

国家自然科学基金项目（编号：61402424）资助
中国地质大学（武汉）中央高校基本科研业务费专项资金项目资助



高光谱遥感数据 降维与分类

GAO GUANG PU YAO GAN SHU JU
JIANG WEI YU FEN LEI

姜鑫维 李程俊 刘 然 等著



中国地质大学出版社
ZHONGGUO DIZHI DAXUE CHUBANSHE

国家自然科学基金项目(编号:61402424)资助

中国地质大学(武汉)中央高校基本科研业务费专项资金项目资助

高光谱遥感数据降维与分类

GAO GUANG PU YAO GAN SHU JU JIANG WEI YU FEN LEI

姜鑫维 李程俊 刘 然 等著

图书在版编目(CIP)数据

高光谱遥感数据降维与分类/姜鑫维等著. —武汉:中国地质大学出版社,2016. 12

ISBN 978-7-5625-3964-3

I. ①高…

II. ①姜…

III. ①遥感数据-数据处理

IV. ①TP751.1

中国版本图书馆 CIP 数据核字(2016)第 327090 号

高光谱遥感数据降维与分类

姜鑫维 李程俊 刘 然 等著

责任编辑:王 敏 王凤林

责任校对:徐蕾蕾

出版发行:中国地质大学出版社(武汉市洪山区鲁磨路 388 号)

邮政编码:430074

电 话:(027)67883511

传 真:67883580

E-mail:cbb@cug.edu.cn

经 销:全国新华书店

<http://www.cugp.cug.edu.cn>

开本:787 毫米×1092 毫米 1/16

字数:186 千字 印张:7.25

版次:2016 年 12 月第 1 版

印次:2016 年 12 月第 1 次印刷

印刷:武汉珞南印务有限公司

ISBN 978-7-5625-3964-3

定价:30.00 元

如有印装质量问题请与印刷厂联系调换

前言

高光谱遥感图像分类技术可以实现比普通图像分类技术更精准的识别地表物质的类别信息,在城镇土地规划、管理和精细农业等方面有重要应用。但是,在高光谱遥感图像的分类过程中也面临一些问题:高光谱遥感数据的快速增长,导致对图像进行标注需要大量的人力和物力,因此作为训练的标注数据常常非常有限;在进行地物分类过程中,某些类别对应的特征不够显著会造成后续分类器精度的降低;当训练样本有限时,分类器最终给出的总体分类精度会随着特征维数的增加而降低,即“维数灾难”问题。综合以上问题,在前两个问题无法有效解决的情况下,我们致力于研究“维数灾难”问题,针对高光谱数据的特点基于降维模型进行特征提取,通过学习表达能力强的低维特征来提高高光谱遥感图像分类模型的精度。

尽管国内外研究者针对上述问题已有很多相关研究,提出了很多基于降维模型的高光谱数据处理模型,但尚未找到一本专著能够全面地从降维模型的角度去综合分析对比不同的降维模型对高光谱遥感图像分类的影响。高光谱遥感图像信息量丰富,每个像素点不仅存在光谱特征信息,而且还包含空间特征信息,因此,针对高光谱数据的上述特点,本书不仅考虑对光谱特征进行降维,更进一步考虑对空间特征进行降维。

针对基于降维的高光谱遥感图像分类,本书主要的内容包括:

(1)综述了高光谱遥感图像分类的流程,对每个流程的具体内容和其中的典型模型进行简述和回顾。

(2)综述了一些常见的高维数据降维算法,并从线性降维和非线性降维、有监督降维和无监督降维的角度对典型的降维算法进行了简述和回顾。

(3)分别基于高光谱数据的谱间相关性和空间相关性对地物进行分类,将典型的线性和非线性降维算法应用于高光谱图像分类,并对这些常用的降维算法在高光谱图像分类问题上进行了分析和对比。实验部分选取3个高光谱图像样本集,采用两种降维方式,即直接将所有数据降维、采用样本外点的计算和估计函数的降维,选择了13个降维算法,分类算法选择的是 k 最近邻算法,主要以总体分类精度为评价标准,完成参数 k 与总体分类精度的对比曲线图和维度与总体分类精度的对比曲线图,统计每个降维算法对测试样本进行降维、分类所需的时间,绘制测试样本的真实地物分割图,对比降维算法的效果,得出13个降维算法的特点。

(4)介绍近期提出的两个新的有监督降维算法,并应用于高光谱遥感图像分类任务中,在7个高光谱遥感图像分类数据集上的实验显示,这两个新提出的降维模型非常适宜于高光谱遥感图像降维任务,在实验中显示出了卓越的性能。

(5)最后对内容进行总结,进一步展望今后的研究计划。

作者

2016年11月

目 录

1 绪 论	(1)
1.1 高光谱遥感技术简介	(1)
1.1.1 遥感技术	(1)
1.1.2 遥感数字图像处理	(1)
1.1.3 高光谱遥感的发展	(1)
1.1.4 高光谱遥感图像数据的特点	(2)
1.2 高光谱数据降维发展综述	(4)
1.3 高光谱数据分类发展综述	(7)
1.4 研究内容	(9)
1.5 框架安排	(10)
2 高光谱图像分类	(11)
2.1 高光谱图像分类的流程	(11)
2.1.1 数据预处理	(12)
2.1.2 标记训练样本	(12)
2.1.3 特征提取与特征选择(降维)	(12)
2.1.4 分类识别	(13)
2.1.5 精度评价	(13)
2.2 高光谱遥感图像分类方法	(14)
2.2.1 无监督分类方法	(14)
2.2.2 有监督分类方法	(15)
2.2.3 适合于高光谱图像分类的算法	(21)
3 高维数据降维技术	(23)
3.1 无监督降维技术	(24)

