

# 改进的 FCM 算法在 UASN 分簇中的应用

江萌萌, 刘广钟

(上海海事大学 信息工程学院, 上海 201306)

**摘 要:**分簇路由算法是水声传感器网络的关键技术之一,在具有严格能量约束的水声传感器网络中,如何达到节点在网络中均匀分布并且延长网络寿命一直是水声传感器网络路由协议中的难题。针对改善网络中的能量消耗和延长网络生命周期这两个目的,将水声传感器节点分簇路由问题建模为样本空间的模糊聚类划分问题,采用一种基于改进的模糊 C-均值聚类算法竞选簇头节点完成分簇,在计算初始化聚类中心过程中充分考虑节点能量衰减和节点的距离这两个因素,有效解决了原算法中初始聚类中心的随机性问题,避免陷入局部最优。同时结合聚类粒度原理选取最佳簇头个数。仿真结果表明,与基于模糊控制的 CEFL 算法相比,该算法明显改善了网络中能耗的均衡性,有效地生成了合理的簇结构,并且延长了网络生命周期。

**关键词:**水声传感器网络;分簇算法;簇头;模糊聚类;能量

**中图分类号:**TP39

**文献标识码:**A

**文章编号:**1673-629X(2018)12-0147-05

doi:10.3969/j.issn.1673-629X.2018.12.031

## Application of Improved FCM Algorithm in UASN Clustering

JIANG Meng-meng, LIU Guang-zhong

(School of Information Engineering, Shanghai Maritime University, Shanghai 201306, China)

**Abstract:** The clustering routing algorithm is one of the key technologies of the underwater acoustic sensor networks. In the underwater acoustic sensor network with strict energy constraints and mobile nodes, how to achieve uniform distribution of nodes in the network and extend the network lifetime in the routing protocols is always a difficult problem. In order to improve the network energy consumption and prolong the network life cycle, the underwater sensor nodes are modeled as fuzzy clustering problem in sample space, and the election of cluster head will use the improved fuzzy C-means clustering algorithm to complete the clustering. Two factors of energy attenuation and node distance are fully considered in the process of initializing the cluster center, which effectively solves the problem about the randomness of initial cluster centers in the original algorithm and avoids falling into local optimum. And combined with cluster granularity principle, the optimal number of cluster heads is selected. The simulation demonstrates that compared with the CEFL algorithm based on fuzzy control, the algorithm proposed can improve the balance of energy consumption within the network, generate reasonable cluster structure and prolong the network life cycle effectively.

**Key words:** UASN; clustering algorithm; cluster head; fuzzy clustering; energy

## 0 引言

水声传感器网络 (underwater acoustic sensor networks, UASN) 由于水下环境的潜在利益和独特的挑战而受到学术界和工业界的高度重视。UASN 允许大量的应用程序变得可行又有效,包括商业开发,海洋学数据收集和海岸线保护关于水声传感器网络方面的一些重要技术的研究引起了广泛重视<sup>[1-2]</sup>。

UASN 由大量便宜的便携式传感器节点以自组织

的方式组成,具有有限的功率,存储和计算能力。由于水下信道的复杂性,水声传感器网络环境下的数据传输速率和网络生存时间等都会受到严重影响,同时在水下工作想要更换节点电池是不可行的,所以节点的能量消耗必然引起人们的重视<sup>[3]</sup>。提出的分簇路由协议,可以通过仅允许一些节点与基站通信来减少能量消耗。这些称为簇头的节点收集该簇中的每个节点发送的数据,并将其融合数据发送到基站<sup>[4]</sup>。

收稿日期:2018-01-06

修回日期:2018-05-15

网络出版时间:2018-07-04

基金项目:国家自然科学基金(61202370);中国博士后科学基金(2014M561512);上海市教委科研创新项目(14YZ110)

作者简介:江萌萌(1992-),女,硕士研究生,研究方向为水声传感器网络、大数据和智能决策方法;刘广钟,博士,教授,CCF 高级会员(E200006996S),研究方向为水声传感器网络、分布式人工智能、计算机网络技术、网络计算、CIMS 技术、物流信息化技术等。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20180703.1511.028.html>

