

Mean Shift 算法的收敛性分析*

文志强^{1,2+}, 蔡自兴¹

¹(中南大学 信息科学与工程学院,湖南 长沙 410083)

²(湖南工业大学 计算机科学与技术系,湖南 株洲 412008)

Convergence Analysis of Mean Shift Algorithm

WEN Zhi-Qiang^{1,2+}, CAI Zi-Xing¹

¹(School of Information Science and Engineering, Central South University, Changsha 410083, China)

²(Department of Computer Science and Technology, Hu'nan University of Technology, Zhuzhou 412008, China)

+ Corresponding author: Phn: +86-731-2655993, E-mail: zhqwen20001@163.com

Wen ZQ, Cai ZX. Convergence analysis of Mean Shift algorithm. *Journal of Software*, 2007,18(2):205–212.
<http://www.jos.org.cn/1000-9825/18/205.htm>

Abstract: The research of its convergence of Mean Shift algorithm is the foundation of its application. Comanicu and Li Xiang-ru have respectively provided the proof for the convergence of Mean Shift but they both made a mistake in their proofs. In this paper, the imprecise proofs existing in some literatures are firstly pointed out. Then, the local convergence is proved in a new way and the condition of convergence to the local maximum point is offered. Finally, the geometrical counterexamples are provided for explanation about convergence of Mean Shift and the conclusion is further discussed. The results of this paper contribute to further theoretical study and extensive application for Mean Shift algorithm.

Key words: Mean Shift algorithm; convergence; kernel function; kernel density estimation; gradient ascent method

摘要: 作为迭代算法,Mean Shift 的收敛性研究是应用的基础,而 Comanicu 和李乡儒分别证明了 Mean Shift 的收敛性,但证明过程存在错误.首先指出了 Comanicu 和李乡儒的证明过程存在错误;然后,从数学上重新证明了 Mean Shift 算法的局部收敛性,并指出其收敛到局部极大值的条件;最后,从几何上举反例分析了 Mean Shift 的收敛性,并进行了深入比较和讨论.这为 Mean Shift 算法的深入研究及应用奠定了基础.

关键词: Mean Shift 算法;收敛性;核函数;核密度估计;梯度上升方法

中图法分类号: TP301 文献标识码: A

Mean Shift 算法是由 Fukunaga 和 Hostetler 在 1975 年提出来的,直到 1995 年,Cheng^[1]的论文发表才引起人们的研究兴趣,掀起了研究和应用 Mean Shift 算法的热潮.作为一种有效的统计迭代算法,近年来,Mean Shift 算法已广泛应用于计算机视觉与模式识别等领域,例如目标跟踪^[2-4]、图像分割^[5,6]、模式识别与聚类分析^[7,8]、滤

* Supported by the National Natural Science Foundation of China under Grant Nos.60234030, 60404021 (国家自然科学基金); the Basic Research Project of the 11th Five-Year-Plan of China under Grant No.A1420060159 (国家“十一五”基础研究项目); the Academician Foundation Project of Hu'nan Province of China under Grant No.06IJJY3035 (湖南省院士基金项目)

Received 2005-11-28; Accepted 2006-04-27

波^[9]、信息融合^[10]、特征空间分析^[11]等.除了 Mean Shift 算法的应用研究以外,Mean Shift 算法的基础研究主要集中在以下 3 个方面:

(1) Mean Shift 算法的形式.学者们提出了不同的 Mean Shift 算法,如 Cheng^[11]提出了如下 Mean Shift 算法:

$$y_{t+1} = \sum_{i=1}^n w_i k(\beta \|x_i - y_t\|^2) x_i / \sum_{i=1}^n w_i k(\beta \|x_i - y_t\|^2).$$

$k(x)$ 是文献[1]定义的轮廓函数.文献[11]的 Mean Shift 算法在上式中引入了窗宽(又称带宽 bandwidth)参数 h ,表示如下:

$$y_{t+1} = \sum_{i=1}^n x_i g\left(\left\|\frac{y_t - x_i}{h}\right\|^2\right) / \sum_{i=1}^n g\left(\left\|\frac{y_t - x_i}{h}\right\|^2\right).$$

文献[12]在 Mean Shift 算法中引入权重参数,提出了如下 Mean Shift 算法:

$$y_{t+1} = \sum_{i=1}^n w_i c_{k,i,h} g\left(\left\|\frac{y_t - x_i}{h}\right\|^2\right) x_i / \sum_{i=1}^n w_i c_{k,i,h} g\left(\left\|\frac{y_t - x_i}{h}\right\|^2\right).$$

其中: $g(x)=-k'(x)$; $c_{k,i,h} = 1/\int k((x-x_i)^T \sum_i^{-1}(x-x_i)/h^2) dx$; $h>0$,为常数; $k(x)$ 是核函数.实际上,文献[12]的核函数和文献[1,11]的轮廓函数都满足相同的条件:函数 $k(x)$ 在 $[0,\infty)$ 上非负,单调递减、连续并且 $\int_0^\infty k(x)dx < +\infty$. 这样的

函

数有很多,比较常用的函数有: 高斯核函数 $k(x)=e^{-x}, x \geq 0$; 截断的高斯函数; 单位平坦(unit flat)函数

$$k(x) = \begin{cases} 1, & \text{if } |x| \leq 1 \\ 0, & \text{if } |x| > 1 \end{cases}.$$

本文沿用文献[12]的核函数概念.

