信息与控制　2016 年　第 45 卷　第 4 期：463 ~ 470

DOI：10.13976/j.cnki.xk.2016.0463　　文章编号：1002 - 0411(2016) - 04 - 0463 - 09

基于最大均值差异的多标记迁移学习算法

姜海燕[1,2]，刘昊天[1]，舒　欣[1]，徐　彦[1]，伍艳莲[1,2]，郭小清[1]

1. 南京农业大学信息科技学院，江苏 南京　210095；　2. 国家信息农业工程技术中心，江苏 南京　210095

基金项目：国家自然科学基金资助项目(30971697，61403205)；国家 863 计划资助项目(2013AA100404)；江苏省农业科技自主创新资金(CX(16)1039)

通信作者：姜海燕，jianghy@njau.edu.cn　收稿/录用/修回：2015 - 08 - 12/2016 - 01 - 20/2016 - 01 - 29

摘要

针对多标记迁移学习中源领域与目标领域的特征分布差异会导致源领域数据无法被目标领域利用的问题，提出了一种基于最大均值差异的多标记迁移学习算法(Multi-Label Transfer Learning via Maximum mean discrepancy，M-MLTL)，算法通过分解关系矩阵构造共享子空间，并采用最大均值差异(maximum mean discrepancy)作为评价指标，最小化子空间特征的分布差异，从而使源领域与目标领域的特征分布尽可能相似. 多标记图像分类实验的结果表明，新算法比同类算法有更高的精度和计算效率.

关键词

多标记

迁移学习

最大均值差异

共享子空间

中图分类号：TP181

文献标识码：A

Multi-label Transfer Learning via Maximum Mean Discrepancy

JIANG Haiyan[1,2], LIU Haotian[1], SHU Xin[1], XU Yan[1], WU Yanlian[1,2], GUO Xiaoqing[1]

1. *College of Information Science and Technology*, *Nanjing Agricultural University*, *Nanjing* 210095, *China*;

2. *National Engineering and Technology Center for Information Agriculture*, *Nanjing* 210095, *China*

Abstract

Due to the different distribution of features between the source and target domains in a multi-label transfer learning problem, source domain data cannot exert any effect. To resolve this problem, here we propose novel multi-label transfer learning via the maximum mean discrepancy. The proposed algorithm decomposes a relational matrix to learn a common subspace. Furthermore, we incorporate the empirical maximum mean discrepancy into the objective function of matrix factorization to minimize the probability distance between different domains. Experimental results from multi-label classification demonstrate that the proposed approach achieves better performance than other similar algorithms in terms of accuracy and efficiency.

Keywords

multi-label;

transfer learning;

maximum mean discrepancy;

shared subspace

1　引言

迁移学习是运用已存在的知识对不同但相关领域问题进行求解的一类新型机器学习方法[1]. 相比于传统的机器学习方法，迁移学习的训练数据和测试数据间没有同分布要求，更适用于大数据时代数据变化快、样本时效性强的特点，是近年来机器学习领域的研究热点[2-5]. 基于特征的迁移是迁移学习的常用方法，根据对样本特征的处理又可以分为基于特征选择和特征映射方法：基于特征选择的方法首先提取不同领域的共有特征，再利用共有特征进行知识迁移[6-7]；而基于特征映射的方法将不同领域样本的特征由原始高维特征空间映射到低维潜在特征子空间(latent feature subspace)中，使得不同领域样本的特征分布在子空间内相似，进而可以基于子空间特征训练得到同时适用于源领域和目标领域的预测模型[8-11].

尽管迁移学习算法理论研究已趋于成熟，但大多数迁移学习算法只能够解决单个类别标记的“单标记数据”迁移问题，而现实世界中还包含许多被标注了多个类别标记的“多标记数据”[12]. 如何从相关领域的多标记数据中提取知识迁移到目标问题域中是迁移学习急需解决的问题.

目前针对多标记数据的迁移学习主要有两类方法：一类方法将多标记迁移学习问题转化为多个单标记迁移学习问题处理[13]. 该方法对多标记数据中的每一个类别标记采用独立的单标记迁移学习算法学习并预测，不同类别标记的分类算法间彼此独立. 然而，一个多标记数据包含的不同标记之间往往存在相关联的语义，研究发现忽视标记

间的关联关系而对每个标记独立学习不利于算法的学习和预测[14]. 另一类方法将已有的单标记迁移学习算法或多标记传统机器学习算法改进为多标记迁移学习算法，使其能够处理多标记迁移学习问题. Cheng[15]针对多模态多标记数据的迁移学习问题，将 Lasso 模型改进为多模态多标记特征选择模型，模型提取不同模态多标记数据间的公共特征，并采用多核相关向量机(relevance vector machine，RVM)基于公共特征进行预测. Fu[16]针对多标记数据的零样本学习问题，采用深度学习构造样本的文本语义表示，并通过构造标记关系的幂集与样本语义文本的 KNN 图建立样本与标记间的关系. Han[17]针对多标记迁移学习中特征向量维数过高的问题提出基于稀疏表示的多标记迁移学习算法，将源领域特征与标记的关联关系映射到低维子空间中，目标领域通过与源领域共享子空间的方式学习源领域的知识. 上述多标记迁移学习算法均通过共享子空间的方法实现知识迁移，但由于源领域与目标领域样本的原始特征分布不同，相应子空间内的潜在特征分布也可能不同，此时源领域子空间中的潜在特征不一定适用于目标领域，无法促进目标领域的学习.

为了减小源领域与目标领域的多标记数据在子空间内的分布差异，本文提出一种基于最大均值差异的多标记迁移学习算法(Multi-Label Transfer Learning via Maximum mean discrepancy，M-MLTL). 算法通过分解源领域的“特征—标记”关系矩阵将标记间的关联关系嵌入潜在特征子空间，目标领域与源领域通过共享子空间迁移知识；采用最大均值差异(maximum mean discrepancy，MMD)[18]作为评价指标，最小化不同领域的子空间特征概率分布差异，从而充分利用源领域的数据.

2 最大均值差异

假设分别存在一个满足$\mathbb{P}$分布的源领域 $\boldsymbol{X}^{(\mathrm{s})}=[\boldsymbol{x}_1^{(\mathrm{s})},\cdots,\boldsymbol{x}_{n_\mathrm{s}}^{(\mathrm{s})}]$和一个满足$\mathbb{Q}$分布的目标领域 $\boldsymbol{X}^{(\mathrm{t})}=[\boldsymbol{x}_1^{(\mathrm{t})},\cdots,\boldsymbol{x}_{n_\mathrm{t}}^{(\mathrm{t})}]$. 令$\mathcal{H}$表示再生核希尔伯特空间(reproducing kernel Hilbert space，RKHS)，$\boldsymbol{\phi}(\cdot):\mathbb{X}\to\mathcal{H}$表示原始特征空间映射到 RKHS 的映射函数，当 n_s，$n_\mathrm{t}\to\infty$ 时 $\boldsymbol{X}^{(\mathrm{s})}$ 和 $\boldsymbol{X}^{(t)}$ 在 RKHS 中的最大均值差异(maximum mean discrepancy，MMD)[18]可以表示为

$$f(\boldsymbol{X}^{(\mathrm{s})},\boldsymbol{X}^{(\mathrm{t})})=\left\|\frac{1}{n_\mathrm{s}}\sum_{i=1}^{n_\mathrm{s}}\boldsymbol{\phi}(\boldsymbol{x}_i^{(\mathrm{s})})-\frac{1}{n_\mathrm{t}}\sum_{j=1}^{n_\mathrm{t}}\boldsymbol{\phi}(\boldsymbol{x}_j^{(\mathrm{t})})\right\|_{\mathcal{H}} \tag{1}$$

MMD 度量就是使用源域数据集与目标域数据集的总体均值之差来表示源域与目标域之间的分布差异. 目前 MMD 已被广泛应用于单标记迁移学习研究中[19-21]，但仍没有被用于多标记迁移学习领域中.

