

文章编号: 2095-2163(2021)04-0017-04

中图分类号: U15

文献标志码: A

利用 Kmeans 与蚁群算法的路径寻优方法

彭熙舜, 陆安江, 贾明俊, 卢学敏

(贵州大学 大数据与信息工程学院, 贵阳 550025)

摘要: 科技时代的到来丰富了人民的日常生活,线上购买物品已经成为了一种新的趋势,但随着网络用户的增加,物流配送成了一块大难题。本文提出了利用 Kmeans 算法与蚁群算法的特性,先对客户住址进行聚类分析,再计算出最优路径,以此来降低配送成本,提高配送效率。最终结果表明该组合算法具有可实施性。

关键词: Kmeans 算法; 蚁群算法; 最优路径

Path optimization method using Kmeans and ant colony algorithm

PENG Xishun, LU Anjiang, JIA Mingjun, LU Xuemin

(College of big data and information engineering, Guizhou University, Guiyang 550025, China)

[Abstract] The advent of the technological age has enriched people's daily life. Online shopping has become a new fashion, but with the increase of Internet users, logistics and distribution have become a major problem. This paper proposes to use the characteristics of Kmeans and ant colony algorithm to first perform cluster analysis on customer addresses, and then calculate the optimal route, so as to reduce distribution costs and improve distribution efficiency. The final result shows the feasibility of the combined algorithm.

[Key words] Logistics distribution efficiency; Kmeans algorithm; Ant colony algorithm

0 引言

在当今时代,网购已经成为了一种新的潮流,每年由电商举办的双十一、六一八更是风靡盛行。随着快递数目的剧烈增长,物流的配送工作越来越重要。人们购买的物件也是花样繁多,生鲜快递必须满足时效性,装饰类物品需要避免发生碰撞。物流配送过程中常规应用 GPS 导航,而高效的物流管理取决于两个关键因素,车辆问题(VRP)与资源配置问题,因此路径规划开始被广泛应用于物流管理中。本文提出了利用 Kmeans 聚类将客户购买物品按照特性进行分类,再利用蚁群算法进行路径规划。

1 研究背景与意义

路径规划是近代兴起的新型技术,被广泛应用于机器人避障,无人机避障,防空导弹系统中。人们的日常生活也开始利用这种技术,比如城市交通规划,GPS 智能导航,物流配送管理系统。通常情况下,按照周围环境信息的布局规划,可以将其分为全局和局部两种路径规划^[1]。前者需要在实践之前掌握所有的环境数据,提前作出路径规划判断;后者主要是采取传感器实时得来的局部环境信息,及时

的给出路径规划。本文主要研究解决车辆路径问题(VPR),以物流中心为起点,派出车辆向不同位置的客户进行物资配送,配送任务结束后返回起点,而其中的关键在于将配送效率最大化,车辆使用率最大化。

路径规划一般是由环境建模,路径搜索,路径平滑这3步所组成^[2]。环境建模是起始环节,搭建一个方便计算机规划具体路径的模型,实质上是将物理信息转为数字信息,在空间上相互映射;路径搜索是指在搭建的环境中寻求到达目标的路径信息;路径平滑是在所有的路径信息中选取最优路径。

2 智能算法

2.1 Kmeans 算法简介

网络用户的大幅度增长也带动了数据的产生,面对愈来愈多的用户数据,根据其之间的相似性,进行聚类分析(cluster analysis)。所谓聚类并不是通常意义上的按规格分类,空间中点的汇聚称为一个类簇,不同类簇中任意的两点距离应该大于同一类簇中的任意两点距离^[3]。具体过程可以描述为,第一步设置一个衡量标准,将系统中的个体数据特性计算区别;第二步使用算法将个体汇编并定义为一

作者简介: 彭熙舜(1997-),男,硕士研究生,主要研究方向:信息与信号处理;陆安江(1978-),男,博士,副教授,主要研究方向:图像处理与分析、MEMS 微机电系统。

