

中图分类号: TP751.1 文献标志码: A 文章编号: 1006-8961(2011)11-2030-06

论文索引信息: 李娜, 赵慧洁, 贾国瑞. 因子分析模型的高光谱数据降维方法 [J]. 中国图象图形学报, 2011, 16(11): 2030-2035

# 因子分析模型的高光谱数据降维方法

李娜, 赵慧洁, 贾国瑞

(北京航空航天大学仪器科学与光电工程学院, 精密光机电一体化技术教育部重点实验室, 北京 100191)

**摘要:** 为解决高光谱遥感数据量大且波段间相关性高等问题, 提出基于因子分析模型的高光谱数据降维方法。该方法通过因子载荷矩阵求解、模型参数求解、旋转矩阵计算以及因子得分估计, 得到表征高光谱图像的本征维数。该方法可以找出少数的几个综合因子来代表众多因子, 而这少数几个综合因子不仅能主要反映原来的众多因子的信息, 而且彼此独立, 从而实现高光谱数据的降维。通过利用航空推扫型成像光谱仪(PHI)数据进行本文方法的性能验证, 结果表明, Kappa 系数从未降维数据的 0.744 提高到 0.821, 满足了得到数据本征维数的同时最大程度地保留数据有用信息、消除波段间的相关性和增大类间的可分性的应用需求。

**关键词:** 高光谱遥感; 降维; 因子分析模型; 推扫型成像光谱仪(PHI)

## Dimensional reduction method based on factor analysis model for hyperspectral data

Li Na, Zhao Huijie, Jia Guorui

(School of Instrument Science and Opto-electronics Engineering, Beihang University, Precision Opto-mechanics Technology Key-Laboratory of Education Ministry, Beijing 100191 China)

**Abstract:** A dimensional reduction method based on the factor analysis model is proposed for hyperspectral data to resolve the problems of high relativity of bands and large volumes of data. The intrinsic dimensions of hyperspectral data can be obtained by our method through further processing, including solving the factor payload matrix, computation of model parameters and rotated matrix, and the estimation of the factor contribution. Less composite factors can be found to replace data of all bands, which can not only represent almost information of original data, but is also factor independent. Push Hyperspectral Imager (PHI) data is used to evaluate the performance of our proposed method. The result illuminates Kappa parameter is improved from 0.744 to 0.821, and all useful information of data is reserved, relativity among bands is removed, and class separability is increased after dimensional reduction.

**Keywords:** hyperspectral remote sensing; dimensional reduction; factor analysis model; push hyperspectral imager (PHI)

## 0 引言

高光谱数据维数高、数据量大; 光谱分辨率较高, 各个波段之间具有较高的相关性, 存在大量的

冗余信息; 并且随着波段数据的增加, 数据处理量呈指数上升。因此, 如何将高光谱数据降维, 消除波段间的相关性, 是数据处理的需要解决的关键问题之一。

目前针对高光谱数据降维已经提出很多方

收稿日期: 2010-07-13; 修回日期: 2010-11-29

基金项目: 国家高技术研究发展计划项目(2008AA121102, 2009AA12Z119, 2007AA12Z167); 中国地质调查局项目(1212010816033-3); 长江学者和创新团队发展计划项目(IRT0705)。

第一作者简介: 李娜(1978—), 女, 讲师。2009年于北京航空航天大学获测试计量技术及仪器专业博士学位, 主要研究方向为遥感信息建模与应用、高光谱遥感数据处理。E-mail: lina\_17@buaa.edu.cn。



$$\begin{aligned}
 V &= V_1 + V_2 + \dots + V_m = \\
 &\frac{1}{p} \sum_{i=1}^p (d_{i1}^2 - \bar{d}_1)^2 + \dots + \\
 &\frac{1}{p} \sum_{i=1}^p (d_{im}^2 - \bar{d}_m)^2
 \end{aligned} \tag{6}$$

所以,由  $\frac{\partial V}{\partial \theta} = 0$ , 得

$$\tan 4\theta = \frac{D - 2AB/p}{C - (A^2 - B^2)/p} \tag{7}$$

式中,  $A = \sum_{i=1}^p u_i, B = \sum_{i=1}^p v_i, C = \sum_{i=1}^p (u_i^2 - v_i^2),$

$$D = 2 \sum_{i=1}^p u_i v_i, u_i = \left(\frac{a_{i1}}{h_i}\right)^2 + \left(\frac{a_{i2}}{h_i}\right)^2, v_i = 2 \frac{a_{i1} a_{i2}}{h_i^2}$$