在模糊聚类算法<sup>[5-6]</sup>中,样本按一定的隶属度分类,得到了样本属于各个类别的不确定性程度,这样更能准确地反映现实世界。水声传感器网络的分簇过程类似模糊聚类分析,将水声传感器节点分簇路由问题建模为样本空间的模糊聚类划分问题。整个 UASN 看作一个模糊聚类的对象集合,每个传感器节点作为该集合中的一个样本,最接近模糊聚类得到的聚类中心就是 UASN 中的簇头节点,聚类得到的子集就是 UASN 中的每个簇。

1 相关工作

目前已有的分簇算法是基于节点的剩余能量,簇头的位置分布和覆盖度等准则提出的。典型的分簇路由算法有 LEACH<sup>[4]</sup>、HEED<sup>[7]</sup>、APTEEN<sup>[8]</sup>、DAEA<sup>[9]</sup>、TEEN<sup>[10]</sup>等。这里主要介绍基于模糊逻辑控制的 CEFL 分簇算法<sup>[11]</sup>。

在 CEFL 算法中,使用模糊逻辑控制模型中最常用的模糊推理技术,提出了基于能量、集中度和中心性三个描述符的簇头选举分簇算法,通过精确地修改每个模糊集的形状,进一步改善网络寿命和能量消耗。模糊规则库目前包括以下规则:如果能量高,集中度高,中心性接近,节点当选机会大。但该算法存在一些缺点:首先,每个节点概率地决定是否成为簇头,所以可能选出彼此相邻的两个簇头,增加了网络中耗尽的总能量。其次,CEFL 算法最后生成的簇头节点的数量不是固定的,所以有时它可能大于或小于优选值。

针对上述问题,提出一种基于改进的模糊 C-均值聚类算法的水声传感器网络分簇路由协议(FCM-L)。采用 FACM-L 对水声传感器节点进行聚类,在计算初始化聚类中心时通过考虑节点间的相对距离和能量衰减率这两个因素,避免了 FCM 算法对初始化聚类中心敏感这一缺点,并构造一个有效性函数来选定最佳聚类类别数,根据已产生的最佳聚类数完成水声传感器节点分簇过程。

2 网络模型

在分簇的 UASN 架构中,基站远离传感器节点并且是静止的,簇头节点负责控制簇内节点的协同工作,簇内节点向相应的簇头发送数据,簇头节点将收集到的数据信息压缩融合处理后将其发送到基站。

文中对网络模型做如下假设:

- (1)假设水声传感器网络是同构网络,每个节点都有唯一的 ID。
- (2)主要针对二维静态水声传感器网络,不考虑节点移动性。
- (3)每个节点都有可能竞选簇头。

(4)每个节点都具有相同初始能量,具有数据融合功能,而且节点能根据距离的远近来调整发射功率。

(5)基站的能量是无限的,而且能覆盖整个网络。

由于簇头节点不仅要完成通信,还要进行数据融合传输信息量,所以簇头节点的能量直接导致节点间的信息传输。这里将构建水声通信系统模型考虑节点在发送接收过程中确定节点传输数据所消耗的能量。引用文献[12-13]中的能耗模型,定义如下:

发送节点的最低发送功率为:

$$P_0' = P_0 \alpha^d d^k \tag{1}$$

其中,  $P_0$  表示节点接收端正常接收一个数据包的最低功率;  $k$  为能量扩展因子,通常取 1.5;  $\alpha$  为与频率有关的能力衰减系数,通常  $\alpha = 10^{\alpha(f)/10}$ ,  $\alpha(f)$  为吸收损耗系数:

$$\alpha(f) = \frac{0.11f^2}{1+f^2} + \frac{44f^2}{4\ 100+f^2} + \frac{2.75f^2}{10\ 000} + 0.03 \tag{2}$$