3 基于最大均值差异的多标记迁移学习算法

首先描述本文的问题，假设存在已标记源领域样本集 $D^{(\mathrm{s})}=\{(\boldsymbol{x}_1^{(\mathrm{s})},\boldsymbol{y}_1^{(\mathrm{s})}),\cdots,(\boldsymbol{x}_{n_\mathrm{s}}^{(\mathrm{s})},\boldsymbol{y}_{n_\mathrm{s}}^{(\mathrm{s})})\}$和未标记的目标领域样本集 $D^{(\mathrm{t})}=\{\boldsymbol{x}_1^{(\mathrm{t})},\cdots,\boldsymbol{x}_{n_\mathrm{t}}^{(\mathrm{t})}\}$，构造用于预测目标领域的样本标记 $\boldsymbol{Y}^{(\mathrm{t})}=[\boldsymbol{y}_1^{(\mathrm{t})},\cdots,\boldsymbol{y}_{n_\mathrm{t}}^{(\mathrm{t})}]$的预测模型，其中 $D^{(\mathrm{s})}$ 和 $D^{(\mathrm{t})}$ 的样本特征概率分布不同. 为叙述方便，表 1 给出了本文中的常用符号.

表 1 本文的常用符号及描述

Tab.1 Summary of the notations used in this paper

符号	描述
$D^{(\mathrm{s})}$，$D^{(\mathrm{t})}$	源/目标领域样本集
$\boldsymbol{X}^{(\mathrm{s})}$，$\boldsymbol{X}^{(\mathrm{t})}$	源/目标领域样本特征矩阵
$\boldsymbol{Y}^{(\mathrm{s})}$，$\boldsymbol{Y}^{(\mathrm{t})}$	源/目标领域样本标记矩阵
$\boldsymbol{Z}^{(\mathrm{s})}$，$\boldsymbol{Z}^{(\mathrm{t})}$	源/目标领域样本潜在特征矩阵
n_s，n_t	源/目标领域样本数量
d	样本输入特征数量
m	样本类别标记数量
g	低维子空间特征数量

3.1 目标函数

本文通过构造一个理想的潜在子空间，将标记间关系嵌入潜在特征中并保证源领域与目标领域的潜在特征分布相似. 如何将标记间关联关系嵌入潜在子空间及如何最小化子空间中的分布差异是本算法的两个关键问题.

3.1.1 潜在子空间嵌入

潜在特征表达了介于样本输入特征的低级语义和类别标记的高级语义间的中间语义，因此必须建立输入特征和类别标记间的关系. 基于上述考虑，算法构造“特征 — 标记”关系矩阵$\boldsymbol{G}=\boldsymbol{X}^{(\mathrm{s})}\boldsymbol{Y}^{(\mathrm{s})\mathrm{T}}\in\mathrm{R}^{d\times m}$. $\boldsymbol{G}_{ij}=\sum_k x_{ik}^{(\mathrm{s})}\cdot y_{jk}^{(\mathrm{s})}$，其中 $x_{ik}^{(\mathrm{s})}\geqslant 0$ 是第 k 个样本中的第 i 个输入特征值，与之相关的 $n_j^{(i)}=\sum_k y_{jk}^{(\mathrm{s})}$ 是含有第 j 个标记的样本的数量. 如果 $\boldsymbol{G}_{ij}$ 的值较大，则 $x_{ik}^{(\mathrm{s})}$ 和 $n_j^{(i)}$ 的值也较大，因此 $\boldsymbol{G}_{ij}$ 反映了第 i 个输入特征和第 j 个类别标记间的相关性.

受潜在语义分析(latent semantic analysis，LSA)[22]启发，算法通过分解关系矩阵 $\boldsymbol{G}$ 将训练样本的输入特征和类别标记关系嵌入潜在子空间：

$$\boldsymbol{G}=\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}} \tag{2}$$

其中 $\boldsymbol{U}\in\mathrm{R}^{d\times g}$，$\boldsymbol{V}\in\mathrm{R}^{m\times g}$，$g$ 是潜在特征的个数，矩阵 $\boldsymbol{U}$ 的行向量 $\boldsymbol{u}_i$ 是第 i 个输入参数的潜在特征表示. 类似地，矩阵 $\boldsymbol{V}$ 的行向量 $\boldsymbol{v}_j$ 是第 j 个标记的潜在特征表示. 为方便求解，式(2)的矩阵分解问题可以表示成式(3)的优化问题：

$$\min_{\boldsymbol{U},\boldsymbol{V}}\left(\left\|\boldsymbol{G}-\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\right\|_{\mathrm{F}}^2+\gamma_1\left\|\boldsymbol{U}\right\|_{\mathrm{F}}^2+\gamma_2\left\|\boldsymbol{V}\right\|_{\mathrm{F}}^2\right) \tag{3}$$

其中，$\|\cdot\|_{\mathrm{F}}$ 表示矩阵的 Frobenius 范数，$\|\boldsymbol{U}\|_{\mathrm{F}}^2$ 和 $\|\boldsymbol{V}\|_{\mathrm{F}}^2$ 用于控制正则项的复杂度，γ_1 和 γ_2 是非负参数，本文中设 $\gamma_1=\gamma_2=1$.

根据上述分解，可以得到源领域和目标领域数据在基 $\boldsymbol{U}$ 下的潜在特征，即 $\tilde{\boldsymbol{x}}_i^{(\mathrm{s})}=\boldsymbol{\psi}(\boldsymbol{x}_i^{(\mathrm{s})})=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_i^{(\mathrm{s})}$，$\tilde{\boldsymbol{x}}_i^{(\mathrm{t})}=\boldsymbol{\psi}(\boldsymbol{x}_i^{(\mathrm{t})})=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_i^{(\mathrm{t})}$.

