

文章编号:1001-9081(2015)06-1611-06

doi:10.11772/j.issn.1001-9081.2015.06.1611

基于多图的交替优化图直推方法

修宇^{1*}, 王骏², 王忠群¹, 刘三民¹

(1. 安徽工程大学 计算机与信息学院,安徽 芜湖 241000; 2. IBM T J Watson 研究中心,纽约州 约克城高地 10598)

(*通信作者电子邮箱 xiuyu1860@126.com)

摘要:针对基于单图的半监督学习(GSSL)算法的性能受单个图质量的影响,且在单视图数据下,大多数基于多图的CSSL算法难以使用的问题,提出了一种基于多图的交替优化图直推方法(MG-GTAM)。首先,使用不同的图构建参数来构建单视图数据下的多个图,利用多个图来表达数据间关系;然后,借助交替迭代方式综合多个图的信息,选择置信度高的未标记样本进行伪标记并通过权重权衡各图的重要程度,以优化多图上的预测函数的一致性和平滑性;最后通过组合每个图的预测函数完成对所有未标记样本的标记。仿真实验表明,与经典的局部和全局一致(LGC)、高斯随机场和调和函数(GFHF)、交替优化直推(GTAM)、组合图拉普拉斯(CGL)算法相比,在 COIL20 目标物体数据集和 NEC Animal 数据集上, MG-GTAM 的分类错误率比这些经典算法均有下降,表明了该方法具有良好的性能。实验结果表明, MG-GTAM 能有效地利用多个图来表达数据之间的关系,获得更低的分类错误率。

关键词:图半监督学习;图直推;图构建;多图;交替优化

中图分类号: TP181 **文献标志码:**A

Graph transduction via alternating minimization method based on multi-graph

XIUYU^{1*}, WANG JUN², WANG ZHONGQUN¹, LIU SANMIN¹

(1. College of Computer and Information Engineering, Anhui Polytechnic University, Wuhan Anhui 241000, China;

2. IBM T. J. Watson Research Center, Yorktown Heights New York 10598, USA)

Abstract: The performance of the Graph-based Semi-Supervised Learning (GSSL) method based on one graph mainly depends on a well-structured single graph and most algorithms based on multiple graphs are difficult to be applied while the data has only single view. Aiming at the issue, a Graph Transduction via Alternating Minimization method based on Multi-Graph (MG-GTAM) was proposed. Firstly, using different graph construction parameters, multiple graphs were constructed from data with one single view to represent data point relation. Secondly, the most confident unlabeled examples were chosen for pseudo label assignment through the integration of a plurality of map information and imposed higher weights to the most relevant graphs based on alternating optimization, which optimized agreement and smoothness of prediction function over multiple graphs. Finally, more accurate labels were given over the entire unlabeled examples by combining the predictions of all individual graphs. Compared with the classical algorithms of Local and Global Consistency (LGC), Gaussian Fields and Harmonic Functions (GFHF), Graph Transduction via Alternation Minimization (GTAM), Combined Graph Laplacian (CGL), the classification error rates of MG-GTAM decrease on data sets of COIL20 and NEC Animal. The experimental results show that the proposed method can efficiently represent data point relation with multiple graphs, and has lower classification error rate.

Key words: Graph-based Semi-Supervised Learning (GSSL); graph transduction; graph construction; multi-graph; alternating minimization

0 引言

在机器学习的很多实际问题中,易于得到大量的样本但较难获取这些样本的标记。半监督学习(Semi-Supervised Learning, SSL)能综合利用有标记样本和无标记样本,在较少标记的情况下构建性能较好的学习器,因而得到了广泛关注和研究^[1-2]。基于图的半监督学习(Graph-based Semi-

Supervised Learning, GSSL)是半监督学习方法的一个重要范型^[3]。以往对GSSL方法的研究大多关注单个图下的学习,研究表明单个图的质量是影响这些方法的关键因素^[4-5]。近年来,数据的多视图性使得基于多图的图半监督学习方法开始受到关注。数据具有多视图是一种常见现象,通过映射可以直接将数据的多个视图映射为多个图,利用多个图上的信息可提升GSSL的性能,不少研究者提出了基于多图的学习

收稿日期:2014-12-31;修回日期:2015-04-01。

基金项目:国家自然科学基金资助项目(71371012, 71171002);教育部人文社科规划项目(13YJA630098);安徽省优秀青年人才基金重点项目(2013SQRL034Z);安徽省高校省级科学研究项目(TSKJ2014B10);安徽工程大学计算机应用技术重点实验室开放基金资助项目(JSJKF201510)。

作者简介:修宇(1976-),男,山东单县人,讲师,硕士,CCF会员,主要研究方向:机器学习、数据挖掘;王骏(1976-),男,安徽芜湖人,研究员,博士,主要研究方向:机器学习、数据挖掘;王忠群(1965-),男,安徽芜湖人,教授,主要研究方向:商务智能;刘三民(1978-),男,安徽岳西人,副教授,博士,主要研究方向:模式识别、网络安全。

