0 \Rightarrow \frac{n}{\gamma} \sum_{i=1}^{N_j^S} \alpha_{ij}^S y_{ij}^S = 1 \quad (7)$$

$$\frac{\partial L}{\partial \xi_{ij}^S} = 0 \Rightarrow C - \alpha_{ij}^S - \alpha_{ij}^S = 0 \quad (8)$$

将式(4)~式(8)代入式(3), 化简可得式(9)

$$L(\alpha_{1,1}, \dots, \alpha_{N_s,n}) =$$

$$\begin{aligned}
& - \frac{1}{2} \sum_{k=1}^n \sum_{l=1}^n \sum_{i=1}^{N_k^S} \sum_{j=1}^{N_l^S} \alpha_{j,l} \alpha_{i,k} \gamma_{j,l}^S \gamma_{i,k}^S \varphi(x_{j,l}^S) \varphi(x_{i,k}^S) \\
& - \frac{n}{2\gamma} \sum_{l=1}^n \sum_{i=1}^{N_k^S} \sum_{j=1}^{N_l^S} \alpha_{i,k} \alpha_{j,k} \gamma_{i,k}^S \gamma_{j,k}^S \varphi(x_{j,k}^S) \varphi(x_{i,k}^S) \\
& + \sum_{j=1}^n \sum_{i=1}^{N_j} \alpha_{i,j}
\end{aligned} \quad (9)$$

将式(9)写成对偶形式即得到式(2),定理得证。

对于式(2)的二次规划,通常需核矩阵 \mathbf{H} 正定半正定,为此需要其满足 Mercer 核条件,给出如下定理:

引理 1^[14] 设 \mathbf{X} 是 \mathbf{R}^n 上的一个紧集,若 $\mathbf{H}(\mathbf{x}_i, \mathbf{x}_j)$ 是 $\mathbf{X} \times \mathbf{X}$ 上的连续对称函数且关于任意 $\mathbf{x}_i \in \mathbf{X}$ 的 Gram 矩阵半正定,则 $\mathbf{H}(\mathbf{x}_i, \mathbf{x}_j)$ 是 Mercer 核。

定理 2 形如式(3)中 $\mathbf{H} \in \mathbf{R}$ 所表达的核函数是 Mercer 核。

证明 由定理 1 可知 \mathbf{H} 为实数矩阵,对于 $\tilde{\mathbf{H}}_k = [\hat{\mathbf{h}}_{ij}]_{N_i^S \times N_j^S}$, $\tilde{\mathbf{h}}_{ij} = \frac{n}{\gamma} \gamma_{i,k}^S \gamma_{j,k}^S \varphi(x_{j,k}^S)^T \varphi(x_{i,k}^S)$, $\hat{\mathbf{H}}_{k,l} = [\hat{\mathbf{h}}_{ij}]_{N_i^S \times N_j^S}$, $\hat{\mathbf{h}}_{ij} = \gamma_{i,l} \gamma_{j,k} \varphi(x_{j,k}^S)^T \varphi(x_{i,k}^S)$, 由于 $\varphi(x_{j,k}^S)^T \varphi(x_{i,k}^S)$ 为实对称阵,因此 $\tilde{\mathbf{H}}_k$ 是实对称矩阵;同理 $\hat{\mathbf{H}}_{k,j}$ 也是实对称矩阵. 因此 \mathbf{H} 矩阵是一个对称矩阵. 其次,对任意的 α_i , 则:

$$\begin{aligned}
& \sum_{i,j} \alpha_i \alpha_j (\varphi(x_{j,k}^S)^T \varphi(x_{i,k}^S)) \\
& = \left(\sum_{i=1}^{N_i} \alpha_i \varphi(x_{j,k}^S) \right)^T \left(\sum_{i=1}^{N_i} \alpha_i \varphi(x_{j,k}^S) \right) \geq 0 \\
& \sum_{i,j} \alpha_i \alpha_j (\varphi(x_{j,k}^S)^T \varphi(x_{i,k}^S)) \\
& = \left(\sqrt{\frac{n}{\gamma}} \sum_{i=1}^{N_i} \alpha_i \varphi(x_{j,k}^S) \right)^T \left(\sqrt{\frac{n}{\gamma}} \sum_{i=1}^{N_i} \alpha_i \varphi(x_{j,k}^S) \right) \geq 0
\end{aligned} \quad (10)$$

