有限混合密度模型及遥感影像 EM 聚类算法

骆剑承 周成虎
(中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101)
梁 怡 (香港中文大学地理系, 香港)
马江洪 (长安大学数学教研室, 西安 710049)

摘 要 遥感信息是地球表层信息的综合反映. 由于地球表层系统的复杂性和开放性, 地表信息是多维的、无限的, 遥感信息传递过程中的局限性以及遥感信息之间的复杂相关性, 决定了遥感信息其结果的不确定性和多解性. 遥感信息具有一定的统计特性, 同时又具有高度的随机性和复杂性, 在特征空间中往往表现为混合密度分布. 针对遥感信息这种统计分布的复杂性, 提出了有限混合密度的期望最大(EM)分解模型, 该模型假设总体分布可分解为有限个参数化的密度分布, 通过 EM 迭代计算可估计出各密度分布的最大似然参数集; 将有限混合 EM 聚类算法应用于遥感影像的聚类分析中, 并与传统统计聚类方法进行了比较, 比较结果表明, 其对复杂地物的区分具有优势, 另外在融合专家知识、初始化等方面具有扩展能力.

关键词 混合模型 EM 算法 聚类 遥感数据
中图法分类号: TP751 **文献标识码**: A **文章编号**: 1006-8961(2002)04-0336-05

Finite Mixture Model and Its EM Clustering Algorithm for Remote Sensing Data

LUO Jian-cheng[1], ZHOU Cheng-hu[1], LEUNG Yee[2], MA Jiang-hong[3]
[1](*LREIS, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing* 100101)
[2](*Department of Geography, the Chinese University of Hong Kong, Hong Kong*)
[3](*Changan University, Xi'an* 710049)

Abstract Generally, the analyzed results from remote sensing data are uncertain and multi-solution, which is determined by the characteristics of global surface information being multi-dimensional and infinite. Therefore, remote sensing information has some degree of definite statistical characteristic, but as well as holds the high randomness and complexity, which generally behaves as mixture density distribution in feature space. In allusion to the complexity of statistical distribution of remote sensing information, in this study we firstly introduce into the finite mixture model and its expectation maximization(EM) algorithm for decomposing the mixture distribution into finite parametric density distributions in order to simulate or approach the whole mixture distribution. By the model it should be firstly assumed that whole distribution could be separated into infinite parametric density distributions, then by EM iterative computation the maximum likelihood parameters of each proportional distribution can be estimated. Furthermore, the finite mixture model and its EM algorithm are extended to clustering algorithm for remotely sensed data. By the experimental case, the EM clustering algorithm is synthetically compared with conventional statistical clustering algorithm. The results show that the EM algorithm has several particular advantages such as self-adaptive decision for clustering number, extensibility of prior-knowledge integration and free initialization, etc.

Keywords Mixture model, The EM algorithm, Clustering, Remote sensing data

基金项目: 中国科学院创新项目(KZCX1-Y-02)
收稿日期: 2001-05-16; **改回日期**: 2001-07-06

0　引　言

遥感影像成像过程受多方面随机变化因素的影响, 其中主要来自于成像过程的随机因素和成像对象的复杂性及不确定性, 从而导致获得的影像数据具有一定的随机性; 同时, 遥感影像作为一个整体, 反映的是地球表层系统中区域性地物的电磁波辐射量的情况, 因此其又具有总体综合的信息特征. 作为空间信息的重要组成部分, 遥感信息的分析处理过程具有空间信息复杂性和不确定性的特征. 统计分析一直是遥感数据分析处理的基本方法, 如大多数分类器都是建立在统计方法之上的, 包括 C 均值、最大似然、模糊分类、基于距离的统计聚类分析等[1~3], 这些分类器都要求样本数据遵循一定的统计特征分布. 根据密度分布模型(PDF) 是否可定义为参数化形式, 其可分为参数化统计模型和非参数化统计模型, 如 Bayes 最大似然分类器(MLC), 要求预先假设各类别在特征空间的密度分布服从高斯密度分布(GDD), 所以属于典型的参数化模型. 但很多具体地学分析问题, 参数化密度分布的假设并不一定能满足, 一些常见的参数化分布形式, 如高斯分布, 并不一定适合实际情况的复杂密度分布. 传统的参数化密度分布要求都是单峰形式的, 即只有一个极大值, 而实际问题中, 可能包含多峰的密度分布, 如在遥感影像中的水体信息, 由于其深浅、浑浊、表面波浪等程度或外界因素的差异, 其在特征空间上的分布往往表现为多种峰式分布; 另外, 不同类别的地物在特征空间上的分布可能存在重叠或相互交错, 因此就不能把这种复杂的分布通过简单的参数化形式表达出来.

