

# 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	1
1.1	研究背景及意义 . . . . .	1
1.2	研究现状 . . . . .	2
1.3	本文主要工作 . . . . .	2
1.4	本文组织结构 . . . . .	2
2	方案论证	2
3	过程论述	3
4	实验与分析	4

## 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  $\text{\TeX}$  系统，毫不夸张地评价， $\text{\TeX}$  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

半监督学习是利用....

### 1.3 本文主要工作

本文主要针对 HSI 问题的，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，把正则项的中的范数改为 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

## 2 方案论证

这部分主要介绍协同表达模型特点，以及在 HSI 领域的应用情况

**定理 2.1** (Lévy). 令  $F(x), \varphi(t)$  分别为随机变量  $X$  的分布函数和特征函数。假定  $F(x)$  在  $a+h$  和  $a-h(h>0)$  处连续，则有

$$F(a+h) - F(a-h) = \lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{-T}^T \frac{\sin ht}{t} e^{-ita} \varphi(t) dt \quad (1)$$

证明. 从略。感兴趣的读者可以参考……。

□

**推论 1.** 密度函数和特征函数之间有如下的关系。

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi(t) dt \quad (2)$$

证明. 由公式 (1) 和 Lebesgue 定理，我们有

$$\begin{aligned} \frac{F(x+\Delta x) - F(x)}{\Delta x} &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\sin(t\Delta x/2)}{t\Delta x/2} e^{-it(x+\Delta x/2)} \varphi(t) dt \\ f(x) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \lim_{\Delta x \rightarrow 0} \frac{\sin(t\Delta x/2)}{t\Delta x/2} e^{-it(x+\Delta x/2)} \varphi(t) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi(t) dt \end{aligned} \quad \square$$

我们知道特征函数的定义是

$$\varphi(t) = E(e^{itX}) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx \quad (3)$$

Lévy 定理在分布函数和特征函数之间搭建了一座桥梁。对比 (2) 和 (3) 可见，密度函数和特征函数之间的关系非常巧妙。

在  $\text{\TeX}$  环境里，数学公式的表达是很自然的，绝大多数命令就是英文的数学专有名词或它们的缩写，如果你以前读过英文的数学文献，记忆这些命令是不难的。如果你没读过，正好通过记忆这些命令来了解术语。

手头有个命令快速寻查表是很方便的，我用的是 Hypertext Help with  $\text{\TeX}$ ，网上可以搜到，是免费的。

贝叶斯方法 [Gelman et al., 2004] 主要用于小样本数据分析，它利用参数先验分布和后验分布之差异进行统计推断，其一般步骤是：

1. 构建概率模型，包括参数的先验分布。
2. 给定观察数据，计算参数的后验分布。
3. 分析模型的效果，如有必要，回到第一步。

**例 1.** 下面，我们给一个表格的例子，一个图形的例子。

表 1: 二维随机向量  $(X, Y)$  的边缘分布

$X \backslash Y$	$y_1$	$y_2$	$\cdots$	$y_j$	$\cdots$	
$x_1$	$p_{11}$	$p_{12}$	$\cdots$	$p_{1j}$	$\cdots$	$p_{1\cdot}$
$x_2$	$p_{21}$	$p_{22}$	$\cdots$	$p_{2j}$	$\cdots$	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_i$	$p_{i1}$	$p_{i2}$	$\cdots$	$p_{ij}$	$\cdots$	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	1

在表 1 中， $p_{\cdot j} = \sum_i p_{ij}$ ，类似地， $p_{i\cdot} = \sum_j p_{ij}$ 。

### 3 过程论述

这个部分包括模型改进介绍，包括已有模型、改进后的模型、改进后的算法，改进后的实现

## 4 实验与分析

这个 TeX 模板只是为了提供一个学习 TeX 的参考，各节的内容并没有关联性。欢迎读者使用并改进该模板，并祝学习 TeX 愉快！

Knuth 大师最初设计 TeX 的时候并没有想到中文化，TeX 排版系统的中文化始终令初学者望而却步、云山雾罩。类 UNIX 系统下的 teTeX 和 Windows 系统下的 MikTeX，都是 TeX 知名的发行版。然而，teTeX 已经停止研发五年之久，基于 MikTeX 的中文发行版 CTeX 虽然如火如荼，但依然挡不住 TeXLive 一统江湖的大趋势。

虽然 TeXLive 还未入住 FreeBSD 的 ports tree，但 teTeX 的远去，令 FreeBSD 之下的很多 ports 不得不面临改换门庭的窘境。例如，auctex、latex-cjk 等等。

TeX 的中文化可以有多种途径，xelatex 是最简单的（不见得是最美观的）。在 TeXLive 2011 之下，不需要有任何更多的设置，甚至不用考虑中英文混排，xelatex 能满足绝大多数中文化要求。这对于初学者来说，无疑是一个福音。

## 参考文献

- [Gelman et al., 2004] Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2004) Bayesian Data Analysis (Second Edition). Chapman & Hall/CRC.