(2) 窗宽的选择.对于 Mean Shift 算法的应用来说,窗宽的选择很重要.核窗宽的大小起着非常重要的作用.它决定了参与 Mean Shift 迭代的样本数量,直接影响迭代速度.有许多学者对此进行了大量的研究,得到一些好的成果,可参考文献[13-15].

(3) 收敛性. Mean Shift 算法作为一种迭代算法,其收敛性的研究是非常重要的.Cheng^[11]讨论了 3 种形式的 Mean Shift 算法,其中主要对模糊过程(blurring process)作了理论研究.假设有限样本集 S 的数据是 d 维 Euclidean 空间 $X, T \subset X$ 为迭代过程中产生的有限数据集.当 $T=S$ 时,则该 Mean Shift 算法为模糊过程,即每迭代一次后循环修改输入数据.为此,Cheng 给出了数据半径和直径的概念,并证明其数据直径的收敛速度及模糊过程经过有限次迭代后到达一个稳定点.然而,在整个 Mean Shift 算法迭代过程中,当 S 固定、 T 初始化为 S 时就不再是 blurring process,超出了 Cheng 的理论范围.文献[3,11]从密度估计方法出发推导出 Mean Shift 算法,并证明了它的单调递增收敛性.文献[14]证明了可变带宽 Mean Shift 算法的收敛性.在 Mean Shift 算法迭代过程中, S 固定, T 不等于 S ,这恰好是 Cheng 未完成的理论.文献[12]发现文献[11]在证明 Mean Shift 算法单调递增收敛性时,其关键步骤存在失误,并举反例进行了说明;在 Comanicu 的文献[3,14]中也存在着相同的问题.针对这些问题,文献[12]在 Mean Shift 算法中引入权重参数,并重新从数学上严格地证明了更一般意义下的 Mean Shift 算法的收敛性.事实上,文献[3,11,12,14]的单调递增收敛性证明的关键步骤都存在错误.文献[12]给出的收敛条件是:核函数 $k(x)$ 是凸的,满足 $k(x_2)-k(x_1) > k'(x_1)(x_2-x_1), x_1 > 0, x_2 > 0, k'(x_1)$ 为 $k(x)$ 在 x_1 的导数且连续、有界,该条件的实质是 $k(x)$ 为凸函数;文献[3,11,14]的假设条件是: $k(x)$ 为凸的且单调递减的轮廓函数,其实质也是 $k(x)$ 为凸函数.文献[3,11,12,14]在证明 $\hat{f}_{h,k}(y_t), t=1,2,\dots$ 序列收敛并且严格单调递增时,利用 $k(x)$ 的凸函数性质来进行单调性证明,但其实际的表达形式为 $k(\|y_j-x_i\|^2)$,而 $k(\|y_j-x_i\|^2)$ 不一定是凸函数,可能为凹函数或两者都不是,而当 $k(x)$ 为凸函数时,序列 $\hat{f}_{h,k}(y_t)$ 也不一定能收敛到局部极大值.

本文针对 Mean Shift 算法的收敛性,首先指出文献[11,12]在 $\hat{f}_{h,k}(y_t), t=1,2,\dots$ 序列收敛并且严格单调递增的证明过程中存在错误,并从理论上进行了证明;然后,从数学上重新证明了 Mean Shift 算法的收敛性,指出其收敛

到局部极大值的条件;最后,从几何上分析了本文结论的正确性,为 Mean Shift 算法的深入研究及应用奠定了基础.

本文第 1 节给出 Mean Shift 算法.第 2 节指出文献[11,12]的错误,并重新给出 Mean Shift 算法的收敛性证明.第 3 节讨论第 2 节的结论.第 4 节为全文的结束语.

1 Mean Shift 算法

核密度估计(又名 Parzen 窗估计)是一个很流行的密度估计方法,给定在 d 维空间 R^d 的 n 个样本数据 $x_i, i=1, \dots, n$, 多维变量的核密度估计可写成

$$\hat{f}_{h,k}(x) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^n |H|^{-1/2} k\left(\left\|\frac{x-x_i}{H}\right\|^2\right),$$

其中: $c_{k,d}$ 为归一化常数; $k(x)$ 在文献[12]中称为核函数,在文献[11]中称为轮廓函数(profile); H 为 $d \times d$ 维的带宽矩阵(bandwidth matrix).完整的参数表示 H 会增加估计的复杂性,在实际中, H 可以为对角矩阵 $H = \text{diag}[h_1^2, \dots, h_d^2]$ 或为 $h^2 I, I$ 为 $d \times d$ 单位矩阵.为了简单起见,本文使用后一种 H ,则核密度估计为

$$\hat{f}_{h,k}(x) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^n k\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \tag{1}$$

其中,核函数 $k(x)$ 满足下列条件:

- 1) $k(x)$ 为非负;
- 2) $k(x)$ 非单调递增:假如 $a < b$,则 $k(a) \geq k(b)$;
- 3) $k(x)$ 为连续函数且有界: $\int_0^\infty k(t)dt < \infty$.

