

基于 p 模的协同表达高光谱分类模型

雷开宇，中国农业大学数学与应用数学系

2018 年 4 月 7 日

目录

1	绪论	2
1.1	研究背景及意义	2
1.2	研究现状	2
1.2.1	HSI 图像特征	2
1.2.2	常见的 HSI 分类问题的解决方案	2
1.3	本文主要工作	2
1.4	本文组织结构	2
2	相关知识与理论基础	3
2.1	稀疏表达分类模型	3
2.2	协同表达分类模型	4
3	基于 l_p 范数的协同表达 HSI 分类模型	4
3.1	岭回归与 l_p 范数	4
3.2	Tikhonov 回归	5
3.3	降维方法	6
3.3.1	主成分分析 PCA	6
3.3.2	线性判别分析 LDA	7
3.4	基于 l_p 范数的协同表达分类模型 (pCRC)	7
3.4.1	空间滤波与数据预处理	7
3.4.2	pCRC-1	7
3.4.3	pCRC-2	8
4	实验与分析	9
4.1	数据集介绍	9
4.1.1	Indian Pines 数据集	9
4.1.2	Pavia 数据集	9
4.1.3	Salinas Scene 数据集	9
4.2	数据预处理	10
4.3	数值实验	11

摘要

TODO: 补充摘要，中文摘要 300 字左右，英文摘要 250 词左右

公元 1974 年, ACM 图灵奖授予了 Stanford 大学教授 Donald E. Knuth (高德纳), 表彰他在算法和程序语言设计等多方面杰出的成就。他的巨著 The Art of Computer Programming 令人震撼。另外, Knuth 的突出贡献还包括 T_EX 系统, 毫不夸张地评价, T_EX 给排版带来了一场革命。

1 绪论

在信息时代....

1.1 研究背景及意义

高光谱图像 (Hyperspectral Imagery, HSI), 是一种通过遥感技术获得的图像。高光谱影像收集及处理整个跨电磁波谱的信息, 这些信息是由高光谱传感器 (Hyperspectral image sensors) 获取并整理。正如人眼一样, 人眼能够看到物体, 是因为物体反射的光, 或者自身发出的光进入眼睛, 大脑感受到信号, 生成物体影响。与人眼不同的是, 人眼只能分辨出可见光, 而高光谱图像则可以延伸至红外、紫外, 甚至整个电磁波谱上。

HSI 被广泛应用于农业、地质、天文、化学、食品工程和环境工程等领域, 在军事上也有着重要的作用。研究者通过分析 HSI 来达到不同的目的。例如在地质勘测中, 可以通过遥感获取的 HSI 图像来定位矿区的主要矿种和位置, 在环境工程中, 通过对有毒气体泄漏区域的空气进行 HSI 分析, 可以确定毒气成分和扩散趋势。

HSI 分析中, 分类问题是研究的一个热门。在现实中, 人们往往能通过专业领域知识来确定出 HSI 中某个区域的物质构成, 以及少量的分布信息。例如在一个利用 HSI 进行农产品种植结构分析的例子中, 人们往往可以通过各地农业部门上报的信息, 来确定 HSI 中存在几种农作物, 并且可以根据采样调查知道某个区域农作物的具体分布情况。如何通过这些有限的信息来对整个 HSI 中的物质进行分类, 这是一个值得研究的问题。

机器学习在 HSI 分类中有着广泛的应用。当人们获取到 HSI 时, 由于人类自身生理结构和极限的限制, 直接对 HSI 进行分析无疑需要大量的人力物力, 成本过于高昂。同时, 人工方式分类 HSI 还会受到人类主观心理的影响, 例如对图像中特定区域, 不同的人可能认为该区域的成分分布不同。随之 HSI 传感器分辨率越来越高, 通过人力直接进行分类变得不再可行, 因此使用计算机来对 HSI 进行分析变得越来越重要。在分类问题中, 机器学习能够自主地通过已知信息来学习, 因此十分适合 HSI 分类问题。

从样本数量来说, 机器学习分为三类: 监督学习、半监督学习和无监督学习。监督学习是将已知标签的样本作为训练集, 对样本特征进行学习。无监督学习是指作为训练集的样本没有标签信息, 典型的无监督学习有聚类问题。半监督学习是指训练集中既有有标签的样本点, 又有无标签的样本点。在本文的假定下, HSI 分类问题是一个半监督学习问题, 通过数量较少的有标签样本点和部分无标签的样本点构造算法, 实现对 HSI 图像的自动分类。

1.2 研究现状

1.2.1 HSI 图像特征

半监督学习是利用....