3.1.2 最小化子空间分布差异

由于源领域和目标领域的样本特征分布不同，源领域和目标领域在子空间中的潜在特征 $\tilde{\boldsymbol{X}}^{(\mathrm{s})}=[\boldsymbol{x}_1^{(\mathrm{s})},\cdots,\boldsymbol{x}_{n_\mathrm{s}}^{(\mathrm{s})}]$ 和 $\tilde{\boldsymbol{X}}^{(\mathrm{t})}=[\boldsymbol{x}_1^{(\mathrm{t})},\cdots,\boldsymbol{x}_{n_\mathrm{t}}^{(\mathrm{t})}]$也可能出现分布不同的情况. 由 MMD 定义的 $\tilde{\boldsymbol{X}}^{(\mathrm{s})}$ 和 $\tilde{\boldsymbol{X}}^{(t)}$ 之间的概率距离可记为

$$f^2(\widetilde{\boldsymbol{X}}^{(s)}, \widetilde{\boldsymbol{X}}^{(t)}) = \left\| \frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{\phi}(\tilde{\boldsymbol{x}}_i^{(s)}) - \frac{1}{n_t}\sum_{j=1}^{n_t}\boldsymbol{\phi}(\tilde{\boldsymbol{x}}_j^{(t)}) \right\|_{\boldsymbol{H}}^2$$
$$= \left\| \frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{\phi}(\boldsymbol{\psi}(\boldsymbol{x}_i^{(s)})) - \frac{1}{n_t}\sum_{j=1}^{n_t}\boldsymbol{\phi}(\boldsymbol{\psi}(\boldsymbol{x}_j^{(t)})) \right\|_{\boldsymbol{H}}^2 \tag{4}$$

其中，$\boldsymbol{\phi}(\cdot)$是特征映射核函数，$\boldsymbol{\phi}(\boldsymbol{\psi}(\cdot))$也可以看作一个特征映射核函数. 这里构造一个线性映射核函数 $\boldsymbol{\phi}(\boldsymbol{\psi}(\boldsymbol{x})) = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}$，故有：

$$\min_{\boldsymbol{\phi}(\boldsymbol{\psi}(\cdot))} f^2(\widetilde{\boldsymbol{X}}^{(s)}, \widetilde{\boldsymbol{X}}^{(t)}) = \min_{\boldsymbol{U}} f^2(\widetilde{\boldsymbol{X}}^{(s)}, \widetilde{\boldsymbol{X}}^{(t)})$$

其中$f^2(\widetilde{\boldsymbol{X}}^{(s)}, \widetilde{\boldsymbol{X}}^{(t)})$可以通过如下步骤化简：

$$f^2(\widetilde{\boldsymbol{X}}^{(s)}, \widetilde{\boldsymbol{X}}^{(t)})$$
$$= \left\| \frac{1}{n_s}\sum_{i=1}^{n_s}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_i^{(s)} - \frac{1}{n_t}\sum_{j=1}^{n_t}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_i^{(t)} \right\|^2$$
$$= \left\| \frac{1}{n_s}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{(s)}\boldsymbol{I}^{(s)} - \frac{1}{n_t}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{(t)}\boldsymbol{I}^{(t)} \right\|^2$$
$$= \boldsymbol{U}^{\mathrm{T}}\left(\frac{1}{n_s^2}\boldsymbol{X}^{(s)}\boldsymbol{I}^{(s)}\boldsymbol{I}^{(s)\mathrm{T}}\boldsymbol{X}^{(s)\mathrm{T}} - \frac{1}{n_s n_t}\boldsymbol{X}^{(s)}\boldsymbol{I}^{(s)}\boldsymbol{I}^{(t)\mathrm{T}}\boldsymbol{X}^{(t)\mathrm{T}} - \frac{1}{n_s n_t}\boldsymbol{X}^{(t)}\boldsymbol{I}^{(t)}\boldsymbol{I}^{(s)\mathrm{T}}\boldsymbol{X}^{(s)\mathrm{T}} + \frac{1}{n_t^2}\boldsymbol{X}^{(t)}\boldsymbol{I}^{(t)}\boldsymbol{I}^{(t)\mathrm{T}}\boldsymbol{X}^{(t)\mathrm{T}} \right)\boldsymbol{U}$$
$$= \mathrm{tr}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U}) \tag{5}$$

其中，$\boldsymbol{X} = [\boldsymbol{X}^{(s)}, \boldsymbol{X}^{(t)}] \in \mathrm{R}^{d\times(n_s+n_t)}$，$\boldsymbol{I}^{(s)} = [1, \cdots, 1]^{\mathrm{T}} \in \mathrm{R}^{n_s\times 1}$，$\boldsymbol{I}^{(t)} = [1, \cdots, 1]^{\mathrm{T}} \in \mathrm{R}^{n_t\times 1}$，参数矩阵 $\boldsymbol{M}$ 构造如下：

$$M_{ij} = \begin{cases} \dfrac{1}{n_s^2}, & \boldsymbol{x}_i, \boldsymbol{x}_j \in D^{(s)} \\ \dfrac{1}{n_t^2}, & \boldsymbol{x}_i, \boldsymbol{x}_j \in D^{(t)} \\ \dfrac{-1}{n_s n_t}, & \text{otherwise} \end{cases} \tag{6}$$

结合式(3)和式(5)，所提模型可描述为如下的优化问题：

$$\min_{\boldsymbol{U},\boldsymbol{V}} F(\boldsymbol{U}, \boldsymbol{V}) = \min_{\boldsymbol{U},\boldsymbol{V}} \left(\|\boldsymbol{G} - \boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\|_{\mathrm{F}}^2 + \mu\mathrm{tr}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U}) + \gamma_1\|\boldsymbol{U}\|_{\mathrm{F}}^2 + \gamma_2\|\boldsymbol{V}\|_{\mathrm{F}}^2 \right) \tag{7}$$

其中$\mu \geqslant 0$.

3.2　优化问题求解

优化问题(7)有两个待优化变量 $\boldsymbol{U}$、$\boldsymbol{V}$，当固定 $\boldsymbol{U}$ 时目标问题是一个关于 $\boldsymbol{V}$ 的凸优化问题，反之亦然. 因此，本文拟采用交替迭代算法来求解，其求解过程如下：

步骤 1　固定 $\boldsymbol{U}$ 求解 $\boldsymbol{V}$. 令 $F(\boldsymbol{U}, \boldsymbol{V})$关于 $\boldsymbol{V}$ 的偏导数为 0，得到：

$$\frac{\partial F(\boldsymbol{U}, \boldsymbol{V})}{\partial \boldsymbol{V}} = 0 \tag{8}$$

经过化简，得到关于 $\boldsymbol{V}^{\mathrm{T}}$ 的线性方程组：

$$(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} + \gamma_2\boldsymbol{I})\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{G} \tag{9}$$

由于 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} + \gamma_2\boldsymbol{I}$ 是对称正定的，该方程可以采用共轭梯度法(Conjugate Gradient method，CG)[23]求解.

步骤 2　固定 $\boldsymbol{V}$ 求解 $\boldsymbol{U}$. 与求解 $\boldsymbol{V}$ 类似，令 $F(\boldsymbol{U}, \boldsymbol{V})$关于 $\boldsymbol{U}$ 的偏导数为 0，得到：

$$\frac{\partial F(\boldsymbol{U}, \boldsymbol{V})}{\partial \boldsymbol{U}} = 0 \tag{10}$$

通过矩阵变换得到：

$$(\mu\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^{\mathrm{T}} + \gamma_1\boldsymbol{I})\boldsymbol{U} + \boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{G}\boldsymbol{V} \tag{11}$$

上述方程是一个 Sylvester 方程，可以利用 Bartels-Stewart 算法[24]求解.

