doi:10.3788/gzxb20144306.0630002

高光谱数据基于流形的半监督特征选择

魏峰,何明一,申志明,李旭
(西北工业大学 陕西省信息获取与处理重点实验室, 西安 710129)

摘 要:传统的高光谱数据特征选择方法分为监督和无监督模式,然而在高光谱数据实际处理中,大量无标记和少量有标记数据并存.此外,传统方法忽视了真实数据嵌入在高维空间中的流形结构.本文提出一种基于流形的半监督特征选择算法,定义一个合理的特征评判准则,考虑标记样本的先验信息以及高维数据局部和非局部结构的不变特性.通过对标记数据类间方差和类内方差的极大化和极小化,优化数据的判别结构;同时通过构建局部 graph 和非局部 graph,挖掘高维数据的流形结构.然后,选择一组有效的特征子集,实现高维数据的特征选择.最后,通过对特征选择后的真实高光谱数据进行分类实验,结果显示本文方法可以很好地对高光谱数据实现降维并且保留数据的主要结构.

关键词:高光谱;遥感;半监督;特征;监督学习;光谱分辨率;分类;谱理论
中图分类号:TP701 **文献标识码**:A **文章编号**:1004-4213(2014)06-0630002-5

Manifold based Semi-supervised Feature Selection for Hyperspectral Data

WEI Feng, HE Ming-yi, SHEN Zhi-ming, LI Xu
(*Shaanxi Key Lab of Information Acquisition and Processing, Northwestern Polytechnical University, Xi'an* 710129, *China*)

Abstract: The traditional Feature Selection methods of hyperspectral data include supervised and unsupervised modes, is not efficient for the real processing system in which abundant unlabeled and few labeled data co-exist. Additionally, most of existing feature selection methods ignore that real data has a manifold structure which embedded into the high dimensional space. In order to solve these problems, a Manifold based Semi-supervised Feature Selection (MSFS) algorithm was proposed. Considering the prior information of labeled data with the local and non-local invariance of the whole data, the discriminate structure is optimized through simultaneously maximizing between-class and minimizing within-class variances. Meanwhile, the manifold structure is exploited from constructing local and non-local graphs for the whole data. Then, the efficient features is selected by defining an appropriate evaluation criterion. Finally, through performing the classification experiment on the selected features of real hyperspectral data, it demonstated that our method is able to retain the main structure of data after dimensionality reduction well.

Key words: Hyperspectral; Remote Sensing; Semi-upervised; Feature; Supervised learning; Spectral resolution; Classification; Graph theory
OCIS Codes: 300.0300; 280.1310; 070.4560; 070.5010

0 引言

高光谱遥感技术是基于电磁波谱理论,在可见光、近红外、中红外和热红外波段的范围内,采集很窄的连续光谱影像数据[1].高光谱遥感在为观测目标提供细致丰富的特性描述的同时,其光谱波段间的大量冗余信息也给数据的实际处理带来了困难[2].由于过多的特征(光谱波段)和不足的训练样本所带来的维度灾难问题,在对数据分类和识别处理之前,需要进行特征约简.

基金项目:国家自然科学基金(Nos. 61171154, 61301195)资助
第一作者(通讯作者):魏峰(1987—),男,博士,主要研究方向为高光谱遥感数据处理. Email:weifengg@163.com
导 师:何明一(1958—),男,教授,博士,主要研究方向为高光谱遥感信息获取、处理与传输. Email:myhe@nwpu.edu.cn
收稿日期:2014-01-23;**录用日期**:2014-05-07

特征提取被广泛应用到高光谱数据特征约简的过程中，典型的方法有：主成分分析(Principal Component Analysis, PCA)[3]，Fisher 判别分析(Fisher Discriminant Analysis, FDA)[4]，局部保留投影(Locality Preserving Projection, LPP)[5]等，它们把原始数据从高维空间变换到新的低维空间，同时能够保留数据的主要信息.然而，这样的空间变换改变了原来光谱波段的物理意义.与特征提取相比，特征选择在寻找一组特征子集，构成一个有利于聚类、分类或检索的低维子空间时，可以很好地保留原来的光谱波段，不会破坏数据的物理意义.

