

基于初始质心的 K-Means 算法优化

何嘉伦, 马 冲

(新疆大学软件学院, 新疆 乌鲁木齐 830000)

摘要: 由于传统 K-Means 聚类算法对初始质心的赋值具有随机性, 使得模拟结果有着极大的波动。针对这一问题, 文章采用均分法, 首先将样本数据清洗, 将偏差值较大的数据剔除, 然后把处理后的样本数据在二维平面内均匀划分成若干等分, 进行排序归纳计算出最佳初始质心用于聚类算法的初次迭代, 并调用 SSE 等度量指标对迭代的质心进行更新, 最后将样本数据划分成有意义的簇。实验结果表明, 文章针对 K-Means 算法的优化在一定程度上减少了质心的迭代次数, 节省了 time 同时提高了准确率, 对初始质心赋值优化具有有效性和实用性的特点。

关键词: K-Means 算法; 欧氏距离; SSE; 轮廓系数; 方差比准则; DB

中图分类号: TP311.13

文献标识码: A

文章编号: 2096-9759(2023)06-0069-05

K-Means algorithm optimization based on initial centroid

HE Jialun, MA Chong

(School of Software of Xinjiang University, Xinjiang Urumqi 830000)

Abstract: Due to the randomness of the initial centroid assignment of the original K-Means clustering algorithm, the simulation results fluctuate greatly. However, the current research has some deficiencies in the assignment of the initial centroid. In order to solve this problem, this paper uses the mean method to clean the sample data and screen out the data with large deviation, then evenly divide the processed sample data into several equal parts in the two-dimensional plane, sort and calculate the best initial centroid for the first iteration of the clustering algorithm, and call metrics such as SSE to update the centroid of the iteration, and finally divide the sample data into meaningful clusters. The experimental results show that the optimization of the K-Means algorithm reduces the number of iterations of the centroid to a certain extent, saves time and improves the simulation accuracy, and has the effectiveness and practicability of optimizing the initial centroid assignment.

Key words: K-Means algorithm; Euclidean distance; SSE; Profile factor; Variance ratio guidelines; DB

1 引言

K-Means 算法是基于原型的聚类技术, 具有简单、快速并有效处理大规模数据等诸多优点, 是应用最广泛的聚类方法之一^[1]。该算法在文档分类、物品传输优化、消费客户的分类、乘车数据分析等领域都有着广泛且深入的应用, 渗透到生活的方方面面。而传统的 K-Means 算法在质心的初始值设定具有一定的随机性, 在处理大量数据时会引起由初始点定位的不确定性引发的聚类效果与预期值存在偏差, 当初始点的设置过于偏离预期值时会导致算法运行时间过长而降低数据处理的效率。

文献[2]提出从空圆的圆周点中选择初始质心, 则与传统的 k 均值进行随机初始化相比, 可以显著提高 k 均值的性能。由于样本数据不同, k 值会随样本数据的变化而变化, 文献[3]中提出的肘法则思想, 将误差平方和 SSE 作为测量聚类效率的指标, 随着聚类数 k 逐渐增大, 样本数据会更加精细, 每个簇收敛的效果也会逐渐提高。在得到合适的聚类数 k 后, 以方差作为样本离散程度的度量, 选取方差最小且距离大于样本平均距离的 k 个数据点作为数据集的初始聚类中心。

本文采用肘法则思想, 以初始质心的敏感性作为 K-Means 算法优化的突破口, 将初始质心的值设定在一个合理的范围内, 可以有效减少聚类算法处理数据的时间。利用肘法则思想相关方法并且对初始质心合理范围进行设定。基于初始质心的 K-Means 算法优化设定 2k 个初始质心, 在优化排序后筛

选出 k 个质心作为初始质心进行聚类, 既避免了由于初始质心的随机赋值导致聚类不理想的风险, 也在一定程度上避免了不必要的误差平方和计算, 大大节省了 K-Means 算法运行的时间。