第一轮旋转后的因子载荷矩阵为  $A^{(1)}$ , 然后开始下一轮的旋转, 得到一系列的因子载荷矩阵为:  $A^{(1)}, A^{(2)}, \dots, A^{(s)}, \dots$ , 则必有:  $V^{(1)} \leq V^{(2)} \leq \dots \leq V^{(s)} \leq \dots, V^{(s)}$  为  $A^{(s)}$  各列元素平方的相对方差之和, 实际应用中, 当  $V^{(s)}$  值变化不大时, 即可停止旋转。

### 1.4 因子得分估计

在得到  $A$  和  $D$  的基础上, 采用加权最小二乘对  $F$  进行求解, 首先将数据中心化, 即  $\mu = 0, \varepsilon$  看做

误差, 则用误差方差的倒数作为权重的误差平方和, 得到了  $F$  的加权最小二乘估计:

$$\hat{F} = (\hat{A}^T \hat{D}^{-1} \hat{A})^{-1} \hat{A}^T \hat{D}^{-1} X \tag{8}$$

利用因子得分估计即可得到表征高光谱数据的本征维数, 从而实现高光谱数据降维<sup>[7]</sup>。

## 2 实验与结果分析

### 2.1 实验数据

为了更好地说明基于因子分析模型的高光谱数据降维方法的性能与优点, 利用该方法进行江苏方麓茶场地区的 PHI 航空高光谱数据的数据降维。推帚式超光谱成像仪 (PHI) 由中科院上海技术物理研究所研制, 具有 124 个波段, 波长范围为 417.3 ~ 854.4nm, 试验过程中所使用的方麓茶场地区高光谱数据中包含的地物类型为水稻 C4、香菜 V13、红薯 V2、茶树 T7、竹子 T6、水体 W2、裸土 S2。

### 2.2 降维结果比较

采用基于因子分析模型的高光谱数据降维方法, 首先进行因子分析建模, 再进行因子载荷矩阵求解与因子载荷矩阵求解, 最后利用因子得分估计结果实现数据降维, 结果如图 1 所示。为