其中,  $f$  表示节点工作频率。

所以,当节点的传输距离为  $d$  时,发送  $b$  bit 的数据包的能耗为  $E_t = bP_0'$ , 节点接收  $b$  bit 数据包的能耗为  $E_r = P_r b$ ,  $P_r$  是与设备有关的常数。节点数据融合的能耗为  $E_d = bE_d$ ,  $E_d$  是一个常数。

节点能量衰减率为:

$$e_i = \frac{E_t + E_r + E_d}{E} \tag{3}$$

符号含义见表 1。

表 1 符号含义

符号	含义
$X$	分类对象集
$R$	模糊相似矩阵
$W$	聚类中心矩阵
$n$	节点个数
$c$	聚类个数
$E_t$	发送数据能耗
$E_r$	接收数据能耗
$x_{ij}$	第 $i$ 个分类对象的第 $j$ 个特性
$r_{ij}$	隶属度
$w_i$	聚类中心向量
$e_i$	能量衰减率

3 基于 FCM-L 的分簇路由算法设计

3.1 基于 FCM-L 的模糊聚类模型

文中提出一种基于改进的模糊 C-均值聚类算法的水声传感器网络分簇路由协议(FCM-L)。主要思想为将水声传感器节点作为分类对象,将其分为  $c$  个部分,其中每个部分都有一个聚类中心。这里需要利

用 FCM 算法给出一个目标函数来调整聚类中心,当该目标函数值小于给定阈值时,停止迭代,得到最后的聚类结果。设被分类对象的集合为  $X = \{x_1, x_2, \dots, x_n\}$ ,  $X$  为节点集合,  $x_i$  表示每个节点。

### 3.1.1 初始化聚类中心

考虑 FCM 算法对初始化聚类中心敏感,容易出现局部最优,导致簇头节点在网络中分布不均匀。所以通过计算分析数据对象两两之间的距离与能量衰减率倒数的乘积,能量衰减越快,倒数越小,所以对于下式结果越小越好,以便得到一个较好的初始聚类中心。

$$\beta = \left( \frac{e_i + e_j}{2} \right)^{-1} \cdot \sqrt{(x_i - x_j)^2} \quad (4)$$

比较分析找出两个节点之间最小的  $\beta$ , 创建新的数据对象集  $Y_1$ , 使这两个节点归入  $Y_1$ , 即  $(x_i, x_j) \in Y_1$ 。然后在原数据集中删除  $x_i, x_j$ , 原数据对象集变为  $(x_1, x_2, \dots, x_{n-2})$ , 接着计算  $Y_1$  中每一个节点与数据集  $X$  中每一个节点的距离与能量衰减率, 找出最小的  $\beta$ , 将该节点归入  $Y_1$  中, 设置阈值  $\gamma$ , 当  $Y_1$  中的节点个数达到一定阈值, 新建数据对象集  $Y_2$ , 重复数据对象集  $Y_1$  的形成过程, 直到形成  $k$  个数据对象集。将这  $k$  个聚类中心作为  $K$  均值聚类算法的初始聚类中心。

根据相关资料, 最佳聚类数  $2 < c \leq \sqrt{n}$ 。在以往有些研究方法中聚类数是根据系统预先给定, 这往往不能代表最终的聚类结果, 所以文中将先选出一组聚类数  $c$ , 然后构建一个有效性函数来确定最佳的聚类数。设  $c \in (2, \sqrt{n})$ , 这个范围可根据实际网络规模的大小决定, 如果网络规模较大, 可适当增大  $c$  的下界。首先取  $c = \sqrt{n}$ 。

于是把数据对象集  $X$  分成  $c$  类, 设  $c$  个聚类中心向量构成矩阵  $W$ :

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_c \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{c1} & w_{c2} & \cdots & w_{cm} \end{bmatrix} \quad (5)$$

### 3.1.2 构造模糊相似矩阵

模糊分类中被分类的对象集合  $X$  中的对象  $x_i$  以一定的隶属度属于某一类, 因此每一类就认为是对象集合  $X$  上的一个模糊子集, 每一种模糊分类就是一个  $c \times n$  的模糊矩阵  $R$ 。使用夹角余弦法求出隶属度  $r_{ij}$ , 具体求解过程参考文献[14]。每个  $r_{ij} \in [0, 1]$ , 于是构造模糊相似矩阵  $R = (r_{ij})_{c \times n}$ 。