2 传统 K-Means 算法

2.1 K-Means 算法定义

K-Means 算法属于无监督学习的聚类算法, 将一组数据样本按照给定的 k 值聚为 k 个簇。其中 means 又被称之为质心, 传统 K-Means 算法将每个类中数据的均值作为质心, 用来描述该簇。算法处理数据的思路大致为: 从初始值中读取给定的 k 值作为质心, 随机给这 k 个质心赋值作为初始值, 计算每个质心与样本数据的距离, 按照计算返回的结果将样本划分到距离最近的质心所在簇中去, 以此进行循环计算, 直到触发终止条件或“簇中心”不再移动, 生成新的质心即新的“簇中心”。

2.2 K-Means 算法分析

2.2.1 K 值的选择

k 值一般情况下由用户直接给出, 或者根据数据处理的需求进行决定。需要注意的是: k 值也称为质心, 即最后数据处理的结果得到的簇数量也是 k 个。k 值在整个聚类的过程中是不变的, 每次只更新 k 个质心的坐标, 质心的数目并不会随着聚类算法的运行而变化。最终数据处理的结果簇个数和用户初始指定的 k 值是相等的, 并且每个样本只能存在于 k 个

收稿日期: 2023-02-14

作者简介: 何嘉伦(2001-), 男, 湖北随州人, 本科, 主要研究方向: 软件工程。

通讯作者: 马冲(1993-), 男, 籍贯河南, 硕士, 新疆大学软件学院讲师, 主要研究方向: 大数据处理、智能电网。

簇中的一个,不能出现“一点多簇”的情况。

2.2.2 度量样本间距离

K-Means 聚类算法的核心思想是采用邻近的度量策略将样本分配到距离自己最近的质心周围,组成一个簇。距离度量的公式大致分为以下三种:

欧式距离

$$d(x, y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1)$$

曼哈顿距离

$$d_{12} = \sum_{k=1}^n |x_{1k} - x_{2k}| \quad (2)$$

余弦相似度距离

$$\cos \theta = \frac{A \cdot B}{|A| \cdot |B|} \quad (3)$$

余弦相似度距离一般用于文档的处理中,有时也可以采用曼哈顿距离作为度量来处理文档。本次的算法优化主要是在欧式空间中,所以采用欧式距离来计算。

2.2.3 更新质心

在计算样本间欧式距离循环完成一次后,需要更新每个簇的质心。传统 K-Means 算法采用均值法来更新质心来进行下一次的迭代。取每个簇样本的坐标求平均值,将产生的均值赋给质心再次计算样本到各质心的欧氏距离。计算每个聚类簇集合 $S_i (1 \leq i \leq K)$ 中数据点的均值为^[4]:

$$c'_i = \frac{\sum_{p \in S_i} p}{|S_i|} \quad (4)$$

对于多次运行 K 均值产生的簇集,选择误差平方和最小的那个,使得聚类的质心可以更好地代表簇中的点,其中 c_i 为簇 C_i 的聚类中心, x 为该簇中的一个样本^[5]。将已更新的均值 c'_i 作为新的质心,以此来计算每个簇的误差平方和 SSE (Sum of Squares Due to Error):

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} \text{dist}(c_i, x)^2 \quad (5)$$

2.2.4 停止运算

在没有给定停止条件下, K-Means 算法默认为当质心不再被更新时,计算结束,返回每个簇的最终质心,输出聚类算法处理的结果。同时,也可以人为设定一系列的条件,包括但不限于给定最大循环次数 Loop_max、目标函数 SSE 收敛等。

3 改进 K-Means 算法

3.1 样本的划分与统计

通过分析二分 K 均值聚类算法的描述可知,其考虑到了随机选取初始质心的 K 均值聚类算法的不确定性,容易使聚类结果走向局部最优化,采用多次使用 K 均值聚类算法进行聚类,选择具有最小 sse 的簇集作为二分结果^[6]。传统 K-Means 算法对初始质心的随机赋值,也会导致由于每次分布过于随机,聚类结果产生偏移,随之导致同一个样本聚类结果存在一定的差异性和不确定性。而二分聚类将整个样本看成一个簇,每次通过在给定 k 值为二的簇上进行 K-Means 聚类,然后计算将该簇一分为二后的总误差,选取较小的误差的簇再次进行划分,直到筛选出最终的质心。受文献[6]启发,分布在二维平面上的样本在纵轴上被一分为二,再将横轴均匀分成 k 份,使得整个样本面被等分成 $2k$ 个格子。其中二维平面尺寸的确定通过遍历一次样本,找到样本的边界点,以边界点作为终止点计算出横轴和纵轴的长度,为接下来划分网格确定范围和计算网格覆盖样本数做好准备。在样本划分工作完成后,

