

DOI: 10.11992/tis.201505024

网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20151229.0837.016.html>

# 基于最小最大概率机的迁移学习分类算法

王晓初<sup>1,2</sup>, 包芳<sup>2,3</sup>, 王士同<sup>1</sup>, 许小龙<sup>1</sup>

(1. 江南大学 数字媒体学院, 江苏 无锡 214122; 2. 江阴职业技术学院 江苏省信息融合软件工程技术研发中心, 江苏 江阴 214405; 3. 江阴职业技术学院 计算机科学系, 江苏 江阴 214405)

**摘要:**传统的迁移学习分类算法利用源域中大量有标签的数据和目标域中少量有标签的数据解决相关但不相同目标域的数据分类问题,但对于已知源域的不同类别数据均值的迁移学习分类问题并不适用。为了解决这个问题,利用源域的数据均值和目标域的少量标记数据构造迁移学习约束项,对最小最大概率机进行正则化约束,提出了基于最小最大概率机的迁移学习分类算法,简称 TL-MPM。在 20 News Groups 数据集上的实验结果表明,目标域数据较少时,所提算法具有更高的分类正确率,从而说明了算法的有效性。

**关键词:**迁移学习;最小最大概率机;分类;源域;目标域;正则化

**中图分类号:**TP391.4 **文献标志码:**A **文章编号:**1673-4785(2016)01-0084-09

中文引用格式:王晓初,包芳,王士同,等.基于最小最大概率机的迁移学习分类算法[J].智能系统学报,2016,11(1):84-92.

英文引用格式:WANG Xiaochu, BAO Fang, WANG Shitong, et al. Transfer learning classification algorithms based on minimax probability machine[J]. CAAI Transactions on Intelligent Systems, 2016, 11(1): 84-92.

## Transfer learning classification algorithms based on minimax probability machine

WANG Xiaochu<sup>1,2</sup>, BAO Fang<sup>2,3</sup>, WANG Shitong<sup>1</sup>, XU Xiaolong<sup>1</sup>

(1. School of Digital Media, Jiangnan University, Wuxi 214122, China; 2. Information Fusion Software Engineering Research and Development Center of Jiangsu Province, Jiangyin Pdyteehnie College, Jiangyin 214405, China; 3. Department of Computer Science, Jiangyin Pdyteehnie College, Jiangyin 214405, China)

**Abstract:** Traditional transfer learning classification algorithms solve related (but not identical) data classification issues by using a large number of labeled samples in the source domain and small amounts of labeled samples in the target domain. However, this technique does not apply to the transfer learning of data from different categories of learned source domain data. To solve this problem, we constructed a transfer learning constraint term using the source domain data and the limited labeled data in the target domain to generate a regularized constraint for the minimax probability machine. We propose a transfer learning classification algorithm based on the minimax probability machine known as TL-MPM. Experimental results on 20 Newsgroups data sets demonstrate that the proposed algorithm has higher classification accuracy for small amounts of target domain data. Therefore, we confirm the effectiveness of the proposed algorithm.

**Keywords:** transfer learning; minimax probability machine; classification; source domain; target domain; regularization

迁移学习是机器学习领域的一个新方向,是对机器学习能力的一次拓展,具有很重要的研究价值。文献[1]讨论了迁移学习的应用,现阶段研究进展

及对未来迁移学习潜在的问题。文献[2]推广了传统的 AdaBoost 算法,提出 Tradaboosting 迁移学习算法,该算法思想是利用 boosting 的技术来过滤掉辅助数据中那些与源训练数据最不相像的数据。其中 boosting 的作用是建立一种自动调整权重的机制,于

收稿日期:2015-05-11. 网络出版日期:2015-12-29.

基金项目:国家自然科学基金资助项目(61170122, 61272210).

通信作者:王晓初. E-mail: [icnice@yeah.net](mailto:icnice@yeah.net).

是重要的辅助训练数据的权重将会增加,不重要的辅助训练数据的权重将会减小。调整权重之后,这些带权重的辅助训练数据将会作为额外的训练数据,与源训练数据一起来提高分类模型的可靠度。文献[3]通过维数约简的思想,设法学习出一个低维隐特征空间,使得源域的数据和目标域的数据在该空间的分布相同或接近对方。文献[4-7]将协同学习、正则化、模糊理论等用于迁移学习,取得了很好的效果。文献[8]同样提出了基于正则化的迁移学习模型,文献[9]提出了图协同正则化的迁移学习。文献[10]将迁移学习的思想用于广告显示。

前面对迁移学习的研究,大多是在源域的数据样本具体标签已知的假设下进行的。本文针对源域中不同类别的数据均值和目标域中少量有标签的数据已知的分类问题,以最小最大概率机分类算法模型为基础,提出 TL-MPM 算法,本算法最大的优点是并不知道源域的数据训练样本的具体标签,仅知道均值信息,从而减小了对标记源域的数据所需花费的代价。

## 1 最小最大概率机

### 1.1 线性部分

假设  $\mathbf{x}$  和  $\mathbf{y}$  为 2 维分类问题的 2 个随机变量,服从某种分布  $\mathbf{x} \sim (\bar{\mathbf{x}}, \Sigma_{\mathbf{x}})$ ,  $\mathbf{y} \sim (\bar{\mathbf{y}}, \Sigma_{\mathbf{y}})$ , 这里  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ ,  $\Sigma_{\mathbf{x}}, \Sigma_{\mathbf{y}} \in \mathbb{R}^{d \times d}$ ,  $\bar{\mathbf{x}}, \bar{\mathbf{y}}, \Sigma_{\mathbf{x}}, \Sigma_{\mathbf{y}}$  分别为随机变量  $\mathbf{x}$  和  $\mathbf{y}$  的均值和方差。最小最大概率机实现分类的目的就是找到一个超平面  $\mathbf{a}^T \mathbf{z} - b = 0$ , ( $\mathbf{a}, \mathbf{z} \in \mathbb{R}^d$ ,  $b \in \mathbb{R}$ )<sup>[11]</sup>, 将 2 类样本在样本估计的均值和方差的前提下,按照最大概率分离,所以分类问题可以描述为

$$\begin{aligned} \max_{\alpha, a, b} \quad & \alpha \text{ s.t. } \inf \Pr\{\mathbf{a}^T \mathbf{x} \geq b\} \geq \alpha \\ & \inf \Pr\{\mathbf{a}^T \mathbf{y} \leq b\} \geq \alpha \end{aligned} \quad (1)$$

式中:  $\alpha$  为正确分类概率,  $\inf$  表示下确界。主要是保证在错分概率最小的情况下,通过一个核函数将特征映射到高维空间,并构造一个最优分类超平面,从而实现特征分类。(1)式可归结为一个带约束的优化问题<sup>[12-13]</sup>:

$$\begin{aligned} \max_{k, a, b} \quad & k \text{ s.t. } -b + \mathbf{a}^T \bar{\mathbf{x}} \geq k \sqrt{\mathbf{a}^T \Sigma_{\mathbf{x}} \mathbf{a}} \\ & b - \mathbf{a}^T \bar{\mathbf{y}} \geq k \sqrt{\mathbf{a}^T \Sigma_{\mathbf{y}} \mathbf{a}} \end{aligned} \quad (2)$$

