

基于异质信息融合的网络图像半监督学习方法

杜友田¹ 李谦¹ 周亚东¹ 吴陈鹤¹

摘要 网络图像通常包含文本、颜色和纹理等异质信息. 本文提出了一种基于多类异质信息融合的网络图像半监督学习方法——局部协同训练 (Local co-training, LCT). 该方法在每个视图 (对应一类信息) 上对每个样本点的邻域构建线性局部模型, 利用一组局部模型来表示数据关系; 基于信息传播和协同训练对模型进行增量式迭代更新. 该算法在协同训练和基于图正则化的方法这两类半监督学习算法间建立了桥梁. 局部协同训练算法能够准确地描述样本的复杂分布, 并且可以进行高效的增量学习, 有利于大规模网络图像的在线学习. 在 Corel, Pascal 和 ImageNet 数据集上的实验结果表明该方法具有良好的性能.

关键词 网络图像分类, 异质信息, 局部协同训练, 机器学习

引用格式 杜友田, 李谦, 周亚东, 吴陈鹤. 基于异质信息融合的网络图像半监督学习方法. 自动化学报, 2012, 38(12): 1923–1932

DOI 10.3724/SP.J.1004.2012.01923

Web Image Semi-supervised Learning Method Based on Heterogeneous Information Fusion

DU You-Tian¹ LI Qian¹ ZHOU Ya-Dong¹ WU Chen-He¹

Abstract Web images generally consist of heterogeneous information including texts, colors and textures. This paper proposes a new method, called local co-training (LCT), for semi-supervised classification of web images based on fusion of heterogeneous information. The proposed method employs a set of local linear models to represent data points of each view, and incrementally refines these models by exploiting unlabeled data with information propagation and co-training. The local co-training builds a bridge between graph-based methods and co-training. The local co-training can model the instance distribution accurately in the high-dimensional space, and learn local models incrementally, which benefits the online classification of large scale of web images. Experiments on Corel, Pascal and ImageNet datasets demonstrate that the local co-training can effectively improve the classification performance of learners by exploiting multiple attribute sets and unlabeled data.

Key words Web image classification, heterogeneous information, local co-training (LCT), machine learning

Citation Du You-Tian, Li Qian, Zhou Ya-Dong, Wu Chen-He. Web image semi-supervised learning method based on heterogeneous information fusion. *Acta Automatica Sinica*, 2012, 38(12): 1923–1932

对互联网上涌现的海量多媒体数据进行学习和理解已成为备受关注的研究问题. 网络多媒体数据有三个特点: 1) 规模巨大; 2) 多种异质数据耦合, 如文本、图像及音视频等; 3) 数据内容分散. 因此,

对网络多媒体数据进行自动理解具有很大的挑战性. 本文主要针对网络图像进行研究.

底层视觉特征是一种广泛使用的图像表示形式, 其中包括颜色、纹理、角点等. 以往工作通常采用多种视觉信息融合的方法来解决图像理解问题, 如前融合 (Early fusion) 和后融合 (Late fusion)^[1]. 对于互联网上大规模涌现的图像而言, 除了视觉特征本身, 还通常附有标题、评论、注释、类别标签以及用户名等文本信息^[2]. 视觉信息和文本信息结构不同, 表征了图像数据的不同属性并相互补充: 前者侧重于表象, 后者侧重于语义. 如何有效融合这些异质信息对网络图像进行半监督学习具有重要意义.

多视图半监督学习将目标对象的多个特征集合进行融合, 并利用未标记样本提高学习性能, 是近年来备受关注的学习范式^[3]. 该学习范式主要包含两类算法: 1) 协同训练 (Co-training) 算法^[4–8]; 2) 基于图正则化 (Regularization) 的算法^[9]. 协同训练算

收稿日期 2012-03-15 录用日期 2012-06-27
Manuscript received March 15, 2012; accepted June 27, 2012
国家自然科学基金 (60905018, 61202392, 60921003, 61175039), 十二五国家科技支撑重点项目 (2011BAK08B02), 教育部博士点基金 (20090201120032), 中央高校基本科研业务费 (xjj2009041, 2012jdhz08) 资助
Supported by National Natural Science Foundation of China (60905018, 61202392, 60921003, 61175039), Key Projects in the National Science and Technology Pillar Program (2011BAK08B02), Research Fund for Doctoral Program of Higher Education (20090201120032), and Fundamental Research Funds for the Central Universities (xjj2009041, 2012jdhz08)

本文责任编辑 刘成林
Recommended by Associate Editor LIU Cheng-Lin
1. 西安交通大学智能网络与网络安全教育部重点实验室 西安 710049
1. Key Laboratory for Intelligent Networks and Network Security of Ministry of Education, Xi'an Jiaotong University, Xi'an 710049

法最初假设数据包两个充分冗余的视图, 基于每个视图构造一个学习器, 在学习过程中分别选择置信度高的未标记示例进行相互标记, 从而扩大训练集并使模型得到更新^[4]. 此外, 还有其他协同训练算法采用结构或参数不同的多个学习器对同一数据集进行互相学习, 克服了某些数据对象只具有单一视图的问题^[10]. 协同训练算法通过不断增加训练数据来迭代地提高学习器性能, 适合于网络数据的在线分析. 基于图正则化的方法^[3, 8, 11], 根据训练样本和某种相似度量建立一个图 (Graph), 利用决策函数在图上的光滑性作为目标函数的正则化项来求取最优模型参数. 目前, 利用图正则化的方法进行多视图学习的研究也开始出现, 多采用增加正则化项来保证多个视图上结果一致的方法, 如 Co-regularization^[12-14]. 除以上两类算法外, 还有多核学习的方法等^[15]. 通常, 样本在高维空间呈现复杂的分布, 采用单一的全局模型对样本分布进行准确建模比较困难. 图上的学习 (或局部学习) 考虑了相邻样本点之间的关系, 能够较为准确地描述数据分布. Bottou 等^[16] 指出, 局部学习方法通常优于全局学习方法. 通过以上分析可知两类学习算法各有优缺点. 对于大规模网络异构数据分析来说, 将局部学习用于迭代的在线多视图半监督学习具有重要价值.

本文以局部学习和协同训练为基础, 提出了一种基于局部协同训练 (Local co-training, LCT) 的网络图像半监督学习算法. 该方法将网络图像包含的多种异质数据看作多个视图; 在每个视图上对每个数据点的邻域构建一个线性局部模型, 并采用一组局部模型来表示每个视图上的数据关系; 然后基于图上的信息传播理论和协同训练策略对模型进行增量式迭代更新. 该方法在协同训练和基于图正则化的方法之间建立了桥梁. 局部协同训练方法具有如下优点: 1) 比较准确地描述了样本的复杂分布; 2) 可以进行高效的增量学习; 3) 适合于大规模异构数据的在线学习. 本文的主要工作归纳如下:

1) 建立微分方程, 定量描述了局部学习中类别标签传播的衰减性规律: 类别标签只在标记样本的局部区域内进行有效传播;

2) 在标签传播的衰减性规律基础上, 给出单一视图上的类别标签增量预测算法: 在已经完成预测的图上加入新的标记样本, 则新的标签预测可以通过增量学习得到;

3) 提出针对异质数据学习的局部协同训练算法, 该算法在每次迭代中只需要更新少量的主导局部模型, 并验证了该算法在网络图像在线半监督学习中的性能.