遥感信息统计密度分布具有复杂性和多样性, 在特征空间中往往表现为多种密度分布的混合, 很难用传统的基于距离的统计方法来进行混合密度分布的参数估计[4]. 但对于这种混合模型可以采用密度分布降解的办法, 用有限个参数化密度分布的组合来逼近整体复杂的混合密度分布[5], 期望最大(Expectation Maximization, EM) 算法就是针对这种混合密度分布模型的一种有效参数化分解方法. 本文引进有限混合密度的期望最大分解算法, 并提出其应用于遥感影像聚类的具体算法. EM 模型首先假设遥感影像数据集由有限个参数化高斯密度分布, 根据一定的比例构成, 通过迭代计算, 得出各密度分布的最大似然参数估计, 最后通过密度分布的概率大小来确定类别的归属[6,7].

1　混合密度模型及其 EM 分解方法

1.1　混合密度

模式识别要解决的问题是如何对一系列过程、现象或对象进行识别和描述, 这些对象通常由有限个特征集或性质来决定. 遥感图象信息处理与分析正是这样一个模式识别问题. 因为实际数据往往受到噪声的污染甚至还可能丢失, 所以, 不可避免地要用到许多数学特别是统计学的有关方法. 为说明有限混合分布方法在图象模式识别中的应用, 考虑如下的具体问题: 假定在一个图象平面上有许多散点, 这些散点聚类形成若干对象, 如何识别和描述它们呢? 有限混合分布理论的方法就是对整个数据点拟合一个加权混合的概率密度函数, 使每个对象对应混合密度的一个分支密度, 而相应的权重正是该对象的数据点在整个数据集里所占的比例. 通过对混合密度的参数进行统计推断, 从而达到模式识别的目的. 实践已经证明这种方法是十分有效的, 并已得到了广泛的应用[2,8], 其模型参数估计一般采用从不完全数据中找局部极大似然估计的 EM 算法[6], 这个算法在统计学、神经网络等领域已有很多研究.

1.2　混合密度的 EM 参数估计算法

期望最大算法是迭代计算最大似然估计(MLE) 的一个常用方法, 可用在混合密度的参数识别中. 首先, 给出有限混合密度模型的定义. 假设数据 $\boldsymbol{x}(\boldsymbol{x}\in \mathbf{R}^m)$ 来自多个分布的混合体, 那么, 其概率密度就可表示为

$$f(\boldsymbol{x};\theta)=\sum_{i=1}^{g} r_i f_i(\boldsymbol{x};\theta_i)$$

其中, g 为密度分支的个数, $r_1,\cdots,r_g$ 是各混合密度分支点总体分布的比例(通常未知), $\sum_{i=1}^{g} r_i=1(r_i\geq 0, i=1,\cdots,g)$, f_i 是第 i 个分支的密度, θ_i 是相应分支的未知参数, 整个混合密度的参数为 $\theta=(r_1,\cdots,r_g,\theta_1,\cdots,\theta_g)$. 当有了观测数据集$\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$之后, 为识别混合密度参数 θ, 自然考虑用极大似然法来求其 MLE

$$\hat{\theta}_{\mathrm{MLE}}=\arg\max_{\theta}\prod_{i=1}^{n} f(\boldsymbol{x}_i;\theta)$$

不过, 由于密度函数的复杂性, 直接计算 MLE 往往具有一定的困难, 特别是对混合密度, 因此, 可借助

于从“不完全数据”x 来计算 $\hat{\theta}_{MLE}$的 EM 算法.