式中:  $k = \sqrt{\alpha / (1 - \alpha)}$ <sup>[14]</sup>。化简式(2)并削去  $k$  且不失一般性地令  $\mathbf{a}^T (\bar{\mathbf{x}} - \bar{\mathbf{y}}) = 1$ , 最大化  $k$  等价于:

$$\min_a \sqrt{\mathbf{a}^T \Sigma_{\mathbf{x}} \mathbf{a}} + \sqrt{\mathbf{a}^T \Sigma_{\mathbf{y}} \mathbf{a}} \text{ s.t. } \mathbf{a}^T (\bar{\mathbf{x}} - \bar{\mathbf{y}}) = 1 \quad (3)$$

目标函数(3)为凸优化问题<sup>[15]</sup>, 假设求得最优

解为  $\mathbf{a}_*$ , 对应  $k$  取得最小值  $k_*$ , 此时(2)式中的不等号变为等号, 可得

$$b_* = \mathbf{a}_*^T \bar{\mathbf{x}} - k_* \sqrt{\mathbf{a}_*^T \Sigma_{\mathbf{x}} \mathbf{a}_*} = \mathbf{a}_*^T \bar{\mathbf{y}} - k_* \sqrt{\mathbf{a}_*^T \Sigma_{\mathbf{y}} \mathbf{a}_*} \quad (4)$$

对于测试样本  $\mathbf{x}^{\text{new}}$ , 若  $\mathbf{a}_*^T \mathbf{x}^{\text{new}} - b_* \geq 0$  则  $\mathbf{x}^{\text{new}}$  为正样本; 反之  $\mathbf{x}^{\text{new}}$  为负样本。

### 1.2 非线性部分

样本映射到非线性:  $\mathbf{x} \mapsto \boldsymbol{\varphi}(\mathbf{x}) \sim (\bar{\boldsymbol{\varphi}}(\mathbf{x}), \Sigma_{\boldsymbol{\varphi}(\mathbf{x})})$   
 $\mathbf{y} \mapsto \boldsymbol{\varphi}(\mathbf{y}) \sim (\bar{\boldsymbol{\varphi}}(\mathbf{y}), \Sigma_{\boldsymbol{\varphi}(\mathbf{y})})$ , 这里  $\boldsymbol{\varphi}: \mathbb{R}^d \mapsto \mathbb{R}^n$ , 核函数可写为  $K(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x}_j)$ 。核函数用于最小最大概率机, 决策超平面写为  $\mathbf{a}^T \boldsymbol{\varphi}(\mathbf{z}) - b = 0$ , 其中  $\mathbf{a}, \boldsymbol{\varphi}(\mathbf{z}) \in \mathbb{R}^n$ 。表达式(3)非线性形式为

$$\begin{aligned} \min_a \quad & \sqrt{\mathbf{a}^T \Sigma_{\boldsymbol{\varphi}(\mathbf{x})} \mathbf{a}} + \sqrt{\mathbf{a}^T \Sigma_{\boldsymbol{\varphi}(\mathbf{y})} \mathbf{a}} \\ \text{s.t.} \quad & \mathbf{a}^T (\boldsymbol{\varphi}(\bar{\mathbf{x}}) - \boldsymbol{\varphi}(\bar{\mathbf{y}})) = 1 \end{aligned} \quad (5)$$

如果  $\mathbf{a}$  中有一分量正交于  $\boldsymbol{\varphi}(\mathbf{x}_i)$ ,  $i = 1, 2, \dots, N_x$  和  $\boldsymbol{\varphi}(\mathbf{y}_i)$ ,  $i = 1, 2, \dots, N_y$  的子空间,  $N_x, N_y$  分别为正负标记样本的个数, 那么这个分量不会影响到目标函数的结果, 因此可以将  $\mathbf{a}$  写成

$$\mathbf{a} = \sum_{i=1}^{N_x} \alpha_i \boldsymbol{\varphi}(\mathbf{x}_i) + \sum_{i=1}^{N_y} \beta_i \boldsymbol{\varphi}(\mathbf{x}_i) \quad (6)$$

设  $\boldsymbol{\gamma} = [\alpha_1 \alpha_2 \dots \alpha_{N_x} \beta_1 \beta_2 \dots \beta_{N_y}]$ ,  $[K_x]_i = K(\mathbf{x}_j, \mathbf{z}_i)$  为核函数,  $[\tilde{\mathbf{k}}_x]_i = \frac{1}{N_x} \sum_{j=1}^{N_x} K(\mathbf{x}_j, \mathbf{z}_i)$ ,  $[\tilde{\mathbf{k}}_y]_i = \frac{1}{N_y} \sum_{j=1}^{N_y} K(\mathbf{y}_j, \mathbf{z}_i)$  则约束条件可表示为

$$\mathbf{a}^T \boldsymbol{\varphi}(\mathbf{x}) - \mathbf{a}^T \boldsymbol{\varphi}(\mathbf{y}) = \boldsymbol{\gamma}^T \tilde{\mathbf{k}}_x - \boldsymbol{\gamma}^T \tilde{\mathbf{k}}_y \quad (7)$$

又假设  $\mathbf{I}_m$  为  $m$  维全 1 列向量, 且  $\tilde{\mathbf{K}}_x = \mathbf{K}_x - \mathbf{I}_{N_x} \tilde{\mathbf{k}}_x^T$ , 同样形式  $\tilde{\mathbf{K}}_y = \mathbf{K}_y - \mathbf{I}_{N_y} \tilde{\mathbf{k}}_y^T$ , 其中  $i = 1, 2, \dots, N_x$  时  $\mathbf{z}_i = \mathbf{x}_i$ ,  $i = N_x + 1, \dots, N_x + N_y$  时  $\mathbf{z}_i = \mathbf{y}_{i-N_x}$ , 那么正负样本的协方差表达式可表示为

$$\begin{aligned} \mathbf{a}^T \Sigma_{\boldsymbol{\varphi}(\mathbf{x})} \mathbf{a} &= \frac{1}{N_x} \boldsymbol{\gamma}^T \tilde{\mathbf{K}}_x^T \tilde{\mathbf{K}}_x \boldsymbol{\gamma} \\ \mathbf{a}^T \Sigma_{\boldsymbol{\varphi}(\mathbf{y})} \mathbf{a} &= \frac{1}{N_y} \boldsymbol{\gamma}^T \tilde{\mathbf{K}}_y^T \tilde{\mathbf{K}}_y \boldsymbol{\gamma} \end{aligned} \quad (8)$$