特征选择方法可以分为监督和无监督两种模式.随着高光谱遥感数据获取、传输技术的发展，采集大量无类别标记的数据比较容易，而为其提供标记信息则相对困难[6-7].所以，在高光谱数据处理的实际应用过程中，大多数时候是大量无标记数据和少量有标记数据并存的情况.因此，如何同时有效地利用这两种数据，成为当前高光谱数据处理倍受关注的重要问题之一[8-9].本文基于半监督学习准则，利用少量标记和大量无标记的样本，考虑高维数据特征选择的特点，寻找一组优化的特征子集.

文献[10]表明，很多数据都是采样自一个嵌入在高维空间中的非线性低维流形结构.然而，大多数特征选择方法并没有挖掘这种蕴含在数据内部的几何结构.Laplacian Score[11]是一种无监督特征选择算法，通过计算每一个特征对局部几何结构保留的能力来评估该特征的性能，然而没有充分利用数据宝贵的标记信息.

本文同时考虑高光谱数据局部和非局部流形结构的不变特性[12]以及标记样本的先验信息，提出基于流形的半监督特征选择(Manifold based Semi-supervised Feature Selection, MSFS)方法，发掘数据空间的几何和判别结构.该方法具有如下特点：1)针对高光谱数据的后续处理，采用一种具有判别性质的思想设计，能够使得特征约简之后的高光谱数据几何结构更加优化.2)同时利用少量的标记样本和大量的无标记的样本，获得的模型具有更强的泛化性能.3)与特征提取相比，该方法不需要解决复杂的特征值问题，具有更高的处理效率和更强的实用性.4)该方法能够很好保留原来的光谱波段，从而不会对数据的物理意义进行破坏，有利于高光谱数据的进一步处理和信息解译.

1 Fisher Score

Fisher Score 是一种典型的监督特征选择算法，给定一组高维数据及它们的类别标记信息，(x_i, y_i)，$y_i \in (1,\cdots,c)$，$i=1,\cdots,n$，c 是类别的个数.在第 r 个特征上，定义 n_i 为第 i 个类别样本点的个数，μ_i 和 σ_i^2 为第 i 类的均值和方差，其中 $i=1,\cdots,c$.定义 μ 和 σ^2 为所有样本的均值和方差.则 Fisher Score 为

$$F_r=\frac{\sum_{i=1}^{c} n_i(\mu_i-\mu)^2}{\sum_{i=1}^{c} n_i\sigma_i^2}=\frac{V_b}{V_w}$$
$$V_b=\sum_{i=1}^{c} n_i(\mu_i-\mu)^2$$
$$V_w=\sum_{i=1}^{c} n_i\sigma_i^2 \tag{1}$$

式中 V_b 和 V_w 分别定义为第 r 个特征上标记样本的类间方差和类内方差.通过计算每个特征对应的分值，可以选择出一组具有高分值的特征子集.

这种相似的准则也能够推广到特征提取方法中，即为 FDA 算法[4]，设 $\boldsymbol{a}$ 为投影向量，则 FDA 的目标函数为

$$\max_{a}\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{a}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{a}}$$
$$\boldsymbol{S}_b=\sum_{i=1}^{c} n_i(\mu^i-\mu)(\mu^i-\mu)^{\mathrm{T}}$$
$$\boldsymbol{S}_w=\sum_{i=1}^{c}\left(\sum_{j=1}^{n_i}(x_j^i-\mu^i)(x_j^i-\mu^i)^{\mathrm{T}}\right) \tag{2}$$

式中 μ 为总体样本均值，μ^i 为第 i 类样本的均值，x_j^i 为第 i 类中的第 j 个样本，$\boldsymbol{S}_w$ 为类内离散度矩阵，$\boldsymbol{S}_b$ 为类间离散度矩阵.通过求解下面的广义特征值问题可以得到优化的投影向量，即

$$\boldsymbol{S}_b\boldsymbol{a}=\lambda\boldsymbol{S}_w\boldsymbol{a} \tag{3}$$

2 MSFS 方法

MSFS 算法对每一个特征进行评估：通过极大化标记数据的类间方差和极小化标记数据的类内方差，优化数据的判别结构；同时通过构建局部和非局部 graph，挖掘所有数据(包括标记和非标记)的流形结构.通过定义一个合理的特征评判准则，获得一组优化的特征子集.具体来讲，在数据分布的低维特征子空间中，对于标记样本，希望使得同一类别的数据分布更加集中，不同类别的数据分布更加分散；对于整体样本，原始高维空间中距离相近的两个数据点 x_i 和 x_j，其对应的低维表达 y_i 和 y_j 仍然很近，反之，如果距离相远的点，其对应的低维表达变得更远.即高维数据经过特征约简后在低维空间中其几何结构更加优化.MSFS 综合考虑了特征选择、流形结构和半监督学习的共同特点.