1.2.2 常见的 HSI 分类问题的解决方案

1. 稀疏表达
2. 协同表达
3. SVM
4. 深度学习

1.3 本文主要工作

本文主要针对 HSI 分类问题，对协同表达算法进行了改进，使之在 HSI 分类中表现更好。主要工作有以下几点：

- (1) 在 CR 模型的基础上，引入了 p 模技术，把正则项中的范数改为 p 范数，取得更好的稀疏性。
- (2) 在 JCR 模型的基础上，引入了 p 模技术，以得到更好的稀疏性和精确度。
- (3) 将本文改进后的 p-CRC 模型、p-JRC 模型分别在 Indian、Pavia、Salinas 等数据集上，与 SRC、CRC、JCR 和 SVM 方法进行比较，分析实验结果。

1.4 本文组织结构

本文组织结构分为六部分，分别为：绪论、相关知识与理论基础、基于 p 模的协同表达 HSI 分类模型、基于 p 模的联合协同表达 HSI 分类模型、实验设计与分析、结论与展望。

2 相关知识与理论基础

2.1 稀疏表达分类模型

稀疏表示最初是应用在信号处理领域的一种压缩感知方法。稀疏表示的目的是在给定的字典中用尽可能少的元素来表示原信号。但是在近些年，稀疏表示渐渐地与信号处理背景相分离，逐渐成为发展成为了一种称为“字典学习”的机器学习算法。以稀疏表示理论基础的字典学习算法称为“稀疏字典学习 (Sparse Representation)”。

设样本集为：

$$T = \{(x_1, y_1), \dots, (x_m, y_l), x_{m+1}, \dots, x_{n+m}\} \quad (1)$$

其中， $x_i \in R^d, 1 \leq i \leq m$ 为输入的数据样本， m 为已知标签样本点的数量， n 为未知标签样本点的数量， $y_i, 1 \leq i \leq l$ 为某个已知标签样本点的标签信息。在稀疏字典学习中，一般将已知标签样本点作为训练集，而将未知标签样本点作为测试集。

对于给定训练集，我们将类别为 l 的已知标签样本集并列到一个矩阵中，组成一个“字典”。记这个字典为 D_l ：

$$D_l = [x_1, x_2, \dots, x_{n_l}] \quad (2)$$

对于测试集中的样本点 y ，求解以下优化模型：

$$\arg \min_{\alpha} \|y - D_l \alpha\|_2^2 + \lambda \|\alpha\|_0 \quad (3)$$

其中， α 是线性表达系数， $\lambda \|\alpha\|_0$ 称为稀疏项， λ 称为稀疏系数。稀疏项使用 l_0 模，该模的定义为：

$$\|x\| = n \quad (4)$$

其中， n 为 x 中非零元素的个数。根据 l_0 模的定义，由于稀疏项 $\lambda \|\alpha\|_0$ 的存在，整个优化问题(3)是一个非凸规划。在优化问题中，非凸规划的求解是十分困难的，甚至是无法求解的。值得庆幸的是，人们通过研究发现，在很多情况下，可以通过 l_1 模来凸近似 l_0 范数，因此稀疏表达的优化问题就变成了：

$$\arg \min_{\alpha} \|y - D_l \alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (5)$$

其中：

$$\|x\|_1 = \sum_{i=1}^n |x_i| \quad (6)$$

其中 n 为 x 的维度令 $r_l = \|y - D_l \hat{\alpha}\|_2$ ，其中 $\hat{\alpha}$ 为优化问题(5)的解，则测试样本 y 类别判断函数为：

$$\text{label}(y) = \text{identity}(\arg \min_l \{r_l\})$$

对于(5)，我们需要通过迭代的方法来求解近似解。其中，最简单的方法是近端梯度下降法 (Proximal Gradient Descent, PGD)。令 ∇ 为微分算子符号，对于优化问题(5)：

$$\arg \min_{\alpha} \|y - D_l \alpha\|_2^2 + \lambda \|\alpha\|_1$$

令 $f(x) = \|y - D_l \alpha\|_2^2$ ，显然， $f(x)$ 可微且 $\nabla f(x)$ 满足如下的 *Lipschitz* 条件：

定义 1. $\exists \beta > 0$ ，使得 $\forall x_1, x_2$

$$\|\nabla f(x_1) - \nabla f(x_2)\| < \beta \|x_1 - x_2\|$$

因此，在定义域上某点 x_k 附近， $f(x)$ 可通过展开为二阶泰勒公式：

$$\hat{f}(x) \simeq f(x_k) + \nabla f(x_k)^T (x - x_k) + o(\|x - x_k\|_2^2) \quad (7)$$

由上述 *Lipschitz* 条件可以得出， $\exists L > 0$ ，使得：

$$\frac{\|\nabla f(x_1) - \nabla f(x_2)\|_2^2}{\|x_1 - x_2\|_2^2} < L$$

则(7)就等价于：

$$\hat{f}(x) \simeq f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{L}{2} \|x - x_k\|_2^2 \quad (8)$$

$$= \frac{L}{2} \|x - (x_k - \frac{1}{L} \nabla f(x_k))\|_2^2 + \text{const} \quad (9)$$