当 $k(x)(x > 0)$ 可微时,从式(1)可得到的梯度:

$$\nabla \hat{f}_{h,k}(x) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^n (x-x_i)k'\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \tag{2}$$

定义 $g(x) = -k'(x)$,式(2)中引入 $g(x)$ 变为

$$\nabla \hat{f}_{h,k} = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^n (x_i - x)g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) = \frac{2c_{k,d}}{nh^{d+2}} \left[\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \right] \cdot \left[\frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \right] \tag{3}$$

式(3)的第 2 部分就是 Mean Shift 算法,设

$$\nabla m_{h,g}(x) = \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x, \quad \hat{f}_{h,g}(x) = \frac{c_{g,d}}{nh^d} \sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right),$$

其中, $c_{g,d}$ 为归一化常数,则式(3)可变为

$$\nabla \hat{f}_{h,k}(x) = \hat{f}_{h,g}(x) \frac{2c_{k,d}}{h^2 c_{g,d}} \nabla m_{h,g}(x),$$

即

$$\nabla m_{h,g}(x) = \frac{1}{2} h^2 c \frac{\nabla \hat{f}_{h,k}(x)}{\hat{f}_{h,g}(x)} \tag{4}$$

其中, $c = c_{g,d}/c_{k,d}$.式(4)说明局部均值朝附近数据样本密集区域移动,所以有迭代公式:

$$y_{t+1} = y_t + \nabla m_{h,g}(y_t) \quad (5)$$

式(5)经过一些转换得到 Mean Shift 算法的迭代式(6),具体过程见算法 1.

$$y_{t+1} = \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{y_t - x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{y_t - x_i}{h}\right\|^2\right)} \quad (6)$$

式(4)代入式(5),得到

$$y_{t+1} = y_t + \lambda_t d_t \quad (7)$$

其中: $\lambda_t = h^2 c / 2 \hat{f}_{h,g}(y_t) > 0$; $d_t = \nabla \hat{f}_{h,k}(y_t)$. 式(7)表明, Mean Shift 算法沿梯度方向迭代,使每个待处理的点“漂移”到分布密度函数的局部极大值点处,其步长 λ_t 随迭代过程自适应地变化:在低密度区,迭代步长较长;在局部极大值附近,迭代步长较短.

算法 1.

- 1) 设置初始值 y_0 , 结束条件 ε ;
- 2) 用式(6)计算 y_{t+1} 的值;
- 3) 判断是否满足 $|y_{t+1} - y_t| \leq \varepsilon$, 如果满足, 则退出;
否则, 用 y_{t+1} 替代 y_t , 转至 2).

2 Mean Shift 算法的收敛性

文献[11,12]在证明 $\hat{f}_{h,k}(y_t)$, $t=1,2,\dots$ 序列收敛并且严格单调递增时,假设其核函数 $k(x)$ 是凸的,在其证明过程中利用凸函数性质: $k(x_2) - k(x_1) > k'(x_1)(x_2 - x_1)$, $x_1 > 0, x_2 > 0$ (这里, x_1, x_2 为单个变量,其他情况为向量).但是,在证明过程中,核函数的形式为 $k(\|y_j - x_i\|^2)$,这时, $k(\|y_j - x_i\|^2)$ 不一定是凸的,可能是凹的或两者都不是(x_i 为常数).下面给出严格的证明,有关凸集、凸函数、凹函数的定义及相关性质请参考文献[16]或其他相关文献.

定理 1. 如果函数 $k(x)$ 是凸函数,则核函数 $k(\|x-c\|^2)$ 是凸的,或凹的,或两者都不是,其中, c 为常数向量.

证明:函数 $k(x)$ 在定义域内的 x_1 处用 Taylor 公式二阶展开,然后用 x_2 代替 x (不妨设 $x_1 < x_2$),得到:

$$k(\|x_2 - c\|^2) = k(\|x_1 - c\|^2) + \nabla k(\|x_1 - c\|^2)^T (x_2 - x_1) + (x_2 - x_1)^T \nabla^2 k(\|\theta\|^2) (x_2 - x_1) / 2,$$

其中, $x_1 < \theta < x_2$.

当 $\nabla^2 k(\|\theta\|^2)$ 为半正定矩阵时, $k(\|x_2 - c\|^2) \geq k(\|x_1 - c\|^2) + \nabla k(\|x_1 - c\|^2)^T (x_2 - x_1)$, 则 $k(\|x - c\|^2)$ 为凸函数;

当 $\nabla^2 k(\|\theta\|^2)$ 为负定矩阵时, $k(\|x_2 - c\|^2) < k(\|x_1 - c\|^2) + \nabla k(\|x_1 - c\|^2)^T (x_2 - x_1)$, 则 $k(\|x - c\|^2)$ 为凹函数;

当 $\nabla^2 k(\|\theta\|^2)$ 为其他情况时, $k(\|x - c\|^2)$ 既不是凸函数也不是凹函数.

