

包装技术与工艺

基于数据挖掘的动态货位指派系统

徐翔斌，李秀

(华东交通大学，南昌 330013)

摘要：目的 为提高需求快速变化、波动较大的在线零售企业的仓库货位优化效率。**方法** 利用数据仓库和数据挖掘法，研究基于复合规则的动态货位指派策略。对该货位指派的数据集成分析、指标计算、规则生成和货位指派等 4 个模块进行分析，并设计库区标定算法和规则生成算法来生成货位指派规则集。**结果** 基于复合规则动态货位指派不仅能够节约拣货距离，而且拣货效率受需求变化的影响非常小。**结论** 数值实验表明，与传统的货位指派策略相比，基于复合规则动态货位的指派系统可以得到更好的结果，并且在平均订单规模较大和需求偏度大的情况下效果更加明显。

关键词：数据挖掘；动态货位指派；复合规则；库区标定算法；决策树

中图分类号：TB482 文献标识码：A 文章编号：1001-3563(2017)19-0128-05

Data Mining-based Dynamic Storage Location Assignment System

XU Xiang-bin, LI Xiu

(East China Jiaotong University, Nanchang 330013, China)

ABSTRACT: The work aims to improve the optimization efficiency of storage location of the online retail business subject to the rapid change in demand and large fluctuation. The composite rules-based dynamic storage location assignment strategy was studied based on the data warehouse and data mining technology. The date integration analysis, index calculation, rule generation and storage location assignment of such storage location assignment were analyzed. Moreover, the zone calibration algorithm and rule generation algorithm were designed to generate the storage location assignment rule set. The composite rules-based dynamic storage location assignment could save the picking distance, and its picking efficiency was insensitive to the change in demand. Numerical experiments show that, compared with the traditional storage location assignment strategy, the composite rules-based dynamic storage location assignment system can get better results, and especially in the larger-sized average order with high demand skewness, the effect is more obvious.

KEY WORDS: data mining; dynamic storage location assignment; composite rules; zone calibration algorithm; decision tree

在线零售业的迅猛发展使得拣货和配送效率成为焦点话题，客户对配送的时效性要求越来越高，如何以更低的成本和更快的速度将商品送到客户手中，是在线零售企业获得竞争优势的关键，使得在线零售企业不断提升其拣货效率。拣货系统是典型的劳动密集型作业过程，其作业劳动量占仓库作业量的 60%以上，其中货位指派是提高仓库效率的重要途径之一^[1]。货位指派指将每种品项（SKU，复数为 SKUs）分配到合适的货位，使订单分拣过程的行走距离最短，科学的货位指派方法会缩短行走距离，降低搜寻时间，

大幅度提高仓库分拣效率。

Hausman 等最早提出定位存储、随机存储和分类存储这 3 种基本货位指派策略^[2]，随机存储能够提高仓库空间利用率，定位存储可有效地降低仓库的拣货成本，分类存储则同时考虑存储空间利用率和拣货成本进行货位优化；随后大量的文献对基于出货量^[3]、库存周转率^[4]、拣货强度^[5]以及订单体积指数法^[6]的分类存储问题进行研究。由于基本货位指派没有考虑 SKUs 之间需求或结构关联性，Frazelle 提出基于需求关联性的货位优化^[7]，将具有需求关联性的 SKUs 放

收稿日期：2017-04-05

基金项目：国家自然科学基金（71761013, 71540039）；江西省自然科学基金（20151BAB207060）

作者简介：徐翔斌（1975—），男，博士，华东交通大学教授，主要研究方向为供应链管理。

在相近的货位中, 肖建^[8]则考虑 SKUs 结构相关性, 获得更好的货位优化效果。李英德^[9]进一步指出 SKUs 的需求相关强度越高, 货位优化对提高分拣效率更具有积极的影响。冯乾等^[10—12]利用聚类的方法来最大化 SKUs 之间的需求关联性, 在此基础上进行货位指派。总的来说, 这些货位指派方法基于 SKUs 销售和出库的历史数据, 利用统计、规划求解和聚类分析等方法分析得到其需求模式, 在此基础上进行货位指派, 适合传统 SKUs 品项稳定、需求相对平稳的零售和工业企业仓库拣货系统的货位优化。