其中 $const$ 是与 x 无关的常数。可以看出，使用梯度下降法对原函数 $f(x)$ 进行最小化，实际上就等价于对 $\hat{f}(x)$ 进行最小化，即对式(9)进行最小化。式(9)的最小值可通过求导解出，其最小值在 x_{k+1} 处获得：

$$x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k)$$

将上述思路带回优化问题(5)中，可得 PGD 的每一步迭代应为：

$$\alpha_{k+1} = \arg \min_{\alpha} \frac{L}{2} \|\alpha - (\alpha_k - \frac{1}{L} \nabla f(\alpha_k))\|_2^2 + \lambda \|\alpha\|_1 \quad (10)$$

稀疏表达的本质，其实是利用尽可能少的资源，来表示尽可能多的知识。这样可以带来的附加好处，就是会缓解存储压力。事实上，在“字典学习”中，当样本的数量级十分巨大时，字典学习的任务还包括对字典 D 进行学习，通过特征选择去除一些与学习任务无关的特征。这种情况下算法更为复杂，常见的算法有 $K-SVD$ 算法。在 HSI 分类问题中，相比样本数量，样本的维度较小，因此使用(5)所表示的优化问题即可。

2.2 协同表达分类模型

协同表达分类模型与稀疏表达分类模型思路相似，均为通过字典来表示测试点。不同于稀疏表达，协同表达分类模型中，使用的是 l_2 正则化，而不是 l_1 正则化。

协同表达中，对前述的 T, D_l, y ，求解优化问题：

$$\arg \min_{\alpha} \|y - D_l \alpha\|_2^2 + \lambda \|\alpha\|_2^2 \quad (11)$$

令 $\hat{\alpha}$ 为优化问题(11)的最优解，记 r_l 为测试点 y 与其近似线性表达的残差：

$$r_l = \frac{\|y - D_l \hat{\alpha}\|_2}{\|\hat{\alpha}\|_2}$$

测试样本点 y 的判别函数为：

$$\text{label}(y) = \text{identity}(\arg \min_l \{r_l\})$$

不同于 SRC, CRC 的求解过程是十分简单的。由于 l_2 范数具有可微性，因此优化问题(11)存在解析解：

$$\hat{\alpha} = (D_l^T D_l + \lambda I)^{-1} D_l^T y \quad (12)$$

以下是协同表达分类模型的算法：

Algorithm 1 Collaborative Representation Classifier

输入： 测试样本点 y ，训练样本集 D_l ，参数 λ ；

输出： 样本类别 $\text{identity}(y)$ ；

令 $l = 1$

循环：

 选取 D_l ，计算 $U = (D_l^T D_l + \lambda I)^{-1} D_l^T$

 计算 $\hat{\alpha} = Uy$

 计算 $r_l = \frac{\|y - D_l \hat{\alpha}\|_2}{\|\hat{\alpha}\|_2}$

循环停止条件： 训练样本全部计算完毕

 计算 $\text{identity}(y) = \arg \max_l \{r_l\}$

文献 [1] 中, Zhang Lei 将 SRC 与 CRC 进行了对比, 结果表明, 在人脸识别领域, CRC 能够达到与 SRC 相近的分类效果, 甚至在面对面部有遮挡的情况也有着较为优秀的效果。在 HSI 分类领域, CRC 也取得了较为成功的应用。众多研究者们通过不断对 CRC 进行改进, 使之在 HSI 分类问题中有了很好的表现。

3 基于 l_p 范数的协同表达 HSI 分类模型

3.1 岭回归与 l_p 范数

我们首先考虑一个回归问题。假设对训练集 $A = \{(x_1, y_1), \dots, (x_m, y_m)\}$ 进行线性回归, 其中 $x_i \in \mathbb{R}^d, y_i \in \mathbb{R}, 1 \leq i \leq m$ 。我们希望能够得出一条直线 $w^T x = b$ 。

我们知道, 在使用最小二乘法进行回归时, 要求解如下的二次优化问题:

$$\min_w \sum_{i=1}^m (y_i - w^T x_i)^2 \quad (13)$$

在回归问题中, 常常会遇到病态问题, 这种病态问题对输入数据十分敏感, 往往初始数据的输入差之毫厘, 最终的结果却谬以千里。病态的回归问题会导致模型的过拟合/欠拟合现象。在机器学习中, 对过拟合现象更为重视, 如果一个模型是过拟合的, 那么这个模型过度关注了训练集上的误差, 将训练集自身的特性当成了共有的特性, 导致模型在训练集上的误差很小, 而在测试集上的误差很大的情况。过拟合一般发生在**样本点 (相对样本特征) 不足**的情况。在机器学习中, 人们往往希望通过增加训练集来解决过拟合问题, 但是在很多场景中无法做到这点。为了在不增加训练集的情况下防止过拟合的发生, 研究者引入了多种正则化项。这些正则化项中, 最常见的正则化项是 l_2 正则项, 即在优化问题(13)目标函数中增加 w 的 l_2 范数:

$$\min_w \sum_{i=1}^m (y_i - w^T x_i)^2 + \lambda \|w\|_2^2 \quad (14)$$

其中, 参数 λ 称为正则化参数, 一般 $\lambda > 0$ 。优化问题(14)被称为“**岭回归 (Ridge Regression)**”, 由 Tikhonov 与 Arsenin 在 1977 年提出。

在 l_2 正则项的基础上, 人们将其推广为 l_p 正则项, 即使用 l_p 范数进行正则化, 解决过拟合现象。数学中, l_p 范数被定义为:

$$\|x\|_p = \sqrt[p]{\sum_{i=1}^n x_i^p}$$

优化问题(14)就变成了:

$$\min_w \sum_{i=1}^m (y_i - w^T x_i)^2 + \lambda \|w\|_p^p \quad (15)$$

由以上定义可得, 稀疏表达中稀疏项所使用的 l_1 范数、协同表达中的正则项使用的 l_2 也属于 l_p 范数。当使用 l_1 范数时, 优化问题(15)称为 LASSO (Least Absolute Shrinkage and Selection Operator) 回归, 由 Tibshirani 于 1996 年提出。

相比 l_2 范数, 使用 l_p 范数更易得到“稀疏解”, 其直观表示如下图所示

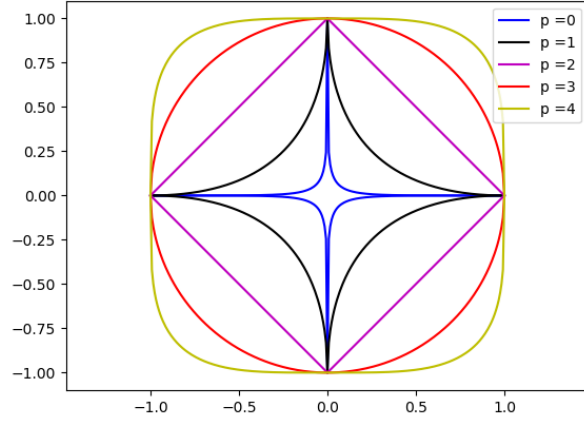


图 1: l_p 范数

显然，当 p 在 0 到 2 之间时， l_p 范数能够计算出较为优异的稀疏解。

3.2 Tikhonov 回归

Tikhonov 回归是另一种防止过拟合的方法。上世纪五六十年代，前苏联科学家 Andrey Nikolayevich Tikhonov 提出了通过正则化的方法来解决病态回归问题的方法，后来在机器学习领域得以广泛应用。

假设存在线性回归病态问题 $Dw = y$ ，其中 w 为所求回归超平面的系数， y 为 D 中样本点对应的值。正如前文所述，在使用最小二乘法时，要构造优化问题：

$$\arg \min_w \|Dw - y\|_2^2$$

为了解决病态问题导致的过拟合现象，Tikhonov 同样在优化函数中增加了一个正则项，将优化问题构造成为：

$$\|Dw - y\|_2^2 + \lambda \|\Gamma w\|_2^2$$

式子中， Γ 称为 Tikhonov 矩阵，这种正则化方法称为 Tikhonov 正则化 (Tikhonov Regularization)。在很多情况下，Tikhonov 矩阵会选择单位矩阵，此时 Tikhonov 正则化就是 l_2 正则化。除此之外，Tikhonov 矩阵还会有其他选择。广义的 Tikhonov 正则化中，Tikhonov 矩阵通常选择为 w 的协方差矩阵的转置矩阵。

3.3 降维方法

在数据挖掘与机器学习中，由高维空间引起的维数灾难常常是令人头痛的。维度灾难会使得计算复杂度呈指数型向上升，因此，在高维度数据挖掘中，进行特征提取，降低原数据的维度成为了必要的步骤。

定义 2. 对样本集合 $X = \{x, \dots, x_N\} \in R^{m \times N}$, $x_i \in R^m$ ，降维问题指的是：建立一个 $R^m \mapsto R^d, m > d$ 的一个映射矩阵 $A \in R^{m \times d}$ ，使得原样本通过映射 $z_i = A^T x_i$ 对应到低维子空间 R^d 上的点。

