

零售4.0时代先行发货问题的大数据分析

马丽娜,万美好,杜玉申

(吉林大学 商学院,吉林 长春 130012)

摘要:在零售4.0时代,渠道的多样化不仅丰富了数据源,还能迅速生成大量数据,需要通过分析大数据,为决策提取有意义的信息,通过分析先行发货的重要性,提出了一种基于遗传算法(GA)的优化模型,预测顾客何时购买,然后在顾客线下单前将产品运送到距顾客最近的配送中心,解决先行发货中存在的问题。研究认为,需要先部署云计算来存储所有渠道生成的大数据,再应用基于集群的关联规则挖掘研究顾客的购买行为,根据“如果-那么”预测规则预测未来的采购情况,最后利用修正的遗传算法生成最优的先行发货计划;这种遗传算法考虑了其在运输成本和运输距离之外,还有预测规则的置信度,利用大量的数值实验权衡了先行发货中的不同因素,验证了模型的最优可靠性。

关键词:零售4.0时代;全渠道零售;大数据;供应链;先行发货;遗传算法;关联规则挖掘

中图分类号:F724.2;F724.6 **文献标志码:**A **文章编号:**1671-6248(2020)05-0075-11

Big data analysis on the issue of dispatch in advance in the retail 4.0 era

MA Lina, WAN Meiyu, DU Yushen

(School of Business, Jilin University, Changchun 130012, Jilin, China)

Abstract: In the retail 4.0 era, the diversification of retail sales channels not only enriches the data source, but also can generate huge amounts of data. As the big data needs to be analyzed to provide meaningful information for decision making, this paper proposes an optimized model based on the Genetic Algorithm (GA) through analyzing the importance of dispatch in advance. The model predicts when customers will make the purchase, and dispatches the products to the nearest distribution center to the

收稿日期:2020-05-30

基金项目:教育部人文社会科学规划基金项目(10YJA630029);吉林省科技厅软科学项目(20190601086FG);吉林大学哲学社会科学研究项目(2018QY036)

作者简介:马丽娜(1975-),女,吉林四平人,副教授,管理学博士。

customer before they make the order online, in order to solve the issues in dispatch in advance. The research finds that cloud computing should be deployed in advance to store the big data generated through all retail sales channels, then the Association Rule Mining based on clusters should be applied to study the purchase behaviors of the customers, and their future purchases will be predicted according to the if-then prediction rule, and finally the revised Genetic Algorithm is used to produce the optimized plan for dispatch in advance. This type of Genetic Algorithm takes the credibility of the prediction rule into account, in addition to the transportation cost and transportation distance, and utilizes a large number of numerical experiments to measure and weigh various factors in dispatch in advance, thus verifying the reliability of the optimized model.

Key words: retail 4.0 era; omni-channel retail; big data; supplychain; dispatch in advance; Genetic Algorithm; Association Rule Mining

随着信息技术的发展,购物渠道越来越多,促使许多零售商实施多渠道战略。但是,由于数字化的不断发展,线上和线下购买渠道之间的区别正在消失。在这种情况下,传统的分销渠道正在发生变化,全渠道零售成为整个零售业发展的关键路径^[1],多渠道零售正在转向全渠道零售,迎来了第四代零售革命,即零售4.0时代。

全渠道零售是为顾客提供了一种无缝连接购买渠道,让顾客通过所有可行渠道体验购物,如线下渠道、线上渠道和移动渠道^[2]。同时,购买渠道的多样化使数据源越来越多,产品数据和顾客数据的数量呈爆炸式增长,使数据聚合更困难。由于顾客个体不断飞速地产生数据,因此需要利用大数据分析来从数据中提取隐含信息,从而为商业决策提供帮助^[3]。大数据分析在零售业虽然已经广泛应用,但由于零售业本身数据种类和来源过于繁杂,使得大数据在零售行业的切实应用存在一定的困难。本文通过基于遗传算法优化的模型,完善了大数据分析在零售业中的应用和能产生的相关指导作用。

大数据的应用主要包括云计算、关系数据库以及数据挖掘技术等^[4]。大数据分析在零售4.0时代中的重要性表现在以下两个方面。其一,在全渠道零售商业中,零售商必须以顾客为导向,与顾客紧密联系。例如,他们需要将所有可行的购买渠道与相匹配的顾客和产品信息连接起来,并汇总所有数