经过上面核化后, 约束条件和协方差的表达式变为(7)式和(8)式的形式, 带入  $\mathbf{a}$  表示的非线性表达式(5)式可得  $\boldsymbol{\gamma}$  表示的非线性目标函数:

$$\begin{aligned} \min_{\boldsymbol{\gamma}} \quad & \sqrt{\frac{1}{N_x} \boldsymbol{\gamma}^T \tilde{\mathbf{K}}_x^T \tilde{\mathbf{K}}_x \boldsymbol{\gamma}} + \sqrt{\frac{1}{N_y} \boldsymbol{\gamma}^T \tilde{\mathbf{K}}_y^T \tilde{\mathbf{K}}_y \boldsymbol{\gamma}} \\ \text{s.t.} \quad & \boldsymbol{\gamma}^T (\tilde{\mathbf{k}}_x - \tilde{\mathbf{k}}_y) = 1 \end{aligned} \quad (9)$$

目标函数(9)也为凸优化问题。

显然, 与线性部分算法一致。当  $\boldsymbol{\gamma}$  取最优解  $\boldsymbol{\gamma}_*$ , 对应  $k$  取得最小值  $k_*$ , 可得

$$b_* = \gamma_*^T \tilde{k}_x - k_* \sqrt{\frac{1}{N_x} \gamma_*^T \tilde{K}_x \tilde{K}_x \gamma_*} = \gamma_*^T \tilde{k}_y - k_* \sqrt{\frac{1}{N_y} \gamma_*^T \tilde{K}_y \tilde{K}_y \gamma_*} \quad (10)$$

对于测试样本  $\mathbf{x}^{\text{new}}$ , 若  $\mathbf{a}_*^T \boldsymbol{\varphi}(\mathbf{x}^{\text{new}}) - b_* = \sum_{i=1}^{N_x+N_y} [\gamma_*]_i K(\mathbf{z}_i, \mathbf{x}^{\text{new}}) \geq 0$ , 则  $\mathbf{x}^{\text{new}}$  为正样本; 反之  $\mathbf{x}^{\text{new}}$  为负样本。

## 2 基于最小最大概率机的迁移学习分类算法

### 2.1 TL-MPM 算法的应用背景

实际问题中, 有时候当前分类目标的标记样本并不充分, 这样对于分类结果的预测会带来很大偏差。而当前分类目标往往和上一阶段的分类目标比较, 既有新的变化, 又具有某些类似的特性。比如超市不同季度的营业额、银行不同月份的贷款数、公司每个阶段的出货量、港务每年吞吐量等, 一般情况每个阶段虽然有波动, 不会有太大的偏差, 但由于种种原因, 上个阶段的样本信息, 并不能完全得知; 即使得知, 也不能直接将所有信息用在当前阶段中, 因为上一阶段的均值往往是易于得知, 并且各个阶段数据均值波动基本很小。所以上一阶段均值是对当前阶段极为有用的信息。这样在当前阶段信息不充分的情况下可以充分利用上个阶段的均值信息进行迁移学习。

### 2.2 TL-MPM 算法的理论依据

在最小最大概率机训练的过程中, 如式(3)所描述的, 通过最小化  $\sqrt{\mathbf{a}^T \boldsymbol{\Sigma}_x \mathbf{a}}$  和  $\sqrt{\mathbf{a}^T \boldsymbol{\Sigma}_y \mathbf{a}}$  之和, 使得样本在超平面所在空间分布为线性可分的形式, 但在样本很少的情况下, 协方差和均值并不能很好地代表整体的协方差和均值, 那么训练出来的  $\mathbf{a}$  值实际上并不是最优的。对于相关但不相同的数据, 源域的数据和目标域的数据具有相似性, 所以均值波动性并不大, 可以预测目标域的数据均值与源域的数据均值或多或少是相近的, 所以在训练过程中, 可以利用源域的数据均值和目标域的数据均值实现数据迁移学习。在具体的实现中, 是通过最小化源域的数据均值和目标域的数据均值在超平面所在空间的欧氏距离来实现的。假设源域的数据正负类样本的均值分别为  $\bar{\mathbf{x}}_t$  和  $\bar{\mathbf{y}}_t$ , 目标域少量数据的样本的均值分别为  $\bar{\mathbf{x}}_t$  和  $\bar{\mathbf{y}}_t$ , 可得超平面所在空间的均值之间的线性距离可以表示为

$$L = \|f(\bar{\mathbf{x}}_t) - f(\bar{\mathbf{x}}_s)\|^2 + \|f(\bar{\mathbf{y}}_t) - f(\bar{\mathbf{y}}_s)\|^2 \quad (11)$$

均值之间的非线性距离表示为

$$L_k = \|f(\boldsymbol{\varphi}(\bar{\mathbf{x}}_t)) - f(\boldsymbol{\varphi}(\bar{\mathbf{x}}_s))\|^2 + \|f(\boldsymbol{\varphi}(\bar{\mathbf{y}}_t)) - f(\boldsymbol{\varphi}(\bar{\mathbf{y}}_s))\|^2 \quad (12)$$

TL-MPM 算法的理论依据是: 若两个领域相关, 源域的数据和目标域的数据在超平面所在空间的均值应相近。通过在 MPM 线性目标式(3)中增加  $\lambda L$ , 非线性目标式(5)中增加  $\lambda L_k$  实现两个域之间的迁移学习。加入迁移学习项后的线性目标函数可以写为

$$\min_a \sqrt{\mathbf{a}^T \boldsymbol{\Sigma}_x \mathbf{a}} + \sqrt{\mathbf{a}^T \boldsymbol{\Sigma}_y \mathbf{a}} + \lambda L \quad \text{s.t.} \quad \mathbf{a}^T (\bar{\mathbf{x}}_t - \bar{\mathbf{y}}_t) = 1 \quad (13)$$

非线性目标函数可写为

$$\min_a \sqrt{\mathbf{a}^T \boldsymbol{\Sigma}_{\varphi(x)} \mathbf{a}} + \sqrt{\mathbf{a}^T \boldsymbol{\Sigma}_{\varphi(y)} \mathbf{a}} + \lambda L_k \quad \text{s.t.} \quad \mathbf{a}^T (\boldsymbol{\varphi}(\bar{\mathbf{a}}_t) - \boldsymbol{\varphi}(\bar{\mathbf{y}}_t)) = 1 \quad (14)$$

式中:  $L$  和  $L_k$  表示两个域的差异程度, 参数  $\lambda$  控制惩罚程度。本算法对  $\lambda$  在  $10^{-3} \sim 10^3$  范围内进行交叉验证选优。当  $\lambda$  取值较小时, 说明两个领域之间均值相关性较高; 反之, 两个领域之间均值相关性较小。TL-MPM 算法的原理可用图 1 表示。