方法。例如:Sindhwani 等^[6-7]提出利用图正则化进行多视图学习的协同正则法(Co-regularization);Zhang 等^[8]将局部学习推广用于解决多视图数据下的分类问题,设计并利用局部多视图模型来提高分类性能;Du 等^[9]提出一种基于多图半监督学习方法,在每个网络图像的视图上对每个样本点的邻域构建线性局部模型,通过多个图分类器之间的局部协同训练来提高网络图像的分类性能。上述方法均假设数据包括多个视图,然而在一些实际应用中,数据往往仅具有一个视图,这使得上述多图方法很难用于这些单视图数据。不同于以上研究者的思路,Argyriou 等^[10]提出一种组合图拉普拉斯半监督学习方法(Combined Graph Laplacian,CGL),使用不同的构图参数构建多个图,通过迭代求解图核的最优凸组合问题来进行分类。CGL 在每次迭代优化过程中都需计算图核矩阵组合,导致算法计算量很大,很难应用于实际应用中。

本文基于交替优化图直推算法(Graph Transduction via Alternating Minimization method,GTAM),提出了一种基于多图的交替优化图直推算法(Multi-Graph-based Graph Transduction via Alternating Minimization method,MG-GTAM)。该方法的优点在于:1)在单视图下使用不同的构图参数来生成多个表达数据间关系的图;2)在迭代过程中通过伪标记的策略进行多个图上信息的交互,并通过权重自动权衡每个图的重要程度;3)在少量已标记样本下能获得较高的分类性能。

1 多图的构建

基于图的半监督学习问题基本设置如下:给定 n 个数据的集合 $X = \{x_1, x_2, \dots, x_n\}$, 其中 $X_l = \{x_i\}_{i=1}^l, X_u = \{x_i\}_{i=l+1}^n$, 分别代表已标记样本集合未标记样本集,通常 $l \ll n$, X_l 的标记与已标记样本标签集 $Z_l = \{z_1, z_2, \dots, z_l\}$ 关联, 其中 $z_i \in \{1, 2, \dots, c\}, i = 1, 2, \dots, l, c$ 代表类别。学习的目标是推导未标记样本集 $\{x_i\}_{i=1}^n$ 的标签集 $\{z_{l+1}, z_{l+2}, \dots, z_n\}$ 。基于流形假设,可采用一定的图构建方法构建图 $G = \{X, e\}$, 使已标记样本的信息和未标记样本信息有效地结合,利用图上的邻接关系将已标记样本 X_l 的标签信息向未标记样本 X_u 传播,完成对未标记样本的标记。

图构建包括三个基本步骤^[1-3]:1) 选择合适的距离函数或相似度函数计算图上样本间的距离或相似度。可供选择的距离函数有欧氏距离、马氏距离、曼哈顿距离、切比雪夫距离等。该步骤将产生一个全邻接矩阵 $A \in \mathbb{R}^{n \times n}$ 。2) 进行图的稀疏化,确定图的邻接结构。稀疏化可以并按照某种准则建立距离最近节点连接,同时移除 A 中不必要的边,生成稀疏图。常用的稀疏化方法有 ε 近邻(ε NN)、 k 近邻(k NN)、 b 匹配(b -matching)等。 ε 近邻法和 k 近邻法分别需要选择合适的半径 ε 和近邻数 k ,而 b 匹配法则需要调整每个节点仅能连接的邻居个数 b ,这些参数的选择往往需根据经验或实验来确定。3) 图上边的赋权。图的稀疏化完成后,可采用一定方法给稀疏图上的边赋权得到权重矩阵 W ,以反映边所连接的两个节点的相似程度。常用的方法有:二值置权法、核函数置权法、局部线性重构置权法。核函数置权法中可选的核函数有线性核、多项式核、高斯核、双曲正切核、样条核等。目前在基于图的半监督学习中使用比较多的边赋权值方法是高斯核函数置权法,样本点 x_i 和 x_j 间的权重一般采用如下公式计算:

$$w_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2 / 2\sigma^2), & x_i \text{ 和 } x_j \text{ 是邻居} \\ 0, & x_i \text{ 和 } x_j \text{ 非邻居} \end{cases}$$

其中核宽度参数 σ 的值通常设置为样本与其 k 个近邻距离的平均值,也可以根据领域知识人工设置。

根据图构建的基本步骤可以看出,采用不同的边度量方法、稀疏化方法和边赋权策略,会得到不同结构的图^[1-3]。实践表明,图的构造质量直接影响基于图的学习方法的性能^[4-5,12]。具体问题中往往需通过经验或试凑法确定图的构造参数,如当问题规模较大时,为了保证每个连通子图中至少有 1 个有标记样本,需要花很大的时间开销选择图的构造参数^[13];此外当已标记样本数较少时并没有可靠的方法来寻找合适的构图参数。文献[14]指出基于图的聚类方法可以在多个不同的图上都获得最优的性能,这说明可通过多个图来表示数据的内在分布和几何结构关系。文献[5]通过对比实验研究了图构造参数对五种经典的图半监督学习算法的性能影响,实验结果显示对于某些数据集很难选择特定的构图参数使得图半监督学习方法性能达到最优,这表明对于某些数据集仅采用单个图无法准确描述数据内在分布和几何结构关系。Argyriou 等^[10]提出的 CGL 方法的目的是避免图构造参数的选择问题,其方法虽然有效但在很长一段时间内其奏效的内在原因并不清楚。周志华[11]在后续研究中从分歧学习的角度指出 CGL 方法之所以奏效的可能原因在于其利用构图参数的差异产生了多个具有差异的图传播结构,从而为在单视图条件下进行多视图学习提供了必要条件。因此,本文认为可以利用不同的图构建参数来构造多个图以反映数据之间的复杂关系。其好处是:一方面通过多个图可综合反映数据的内在分布和几何结构;另一方面图构建参数的差异使得每个图上的传播结构不同,这为在单视图条件下进行多视图学习提供了可能。