据进行分析,以便与顾客进行个性化交流。大数据挖掘可以完成购物篮分析,从而帮助企业预测顾客购买趋势并设计营销促销策略。大数据分析可以洞察顾客的需求,预测顾客的行为,有效帮助企业做出满足顾客个性化需求的决策^[5]。其二,无论订单是在线上还是在线下完成,零售商提供全套发货和提货方式非常重要。顾客希望在下单后快速交付,所以交付绩效是影响全渠道零售交易中顾客购物体验的关键因素之一。例如,亚马逊申请了先行发货专利,用于预测顾客何时进行购买并在顾客在线下订单之前将产品运送到最近的配送中心^[6]。

通常,衡量交付绩效的两个一般因素是成本和时间。近期的相关研究中,学者们考虑用成本和时间来改善配送网络,但是较成熟的模型通常假设已知顾客的需求,它不适用于先行发货,因为不同产品的顾客需求完全是根据顾客行为预测的结果。因此,应该考虑预测结果的置信度,因为它决定是否应该在收到任何订单之前把产品预先运送到特定配送中心(DC),而且它可以有效解决当优化一个因素时导致其他因素改变产生的问题。例如,当将某类产品运送到配送中心A所需的成本低于配送中心B时,但预计在配送中心B附近的购买置信度比配送中A附近的购买置信度更大时,零售商不得不面对交易成本与预测置信度之间的权衡。如果产品运送到购买置信度更高的地点,预计投资回报率可能会更好。

本文的目的是分析预测顾客的购买行为,并根据多种因素,包括运输成本、运输时间和预测结果置信度,提出一种用来确定将产品分配至不同配送中心的优化模型。

一、文献综述

(一)全渠道零售及大数据分析研究

互联网的出现和发展及其在供应链管理中的重要性,打开了电子物流发展新机会的大门,由于电子商务的拓展,除传统零售外,顾客能够通过多种渠道进行产品购买^[7],例如商店、产品目录、网络、移动终端等包括在内的所有渠道^[8]。顾客无论是亲自购物,还是通过目录、手机、在线等其他方式购物,都会对如何购买产品做出决定^[9]。随着一些电商分销商对产品不同程度的主导,对传统零售商产生了不同的影响,甚至会导致其退出销售渠道^[10]。这种情况下,传统零售商将面临复杂的电子物流带来的挑战,他们在制定零售策略时不得不考虑技术问题,因为一项新技术会为零售业创造新渠道,零售商开始采用多渠道零售策略,学者们已将线上和线下渠道融入零售业视为零售商较有前景的策略。然而,近几十年来信息技术的发展为全渠道零售提供了动力,而全渠道零售业逐渐成为零售业4.0时代的主流战略。

零售业4.0依托于工业4.0,利用大数据进行计算和建模,进行针对性的有效营销,并结合全渠道服务模式,为消费者提供了全新购物体验^[11]。中国的零售业在经历了计划经济下的传统零售1.0时代,市场经济下客户主导的终端零售2.0时代,以及电商线上零售的3.0时代后,逐步进入零售业4.0时代^[12]。零售业4.0推崇各零售渠道以及物流之间协作,通过形成全渠道新零售模式,为消费者提供跨渠道无差别服务^[11]。零售渠道发展的高级阶段是全渠道零售,它是“以消费者为中心”为主要特征,实现各渠道之间的无缝链接,为顾客提供更高的消费体验价值^[13]。

尽管全渠道零售商业模式可以让企业通过较多的渠道向顾客销售商品或提供服务,但它也存在

潜在的消极因素。一个主要的现象是“展厅现象”,即顾客在商店中打量商品,但同时在网上搜索更多信息并以更具吸引力的价格在竞争对手的网站上购买^[14]。同样,“反展厅现象”与“展厅现象”相反,出现了顾客在网上寻求信息但在线下购买商品的现象。因此,多渠道的快速扩散,可能会造成顾客跨渠道搭便车的问题。我们鼓励零售商投入资金来分析顾客的购买行为,顾客大数据分析是从大数据中提取关于顾客行为的隐含信息,并用有效的解释来剖析该隐含信息^[15]。