收稿日期: 2020-12-12

个新的集群。如今存在着多种不同的聚类分析方法,如何根据自身要求选择一种合适的算法来进行数据分析成了当务之急。

Kmeans 又称为 K 均值,是一种划分聚类的算法,具有高效简洁的特点普及率高。其算法步骤清晰易懂,首先根据客户需求设置值为 k 的 n 个初始质心;其次,将质心散开,将系统中的数据点根据自身特性匹配到与其距离最近的质心,同一个质心中的数据点构成了簇。分析簇中的数据点,更新质心的值。按照以上步骤重复进行,直到质心的值不再变化或者达到了设定的终止条件为止。

为了将数据点匹配到与之最近的质心,需要设定方法来测算距离。给定一个样本空间 $x^i = \{x_1^i, x_2^i, \dots, x_n^i\}$ 和 $x^j = \{x_1^j, x_2^j, \dots, x_n^j\}$,将 i, j 设置为样本数, n 表示数据特征。首先计算有序距离度量,根据闵可夫斯基距离公式(1):

$$\text{dist}_{mk}(x^i, x^j) = \left(\sum_{u=1}^n |x_u^i - x_u^j|^p \right)^{\frac{1}{p}}. \quad (1)$$

其次,介绍无序属性距离度量,式(2):

$$\text{VDM}_p(x_u^i - x_u^j) = \sum_{z=1}^k \left| \frac{m_u, x_u^i, z}{m_u, x_u^i} - \frac{m_u, x_u^j, z}{m_u, x_u^j} \right|^{\frac{1}{p}}. \quad (2)$$

其中, m_u, x_u^i 代表在特征属性 u 上取值为 x_u^i 的样本数; z 代表了第 z 个样本簇; $\text{VDM}_p(x_u^i - x_u^j)$ 代表的是在 u 上的两个离散值 x_u^i, x_u^j 的无序属性距离。

Kmeans 在空间中只需考虑数据点与质心问题,它的空间复杂度不高。空间的存储量设定 $O((m+k)n)$, 其中的 m 是数据点数, n 代表了属性数。它的时间复杂性与数据点的数目有关,所需要的时间于 m 呈线性表示,为 $O(I \times K \times m \times n)$, I 代表了收敛的迭代次数。

2.2 蚁群算法

蚁群算法,顾名思义是模仿生物界中蚂蚁特性的算法。这种算法模仿的是蚂蚁从蚁窝中寻找食物所自动生成的最佳路径的过程。在蚂蚁行走的过程中,它们会释放一种被称为信息素的物质,以此来标识自己的行走路径,路径越长的地方,信息素越少,而路径越短的地方,信息素越多,表示了蚂蚁选择走该路径的数量居多。久而久之,信息素少的地方自然也就被淘汰,最后留下来的就是一个优化路线^[4]。

蚁群算法在离散空间中是通过离散的点状散布来选取的信息量的最优解。连续空间中的解空间与离散的不同,是以区域性的方式进行表示,不是用离散的点集合表示。连续空间与离散空间有着3处不

同,首先是观察蚁群的信息量留存方式;其次蚁群在解空间中寻找最优路径方法;以及最后的群体前进策略。

蚁群算法在连续空间下的寻优方法是基于蚁群的初始分布状态,蚂蚁路途中释放的信息量分布状态,整个蚁群的行进方向策略。因此,可以用数学表达式模拟出基本过程:第一步是将群体的初始分布状态表示,根据问题来设置蚁群的大小,例如设置共有 N 个小蚂蚁;将问题的定义区间均等分为 N 个子空间,每个子空间匹配一只小蚂蚁,编号为 i 。因为蚂蚁会变换自己的活动区域,所以所规定的子空间也是随着蚂蚁同步变化的,一一对应。当小蚂蚁随机移动时,可以发现它所携带的子空间会与其它2个相邻的子空间重叠,根据这个重叠的程度,可以推算出2个相邻子空间内的实际蚂蚁变化程度^[5-6]。

按照以上叙述方式,设置提出问题的定义区间是 $[A, Z]$,则当种群内的蚂蚁数目为 N 时,可以得到各个子空间的长度为式(3):

$$L = \frac{Z - A}{N}. \quad (3)$$

因为蚂蚁有着自身运动的移动子空间,而这个移动子空间的长度与基本子空间是没有区别的,所以有 $L_R = L$, 那么蚁群的初始点的坐标分布可以定义为式(4):

$$x_i = A + \frac{\alpha i}{\beta N} \frac{1}{2} \frac{\ddot{\theta}}{\theta} L. \quad (4)$$