根据定理 1,文献[11,12]的单调性证明是错误的,原因是不能使用 $k(x)$ 为凸函数的条件来证明 Mean Shift 算法收敛.因此我们可以看出, Mean Shift 算法的收敛应是有条件的收敛.下面从数学上严格证明文献[11,12]的单调收敛证明是错误的.

定理 2. 文献[11,12]的定理(文献定理[11]中的定理为:如果核函数是凸的,则 $\hat{f}_{h,k}(y_t)$, $t=1,2,\dots$ 序列收敛并且严格单调递增;文献[12]有类似的定理)证明过程是错误的.

证明:可参考文献[11,12]中的详细证明过程.假设文献[11,12]的证明过程成立,则序列 $\hat{f}_{h,k}(y_t)$ 收敛并且严格单调递增的定理成立.由文献[11,12]知 y_t 收敛,设收敛于 \bar{y} , 则有 $\nabla \hat{f}_{h,k}(\bar{y}) = 0$.

文献[12]证明时根据凸函数的性质: $k(x_2) - k(x_1) > k'(x_1)(x_2 - x_1)$, 有下式成立(为了更清楚证明过程,这里使用原文的表达式):

$$\begin{aligned} \hat{f}(y_{j+1}) - \hat{f}(y_j) &= \sum_{i=1}^n w_i c_{k,i,h} [k(\|y_{j+1} - x_i\|_{H_i}^2) - k(\|y_j - x_i\|_{H_i}^2)] \\ &\geq \sum_{i=1}^n w_i c_{k,i,h} k'(\|y_j - x_i\|_{H_i}^2) \cdot [\|y_{j+1} - x_i\|_{H_i}^2 - \|y_j - x_i\|_{H_i}^2]. \end{aligned}$$

这里,无意识地认为 $k(\|y_{j+1} - x_i\|_{H_i}^2)$ 是凸函数(文献[11]的证明与文献[12]类似),写成本文表达方式: $k(\|(y - x_i)/h\|^2)$ 为凸函数.根据凸函数的性质, $\hat{f}_{h,k}(y_i)$ 为凸函数且 $\nabla^2 \hat{f}_{h,k}(y_i)$ 为半正定矩阵.

对 $\hat{f}_{h,k}(y)$ 在 \bar{y} 处用二阶 Taylor 公式展开,得到:

$$\hat{f}_{h,k}(y) = \hat{f}_{h,k}(\bar{y}) + \nabla \hat{f}_{h,k}(\bar{y})^T (y - \bar{y}) + \frac{1}{2} (y - \bar{y})^T \nabla^2 \hat{f}_{h,k}(\hat{y}) (y - \bar{y}), \hat{y} = \bar{y} + \eta(y - \bar{y}), 0 < \eta < 1.$$

取 $y=y_t$,得到:

$$\hat{f}_{h,k}(y_t) - \hat{f}_{h,k}(\bar{y}) = \frac{1}{2} (y_t - \bar{y})^T \nabla^2 \hat{f}_{h,k}(\hat{y}) (y_t - \bar{y}).$$

因为 $\hat{f}_{h,k}(\bar{y}) > \hat{f}_{h,k}(y_t)$,故 $(y_t - \bar{y})^T \nabla^2 \hat{f}_{h,k}(\hat{y}) (y_t - \bar{y}) < 0$,

则有 $\lim_{\eta \rightarrow 1} (y_t - \bar{y})^T \nabla^2 \hat{f}_{h,k}(\hat{y}) (y_t - \bar{y}) = (y_t - \bar{y})^T \nabla^2 \hat{f}_{h,k}(y_t) (y_t - \bar{y}) < 0$ (考虑连续性).

所以,存在 y_t ,使得 $\nabla^2 \hat{f}_{h,k}(y_t)$ 为负定矩阵.这与 $\nabla^2 \hat{f}_{h,k}(y_t)$ 为半正定矩阵矛盾.所以定理 2 成立.

从定理 2 的证明我们可以猜想,如果 $\hat{f}_{h,k}(y_t), t=1,2,\dots$ 序列收敛并且严格单调递增,则相应的 Hessian 矩阵 $\nabla^2 \hat{f}_{h,k}(y_t)$ 为负定矩阵.因此有下面的定理 3.为了证明该定理,首先给出引理.