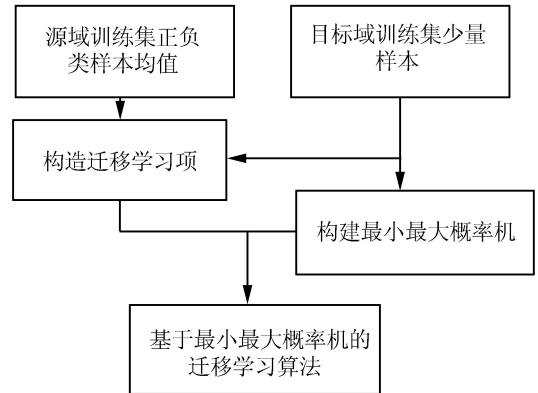


图 1 TL-MPM 算法

Fig.1 Flowchart of TL-MPM

为了更清晰地展示 TL-MPM 算法的作用原理及效果, 在二维数据上做了线性算法的模拟实验。图 2 为源域样本分布, 横轴  $x$  表示样本的一维, 纵轴  $y$  表示样本的二维, 图中“ $\times$ ”表示目标域正样本, “+”号表示目标域负样本, 菱形和正方形分别表示已知的目标域正样本和负样本, 实心圆表示目标域正样本和负样本均值, 空心圆表示源域正样本和负样本均值。虚线为目标域少量标记样本在 MPM 算法下得到的分类超平面, 实线为 TL-MPM 算法下得到的分类超平面。

由图 3 可以明显看出, 源域正负样本的均值和目标域样本的均值相差并不是很大。正是由于这种源域样本与目标域样本的相关性, 经过源域的数据均值与目标域的数据均值的迁移学习, 得到的目标域的分类超平面更为准确。

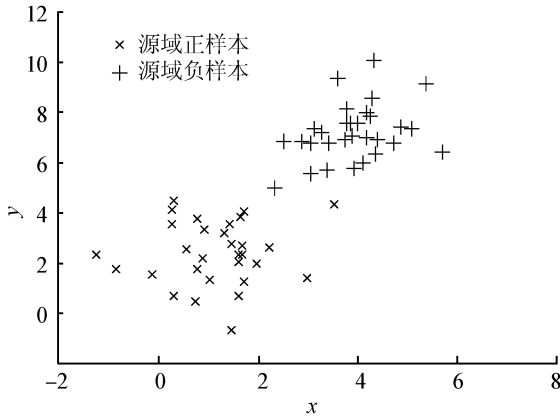


图2 源域样本分布

Fig.2 Source domain samples

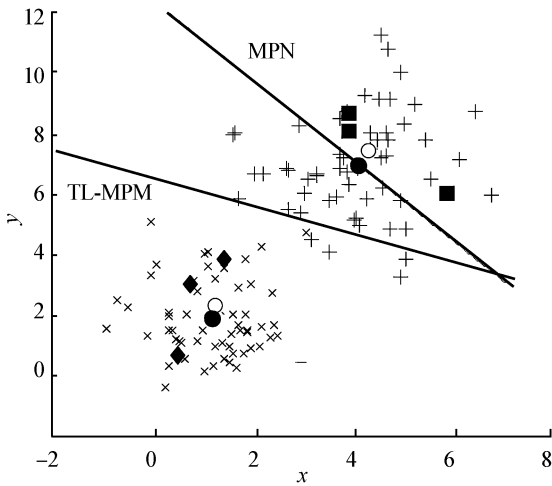


图3 二维数据模拟实验

Fig.3 Two-dimensional simulation experiment

## 2.3 具体推导过程

### 2.3.1 线性部分

由于分类超平面为  $\mathbf{a}^T \mathbf{x} - b = 0$ , 那么样本在超平面所在空间的分布可以写为  $f(\mathbf{x}) = \mathbf{a}^T \mathbf{x} - b$ , 所以对线性迁移学习约束项中的前半部分可推导为

$$\begin{aligned} \|f(\bar{\mathbf{x}}_t) - f(\bar{\mathbf{x}}_s)\|^2 &= \|\mathbf{a}^T \bar{\mathbf{x}}_t - \mathbf{a}^T \bar{\mathbf{x}}_s\|^2 = \\ &= \mathbf{a}^T (\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_s) (\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_s)^T \mathbf{a} = \\ &= \mathbf{a}^T (\bar{\mathbf{x}}_t \bar{\mathbf{x}}_t^T + \bar{\mathbf{x}}_s \bar{\mathbf{x}}_s^T - \bar{\mathbf{x}}_t \bar{\mathbf{x}}_s^T - \bar{\mathbf{x}}_s \bar{\mathbf{x}}_t^T) \mathbf{a} = \mathbf{a}^T \mathbf{D}_x \mathbf{a} \end{aligned} \quad (15)$$

线性迁移学习约束项中的后半部分可推导为

$$\begin{aligned} \|f(\bar{\mathbf{y}}_t) - f(\bar{\mathbf{y}}_s)\|^2 &= \|\mathbf{a}^T \bar{\mathbf{y}}_t - \mathbf{a}^T \bar{\mathbf{y}}_s\|^2 = \\ &= \mathbf{a}^T (\bar{\mathbf{y}}_t - \bar{\mathbf{y}}_s) (\bar{\mathbf{y}}_t - \bar{\mathbf{y}}_s)^T \mathbf{a} = \\ &= \mathbf{a}^T (\bar{\mathbf{y}}_t \bar{\mathbf{y}}_t^T + \bar{\mathbf{y}}_s \bar{\mathbf{y}}_s^T - \bar{\mathbf{y}}_t \bar{\mathbf{y}}_s^T - \bar{\mathbf{y}}_s \bar{\mathbf{y}}_t^T) \mathbf{a} = \mathbf{a}^T \mathbf{D}_y \mathbf{a} \end{aligned} \quad (16)$$

那么, 基于最小最大概率机的迁移学习算法对应的

线性迁移学习约束项有简单的形式:

$$L = \mathbf{a}^T \mathbf{D}_x \mathbf{a} + \mathbf{a}^T \mathbf{D}_y \mathbf{a} = \mathbf{a}^T \mathbf{D} \mathbf{a}$$

其中:

$$\begin{aligned} \mathbf{D}_x &= \bar{\mathbf{x}}_t \bar{\mathbf{x}}_t^T + \bar{\mathbf{x}}_s \bar{\mathbf{x}}_s^T - \bar{\mathbf{x}}_t \bar{\mathbf{x}}_s^T - \bar{\mathbf{x}}_s \bar{\mathbf{x}}_t^T \\ \mathbf{D}_y &= \bar{\mathbf{y}}_t \bar{\mathbf{y}}_t^T + \bar{\mathbf{y}}_s \bar{\mathbf{y}}_s^T - \bar{\mathbf{y}}_t \bar{\mathbf{y}}_s^T - \bar{\mathbf{y}}_s \bar{\mathbf{y}}_t^T \\ \mathbf{D} &= \mathbf{D}_x + \mathbf{D}_y \end{aligned}$$

最小化  $\mathbf{a}^T \mathbf{D} \mathbf{a}$  等价于最小化  $\sqrt{\mathbf{a}^T \mathbf{D} \mathbf{a}}$ , 为了便于运算, 在不影响结果的情况下, 所以基于最小最大概率机的迁移学习算法可以化简成下面形式的目标函数:

$$\begin{aligned} \min_a & \sqrt{\mathbf{a}^T \Sigma_x \mathbf{a}} + \sqrt{\mathbf{a}^T \Sigma_y \mathbf{a}} + \lambda \sqrt{\mathbf{a}^T \mathbf{D} \mathbf{a}} \\ \text{s.t. } & \mathbf{a}^T (\bar{\mathbf{x}}_t - \bar{\mathbf{y}}_t) = 1 \end{aligned}$$