计算出每个格子覆盖的样本数目,将覆盖样本的数目进行统计并按照从大到小的顺序进行排列。计算网格中样本数的算法流程图如图 1 所示:

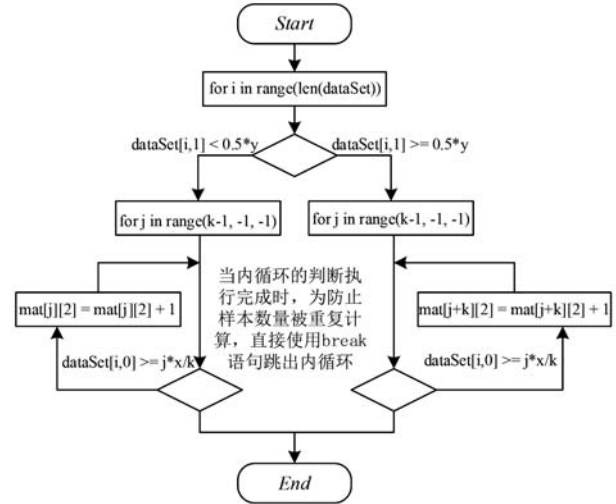


图 1 样本统计算法流程图

考虑到样本划分存在样本处于划分边界的情况,改进的算法统一设置为落在网格左侧与网格下侧,落于网格右上边界的样本不计入当前网格中。由于聚类的簇数可能存在奇数,基于簇数的两倍来划分样本区域一方面考虑到划分的完整性,保证了无论簇数的奇偶,样本总是可以被均匀划分成偶数分,其次,选择在纵轴上均匀的一分为二而不是一分为多,可以有效避免样本中某一簇范围过大导致初始质心集中分布在该簇周围,造成其他簇周围没有邻近的质心增大运算量和计算时间。基于给定质心数目的两倍数量来划分网格就可以很好地避免因某一簇涵盖样本过多而导致质心分配不均的问题,同时根据划分的网格排序后选取前半部分作为初始质心的赋值范围在很大程度上避免了网格数目过多而增加无效的计算,前半部分的网格已经揽括了该样本大部分的簇,剩余所含样本较少的簇也会在接下来的循环更新质心中收敛。

3.2 初始质心的赋值

系统选取前 k 个格子的中心赋值给初始质心。为减少由算法冗余度,在计算网格中心时采用一次循环遍历所有网格,将计算循环的结果直接赋值给类型为 list 的 mat 集合,在循环结束后,调用 sort 函数,排序并筛选出 k 个坐标存储在 new_mat 集合中用于聚类算法对初始质心的赋值操作。通过实验分析发现,网格中心坐标计算符合递增数列的规律,以二维平面网格左下角第一个中心为例,与之相邻的三个网格分别在横坐标、纵坐标、横纵坐标呈现递增趋势。换言之,与第一个中心相邻的网格在纵轴的递增量为边界点纵坐标单位长度的三倍,而横坐标在横轴上与上一中心点递增值是边界点横坐标单位长度的两倍。设定变量 i 作为纵坐标的循环值,由于对所有样本点都采取纵坐标一分为二的操作,故外循环只用循环两次,而内循环次数取决于聚类簇的数量,定义变量 j 来接收循环次数 k。由于网格中心的计算都是基于左下角初始中心依次叠加,在循环中初始中心被单独列出来计算。

3.3 算法优化评价指标

聚类算法的性能度量指标可以用来评价聚类结果的契合度, 通过性能度量指标也可以从侧面更好的设定聚类算法中参数的设置。在无监督类学习算法中最常用到的就是误差平方和 (SSE: Sum of squares of errors), 在本次 K-Means 算法优化中共计选取轮廓系数、Calinski-Harabasz Index (CH)、Davies-Bouldin Index (DB) 三个指标用来度量算法优化的效果。