在线零售企业具有商品生命周期短、营销策略多样性和季节性需求明显等特点, 其销售商品的品项、数量和需求关联性快速变化, 并且需求波动性极大, 由历史数据统计得到的商品需求模式往往并不能精确、实时地反映商品的动态变化, 传统货位指派方法难以获得很好的优化效果, 数据挖掘的方法在动态环境下能够有效地捕捉到商品需求的变化, 并且可以挖掘深层次的需求模式^[13—14], 随着ERP、POS、WMS 和在线销售系统的广泛应用以及数据挖掘技术的日益成熟, 从这些系统中获取SKUs的特征、需求和分拣信息是非常方便的事情, 这为利用数据挖掘技术构建智能货位指派方案提供了可能, 鉴于此, 文中以SKUs 品项和需求快速变化的在线零售企业仓库为研究对象, 提出一种基于数据挖掘智能货位指派方法, 设计了基于复合规则的动态货位指派策略(DRSAS), DRSAS一方面将传统的静态货位指派调整为多阶段动态调整过程, 并且采用数据挖掘技术自主学习得到货位指派规则, 文中对其系统架构及核心算法进行研究, 并在线零售企业运营数据采集的基础上进行数值仿真比较, 相对于传统的货位指派, DRSAS优化效果较为明显。

1 DRSAS 系统架构

DRSAS是一个集成货位指派信息结构化存储和分析、货位指派规则生成和实施的平台, 它在企业信息集成的基础上, 自主学习得到货位指派规则集, 为每个待上架批次的SKUs进行货位指派及路径生成, 它由数据集成和分析、指标计算、规则生成和货位指派等4个模块构成, 其系统架构见图1。

1.1 数据集成和分析模块(DICAM)

DICAM从WMS/ERP/POS等系统中抽取与货位指派相关的数据, 包括SKUs的基本特征、销售、库存和分拣等方面的信息。DICAM包括数据抽取、数据立方体和联机分析处理(OLAP)3个组件。数据抽取组件负责导入系统数据, 它提供和WMS、ERP和POS等系统数据连接的接口, 定期抽取DRSAS所需的数

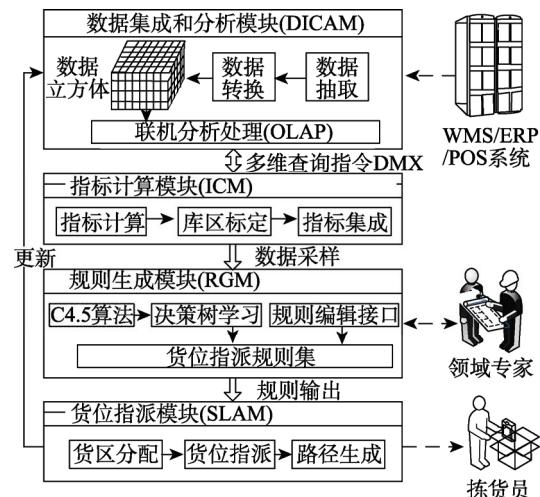


图1 基于复合规则动态规则货位指派系统架构

Fig.1 The architecture of compound rules-based dynamic storage location assignment system

据, 包括SKU的基本特征数据(体积、有效期等)和业务数据(营销策略、销售记录、出入库记录、存储时间和分拣时间等), 并进行数据清洗和转换, 以保证数据的一致性; 数据立方体组件负责DRSAS数据的结构化存储, 它从多个维度对货位指派相关的信息进行描述; OLAP组件则负责多维数据的统计和查询处理, 它解析来自指标计算模块的多维查询指令(DMX), 根据DMX指令将相关的信息从数据立方体抽取出来, 提供给指标计算模块进行进一步处理。

1.2 指标计算模块(ICM)

不同于传统的单一属性货位指派, DRSAS需要在综合考虑到存储天数、出货量、周准率、分拣时间和分拣密度等的基础上进行货位指派, 这些指标在SKUs的销售和分拣过程中高频动态变化, 需要实时计算得到, ICM负责这些指标的计算: 它定时发送DMX给DICAM, 从其数据立方体抽取所需的数据, 并将这些指标实时计算出来, 用于货位指派规则的生成, 参照文献的研究^[3—5,10], 并调研影响在线零售企业仓库拣货效率的关键因素, 得到DRSAS的货位指派决策指标及计算方式如下所述(决策变量中的下标*i*表示品项的编号)。