$$\begin{aligned}
& = \left(\sqrt{\frac{n}{\gamma}} \sum_{i=1}^{N_i} \alpha_i \varphi(x_{j,k}^S) \right)^T \left(\sqrt{\frac{n}{\gamma}} \sum_{i=1}^{N_i} \alpha_i \varphi(x_{j,k}^S) \right) \geq 0
\end{aligned} \quad (11)$$

由式(10),式(11)可知, \mathbf{H} 为半正定矩阵且在 $\mathbf{X} \times \mathbf{X}$ 上的连续对称函数,由引理 1 可知,核 \mathbf{H} 为 Mercer 核矩阵。

2.3 CDV-MSTL 算法实现

算法 1 CDV-MSTL 算法

输入:多源数据,初始值;正则化参数 C ,平衡参数 γ 以及核宽度参数 ν ;

Step 1:构造核矩阵 \mathbf{H} ,由式(3)得到各个领域的支持向量 τ ;

Step 2:由式(4)得到 \mathbf{w}_i ,并据此计算 $b_i^* = y_j - \sum_{i=1}^{N_j^S} y_i \tau_i \mathbf{H}(\mathbf{x}_j, \mathbf{x}_i)$;

Step 3: (\mathbf{w}_i , b_i) 核化得到目标领域分类器 $f_i(\mathbf{x}) = \text{sgn}(\sum_{i=1}^{N_j^S} \tau_i^* y_i \mathbf{H}(\mathbf{x}, \mathbf{x}_i) + b_i^*)$, 通过比较测试精度,选择最优参数组合 C, γ, ν ;

输出:目标领域决策函数。

3 基于共同决策方向矢量的多源迁移快速学习方法

最小包含球(Minimal Enclose Ball, MEB)作为一种解决大样本数据的方法在模式识别中得到了广泛研究和应用,通过构造一个球心为 c 和半径 R 的球 $B(c, R)$, 选出样本的核心数据集,借助适当的核函数 φ 将原始样本空间映射到高维特征空间中,有效地降低算法的复杂度,提升了算法的效率。

3.1 最小包含球理论

Tsang 等指出^[15]硬边界 SVDD 问题、软边界单类及二类的 I2-SVM 问题都可以转化为核化 MEB 问题,并基于 MEB 的核心集(Core-set)技术提出的核心向量机(Core Vector Machines, CVM)算法. Tsang 等^[16]又通过引入一个属性项 $\delta_i \in \mathbf{R}$ 对 CVM 进行了扩展,又提出了中心约束最小包含球(CCMEB)算法,核空间中任意样本点 $\varphi(x_i)$ 附加 1 维新特征 $\delta_i \in \mathbf{R}$, 形成新样本 $\begin{pmatrix} \varphi(x_i) \\ \delta_i \end{pmatrix}$, 其优化的对偶形式可以表示为:

$$\begin{aligned}
& \arg \max_{\alpha} (\alpha^T \text{diag}(\mathbf{K}) + \Delta) - \alpha^T \mathbf{K} \alpha \\
& \text{s.t. } \alpha^T \mathbf{1} = 1, \quad \alpha \geq 0
\end{aligned} \quad (12)$$

其中 $\Delta = [\delta_1^2, \delta_2^2, \dots, \delta_N^2]^T \geq 0$, $\eta \in \mathbf{R}$ 为常数;通过求解式(11)的最优解,可以得到最小包含球的中心 c 与半径 R :

$$c = \sum_{i=1}^N \alpha_i \varphi(x_i), R = \sqrt{\alpha^T (\text{diag}(\mathbf{K}) + \Delta) - \alpha^T \mathbf{K} \alpha} \quad (13)$$