(二)关联规则挖掘相关研究

数据挖掘技术,如聚类和关联,可以从大数据中提取隐含信息来实现目标^[16]。聚类是基于对象的相似点将其分成多个类的非监督分类过程,由于聚类所要划分的类是未知的,仅凭数据自然分类,故聚类分析又称作非监督分类,例如高聚类内相似性和低聚类间相似性,物体之间的相似性通常通过距离来衡量,而顾客可以通过具有相似的特征,如兴趣、收入水平和支出相似,组合为一个集群。聚类算法通常采用层次聚类方法^[17]和K-means聚类方法,例如采用K-means聚类方法对游客分布分类、对消费者客户分类等^[18-20]。此外,关联规则挖掘有助于找出所有项集之间的相关性,并用关联规则来表达它们。关联规则以 $P \rightarrow Q$ 的形式表示,其中 P 和 Q 分别定义为表示规则的“如果”部分和“那么”部分的项集。Agrawal et al.提出了Apriori算法,一种挖掘关联规则的频繁项集算法,Chung et al.在研究中指出该算法可以用于发现数据集中所有项集之间的关联规则^[21]。戴稳胜等使用数据挖掘技术发现交叉销售机会,即研究顾客购买某公司的一种产品后,又购买该公司另一件产品的关联性,通过数据挖掘构建顾客需求预测模型^[22]。张德鹏等利用企业数据资源,提出利用关联规则技术制定商品的营销组合,为商品提供外在推动力^[23]。何黎等利用相关性分析和关联规则分析对平台用户进行分析,实现个性化营销^[24]。郭燕等认为数据挖掘为零售业创新提供了新工具,对于传统零售业而言,充分运用大数据成为创新制胜的必由之路^[25],范生

万等采用 Apriori 算法对通过聚类分析的企业客户和商品进行关联规则分析,从而指导商品精准营销^[18]。

(三) 遗传算法相关研究

顾客在下订单后希望快速收到产品,一些新型交付方式,诸如多峰性和时窗方式等越来越重要^[9]。另外,改善配送网络也成为提高物流绩效的关键,配送网络通过不同的配送中心把来自不同来源点(供应商)的产品配送到不同的需求点(顾客)。物流的两种主要属性是交付速度和交付可靠性,前者是指缩短订单接收和顾客交付之间时间的能力,而后者则指准确报价或预期交付日期和数量的能力。遗传算法(GA)是启发式方法之一,广泛使用于配送网络问题。相比于其他算法,遗传算法对所优化目标函数的连续性和是否可微没有严格的要求,并且在一定时间内能够从大规模优化问题中做出选择^[26]。在遗传算法中,每个单独的解决方案都用染色体的形式表示,染色体通过使用遗传算子进行修改,如交叉和变异。遗传算法的目标是创建新一代可行的解决方案,使其趋于函数的最优解。Lin et al. 开发了一个遗传算法模型来确定交付和回收商品的最佳提货策略^[27];Ko et al. 采用遗传算法优化了新型电动汽车充电基础设施的分配,他们考虑了电力变送器成本这一非线性成本^[28];Dib et al. 结合了遗传算法和可变邻域搜索算法,解决道路网中的最短路径问题^[29];赵泉午等在优化大型零售企业城市配送网络的研究中,运用遗传算法时,考虑了配送中心到需求点的配送距离^[30]。

本文创新性地将遗传算法和基于聚类的关联规则挖掘相结合,研究了多因素对配送网络的影响,并提出了先行发货的优化方法。

二、基于遗传算法的优化模型

图 1 是先行发货的遗传算法优化模型。图 1 构建了基于遗传算法优化模型的基本框架,以此分析顾客行为和先行发货,它由 3 个模块组成:(1) 大数据管理模块;(2) 顾客预测分析模块;(3) 先行发货

优化模块。大数据中的有些数据更能显著地体现出顾客在全渠道零售中的购买行为,为了筛选出这些数据,需要把各个顾客的数据集远程存储在大数据管理模块的云端中,零售商能够从云端获取大数据分析的数据。在利用云端管理数据的模式下,大数据管理模块为全渠道零售供应链各方共享信息提供了一个公共平台。顾客预测分析模块嵌入了大数据挖掘技术,以确保所有信息源均有效满足挖掘目标。换言之,模型中完成的大数据挖掘技术的功能包括聚类 and 关联。在预先发货优化模块中,遗传算法用于为先行发货生成近似最优配送解决方案,以下介绍各个模块的详细信息。