2 基于多图的交替优化直推方法

交替优化图直推方法(GTAM)^[14]是一种基于双变量的图半监督学习方法,较好地解决了标记不平衡问题和噪声问题,具有优越的分类性能。后续研究^[15]表明该方法的求解过程等同于求解图上的最大割问题。限于篇幅,GTAM 方法的具体细节可见文献[14]。假设采用不同的图构建参数构建了 k 个图 $\{G_1, G_2, \dots, G_k\}$, 基于 GTAM 的代价函数 $Q(\mathbf{F}, \mathbf{Y})$ ^[14], 可定义基于多图的交替优化图直推方法(MG-GTAM)的目标函数如下:

$$\begin{aligned} \mathbf{F}^* = \arg \min_{\mathbf{F} \in \mathbb{R}^{n \times c}, \mathbf{Y} \in \mathbb{R}^{n \times c}, \boldsymbol{\alpha} \in \mathbb{R}^k} Q(\mathbf{F}, \mathbf{Y}, \boldsymbol{\alpha}) = \\ \arg \min \sum_{q=1}^k \alpha_q \|\mathbf{F}_q\|^2 + \frac{\mu}{2} \sum_{q=1}^k \alpha_q \|\mathbf{F}_q - V\mathbf{Y}\|^2 \end{aligned} \quad (1)$$

$$\begin{aligned} \text{s. t. } \sum_{q=1}^k \alpha_q = 1, \alpha_q \geq 0, \forall q \in \{1, 2, \dots, k\} \\ Y_{ij} \in \{0, 1\}, \sum_j Y_{ij} = 1 \\ \forall i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, c\} \end{aligned}$$

其中: $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_k]$ 代表 $\{G_1, G_2, \dots, G_k\}$ 中每个图的权重值,图 G_q 对应的 α_q 值越大,代表该图对标签传播的影响越大。

式(1)中的第一部分 $\sum_{q=1}^k \alpha_q \|\mathbf{F}_q\|^2$ 代表对多图实施光滑性约束,保证每个图上相邻样本之间的标签具有一定的相似

性从而使预测函数具有足够的平滑性。 $\sum_{q=1}^k \alpha_q \|F_q\|^2$ 的具体形式表示如下:

$$\begin{aligned}\sum_{q=1}^k \alpha_q \|F_q\|^2 &= \frac{1}{2} \left(\sum_{q=1}^k \alpha_q F^T L^q F \right) = \\ &\frac{1}{2} \text{tr} \left[F^T \left(\sum_{q=1}^m \alpha_q L^q \right) F \right] = \frac{1}{2} \text{tr} \left[F^T \mathcal{L} F \right]\end{aligned}\quad (2)$$

不同于 CGL 算法^[10]中采用的固定形式的拉普拉斯矩阵组合 $\mathcal{L} = \sum_{q=1}^k L^q$, 式(2)中 \mathcal{L} 定义为 $\mathcal{L} = \sum_{q=1}^k \alpha_q L^q$, 即通过权重的来权衡每个图的重要程度。

式(1)中第二部分 $\sum_{q=1}^k \alpha_q \|F_q - VY\|^2$ 用来惩罚多个图上的直推分类器的预测标签不等于给定标签的情况, 保证已标记样本能够得到正确的分类。其中 $V = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ 为标签规范化矩阵, 其作用为避免不平衡标签或不平衡类对标签传播的影响。在多图情况下定义 V 如下:

$$\lambda_i = \begin{cases} \sum_{q=1}^m \alpha_q \frac{D_{ii}^q}{\sum_r Y_{ij} D_{rr}}, & Y_{ij} = 1 \\ 0, & \text{其他} \end{cases} \quad (3)$$

不同于 GTAM 中标签规范化系数定义, 式(3)综合考虑到被标记点在每个图中传播中的重要程度, 从而更好地在迭代过程中减轻类不平衡性对标签传播的影响。令 $\bar{Y} = VY$, \bar{Y} 代表规范化的标签矩阵, 则 $\sum_{q=1}^k \alpha_q \|F_q - VY\|^2$ 可写成如下形式:

$$\sum_{q=1}^k \alpha_q \|F_q - VY\|^2 = \sum_{q=1}^k \alpha_q \|F_q - \bar{Y}\|^2 = \sum_{q=1}^k \alpha_q (F_q - \bar{Y})^T (F_q - \bar{Y}) \quad (4)$$

MG-GTAM 目标函数可进一步写成如下形式:

$$\begin{aligned}Q(F, \bar{Y}, \alpha) &= \frac{1}{2} \sum_{q=1}^k \alpha_q \text{tr}(F_q^T L^q F_q) + \\ &\frac{\mu}{2} \sum_{q=1}^k \alpha_q (F_q - \bar{Y})^T (F_q - \bar{Y})\end{aligned}\quad (5)$$

采用梯度下降法对目标函数(5)交替迭代求解, 推导步骤如下:

对 $Q(F, \bar{Y}, \alpha)$ 中每个 F_q 求偏导有:

$$\frac{\partial Q}{\partial F_q} = \alpha_q L_q F_q + \mu \alpha_q (F_q - \bar{Y}) \quad (6)$$