通过降维操作，我们能够显著地减小模型的计算复杂度。除此之外，降维还可以剔除原样本中的冗余信息，凸显出对数据影响最大的一些特征。

一般来讲，降维分为特征选择和特征提取两种，前者是在原有特征（后者称为属性）中选择出一些主要特征，而后者则是将原特征进行综合考虑，形成一些新的特征。在 HSI 中，每个像素背后都对应着一条光谱带，通常为 220 维，我们要确定某类物质，往往并不需要对全部的 220 个特征进行判断，因此我们可以通过降维来减少计算量，以加快运算速度。

在我们的模型中，我们主要使用主成分分析 (PCA) 方法与线性判别法 (LDA) 来降维。

3.3.1 主成分分析 PCA

主成分分析 (Principal components analysis) 是最常用的一种特征提取方法，其主要思想在于从数据的原始空间中提取主要特征，即主成分，降低样本维度，同时尽可能保持原始空间中数据的相对位置。

设有二维样本集 $X = \{x_1, \dots, x_N\}$, $x_i \in R^2$ ，我们希望通过 PCA 将该样本集降维到一维空间，即一条直线上。如下图所示，

我们想要寻找的这条直线应该满足这样的条件：

- (1) 最近重构性：样本点 x_1, \dots, x_n 到直线的距离应该尽可能小；
- (2) 最大可分性：样本点 x_i, \dots, x_n 在直线上的投影应该尽可能散落。

设样本点 x_i 在直线上的投影为 z_i ，则满足条件 (1) 的直线可通过如下优化问题来求解：

$$\min \sum_{i=1}^N \|x_i - z_i\|^2 \quad (16)$$

设直线的单位方向向量为 $w = a, b$ ，则 z_i 可以表示为： $z_i = w^T x_i$ 。

为了使直线满足条件 (2)，我们需要构建优化问题：

$$\max \sum_{i=1}^N \|z_i - \bar{z}\|^2 \quad (17)$$

其中 \bar{z} 是 z_i 的平均值，即 $\bar{z} = \frac{1}{N} \sum_{i=1}^N z_i$ 。

通过线性代数的推导，我们可以发现，优化问题(16)与(17)的目标函数其实是等价的，二者都可以转变成：

$$\max \text{tr}(wSw) \quad (18)$$

其中， $S = \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T$ ， \bar{x} 表示 x_i 的平均值。 S 为原样本的协方差矩阵，用来表示原样本的整体离散度。

优化问题(17)可以直接推广到高维空间，且易知 $S = XX^T$ 。设 $X = x_1, \dots, x_N$, $x_i \in R^{m \times N}$ 我们可以在高维空间上构造出同样的优化问题：

$$\begin{aligned} \arg \min_W \quad & \text{tr}(W^T XX^T W) \\ \text{s.t.} \quad & W^T W = I \end{aligned} \quad (19)$$

其中, $W = w_1, \dots, w_m$ 为投影变换后的坐标系且 w_i 为该坐标系的标准正交基向量。计算优化问题(19), 得到新的坐标系 W 。

根据线性空间理论和拉格朗日乘方法, 优化问题(19)等价于:

$$\arg \min_W \text{tr}(W^T X X^T W - \Lambda(W^T W - I)) \quad (20)$$

其中 $\Lambda = \lambda_1, \dots, \lambda_d$ 为拉格朗日系数。

优化问题(20)可直接通过求导来得到其极值:

令 $F(W) = W^T X X^T W - \lambda(W^T W - I)$, 当 $F(W)$ 的导数 $dF(W)$ 为零时, 有

$$\begin{aligned} 0 &= dF(W) \\ &= 2X X^T W - 2\Lambda W \\ &= X X^T W - \Lambda W \end{aligned}$$

即:

$$X X^T w_i = \lambda_i w_i \quad (21)$$

显然, 满足(21)的 λ_i, w_i 实质为矩阵 $X X^T$ 的特征值与特征向量。因此优化问题(19)就转变成了对协方差矩阵 S 进行特征值分解, 我们将特征值按照从小到大的顺序重新排列为 $\lambda_1, \dots, \lambda_d$, 按照特征值的顺序排列特征向量, 得到的

$$W = w_1, \dots, w_d$$

即为(19)的解, 原样本的映射为 $z_i = W^T x_i$ 。

这样得到的新的坐标系同原样本空间的维度是相同的, 为了达成降维的目的, 还需要舍去 W 中的部分坐标。通常降维后的空间维度 d' 是根据重构阈值 t 来确定:

$$\frac{\sum_{i=1}^{d'} \lambda_i}{\sum_{i=1}^d \lambda_i} \geq t \quad (22)$$

最终的 d' 就是使上式成立的最小的 d' 。

主成分分析降维算法如下:

Algorithm 2 主成分分析: Principal components analysis

输入: 原样本 $X = \{x_i\}$;

步骤:

1. 中心化样本: $x_i := x_i - \frac{1}{m} \sum_{i=1}^m x_i$;
 2. 计算协方差矩阵 $S = X X^T$;
 3. 对 S 进行特征值分解, 并将特征值从大到小进行排列;
 4. 根据公式(22)确定 d' ;
 5. 计算前 d' 个特征值的特征向量 w_i , 输出投影矩阵 $P = \{w_1, \dots, w_{d'}\}$;
 6. 计算降维后的投影 $z_i = P^T x_i$ 。
-

3.3.2 线性判别分析 LDA

线性判别分析

3.4 基于 l_p 范数的协同表达分类模型 (pCRC)

3.4.1 空间滤波与数据预处理

3.4.2 pCRC-1

正如前文所述, 相比 l_2 范数, l_p 范数能够增加模型的稀疏性, 得到更好的学习表现, 同时减少内存存储压力; 相比 l_1 范数, l_p 范数一定程度上又能加速训练速度, 因此, 我们不妨考虑使用 l_p 范数作为协同表达的正则项。

对比优化问题(5)与(11), 我们可以对比二者得到的表达系数 α 。图(2)表示的是在测试点 y 与训练集 X 在同样参数 λ 下的测试结果。

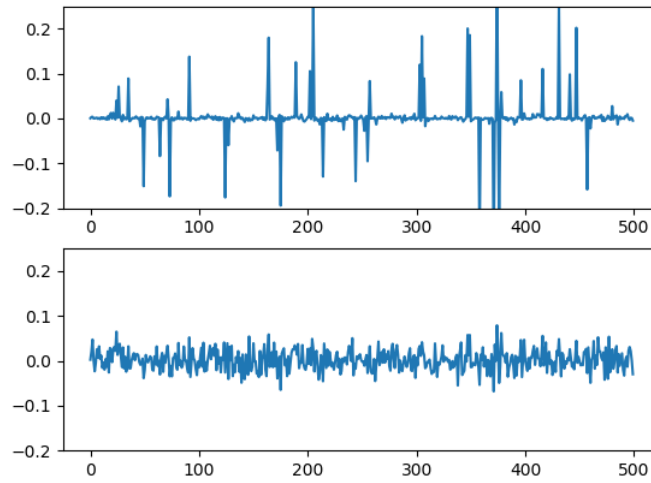


图 2: l_1 范数与 l_2 范数对比

其中, 上图表示的是 l_1 范数得到的 α , 下图表示的是 l_2 范数下得到的 α 。显然, l_1 范数得到的结果更加稀疏。在字典学习中, 稀疏的表达系数往往能够体现出与测试样本相关性比较大的训练样本。根据线性空间理论, 如果一个样本处于某个线性子空间中, 该样本可以表示成该子空间一组极大线性无关组的线性组合。在 HSI 分类问题中, 我们认为, 高光谱数据的每一类样本处于同一线性子空间中, 因此, 我们期望最终的线性表达系数能够体现出样本所处的线性子空间的特征。

理想中的线性表达系数应该满足这样的特征:

- (1) 对样本点的线性表达应该尽可能接近 y , 使得 $\|y - D\alpha\|_2$ 尽可能小;
- (2) 线性表达系数中, 非零项应该集中在与测试点处于同一子空间中的那些点。

事实证明, 线性表达系数 α 越稀疏, $D\alpha$ 就越接近 y , 即残差 $\|y - D\alpha\|^2$ 越小。下表是在不同情况下对残差的计算, 很明显, l_1 范数下的残差要小于 l_2 。但若表达系数过于稀疏, 则可能不足以准确地表达出样本所处的子空间, 因此, 我们希望的是得到一个较为稀疏的解。 l_1 得到的解过于稀疏, 因此我们需要用 $p \in (1, 2)$ 来取得适当的稀疏性。

我们构造优化问题:

$$\arg \min_w \|y - D\alpha\|_2^2 + \lambda \|\alpha\|_p^p \quad (23)$$

解决优化问题(23)，需要通过迭代来计算，具体步骤如下：

Algorithm 3 基于 l_p 范数的协同表达模型的迭代算法：pCRC

- 记 $1-$ 向量为 L ，初始化向量 $\beta_0 = L$ ， β_0 的维度与 α 相同
1. 计算 $\alpha = (D^T D + \lambda \beta^T \beta)^{-1} D^T \cdot y$;
 2. 计算 $\beta_1 = p(|\alpha| + L)^{p-2}$;
 3. 计算 $error = \|\beta_0 - \beta_1\|$ ，并令 $\beta_0 = \beta_1$;
 4. 重复上 (1) ~ (3)，直到 $error$ 小于某个值，或迭代次数到达最大值;
-

3.4.3 pCRC-2

事实上，对于优化问题(11)，我们可以将其作为一个以单位矩阵为 Tikhonov 矩阵的优化问题：

$$\arg \min_{\alpha} \|y - D_l \alpha\|_2^2 + \lambda \|\Gamma_l \alpha\|_2^2, \quad \Gamma = I$$