3.1.1	基于谱方法的无监督降维方法	(24)
3.1.2	基于隐变量模型的无监督降维方法	(29)
3.2	有监督降维技术	(33)
3.2.1	基于谱方法的有监督降维方法	(34)
3.2.2	基于隐变量模型的有监督降维方法	(37)
4	基于降维的高光谱图像分类	(41)
4.1	面向高光谱数据的降维模型	(41)
4.1.1	基于光谱特征的降维模型	(41)
4.1.2	基于空间特征的降维模型	(42)
4.1.3	两种降维策略	(42)
4.2	实验设计	(43)
4.2.1	实验设计	(43)
4.2.2	数据集介绍	(45)
4.3	实验结果与分析	(47)
4.3.1	基于光谱特征的降维与分类	(47)
4.3.2	基于光谱特征的降维与分类实验结果分析	(65)
4.3.3	基于空间特征的降维与分类	(67)
4.3.4	基于空间特征的降维与分类实验结果分析	(75)
5	基于有监督降维算法的高光谱图像分类	(76)
5.1	有监督隐线性高斯过程隐变量降维模型	(76)
5.1.1	模型表示	(76)
5.1.2	算法推导	(78)
5.1.3	算法分析	(82)
5.1.4	算法可视化验证	(83)
5.2	面向有监督降维的非参数指导的自编码器模型	(84)
5.2.1	模型表示	(84)
5.2.2	算法推导	(87)
5.2.3	算法分析	(88)
5.2.4	算法可视化验证	(88)

5.3 基于有监督降维模型的高光谱遥感图像分类..... (89)

5.3.1 实验设计..... (89)

5.3.2 数据集介绍..... (89)

5.3.3 基于光谱特征的降维与分类..... (93)

6 总结与展望..... (98)

6.1 总结..... (98)

6.2 展望..... (98)

主要参考文献..... (100)

致 谢 (105)

1 绪 论

1.1 高光谱遥感技术简介

1.1.1 遥感技术

兴起于 20 世纪 60 年代的遥感技术,基于不同的物体对光谱的吸收、反射和辐射的特性不同的由磁波理论来探测地物的性质,识别地表物质。它是一种检测技术,包括信息传输、接收设备、遥感图像处理平台等设备,主要利用航天飞行器上的传感器对地面各种景物进行检测、传递、记载和分析。遥感技术能从不同高度、大范围、快速和多谱段地进行探测,周期性、实时性地获取地物的大量信息(吴静,2010)。遥感技术在各领域中的应用大大推动了人类的进步,比如自然灾害的实时监测,植被研究中的作物类型识别、森林树种识别,对城镇道路进行识别,对地球上现有资源进行普查、军事动态监察和测绘等。

1.1.2 遥感数字图像处理

利用计算机技术,对遥感图像采取的一系列图像恢复、图像数据压缩、提取图像信息和图像增强等操作,来达到某种想要的预期效果,称为遥感数字图像处理。对图像处理过程中的错误数据、畸变、失真和噪声等一系列问题进行校正的操作,可以更好地恢复图像原始数据;对于海量的遥感图像数据,可以通过对数据进行压缩来减小数据量从而提高处理、输送和存储数据的效率;对数据进行彩色增强、边缘增强、比值计算等影像增强操作可以提高影像的目视质量;采用统计分析、集群分析等信息识别和分类技术来获取有用的遥感信息,从而完成信息提取。遥感数字图像处理在很多领域都被广泛使用,特别是遥感图像的地物分类。

1.1.3 高光谱遥感的发展

高光谱遥感(Hyperspectral Remote Sensing),也称为成像光谱遥感,是指利用很多很窄的电磁波段从感兴趣的物体获取许多连续并且窄的光谱曲线的一种探测技术(吴静,2010)。高光谱遥感的概念始于 20 世纪 80 年代中期,是目前科学家用于矿物鉴定和检测地面植被的一个相对较新的技术。高光谱遥感主要应用的仪器是成像光谱仪。成像光谱仪是引导遥感和地理信息的最新进展的发展方向,它能够获取许多波段狭窄且连续的影像,与普通遥感相比具有更高的分辨率,可以达到纳米级别。

通过高光谱遥感获取到的是能够充分反映目标的光谱特征的数据立方体。高光谱图像指

的是光谱分辨率在 10nm (即 $10^{-2}\lambda$) 数量级范围内的光谱图像(何同弟, 2014), 更严格意义上的高光谱图像指的是通过高光谱成像技术获得的图像, 相对于通过多光谱成像方法获取的多光谱图像而言, 高光谱图像在光谱的丰富性等方面具有优越性, 主要表现在高光谱图像不仅具有非常丰富光谱信息, 同时也拥有丰富的空间信息和辐射强度。在对地物进行分析时, 只需要对它相应的高光谱图像的光谱信息、空间信息以及辐射强度等信息进行分析, 就可以得到详细的地物特征。