令式(6)等于 0, 则有:

$$F_q^* = (L_q / \mu + I)^{-1} \bar{Y} = P_q \bar{Y} \quad (7)$$

其中定义 $P_q = (L_q / \mu + I)^{-1}$ 。当已知部分样本标签矩阵 \bar{Y} 和图 G_q 上的 P_q 矩阵, 可由式(7)计算每个图的预测函数, 因此 P_q 可视为图 G_q 上的传播矩阵。将式(7)代入式(5)中可得:

$$Q(\bar{Y}, \alpha) = \frac{1}{2} \text{tr} \left[\bar{Y}^T \left(\sum_{q=1}^k \alpha_q (P_q^T L_q P_q + \mu (P_q - I)^2) \right) \bar{Y} \right] \quad (8)$$

定义常量 $A_q = P_q^T L_q P_q + \mu (P_q - I)^2$, 可看出 A_q 的值取决于由图 G_q 的结构, 且每个图的 A_q 能离线计算。式(8)可进一步表示为:

$$\begin{aligned}Q(\bar{Y}, \alpha) &= \frac{1}{2} \text{tr} \left(\bar{Y}^T \left(\sum_{q=1}^k \alpha_q A_q \right) \bar{Y} \right) = \\ &\frac{1}{2} \sum_{q=1}^k \alpha_q \text{tr} (\bar{Y}^T A_q \bar{Y})\end{aligned}\quad (9)$$

式中代价函数 $Q(\bar{Y}, \alpha)$ 最小化问题可视为求解 $\{A_q\}$ ($q = 1, 2, \dots, k$) 的凸组合问题, 因而可以采用梯度下降法分别对 \bar{Y} 和 α 交替优化求解。

对 $Q(\bar{Y}, \alpha)$ 中 \bar{Y} 和 α 分别求偏导有:

$$\frac{\partial Q}{\partial \bar{Y}} = \sum_{q=1}^k \alpha_q A_q \bar{Y} \quad (10)$$

$$\frac{\partial Q}{\partial \alpha_q} = \frac{1}{2} (\bar{Y}^T A_q \bar{Y}) \quad (11)$$

式(10)中 \bar{Y} 为规范化的标签矩阵, 其元素与标签矩阵 Y 中的元素一一对应, 因此对 \bar{Y} 更新等同于对 Y 进行更新。 $\frac{\partial Q}{\partial \bar{Y}}$ 梯度矩阵中对应未标记样本元素值的大小表示了影响代价函数值 $Q(\bar{Y}, \alpha)$ 的下降程度。文献[14]指出在 GTAM 中该值越小代表其变化时影响代价函数值下降的程度越高, 同时属于某一类的置信度越大。GTAM 仅根据单个图的信息来判断未标记样本属于某一类的置信度, 而 MG-GTAM 通过权重 α 综合考虑多个图的信息来判断未标记样本属于某一类的置信度。采用类似 GTAM 在每次迭代中通过寻找梯度矩阵 $\nabla_{\bar{Y}} Q$ 中最小的元素值的方法来完成对未标记集 X_u 中样本 x_i 的标记, 即:

$$(i^*, j^*) = \arg \min_{x_i \in X_u, 1 \leq j \leq c} \frac{\partial Q}{\partial \bar{Y}} \quad (12)$$

$$y_{i^*, j^*} = 1 \quad (13)$$

通过式(10)、(12)和式(13), MG-GTAM 在每次迭代中综合多个图上的标签信息来选择置信度高的未标记样本进行伪标记, 并将置信度较高的样本加入到所有其他图直推分类器中, 利用伪标记完成多图之间信息的交换。该策略类似于协同训练中多学习器之间的信息交互, 不同于协同训练中的伪标记由每个学习器给出, MG-GTAM 通过综合多个图上的标签信息来确定未标记样本的伪标记。

利用式(13)更新标签矩阵 Y 后, 通过式(3)计算 V 并更新规范化的标签矩阵 \bar{Y} 。基于式(9)和(10), 图的权重值 α 采用梯度下降法方法更新, 即:

$$\alpha = \alpha_q - \eta \frac{\partial Q}{\partial \alpha_q} \quad (14)$$

其中 η 为步长参数。由式(1)知 α_q 的值受 $\alpha_q \geq 0$ 和 $\sum_{q=1}^k \alpha_q = 1$ 的约束, 因此需在 α_q 每次更新后进行归一化, 并对于值小于 0 的 α_q 赋值为 0。通过迭代过程中 α_q 值的更新, 可自动根据每个图对标签传播的贡献调整图的权重值。考虑到对于每个图 G_q , 当 Y, α 固定时, 通过优化求解 $\arg \min_F Q(F, Y, \alpha)$ 得到 $\{F_q\}$ ($q \in \{1, 2, \dots, k\}$)。经过 T 次迭代后, 预测函数 F^* 由每个图的 $\{F_q\}$ 通过每个图的权重值 $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_k]$ 线性组合计算获得, 即:

$$F^* = \sum_{q=1}^k \alpha_q F_q \quad (15)$$

在以上分析的基础上, 本文利用第 1 章的多图生成策略在单视图数据上构造 k 个图 $\{G_1, G_2, \dots, G_k\}$, 进而提出 MG-GTAM 算法, 其主要步骤如下:

输入 数据集 $X = \{X_l, x_u\}$, 权重矩阵 $\{W_q\}$, 度矩阵 $\{D_q\}$, 图拉普拉斯矩阵 $\{L_q\}$, 传播矩阵 $\{P_q\}$ ($q = 1, 2, \dots, k$), 初始标签矩阵 Y^0 , 已标记样本标签集 $Z_l = \{z_1, z_2, \dots, z_l\}$, 迭代次数 T 。

