

论文

基于仿射子空间稀疏表示的半监督分类

古楠楠^①, 樊明宇^{②*}, 王迪^②, 贾立好^③, 杜亮^④^① 首都经济贸易大学统计学院, 北京 100070^② 温州大学数学与信息科学学院, 温州 325035^③ 中国科学院自动化研究所精密感知与控制研究中心, 北京 100190^④ 山西大学计算机与信息技术学院, 太原 030006

* 通信作者. E-mail: fanmingyu@wzu.edu.cn

收稿日期: 2015-05-15; 接受日期: 2015-07-10; 网络出版日期: 2015-07-21

国家自然科学基金项目 (批准号: 11426159, 61473212, 61203241, 61472285, 61305035)、浙江省自然科学基金项目 (批准号: LY15F030011, LQ13F030009)、浙江省科技厅项目 (批准号: 2014C1062)、北京市自然科学基金项目 (批准号: 4154087) 和首都经济贸易大学校级科研项目 (批准号: 2015XJQ011) 资助

摘要 基于图的半监督分类是近年来机器学习与数据挖掘领域的研究热点之一. 该方法一般通过构造图来挖掘数据中所蕴含的本质结构, 并进一步利用图的结构信息帮助对无标签样本进行分类. 一般来说, 基于图的半监督分类方法的效果高度依赖于其构造的图. 本文提出了一种基于仿射子空间稀疏表示的图构造方法, 该稀疏编码方法在最小化输入信号重构误差时考虑了 3 个约束条件: (1) 输入信号能够被字典矩阵的仿射组合近似表示; (2) 线性表示系数的非负性约束; (3) 线性表示系数的稀疏性约束. 根据这 3 个约束, 我们构造了基于 l_0 -范数的稀疏编码的约束优化问题, 提出相应近似求解方法, 并进而构造了数据的 l_0 -图. 最后, 在正则化学习理论框架下, 通过引进度量 l_0 -图中结构保持误差的正则项, 提出了一种新的半监督学习方法. 该方法具有显性的多类分类函数, 同时也继承了由数据稀疏编码所得 l_0 -图中蕴含的强判别信息, 因此对外样本具有快速和准确的分类能力. 一系列人工数据与现实采集的数据集上的实验结果验证了所提半监督分类方法的有效性.

关键词 分类算法 数据稀疏分布 图方法 信号分类 最小二乘方法

1 引言

在很多实际应用中, 如图像分析、药物研究及网页分析等, 有标签样本比较缺乏且难以获得, 而无标签数据则比较丰富且相对容易获得, 因此近年来出现了半监督分类方法^[1~3]. 半监督分类的目的是同时利用少量有标签数据与大量无标签数据中蕴含的信息训练分类器, 提高分类器的泛化性能. 由于半监督分类需要更少的人力, 且能获得更高的准确率, 所以它已经吸引了数据挖掘以及机器学习领域众多研究者的广泛关注. 迄今为止, 已有多种半监督分类方法被提出^[4,5], 如半监督生成模型的期望最大化算法^[6,7]、自训练^[8,9]、协同训练^[10,11]、直推式支持向量机^[12,13]等, 这些方法已在基因数据分析^[14]、图像分类^[15]、视频分割^[16]等领域有了初步成功应用.

在众多半监督分类方法中, 基于图的方法研究是最热点的研究领域之一. 这类方法首先在所有训练数据上构造一个邻域图, 图的结点为训练数据, 边的权重为它所连接的两个数据之间的相似度; 然

后在此邻域图上基于某种假设 (如聚类假设或流形假设^[4]) 进行分类. 代表性的基于图的半监督分类方法包括: Gauss 随机场与调和函数 (Gaussian random fields and harmonic functions, GRF)^[17]、局部全局一致性 (local and global consistency, LGC)^[18]、线性近邻传播 (local neighborhood propagation, LNP)^[19]、Laplace 正则化最小二乘分类 (Laplacian regularized least square classification, LapRLSC)^[20] 等.

尽管基于图的半监督分类方法已被成功地应用在很多领域中^[5], 然而其中仍然有一些问题尚未得到完全合适的解决. 主要考虑如下 3 个方面:

(1) 数据近邻选择. 许多基于图的半监督分类算法所采用的传统的邻域图构造方法, 如 k -近邻法或 ϵ -近邻法等, 主要依赖于成对的欧式距离, 通过一组重叠的局部片来构造邻域图. 这会带来邻域尺寸的参数选择问题; 且固定的邻域尺寸无法构造数据自适应的邻域, 因此不适用于分布不均匀的数据; 此外, 这些方法对数据局部噪音与误差比较敏感.

(2) 模型假设. 半监督分类大多基于聚类假设或流形假设. 聚类假设指处在同一聚类中的样本有较大可能拥有相同的标签; 流形假设则认为沿着数据所在流形相近的样本, 其标签也应相近. 然而现实中的数据是复杂且多样的, 希望利用新的模型假设, 更好地提取数据中的判别信息进行半监督分类.

(3) 显性的多类分类函数. 许多基于图的分类算法不能提供显性的多类分类函数, 于是不能在线地对新样本点进行分类. 这一问题限制了这些算法在在线决策场合中的应用.

为了解决上述问题, 本文提出在正则化学习理论框架下利用数据稀疏表示中蕴藏的判别信息进行半监督分类. 数据的稀疏表示^[21]是指用字典中的一小部分基本数据的线性组合来重构某数据点, 具有很好的判别性. Wright 等^[22]利用数据的稀疏表示构造分类器, 得到了一种对图像噪声及遮挡问题具有较好稳健性的人脸识别方法. Cheng 等^[23]提出利用数据稀疏表示构建 l_1 -图, 该图能自适应选择数据近邻且展现出很好的识别效果. 在文献 [24] 中, Yan 等认为 l_1 -图中边的权重大的数据对, 其类别标签也应该相似, 由此进行半监督分类. 这些方法是在线性子空间假设基础之上构造稀疏表示问题的, 即认为同类数据位于同一个线性子空间, 但该假设并不总是成立的. 此外, 在这些方法中, 对于每个训练数据, 为了获得其稀疏表示系数, 需将相应 l_0 -范数最小化问题松弛为 l_1 -范数最小化问题进行求解. 但是这种凸松弛并不能保证所得解 (即子空间重构系数向量) 总是有意义的, l_1 优化问题可能得到稠密的稀疏表示系数, 此时违背了子空间重构系数的稀疏性的设想.

不同于上述方法的线性子空间的假设, 本文提出一种基于仿射子空间线性重构系数稀疏化表示的图构造方法, 并进而在正则化学习理论框架下利用所构造的图进行半监督分类. 所构造的稀疏编码问题在最小化输入数据重构误差时考虑了 3 个约束条件: 输入数据应当位于基底向量生成的仿射子空间中; 线性表示系数的非负性约束; 线性表示系数的基于 l_0 -范数的稀疏性约束. 我们针对此 l_0 优化问题提出一种近似求解算法, 然后利用仿射子空间的稀疏表示系数构造 l_0 -图, 并在流形正则化理论框架下^[20]基于所构造的图得到半监督分类器. 所提的半监督分类方法具有显性的多类分类函数, 能够快速对新来样本点进行分类; 而且还继承了稀疏表示的能够自适应进行邻域选择以及具有较高判别性的优点. 一系列人工数据及现实数据集上的实验结果验证了所提方法具有较强的分类能力.

本文的结构安排如下: 第 2 节回顾相关工作; 第 3 节介绍本文所提的基于仿射子空间的稀疏 l_0 -图构造方法, 并导出相应半监督分类方法; 第 4 节给出所提方法在一系列数据集上的实验效果; 最后在第 5 节对全文进行总结.

2 相关工作

为了避免混淆, 在此给出本文所使用的部分主要符号. 矩阵用大写字母表示, 数据点都是列向量形式并用小写字母表示. 我们将半监督分类问题的训练数据表示为 $\{(x_i, z_i), x_{l+j}, i = 1, \dots, l, j = 1, \dots, N-l\}$, 其中 l 与 $(N-l)$ 分别是有标签数据与无标签数据的数目, $x_i \in \mathbb{R}^D$ 为数据点, $z_i \in \{1, \dots, C\}$ 为 x_i 的类别标签, C 为数据的类别总数, $X = [x_1, \dots, x_l, \dots, x_N] \in \mathbb{R}^{D \times N}$ 为训练数据矩阵. 向量 $x_j = [x_{1j}, \dots, x_{Dj}]^T \in \mathbb{R}^D$ 的 l_2 -范数定义为 $\|x_j\|_2 = \sqrt{\sum_{i=1}^D (x_{ij})^2}$, 简便起见, 本文将 l_2 -范数用 $\|\cdot\|$ 表示; 向量 $x_j \in \mathbb{R}^D$ 的 l_1 -范数定义为 $\|x_j\|_1 = \sum_{i=1}^D |x_{ij}|$; 向量 $x_j \in \mathbb{R}^D$ 的 l_0 -范数定义为 x_j 中非零元素的个数.