高光谱遥感将图像和光谱信息相对融洽地结合在一起, 具有“图谱合一”的特性, 同时包含大量的数据集, 显示出巨大的信息潜力, 所以迫切需要新的处理方法。高光谱遥感技术的发展一直伴随着提高空间分辨率和时间分辨率两方面内容的发展而不断提高与进步(吴静, 2010), 因此如何提高高光谱图像的空间分辨率和光谱分辨率, 从如此大量的信息中高效地获取数据, 做到省时省力, 一直是当前科学工作者的研究热点, 以及一直积极解决的问题。

高光谱遥感技术无论在军事、民用和农业领域等方面都有很大的应用价值, 例如美国利用高光谱遥感来探测全球的臭氧层变化情况, 将高光谱遥感与激光雷达遥感相融合侦察云层的表面温度、湿度等一系列参数为气象部门服务(何同弟, 2014)。而高光谱遥感在国内的应用领域主要是赤潮监测、内陆水环境调查、矿产探测、区域性植物生态监测与精细农业等。虽然国内的高光谱遥感技术已取得一定的成果, 但是与国际上的先进技术相比, 国内的高光谱遥感技术仍然存在一定的差距, 有待进一步提高研究水平(吴静, 2010)。

高光谱遥感图像的应用实质上就是高光谱遥感图像分类, 如何对高光谱图像中每个像素的类别进行准确判别是这个问题的核心, 在解决这个问题的过程中有以下困难需要克服:

(1) 高光谱遥感图像的高光谱维度的特性会给数据存储和处理带来挑战, 不仅会增加计算复杂度, 而且也会使分类精度降低, 为此高维数据降维技术常常被应用于高光谱遥感图像的数据预处理, 通过特征提取或者特征选取, 能够有效地降低数据存储和后续的分类等模型的开销, 并提高模型分类精度。

(2) 对于高光谱遥感图像分类问题, 给定分类模型, 训练样本的选取也很重要, 选取地物类别的训练样本时, 训练样本的好坏对分类模型的泛化能力以及最后的分类精度有很大的影响。

(3) 不同的光谱曲线可能出现在同种地物上, 相同或者相似的光谱曲线可能出现在不同的物体, 这就是所谓的“异谱同物”或者“异物同谱”现象, 这给高光谱遥感图像分类带来很大的困扰。

(4) 成像光谱仪的高分辨率引入了较多的类别数量, 处理复杂度加大(高恒振, 2011)。

1.1.4 高光谱遥感图像数据的特点

高光谱图像通常是一个三维的数据立方体, 将空间信息在 $X-Y$ 平面上体现出来, 光谱信息在 Z 方向上体现出来, 即将遥感数据的图像维与光谱维信息有机结合(高恒振, 2011), 高光谱遥感图像中每一个像元都具有特殊的与它相对应的光谱曲线。高光谱遥感获得的每个谱段的信息, 可以用来研究其空间特征, 获得的光谱信息可以被用来作为研究物理特征的基础。高光谱数据具有大数据量、高维度、小样本分类和数据不确定性的特点, 因此必须针对高光谱数据的特性, 进行全面的分析, 才能更加高效地获取有用的信息。

高光谱图像的数据结构从空间特征的角度来说, 可以用如图 1-1 所示的一个“三维数据块”来形象生动地描述。

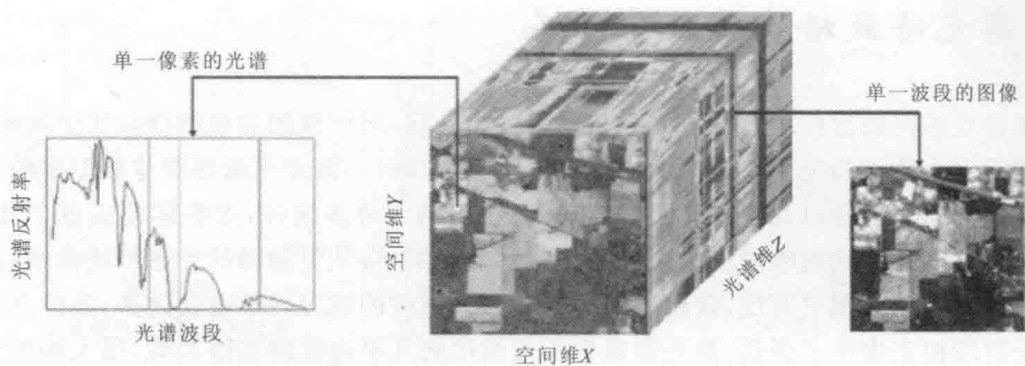


图 1-1 高光谱图像的结构

如图 1-1 所示,这是一个典型的高光谱图像的三维立体结构示意图,简单来说,高光谱图像的数据立方体就是在原本二维的平面信息的基础上,加上了光谱信息,使它从二维变为三维。高光谱图像的样本既包含图像本身具有的图像信息,同时也可以从中提取出丰富的光谱信息。通过对图像信息的观察和分析,我们可以得到关于图像样本的规模、形状、外观等外部的品质特征;而通过对高光谱图像的光谱信息的分析,可以得到关于样品的成分、内部的物理结构等内部品质特征。