具体求解过程见本文第4部分。由于加入了约束项,  $b_*$  的表达发生了变化。注意到引入项对正负类的贡献不同, 用  $\eta$  ( $0 \leq \eta \leq 1$ ) 表示正类在整个样本中所占的比率。一般情况下  $\eta$  可设置为 0.5, 那么判别式可由下面公式推导求出:

$$\begin{aligned} \max_{k, a, b} k \text{ s.t. } & -b + \mathbf{a}^T \bar{\mathbf{x}}_t \geq k(\sqrt{\mathbf{a}^T \Sigma_x \mathbf{a}} + \eta \sqrt{\lambda \mathbf{a}^T \mathbf{D} \mathbf{a}}) \\ & b - \mathbf{a}^T \bar{\mathbf{y}}_t \geq k(\sqrt{\mathbf{a}^T \Sigma_y \mathbf{a}} + (1 - \eta) \sqrt{\lambda \mathbf{a}^T \mathbf{D} \mathbf{a}}) \end{aligned} \quad (17)$$

当  $\mathbf{a}$  取最优解  $\mathbf{a}_*$  时, 对应  $k$  取得最小值  $k_*$ , 式(17)中不等式变为等式, 可得

$$\begin{aligned} b_* &= \mathbf{a}_*^T \bar{\mathbf{x}}_t - k_*(\sqrt{\mathbf{a}_*^T \Sigma_x \mathbf{a}_*} + \eta \sqrt{\lambda \mathbf{a}_*^T \mathbf{D} \mathbf{a}_*}) = \\ &= \mathbf{a}_*^T \bar{\mathbf{y}}_t + k_*(\sqrt{\mathbf{a}_*^T \Sigma_y \mathbf{a}_*} + (1 - \eta) \sqrt{\lambda \mathbf{a}_*^T \mathbf{D} \mathbf{a}_*}) \end{aligned}$$

对于测试样本  $\mathbf{x}^{\text{new}}$ , 若  $\mathbf{a}_*^T \mathbf{x}^{\text{new}} - b_* \geq 0$ , 则  $\mathbf{x}^{\text{new}}$  为正样本; 反之  $\mathbf{x}^{\text{new}}$  为负样本。

### 2.3.2 非线性部分

核表示方式是将数据映射到高维空间来增加线性学习器的学习能力。通过选择恰当的核函数来替代内积, 可以隐式地将训练数据非线性映射到高维空间。

对应于线性分布,  $f(\varphi(\mathbf{x})) = \mathbf{a}^T \varphi(\mathbf{x}) - b$  为样本在非线性超平面所在空间的分布, 且  $\mathbf{a}^T \varphi(\mathbf{x}) = \gamma^T \tilde{\mathbf{k}}_x$ , 非线性迁移学习约束项的前半部分可以推导为

$$\begin{aligned} \|f(\varphi(\bar{\mathbf{x}}_t)) - f(\varphi(\bar{\mathbf{x}}_s))\|^2 &= \|\gamma^T \tilde{\mathbf{k}}_{\bar{\mathbf{x}}_t} - \gamma^T \tilde{\mathbf{k}}_{\bar{\mathbf{x}}_s}\|^2 = \\ &= \gamma^T (\tilde{\mathbf{k}}_{\bar{\mathbf{x}}_t}^T - \tilde{\mathbf{k}}_{\bar{\mathbf{x}}_s}^T)^T (\tilde{\mathbf{k}}_{\bar{\mathbf{x}}_t} - \tilde{\mathbf{k}}_{\bar{\mathbf{x}}_s}) \gamma = \\ &= \gamma^T (\tilde{\mathbf{k}}_{\bar{\mathbf{x}}_t}^T \tilde{\mathbf{k}}_{\bar{\mathbf{x}}_t} + \tilde{\mathbf{k}}_{\bar{\mathbf{x}}_s}^T \tilde{\mathbf{k}}_{\bar{\mathbf{x}}_s} - \tilde{\mathbf{k}}_{\bar{\mathbf{x}}_t}^T \tilde{\mathbf{k}}_{\bar{\mathbf{x}}_s} - \tilde{\mathbf{k}}_{\bar{\mathbf{x}}_s}^T \tilde{\mathbf{k}}_{\bar{\mathbf{x}}_t}) \gamma = \\ &= \gamma^T \mathbf{H}_x \gamma \end{aligned}$$



非线性迁移学习约束项的后半部分可推导为

$$\begin{aligned} & \|f(\varphi(\bar{y})) - f(\varphi(\bar{y}_l))\|^2 = \\ & \|\gamma^T \tilde{k}_{\bar{y}_l}^T - \gamma^T \tilde{k}_{\bar{y}_s}^T\|^2 = \\ & \gamma^T (\tilde{k}_{\bar{y}_l}^T - \tilde{k}_{\bar{y}_s}^T) (\tilde{k}_{\bar{y}_l}^T - \tilde{k}_{\bar{y}_s}^T)^T \gamma = \\ & \gamma^T (\tilde{k}_{\bar{y}_l}^T \tilde{k}_{\bar{y}_l}^T + \tilde{k}_{\bar{y}_s}^T \tilde{k}_{\bar{y}_s}^T - \tilde{k}_{\bar{y}_l}^T \tilde{k}_{\bar{y}_s}^T - \tilde{k}_{\bar{y}_s}^T \tilde{k}_{\bar{y}_l}^T) \gamma = \\ & \gamma^T H_y \gamma \end{aligned}$$

那么,基于最小最大概率机的迁移学习算法对应的迁移学习线性约束项有简单的形式:

$$L_k = \gamma^T H_x \gamma + \gamma^T H_y \gamma = \gamma^T H \gamma$$

其中:

$$\begin{aligned} H_x &= \tilde{k}_{\bar{x}_l}^T \tilde{k}_{\bar{x}_l}^T + \tilde{k}_{\bar{x}_s}^T \tilde{k}_{\bar{x}_s}^T - \tilde{k}_{\bar{x}_l}^T \tilde{k}_{\bar{x}_s}^T - \tilde{k}_{\bar{x}_s}^T \tilde{k}_{\bar{x}_l}^T \\ H_y &= \tilde{k}_{\bar{y}_l}^T \tilde{k}_{\bar{y}_l}^T + \tilde{k}_{\bar{y}_s}^T \tilde{k}_{\bar{y}_s}^T - \tilde{k}_{\bar{y}_l}^T \tilde{k}_{\bar{y}_s}^T - \tilde{k}_{\bar{y}_s}^T \tilde{k}_{\bar{y}_l}^T \\ H &= H_x + H_y \end{aligned}$$