所谓“不完全数据”是相对“完全数据”而言的. 由于对所观察到的数据 x, 并不知道它来自哪个分支, 因而, 可看成是不完全的, 而完全数据应当是$(\boldsymbol{x}, y)$, 其中 y 表示 x 所属分支的标签, 取值为 $y\in\{1,\cdots,g\}$. EM 算法的最大特点在于: 它是通过对完全数据的处理来解决不完全数据的问题.

EM 算法本质上是个迭代算法. 它从初始解 $\theta^{(0)}$ 开始, 迭代地得到解 $\theta^{(1)},\cdots,\theta^{(t)}$, 在每步迭代中, 似然函数单调增加. 算法具体实施如下:

(1) 给定初始解 $\theta^{(0)}$;

(2) 对 $t=0,1,2,\cdots$, 重复下面两步:

E–步: 在观察样本和当前解 $\theta^{(t)}$ 给定的条件下, 计算完全数据的对数似然函数期望值

$$Q(\theta|\theta^{(t)})=\sum_{i=1}^{n}E_y\left[\log f(x,y;\theta)\,|\,x_i,\theta^{(t)}\right]$$

其中, $f(\boldsymbol{x},y;\theta)$ 是完全数据$(\boldsymbol{x},y)$ 的概率密度函数, E_y 表示关于随机变量 y 求期望.

M–步:

$$\theta^{(t+1)}=\arg\max_{\theta}Q(\theta|\theta^{(t)})$$

最后, 通过适当的停止规则终止迭代. 理论上已经证明, 在相当一般的条件下, EM 算法给出的最终值就是 $\hat{\theta}_{MLE}$.

特别地, 如果混合分支的密度均假定是高斯密度, 即

$$f_i(\boldsymbol{x};\theta)=2\pi^{-d/2}|\Sigma|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\mu_i)^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\mu_i)\right]$$

$$i=1,\cdots,g$$

其中, θ 代表均值向量 μ_i 和协方差矩阵 Σ, 则上述 EM 算法可简化为[7]:

E–步:

$$\tau_{ij}^{(t+1)}=\tau_i(\boldsymbol{x}_j;\theta_i^{(t)})=\frac{r_i^{(t)}f_i(\boldsymbol{x}_j;\theta_i^{(t)})}{\sum_{k=1}^{g}r_k^{(t)}f_k(\boldsymbol{x}_j;\theta_k^{(t)})}$$

即

$$\tau_{ij}^{(t+1)}=\frac{r_i^{(t)}|\Sigma_i^{(t)}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}_j-\mu_i^{(t)})^{\mathrm{T}}\Sigma_i^{(t)-1}(\boldsymbol{x}_j-\mu_i^{(t)})\right]}{\sum_{k}r_k^{(t)}|\Sigma_k^{(t)}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}_j-\mu_k^{(t)})^{\mathrm{T}}\Sigma_k^{(t)-1}(\boldsymbol{x}_j-\mu_k^{(t)})\right]}$$

其中, $\tau_{ij}^{(t+1)}$ 是在第 t 步迭代中, $\boldsymbol{x}_j$ 属于第 i 个分支(类)的后验概率.

M–步:

$$r_i^{(t+1)}=\frac{1}{n}\sum_{j=1}^{n}\tau_{ij}^{(t)}$$

$$\mu_i^{(t+1)}=\frac{1}{nr_i^{(t)}}\sum_{j=1}^{n}\tau_{ij}^{(t)}\boldsymbol{x}_j$$

$$\Sigma^{(t+1)}=\frac{1}{nr_i^{(t)}}\sum_{j=1}^{n}\tau_{ij}^{(t)}(\boldsymbol{x}_j-\mu_i^{(t)})(\boldsymbol{x}_j-\mu_i^{(t)})^{\mathrm{T}}$$

1.3 遥感图象 EM 聚类算法

根据是否需要先验知识, 分类可分为监督分类和非监督分类. 在实际应用中, 往往没有已知类别的样本可供利用, 甚至应将提供的样本分为几类也不清楚, 此时往往需要用非监督的方法进行划分. 聚类属于非监督分类的一种, 即根据样本间的相似程度, 在无预知其类别的情况下对数据集进行自动划分.