引理. 当 $k(\cdot)$ 为高斯核函数时,令由 Mean Shift 算法产生的点列为 $\{y_t\}$,则有下式成立:

$$\nabla \hat{f}_{h,k}(y_t)^T \nabla \hat{f}_{h,k}(y_{t+1}) > 0.$$

证明:由文献[11]可知,当 $k(\cdot)$ 为高斯核函数时,有

$$\frac{\nabla m_{h,g}(y_t)^T \nabla m_{h,g}(y_{t+1})}{\|\nabla m_{h,g}(y_t)\| \cdot \|\nabla m_{h,g}(y_{t+1})\|} > 0.$$

将式(4)代入,得到:

$$\frac{1}{2} h^2 c \left[\frac{\nabla \hat{f}_{h,k}(y_t)}{\hat{f}_{h,k}(y_t)} \right]^T \cdot \frac{1}{2} h^2 c \frac{\nabla \hat{f}_{h,k}(y_{t+1})}{\hat{f}_{h,k}(y_{t+1})} > 0,$$

即 $\nabla \hat{f}_{h,k}(y_t)^T \cdot \nabla \hat{f}_{h,k}(y_{t+1}) > 0$.

该引理说明:相邻两步之间迭代方向的夹角在 $(0^\circ, 90^\circ)$ 内,保证沿着梯度上升的方向收敛.

定理 3(收敛性定理). 设 $S \subseteq R^d$ 为非空开凸集, $\hat{f}_{h,k}: S \rightarrow R$ 在 S 上具有二阶连续偏导数,如果 $\forall y_t \in S, k(\cdot)$ 为高斯核函数且 $\hat{f}_{h,k}$ 在 y_t 处的 Hessian 矩阵 $\nabla^2 \hat{f}_{h,k}(y_t) (t=1,2,\dots)$ 为负定矩阵,则 $\hat{f}_{h,k}(y_t)$ 序列收敛并且严格单调递增,而且序列 $\{y_t\}$ 收敛.

证明:从核函数的性质可知 $k(x)$ 有界,由式(1)可知 $\hat{f}_{h,k}$ 有界,只需证明 $\hat{f}_{h,k}(y_t)$ 单调递增,就已说明 $\hat{f}_{h,k}(y_t)$ 的收敛性.

对 $t \geq 0$, $\hat{f}_{h,k}(y_t)$ 在 y_t 处一阶 Taylor 公式展开,并令 $y=y_t + \lambda_t d_t \in S$,得到:

$$\hat{f}_{h,k}(y_t + \lambda_t d_t) = \hat{f}_{h,k}(y_t) + \lambda_t \nabla \hat{f}_{h,k}(y_t + \theta \lambda_t d_t)^T d_t, 0 < \theta < 1.$$

设 $\varphi(\theta) = \hat{f}_{h,k}(y_t + \lambda_t d_t) - \hat{f}_{h,k}(y_t) = \lambda_t \nabla \hat{f}_{h,k}(y_t + \theta \lambda_t d_t)^T d_t$,则下列极限成立:

$$\lim_{\theta \rightarrow 0} \varphi(\theta) = \lim_{\theta \rightarrow 0} \lambda_t \nabla \hat{f}_{h,k}(y_t + \theta \lambda_t d_t)^T d_t = \lambda_t \|\nabla \hat{f}_{h,k}(y_t)\|^2 > 0,$$

$$\lim_{\theta \rightarrow 1} \varphi(\theta) = \lim_{\theta \rightarrow 0} \nabla \hat{f}_{h,k}(y_t + \lambda_t d_t)^T \lambda_t d_t = \lambda_t \nabla \hat{f}_{h,k}(y_{t+1})^T \nabla \hat{f}_{h,k}(y_t) > 0 \text{ (引理)}.$$

$\varphi(\theta)$ 连续且可导,则

$$\varphi'(\theta) = \lambda_i^2 d_i^T \nabla^2 \hat{f}_{h,k}(y_i + \theta d_i)^T d_i = (y_{i+1} - y_i)^T \nabla^2 \hat{f}_{h,k}(y_i + \theta d_i)^T (y_{i+1} - y_i) < 0 \quad (\text{易知 } y_i + \theta d_i \in S).$$

易知 $\varphi(\theta)$ 单调递减且 $\varphi(\theta) > 0, \forall \theta \in (0, 1)$, 即 $\hat{f}_{h,k}(y_{i+1}) > \hat{f}_{h,k}(y_i)$.

所以, $\hat{f}_{h,k}(y_i)$ 序列严格单调递增且收敛, 设收敛于 $\hat{f}_{h,k}(\bar{y})$ 点, $\bar{y} \in S$, 故 $\lim_{\hat{f}_{h,k}(y_i) \rightarrow \hat{f}_{h,k}(\bar{y})} = 0$.

又 $0 < \varphi(1) < \varphi(\theta)$, 得到:

$$\lim_{\hat{f}_{h,k}(y_i) \rightarrow \hat{f}_{h,k}(\bar{y})} \nabla \hat{f}_{h,k}(y_{i+1})^T \nabla \hat{f}_{h,k}(y_i) = 0 \quad (8)$$

$$\lim_{\hat{f}_{h,k}(y_i) \rightarrow \hat{f}_{h,k}(\bar{y})} \nabla \hat{f}_{h,k}(y_i)^T \nabla \hat{f}_{h,k}(y_{i+1}) = 0 \quad (9)$$

式(8)和式(9)的左边相乘, 得到:

$$\lim_{\hat{f}_{h,k}(y_i) \rightarrow \hat{f}_{h,k}(\bar{y})} \|\nabla \hat{f}_{h,k}(y_{i+1})\|^2 \cdot \|\nabla \hat{f}_{h,k}(y_i)\|^2 = 0,$$

则 $\lim_{\hat{f}_{h,k}(y_i) \rightarrow \hat{f}_{h,k}(\bar{y})} \nabla \hat{f}_{h,k}(y_i) = 0$, 即 $\lim_{\hat{f}_{h,k}(y_i) \rightarrow \hat{f}_{h,k}(\bar{y})} (y_{i+1} - y_i) = 0$, 所以, y_i 收敛.