1 局部协同训练算法

给定 l 个标记样本 $\mathcal{L} = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots,$

$(\mathbf{x}_l, \mathbf{y}_l)\} \subset X \times Y$ 和 u 个未标记样本 $\mathcal{U} = \{\mathbf{x}_{l+1}, \mathbf{x}_{l+2}, \dots, \mathbf{x}_{l+u}\} \subset X$, 其中, $\mathbf{x}_i \in X \subset \mathbf{R}^d$ ($1 \leq i \leq u+l$) 是 d 维输入样本, 为输入空间; $\mathbf{y}_i = [y_{ic}]_{1 \times C} \in \mathbf{R}^C$ ($1 \leq i \leq l$) 是类别标签向量, 若 \mathbf{x}_i 属于第 c 类, $y_{ic} = 1, y_{ic'} = 0$ ($c' \neq c$). 在多视图问题中, X 分解成多个子空间, 即 $X = X^{(1)} \times X^{(2)} \times \dots \times X^{(V)}$, $\mathbf{x} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(V)})$, 其中 $\mathbf{x}^{(v)} \in X^{(v)} \subset \mathbf{R}^{d_v}$, $v = 1, 2, \dots, V$.

1.1 单一视图上的局部模型构建

假设样本 $\{\mathbf{x}_i\}_{i=1}^{l+u}$ 中每个样本点的邻域满足线性模型^[11]: $o_i(\mathbf{x}) = \mathbf{w}_i^T(\mathbf{x} - \mathbf{x}_i) + b_i$, 其中 $o_i(\cdot)$ 表示该模型的输出函数, 在两类问题中为标量; \mathbf{x} 是 \mathbf{x}_i 的邻域点, 即 $\mathbf{x} \in N(\mathbf{x}_i)$. 本文将上式推广至向量输出函数并将其用于多类问题. 在视图 $X^{(v)}$ 上, 样本 $\mathbf{x}_i^{(v)}$ 对应的局部线性模型表示如下:

$$\mathbf{o}_i^{(v)}(\mathbf{x}^{(v)}) = W_i^{(v)T}(\mathbf{x}^{(v)} - \mathbf{x}_i^{(v)}) + \mathbf{b}_i^{(v)} \quad (1)$$

其中, $W_i^{(v)}$ 和 $\mathbf{b}_i^{(v)}$ 为参数; $\mathbf{x}^{(v)}$ 是 $\mathbf{x}_i^{(v)}$ 的邻域点, 即 $\mathbf{x}^{(v)} \in N_v(\mathbf{x}_i^{(v)})$. 本节及第 1.2 节和第 1.3 节都是针对单个视图 $X^{(v)}$ 分析的, 故将变量上标或下标 v 去掉以简化表述. 式 (1) 中的参数通过求解如下优化问题来获得:

$$\min_{W_i \in \mathbf{R}^{d \times C}, \mathbf{b}_i \in \mathbf{R}^C} \lambda \|W_i\|_{\Omega}^2 + \sum_{\mathbf{x}_j \in N(\mathbf{x}_i)} \|W_i^T(\mathbf{x}_j - \mathbf{x}_i) + \mathbf{b}_i - \mathbf{f}_j\|_2^2 \quad (2)$$

其中, $\|W_i\|_{\Omega} = [\text{tr}(W_i^T W_i)]^{\frac{1}{2}}$, \mathbf{f}_j 是 \mathbf{x}_j 的实值预测标签向量, λ 为参数. 综合式 (1) 和式 (2) 可得:

$$\begin{aligned} \mathbf{o}_i(\mathbf{x}_i) &= [\alpha_i F_i]^T \\ \alpha_i &= \frac{\mathbf{1}^T - \mathbf{1}^T Q_i (\lambda I + Q_i)^{-1}}{n_i - \mathbf{1}^T Q_i (\lambda I + Q_i)^{-1} \mathbf{1}} \end{aligned} \quad (3)$$

其中, $F_i = [f_j]^T \in \mathbf{R}^{n_i \times C}$, $Q_i = X_i^T X_i$, $X_i = [\mathbf{x}_j - \mathbf{x}_i] \in \mathbf{R}^{d \times n_i}$, I 为单位矩阵, $\mathbf{1}$ 是分量均为 1 的列向量, n_i 是 \mathbf{x}_i 的邻域点个数.

1.2 类别标签传播规律分析

定义如下目标函数:

$$Q(F) = \sum_c \left(\sum_i \left((f_{ic} - o_{ic})^2 + e_i (f_{ic} - y_{ic})^2 \right) \right) \quad (4)$$

其中, f_{ic} , o_{ic} 和 y_{ic} 分别是 \mathbf{x}_i 的预测标签向量 \mathbf{f}_i , 模型输出 \mathbf{o}_i 和真实标签向量 \mathbf{y}_i 对应于类别 c 的分量, $F = [\mathbf{f}_i]_{n \times 1} = [f_{ic}]_{n \times C}$ 为预测标签矩阵; e_i 是常数, 当 $1 \leq i \leq l$ 时, $e_i = E_l > 0$; $l+1 \leq i \leq n$

($n = l + u$) 时, $e_i = E_u > 0$. 最小化式 (4) 可得:

$$\begin{aligned} F &= (M + E)^{-1} E Y \\ \tilde{\mathbf{f}}_c &= (M + E)^{-1} E \tilde{\mathbf{y}}_c \end{aligned} \quad (5)$$

其中, $Y = [y_{ic}]_{n \times C}$, $\tilde{\mathbf{f}}_c$ 和 $\tilde{\mathbf{y}}_c$ 分别为所有样本的预测标签向量和真实标签向量中第 c 个分量构成的列向量; $E = \text{diag}\{e_1, e_2, \dots, e_n\}$; $M = (I - A)^T (I - A) \in \mathbf{R}^{n \times n}$, 其中相似度矩阵 $A = [a_{ij}]_{n \times n}$ 定义如下: 若 $\mathbf{x}_j \in N(\mathbf{x}_i)$, 则 a_{ij} 等于 α_i 中对应元素的值, 否则 $a_{ij} = 0$. $\tilde{\mathbf{y}}_c$ 可作如下分解: $\tilde{\mathbf{y}}_c = \sum_{k: y_{kc}=1} \tilde{\mathbf{y}}_c^k$, 其中 $\tilde{\mathbf{y}}_c^k$ 是第 k 个分量为 1、其他分量为 0 的列向量. 因此, 预测结果 $\tilde{\mathbf{f}}_c$ 为第 c 类中的标记样本在整个样本 \mathbf{x}_k 空间上的预测结果 $\tilde{\mathbf{f}}_c^k$ 的线性叠加.

首先分析第 c 类标记样本 \mathbf{x}_k 引起的标签传播结果. 将式 (3) 代入式 (4), 并对 f_{ic}^k 求导来最小化 $Q(F)$:

$$\begin{aligned} f_{ic}^k - \sum_j a_{ij}^k f_{jc}^k - \sum_m \left(f_{mc}^k - \sum_{m_q} a_{mm_q}^k f_{m_q c}^k \right) \times \\ a_{mi}^k + E_u f_{ic}^k = 0 \end{aligned} \quad (6)$$

其中, f_{ic}^k 表示由 \mathbf{x}_k 引起的未标记样本 \mathbf{x}_i 的标签预测值. 假设在 \mathbf{x}_i 周围的局部区域内样本的分布近似均匀, $\sum_m a_{mi}^k \approx 1$. 定义 $(\Delta_A f_c^k)(\mathbf{x}_i) = \sum_{j: \mathbf{x}_j \in N(\mathbf{x}_i)} a_{ij} f_{jc}^k - f_{ic}^k$, 式 (6) 可简写为

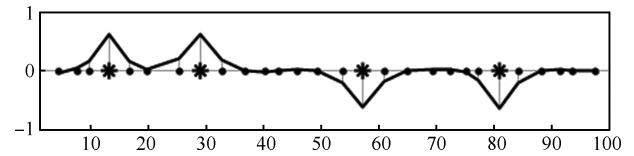
$$(\Delta_A^2 f_c^k)(\mathbf{x}_i) + E_u f_{ic}^k = 0 \quad (7)$$