(一) 大数据管理模块

云计算是管理大数据的新兴技术之一,是减少运营和资本成本较有前景的工具,云计算系统的体系结构要求的细节可以参考 Rimal et al. 的研究^[31]。云计算是基于互联网的计算,通过计算机或其他设备之间提供共享资源、软件和信息,在大数据管理模块中,把从物理渠道、商品目录渠道、手机移动渠道和电脑在线渠道收集的数据整合并存储在云中。由于现收现付成本结构,其资源成本取决于使用量,可减慢新渠道的增加,为零售商提供更灵活多变的渠道整合技能。

(二) 顾客预测分析模块

把多种渠道中的数据汇总到云端之后,进行顾客预测分析,进而从数据中发现信息,以利于提取顾客隐含的行为模式并预测未来的购买行为。在顾客预测分析模块中,数据分为两个级别:网络层级和集群层级。在网络层级,应用关联规则挖掘探究整个配送网络中顾客与已购物品之间的关系。虽然在某些情况下,不同地理位置的顾客可能会有不同的购买行为,但在网络层级的应用关联规则挖掘,能同时考虑到一些适用于整个网络的规则。在集群层级,顾客购买的需求点进一步分为不同的集群,同一集群内的需求将由同一个配送中心实现。因此,配送中心的需求是同一集群内所有需求点的总需求,关联规则挖掘应用于研究每个集群中顾客

与已购物品之间的关系。在这个模块中,Apriori 算法为预测目标生成一组关联规则,输出则采用“如果-那么”的形式,预测当某商品被购买时顾客对其他商品的购买行为。利用规则置信度衡量已生成规则的有效性,规则置信度也可以作为预测结果置信度的指标,如果相应的预测规则具有较高的置信度,则购买预测产品的置信度更高。每条规则的置信度是包含 P 和 Q 的百分比。它可以表示如下

$$Confidence(P \rightarrow Q) = P(P \mid Q) \tag{1}$$

式中: P 为关联规则的“如果”部分的项集; Q 为关联规则的“那么”部分项集。置信度临界值为用户预先定义的值,只有大于或等于临界值的规则才会被视为有用的规则。在先行发货优化模块中,当优化总体规则置信度时,通过遗传算法生成预期运输计划,并最大限度地提高对所考虑规则的整体置信度。