2.1 流形正则化框架

流形正则化 (manifold regularization, MR) 框架是由 Belkin 等^[20] 基于再生核希尔伯特空间理论提出的. 令 \mathcal{H}_K 为半正定核函数 $k(\cdot, \cdot)$ 所生成的再生核希尔伯特空间, 若欲求的分类函数 f 为 \mathcal{H}_K 中的实值函数, 则流形正则化框架可以表示为

$$f^* = \arg \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{l} \sum_{i=1}^l V(x_i, z_i, f) + \gamma_K \|f\|_K^2 + \gamma_I \|f\|_I^2 \right\}, \quad (1)$$

其中, V 为损失函数, γ_K 及 γ_I 为两个惩罚项参数. 在式 (1) 中, 目标函数的第 1 项为拟合误差项, 它用来度量已知的类别标签与通过 f 所得预测标签之间的差异; 目标函数的第 2 项为复杂度惩罚项, 它为 f 在再生核希尔伯特空间 \mathcal{H}_K 中的范数的平方, 用来度量分类函数 f 在 \mathcal{H}_K 中的函数复杂度; 目标函数的第 3 项为光滑度惩罚项, 它用来度量分类函数 f 关于数据分布的光滑度. 在文献 [20] 中, Belkin 等定义了如下的基于流形的光滑度惩罚项, 度量了 f 在数据所在流形上的光滑度,

$$\|f\|_I^2 = \frac{1}{N^2} \sum_{i,j=1}^N w_{ij} \|f(x_i) - f(x_j)\|^2, \quad (2)$$

其中 w_{ij} 为数据邻域图中边的权重. 此时, 分类函数 f 可显性地由训练数据的核函数线性表出.

广义来讲, 式 (1) 中的损失函数 V 和光滑度惩罚项 $\|f\|_I^2$ 可采用不同方式来定义, 由此可导出不同的分类算法. 例如, 若令式 (1) 中的损失函数 V 为平方损失函数 $V(x_i, z_i, f) = \|z_i - f(x_i)\|^2$, 并将 $\|f\|_I^2$ 定义为式 (2) 的形式, 则可得到 LapRLSC 分类器; 若令式 (1) 中的损失函数为 Hinge 损失函数^[20], 并且同样把 $\|f\|_I^2$ 定义为式 (2) 的形式, 则流形正则化框架给出 Laplace 正则化支持向量机分类器.

很多基于图的半监督分类算法都可以被纳入流形正则化框架. 然而, 其中仍然存在一些问题需要解决. 例如, 这些方法经常使用一个固定的邻域尺寸来定义邻域图, 这会出现参数选择这个难题, 且固定的邻域尺寸不适用于分布不均匀的数据. 另外, 这些方法基于流形假设, 即假定高维数据分布在低维流形上, 但是此假设并不总是成立的. 因此, 本文提出基于数据稀疏表示的半监督分类方法, 该方法能够自适应建立数据之间的关系, 且基于稀疏化假设, 具有较好的判别性能.

2.2 基于稀疏表示的分类方法

稀疏表示是指用很小数目的基本信号的线性组合来表示信号中绝大部分信息. 具体来说, 若 $x \in \mathbb{R}^D$ 为输入数据, 给定一个由过完备基向量组构成的字典 $\Phi \in \mathbb{R}^{D \times M}$, 则可通过优化问题 (3) 来寻找 x

的稀疏表示系数 α^* :

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad \|x - \Phi\alpha\| \leq \varepsilon, \quad (3)$$

其中 $\varepsilon > 0$ 为噪声容忍度. 然而式 (3) 的确切解只能通过组合搜索的方法获得, 计算复杂度非常高. 针对此问题, 相关研究者提出了很多替代算法, 如正交匹配追踪 (orthogonal matching pursuit, OMP) [25] 与迭代硬阈值算法 (iterative hard thresholding, IHT) [26] 等. 另一方面, 相关理论结果证明, 当满足一定条件时, l_1 - 与 l_0 - 范数最小化问题是等价的 [27]. 因此, 数据的稀疏表示也可用如下问题求得:

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad \|x - \Phi\alpha\| \leq \varepsilon. \quad (4)$$

针对此凸优化问题, 内点法 (interior point method) [28] 与同伦法 (homotopy method) [29,30] 是两种经典的求解方法. 然而它们的计算复杂度比较高, 因此近年来相关研究者提出了很多其他方法, 如快速迭代收缩阈值算法 (fast iterative shrinkage-thresholding algorithm, FISTA) [31], 增广 Lagrange 方法 (augmented Lagrangian methods, ALM) [32], 原始增广 Lagrange 方法 (primal ALM, PALM) [33] 等.

稀疏表示技术最开始被用于信号的压缩和编码, 但由于其中包含很强的判别信息, 所以可被用来训练分类器. 在文献 [22] 中, Wright 等假设同类数据位于同一个线性子空间, 在此假设下, 图像数据的稀疏表示是具有判别性的, 有利于数据分类. 具体来说, 对于测试样本 x , 通过如下的 l_1 - 范数最小化问题来获得 x 的稀疏表示:

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad \|x - X\alpha\| \leq \varepsilon, \quad (5)$$

其中 $X = [x_1, \dots, x_N] \in \mathbb{R}^{D \times N}$ 为训练数据矩阵. 随后 Wright 等利用数据的稀疏表示构造分类器, 得到了一种对图像噪声及遮挡问题具有较好稳健性的人脸识别方法. Yan 等 [24] 基于稀疏表示提出了一种半监督学习方法. 该方法首先利用数据稀疏表示构建 l_1 - 图, 图的顶点为所有的样本, 与数据相关联的边的权重刻画了用其余点表示该点的稀疏表示系数. 相对而言, 传统的 k - 近邻图和 ε - 邻域图的构建需要有近邻搜索与边的权重选择这两个独立的步骤, 而 l_1 - 图的近邻结构与边的权重可以同时获得, 且不需要参数. 获得 l_1 - 图后, 文献 [24] 通过保持图中数据之间的权重来进行半监督学习.

这些基于稀疏表示的分类方法是建立在线性子空间假设基础上的, 即认为同类数据位于同一个线性子空间. 另外, 在这些方法中, 对于每个训练数据, 需要将相应 l_0 - 范数最小化问题松弛为 l_1 - 范数最小化问题, 获得稀疏表示系数. 但是这种凸松弛可能会得到稠密的表示系数, 此时违背了子空间重构系数的稀疏性的设想. 因此, 本文提出了基于仿射子空间线性重构系数稀疏化表示的 l_0 - 图构造方法, 并进而在正则化学习理论框架下利用所构造的图进行半监督分类.

3 基于仿射子空间稀疏表示的半监督分类方法

本节首先提出了数据分布的仿射子空间假设, 及该假设下的线性稀疏编码问题; 随后给出了该问题的求解算法与相应稀疏 l_0 - 图构造方法; 最后, 提出基于 l_0 - 结构图正则化的半监督分类算法.

3.1 基于仿射子空间模型的稀疏表示

线性子空间模型假设同类数据位于同一个线性子空间, 而异类数据分布于不同的线性子空间. 对于训练样本 x_i , 基于线性子空间模型假设的稀疏表示数学模型为

$$\alpha_i^* = \arg \min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad \|x_i - \hat{X}^i \alpha\| \leq \varepsilon. \quad (6)$$

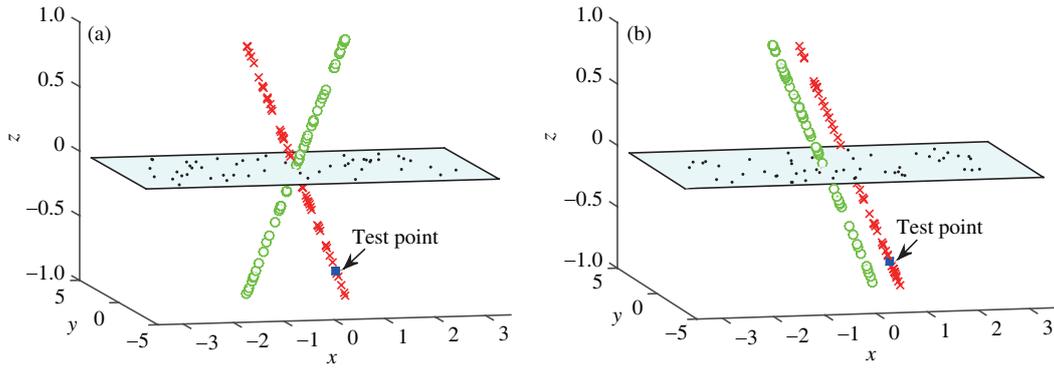


图 1 (网络版彩图) 线性子空间与仿射子空间示意图。(a) 中数据位于三个线性子空间：两条线和一个平面上；(b) 中数据位于三个仿射子空间：两条线状仿射子空间和一个平面子空间上

Figure 1 (Color online) Illustrations for linear subspaces and affine subspaces. (a) Data points lying on three linear subspaces: two lines and one plane; (b) data points lying on three affine subspaces: two lines and one plane.