其中, $\Delta_A^2 = \Delta_A \circ \Delta_A$ 是关于矩阵 A 的离散 bi-Laplacian 算子. 当样本较多时, 离散 bi-Laplacian 算子可用连续算子来近似刻画^[17]. 由于标签传播结果与样本分布相关, 其规律的准确刻画非常困难. 本文分析在样本近似为均匀分布的情况下的传播规律. 假设存在一个光滑函数 $\varphi_c^k: \mathbf{R}^d \rightarrow \mathbf{R}$, 使得 $\varphi_c^k(\mathbf{x}_i) = f_{ic}^k$, 则可通过下式来分析径向上的标签传播规律:

$$\begin{aligned} E_\rho (\Delta^2 \varphi^k) + E_u \varphi^k = \\ E_\rho \left(\frac{\partial^4 \varphi^k}{\partial r^4} + a_3 r^{-1} \frac{\partial^3 \varphi^k}{\partial r^3} + a_2 r^{-2} \frac{\partial^2 \varphi^k}{\partial r^2} + \right. \\ \left. a_1 r^{-3} \frac{\partial \varphi^k}{\partial r} \right) + E_u \varphi^k = 0 \end{aligned} \quad (8)$$

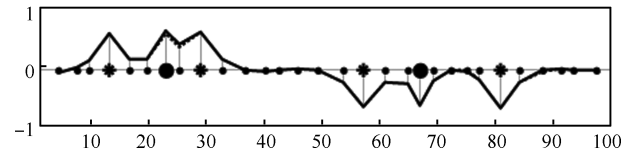
其中, Δ^2 为连续的 bi-Laplacian 算子, E_ρ 为常数, $a_1 = -(d-1)(d-3)$, $a_2 = (d-1)(d-3)$, $a_3 = 2(d-1)$. 式 (8) 的封闭解难以求得, 但通过分析可得到符合半监督学习的可行解性质, 如 $\lim_{r \rightarrow \infty} \varphi^k = 0$, $\lim_{r \rightarrow \infty} \frac{\partial^4 \varphi^k}{\partial r^4} = 0$. 特别地, 当 $d = 1$ 时, 一个可行解为 $\varphi^k(\mathbf{x}_i) = \exp(-E_\Delta r)(\cos E_\Delta r + i \sin E_\Delta r)$, 即

由 \mathbf{x}_k 引起的标签传播随着距离 r 增大而呈指数衰减, 其中 $E_\Delta = \sqrt[4]{E_u / (4E_\rho)}$, $r = \|\mathbf{x} - \mathbf{x}_k\|_2$. 通过以上定量分析, 我们得到如下结论: 在第 1.1 节描述的局部学习中, 当加入一个新的标记样本时, 标签信息主要在以其为中心的局部域内进行有效传播, 该局部域大小与 E_l 和 E_u 有关. 该性质对于迭代的增量在线学习具有重要意义. 图 1(a) 验证了在 1 维数据上的标签传播随着距离增大而衰减的规律. 尽管实际样本在高维空间中分布不一定均匀和稠密, 但是标签传播的衰减性质仍然成立, 在实验部分我们将验证在 ImageNet 数据上标签传播的衰减规律.



(a) t 时刻的标签预测结果

(a) Prediction results at time t



(b) $t + 1$ 时刻加入新的样本时的标签预测结果

(b) Prediction results when new labeled data

are added at time $t + 1$

图 1 在 1 维样本上的标签估计结果 (小黑点、星号和大黑点分别表示未标记样本、已标记样本和新加入的已标记样本, 图 1(b) 中实线和虚线分别表示通过本文的增量算法和根据式 (5) 在整个样本集上重新计算得到的标签预测结果)

Fig.1 The label prediction results on a 1-dimensional dataset (The small points, stars, and large points denote unlabeled data, labeled data, and new joining data, respectively. In Fig.1(b), the solid line denotes the results of the incremental learning algorithm, and dashed line denotes the results directly computed with (5).)

1.3 样本类别标签的增量预测

假定在 t 时刻数据表示为图 $\langle \mathcal{X}_t, A_t \rangle$ 表示样本集合, 在该图上的预测结果为 $F_t = [\tilde{\mathbf{f}}_{1,t}, \tilde{\mathbf{f}}_{2,t}, \dots, \tilde{\mathbf{f}}_{C,t}]$. 在 $t + 1$ 时刻加入一个标记样本 $(\mathbf{x}_i, \mathbf{y}_i)$, 则 $\mathcal{X}_{t+1} = \mathcal{X}_t \cup \mathbf{x}_i$ 上的预测结果为 F_{t+1} 可按如下求解. 首先在 \mathcal{X}_t 中寻找 \mathbf{x}_i 的 K 个邻域样本, 并根据式 (3) 计算 α_i . 由于 \mathbf{x}_i 的加入使得 \mathcal{X}_{t+1} 上的邻域关系发生了变化, 故 $t + 1$ 时刻的相似度矩阵需重新计算. 一种办法是将 α_i 中的相应元素跟 A_t 直接组合, 构成新的相似度矩阵, 记为 \hat{A}_{t+1} . \hat{A}_{t+1} 在对于 \mathbf{x}_i 附近样本点的邻域关系描述不够精确. 本文采用如下办法: 求解以 \mathbf{x}_i 为中心的局部区域 $LR(\mathbf{x}_i)$ 内的

每个样本 \mathbf{x}_j 对应的 $\alpha_j, LR(\mathbf{x}_i)$ 之外的相似度较 A_t 的对应元素不变, 最后将 $\{\alpha_j\}$ 和 A_t 组合得到矩阵 A_{t+1} . 显然 A_{t+1} 比 \hat{A}_{t+1} 更准确. 通常, \hat{A}_{t+1} 和 A_t 比较接近, 即 $\|A_t - \hat{A}_{t+1}\|$ 较小. 计算从 K 个邻域点线性传播至 \mathbf{x}_i 的标签值: $\mathbf{g}_i = \alpha_i F_{i,t}$, 其中 $F_{i,t}$ 为 t 时刻 \mathbf{x}_i 邻域样本的标签预测向量构成的矩阵. 根据式 (5), $t+1$ 时刻在 $\langle \mathcal{X}_{t+1}, A_{t+1} \rangle$ 上的预测结果为

$$F_{t+1} = (M_{t+1} + E_{t+1})^{-1} E_{t+1} [Y_t; \mathbf{y}_i] = F'_{t+1} + F''_{t+1} \quad (9)$$

其中

$$F'_{t+1} = (M_{t+1} + E_{t+1})^{-1} E_{t+1} [Y_t; \mathbf{g}_i] \quad (10)$$

$$F''_{t+1} = (M_{t+1} + E_{t+1})^{-1} E_{t+1} [\mathbf{0}; \mathbf{y}_i - \mathbf{g}_i] \quad (11)$$

考虑另外一个图 $\langle \mathcal{X}_{t+1}, \hat{A}_{t+1} \rangle$, 令 $\hat{F}'_{t+1} = (\hat{M}_{t+1} + E_{t+1})^{-1} E_{t+1} [Y_t; \mathbf{g}_i]$, 由于 \mathbf{x}_i 的加入没有改变该图中其他样本的邻域信息, 故 $\hat{F}'_{t+1} = [F_t; \mathbf{g}_i]$. 由前面所知 \hat{A}_{t+1} 和 A_t 接近, 根据式 (10) 可得到如下近似: $F'_{t+1} \approx \hat{F}'_{t+1} = [F_t; \mathbf{g}_i]$. 根据类别样本传播具有衰减性, $(\mathbf{x}_i, \mathbf{y}_i)$ 加入而引起预测结果的变化主要发生以 \mathbf{x}_i 为中心的局部区域 $LR'(\mathbf{x}_i)$ 内, 故 F''_{t+1} 只需在此局部区域内计算即可. 在实际计算中, 可以令 $LR'(\mathbf{x}_i)$ 和 $LR(\mathbf{x}_i)$ 相同. 在第 3.2 节通过实验估计了 $LR(\mathbf{x}_i)$ 的大小.

