

基于随机点模式有限混合模型的 Gibbs 参数采样算法

王志, 刘伟峰, 黄梓龙, 丁禹心

杭州电子科技大学自动化学院, 杭州, 310018

摘要: 针对于随机点模式有限混合模型参数估计问题, 本文提出了一种基于随机点模式有限混合模型的 Gibbs 采样算法。首先, 基于随机有限集理论构建随机点模式有限混合模型, 其次通过 Gibbs 采样算法对混合模型的参数进行估计, 进一步利用 BIC 准则来判断算法模型拟合数据的真实程度。仿真实验以高斯混合分布作为随机点模式的特征分布, 仿真实验考虑了若干组随机点模式分布, 并与传统的有限混合分布算法进行了比较。结果表明, 本文提出的 RPP-FMM 的 Gibbs 采样算法相比较于传统 FMM 的 Gibbs 采样算法估计精度得到显著提高。

关键词: 随机点模式, RPP-FMM, Gibbs 采样, 有限混合模型

Gibbs parameter sampling algorithm based on finite mixture model of random point pattern

WANG Zhi, LIU Wei-feng, HUANG Zi-long, DING Yu-xin

School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China

Abstract: A parameter sampling algorithm for finite mixture model of random point pattern by using the Gibbs sampling is proposed. First, the random finite set theory is used to construct finite mixture model of random point pattern. Secondly, using Gibbs sampling algorithm to estimate the parameters of mixture model. Finally, using BIC criterion to judge the effect of algorithm model fitting real data. In the simulations Gaussian mixture distribution is used as the characteristic distribution of random point pattern, several groups of random point pattern distribution are considered and compared with the traditional finite mixed model algorithm. The results show that the RPP-FMM algorithm proposed in this paper is more accurate than the traditional FMM algorithm.

Key words: random point pattern, RPP-FMM, Gibbs sampling, finite mixture model

1 引言

有限混合模型(FMM)用于描述多个随机源产生的数据问题, 当随机源个数有限时, 每个随机源以一定的概率产生观测数据, 最终所获得的观测数据集符合该 FMM^[1-3], 例如最常见模型是高斯混合模型^[4-5]。对它的研究最早可以追溯到一个世纪之前, 1894 年 Pearson 采用两个单变量高斯混合模型对一组观测数据进行了拟合, 并采用矩阵估计方法估计混合分布的

参数。1977 年 Dempster 等把 FMM 看作为一个不完全数据(incomplete data)的参数估计问题,提出了计算不完全数据的极大似然估计方法,即著名的期望似然函数最大化方法(EM)^[6-7]。目前,有限混合模型方法在诸多的领域得到了应用,例如机器学习^[8],目标跟踪^[9],图像处理^[10],生物医疗^[11]等应用领域。

FMM 的研究主要包括两个方面:混合模型的分布元个数和相应的模型参数估计。实际获得数据大多具有非高斯特性,通常采用高斯混合模型进行近似。有限混合模型的参数估计可以通过参数的学习算法来获得。依据混合模型个数是否已知,可以把该学习算法分为有监督学习,无监督学习。按照学习算法的特点,可分为确定性学习算法和非确定学习算法。确定性学习算法以 EM 为代表的极大似然估计算法,非确定学习算法以 Markov 链为代表的 Bayes 学习算法^[12]。目前,传统 FMM 分布均假设各个数据点之间是独立的,因此,数据似然函数模型是所有数据点似然函数的乘积获得,这种数据似然函数无法刻画数据基数(数据点个数)的随机特性,甚至在某些情况下,会产生矛盾的估计结果。

在种情况下,引入随机点模式有限混合模型(RPP-FMM)更为准确。同时,随机点模式^[13-15]的参数估计也带来新难题,相较于传统 FMM 仅描述了数据的特征随机性,随机点模式分布函数还描述了数据的基数随机性。以往的相关研究中提出采用 EM 算法和 Gibbs 采样算法^[16-17]解决相关问题,EM 算法容易受初值影响;另外,由于 EM 算法属于一种确定性的算法,对于给定的初值,参数估计可能一直陷入局部极值点,不能得到全局极值点。Gibbs 采样算法属于随机采样算法,初值的影响相对较小。基于这种认识,本文提出了一种基于随机点模式有限混合模型(RPP-FMM)的 Gibbs 参数采样算法,基本思想是构造随机点模式的马尔可夫链来获取模型参数,进一步提高建模精度和参数估计精度;

2 背景和问题描述

基于随机点模式的有限混合分布条件下的采样问题不同于传统有限混合分布采样,因为随机点模式数据,相比较一般的传统点数据模型分布多出了一个基数分布,所以采样模型参数相应的难度较大。在本章中首先分析标准 Gibbs 采样算法,并对 RPP-FMM 进行描述,在此基础上把现有标准 Gibbs 推广到随机点模式条件下 Gibbs 采样问题。

2.1 Gibbs 采样

Gibbs 采样算法是利用马尔可夫链机制估计状态空间以生成样本的方法^[18-19]。可以看作是 MH 算法^[20-21]的特例，但解决了接受率为 1 的 MH 算法面临数据特征多，特征维度大而不适用问题。当难以直接从联合分布中采样，Gibbs 采样是对于高维空间的每一维进行采样，逐步逼近高维采样点。这种采样方法的采样难度相对较低，但增加了采样进入稳态时的采样次数。基于采样难度相对较低，本文采用基于 Gibbs 采用的方法，简单描述如下：假设我们需要从 N 维联合分布 $p(X_1, \dots, X_N)$ 中抽取 N 个样本，记第 n 个样本 $X_n = (x_1, \dots, x_i)$ 。

