

一种 SVM-RFE 高光谱数据特征选择算法

张 睿^{1,2} 马建文³

(1 中国科学院遥感应用研究所,北京市大屯路甲 11 号,100101)

(2 中国科学院研究生院,北京市玉泉路甲 19 号,100049)

(3 中国科学院对地观测与数字地球科学中心,北京市中关村北一条 9 号,100190)

摘 要:提出了一种基于一对一(one-verse-one, OVO)多类策略的支持向量机递归特征约减算法(support vector machine recursive feature elimination, SVM-RFE)用于高光谱数据的特征选择。对比分析了该算法所选择波段与基于一对多(one-verse-all, OVA)策略的 SVM-RFE 算法、MSVM-RFE 算法以及 OneR、InfoGain、ReliefF 等 3 种基于特征排序的方法所选择波段在高光谱数据分类中的精度表现。结果显示,OVO SVM-RFE 算法是一种可靠有效的高光谱数据特征选择算法,并且所选择波段在分类精度方面优于 5 种对比算法。

关键词:SVM-RFE;特征选择;高光谱

中图法分类号:P237.4

高光谱数据具有光谱分辨率高,波段数目多的特点。目前,有多种方法可应用于高光谱数据的特征选择^[1-5]。SVM-RFE 是一种将支持向量机用于高维数据特征选择的算法,最先用于分子生物学领域^[6],后来在遥感领域中也有应用^[7]。SVM 算法是一种基于两类问题分类的算法,为了将该算法扩展到适合遥感应用的多类分类,一般的解决思路是同时训练多个 SVM 分类器,基于 OVO 或 OVA 策略处理多类问题。SVM-RFE 算法同样使用多类问题策略。本文提出了一种使用 OVO 策略的 SVM-RFE 算法,并且将其应用于航空 AVIRIS 高光谱数据的特征选择。将该方法所选择波段与 OVA SVM-RFE、MSVM-RFE、OneR、InfoGain、ReliefF 等 5 种特征选择算法所选波段进行了分类实验,验证了该算法的有效性。

1 算法原理

1.1 SVM-RFE 算法

SVM-RFE 算法是一种基于 SVM 中最大间隔原理的后向序列约减算法。SVM-RFE 算法不依赖于特定的分类器,而是使用 SVM 算法中的

权重 $\|W\|^2$ 对所有波段进行排序,并评估每一波段对分类的影响,按排序准则 R_c 移除对分类器作用最小的波段。排序准则 R_c 为完整 SVM 问题权重 $\|W\|^2$ 与假设移除某一波段后的 SVM 问题权重 $\|W^{(-p)}\|^2$ 差的绝对值。

$$\|W\|^2 = \sum_{i,j=1}^N \alpha_i^* \alpha_j^* y_i y_j K(x_i, x_j) \quad (1)$$

$$R_c = \left| \|W\|^2 - \|W^{(-p)}\|^2 \right| = \frac{1}{2} \left| \sum_{i,j=1}^N \alpha_i^* \alpha_j^* y_i y_j K(x_i, x_j) - \sum_{i,j=1}^N \alpha_i^{*(-p)} \alpha_j^{*(-p)} y_i y_j K(x_i^{(-p)}, x_j^{(-p)}) \right| \quad (2)$$

式中, i, j 表示循环变量; y 表示类别标记; N 表示样本数目; $x_i^{(-p)}$ 表示移除了第 p 波段后的训练点; $K(x_i, x_j)$ 表示核函数; α_i^* 和 $\alpha_i^{*(-p)}$ 由解算 SVM 对偶优化问题得到; S 表示现有波段列表。在得到了所有波段的 R_c 后,约减最小的值所对应的波段。重复上面的过程,直到原始波段列表为空或选择波段达到预先设定的波段数目。