步骤 3　重复步骤 1 ~ 2，直到收敛.

3.3　算法步骤

经上述分析，所提算法详细步骤描述如下：

input 源领域样本输入矩阵 $\boldsymbol{X}^{(s)} \in \mathrm{R}^{d\times n_s}$和标记矩阵 $\boldsymbol{Y}^{(s)} \in \{0, 1\}^{m\times n_s}$；目标领域样本输入特征矩阵 $\boldsymbol{X}^{(t)} \in \mathrm{R}^{d\times n_t}$；控制参数 μ、γ_1 和 γ_2，潜在特征的数量 g.

Step 1　计算子空间矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$：

$$\boldsymbol{G} = \boldsymbol{X}^{(s)}\boldsymbol{Y}^{(s)\mathrm{T}}$$

随机初始化矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$；

repeat

　$\boldsymbol{U}$ 不变，用 CG 算法求解式(9)中的矩阵 $\boldsymbol{V}$；

　$\boldsymbol{V}$ 不变，用 Bartels-Stewart 算法求解式(11)中的矩阵 $\boldsymbol{U}$；

until 收敛.

Step 2　计算样本的潜在特征 $\widetilde{\boldsymbol{X}}$：

$$\widetilde{\boldsymbol{X}}^{(s)} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{(s)}, \ \widetilde{\boldsymbol{X}}^{(t)} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{(s)}.$$

Step 3　采用基分类算法预测目标领域样本类别标记 $\boldsymbol{Y}^{(t)}$. 采用 RankSVM[17]基于 $\widetilde{\boldsymbol{X}}^{(s)}$ 和 $\boldsymbol{Y}^{(s)}$ 训练分类器，并预测 $\widetilde{\boldsymbol{X}}^{(t)}$ 的标记 $\boldsymbol{Y}^{(t)}$.

output 目标领域样本的预测标记矩阵 $\boldsymbol{Y}^{(t)}$.

3.4　时间复杂度分析

所提模型的求解主要包括式(9)中线性方程的求解与式(11)中 Sylvester 方程的求解两个部分. 其中 CG 时间复杂度为 $O(g\sqrt{k})$[23]，k 是矩阵 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} + \gamma_2\boldsymbol{I}$ 的条件数. 但求解式(9)时要将原方程分解为 m 个关于 $v_i(\boldsymbol{V}^{\mathrm{T}} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m))$的线性方程，并对每个方程采用 CG 求解，所以求解式(9)的时间复杂度为 $O(mg\sqrt{k})$. 求解式(11)的 Bartels-Stewart 算法时间复杂度为 $O(d^3)$[24]. 因此，本文算法每一次迭代的时间复杂度为 $O(\max(d^3, mg\sqrt{k}))$. 设算法最大迭代次数为 T，则算法时间复杂度为 $O(\max(d^3, mg\sqrt{k})T)$，可见本算法关键环节的时间复杂度不受训练样本数量的影响，能够应付大样本情况下的学习问题.

4　实验及结果分析

本文采用两个已被广泛认可的多标记图像数据集 Corel5k[25]和 ESPGame[26]构造实验数据集，选择 4 种对比算法基于 5 项评价指标验证本文所提模型的预测精度和计算时间. 本节首先对实验数据集、评价指标和对比算法做详细介绍，再结合实验结果从分类精度、关键参数和样本数量对算法的影响及计算时间三个方面分析所提算法的性能.

4.1　实验数据集构造

本文采用 Corel5k 和 ESPGame 交替作为源领域和目标领域组成实验数据集“Corel5k vs ESPGame”和“ESPGamevs Corel5k”. 其中，Corel5k 包含了 5 000 张图像和 260 个类别关键字，每张图像被手工标注 1 ~ 5 个关键字. ESPGame

数据集包含的图像比 Corel5k 更多样，除了照片还包括商标、涂鸦等类型的图像.

本文实验选取上述两个数据集共有的 10 个标记作为目标标记，图像的特征采用文[27]中提取的 Dense SIFT 视觉描述子. 从两个样本集中各随机抽取 1 200 个样本作为实验样本.

4.2 评价指标

与单标记学习算法不同，多标记学习算法不仅会输出一个分类函数 $h: \boldsymbol{X} \rightarrow 2^m$，通常还会输出一个标记排序函数 $\mathrm{rank}(\boldsymbol{x}_i, y_{ji})$，一个理想的函数 rank 对于所有 $y_{ji}=1$ 且 $y_{hi}=0$ 的情况都满足 $\mathrm{rank}(\boldsymbol{x}_i, y_{ji}) < \mathrm{rank}(\boldsymbol{x}_i, y_{hi})$. 为了能够从多个方面综合反映算法的预测效果，本文选取了 5 个目前被广泛采用的多标记学习评价指标[19, 28]：Hamming Loss、Ranking Loss、One Error、Coverage、Average Precision. 其中，Hamming Loss 是针对分类函数 h 准确率的评价指标，Hamming Loss 值越小表示算法的分类越准确；其余 4 个指标用于评价排序函数 rank 的效果，Average Precision 值越大表示函数 rank 的预测越好，其余的 3 个指标(Ranking Loss、One Error、Coverage)值越小说明排序函数 rank 的预测越好.

4.3 对比算法介绍

为验证本文所提算法的性能，本文选择了以下 4 个多标记算法(见表 2)作为对比. 各算法分别采用不同的降维方法，但都采用线性核函数的 RankSVM[29] 作为基础分类器. 其中 BL(base line)是原始 RankSVM 分类算法，该算法直接利用图像的 Dense SIFT 视觉描述子作为输入特征进行分类；LIFT(multi-label learning with Label spec IficFeaTures)[30]是目前最新的多标记学习算法，算法通过构造新的标记特征表示标记间关系；S-MLTL(Multi-Label Transfer Learning with Sparse representation)[17]是典型的多标记迁移学习算法，算法通过共享潜在子空间迁移知识；M-MLTL 是本文提出的多标记迁移学习算法，本算法相较于 S-MLTL 进一步考虑了源领域与目标领域的分布差异；为了验证 MMD 正则项对算法学习效果的影响，在实验中令 $\mu=0$ 得到M-MLTL 算法的对比算法 MLTL. MLTL 算法不包含 MMD 正则项，仅通过分解关系矩阵构造低维子空间. 本实验中 M-MLTL 算法的参数 μ 取值为 30，其它对比算法的参数设置均采用原文中的默认值.

4.4 多标记图像分类

5 种多标记算法的图像分类结果如表 2 所示，评价指标的最优结果被加粗表示. 由表 2 可知，在数据集“Corel5k vs ESPGame”上，新算法的 Ranking Loss、One Error、Coverage、Average Precision 四个评价指标都取得了最好的效果，但 Hamming Loss 指标上 M-MLTL 与 MLTL、LIFT 和 BL 算法相比没有明显的优势. 在数据集“ESPGame vs Corel5k”上，新算法的 Hamming Loss、Ranking Loss、Coverage 三个评价指标取得了最好的效果，但 S-MLTL 算法在 One Error 和 Average Precision 两个评价指标上略优于 M-MLTL.