综上, 在该标签增量预测算法中, $t+1$ 时刻所有样本的标签传播结果由两部分叠加构成: 1) t 时刻的传播结果 (增加 \mathbf{x}_i 的线性预测结果 \mathbf{g}_i); 2) 由 $\mathbf{y}_i - \mathbf{g}_i$ 在局部区域内产生的类别标签增量 F''_{t+1} . 图 1(b) 显示了在 1 维数据上该增量算法得到的结果与直接采用式 (5) 得到的真实结果非常接近. 此外, 通过实验我们也发现在实际的网络图像数据中二者也能够很好的吻合.

1.4 多视图上局部模型的协同学习

基于第 v 个视图上构建的局部模型集合诱导出对应的分类器 $h^{(v)}$ ($v = 1, 2, \dots, V$). 根据文献 [4] 可知, 若视图之间满足充分冗余假设, 协同训练可以使得弱学习器性能得到任意提高. 在大多不严格满足该假设的实际问题中, 该策略也通常能得到较好的结果. 本节分析如何对多个视图上的局部模型进行协同学习并提高其分类性能, 算法 1 详细描述了本文提出的局部协同训练算法.

算法 1. 局部协同训练算法

$[h^{(1)}, h^{(2)}, \dots, h^{(V)}] = \text{LCT}(\mathcal{L}, \mathcal{U}, \mathcal{U}_p, T, K)$

输入. 初始分类器 $h_0^{(1)}, h_0^{(2)}, \dots, h_0^{(V)}$, 已标记数据 \mathcal{L} , 未标记数据 \mathcal{U} 和 \mathcal{U}_p , 邻域大小 K , 迭代次数 T .

输出. 分类器 $h_T^{(1)}, h_T^{(2)}, \dots, h_T^{(V)}$, 标签预测结果 F .

初始化. $\mathcal{X}_0 = \mathcal{L} \cup \mathcal{U}$.

步骤 1. \mathcal{U}_p 中随机选取 u' 个未标记样本构成集合 \mathcal{U}'_p ;

步骤 2. for $t = 1, 2, \dots, T$ do

步骤 3. 根据式 (5) 用分类器 $h_t^{(v)}$ 来对 \mathcal{U}'_p 中的样本标签进行预测, 从中选择 s 个样本 X_s 添加至 \mathcal{X}_t , 其中从预测为第 c 类的样本中选择 m_{cv} 个, $s = \sum_{c=1}^C \sum_{v=1}^V m_{cv}$, $\mathcal{X}_{t+1} = \mathcal{X}_t \cup X_s$, 计算 \mathbf{A}_{t+1} ;

步骤 4. 寻找分类器 $h_t^{(v)}$ 中对应于 X_s 的主导局部模型 DLM, 通过增量预测算法计算 $\langle \mathcal{X}_{t+1}^{(v)}, \mathbf{A}_{t+1}^{(v)} \rangle$ ($v = 1, 2, \dots, V$) 上的类别标签预测值 $\mathbf{F}_t^{(v)}$;

步骤 5. 从 \mathcal{U}_p 中随机选择 s 个样本并将其移入 \mathcal{U}'_p ;

步骤 6. end for

步骤 7. 根据式 $\mathbf{F} = \sum_{v=1}^V \beta_v \mathbf{F}_T^{(v)}$ 来融合多个视图上的分类器, β_v 为融合系数.

在算法 1 中, $\mathcal{L} \cup \mathcal{U}$ 是训练集, 额外定义一个未标记数据集 \mathcal{U}_p , 其作用是在算法执行过程中提供“好”的样本来扩充训练集. 首先, 在 t 时刻根据图 $\langle \mathcal{X}_t^{(v)}, A_t^{(v)} \rangle$ 及其标签传播结果 $F_t^{(v)}$ 来预测 $\mathcal{U}'_p \subset \mathcal{U}_p$ 中样本的标签值. 具体做法是在 $\mathcal{X}_t \cup \mathcal{U}'_p$ 上根据第 1.3 节介绍的增量预测算法进行计算 \mathcal{U}'_p 中样本的标签预测结果. 在实际的在线分类问题中, 该预测环节与分类目的一致, 故不产生额外的计算复杂度.

可以认为, 标签预测值 $f_{ic}^{(v)}$ 越大, $\mathbf{x}_i \in \mathcal{U}'_p$ 在第 v 个视图上被判断为 c 类样本的置信度也越高. 另一方面, 若 $f_{ic}^{(v)}$ 值过大 (接近于 1), 则在不发生错分的情况下 $\mathbf{f}_i^{(v)} = [f_{i1}^{(v)}, f_{i2}^{(v)}, \dots, f_{iC}^{(v)}]$ 越接近实际的标签向量 $\mathbf{y}_i = [0, \dots, 0, 1, 0, \dots, 0]$, 即 $\mathbf{g}_i^{(v)}$ 也越接近 \mathbf{y}_i (由于 $A_{t+1}^{(v)}$ 和 $\hat{A}_{t+1}^{(v)}$ 很接近). 根据式 (11) 可知, \mathbf{x}_i 的加入对 t 时刻的标签预测结果改变轻微, 即对模型性能的改善帮助很小. 故未标记样本的选取需要同时考虑这两个方面. 在实验分析基础上, 局部协同训练算法按照如下方式选择 \mathcal{U}'_p 中的样本: 将 $f_{ic}^{(v)}$ 从大到小排序, 取不在前 5% 中的前 m_{cv} 个样本作为第 c 类样本加入至训练集, $c = 1, 2, \dots, C$, $v = 1, 2, \dots, V$. 增加的样本当作标记样本用于第 1.3 节中类别标签的增量预测算法.

在局部协同训练算法中, 引入如下定义: 集合 $X_\alpha = \{\mathbf{x} | \mathbf{x} \in X_s \text{ or } N(\mathbf{x}) \cap X_s \neq \emptyset\}$ 中每个样本对应的局部模型称为主导局部模型 (Dominant local model, DLM), 其中 X_s 表示每次迭代中新加入的样本. 假定 $n_i^{(v)} = n_i$, 根据式 (3) 可知 $\alpha_i^{(v)}$ 求解的时间复杂度为 $O(n_i^3)$. 在每次迭代中, 只有 X_α 中的样本因为 X_s 的加入而改变了邻域关系, 其对应的 $\alpha_i^{(v)}$ 需要重新计算, 即只有 DLMs 才需要更新. 因此, 在每次迭代中 $A^{(v)}$ 的计算复杂度取决于 X_s 的大小 s , 邻域大小 K 以及样本的分布情况. 在实际计算中 X_α 可用 $\cup_{\mathbf{x} \in X_s} LR(\mathbf{x})$ 来近似代替. 假设 $LR(\mathbf{x})$ 的大小为 M , 则局部协同训练算法的时间复杂度为 $C_1 = O(nK^3 + n^3 + T(sMK^3$

$+ (sM)^3)V$, 其中 T 是迭代次数. 忽略低阶项, 则 $C_1 \approx O(n^3 + T(sM)^3)V$. 如果直接在 \mathcal{X}_T 上进行训练分类器, 则复杂度近似为 $C_2 \approx O((n + Ts)^3V)$. 显然当 T 很大时, $C_1 < C_2$, 即迭代的局部协同训练适合于在线学习问题.

