基于 DL1 图和 KNN 图叠加图的高光谱图像半监督分类算法

王君言¹, 张春梅^{1*}, 张云斌², 刘瑶瑶¹, 王晓峰¹, 李丹¹, 张少敏¹, 白静¹

1. 北方民族大学计算机科学与工程学院, 银川 750021

2. 中软国际有限公司欧美线 speech 组, 北京 100070

* 通信作者. E-mail: chunmei66@hotmail.com

收稿日期: 2017-03-07; 接受日期: 2017-05-08; 网络出版日期: 2017-08-30

国家自然科学基金 (批准号: 61461002)、宁夏自然科学基金 (批准号: NZ15105)、北方民族大学校级科研项目 (批准号: JSKY06) 和北方民族大学研究生创新项目 (批准号: YCX1657) 资助

摘要 基于少数已标记样本的高光谱图像分类是一个具有挑战的任务. 本文将概率矩阵与 L1 图的权值矩阵叠加, 形成了强鉴别力的 DL1 图. 将空间的局部信息与光谱的全局信息通过 KNN 图和 DL1 图结合在一起, 构建了空谱信息联合的图框架结构, 使用该框架构建的图, 能更精细地反映高光谱图像数据的图谱结构. 利用图的标记传播达到半监督分类的目的, 以此提高小样本高光谱图像自动分类的精度, 实验表明, 在标记样本比例为 5% 时, 分类精度提升亦非常显著.

关键词 高光谱图像, 半监督分类, 稀疏图, KNN 图, 标记传播

1 引言

高光谱图像包含地物的高分辨率的光谱信息, 对于用相似光谱信息区分不同地物类别是很有吸引力的. 然而, 高光谱图像分类通常面临着有标记样本数量有限的问题, 原因是收集大量的训练样本代价昂贵, 费时费力^[1~4]. 半监督学习, 能够同时使用少量的已标记样本和大量的未标记样本, 因而受到了机器学习领域的广泛关注^[5]. 在当前的半监督学习方法中, 由于在实践中实证性的成功和较高的计算效率, 基于图的半监督学习更受欢迎^[6]. 基于图的半监督学习使用一个图 $G = (V, E)$ 代表数据结构, 其中 V 是顶点集 (每个顶点对应一个数据样本), E 是与权值矩阵 W 关联的边集. 一个样本子集的标记信息可以通过图以某种方式传播到剩余的未标记样本, 利用此特性, 可以达到半监督学习的目的. 为了得到期望的分类精度, 构造一个好的图对于基于图的半监督方法是很关键的, 然而, 如何构造一个好的图仍然是一个开放的问题^[7,8].

Cheng 和 Yan 等^[9,10] 通过稀疏表示解决 L1 范数的最优化问题构造了一个 L1 范数图, 并被用于目标跟踪. 其核心思想就是通过解决 L1 优化问题, 每个样本能被其余样本稀疏线性叠加编码. 通过

引用格式: 王君言, 张春梅, 张云斌, 等. 基于 DL1 图和 KNN 图叠加图的高光谱图像半监督分类算法. 中国科学: 信息科学, 2017, 47: 1662–1673, doi: 10.1360/N112016-00265
Wang J Y, Zhang C M, Zhang Y B, et al. Semi-supervised classification algorithm of hyperspectral image based on DL1 graph and KNN superposition graph (in Chinese). Sci Sin Inform, 2017, 47: 1662–1673, doi: 10.1360/N112016-00265

这种方式, 图的邻接关系和权值能够同时自动获得. 与传统方法相比, 如 K 近邻图 (KNN 图) 和局部线性嵌入图 (LLE 图)^[11, 12], L1 图利用了数据之间更高阶的关系, 因此具有天然的区分能力. 然而, L1 图对每个样本稀疏表示是通过无监督的方法, 对样本间的相似性进行编码时缺乏效率.

受上述见解的启发, 我们提出将 L1 图和部分标记信息组合起来构造一个区分能力更强的 L1 图, 即 DL1 图. 对于 DL1 图, 当两个样本属于不同的聚类时可以降低这两个样本之间的权值.

为了解决高光谱数据分类中空谱结合的问题, 本文延用了文献 [13] 的构图思想: 首先构造一个 DL1 范数图, 作为主图, 以获得具有全局性的高光谱数据的光谱结构图, 然后构造一个 KNN 图, 作为辅助图, 以获得能反映高光谱数据局部特征的空间结构图. 最后给予辅助图一个适当比例, 将其与主图叠加得到 DL1KNN 图.