表 2 算法分类结果(平均值方差)

Tab.2 Experimental results of the algorithms (mean ± std)

数据集	评价指标	算法				
		M-MLTL	MLTL	S-MLTL	LIFT	BL
Corel5k vs ESPGame	Hamming Loss ↓	0.110 ±0.000	0.110 ±0.000	0.198 ±0.035	0.110 ±0.000	0.115 ±0.000
	Ranking Loss ↓	0.449 ±0.005	0.480 ±0.018	0.545 ±0.000	0.534 ±0.020	0.576 ±0.000
	One Error ↓	0.873 ±0.004	0.903 ±0.018	0.941 ±0.000	0.927 ±0.008	0.940 ±0.000
	Coverage ↓	4.218 ±0.046	4.506 ±0.161	5.118 ±0.000	4.946 ±0.171	5.351 ±0.000
	Average Precision ↑	0.326 ±0.003	0.319 ±0.003	0.270 ±0.000	0.266 ±0.005	0.240 ±0.000
ESPGame vs Corel5k	Hamming Loss ↓	0.103 ±0.000	0.107 ±0.001	0.179 ±0.000	0.107 ±0.006	0.107 ±0.000
	Ranking Loss ↓	0.602 ±0.009	0.613 ±0.006	0.677 ±0.018	0.657 ±0.026	0.679 ±0.000
	One Error ↓	0.950 ±0.006	0.958 ±0.000	0.884 ±0.018	0.951 ±0.015	0.971 ±0.000
	Coverage ↓	5.459 ±0.078	5.557 ±0.055	6.130 ±0.161	5.945 ±0.246	6.149 ±0.000
	Average Precision ↑	0.227 ±0.005	0.218 ±0.000	0.242 ±0.003	0.203 ±0.018	0.186 ±0.000

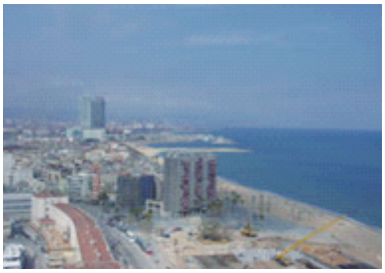
表 3 算法实验结果排序

Tab.3 Experimental results order of the algorithms

数据集	评价指标	算法 A_1: M-MLTL, A_2: MLTL, A_3: S-MLTL, A_4: LIFT, A_5: BL
Corel5k vs ESPGame	Hamming Loss	$A_1=A_2=A_4=A_5>A_3$
	Ranking Loss	$A_1>A_2>A_4>A_3>A_5$
	One Error	$A_1>A_2>A_4>A_5>A_3$
	Coverage	$A_1>A_2>A_4>A_3>A_5$
	Average Precision	$A_1>A_2>A_3>A_4>A_5$
ESPGame vs Corel5k	Hamming Loss	$A_1>A_5>A_2>A_4>A_3$
	Ranking Loss	$A_1>A_2>A_4>A_3>A_5$
	One Error	$A_3>A_1>A_4>A_2>A_5$
	Coverage	$A_1>A_2>A_4>A_3>A_5$
	Average Precision	$A_3>A_1>A_2>A_4>A_5$
综合排序		$A_1(38)>A_2(27)>A_4(19)>A_3(17)>A_5(8)$

表 4　分类结果的典型样例

Tab.4　Typical examples of classification results

源领域训练样本	 sand，sea	 house，sea
	 sand	 house，light
目标领域测试样本	 sand，sea	 house，sea
M-MLTL	sand，sea	house，sea
MLTL	sand，sea	—
S-MLTL	sand，sea	—
LIFT	sand	—
BL	sand	—

为了更直观地比较各项评价指标上算法的表现，本文综合两个数据集的结果，对每项评价指标上所有算法的结果排序，用 $A_1 \succ A_2$ 表示在该项指标下算法 A_1 优于算法 A_2. 根据排序结果对每个算法打分，算法在某个指标中排第 i 名则获 5-i 分，最后以算法在所有指标上的得分总和作为该算法的综合评分，排序结果见表 3. 算法综合表现排名由高到低依次为 M-MLTL、MLTL、LIFT、S-MLTL、BL，其中 S-MLTL 与 LIFT 差距不大. 值得注意的是，M-MLTL 算法的各项指标都优于 MLTL 算法，说明 MMD 正则项对算法的学习效果起到了积极作用. LIFT 和 S-MLTL 算法在不同数据集的各项指标上各有优劣但综合排名相近，虽然均高于没有经过降维操作的 BL 算法但综合表现低于本文所提算法，这是因为这 2 种算法在降维的过程中没有考虑不同领域数据在子空间中的特征分布差异，所以分类效果不理想.

表 4 提供了两组实验样例及各算法的分类结果. 在左边一组样例中，测试样本的期望标记是“sand，sea”，该测试样本与该组第 2 张训练样本相似，与第 1 张训练样本差异较大. 因此，LIFT 和 BL 作为普通多标记的学习算法仅识别了与第 2 张训练样本相同的标记“sand”；而 M-MLTL、MLTL 和 S-MLTL 三种多标记迁移学习算法都采用了分解“特征—标记”关系矩阵的子空间构造方法，可以通过对该组第 1 张训练样本的学习放大“sand”和“sea”在子空间的关联特征，进一步识别“sea”标记. 在右边一组样例中，训练样本普遍较灰暗，而测试样本较明亮，这样的特征分布差异导致没有考虑不同领域特征分布差异的 MLTL 等算法未能识别任何标记，而 M-MLTL 引入 MMD 作为指标，通过缩小训练样本与测试样本在子空间中的特征分布差异，成功识别所有标记. 上述两组实验样例说明：通过分解“特征—标记”关系矩阵构造子空间的方法能够发掘标记间的关联特征；通过引入 MMD 正则项减小训练样本与测试样本在子空间中的分布差异，能够有效提高迁移学习算

法的分类精度. MMD 作为指标，通过缩小训练样本与测试样本在子空间中的特征分布差异，成功识别所有标记. 上述两组实验样例说明：通过分解“特征—标记”关系矩阵构造子空间的方法能够发掘标记间的关联特征；通过引入 MMD 正则项减小训练样本与测试样本在子空间中的分布差异，能够有效提高迁移学习算法的分类精度.