其中, $\hat{X}^i = [x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N] \in \mathbb{R}^{D \times (N-1)}$ 为除了 x_i 之外的所有训练数据构成的字典矩阵, $\varepsilon > 0$ 为噪声容忍度, $\alpha \in \mathbb{R}^{N-1}$ 为线性表示系数构成的向量. 则在线性子空间假设下, 所获数据稀疏表示系数 α_i^* 是具有判别性的, 可以用来训练分类器 [22].

已有的利用稀疏表示的分类方法大多基于线性子空间假设, 本文则考虑数据分布的仿射子空间模型, 即假设同类数据位于同一个仿射子空间, 而异类数据分布于不同的仿射子空间. 需要注意的是, 对于一个 D 维的仿射子空间 S_D , 它可以被认为是包括了 S_D 与原点的 $D+1$ 维线性空间的一个子集. 因此, 当同类数据位于同一个仿射子空间上时, 采用线性子空间模型可能会将位于不同仿射子空间上的数据混淆. 例如, 图 1(a) 所示的 3 类数据 (红色“x”状点、绿色圆圈、黑色圆点所示) 位于 3 个可分的线性子空间: 两条线和一个平面, 它们的方程分别为

$$x = -z = \frac{y}{0}, \quad x = z = \frac{y}{0}, \quad z = 0. \tag{7}$$

因为这 3 个线性子空间是可分的, 所以已有基于线性子空间假设的稀疏表示方法能够正确地对测试点进行稀疏编码. 另一方面, 图 1(b) 所示的 3 类数据位于 3 个仿射子空间: 两条线和一个平面. 它们的方程分别为

$$x - \frac{1}{4} = -z = \frac{y}{0}, \quad x + \frac{1}{4} = -z = \frac{y}{0}, \quad z = 0. \tag{8}$$

因为这两条线所对应的线性子空间是完全相同的, 所以基于线性子空间假设的稀疏表示方法可能会利用另一个仿射子空间中的数据点 (绿色圆圈) 来对测试点进行稀疏编码. 在第 4 节中, 我们针对图 1(b) 所示数据进行了实验, 实验结果显示我们方法能够较好地判别出数据点之间的仿射关系.

在仿射子空间模型假设下, 若每类有足够多的训练样本, 记为 $A_k = [x_{k,1}, \dots, x_{k,n_k}]$ (其中 n_k 为第 k 类训练数据的个数), 则任意训练数据 $x_{k,i}$ 近似位于本类中其他训练数据所生成的仿射子空间, 即

$$x_{k,i} = \alpha_{k,1}x_{k,1} + \dots + \alpha_{k,i-1}x_{k,i-1} + \alpha_{k,i+1}x_{k,i+1} + \dots + \alpha_{k,n_k}x_{k,n_k}, \quad e^T \alpha^k = 1, \tag{9}$$

其中, $\alpha^k = [\alpha_{k,1}, \dots, \alpha_{k,i-1}, \alpha_{k,i+1}, \dots, \alpha_{k,n_k}]^T$ 为重构系数向量, e 为元素全为 1 的向量. 进一步地, 我们可将 $x_{k,i}$ 表示为所有的 C 类训练数据的仿射组合:

$$x_{k,i} = X\alpha, \quad e^T \alpha = 1, \tag{10}$$

其中 $X = [A_1, \dots, A_C] \in \mathbb{R}^{D \times N}$ 为训练数据矩阵, $\alpha = [0, \dots, 0, \alpha_{k,1}, \dots, \alpha_{k,i-1}, 0, \alpha_{k,i+1}, \dots, \alpha_{k,n_k}, 0, \dots, 0]^T \in \mathbb{R}^N$ 为系数向量. 此时 α 自然地具有稀疏性, 因为只有与 $x_{k,i}$ 同类的那些点对应的分量值非零. 若 $D < N$, 则式 (10) 为欠定的, 此时可通过如下的稳健稀疏表示问题来获得数据 x_i 的表示系数:

$$\alpha_i^* = \arg \min_{\alpha} \|x_i - \hat{X}^i \alpha\|^2, \quad \text{s.t.} \quad e^T \alpha = 1, \quad \|\alpha\|_0 \leq K, \quad (11)$$

其中 $\hat{X}^i = [x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N] \in \mathbb{R}^{D \times (N-1)}$, $\alpha_i = [\alpha_{i,1}, \dots, \alpha_{i,N-1}]^T \in \mathbb{R}^{N-1}$ 为表示系数, K 为稀疏度参数. 可以看出, 通过式 (11) 寻找数据的稀疏表示能够自动判别训练集中的不同类别. 因此基于仿射子空间模型的稀疏表示具有较好判别性能, 有利于解决目标分类任务.

进一步地, 为了利用稀疏表示系数刻画数据关系, 定义数据邻域图中边的权重, 我们在式 (11) 中加入非负约束, 得到如下基于仿射子空间假设的稀疏编码问题:

$$\alpha_i^* = \arg \min_{\alpha} \|x_i - \hat{X}^i \alpha\|^2, \quad \text{s.t.} \quad e^T \alpha = 1, \quad \alpha \geq 0, \quad \|\alpha\|_0 \leq K. \quad (12)$$

所构造的稀疏表示问题满足 3 个约束: 第 1 个约束是令稀疏表示系数向量中各元素之和为 1, 即认为输入数据应当近似位于其余训练样本生成的仿射子空间中; 第 2 个约束保证了稀疏表示系数向量中各元素的非负性, 以便于在半监督学习中利用稀疏表示系数定义数据邻域图中边的权重; 第 3 个约束则限制了所得表示系数中非零元素的个数, 以得到稀疏的解. 需要一提的是, 稀疏编码问题 (12) 不仅可以处理同类训练数据位于同一个仿射子空间的情况, 也可以处理同类训练数据位于同一个线性子空间的情况, 因为线性子空间是一个特殊的仿射子空间.

3.2 基于仿射子空间的稀疏 l_0 - 图

第 3.1 小节中构造的稀疏编码问题 (12) 是非凸且 NP 难的, 本小节将针对该问题提出一种简单但有效的近似求解算法. 随后将给出基于仿射子空间的稀疏 l_0 - 图的构造方法.

对于优化问题 (12), 其目标函数可以作如下变换:

$$Q(\alpha) = \|x_i - \hat{X}^i \alpha\|^2 = \|x_i e^T \alpha - \hat{X}^i \alpha\|^2 = \|(x_i e^T - \hat{X}^i) \alpha\|^2 = \alpha^T (x_i e^T - \hat{X}^i)^T (x_i e^T - \hat{X}^i) \alpha. \quad (13)$$

因此, 优化问题 (12) 可以变为

$$\alpha_i^* = \arg \min_{\alpha} \alpha^T H_i \alpha, \quad \text{s.t.} \quad e^T \alpha = 1, \quad \alpha \geq 0, \quad \|\alpha\|_0 \leq K, \quad (14)$$

其中 $H_i = (x_i e^T - \hat{X}^i)^T (x_i e^T - \hat{X}^i)$, 易看出, H_i 为半正定矩阵.

根据以上分析, 可以构造稀疏编码问题 (12) 的近似求解方法, 如算法 1 中第 2~11 行所示. 我们对其中的记号作一些说明:

- 第 7 行中的 $\text{supp}_K(\cdot)$ 为硬阈值指标算子. 给定向量 $\alpha = [\alpha_{(1)}, \dots, \alpha_{(m)}]^T \in \mathbb{R}^m$, 指标集 $\text{supp}_K(\alpha) = \{j | \alpha_{(j)} \text{ 为 } \alpha_{(1)}, \dots, \alpha_{(m)} \text{ 中具有最大绝对值的 } K \text{ 个之一}\}$.