基于上述描述, 我们首先使用 SR (sparse representation) 方法估计两个像素属于同一类别的概率矩阵, 然后将概率矩阵整合到 L1 图中, 从而得到 DL1 图, 然后利用权值融合方法将数据的 DL1 图和 KNN 图进行加权线性叠加, 在半监督学习框架下构建一种能够同时揭示数据全局结构信息和局部结构信息 (空-谱结合) 且具有较高鉴别能力的 DL1 图和 KNN 图叠加图, 即 DL1KNN 图, 完成高光谱图像的半监督分类任务.

在 Indiana Pines 高光谱数据集上的实验结果表明, 与 L1 图、DL1 图、L1KNN 图方法相比, 在同样比例的标记样本下, 本文图算法分类精度均最高. 尤其是在最低比例的标记样本情况下, 本文图算法给出很好的结果.

2 DL1KNN 图的构建

2.1 构造 L1 图

受传统图和稀疏表示的影响, Yan 等^[14] 将每个样本用所有的其他样本稀疏表示, 利用稀疏表示系数作为图连接, 构造了一个 L1 范数图. 具体步骤是: 首先得到一个高维度表示 (每个样本和其他所有样本的线性关系) 的 L1 范数图, 然后进行再编码, 选择最有利最稀疏的样本组合, 得到一个稀疏表示图. 为了简化后续工作需要, 本文没有进行再编码 (稀疏编码), 直接将稀疏表示系数作为图连接, L1 范数图的具体构造如下所示.

对于样本集 $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{m \times n}$ 中的任意一个样本 x_i , 可以通过解决以下的 L1 范数优化问题求解它的稀疏系数:

$$\min_{\alpha^i} \|\alpha^i\|_1 \quad \text{s.t.} \quad x_i = D^i \alpha^i = \sum_{j \neq i} x_j \alpha_j^i, \quad \alpha^i \geq 0, \quad (1)$$

其中字典 $D^i = [x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n] \in \mathbb{R}^{m \times (n-1)}$ 表示去除样本 x_j ($j = i$) 的字典. $\alpha^i = [\alpha_1^i, \dots, \alpha_{j-1}^i, \alpha_{j+1}^i, \dots, \alpha_n^i] \in \mathbb{R}^{n-1}$, $j = [1, 2, \dots, n]$, $j \neq i$, α^i 表示稀疏系数矩阵, α_{j-1}^i 表示对应样本 x_{j-1} 的系数. 用 $G_1 = (X, W)$ 表示 L1 范数图, 样本集 X 作为图的顶点, W 作为图的权值矩阵, 则 W 的计算式如下:

$$W_{(L1)_{ij}} = \begin{cases} \alpha_j^i, & \text{if } i > j, \\ \alpha_{j-1}^i, & \text{if } i < j. \end{cases} \quad (2)$$

与 KNN 图和 LLE 图相比, 本文构建的稀疏图 (没有进行再编码, 直接将稀疏表示系数作为图连接权值) 可以同时获得全局关系和边权值, 并且具有区分能力. 然而, 这样构图并未考虑先验知识. 有

些时候事先知道某些边的标记, 并且希望在最终的图中将这些标记利用起来. 因此, 通过考虑部分已标记信息, 可以构造一个 DL1 图.

2.2 构造 DL1 图

本节采用非均匀类概率构建 L1 范数图, 这样构建的 L1 范数图被称为 DL1 范数图. 在这样一种图结构中, 每一对节点依据它们属于同一类别的概率被区别对待, 与 L1 图不同, 它强化了数据样本之间的关系, 因此具有更强的区分能力.

首先, 给出一种估计无标记样本类概率矩阵的方法, 然后给出 DL1 范数图的定义.

对于有标记样本, 它们与某一类别间的关系是确定的. 而无标记样本与类别间的关系是不确定的, 幸运地是, 我们可以通过部分标记信息估计无标记样本的类概率. 依据稀疏表示分类器 SRC (sparse representation-based classifier)^[15], 无标记样本中的一个测试样本可以被表示成训练样本的稀疏线性叠加, 如果在分解式中, 两个样本彼此间都含有对方的非零表示系数, 则意味着二者很可能属于同一类, 系数的大小决定了二者的相似程度. 基于这种特性, SRC 被用来估计无标记样本的类概率.