1.2 OVO SVM-RFE 算法

OVO 策略将每一类训练样本 O_1 与另一类训练样本 O_2 构建二类分类子问题,每一类都会

与其他所有的类构建一个子问题,如果待分类数为 C_n ,则共有 $T=C_n \times (C_n - 1)/2$ 个子问题。在每一次 SVM-RFE 迭代过程中,都需要求解 T 个子问题,得出 T 个代表排序准则 R_c 值的列表。分别找出 T 个问题中最小 R_c 所对应的波段 q ,并对 T 个问题的 q 进行统计 $K = \sum_T q$ 。如果某个波段的 K 明显大于其余波段,则说明该波段对几乎所有 T 个子问题所得到的 R_c 值都最小,即该波段最应该被约减掉。如果在 T 个子问题中,没有任一波段的 K 明显大于其余波段,或者类别数目 T 较小,则需要使用一定的容差范围来进一步选择。依据容差值选取 K 较大的几个波段组成波段集合 X ,对这些波段的 R_c 值进行求和 $J = \sum_X R_c$,选择 J 最小的波段为移除波段。容差的使用是为了减小偶然误差对 K 计算的影响,本文使用的容差值为:

$$t = \begin{cases} 3, r \leq T \\ 1, T < r \leq 2T \\ 0, \text{其他} \end{cases} \quad (3)$$

式中, r 是指将原始波段列表中波段迭代约减后剩余的波段数目。

2 数据与实验

2.1 高光谱数据

本文使用的高光谱数据为 1992 年 6 月获取的航空 AVIRIS 影像,为美国印第安纳州农作物覆盖区域。图像大小为 145 像素 \times 145 像素,光谱范围为 0.4~2.5 μm ,共有 220 波段。该数据可以从网上公开获取,地面真实验证数据可同时得到。原始图像中有 16 种地物,选取面积较大的 9 种地物进行实验^[8]。训练样本和验证样本数目分别为 445 和 1 783。

2.2 实验方法

由于 104~108,150~163,220 波段成像受大气水分吸收影响较大,这些波段可认为是噪声^[8],在预处理阶段去除这些波段,实验过程中使用剩

余的 200 波段。依据图像各波段极值,将训练样本及验证样本归一化到 $[0, 1]$ 范围。使用 SVM 分类算法对训练样本进行训练,核函数选择径向基函数 RBF,其公式为 $K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$ 。在使用 SVM 进行分类的过程中,由于 SVM 模型选择问题会对分类精度产生很大影响,因此,为了精确地获得超参数,本文使用了 Leave-One-Out 交叉验证(LOOCV)^[9]方法进行网格式搜索超参数。对于本文所使用的分类器和核函数,所需选择的超参数为惩罚系数 C 和核函数宽度 γ 。

SVM-RFE 算法在每次迭代中约减一个波段,直到所有波段都被约减。算法运行结束后,会得到一个根据排序准则降序排列的波段列表。分别将前 150、100、50、20、10、5 波段作为特征选择的结果进行后续的分类实验。使用 SVM 分类器评估特征选择的结果,较高的分类精度意味着较好的特征选择效果。为了进行对比研究,本文还实现了 OVA SVM-RFE,以及一种基于多 SVM-RFE 算法的特征选择方法(MSVM-RFE)^[10]。MSVM-RFE 算法同样使用 OVO 策略将 SVM-RFE 扩展成多类问题,但在算法运行过程中,它需要统计波段的均值和方差,并用这些信息构造新的排序准则,进而选择移除的波段。这两种算法都属于 SVM-RFE 系列算法,只是在计算过程中与本文方法有所不同。实验中还选择了 OneR、InfoGain 和 ReliefF 等 3 种基于特征排序的算法进行比较。OneR 算法基于一种简单的分类器进行特征排序;InfoGain 对移除某波段引起的信息量变化值进行度量并排序,信息量的度量使用熵值;而 ReliefF 算法借用了最近邻算法的思想,认为好的特征应使得最近邻的同类样本之间特征值相近,据此对每个特征赋权值并进行排序。