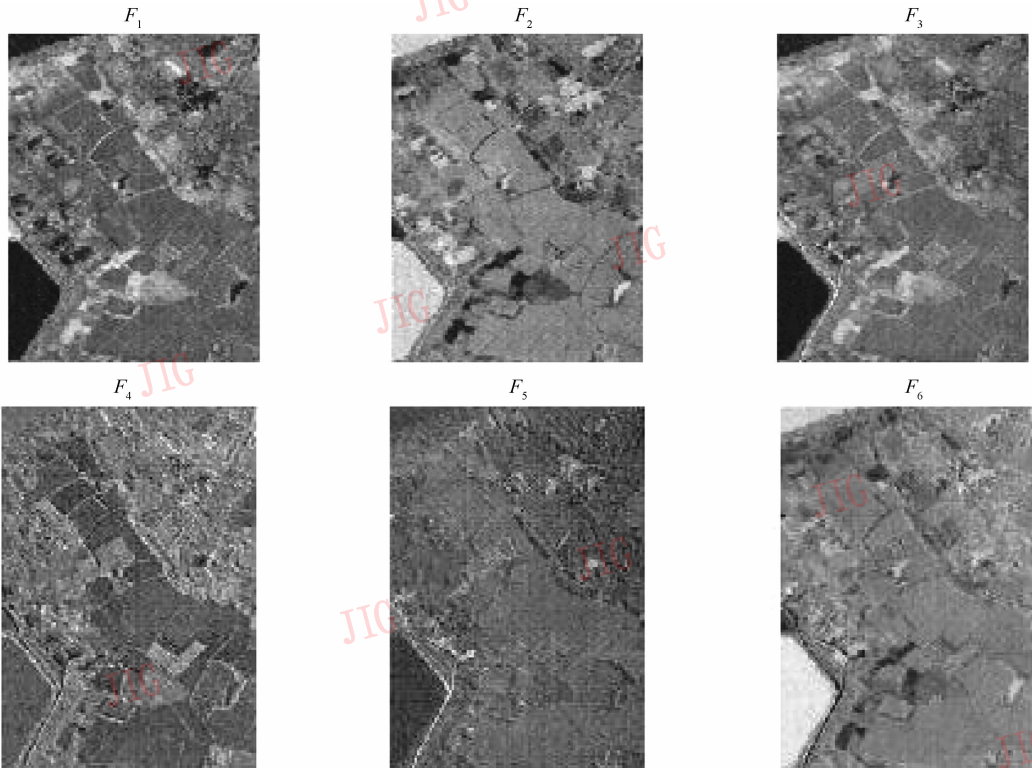


图 1 基于因子分析模型的高光谱数据降维结果

Fig. 1 Data dimension reduction based on factor analysis model

了较好地分析高光谱数据经过该方法的数据降维增大地物之间的可分性,采用光谱角制图(SAM)方法进行红薯、水稻、香菜、水体、裸土等不同地物间的光谱相似性的综合度量,原始数据的 SAM 结果与降维后数据的 SAM 结果分别如表 1 和表 2 所示。

表 1 原始数据的不同类型地物之间的 SAM 结果

Tab.1 SAM result of raw data

	红薯	水稻	香菜	水体	裸土
红薯	0	0.077 9	0.344 4	0.467 5	0.458 6
水稻	0.077 9	0	0.302 6	0.410 3	0.418 2
香菜	0.344 4	0.302 6	0	0.199 8	0.147
水体	0.467 5	0.410 3	0.199 8	0	0.184 9
裸土	0.458 6	0.418 2	0.147	0.184 9	0
平均值	0.337 1	0.302 918	0.248 45	0.315 625	0.302 175

表 2 因子分析模型降维数据不同类型地物之间的 SAM

Tab.2 SAM result of dimension reduction data based on factor analysis model

	红薯	水稻	香菜	水体	裸土
红薯	0	0.312 9	0.503 6	2.141 5	0.566 4
水稻	0.312 9	0	0.687 5	1.889 9	0.776 7
香菜	0.503 6	0.687 5	0	2.099 9	0.426 9
水体	2.141 5	1.889 9	2.099 9	0	1.709 9
裸土	0.566 4	0.776 7	0.426 9	1.7099	0
平均值	0.906 10	0.916 75	0.929 475	1.960 3	0.869 975

由表 1 和表 2 可以看出,降维后的数据不同地物之间光谱角明显增加,原始数据不同地物之间的平均光谱角最大为 0.337 1,而降维后不同地物之间的平均光谱角最小为 0.869 975,表明经过数据降维后的数据不同类型之间的光谱特征可分性大大增加。

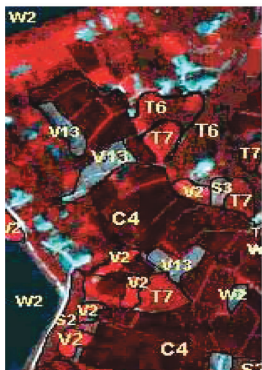
### 2.3 分类效果比较

为了验证本文方法的降维效果,利用误差平方和最小准则的 K 均值分类器对未降维数据、因子分析模型降维数据进行分类研究。在分类结果量化评价中,选择的测试样本以及不同地物类型的标识如表 3 所示,地面参考信息图以及得到的分类结果比较如图 2 所示。未降维数据、因子分析模型降维数据的分类结果统计参数分别如表 4 和表 5 所示。混淆矩阵中的第 i 行第 j 列元素 (i, j) 表示:第 i 类地物被分到第 j 类地物中的样本数目;并且计算 Kappa 系数用于进行

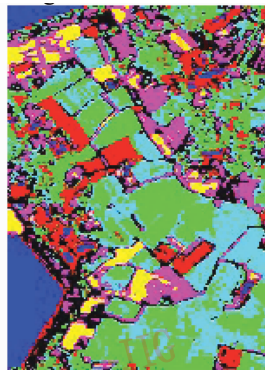
表 3 用于农业精细分类的各类别测试样本

Tab.3 Test samples for crop classification

	地物标识	测试样本数目
茶树	T7	306
竹子	T6	255
红薯	V2	197
香菜	V13	246
水稻	C4	743
水体	W2	468
裸土	S2	165
总和		2 380



(a) 地面参考信息图



(a) 未降维数据分类结果



(c) 因子分析模型降维数据分类结果

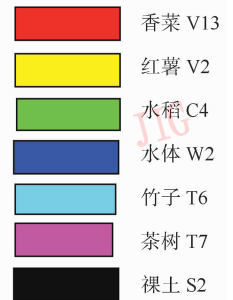


图 2 地面参考信息与不同数据 K 均值分类结果

Fig.2 Comparison of K-means classification results

表 4 未降维数据  $K$  均值的分类结果统计Tab. 4 Confusion matrix of  $K$ -means classification results for raw data