- 第 8 行中的 $T(\cdot, \cdot)$ 为修剪算子. 给定向量 $\alpha = [\alpha_{(1)}, \dots, \alpha_{(m)}]^T \in \mathbb{R}^m$ 与指标集 $I = \{i_1, \dots, i_K\}$ ($K < m$), 修剪算子 $T(\alpha, I)$ 输出向量 $\hat{\alpha} = [\hat{\alpha}_{(1)}, \dots, \hat{\alpha}_{(m)}] \in \mathbb{R}^m$, 使得 $\hat{\alpha}$ 是由 α 依据指标集 I 修剪而得, 即

$$\hat{\alpha}_{(i)} = \begin{cases} \alpha_{(i)}, & \text{若 } i \in I \\ 0, & \text{其他} \end{cases} \quad (i = 1, \dots, m). \quad (15)$$

具体地说, 为了获得稀疏表示问题 (12) 的解, 首先通过二次规划问题 (如算法 1 中第 6 行所示) 获得一个稠密的表示系数 $\alpha_i^*(t+1)$, 然后通过修剪算子将 $\alpha_i^*(t+1)$ 中的绝对值较小的元素变为 0 (如第 7, 8 行所示) 以满足稀疏约束 $\|\alpha_i\|_0 \leq K$, 并将新的表示系数归一化 (如第 9 行所示) 以满足仿射约束 $e^T \alpha_i = 1$, 由此得到表示系数 $\alpha_i(t+1)$; 随后, 将 $\alpha_i(t+1)$ 传递到下一次迭代过程, 重新进行以上步骤, 以此得到越来越接近于式 (12) 的真实解的稀疏表示系数.

算法 1 基于仿射子空间的稀疏 l_0 - 图构造算法

输入: 训练样本矩阵 $X = [x_1, \dots, x_N] \in \mathbb{R}^{D \times N}$, 参数 K , 最大迭代步数 T

输出: 稀疏 l_0 - 图的权重矩阵 W

```

1: for  $i = 1$  to  $N$  do
2:    $\hat{X}^i = [x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N]$ ;
3:    $H_i = (x_i e^T - \hat{X}^i)^T (x_i e^T - \hat{X}^i)$ ;
4:    $t = 1, \alpha_i(t) = 0$ ;
5:   while  $t \leq T$  且不收敛 do
6:      $\alpha_i^*(t+1) = \arg \min_{\alpha} (\alpha^T H_i \alpha - \alpha_i^T(t) H_i \alpha), \quad \text{s.t.} \quad e^T \alpha = 1, \quad \alpha \geq 0$ ;
7:      $I = \text{supp}_K(\alpha_i^*(t+1))$ ;
8:      $\hat{\alpha}_i(t+1) = T(\alpha_i^*(t+1), I)$ 
9:      $\alpha_i(t+1) = \frac{\hat{\alpha}_i(t+1)}{e^T \hat{\alpha}_i(t+1)}$ ;
10:     $t = t + 1$ .
11:  end while
12:   $\alpha_i = \alpha_i(t)$ ;
13:  for  $j = 1$  to  $N$  do
14:    if  $j < i$  then
15:       $W_{ij} = \alpha_{i,j}$ ;
16:    else if  $j = i$  then
17:       $W_{ij} = 0$ ;
18:    else
19:       $W_{ij} = \alpha_{i,(j-1)}$ ;
20:    end if
21:  end for
22: end for

```

性质 1 对于数据 $x_i (i = 1, \dots, N)$, 有 $Q(\alpha_i^*(t+1)) \leq Q(\alpha_i(t)) (t \geq 1)$, 其中 $Q(\alpha) = \|x_i - \hat{X}^i \alpha\|^2$.

证明 简便起见, 本证明中将 $\alpha_i(t)$ 记为 $\alpha(t)$, $\alpha_i(t+1)$ 记为 $\alpha(t+1)$, $\alpha_i^*(t+1)$ 记为 $\alpha^*(t+1)$.

从算法中可知, $\alpha^*(t+1)$ 是如下优化问题的解:

$$\alpha^*(t+1) = \arg \min_{\alpha} L(\alpha; \alpha(t)), \quad \text{s.t.} \quad e^T \alpha = 1, \quad \alpha \geq 0, \tag{16}$$

其中 $L(\alpha; \alpha(t)) = \alpha^T H_i \alpha - \alpha^T(t) H_i \alpha$. 由式 (13) 与 (14) 易得

$$Q(\alpha) - Q(\alpha(t)) = 2L(\alpha, \alpha(t)) - (\alpha - \alpha(t))^T H_i (\alpha - \alpha(t)). \tag{17}$$

由于 H_i 是半正定矩阵, 所以 $(\alpha - \alpha(t))^T H_i (\alpha - \alpha(t)) \geq 0$, 于是 $Q(\alpha) - Q(\alpha(t)) \leq 2L(\alpha; \alpha(t))$. 又由 $L(\alpha^*(t+1); \alpha(t))$ 是 $L(\alpha; \alpha(t))$ 关于 α 的最小值, 可得 $L(\alpha^*(t+1); \alpha(t)) \leq L(\alpha(t); \alpha(t)) = 0$. 因此, $Q(\alpha_i^*(t+1)) \leq Q(\alpha_i(t))$.

算法 1 描述了基于仿射子空间的稀疏 l_0 - 图的构造方法. 具体来说, 对于每一个训练样本 $x_i (i = 1, \dots, N)$, 我们可通过稀疏编码问题 (12) 获得其余训练样本对该样本的稀疏表示系数 $\alpha_i = [\alpha_{i,1}, \dots,$

$\alpha_{i,N-1}]^T \in \mathbb{R}^{N-1}$. 若记 $\beta_i = [\alpha_{i,1}, \dots, \alpha_{i,i-1}, 0, \alpha_{i,i}, \dots, \alpha_{i,N-1}]^T \in \mathbb{R}^N$, 则 β_i 的第 j 个元素 $\beta_{i,j}$ 可作为数据 x_i 与 x_j 之间关系的一种度量, 并可由此构造基于仿射子空间的稀疏 l_0 -图 $G = (V, W)$. 邻域图 G 的顶点集 V 由所有训练数据 x_1, \dots, x_N 构成; 同时, 若 $\beta_{i,j} \neq 0$, 则在顶点 x_i 与 x_j 之间建立一条直接相连的有向边, 边的权重为 $W_{ij} = \beta_{i,j}$. 与传统的 k -近邻图或 ε -近邻图相比, 此处构造的 l_0 -图的优势在于: 首先, l_0 -图能够自适应选择数据近邻; 而传统的 k -近邻图或 ε -近邻图采用固定的邻域尺寸, 无法构造数据自适应的邻域, 因此不适用于数据分布不均匀及可用数据不充足的情形. 其次, l_0 -图具有稀疏性, 在大规模数据的应用中, 能够大幅减少数据存储量. 最后, 传统的 k -近邻图或 ε -近邻图仅根据数据间欧氏距离关系确定数据近邻, 而 l_0 -图深层挖掘了数据中所蕴含的本质稀疏结构信息, 有利于帮助有标签样本对无标签样本进行分类, 具有很好的判别性能.

3.3 基于仿射子空间的稀疏正则化半监督分类方法

对于流形正则化学习框架:

$$f^* = \arg \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{l} \sum_{i=1}^l V(x_i, y_i, f) + \gamma_K \|f\|_K^2 + \gamma_I \|f\|_I^2 \right\}, \quad (18)$$

在此令 $f(\cdot)$ 为 C 类判别函数 $f(x) = [f_1(x), f_2(x), \dots, f_C(x)]^T$, 则数据 x 所属的类别可定义为 $\arg \max_{i=1, \dots, C} f_i(x)$. 相应地, 需将已知类别标签 $z_i \in \{1, \dots, C\}$ ($i = 1, \dots, l$) 转化为 C 维的标签向量 $y_i = [y_{1i}, \dots, y_{Ci}]^T$:

$$y_{ji} = \begin{cases} 1, & \text{若 } z_i = j \\ 0, & \text{其他} \end{cases} \quad (j = 1, \dots, C). \quad (19)$$

对于式 (18) 中目标函数的第 1 项: 拟合误差项, 将损失函数 V 定义为平方损失, 即

$$V(x_i, y_i, f) = \|y_i - f(x_i)\|^2. \quad (20)$$

对于式 (18) 中目标函数的第 2 项: 复杂度惩罚项, 利用 f_s ($s = 1, \dots, C$) 在与半正定核函数 $k(\cdot, \cdot)$ 对应的再生核希尔伯特函数空间 \mathcal{H}_K 中的范数来定义. 若 f_s 可以表示为 $f_s = \sum_i \eta_i^{(s)} k(w_i, \cdot)$, 则对于向量函数 $f(x) = [f_1(x), f_2(x), \dots, f_C(x)]^T$, $\|f\|_K^2$ 可以被定义为

$$\|f\|_K^2 = \sum_{s=1}^C \|f_s\|_K^2 = \sum_{s=1}^C \sum_{i,j} \eta_i^{(s)} \eta_j^{(s)} k(w_i, w_j). \quad (21)$$