在文献 [2] 中，Li Wei 等学者将 Tikhonov 矩阵进行了更改，在文章中，Tikhonov 矩阵被定义为一个对角矩阵，其中每个对角元为测试点 y 与 D_l 中的每个训练点的距离，构造了联合协同表达 (Joint Collaborative Representation Classifier)：

$$\Gamma_l = \begin{bmatrix} \|y - x_{l,1}\|_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \|y - x_{l,m}\|_2 \end{bmatrix} \quad (24)$$

文章通过改变 Tikhonov 矩阵，将测试点与训练点的相对位置关系考虑到了模型中，优化了模型的表现。实验结果显示，该模型相比传统的模型，分类精确度有所提升。

我们认为，上述的 Tikhonov 能够有效地将样本点的位置信息考虑进入模型中，因此我们考虑将该 Tikhonov 矩阵引入 pCRC 模型中。通过增加 Tikhonov 矩阵，我们构造出了优化问题(25)：

$$\arg \min_w \|y - D\alpha\|_2^2 + \lambda \|\Gamma\alpha\|_p^p \quad (25)$$

Γ 的定义同上。

优化问题(25)的求解类似于 pCRC-1：

Algorithm 4 基于 l_p 范数的联合协同表达模型的迭代算法：pCRC-2

- 记 $1-$ 向量为 L ，初始化向量 $\beta_0 = L$ ， β_0 的维度与 α 相同
1. 计算 $\alpha = (D^T D + \lambda \beta^T \beta \Gamma^T \Gamma)^{-1} D^T \cdot y$;
 2. 计算 $\beta_1 = p(|\alpha| + L)^{p-2}$;
 3. 计算 $error = \|\beta_0 - \beta_1\|$ ，并令 $\beta_0 = \beta_1$;
 4. 重复上 (1) ~ (3)，直到 $error$ 小于某个值，或迭代次数到达最大值;
-

4 实验与分析

为了验证上述模型能够取得更好的表现，我们使用了 Indian Pines, University of Pavia 与 Salinas Scene 三组数据集来进行实验，分别对比 $PCA-CRC$, $PCA-JCRC$, $PCA-pCRC1$, $PCA-pCRC2$, $LDA-CRC$, $LDA-JCRC$, $LDA-pCRC1$, $LDA-pCRC2$ 在这三个数据集上的分类精确度，以此判断算法的优劣。

4.1 数据集介绍

我们选择了三个训练集来进行实验，这三个数据集分别为 Indian Pines, University of Pavia 与 Salinas Scene。

4.1.1 Indian Pines 数据集

Indian Pines 高光谱图像包含 145×145 个像素点，每个像素点是一条 224 维的光谱带，波长范围 $0.4 \sim 2.5 \times 10^{-6}$ 。该数据集取自 Indiana 西北部的高光谱图像。Indian Pines 高光谱图像中，有三分之二的部分为农业区域，另外三分之一的区域为森林与其他多年生植物。另外，图像中还包含了两条双车道高速公路，一条铁路，以及一些低矮的建筑物和乡间小路。整个数据集上的样本被分为 17 类，其中包括 16 类有效样本与背景样本，在实验中，我们只采用 16 类有效样本，而将背景点舍弃。

在 Indian Pines 数据集中，每一类样本的数量分布如下：

表 1: Indian Pines 数据集样本分布

编号	类别名称	样本数量
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-Grass-Trees-Drives	386
16	Stone-Steel-Towers	93

Indian Pines 高光谱图像是常用的高光谱图像。由于该图像拍摄于 6 月份，正值农作物生长旺盛的阶段，农作物覆盖了第 17 中有效物质：水。因此，在实验中常常用修正过的数据集。修正后的数据集剔除了光谱中属于水的信息，由 224 维减少到 220 维。

4.1.2 Pavia 数据集

Pavia 高光谱图像包含两个场景，一个场景是意大利北部城市 Pavia 的市中心，另一个场景是 Pavia 大学。二者均由飞机搭载高光谱相机拍摄，其中前者光谱带有 102 维，后者光谱带 103 维。Pavia 市中心是一个 1096×1096 的图像，Pavia 大学规模稍小，有 610×610 个像素。两个图像中都存在一些不包含任何有用信息的点，在收集数据时我们已经将这些点剔除。

修正后的 Pavia 市中心与 Pavia 大学中都有 9 类点，两个数据集各类点分布如下：

Pavia 市中心数据集

表 2: Pavia Centre 数据集样本分布

编号	类别名称	样本数量
1	Asphalt	816
2	Bitumen	808
3	Bare Soil	820
4	Meadows	824
5	Self-Blocking Bricks	808
6	Shadows	476
7	Tiles	1260
8	Trees	820
9	Water	824