(1) 轮廓系数

轮廓系数是一种描述 K-Means 算法聚类后各簇轮廓清晰度的性能度量指标。轮廓系数还包含反映样本点与类内元素的紧密程度的内聚度因素和反映样本点与类外元素的紧密程度的分离度因素。设定: a_i = 样本 d_i 与其同簇其他样本的平均聚类; 对于其他非簇 A 的簇 C 而言则令: $D(iC)$ = 样本 d_i 与聚 C 中所有样本的平均距离, 则 $b_i = \min_{C \neq A} \{D(iC)\}$ 假设样本与簇 B 中所有样本的平均距离取得该最小值 b_i , 针对数据集里样本 d_i 假设样本 d_i 被聚类到簇 A 其轮廓系数 S_i 定义如下^[7]:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (6)$$

(2) Calinski-Harabasz Index (CH)

Calinski-Harabasz 又被称之为方差比准则, CH 的度量指标是每个簇之间的距离和簇内距离的比值。该准则可以有效确定 k 值, 值得注意的是, 由于每个质心到全局簇中心距离只计算一次, 为了消除在簇内距离的差异性, 在计算比值时需要乘上每个簇所包含的样本数量, 并且采用无偏估计来计算每个簇之间的距离和簇内距离的比值。CH 度量指标公式如下:

$$S = \frac{tr(B_k)(N-K)}{tr(W_k)(K-1)} \quad (7)$$

上述方差比公式中 B_k 代表类间的协方差矩阵, W_k 为内数据的协方差矩阵, 详细公式如下所示:

$$B_k = \sum_{q=1}^k n_q (C_q - C_e)(C_q - C_e)^T \quad (8)$$

$$W_k = \sum_{q=1}^k \sum_{x \in C_q} (x - C_q)(x - C_q)^T \quad (9)$$

其中 C_q 表示类 q 的中心点, C_e 表示数据集的中心点, n_q 表示类 q 中的数据的数目, C_q 表示类 q 的数据集合。方差比的分值越大代表着聚类效果越好, 等同于类别内部数据的距离平方和越小越好, 类别之间的距离平方和越大越好。

(3) Davies-Bouldin Index (DB)

该度量指标由大卫 L · 戴维斯和唐纳德 · Bouldin 提出, 也是计算距离和的比值, 区别方差比准则的是, 方差比准则计算每个簇之间的距离和簇内距离的比值, 而 DB 是计算类内距离之和与类间距离之比, 通过对 k 值的优化操作, 避免算法在处理数据过程中生成局部最优解的状况。DB 指标以类间的离散程度和类内的紧密程度为依据来评估聚类效果。同一类内的数据对象相似程度越高, 不同类间的相似程度越低, DB 指标值就越小, 分类效果越好^[8]:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j, i, j \in [1, K]} \frac{S_i + S_j}{M_{ij}} \quad (10)$$

DB 在衡量指标中 S_i 代表分散度, 表示在第 i 个类中度量数据点的分散程度, 而 M_{ij} 则表示第 i 类和第 j 类中心之间的距离, 公式如下所示:

$$S_i = \left\{ \frac{1}{n} \sum_{j=1}^n |X_{ij} - A_i|^q \right\}^{\frac{1}{q}} \quad (11)$$

其中 n 表示第 i 类中样本点的个数, X_{ij} 表示第 i 类中的第

j 个样本点, A_i 代表着第 i 类的质心, 当 q 的取值为 1、2 时, 分别表示样本数据到质心的距离平均值和样本点到质心距离的标准差, 该公式也可用于来估计样本的分散程度。