考虑到连续性, 易知 y_i 收敛于 \bar{y} , 且 $\nabla \hat{f}_{h,k}(\bar{y}) = 0$.

3 结果讨论

定理 3 假设条件的实质是要求 $\hat{f}_{h,k}(y)$ 为凹函数, $\forall y \in S$, 对 $k(\cdot)$ 的凸凹性不作要求; 而文献[11,12]的假设条件是 $k(\cdot)$ 为凸函数, 在 Mean Shift 算法中, $k(\|x-x_i\|^2)$ 可能为凸函数, 或凹函数, 或两者都不是. 定理 3 的假设条件保证初始点 $\forall y_0 \in S$, 则 Mean Shift 算法一定会收敛到局部极大值; 如果不具有定理 3 的假设条件, 否则 Mean Shift 算法不一定会收敛到局部极大值, 可能收敛到鞍点. 下面从几何的角度进一步说明上面的定理的正确性.

假设核函数 $k(x)$ 为高斯函数 $\exp(-x)$, 则 $y = \exp(-x)$ 在定义域内为凸函数, 如图 1(a)所示. 函数为 $y = \exp(-x^2)$ 时, 当 $x < a$ 和 $x > b$ 时分别为凸函数, $x \in (a, b)$ 区间为凹函数, 如图 1(b)所示, 其中, $a = -\sqrt{2}/2, b = \sqrt{2}/2$. 当 $x=0$ 时, y 取极大值, 此时, $x \in (a, b)$. 函数为 $y = \exp(-(x-3)^2)$ 时, 在 $x \in (-\infty, c)$ 和 $x \in (d, +\infty)$ 区间分别为凸函数, 在 $x \in (c, d)$ 区间为凹函数, 如图 1(c)所示, 其中, $c = 3 - \sqrt{2}/2, d = 3 + \sqrt{2}/2$. 当 $x=3$ 时, y 取极大值, 此时, $x \in (c, d)$. 函数为 $y = \exp(-(x_1-3)^2 - (x_2-7)^2)$ 的三维图形如图 2(d)所示. 当 $2.5 < x_1 < 3.5, 6.5 < x_2 < 7.5$ 时为凹函数, 如图 1(d)深色部分所示; 当 $x_1=3, x_2=7$ 时, y 取极大值, 此时, $2.5 < x_1 < 3.5, 6.5 < x_2 < 7.5$. 从图上看, 当初始点落在凹函数区间内时, 则序列 $\hat{f}_{h,k}(y_i)$ 沿着梯度方向递增且一定会收敛到局部极大值. 而当初始点落在凸函数区间内时, 序列 $\hat{f}_{h,k}(y_i)$ 也会沿着梯度方向递增, 如果沿着梯度方向能进入凹函数区, 则会收敛到局部极大值, 否则会收敛到鞍点. 上面是从单个数据点的角度用几何的方法来分析收敛条件, 没有考虑 h 对图形的影响. 对于有多个数据点的情况, 有相同的收敛结论, 只不过多个数据点的图形更为复杂, 在定义域内将会有多个峰值, 也可能有多个鞍点. 如图 1(e)和图 1(f)所示为 $h=2$ 和 $h=2.5$ 的多个数据点的二维高斯核函数的密度估计图形, 灰色点表示局部极值点, 黑色点表示鞍点, 深色部分表示 $\hat{f}_{h,k}(y_i)$ 在该区域为凹函数. 带宽 h 仅仅决定观测到的峰值个数^[7]和 $\hat{f}_{h,k}(y_i)$ 为凹函数区域的大小. 一般来说, 峰值的个数随着 h 的增大而减少, 而 $\hat{f}_{h,k}(y_i)$ 为凹函数区域随 h 的增大而向四周扩张. 如图 1(f)的深色部分在 x_1, x_2 平面的投影区大于图 1(e)在 x_1, x_2 平面的投影区. 图 1 中的图形显然是文献[11,12]的假设条件的反例.