### 3.1.3 聚类求解

聚类准则是通过不断迭代使如下目标函数达到最小值, 得到最终的聚类结果:

$$\min J_{\text{FCM}} = \sum_{i=1}^c \sum_{k=1}^n (r_{ik})^m \|x_k - w_i\|^2 \quad (6)$$

约束条件满足  $\sum_{i=1}^c x_{ik} = 1$ 。  $\|x_k - w_i\|$  表示  $x_k$  对应的特征向量与第  $i$  类聚类中心向量的距离,  $m$  表示聚类结果模糊度的权重指数, 通常取 2。利用 Lagrange 法求解式 6, 具体数学求解过程参考文献[15], 得到模糊划分矩阵  $R$  和聚类原型  $W$  的更新公式:

$$w_i = \frac{\sum_{k=1}^n (r_{ik})^m x_k}{\sum_{k=1}^n (r_{ik})^m} \quad (7)$$

$$r_{ik} = \left[ \sum_{j=1}^c \left( \frac{\|x_k - w_i\|}{\|x_k - w_j\|} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (8)$$

在最小化目标函数时需要不断迭代调整聚类中心, 给定阈值  $\varepsilon$  ( $0.001 < \varepsilon < 0.01$ ),  $w_i^{(t+1)} - w_i^t = \varepsilon$ , 如果目标函数值大于阈值  $\varepsilon$ , 则进行迭代重新更新隶属度, 否则停止迭代。

### 3.1.4 确定最佳聚类数

通过上述聚类方法得到一组聚类数  $c$ , 接下来将构建一个有效性函数来确定一个最佳聚类数。根据文献[16], 衡量聚类结果的好坏可根据簇内的紧凑度和簇间的分离度决定, 即簇内的对象尽可能紧凑, 簇间尽可能分离。具体步骤如下:

首先选取得到的不同聚类类别数  $c$ , 利用簇内紧凑度和簇间分离度决定最终聚类结果。

求解簇内紧凑度:

$$C(c) = \sum_{i=1}^c \frac{\sum_{j=1}^n x_{ij}^m \|x_k - w_i\|^2}{n}, i=1, 2, \dots, c, j=1, 2, \dots, n \quad (9)$$

求解簇间分离度:

$$S(c) = \frac{\sum_{i,k=1, i \neq k}^c \|x_k - w_i\|^2}{\sum_{i=1}^c c_i/n}, i, k=1, 2, \dots, c \quad (10)$$

根据紧凑度量簇内的紧致性, 紧凑度越小, 紧致性越好; 分离度量簇间的分离性, 分离度越大, 分离性越好。于是构造以下聚类有效性函数:

$$CS(c) = \frac{C(c)}{S(c)} \quad (11)$$

最后参考  $CS(c)$ , 它越小代表较好的聚类结果与  $CS(c)$  最小值相对应的  $c$  值就是最佳的簇头节点数。

## 3.2 算法描述

Step1: 输入对象集矩阵—水声传感器节点和节点特性指标矩阵。

Step2: 根据上述改进算法计算初始化聚类中心。

Step3: 构造模糊相似矩阵  $R = (r_{ij})_{c \times n}$ 。

Step4: 根据式 7 更新聚类中心  $w$ 。

Step5: 根据式 8 更新模糊分类矩阵  $R$ 。

Step6: 当迭代运算过程达到收敛  $\max\{|w_{ik}^{(l+1)} - w_{ik}^{(l)}|\} \leq \varepsilon$ , 则得到所要的  $R$  和  $W$ , 设停止参数  $\varepsilon = 0.01$  (通常取 0.01), 否则继续迭代, 转向 Step4。