传统的聚类方法可分成分层法和划分法这两种主要的类型. 在分层法中, 不需事先规定类的个数, 也不存在初始化和局部极值引发的问题. 不过, 因为分层法在每步中只考虑局部区域, 所以, 它们不能加入有关类的总体形状或大小的先验知识, 从而它们也就不能分离重叠的类; 另外, 分层聚类法是静态的, 所以前面阶段指派的一个给定类不能移到不同的类中. 基于原型的划分聚类法是动态的, 且每个点可从一个类移到另一个类, 通过用适当的原型和距离度量, 可在划分法中加入关于类形状或大小的知识, 这些算法已扩展到探测直线、平面、圆、椭圆、曲线和曲面的情形. 常用的划分方法是交替优化法(即迭代法), 而迭代使得算法对初始化敏感, 并陷入局部极小, 划分法的另外两个缺点是: 类的个数难以确定以及对噪声和异常值的敏感性.

聚类分析是空间数据挖掘的最重要的方法之一, 其中, 在图象分类应用中, 就有多种聚类方法, 典型的有: 基于最近邻规则的试探法、最大最小距离法、K 均值算法、迭代自组织的数据分析法(ISODATA)、KOHONEN 聚类网络、ART 等. 这些聚类方法应用于遥感影像分类中, 都存在上述传统聚类方法的一些缺陷, 为此, 引进 EM 算法应用于混合模式的空间数据聚类. 遥感影像 EM 聚类算法主要包括以下几个过程:

(1) 样本数据的选择 从原始遥感影像中选取若干样本 $\boldsymbol{x}(\boldsymbol{x}\in\mathbf{R}^m)$, 可以随机选取, 也可在目视辅助下有选择地选取样本.

(2) 初始化 包括设定初始类别的个数 g, 各类别 i 的均值向量 μ_i 和协方差矩阵 Σ. 一般初始类别的数目根据实际情况给定一个较大的数, 可随机选择 g 个样本为 μ_i, 然后根据协方差的计算来设置 Σ.

(3) 最大似然参数估计　采用 EM 迭代运算不断进行各类别最大似然参数的估计.

(4) 类别个数的确定　当迭代过程相对进入一个稳定期时, 根据当前迭代步骤中得到的类别 i 的比例值 r_i 及该类别包含的样本个数来确定类别的个数. 具体方法是: 首先通过最大似然判断样本集中各样本归属于当前类别的个数, 判断最小包含样本个数的类别 j 的比例值 r_j 是否小于一个给定的值, 若是, 则删除该类别及其参数集(μ_j, Σ_j), 重新设置参数进行第 3 步的最大似然估计; 若否, 则继续进行第 3 步的 EM 迭代计算, 直到运行到一个长期稳定的阶段, 停止.

(5) 未知样本的类别归属　得到各类别的最大似然参数后, 逐点从图象中读入未知样本 $\boldsymbol{x}$, 计算 $\boldsymbol{x}$ 属于各类别 $\tilde{\omega}$ 的后验条件概率密度值

$$p(\boldsymbol{x}/\tilde{\omega}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_i|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\mu_i)^{\mathrm{T}}\Sigma_i^{-1}(\boldsymbol{x}-\mu_i)\right]$$

其中, $|\Sigma_i|$ 为协方差矩阵 Σ_i 的行列式. 判断 $\boldsymbol{x}$ 归属于 p 为最大的类别.

与传统的聚类方法相比较, EM 算法具有几个独特的性质, 因此, EM 算法在一定程度上克服了传统聚类算法的主要缺陷, 而且算法简单, 便于实施. 其主要优点是:

(1) 稳定性和可靠性　EM 算法的每一步迭代的类别似然值总是在不断增大, 具有稳定的单调收敛性, 而且在相当一般的初始条件下, 也有可靠的整体收敛性, 因此 EM 算法对初始条件要求很低.

(2) 计算简单性　似然函数不需要涉及到求导, 而是通过迭代来实现期望最大, 故便于实施.

(3) 易于解析表达　因为极大化问题一般涉及的是完全数据, 所以对数似然的条件期望易于计算, 使得 EM 算法在 E-步容易得到实施.

(4) 可确定类别个数　如上面所提出的, 可根据得到的各类别的比例值自适应地确定类别的最终个数.

