

# K-Means 聚类算法和人工鱼群算法应用于图像分割技术<sup>①</sup>

楚晓丽

(广东农工商职业技术学院, 广州 510507)

**摘要:** 提出一种基于 K-Means 聚类的人工鱼群算法, 该算法利用人工鱼群算法鲁棒性较强且不易陷入局部最优值的特点, 动态的确定了聚类的数目和中心, 解决了 K-Means 聚类初始点选择不稳定的缺陷, 在此两种算法融合的基础上进行图像分割处理, 经试验证明该算法效果理想。

**关键词:** 图像分割技术; K-Means 聚类算法; 人工鱼群算法

## K-Means Clustering Algorithm and Artificial Fish Swarm Algorithm Applied in Image Segmentation Technology

CHU Xiao-Li

(Guangdong AIB Polytechnic College, Guangzhou 510507, China)

**Abstract:** The paper presents an artificial fish swarm algorithm based on K-Means clustering. The algorithm uses the feature of having artificial fish swarm algorithm's strong robustness and being not easy to fall into local optimum value, and hence dynamically determines the number of clusters and center, overcoming the defects of K-Means clustering initial point selected unstable. The image segmentation is processed based on the fusion of two algorithms. The test proves the algorithm is ideal.

**Key words:** image segmentation; K-Means clustering algorithm; artificial fish swarm algorithm

图像分割是把人们感兴趣的对象从复杂场景中提取的一种图像技术, 图像分割有许多种办法, 包括边缘检测法、阈值分割法、区域分割法、运动分割法等, 但是各种方法各有优劣势的存在, 并没有公认的分割标准。近些年来研究学者提出了将一些特定的理论、方法、工具相结合的方式, 达到优势互补的效果, 实现图像分割。

人工鱼群算法是一种新兴的仿生学算法, 它利用自上而下的寻优模式, 模仿自然界鱼群的觅食方式, 通过建立鱼的觅食、群聚和追尾行为, 实现个体底层行为的构造。通过鱼群中每个个体的寻优, 满足群体行为在全局寻找最优值的目的。其主要特点是不需要了解问题的特殊信息, 只需要对问题进行优劣的比较, 通过各人工鱼个体的局部寻优行为, 最终使全局最优值突显出来, 有着较快的收敛速度<sup>[1]</sup>。并且算法中只使用目标函数的函数值, 不需要其梯度值等特殊信息,

对搜索空间具有一定的自适应能力, 算法对初值无要求, 对各参数的选择也不很敏感<sup>[2]</sup>关于人工鱼群算法的研究目前逐渐增多, 应用的领域也日渐广阔, 广大学者在图像处理、电力系统、前向神经网络优化、数学的组合优化, 时变系统的在线辨识, 鲁棒 PD 的参数整定, 短期负荷预测中, 均有了较为深入的研究, 且取了一定的成果。

K-Means 聚类算法及改进方法已在图像分割中得到了广泛的应用, 该分割方法将图像作为一个图像特征向量几何, 把图像分割任务转化成对数据集合的聚类任务<sup>[3]</sup>。k-means 聚类算法的初始点选择不稳定, 是随机选取的, 这就引起聚类结果的不稳定, 本实验中虽是经过多次实验取的平均值, 但是具体初始点的选择方法还需进一步研究<sup>[4]</sup>本文把人工鱼群算法引入到 K-Means 聚类算法中, 用人工鱼群算法代替 K-Means 聚类算法中的迭

① 收稿时间:2012-09-12;收到修改稿时间:2012-10-29

代过程, 优化初始点的选定, 改善 K-Means 聚类算法的不稳定性特点, 并且运用到图像分割中.

## 1 人工鱼群算法(AFSA)<sup>[1,5,6]</sup>

人工鱼群算法(Artificial Fish Swarm Algorithm, AFSA)是李晓磊等人在2002年模仿鱼类行为方式提出的一种基于动物自治体的优化方法<sup>[6-8]</sup>. 人工鱼的个体状态可表示为向量  $X=(x_1, x_2, \dots, x_n)$ , 其中  $x_i(i=1,2,\dots,n)$  为欲寻优的变量; 人工鱼当前所在位置的食物浓度表示为  $Y=f(x)$ , 其中  $Y$  为目标函数值; 人工鱼个体之间的距离表示为  $d_{ij}=\|X_i-X_j\|$ ;  $Visual$  表示人工鱼的感知范围;  $\delta$  表示拥挤度因子;  $Step$  表示人工鱼移动的步长, 人工鱼典型行为的实现原理如下:

### 1.1 觅食行为

设人工鱼当前状态为  $X_i$ , 在其感知范围内随机选择一个状态  $X_j(d_{ij}\leq Visual)$ , 如果  $X_j$  优于  $X_i$  则向  $X_j$  前进一步; 反之, 再重新随机选择状态  $X_j$ , 判断是否满足前进条件; 反复几次后, 如果仍不满足前进条件, 则随机移动一步.