2 网络图像的多模态特征提取

网络图像包含了图像本身的视觉信息和用户标注的文本信息, 涉及到颜色、纹理、角点、图像名称及图像描述等多种特征. 本文采用的特征按照其特点划分为如下三类:

2.1 整体视觉特征

整体视觉特征 (Global visual feature, GVF) 基于整幅图像的颜色直方图和梯度直方图来建立. 在颜色直方图中, R、G、B 三个颜色分量相互独立, 每个分量被平均分为 16 个小区间 (bin), 故得到 48 维的特征. 梯度直方图 (Histogram of gradient, HOG) 计算过程如下: 1) 针对每个颜色通道分别求解像素 $p_{i,j}$ 对应的梯度向量 $\mathbf{g}_{ij} = (x_{ij}, y_{ij})^T$, 其中, $x_{ij} = p_{i+1,j} - p_{i-1,j}$ 和 $y_{ij} = p_{i,j+1} - p_{i,j-1}$ 分别是横向和纵向的梯度分量; 2) 求解 \mathbf{g}_{ij} 的幅值 $\|\mathbf{g}_{ij}\|_2$ 和方向 $\phi_{ij} \in [0^\circ, 360^\circ)$; 3) 将方向角 ϕ_{ij} 平均分成 18 个 bins, 幅值 $\|\mathbf{g}_{ij}\|_2$ 在 $[0, 127]$ 中平均分成 31 个 bins, $(127, 255\sqrt{2}]$ 作为第 32 个 bin. 由于梯度幅值大的像素所占比例较小, 故上述梯度幅值空间未等间距分割. 最后, 将以上特征串联成一个 198 维的 GVF 向量.

2.2 局部视觉特征

局部视觉特征 (Local visual feature, LVF) 由局部颜色、局部纹理及 SIFT 描述子组成, 提取步骤如下: 1) 将每幅图像平均分成 $H \times W$ 个小块, 本文采用 4×3 (或 3×4) 和 6×4 (或 4×6); 2) 按照如下方法提取局部颜色特征: 对于每个小块, 分别在 RGB 空间和 HSV 空间中计算每个颜色分量的均值、标准差和偏态 (Skewness); 3) 按照如下方法提取局部纹理特征: 在灰度图像基础上采用 Gabor 滤波器在 3 个尺度 (频率为 0.125, 0.125/2 和 0.125/4) 和 4 个方向 ($0^\circ, 45^\circ, 90^\circ$ 和 135°) 上对每个小块进行计算; 4) 令 \mathbf{x}_r 表示每个小块的 30 维原始特征向量, 采用 K-均值算法对特征向量聚类, 得到中心为 \mathbf{c}_j 的 m 个簇 ($j = 1, 2, \dots, m$). \mathbf{x}_r 与第 j 个簇的相似性表示为

$$x_j = \left(\frac{d(\mathbf{x}_r, \mathbf{c}_j)}{\max_i d(\mathbf{x}_r, \mathbf{c}_i)} + \Delta \right)^{-1} \quad (12)$$

其中, Δ 为正则化因子, 本文令 $\Delta = 0.1$. 通过实验

分析, 令 $m = 300$. 此外, 局部特征还包括用来描述局部关键点的 SIFT 描述子^[18]. 在 HSV 颜色空间中, 每个关键点采用维的 SIFT 向量表示. 通过 K-均值算法获得 900 个视觉关键词, 则每幅图像采用直方图方式进行特征表示.

2.3 文本特征

文本特征 (Textual feature, TF) 基于图像的文本标题和描述进行提取. 首先将词的前缀、后缀及停用词去掉; 其次, 基于维基百科来将与每个词相关的词汇通过其自带的 API 接口函数进行返回, 用以扩充短文本, 降低稀疏性^[19], 该环节能使向量中非零元素增加为原来的 2.8 倍左右; 采用 TF-IDF 来表示文档:

$$t_{i,m} = tf(d_i, tm_m) \cdot \log \frac{N_d}{tsf(tm_m)} \quad (13)$$

其中, d_i 表示第 i 个文档, tm_m 表示第 m 个词 (term), $f(\cdot, \cdot)$ 表示词频, $tsf(\cdot, \cdot)$ 表示逆向文档频率, N_d 表示文档数目. 文本集合可以用矩阵 $T_D = [t_{i,m}]_{N_d \times M}$ 来表示. 网络多媒体中的文本较短且内容比较分散, 故直接使用式 (13) 得到的特征矩阵 T_D 维数高且比较稀疏. 本文采用潜在语义分析方法^[20] 将其映射至低维语义空间: $T_D = USV^T$, 其中 U 和 V 均为正交矩阵, S 为 $T_D T_D^T$ 特征值的非负平方根 $\{\sigma_i\}$ 组成的矩阵. 选取 M' 个最大值使得 $\sum_{i=1}^{M'} \sigma_i / \sum_{j=1}^M \sigma_j > \tau$, 则

$$T_D \approx U_{M'} S_{M'} V_{M'}^T \quad (14)$$

本文取 $\tau = 0.95$, 其中 $U_{M'}$, $S_{M'}$ 和 $V_{M'}^T$ 分别为 U , S 和 V^T 中的前 M' 列、前 M' 个对角元素和前 M' 行构成的矩阵. 第 j 个文档在语义空间中的坐标为 $\mathbf{d}_j V_{M'}^T$, 其中 \mathbf{d}_j 为通过式 (13) 求得的该文档对应的行向量.

3 实验结果及分析

3.1 实验数据描述

本文实验基于 Corel, Pascal VOC2007 和 ImageNet 三个数据集进行. 在 Corel 数据集中, 每个样本包含一幅图像和一段文本注释; 在 Pascal 和 ImageNet 数据集中只包含图像. 本文采用整体视觉特征 (GVF)、局部视觉特征 (LVF) 和文本特征 (TF) 来表示样本, 分别对应视图 $X^{(1)}$, $X^{(2)}$ 和 $X^{(3)}$. 类别选取主要考虑该类样本是否数量较多以及在互联网上是否普遍等因素. 由于 Pascal 和 ImageNet 中没有针对单幅图片的文本描述, 故在实验中不采用文本特征. 表 1 给出了数据集的描述.

3.2 类别标签传播分析结果

通过第 1.2 节分析可知, 在高维空间中样本类别标签随着与已标记样本的距离增大而减小. 本节在 ImageNet 数据的整体视觉特征空间中进行了实验验证. 共进行 20 次独立实验, 每次随机选取 5 000 个样本, 其中包含 1 个随机指定的标记样本, 其他均为未标记样本. 参数为 $K = 20$, $E_l = 10$, $E_u = 0.1$, $\lambda = 5$. 根据未标记样本个数随着与标记样本距离的变化情况, 我们将距离分为 14 个不等长区间, 在每个区间中计算标签估计值的均值和标准差. 图 2 表明标签预测结果随着未标记样本与标记样本距离的增大而衰减, 与第 1.2 节分析结果一致. 在该实验中通过计算发现, 在类别标签传播过程中, 平均有 67 个样本的预测值大于 0.025, 换句话说, 在第 1.3 节的增量学习算法中, 可以令样本标签传播局部区域 $LR(\mathbf{x})$ 包含 \mathbf{x} 周围的 67 个样本就能够得到准确的标签估计结果. 在实验中, 令 $LR(\mathbf{x})$ 的大小为 $3.5K$.

表 1 数据集描述

Table 1 Description of data sets

数据集	样本数	类别数	d_1	d_2	d_3
Corel	5 000	10	198	1 200	374
Pascal	9 963	20	198	1 200	-
ImageNet	45 000	8	198	1 200	-

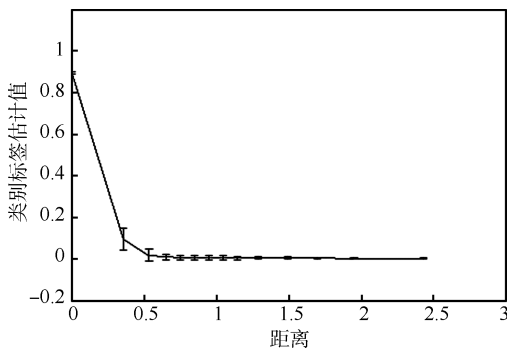


图 2 在 ImageNet 数据集的整体视觉特征空间中样本标签传播随着距离的变化规律

Fig. 2 The characteristics of label propagation with respect to the distance in global visual space on ImageNet

3.3 网络图像分类结果与分析

本文实验通过以下五种方法的比较来评估局部协同训练算法的性能:

- 1) LCT: 本文提出的局部协同训练算法;
- 2) CT-SVM: 以支持向量机 (Support vector machine, SVM) 为基本分类器的 Co-training 算法;
- 3) LL-CoReg: 基于局部模型的 Co-regulariza-

tion 方法;

4) La-CoReg: 基于 Laplacian 图的 Co-regularization 方法;

5) MKL-SVM: SVM 的多核学习 (Simple multiple kernel learning).