由于 $\alpha^T \mathbf{1} = 1$, 故在式(13)的目标函数中增一项 $-\alpha^T \eta \mathbf{1}$, $\eta \in \mathbf{R}$, 得式(14):

$$\begin{aligned}
& \arg \max_{\alpha} \alpha^T (\text{diag}(\mathbf{K}) + \Delta - \eta \mathbf{1}) - \alpha^T \mathbf{K} \alpha \\
& \text{s.t. } \alpha^T \mathbf{1} = 1, \quad \alpha \geq 0
\end{aligned} \quad (14)$$

3.2 CDV-CVM 算法

本文借鉴最小包络球思想,结合迁移学习策略,提出了基于共同决策方向矢量的快速学习方法,仍然以目标领域的决策方向矢量为基准,通过优化各源领域和目标领域之间的决策方向矢量之间的距离,将其转化为 CVM 的形式,给出如下的核空间迁移学习目标函数:

$$\begin{aligned}
& \min_{\mathbf{w}_i, \Delta \mathbf{w}_i, \xi_j, \rho_j} \frac{1}{2} \|\mathbf{w}_i\|^2 + \frac{1}{2} b_i^2 + \frac{\gamma}{2n} \sum_{j=1}^n (\lambda_j \Delta \mathbf{w}_j)^2 \\
& + \frac{\gamma}{n} \sum_{j=1}^n \lambda_j \Delta b_j^S + C \sum_{j=1}^n \sum_{i=1}^{N_j^S} (\xi_{ij}^S)^2 - \nu \rho \\
& \text{s.t. } y_{ij}^S [(\mathbf{w}_i + \lambda_j \Delta \mathbf{w}_j) \varphi(\mathbf{x}_{ij}^S) + (b_i + \lambda_j \Delta b_j^S)] \geq \rho - \xi_{ij}^S \\
& i = 1, 2, \dots, N_j^S; j = 1, 2, \dots, n
\end{aligned} \quad (15)$$

定理 3 对于式(15)所示优化问题,其对偶问题可

表示为如下的二次规划优化形式:

$$\begin{aligned} \arg \max_{\zeta} & -\frac{1}{2} \zeta^T Q \zeta - \frac{\zeta}{4C} \\ \text{s.t.} & \quad \zeta^T \mathbf{1} = 1, \quad \zeta \geq 0 \end{aligned} \quad (16)$$

其中, $\zeta = (\underbrace{\beta_{1,1}, \dots, \beta_{N_{S_1},1}}_{N_{S_1}}, \dots, \underbrace{\beta_{1,n}, \dots, \beta_{N_{S_n},n}}_{N_{S_n}})^T =$

$(\zeta_1, \dots, \zeta_n)^T$ 为拉格朗日算子. $\mathbf{l} = (\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_n)$ 为标

签; $\mathbf{Q} = \begin{pmatrix} \hat{\mathbf{Q}}_1 + \hat{\mathbf{Q}}_{1,1} & \hat{\mathbf{Q}}_{2,1} & \cdots & \hat{\mathbf{Q}}_{n,1} \\ \hat{\mathbf{Q}}_{1,2} & \hat{\mathbf{Q}}_2 + \hat{\mathbf{Q}}_{2,2} & \cdots & \hat{\mathbf{Q}}_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{Q}}_{1,n} & \hat{\mathbf{Q}}_{2,n} & \cdots & \hat{\mathbf{Q}}_n + \hat{\mathbf{Q}}_{n,n} \end{pmatrix}_{n \times n}$ 为

核矩阵, 其中 $\hat{\mathbf{Q}}_k = [\hat{\mathbf{q}}_{ij}]_{N_i^S \times N_j^S}$, $\hat{\mathbf{q}}_{ij} = \frac{n}{\gamma} y_{i,k} y_{j,k} \varphi(x_{j,k}^S)^T \varphi(x_{i,k}^S)$, $\tilde{\mathbf{Q}}_{k,l} = [\tilde{\mathbf{q}}_{ij}]_{N_i^S \times N_j^S}$, $\tilde{\mathbf{q}}_{ij} = y_{i,l} y_{j,k} \varphi(x_{j,k}^S)^T \varphi(x_{i,k}^S)$. (证明略)

定理 4 式(16)等价于 CC-MEB 问题.