流形(manifold)是对一般几何对象的总称，可以近似视为一些相对简单空间的结构.真实的高光谱数据是一组维度很高的数据，可以将其看作嵌入在高维空间中的一个低维流形[13].

构建局部和非局部 graph 结构：近年来的 spectral graph 理论和流形学习理论表明[14-15]，对于分布在原始空间中的一组数据，其局部几何结构能够通过一个最近邻 graph 结构来有效地近似，其中的每个节点 i 对应相应的数据点 x_i.每一个 x_i 都和 K 个最近的点相连，

这样，局部 graph 结构可以通过定义权值矩阵 $\boldsymbol{G}$ 来表示为

$$\boldsymbol{G}_{ij}=\begin{cases}1 & \text{if } x_j\in N_i \text{ or if } x_i\in N_j\\ 0 & \text{other}\end{cases} \tag{4}$$

式中，N_i 和 N_j 分别为 x_i 和 x_j 的 K 个最近邻点集. 明显地，权值矩阵 $\boldsymbol{G}$ 是一个对称矩阵，其每一个值 $\boldsymbol{G}_{ij}$ 可以测量两个样本点 x_i 和 x_j 的相近程度.

对比局部 graph 权值矩阵的特点，定义非局部 graph 权值矩阵为

$$\boldsymbol{G}'_{ij}=\begin{cases}1 & \text{if } x_j\in N'_i \text{ or if } x_i\in N'_j\\ 0 & \text{other}\end{cases} \tag{5}$$

式中，N'_i 和 N'_j 分别为 x_i 和 x_j 的 K 个最远点集. 定义 D 和 D' 为对角线矩阵，其中每一个值分别为 $\boldsymbol{G}$ 和 $\boldsymbol{G}'$ 的列和（或者行和，因为 $\boldsymbol{G}$ 和 $\boldsymbol{G}'$ 都是对称矩阵），即

$$D_{ii}=\sum_j \boldsymbol{G}_{ij},\ D'_{ii}=\sum_j \boldsymbol{G}'_{ij} \tag{6}$$

则 $\boldsymbol{G}$ and $\boldsymbol{G}'$ 的 graph Laplacians 可以由式(7)计算.

$$L=D-\boldsymbol{G},\ L'=D'-\boldsymbol{G}' \tag{7}$$

在应用标记样本的先验信息时，Fisher Score 的一个主要限制是在一组高维数据中，其假设每类样本均为高斯分布. 然而，实际高光谱数据分布具有很强的多峰特性. 局部 Fisher 判别分析[16]是一种新颖的特征提取方法，可以有效地处理具有多峰非高斯分布的数据，从而很好地保留数据的潜在结构. 基于这种思想，对标记样本的类间和类内方差进行修改，首先定义两个权值矩阵 $\boldsymbol{W}_b$ 和 $\boldsymbol{W}_w$ 为

$$\boldsymbol{W}_{b,ij}=\begin{cases}A_{ij}(1/n-1/n_{l(x_i)}) & \text{if } l(x_i)=l(x_j)\\ 1/n & \text{if } l(x_i)\neq l(x_j)\end{cases}$$

$$\boldsymbol{W}_{w,ij}=\begin{cases}A_{ij}/n_{l(x_i)} & \text{if } l(x_i)=l(x_j)\\ 0 & \text{if } l(x_i)\neq l(x_j)\end{cases} \tag{8}$$

式中 n 为标记样本的个数，$l(x_i)$ 表示 x_i 的类别标记，A_{ij} 为 x_i 和 x_j 基于局部 scaling heuristic 的连接值，文献[16]给出了具体的定义. 设 V'_{b} 和 V'_{w} 为修改后第 r 个特征上标记样本的类间方差和类内方差，x_i^r 和 x_j^r 分别为 x_i 和 x_j 在第 r 个光谱特征上的值，则

$$V'_{\mathrm{b}}=\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{W}_{b,ij}(x_i^r-x_j^r)^2$$

$$V'_{\mathrm{w}}=\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{W}_{w,ij}(x_i^r-x_j^r)^2 \tag{9}$$

本文希望找到一组特征子集，使得原来高光谱数据在这组特征子集组成的低维空间中的分布更加优化. 即同一类别的样本更加集中，不同类别的样本间距变大，同时所有样本的近邻结构和非近邻结构在新的低维空间中得到了很好的保留.