3 实验结果与分析

表 1 给出了上述 6 种算法的特征选择结果,只显示前 20 波段的特征选择结果。

表 1 特征选择结果

Tab. 1 Result of Feature Selection

算法	波段(5,10,20)
OVO SVM-RFE	26,27,28,30,31,42,43,61,62,63,70,74,76,110,124,125,126,149,164,172
OVA SVM-RFE	6,26,27,29,30,31,33,37,38,61,73,75,81,83,88,91,99,100,101,134
MSVM-RFE	27,29,36,40,42,43,74,75,76,82,107,108,110,112,126,127,149,151,162,177
OneR	16,18,23,25,26,27,28,29,30,119,124,127,130,132,134,135,136,138,140,141
InfoGain	134,135,138,140,149,159,160,162,163,165,166,167,168,169,170,171,173,177,178,179
ReliefF	148,149,150,153,154,155,156,157,158,159,160,162,163,164,165,166,167,169,172,173

从表1可以看出,3种基于SVM-RFE的方法所选择出的波段分布范围相似,而InfoGain和ReliefF所选波段范围相似。特征选择前20波段的分布范围如表2所示。

表2 各波段范围内波段数目
Tab.2 Band Number for Ranges

波段范围	OVO	OVA	MSVM-RFE	OneR	InfoGain	ReliefF
1~40	5	9	4	9	0	0
41~80	8	3	5	0	0	0
81~120	1	7	5	1	0	0
121~160	4	1	4	10	7	11
161~200	2	0	2	0	13	9

表3 所选择波段的分类结果/%

Tab.3 Result of Classification for Selected Bands/%

算法	150	100	50	20	10	5
OVO SVM-RFE	95.008 4	95.961 9	95.793 6	94.784 1	95.008 4	92.596 7
OVA SVM-RFE	95.008 4	94.896 2	95.625 4	93.774 5	92.989 3	88.278 2
MSVM-RFE	93.942 8	94.111 0	93.886 7	93.830 6	93.550 2	91.643 3
OneR	95.457 1	93.998 9	93.157 6	86.483 5	80.370 2	75.322 5
InfoGain	95.401 0	93.269 8	91.587 2	81.940 5	69.770 1	68.760 5
ReliefF	95.232 8	92.765 0	85.642 2	70.947 8	65.675 8	61.020 8

从表3可以看出,3种SVM-RFE算法所选择波段的分类精度要明显高于其他3种算法,特别是当选择波段只有20、10、5波段时,对比算法的分类精度与SVM-RFE系列算法的精度差距迅速拉大至10%~20%,说明这3种对比算法所选择出的波段并不是区分各个类别的最有效的波段;而3种SVM-RFE算法选择出的波段,分类精度一直比较稳定,并没有随着波段数的减少而明显下降,说明该类算法比较成功地找到了区分判别上述地物的波段。

表3显示,OVO SVM-RFE算法分类精度要稍高于OVA SVM-RFE和MSVM-RFE算法,特别是当选择波段至20、10、5时,OVO SVM-RFE算法相比较OVA SVM-RFE算法和MSVM-RFE算法,在分类精度上分别提高1%、2%、4.3%和0.9%、1.5%、0.9%。这说明本文使用OVO思想将SVM-RFE扩展到支持多类数据的特征选择和分类的过程是有效的,它比基于OVA策略的SVM-RFE算法在分类精度方面有一定优势。对OVO SVM-FRE算法,150、100、50、20、10波段的数据的分类精度都优于原始200波段数据的分类精度,这说明对于SVM等较少受维数灾难影响的分类器,特征选择对于后续的数据分类仍然是有益的。

从表2可以看出,3种基于SVM-RFE的算法在各个波段范围几乎都有分布,这说明对于前20波段,这些方法比较好地选择出了不同波谱区间中有代表性的波段。其中,MSVM-RFE方法选取的波段范围中波段数目最为平均,OVO其次,而OneR、InfoGain和ReliefF算法所选择波段的波谱范围分布不均匀,这意味着对于某些地物类型,这些方法不可能很好地加以区分。

表3给出了上述6种算法选择波段用于AVIRIS数据的部分选择波段分类验证结果,分类使用SVM分类器及RBF核函数。

4 结 语

1) 本文提出的OVO SVM-RFE算法对于高光谱数据的特征选择非常适用,在从150波段到5波段的分类实验中,分类精度一直保持比较稳定,没有明显的下降,说明该算法稳定可靠,所选择波段真正包含了区分地物类别的信息。