对于式 (18) 中目标函数的第 3 项: 光滑度惩罚项, 利用 3.2 小节所提的稀疏 l_0 -图来定义. l_0 -图中的权重 W_{ij} 度量了样本 x_j 对样本 x_i 稀疏表示的贡献度, 其中含有强判别信息. 假设判别函数在数据集样本点上的输出也满足稀疏编码的重构关系, 这样每个样本点的标签能够被其他样本点的标签所稀疏表示. 基于此, 定义如下的惩罚项, 来度量判别函数 f 保持数据稀疏表示系数的能力:

$$\|f\|_I^2 = \frac{1}{N^2} \sum_{i=1}^N \left\| f(x_i) - \sum_{j=1}^N W_{ij} f(x_j) \right\|_2^2. \quad (22)$$

此处构造的基于稀疏表示的光滑度惩罚项 (22) 与流形正则化框架中基于流形的光滑度惩罚项 (2) 的区别在于: 首先, 式 (2) 常利用传统的 k -近邻图或 ε -近邻图选取数据近邻, 然后只对近邻数据对赋予

非零权重, 固定的邻域尺寸及相应权重不适用于分布不均匀的数据, 无法很好地刻画数据类别标签间的关系; 而式 (22) 根据数据稀疏表示自动获得近邻结构与边的权重, 利用数据的稀疏表示信息对数据类别标签 $f(x_i)$ 进行重构, 具有更好的判别性能. 其次, 式 (2) 是建立在流形假设基础上的, 认为沿着数据所在流形相近的样本其标签也应相近. 然而现实中的数据是复杂且多样的, 本文所提光滑度惩罚项 (22) 利用新的仿射子空间稀疏模型假设, 能更好地提取数据中的判别信息进行半监督分类.

将上面所定义的判别函数、拟合误差项 (20)、复杂度惩罚 (21)、光滑度惩罚项 (22) 代入流形正则化框架 (18) 中, 可获得我们的基于仿射子空间的稀疏正则化半监督分类方法 (semi-supervised classification based on affine subspace sparse regularization, SSC-ASSR):

$$f^* = \arg \min_{\substack{f_s \in \mathcal{H}_K \\ s=1, \dots, C}} \left\{ \frac{1}{l} \sum_{i=1}^l \|y_i - f(x_i)\|^2 + \gamma_K \sum_{s=1}^C \sum_{i,j} \eta_i^{(s)} \eta_j^{(s)} k(w_i, w_j) + \frac{\gamma_I}{N^2} \sum_{i=1}^N \left\| f(x_i) - \sum_{j=1}^N W_{ij} f(x_j) \right\|_2^2 \right\}, \quad (23)$$

其中 γ_K 与 γ_I 是惩罚项参数.

定理1 优化问题 (23) 的解具有如下形式:

$$f(x) = \sum_{i=1}^N b_i k(x_i, x), \quad (24)$$

其中 $b_i = [b_{1i}, \dots, b_{Ci}]^T \in \mathbb{R}^C$.

证明 请参考文献 [20] 的 3.4 小节中定理 2 的证明.

将表达式 (24) 代入式 (23) 并进行整理, 可获得如下的凸优化问题:

$$B^* = \arg \min_{B \in \mathbb{R}^{C \times N}} \left\{ \frac{1}{l} \text{tr}((Y - BKJ)(Y - BKJ)^T) + \gamma_K \text{tr}(BKB^T) + \frac{\gamma_I}{N^2} \text{tr}(BK(I - W)^T(I - W)KB^T) \right\}. \quad (25)$$

其中, $Y = [y_1, \dots, y_l, \mathbf{0}, \dots, \mathbf{0}] \in \mathbb{R}^{C \times N}$ 为标签矩阵, $\mathbf{0} \in \mathbb{R}^C$ 为零向量; $B = [b_1, \dots, b_N] \in \mathbb{R}^{C \times N}$ 为核函数的系数矩阵; $K = (k(x_i, x_j)) \in \mathbb{R}^{N \times N}$ 为核矩阵; $J \in \mathbb{R}^{N \times N}$ 为前 l 个对角元素为 1, 其余元素为 0 的对角矩阵; $W \in \mathbb{R}^{N \times N}$ 为稀疏 l_0 - 图的权重矩阵; I 为 N 阶的单位阵. 令目标函数对 B 的导数为 0, 可得到式 (25) 的最优解:

$$B^* = Y \left(KJ + \gamma_K lI + \frac{\gamma_I l}{N^2} K(I - W)^T(I - W) \right)^{-1}. \quad (26)$$

则问题 (23) 的解为

$$f^*(x) = (f_1^*(x), f_2^*(x), \dots, f_C^*(x))^T = \sum_{i=1}^N b_i^* k(x_i, x), \quad (27)$$

其中 b_i^* 为 B^* 的第 i 列. 因此, 我们的 SSC-ASSR 分类器为

$$g^*(x) = \arg \max_{s=1,2,\dots,C} \{f_s^*(x)\}. \quad (28)$$

基于以上思想, 将对应的 SSC-ASSR 算法进行总结, 如算法 2 所示.

算法 2 基于仿射子空间的稀疏正则化半监督分类方法 (SSC-ASSR)

输入: 数据集 $\{(x_1, z_1), \dots, (x_l, z_l), x_{l+1}, \dots, x_N\}$, 惩罚项参数 γ_K, γ_I , 核函数 $k(\cdot, \cdot)$, 稀疏编码参数 K , 稀疏编码最大迭代步数 T

输出: 半监督分类函数 $g^*(x) = \arg \max_{s=1,2,\dots,C} \{f_s^*(x)\}$

- 1: 由算法 1 计算稀疏 l_0 - 图的权重矩阵 W ;
- 2: 计算核矩阵 $K = (k(x_i, x_j)) \in \mathbb{R}^{N \times N}$;
- 3: 根据式 (26) 计算核矩阵 K 的系数矩阵 B^* ;
- 4: 根据式 (27) 计算判别函数 $f^*(x)$.

4 实验

为了评价所提算法的分类性能, 本节在图 1(b) 所示人工数据以及如下 6 个标准数据集上进行了分类实验: USPS 手写数字^[34] 与 ORL 人脸^[35] 数据集¹⁾; 来自于 IDA 标准数据库的 Splice 与 Waveform 数据集²⁾; 来自于 UCI 机器学习库的 Tic-Tac-Toe Endgame 与 Pima Indians Diabetes 数据集³⁾. 人工数据集包含了 3 类数据, 每类有 50 个点. 对于 USPS 数据库, 我们从中随机选取了手写数字“0”与“1”的各 500 张图像, 构成了 USPS 数据集. ORL 数据集包含了 400 张图像, 它们属于 40 个不同的类. 我们所取的 Splice 与 Waveform 数据集各包含了 1000 个数据; Tic-Tac-Toe Endgame 与 Pima Indians Diabetes 数据集各包含了 958 和 768 个数据. 为了减轻不同维度的不同尺度所造成的负面影响, 对于所有数据集, 将训练数据矩阵 $X \in \mathbb{R}^{D \times N}$ 的每一行都进行归一化使该行的最大元素为 1.

在实验中参与比较的算法为: 本文所提的 SSC-ASSR, LapRLSC^[20], GRF^[17], 1- 近邻 (1-NN), LNP^[19], SRC^[22], l_1 -graph SSC^[24]. 对于参数选择问题, SSC-ASSR, LapRLSC, GRF 均将 Gauss 核参数 σ 作为算法实施的关键参数. 在实验中, 从 $\{0.01\sigma_0, 0.1\sigma_0, \sigma_0, 10\sigma_0, 100\sigma_0\}$ 中通过交叉验证法选择 σ , 其中 σ_0 为训练集中数据点对之间的 l_2 距离的均值. SSC-ASSR 与 LapRLSC 还需要指定惩罚项参数 γ_K 与 γ_I . 令 $CK = \gamma_K l$, $CI = \frac{\gamma_I l}{N^2}$, 可以发现算法对于这两个参数并不十分敏感, 因此在所有的实验数据中固定 $CK = 0.005$, $CI = 1$. SSC-ASSR 算法在计算基于核的稀疏表示时还需要设置参数 K , 在实验中将其设置为 $0.8 \times \min\{n_1, \dots, n_c\}$, 其中 n_i 为第 i 类训练数据所包含样本个数. SRC 算法只需要一个参数: 误差容忍度 ϵ . 在文献 [22] 中, 该参数被固定为 $\epsilon_0 = 0.05$, 因此在实验中我们将采取同样的设置. 对于 LapRLSC, GRF, LNP 所需要的邻域尺寸参数, 在 ORL 数据集中将其设置为 4 (ORL 训练数据集中每类只包含 9 个点), 在其他数据集中设为 7.

