

基于聚类算法的订单分批策略研究

Research on order batching strategy based on clustering algorithm

秦馨¹, 赵剑道², 任楠²

QIN Xin¹, ZHAO Jian-dao², REN Nan²

(1.北京机械工业自动化研究所, 北京 100120; 2.北自所(北京)科技发展有限公司, 北京 100120)

摘要: 订单分批是物流中心提高作业分拣效率的一个有效方法。主要针对“货到人”模式下的订单分批问题, 构造了基于聚类算法的数学模型, 并以最小化货物搬运次数为目标函数进行求解。通过在不同规模的订单数据下, 比较基于聚类算法的分批策略与采用先到先分批策略、不分批策略在拣选作业中的实际效果, 表明该算法能有效提高分拣任务的作业效率。

关键词: 聚类算法; 订单分批; 拣选效率

中图分类号: TP27

文献标识码: A

文章编号: 1009-0134(2021)01-0134-04

0 引言

近年来, 在新经济发展环境下, 市场逐渐变得多元化, 消费需求因此开始向多样化进行转变, 商品的供应订单也随之呈现出碎片化、零散化的趋势。越来越多的企业倾向于通过一个集订单处理、仓储管理、拣货配送于一体的物流配送中心来提高整体物流供应链的供应效率。而在配送中心的所有作业中, 拣货作业的自动化程度相对很难提高, 作业时间占据比重较大, 是制约整个物流配送流程效率的关键因素。因此为了提高仓储运作效率、缩短事务处理时间与降低物流成本, 如今物流中心进行配送管理最重要的目标之一是规划合理的拣货方法。而在对拣货作业流程进行控制优化的策略中, 选择合理的订单分批方法是最有效且最关键的研究重点之一。

1 订单分批

订单分批是指将大批量的订单按照一定的规则和方法进行统一的分类整合, 最终形成不同批次, 每一个批次都包含多个订单, 然后按照不同的批次合并来进行货物分拣作业。订单分批通过缩短整体的拣选路径、减少货物的搬运次数, 显著的提高了拣选效率、节省了劳动力、减少了拣选成本。

早期的订单分批问题研究主要根据项目应用经验进行优化, 一般按照某种固定的优先级来确定分批策略, 最为常用的为先到先分批策略。这种传统策略没有考虑订单之间的相似性, 效率优化效果有限。随着订单数目的增加, 在拣选方式上, 大部分学者开始采用启发式算法对这一方向进行了深入的研究, 例如种子算法、包络算法、节约算法、k-means算法及其各种组合算法。但

是这些算法或受初始订单的选择影响较大, 分批效果不稳定, 或对订单的本身属性考虑不足。而且大部分研究都是针对“人工拣货”的应用场景, 对于“货到人”模式的适用性不足, 因此对“货到人”拣选系统的订单分批优化策略值得探究。

本文基于“货到人”的实际应用场景, 采用基于凝聚型聚类算法的订单分批策略, 以最小化货物搬运次数为目标构建了订单分批问题的数学模型, 通过计算相似系数, 在一定的约束条件下得到了有效的分批结果。

2 模型与假设

假设某段时间内存在 N 个订单 X_i ($i=1, 2, \dots, n$) 需要进行分拣作业, 其中, 每个订单中有若干个品项, 相同品项的存储位置视为同一储位。现根据货到人系统的实际工作场景, 无需考虑人员或设备的行走距离以及不同货物间的储位关联程度, 所有订单的拣选时间和拣选距离与订单品项的搬运次数可近似于线性相关, 同时每个拣选工位可同时拣选的订单数固定。因此, 本文中订单分批的总目标为, 在每个订单不允许分割且每批拣选订单数固定的情况下, 将 N 个订单进行分批拣选, 使得分批后货物总的搬运次数最少。假设每批订单数固定为 C , 分批后的批次总数为 T , 第 j 批订单的搬运次数为 L_j , 总的搬运次数为 L 。则该订单分批问题建立的数学模型为:

目标函数:

$$L = \min \sum_{j=1}^T L_j \quad (1)$$

约束条件:

收稿日期: 2019-05-25

作者简介: 秦馨(1996-), 女, 山西吕梁人, 硕士研究生, 研究方向为控制理论与控制工程。

$$\sum_{j=1}^T x_{ij} = 1, \quad i = 1, 2, \dots, n \quad (2)$$

$$\sum_{i=1}^n x_{ij} \begin{cases} = C, & j = 1, 2, \dots, T-1 \\ \leq C, & j = T \end{cases} \quad (3)$$

$$x_{ij} = \begin{cases} 1, & X_i \in T_j \\ 0, & X_i \notin T_j \end{cases} \quad (4)$$

