

基于最大间隔和流形假设的半监督学习算法

戴伟, 柴晶, 刘雅娇

引用本文

戴伟, 柴晶, 刘雅娇. 基于最大间隔和流形假设的半监督学习算法[J]. 计算机科学, 2024, 51(2): 259-267.

DAI Wei, CHAI Jing, LIU Yajiao. [Semi-supervised Learning Algorithm Based on Maximum Margin and Manifold Hypothesis](#) [J]. Computer Science, 2024, 51(2): 259-267.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多教师网络模型的半监督语义分割方法](#)

Semi-supervised Semantic Segmentation Method Based on Multiple Teacher Network Model
计算机科学, 2023, 50(12): 279-284. <https://doi.org/10.11896/jsjcx.221000245>

[基于分类不确定性最小化的半监督集成学习算法](#)

Classification Uncertainty Minimization-based Semi-supervised Ensemble Learning Algorithm
计算机科学, 2023, 50(10): 88-95. <https://doi.org/10.11896/jsjcx.230600048>

[基于序贯三支决策的半监督目标检测算法](#)

Semi-supervised Object Detection with Sequential Three-way Decision
计算机科学, 2023, 50(10): 1-6. <https://doi.org/10.11896/jsjcx.230600035>

[双编码半监督异常检测模型](#)

Dually Encoded Semi-supervised Anomaly Detection
计算机科学, 2023, 50(7): 53-59. <https://doi.org/10.11896/jsjcx.220900027>

[基于FlexUDA模型的SQL注入检测研究](#)

Study on SQL Injection Detection Based on FlexUDA Model
计算机科学, 2023, 50(6A): 220600172-6. <https://doi.org/10.11896/jsjcx.220600172>

基于最大间隔和流形假设的半监督学习算法

戴伟 柴晶 刘雅娇

云南大学信息学院 昆明 650500

(daiwei@mail.ynu.edu.cn)

摘要 半监督学习是一种介于监督学习和无监督学习之间的弱监督学习模式,其在学习过程中将少量标记示例和大量未标记示例结合起来构建模型,以期取得比监督学习仅使用标记示例更高的学习精度。在该学习模式下,文中提出了一种将最大间隔准则和示例空间的流形假设思想相结合的半监督学习算法。该算法在利用示例流形结构估计未标记示例标记置信度的同时利用最大间隔准则构建分类模型,并采用交叉优化方法以迭代的方式交替地求解分类模型参数和标记置信度。在12个UCI数据集和4个由MNIST手写数字集生成的数据集上的实验结果表明,采用半监督直推学习方式,该算法的性能优于其他对比算法的情况为60.5%;采用半监督归纳学习方式,该算法的性能优于其他对比算法的情况为42.6%。

关键词: 半监督学习;最大间隔;流形假设;标记置信度;支持向量机

中图分类号 TP181

Semi-supervised Learning Algorithm Based on Maximum Margin and Manifold Hypothesis

DAI Wei, CHAI Jing and LIU Yajiao

School of Information Science and Engineering, Yunnan University, Kunming 650500, China

Abstract Semi-supervised learning is a weakly supervised learning pattern between supervised learning and unsupervised learning. It combines a small number of labeled instances with a large number of unlabeled instances to build a model during the process of learning, hoping to achieve better learning accuracy than supervised learning using only labeled instances. In this learning pattern, this paper proposes a semi-supervised learning algorithm that combines the maximum margin with manifold hypothesis of the instance space. The algorithm utilizes the manifold structure of instances to estimate the labeling confidence over unlabeled instances, at the same time utilizes the maximum margin to derive the classification model. And alternating optimization is adopted to address the quadratic programming problem of the model parameters and the labeling confidence in an iterative manner. On 12 UCI datasets and 4 datasets generated by the MNIST database of handwritten digits, in semi-supervised transductive learning, the proposed algorithm's performance outperforms the comparison algorithms for 60.5% of the configurations in semi-supervised inductive learning, the proposed algorithm's performance outperforms the comparison algorithms for 42.6% of the configurations.

Keywords Semi-supervised learning, Maximum margin, Manifold hypothesis, Labeling confidence, Support vector machine

1 引言

半监督学习(Semi-supervised Learning, SSL)^[1-4]是介于监督学习(Supervised Learning)和无监督学习(Unsupervised Learning)之间的一个机器学习的重要分支。无监督学习^[5]对事先未给定标记的示例进行学习建模,由于示例没有监督信息,因此其学习建模通常具有一定困难。监督学习^[6-7]则通过对示例及其相应标记组成的标记示例集进行学习建模,力图构建出能够对未见示例进行准确预测的分类模型或回归模型。监督学习需要大量的标记示例进行学习建模,但在许多现实

任务中,由于标记示例通常涉及必要的领域知识,收集起来可能费时费力。例如在计算机辅助医疗诊断^[8]中,可以从病人的常规检查中获得大量的未标记医学图像,但是要求医学专家给出所有图像的病灶标记则存在一定困难。相比之下,在许多任务中,获取大量未标记示例更容易。半监督学习可以在标记示例有限的情况下,充分利用大量未标记示例进行建模,这在很大程度上降低了对标记示例的需求量。半监督学习已经被证实许多实际应用中有有效,如计算机辅助医疗诊断^[8-9]、文本分类^[10]、图像检索^[11]、生物信息学^[12]、新冠(COVID-19)检测^[13]、自然语言处理^[14]和邮件分类^[15]。

到稿日期:2022-11-15 返修日期:2023-04-03

基金项目:国家自然科学基金(62166046);云南省智能系统与计算重点实验室开放课题(ISC23Y01)。

This work was supported by the National Natural Science Foundation of China(62166046) and Open Project Program of Yunnan Key Laboratory of Intelligent Systems and Computing(ISC23Y01).

通信作者:柴晶(jingchai@aliyun.com)

流形假设是半监督学习中的常见假设,即假设示例分布在某个低维的流形结构上,流形相近的示例拥有相似的标记。在具体实现时,通常使用图的方法来近似流形结构,从而衍生出了一些基于图的半监督学习(Graph-based Semi-supervised Learning, GSSL)算法^[16-21]。近似流形结构时所构建的图的质量对于基于流形假设的算法性能而言至关重要^[3],一个不合适的图可能导致基于图的半监督学习算法比仅使用标记示例的监督学习算法性能更差。然而,如何构建一个高质量的流形结构图仍是一个问题。

支持向量机(SVM)^[22]基于最大间隔准则,其目标是寻找具有最大间隔的划分超平面。半监督支持向量机(Semi-supervised Support Vector Machine, S3VM)^[10,23]是支持向量机在半监督学习问题上的推广。S3VM基于大间隔(Large Margin)或低密度划分(Low Density Assumption)原则^[2],试图寻找一个划分超平面使得所有标记示例和未标记示例到划分超平面之间都具有最大间隔。但是, Li等^[24]指出,当给定少量标记示例和大量未标记示例时,示例特征空间中可能会存在多个大间隔划分。在缺乏充足领域知识对这些大间隔划分加以选择的情况下,选择其中一种错误划分可能会严重影响算法性能。

通过对上述单独运用流形假设或最大间隔原则建模的问题进行分析并受文献^[25]中算法的启发,本文将最大间隔准则和示例空间的流形假设思想相结合,提出基于最大间隔和流形假设的半监督学习算法(semi-supervised learning algorithm based on MAximum MAargin and MAnifold Hypothesis, MA3H)。MA3H在构建最大间隔划分超平面的同时,通过流形正则化项将标记示例的标记置信度和未标记示例的标记置信度保持在流形结构上,并采用交叉优化方法交替地求解最大间隔划分超平面的权重系数和未标记示例的标记置信度。