4 实验结果与分析

K-Means 聚类算法优化共选取六个样本数据集测试优化的效果, 六个样本数据集包含了不同的簇数, 有人工生成的数据集, 也包含 UCI 数据集。将数据集清洗分类处理, 分别带入传统 K-Means 算法与优化改进后的算法中作比较, 从算法的迭代次数、优化前后算法运行同一组数据的时间损耗、最终迭代的结果和理论结果的差值三个维度对改进算法进行测试分析。如图 1 所示, 将 K-Means 优化前后的对不同样本迭代处理的时间进行对比, 在处理样本 1 和样本 2 时, 实验数据显示出优化后的 K-Means 算法迭代时间比传统算法略高, 不同于优化算法的初始质心分区赋值, 传统算法的随机赋值随机性较强, 在对样本处理时存在一定的偶然因素。初始聚类算法对初始质心的定位和理论簇的簇心距离较近, 导致最终的迭代次数较少, 能够在短时间内模拟出理想的效果。

K 个质心的选择可以采用随机方式, 质心的初始位置会极大地影响最终的聚类结果及运行时间, 质心之间的距离不应太近, 否则会影响聚类效果^[9]。在图 2 中也能体现出由传统 K-Means 算法偶然性导致的模拟效果优于改造后算法的原因, 在样本 1 中, 优化后算法的迭代次数是传统算法迭代时间的两倍。由于迭代次数和迭代时间是成正相关的, 随着迭代次数的增加, 迭代时间也逐渐变长, 因此对样本 1 的处理, 优化后的聚类算法迭代时间是传统聚类算法的两倍。在对样本 2 进行处理时, 虽然优化前算法的效率高于优化后的算法, 但由于迭代次数相同, 二者在迭代时间上的差距并不大, 相隔了 0.002 秒, 在一定程度上可以忽略不计。

随着样本数据量的增大, 传统算法的随机性优势逐渐变为劣势, 优化后的聚类算法无论是在迭代次数上还是在迭代时间上均优于传统算法。在图 2 中对表格中迭代时间, 优化前的折线和优化后的折线差距随着样本数据的增大差距开始逐渐拉大:

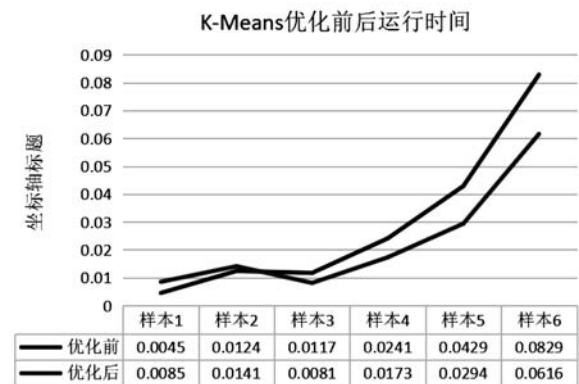


图 2 算法优化前后运行时间折线图

而传统 K-Means 算法的随机性在图 3 中也得到了明显的体现, 相比于优化后聚类算法迭代次数的稳定性, 传统算法的迭代次数处于上下波动的增长的趋势。在研究选取的六个样本中, 传统算法处理数据的结果具有很强的随机性, 同一组数据不同时间运行的结果也相差甚远, 不同于优化后的聚类算法, 不会随着运行时间的不同而输出不同的结果, 处理样本数据较稳定, 产出的数据基本处于同一水平线上。

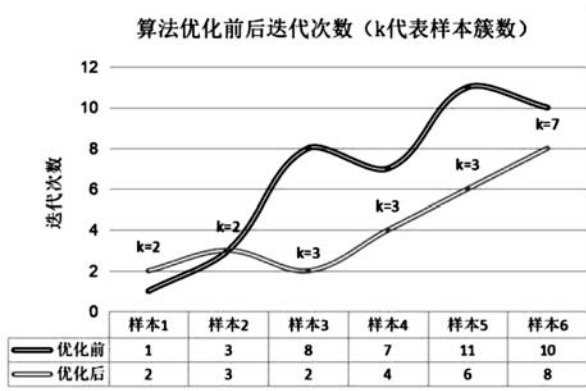


图3 算法优化前后迭代次数折线图

影响聚类结果有效性的因素有三个：其一是聚类方法强加给数据的结构模型是否有效，称为假设结构有效性；其二是在假设结构下特定的聚类方法能否导出与数据匹配度很高的结果，即算法有效性；其三是先验知识的有效性，如用户给定的类别数是否正确^[10]。针对上述影响因素改进的 K-Means 算法不同样本数据的处理结果如图4所示，通过算法处理的输出结果对比理想数据值，计算出传统算法与优化后算法处理结果的准确率进行对比，颜色较深的一组为优化前的聚类算法处理样本数据结果的准确率，浅色代表优化后的聚类算法处理样本数据结果，雷达图涵盖了六个样本的处理结果，可以明显看出优化后的 K-Means 算法占面积较大，从另一侧面反映出优化后的聚类算法是有效果的，处理样本数据的能力要优于传统聚类算法。