LL-CoReg 方法通过下式进行预测:

$$\begin{aligned} (\mathbf{f}^{(1)*}, \dots, \mathbf{f}^{(V)*}) = \arg \min_{\mathbf{f}^{(1)}, \dots, \mathbf{f}^{(V)}} \sum_{v=1}^V \alpha_v \times \\ \left(\|\mathbf{f}^{(v)} - \mathbf{A}^{(v)} \mathbf{f}^{(v)}\|^2 \|\mathbf{C}_1^{(v)\frac{1}{2}} (\mathbf{f}^{(v)} - \mathbf{y})\|^2 \right) + \\ \alpha_C \sum_{v,w} \|\mathbf{f}^{(v)} - \mathbf{f}^{(w)}\|^2 \end{aligned} \quad (15)$$

其中, α_v , α_C 和 \mathbf{C}_1 为参数. La-CoReg 方法通过下式进行预测:

$$\begin{aligned} (\mathbf{f}^{(1)*}, \dots, \mathbf{f}^{(V)*}) = \arg \min_{\mathbf{f}^{(1)}, \dots, \mathbf{f}^{(V)}} \sum_{v=1}^V \gamma_v \times \\ \left((\mathbf{f}^{(v)\text{T}} \mathbf{L}^{(v)} \mathbf{f}^{(v)})^2 + \|\mathbf{C}_2^{(v)\frac{1}{2}} (\mathbf{f}^{(v)} - \mathbf{y})\|^2 \right) + \\ \gamma_C \sum_{v,w} \|\mathbf{f}^{(v)} - \mathbf{f}^{(w)}\|^2 \end{aligned} \quad (16)$$

其中, γ_v , γ_C 和 \mathbf{C}_2 为参数, $\mathbf{L}^{(v)}$ 为 Laplacian 矩阵. 式 (15) 和式 (16) 中右侧的三项分别对应于预测函数的光滑性、准确性和不同模态上的一致性. 在 LL-CoReg 和 La-CoReg 算法的参数选取中, 令 $\alpha_C = 1$, $\gamma_C = 1$ 和 γ_v 通过在 1 维网格 [0.01, 0.1, 1, 10, 100] 上搜索获得, $\mathbf{C}_1^{(v)}$ 和 $\mathbf{C}_2^{(v)}$ 的选取同 LCT 中的 $\mathbf{E}^{(v)}$. 对于 MKL-SVM 方法^[15], 对于视觉模态选用径向基函数 (Radical basis function, RBF), 文本模态选用线性核函数, 采用凸组合方式对核进行融合: $k_c(\cdot, \cdot) = \sum_{v=1}^V k_v(\cdot, \cdot)$. 在未标记训练样本的利用中, 我们在标记样本集 \mathcal{L} 上训练 MKL-SVM 分类器, 并对未标记样本集 \mathcal{U}_p 进行预测, 将落在分类间隔之外的样本加入到 \mathcal{L} 中并重新对 MKL-SVM 进行训练和测试. 对于本文提出的 LCT 算法, 局部模型参数通过在以下网络上进行搜索确定: $n_i^{(v)} \in \{5, 10, 15, 20, 50\}$, $\lambda \in \{0.1, 1, 5, 20, 50, 100\}$, $E_l^{(v)} \in \{1, 5, 10, 50\}$, $E_u^{(v)} = 0.1$, $v = 1, 2, \dots, V$, 融合系数 β_v 通过 5 倍交叉验证来选取. CT-SVM 算法基于 LibSVM 库^[21] 来实现. 文本数据采用线性核函数, 视觉数据采用 RBF 核函数会得到最好的效果, 采用 5 倍交叉验证进行参数估计.

在样本集中, \mathcal{L} 占总样本的比例定义为标记率 r_L . 对每个 r_L 进行 20 次独立实验, 每次实验都对 \mathcal{L} , \mathcal{U} , \mathcal{U}_p 和 \mathcal{U}_t 进行随机划分, 其中 \mathcal{U}_t 为测试数据, $\mathcal{L} \cup \mathcal{U}$, \mathcal{U}_p 和 \mathcal{U}_t 占的比例分别为 30%, 25% 和 45%.

算法性能采用准确率 (Classification accuracy, CA) 来衡量:

$$CA = \frac{N_C}{N_T} \quad (17)$$

其中, N_C 和 N_T 分别表示正确分类的样本数和总样本数. 在算法 1 的协同学习环节, 每次迭代都从 \mathcal{U}_p' 中第 c 类样本的第 v 个视图选取 m_{cv} 样本添加到训练集 $\mathcal{L} \cup \mathcal{U}$ 中. 对于 Corel 和 ImageNet, 每个类别的数据量接近, 故令 $m_{1v} = m_{2v} = \dots = m_{Cv}$, 可用向量 $(m_{1v})_{1 \times V}$ 描述样本选取方式; 对于 Pascal, 由于类别很不均衡, (m_{cv}) 的选取与每类样本数量成正比. 对于每个数据集至少选取 5 种以上的样本选取方式, 下面的实验结果列出了两种较好的方式.

图 3、表 2 和表 3 比较了不同算法的部分学习结果. 其中, LL-CoReg 和 La-CoReg 算法均采用两种学习方式: 1) 将 \mathcal{U}_p 中的全部样本添加至训练集

中, 记为 LL-CoReg1 和 La-CoReg1; 2) 从 \mathcal{U}_p 中选取部样本添加至训练集, 其数量等同于 LCT 迭代过程中选取的样本, 记为 LL-CoReg2 和 La-CoReg2. 由于 LL-CoReg 和 La-CoReg 不需要迭代地标记样本, 故实验中只给出了两个结果: 在 $\mathcal{L} \cup \mathcal{U}$ 上和 $\mathcal{L} \cup \mathcal{U}$ 增加未标记样本后的测试结果, 分别标记在 $t = 0$ 和 $t = 30$ 处. 随着样本数量不断增加, LCT 和 CT-SVM 的学习性能有着不同程度的改变. 随着迭代次数的增加, LCT 算法的分类性能提高的较为明显, 而 CT-SVM 算法的分类性能提升幅度较小, 有时候会有一定下降, 如表 2. 对于 LL-CoReg 和 La-CoReg 两种正则化算法来说, 未标记样本的增多对学习器性能的提高有一定帮助. 同 LCT 算法比较可以看出, 在同样大小的训练集上, LCT 的性能要优于两种正则化算法. 此外, LL-CoReg1 和 La-CoReg1 尽管利用了更多的未标记训练数据, 但