2) SVM-RFE系列算法是一种非常有效的高光谱数据特征选择方法,通过3种不同的策略得到的不同的SVM-RFE算法整体的分类精度都比较高,说明SVM-RFE通过循环迭代的过程,得到的各个波段分类能力的排序列表可以很好地判别各个波段的分类能力。

3) 即使对于SVM等较少受维数灾难影响的分类器,特征选择过程作为高光谱数据的预处理步骤,也可以对后续的分类过程带来有利影响。

参 考 文 献

- [1] Du Qian, Chang C I. A linear Constrained Distance-Based Discriminate Analysis for Hyperspectral Image Classification[J]. Pattern Recognition, 2001, 34(2):361-373
- [2] Harsanyi J C, Chang C I. Hyperspectral Image Classification and Dimensionality Reduction: an Orthogonal Subspace Projection Approach[J]. IEEE Transactions on Geoscience and Remote Sensing,

- 1994, 32(4):779-785
- [3] Heesung K, Nasrabadi N M. Kernel Orthogonal Subspace Projection for Hyperspectral Signal Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2005, 43(12):2 952-2 962
- [4] Du Hongtao, Qi Hairong, Wang Xiaoling, et al. Band Selection Using Independent Component Analysis for Hyperspectral Image Processing[C]. Applied Imagery Pattern Recognition Workshop, Washington D C, USA, 2003
- [5] Kim D H, Finkel L H. Hyperspectral Image Processing Using Locally Linear Embedding[C]. The First International IEEE EMBS Conference on Neural Engineering, Capri, Italy, 2003
- [6] Guyon I, Weston J, Barnhill S. Gene Selection for Cancer Classification Using Support Vector Machines[J]. Machine Learning, 2002, 46(1): 389-422
- [7] Bazi Y, Melgani F. Toward an Optimal SVM Classification System for Hyperspectral Remote Sensing Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2006, 44(11): 3 374-3 385
- [8] Camps-Valls G, Bruzzone L. Kernel-Based Methods for Hyperspectral Image Classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2005, 43(6):1 351-1 362
- [9] Joachims T. Estimating the Generalization Performance of a SVM Efficiently[C]. The 17th International Conference on Machine Learning, San Francisco, USA, 1999
- [10] Duan K B, Rajapakse, J C, Wang H, et al. Multiple SVM-RFE for Gene Selection in Cancer Classification With Expression Data[J]. IEEE Transactions on Nanobioscience, 2005, 4(3):228-234

第一作者简介:张睿,博士生,主要从事遥感数据智能处理研究。
E-mail:david.zhangrui@gmail.com

A Feature Selection Algorithm for Hyperspectral Data with SVM-RFE

ZHANG Rui^{1,2} MA Jianwen³

(1 Institute of Remote Sensing Applications, CAS, 11A Datun Road, Beijing 100101, China)

(2 Graduate University of CAS, 19A Yuquan Road, Beijing 100049, China)

(3 Center for Earth Observation and Digital Earth, CAS, 9 Beiyitia, Zhongguancun, Beijing 100190, China)

Abstract: Many conventional classification algorithms have difficulties to be applied to hyperspectral data directly due to huge band number and high correlation among bands. Hence, how to reduce the band number and preserve the information of original data as much as possible simultaneously is an on-going issue. An algorithm of feature selection for hyperspectral data based on one-verse-one support vector machines recursive feature eliminate (OVO SVM-RFE) is proposed. The AVIRIS hyperspectral data was utilized to summarize the principle and characteristic in feature selection. The algorithm was compared to the one-verse-all SVM-RFE (OVA SVM-RFE), MSVM-RFE, OneR, InfoGain and ReliefF approach for feature selection in classification accuracy using support vector machines classifier. Experimental results indicate that the OVO SVM-RFE is a reliable and effective approach of feature selection for hyperspectral data and is superior to other five algorithms in accuracy of classification for selected bands.

Key words: SVM-RFE; feature selection; hyperspectral data