本文第2章介绍相关工作;第3章阐述本文的算法原理和优化流程;第4章介绍相关实验结果;最后总结全文并展望未来。

2 相关工作

在过去的二十年里,学者们提出了大量的半监督学习算法。根据工作方法的的不同,它们可大致分为4类主流的学习方法:基于生成式模型的方法^[26-31]、基于分歧的方法^[9,32-35]、基于图的方法和基于半监督支持向量机的方法。为了给MA3H奠定基础,本章主要介绍后两类。

基于图的半监督学习算法是半监督学习的主流方法之一,该类算法直接或间接地利用了流形假设思想。Zhu等^[17]提出了基于半监督学习的标记传播算法(Label Propagation Algorithm, LPA), LPA首先构造一个图来近似流形结构并计算示例之间的相似度得到图的权重矩阵,由权重矩阵得到概率转移矩阵,然后以随机行走的方式将标记信息迭代传播到未标记示例中。在此基础上, Zhu等^[19]结合高斯随机场和谐波函数,利用图上的高斯随机场的谐波特性,提出了基于高斯随机场与谐波函数(Gaussian Fields and Harmonic Function, GFHF)的半监督学习算法,由于GFHF将真正的标记固定在

标记示例上,因此其不能很好地处理标记噪声。Zhou等^[16]提出了一种与GFHF密切相关的局部与全局一致性算法(Local and Global Consistency, LGC), LGC利用局部和全局一致性条件进行标记传播来弥补GFHF方法中的不足。Wang等^[36]提出了线性近邻重构算法(Linear Neighborhood Propagation, LNP), LNP假设每个示例都可以由其近邻示例的加权线性组合进行重构,并使用示例的局部线性重构计算示例之间的权重,保证了图中局部和全局流形结构的平滑性。Gong等^[37]提出了图拉普拉斯形变算法(Label Prediction algorithm via DGL, LPDGL), LPDGL将示例的局部信息加入平滑正则化中,实现了全局最优预测模型。Calder等^[38]基于图的半监督学习提出了Poisson learning框架,用于在标记示例比率非常低的情况下基于图的半监督学习。

半监督支持向量机是半监督学习的另一个主流方法, Bennett等^[39]提出了最初的S3VM,为了能够利用未标记示例的信息,其在SVM的基础上,对训练集中每个未标记示例引入两个松弛项,用于惩罚对未标记示例的错误分类并计算其可能的错分风险,从而将未标记示例归类到其错分风险最小的类中。Joachims^[10]提出了直推式支持向量机(Transductive Support Vector Machine, TSVM),其通过添加新的约束项将未标记示例的标记指派(Label Assignment)加入到算法的优化函数中,从而寻找一个在标记示例和进行了标记指派的未标记示例上间隔最大化的决策边界。S3VM算法中的优化问题是混合整数规划问题,具有较高的计算复杂度。针对该问题, Li等^[40]通过对未标记示例类中心的估计,提出了MeanS3VM。MeanS3VM使两个类别的未标记示例的类中心之间的间隔最大化;通过考虑类中心,不再考虑S3VM算法中对每个未标记示例的限制。因此, MeanS3VM提升了S3VM的学习效率。Li等^[41]在MeanS3VM的基础上引入代价敏感学习,提出了代价敏感的半监督支持向量机(Cost-Sensitive Semi-supervised Support Vector Machine, CS4VM), CS4VM把代价敏感学习和MeanS3VM相结合,对未标记示例不同的错误分类采用不同的代价,赋予了S3VM处理不均衡代价的能力。

此后,针对在某些情况下,半监督学习可能出现性能下降的情况, Li等^[8]提出了安全半监督学习(Safe Semi-supervised Support Vector Machine, S4VM)。S3VM试图寻找一个具有最大间隔的划分超平面, S4VM则同时寻找多个具有大间隔的划分超平面,再通过优化求解最优的决策边界。

基于上述研究现状,本文提出了一种将最大间隔准则和流形假设思想相结合的半监督学习算法MA3H,下文将对MA3H的具体原理和流程进行详细描述。

3 MA3H 算法

3.1 问题定义

设 $\mathcal{X}=\mathbb{R}^n$ 表示 n 维特征空间, $\mathcal{Y}=\{y_1, y_2, \dots, y_q\}$ 表示含有 q 类标记的类别空间。给定标记示例集 $\mathcal{D}_L=\{(\mathbf{x}_i, y_i) \mid 1 \leq i \leq l\}$ 和未标记示例集 $\mathcal{D}_U=\{x_i \mid l+1 \leq i \leq l+u\}$, 其中 $\mathbf{x}_i \in \mathcal{X}$ 是示例的 n 维特征向量 $[x_{i1}, x_{i2}, \dots, x_{in}]^T$, $y_i \in \mathcal{Y}$ 是示例 \mathbf{x}_i 的类别标记, 通常令 $l \ll u$ 。半监督学习的目标是从示例集合 $\mathcal{D}_L \cup$

\mathcal{Q}_U 中学习一个分类器 $f: \mathcal{X} \rightarrow \mathcal{Y}$.

3.2 模型构建

设 $f_i = [f_{i1}, f_{i2}, \dots, f_{iq}]^T$ 表示示例 $x_i (1 \leq i \leq l+u)$ 的标记置信度向量。未标记示例的置信度矩阵为 $F_U = [f_{l+1}, f_{l+2}, \dots, f_{l+u}]^T \in [0, 1]^{u \times q}$ 且 $\sum_{c=1}^q f_{ic} = 1$; 标记示例的置信度矩阵为 $F_L = [f_1, f_2, \dots, f_l]^T \in (0, 1)^{l \times q}$ 。由于标记示例的类别标记已知, 因此对于任意标记示例 $(x_i, y_i) (1 \leq i \leq l)$, 有:

$$f_{ic} = \begin{cases} 1, & \text{if } c \in \mathcal{Y} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

定义特征空间映射函数 $\Phi(x, y): \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}^{nq}$ 如下:

$$\Phi(x, y) = \begin{bmatrix} x \cdot \mathbb{I}(y=y_1) \\ x \cdot \mathbb{I}(y=y_2) \\ \dots \\ x \cdot \mathbb{I}(y=y_q) \end{bmatrix} \quad (2)$$