(5) 由于 EM 算法的目的是对最大似然参数进行估计, 因此可通过 Bayes 扩展进行先验知识的融合, 另外, EM 算法根据实际情况具有很强的扩展能力, 其得到的结果反过来对聚类过程又具有很强的解释能力.

2　实例分析

2.1　实验区

试验区为香港港岛地区(经度 114°07′~ 114°16′, 纬度 22°11′~ 22°17′), 位于香港特别行政区的最南端, 维多利亚海湾之南, 北与九龙半岛隔海相望. 地势陡峭崎岖, 多丘陵山地, 其中以西北太平山为最高, 向四周辐射延伸, 主要以东西走向为主. 地形由北向南逐渐降低, 近海为狭窄的带状平地. 岛北是香港的主要市区之一, 是香港地区的行政、金融中心. 历史上最早就是从港岛北部的维多利亚港附近的平地和低丘区开埠, 原来仅跑马地附近为平地, 后来向东西方向通过移山填海工程, 密集的建筑群在港岛北部狭长的海岸带随地势变化基本上呈东西向延伸, 发展成为繁华的市区. 岛上山坡基本被茂密的植被所覆盖, 零星建筑物(主要是一些富豪的私家住宅和观光建筑) 散布在其中. 主要地物类型包括森林、草地、裸岩、水体、居民建筑用地、道路、沙滩等.

实验选用的遥感资料为 1996-03-03 的 Landsat-TM 数据. 如图版II 图 1 所示, 图象大小为 400 行× 600 列, 空间分辨率约为 30m, 覆盖大约 220km^2. 为了突出分类数据的多维性, 选用了 TM 除热红外波段的 6 个波段(CH1、CH2、CH3、CH4、CH5、CH7), 其中 CH1 是蓝色波段, CH2 是绿色波段, CH3 是红色波段, CH4 为近红外波段, CH5 (1.55μm~ 1.75μm) 和 CH7(2.08μm~ 2.35μm) 为中红外波段. 实验是用自行开发的遥感图象理解系统(GRS2000) 进行的, 全套系统是基于 VISUAL C^{++} 6.0 设计开发的.

2.2　结果分析

对影像的典型样本数据集进行聚类处理, 然后将得到的类别最大似然参数推广到整个图象. 一般随机样本的采集可以通过在图象上, 以一定行列间距采集, 也可以视地物复杂程度进行有选择性地采集, 一般在复杂区域选择的样本多, 原则上要求样本具有普遍性、多样性和均匀性. 考虑到算法的计算量, 样本采集的数量必须适中, 如果太多, 则影响计算速度, 而太少, 则影响分类的精度. 实验一共选取了1 000个样本数据.

通过 EM 迭代自适应计算, 最后得到 8 个大类别的分类图象(如图版II 图 2(a) 所示). 与实际土地覆盖情况的对照和分析可知, 8 个类别分别对应于: C1—清水体, C2—混浊水(排水区), C3—密集城区, C4—零星建筑(别墅群), C5—一般建筑(道路、低矮建筑), C6—裸露地(填海地、建筑工地、沙滩等), C7—茂密林地, C8—稀疏植被(灌木或草地). 这 8 个类别包括了研究区主要的地表景观类型. 同样, 用

传统的 ISODATA 方法也得到 8 个类别的聚类结果(如图版II 图 2(b) 所示).

通过比较,发现 EM 算法在以下几个方面比一般聚类方法更接近实际情况:

(1) 城市的区分　在ISODATA 方法得到的结果(图版II 图 2(a)) 中,很多山地的阴影与零星建筑(C4) 混淆在一起,另外,裸露地(C6) 和一般建筑(C5)、密集城市区域(C3) 等也容易混淆;而通过 EM 算法可以得到较好的区分.

(2) 植被的区分　由图版II 图 2(a) 可见,对于林地区域,ISODATA 由于受山坡阴阳坡的影响而进行了错误划分,另外,C5 和 C8 之间又引起了混淆;而通过 EM 算法只根据密度分布划分出 C6 和 C7 两类.