### 1.2 聚群行为

设人工鱼当前状态为  $X_i$ , 探索可见邻域内(即  $d_{ij}\leq Visual$ )的人工鱼数目为  $n_f$ , 如果  $n_f < \delta N$ , 表明伙伴中心有较多的食物并且不太拥挤, 如果此时  $Y_c$  优于  $Y_i$ , 则朝伙伴的中心位置方向向前进一步; 否则执行觅食行为.

### 1.3 追尾行为

设人工鱼当前状态为  $X_i$ , 探索其在  $Visual$  范围内状态最优的邻居  $X_{max}$ , 如果  $Y_{max}$  优于  $Y_i$ , 并且  $X_{max}$  的邻域内伙伴的数目  $n_f$  满足  $n_f < \delta N$ , 表明  $X_{max}$  的附近有较多的食物并且不太拥挤, 则向  $X_{max}$  的位置前进一步; 否则执行觅食行为.

### 1.4 随机行为

人工鱼在视野中随机选择一个状态, 然后向该方向移动, 它是觅食行为的一个缺省行为. 在觅食行为中, 当重复尝试次数  $try\_number$  较小时, 为人工鱼提供了随机游动的机会, 增加了种群的多样性, 以实现跳出局部极值.

### 1.5 约束行为

在寻优过程中, 由于聚群行为、随机行为等操作的作用, 容易使得人工鱼的状态变得不可行, 这时就需要加入相应的约束行为对其进行规整, 使它们由无效状态或不可行状态转变成可行状态.

## 1.6 公告板

公告板用来记录最优人工鱼个体的状态  $X_{best}$ . 各人工鱼个体在寻优过程中, 每次行动完毕就检验自身状态与公告板的状态, 如果自身状态优于公告板状态, 就将公告板的状态改写为自身状态, 这样就使公告板记录下历史最优状态.

## 2 K-Means聚类算法

K-Means 算法以  $k$  为参数, 把  $n$  个对象分成  $k$  个簇, 使簇内具有较高的相似度, 而簇间的相似度较低<sup>[4]</sup>. 假如数据集为:  $X=\{x_m | m=1,2,3,\dots, total\}$ , 其中  $X$  中的样本用  $d$  个描述属性的  $A_1, A_2, \dots, A_d$  来表示, 并且  $d$  个描述属性都是连续型属性. 数据样本  $x_i=(x_{i1}, x_{i2}, \dots, x_{id})$ ,  $x_j=(x_{j1}, x_{j2}, \dots, x_{jd})$ , 其中,  $x_{i1}, x_{i2}, \dots, x_{id}$  和  $x_{j1}, x_{j2}, \dots, x_{jd}$  分别是样本  $x_i$  和  $x_j$  对应  $d$  个描述属性  $A_1, A_2, \dots, A_d$  的具体取值. 样本  $x_i$  和  $x_j$  之间的相似度通常用它们之间的距离  $d(x_i, x_j)$  来表示, 欧式距离公式如下:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^d (x_{ik} - x_{jk})^2} \quad (1)$$

很明显, 距离越小, 样本  $x_i$  和  $x_j$  越相似, 差异度越小; 距离越大, 样本  $x_i$  和  $x_j$  越不相似, 差异度越大.

K-Means 算法的工作原理是: 算法首先随机从数据集中选取  $K$  个对象, 每个对象初始地代表了一个簇的平均值或中心, 计算各个样本到聚类中的距离, 把样本归到离它最近的那个聚类中心所在的类. 对于剩余的其他对象, 根据其距各簇的距离, 归到最近的簇, 继而再次计算每个簇的平均值, 新形成的每一个聚类的数据对象重复迭代, 不断得到新的聚类中心, 直到函数收敛. 通常采用的准则为平方误差, 定义如下:

$$E = \sum_{i=1}^k \sum_{p \in X} \|p - m_i\|^2 \quad (2)$$

## 3 基于K-Means聚类的人工鱼群算法

该算法的主要建模方式为: 每一幅图像在图像处理工具中被处理为一个  $M \times N$  的矩阵, 其中  $M$  和  $N$  分别是图像行像素数和列像素数. ( $0 < i < M$ ,  $0 < j < N$ ) 每一幅图的像素矩阵都有一个随机初始位置, 每一条鱼能够在像素矩阵上移动<sup>[7]</sup>. 根据目标函数将实现自组织聚类过程. 实施步骤如下:

Step1: 初始化群体, 在可控制域内形成  $n_f$  条人工

鱼, 初始化聚类中心个数  $k$ , 最大迭代次数, 人工鱼感知范围, 拥挤度因子等参数; 目标函数值  $Y$  为公式(1).