分类实验是按照如下方式进行的. 对于每个数据集 \mathcal{X} , 利用十倍交叉验证法进行测试. 令 \mathcal{X}_1 表示每次的训练集, \mathcal{X}_2 表示测试集. 随机从 \mathcal{X}_1 中选出 m 个数据点, 假定它们的类别标签是已知的, 由此得到有标签集 \mathcal{L} ; \mathcal{X}_1 中剩下的数据认为其标签未知, 由此得到无标签集 \mathcal{U} . 那么, 对于 1-NN, SRC 这两个有监督算法, 将有标签集 \mathcal{L} 作为训练数据集; 而对于 LapRLSC, GRF, LNP, l_1 -graph SSC 以及所提的 SSC-ASSR 这几种半监督算法, 将集合 $\mathcal{X}_1 = \mathcal{L} \cup \mathcal{U}$ 作为训练数据集, 即同时利用有标签数据与无标签数据进行训练. 然后首先对无标签数据集 $\mathcal{U} \subset \mathcal{X}_1$ 中的数据进行分类, 再对测试数据集 \mathcal{X}_2 中的所有数据进行分类.

各算法在训练数据集中的无标签数据 $\mathcal{U} \subset \mathcal{X}_1$ 上的分类结果如图 2 所示, 在测试数据集 \mathcal{X}_2 上的分类效果如图 3 所示, 其中 m 为 \mathcal{X}_1 中有标签数据的数目. 可以看出, 在人工数据集、USPS 数据集

1) <http://www.cad.zju.edu.cn/home/dengcai/Data/data.html>.

2) http://mldata.org/repository/tags/data/IDA_Benchmark_Repository/.

3) <http://archive.ics.uci.edu/ml/datasets.html>.

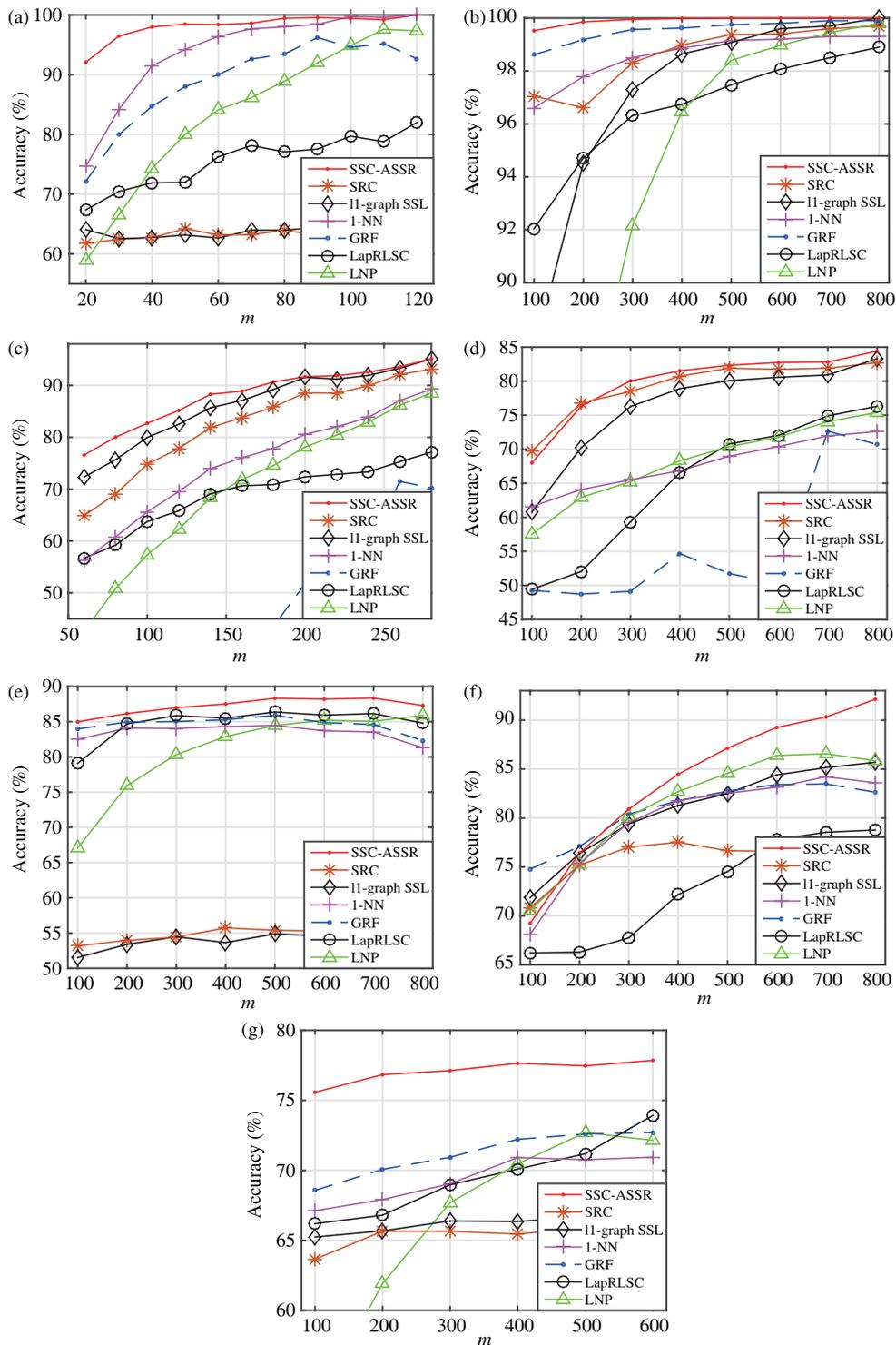


图 2 (网络版彩图) 各算法在无标签数据集 $U \subset \mathcal{X}_1$ 上的分类准确度, 其中 m 为 \mathcal{X}_1 中有标签数据的数目
Figure 2 (Color online) Classification accuracies of the algorithms on the unlabeled data set $U \subset \mathcal{X}_1$, where m is the number of labeled data points in \mathcal{X}_1 . (a) Synthetic Data; (b) USPS; (c) ORL; (d) Splice; (e) Waveform; (f) Tic-Tac-Toe Endgame; (g) Pima Indians Diabetes