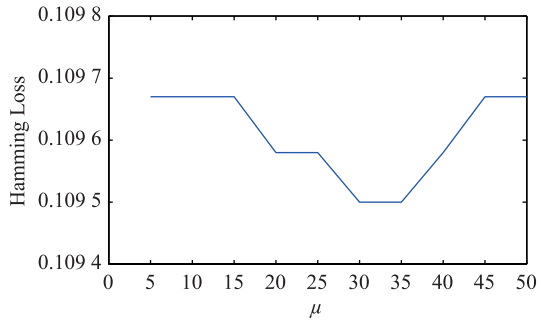
4.5 参数 μ 对分类效果的影响

MMD 正则项是本模型的关键成分，其对目标函数值的影响受参数 μ 控制. 为了讨论参数 μ 对算法分类效果的影响，本次实验令 μ 的取值以 5 为步长由 5 逐步增加到 50. 算法在“Corel5k vs ESPGame”数据集上的各项指标变化见图 1. 其中，参数 μ 取值的变化对所有指标均有不同程度的影响，具体表现如下：

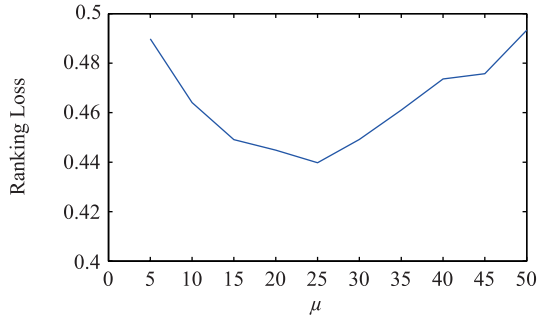
1）当 $\mu<25$ 时，随着 μ 的增大所有指标都有提升，这个阶段 μ 取值较小，MMD 正则项没有完全发挥作用；

2）当 $\mu>30$ 时，随着 μ 的增大所有指标都不断下降，这个阶段 μ 取值过大，使得目标函数中除 MMD 正则项的其它部分（如“特征—标记”关系矩阵分解）被忽视.

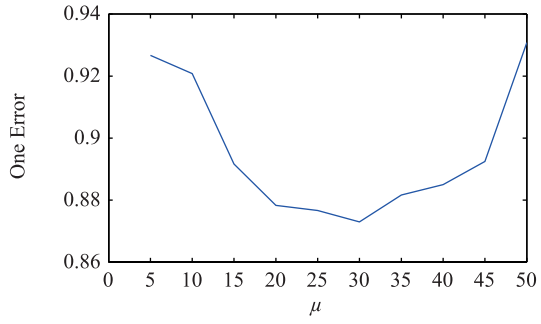
综上，在该数据集上，当 $\mu\in[25, 30]$ 时算法性能最佳. 综合考虑各评价指标，本文实验中可取 $\mu=30$.



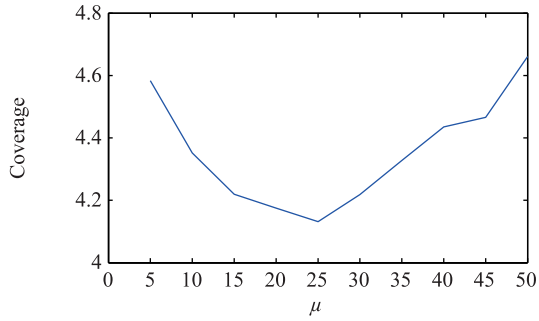
(a) Hamming Loss



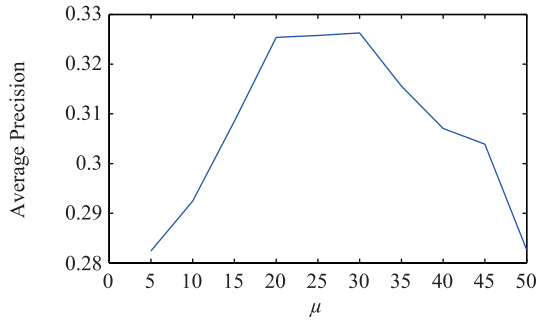
(b) Ranking Loss



(c) One Error



(d) Coverage



(e) Average Precision

图 1 参数 μ 对分类效果的影响

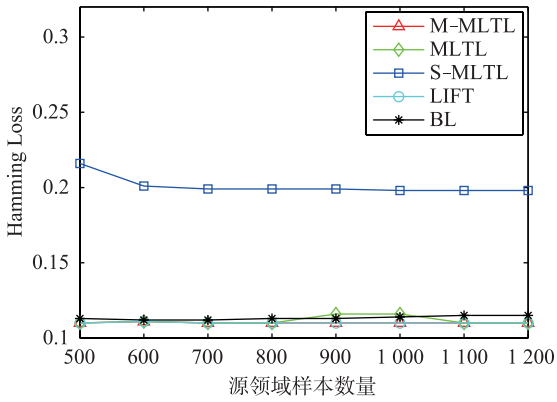
Fig.1 Influence on the classification results with different parameter μ

4.6 源领域训练样本数量对分类效果的影响

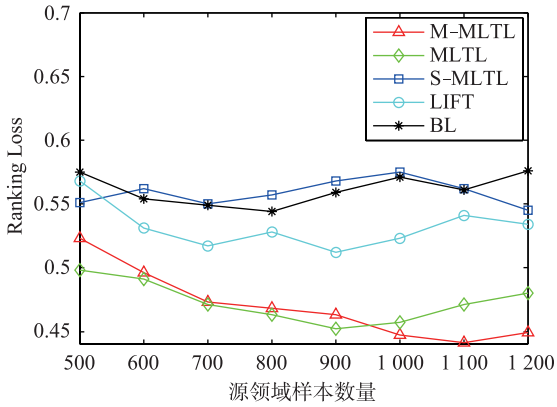
源领域样本数量是迁移学习算法效果的重要影响因素，因此当源领域样本数量增加时，算法分类效果的变化. 实验采用“Corel5k vs ESPGame”数据集，随机抽取 500 ~ 1 200个源领域样本组成训练数据集训练上述的 5 种多标记算法，并测试算法在目标领域的预测效果. 算法的 5 项评价指标随样本数量的变化情况见图 2. 如图所示，随着源领域样本数的增加，BL、LIFT 和 S-MLTL 的各项指标相对于其它 2 个算法较稳定；MLTL 的多数指标没有明显的变化，但 One Error 有增大的趋势，这是因为随着样本数量增加，源领域与目标领域特征的分布差异更明显，源领域的特征越来越不适用于目标领域；M-MLTL 的 Hamming Loss 没有明显变化，但其它指标均有不同程度的提升，这说明对源领域知识的学习提高了目标领域的预测效果. 值得注意的是，当训练样本数为 500 时 MLTL 算法的各项指标都优于 M-MLTL，这是由于本文中的 MMD 正则项采用的是线性核函数，即采用源域数据集和与目标域数据集的总体均值之差评价分布差异，而小数量样本可能存在选择性偏差（sample selection bias），不能代表整个源领域的分布情况，此时可以将控制参数 μ 的取值降低以提升算法的分类精度. 当样本数大于 1 000 时，M-MLTL 的各项指标都赶超 MLTL. 因此，M-MLTL 算法在大样本情况下更有优势.