给出初始的标签矩阵 $Y_l \in \mathbb{R}^{l \times c}$, 如果数据 x_i 属于类别 j , $j \in [1, 2, \dots, c]$, 则有 $Y_{ij} = 1$, 否则 $Y_{ij} = 0$. D 是训练样本, $x_i \in X_u$ 是测试样本, 通过求解下面的最小化问题, 可以得到稀疏向量 $A \in \mathbb{R}^{l \times 1}$, 稀疏向量 A 表示测试样本 x_i 和 l 个训练样本之间的相似性:

$$\min \|A\|_1 \quad \text{s.t.} \quad DA = x_i, \quad (3)$$

其中, $\|A\|_1$ 是 L1 范数, 即 A 中所有成分的绝对值之和, x_i 的类概率向量由下式计算:

$$P_i = A^T Y_l, \quad (4)$$

其中 $P_i = (P_{i1}, P_{i2}, \dots, P_{ic}) \in \mathbb{R}^{1 \times c}$, 向量中的项 P_{ic} 代表 x_i 属于类别 c 的概率. 因而, x_i 和 x_j 属于同一类别的概率可由下式计算:

$$P_{ij} = \begin{cases} 1, & i = j, \\ P_i P_j^T, & i \neq j, \end{cases} \quad (5)$$

DL1 图中两个样本 x_j 和 x_i 的权值由下式给出:

$$W_{(\text{DL1})_{ij}} = W_{(\text{L1})_{ij}} P_{ij}, \quad (6)$$

与 L1 图不同, DL1 图利用了样本之间的先验类别信息, 区分能力更强.

2.3 构造 DL1 图和 KNN 图叠加图

KNN 图构造方法: 首先求取所有样本的空间最近邻集, 然后, 对于任意一个样本, 判断它是否属于某一个样本的空间最近邻集. 如果属于, 那么这两个样本在图中对应的边的权值为 1, 否则为 0. KNN 图的具体构造过程如下所示.

用 $G_2 = (X, K)$ 表示 KNN 图, 样本集 $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{m \times n}$ 作为图的顶点, K 作为图的权值矩阵. 假设 K_j 表示样本 x_j 的 K 近邻节点的集合, 对于任意一个样本 x_i , 如果 $x_i \in K_j$ 且 $i \neq j$ 则 $K_{ij} = 1$, 否则 $K_{ij} = 0$. KNN 图权值矩阵的计算式如下:

$$K_{ij} = \begin{cases} 1, & x_i \in K_j, i \neq j, \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

为了能够很好地获得数据的内在结构, 将 DL1 图和 KNN 图按比例进行叠加. 具体的叠加算法如下: 对应给定的样本集合 $X = L \cup U$, 其中, L 为有标记样本集, U 为无标记样本集, $l + u = n$, l 为有标记样本数, u 为无标记样本数, 构造一个 DL1 图和 KNN 图, 将 DL1 图作为主图, KNN 图作为辅助图, 通过一个固定比例 β 将二者叠加. 则叠加图可以表示为 $G_3 = (X, W_3)$, 样本集 X 为图的顶点, W_3 为图的权值矩阵. 权值矩阵 W_3 的计算式如下:

$$W_3 = W_{(\text{DL1})} + \beta K, \quad (8)$$

其中, β 是一个平衡数据全局性和局部性 (高光谱数据的光谱结构和空间结构) 的参数, 其取值大小通过实验确定.

可以通过算法 1 构造 DL1KNN 图.

算法 1 DL1KNN 图构造算法

- 1: 输入高光谱图像, 其中 l 个标记样本 $X_l = [x_1, x_2, \dots, x_l]$, u 个无标记样本 $X_u = [x_{l+1}, x_{l+2}, \dots, x_{l+u}]$, 初始的标记矩阵 $Y_l \in \mathbb{R}^{l \times c}$;
 - 2: 预处理样本: 归一化样本 $x_i = x_i / \|x_i\|_2$, 去掉样本 x_i 得到预处理样本 $X' = [x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n]$;
 - 3: 通过式 (2) 得到 L1 范数图权值矩阵 $W_{L1} = \{W_{ij}\}_{n \times n}$;
 - 4: 通过式 (4) 和 (5) 得到类概率矩阵 $\{P_{ij}\}_{n \times n}$;
 - 5: 根据式 (6) 获得 DL1 图的权值矩阵 $W_{(\text{DL1})}$;
 - 6: 通过式 (7) 得到 K 近邻矩阵 $K = \{K_{ij}\}_{n \times n}$;
 - 7: 通过式 (8) 将 DL1 图和 KNN 图叠加得到叠加矩阵 W_3 , 根据实验设置 β 的值为 0.2;
 - 8: 输出叠加图 $G_3 = (X, W_3)$.
-