证明 这里取 $\Delta = \eta \mathbf{1} - \text{diag}(\mathbf{Q}) - \frac{1}{2C}$. 则式(16)可以重新表示为:

$$\begin{aligned} \arg \max_{\zeta} & \frac{1}{2} (\zeta^T (\text{diag}(\mathbf{Q}) + \Delta - \eta \mathbf{1}) - \zeta^T Q \zeta) \\ \text{s.t.} & \quad \zeta^T \mathbf{1} = 1, \quad \zeta \geq 0 \end{aligned} \quad (17)$$

取 $\eta = \max |\text{diag}(\mathbf{Q})| + \left| \frac{1}{2C} \right|$ 时, $\Delta \geq 0$. 证毕.

因此, 利用样本核心集训练获得判别函数, 从而快速解决大样本问题.

算法 2 CDV-CVM 算法

输入: 数据, 核心集逼近精度 ϵ 以及 η 等参数

Step 1: 初始化 $t = 0$, S_t , c_t 和 R_t ;

Step 2: 所有样本点均被 $B(c_t, (1 + \epsilon) R_t)$ 包围, 则 $S = S_t$ 转入 Step 6; 否则转入 Step 3;

Step 3: 找出离 c_t 最远的点 z , 加入核心集中, 即 $S_{t+1} = S_t \cup \{z\}$;

Step 4: 利用式(14) 求解新的球心 c_{t+1} 和半径 R_{t+1} , 即新的 MEB: $B(c_{t+1}, R_{t+1})$;

Step 5: $t = t + 1$, 转 Step 2;

Step 6: 终止训练, 返回最终的核心集 S , 采用 CDV 对核心集 S 训练

输出: 目标领域分类决策函数.

4 实验分析

为了验证文本所提 CDV-MSTL 和 CDV-CVM 算法的有效性, 将在不同类型数据集上对其进行性能评估, 所采用的数据集主要包括如下两大类: (1) 模拟数据集; (2) 真实数据集. 实验中将 CDV-MSTL 和 CDV-CVM 与相关的方法进行了性能比较, 采用的比较方法有: CCR^[3],

LSSVM^[17], TrAdaBoost^[18], MultiSource-TrAdaBoost^[19], TL-SSVM^[20], MT-SVM^[21] 和 CoCC^[22]. 算法参数均根据文献的具体应用通过 5 重交叉验证获得.

本文算法采用高斯核函数, 参数值通过交叉验证获得取.

4.1 双月模拟数据集

4.1.1 双月模拟数据集构造

本节实验模拟生成的双月多源数据集是由 4 个分布相似但存在一定差异的 2 维双月数据集组成. 通过旋转不同角度并添加一定量的高斯噪声来模拟不同的源数据. 本文旋转角度分别取 $\theta = \{-5^\circ, 5^\circ, 10^\circ, 20^\circ, 30^\circ\}$ 以及噪声参数分别选择 $\text{noise}_\sigma = \{0.5, 2, 10\}$, 通过选择不同的参数组合构造了 6 个多领域数据集.