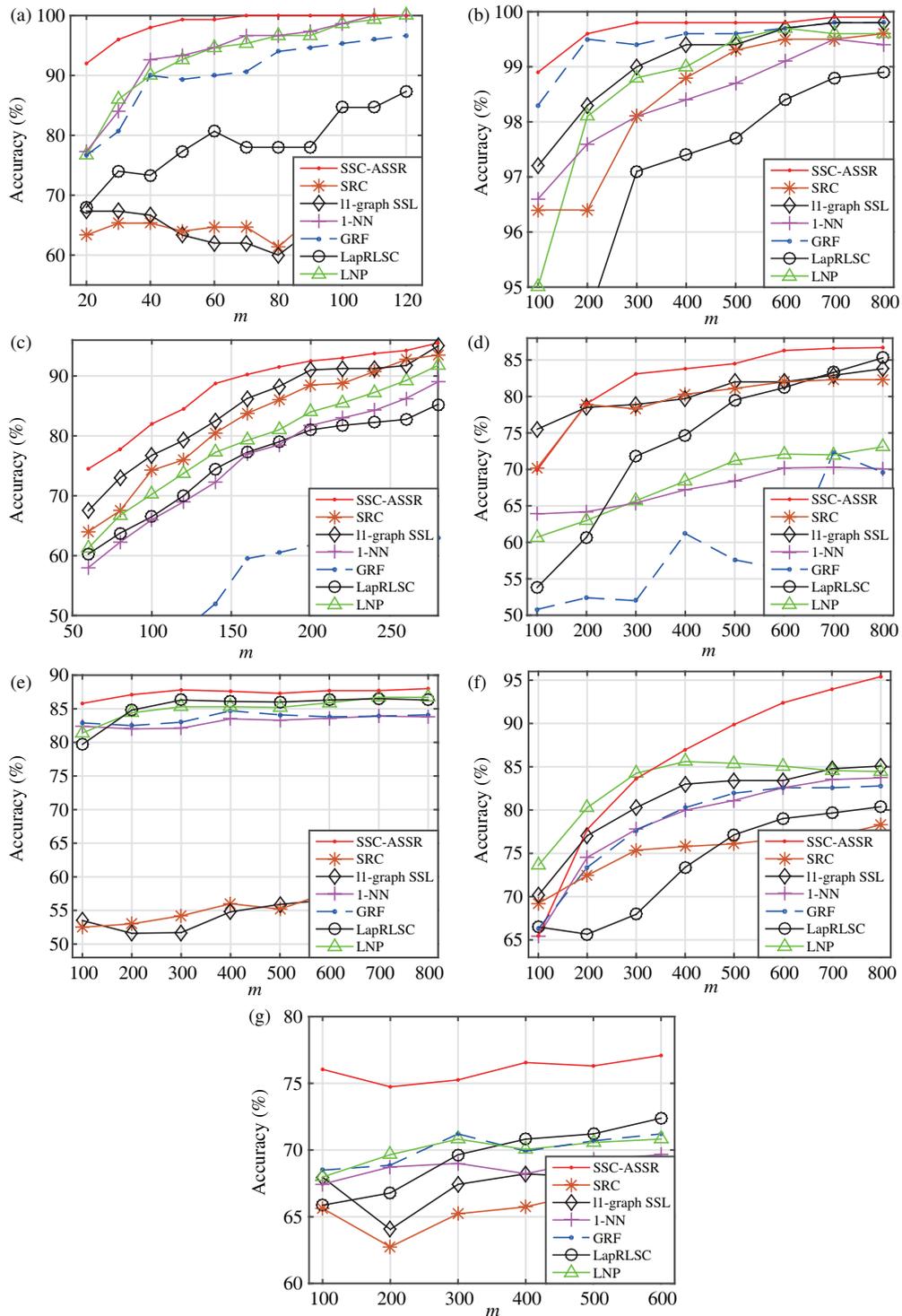


图 3 (网络版彩图) 各算法在测试数据集 \mathcal{X}_2 上的分类准确度, 其中 m 为 \mathcal{X}_1 中有标签数据的数目

Figure 3 (Color online) Classification accuracies of the algorithms on the test set \mathcal{X}_2 , where m is the number of labeled data points in \mathcal{X}_1 . (a) Synthetic Data; (b) USPS; (c) ORL; (d) Splice; (e) Waveform; (f) Tic-Tac-Toe Endgame; (g) Pima Indians Diabetes

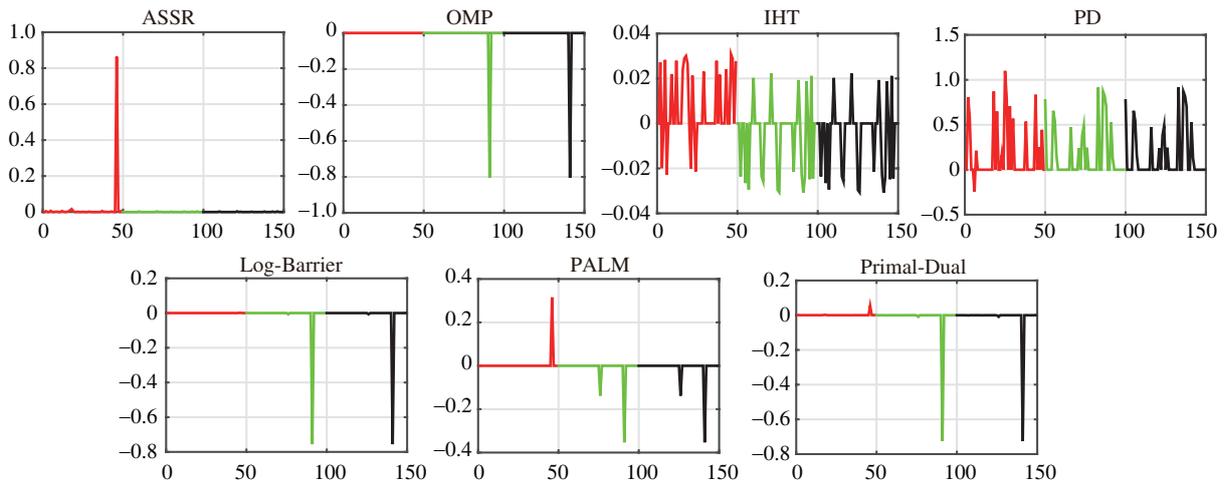


图 4 (网络版彩图) 对于人工数据集中的测试点, 将其余训练数据 (横轴所示) 作为字典所获的稀疏表示系数 (纵轴所示). 红色线对应的点 (即横坐标为 1 ~ 49 的点) 为与测试数据同类的点

Figure 4 (Color online) The sparse representation coefficients (as shown by vertical axis) of the test point in the synthetic data set, with the rest training data (as shown by horizontal axis) being the dictionary. The points corresponding to the red lines (i.e., the points with the horizontal coordinates of 1 ~ 49) are in the same class of the test point.

与 ORL 数据集上, 当有标签数据个数 m 较小时, 所提 SSC-ASSR 方法具有明显优势. 在 Splice 与 Tic-Tac-Toe Endgame 数据集上, 当有标签数据个数 m 较大时, 所提 SSC-ASSR 方法具有明显优势; 在 Pima Indians Diabetes 与 Waveform 数据集上, 所提 SSC-ASSR 方法一直具有明显优势. 对于基于数据稀疏表示的 SRC 与 l_1 -graph SSC 方法, 它们在人工数据集、Waveform 数据集、Pima Indians Diabetes 数据集上具有最差的效果, 这可能是由于这几个数据集违背了线性子空间的假设.

此外, 对图 1(b) 所示人工数据集中的测试点 x^* (蓝色正方形点), 将其余训练数据作为字典获得 x^* 的稀疏表示 $\alpha^* = [\alpha_1^*, \dots, \alpha_{N-1}^*]^T$, 结果如图 4 所示. 参与比较的算法包括: 本文所提基于仿射子空间的稀疏表示方法 (affine subspace based sparse representation, ASSR), OMP^[25], IHT^[26], Penalty Decomposition (PD)^[36]; Log-Barrier 内点方法^[37], PALM^[33], 基追踪问题的原始—对偶方法 (primal-dual algorithm)^[37]. 其中, 前 4 种为 l_0 稀疏表示问题求解算法, 后 3 种为 l_1 稀疏表示问题求解算法. 在图 4 中, 横坐标表示各训练数据的指标 i , 纵坐标表示第 i 个数据所对应的稀疏表示系数 α_i^* . 红色线对应的点 (即横坐标为 1 ~ 49 的点) 为与测试数据同类的点, 绿色与黑色线对应的点 (即横坐标为 50 ~ 149 的点) 为其他两类数据. 从图中可以看出, 对于所构造的满足仿射子空间假设的数据, 本文所提方法获得的稀疏表示系数能较准确地找到与测试数据 x^* 同类的点, 能够判断出数据间的仿射关系, 因此具有较好的判别性.

5 结论

本文提出了一种基于仿射子空间的稀疏正则化的半监督分类方法. 该方法首先将每个训练数据用其余训练数据稀疏编码, 该稀疏编码问题在最小化重构误差的同时考虑 3 个约束: 仿射组合约束, 非负性约束与稀疏约束; 然后基于所得稀疏表示系数自适应建立数据间近邻关系, 由此得到训练数据集的稀疏 l_0 -图; 最后在流形正则化理论框架下, 通过保持 l_0 -图中数据之间的稀疏重构关系, 获得显性的多类分类函数. 一系列人工数据与实际数据集上的实验验证了所提半监督分类方法具有较好的判别

性能. 然而, 该方法也存在一些问题. 例如, 算法需要对每个训练样本进行稀疏编码, 这使得算法运行效率比较低; 未来我们将研究快速的基于稀疏表示的半监督分类方法. 另外, 对于所提半监督分类算法的 Gauss 核函数中的参数 σ , 本文在实验中采用交叉验证法进行选取; 在未来的研究中, 将考虑使用更加快速有效的选取方法^[38].