其中, $\mathbb{I}(\cdot)$ 为指示函数, $\Phi(x, y)$ 将示例的原始特征映射为特定于示例类别标记的联合特征。

3.2.1 流形假设

MA3H 基于流形假设来保持未标记示例和标记示例的标记置信度处于流形结构上, 并使用图的方法来逼近该流形。首先在 \mathcal{Q}_L 和 \mathcal{Q}_U 上构建一个加权图 $G=(L, U, E, W)$, 其中顶点集 L 和 U 分别表示标记示例 $x_i (1 \leq i \leq l)$ 和未标记示例 $x_i (l+1 \leq i \leq l+u)$ 在图中的顶点集合。边集 E 包含从 L 到 U 的有向边集合, 对于每个未标记示例 $x_i \in \mathcal{Q}_U$, 找到其在标记示例集 \mathcal{Q}_L 中的 k 近邻 $\mathcal{N}(x_i)$, 则图 G 中的边集 E 可表示为 $E = \{(x_j, x_i) \mid x_j \in \mathcal{N}(x_i), l+1 \leq i \leq l+u\}$ 。 W 表示 E 中未标记示例和标记示例之间的相似度权重矩阵, $W = [\omega_{i,j}]_{u \times l}$ (如果 $(x_j, x_i) \in E$, 则 $\omega_{i,j} \geq 0$, 否则 $\omega_{i,j} = 0$)。本文采用文献[42]中的方法计算 W , 该方法通过计算示例在近邻空间的加权重构来确定相应权重。设 $\hat{w}_i = [\hat{w}_{i,j_1}, \dots, \hat{w}_{i,j_k}]^T (j_a \in \mathcal{N}(x_i), 1 \leq a \leq k)$ 表示未标记示例 x_i 和其在标记示例中 k 近邻 $\mathcal{N}(x_i)$ 之间的权重向量, $\mathcal{N}(x_i)$ 中示例 x_{j_a} 在重构 x_i 时的权重系数 ω_{i,j_a} 可由以下优化问题求得:

$$\min_{w_i} \left\| x_i - \sum_{a=1}^k \omega_{i,j_a} \cdot x_{j_a} \right\| \quad (3)$$

s. t. $\omega_{i,j_a} \geq 0 (j_a \in \mathcal{N}(x_i), 1 \leq a \leq k)$

如式(3)所示, 通过拟合具有非负约束的线性最小二乘问题来优化权重向量 w_i , 该问题可以通过二次规划求解器求解。进一步对 W 行归一化得到归一化后的权重矩阵 $S = [s_{ij}]_{u \times l}$:

$$S = D^{-1}W \quad (4)$$

其中, $D = \text{diag}[d_1, d_2, \dots, d_u]$, $d_i = \sum_{j=1}^l \omega_{i,j}$ 。

算法 1 给出了构建加权图的具体算法流程。

算法 1 加权图构建算法

输入: $(\mathcal{Q}_L, \mathcal{Q}_U, k)$

输出: S

1. 初始化权重矩阵 $W = [\omega_{i,j}]_{u \times l}$
2. for $i=1$ to u do
3. 找到未标记示例 x_i 在 \mathcal{Q}_L 中的 k 近邻 $\mathcal{N}(x_i)$;
4. 通过 x_i 和 $\mathcal{N}(x_i)$ 求解式(3)确定权重矩阵 $\hat{w}_i = [\hat{w}_{i,j_1}, \dots, \hat{w}_{i,j_k}]^T$;

5. for $j_a \in \mathcal{N}(x_i)$ do

6. set $\omega_{i,j_a} = \hat{w}_{i,j_a}$

7. end for

8. end for

9. 对权重矩阵 W 行归一化得到 $S = D^{-1}W$

最后, 为保持标记示例和未标记示例的标记置信度处在图 G 的流形结构上, 需最小化如下流形正则化项:

$$\sum_{i=1}^u \sum_{j=1}^l s_{ij} \|f_{l+i} - f_j\|_2^2 \quad (5)$$

其中, s_{ij} 是图 G 中未标记示例顶点和标记示例顶点之间的相似度权重值。如果两个顶点由具有较大权重的边进行连接, 则它们的置信度向量应趋于一致, 以保持流形结构的平滑性。

3.2.2 MA3H 模型

联合优化最大间隔划分超平面权重向量 $w \in \mathbb{R}^{nq}$ 和流形正则化项 $\sum_{i=1}^u \sum_{j=1}^l s_{ij} \|f_{l+i} - f_j\|_2^2$, 得到基于最大间隔和流形假设的半监督学习模型 MA3H 如下:

$$\begin{aligned} \min_{w, \Xi, F_U} & \frac{1}{2} \|w\|_2^2 + \frac{\lambda}{l} \sum_{i=1}^u \sum_{c=1}^q f_{ic} \xi_c + \frac{\mu}{u} \sum_{i=l+1}^{l+u} \sum_{c=1}^q f_{ic} \xi_c + \\ & \gamma \sum_{i=1}^u \sum_{j=1}^l s_{ij} \|f_{l+i} - f_j\|_2^2 \\ \text{s. t. } & w^T \Phi(x_i, y_c) - \max_{y_{c'} \neq y_c} w^T \Phi(x_i, y_{c'}) \geq 1 - \xi_c, \\ & (1 \leq i \leq l+u, 1 \leq c \leq q) \\ & \xi_c \geq 0, (1 \leq i \leq l+u, 1 \leq c \leq q) \\ & f_{ic} \geq 0, (l+1 \leq i \leq l+u, 1 \leq c \leq q) \\ & \sum_{c=1}^q f_{ic} = 1, (l+1 \leq i \leq l+u) \end{aligned} \quad (6)$$

其中, $\Xi = [\xi_c]_{(l+u) \times q}$ 是一系列用于调整分类间隔的松弛变量, 每个示例 $x_i (1 \leq i \leq l+u)$ 中有 q 个松弛变量 $\xi_c (1 \leq c \leq q)$; y_c 是 x_i 的真实类别标记。第一项是划分超平面权重向量 w ; 第二项和第三项是示例的松弛变量 ξ_c , 其由示例标记置信度 f_{ic} 进行加权调节; 第四项是标记置信度的流形正则化项, 用于保持未标记示例和标记示例的标记置信度处于平滑的流形结构上。

3.3 交叉优化

MA3H 可以采用文献[25]中的交叉优化的方法以迭代的方式交替更新 w 和 F_U 。

3.3.1 固定 w , 优化 F_U

当 w 固定时, 松弛变量 ξ_c 根据式(6)的第一项和第二项约束可由式(7)计算得到。

$$\xi_c = \max(0, 1 + \max_{y_{c'} \neq y_c} w^T \Phi(x_i, y_{c'}) - w^T \Phi(x_i, y_c)) \quad (7)$$

则式(6)可被转换成下列二次规划问题:

$$\begin{aligned} \min_{F_U} & \frac{\mu}{u} \sum_{i=l+1}^{l+u} \sum_{c=1}^q f_{ic} \xi_c + \gamma \sum_{i=1}^u \sum_{j=1}^l s_{ij} \|f_{l+i} - f_j\|_2^2 \\ \text{s. t. } & f_{ic} \geq 0, (l+1 \leq i \leq l+u, 1 \leq c \leq q) \\ & \sum_{c=1}^q f_{ic} = 1, (l+1 \leq i \leq l+u) \end{aligned} \quad (8)$$

式(8)是一个含有 uq 个变量和 $u(q+1)$ 个约束的二次规划问题。为降低计算复杂度, 在优化 F_U 中未标记示例 x_i 标记置信度 f_i 的同时固定 F_U 中的其他分量, 可进一步将式(8)分解成 u 个下列二次规划子问题:

$$\begin{aligned} \min_{f_i} & \gamma f_i^T f_i + \left(\frac{\mu}{u} \xi_i - 2\gamma \sum_{j=1}^l s_{ij} f_j \right)^T f_i \\ \text{s. t.} & f_{ic} \geq 0, (1 \leq c \leq q) \\ & \sum_{c=1}^q f_{ic} = 1 \end{aligned} \quad (9)$$