Gibbs 采样的过程则为：确定初始值 $X_n^{(0)}$ ，假设已知样本 $X_n^{(t)}$ ，记下一个样本为

$X_n^{(t+1)} = (x_1^{(t+1)}, x_2^{(t+1)}, \dots, x_i^{(t+1)})$ ，于是可将其看作成一个向量，对向量中的某一分量 $x_j^{(t+1)}$ ，可通过在其他分量已知的条件下利用分量的概率分布来采样该分量。对于该分量的条件概率可根据样本 $X_n^{(t+1)}$ 中已得到的分量 $x_1^{(t+1)}, \dots, x_{j-1}^{(t+1)}$ 以及上一个样本 $X_n^{(t)}$ 的分量 $x_{j+1}^{(t)}, \dots, x_i^{(t)}$ 得到，即 $p(x_j^{(t+1)} | x_1^{(t+1)}, \dots, x_{j-1}^{(t+1)}, x_{j+1}^{(t)}, \dots, x_i^{(t)})$ 。重复上述过程直到采样出样本 $X_n^{(t)}$ 的所有分量，这样我们就完成了一次 $X_n^{(t)}$ 到 $X_n^{(t+1)}$ 的转移，重复以上过程直到达到 $X_n^{(0)}$ 稳态，从而获取了一条 Markov 链。

2.2 随机点模式

随机点模式在自然界中普遍存在，比如城市的位置（地理学）（统计生态学），血液中的红细胞的位置（生物学），恒星和星系的位置（天文学）等等，参见图 1 场景所示：

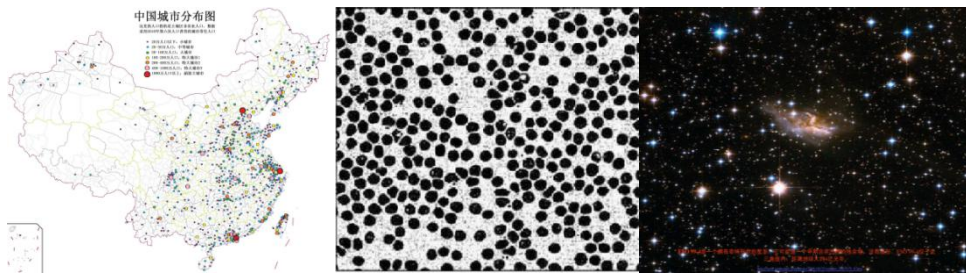


图 a 中国城市分布

图 b 血液中的红细胞

图 c 恒星和星系分布

图 1 随机点模式场景

随机点模式是无序点的集合或者(multi-set)多集合，多集合可以包含重复的元素，但多集可以被一个集合等价的表示。具体的说， N_1, \dots, N_m 分别为元素 x_1, \dots, x_N 的重复次数，

组成的多集的可以被集合 $\{(x_1, N_1), (x_2, N_2), \dots, (x_m, N_m)\}$ 表示。在特征空间 \mathcal{X} 的一个计数测度 N 被用来描述随机点模式 X ，对于每个属于特征空间中的集合 $A \subseteq \mathcal{X}$ ， $N(A)$ 表示随机点模式 X 分布在特征空间 \mathcal{X} 的紧子集 A 上点的个数。在形式上，一个点过程可以被定义为一个随机计数测度，随机计数测度 N 可以被视为依赖集合 $A \subseteq \mathcal{X}$ 的一个随机变 $N(A)$ 。如果一个点过程的实现总是简单且有限的，那么它就是简单且有限的点过程，在此我们只考虑在 R^d 的紧子集 \mathcal{X} 上的点过程。

在自然语言处理和信息检索中，点模式数据被定义为包(bag)^[22]，表示将每个文档视为单词的多集，数据的似然被认为是包中元素的联合密度。基于随机点模式数据的机器学习是一个新兴领域，通常称为多示例(MI)学习^[23-24]。由于随机点模式数据的丰富性和重要性，对于数学概念的似然函数的研究就变得非常重要，用于对点模式进行建模和分析的工具可以从点过程理论获得，并已在许多不同领域得到应用。

2.3 随机有限集

通常，随机点模式的概率密度可能不存在为了确保概率密度可用，本文针对简单有限的随机点模式进行研究，简单指的是随机点模式的集合中元素不重复，而有限指的是随机点模式的基数分布有限。简单且有限的随机点模式可等价于一个随机有限集^[25-26]。随机有限集具有基数分布和特征分布两大性质。随机变量在特征空间 \mathcal{X} 的有限子集空间 $\mathcal{F}(\mathcal{X})$ 中取值， $\mathcal{F}(\mathcal{X})$ 被称为随机有限集。 $X_n = \{x_1, \dots, x_i\} \in \mathcal{F}(\mathcal{X})$ 。通常关于特定测度 u 的随机有限集 $\mathcal{F}(\mathcal{X}) \rightarrow [0, \infty)$ ， u 的定义如下：

$$u(T) = \sum_{i=0}^{\infty} \frac{1}{i!U!} \int 1_T(\{x_1, \dots, x_i\}) d(x_1, \dots, x_i), \quad (1)$$

其中， U 是特征空间 \mathcal{X} 上的超体积单位， $1_T(\{x_1, \dots, x_i\})$ 是 \mathcal{T} 的指示函数，指示函数是定义在随机点模式 $\{x_1, \dots, x_i\}$ 上的函数，表示点模式有哪些元素属于子集 \mathcal{T} 。当 $x_i \in \mathcal{T}$ 时， $1_T(x_i) = 1$ ，否则 $1_T(x_i) = 0$ 。函数 $f(X_n)$ 关于特定的测度 u 的积分定义如下：

$$\int f(X)u(dX) = \sum_{i=0}^{\infty} \frac{1}{i!U^i} \int f(\{x_1, \dots, x_i\})d(x_1, \dots, x_i), \quad (2)$$

一个关于 u 在指定排序 $\{x_1, \dots, x_i\}$ 处的随机有限集的概率密度如下：

$$f(\{x_1, \dots, x_i\}) = p_c(i)i!U^i f_i(\{x_1, \dots, x_i\}) \quad (3)$$

其中， $p_c(i)$ 是关于 i 的基数分布函数， $f_i(x_1, \dots, x_i)$ 是一个对称函数，表示在给定基数 i 下随机点模式 $\{x_1, \dots, x_i\}$ 的联合概率密度，描述了基数信息以及特征之间的依赖关系。 U^i 抵消了概率密度的单位，从而避免了单位失配。

2.4 Poisson 点过程

Poisson 点过程^[27]是有限点过程之一，有限点过程的概率密度完全由其强度函数决定，Poisson 点过程模型完全由强度函数 $\lambda=\rho$ 和泊松基数分布唯一确定，泊松基数分布是由一个非负数字 ρ 描述，因此，在 Poisson 点过程模型基数分布的选择中只有一个自由度。假设随机点模式的元素之间互相独立，并且势分布满足强度为 ρ 的 Poisson 分布，那么随机点模式分布函数^[28]定义如下：

$$f(\{x_1, \dots, x_i\}) = \rho^i e^{-\rho} \prod_{l=1}^i U \cdot p_f(x_l), \quad (4)$$