4.1.2 双月模拟数据集实验结果及分析

将 CDV-MSTL 方法分别与 LSSVM, 单源迁移 TrAdaBoost 和 TL-SVM 算法, 多源迁移学习方法 MultiSource-TrAdaBoost, CoCC 以及多领域适应算法 CCR 进行比较. 实验结果如表 1 所示: (1) 传统的 LSSVM 方法没有学习领域间的知识, 因而分类精度逊于其他 5 种迁移学习方法, 证明迁移学习提供知识支撑, 可以帮助提高目标领域的学习效率; 单源迁移学习算法 TrAdaBoost 和 TL-SSVM 的分类性能逊于 MultiSource-TrAdaBoost, CoCC, CCR 和 CDV-MSTL. 这进一步说明, 在充分学习各领域间的知识和信息后, 有利于提高算法的泛化性能. (2) 由表 1 可看出, 在噪音参数 noise_σ 和旋转角度 θ 的数值相对较小时, TrAdaBoost, TL-SVM, MT-SVM 和 CCR 方法分类精度随参数变化呈现下降趋势, 但 CDV-MSTL 方法的分类性能仍优越于其他 4 种迁移方法; 当 noise_σ 和 θ 的变化较大时, 由于领域间的相似性减小, 大多数迁移学习算法的性能都受到一定影响. TrAdaBoost 算法通过选择源领域相似的数据来促进算法性能的提高, 但是随着 noise_σ 和 θ 值增大, 领域间相似的程度逐渐减小, 适用于目标领域的源领域的的数据很少, 所以, TrAdaBoost 算法性能下降明显; CoCC 算法由于依赖前期协同聚类效果, 因而其分类精度显著受到数据集间的差异影响; CCR 算法要求不同任务之间的分布尽量相近, 但随着噪声的增大使得领域间差异变大进而辅助数据减少, 因而分类性能减小; MultiSource-TrAdaBoost 方法要求事先确定好领域间的权重, 再通过优化目标领域的决策方法, 无法实现全局上的最优要求, 因而分类精度受到一定的影响. CDV-MSTL 算法是通过约束共享的决策方向矢量来减小领域间的差异来实现不同领域间的学习, 虽然随着噪声增大领域间的相似性减弱, 但借助决策矢量的判别信息, 使得 CDV-MSTL 仍能保持较好的迁移学习性能.

4.2 图像数据集

4.2.1 数据构造与预处理

图像多源数据集^[23],通过不同父类下的子类交叉

构成.本文选择 Animal,Building,Bird 和 Traffic 等四个父类作为分类目标,每个父类下又选择了相近的四个子类构成源域数据和目标域数据,见图 1.

表 1 各种比较算法的分类精度(%)

Domain	Method						
	LSSVM	TrAdaBoost	TL-SSVM	CoCC	CCR	MultiSource-TrAdaBoost	CDV-MSTL
DS_1	81.25	85.14	86.48	87.59	91.25	90.18	92.58
DS_2	80.44	81.47	85.42	86.72	89.25	88.74	91.25
DS_3	79.56	83.14	82.15	85.92	88.53	86.92	87.14
DS_4	76.45	80.07	81.46	85.13	87.62	88.02	89.36
DS_5	79.36	82.16	81.32	83.78	85.47	84.36	82.51
DS_6	76.42	79.48	77.59	79.09	81.58	82.58	83.95



图1 图像数据示意图

4.2.2 图像数据集实验结果及分析

从表 2 中实验结果可以给出如下观察:对于不同父类的多领域数据集虽然局部有一定的变化,但是对于同一父类数据的主要信息在某些属性上还是具有很大相似性,通过提取这部分相似信息对提高领域间的学习能力将起到积极的作用;CDV-MSTL 算法在训练之前利用 PCA 将不同领域间的相似属性信息提取出来,再基于共同决策矢量在多个领域之间迁移学习,这样迁移学习就更加具有针对性,适应性和灵活性,从表 2 的实验结果可以发现 CDV-MSTL 算法比 TrAdaBoost, TL-

SSVM, CCR, MultiSource-TrAdaBoost 等方法具有更好的分类性能,说明了在考虑了共同决策矢量的基础上,通过强化多个相似领域分类器的方向来实现在相似特征上的迁移学习,更加符合实际场景,因而取得了相对较好的识别性能.

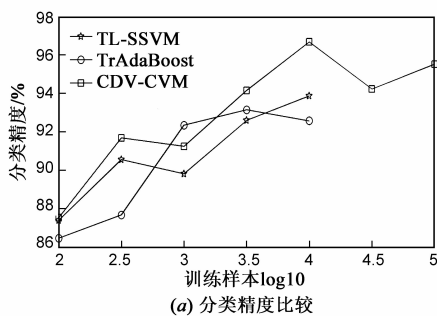
4.3 CDV-CVM 算法

本节实验从大样本的视角验证 CDV-CVM 算法的快速学习能力,选择更大规模的垃圾邮件数据集(Email Spam,ES)进行实验,由于每次计算初始值对核心集有一定的影响,而 CDV-CVM 算法的快速学习主要是在核心集上进行,因此我们这里选择平均分类精度作为算法的性能指标.见图 3.并得出如下观察:(1)通过实验我们发现,在多源迁移学习过程中,虽然多源数据量很大,但 CDV-CVM 算法实际叠代次数及运行时间低,并且运行速度明显优于其他的多源迁移算法,这也说明了 CDV-CVM 算法对大数据集的处理是非常有效的,适合大样本数据和高维数据运算.(2)当实验样本数目大于 10000 时 TrAdaBoost 和 TL-SSVM 算法已无法进行核矩阵的运算(内存溢出错误),而在相同条件下 CDV-