最小化  $\gamma^T H \gamma$  等价于最小化  $\sqrt{\gamma^T H \gamma}$ , 为了便于运算,在不影响结果的情况下,基于最小最大概率机的迁移学习算法可以化简成下面形式的目标函数:

$$\begin{aligned} \min_{\gamma} & \sqrt{\frac{1}{N_x} \gamma^T \tilde{K}_x^T \tilde{K}_x \gamma} + \sqrt{\frac{1}{N_y} \gamma^T \tilde{K}_y^T \tilde{K}_y \gamma} + \lambda \sqrt{\gamma^T H \gamma} \\ \text{s.t.} & \gamma^T (\tilde{k}_{\bar{x}_l} - \tilde{k}_{\bar{y}_s}) = 1 \end{aligned}$$

显然非线性目标函数求解方法与线性一致。当  $\gamma$  取最优解  $\gamma_*$  时,对应  $k$  取得最小值  $k_*$ , 求取  $b_*$  推理与线性部分类似,得到  $b_*$  的表达式为

$$\begin{aligned} b_* &= \gamma_*^T \tilde{k}_{\bar{x}_l} - k_* \left( \sqrt{\gamma_*^T \frac{1}{N_x} \tilde{K}_x^T \tilde{K}_x \gamma_*} + \eta \lambda \sqrt{\gamma_*^T H_k \gamma_*} \right) = \\ & \gamma_*^T \tilde{k}_{\bar{y}_l} + k_* \left( \sqrt{\gamma_*^T \frac{1}{N_y} \tilde{K}_y^T \tilde{K}_y \gamma_*} + (1 - \eta) \lambda \sqrt{\gamma_*^T H_k \gamma_*} \right) \end{aligned}$$

对于测试样本  $x^{\text{new}}$ , 若:  $a_*^T \varphi(x^{\text{new}}) - b_* = \sum_i^{N_l+N_u} [\gamma_*]_i K(z_i, x^{\text{new}}) \geq 0$ , 则  $x^{\text{new}}$  为正样本; 反之  $x^{\text{new}}$  为负样本。

### 3 TL-MPM 算法解法流程

线性部分和非线性部分目标函数形式相同,都是非线性有约束优化问题,并且是凸优化问题。有好多方法可以选择<sup>[16-18]</sup>,比如拟牛顿法, Rosen 梯度投影法等,这里我们用递归最小二乘法对修正优化问题进行求解,下面只解析线性目标函数求解过程。

1) 令  $a = a_0 + Fv$ ,  $F$  的列正交于  $\bar{x}_l - \bar{y}_l$ ;

2) 式(16)可写为

$$\begin{aligned} \min_{\beta, \delta, \varepsilon, v} & \beta_k + \frac{1}{\beta_k} \|\Sigma_x^{1/2} (a_k + Fv_k)\|_2^2 + \delta_k + \\ & \frac{1}{\delta_k} \|\Sigma_y^{1/2} (a_k + Fv_k)\|_2^2 + \varepsilon_k + \frac{\lambda}{\varepsilon_k} \|D^{1/2} (a_k + Fv_k)\|_2^2 \end{aligned} \quad (18)$$

3) 令  $a_0 = (\bar{x}_l - \bar{y}_l) / \|\bar{x}_l - \bar{y}_l\|_2$ ,  $\beta_0 = 1$ ,  $\delta_0 = 1$ ,  $\varepsilon_0 = 1$  并带入(18)得到最小二乘问题,求解得  $v_0$ ;

4)  $k = 1$ ;

5)  $a_k = a_{k-1} + Fv_{k-1}$ ,  $\beta_k = \sqrt{a_{k-1}^T \Sigma_x a_{k-1}}$ ,  $\delta_k = \lambda \sqrt{a_{k-1}^T \Sigma_y a_{k-1}}$ ,  $\varepsilon_k = \sqrt{a_{k-1}^T D a_{k-1}}$  并带入(18)式求解得  $v_k$ ;

6)  $k = k + 1$ ;

7) 重复步骤 5、6 直到收敛或满足停止条件;

8) 最后求得:  $a_* = a_k$ ,  $k_* = 1/(\beta_k + \delta_k + \varepsilon_k)$ ,  $b_* = a_*^T \bar{x} - k_* (\beta_k + \eta \varepsilon_k)$ 。

## 4 实验结果与分析

为了验证 TL-MPM 算法的有效性,在常用的真实数据集: 20 News Groups 数据集上进行实验。20 News Groups 数据集的信息如表 1 所示。

表 1 20 News Groups 数据集  
Table 1 20 News Groups Data

大类	小类	样本个数
comp	comp.graphics	997
	comp.windows.x	998
	comp.os.mswindows.misc	992
	comp.sys.ibm.pc.hardware	997
	comp.sys.mac.hardware	996
rec	rec.motorcycles	997
	rec.autos	998
	rec.sport.baseball	998
	rec.sport.hockey	998
	talk.politics.mideast	1 000
talk	talk.politics.misc	998
	talk.politics.guns	1 000
	talk.religion.misc	999
sci	sci.crypt	998
	sci.med	998
	sci.space	999
	sci.electronics	999

为了方便实验的描述,对实验中涉及的有关符号给出相应的定义,如表 2 所示。

表 2 实验中涉及的符号定义

Table 2 Definition of symbols in the experiment	
符号	定义
Ds	源域数据集
Dt	目标域数据集
T-LMPM	仅利用目标域少量数据训练得到的目标域 MPM 线性分类器
T-KMPM	仅利用目标域少量数据训练得到的目标域 MPM 非线性分类器
LTL-MPM	本文提出的 TL-MPM 算法得到的线性分类器
KTL-MPM	本文提出的 TL-MPM 算法得到的非线性分类器

4.1 20News Groups 数据集预处理

首先对数据集分成源域训练数据和目标域训练数据,具体的处理结果如表 3 所示。

表 3 20News Groups 数据集处理结果

Table 3 Results of 20News Groups Data set processing		
Datasets	Ds	Dt
comp vs rec	comp.graphics	comp.windows.x
	rec.motorcycles	rec.autos
comp vs sci	comp.os.mswindows.misc	comp. sys. ibm. pc.
	sci.crypt	hardware
comp vs talk	comp.sys.mac.hardware	sci.med
	talk.politics.mideast	comp. os. mswindows.
rec vs talk	rec.autos	misc
	talk.politics.misc	talk.politics.guns
rec vs sci	rec.autos	rec.sport.baseball
	sci.space	talk.religion.misc
sci vs talk	sci.med	rec.sport.hockey
	talk.religion.misc	sci.electronics
		sci.space
		talk.politics.mideast