Step7: 计算有效性函数  $CS(c)$ , 选择最小的  $CS(c)$  生成最佳聚类数  $c$ , 利用聚类中心得到簇头, 与聚类中心最近的节点当选簇头节点, 并根据聚类结果形成  $c$  个簇。

#### 4 仿真实验

使用 MATLAB 进行仿真实验并与基于模糊控制理论的 CEFL 方法进行比较。仿真实验中将采用理想的环境, 主要考虑传感器节点发送数据、接收数据和进行数据融合所消耗的能量, 通过分析网络中的总能量消耗和存活节点数目评定该协议的性能。

仿真实验是由 100 个水声传感器节点随机分布在  $100\text{ m} \times 100\text{ m}$  的范围内。将从簇头节点分布和节点平均剩余能量两个方面比较文中算法与 CEFL 的性能。利用水声通信能量模型,  $P_0 = 2 \times 10^{-3}\text{ J/b}$  作为数据能被接收的最低功率, 融合功率  $E_d = 5 \times 10^{-4}\text{ J/b}$ , 接收功率  $P_r = 0.2 \times 10^{-3}\text{ J/b}$ , 每个节点的初始能量为  $0.5\text{ J/node}$ 。

从图 1 (圆圈表示簇头节点) 可以看出, 实验仿真 100 次后, CEFL 分簇算法得到的簇头节点分布不够均匀, 有些距离较远的节点可能无法与簇头节点通信, 容易使节点能耗不均匀, 网络生存周期比较短, 这是由于 CEFL 算法中每个节点概率地决定是否成为簇头, 可能存在两个簇头彼此相邻选择的情况。而 FCM-L 算法采用改进的模糊 C-均值聚类算法初始化聚类中心, 能有效控制分簇结果陷入局部最优。从图 2 可看出, 同样在仿真 100 次后, FCM-L 选出的这 5 个簇头节点相较于图 1 分布均匀, 而且簇头节点的个数也比较接近最佳簇头个数, 相对来说能有效延长网络生存周期。

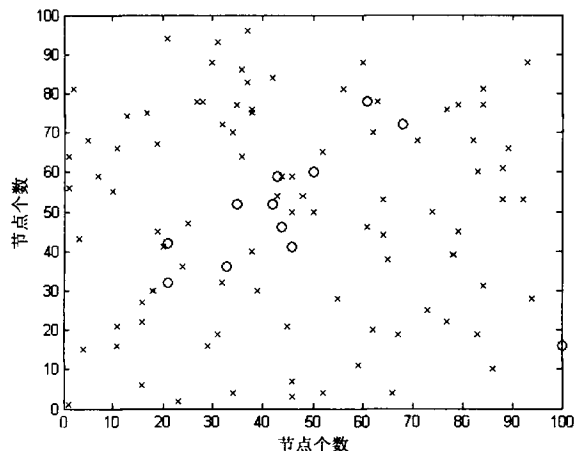


图 1 CEFL 簇头选举结果

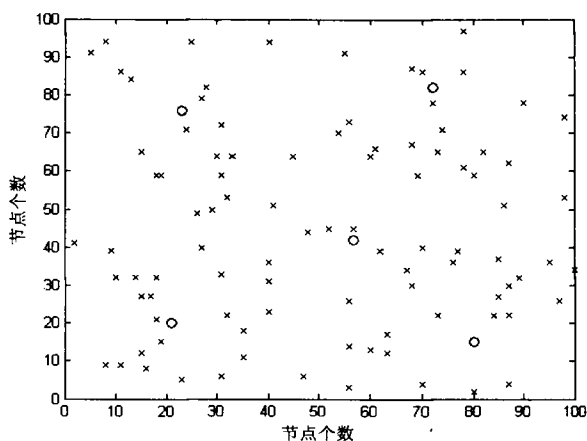


图 2 FCM-L 簇头选举结果

图 3 是两种算法的节点总能量消耗比较。FCM-L 算法相较于 CEFL 算法, 它的能量消耗趋势相对缓慢一些, 因为 FCM-L 算法根据节点间的距离与能量衰减率得到初始化聚类中心, 保证了簇头节点能均匀分布在网络中, 并权衡了簇头节点的能量消耗问题。所以可以看 FCM-L 算法得到的总能量消耗趋势比较缓慢, 而且能量消耗结束的轮数也有一定的延迟。