- 1) 出货量计为 p_i , 为品项*i*的出库数量的总和。
- 2) 分拣密度计为 e_i , 为品项*i*分拣出数量占全部分拣出品项数量比率, 即品项*i*的出货量/全部品项的出货量总和。
- 3) 平均库存计为 v_i , 为品项*i*的(期初库存+期末库存)/2。
- 4) 库存周转率计为 r_i , 为品项*i*的总出货成本/平均库存。
- 5) 平均分拣时间计为 t_i , 为所有包含品项*i*订单的分拣时间/品项*i*出货量。

1.3 规则生成模块(RGM)

采用机器学习领域中的决策树模型自动生成货位指派规则。实现决策树模型的算法有很多，文中采用C4.5算法，因为它能够处理连续值类型的属性，C4.5算法生成决策树后，每条从树根到树叶之间的路径，就是一条可用于货位指派的决策规则（根节点和中间节点是判定节点，叶节点是决策节点），相对于传统的专家系统来说，RGM模块优点为：规则自动生成，不需要知识工程师和领域专家们预先输入知识，降低了系统的实施成本；决策树学习得到的规则不需要进行复杂的推理就可直接用于货位指派，系统的效率高；SKUs决策指标是动态变化的，决策树生成的货位指派规则能够自适应SKUs需求和库存的变化；此外，RGM还提供一个人机接口，领域专家们可以通过该接口对货位指派规则进一步调整和优化。

1.4 货位指派模块(SLAM)

对于待上架批次中所有SKUs，SLAM根据SKU的存储天数、出库量、周准率、分拣时间和分拣密度等指标，在货位指派规则集中进行规则匹配，将SKU指派到相应的库区；在库区内部，按照距离出入口的从小到大的顺序，为SKU指定最佳的存储货位；在指派完该批次所有SKUs后，SLAM还生成SKUs最短的上架路径，拣货员按该路径完成该批次SKUs上架，最后SLAM还要更新WMS系统中SKUs存储信息。

2 货位指派决策算法

2.1 SKUs 库区标定算法

该算法标定所有SKUs的库区，描述如下所述。
算法名称：库区标定算法。输入：SKUs分拣记录。
输出：SKUs库区集合。算法实现步骤如下所述。

- 1) 生成一个空的库区集合Zones={ }。
- 2) 仓库库区划分。按中心位置距离仓库出入口从小到大的顺序将整个仓库划分为Good(G), Average(A)和Bad(B)这3个库区。
- 3) 分拣时间区间计算。计算所有品项的平均分拣时间，并找到最短和最长的分拣时间 t_{\min} 和 t_{\max} ，得到分拣时间区间 $L=[t_{\min}, t_{\max}]$ 。
- 4) 分拣时间和库区的映射关系构建。将步骤3)计算得到的分拣时间区间 L 分为3个区段，分别应仓库的G, A和B这3个库区，得到分拣时间和库区映射表。
- 5) SKUs库区标定。对每个品项*i*，由其平均分拣时间 t_i 确定其在分拣时间区间*I*中的区段，然后从分拣时间和库区映射表找到所在的库区，并将SKU编号及其库区添加到Zones中。
- 6) 返回Zones。

需要特别说明的是，该过程是动态自适应的，算

法能够根据SKUs需求和库存的变化，实时、动态地为SKUs标定库区。算法标定所有SKUs的库区属性后，将其合并到ICM计算得到SKUs的出货量、周转率、体积、剩余库存天数和分拣密度等决策指标中，删除SKUs的名称和标示属性后进行采样，构成了规则生成算法的训练样本集。