其次,由于 20News Groups 文本数据集的特征数很大,为了实验的方便,用主元分析法对数据集的特征做降维预处理,取前 100 维构成新的实验数据。

4.2 少量目标域训练样本上的实验结果与分析

实验中取源域中的全部数据,分别取目标域中的 1%、5%、10%的数据训练 TL-MPM 算法分类器和 MPM 算法分类器。在 T-MPM 和 TL-

MPM 算法中的核函数都选取高斯核函数,参数  $\sigma$  在 (0.001,0.01,0.1,1,10,100,1 000) 中选取。TL-MPM 算法中的参数在 (0.001,0.01,0.1,1,10,100,1 000) 中选取。对表 3 处理好的 6 个真实分类数据集分别随机独立重复 10 次实验,选取最优参数下的取值,表 4、表 5 和表 6 分别记录了 10 次独立重复实验正确率的均值。

表 4 目标域样本的 1%训练结果

Table 4 Results of 1% training samples in the target domain/ %				
数据集	T-LMPM	T-KMPM	LTL-MPM	KTL-MPM
comp vs rec	71.71	72.96	72.56	78.42
comp vs sci	66.46	74.16	72.22	76.49
comp vs talk	90.96	99.15	99.48	99.75
rec vs talk	60.51	71.29	66.26	76.14
rec vs sci	62.49	71.00	64.17	76.48
sci vs talk	68.21	77.24	67.98	76.44

表 5 目标域样本的 5%训练结果

Table 5 Results of 5% training samples in the target domain/ %				
数据集	T-LMPM	T-KMPM	LTL-MPM	KTL-MPM
comp vs rec	61.46	80.92	64.19	83.94
comp vs sci	58.72	76.33	63.29	79.56
comp vs talk	94.13	99.41	97.74	99.83
rec vs talk	57.36	76.30	60.80	78.60
rec vs sci	57.12	78.81	57.49	81.50
sci vs talk	57.42	86.34	63.98	86.75

表 6 目标域样本的 10%训练结果

Table 6 Results of 10% training samples in the target domain/ %				
数据集	T-LMPM	T-KMPM	LTL-MPM	KTL-MPM
comp vs rec	76.46	85.51	80.05	86.06
comp vs sci	70.87	84.75	75.00	83.70
comp vs talk	96.16	99.32	98.66	99.69
rec vs talk	65.30	81.64	66.74	81.56
rec vs sci	64.93	84.89	70.17	85.90
sci vs talk	70.49	90.18	72.17	89.09

观察并比较表 4、表 5、表 6 中 T-LMPM 和 T-KMPM、LTL-MPM 和 KTL-MPM 的列。发现 T-KMPM 在各数据集上的分类结果优于 T-LMPM, KTL-MPM 在各数据集上的分类结果优于 LTL-MPM。说明该真实数据集是非线性的,线性分类方法有一定的局限性。另外,LTL-MPM 在各数据集上的分类结果优于 T-LMPM,KTL-MPM 在各数据集上的分类结果优于 T-KMPM,说明源域对目标域的分类起到了一定程度的作用。但提高的多少取决于两域相关性。注意到,在 comp vs rec,comp vs sci,rec vs talk 和 rec vs sci 四个数据集中,本文算法对实验结果有很大程度的提高,说明在这四类数据集上源域与目标域数据的相似性比较高。在 comp vs talk 数据集上提高程度很小,说明该数据集上源域和目标域相似度不高。在 sci vs talk 数据集上没有提高,反而下降,源域对目标域数据起了反作用,说明该数据集上源域和目标域相似度很低。

#### 4.3 目标域训练样本变化对分类结果的影响

少量目标域训练样本上实验说明该真实数据集是非线性数据,线性算法并不能很好反应数据的分类情况,为了观察 TL-MPM 算法在样本变化时的有效性,实验中用其非线性算法训练分类器。实验中取源域中的全部数据,取目标域中的 1%、5%、10%、20%、30%、40%、50% 的数据训练 KTL-MPM 算法分类器,取目标域中的 1%、5%、10%、20%、30%、40%、50% 的数据训练 T-KMPM 算法分类器。其他设置同 4.2 节。图 4~9 分别为 6 个数据集上的实验结果绘制的折线图。

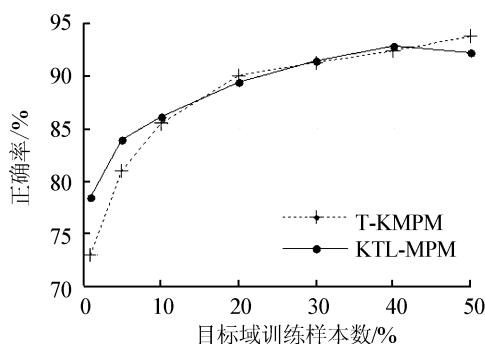


图 4 comp vs rec 分类精度变化

Fig.4 Classification accuracy trends of comp vs rec

观察发现,在目标域训练样本增加时,两种算法的精确度都有提高,在目标域训练样本量较少时

(10%以内),除 sci vs talk 数据集,其他数据集上 KTL-MPM 算法都好于 T-KMPM 算法,说明样本较少时,本文提出算法的优势很明显。在目标域训练样本增加到一定程度时,两算法的精度差距在逐渐缩小,最后几乎一致,甚至被 T-KMPM 算法反超,说明样本增大时,目标域的训练数据的信息越来越多,偏差越来越小,迁移过来的源域均值信息已经无法矫正该偏差,因为这个偏差已经小于了源域与目标域之间的偏差。这时迁移过来的源域均值信息甚至转化为了噪声信息,所以精确度反而变低。