4.7 算法耗时

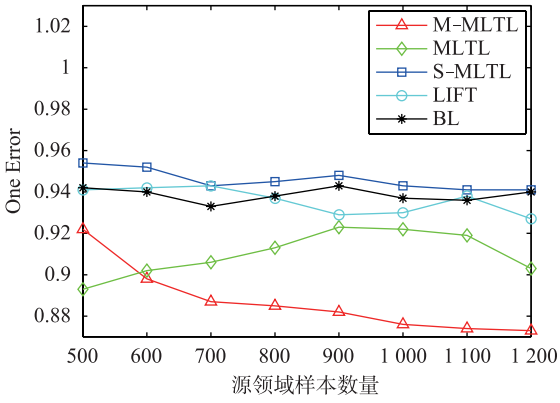
低维子空间特征数量和训练样本的个数都会影响算法的计算时间. 实验采用“Corel5k vs ESPGame”数据集，测试不同的潜在子空间特征数和训练样本个数影响下 3 种降维算法的计算时间.



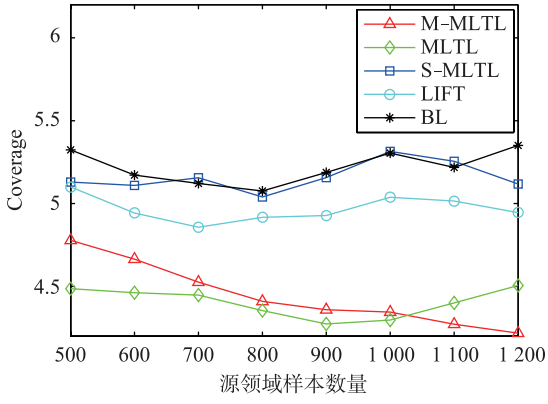
(a) Hamming Loss



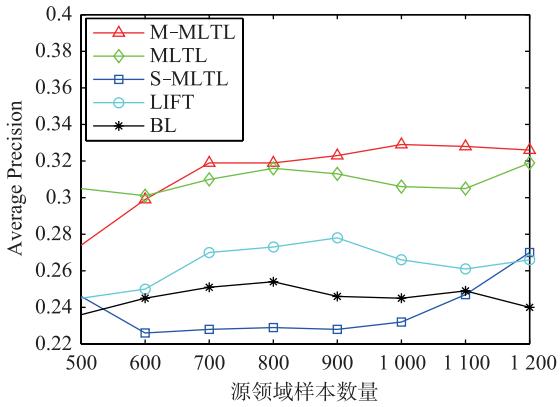
(b) Ranking Loss



(c) One Error



(d) Coverage



(e) Average Precision

图 2　源领域样本数量对分类效果的影响

Fig.2　Influence on the classification results with differentsource domain data number

M-MLTL、MLTL 和 S-MLTL 的计算时间随子空间特征数的变化情况见图 3. 3 种算法的计算时间都随子空间特征数增加而增长，其中 S-MLTL 的计算时间随子空间特征数增加阶梯上涨，MLTL 的计算时间变化与 S-MLTL 类似. 而 M-MLTL 计算时间与子空间特征数线性相关，这与 3.4 节的时间复杂度分析相符. 当子空间特征数小于 200 时 S-MLTL 和 MLTL 的计算耗时都小于 M-MLTL，但 M-MLTL 的计算时间随子空间特征数的增加变化相对较小，当子空间特征数大于 200 时 M-MLTL 计算时间小于其它两种算法.

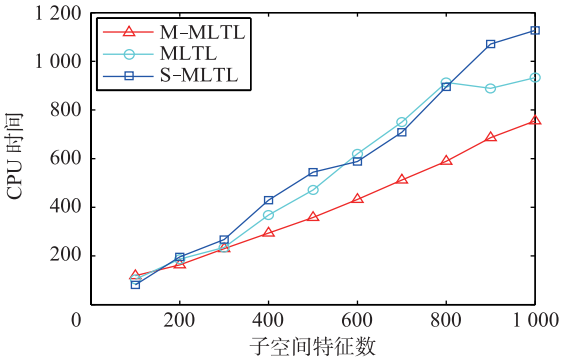


图 3　不同子空间特征数的计算时间

Fig.3　Computational time of different subspace feature number

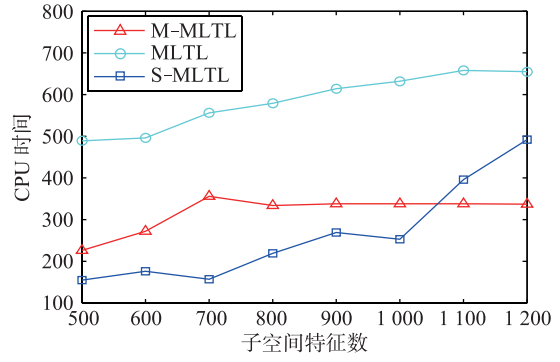


图 4　不同训练样本数量的计算时间

Fig.4　Computational time ofdifferent training data number

图 4 显示了算法计算时间随训练样本增加的变化情况. 由图可知, 当样本数量小于 700 时, 3 种算法的计算时间都随样本数增加而增长; 但当样本数量大于 700 时, M-MLTL 的计算时间不再受样本数量的影响, 而另 2 种算法的计算时间仍在持续增长, 这与 3.4 节的结论相符, 因此本文的算法在大样本学习中计算效率更有优势.

5 结论

本文提出了一种新多标记迁移学习算法, 通过分解"特征—标记"关系矩阵将源领域样本的标记间关联关系嵌入共享的潜在子空间, 同时引入最大均值差异最小化源领域与目标领域在共享子空间中的特征分布差异. 新算法着重解决了多标记迁移学习中共享子空间的领域适应问题. 在多标记图像分类实验中, 新算法比目前已有的同类算法有更好的分类效果和较高的计算效率. 同时, 新算法关键计算的耗时不受样本数量的影响, 因此新算法可以应对大样本的学习问题.

然而当训练样本数量较少时, 由于存在选择性偏差, 训练样本不能代表整个领域的分布情况, 新算法的优势不明显. 如何解决小样本多标记迁移中存在的选择性偏差问题将是下一步的研究方向.

参考文献

[1] 庄福振, 罗平, 何清, 等. 迁移学习研究进展[J]. 软件学报, 2015, 26(1): 26 – 39.
Zhuang F Z, Luo P, He Q, et al. Survey on transfer learning research[J]. Journal of Software, 2015, 26(1): 26 – 39.

[2] Shao L, Zhu F, Li X. Transfer learning for visual categorization: A survey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 26(5): 1019 – 1034.

[3] Jia C, Kong Y, Ding Z, et al. Latent tensor transfer learning for RGB-D action recognition[C]//Proceedings of the ACM International Conference on Multimedia. New York, NJ, USA: ACM, 2014: 87 – 96.

[4] 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.

[5] 许敏, 王士同, 顾鑫. TL-SVM: 一种迁移学习算法[J]. 控制与决策, 2014, 29(1): 141 – 146.
Xu M, Wang S T, Gu X. TL-SVM: A transfer learning algorithm[J]. Control and Decision, 2014, 29(1): 141 – 146.

[6] Zhuang F, Luo P, Xiong H, et al. Exploiting associations between word clusters and document classes for cross-domain text categorization[J]. Statistical Analysis and Data Mining: The ASA Data Science Journal, 2011, 4(1): 100 – 114.