Mean Shift 算法的收敛性在文献[1]中也得到讨论, 但该文献所有的讨论都是针对 blurring process 的: 每一步迭代后, 要循环地修改输入, 与本文讨论的 Mean Shift 算法不同. Mean Shift 算法的收敛性归功于式(7)的梯度上升属性, 但是, 朝局部梯度方向移动仅仅保证以无限小迭代步收敛. 迭代算法的步长尺寸对于算法的整体性能是非常重要的: 假如步长太大, 算法将会发散; 步长太小, 则收敛速度太慢. 而 Mean Shift 算法采用自适应步长, 不需要增加另外的步骤来搜索合适的步长(如最速下降法需要搜索最优步长), 这是 Mean Shift 算法的优点. 文献[17,18]指出: 极大熵聚类算法(maximum entropy clustering algorithm)不一定收敛到局部极值点, 有时收敛到鞍点, 这与本文得出的结论是一致的. 这主要是因为它们都是迭代算法, 形式上很相似, 且文献[1]指出最大熵聚类算法是 Mean Shift 算法的一种特殊情况. Comaniciu 在文献[7]中指出: Mean Shift 保证在密度估计的零梯度附近

收敛,需要一种方法来检验是否收敛于局部极大;而且他还提供了判断是否为鞍点的算法,这进一步为我们的结论提供了佐证.文献[1]认为, K -均值聚类算法是 Mean Shift 算法的一种极限形式,但众所皆知, K -均值聚类是一种收敛到局部稳定点的算法,且很多学者试图采用优化算法来使 K -均值聚类算法收敛于全局稳定点,如文献[19]采用遗传算法来选择初始聚类中心.从这一点来看,Mean Shift 算法和 K -均值聚类算法应具有相同的局部收敛性,而本文 Mean Shift 算法的收敛性定理为此提供了理论.

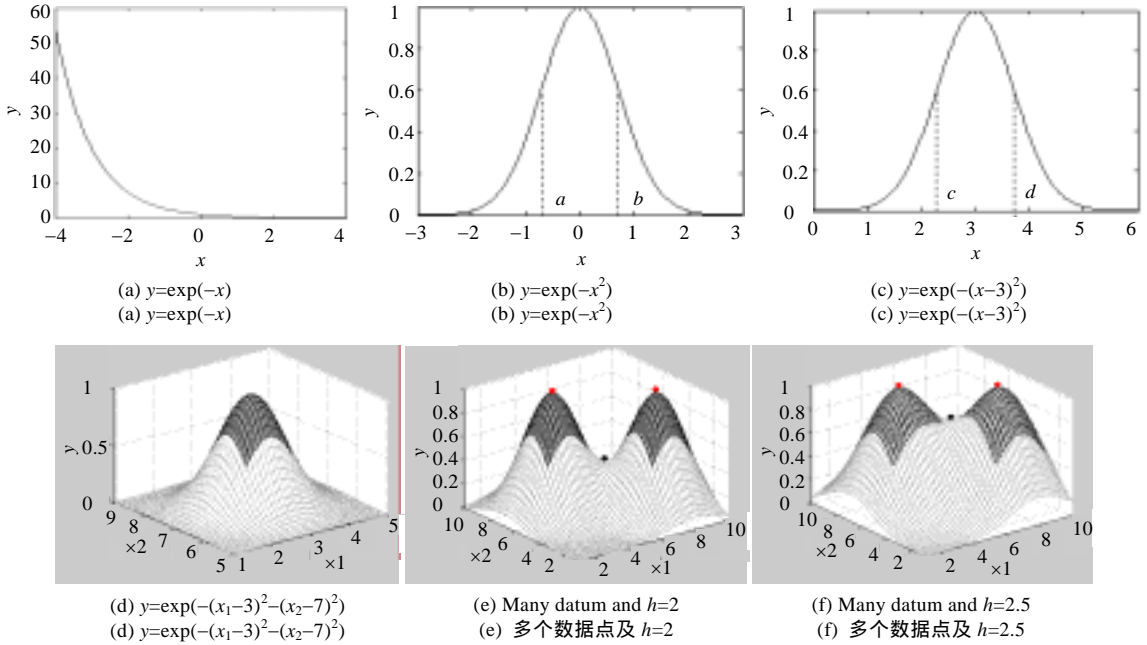


Fig.1 Geometrical graph

图 1 几何图形

4 结束语

本文针对 Mean Shift 算法的收敛性,首先指出文献[11,12]证明过程的错误,并从理论上进行了证明;然后,从数学上重新证明了 Mean Shift 算法的收敛性,指出其收敛到局部极大值的条件;最后,通过几何举例和对比讨论分析了本文的结论,为 Mean Shift 算法的深入研究及应用奠定了基础.

当然,该算法还有许多需要进一步研究的问题.例如:Cheng^[1]指出,Mean Shift 是自适应步长的梯度增长方法,但是步长的理论一直不清楚;另外,文献[20]证明了基于分段常数轮廓函数(piecewise constant profile)的 Mean Shift 算法是 Newton 的迭代,但基于高斯函数的 Mean Shift 的收敛速度一直不清楚;最后,能否在 Mean Shift 算法的迭代过程中增加某些限制来加快算法的收敛速度.到目前为止,人们对 Mean Shift 算法的内部机制还缺乏了解,需要有更多的研究成果来推动 Mean Shift 算法的发展.我们期待本文的成果会推动 Mean Shift 算法的深入研究.