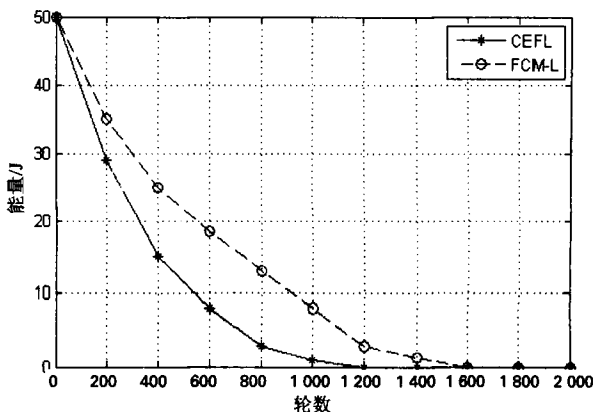


图 3 节点总能量消耗

图 4 是两种算法的死亡节点数量变化的比较。可看到在第 230 轮左右, 利用 FCM-L 算法第一个节点开始死亡, 比 CEFL 算法推迟了 140 轮, 而且在 FCM-L 算法中节点的死亡趋势较为平缓, 所以它的生命周期较 CEFL 算法延迟了 500 轮左右。这是由于 FCM-L 算法以合理的簇头个数进行分簇, 得到一个负载均衡的网络结构, 使得整个生命周期得到延迟。而 CEFL 算法产生的簇头个数具有不稳定性, 导致节点死亡趋势相对较快和不稳定。

图 5 是在不同簇头个数下所有节点的总能量消耗比较。首先  $c = 4$  时, 明显看出它在整个网络中的能量消耗比  $c = 5$ 、 $c = 6$  时的趋势较快, 而且在 1200 轮时能量已消耗殆尽, 簇头个数为 5 和 6 时消耗趋势相差不多, 但是当实验进行到第 800 轮以后, 簇头个数为 5 时的能量消耗趋势逐渐缓慢, 充分说明如果能选出一

个最优的簇头个数将有效地减缓能量的消耗。

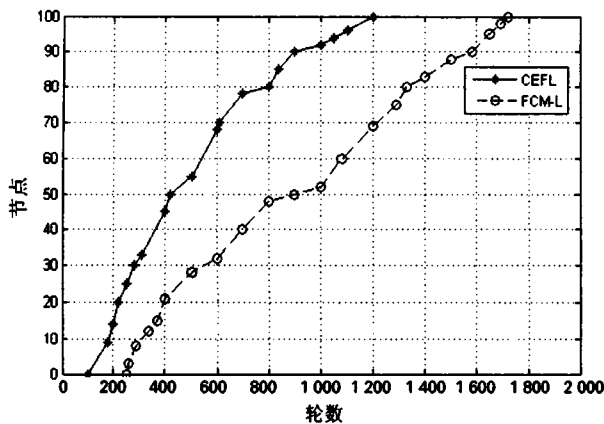


图4 死亡节点趋势图

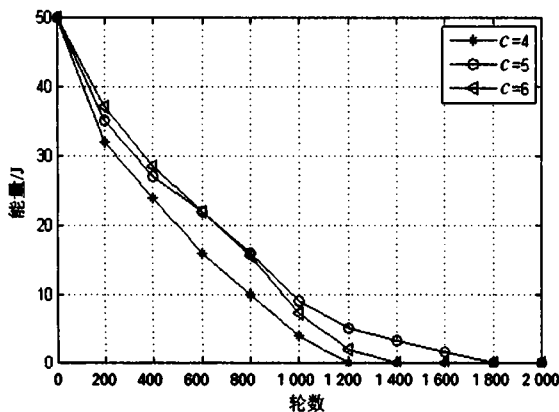


图5 不同簇头个数的总能量消耗

## 5 结束语

提出了一种水声传感器网络簇头选举方法,将水声传感器网络的分簇过程建模为样本空间的模糊聚类划分问题。使用FCM-L对水声传感器网络节点进行聚类分簇,构造一个有效性函数确定划分最佳簇头数。与CEFL相比,FCM-L算法产生的簇头节点在网络中分布更加均匀,从而实现了网络寿命的显著增加,进一步改善了能量消耗。