其中, $\xi_i = [\xi_{i1}, \xi_{i2}, \dots, \xi_{iq}]$ 是 x_i 的松弛变量。

3.3.2 固定 F_U , 优化 w

当 F_U 固定时, 式(6)的优化问题可以转换为如下问题:

$$\begin{aligned} \min_{w, \xi} & \frac{1}{2} \|w\|_2^2 + \frac{\lambda}{l} \sum_{i=1}^l \sum_{c=1}^q f_{ic} \xi_{ic} + \frac{\mu}{u} \sum_{i=l+1}^{l+u} \sum_{c=1}^q f_{ic} \xi_{ic} \\ \text{s. t.} & w^T \Phi(x_i, y_c) - \max_{y_c' \neq y_c} w^T \Phi(x_i, y_c') \geq 1 - \xi_{ic}, \\ & (1 \leq i \leq l+u, 1 \leq c \leq q) \\ & \xi_{ic} \geq 0, (1 \leq i \leq l+u, 1 \leq c \leq q) \end{aligned} \quad (10)$$

为求解方便, 进一步将式(10)中的两个约束条件写为: $w^T \Phi(x_i, y_c) + \delta_{cr} - w^T \Phi(x_i, y_r) \geq 1 - \xi_{ic}$ ($1 \leq i \leq l+u, 1 \leq c, r \leq q$), 其中, 如果 $c=r$ 则 $\delta_{cr}=1$, 否则 $\delta_{cr}=0$ 。由拉格朗日乘法可得式(10)的拉格朗日函数为:

$$\begin{aligned} L(w, \alpha) = & \frac{1}{2} \|w\|_2^2 + \frac{\lambda}{l} \sum_{i=1}^l \sum_{c=1}^q f_{ic} \xi_{ic} + \frac{\mu}{u} \sum_{i=l+1}^{l+u} \sum_{c=1}^q f_{ic} \xi_{ic} \\ & + \sum_{i=1}^{l+u} \sum_{c=1}^q \sum_{r=1}^q \alpha_{cr}^i (w^T \Phi(x_i, y_r) - w^T \Phi(x_i, y_c) \\ & - \delta_{cr} + 1 - \xi_{ic}) \end{aligned} \quad (11)$$

其中, $\alpha = [\alpha_{11}^1, \dots, \alpha_{lr}^l, \dots, \alpha_{(l+u)r}^{l+u}]^T$ 是相应示例的拉格朗日乘子且 $\alpha_{cr}^i \geq 0$ ($1 \leq i \leq l+u, 1 \leq c, r \leq q$)。令 $L(w, \alpha)$ 关于 w 和 α 的偏导数为 0, 可得式(10)的对偶问题如下:

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \sum_{i=1}^{l+u} \sum_{j=1}^{l+u} x_i^T x_j \sum_{c=1}^q \left(\sum_{r=1}^q \alpha_{cr}^i - \sum_{r=1}^q \alpha_{rc}^i \right) \left(\sum_{r=1}^q \alpha_{cr}^i - \sum_{r=1}^q \alpha_{rc}^i \right) + \\ & \sum_{i=1}^{l+u} \sum_{c=1}^q \sum_{r=1}^q \alpha_{cr}^i \delta_{cr} \\ \text{s. t.} & \alpha_{cr}^i \geq 0, 1 \leq i \leq l+u, 1 \leq c, r \leq q \end{aligned} \quad (12)$$

式(12)是一个含有 $(l+u)q^2$ 个变量和 $(l+u)q^2$ 个约束的二次规划问题。为求解方便, 进一步将其分解成对每个示例 x_i 的拉格朗日乘子 $\alpha^i = [\alpha_{cr}^i]_{q \times q}$ 单独处理的 $l+u$ 个子问题并引入矩阵 $M \in \{0, 1\}^{q \times q^2}$ 和 $N \in \{0, 1\}^{q \times q^2}$ 。

$$M = [I_{q \times q}, \dots, I_{q \times q}] \quad (13)$$

$$N = \begin{bmatrix} I_{1 \times q} & \mathbf{0}_{1 \times q} & \cdots & \mathbf{0}_{1 \times q} \\ \mathbf{0}_{1 \times q} & I_{1 \times q} & \cdots & \mathbf{0}_{1 \times q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{1 \times q} & \mathbf{0}_{1 \times q} & \cdots & I_{1 \times q} \end{bmatrix} \quad (14)$$

其中, $I_{q \times q}$ 是单位矩阵。优化 α^i 时, 固定 α 中的其他拉格朗日乘子, 则式(12)可以写为:

$$\begin{aligned} \min_{\alpha^i} & \frac{1}{2} x_i^T x_i \text{vec}(\alpha^i)^T C^T C \text{vec}(\alpha^i) + \left(\sum_{j \neq i} x_j^T x_j C^T C \text{vec}(\alpha^i) + \right. \\ & \left. \text{vec}(I_{q \times q}) \right)^T \text{vec}(\alpha^i) \\ \text{s. t.} & \alpha_{cr}^i \geq 0, (1 \leq c, r \leq q) \end{aligned} \quad (15)$$

其中, $\text{vec}(\cdot)$ 将矩阵拉伸为长向量, $C = M - N$ 。

当交叉优化结束后, 根据 Karush-Kuhn-Tucker(KKT) 条件可得划分超平面权重向量 $w = [w_1; w_2; \dots; w_q]$ 为:

$$w_r = \sum_{i=1}^{l+u} \left(\sum_{c=1}^q \alpha_{rc}^i - \sum_{c=1}^q \alpha_{cr}^i \right) x_i \quad (1 \leq r \leq q) \quad (16)$$

最后, 对于训练集中未标记示例或者未见示例 x_* , MA3H 对该示例的预测类别为:

$$y_* = \underset{y \in \mathcal{Y}}{\text{argmax}} w^T \Phi(x_*, y) \quad (17)$$

MA3H 算法的伪代码如算法 2 所示。

算法 2 MA3H 算法

输入: $(\mathcal{D}_L, \mathcal{D}_U, \lambda, \mu, \gamma, x_*)$

输出: y_*

1. 初始化未标记示例的置信度矩阵权重矩阵 F_U ;
2. 重复
3. 通过求解式(15)中的一系列二次规划子问题求解 α ;
4. 通过求解式(16)更新 w ;
5. 通过求解式(8)中的一系列二次规划子问题更新 F_U ;
6. 直到收敛
7. 通过求解式(17)返回 y_* 。

4 实验

本章首先介绍实验中所使用的数据集和对比算法, 然后从半监督的直推式和归纳式两种学习方式来进行对比实验。

4.1 实验数据集

实验所用数据集来源于 12 个 UCI 数据集¹⁾ 和 4 个由 MNIST 手写数字数据集²⁾ 生成的二分类数据集。MNIST 一共包含 45 对二分类问题, 本文选择其中 4 对 (mnist4vs9, mnist3vs8, mnist7vs9, mnist1vs7) 分类难度最大的二分类问题^[8], 采用文献[8]中的做法, 去除上述 4 对数据集中特征全 0 的维度。表 1 列出了这 16 个半监督数据集的主要特征。

表 1 实验数据集的特征

Table 1 Characteristics of experimental datasets

数据集	特征维度	示例数目		
		正示例	负示例	总计
house	16	108	124	232
haberman	14	81	225	306
isolet	51	300	300	600
australian	42	383	307	690
german	59	300	700	1000
optdigits	42	572	571	1143
digit1	241	766	734	1500
ethn	30	1310	1320	2630
sat	36	1508	1533	3041
spambase	57	1813	2788	4601
mushroom	21	3916	4208	8124
mnist4vs9	629	6824	6958	13782
mnist3vs8	631	7141	6825	13966
mnist7vs9	600	7293	6958	14251
mnist1vs7	652	7877	7293	15170
magic	10	12332	6688	19020