高光谱遥感研究的光谱波长范围包括:可见光、近红外(VIS - NIR)、短波红外(SWIR)、中热红外(MIR)和热红外(MIR - TIR)波段,由于波长范围非常大,高光谱图像数据中,描述模型非常多。常见的 3 种描述模型为:图像模型[图 1-2(a)]、光谱曲线模型[图 1-2(b)]与特征向量模型[图 1-2(c)]。图像模型是单一/多个波段构成的图像,包含丰富的空间信息;光谱曲线模型是每一个像素对应的光谱波段,包含高维的、具有判别性的光谱信息;特征向量模型则包含经过降维模型处理后的特性,一个好的降维算法能够使得那些属于同一类的地物在特征向量空间相互聚集,而属于不同类别的地物则相互远离。正是高光谱图像具有如此丰富的特性,大量的降维算法被应用于高光谱图像的处理中,而且由于这 3 类特征模型带来的分类精度上的不同,也产生了大量的基于特征融合的分类技术。

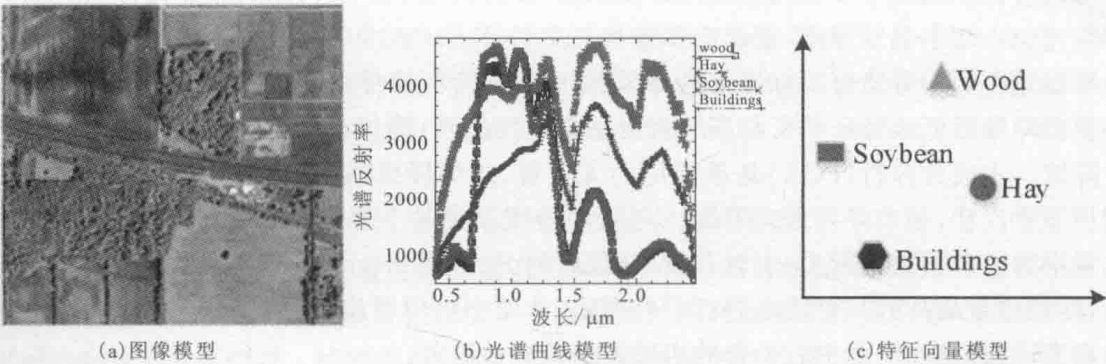


图 1-2 高光谱图像的 3 种描述模型

1.2 高光谱数据降维发展综述

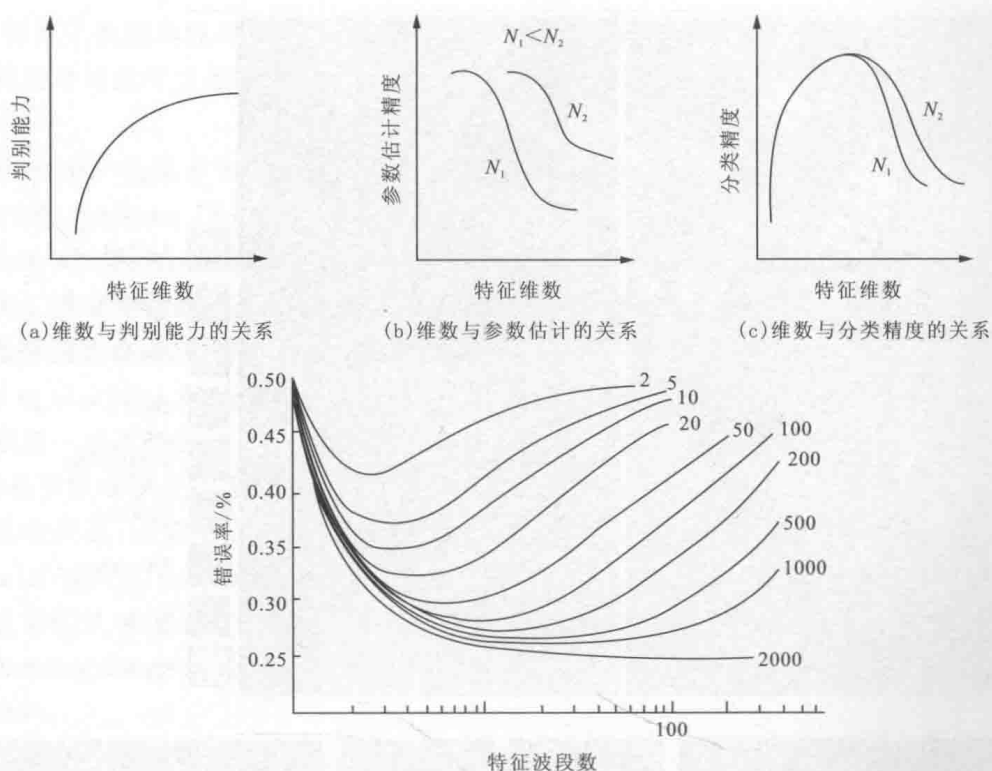
高光谱遥感将成像技术与细分光谱技术融合在一起,对所探测的地物获取图像的同时,每个空间像元也由几十乃至几百个窄波段构成(高恒振,2011),高光谱遥感图像数据中包含了丰富的空间信息和光谱信息,为精确的地物分类问题提供了更多途径,主要体现在以下几方面:高光谱分辨率能检测到地表物质中原来无法辨别的类别,得到更精确的光谱特征曲线;高光谱遥感器覆盖波长的范围更宽泛,获取的地表物质对电磁波的响应特征也就越多;波段众多为相互直接的波段校正提供了渠道;高光谱遥感图像给出的几乎连续的波谱曲线,很大程度地减少了分类过程中由于波谱数据信息量的不足造成的不确定现象,为分类提供了方便(高恒振,2011)。