因此，基于这样的假设，定义 S_r 为第 r 个特征的 MSFS Score，可以通过式(10)选择一组具有判别性质的特征.

$$S_r=\lambda\frac{V'_{\mathrm{b}}}{V'_{\mathrm{w}}}+\beta\frac{f_r^{\mathrm{T}}L'f_r}{f_r^{\mathrm{T}}Lf} \tag{10}$$

可以看出 MSFS 同时利用了标记样本和无标记样本，获得的模型具有更强的泛化性能. 此外，与常用的特征提取方法相比，其不用解决复杂的特征值问题；同时能够很好地保留原始的光谱波段，不改变高光谱数据特征的物理意义.

3 MSFS 方法的证明

通过对保留高光谱数据流形结构和判别结构的评估可以衡量对应特征的重要性.

定义 $f_r=(f_{r1},f_{r2},\cdots,f_{rN})^{\mathrm{T}}$，其中 N 为总体样本的个数，f_{ri} 为第 r 个特征上的第 i 个样本，因此，一个用于选择好的特征的合理准则即为

$$\min_r\sum_{ij}(f_{ri}-f_{rj})^2G_{ij}=\min_r\sum_{ij}(f_{ri}-f_{rj})^2G_{ij}=$$
$$\min_r(f_r^{\mathrm{T}}Df_r-f_r^{\mathrm{T}}Gf_r)=\min_r f_r^{\mathrm{T}}Lf_r \tag{11}$$

通过最小化该目标函数，使得 G_{ij} 越大，$|f_{ri}-f_{rj}|$ 越小，从而很好的保留数据的局部 graph 结构.

对于非局部 graph 结构，一个好的特征选择的合理准则如式(12)，从而很好地保留数据的非局部 graph 结构.

$$\max_r\sum_{ij}(f_{ri}-f_{rj})^2G'_{ij}=\max_r\sum_{ij}(f_{ri}-f_{rj})^2G'_{ij}=$$
$$\max_r(f_r^{\mathrm{T}}D'f_r-f_r^{\mathrm{T}}G'f_r)=\max_r f_r^{\mathrm{T}}L'f_r \tag{12}$$

对于标记样本，希望同一类别更加集中，不同类别间距变大，因此选择特征的合理准则就是具有较大值 F'_r，即

$$F'_r=\frac{V'_{\mathrm{b}}}{V'_{\mathrm{w}}} \tag{13}$$

从而很好地优化了标记数据的类内和类间结构，同时结合目标函数式(11)和式(12)，对于每个特征，可以对其进行评估，即

$$S_r=\lambda\frac{V'_{\mathrm{b}}}{V'_{\mathrm{w}}}+\beta\frac{f_r^{\mathrm{T}}L'f_r}{f_r^{\mathrm{T}}Lf} \tag{14}$$

λ 和 β 为权值系数，用来调节样本的流形结构和类间、类内方差分别对 MSFS 算法产生的影响，取值 0 到 1 之间. 通过选择出一组可以同时极大化 V'_{b} 和 Φ_2，极小化 V'_{w} 和 Φ_1 的特征子集，既可以使标记样本的类间方差和类内方差得到优化，也可以确保高维空间中数据的整体几何结构得到优化. 结合这种基于判别性质和的流形结构的特征选择思想，即为本文方法.

4 实验结果与分析

本文采用 224 波段 AVIRIS 高光谱遥感数据进行试验，对不同的特征选择算法进行比较. 该数据取自 1998 年 10 月的美国加利福尼亚州（California）的萨利纳斯谷（Salinas Valley）测试区农场和森林交杂的地

带，数据的光谱波段范围为 0.4～2.5 μm 之间，其空间分辨率达到 3.7 m. 实验中去除受空气中水汽和臭氧影响比较严重的波段，保留 200 个信噪比比较高的光谱波段构成数据集合(图 1).

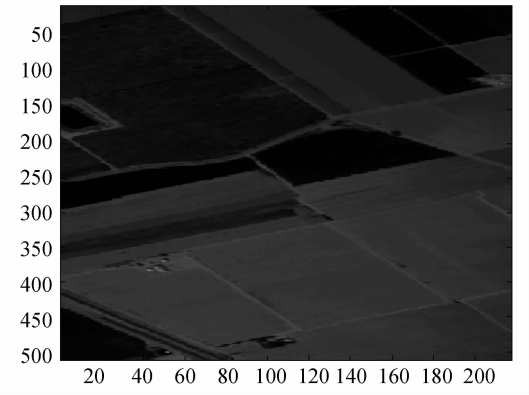

图 1 原始 Salinas Valley 光谱数据

Fig. 1 Original Salinas Valley spectral data

该数据集共有 512×217 个数据点，分别归属 16 类地物，地物分类真实信息如图 2. 试验选取具有代表性的8个类别作为样本测试集，所选取的样本集如表1.