蚂蚁所处子空间 i 的左边界为 x_{iL} , 右边界为 x_{iR} , 可以得到式(5):

$$x_{iL} = A + (i - 1)L, \quad x_{iR} = A + i \cdot L. \quad (5)$$

当蚂蚁随机移动 k 时,移动子空间与相邻子空间的重叠度也为 k , 那么可以定义两个相邻的子空间内对应当前蚂蚁的实际个数 N_1 的变化为式(6):

$$\Delta n = \frac{\Delta x}{L_R} = \frac{\Delta x}{L}. \quad (6)$$

所以,当小蚂蚁向右行进时,右边子空间内实际的蚂蚁数目多出了 Δn , 与之相对,左边的子空间实际蚂蚁数目就减少了 Δn 。

接下来根据蚂蚁的分布情况来确定空间中的信息量分布密度。若蚁群在 x_i 处的函数值为 $f(x_i)$, 那么可以规定此时蚂蚁留下的信息量峰值为 M_i , 这样可以根据函数与信息量的大小得出最优路径。假设在某一区间内去实现寻找函数的最小值,那么可以得到相应的信息量分布函数,式(7):

$$M_i = R - f(x_i). \quad (7)$$

其中, R 是设定的常数, 规定 $R > f(x_i)$ 。此时可以发现, 函数值越小, 信息量的分布函数峰值越大。相对的, 如寻找函数的最大值, 当 $f(x_i) > 0$, 可以得到式(8):

$$M_i = R \cdot f(x_i). \quad (8)$$

因此, 可以得到单个蚂蚁所对应的信息量分布函数为式(9):

$$F_i(x) = \frac{M_i e^{-k_i(x-x_i)}}{[1 + e^{-k_i(x-x_i)}]^2}. \quad (9)$$

在得到蚂蚁群分布的总信息量后, 需要确定各子区间内的实际蚂蚁数目^[7]。根据信息量分布函数可以得出当前蚁群在各个子空间内分布积分和为式(10):

$$S N_i = \int_{x_{iL}}^{x_{iR}} \sum_{i=1}^N F_i(x) dx. \quad (10)$$

实际上的各子空间总信息量为式(11):

$$S_j = S N_i + q - p. \quad (11)$$

其中, q 代表上上次的总信息量遗留部分, p 代表信息量的挥发量。所以可以得出实际总空间中的总信息量为式(12):

$$S_\Sigma = \sum_{i=1}^N S_j. \quad (12)$$

按照子空间中的实际总信息量与总空间的信息量可以推算出蚂蚁的分布情况, 得出蚂蚁在某一子空间中的数目为式(13):

$$N_{iM} = \frac{S_j}{S_\Sigma} N. \quad (13)$$

3 仿真分析

利用 Kmeans 算法与蚁群算法思想, 事先构思出聚类规划和寻优路径的新型路径配送方式。将文本数据导入程序, 设 N 为样本数量, n 是样本中的属性数; 将蚂蚁群分成 4 个小组; 蚂蚁总数目设为 R , 最大迭代次数用 T_{\max} 表示, 表 1 为不同组参数设置表, 其中 Min 代表蚂蚁到其对应的聚类中心的距离最小值。

此外, 还需要注意偏离误差的计算, 即为蚂蚁到其对应的聚类中心的距离 Min , Min 越小, 说明蚂蚁越集中, 聚类水平高。倘若得到每一只蚂蚁的 Min 值, 那么从中挑选出最小的 Min , 其所对应的路径就是本次迭代的最佳路径。而迭代的方法就是利用循环, 更新蚁群的信息素, 根据新的参数进行寻优计算, 直到满足要求为止。当 $R = 100$, $T_{\max} = 1\ 000$