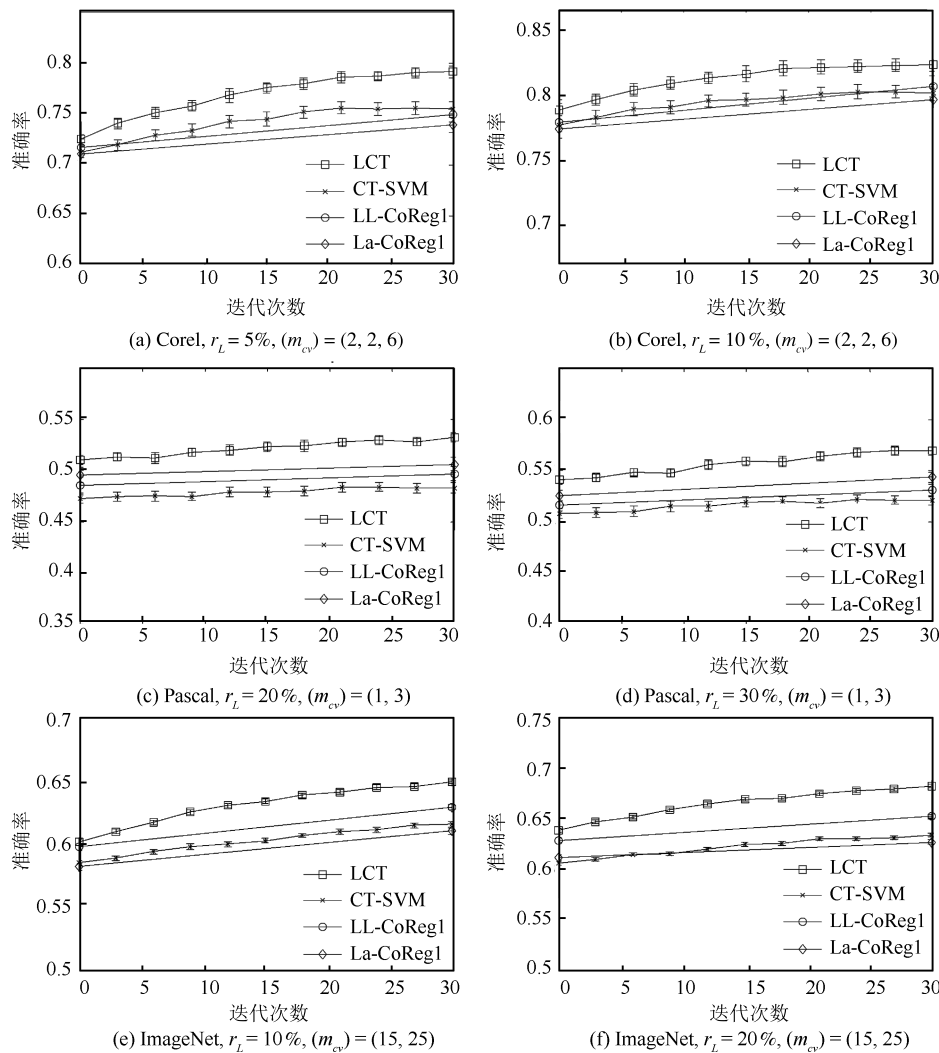


图 3 网络图像分类准确率比较结果 (图中每条曲线的错误条表示 ± 1 的标准错误偏差)

Fig. 3 The comparison of web image classification (The error bars indicate ± 1 standard error deviation.)

表 2 LCT 和 CT-SVM 算法性能比较
Table 2 Comparison of LCT and CT-SVM

Dataset	Algorithm	r_L (%)	(m_{cv})	CA_I	CA_F	ΔCA (%)	CA_A
Corel	LCT	5	(2, 2, 6)	0.725	0.791	24.0	0.769
		5	(2, 3, 5)	0.725	0.776	18.6	0.763
		10	(2, 2, 6)	0.792	0.823	16.7	0.816
		10	(2, 3, 5)	0.792	0.819	12.9	0.811
	CT-SVM	5	(2, 2, 6)	0.713	0.755	14.8	0.742
		5	(2, 3, 5)	0.713	0.722	3.1	0.720
		10	(2, 2, 6)	0.779	0.805	11.6	0.798
		10	(2, 3, 5)	0.779	0.783	1.8	0.783
Pascal	LCT	20	(1, 3)	0.508	0.532	4.9	0.522
		20	(2, 2)	0.508	0.525	3.5	0.518
		30	(1, 3)	0.541	0.569	6.2	0.557
		30	(2, 2)	0.541	0.563	4.6	0.552
	CT-SVM	20	(1, 3)	0.472	0.483	2.0	0.479
		20	(2, 2)	0.472	0.470	-0.4	0.472
		30	(1, 3)	0.508	0.520	2.6	0.516
		30	(2, 2)	0.508	0.496	-2.5	0.506
ImageNet	LCT	10	(20, 20)	0.602	0.646	11.1	0.628
		10	(15, 25)	0.602	0.650	12.1	0.631
		20	(20, 20)	0.638	0.677	10.8	0.660
		20	(15, 25)	0.638	0.682	12.2	0.665
	CT-SVM	10	(20, 20)	0.585	0.609	5.8	0.597
		10	(15, 25)	0.585	0.616	7.5	0.602
		20	(20, 20)	0.605	0.630	6.3	0.620
		20	(15, 25)	0.605	0.633	7.1	0.621

表 3 LCT 与正则化方法和多核学习方法的最终分类准确率比较
Table 3 Comparison of LCT, regularization, and multiple kernel methods

	r_L (%)	LCT	LL-CoReg1	LL-CoReg2	La-CoReg1	La-CoReg2	MKL-SVM
Corel	5	0.791	0.758	0.743	0.747	0.735	0.768
	10	0.823	0.809	0.797	0.801	0.786	0.812
Pascal	20	0.532	0.498	0.495	0.506	0.500	0.503
	30	0.569	0.531	0.523	0.544	0.538	0.542
ImageNet	10	0.650	0.629	0.618	0.610	0.598	0.627
	20	0.682	0.652	0.640	0.626	0.619	0.662

准确率仍旧提升不明显。

在表 2 和表 3 中, CA_I , CA_F , ΔCA 和 CA_A 分别表示学习器的初始准确率、最终准确率、错误率下降比例和迭代过程中的平均准确率. 从中可以看出 LCT 整体上具有一定优势. 例如, LCT 最终的分类型准确率总体上比其他比较算法高约 1.1% ~ 6.7%. 从表 3 可以看出, La-CoReg1 和 LL-CoReg1 的分类性能相比 La-CoReg2 和 LL-CoReg2 有轻微提高, 这说明未标记样本的增多可以提高分类准确率, 但是直接使用未标记样本会引入较多噪声. 而 LCT 利用了类别标记置信度较高的未标记样本来扩充训练集, 引入的噪声较小, 对于提升半监督学习的性能有较好的帮助. 此外, 从表 3 可以看出 MKL-SVM 的性能也较好, 在某些情况下优于正则化方法. 一般地, 半监督算法的学习性能会受某些因素影响, 如样本标记率 r_L 、未标记样本选取方式 (m_{cv}) 等. 如果样本分布比较复杂, r_L 过小则会使得分类器性能较差, 进而在对未标记样本的预测上会产生比较大的错误率并会引入较大的噪声, 造成学习性能提升不明显甚至恶化. 此外, 不同视图对目标表示能力的差异使得未标记样本的选取方式也会影响学习性能. 例如, 对于 Corel 数据, 当 (m_{cv}) = (2, 2, 6) 时, LCT 算法的分类性能提高要略优于 (m_{cv}) = (2, 3, 5), 其性能差别主要源于文本信息对样本的表示更准确.

LCT 算法在网络图像半监督学习中优于 CT-SVM, 主要原因如下: 1) 局部学习通常要比全局学习有更好的性能, 尤其是在样本分布复杂情况下可以带来较低的错误率; 2) LCT 在迭代过程中若错误地标记了样本, 不良影响主要限制在局部区域, 而 CT-SVM 可能会对整个空间的分类性能产生影响. 此外, 对于 CT-SVM 或其他以全局学习为基础的 Co-Training 算法来说, 每次迭代选取的未标记样本通常远离分类面, 包含新的分类信息通常较少, 不利于快速提升学习器性能. 这些因素也是 LCT 优于 MKL-SVM 的重要原因. 由 LCT 和正则化算法的比较来看, 选择标记置信度高的未标记样本要比随机选取更有利于分类器学习, 后者在提高分类性能同时更容易引进噪声.