实验在每个数据集上均采用十折交叉验证法对各算法进行评估, 同时记录各算法的平均准确度和标准差。给定数据集 \mathcal{D} , 首先将其平均分成 10 份, 每次轮流将其中 1 份作为测试集 $\mathcal{D}_{\text{test}}$, 其余 9 份作为训练集 $\mathcal{D}_{\text{train}}$ 。采用文献[8]中划分标记示例和未标记示例的做法: 在 $\mathcal{D}_{\text{train}}$ 中随机选择 l ($l=10, 100$) 个示例作为标记示例集 \mathcal{D}_L , 其余作为未标记示例集 \mathcal{D}_U 。为分析 MA3H 算法的性能, 分别以直推式和归纳式两种

¹⁾ <https://archive.ics.uci.edu/ml/datasets.php>

²⁾ <http://yann.lecun.com/exdb/mnist/>

半监督学习方式对比实验,直推式在训练集 $\mathcal{D}_{\text{train}}$ 上训练,对 \mathcal{D}_{U} 中示例进行预测,归纳式则对 $\mathcal{D}_{\text{test}}$ 中示例进行预测。

4.2 对比算法

将 MA3H 算法与 4 个经典的半监督学习算法和 1 个基准的监督学习算法进行对比。SVM(只使用 \mathcal{D}_{L} 训练)为基准监督学习算法,4 个半监督算法包括:基于图的学习算法 LNP^[36]和 LGC^[16]、基于半监督支持向量机的 S3VM^[10]和安全半监督支持向量机算法 S4VM^[8]。

实验中各对比算法参数设置如下:监督 SVM 和 S3VM 算法中的正则化参数按照文献[2]中建议的参数进行设置;S4VM 中正则化参数 C_1 和 C_2 、采样数量 N 、分类器数量 T 和风险系数 λ 按照文献[8]中建议的参数进行设置;LGC 中平衡因子 α 按照文献[17]中建议的参数进行设置,近邻数目 k

设置为 3(与 MA3H 相同);LNP 中平衡因子 α 按照文献[37]中建议的参数进行设置,近邻数目 k 设置为 3(与 MA3H 相同)。如算法 1 所示,MA3H 中近邻数目 k 设置为 3(与 LGC 和 LNP 相同),如算法 2 所示,MA3H 中参数 λ 和 μ 从 $\{0.001,0.01,0.1,1,5,10,50,100\}$ 中选取,参数 γ 固定为 0.01。

4.3 实验结果

表 2 和表 3 分别列出了在标记示例数目 $l=10$ 和 $l=100$ 时,MA3H 与对比算法的直推式学习实验结果。在每个数据集上性能最优的结果均加粗表示。基于显著性水平为 0.05 的配对 t 检验,使用 \cdot/\circ 表示 MA3H 是否明显优于/差于对比算法。表 4 列出了在标记示例 $l=10$ 和 $l=100$ 的情况下,MA3H 算法优于(win)/持平(tie)/差于(loss)对比算法的实验次数。

表 2 MA3H 与对比算法在 10 个有标记示例($l=10$)直推式学习方式上的分类准确度

Table 2 Classification accuracy of MA3H and each comparison algorithm on 10 labeled instances($l=10$) in transductive learning

数据集	SVM	LGC	LNP	S3VM	S4VM	MA3H
house	0.918±0.030	0.897±0.014	0.901±0.018	0.851±0.063	0.919±0.026	0.914±0.027
haberman	0.634±0.068	0.619±0.122	0.601±0.119	0.613±0.058	0.646±0.067	0.636±0.121
isolet	0.871±0.071	0.846±0.085	0.869±0.053	0.860±0.110	0.890±0.098	0.990±0.006
australian	0.686±0.167	0.646±0.096	0.645±0.103	0.701±0.184	0.684±0.173	0.709±0.119
german	0.639±0.038	0.660±0.041	0.646±0.037	0.612±0.046	0.640±0.038	0.667±0.032
optdigits	0.931±0.058	0.900±0.101	0.940±0.051	0.896±0.095	0.941±0.073	0.976±0.025
digit1	0.764±0.082	0.721±0.097	0.746±0.087	0.825±0.055	0.764±0.082	0.833±0.049
ethn	0.694±0.057	0.606±0.063	0.666±0.067	0.777±0.096	0.729±0.089	0.687±0.185
sat	0.982±0.023	0.989±0.007	0.990±0.007	0.901±0.089	0.946±0.060	0.995±0.001
spambase	0.609±0.074	0.686±0.069	0.687±0.062	0.602±0.123	0.600±0.119	0.670±0.084
mushroom	0.769±0.116	0.697±0.096	0.772±0.083	0.731±0.122	0.735±0.119	0.783±0.103
mnist4vs9	0.675±0.094	0.618±0.094	0.648±0.086	0.692±0.084	0.691±0.098	0.737±0.129
mnist3vs8	0.763±0.077	0.727±0.064	0.754±0.073	0.778±0.067	0.786±0.079	0.840±0.061
mnist7vs9	0.769±0.055	0.708±0.053	0.754±0.047	0.764±0.069	0.772±0.083	0.837±0.046
mnist1vs7	0.903±0.046	0.897±0.106	0.934±0.035	0.876±0.094	0.890±0.090	0.945±0.021
magic	0.576±0.150	0.654±0.122	0.637±0.127	0.642±0.131	0.637±0.156	0.666±0.123

表 3 MA3H 与对比算法在 100 个有标记示例($l=100$)直推式学习方式上的分类准确度

Table 3 Classification accuracy of MA3H and each comparison algorithm on 100 labeled instances($l=100$) in transductive learning

数据集	SVM	LGC	LNP	S3VM	S4VM	MA3H
house	0.925±0.038	0.930±0.022	0.929±0.025	0.919±0.032	0.925±0.034	0.950±0.024
haberman	0.739±0.043	0.686±0.039	0.662±0.041	0.698±0.045	0.725±0.033	0.743±0.032
isolet	0.989±0.005	0.971±0.011	0.973±0.008	0.938±0.039	0.982±0.019	0.993±0.005
australian	0.805±0.022	0.783±0.024	0.773±0.023	0.796±0.025	0.813±0.026	0.848±0.011
german	0.645±0.032	0.683±0.015	0.680±0.018	0.658±0.027	0.650±0.030	0.699±0.015
optdigits	0.992±0.002	0.993±0.004	0.994±0.003	0.955±0.026	0.991±0.008	0.997±0.001
digit1	0.897±0.013	0.926±0.016	0.933±0.014	0.899±0.015	0.900±0.011	0.929±0.018
ethn	0.884±0.022	0.824±0.038	0.858±0.036	0.909±0.024	0.886±0.021	0.919±0.030
sat	0.997±0.002	0.995±0.001	0.995±0.002	0.961±0.031	0.996±0.004	0.996±0.001
spambase	0.760±0.043	0.806±0.019	0.807±0.021	0.827±0.016	0.823±0.013	0.824±0.015
mushroom	0.926±0.024	0.940±0.010	0.961±0.013	0.921±0.024	0.922±0.020	0.920±0.012
mnist4vs9	0.910±0.013	0.843±0.025	0.856±0.023	0.897±0.017	0.916±0.013	0.931±0.010
mnist3vs8	0.919±0.010	0.920±0.012	0.924±0.010	0.908±0.016	0.926±0.009	0.914±0.012
mnist7vs9	0.911±0.011	0.901±0.017	0.908±0.017	0.892±0.040	0.909±0.035	0.919±0.007
mnist1vs7	0.979±0.006	0.976±0.006	0.979±0.005	0.963±0.027	0.967±0.021	0.969±0.004
magic	0.749±0.024	0.760±0.013	0.747±0.016	0.774±0.011	0.773±0.014	0.775±0.006