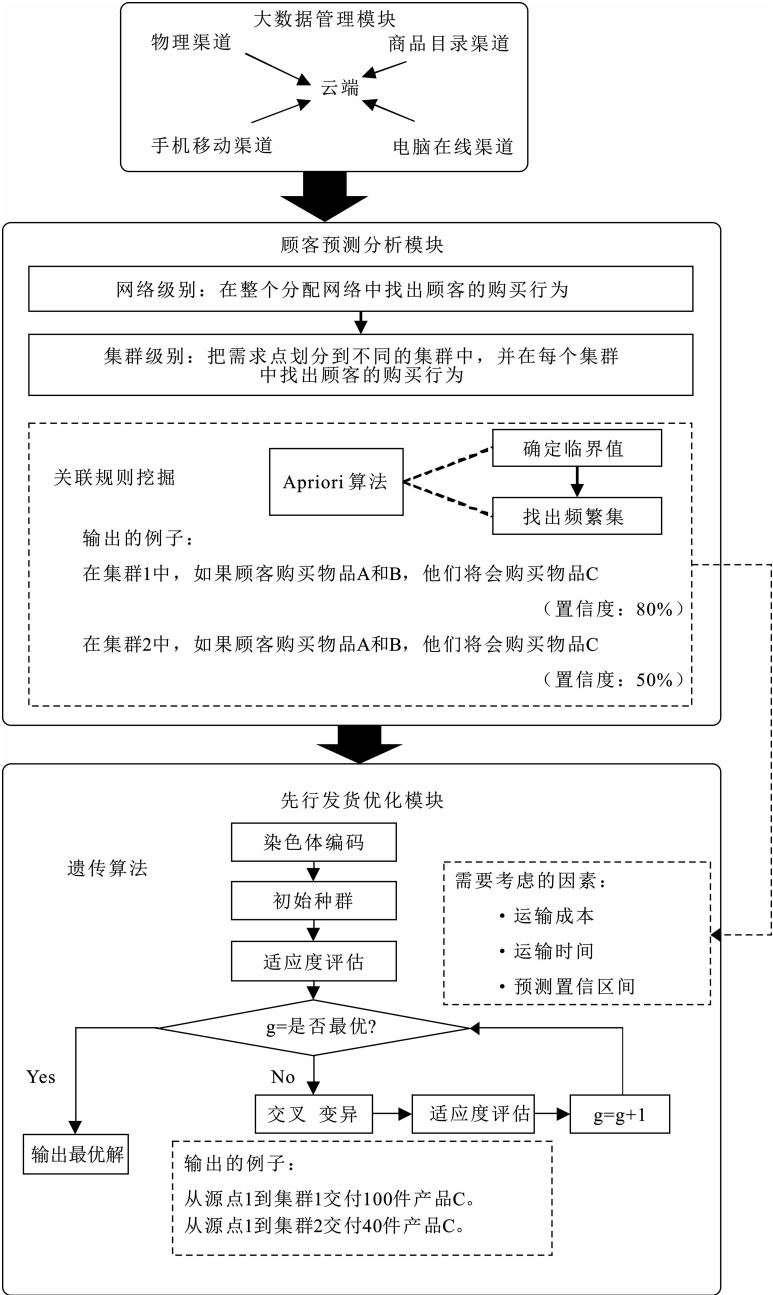


图1 先行发货的遗传算法优化模型

(三) 预先发货优化模块

1. 问题描述

表 1 参数和决策变量的符号

符号	解释
n	产品类型的数量
$N = \{1, 2, \dots, n\}$	产品类型
h	配送中心的数量
$H = \{1, 2, \dots, h\}$	配送中心
s	源点的数量
$S = \{1, 2, \dots, s\}$	源点
C_{ijk}	将产品 i 从源点 j 发送到配送中心 k 所需的单位运输成本, $\forall i \in N, \forall j \in S, \forall k \in H$
t_{ijk}	将产品 i 从源点 j 发送到配送中心 k 所需的运输时间, $\forall i \in N, \forall j \in S, \forall k \in H$
r_{ik}	产品 i 在配送中心 k 附近的预测规则置信度, $\forall i \in N, \forall j \in S, \forall k \in H$
P_k	配送中心 k 的容量, $\forall k \in H$
d	需求点的数量
$D = \{1, 2, \dots, d\}$	需求点
y_{ik}	配送中心 k 中产品 i 的预测需求 $\forall i \in N, \forall k \in H$
z_{ij}	源点 j 处可用于发出产品 i 的数量 $\forall i \in N, \forall j \in S$
L_r	预测规则置信度对应的服务级别
C_0	运输成本的近似最优值
T_0	运输时间的近似最优值
R_0	预测规则置信度的近似最优值
决策变量 μ_{ijk}	$=1$, 如果产品 i 从源点 j 发送到配送中心 k $=0$, 其他, $\forall i \in N, \forall j \in S, \forall k \in H$
Q_{ijk}	从源点 j 发送到配送中心 k 的产品 i 的数量, $\forall i \in N, \forall j \in S, \forall k \in H$

表 1 中的符号描述了与先行发货相关的问题,图 2 中,配送网络由所有需求点 D 及其包含的各类产品 N 组成。在先行发货期间,不同类型产品将从一组源点 S 运送到一组配送中心 H ,当收到顾客订单时,每个配送中心只会将产品分配到集群内的需求点。因此,本文不研究产品如何从总部向需求点交付,产品 i 运送到配送中心 k 是否会受到相关预测规则置信度 r_{ik} 的影响。如果某类产品运送到一个配送中心,则发货数量将能够满足服务水平 L_r ,并且从每个源点 Q_{ijk} 运出的每种产品 i 的数量不超过源点中可发出数量 z_{ij} ,每个配送中心 k 具有容量限度 P_k 。把