虽然高光谱数据的丰富性给图像分类带来了很多契机,但是对于高维数据,当训练样本有限,光谱波段增加时,分类精度会随着特征维数的增加出现降低的现象,这就是所谓的“维数灾难”(Curse of Dimensionality),也被称作 Hughes 现象。对于“维数灾难”产生的原因,如图 1-3 所示,对于图 1-3(a)中的情况,当训练样本足够多,随着维数的递增,样本类别的可判别能力随之增加的;当在训练样本有限的情况下,如图 1-3(b)中, N_1 、 N_2 分别代表两组训练样本的数目,训练样本数目越多,测试样本的分类精度越大,针对 N_1 、 N_2 的精度曲线来看,当训练样本数目有限时,参数估计精度会随着特征维数的递增而逐渐下降;将图 1-3(a)、(b)合并最后得到维度与分类精度的关系,图 1-3(c)表示的是当训练样本有限的时候,分类精度会随着特征维数的递增而增长到一个最大值,然后将不再变大,接着可能会稳定一段时间后再随着特征维数的递增分类精度开始急剧下降。例如,图 1-3(d)所示的横坐标表示特征波段数,纵坐标表示分类的错误率,而图中间的数字表示训练样本的个数。对于一个维数为 100 维的高光谱图像,学习阶段训练样本数目少于 2000,如果采用传统的统计分类方法来对高光谱数据进行分类,波段的数目增加到一定程度反而会带来分类错误率的上升(苏红军等,2008)。

为了解决“维数灾难”问题,许多降维算法被应用到高光谱图像分类中,用于对高光谱图像的特征进行降维,从而提高分类精度。从特征表示的角度看,降维算法实际上等价于特征学习。降维模型包含特征提取和特征选取,前者进行特征融合后再降低特征维度,而后者仅通过从现有的特征里选择部分有代表性的特征来实现降维,由于本书关注的焦点是特征提取,常规的特性选取未在本书作介绍,请读者参考相关文献(Jia et al,2014)。

根据是否利用有监督的标签信息,降维模型可分为有监督降维、无监督降维以及半监督降维。根据降维后低维特征是否与高维观测呈现线性关系,降维模型也可被分为线性降维和非线性降维。主成分分析(PCA)是最常用的无监督、线性降维模型,在高光谱图像分类和处理中应用非常广泛,基于方差最大准则或者线性重建误差最小原则,PCA 能够尽可能地减少降维过程中的信息丢失。图 1-4 对 PCA 处理后的 Pavia University 高光谱数据集的前三个主成分,即信息量最高的 3 个维度进行了可视化,从图中可以看出,不同的主成分对不同的地物光谱有不同的反映,图像中右上角的房屋表现得非常显著。

著作“Remote Sensing Digital Image Analysis: An Introduction”(JA Richards et al, 1986)对基于 PCA 的高光谱图像处理进行了详细的介绍。



(d) 分类模型的错误率随着特征波段数的增加呈现先下降再升高的波动曲线

图 1-3 “维数灾难”形成原因

虽然 PCA 能够较为有效地从高光谱图像中提取有意义的特征,但受限于其线性表示结构,对于一些复杂的、低维特征与高维观测之间存在复杂的非线性关系的结构,PCA 就会因为无法建模而导致降维后特征的判别性降低,最终导致分类模型精度的下降。因此,一些基于 PCA 的扩展模型也被纷纷应用于高光谱遥感图像的处理中,例如 Maximum Noise Fraction (MNF) (Green et al, 1988), Kernel PCA (KPCA) (Mathieu et al, 2008), Sparse PCA (Chaib et al, 2015), Independent Component Analysis (ICA) (Falco et al, 2014) 等。此外,一些非线性的流型降维模型也被应用,例如 Locally Linear Embedding (LLE) (Yu et al, 2014), ISOMAP, Laplacian Eigenmaps (LE), Sparse Graph Embedding (SGE) (He et al, 2016) 等,由于这些流型降维模型通常都面临样本外延计算问题(out of sample problem),一些通过近似线性化流型模型的降维技术也被提出,例如 Locality Preserving Projections (LPP) 和 Neighborhood Preserving Embedding (NPE) 等。

虽然这些模型能够线性或非线性地从高光谱图像中提取光谱/空间特征,但由于这些模型均属于无监督降维模型,也意味着它们不能有效地利用额外的标签信息,这对于包含标签的训练数据来说无疑是巨大的信息浪费,为此,一些在降维的同时也考虑标签信息的降维模型被提出,可被统称为有监督降维。线性判别分析(LDA)是最典型的有监督、线性降维模型,其模型基于最大化样本类间距离和最小化样本类内距离的原则进行降维,通常能够得到比 PCA 更加有判别性的特征。为了克服有监督的线性降维模型表达能力弱的问题,许多研究者也提出了非线性的有监督降维算法,例如基于 LDA 的扩展模型,Regularized LDA (Bandos et al, 2009) 等,