表 4 MA3H 与各对比算法在直推式学习方式上的胜/平/负统计

Table 4 Win/tie/loss counts between MA3H and each comparison algorithm in transductive learning

标记示例 数目 l	MA3H 对比				
	SVM	LGC	LNP	S3VM	S4VM
$l=10$	11/5/0	11/5/0	9/7/0	10/5/1	9/7/0
$l=100$	9/6/1	11/3/2	11/2/3	9/7/0	7/8/1
总计	20/11/1	22/8/2	20/9/3	19/12/1	16/15/1

在总共 32 个(16 个数据集 \times 2 个 l 值)直推式学习结果中,经统计发现:

(1)对比基准 SVM 算法,MA3H 优于 SVM 的情况为 62.5%,与 SVM 持平的情况为 34%。说明采用直推式学习时,在大多数情况下 MA3H 不会比仅使用标记示例的 SVM 性能更差。

(2)对比基于图的半监督学习算法,MA3H 优于 LGC 的情况为 68%,与 LGC 持平的情况为 25%。MA3H 优于 LNP 的情况为 62.5%,与 LNP 持平的情况为 25%。

(3)对比基于支持向量机的半监督算法,MA3H 优于 S3VM 的情况为 59.4%,与 S3VM 持平的情况为 37.5%。MA3H 优于 S4VM 的情况占 50%,与 S4VM 持平的情况为 46.9%。

为评估 MA3H 归纳式学习性能,本文进一步采用归纳式学习测试 MA3H 和对比较算法的分类性能。鉴于 LGC 与 LNP 没有较好的归纳式学习实现方法,且 S3VM 和 S4VM 的

归纳式学习在数据规模较大时计算复杂度太高,本文仅在较小规模的数据集上对基准 SVM,S3VM 和 S4VM 进行归纳式学习。

表 5 和表 6 分别列出了当标记示例数目 $l=10$ 和 $l=100$ 时 MA3H 与对比算法的归纳式学习实验结果。在每个数据集上性能最优的结果均加粗表示。基于显著性水平为 0.05 的配对 t 检验,使用 \cdot/\circ 表示 MA3H 是否明显优于/差于对比算法。表 7 列出了在标记示例 $l=100$ 和 $l=100$ 情况下,MA3H 算法优于(win)/持平(tie)/差于(loss)对比算法的实验次数。

表 5 MA3H 与对比算法在 10 个有标记示例($l=10$)归纳式学习方式上的分类准确度

Table 5 Classification accuracy of MA3H and each comparison algorithm on 100 labeled instances($l=10$) in inductive learning

数据集	SVM	S3VM	S4VM	S4VM
house	0.927±0.057	0.840±0.096 ·	0.927±0.053	0.936±0.049
haberman	0.631±0.113	0.631±0.098	0.644±0.105	0.653±0.112
isolet	0.860±0.112 ·	0.852±0.113 ·	0.885±0.117 ·	0.988±0.016
australian	0.658±0.170	0.657±0.186	0.654±0.180	0.684±0.126
german	0.652±0.058 ·	0.638±0.052 ·	0.655±0.059 ·	0.688±0.054
optdigits	0.940±0.042 ·	0.897±0.093 ·	0.953±0.052	0.973±0.030
digit1	0.758±0.086	0.829±0.061	0.759±0.086	0.819±0.059
ethn	0.687±0.069	0.786±0.097 ·	0.711±0.107	0.671±0.185
sat	0.983±0.019	0.900±0.090 ·	0.961±0.046	0.995±0.005

表 6 MA3H 与对比算法在 100 个有标记示例($l=100$)归纳式学习方式上的分类准确度

Table 6 Classification Accuracy of MA3H and each comparison algorithm on 100 labeled instances($l=100$) in inductive learning

数据集	SVM	S3VM	S4VM	S4VM
house	0.922±0.044 ·	0.905±0.082 ·	0.922±0.044 ·	0.961±0.038
haberman	0.744±0.096	0.675±0.112 ·	0.728±0.087	0.744±0.079
isolet	0.995±0.008	0.952±0.058	0.997±0.007	0.995±0.008
australian	0.803±0.057 ·	0.801±0.063 ·	0.826±0.059	0.851±0.050
german	0.659±0.061	0.655±0.053 ·	0.651±0.047 ·	0.701±0.039
optdigits	0.990±0.009 ·	0.962±0.028 ·	0.994±0.008	0.997±0.004
digit1	0.905±0.026 ·	0.903±0.031 ·	0.911±0.023	0.923±0.023
ethn	0.892±0.038	0.915±0.038	0.892±0.038	0.908±0.029
sat	0.997±0.003	0.958±0.035 ·	0.997±0.002	0.995±0.004

表 7 MA3H 与各对比算法在归纳式学习方式上的胜/平/负统计

Table 7 Win/tie/loss counts between MA3H and each comparison algorithm in inductive learning

标记示例数目 l	MA3H 对比		
	SVM	S3VM	S4VM
$l=10$	3/6/0	5/3/1	2/7/0
$l=100$	4/5/0	7/2/0	2/7/0
总计	7/11/0	12/5/1	4/14/0

在总共 18 个(9 个数据集 \times 2 个 l 值)归纳式学习结果中,经统计发现:

(1)相比基准 SVM 算法,MA3H 优于 SVM 的情况为 38.9%,与 SVM 持平的情况为 61.1%。

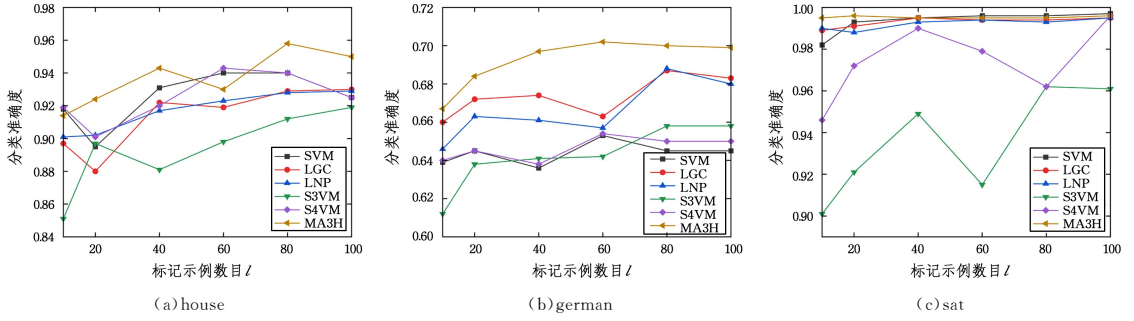
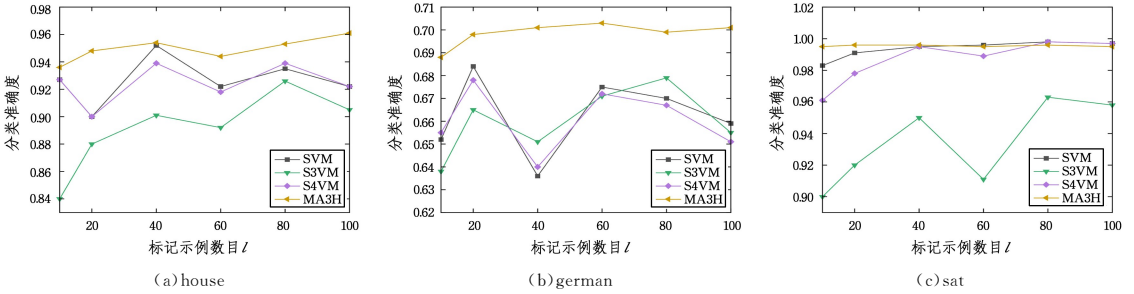
(2)相比 S3VM 算法,MA3H 优于 S3VM 的情况为 66.7%,与 S3VM 持平的情况为 27.8%。