2.2 规则生成算法

C4.5算法是一个递归过程，先从训练样本集中，选择一个指标作为根节点来区分这些样本，对指标的每一区间值产生一个分支，其下的样本子集被移到新生成的子节点上，然后在每个子节点上递归运用该算法，直到所有的样本都被分配到某个库区中为止，在递归过程中，决策指标选择的标准是最大信息增益率，计算过程见式(1—3)。

$$Info(S) = - \sum_{i=1}^k ((freq(C_i, S) / |S|) \times \log_2(freq(C_i, S) / |S|)) \quad (1)$$

$$Info_x(S) = - \sum (|S_i| / |S|) \times Info(S_i) \quad (2)$$

$$GainRatio(X) = \frac{Info(S) - Info_x(S)}{Info_x(X)} \quad (3)$$

式中： $freq(C_i, S)$ 为训练样本集中 S 中库区为 C_i 的样本集合； $|S|$ 为 S 的样本数量，规则生成算法描述如下所述。

算法：规则生成算法。输入：货位指派训练数据集 S 。输出：货位指派规则集。算法实现步骤如下所述。

- 1) 生成一颗空的树Tree={ }和空的货位指派规则集合Rules={ }，并令 $D=S$ 。
- 2) 如果训练样本集 D 中所有的样本都属于一个库区，转到步骤6)。
- 3) 对于训练样本集 D 中的每个属性 x ，分别计算其信息增益率 $GainRatio(x)$ 。
- 4) 找到信息增益率最大的属性 a_{best} 作为Tree的根节点，并根据 a_{best} 划分得到子样本集合 D_a 。
- 5) 对于 D_a 中的每个子样本集 D_a^i ，令 $D = D_a^i$ 转到步骤2)。
- 6) 沿Tree根节点到叶节点遍历所有路径，将每条路径作为一条货位指派规则加入到Rules中。
- 7) 返回Rules。

3 仿真试验与分析

文中研究背景是经营女性化妆品的某小型在线零售企业，SKUs品项2000种左右，日均订单3000左右，由于化妆品的销售连带率较高，平均订单规模(即订单所包含SKU的数量)为15，并且呈正态分布，由于经常进行品类促销，SKUs的需求偏度较大。该电

商为单区仓库，共10个拣货巷道，巷道宽度为200 cm，货架的和长度宽度为80 cm，货架共3层，每层20个货位，仓库采用“一单一拣”的拣货方式，由于仓库的运营效率无法适应SKUs品项和销售的快速变化，并考虑到信息集成的需要，该电商企业构建了DRSAS系统，其OLAP服务器采用Mondrian多维引擎。为模拟优化效果，参照文献[16]提出的CMS4S状态图方法，基于Anylogic平台构建了多主体仿真系统，将拣货员建模为Agent，整个仿真系统包括仓库生成、订单生成、路径规划、规则生成、货位调整及数据分析等模块，最为关键规则生成算法扩展自Weka机器学习库中的C4.5算法。系统测试阶段在调研和数据采集的基础上，模拟在线订单生成过程如下所述。

- 1) 生成一个空订单 $O=\{\}$ 。
- 2) 生成订单规模 N 。以平均订单规模 N 为均值，变异系数为0.1，按照正态分布模型生成订单大小。
- 3) 生成下一个SKU。按照预定的需求偏度生成一个编号*i*的SKU。
- 4) 生成SKU的数量 n_i 。按照均值为1.5的二项分布生成品项*i*的数量 n_i ，且 $1 \leq n_i \leq 5$ 。
- 5) 若 $\sum n_i < N$ ，将*i*: n_i 添加到订单 O 中，并转到步骤3)，否则输出订单 O 。

DICAM模块每24 h从在线网站和WMS中抽取数据，ICM计算指标后，RGM生成货位指派规则集。为消除随机误差，每个数据点取30次实验的平均值（以下同），在不同订单规模 N 下，DRSAS和基于出货量的货位指派(VSLA)2种策略的订单分拣距离比较见图2。

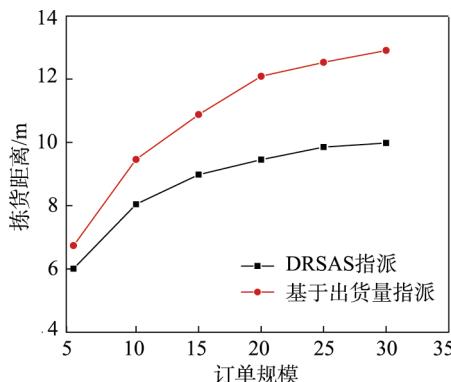


图2 DRSAS 和基于出货量货位指派的比较

Fig.2 Comparison of DRSAS and VSLA

在仿真中发现当SKUs的品类和数量快速变化时，基于出货量的货位指派会出现位于仓库IO附近的黄金货位中的SKUs长时间滞留，随后出现的畅销SKU将被迫指派到远离IO的货位，并且在订单规模较大时，SKUs几乎均匀地分布在所有的巷道中，这意味着拣货员几乎要穿越所有的巷道才能完成订单的分拣。另外仿真还发现在需求变动较大的时候，严格按照出货量这一单一属性进行指派会造成SKUs来