表 2 各种比较算法的分类精度(%)

Domain	Method				
	TrAdaBoost	TL-SSVM	MT-SVM	CCR	CDV-MSTL
Animal vs. Building	85.36	83.62	84.52	90.25	94.58
Animal vs. Traffic	81.57	82.54	86.59	89.48	90.26
Animal vs. Bird	78.59	80.25	84.57	88.16	87.59
Building vs. Bird	84.39	83.49	86.58	83.29	86.21
Building vs. Traffic	86.82	85.78	84.29	87.68	89.50
Bird vs. Traffic	82.69	84.29	87.13	89.18	90.28

CVM 算法仍然可在大规模数据条件下运行且分类精度更高. 表明了 CDV-CVM 算法即适用于样本较少的数据



又适用于大样本数据, 而图 2(b)则显示 CDV-CVM 时间复杂度和样本规模呈近似线性关系.

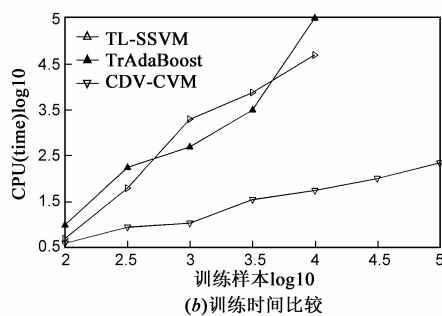


图2 三种迁移算法的分类精度和训练时间比较

5 结论

本文针对多源迁移学习算法在再生核空间核化复杂及计算开销大等问题, 提出了基于共同决策方向矢量的多源迁移学习方法. 相对于已有的方法, CDV-MSTL 方法可以将多个相似领域的决策矢量共同嵌入到支持向量机的训练过程中, 通过特征提取强化了源领域和目标领域间知识迁移, 提高了目标领域分类器的分类性能. 利用 CVM 理论, 将算法扩展为 CDV-CVM, 实现对大样本的快速训练和分类. 实验分析验证了所提方法在分类问题上的良好适应性. 虽然本文方法展现出了较好的有效性和稳定性, 其依然面临一些进一步需要探讨的问题. 例如: 如何进一步设计更好的兼顾不同特征的多源迁移学习策略将是一个非常有意义的工作, 我们将在未来的工作中作更深入地研究.