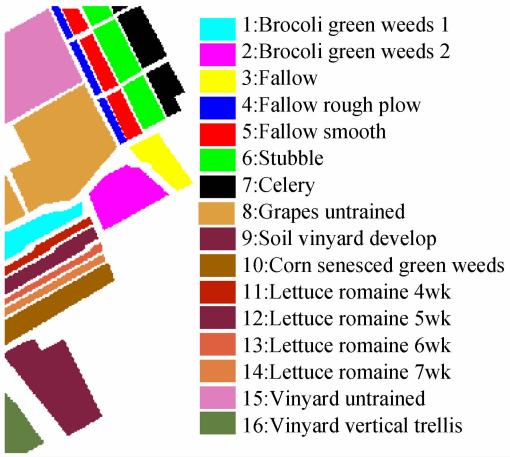

图 2 Salinas Valley 地物分类信息

Fig. 2 Salinas Valley ground truth

表 1 Salinas Valley 样本集

Table 1 Salinas Valley sample set

Category	Sample size
1	2 099
3	1 976
4	1 394
11	1 068
12	1 927
13	916
14	1 070
16	1 807

4.1 近邻分类误差

采用最近邻分类误差准则对多种不同特征选择方法的性能进行评估，该准则广泛地应用在各种特征约简算法的分析中. 其基本原理为：在对应的特征空间中，对每一个样本数据 x_i，寻找距离其最近的样本点 x_i'，假设 $l(x_i)$ 为 x_i 的类别标记，则最近邻分类误差率 ER 定义为

$$\mathrm{ER}=1-\frac{1}{N}\sum_{i=1}^{N}\delta(l(x_i),l(x_i')) \tag{15}$$

式中 N 为所有样本点的个数，当 $a=b$ 时，$\delta(a,b)=1$，否则为 0.

4.2 实验参量选择

在 MSFS 算法实验过程中，随机选取表 1 中每类样本的 30%作为标记样本，来衡量每个特征上标记样本的类间方差和类内方差，然后通过计算每个特征的 MSFS Score，选取最大的一组作为优化的特征子集.

该算法需要确定每个样本数据点局部邻域的样本个数，即任意样本点 x_i 的局部邻域 N_i 内样本点的个数 K. 在设置不同 K 值的情况下，计算波段选择后的数据的最近邻分类误差率 ER. 实验结果如图 3 所示，可以看出：当 K 值较小和较大时，分类误差有所上升.

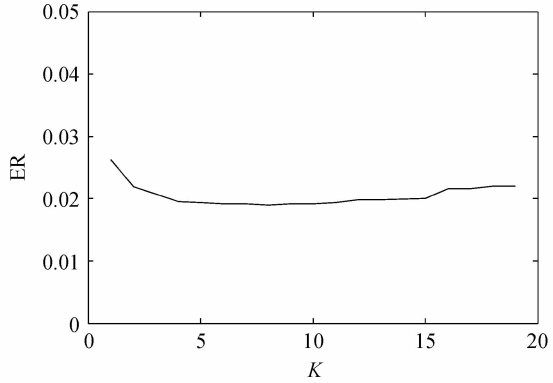

图 3 K 值对 MSFS 方法的影响

Fig. 3 Classification accuracy of MSFS under different K

4.3 MSFS 算法的性能比较

为了验证本文算法的有效性，实验在选择出不同数目特征的情况下验证 MSFS 算法的分类准确度，并将其与两种常见的特征选择算法，Fisher Score 和 Maximum Variance 进行性能比较，同时保留数据的所有特征作为 baseline，直接进行最近邻分类.

1)Fisher Score，旨在寻找一组优化的特征子集，使得类间方差变大，类内方差变小.

2)Maximum Variance，最大化数据方差.

3)Baseline，保留所有的特征.

实验结果如图 4.