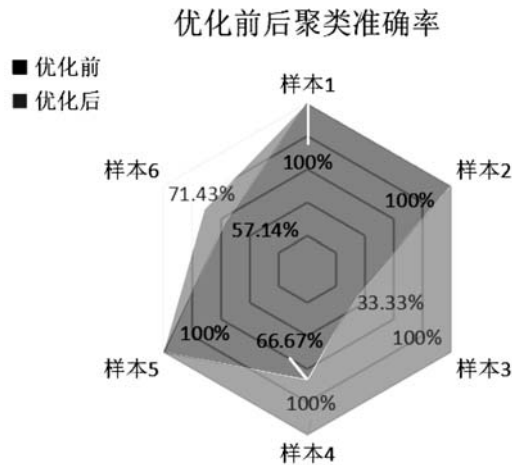


图4 算法优化前后准确率雷达图

4 亟待解决的问题

本次研究 K-Means 算法优化只针对于初始质心的定位与赋值，在样本数据模拟的结果来看，当样本研究中理论簇数较少时，优化后的聚类算法模拟的结果基本与理论值相符合。如图5所示，样本五的理论簇为三，设定好 k 值后，算法通过六次迭代，模拟的结果与理论值高度契合，三个质心均落在簇中央。但随着样本数据量的增大，理论簇数较多时，实际计算结果与理论值开始出现偏差。如图6所示，优化后的算法只有四个质心准确落到了簇中央，有两个簇心出现在同一簇上，还有一个质心落在了两簇之间。根据模拟输出的结果可知，该算法优化针对范围较大的簇有着较好的收敛效果，当样本数据中出现大小不一的簇时，由于算法网格划分从大到小划分

的属性，前 k 个质心初始化时有概率落到样本数据较大的簇中，出现某一簇拥有多个质心，而其他簇没有质心的情况。随着算法的迭代，可能出现多个质心均落在同一簇，并且不同质心的度量指标均在合理范围内，最终造成“一簇多心”的情况，如同图6所示，两个质心落在了同一簇。

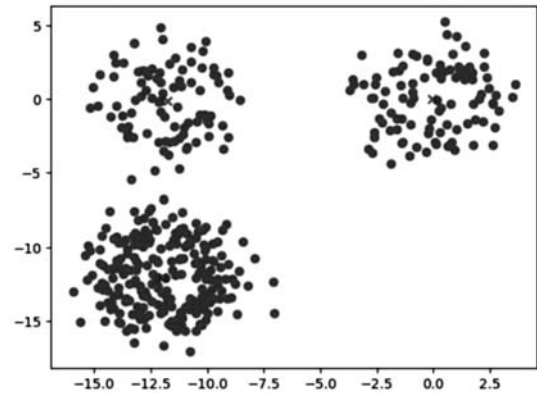


图5 样本五聚类模拟结果

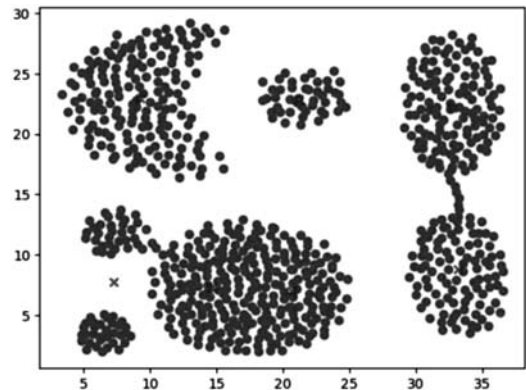


图6 出现“一簇多心”的模拟结果

5 结语

无监督机器学习的聚类算法在商业选址、用户分类、图像分割等领域都有着极为广泛的应用与发展，本次基于初始质心的 K-Means 算法优化研究在一定程度上缓解了初始质心随机赋值的不确定性，保证了样本数据处理结果的稳定性，用更少的迭代次数减少样本数据的处理时间，最大程度上确保数据收敛的准确度。但优化初始质心的算法也存在簇心定位不精确，初始质心容易聚集在含有较大样本数据的簇中等亟待解决的问题需要后期不断地修改优化，解决“一簇多心”模拟结果，带入更多的样本数据测试优化现有算法并进行进一步的研究探讨。