回频繁移动而影响货位优化效果^[15]，这说明传统的货位指派在需求变化较大时难以适应SKUs的品类和数量变化，其货位优化效果几乎近似于随机指派。DRSAS并没有严格按照出货量来进行指派，DRSAS的货位调整规则会综合考虑SKUs的分拣密度、出货量等决策指标来进行货位指派，在需求变化较大的情况下，DRSAS自主学习得到货位指派规则的根节点为分拣密度，其次为出货量，出货量大的SKUs只是按照一定的比例分配在距离仓库IO较近的货区中，这可有效地避免SKUs在黄金货位滞留和频繁移动这2个现象，在订单规模较大时，即使需求变化频繁，拣货员也不需要穿越所有的巷道就能够完成订单的拣选，从而大幅度降低拣货路径长度。

考虑到该电商SKUs的需求偏度和订单规模波动较大，取平均订单规模 N 为10，在订单生成过程的步骤3) 考虑80/20, 70/30, 60/40和50/50等4种需求偏度（80/20表示20%的SKU产生80%的出货量，余此类推），进行数值仿真，结果见图3a，相对于基于出货量的货位指派，需求偏度越大，DRSAS节约的拣货距离越大，这表明SKUs需求偏度越大，DRSAS货位优化越明显。在订单生成过程的步骤2) 中考虑订单规模 N 的变化（方差分别为平均订单规模 N 的0.1, 0.2, 0.4, 0.6倍，并且最小值取平均订单规模 N 的0.5倍，最大值取平均订单规模 N 的1.5倍）进行数值仿真，结果见图3b。相对于基于出货量的货位指派来说，DRSAS不仅能够节约拣货距离，而且拣货效率几乎不受需求变化的影响。

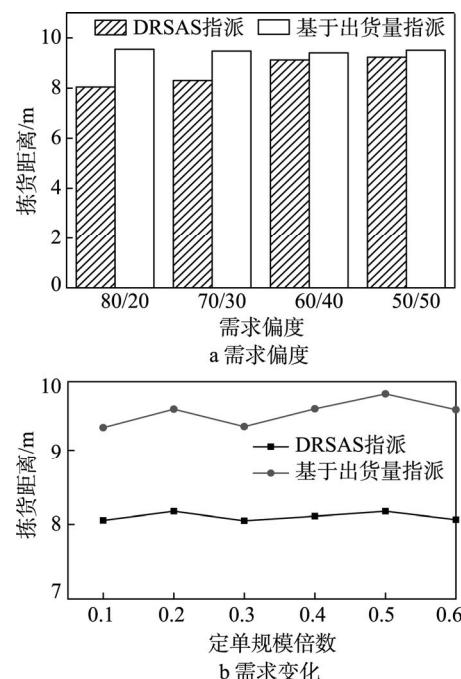


图3 考虑需求变化时 DRSAS 和基于出货量的货位指派的比较

Fig.3 Comparison of DRSAS and VSLA considering the demand fluctuations

在平均订单规模较大、需求偏度大时，并且需求变化幅度较大的情况下，DRSAS能节约分拣成本，并且能够很好地降低仓库作业的波动性，这表明了DRSAS能够自适应客户需求的变化，在需求快速变化、波动频繁的电商的仓库中具有较大的优势。

4 结语

考虑在线零售普遍存在销售商品的品种、数量和需求关联性快速变化这一情况，提出并研究基于数据挖掘的智能货位指派问题，构建基于复合规则的动态货位指派策略，并对其关键算法进行研究，通过和传统的货位指派方法进行数值仿真比较，发现在需求快速变化较大的情况下，基于数据挖掘的智能货位指派模型可大幅度降低拣货路径长度，并能够快速响应客户需求变化，货位优化效果较为明显。大量的文献研究表明考虑 SKUs 关联性能获得更好的货位优化结果，利用数据挖掘的关联规则算法^[17]来挖掘 SKUs 的需求关联性，在此基础上对 DRSAS 进行优化是下一步的研究方向。