3 基于 DL1KNN 图的高光谱图像半监督分类

基于图的半监督学习方法, 首先需要定义一个图, 然后利用基于图的半监督学习方法进行分类或者聚类. 本次实验使用了较为流行的基于图的半监督学习传播算法对 Indiana Pines 数据进行分类. 选择的方法是: Gauss 随机场和谐波函数 GHF (Gaussian fields and harmonic functions)^[16]. 在实验中, 比较 DL1KNN 图和其他图 (L1 图、DL1 图、L1KNN 图) 的分类效果. 将 $Y = [Y_l, Y_u]^T \in \mathbb{R}^{|V| \times c}$ 设为类标记矩阵. 如果样本 x_i 属于类 j , $j \in [1, 2, \dots, c]$, $Y_{ij} = 1$, 否则 $Y_{ij} = 0$. $V = \{v_i\}_{i=1}^n$ 是顶点集, c 为类别数. GHF 通过分类函数 $F = [F_l, F_u] \in \mathbb{R}^{|V| \times c}$ 实现标记传播. 它们利用图和未标记样本, 通过优化不同的预设能量函数重新获得连续的分类函数. 为了重新获得分类函数 F , GHF 结合 Gauss 随机场和谐波函数来优化下面的能量函数:

$$\min_{F \in \mathbb{R}^{|V| \times c}} \text{tr}(F^T \Delta F), \quad \Delta F_{UU} = 0, \quad F_L = Y_L, \quad (9)$$

其中, 样本集合 $X = L \cup U \in \mathbb{R}^{m \times n}$, L 为有标记样本集, U 为无标记样本集. $\Delta = D - W$ 为 Laplace 矩阵, $D_{ii} = \sum_j W_{ij}$ 为对角矩阵, 谐波的性质是每一个无标记样本 x_i 的 $F(x_i)$ 值是它在图中近邻的平均值,

$$F(x_i) = \frac{1}{D_{ii}} \sum_{j \sim i} W_{ij} F(x_j), \quad \text{for } x_i \in X_u, \quad (10)$$

其中, $0 \leq F(x_i) \leq 1$, $x_i \in X_u$ ($F(x_i) = 0$ or 1 , $x_i \in X_l$) 为了计算谐波解, 将权重矩阵 W 分成四块:

$$W = \begin{pmatrix} W_{LL} & W_{LU} \\ W_{UL} & W_{UU} \end{pmatrix}, \quad (11)$$

$$\Delta F = 0, \quad F_L = Y_L. \quad (12)$$

通过式 (12) 可以求得谐波解:

$$\begin{aligned} F_u &= (D_{UU} - W_{UU})^{-1} W_{UL} Y_L \\ &= -(\Delta_{UU})^{-1} \Delta_{UL} Y_L \\ &= (I - P_{UU})^{-1} P_{UL} Y_L, \end{aligned} \quad (13)$$

其中, $P = D^{-1}W$ 为图的转换矩阵, I 为单位矩阵. 令 $F^* = F_u$, 可以用下面的公式预测每个无标记样本 x_i ($x_i \in X_u$) 的类标记:

$$y_i = \arg \max_{j \leq c} F_{ij}^*. \quad (14)$$

4 实验和分析

通过设置相同的半监督学习条件, 本文在 Indiana Pines 高光谱数据上进行实验, 比较本文图与其他图的性能. 实验中对于数据子集中的每类样本随机选择 5%, 10%, 15%, 20% 和 25% 作为标记样本, 剩余的样本作为测试样本. 所有算法都通过 MATLAB 2012a 编程实现. 所有实验都在装有 Windows 7 系统电脑上运行 10 次, 电脑的处理器的 Intel(R)I7 2600K-Core 3.4 GHz, 运行内存是 4.00 G.