产品从一个点运送到另一个点时,将会产生单位运输成本 c_{ijk} 和运输时间 t_{ijk} 。

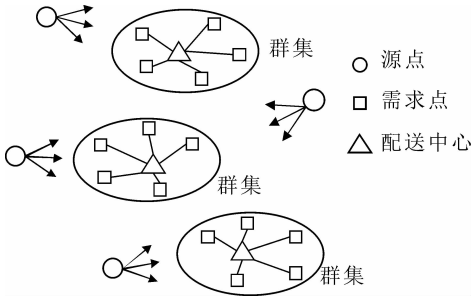


图 2 先行发货的配送网络

$$f(x) = \min[\alpha(C) + \beta(T) - \gamma(R)] \tag{2}$$

式中: α 为运输成本; β 为运输时间; γ 为预测规则置信度的权重。式(2) 是考虑多因素的目标函数,其目的是使运输成本和运输时间最小化,并使预测规则置信度最大化。

$$C = (\sum_{i=1}^n \sum_{j=1}^s \sum_{k=1}^h \mu_{ijk} c_{ijk} Q_{ijk} - C') / (C_{\max} - C') \tag{3}$$

$$T = (\sum_{i=1}^n \sum_{j=1}^s \sum_{k=1}^h \mu_{ijk} t_{ijk} - T') / (T_{\max} - T') \tag{4}$$

$$R = (\sum_{i=1}^n \sum_{j=1}^s \sum_{k=1}^h \mu_{ijk} r_{ik}) / (\sum_{i=1}^n \sum_{j=1}^s \sum_{k=1}^h \mu_{ijk}) \tag{5}$$

式中: C 为总运输成本; T 为总运输时间; R 为预测规则置信度, C' 和 T' 分别表示运输成本、运输时间的适应度值。虽然 R 总是介于 0 和 1 之间,但 C 和 T 可以处于不同的等级。 C 和 T 在式(3) 和式(4) 中分别在 0 和 1 之间分别归一化,这可以使它们的聚合在相同维度下进行,消除了对目标结果的不合理影响。

另一方面,这个问题受到以下限制

$$P_k \geq \sum_{i=1}^n \sum_{j=1}^s \mu_{ijk} Q_{ijk} \quad k = 1, 2, \dots, h \tag{6}$$

式(6) 确保满足每个配送中心的容量。

$$\sum_{j=1}^s \mu_{ijk} Q_{ijk} \geq L_r y_{ik} \quad i = 1, 2, \dots, n; k = 1, 2, \dots, h \tag{7}$$

式(7) 保证了运送到配送中心 k 的数量可以达到与预测规则置信度相对应的服务水平。如果预测规则的置信度大于或等于“高置信度”的最小临界值,则

认为服务水平较高(如 80%)。例如,在配送中心 k 的产品 i 的预测需求是1 000 单位,并且根据预定临界值将预测规则置信度视为高置信度,那么从所有源点到配送中心 k 的产品 i 的发货量应当是至少 $80\% \times 1\,000 = 800$ 个单位。

$$z_{ij} - \sum_{k=1}^h \mu_{ijk} Q_{ijk} \geq 0 \quad i = 1, 2, \dots, n; j = 1, 2, \dots, s \tag{8}$$

式(8) 确保配送中心中来自各源点的每类产品数量不会超过该类产品的可发货数量。

为了简化模型,有两个假设:所有计划产品一次发货;无论数量如何,每次运输都有固定时间。

在这样的假设下, T 与 Q_{ijk} 独立, T 既不影响顾客也不影响总运输成本。因此, T 在模型中的重要性较低,因此预期在目标函数中 β 应该不大。另外, R 在目标函数中与 Q_{ijk} 独立,但是 R 会影响 z_{ij} ,从而影响 Q_{ijk} 的最小值。

2. 引入遗传算法

在这个模块中,使用遗传算法确定一个近似最优的方案,解决如何从不同的源点将不同类型的产品分配到不同的配送中心进行先行发货。遗传算法的最终目标是确保当顾客在线下单时,该产品已经在顾客附近的配送中心有可发货的库存。