目标函数式(1)表示订单分批后, 每批订单搬运货物次数总和最小, L_j 为第j批订单的搬运货物次数(即为同一批中各个订单合并后所有品项的搬运次数, 同一品项按照搬运一次计算)。约束条件式(2)表示每个订单在分批后只允许在其中一个批次出现, 既不能分割订单, 也不能重复分配订单。式(3)表示订单分批后每一批订单数量为固定值, 前T-1批订单数量为C个, 第T批订单数量不大于C个。式(4)为订单分批结果, x_{ij} 表示订单 X_i 的分批结果, 若该订单分配到第j批中, 取值为1; 若不存在, 则取值为0。

3 算法

聚类算法是一种常见的分类算法, 已经大量应用于自然科学和社会科学等不同的领域中, 如统计学、信息检索、生物学和机器学习等。不同于决策树、支持向量机等分类算法, 聚类是一种典型的非监督型的学习算法, 是对于已经获取的没有预先进行标记和确定其类别的数据, 根据数据本身的相似性来将样本分到不同的组别, 尽可能使得每组内的数据之间是相似的(相关的), 而不同组的数据之间是不同的(不相关的)。同一组内的数据间相似性越高, 距离越小, 不同组间的数据差别越大, 距离越大, 聚类效果就越好。

本文采取的聚类算法为凝聚型层次聚类, 也称为合并型聚类算法, 是一种自底向上的算法。它是从每个点作为一个个体簇开始, 然后重复地将两个最靠近的簇合并为一个, 直到获取满意的聚类结果为止。凝聚型层次聚类的算法原理是:

- 1) 定义簇的邻近性概念, 计算邻近度矩阵;
- 2) Repeat;
- 3) 将邻近度最高的两个簇合并为一个簇;
- 4) 按照最新的簇的划分来进行邻近度计算, 确定其邻近性;
- 5) Until只剩下符合要求的簇。

订单分批问题的聚类分析就是通过一定的邻近度计算, 将所有订单分成不同的批次, 然后对每一批的订单进行集中拣选, 其目的是减少货物的搬运次数。

3.1 特征向量

邻近度的计算可以通过特征向量来实现。在订单分批问题中, 每一个订单包含多个货物品项, 每一个品项的储存位置已知。由于本文的订单分批基于货到人系统, 人无需去货架中行走拣选货物, 而且在自动化仓库的出库模式下, 每次只能搬运一种货物出库, 因此在进行邻近度计算时, 不需要考虑不同品项储存位置之间的相似性以及行走距离的远近。

利用订单中包含的货物种类来生成每个订单的特征向量, 将每一个品项定义为特征向量的一个维度, 根据单个订单中是否包含该品项定义该维度的值。假设定义总的货物品项数目为m, 每种货物索引为 I_j ($j=1, 2, \dots, m$), V_i 为订单 X_i ($i=1, 2, \dots, n$)的特征向量, 则定义:

$$V_i = \{v_{i1}, v_{i2}, \dots, v_{im}\}$$

其中:

$$v_{ij} = \begin{cases} 1, & I_j \in X_i \\ 0, & I_j \notin X_i \end{cases}$$

3.2 邻近度

邻近度是对两个对象相关性的量化描述, 两个数据越相似, 邻近度越高。聚类算法中, 邻近性度量的选择与计算非常重要, 对聚类的效果好坏有着极大的影响。目前最为常用的邻近性度量计算方法是基于欧几里得距离和余弦相似性。

3.2.1 欧几里得距离

欧几里得距离也称欧氏距离, 是一个通常采用的两点间的距离定义, 表示在多维空间之中这两个点之间的真实距离, 两个对象之间的欧氏距离越大, 则相似性越低。使用欧氏距离计算任意两个订单间邻近性的公式为:

$$s_{ij} = \frac{1}{\sqrt{\sum_{k=1}^m (v_{ik} - v_{jk})^2}}$$

其中, v_{ik} , v_{jk} 分别表示第i、j个订单所对应的特征向量之中第k维的特征值。

3.2.2 余弦相似性

余弦相似性是通过计算两个向量所形成的夹角余弦值的大小来衡量它们之间的邻近性, 邻近性越高的两个对象, 它们的余弦相似性越接近于1, 邻近度越低, 他们的余弦相似性越接近零。余弦相似性的值是和向量本身的长度无关的, 只和向量的指向方向相关。使用余弦相似性计算任意两个订单间邻近性的公式为:

$$s_{ij} = \frac{V_i \cdot V_j}{\|V_i\| \|V_j\|} = \frac{\sum_{k=1}^m v_{ik} \times v_{jk}}{\sqrt{\sum_{k=1}^m (v_{ik})^2} \times \sqrt{\sum_{k=1}^m (v_{jk})^2}}$$