4.1 实验设置

AVIRIS Indiana Pines 是高光谱分类实验中一个常用的数据, 它是由美国国家航天局的机载可见红外成像光谱仪 (AVIRIS) 对美国 Indiana 州西北部印第安遥感实验区的成像, 于 1992 年获取. 它包含了农作物、草地及森林植被的混合区, 共 16 类地物. 整幅图像大小为 145×145 像素, 空间分辨率 $20 \text{ m} \times 20 \text{ m}$, 去掉 20 个杂波波段后剩余 200 个波段, 波长范围为 $400 \sim 2500 \text{ nm}$. 本文采用的图像是这幅 Indiana Pines 图的子场景, 使用像素 $[27 - 94] \times [31 - 116]$ 的一个子场景进行实验, 大小为 68×86 . 该子场景包含 4 类地物, 分别是: Soybean-mintill (在全图中 2468 个样本)、Corn-notill (在全图中 1434 个样本)、Soybean-notill (在全图中 968 个样本) 和 Grass-trees (在全图中 747 个样本). 子场景伪彩色图像及 4 类植被信息如图 1 所示, 实验参与比较的图如下所述.

L1 图: 按照文献 [16] 构造一个 L1 范数图. 在本文实验中求解 L1 优化问题均采用内点法 (interior point method), 我们使用 `l1_ls` 包 [17] 求解 L1 正则化最小平方问题, 求解过程中正则化参数 λ 的取值参见文献 [17] 中的建议及参考引用此包的相关论文中的取值. 内点法求解 L1 优化问题的主要参数参见文献 [18]. 内点法中主要有 6 个参数, 设置初始值 $t = \frac{1}{\lambda}$, μ 用来更新参数 t , $\mu = 2$; 内点法最大迭代次数 $\text{MAX_NT_ITER} = 4 \times 10^3$; 线性搜索时的参数有: 目标对象的最小下降比例 $\alpha = 0.01$, 步长递减系数 $\beta = 0.5$, 回溯线性搜索最大迭代次数 $\text{MAX_LS_ITER} = 10^3$. 由文献 [18], 初始值和参数的选择对算法性能的影响不大.

DL1 图: 在 L1 范数图中加入类概率矩阵, 直接比较类概率矩阵对 L1 图的分类精度提升效果.

图 1 AVIRIS Indiana Pines 子场景的 (a) 伪彩色图像及 (b) 4 类植被
Figure 1 (a) Pseudo color image and (b) four kinds of vegetation of the sub scene of AVIRIS Indiana Pines

表 1 在不同比例的标记样本下不同的图结合 GHF 标记传播方法的分类正确率 (%)

Table 1 Classification accuracy (%) of various graphs combined with the GHF label propagation method under different percentages of labeled samples

Indiana Pines (%)	L1 图	DL1 图	L1KNN 图				DL1KNN 图			
			$k = 5$	$k = 8$	$k = 10$	Average	$k = 5$	$k = 8$	$k = 10$	Average
5	0.556	0.744	0.792	0.891	0.883	0.855	0.818	0.876	0.906	0.867
10	0.771	0.873	0.887	0.909	0.873	0.890	0.881	0.910	0.880	0.890
15	0.715	0.868	0.932	0.904	0.898	0.911	0.913	0.909	0.918	0.913
20	0.769	0.871	0.931	0.924	0.918	0.924	0.931	0.923	0.928	0.927
25	0.794	0.880	0.939	0.916	0.923	0.926	0.948	0.920	0.923	0.930

L1KNN 图: 将 L1 图作为主图, KNN 图作为辅图, 通过一个固定比例 β 将他们叠加, 得到 L1KNN 图. 根据实验设置参数 β 的值为 0.2. 这里构造 3 个 L1KNN 图 (最近邻的个数为 5, 8 和 10), L1KNN 图分类结果取 3 个不同 L1KNN 图的平均值.

DL1KNN 图: 由 DL1 图与 KNN 图叠加得到. DL1KNN 图最近邻的个数分别为 5, 8 和 10, DL1KNN 图的分类结果取 3 个不同 DL1KNN 图的平均值.