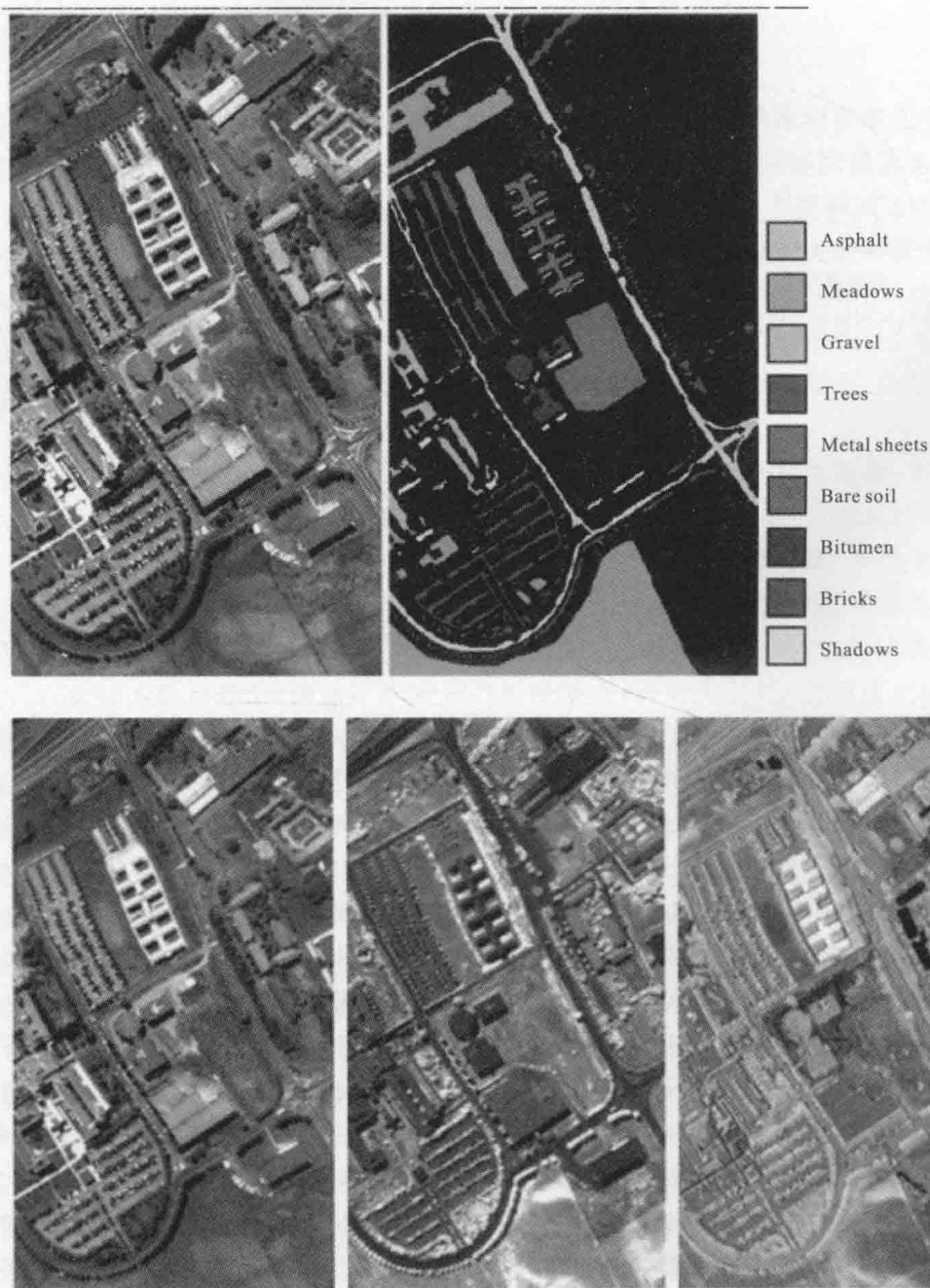


图 1-4 上排两个图像显示的是 Pavia University 高光谱数据集的三通道假彩图,参考图像及颜色标签;下排三个图像是 PCA 对 Pavia University 高光谱数据集降维后的可视化结果,从左到右依次为对第一、第二、第三个主成分进行可视化的图示

(Ghamisi et al, 2015)

Discriminant Analysis Feature Extraction (DAFE) (Landgrebe et al, 2005), Nonparametric Weighted Feature Extraction (NWFE) (Kuo et al, 2005), Kernel NWFE (Kuo et al, 2009) 等。此外,也有一些研究者尝试将无监督的非线性降维模型扩展到有监督降维模型并应用于高光谱图像的处理中,例如 Supervised Locally Linear Embedding (SLLE) (Cao et al, 2016),

Discriminant Locality Preserving Projection (DLPP)(Feng et al, 2015)等。2013年 Xiuping Jia 等从特征工程的角度对一些降维模型在高光谱图像分类任务中的性能做了较全面的综述,有兴趣的读者可参考文献《Feature Mining for Hyperspectral Image Classification》(Jia et al, 2013)。

此外,还有一些基于光谱和空间特征融合的降维模型,例如 Spectral - Spatial Kernel - Based Multivariate Analysis (KMVSSA)(Borhani et al, 2015), Spectral - Spatial LDA(Yuan et al, 2014), Spatial - Spectral Similarity Measure based DR(Huang et al, 2015)等。

作为三维立方体的高光谱图像,也可很自然地与张量(Tensor)联系起来,因此一些基于张量的降维模型也在高光谱降维中得到应用,例如 Lower Rank Tensor Approximation(LRTA)(Renard et al, 2009), Group - based Tensor DR(An et al, 2016)等。