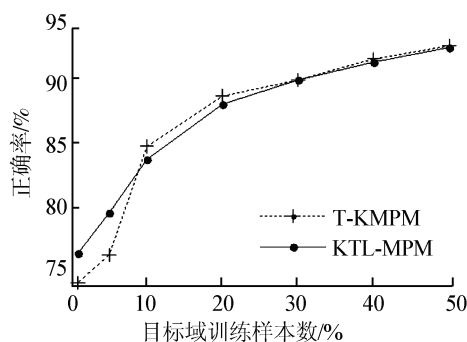


图 5 comp vs sci 分类精度变化

Fig.5 Classification accuracy trends of comp vs sci

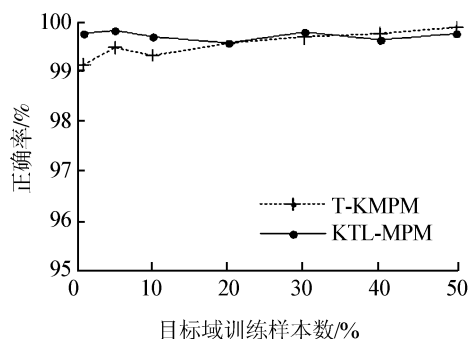


图 6 comp vs talk 分类精度变化

Fig.6 Classification accuracy trends of comp vs talk

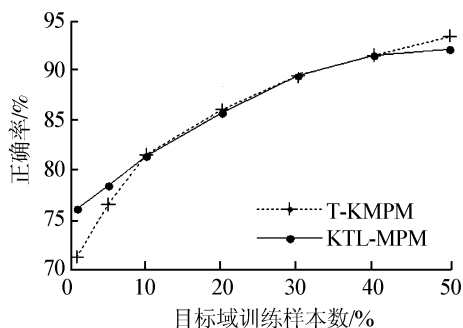


图 7 rec vs talk 分类精度变化

Fig.7 Classification accuracy trends of rec vs talk

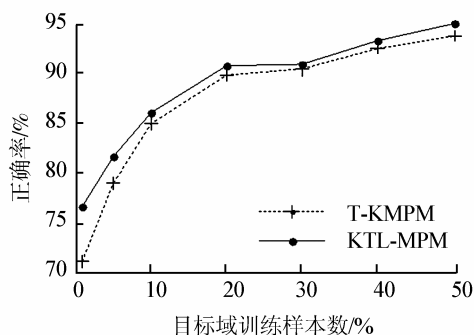


图8 rec vs sci 分类精度变化

Fig.8 Classification accuracy trends of rec vs sci

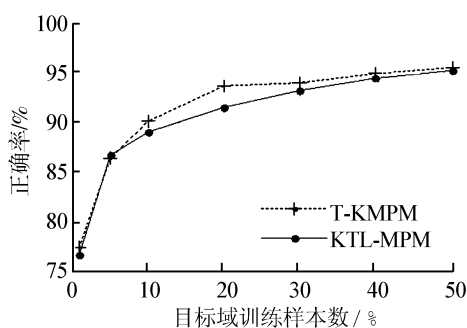


图9 sci vs talk 分类精度变化

Fig.9 Classification accuracy trends of sci vs talk

另外,本文算法并没有增加算法的时间复杂度,分析算法部分可知,该算法线性部分的时间复杂度为  $O(kd^3)$ ,非线性部分的时间复杂度为  $O(kn^3)$ ,这是因为在最小二乘递归过程中需要求矩阵的逆,其中  $k$  为迭代次数, $d$ 、 $n$  分别为目标域训练样本的维数和个数。

## 5 结束语

本文针对目标域标记数据个数很少,而源于数据仅知道均值的背景下,提出了基于最小最大概率机的迁移学习分类算法,将源域的均值数据用来修正目标域因训练数据少而带来的分类偏差。实验表明,在目标域训练样本个数很少的情况下,取得了很好的效果,另外将最小最大概率机分类算法扩展到了迁移学习的领域。由于该算法仅利用了源域数据的均值信息,并不能从多方面挖数据的相似之处,这既是该算法的优点,也是本算法的不足之处。

## 参考文献:

[1] PAN S J, YANG Qiang. A survey on transfer learning[J].

IEEE transactions on knowledge and data engineering, 2010, 22(10): 1345-1359.

[2] DAI Wenyuan, YANG Qiang, XUE Guirong, et al. Boosting for transfer learning[C]//Proceedings of the 24<sup>th</sup> International Conference on Machine learning. New York, NY, USA, 2007: 193-200.

[3] PAN S J, KWOK J T, YANG Qiang. Transfer learning via dimensionality reduction[C]//Proceedings of the 23<sup>rd</sup> National Conference on Artificial Intelligence. AAAI, 2008, 2: 677-682.

[4] 张景祥, 王士同, 邓赵红, 等. 具有协同约束的共生迁移学习算法研究[J]. 电子学报, 2014, 42(3): 556-560. ZHENG Jingxiang, WANG Shitong, DENG Zhaohong, et al. Symbiosis transfer learning method with collaborative constraints[J]. Acta electronica sinica, 2014, 42(3): 556-560.

[5] 于重重, 吴子珩, 谭励, 等. 非平衡集成迁移学习模型及其在桥梁结构健康监测中的应用[J]. 智能系统学报, 2013, 8(1): 46-51.

YU Chongchong, WU Zijun, TAN Li, et al. Unbalanced integrated transfer learning model and its application to bridge structural health monitoring[J]. CAAI transactions on intelligent systems, 2013, 8(1): 46-51.

[6] 毕安琪, 王士同. 基于 SVC 和 SVR 约束组合的迁移学习分类算法[J]. 控制与决策, 2014, 29(6): 1021-1026.

BI Anqi, WANG Shitong. Transfer classification learning based on combination of both SVC and SVR's constraints [J]. Control and decision, 2014, 29(6): 1021-1026.

[7] 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.

[8] LONG Mingsheng, WANG Jianmin, DING Guiguang, et al. Adaptation regularization: a general framework for transfer learning[J]. IEEE transactions on knowledge and data engineering, 2014, 26(5): 1076-1089.

[9] LONG Mingsheng, WANG Jianmin, DING Guiguang, et al. Transfer learning with graph co-regularization [J]. IEEE transactions on knowledge and data engineering, 2014, 26(7): 1805-1818.



- [10] PERLICH C, DALESSANDRO B, RAEDER T, et al. Machine learning for targeted display advertising: transfer learning in action[J]. Machine learning, 2014, 95(1): 103-127.
- [11] LANCKRIET G R G, EL GHAOU L, BHATTACHARYYA C, et al. A robust minimax approach to classification[J]. The journal of machine learning research, 2003(3): 555-582.
- [12] ISII K. On the sharpness of Chebyshev-type inequalities [J]. Annals of the institute of statistical mathematics, 1963(14): 185-197.
- [13] BERTSIMAS D, SETHURAMAN J. Moment problems and semidefinite optimization[M]//WOLKOWICZ H, SAIGAL R, VANDENBERGHE L. Handbook of Semidefinite Programming. 3<sup>rd</sup> ed. US: Springer, 2000: 469-509.
- [14] CHERNOFF H. The selection of effective attributes for deciding between hypotheses using linear Discriminant functions[C]//The Proceedings of the International Conference on Frontiers of Pattern Recognition. New York, USA, 1972: 55-60.
- [15] 陈宝林. 最优化理论与算法[M]. 2 版. 北京: 清华大学出版社, 2005.
- CHEN Baolin. Optimization theory and algorithm[M]. 2<sup>nd</sup> ed. Beijing: Tsinghua University Press, 2005.
- [16] PLAN Y, VERSHYNIN R. Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach[J]. IEEE transactions on information theory, 2012, 59(1): 482-494.
- [17] BENSOUSSAN A, TEMAM R. Equations aux derivees partielles stochastiques non lineaires [J]. Israel journal of mathematics, 1972, 11(1): 95-129.
- [18] BOYD S, VANDENBERGHE L, FAYBUSOVICH L. Convex optimization[J]. IEEE transactions on automatic control, 2006, 51(11): 1859.

#### 作者简介:



王晓初,男,1987 年生,硕士研究生,主要研究方向为人工智能与模式识别、数字图像处理。



包芳,女,1970 年生,教授,博士,主要研究方向为人工智能、模式识别、图像视觉。



王士同,男,1964 年生,教授、博士生导师,主要研究方向为人工智能。主持或参加过 6 项国家自然科学基金项目,1 项国家教委优秀青年教师基金项目,其他省部级科研项目 10 多项。先后获国家教委、中船总公司和江苏省省部级科技进步奖 5 项。发表学术论文近百篇,被 SCI、EI 检索 50 余篇。