其中， $p_f(x_l)$ 是随机点模式中独立且同分布的元素 x_l 的概率密度函数。则随机点模式似然函数 $f(X_{1:N} | \Theta)$ 定义如下：

$$f(X_{1:N} | \Theta) = \prod_{n=1}^N \sum_{k=1}^K \pi_k f(X_n | \theta_k) \quad (5)$$

其中， π_k 为混合权重， $f(X_n | \theta_k)$ 为随机点模式 X_n 的后验概率密度函数，其中 N 个随机点模式属于 K 个随机点模式类别。

3 混合模型 Bayes 分布

3.1 随机点模式有限混合模型

随机点模式有限混合模型常用于描述一组来自各种随机源的观测 $X = \{X_1, \dots, X_N\}$ ，

$f(X_n | \Theta)$ 为混合密度, 这里 Θ 是一个参数集。那么具有 K 个随机源的随机点模式混合模型可以表示为如下形式:

$$f(X_n | \Theta) = \pi_1 f(X_n | \theta_1) + \pi_2 f(X_n | \theta_2) + \cdots + \pi_K f(X_n | \theta_K), \quad (6)$$

其中, $\Theta = \{\pi_1, \dots, \pi_K, \theta_1, \dots, \theta_K\} \in (R_+ \times \Theta)^K$, $X_n \in \mathcal{F}(R^d)$ 是观测数据, $\{\theta_1, \dots, \theta_K\}$ 是随机点模式分布函数中的参数变量, $\{\pi_1, \dots, \pi_K\}$ 为相应的混合权重, π_k 是第 k 个分布元的混合权重。对于 N 个独立随机点模式特征数据 $X_n = \{x_{n,1}, \dots, x_{n,l_n}\}$, x_{n,l_n} 表示随机点模式 X_n 中第 l_n 个单点数据, 那么随机点模式的似然函数可以表示成:

$$L_{X_n}(\Theta) = L_{X_n}(\Theta | X_n) = \sum_{k=1}^K f(X_{1:N}, e_{1:N} | \pi_{1:K}, \theta_{1:K}) = \prod_{n=1}^N \sum_{k=1}^K \pi_k f(X_n | \theta_k), \quad (7)$$

其中 e_n 为缺失变量, 与点模式数据 X_n 组成完全数据 (X_n, e_n) , $X_{1:N} \triangleq \{X_1, X_2, \dots, X_N\}$, $e_{1:N} \triangleq \{e_1, e_2, \dots, e_N\}$, 根据上式 (7), 点模式混合模型的参数后验分布可描述为:

$$P(\Theta | X) = \frac{1}{C} L_X(\Theta) P(\Theta), \quad (8)$$

其中, $C = \int L_X(\Theta) p(\Theta) d(\Theta)$ 是正则常量, $P(\Theta)$ 是 RPP-FMM 参数的先验分布。对于每一个得到的单点观测数据 $x_{n,i} \in X_n$, 我们并不知道该观测是由哪个分布元产生。因此, 定义了一个 K 维的指示变量, 每一维指示了混合分布中的一个分布元。显然, 一个观测只能由一个分布元产生, 所以 K 维的缺失变量只有一维为 1, 其他各维为 0, 即表示为

$e_n = [e_{n,1}, \dots, e_{n,K}]^T$, 并且满足 $e_{n,k} \in \{0, 1\}$, $\sum_{k=1}^K e_{n,k} = 1$, 其中 $e_{n,k} = 1$ 表示数据 x_n 由分布元

$f(x_n | \theta_k)$ 产生。由于高斯混合分布具有良好的拟合性能, 本文以高斯混合分布作为点模式的特征分布为例, 相应点模式参数集为 $\theta_k = \{\rho_k, \mu_k, \Sigma_k\}$, 其中 ρ_k 是基数, μ_k 是均值, Σ_k 是协方差。

3.2 随机点模式有限混合模型参数先验分布

考虑公式 (8), 首先分析如何获得先验分布 $p(\Theta)$, 在高斯混合情况下, 先验参数为 $\bigcup_{k=1}^K \{\pi_k, \rho_k, \mu_k, \Sigma_k\}$, 模型参数先验是联合分布, 直接估计单个参数难度较大, 所以采用了条件分布的形式, 推导先验分布 $p(\Theta)$ 如下公式所示:

$$\begin{aligned} p(\Theta) &= p(\pi_{1:K}, \rho_{1:K}, \mu_{1:K}, \Sigma_{1:K}) = p(\pi_{1:K})p(\rho_{1:K}, \mu_{1:K}, \Sigma_{1:K} | \pi_{1:K}) \\ &= p(\pi_{1:K})p(\rho_{1:K} | \pi_{1:K})p(\Sigma_{1:K} | \rho_{1:K}, \pi_{1:K}, \mu_{1:K})p(\mu_{1:K} | \rho_{1:K}, \pi_{1:K}, \Sigma_{1:K}), \end{aligned} \quad (9)$$

(1) 混合权重

由于混合权重反映了各组观测个数的比例，通常采用分类分布作为混合权重的先验分布模型。因此，采用 Dirichlet 分布^[29]作为分类分布的共轭先验。如果各分布元比例未知，最简单的先验分布可以采用等值 Dirichlet 分布：

$$p(\pi_{1:K}) \sim \text{Dir}(\frac{1}{K}, \dots, \frac{1}{K}), \quad (10)$$

(2) 基数(强度)

$$p(\rho_{1:K}) = \prod_{k=1}^K p(l_k) = \prod_{k=1}^K \text{Poisson}(l_k), \quad (11)$$

RPP-FMM 的基数的先验分布服从泊松分布，此时的基数 l_k 就是泊松分布的强度。

(3) 协方差

$$p(\Sigma_{1:K}^{-1} | \rho_{1:K}, \pi_{1:K}) = W(V, \beta), \quad (12)$$

RPP-FMM 的协方差的逆矩阵的先验服从 Wishart 分布， $W(V, \beta)$ 表示一个 Wishart 分布，其中 V 是一个正定矩阵， β 是自由度。