其中， v_{ik} ， v_{jk} 分别表示第 i 、 j 个订单所对应的特征向量之中第 k 维的特征值。

3.3 算法流程

根据以上分析以及邻近性度量系数的选择，使用聚类算法进行订单分批的具体算法流程如下：

1) 根据每个订单所包含的品项，推导每个订单的特征向量 V_i 。

2) 根据订单的特征向量计算出每两个订单之间的相似系数 s_{ij} 。

3) 将所有的相似系数 s_{ij} 倒序排列，若存在相同的相似系数，则按照两个订单共同品项和总品项的数量进行排序。

4) 选取相似系数最大的两个订单合并为同一订单，判断新订单是否满足约束条件，若不满足，选择下一个 s_{ij} ，直到满足所设定的约束条件，计算新订单的特征向量。

5) 判断新订单包含订单数是否达到一批订单上限，若没有，将新订单并入原有的订单数据，若达到上限，将分配完成的一批订单输出，然后重复第2)、3)、4)步，直到所有订单分批结束。

4 仿真

4.1 实验方案

实验数据来源于随机生成的2000个订单数据，根据测试需要，只选取其中的订单号和包含品项的信息。其中，总的品项数量定义为50种，依据现实使用场景，单个订单包含品项数基本集中在2~20种之间，订单分批后每批订单的数量固定为5个。将2000个订单中每200个订单分为一组，共随机分为10组。

本文采用基于聚类算法的分批策略进行研究，衡量该算法在订单分批问题中的优化效果，并且分别采用欧几里得距离和余弦相似性进行相似系数的计算，比较在使用聚类算法来进行订单分批时，不同的邻近度计算方法在这一问题应用中的优劣。并且为了更好的验证本文提出的聚类算法在订单分批问题中的优势，将相同的数据分别采用不分批的策略和先到先分批的策略进行计算，作为本文所提出算法的对照参考组。

每次实验都从10组数据（每组数据包含200个订单）中分别随机抽取20、50、100、150、200个订单来

进行分批计算，然后将10组数据的计算结果取其平均值作为最终分批结果。

4.2 实验结果分析

实验采用了基于欧几里得距离的聚类算法、基于余弦相似性的聚类算法、先到先分批的策略和不分批策略四种方案进行计算，对比验证不同方案的拣选效率。通过计算分批问题的目标函数值，即所有订单品项总的搬运次数作为衡量分拣效率的指标。对10组数据的实验结果计算其平均值如表1所示。由表1可得，在不同规模的数据集下，不采用订单分批策略的方案最终的品项搬运次数最高；采用先到先分批策略进行订单分批后，搬运次数有所下降；而基于聚类算法的分批策略，使得订单总的搬运次数最少。因此，基于聚类算法的分批策略在订单分拣作业中，对拣选效率有着很大提升。

表1 分拣作业总的搬运次数

单位：次

| 算法类型 \ 订单数量 (个) | 20 | 50 | 100 | 150 | 200 |
|-----------------|-----|-----|------|------|------|
| 基于欧氏距离的聚类算法 | 125 | 294 | 556 | 814 | 1062 |
| 基于余弦相似性的聚类算法 | 124 | 286 | 536 | 783 | 1024 |
| 先到先分批算法 | 143 | 350 | 692 | 1041 | 1391 |
| 不分批算法 | 217 | 536 | 1052 | 1581 | 2111 |

图1为不同的分批策略下，最终的货物搬运次数与不分批策略搬运次数的比值。由图1可得，在不同规模的订单数据集下，先到先分批策略对于拣选效率的优化效果比较稳定，在65%~70%之间，而随着订单规模的增加，基于聚类算法的分批策略对拣选作业的优化效果有着显著提高，当订单数达到200个时，货物搬运次数可降至不分批策略的50%左右。其中，不论订单规模的大小，采用余弦相似性计算邻近度的聚类算法在拣选效率的优化上都要更优于基于欧氏距离的聚类算法。