致谢 在此,我们向对本文的工作给予支持和建议的同行,尤其是中南大学智能机器人实验室的全体老师和同学表示感谢.

References:

[1] Cheng YZ. Mean shift, mode seeking, and clustering. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1995,17(8): 790-799.

- [2] Comaniciu D, Ramesh V. Mean shift and optimal prediction for efficient object tracking. In: Mojsilovic A, Hu J, eds. Proc. of the IEEE Int'l Conf. on Image Processing (ICIP). 2000. 70–73. <http://www.caip.rutgers.edu/~comanici/>
- [3] Comaniciu D, Ramesh V, Meer P. Real-Time tracking of non-rigid objects using mean shift. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2000. 142–149. <http://www.caip.rutgers.edu/~comanici/>
- [4] Collins RT. Mean shift blob tracking through scale space. In: Proc. of the Conf. on Computer Vision and Pattern Recognition (CVPR). 2003. 18–20. <http://csdl.computer.org/comp/proceedings/cvpr/2003/1900/02/190020234abs.htm>
- [5] Comaniciu D. Image segmentation using clustering with saddle point detection. In: Proc. of the IEEE Int'l Conf. on Image Processing (ICIP). 2002. 297–300. <http://www.caip.rutgers.edu/~comanici/>
- [6] Wang J, Thiesson B, Xu YQ, Cohen M. Image and video segmentation by anisotropic kernel mean shift. In: Proc. of the European Conf. on Computer Vision (ECCV). 2004. <http://research.microsoft.com/~cohen/MeanShiftECCV.pdf>
- [7] Comaniciu D, Ramesh V, Bue AD. Multivariate saddle point detection for statistical clustering. In: Proc. of the European Conf. on Computer Vision (ECCV). 2002. 561–576. <http://www.caip.rutgers.edu/~comanici/>
- [8] Georgescu B, Shimshoni I, Meer P. Mean shift based clustering in high dimensions: A texture classification example. In: Proc. of the ICCV. 2003. 456–463. <http://www.caip.rutgers.edu/riul/research/papers/pdf/hims.pdf>
- [9] Comaniciu D, Meer P. Mean shift analysis and applications. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV). 1999. 1197–1203. <http://www.caip.rutgers.edu/~comanici/>
- [10] Comaniciu D. Nonparametric information fusion for motion estimation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2003. 59–66. <http://www.caip.rutgers.edu/~comanici/>
- [11] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002,24(5):603–619.
- [12] Li XR, Wu FC, Hu ZY. Convergence of a mean shift algorithm. Journal of Software, 2005,16(3):365–374 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/16/365.htm>
- [13] Peng NS, Yang J, Liu Z, Zhang FC. Automatic selection of kernel-bandwidth for mean shift object tracking. Journal of Software, 2005,16(9):1542–1550 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/16/1542.htm>
- [14] Comaniciu D, Ramesh V, Meer P. The variable bandwidth mean shift and data-driven scale selection. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV). 2001. 438–445. <http://www.caip.rutgers.edu/~comanici/>
- [15] Comaniciu D. An algorithm for data-driven bandwidth selection. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003, 25(2):281–288. <http://www.caip.rutgers.edu/~comanici/>
- [16] Xie Z, Li JP, Tang ZY. Non-Linear Optimization. Changsha: Publishing House of National University of Defense Technology, 2003. 167–174 (in Chinese).
- [17] Yu J, Shi HB, Huang HK, Sun XC, Cheng QS. Counterexamples to convergence theorem of maximum-entropy clustering algorithm. Science in China Series F, 2003,46(5):321–326.
- [18] Mu YM, Yu J. On convergence of the maximum entropy clustering algorithm. Journal of Northern Jiaotong University, 2003,27(5): 26–29 (in Chinese with English abstract).
- [19] Laszlo M, Mukherjee S. A genetic algorithm using hyper-quadtrees for low-dimensional k -means clustering. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006,28(4):533–543.
- [20] Fashing M, Tomasi C. Mean shift is a bound optimization. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005,27(3): 471–474.

附中文参考文献:

- [12] 李乡儒,吴福朝,吴战义.均值漂移算法的收敛性.软件学报,2005,16(3):365–374. <http://www.jos.org.cn/1000-9825/16/365.htm>
- [13] 彭宁嵩,杨杰,刘志,张凤超.Mean Shift 跟踪算法中核函数窗宽的自动选取.软件学报,2005,16(9):1542–1550. <http://www.jos.org.cn/1000-9825/16/1542.htm>
- [16] 谢政,李建平,汤泽滢.非线性最优化.长沙:国防科学技术大学出版社,2003.30–54.
- [18] 牟永敏,于剑.极大熵聚类算法的收敛性定理.北方交通大学学报,2003,27(5):26–29.



文志强(1973 -),男,湖南湘乡人,博士生,讲师,主要研究领域为机器视觉,信息融合.



蔡自兴(1938 -),男,教授,博士生导师,CCF高级会员,主要研究领域为人工智能基础,进化计算,智能控制,移动机器人协作与导航.