(4) 均值

$$p(\mu_{1:K} | \Sigma_{1:K}) = N(\mu_{1:K}; \xi_{1:K}, \Sigma_{1:K}) \quad (13)$$

采用高斯分布作为 RPP-FMM 均值的共轭先验，其中 ξ_k 是第 k 个分布元的统计均值。

3.3 随机点模式有限混合模型参数后验分布

在随机点模式观测数据的似然函数中特征分布为高斯分布并且先验分布分别为式(10)，(11)，(12)，(13)的条件下，根据 Bayes 更新公式(8)可得：

$$p(\Theta | X) = \frac{g(X | \Theta)p(\Theta)}{f(X)} = \frac{1}{C} g(X | \Theta)p(\Theta) = \frac{1}{C} g(X | \pi_{1:K}, \theta_{1:K})p(\pi_{1:K}, \theta_{1:K}), \quad (14)$$

后验混合权重服从 Dirichlet 分布，后验均值服从高斯分布，后验方差服从 Wishart 分布。

混合权重 $\{\pi_1, \dots, \pi_K\}$ ，满足下面的 Dirichlet 分布：

$$p(\pi_1, \dots, \pi_K) = \text{Dir}(\alpha_1 + l_1, \alpha_2 + l_2, \dots, \alpha_K + l_K), \quad (15)$$

其中, $\alpha_k > 0$ 是先验分布的常量, l_k 是属于第 k 个随机分布元观测数据的个数。

缺失变量 $\{e_{n,1}, \dots, e_{n,K}\}$, 根据 Bayes 公式估计缺失数据如下:

$$\hat{e}_n = \frac{\pi_k f(X_n | \theta_k)}{\sum_{k=1}^K \pi_k f(X_n | \theta_k)}, \quad (16)$$

其中, $f(X_n | \theta_k)$ 为随机点模式 X_n 的后验概率密度函数如下式所示:

$$f(X_n | \theta_k) = \rho_k^{|X_n|} e^{-\rho_k} \prod_{l_n=1}^{|X_n|} U \cdot p_f(x_{l_n} | \theta_k), \quad (17)$$

基数分布:

$$\rho_k = \frac{\sum_{n=1}^N e_{n,k} |X_n|}{\sum_{n=1}^N e_{n,k}}, \quad (18)$$

其中, $e_{n,k}$ 表示随机点模式观测数据 X_n 的第 k 维指示变量, $|X_n|$ 是表示 X_n 中单点数据的个数, ρ_k 是第 k 的随机点模式的基数。

方差 Σ_k , 方差逆服从 Wishart 分布

$$p(\Sigma_k^{-1}) = W\left(\alpha_0 + l_k / M_0, \beta_0 + \kappa_k^2 / N_0\right), \quad (19)$$

$$\kappa_k^2 = \frac{\sum_{n=1}^N (x_k - \mu_k)(x_k - \mu_k)^T \hat{e}_{n,k}}{\sum_{n=1}^N \hat{e}_{n,k}}, \quad (20)$$

其中 α_0, β_0 是正常数, $M_0, N_0, \alpha_0, \beta_0$ 均是先验参数, 为先验分布的常量。 $M_0, N_0 > 0$ 起调节作用, κ_k^2 为二阶矩值。

均值 μ_k , 均值满足参数为 $\{\xi_k, \Sigma_k\}$ 的高斯分布, 它可以从下式中采样

$$p(\mu_k) = N(\mu_k; \xi_k, \Sigma_k), \quad (21)$$

$$\xi_k = \frac{\sum_{n=1}^N x_k \hat{e}_{n,k}}{\sum_{n=1}^N \hat{e}_{n,k}} \quad (22)$$

其中， ξ_k 是第 k 个分布元的统计均值。

4 结合 BIC 信息准则的模型估计算法

在混合模型的参数估计问题中，如何估计出分布元个数（或者模型阶次）是推断统计的重要内容之一，本节主要解决混合分布中分布元个数的估计问题。可逆跳变的 MCMC 方法采用模型阶次和参数同时采样的方法，利用极大后验准则的方法确定模型阶次，是一种非确定性随机方法。

本文提出的基于随机点模式有限混合模型的 Gibbs 参数采样算法是利用 Gibbs 采样算法和 BIC 信息准则^[31-32]相结合的思路。通过基于 Gibbs 采样的马尔科夫链蒙特卡罗方法，来近似后验分布 $p(\Theta | X)$ 。Gibbs 采样一般需要已知样本的一个属性在其他所有属性下的条件概率，然后利用该条件链式的推导出其他属性的样本值。因此，Gibbs 采样可以在给定的样本数据和各个参数的先验分布已知的条件下采样出参数的后验分布样本。推导出的马尔可夫链可以反映后验分布的概率特征，链中的稳定点往往是分布中的极值点，可以将其作为最终估计，Gibbs 采样算法由下述表 3.1 所示。本文在 Gibbs 采样的基础上，结合贝叶斯信息准则(BIC)来评价 RPP-FMM 和真实数据分布的匹配度，实现了用简单模型来表达更多的信息。BIC 信息准则定义如下：

$$BIC(m_k, \Theta_k, X_k) = -2 \log L(\Theta_k, m_k | X_k) + M_k \ln n_k, \quad (23)$$

其中， M_k 是独立参数的个数， $\log L(\Theta_k, m_k | X_k)$ 表示参数集 Θ_k 和元素个数 m_k 的对数似然函数。

$$M_k = 3m_k + 2 \quad (24)$$

$$L(\Theta_k, m_k | X_k) = \prod_{i=1}^{n_k} f(x_{k,i}, m_k) \quad (25)$$

在一组可供选择的模型类中，使 BIC 达到最小的那个模型是一个可取的模型， m_k 值和参数学习过程相关。模型的参数估计由下式推导：

$$\{m_k, \Theta_k\} = \arg \min_{\{m_k, \Theta_k\}} BIC(m_k, \Theta_k) \quad (26)$$

基于 BIC 准则和 RPP-FMM 的 Gibb 采样的估计算法的基础上，本文实现了对随机点模式有限混合模型的参数和分布元的估计学习。RPP-FMM 的 Gibb 采样算法，BIC 评价步骤如下所示：