时, 仿真分析如图 1, 当 $R = 100$, $T_{\max} = 10\ 000$ 时, 如图 2 所示。

表 1 程序参数设置表

Tab. 1 Program parameter setting table

R	T_{\max}	Min	Time/ S
10	100	50 431	2.100 7
10	1 000	45 045	7.242 4
10	10 000	32 800	68.523 4
100	100	52 109	11.693 2
100	1 000	30 690	47.262 1
100	10 000	19 726	359.875 9

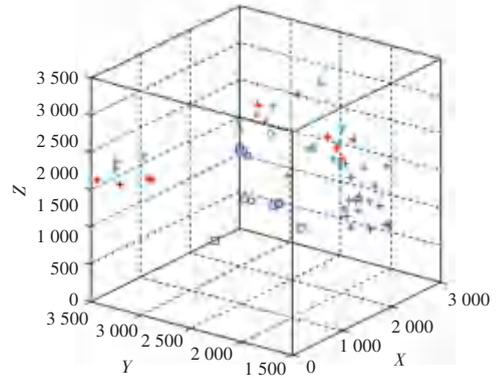


图 1 第一次蚁群聚类结果

Fig. 1 The first ant colony clustering results

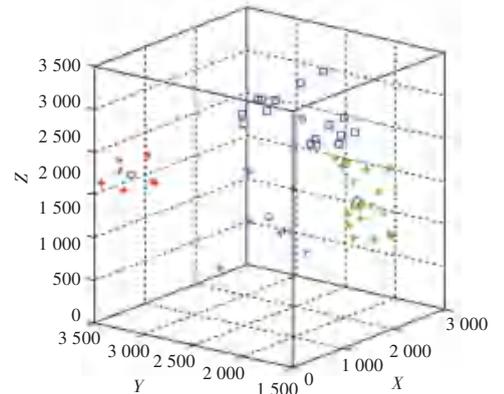


图 2 增大迭代次数后的聚类效果

Fig. 2 The clustering effect after increasing the number of iterations

从图 1 中可以清晰地看到, 共有 $R = 4$ 种蚂蚁组数, 此时的 Min 是 30 690, 达到了一定的聚类效果, 但还不能满足要求, 接下来继续增大迭代次数。将 T_{\max} 增大到 10 000 可得以下结果:

由图 2 知, 此时的 $Min = 19\ 726$, 相比于第一次聚类效果, 在达到良好的聚类分析后, 需要找出一条合适的路径, 从初始点到聚类中心, 接下来需要对路径进行寻优规划。此时用 400 个单位方块模拟环境地图, 边长设置为 1, 以 $(0, 0)$ 作为初始点, $(20, 20)$ 作为终点, 图 3 和图 4 中随机设置障碍物, 并用矩阵存储每一代的每一只蚂蚁的爬行路线长度。

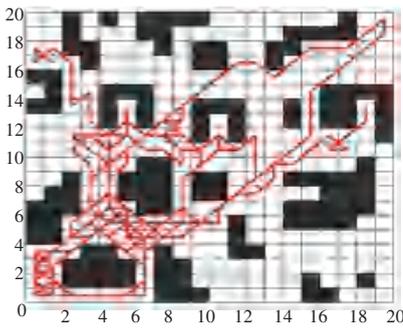


图 3 蚂蚁爬行路线

Fig. 3 Ant crawling route

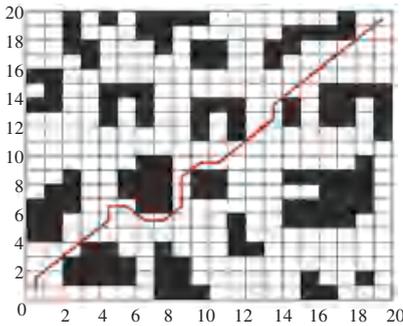


图 4 蚂蚁无碰撞最优路径

Fig. 4 The optimal path of ants without collision

由图 3 可知,蚂蚁从起始点到终点路线数量复杂,尤其是起始点附近,因为蚂蚁刚从蚁穴出来,数量众多,分泌的信息素也多,因为路径更加的繁琐。在接近终点处,由于路程长度,信息素散发稀释,所以蚂蚁的行进路线自然减少。蚂蚁此次路程的最佳