参考文献：

- [1] 张军伟,王念滨,黄少滨,蔺世明.二分K均值聚类算法优化及并行化研究[J].计算机工程,2011,37(17):23-25.
- [2] Biswas,Tuhin Kr. and Giri,Kinsuk.A Novel Approach for Initializing Centroid at K-Means Clustering in Paradigm of Computational Geometry[C].Lecture Notes in Electrical Engineering,2022,(806):545-552.
- [3] Zhao, Huiling.Design and Implementation of an Improved K-Means Clustering Algorithm[J].Mobile Information Systems,2022,(2022):2-5.

(下转第 75 页)

仿真时间为 550s; 节点有效通信距离为 1km; 数据报文长度为 525Byte; MAC 缓冲区长度为 310KB; 节点暂停时间为 0; 发包速率为 5.5pkts/s。搭建完实验环境后, 对本文设计的路由协议性能进行实验检测。

2.2 结果分析

选取集群无人机自组网分组投递率作为本次实验的评价指标, 其计算公式为:

$$P = P_R / P_M \times 100\% \quad (2)$$

其中, P_R 表示集群无人机自组网节点成功接收且成功到达目的地的数据分组数; P_M 表示集群无人机自组网源节点发送的数据分组总数。该评价指标能够有效地评价集群无人机自组网路由协议的性能, P 值越高, 路由协议的性能越好, 反之同理。在集群无人机自组网通信过程中, 设定各个自组网子区域节点传输数据时, 以多跳路由路径传输, 通过中继节点的间接方式, 将数据上传至网络汇聚节点, 进而进行无人机远距离通信数据传输任务。为了使评价结果更加直观, 引入对比分析的实验原理, 将本文提出的基于簇结构的路由协议设置为实验组, 将文献[3]提出的路由协议、文献[4]提出的路由协议分别设置为对照组 1 与对照组 2, 进行对比分析。设定集群无人机自组网节点移动速率分别为 10m/s、20m/s、30m/s、40m/s、50m/s、60m/s、70m/s, 利用 MATLAB 模拟分析软件, 测定节点移动速率逐渐增大的趋势下, 三种路由协议对应的节点分组投递率, 并绘制成图 4 所示的评价指标对比示意图。

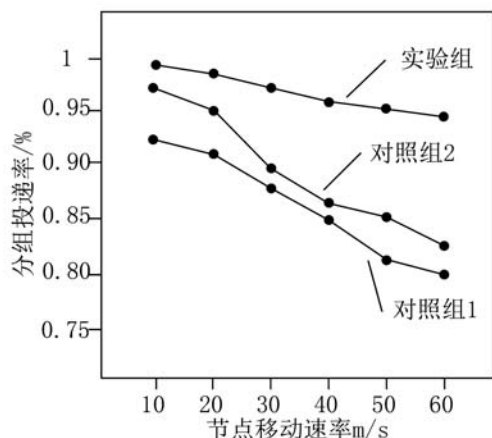


图4 集群无人机自组网节点投递率对比结果

根据图4的对比结果可以看出, 在三种集群无人机自组网路由协议中, 本文设计的基于簇结构的路由协议, 在自组网节点移动速率逐渐增大的趋势下, 节点分组投递率变化较稳定, 无明显波动, 最高分组投递率接近1, 其各个节点对应的分组投递率均高于另外两个对照组。得到如此结果的原因是, 本文设计的路由协议中首先建立了位置预测模型来预测无人