(3)相比 S4VM 算法,MA3H 优于 S4VM 的情况为 22.2%,与 S4VM 持平的情况为 77.8%。

4.4 标记示例数目对分类准确度的影响

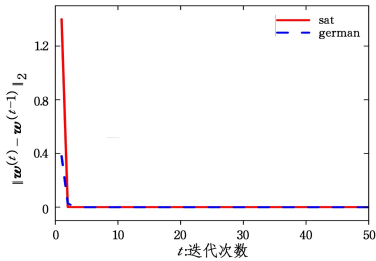
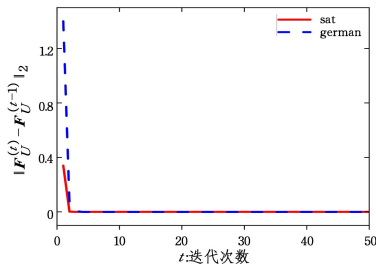
为研究标记示例数目对分类准确度的影响,我们进一步

在 house,german 和 sat 这 3 个数据集上分别进行不同标记数目 $l(l=10,20,40,60,80,100)$ 的直推式学习和归纳式学习实验。图 1 给出了 MA3H 与各对比算法在直推式学习方式上的分类准确度随标记示例数目 l 变化的实验结果。如图 1 所示,MA3H 在标记示例数目较小时,分类准确度具有明显优势。随着标记示例数目增加,各对比算法的分类准确率均呈上升趋势,最终与 MA3H 的分类准确度逐渐趋于一致或略低于 MA3H 的分类准确度。图 2 给出了 MA3H 与各对比算法在归纳式学习方式上的分类准确度随标记示例数目 l 变化的实验结果。如图 2 所示,和直推式学习方式类似,在标记示例数目较小时,MA3H 的分类准确度具有一定的优势。随着标记示例数目的增加,MA3H 的分类准确度趋于稳定,而 SVM 和 S4VM 的分类准确度在 house 和 german 数据集上波动较大,在 sat 数据集上则与 MA3H 逐渐趋于一致。S3VM 在标记示例数目较小时,分类性能明显劣于对比算法。随着标记示例数目的增加,其性能有所提升,但依然劣于大部分对比算法。

图1 MA3H 与对比算法分类准确度在直推式学习方式上随标记示例数目 l 变化的实验结果Fig. 1 Classification accuracy of MA3H and each comparison algorithm changes as numbers of labeled instances(l) in transductive learning图2 MA3H 与对比算法分类准确度在归纳式学习方式上随标记示例数目 l 变化的实验结果Fig. 2 Classification accuracy of MA3H and each comparison algorithm changes as numbers of labeled instances(l) in inductive learning

4.5 算法收敛性和复杂度分析

图3给出了预测模型($\|w^{(t)} - w^{(t-1)}\|$)和未标记示例标记置信度($\|F_U^{(t)} - F_U^{(t-1)}\|$)随交叉优化迭代次数 t 增加的收敛曲线。从图中可以看出,预测模型和未标记示例标记置信度随迭代次数的增加快速收敛。

(a) w (b) F_U 图3 w 和 F_U 在 sat 和 german 数据集上的收敛曲线Fig. 3 Convergence curves of w and F_U on sat and german

时间复杂度是评价算法性能的重要指标,算法 MA3H 的时间复杂度主要由权重矩阵 S 、交叉优化求解预测模型 w 以及未标记示例置信度矩阵 F_U 的运算复杂度决定,MA3H 训练和预测的时间复杂度可分别表示为 $\theta_{\text{train}} = O(u(l+n^2)) + l(\mathcal{F}_{QP}(uq, u(q+1)) + \mathcal{F}_{QP}((l+u)q^2, (l+u)q^2))$ 和 $\theta_{\text{test}} = O(q^2 n^2)$,

其中 $\mathcal{F}_{QP}(a, b)$ 表示计算具有 a 个变量和 b 个约束的二次规划问题的时间复杂度。时间复杂度定性地描述了执行算法所需要的计算工作量,而算法执行所耗费的具体时间则无法从理论上得出。

为分析算法 MA3H 与各对比算法的实际运行效率,本文统计了各算法在相同实验平台(CPU Ryzen7 4800H @ 2.90 GHz, 16 GB RAM, Win10, MATLAB R2018a)下不同数据集上的运行时间,并计算出各算法独立运行 10 次的平均时间,结果如表 8 所列。

表 8 各算法运行时间对比

Table 8 Comparison of running time of each algorithm

datasets	SVM	LGC	LNP	S3VM	S4VM	MA3H
house	0.001	0.002	0.023	0.621	1.079	9.058
german	0.004	0.004	0.123	3.198	103.300	35.180
sat	0.003	0.005	8.011	13.650	145.900	104.200
total	0.008	0.011	8.157	17.470	250.300	148.400

(s)

由表 8 可知,MA3H 的运行时间在大多数情况下比 S4VM 短而比其他对比算法长。S4VM 相较其他算法耗时最多,主要是因为 S4VM 基于低密度假设,使用多个低密度分类器来逼近决策边界,导致其模型复杂度较高。SVM 在训练过程中仅利用了少量标记示例,耗时较少。S3VM 训练过程中的时间复杂度与未标记示例的数目呈超线性增长,所以随着未标记示例数目增加,S3VM 耗时增多。LGC 和 LNP 都是构造流形构图,通过示例空间和标记空间的平滑性来预测未标记示例的类别,两种算法的时间复杂度接近线性时间复杂度,复杂度低,耗时较少。MA3H 求解过程中涉及多个二次规划问题,因此耗时相对较多。

结束语 本文针对半监督学习问题,提出了一种将最大

间隔准则和示例空间流形假设相结合的半监督学习算法。该算法利用示例空间流形假设,在得到平滑标记置信度的同时利用最大间隔准则构建分类模型,并采用交叉优化方法以迭代方式交替地求解分类模型参数和标记置信度。通过 UCI 数据集和 MNIST 手写数字数据集上的实验,验证了本文算法的可行性,其可适用于直推式和归纳式两种半监督学习方式,且能够在标记示例较少的情况下获得相对较高的分类准确度。在下一步工作中,将在半监督多类学习任务中检验所提算法的有效性,并进一步提升该算法的优化效率。