4.2 分类结果分析

将不同的图与 GHF 半监督传播算法结合起来, 统计基于不同图的半监督分类结果. 对于 Indiana Pines 高光谱数据, 从数据中随机抽取 5%, 10%, 15%, 20% 和 25% 的样本作为标记样本, 剩余的作为未标记样本. 分类正确率取 10 次独立分类实验的平均值. 通常, 标记样本的比例越高, 分类精度也会越高, 但在某些应用场合, 如高光谱图像分类问题, 人工标记样本的成本非常昂贵, 在标记样本数目不足的情况下, 半监督学习可以通过机器学习方法扩展标记样本的数目, 进而在小样本情况下提高分类精度. 因此我们对标记样本比例较低的半监督学习算法的分类性能更感兴趣. 算法分类准确度结果如表 1 所示, 数字加粗体表示参与比较的实验结果.

图 2 给出了总的分类精度对比图.

图 2 总的分类精度对比图

Figure 2 The overall classification accuracy comparison chart

图 3 Indiana Pines 子图的分类结果, 标记样本比例 5%, $k=5$. (a) L1 图; (b) DL1 图; (c) L1KNN 图; (d) DL1KNN 图

Figure 3 Classification accuracies of sub scene of Indiana Pines, the percentage of labeled samples is 5%, $k=5$. (a) L1 graph; (b) DL1 graph; (c) L1KNN graph; (d) DL1KNN graph

在表 1 中, 将不同的图与 GHF 传播算法结合进行半监督分类. 对比 L1 图和 DL1 图的分类精度可知, 标记样本比例从 5% 到 25%, 分类精度增加 18.8% (5% 标记样本比例) 到 8.6% (25% 标记样本比例). 这表明与 L1 图不同, DL1 图利用了样本之间的先验类别信息, 区分能力更强.

对比 L1 图和 L1KNN 图的分类结果, 可知标记样本比例从 5% 到 25%, 分类精度增加 29.9% (5% 标记样本比例) 到 13.2% (25% 标记样本比例). 这表明 L1KNN 图可以通过叠加 KNN 图获得更接近真实数据的分布结构, 更适合半监督学习. 同时, L1KNN 图对 L1 图的分类精度提升高于 DL1 图对 L1 图的分类精度提升, 说明空谱信息相结合的图结构与单纯的光谱图改进相比, 能更准确地反映数据的内在关系, 是一种合理高效的图框架结构. DL1KNN 图的分类精度最高.

图 3 和 4 是标记样本比例为 5% 和 25%, $k=5$ 时的分类结果图.

图 5 给出了标记样本比例为 25%, $k=5$ 时, Indiana Pines 子场景分类的混淆矩阵.

图 6 给出了标记样本比例为 25%, $k=5$ 时, Indiana Pines 子场景分类的 OA 和 Kappa 系数.

图 4 Indiana Pines 子图的分类结果, 标记样本比例 25%, $k=5$. (a) L1 图; (b) DL1 图; (c) L1KNN 图; (d) DL1KNN 图

Figure 4 Classification accuracies of sub scene of Indiana Pines, the percentage of labeled samples is 25%, $k=5$. (a) L1 graph; (b) DL1 graph; (c) L1KNN graph; (d) DL1KNN graph

图 5 Indiana Pines 子图的混淆矩阵, 标记样本比例 25%, $k=5$. class1 是 Corn-notill, class2 是 Grass-trees, class3 是 Soybean-notill, class4 是 Soybean-mintill

Figure 5 Confusion matrix of sub scene of Indiana Pines, the percentage of labeled samples is 25%, $k=5$. class1: Corn-notill, class2: Grass-trees, class3: Soybean-notill, class4: Soybean-mintill. (a) L1 graph; (b) DL1 graph; (c) L1KNN graph; (d) DL1KNN graph

图 6 (网络版彩图) 标记样本比例为 25%, $k=5$ 时, Indiana Pines 子场景分类的 OA 和 Kappa 系数

Figure 6 (Color online) Overall accuracy and Kappa factor of Indiana Pines sub scene when percentage of labeled samples is 25%, $k = 5$

图 7 (网络版彩图) 标记样本比例为 25%, $k=5$ 时, Indiana Pines 子场景分类的各类别分类精度

Figure 7 (Color online) Accuracy per class of Indiana Pines sub scene when percentage of labeled samples is 25%, $k = 5$

图 8 (网络版彩图) 标记样本比例为 25%, $k=5$ 时, Indiana Pines 子场景分类的各类别漏分误差

Figure 8 (Color online) Omission of Indiana Pines sub scene when percentage of labeled samples is 25%, $k = 5$

图 7 给出了标记样本比例为 25%, $k = 5$ 时, Indiana Pines 子场景分类的各类别分类精度.