参考文献

- [1] Pan S J L, Kwok J T, Yang Q. Transfer learning via dimensionality reduction[A]. Proceedings of the 23rd International Conference on Artificial Intelligence [C]. California, USA, 2008. 677 – 682.
- [2] Pan S J L, Ni X. Cross-domain sentiment classification via spectral feature alignment [A]. Proceedings of the 19th International Conference on World Wide Web [C]. New York: ACM, 2010. 751 – 760.
- [3] Zhuang F Z, Luo P, Xiong H, Xiong Y H. Cross-domain learning from multiple sources: a consensus regularization perspective[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(12): 1664 – 1678.
- [4] Sun S. Multi-view Laplacian support vector machines[A]. Lecture Notes in Computer Science [C]. Germany: Springer, 2011. 209 – 222.
- [5] Xu Z, Sun S. Multi-view transfer learning with Adaboost[A]. Proceedings of the 23rd International Conference Tools with Artificial Intelligence [C]. New York, 2011. 399 – 402.
- [6] Ling X, Dai W, Xue G R, et al. Spectral domain-transfer learning [A]. Proceedings of the 14th International Conference on Knowledge Discovery and Data Mining [C]. USA: ACM, 2008. 488 – 496.
- [7] Gao J, Fan W, Sun Y, et al. Heterogeneous source consensus learning via decision propagation and negotiation [A]. Proceedings of the 15th International Conference on Knowledge Discovery and Data Mining [C]. USA: ACM, 2009. 339 – 348.
- [8] 蒋亦樟, 邓赵红, 王士同. ML 型迁移学习模糊系统 [J]. 自动化学报, 2012, 38(9): 1393 – 1409.
Jiang Yizhang, Deng Zhaohong, Wang Shitong. Mamdani-Larsen type transfer learning fuzzy system [J]. Acta Automatica Sinica, 2012, 38(9): 1393 – 1409. (in Chinese)
- [9] 胡文军, 王士同, 邓赵红. 适合大样本快速训练的最大夹角间隔核心集向量机 [J]. 电子学报, 39(5): 1178 – 1184.
Hu Wen-jun, WANG Shi-tong, DENG Zhao-hong. Maximum vector angular margin core vector machine suitable for fast training for large datasets [J]. Acta Electronica Sinica, 39(5): 1178 – 1184. (in Chinese)
- [10] Deng Z H, Fu-Lai Chung, Wang S T. FRSDE: fast reduced set density estimator using minimal enclosing ball approximation [J]. Pattern Recognition, 2008, (41): 1363 – 1372.
- [11] Tsang I, Kwok J, Cheung P. Core vector machines: fast SVM training on very large data sets [J]. Journal of Machine Learning Research, 2005, 6: 363 – 392.
- [12] Tax D M J, Duin R P W. Support vector domain description [J]. Pattern Recognition Letters, 1999, 20: 1191 – 1199.
- [13] Tax D M J, Duin R P W. Support vector data description [J]. Machine Learning, 2004, 54: 45 – 66.
- [14] 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.
Deng Nai-yang, Tian Ying-jie. New Method in Data Mining: Support Vector Machine [M]. Beijing: Science Press, 2004. (in Chinese).
- [15] Tsang I, Kwok J, Cheung P. Core vector machines: fast SVM training on very large data sets [J]. Journal of Machine Learning Research, 2005, 6: 363 – 392.

ing Research, 2005, 6: 363 – 392.

- [16] Tsang I, wok J. Generalized core vector machines[J]. IEEE Trans on Neural Networks, 2006, 17(5): 1126 – 1139.
- [17] Suykens J A K, Lukas L, et al. Least squares support vector machine classifiers: a large scale algorithm[A]. Proceedings of the European Conference on Circuit Theory and Design [C]. Stresa, Italy, 1999. 839 – 842.
- [18] Dai W, Yang Q, Xue G, et al. Boosting for transfer learning [A]. Proceedings of the 24th International Conference on Machine Learning [C]. New York: ACM, 2007. 193 – 200.
- [19] Yao Y, Doretto G. Boosting for transfer learning with multiple sources[A]. Proceedings of the 24th International Conference on Computer Vision and Pattern Recognition [C]. New York: ACM, 2010. 1855 – 1862.
- [20] Ablavsky V H, Becker C J. Transfer learning by sharing support vectors [OL]. <http://infoscience.epfl.ch/record/181360>.
- [21] Evgeniou T, Pontil M. Regularized multi-task learning[A]. Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining [C]. New York: ACM, 2004. 109 – 117.
- [22] Dai W, Xue G, Yang Q, et al. Co-clustering based classification for out-of-domain documents[A]. Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining [C]. New York: ACM, 2007: 210 – 219.
- [23] <http://wang.ist.psu.edu/docs/related.shtml>

作者简介



张景祥 男, 1977 年生于吉林通化. 2002 年、2007 年分别在重庆大学、江南大学获得理学学士、工学硕士学位, 2011 年江南大学攻读博士学位, 主要从事模式识别、人工智能和生物信息学等方面的研究.

E-mail: zjx145@163.com



王士同 男, 1964 年生于江苏扬州. 教授、博士生导师、中国计算机学会高级会员. 1984 年、1987 年在南京航空航天大学获得工学学士、硕士学位. 主要从事人工智能、模式识别、模糊系统、医学图像处理和生物信息学等方面的研究工作.