如果进一步从特征表示的角度去看降维,当前随着深度学习(Deep Learning, DL)的兴起,一些基于深度学习的无监督特征学习模型也可被看作是降维模型。近两年在高光谱图像分析中也逐渐兴起,例如 Unsupervised Feature Learning based on Spectral Clustering (UFL - SC)(Hu et al, 2015), Unsupervised Multilayer Feature Learning(Li et al, 2016)等。2016年武汉大学张良培团队对基于深度学习的遥感图像处理做了很好的综述,有兴趣的读者请参考文献《Deep learning for remote sensing data: a technical tutorial on the state of the art》(Zhang et al, 2016)。

虽然这些降维模型都得到了广泛应用,但目前仍缺少一些基于这些降维模型的高光谱图像分类的综述文献,虽然已经进行了一些相关的工作,例如文献(Chen et al, 2015; Hasanlou et al, 2012; 陈宏达, 2013; Jia et al, 2013),但都没有全面覆盖较为常用的降维模型。因此,本书尝试对这些降维模型在高光谱分类任务中进行逐一的实验对比,以实验来验证这些降维模型在高光谱遥感影像处理中的有效性。此外,本书还进一步介绍了近期作者提出的两类有监督的降维算法并应用于高光谱遥感图像处理中。

1.3 高光谱数据分类发展综述

高光谱遥感图像光谱的数量太过庞大,有的时候会有上百个连续的光谱,也就意味着高光谱数据的维度非常高。因此,传统的分类方法已经无法很好地实现对这些高维度的数据的分类处理了,这种算法已经无法在计算速度、准确性等方面达到要求。因此,大量的降维算法被应用到高光谱图像的预处理中,通过降维来实现特征学习。除此之外,鉴于高光谱图像特殊的图谱合一的特点,一些区别于传统图像分类的方法也在高光谱图像分类中得到关注和应用。

伴随着广大研究者不断的研究和探索,这方面的技术已经有了长足的发展,目前已经形成了许多对高光谱图像进行分类的算法。这些算法主要可以分为两个大的类别:①基于地物的光谱特征的分类方法;②基于数据的统计特征的分类方法(高恒振, 2011)。

第一种所说的通过对地物的光谱特征研究进行分类的方法,主要是通过对测地物的光谱曲线进行分析和研究。从光谱曲线中,我们可以得到许多有用的信息,例如地物的光学物理性质等。一般情况下,地物的光谱特征与其本身的种类和所处的条件有关,利用这种原理,可以对地物通过光谱信息进行分类。然后,利用各种匹配算法来进行比较和匹配,来区分图像中不同地物类型,达到分类的目的。常见的分类方法有:欧式距离方法(EMD)、光谱角匹配方法

(SAM)、光谱信息散度(SID)和光谱二值编码(BC)等(高恒振,2011)。

第二种所说的基于数据的特征统计的分类方法,根据在分类过程中是否需要人为指定地物样本的类别,可分为无监督分类和有监督分类两种方法(高恒振,2011)。

无监督分类方法不需要事先对遥感图像地物的属性具备先验知识,可以直接对原始高光谱遥感图像数据在统计上的差别进行“盲目分类”,事后再对分出的各类地物属性进行确认。这种分类方式,没有使用任何关于数据的先验信息,所以分类的效果不是很好,分类精度也往往有所欠缺,但容易实现,也是常用的分类方法之一。常用的无监督分类算法有:K-均值(K-Means)算法、迭代自组织数据分析(Iterative Self-Organizing Data Analysis Technique, ISODATA)、多目标遗传聚类方法(Multi-Objective Genetic Algorithm, MOGA)等。此外,无监督方法还可利用神经网络、随机场等方式实现,如自组织模糊多层神经网络分类器和泛化马尔可夫随机场模型等。无监督分类能够减少用户输入的数据量,但必须要知道类别个数,且由于得到的类别没有标签,还需要人工参与判别,因此这类无监督的分类方法效果通常都不太理想。

与无监督分类相对应的是有监督分类方法,该方法以统计学习理论为基础,依据典型样本训练方法进行分类,即根据已知训练区提供的样本,通过选择特征参数,求出特征参数作为决策规则,然后建立判别函数的图像分类方法,是模式识别的一种。监督分类要求训练区域具有典型性和代表性。判别准则若满足分类精度要求,则此准则成立;反之,需重新建立分类的决策规则,直至满足分类精度要求为止。分类器可以通过概率模型,经验风险最小化原则,结构风险最小化原则等形式来获得(高恒振,2011),常用分类器有:最大似然分类法(也称贝叶斯分类法)、最小距离分类法、马氏距离分类法、平行六面体分类法、K近邻分类法(K-Nearest Neighborhood, KNN)、随机森林分类法(Random Forest, RF)和支持向量机(Support Vector Machine, SVM)(Melgani et al, 2004)分类法和高斯过程分类法(Gaussian Process Classification, GPC)(Bazi, 2010)等。监督分类可以充分利用分类地区的先验知识,预先确定分类的类别,还可以通过对分类器反复检验训练样本来提高分类精度,避免严重错误。虽然需要花费人力、时间进行训练样本的选取与评估,但是分类精度通常比较高。

邻近算法(KNN)是一种比较简单的分类算法,主要依据最邻近的一个或者几个样本的类别来决定待分类样本所属的类别,这种算法实现简单,易于实现,无需训练。如果样本的特征判别性较好,KNN能很好地对高光谱数据进行分类,但是通常高判别性的特征是很难获取的,基于距离匹配的KNN会因为特征的不显著而导致分类精度的下降。