路径如图 4 所示,在无碰撞的情况下保证路程长度。

4 结束语

利用 Kmeans 与蚁群算法能够很好地解决物流配送问题,尤其是当下不同种类的物资需要不同的配送方式。物流中心在接到配送订单时,可以利用该组合算法对配送物品进行聚类分析,可以按照物件的时效性,安全性等作为路径规划,这样就能更好的满足客户的需求,同时也减少了物流配送的时间成本。Kmeans 与蚁群算法对当下的快递行业有着一定的参考意义。

参考文献

- [1] 高健博. 无人驾驶汽车运动规划方法研究综述[J]. 汽车工业研究, 2018 (8): 37-42.
- [2] 熊慕舟, 黎勇. 一种基于多智能体的二层路径规划模型研究[J]. 计算机科学, 2016, 43(6): 59-64.
- [3] 张春涛, 马千里, 彭宏, 等. 基于条件熵扩维的多变量混沌时间序列相空间重构[J]. 物理学报, 2011, 60(2): 118-125.
- [4] 张立. 基于新闻评论数据的 K-means 聚类算法的研究[D]. 太原: 太原理工大学, 2010.
- [5] 罗春芳, 张国华, 刘德华, 等. 基于 Kmeans 聚类的 XGBoost 集成算法研究[J]. 计算机时代, 2020 (10): 12-14.
- [6] 熊达峰, 孙英华, 刘雪庆. 改进状态转移策略的蚁群算法求解 TSP 问题[J]. 青岛大学学报(自然科学版), 2019, 32(1): 112-115, 120.
- [7] 黄敏, 靳婷, 钟声, 等. 基于改进蚁群算法求解连续空间寻优问题[J]. 广西师范大学学报(自然科学版), 2013, 31(2): 34-38.

(上接第 16 页)

(5) 第五类站点: 重要度等级是 5-8 级的站点, 为各线市区段的站点。受是否为换乘站及到城市中心站点的远近影响, 客流也因是否为换乘站及受城市中心站点的辐射程度产生差异, 因此对于重要度有所区分。

(6) 第六类站点: 重要度等级是 9-11 级的站点, 为各线位于市中心附近的站点, 因图论指标和客流指标产生分级。图论指标和客流指标都较大, 因此站点非常重要。

(7) 第七类站点: 重要度等级是 12 级的站点, 为市中心地标性的站点。除了换乘客流还有很多商业观光客流, 位于网络中心, 自身介数和接近中心性指标最大, 级别也最高。

3 结束语

本文通过 Space-L 构建了上海地铁线网模型, 选取图论中衡量节点重要度的 3 个指标, 从自身的连通性、在线网中所处的位置和“流量”方面评价站

点的重要性, 结合现实中客流量因素和城市中心站的辐射影响, 运用 K-means 聚类算法全面的对全网站点进行聚类研究和评估。根据聚类结果, 站点的重要度等级被分为 12 级并将这 12 级归结为 7 类站点。通过研究 7 类站点的特征, 为今后新建车站的资源分配和管理模式的选择提供决策依据。

参考文献

- [1] 袁若岑, 王丽琼, 温志伟, 等. 基于图论的城市轨道交通网络连通性研究[J]. 现代城市轨道交通, 2013 (4): 68-71.
- [2] 陈扶昂, 吴海军. 基于定量分析的城市轨道交通车站分级探讨[J]. 现代城市轨道交通, 2010 (3): 78-80.
- [3] 谢涵洲, 田锋. 基于交通与空间活动功能复合的城市轨道交通站点分级体系研究[J]. 交通科技, 2015 (4): 153-154.
- [4] 李朴, 武子晗. 基于复杂网络理论的城市轨道交通网络结构特性[J]. 都市快轨交通, 2014, 27(5): 35-38.
- [5] 高勃, 秦勇, 肖雪梅, 等. 基于 K-means 的北京地铁路网重要度聚类分析[J]. 交通运输系统工程与信息, 2014, 14(3): 207-213.
- [6] 戴子文, 谭国威, 戴子龙. 城市轨道交通车站分类及等级划分研究[J]. 都市快轨交通, 2016, 29(4): 38-42.