① RPP-FMM 的 Gibb 采样算法

```

t = 0
for t=0 to 100
    t=t+1
     $(\pi_1^{(t)}, \dots, \pi_m^{(t)}) \sim Dir(\alpha_1 + \rho_1^{(t-1)}, \alpha_2 + \rho_2^{(t-1)}, \dots, \alpha_K + \rho_K^{(t-1)})$ 
    for k=1 to K
         $u_{n,k}^{(t)} = (\rho_k^{(t)})^{|X_n|} e^{-\rho_k^{(t-1)}} \cdot \prod_{l=1}^{|X_n|} U \cdot N(x_l; u_k^{(t-1)}, \Sigma_k^{(t-1)})$ 
        
$$e_{n,k}^{(t)} = \frac{\pi_k^{(t-1)} u_{n,k}^{(t)}}{\sum_{k=1}^K \pi_k^{(t-1)} u_{n,k}^{(t)}}$$

        
$$\rho_k^{(t)} = \sum_{n=1}^N e_{n,k}^{(t)}$$

        
$$\xi_k^{(t)} = \frac{\sum_{n=1}^N x_n e_{n,k}^{(t)}}{\sum_{n=1}^N e_{n,k}^{(t)}}$$

        
$$\kappa_k^{(t)2} = \sum_{n=1}^N \frac{(x_n - \xi_k^{(t)})(x_n - \xi_k^{(t)})^T e_{n,k}^{(t)}}{\rho_k^{(t)} = \sum_{n=1}^N e_{n,k}^{(t)}}$$

        
$$\Sigma_k^{(t)-1} \sim W(\alpha_0 + \frac{\rho_k^{(t)}}{2}, \beta_0 + \frac{\kappa_k^{(t)2}}{2})$$

        
$$u_k^{(t)} \sim N(\xi_k^{(t)}, \Sigma_k^{(t)})$$

    end for
end for

```

② BIC 算法

$$BIC(m_k, \Theta_k, X_k) = -2 \log L(\Theta_k, m_k | X_k) + M_k \ln n_k$$

输出： $\{\pi_k, u_k, \Sigma_k, \rho_k\}_{k=1}^K, m_k, BIC(m_k)$

通过上述算法，输出得到权重，均值，协方差，基数和 BIC 值。

表1 基于随机点模式的Gibbs采样算法

Gibbs 采样算法步骤:

- ◆ 初始化 $\theta^0 = (\theta_1^{(0)}, \dots, \theta_K^{(0)})$, $\theta_k = \{\rho_k, \mu_k, \Sigma_k\}$, 从条件密度 $p(\theta_i | \theta_{-i})$ 中采样
- ◆ 从 $p(\theta_1 | \theta_{-1}^{(0)})$ 中采样得到 $\theta_1^{(1)}$
- ◆
- ◆ 从 $p(\theta_K | \theta_{-K}^{(0)})$ 中采样得到 $\theta_K^{(1)}$
- ◆ 从而实现从 $\theta^0 = (\theta_1^{(0)}, \dots, \theta_K^{(0)})$ 到 $\theta^1 = (\theta_1^{(1)}, \dots, \theta_K^{(1)})$ 的跳变
- ◆ 重复上述的步骤, 得到一条 Markov 链

5 仿真实验对比与性能评估

为了验证本文算法的有效性, 本文给出两个实验算例: 三个随机点模式混合算例和五个随机点模式混合算例。在两个算例实验中, 重点分析基于随机点模式有限混合模型的 Gibbs 参数采样算法和 BIC 信息准则, 并与传统 FMM 的 Gibbs 采样算法性能做出对比。

5.1 三个点模式混合算例

见图 2 所示, 考虑 3 个混合高斯分布作为随机点模式的特征分布, 100 个随机点模式观测数据, 其中 60 个作为训练数据, 剩下的 40 个作为测试数据, 真实混合分布模型为:

$$f(X_n | \Theta) = \pi_1 f(X_n | \mu_1, \Sigma_1) + \pi_2 f(X_n | \mu_2, \Sigma_2) + \pi_3 f(X_n | \mu_3, \Sigma_3) \quad (28)$$

其中 $\pi_1 = 0.4, \pi_2 = 0.45, \pi_3 = 0.15$; $\mu_1 = [0; 10], \mu_2 = [3; 8], \mu_3 = [7; 11]$; $\rho_1 = 30, \rho_2 = 24, \rho_3 = 20$; $\Sigma_1 = \begin{bmatrix} 5 & -1 \\ -1 & 3 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 6 & -1 \\ -2 & 3 \end{bmatrix}$;

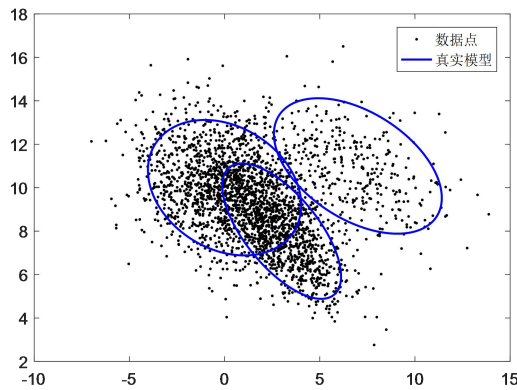


图 2 3 个高斯混合分布

设定的初始化参数取值为：

$$\{\pi_k^{(0)}\}_{k=1}^3 = 1/3; \{u_k^{(0)}\}_{k=1}^3 = [0; 0]; \{\Sigma_k^{(0)}\}_{k=1}^3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix};$$

一般情况下，Gibbs 采样会在 60~70 步进去稳态，本文采用 100 步的采样值作为估计值，对于不同的分布元个数，比较其稳态时的 BIC 优化准则。如图 3 所示分布元个数为 3 时，具有最小 BIC 值。图 4 描述了 BIC 信息准则的优化过程，我们取 100 步稳态迭代作为指标来判断，分布元个数从 2 变化到 5。可以看出，分布元为 3 时具有最小的 BIC 准则值，AIC 准则也几乎对应最小的 AIC 准则值，但是 AIC 准则在分布元个数大于 3 时变化不明显，判断效果远没有 BIC 准则好。分布元个数为 3 时，基于 RPP-FMM 的 Gibbs 采样估计算法估计的混合分布参数为：

$$\pi_1 = 0.4, \pi_2 = 0.46, \pi_3 = 0.14; u_1 = [0.1; 10], u_2 = [2.9; 8], u_3 = [6.9; 10.8];$$

$$\rho_1 = 29.7, \rho_2 = 24.6, \rho_3 = 21.5; \Sigma_1 = \begin{bmatrix} 5 & -1 \\ -1 & 3.1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 3.1 & -2 \\ -2.2 & 2.9 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 5.8 & -2 \\ -2 & 2.9 \end{bmatrix}。$$