输出 未标记样本的标签集 $\{z_{l+1}, z_{l+2}, \dots, z_n\}$ 。

步骤 1 初始化迭代计数变量 $t = 0$, 图的权重值系数 $\alpha_q = 1/k$ ($q = 1, 2, \dots, k$), 标签规范化矩阵 $V^t = \text{diag}(\lambda_1^t, \lambda_2^t, \dots, \lambda_n^t)$ 。

步骤 2 开始迭代:

步骤 2.1 根据式(10)计算 $\frac{\partial Q^t}{\partial Y}$, 并通过式(12)获得未标记样本对应的 $\frac{\partial Q^t}{\partial Y}$ 矩阵中最小值的位置 (i^*, j^*) 。

步骤 2.2 根据式(13)进行伪标记, 即令 $y_{i^*j^*}^{t+1} = 1, z_{i^*} = j^*$, 同时更新 Y 得到 Y^{t+1} 。

步骤 2.3 更新已标记样本集和未标记样本集:

$$X_l^{t+1} \leftarrow X_l^t + x_{i^*}, x_u^{t+1} \leftarrow X_u^t - x_{i^*}$$

步骤 2.4 根据式(3)更新标签规范化系数 V^t 。

步骤 2.5 根据式(11)计算偏导:

$$\left(\frac{\partial Q}{\partial \alpha}\right)^t = \left[\left(\frac{\partial Q}{\partial \alpha_1}\right)^t, \left(\frac{\partial Q}{\partial \alpha_2}\right)^t, \dots, \left(\frac{\partial Q}{\partial \alpha_k}\right)^t\right]$$

步骤 2.6 根据式(14)更新 $\alpha^t = \{\alpha_q^t\}$, 当 $\alpha_q^t < 0$ 时令 $\alpha_q^t = 0$ 。

步骤 2.7 计算 $\alpha_q^t = \alpha_q^t / \sum_{j=1}^k \alpha_j^t$ 使 α^t 归一化, 满足 $\sum_{q=1}^k \alpha_q^t = 1$ 。

步骤 2.8 更新迭代计数变量 $t = t + 1$ 。

步骤 2.9 重复循环直至满足 t 等于 T 。

步骤 3 根据式(7)计算 $\{F_q\}$, 通过图的权重系数 α 对多图的预测函数进行融合并预测未标记样本的标签 $y_{l+1}, y_{l+2}, \dots, y_n$:

$$F = \sum_{q=1}^k \alpha_q F_q$$

$$j^* = \arg \max_{x_i \in X_u, 1 \leq i \leq c} F_{i*}; i = l+1, l+2, \dots, n$$

$$y_{i^*j^*} = 1, z_{i^*} = j^*$$

MG-GTAM 通过 T 次迭代交替优化 V 和 Y 来求解一个局部最优解。对 V 和 Y 的求解需要的计算时间复杂度为 $O(n^2)$, 参数 α 的更新只需要线性计算时间, 因而 MG-GTAM 的时间复杂度为 $O(Tn^2)$, 通常情况下, $T \ll n$ 。MG-GTAM 不需要在每次迭代中重新计算拉普拉斯矩阵, 因此其计算效率较高, 适合在实际应用中使用。

3 实验及分析

本章实验使用文献[15]中涉及的 COIL20 和 NEC Animal 数据集来验证 MG-GTAM 的有效性。所用对比算法有高斯随机场和调和函数(Gaussian Fields and Harmonic Functions, GFHF)^[16]、局部和全局一致(Local and Global Consistency, LGC)^[17]、GTAM, MG-GTAM 和 CGL。其中 GFHF、LGC、GTAM 均为典型的单图上的图直推学习方法, 常被用于图直推学习的对比算法, CGL 算法则为基于多图直推学习方法。采用这两个数据集的原因是, 文献[15]的实验结果显示虽然 GTAM 在多个数据集上取得了比 GFHF、LGC 等方法高的分类精度, 但在这两个数据集上的分类精度仅略优于其他方法, 尤其是在 NEC Animal 数据集上的分类精度只有 0.5 左右。数据集 COIL20 来源于哥伦比亚大学图像数据库^[18], 包括 20 类共 1440 张物体灰度图像, 常被用于目标物体的识别任务。数据集 NEC Animal 来源于 Mabahi 等^[19]从视频中提取的图像, 用于图像相关性分析及分类任务。该数据集包括 60 个类动物

玩具图像, 每个类别的图像取自约 72 个不同角度, 共 4371 张。在实验前, 将 NEC Animal 图像裁剪到大小为 96×72 的灰度图像。

实验采用 Matlab 7.0, 在 Intel i7, 8 GB 内存, Windows 7 64 位系统上运行。算法的评价标准为算法在测试数据上的分类错误率。为避免随机性, 报告中的实验结果为 50 次独立运行结果的平均值。

3.1 不同数量标记下的分类性能

为了分析算法在不同数量标记下的分类精度变化情况, 分别在 COIL20 和 NEC Animal 数据集上执行实验。在这两个实验中, 先设定每个类取一个标记样本, 然后随机地增加每个类别标记样本的数量。MG-GTAM 与 CGL 采用相同的方法产生多图, 即采用不同的构图参数生成不同结构的图。具体设置如下: 采用 k 近邻方法进行图稀疏化, 近邻大小参数 k 从 2 增加到 7。采用二值置权法、高斯核函数置权法两种方法进行图上边的赋权。高斯核函数置权法中核宽度参数 σ 设定为样本与其 k 个近邻距离的平均值。基于以上设置为 MG-GTAM 与 CGL 产生 12 个不同结构的图。GFHF、LGC 和 GTAM 均采用文献[14]推荐的参数进行构图, 即近邻大小参数 $k = 6$, 采用高斯核函数置权。LGC、GTAM 和 MG-GTAM 中参数 μ 设置为 0.99。根据实验经验设定 MG-GTAM 中迭代终止次数参数 $T = 20$, 步长参数 $\eta = 1E-4$ 。MG-GTAM 与其他算法在 2 个数据集上的实验结果分别如图 1、图 2 所示。