参考文献

- 1 Chen K, Wang S H. Semi-supervised learning via regularized boosting working on multiple semi-supervised assumptions. *IEEE Trans Pattern Anal Mach Intell*, 2011, 33: 129–143
- 2 Xiang S M, Nie F P, Zhang C S. Semi-supervised classification via local spline regression. *IEEE Trans Pattern Anal Mach Intell*, 2010, 32: 2039–2053
- 3 Wang Y Y, Chen S C, Zhou Z H. New semi-supervised classification method based on modified cluster assumption. *IEEE Trans Neural Netw Learn Syst*, 2012, 23: 689–702
- 4 Chapelle O, Scholkoph B. *Semi-Supervised Learning (Adaptive Computation and Machine Learning)*. Cambridge: The MIT Press, 2006
- 5 Zhu X J. *Semi-supervised Learning Literature Survey*. Technical Report 1530. Madison: University of Wisconsin, 2006
- 6 Nigam K, McCallum A, Thrun S, et al. Text classification from labeled and unlabeled documents using EM. *Mach Learn*, 2000, 39: 103–134
- 7 Fujino A, Ueda N, Saito K. Semi-supervised learning for a hybrid generative/discriminative classifier based on the maximum entropy principle. *IEEE Trans Pattern Anal Mach Intell*, 2008, 30: 424–437
- 8 Maulik U, Chakraborty D. A self-trained ensemble with semisupervised SVM: an application to pixel classification of remote sensing imagery. *Pattern Recognit*, 2011, 44: 615–623
- 9 Dopido I, Li J, Marpu P R, et al. Semisupervised self-learning for hyperspectral image classification. *IEEE Trans Geosci Rem Sens*, 2013, 51: 4032–4044
- 10 Blum A, Mitchell T. Combining labeled and unlabeled data with co-training. In: *Proceedings of 11th Annual Conference on Computational Learning Theory, COLT'98*, Madison, 1998. 92–100
- 11 Zhang M L, Zhou Z H. CoTrade: confident co-training with data editing. *IEEE Trans Syst Man Cybern Part B-Cybern*, 2011, 41: 1612–1626
- 12 Joachims T. Transductive inference for text classification using support vector machines. In: *Proceedings of 16th International Conference on Machine Learning, Bled*, 1999. 200–209
- 13 Qi Z Q, Tian Y J, Shi Y. Laplacian twin support vector machine for semi-supervised classification. *Neural Netw*, 2012, 35: 46–53
- 14 Kang B Y, Ko S, Kim D W. SICAGO: semi-supervised cluster analysis using semantic distance between gene pairs in gene ontology. *Bioinformatics*, 2010, 26: 1384–1385
- 15 Luo Y, Tao D C, Geng B, et al. Manifold regularized multitask learning for semi-supervised multilabel image classification. *IEEE Trans Image Process*, 2013, 22: 523–536
- 16 Badrinarayanan V, Budvytis I, Cipolla R. Semi-supervised video segmentation using tree structured graphical models. *IEEE Trans Pattern Anal Mach Intell*, 2013, 35: 2751–2764
- 17 Zhu X, Ghahramani Z, Lafferty J. Semi-supervised learning using gaussian fields and harmonic functions. In: *Proceedings of the 20th International Conference on Machine Learning*, 2003. 912–919
- 18 Zhou D, Bousquet O, Lal T, et al. Learning with local and global consistency. In: *Proceedings of the Neural Information Processing Systems Conference, Vancouver*, 2004. 321–328
- 19 Wang F, Zhang C S. Label propagation through linear neighborhoods. *IEEE Trans Knowl Data Eng*, 2008, 20: 55–67
- 20 Belkin M, Sindhwani V, Niyogi P. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. *J Mach Learn Res*, 2006, 7: 2399–2434
- 21 Bruckstein A M, Donoho D L, Elad M. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev*, 2009, 51: 34–81
- 22 Wright J, Yang A, Ganesh A, et al. Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell*, 2009, 31: 210–227

- 23 Cheng B, Yang J C, Yan S C, et al. Learning with ℓ^1 -graph for image analysis. *IEEE Trans Image Process*, 2010, 19: 858–866
- 24 Yan S, Wang H. Semi-supervised learning by sparse representation. In: *Proceedings of SIAM International Conference on Data Mining, SDM'09, Sparks, 2009*. 792–801
- 25 Pati Y C, Rezaiifar R, Krishnaprasad P S. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: *Proceedings of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, Pacific Grove, 1993*. 40–44
- 26 Blumensath T, Davies M E. Iterative hard thresholding for compressive sensing. *Appl Comput Harmon Anal*, 2009, 27: 265–274
- 27 Donoho D. For most large underdetermined systems of linear equations the minimal ℓ^1 -norm solution is also the sparsest solution. *Comm Pure Appl Math*, 2006, 59: 797–829
- 28 Monteiro R, Adler I. Interior path following primal-dual algorithms. Part I: Linear programming. *Math Program*, 1989, 44: 27–41
- 29 Chen S S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit. *SIAM Rev*, 2001, 43: 129–159
- 30 Malioutov D, Cetin M, Willsky A. Homotopy continuation for sparse signal representation. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'05, Pennsylvania, 2005*. 733–736
- 31 Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J Imaging Sci*, 2009, 2: 183–202
- 32 Yang J F, Zhang Y. Alternating direction algorithms for ℓ_1 -problems in compressive sensing. *SIAM J Sci Comput*, 2011, 33: 250–278
- 33 Yang A Y, Zhou Z H, Balasubramanian A G, et al. Fast ℓ_1 -minimization algorithms for robust face recognition. *IEEE Trans Image Process*, 2013, 22: 3234–3246
- 34 Hull J J. A database for handwritten text recognition research. *IEEE Trans Pattern Anal Mach Intell*, 1998, 16: 550–554
- 35 Samaria F S, Harter A C. Parameterisation of a stochastic model for human face identification. In: *Proceedings of the Second IEEE Workshop on Applications of Computer Vision, WACV 1994, Sarasota, 1994*. 138–142
- 36 Lu Z S, Zhang Y. Sparse approximation via penalty decomposition methods. *SIAM J Optim*, 2013, 23: 2448–2478
- 37 Candes E, Romberg J. ℓ_1 -MAGIC: Recovery of Sparse Signals via Convex Programming. Technical Report. 2005
- 38 Xu Z B, Dai M W, Meng D Y. Fast and efficient strategies for model selection of support vector machines. *IEEE Trans Syst Man Cybern Part B-Cybern*, 2009, 39: 1292–1307

Semi-supervised classification based on affine subspace sparse representation

GU NanNan¹, FAN MingYu^{2*}, WANG Di², JIA LiHao³ & DU Liang⁴

¹ *School of Statistics, Capital University of Economics and Business, Beijing 100070, China;*

² *College of Mathematics & Information Science, Wenzhou University, Wenzhou 325035, China;*

³ *Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China;*

⁴ *School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China*

*E-mail: fanmingyu@wzu.edu.cn

Abstract Graph-based semi-supervised classification is one of the hottest research areas in machine learning and data mining. These methods usually model an entire dataset as a graph, then utilize the structure information extracted by the graph to help with the classification of unlabeled data. Generally speaking, the performance of graph-based semi-supervised classification methods highly depends on the constructed graphs. In this paper, we propose a new kind of graph construction method based on affine subspace sparse representation. The proposed sparse coding method minimizes the construction error of the input signal, considering three constraints: (1) the

input signal being approximately reconstructed by the affine combination of the dictionary; (2) the nonnegativity constraint of the reconstruction coefficients; (3) the sparsity constraint of the reconstruction coefficients. Based on the constraints, we present the l_0 -norm constrained optimization problem for sparse coding; then, we propose the algorithm to solve the problem and further construct the l_0 -graph of data. Finally, under the manifold regularization framework, we propose a new kind of semi-supervised classification method by introducing the regularization term that measures the structure preserving error of the l_0 -graph. The proposed semi-supervised classification method has an explicit multiclass classification function and inherits the strong discriminative information from sparse representation. As a result, it has efficient and effective classification ability. Experimental results on artificial and real-world datasets are provided to show the effectiveness of the proposed method.

Keywords classification algorithms, sparse data distributions, graphic methods, classification of signals, least squares method



GU NanNan was born in 1985. She received her Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2012. Currently, she is a lecturer at the School of Statistics, Capital University of Economics and Business, Beijing, China. Her research interests include theory and application of semi-supervised classification, manifold learning, and sparse representation.



FAN MingYu was born in 1984. He received his Ph.D. degree in applied mathematics from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2011. Currently, he is an associate professor at the College of Mathematics and Information Sciences, Wenzhou University, Wenzhou, China. His research interests include theory and application of manifold learning, sparse representation, and image classification.



WANG Di was born in 1984. He received his Ph.D. degree in applied mathematics from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2012. Currently, he is a lecturer at the College of Mathematics and Information Sciences, Wenzhou University, Wenzhou, China. His research interests include theory and application of support vector machines, dictionary learning, and image processing.



JIA LiHao was born in 1980. He received his Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 2011. Currently, he is an assistant professor at the Institute of Automation, Chinese Academy of Sciences. His research interests include computer vision and machine intelligence.