基于传统数据 Gibbs 采样算法估计混合模型的参数为：

$$\pi_1 = 0.35, \pi_2 = 0.37, \pi_3 = 0.28; \mu_1 = [-0.7; 9.5], \mu_2 = [2.8; 9.5], \mu_3 = [5.9; 10.7];$$

$$\Sigma_1 = \begin{bmatrix} 4.6 & -1 \\ -1.1 & 2.8 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 3.2 & -2 \\ -2 & 3.1 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 7 & -3.5 \\ -2.7 & 3.5 \end{bmatrix}。$$

两种算法估计的混合模型和真实混合分布模型对比图分别如图 5，图 6 所示：

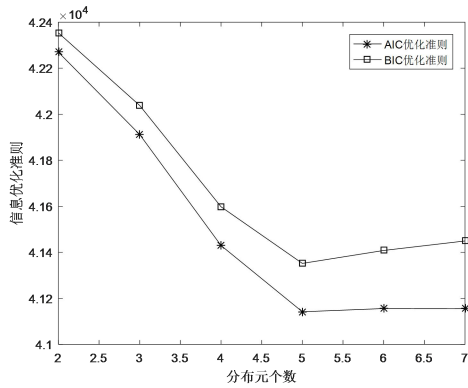


图 3 稳态时分布元个数-优化准则关系

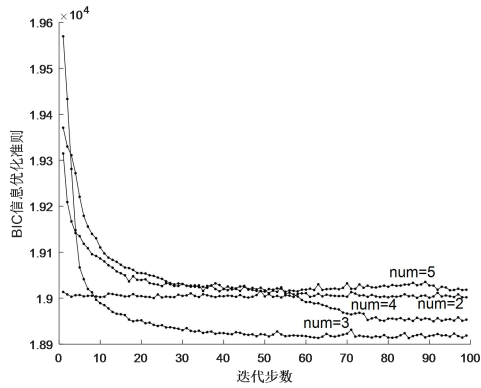


图 4 BIC 信息准则的优化过程

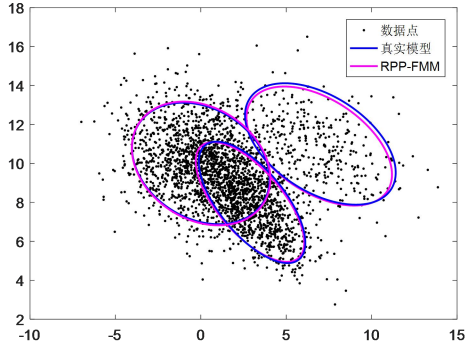


图5 本文算法

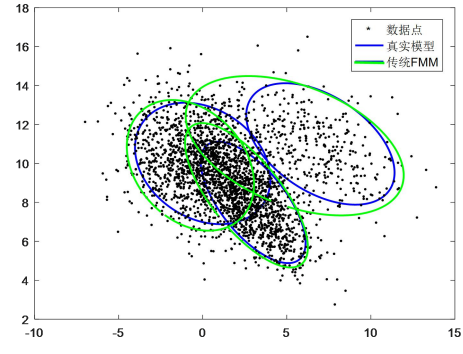


图6 传统 FMM 的 Gibbs 采样算法

5.2 五个点模式混合算

见图 7 所示，本算例中，我们考虑 5 个混合高斯作为随机点模式的特征分布，100 个随机点模式观测数据，其中 60 作为训练数据(带标签)，40 个作为测试数据，真实混合分布模型为：

$$f(X_n | \Theta) = \pi_1 f(X_n | \mu_1, \Sigma_1) + \pi_2 f(X_n | \mu_2, \Sigma_2) + \pi_3 f(X_n | \mu_3, \Sigma_3) + \pi_4 f(X_n | \mu_4, \Sigma_4) + \pi_5 f(X_n | \mu_5, \Sigma_5) \quad (29)$$

其中， $\pi_1 = 0.24, \pi_2 = 0.08, \pi_3 = 0.18, \pi_4 = 0.26, \pi_5 = 0.24; \mu_1 = [0; 8], \mu_2 = [5; 9], \mu_3 = [10; 7], \mu_4 = [2; 8], \mu_5 = [12; 8]; \rho_1 = 30, \rho_2 = 25, \rho_3 = 20, \rho_4 = 16, \rho_5 = 12; \Sigma_1 = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 1.5 \\ 1.5 & 3 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}, \Sigma_4 = \begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}, \Sigma_5 = \begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix}。$

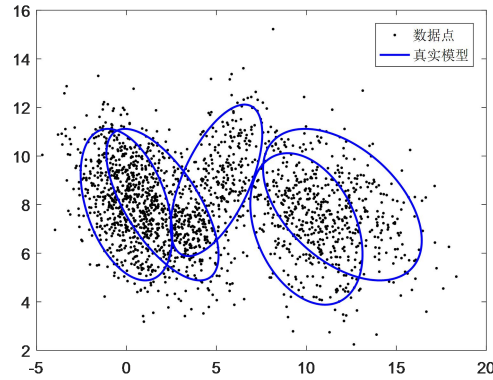


图7 5 个高斯混合分布

设定的初始化参数取值为： $\{\pi_k^{(0)}\}_{k=1}^5 = 1/5; \{\mu_k^{(0)}\}_{k=1}^5 = [0; 0]; \{\Sigma_k^{(0)}\}_{k=1}^5 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ，一般情况下，Gibbs 采样会在 60~70 步进去稳态，本文采用 100 步的采样值作为估计值，对于不同的分布元个数，比较其稳态是的 BIC 优化准则。如图 8 所示分布元个数为 5 时，具有最