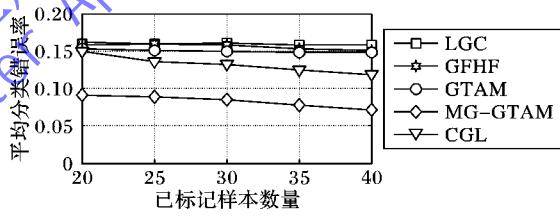


图 1 COIL20 上不同数量已标记样本下的算法分类错误率

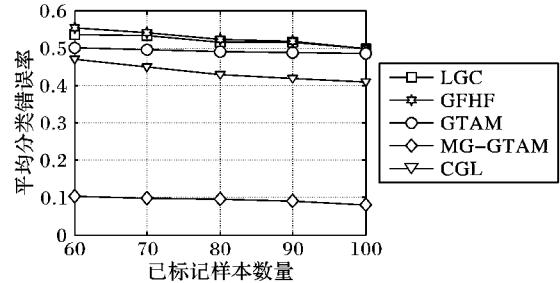


图 2 NEC Animal 上不同数量已标记样本下的算法分类错误率

从图 1 和图 2 中看出, 两个多图上的图直推学习方法在两个数据集上分类错误率低于其他单图上的图直推学习方法, 这是由于利用多个图更好地反映了样本之间的关系。MG-GTAM 在 COIL20 和 NEC Animal 数据集上的分类错误率远低于对比算法, 可见 MG-GTAM 在迭代中通过综合多个图上的信息选择置信度高的未标记样本进行伪标记和通过权重权衡各图的重要程度的策略具有显著成效。此外相比 COIL20 数据集, MG-GTAM 能在 NEC Animal 数据集上取得较大性能提升的原因可能是 NEC Animal 数据集内在分布和几何结构较 COIL20 复杂, 更适合利用多个图来反映了样本之间的关系。虽然 GTAM 分类错误率略优于 GFHF 和 LGC, 但这些算法都是在单个图上进行标签传播, 当图不能完全反映样本点之间的关系时, 其分类性能无法得到显著提升。当已标记样本的数量逐步增加时, 所有算法的分类错误率呈下降

趋势。这说明当已标记样本的数量较少时,增加标记样本数据有利于提升分类的性能。

值得指出的是,即使在已标记样本数量极少时(每个类别一个标记),MG-GTAM 在两个图像数据集上都取得了明显优于其他算法的分类性能。因此 MG-GTAM 适用于已标记样本数量较少的应用领域,如应用于图像检索任务中,根据少量查询来得到更为准确的查询结果。

3.2 迭代次数对分类精度的影响

经过 T 次迭代后, MG-GTAM 通过权重值融合多图的预测函数对未标记样本进行标记,而迭代的次数 T 对 MG-GTAM 的分类性能有一定的影响。图 3 显示了不同数量标记(label)下,NEC Animal 数据集上前 50 次迭代 MG-GTAM 的分类错误率变化情况。

从图 3 中可看出:通过较少次数的迭代, MG-GTAM 的分类错误率得以显著下降;但当迭代次数持续增加时, MG-GTAM 的分类错误率并不能得到显著的降低。出现这一现象的可能原因是,通过伪标记进行图上信息的交换使得各图上标签传播的差异逐步减少,以至于不足以支持进一步分类性能的提升。因此在实际应用中, MG-GTAM 需根据经验设置合适的迭代次数 T 来适时停止算法的迭代。

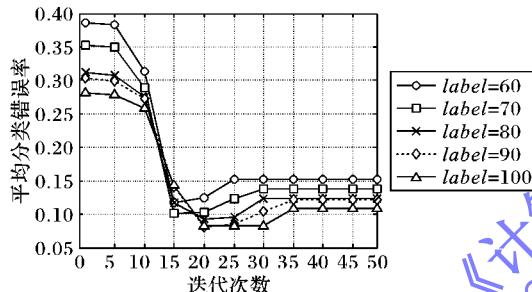


图 3 NEC Animal 上不同迭代次数下的分类错误率

3.3 构图参数对分类精度的影响

本节主要分析 MG-GTAM 性能对构图参数及图的个数的敏感性。实验中,高斯核函数置权法中核宽度参数 σ 设定为样本与其 k 个近邻距离的平均值,迭代终止次数参数 $T = 20$,步长参数 $\eta = 1E - 4$ 。

图上边的赋权策略的差异和 k 近邻法进行稀疏化时近邻的个数对图的结构具有一定影响,为了研究这两个构图参数对 MG-GTAM 分类性能的影响,在 NEC Animal 数据集分别执行两个实验。实验 1 中,分别采用二值置权法、高斯核函数置权法两方法进行图上边的赋权,图稀疏化方法仍采用 k 近邻,其中 $k = \{2, 3, 4, 5, 6, 7\}$,该设置为每次实验产生 6 个不同结构的图,对应的实验结果如图 4。实验 2 中, k 近邻图稀疏化方法分别组合采用 $k = \{2, 3\}, k = \{4, 5\}, k = \{6, 7\}, k = \{2, 7\}$,同时采用二值置权法、高斯核函数置权法两种方法进行图上边的赋权,该设置为每次实验产生 4 个不同结构的图,对应的实验结果如图 5。