Step2: 人工鱼群算法全局搜索最优解. 其过程为:

①公告板赋初值, 计算初始鱼群各人工鱼当前所在位置的食物浓度  $y$ , 取  $y$  最小值进入公告板, 并将此鱼当前状态也赋值给公告板. ②探索当前人工鱼周围的聚类中心位置  $X_c$ . ③根据当前人工鱼与周围聚类中心位置的目标函数的大小分别模拟追尾行为和聚群行为, 选择  $y$  值较优者的行为为实际执行, 缺省行为为觅食行为. ④人工鱼群每行动一次后, 如果人工鱼群最优鱼状态的  $y$  值优于公告板的值, 则更新公告板, 聚类中心也随之更新.

Step3: 终止条件判断. 判断是否已达到最大迭代次数, 满足公式(1)值为  $\min$ . 若不满足则执行步骤 2, 若满足则输出计算结果, 从而得到全局最优解.

Step4: 以输出的最优解作为 K-Means 聚类算法的输入参数, 利用优化后的算法对图像进行分割. 最终输出分割后的图像.

#### 4 仿真实验

为了验证本算法的有效性与可行性, 进行了如下仿真实验. 实验环境为: MATLAB 2007a, cpu 为 AMD athlon5000+, 内存为 2G. 本文进行了两项相关实验, 实验一是采用 UCI 数据库上的经典数据 Iris、Haberman、New-thyriod、Balance scale 作为测试对象, 该四类对象直接反映聚类的质量, 实验二是针对经典的图像处理实例 lenna 图进行图像分割, 直观反映分割效果.

实验一: 传统的 K-means 算法的缺点是稳定性差, 初始中心点具有较强的随机性, 而本算法则在一定程度上有所改进, 本文对传统与改进算法各进行数次实验, 实验数据参照下表, 实验结果表明本改进算法的优异性.

表 1 改进的 AFSA 与 K-means 性能测试结果

算法	次数 序号	Iris (%)	Haberman n (%)	New-thy riod (%)	Balance scale (%)
K-means	1	66.49	55.66	69.69	46.86
	2	75.24	56.38	79.83	48.77
	3	59.31	51.52	75.35	41.69
	4	87.87	50.59	77.70	45.48
	5	84.55	54.61	72.97	51.25
	6	77.62	53.79	70.06	53.74
	7	59.46	55.81	66.89	49.63
	8	63.35	53.63	65.38	47.98
	平均	71.74	54.00	72.23	48.18

基于 K-means 的 AFSA	1	88.57	77.45	79.74	51.52
	2	82.63	74.95	78.94	55.66
	3	78.79	79.66	83.64	54.49
	4	76.58	82.95	85.54	57.82
	5	84.31	81.36	84.72	53.65
	6	89.69	80.97	84.66	49.88
	7	90.74	85.87	81.51	48.63
	8	92.59	76.46	86.94	47.11
	平均	85.49	79.96	83.21	52.35

实验二: 本文使用了经典的图像处理实例 lenna 图作为对象, 图像像素大小为  $175 \times 170$ .

实验参数设置如下: 种群规模为 20, 步长  $\text{step}=0.6$ , 视野  $\text{visual}=2$ , 试探次数  $\text{try\_number}=35$ , 最大迭代次数  $\text{number}=200$ , 通过对 lenna 图像进行实验分割测试, 得到如下效果图, 证明本算法的有效性.