小的 BIC 值。图 9 描述了 BIC 信息准则的优化过程，采用 100 步稳态迭代作为指标来判断，分布元个数从 2 变化到 7。可以看出，分布元为 5 时具有最小的 BIC 准则值，AIC 准则也几乎对应最小的 AIC 准则值，但是 AIC 准则在分布元个数大于 5 时变化十分的不明显，判断效果远没有 BIC 准则好。分布元个数为 5 时，基于 RPP-FMM 的 Gibbs 采样估计算法估计的混合分布参数为：

$$\pi_1=0.23, \pi_2=0.1, \pi_3=0.17, \pi_4=0.28, \pi_5=0.22; \mu_1=[0; 7.9], \mu_2=[5; 8.8],$$

$$\mu_3=[10.3; 7.4], \mu_4=[2; 7.9], \mu_5=[12.3; 7.8]; \rho_1=29.5, \rho_2=26.3, \rho_3=19.2, \rho_4=16.5, \rho_5=12.3;$$

$$\Sigma_1=\begin{bmatrix} 2 & -1 \\ -1 & 2.9 \end{bmatrix}, \Sigma_2=\begin{bmatrix} 2.1 & 1.5 \\ 1.6 & 3.2 \end{bmatrix}, \Sigma_3=\begin{bmatrix} 3.2 & -1 \\ -1 & 3.1 \end{bmatrix}, \Sigma_4=\begin{bmatrix} 3.1 & -2.1 \\ -2.1 & 3 \end{bmatrix}, \Sigma_5=\begin{bmatrix} 6.2 & -2 \\ -2.1 & 2.8 \end{bmatrix}$$

基于传统数据 Gibbs 采样算法估计混合模型的参数为：

$$\pi_1=0.19, \pi_2=0.17, \pi_3=0.15, \pi_4=0.23, \pi_5=0.26;$$

$$\mu_1=[0.2; 7.6], \mu_2=[4.9; 9.3], \mu_3=[9.6; 6.7], \mu_4=[0.3; 8.2], \mu_5=[10.5; 7.8];$$

$$\Sigma_1=\begin{bmatrix} 1.7 & -1 \\ -1 & 2.5 \end{bmatrix}, \Sigma_2=\begin{bmatrix} 2.5 & -0.6 \\ -0.4 & 2.7 \end{bmatrix}, \Sigma_3=\begin{bmatrix} 1.3 & -1.2 \\ -1.4 & 1 \end{bmatrix}, \Sigma_4=\begin{bmatrix} 2.5 & -1.1 \\ -1 & 2.7 \end{bmatrix}, \Sigma_5=\begin{bmatrix} 3.2 & -1 \\ -1 & 2.9 \end{bmatrix}$$

。两种算法估计的混合模型和真实混合分布模型对比图分别如图 10，图 11 所示：

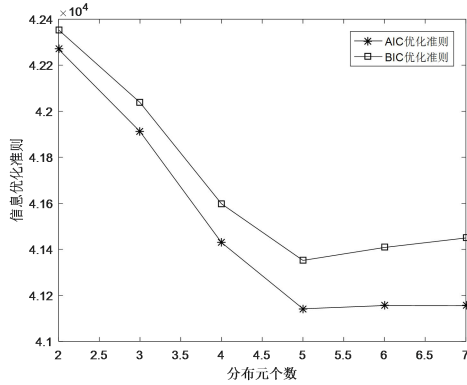


图 8 稳态时分布元个数-优化准则关系

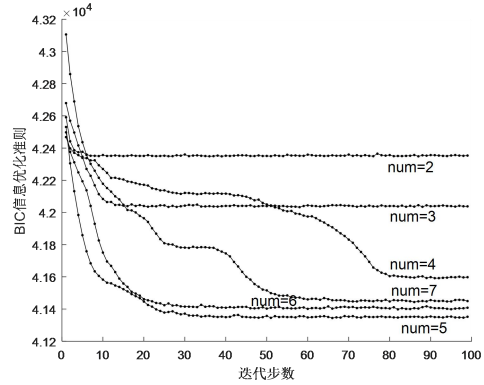


图 9 BIC 信息准则的优化过程

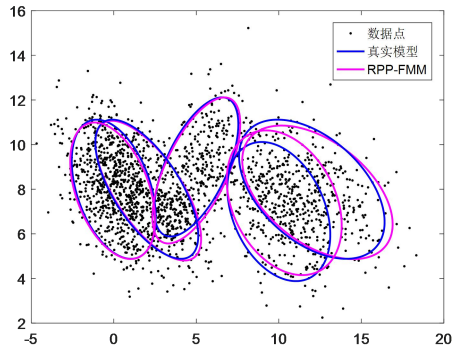


图 10 本文算法

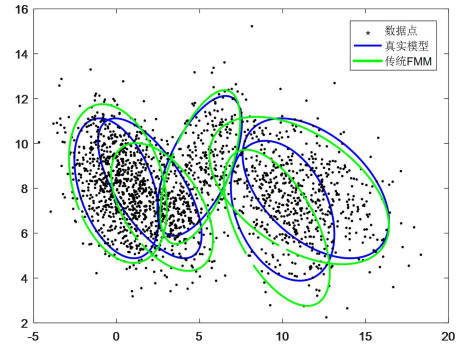


图 11 传统 FMM 的 Gibbs 采样算法

根据两个算例实验结果可以得知，当点模式出现交叠情况时，基于传统 FMM 的 Gibbs 采样算法模型和真实混合分布模型的参数误差较大。由于模型描述更为准确，并且考虑了模型训练的过程，因此，基于本文 RPP-FMM 的 Gibbs 采样估计算法模型与真实混合分布模型的参数非常相近，误差相对较小，性能远远好于传统算法。其中产生相关参数的误差原因主要有三点：（1）样本数据有限 （2）随机算法本身的随机误差 （3）数值计算产生的计算误差。

5 结论

针对随机点模式学习问题，本文首先结合随机点模式的分布模型，给出了随机点模式的有限混合模型，进一步提出了具有点模式的 Gibbs 采样算法，获取点模式的随机源个数估计和分布参数估计。其中引入 BIC 准则来评价 RPP-FMM 和真实数据分布的匹配度，确定点模式随机源个数，点模式的 Gibbs 采样算法估计特征分布参数。参数估计过程包括两个步骤：模型训练过程和模型测试过程。实验仿真结果表明该算法估计 RPP-FMM 参数远远精确于传统 FMM 的 Gibbs 采样算法。算法可用于模式识别与分类，聚类算法，模式检测等研究。