从图 4 可以发现,虽然不同边赋权策略构图使 MG-GTAM 获得的性能具有一定差异,但都取得了比单图下 GTAM 更好的分类性能。同时分别采用边赋权策略构图的 MG-GTAM 的性能都低于同时采用两种边赋权策略构图时 MG-GTAM 取得的分类性能。这不仅说明了边赋权策略构图对 MG-GTAM 性能的影响和采用多图的有效性,而且说明了 MG-GTAM 可利用不同边赋权策略构造的多图更好地获取数据内在分布和几何结构,以提高分类性能。从图 5 可以看出,不同近邻的个数

组合对 MG-GTAM 分类性能有显著影响,当近邻个数较少时 MG-GTAM 取得较好的分类性能。同时 MG-GTAM 在 $k = \{2, 7\}$ 时的分类错误率比在 $knn = \{6, 7\}$ 和 $k = \{4, 5\}$ 时的分类错误率低,这说明 MG-GTAM 通过权重值来权衡每个图对标签传播的重要程度的策略不但有效,而且有效地利用了不同图上的信息来提升算法的分类性能。

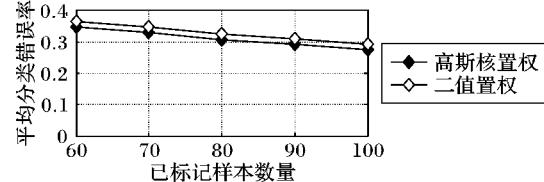


图 4 NEC Animal 上两种边赋权策略构图下 MG-GTAM 分类错误率

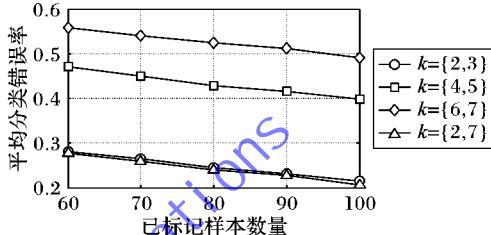


图 5 NEC Animal 上不同 k 值组合构图下 MG-GTAM 分类错误率

4 结语

本文提出了基于多图的交替优化图直推算法(MG-GTAM)。MG-GTAM 使用不同的图构建参数来构建多图,将单个图上二元函数最优化问题转换为多个图上的二元函数最优化问题。该方法借助交替迭代方式选择置信度高的未标记样本进行伪标记,优化多个图上未标记样本标记的一致性和平滑性,通过权重值融合多图的预测函数对未标记样本进行标记。实验结果表明, MG-GTAM 不依赖于多视图假设,无需选择特定的图结构,且与经典的 LGC、GFHF、GTAM、CGL 算法相比具有较好的分类准确率。

图的结构是影响基于图的半监督学习方法的关键因素,因此在未来研究工作中,将致力于研究多个图结构的选择及构造问题。此外,如何设计自适应停止迭代的方法使得 MG-GTAM 适时停止迭代也是未来的研究内容之一。

参考文献:

- [1] ZHU X. Semi-supervised learning literature survey, TR 1530 [R]. Madison: University of Wisconsin-Madison, 2005.
- [2] OLIVER C, BERNHARD S, ALEXANDER Z. Semi - supervised learning [M]. Cambridge: MIT Press, 2006: 9-10.
- [3] LIU J, LIU Y, LUO X. Study on semi-supervised learning methods [J/OL]. Chinese Journal of Computers, 2014, 37. [2014-12-02]. <http://www.cnki.net/kcms/detail/11.1826.TP.20140828.1917.005.html>.
- [4] JEBARA T, WANG J, CHANG S-F. Graph construction and b -matching for semi-supervised learning [C]// Proceedings of the 26th Annual International Conference on Machine Learning. New York: ACM, 2009: 441 - 448.
- [5] de SOUSA C A R, REZENDE S O, BATISTA G E A P A. Influence of graph construction on semi-supervised learning [C]// Proceedings of the 2013 European Conference on Machine Learning and Knowledge Discovery in Databases, LNCS 8190. Berlin: Springer-