为什么 EM 算法得到的聚类结果会优于传统聚类方法呢? 首先,EM 算法是根据含量信息来自适应地调整类别个数,最后得到了与实际分布接近的分布状况,而传统 ISODATA 方法仅根据与均值点之间的距离来调整类别个数,在类别个数相同的条件下,ISODATA 对类别的确定没有严格的依据,造成与实际情况的偏差;其次,EM 算法是根据分布密度去寻找类别,寻找的类别表示区域内比较密集的点集,而 ISODATA 方法仅根据距离来确定点群之间的差别,不够合理;另外,一般传统的聚类方法通常采用普通的欧氏距离来确定特征空间中点之间的偏离,而 EM 算法因为采用无刻度的马氏距离作为密度函数中的距离,消除了各分量之间单位刻度上的不一致,所以 EM 算法得到的聚类结果更接近于实际类别的划分.

3　结论与展望

由于空间数据分布的复杂性和随机性,致使常规的统计聚类方法一般存在以下几方面的缺陷:(1) 很难确定初始化条件;(2) 很难确定全局最优类别个数;(3) 对散点抗干扰能力差;(4) 很难融合地学专家知识. 为此,引进了针对混合密度分布作最大似然参数估计的期望最大(EM) 算法,并提出了基于 EM 算法的遥感影像聚类的具体算法. 由于 EM 算法对初始条件不受限制,而且计算过程中提供了密度分布的参数,因此克服了传统统计聚类方法的上述缺陷. 通过实际例子的分析,EM 算法获得的结果更接近于实际分布. EM 算法可以根据实际情况进行扩展,如,由于 EM 算法是针对密度分布的,因此可直接引进 Bayes 理论进行先验知识的融合;通过引进稳健统计方法来加强 EM 算法抗干扰的能力[8~10].

参 考 文 献

1 Pedrycz W, Waletzky J. Fuzzy clustering with partial supervision [J]. IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, 1997, 27(5): 787~ 795.

2 Bastin L. Comparison of fuzzy c-means classification, linear mixture modeling and MLC probabilities as tools for unmixing coarse pixels[J]. International J. Remote Sensing, 1997, 18(6): 3629~ 3648.

3 Richards J A, Jia X. Remote sensing digital image analysis: An introduction[M], Berlin: Springer, 1999.

4 Stein A, Gorte B. Spatial statistics for remote sensing[M]. New York: Kluwer Academic publishers, 1999.

5 Sclove S C. Application of the conditional population mixture model to image segmentation [J]. IEEE. Trans. Pattern Analysis and Machine Intelligence, 1983, PAMI-5 (2): 428~ 433.

6 Dempster A P, Laird N M, Rubin D B. Maximum likelihood estimation from incomplete data via EM algorithm[J], J. R. Statist. Soc., 1977, B39(1): 1~ 38.

7 McLachlan G J, Basford K E. Mixture models: Inference and applications to clustering [M]. New York: Marcel Dekker, 1988.

8 Tadjudin S, Landgrebe D A. Robust parameter estimation for mixture model[J]. IEEE transactions on geo-science and remote sensing, 2000, 38(1): 439~ 445.

9 McLachlan G J, Krishnan T. The EM algorithm and extensions [M]. New York: John Wiley, 1977.

10 Moon T K. The expectation-maximization algorithm[J]. Signal Processing Mag., 1996, 13(6): 47~ 60.

骆剑承　1970 年生,博士,副研究员,现于资源与环境信息系统国家重点实验室从事基础研究. 主要研究方向包括空间数据挖掘、遥感图象处理、时空信息认知等. 发表论文 20 余篇.

周成虎　1964 年生,博士,研究员,现任中科院资源与环境信息系统国家重点实验室主任,中国科学院与香港中文大学"地理信息科学联合实验室"主任. 主要从事地球信息图谱、时空数据分析、遥感地学分析模型等研究工作,发表论文 70 余篇,出版专著与文集 10 多卷册.

梁　怡　博士,地理学讲座教授,香港中文大学地理系主任,资源与环境信息系统国家重点实验室学术委员会委员. 主要从事空间分析、空间决策、地理专家系统等方面研究.

马江洪　1963 年生,博士,西安长安大学副教授. 主要研究稳健统计、数据挖掘、模式识别等领域. 发表论文 20 余篇.