参考文献：

- [1] DE K, ROODBERGEN K. Design and Control of Warehouse Order Picking: A Literature Review[J]. European Journal of Operational Research, 2007, 182(2): 481—501.
- [2] HAUSMAN W H, SCHWARZ L B, GRAVES S C. Optimal Storage Assignment in Automatic Warehousing Systems[J]. Management Science, 1976, 22(6): 629—638.
- [3] CAPUTO A C, PELAGAGGE P M. Management Criteria of Automated Order Picking Systems in High-rotation High-volume Distribution centers[J]. Industrial Management & Data Systems, 2006, 106(9): 1359—1383.
- [4] PARK B C, FOLEY R D, WHITE J A, et al. Dual Command Travel Times and Miniload System Throughput with Turnover-based Storage[J]. IIE Transactions, 2003, 35(4): 343—355.
- [5] GUE K R, MELLER R D, SKUFCA J D. The Effects of Pick Density on Order Picking Areas With Narrow Aisles[J]. IIE Transactions, 2006, 38(10): 859—868.
- [6] HESKETT J L. Cube-per-order Index-a Key to Warehouse Stock Location[J]. Transportation and Distribution Management, 1963, 3(1): 27—31.
- [7] FRAZELE E A, SHARP G P. Correlated Assignment Strategy can Improve Any Order-picking Operation[J]. Industrial Engineering, 1989, 21(4): 33—37.
- [8] 肖建, 郑力. 检修备品库的货位优化模型[J]. 清华大学学报(自然科学版), 2008, 48(11): 1883—1886.
- [9] XIAO Jian, ZHENG Li. Slotting Optimization Model for Overhaul Warehouses[J]. Journal of Tsinghua University(Natural Science), 2008, 48(11): 1883—1886.
- [10] 李英德. 波次分区拣货时装箱与货位指派问题协同优化的模型与算法[J]. 系统工程理论与实践, 2013, 33(5): 1269—1276.
- [11] LI Ying-de. Model and Algorithm for Cartonization and Slotting Optimization Simultaneously in Wave-picking Zone-based System[J]. System Engineering Theory and Practice, 2013, 33(5): 1269—1276.
- [12] 冯乾, 乐美龙, 赵毅. 物料聚类分析下的仓库货位指派优化[J]. 辽宁工程技术大学学报, 2015, 34(10): 1207—1212.
- [13] FENG Qian, LE Mei-long, ZHAO Yi. Optimization of Storage Location Assignment Based on Materials Clustering Analysis[J]. Journal of Liaoning Technical University(Natural Science), 2015, 34(10): 1207—1212.
- [14] CHUANG Y F, LEE H T, LAI Y C. Item-associated Cluster Assignment Model on Storage Allocation Problems[J]. Computers & Industrial Engineering, 2012, 63(4): 1171—1177.
- [15] WUTTHISIRISART P, NOBLE J S, CHANG C A. A Two-phased Heuristic for Relation-based Item Location[J]. Computers & Industrial Engineering, 2015, 82(1): 94—102.
- [16] CHIANG M H, LIN C P, CHEN M C. Data Mining Based Storage Assignment Heuristics for Travel Distance Reduction[J]. Expert Systems, 2014, 31(1): 81—90.
- [17] BENJAMIN P, BART V, DENIS D. Dynamic ABC Storage Policy in Erratic Demand Environments[J]. Jurnal Teknik Industri, 2003, 5(1): 1—12.
- [18] HEATH B L, CIARALLO F W, HILL R R. An Agent-based Modeling Approach to Analyze the Impact of Warehouse Congestion on Cost and Performance[J]. The International Journal of Advanced Manufacturing Technology, 2013, 67(1): 563—574.
- [19] AGRAWAL R, IMIELIŃSKI T, SWAMI A. Mining Association Rules Between Sets of Items in Large Databases[C]// ACM SIGMOD International Conference on Management of Data, ACM, 1999: 207—216.