图 1 Leana 原图



图 2 K-Means 聚类算法得到的效果图



图 3 本文算法得到的效果图

(下转第 103 页)

输出值不是总体的无偏估计,使总体重心偏小;RPE-filter 算法能够准确剔除随机单点跳跃,用正常时延序列的重心平滑波动的曲线,且能及时感知因网线拓扑的改变而引起的时延阶梯性变化。

对同一个实测时延数据集计算每个时延过滤算法输出的时延值与当前实测时延的绝对误差,按照时间的先后顺序对绝对误差作累加和,得到如图 5 中的三种时延过滤算法的性能对比图,从此图可以看出  $\Delta$ -filter 的性能最差、MP-filter 的性能处于两者之间、RPE-filter 的性能最好。

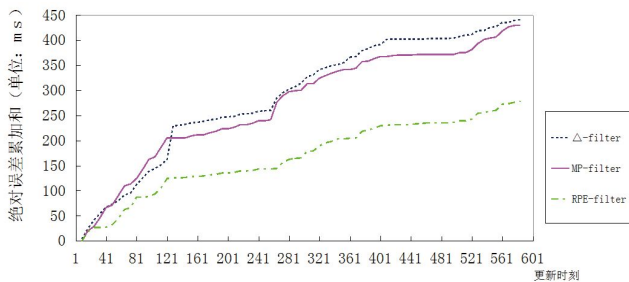


图 5 三种时延预处理算法性能对比

### 3 结语

从理论以及对时延数据集的统计分析得出在实际网络中网络时延出现抖动以及长拖尾现象,因此必须对时延样本进行预处理以避免“失真时延”对构建和维护虚拟网络坐标系统的性能产生负面影响.对已有的时延预处理算法的优缺点进行了分析,提出了递归点估计时延过滤算法,选择核心集中域的重心作为系统的输入时延.将本文算法与已有的两种时延过滤算法的性能进行了对比,本文算法不仅能够去除随机延迟

污染,还能平滑时延的波动,更能实时反映网络拓扑的变化,且绝对误差累加和明显小于已有的算法。

### 参考文献

- 1 T.Ng, Zhang H. Predicting internet network distance with coordinates-based approaches. Proc. IEEE INFOCOM, 2002: 170-179.
- 2 Dabek F, Cox R, Kaashoek F, et al. Vivaldi: A Decentralized Network Coordinate System. Proc. of ACM SIGCOMM'04. Portland: ACM Press, 2004. 15-26.
- 3 Pietzuch P, Ledlie J, et al. Network-Aware Operator Placement for Stream-Processing Systems. ICDE, 2006, Apr.
- 4 Andersen DG, Balakrishnan H, Kaashoek MF, Morris R. Resilient overlay networks. Proc. 18th ACM SOSP, 2001.
- 5 Pietzuch P, et al. Network-Aware Overlays with Network Coordinates. IWDDS, July 2006.
- 6 PlanetLab. <https://www.planet-lab.org/>, 2007-1-1/2012-10-8.
- 7 Ledlie J, Pietzuch P, Seltzer M. Stable and Accurate Network Coordinates. Proc. of International Conference on Distributed Computing Systems. Portugal, Computer Society. 2006. 84-93.
- 8 张敏,周亮,黄琼,阳小龙,隆克平.随机延迟污染对 IP 网络坐标系统的影响及其抑制方法研究.通信学报,2011,32 (12): 29-35.
- 9 李朝辉.数字图像处理及应用.北京:机械工业出版社,2004. 136-165.
- 10 茆诗松,贺思辉.概论论与统计学.武汉:武汉大学出版社, 2010. 235-271.

(上接第 94 页)

本文提出了基于 K-Means 聚类的人工鱼群算法,通过人工鱼群算法的全局搜索能力较强,且不易陷入局部最优的特点,改进了 K-Means 聚类对初始点选择不稳定的算法特性,大大的提高了全局寻优的能力,同时也具备了收敛速度快,效果好,且简单便捷的特点.仿真实验结果表明该算法的有效性与可行性。

### 参考文献

- 1 李晓磊,路飞,田国会,钱积新.组合优化问题的人工鱼群算法应用.山东大学学报(工学版),2004,34(5):64-67.
- 2 王闯.人工鱼群算法的分析及改进[硕士学位论文].大连:

大连海事大学,2008.

- 3 黄宇.基于 MARKOV 随机场 k-means 图像分割算法.电子学报,2009,37(12):2700-2704.
- 4 冯晓蒲,张铁峰.四种聚类方法之比较.微型机与应用, 2010,29(16):1-3.
- 5 李晓磊.一种新型的智能优化方法—人工鱼群算法.杭州:浙江大学,2003.
- 6 李晓磊,邵之江,钱积新.一种基于动物自治体的寻优模式:鱼群算法.系统工程理论与实践,2002,(11):32-38.
- 7 刘白,周永权.一种基于人工鱼群的混合聚类算法.计算机工程与应用,2008,44(18):136-138.