Pavia 大学数据集

表 3: University of Pavia 数据集样本分布

编号	类别名称	样本数量
1	Asphalt	6631
2	Bare Soil	5029
3	Bitumen	1330
4	Gravel	2099
5	Meadows	18649
6	Painted metal sheets	1345
7	Self-Blocking Bricks	3682
8	Shadows	947
9	Trees	3064
10	Water	824

4.1.3 Salinas Scene 数据集

Salinas Scene 数据收集自美国加利福尼亚州 Salinas 山谷。Salinas 山谷是美国著名的农业区，Salinas Scene 数据集同样是农业领域的高光谱图像。该图像分辨率 512×217 ，包括了作物区、裸地和葡萄园。同 Indian Pines 一样，整个数据集有 16 个有效类，光谱维度 224 维。类似的，由于植被覆盖了水，因此舍弃光谱信息中与水有关的光谱带，对源数据进行了修正。

Salinas Scene 数据集类别分布如表(4)所示。

鉴于 Salinas Scene 数据集数据量较大，因此常常采用它的一个子集来做实验。这个子集称为 Salinas-A，选取自原图中 [591:676, 158:240] 的数据，即分辨率为 86×83 。该子数据集包含六类点，如表(5)所示。

4.2 数据预处理

高光谱图像是一种带有空间信息的高维数据。因此，一般来说，我们在处理高光谱数据前要先使用空间滤波对图像进行处理。增加空间滤波的目的，是为了减少噪

表 4: Salinas Scene 数据集样本分布

编号	类别名称	样本数量
1	Brocoli-green-weeds-1	2009
2	Brocoli-green-weeds-2	3726
3	Fallow	1976
4	Fallow-rough-plow	1394
5	Fallow-smooth	2678
6	Stubble	3959
7	Celery	3579
8	Grapes-untrained	11271
9	Soil-vinyard-develop	6203
10	Corn-senesced-green-weeds	3278
11	Lettuce-romaine-4wk	1068
12	Lettuce-romaine-5wk	1927
13	Lettuce-romaine-6wk	916
14	Lettuce-romaine-7wk	1070
15	Vinyard-untrained	7268
16	Vinyard-vertical-trellis	1807

表 5: Salinas Scene 数据集样本分布

编号	类别名称	样本数量
1	Brocoli-green-weeds-1	391
2	Corn-senesced-green-weeds	1343
3	Lettuce-romaine-4wk	616
4	Lettuce-romaine-5wk	1525
5	Lettuce-romaine-6wk	674
6	Lettuce-romaine-7wk	799

声，将整个图像进行平滑化处理。大量数据实验证明，通过采用滤波预处理，能够显著地提高算法的表现。

滤波处理是计算机视觉 (CV) 领域中的概念，常见的滤波器有平滑滤波器、锐化滤波器、统计排序滤波器等。在高光谱中，由于其目的是为了使不同区域之间的边缘变得平滑，减少图像的噪声，因此使用的是平滑滤波器。其中，最为常用的是平均滤波。

设图像中某个像素点 x_{ij} 周围空间的近邻点集合为 R ，用矩阵表示为：(以 3×3 的近邻点为例)

$$R_{3 \times 3} = \begin{bmatrix} x_{i-1,j-1} & x_{i-1,j} & x_{i-1,j+1} \\ x_{i,j-1} & x_{i,j} & x_{i,j+1} \\ x_{i+1,j-1} & x_{i+1,j} & x_{i+1,j+1} \end{bmatrix} \quad (26)$$

我们定义一个卷积核 G :

$$G_{3 \times 3} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (27)$$

令 $x_{i,j}$ 为矩阵 R 与 G 的卷积, 代替原图像中的 $x_{i,j}$:

$$\begin{aligned} x_{i,j} &= g(R_{3 \times 3}, G_{3 \times 3}) \\ &= \frac{1}{9} \sum_{m=i-1}^{i+1} \sum_{n=j-1}^{j+1} x_{m,n} \end{aligned}$$

另外, 当 $x_{i,j}$ 是图像的边界点时, 我们需要先对该点附近的近邻点进行镜像反转, 然后再进行卷积运算。通常我们常用 3×3 和 5×5 作为选取近邻点的窗口。

4.3 数值实验

科学合理地设置实验, 是机器学习领域研究中十分重要的部分。一个好的数值实验应该能够体现算法的性质, 对算法的优劣进行一个评判。在本文中, 我们设置了一系列对照实验, 来达到以上目的。

参考文献

- [1] Zhang L, Yang M, Feng X. Sparse representation or collaborative representation: Which helps face recognition?[C]//Computer vision (ICCV), 2011 IEEE international conference on. IEEE, 2011: 471-478.
- [2] Li W, Du Q. Joint within-class collaborative representation for hyperspectral image classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(6): 2200-2208.