由于在初始化聚类中心时,在每一次创建数据对象集的时候都要计算节点间的距离,虽然这一操作可以避免簇头节点陷入局部最优,但是也带来了一定的计算复杂度。因此,在以后的研究中可以通过降低计算复杂度来重新设置初始化聚类中心,使得这种分簇路由算法更加适用于水声传感器网络。

## 参考文献:

[1] UMAR A, JAVAID N, AHMAD A, et al. DEADS: depth and energy aware dominating set based algorithm for cooperative routing along with sink mobility in underwater WSNs[J].

Sensors, 2015, 15: 14458-14486.

- [2] 陈秋丽, 何明, 王琰, 等. 面向海洋监测的水声传感器网络节点定位算法[J]. 计算机科学, 2015, 42(2): 29-32.
- [3] 李鑫滨, 王贝, 韩松. 基于Stackelberg博弈的双层水下传感器网络功率分配算法[J]. 计算机应用, 2017, 37(3): 730-735.
- [4] HEINZELMAN W, CHANDRAKASAN A, BALAKRISHNAN H. Energy-efficient protocol for wireless microsensor networks[C]//Hawaii international conference on system sciences. [s.l.]: [s.n.], 2000: 3005-3014.
- [5] 陈东辉. 基于目标函数的模糊聚类算法关键技术研究[D]. 西安: 西安电子科技大学, 2012.
- [6] 黄卫春, 刘建林, 熊李艳. 基于样本-特征加权的可能性模糊核聚类算法[J]. 计算机工程与科学, 2014, 36(1): 169-175.
- [7] YOUNIS O, FAHMY S. HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks[J]. IEEE Transactions on Mobile Computing, 2004, 3(4): 366-379.
- [8] MANJESHWAR A, AGRAWAL D P. APTEEN: a hybrid protocol for efficient routing and comprehensive information retrieval in wireless[C]//16th international parallel and distributed processing symposium. Ft. Lauderdale, FL, USA: IEEE, 2002.
- [9] AL-KARAKI J N, UL-MUSTAFA R, KAMAL A E. Data aggregation in wireless sensor networks-exact and approximate algorithms[C]//Workshop on high performance switching & routing. Phoenix, AZ, USA: IEEE, 2004: 241-245.
- [10] MANJESHWAR A, AGRAWAL D P. TEEN: a routing protocol for enhanced efficiency in wireless sensor networks[C]//15th international parallel and distributed processing symposium. San Francisco, CA, USA: IEEE, 2002: 189.
- [11] GUPTA I, RIORDAN D, SAMPALLI S. Cluster-head election using fuzzy logic for wireless sensor networks[C]//3rd annual communication networks and services research conference. Halifax, NS, Canada: IEEE, 2005: 255-260.
- [12] 王鑫, 蔡莘, 蒋华. 负载均衡的水声传感器网络多跳非均匀分簇路由协议[J]. 计算机应用研究, 2017, 34(3): 853-857.
- [13] YU Haitao, YAO Nianmin, LIU Jun. An adaptive routing protocol in underwater sparse acoustic sensor network[J]. Ad Hoc Networks, 2015, 34: 121-143.
- [14] 张小红, 裴道武, 代建华. 模糊数学 Rough 集理论[M]. 北京: 清华大学出版社, 2013.
- [15] 高新波. 模糊聚类分析及其应用[M]. 西安: 西安电子科技大学出版, 2004: 113-119.
- [16] 邱云飞, 狄龙娟. 基于簇间距离自适应的软空间聚类算法[J]. 计算机工程与应用, 2016, 52(21): 88-93.