- Verlag, 2013: 160 – 175.
- [6] SINDHWANI V, NIYOGO P, BELKIN M. A co-regularized approach to semi-supervised learning with multiple views [C] // Proceedings of the 22th International Conference on Machine Learning. Cambridge: MIT Press, 2005: 823 – 831.
- [7] SINDHWANI V, ROSENBERG D S. An RKHS for multi-view learning and manifold co-regularization [C] // Proceedings of the 25th International Conference on Machine Learning. New York: ACM, 2008: 976 – 983.
- [8] ZHANG D, WANG F, ZHANG C, et al. Multi-view local learning [C] // Proceedings of the 23th AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2008, 2: 752 – 757.
- [9] DU Y, LI Q, CAI Z, et al. Multi-view semi-supervised Web image classification via co-graph [J]. Neurocomputing, 2013, 122: 430 – 440.
- [10] ARGYRIOU A, HERBSTER M, PONTIL M. Combining graph Laplacians for semi-supervised learning [C] // Advances in Neural Information Processing Systems 18. Cambridge: MIT Press, 2006: 67 – 74.
- [11] ZHOU Z. Disagreement-based semi-supervised learning [J]. Acta Automatica Sinica, 2013, 39(11): 1871 – 1878. (周志华. 基于分歧的半监督学习 [J]. 自动化学报, 2013, 39(11): 1871 – 1878.)
- [12] MAIER M, von LUXBURG U, HEIN M. Influence of graph construction on graph-based clustering measures [EB/OL]. [2014 – 12 – 03]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.143.8751&rep=rep1&type=pdf>.
- [13] SHE Q, YU Y, JIANG Y, et al. Large-scale image annotation via random forest based label propagation [J]. Journal of Computer Research and Development, 2012, 49(11): 2289 – 2295. (余俏俏, 俞扬, 姜远, 等. 一种基于标记传播的大规模图像分类方法 [J]. 计算机研究与发展, 2012, 49(11): 2289 – 2295.)
- [14] WANG J, JEBARA T, CHANG S-F. Graph transduction via alternation minimization [C] // Proceedings of the 25th Annual International Conference on Machine Learning. New York: ACM, 2008: 1144 – 1151.
- [15] WANG J, JEBARA T, CHANG S-F. Semi-supervised learning using greedy max-cut [J]. Journal of Machine Learning Research, 2013, 14(1): 771 – 800.
- [16] ZHU X, GHAHRAMANI Z, LAFFERTY J. Semi-supervised learning using Gaussian fields and harmonic functions [C/OL] // Proceedings of the 20th International Conference on Machine Learning. [2014 – 12 – 03]. <http://www.aaai.org/Papers/ICML/2003/ICML03-118.pdf>.
- [17] ZHOU D, BOUSQUET O, LAL T N, et al. Learning with local and global consistency [C] // Proceedings of the 2004 Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2004: 321 – 328.
- [18] MURASE H, NENE S K. Visual learning and recognition of 3-D objects from appearance [J]. International Journal of Computer Vision, 1995, 14(1): 5 – 24.
- [19] MOBAHI H, COLLOBERT R, WESTON J. Deep learning from temporal coherence in video [C] // Proceedings of the 26th International Conference on Machine Learning. New York: ACM, 2009: 737 – 744.

(上接第 1599 页)

参考文献:

- [1] SHALEV S-S. Online learning and online convex optimization [J]. Foundations and Trends in Machine Learning, 2012, 4(2): 107 – 194.
- [2] WANG H, BANERJEE A. Online alternating direction method [C/OL] // Proceedings of the 2012 29th International Conference on Machine Learning. [2014 – 12 – 02]. <http://arxiv.org/abs/1206.6448?context=stat.ML>.
- [3] LUO L. Research on large-scale machine learning [J]. Ship Electronic Engineering, 2013, 33(2): 9 – 12 (罗霖. 大规模机器学习问题研究 [J]. 舰船电子工程, 2013, 33(2): 9 – 12.)
- [4] GAO Q. A new online optimization algorithm for non-smooth losses based on ADMM [J]. Computer Technology and Development, 2014, 24(2): 96 – 100 (高乾坤. 一种基于 ADMM 的非光滑损失在线优化算法 [J]. 计算机技术与发展, 2014, 24(2): 96 – 100.)
- [5] SCHIZAS I, RIBEIRO A, GIANNAKIS G. Consensus in Ad Hoc WSNs with noisy links — Part I: distributed estimation of deterministic signals [J]. IEEE Transactions on Signal Processing, 2008, 56 (1): 350 – 364.
- [6] NEDIC A, OZDAGLAR A. Distributed subgradient methods for multi-Agent optimization [J]. IEEE Transactions on Automatic Control, 2009, 54(1): 48 – 61.
- [7] DUCHI J, AGARWAL A, WAINWRIGHT M. Dual averaging for distributed optimization: Convergence analysis and network scaling [J]. IEEE Transactions on Automatic Control, 2012, 57 (3): 592 – 606.
- [8] ZINKEVICH M. Online convex programming and generalized infini-
- tesimal gradient ascent [EB/OL]. [2014 – 12 – 06]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.6680&rep=rep1&type=pdf>.
- [9] YAN F, SUNDARAM S, VISHWANATHAN S, et al. Distributed autonomous online learning: Regrets and intrinsic privacy-preserving properties [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(11): 2483 – 2493.
- [10] HOSSEINI S, CHAPMAN A, MESBAHI M. Online distributed optimization via dual averaging [C] // Proceedings of the 2013 IEEE 52nd Annual Conference on Decision and Control. Piscataway: IEEE, 2013: 1484 – 1489.
- [11] BAZERQUE J A, MATEOS G, GIANNAKIS G B. Group-Lasso on splines for spectrum cartography [J]. IEEE Transactions on Signal Processing, 2011, 59(10): 4648 – 4663.
- [12] SHI W, LING Q, YUAN K, et al. On the linear convergence of the ADMM in decentralized consensus optimization [J]. IEEE Transactions on Signal Processing, 2014, 62(7): 1750 – 1761.
- [13] KEKATOS V, GIANNAKIS G. Distributed robust power system state estimation [J]. IEEE Transactions on Power Systems, 2013, 28(2): 1617 – 1626.
- [14] MOTA J F C, XAVIER J M F, AGUIAR P M Q, et al. D-ADMM: a communication-efficient distributed algorithm for separable optimization [J]. IEEE Transactions on Signal Processing, 2013, 61(10): 2718 – 2723.
- [15] PIATT J C. Fast training of support vector machines using sequential minimal optimization [M]. Cambridge: MIT Press, 1999: 185 – 208.