[7] Dai W, Xue G R, Yang Q, et al. Co-clustering based classification for out-of-domain documents[C]//Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NJ, USA: ACM, 2007: 210 – 219.

[8] 董爱美, 王士同. 共享隐空间迁移 SVM[J]. 自动化学报, 2014, 40(10): 2276 – 2287.
Dong A M, Wang S T. A Shared latent subspace transfer learning algorithm using SVM[J]. Acta Automatica Sinica, 2014, 40(10): 2276 – 2287.

[9] 张倩, 李海港, 李明, 等. 基于马尔可夫逻辑网的关联规则迁移学习[J]. 信息与控制, 2014, 43(6): 715 – 721.
Zhang Q, Li H G, Li M, et al. Association rule transfer learning based on Markov logic network[J]. Information and Control, 2014, 43(6): 715 – 721.

[10] Zhou J T, Pan S J, Tsang I W, et al. Hybrid heterogeneous transfer learning through deep learning[C]//Twenty-Eighth AAAI Conference on Artificial Intelligence. Menlo Park, USA: AAAI, 2014: 2213 – 2219.

[11] Yang P, Gao W. Multi-view discriminant transfer learning[C]//Proceedings of the Twenty-Third international joint conference on Artificial Intelligence. Menlo Park, USA: AAAI, 2013: 1848 – 1854.

[12] Zhang M L, Zhou Z H. A review on multi-label learning algorithms[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819 – 1837.

[13] Mei S. Predicting plant protein subcellular multi-localization by Chou's PseAAC formulation based multi-label homolog knowledge transfer learning[J]. Journal of Theoretical Biology, 2012, 310: 80 – 87.

[14] Tawiah C, Sheng V. Empirical comparison of multi-label classification algorithms[C]//Twenty-Seventh AAAI Conference on Artificial Intelligence. Menlo Park, USA: AAAI, 2013: 1645 – 1646.

[15] Cheng B, Liu M, Zhang D. Multimodal multi-label transfer learning for early diagnosis of Alzheimer's disease[M]//Machine Learning in Medical Imaging. Berlin, Germany: Springer-Verlag, 2015: 238 – 245.

[16] Fu Y, Yang Y, Hospedales T, et al. Transductive multi-label zero-shot learning[C]//British Machine Vision Conference. Durham, UK: BMVA Press, 2015.

[17] Han Y, Wu F, Zhuang Y, et al. Multi-label transfer learning with sparse representation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(8): 1110 – 1121.

[18] Gretton A, Borgwardt K M, Rasch M, et al. A kernel method for the two-sample-problem[C]//Advances in neural information processing systems. New York, NJ, USA: ACM, 2006: 513 – 520.

[19] Pan S J, Kwok J T, Yang Q. Transfer Learning via Dimensionality Reduction[C]//Twenty-Third Conference on Artificial Intelligence. Menlo Park, USA: AAAI, 2008: 677 – 682.

(下转第 478 页)

[21] Chen Y X, Bi J B, Wang J Z. MILES: Multiple-instance learning via embedded instance selection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006: 1931 – 1947.
[22] Wei X S, Wu J, Zhou Z H. Scalable multi-instance learning[C]//14th IEEE International Conference on Data Mining (ICDM'14). Piscataway, NJ, USA: IEEE, 2014: 1037 – 1042.
[23] Zhang W J, Zhou Z H. Multi-instance learning with distribution change[C]//28th AAAI Conference on Artificial Intelligence (AAAI'14). San Francisco, CA, USA: AAAI, 2014: 2184 – 2190.
[24] Wu J S, Huang S J, Zhou Z H. Genome-wide protein function prediction through multi-instance multi-label learning[J]. ACM/IEEE Transactions on Computational Biology and Bioinformatics, 2014, 11(5): 891 – 902.
[25] Zhou Z H, Zhang M L, Huang S J, et al. Multi-instance multi-label learning[J]. Artificial Intelligence, 2012, 176(1): 2291 – 2320.
[26] 雷小锋, 谢昆青, 林帆, 等. 一种基于 K-Means 局部最优性的高效聚类算法[J]. 软件学报, 2008, 19: 1683 – 1692.
Lei X F, Xie K Q, Lin F, et al. An efficient clustering algorithm based on local optimality of K-means[J]. Journal of software, 2008, 19(7): 1683 – 1692.
[27] Zhu Q, Fan X, Feng J. Outlier detection based on k-neighborhood MST[C]//2014 IEEE International Conference on Information Reuse and Integration (IRI). Piscataway, NJ, USA: IEEE, 2014: 718 – 724.

作者简介

钱景辉(1978 –), 男, 硕士, 讲师. 研究领域为机器学习, 数据挖掘和工业过程优化.
窦立阳(1991 –), 男, 硕士生. 研究领域为机器学习方法在工业过程中的应用.
李荣雨(1977 –), 男, 博士, 副教授. 研究领域为工业过程的优化与监控.

(上接第 470 页)

[20] Long M, Wang J, Ding G, et al. Adaptation regularization: A general framework for transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(5): 1076 – 1089.
[21] Yang Y, Zha Z J, Gao Y, et al. Exploiting web images for semantic video indexing via robust sample-specific loss[J]. IEEE Transactions on Multimedia, 2014, 16(6): 1677 – 1689.
[22] Dumais S T. Latent semantic analysis[J]. Annual review of information science and technology, 2004, 38(1): 188 – 230.
[23] Hestenes M R, Stiefel E. Methods of conjugate gradients for solving linear systems[J]. Journal of Research of the National Bureau of Standards, 1952, 49(6): 409 – 436.
[24] Bartels R H, Stewart G W. Solution of the matrix equation $AX + XB = C$[J]. Communications of ACM, 1972, 15(9): 820 – 826.
[25] Duygulu P, Barnard K, deFreitas J F G, et al. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary [M]//Lecture Notes in Computer Science: vol. 2353. Berlin, Germany: Springer-Verlag, 2002: 97 – 112.
[26] Von Ahn L, Dabbish L. Labeling images with a computer game[C]//Proceedings of the SIGCHI conference on Human factors in computing systems. New York, NJ, USA: ACM, 2004: 319 – 326.
[27] Guillaumin M, Mensink T, Verbeek J, et al. Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation [C]//2009 IEEE 12th International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2009: 309 – 316.
[28] Steinwart I. On the influence of the Kernel on the consistency of support vector machines[J]. Journal of Machine Learning Research, 2002, 2(1): 67 – 93.
[29] Elisseeff A, Weston J. A kernel method for multi-labelled classification[C]//Advances in neural information processing systems. Cambridge, Massachusetts, USA: MIT Press, 2001: 681 – 687.
[30] Zhang M L, Wu L. LIFT: Multi-label learning with label-specific features[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(1): 107 – 120.

作者简介

姜海燕(1967 –), 女, 博士, 教授. 研究领域为智能决策支持系统, 数字植物.
刘昊天(1990 –), 男, 硕士生. 研究领域为机器学习与并行计算.
舒　欣(1984 –), 男, 博士, 讲师. 研究领域为机器学习和人工智能.