参 考 文 献

- [1] VAN ENGELEN J E, HOOS H H. A survey on semi-supervised learning[J]. *Machine Learning*, 2020, 109: 373-440.
- [2] CHAPELLE O, SCHOLKOPF B, ZIEN A. *Semi-supervised learning*[M]. Cambridge, MA: MIT Press, 2006.
- [3] ZHU X J. *Semi-supervised learning literature survey*[M]. Madison, USA: Department of Computer Sciences, University of Wisconsin at Madison, 2005.
- [4] ZHU X J, GOLDBERG A B. Introduction to semi-supervised learning[J]. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 2009, 3(1): 1-130.
- [5] CARON M, BOJANOWSKI P, JOULIN A, et al. Deep clustering for unsupervised learning of visual features[C]// *Proceedings of the European Conference on Computer Vision (ECCV 2018)*. Berlin: Springer, 2018: 139-156.
- [6] HASTIE T, TIBSHIRANI R, FRIEDMAN J. *The elements of Statistical Learning: Data Mining, Inference, and Prediction*[M]. Berlin: Springer, 2019.
- [7] FIGUEIREDO M A T. Adaptive sparseness for supervised learning[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(9): 1150-1159.
- [8] LI Y F, ZHOU Z H. Towards making unlabeled data never hurt[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 37(1): 175-188.
- [9] LI M, ZHOU Z H. Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples[J]. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 2007, 37(6): 1088-1098.
- [10] JOACHIMS T. Transductive inference for text classification using support vector machines[C]// *Proceedings of the 16th International Conference on Machine Learning*. San Francisco, USA: Morgan Kaufmann Publishers Inc, 1999: 200-209.
- [11] WANG L, CHAN K L, ZHANG Z. Bootstrapping SVM active learning by incorporating unlabelled images for image retrieval[C]// *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2003: 629-634.
- [12] KASABOV N, PANG S. Transductive support vector machines and applications in bioinformatics for promoter recognition[C]// *Proceedings of 2003 International Conference on Neural Networks and Signal Processing*. Piscataway, NJ: IEEE 1: 1-6.
- [13] SINGLA M, GHOSH D, SHUKLA K K. pin-TSVM: A Robust Transductive Support Vector Machine and its Application to the Detection of COVID-19 Infected Patients[J]. *Neural Processing Letters*, 2021, 53(6): 3981-4010.
- [14] GOUTTE C, DÉJEAN H, GAUSSIER E, et al. Combining labelled and unlabelled data: a case study on Fisher kernels and transductive inference for biological entity recognition[C]// *Proceedings of the 6th Conference on Natural Language Learning*. Stroudsburg, PA: ACL, 2002: 1-7.
- [15] KOCKELKORN M, LÜNEBURG A, SCHEFFER T. Using transduction and multi-view learning to answer emails[C]// *Proceedings of Knowledge Discovery in Databases: PKDD 2003*. Berlin: Springer, 2003: 266-277.
- [16] ZHOU D, BOUSQUET O, LAL T, et al. Learning with local and global consistency[J]. *Advances in Neural Information Processing Systems*, 2003, 16: 321-328.
- [17] ZHU X, GHAMRANI Z. Learning from labeled and unlabeled data with label propagation[R]. Pittsburgh, PA: Carnegie Mellon University, Technical Report: CMU-CALD-02-107, 2002.
- [18] SUN S, HUSSAIN Z, SHAWE-TAYLOR J. Manifold-preserving graph reduction for sparse semi-supervised learning[J]. *Neurocomputing*, 2014, 124: 13-21.
- [19] ZHU X, GHAMRANI Z, LAFFERTY J D. Semi-supervised learning using gaussian fields and harmonic functions[C]// *Proceedings of the 20th International Conference on Machine Learning*. Menlo Park: AAAI Press, 2003: 912-919.
- [20] ZHU X. *Semi-supervised learning with graphs*[M]. Pittsburgh, PA: Carnegie Mellon University, 2005.
- [21] CAI X, WEN G, WEI J, et al. Relative manifold based semi-supervised dimensionality reduction[J]. *Frontiers of Computer Science*, 2014, 8: 923-932.
- [22] VAPNIK V N. An overview of statistical learning theory[J]. *IEEE Transactions on Neural Networks*, 1999, 10(5): 988-999.
- [23] DING S, ZHU Z, ZHANG X. An overview on semi-supervised support vector machine[J]. *Neural Computing and Applications*, 2017, 28(5): 969-978.
- [24] LI Y F, GUO L Z, ZHOU Z H. Towards safe weakly supervised learning[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(1): 334-346.
- [25] WANG W, ZHANG M L. Semi-supervised partial label learning via confidence-rated margin maximization[C]// *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc, 2020: 6982-6993.
- [26] MILLER D J, UYAR H. A mixture of experts classifier with learning based on both labelled and unlabelled data[J]. *Advances in Neural Information Processing Systems*, 1996, 9: 571-577.
- [27] NIGAM K, MCCALLUM A K, THRUN S, et al. Text classification from labeled and unlabeled documents using EM[J]. *Machine Learning*, 2000, 39(2): 103-134.
- [28] SHAHSHAHANI B M, LANDGREBE D A. The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 1994, 32(5): 1087-1095.
- [29] BELKIN M, NIYOGI P. Laplacian eigenmaps and spectral tech-

- niques for embedding and clustering[C]// Proceedings of the 14th International Conference on Neural Information Processing Systems Natural and Synthetic. Cambridge, MA; MIT Press, 2001;585-591.
- [30] HINTON G E,SALAKHUTDINOV R R. Using deep belief nets to learn covariance kernels for Gaussian processes[C]// Proceedings of the 20th International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc,2007;20:1249-1256.
- [31] COATES A,NG A Y. The importance of encoding versus training with sparse coding and vector quantization[C]// Proceedings of the 28th International Conference on International Conference on Machine Learning. Madison, Wisconsin; Omnipress, 2011;921-928.
- [32] BLUM A,MITCHELL T. Combining labeled and unlabeled data with co-training[C]// Proceedings of the 11th Annual Conference on Computational Learning Theory. New York, NY: ACM,1998;92-100.
- [33] NIGAM K,GHANI R. Analyzing the effectiveness and applicability of co-training[C]// Proceedings of the 9th International Conference on Information and Knowledge Management. New York, NY: ACM,2000;86-93.
- [34] GOLDMAN S,ZHOU Y. Enhancing supervised learning with unlabeled data[C]// Proceedings of the 17th International Conference on Machine Learning. San Francisco, CA; Morgan Kaufmann Publishers Inc,2000;327-334.
- [35] ZHOU Z H,LI M. Tri-training: Exploiting unlabeled data using three classifiers[J]. IEEE Transactions on Knowledge and Data Engineering,2005,17(11):1529-1541.
- [36] WANG F,ZHANG C. Label propagation through linear neighborhoods[J]. IEEE Transactions on Knowledge and Data Engineering,2008,20(1):55-67.
- [37] GONG C,LIU T,TAO D,et al. Deformed graph Laplacian for semisupervised learning[J]. IEEE Transactions on Neural Networks and Learning Systems,2015,26(10):2261-2274.
- [38] CALDER J,COOK B,THORPE M,et al. Poisson learning: Graph based semi-supervised learning at very low label rates [C]// Proceedings of the 37th International Conference on Machine Learning. Clearwater Beach, USA; PMLR, 2020; 1306-1316.
- [39] BENNETT K,DEMIRIZ A. Semi-supervised support vector machines[J]. Advances in Neural Information Processing Systems, 1998,11:368-374.
- [40] LI Y F,KWOK J T,ZHOU Z H. Semi-supervised learning using label mean[C]// Proceedings of the 26th Annual International Conference on Machine Learning. New York, NY: ACM,2009; 633-640.
- [41] LI Y F,KWOK J,ZHOU Z H. Cost-sensitive semi-supervised support vector machine[C]// Proceedings of the 24th AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press,2010;500-505.
- [42] WANG Q W,LI Y F,ZHOU Z H. Partial Label Learning with Unlabeled Data[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press,2019;3755-3761.



DAI Wei, born in 1994, postgraduate. His main research interests include machine learning and weakly supervised learning.



CHAI Jing, born in 1983, Ph. D, associate professor. His main research interests include machine learning and weakly supervised learning.

(责任编辑:何杨)