图 8 给出了标记样本比例为 25%, $k = 5$ 时, Indiana Pines 子场景分类的各类别漏分误差.

4.3 DL1KNN 图的参数敏感性实验

本节对 DL1KNN 图的参数敏感性进行了实验, 主要实验参数是叠加比例 β . β 是一个平衡数据全局性和局部性 (高光谱数据的光谱结构和空间结构) 的参数. β 值固定时, 最近邻的个数为 5, 8 和 10, 分类结果取 3 个不同 DL1KNN 图的平均值. 改变参数 β 的值, 统计 DL1KNN 图在不同比例 Indiana Pines 子图上的分类效果, 分类正确率取 10 次实验的平均值, 分类结果如表 2 所示. 根据实验结果, 最终确定 $\beta=0.2$. 表 2 描述了参数 β 和分类正确率之间的关系, 分类正确率最高的结果用数字加粗体表示. 图 9 是标记样本比例不同时, $\beta \in [0, 1]$ 值变化对应的分类正确率 (%) 曲线图.

表 2 标记样本比例不同时 β 值变化对分类正确率 (%) 的影响Table 2 The influence of varying β on the classification accuracy (%) under different percentages of labeled samples

β (%)	0	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	10	100
5	74.4	84.4	86.7	85.5	84.9	85.3	85.2	83.2	82.4	82.8	83.2	84.4	83.0	80.5
10	87.3	86.3	89.0	88.4	87.0	86.6	84.9	83.0	82.7	82.6	82.4	82.5	84.9	84.2
15	86.8	88.8	91.3	88.8	88.0	88.9	86.8	85.1	84.1	84.2	83.5	83.8	81.9	83.3
20	87.1	90.6	92.7	90.6	90.9	89.7	89.5	88.1	87.2	87.9	87.6	87.9	90.6	91.7
25	88.0	89.0	93.0	93.4	90.0	88.0	89.0	88.5	88.7	87.0	87.5	86.7	92.3	89.3

图 9 (网络版彩图) 标记样本比例不同时, 图叠加比例系数 $\beta \in [0, 1]$ 值变化对应的分类正确率 (%) 曲线图

Figure 9 (Color online) The curve graph of classification accuracy (%) with varying scale coefficient of graph under different percentages of labeled samples

5 结束语

本文研究了基于图的高光谱图像半监督学习问题. 该问题广泛存在于许多实际应用中, 直接利用现有图的构造方式所获取的半监督学习分类效果不甚理想^[13,14]. 为增强 L1 图的区分能力, 将 L1 图和部分标记信息组合起来构造一个区分能力更强的 DL1 图, 可有效提升分类精度, 为兼顾数据表示的全局结构和局部结构, 将 DL1 图与 KNN 图叠加, 可将遥感数据的光谱信息和空间信息紧密结合, 进一步提升了分类精度. 尤其是在标记样本比例为 5% 时, 分类精度提升均非常显著, 这对于解决小样本数据的分类难题有积极影响. 本文所提算法除可用于遥感图像分类具体应用之外, 也可推广到其他高维数据的分类应用中.

参考文献

- 1 Haq Q S U, Shi L X, Tao L M. Classification of hyperspectral data using a few training samples. In: Proceedings of IEEE International Conference on Progress in Informatics and Computing, Shanghai, 2010. 220–224
- 2 Zhang C M, Hao X T, Bai J, et al. Small samples hyperspectral data classification using a sparse based new model with