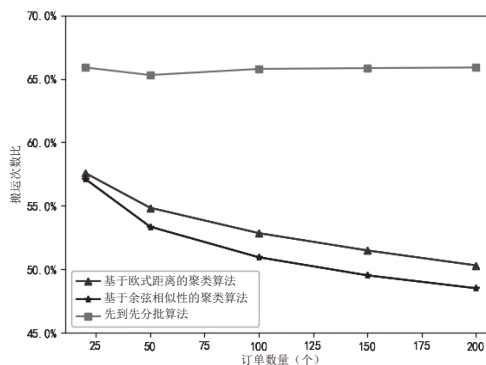


图1 分批策略与不分批策略的搬运次数比

本文研究的订单分批问题是基于实际应用场景的,因此在减少订单总的搬运次数的基础上,应当考虑订单分批的算法计算时间,这同样影响着订单拣选的效率。对10组实验的计算时间计算其平均值如图2所示。由图2可得,对于不同规模的订单,先到先分批策略的计算时间没有明显变化,看忽略不计;而随着订单规模的增加,聚类算法的计算时间增加较明显,其中,基于余弦相似性的聚类算法所消耗的时间始终少于基于欧氏距离的聚类算法,且随着订单数的增加,两种算法计算时间的差距也在逐步增加。

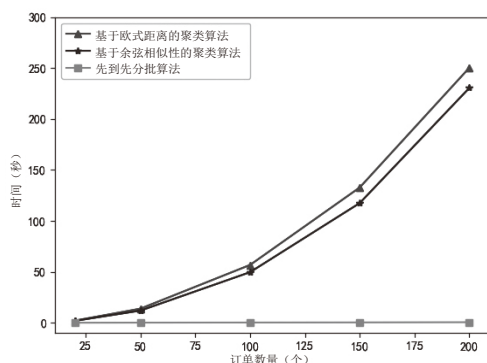


图2 订单分批计算时间

5 结语

本文在自动化仓库“货到人”模式的应用场景下,针对订单分批问题,构造了以最小化订单品项总的搬运

次数为目标函数的数学模型,并且搭建了基于邻近度度量的聚类算法模型,与不分批的策略和传统的采用先到先分批的策略相比,大大减少了货物搬运次数,有效的提高了订单拣选效率,虽然分批算法的计算时间相比较而言有所增加,但仍控制在250s以内,与自动化仓库货物的搬运时间相比存在着较大的差距,因此可忽略不计。而且,在进行邻近度计算时,提出了使用欧几里得距离和余弦相似性两种不同的相似系数的计算方法,其中,采用余弦相似性作为邻近度的聚类算法无论在提高拣选效率方面还是缩短计算时间方面表现都更加优秀。

本文主要针对“货到人”模式下的自动拣选,分拣人员不必在货架间行走,无需考虑行走距离和拣选路径的选择,未来可对“人到货”模式下的拣选作业进行进一步的研究。

参考文献:

- [1] 李诗珍,杜文宏.基于聚类分析的订单分批拣货模型及启发式算法[J].统计与决策,2008(12):53-56.
- [2] 朱友琼,唐思,何进.“货到人”模式下电商物流中心订单分批策略研究[J].现代物流与采购,2019,26.017:34.
- [3] Pang-Ning Tan,Michael Steinbach,Vipin Kumar.Introduction to Data Mining[M].北京:人民邮电出版社,2015:320-327.
- [4] 胡小建,韦超豪.基于Canopy和k-means算法的订单分批优化[J].合肥工业大学学报,2017,40(3):414-419.
- [5] 张彩霞.基于“货到人”模式的电商订单拣选优化研究[D].浙江理工大学,2016.

【上接第125页】

参考文献:

- [1] 杨锋.舟山沿岸渔业现状调查与管理对策研究[D].浙江海洋大学,2017.
- [2] 郭建伟.基于Wi-Fi自组网的渔船视频监控系统的研究[D].集美大学,2014
- [3] 陈明,冯国富,池涛.渔业船联网技术[M].北京:科学出版社,2018:101-102
- [4] 张璘,许自程,冯陈伟.基于树莓派的远程网络视频监控系统[J].电视技术,2018,42(10):92-97.
- [5] 张燕,汪晓红,王晴.基于树莓派云视频流媒体的远程监控系统[J].单片机与嵌入式系统应用,2018,18(11):45-47,54.
- [6] 孟祥飞,李博,刘伟杰.基于ARM11的H.264实时监控系统[J].实验室研究与探索,2018,37(12):87-89,195.
- [7] 屈伟明.基于物联网的智能家居远程控制系统设计与实现[J].电子技术与软件工程,2014(05):100,208.
- [8] 王聪惠.无线音视频同步监控系统的设计与实现[D].重庆邮电大学,2017.
- [9] 李源.基于树莓派的无线视频监控系统关键技术研究[D].郑州大学,2016.
- [10] 韩明.嵌入式视频监控系统的研究和实现[D].西安电子科技大学,2014.
- [11] 周斌.基于ARM的嵌入式视频监控系统[D].安徽理工大学,2017.