除了KNN,神经网络(NN)和支持向量机(SVM)也是比较典型的两个常用于高光谱图像分类的分类算法。神经网络能够自动学习和训练,有效地解决了很多的非线性问题,但是神经网络在处理高光谱数据时也存在一些不足之处:训练时间和训练效果会受到训练样本的影响,而且,神经网络对于有限的训练样本的预测会出现一定的误差,用复杂的预测函数对样本进行学习的效果通常不如用相对简单的预测函数。为此,在统计学的基础上发展起来的(高恒振, 2011)支持向量机(SVM)被用于替换神经网络分类器来对高光谱图像进行分类。SVM较高的模型泛化能力能够提高小样本、高维度高光谱样本数据的分类精度,因此很快地被成功地用来对高光谱图像进行分类。此外SVM也能够很好地处理高光谱数据、高维特征空间和大数据量的问题。但是基于SVM的分类方法中还存在着许多不足之处:SVM是浅层结构的分类模型,随着空间和光谱的变化,不能在训练过程中充分地考虑高光谱数据内在的不变特性,且

与分类标签之间存在复杂的非线性关系,都导致 SVM 分类方法在分类精度上遇到了天花板,无法进一步提高分类精度。

相比而言,基于深度学习的分类器能够实现高光谱图像的特征学习和分类,目前已取得较多的成果(Chen et al,2014;Hu et al,2015;Zou et al,2015;Chen et al,2015)。基于深度学习的分类方法能够自动地学习非线性的层次特征,高效地拟合分类标签与高光谱图像数据特征之间存在着非线性关系,有效地提高了分类精度。近几年,随着深度学习的兴起,大量的深度学习模型被应用到高光谱图像分类问题中,取得了一些令人振奋的效果。但是是否基于深度学习的分类模型在高光谱图像分类问题中一定会优于传统的浅层模型(SVM 等)呢?一些研究者也作了一些对比分析,比如 Peng Liu 等就对比了 SVM 和深度模型 Stacked Auto-Encoder (SAE)(Liu et al,2016),结果发现 SVM 在一些任务中会更加优异,而且训练时间和参数调整都更方便,但我们应该看到,深度学习模型的确也在高光谱分类问题中崭露头角,已逐渐得到广大研究者的关注和认可(Zhang et al,2016)。

1.4 研究内容

高光谱遥感可以在很多相邻的比较狭窄的光谱带收集地物的信息,成像光谱仪能够获取的光谱波段一般都在 10nm 之内,因此能以足够的高分辨率来区分地表物质的光谱特征,从而检测、识别地表物质(高恒振,2011)。但是,由于高光谱遥感数据自身的高维度、高数据量的特点,给数据分析工作者带来了很大的问题,根据数据的特点采取有效的方法来提高数据分析的效率很重要。近几年来,科学研究者根据高光谱数据的特点不断地创造出新的降维算法,以及对降维算法进行改进、整合,数据降维算法层出不穷,每个算法都有自身的优点和缺点,围绕这些已有的降维算法,本书主要介绍了以下工作:

(1)将目前机器学习领域中常见的一些高维数据降维算法作为特征提取模型,应用于高光谱遥感图像分类中并进行对比分析,以实验验证的策略测试出那些适合于高光谱遥感图像分类任务的降维模型。

(2)基于选择出的这些降维算法,选择印第安纳农田数据(Indian pines)、帕维亚中心(Pavia Center)和帕维亚大学(Pavia University)3 组数据分别进行降维,并对降维后的特征基于 KNN 进行分类,同时本书也分别基于数据之间的谱间和空间相关性进行实验,并且采用两种降维策略进行降维。

(3)对降维后的特征使用 KNN 进行分类并进行参数敏感度实验和分析。当数据降维的维度值确定时,针对不同算法,绘制出 k 近邻分类算法中的参数 k 对于测试样本总体分类精度(OA)影响的曲线图;当参数 k 一定时,针对不同的算法,选择恰当的维度值,进行对比实验,绘制出维度参数(dim)对于数据总体分类精度(OA)影响的曲线图,同时对分类结果进行了可视化对比,对基于降维模型的高光谱遥感图像分类模型进行时间复杂度分析和实验对比,最终根据实验结果,得出结论,完成对各个降维算法特点的综述。

(4)基于对已有降维算法的实验对比和分析,介绍两类最新提出的有监督降维模型,并将其应用于高光谱遥感图像分类问题中。实验结果显示,这两类模型在高光谱遥感图像分类任务中要显著地优于已有的其他降维模型。

1.5 框架安排

第2章,主要阐述高光谱遥感图像分类的相关内容,包括图像分类的一些方法,主要是监督分类和非监督分类两种分类方法,以及两种方法自身的特点;着重介绍高光谱遥感图像分类的流程,对每个流程进行分析和综述。

第3章,主要针对高光谱图像分类中可能存在的“维数灾难”的问题,综述一些面向高光谱图像分类的降维算法,主要是介绍和对比线性、非线性降维算法。

第4章,主要根据高光谱数据的两种相关性进行对比实验,针对一些常用的降维算法,采用两种降维策略,对3组实验数据分别进行降维、分类、分析过程,最后得出结论。

第5章,介绍两种新的有监督降维模型并将其应用于高光谱图像降维和分类,7组最优维度和训练集样本个数的实验对这两种新的降维模型进行了验证。

第6章,主要是总结与展望,总结本书的主要内容,展望以后如何进行高光谱遥感图像分类的研究和相关的计划。