参考文献：

- [1] Deb P . Finite Mixture Models[J]. Partha Deb, 2000, 39(4):521–541.
- [2] Hang Li, Kenji Yamanishi. Topic analysis using a finite mixture model[J]. Information Processing & Management, 39(4):511-542.
- [3] 刘伟峰, 杨爱兰. 基于 BIC 准则和 Gibbs 采样的有限混合模型无监督学习算法[J]. 电子学报, 2011, 39(s1):134-139.
- [4] Constantinopoulos C , Titsias M K , Likas A . Bayesian Feature and Model Selection for Gaussian Mixture Models[J]. IEEE Trans Pattern Anal Mach Intell, 2006, 28(6):1013-1018.
- [5] Pernkopf F , Bouchaffra D . Genetic-based EM algorithm for learning Gaussian mixture

- model[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8):1344-1348.
- [6] Heitjan D F. Book Review: The EM Algorithm and Extensions. Geoffrey J. McLachlan and Thriyambakam Krishnan, Wiley, New York, 1997. No. of pages: xvii+274. Price: \$59.95. ISBN 0-471-12358-7[J]. Statistics in Medicine, 1998, 17(10):1187-1187.
- [7] 邢长征, 苑聪. 一种快速、贪心的高斯混合模型 EM 算法研究[J]. 计算机工程与应用, 2015(20):115-119.
- [8] 黄诗华, 陈一民, 陆意骏,等. 基于机器学习的自然特征匹配方法[J]. 计算机工程, 2010, 36(20):182-184.
- [9] Lian F , Han C Z , Liu W F , et al. Joint spatial registration and multi-target tracking using an extended PM-CPHD filter[J]. Science China(Information Sciences), 2012(03):15-25.
- [10] 韩守东, 赵勇, 陶文兵, et al. 基于高斯超像素的快速 Graph Cuts 图像分割方法[J]. 自动化学报, 2011(01):13-22.
- [11] 刘立芳. 生物信息学中的多序列比对与模体识别问题研究[D]. 西安电子科技大学, 2006(03):28-32.
- [12] R. Cogill, "Primal-dual algorithms for discounted Markov decision processes," 2015 European Control Conference (ECC), Linz, 2015, pp. 260-265.
- [13] Amores, Jaume. Multiple instance classification: Review, taxonomy and comparative study[J]. Artificial Intelligence, 2013, 201:81-105.
- [14] MICHAEL F. DACEY. Order Distance in an Inhomogeneous Random Point Pattern[J]. Canadian Geographer, 2008, 9(3):144-153.
- [15] B. Vo, N. Tran, D. Phung and B. Vo, "Model-based classification and novelty detection for point pattern data," 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, 2016, pp. 2622-2627.
- [16] Review by: Debashis Kushary. The EM Algorithm and Extensionsby Geoffrey J. McLachlan; Thriyambakam Krishnan[J]. Technometrics, 1998, 40(3):260.
- [17] JORDAN,M. I. Hierarchical mixtures of experts and the EM algorithm[J]. Neural Computation, 1994,42(06):25-31.
- [18] George Casella and Edward I. George. Explaining the Gibbs Sampler[J]. American Statistician, 1992,46(3):167-174.
- [19] 刘伟峰, 杨爱兰. 基于 BIC 准则和 Gibbs 采样的有限混合模型无监督学习算法[J]. 电子学报, 2011, 39(a03):134-139.
- [20] 曾惠芳, 朱慧明, 李素芳, 等. 基于 MH 算法的贝叶斯分位自回归模型[J]. 湖南大学学报(自然科学版), 2010, 37(2):88-92.

- [21] Chib S , Greenberg E . Understanding the Metropolis-Hastings Algorithm[J]. The American Statistician, 1995, 49(4):327-335.
- [22] Boris Babenko, Ming-Hsuan Yang, Serge Belongie. Robust Object Tracking with Online Multiple Instance Learning[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 33(8):1619-1632.
- [23] Melendez J , Van Ginneken B , Maduskar P , et al. A Novel Multiple-Instance Learning-Based Approach to Computer-Aided Detection of Tuberculosis on Chest X-Rays[J]. IEEE Transactions on Medical Imaging, 2015, 34(1):179-192.
- [24] 黎铭, 薛晓冰, 周志华. 基于多示例学习的中文 Web 目录页面推荐[J]. 软件学报, 2004, 15(9):1328-1335.
- [25] Vo B N , Singh S , Doucet A . Sequential Monte Carlo methods for multitarget filtering with random finite sets[J]. IEEE Transactions on Aerospace and Electronic Systems, 2005, 41(4):1224-1245.
- [26] John Mullane, Ba-Ngu Vo, M.D. Adams, et al. A random-finite-set approach to Bayesian SLAM[J]. IEEE Transactions on Robotics, 2011, 27(2):268-282.
- [27] Sergio Verdú. Asymptotic error probability of binary hypothesis testing for Poisson point-process observations (Corresp.)[J]. IEEE Transactions on Information Theory, 1986, 32(1):113-115.
- [28] Vo, Ba-Ngu, Phung, Dinh, Tran, Quang N. Model-Based Multiple Instance Learning[J]. IEEE Transactions on Aerospace and Electronic Systems, 2017, 44(3):1321-1344.
- [29] Ya Xue, Xuejun Liao, Lawrence Carin, et al. Multi-Task Learning for Classification with Dirichlet Process Priors[J]. Journal of Machine Learning Research, 2007, 8(1):35-63.
- [30] Marie-France Bru. Wishart Processes[J]. Journal of Theoretical Probability, 1991, 4(4):725-751.
- [31] 杨文, 颜卫, 涂尚坦, 等. 基于贝叶斯信息准则的极化干涉 SAR 图像非监督分类[J]. 电子与信息学报, 2012, 34(11).
- [32] Kenneth P Burnham, David R Anderson. Multimodel Inference: understanding AIC and BIC in Model Selection[J]. Sociological Methods & Research, 2004, 33(33):261-304.