	水稻	红薯	香菜	茶树	竹子	水体	裸土	总计	正确分类精度/%	全局 Kappa 系数
水稻	560	0	0	0	183	0	0	743	75.37	
红薯	0	197	0	0	0	0	0	197	100	
香菜	0	0	209	0	0	0	37	246	84.96	
茶树	14	39	0	201	28	0	24	306	65.69	
竹子	105	0	0	11	128	0	11	255	50.2	
水体	0	0	0	0	0	468	0	468	100	
裸土	0	0	0	47	0	0	118	165	71.52	
总计	679	236	209	259	339	468	190	2 380	79.03	0.744

表 5 因子分析模型降维数据  $K$  均值的分类结果统计Tab. 5 Confusion matrix of  $K$ -means classification results for dimension reduction data based on factor analysis model

	水稻	红薯	香菜	茶树	竹子	水体	裸土	总计	正确分类精度/%	全局 Kappa 系数
水稻	689	0	0	0	54	0	0	743	92.73	
红薯	0	197	0	0	0	0	0	197	100	
香菜	0	0	195	0	28	0	23	246	79.27	
茶树	7	12	0	228	47	0	12	306	74.51	
竹子	89	0	0	0	135	0	31	255	52.94	
水体	0	0	0	0	0	468	0	468	100	
裸土	0	0	0	42	0	0	123	165	74.55	
总计	785	209	195	270	264	468	189	2 380	85.5	0.821

分类结果与地面所对应区域的实际类型一致性的概率分析。从图 2 以及表 4 和表 5 所示的各类数据分类结果可以看出,未降维数据的分类结果,总体分类精度达 79.03%,Kappa 分析系数达 0.744,水稻与竹子之间以及茶树与红薯、裸土之间存在较大的误分类;而基于因子分析模型降维数据的分类结果,总体分类精度达 85.5%,Kappa 分析系数达 0.821。

### 3 结 论

针对高光谱数据量大、波段间相关性高等特点,提出基于因子分析模型的数据降维方法,并

利用 PHI 数据从增大类间可分性、提高后续处理精度方面进行了降维效果分析。通过分析可知,经过基于因子分析模型的数据降维方法操作后,得到 6 个表征高光谱数据的本征维数,并且在实现数据降维的过程中,因子载荷矩阵进行了基于方差最大的正交旋转,在 6 个本征维数中,每一类地物只在一个因子中占主导地位,因此,在数据降维后,不仅较好地保存了数据的所有信息,而且较大程度上增大了数据的可分性;通过分类结果比较,经因子分析模型降维后的数据能够有效地提高分类效果,Kappa 系数从未降维数据的 0.744 提高到 0.821,显示了该方法的有效性和实用性,具有良好的应用前景。

## 参考文献 (References)

- [1] Manolakis D, Marden D. Dimensionality reduction of hyperspectral imaging data using local principal components transforms [C]// Proceedings of SPIE, Washington D C, USA: SPIE, 2004, 5425: 393-401.
- [2] Jia Xiuping, Richards J A. Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification [J]. IEEE Trans. on Geoscience and Remote Sensing, 1999, 37(1): 538-542.
- [3] Scholkopf B, Smola A, Müller K R. Nonlinear component analysis as a kernel eigenvalue problem [J]. Neural Computation, 1998, 10(5): 1299-1319.
- [4] Wu Ningning, Wang Jing. Factor analysis based anomaly detection [C]// Proceedings of the IEEE, Workshop on Information Assurance. Washington D C: IEEE Press, 2003, 108-115.
- [5] Liu Xiao gang, Zhao Hui jie, Li Na. Feature extraction based on fractal spectrum for hyperspectral data [J]. Acta Optical Sinic, 2009, 29(3): 844-848. 刘小刚, 赵慧洁, 李娜. 基于多重分形谱的高光谱数据特征提取 [J]. 光学学报, 2009, 29(3): 844-848.
- [6] Zhu Jianping. Applied Multi-statistics Analysis [M]. Beijing: Science Publishing Company, 2006: 15-56. [朱建平, 应用多元统计分析 [M], 北京: 科学出版社, 2006: 15-56.]
- [7] Li Na. Research on accurate modeling and its applications for hyperspectral remote sensing data [D]. Beijing: Beihang University, 2008: 23-33. [李娜, 高光谱遥感数据精确建模与应用处理技术研究 [D]. 北京: 北京航空航天大学, 2008: 23-33.]