从分类结果可以看出，在选择出的特征个数比较少时，各种特征选择算法的分类效果均不理想. 随着特征个数的增加，各种特征选择算法的性能不断增强，分类误差有了显著降低. 与 Fisher Score 和 Maximum Variance 相比，MSFS 具有如下显著的优势：1)在高光谱数据实际处理缺少充足标记样本的情况下，可以同时利用少量标记样本和大量无标记样本进行学习，使

得该算法所构建的模型具有更强的泛化性能;2)算法针对高光谱数据的后续处理,采用一种具有判别性质的思想设计,能够使得高光谱数据其局部机构和全局结构更加优化.

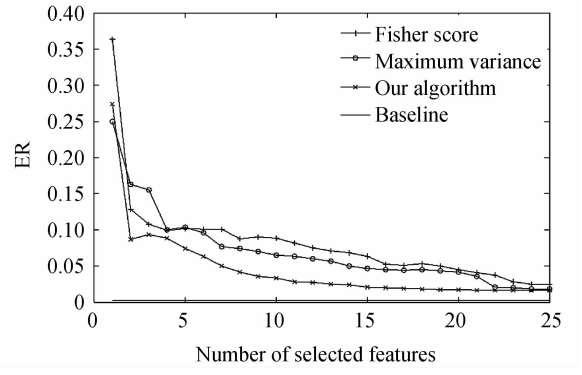

图 4 MSFS 算法的性能比较

Fig. 4 The performance comparison of MSFS and other algorithms

5 结论

提出一种基于流形的半监督特征选择算法,该算法考虑高光谱数据流形结构的不变特性和标记样本的先验信息,选择一组优化的光谱波段子集组成新的低维空间,既能够增强标记数据的判别结构,又能够充分挖掘整体数据的几何结构,有助于后续的相关处理.最后,通过实验,利用真实高光谱数据验证了本文方法的有效性.

参考文献

[1] PLAZA A, BENEDIKTSSON J A, BOARDMAN J W, *et al*. Recent advances in techniques for hyperspectral image processing[J]. *Remote Sensing of Environment*, 2009, **113**: S110-S122.

[2] JIA X, KUO B C, CRAWFORD M M. Feature mining for hyperspectral image classification [J]. *Proceedings of the IEEE*, 2013, **101**(3): 676-697.

[3] RODARMEL C, SHAN J. Principal component analysis for hyperspectral image classification[J]. *Surveying and Land Information Science*, 2002, **62**(2): 115-122.

[4] ETEMAD K, CHELLAPPA R. Discriminant analysis for recognition of human face images[J]. *JOSA A*, 1997, **14**(8): 1724-1733.

[5] HE X, NIYOGI P. Locality preserving projections[C]. Neural Information Processing Systems, 2003, **16**: 234-241.

[6] CAMPS-VALLS G, BANDOS M T, ZHOU D. Semi-supervised graph-based hyperspectral image classification[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2007, **45**(10): 3044-3054.

[7] YANG L X, YANG S Y, JIN P L, *et al*. Semi-supervised hyperspectral image classification using spatio-spectral laplacian support vector machine[J]. *Geoscience and Remote Sensing Letters IEEE*, 2014, **11**(3): 651-655.

[8] ZHAO J, LU K, HE X. Locality sensitive semi-supervised feature selection[J]. *Neurocomputing*, 2008, **71**(10): 1842-1849.

[9] PAL M, FOODY G M. Feature selection for classification of hyperspectral data by SVM [J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2010, **48**(5): 2297-2307.

[10] BELKIN M, NIYOGI P, SINDHWANI V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples[J]. *The Journal of Machine Learning Research*, 2006, **7**: 2399-2434.

[11] HE X, CAI D, NIYOGI P. Laplacian score for feature selection[C]. Advances in Neural Information Processing Systems, 2005: 507-514.

[12] YANG J, ZHANG D, JIN Z, *et al*. Unsupervised discriminant projection analysis for feature extr [C]. International Conference on Pattern Recognition. IEEE, 2006, **1**: 904-907.

[13] BACHMANN C M, AINSWORTH T L, FUSINA R A. Exploiting manifold geometry in hyperspectral imagery[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2005, **43**(3): 441-454.

[14] BELKIN M, NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. *Neural Computation*, 2003, **15**(6): 1373-1396.

[15] CHUNG F R K. Spectral graph theory[M]. Fresno: AMS Bookstore, 1997.

[16] SUGIYAMA M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis[J]. *The Journal of Machine Learning Research*, 2007, **8**: 1027-1061.

[17] DUDA R O, HART P E, STORK D G. Pattern classification [M]. New York: John Wiley & Sons, 2012.