机的位置, 而后划分簇结构, 并对自组网各个簇结构的分簇链路质量进行计算, 提高集群无人机自组网各个链路的质量。而对照组1和2则未能事先预测无人机对应的位置, 使得所构建的路由模型存在一定的误差, 导致各个链路的质量较低。不难看出, 本文设计的路由协议应用后, 自组网节点分组投递率较高, 表明路由协议性能较好, 适用于集群无人机拓扑变化频繁的自组网网络中, 链路的稳定性得到了显著提升, 发生断裂的可能性较低。

3 结语

概而言之, 科学合理的路由协议对集群无人机自组网的稳定运行具有重要的影响, 能够保证集群化无人机协同飞行的能力, 打破信息传输与交互在距离方面的限制, 提升移动信息采集能力。基于此, 本文在传统路由协议的基础上, 引入簇结构, 设计了基于簇结构的集群无人机自组网路由协议。通过本文的研究, 优化了无人机自组网拓扑结构, 提高了路由协议的稳定性与高效性, 降低了由于自组网拓扑频繁导致路由发生断裂的概率。本文设计的路由协议, 在高移动性与高负载性条件下, 均能够表现出良好的性能, 对促进集群无人机自组网的高效运行具有重要研究意义。

参考文献:

- [1] 张珉,董超,杨朋,等.无人机自组网路由协议研究综述[J].数据采集与处理,2022,37(05):952-970.
- [2] 郭晶晶,高华敏,刘志全,等.面向无人机自组网的路由消息完整性保护方法[J].航空科学技术,2022,33(04):28-38.
- [3] 张傲,段续庭,田大新.基于链路质量评估的飞行自组网抗干扰多路径路由协议[J].无人系统技术,2022,5(02):80-92.
- [4] 钟剑峰,王红军.适用于无人机集群应急通信系统分簇路由协议[J].火力与指挥控制,2022,47(02):56-66.
- [5] 姚玉坤,张本俊,周杨.无人机自组网中基于 Q-learning 算法的及时稳定路由策略[J].计算机应用研究,2022,39(02):531-536.
- [6] 梅家栋,南建国.无人机自组网中改进型反应-贪婪-反应路由协议[J].空军工程大学学报(自然科学版),2021,22(04):22-28.
- [7] 孙一凡,米志超,王海,等.基于分簇的拓扑自适应的无人机蜂群 OLSR 路由协议[J].计算机科学,2021,48(06):268-275.
- [8] 李操,杨余旺,肖高权,等.节点密集型无人机自组网路由协议研究[J].舰船电子工程,2021,41(03):93-96+101.
- [9] 王玉,王文灿,白丽,等.基于链路生存时间预测的高动态飞行自组网组播路由协议[J].计算机工程,2021,47(11):198-206.
- [10] 尚建贞.基于拓扑信息的无人机自组网路由协议[J].长江信息通信,2021,34(12):55-57.

(上接第 72 页)

- [4] 孔钰婷,谭富祥,赵鑫,张正航,白璐,钱育蓉.基于差分隐私的 K-means 算法优化研究综述 [J]. 计算机科学, 2022,49(02):162-173.
- [5] 刘广聪,黄婷婷,陈海南.改进的二分 K 均值聚类算法[J].计算机应用与软件,2015,32(02):261-263+277.
- [6] 张军伟,王念滨,黄少滨,蔺世明.二分 K 均值聚类算法优化及并行化研究[J].计算机工程,2011,37(17):23-25.
- [7] 朱连江,马炳先,赵学泉.基于轮廓系数的聚类有效性分析

[J].计算机应用,2010,30(S2):139-141+198.

- [8] 付卫红,马丽芬,李爱丽.基于改进 K-均值聚类的欠定混合矩阵盲估计[J].系统工程与电子技术,2014,36(11):2143-2148.
- [9] 付丽梅.基于 K-means 优化的 SOM 神经网络算法的视频推荐系统 [J]. 软件工程, 2022,25 (10): 17-19+7. DOI: 10.19644/j.cnki.issn2096-1472.2022.010.004.
- [10] 雷小锋,何涛,李奎儒,谢昆青,丁世飞.面向结构稳定性的分裂-合并聚类算法[J].计算机科学,2010,37(11):217-222.