- learning dictionary. In: Proceedings of the 6th Workshop on Hyperspectral Image and Signal Processing, Shanghai, 2014
- 3 Zhou Q Q, Zhang C M, Zhang Y B. Relationship between the number of labeled samples and classification accuracy based on sparse representation. *Proc Sci*, 2015, 1: 97–103
 - 4 Gao Y, Ji R, Cui P, et al. Hyperspectral image classification through bilayer graph based learning. *IEEE Trans Image Process*, 2014, 23: 2769–2778
 - 5 Zhu X. Semi-supervised learning. In: *Encyclop-Edia of Machine Learning*. Berlin: Springer, 2010
 - 6 Liu W, He J, Chang S F. Large graph construction for scalable semi-supervised learning. In: *Proceedings of the 27th International Conference on Machine Learning*, Haifa, 2010. 679–686
 - 7 Zhou D, Bousquet O, Lal T N, et al. Learning with local and global consistency. In: *Proceedings of Conference on Advance Neural Information Processing System*, Tuebingen, 2004. 16: 321–328
 - 8 Zhu X, Lafferty J, Ghahramani Z. Semi-supervised learning using gaussian fields and harmonic functions. In: *Proceedings of the 20th International Conference of Machine Learning*, Washington, 2003. 912–919
 - 9 Cheng H, Liu Z, Yang J. Sparsity induced similarity measure for label propagation. In: *Proceedings of IEEE 12th International Conference of Computer Vision*, Kyoto, 2009. 317–324
 - 10 Yan S, Wang H. Semi-supervised learning by sparse representation. In: *Proceedings of SIAM International Conference of Data Mining*, Sparks, 2009. 792–801
 - 11 Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, 290: 2323–2326
 - 12 Wang F, Zhang C. Label propagation through linear neighborhoods. *IEEE Trans Knowl Data Eng*, 2008, 20: 55–67
 - 13 Zhang Y B, Zhang C M, Zhou Q Q, et al. Semi-supervised classification algorithm based on l1-norm and KNN superposition graph. *Pattern Recogn Artif Intell*, 2016, 29: 850–855 [张云斌, 张春梅, 周千琪, 等. 基于 l1 范数和 k 近邻叠加图的半监督分类算法. *模式识别与人工智能*, 2016, 29: 850–855]
 - 14 Cheng B, Yang J, Yan S. Learning with l1 graph for image analysis. *Image Process*, 2010, 19: 858–866
 - 15 Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell*, 2009, 31: 210–227
 - 16 Han S, Huang H, Qin H. Locality-preserving l1 graph and its application in clustering. In: *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, New York, 2015. 813–818
 - 17 Koh K, Kim S J, Boyd S. l1_ls: a matlab solver for large-scale l1-regularized least squares problems. Stanford: Stanford University, 2008. http://www.stanford.edu/~boyd/l1_ls/
 - 18 Koh K, Kim S, Boyd S. A method for large-scale l1-regularized logistic regression. In: *Proceedings of the 22nd AAAI Conference on Artificial Intelligence*, Vancouver, 2007. 565–571

Semi-supervised classification algorithm of hyperspectral image based on DL1 graph and KNN superposition graph

Junyan WANG¹, Chunmei ZHANG^{1*}, Yunbin ZHANG², Yaoyao LIU¹, Xiaofeng WANG¹, Dan LI¹, Shaomin ZHANG¹ & Jing BAI¹

1. *College of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China;*

2. *The Speech Group of Europe and the United States Business Line, Chinasoft International Co. Ltd, Beijing 100070, China*

* Corresponding author. E-mail: chunmei66@hotmail.com

Abstract The classification of hyperspectral images with a paucity of labeled samples is a challenging task. This paper describes the use of a superpose probability matrix and weight matrix of an L1 graph, thereby forming a strong discriminating DL1 graph. Combining the local information of the space with the global information of the spectrum through the superposition of a KNN graph and a DL1 graph, a graph-based framework is built that combines the spatial and spectral information. This framework of a DL1KNN graph can reflect the more sophisticated structure of hyperspectral image data. Experimental results show that the improvement in classification accuracy is significant when the percentage of labeled samples is 5% through the use of the label propagation of the graph to achieve semi-supervised classification for improving the automatic classification accuracy of hyperspectral data with a small number of samples.

Keywords hyperspectral image, semi-supervised classification, sparsity graph, KNN graph, label propagation

Junyan WANG was born in 1992. She is a postgraduate student at North Minzu University. Her research interests include image processing and analysis and machine learning.

based semi-supervised learning.

Chunmei ZHANG was born in 1964. She received her M.S. degree in communication & electronic systems from Tsinghua University, Beijing in 1991. She is currently a professor at North Minzu University. Her research interests include sparse approximations and mathematical data processing with applications in image and hyperspectral remote sensing data, pattern recognition, computer vision, deep learning, and graph-

Yunbin ZHANG is a graduate student from North Minzu University, China. His research focuses on image processing and machine learning. He is currently working at Chinasoft International, Beijing.

Yaoyao LIU was born in 1991. She received her B.S. degree in information and computing science from Qingdao University of Science and Technology, Qingdao in 2014. She is currently a postgraduate student at North Minzu University. Her area of research is image processing and analysis.