图 4 显示了局部协同训练算法在 Corel 数据集上的计算复杂度. 算法运行在 Intel Core2 Duo 2.53 GHz 处理器、4 G 内存平台、Matlab R2008 环境下. 由该图可知, LCT 算法的计算时间与处理的样本个数 (或算法迭代次数 T) 近似成线性关系, 这与第 1.4 节分析的结果是一致的. 其主要原因是 LCT 能够在线增量地更新学习器, 而正则化方法 (Co-regularization)^[12] 由于不能增量地训练学习器, 故计算复杂度较高, 不利于在线处理新增样本.

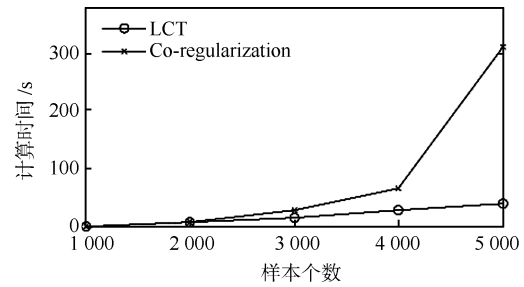


图 4 算法时间复杂度比较 (不含特征提取计算时间)

Fig. 4 The comparison of computation time (excluding the feature extraction cost)

4 结论

本文提出了一种基于局部协同训练的网络图像半监督学习方法. 该方法把网络图像包含的多种异构特征看作多个视图, 在每个视图上基于每个样本点的邻域构建一个线性局部模型, 以一组局部模型表示对应视图上的数据关系, 进而采用协同训练策略对模型进行迭代更新. 在该研究中, 本文着重分析了在加入新的未标记样本后如何有效地重新计算样本的标签估计, 以及如何对多个视图上的局部模型进行协同学习等问题. 局部协同训练算法能够比较准确地描述样本的复杂分布, 同时也能够有效利用多种异构特征和未标记样本在线提高算法性能. 在 Corel, Pascal 和 ImageNet 数据集上的实验结果表明, 局部协同训练算法能够有效地提高学习器的分类性能, 对网络图像的半监督分类有重要帮助.

References

- 1 Wu Y, Chang E Y, Chang K C C, Smith J R. Optimal multimodal fusion for multimedia data analysis. In: Proceedings of the 12th Annual ACM International Conference on Multimedia. New York, USA: ACM, 2004. 572–579
- 2 Jiang T, Tan A H. Learning image-text associations. *IEEE Transactions Knowledge and Data Engineering*, 2009, **21**(2): 161–177
- 3 Zhu X J. Semi-supervised learning literature survey. *Computer Sciences TR 1530*, University of Wisconsin - Madison, USA, 2008, 1–60
- 4 Blum A, Mitchell T. Combining labeled and unlabeled data with co-training. In: Proceedings of the 11th Annual Conference on Computational Learning Theory. Madison, Wisconsin, USA: ACM, 1998. 92–100
- 5 Zhang M L, Zhou Z H. CoTrade: confident co-training with data editing. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 2011, **41**(6): 1612–1626
- 6 Yan R, Naphade M. Semi-supervised cross feature learning for semantic concept detection in videos. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 657–663
- 7 Gupta S, Kim J, Grauman K, Mooney R. Watch, listen & learn: co-training on captioned images and videos. *Lecture Notes in Computer Science*, 2008, **5211**: 457–472

- 8 Kumar A, Daumé H III. A co-training approach for multi-view spectral clustering. In: Proceedings of the 2011 International Conference on Machine Learning. Bellevue, Washington, USA: ACM, 2011. 393–400
- 9 Belkin M, Matveeva I, Niyogi P. Regularization and semi-supervised learning on large graphs. In: Proceedings of the 2004 International Conference on Computational Learning Theory. Banff, Canada: ACM, 2004. 624–638
- 10 Goldman S A, Zhou Y. Enhancing supervised learning with unlabeled data. In: Proceedings of the 17th International Conference on Machine Learning. Stanford, USA: ACM, 2000. 327–334
- 11 Wu M R, Schölkopf B. Transductive classification via local learning regularization. In: Proceedings of the 2007 International Conference Artificial Intelligence and Statistics. San Juan, Puerto Rico, 2007. 628–635
- 12 Sindhwani V, Niyogi P, Belkin M. A co-regularization approach to semi-supervised learning with multiple views. In: Proceedings of the 22nd ICML Workshop on Learning with Multiple Views. Bonn, Germany: ICML, 2005. 321–328
- 13 Zhang D, Wang F, Zhang C S, Li T. Multi-view local learning. In: Proceedings of the 23rd National Conference on Artificial Intelligence. Chicago, IL: AAAI, 2008. 752–757
- 14 Kumar A, Rai P, Daumé H III. Co-regularized multi-view spectral clustering. In: Proceedings of the 2011 Advance in Neural Information Processing Systems. Nevada, USA: MIT Press, 2011. 1413–1421
- 15 Guillaumin M, Verbeek J, Schmid C. Multimodal semi-supervised learning for image classification. In: Proceedings of the 2010 International Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 902–909
- 16 Bottou L, Vapnik V. Local learning algorithms. *Neural Computation*, 1992, **4**(6): 888–900
- 17 Hein M, Audibert J Y, von Luxburg U. From graphs to manifolds-weak and strong pointwise consistency of graph Laplacians. In: Proceedings of the 18th Annual Conference on Learning Theory. New York, USA: ACM, 2005. 470–485
- 18 Lowe D. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, **60**(2): 91–110
- 19 Zhang X, Song Y C, Cao J, Zhang Y D, Li J T. Large scale incremental web video categorization. In: Proceedings of the 1st Workshop on Web-Scale Multimedia Corpus. Beijing, China: ACM, 2009. 33–40
- 20 Landauer T K, Foltz P W, Laham D. Introduction to latent semantic analysis. *Discourse Processes*, 1998, **25**: 259–284
- 21 Chang C C, Lin C J. LIBSVM: a library for support vector machines. [Online], available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, January 1, 2001

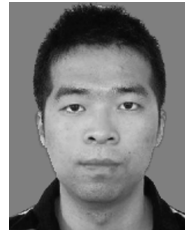


杜友田 西安交通大学电信学院讲师。2008 年获清华大学自动化系博士学位。主要研究方向为网络多媒体理解, 社会网络, 机器学习。

E-mail: duyt@mail.xjtu.edu.cn

(**DU You-Tian** Lecturer at Xi'an Jiaotong University. He received his Ph. D. degree from Tsinghua University

in 2008. His research interest covers social multimedia understanding, social network, and machine learning.)

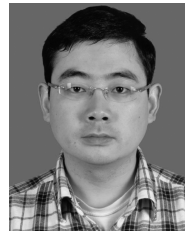


李 谦 西安交通大学电信学院硕士研究生。2010 年获华南理工大学学士学位。主要研究方向为网络多媒体理解。

E-mail: liqian@stu.xjtu.edu.cn

(**LI Qian** Master student at Xi'an Jiaotong University. He received his bachelor degree from South China University of Technology in 2010. His main

research interest is social multimedia understanding.)

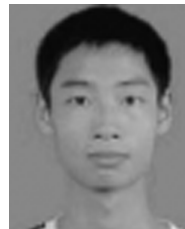


周亚东 西安交通大学电信学院讲师。2011 年获西安交通大学自动化系博士学位。主要研究方向为网络舆情分析, 在线社会网络。本文通信作者。

E-mail: ydzhou@mail.xjtu.edu.cn

(**ZHOU Ya-Dong** Lecturer at Xi'an Jiaotong University. He received his Ph. D. degree from Xi'an Jiaotong University in 2011. His research interest covers public opinion

analysis and social network. Corresponding author of this paper.)



吴陈鹤 西安交通大学电信学院硕士研究生。2010 年获西安交通大学学士学位。主要研究方向为在线社会网络。

E-mail: wuchenhe@stu.xjtu.edu.cn

(**WU Chen-He** Master student at Xi'an Jiaotong University. He received his bachelor degree from Xi'an Jiaotong University in 2010. His main research

interest is online social network.)