(1) 染色体编码。染色体中有两个区域:分配区域和产品数量区域。分配区域是运载 μ_{ijk} 值的二进制区域,而产品数量区域是运载 Q_{ijk} 的值。根据基因数,染色体的长度是 $2 \times n \times s \times h$ 。图3 显示了染色体的一个例子。父代的染色体是随机产生的,轮盘轮法用于染色体选择,父代与子代的大小相等。根据适应值,在子代中,最差的染色体被进化成最优的染色体取代。在每对染色体中,随机选择预定数量的基因进行交叉和突变,如果违反编码方案,染色

体将被修复。

(2) 适应度评估。为了研究不同因素对先行发货分配网络的影响,先用遗传算法优化单一因素。该模型考虑了3 个因素,即运输成本、运输时间和预测规则置信度,式(9) ~ 式(11) 分别计算遗传算法优化运输成本、运输时间和预测规则置信度的适应度值,根据结果得到 C' 、 T' 和 R' ;然后用遗传算法时要同时考虑所有的因素,式(2) 用来计算遗传算法的适应度值。

$$C' = \min \left(\sum_{i=1}^n \sum_{j=1}^s \sum_{k=1}^h \mu_{ijk} c_{ijk} Q_{ijk} \right) \tag{9}$$

$$T' = \min \left(\sum_{i=1}^n \sum_{j=1}^s \sum_{k=1}^h \mu_{ijk} t_{ijk} \right) \tag{10}$$

$$R' = \max \left[\left(\sum_{i=1}^n \sum_{j=1}^s \sum_{k=1}^h \mu_{ijk} r_{ik} \right) / \left(\sum_{i=1}^n \sum_{j=1}^s \sum_{k=1}^h \mu_{ijk} \right) \right] \tag{11}$$

三、优化模型验证和讨论

实验有3 个目的:验证关联规则挖掘在分析顾客购买模式方面的效果;验证遗传算法在考虑配送网络中多重因素时的效果;权衡先行发货因素。案例研究中把收集的数据输入到 XLMiner(Excel 表格数据挖掘插件,能有效对表中数据进行分析挖掘)中以生成一组关联规则,使用 Evolver(Excel 表格的高级最优化插件,适用于 Microsoft Excel 的创新性遗传算法最优化)执行遗传算法进行优化。

(一) 基于集群的关联规则挖掘

为了验证以预测为目的基于集群的关联规则挖掘的可行性,选择一家公司进行案例研究。该公司是国内小型家电制造企业,总部位于广东省,在中

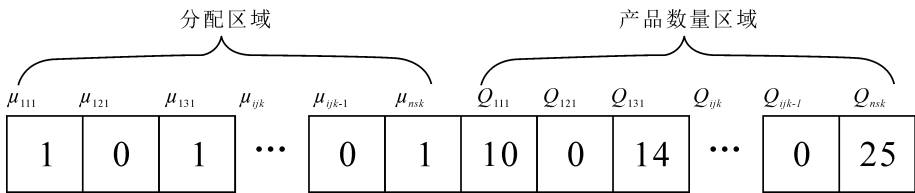


图3 染色体编码

国多个城市有分销中心,拥有线上线下销售平台,为消费者提供全渠道销售服务。在案例中,我们选取该公司华南地区某市的分销供应链,该市有两个分销中心,即配送网络中的发货源点,有 5 个分布在不同位置的直销门店,即配送网络中的配送中心。顾客可以通过线上和线下渠道进行购买,供顾客在产品抵达时提货,每个位置都作为一个集群提供并收集每个集群的交易记录,应用关联规则挖掘来发现隐含模式以预测顾客未来的购买。

Apriori 算法涉及两个主要阶段:从数据中找到高频项目组;使用高频项目组生成关联规则。同时,为了识别高频项目组,先设置最小支持度,并关注高频项目组。在实践中,最小支持度的定义是很严格的,直接从其他相关实证中采用这些值显然不可行。如果阈值设置得太低,可能会挖掘出许多无足轻重或无法解释的规则,从而不能为实践者提供建议;相反,如果阈值设置得太高,则可能难以获得任何规则。因此,阈值的最终选择在很大程度上要使用试错法,直到可以挖掘出一些有用的关联规则为止。

在这个案例研究中,从每个地点总共收集 500 笔交易。小家电产品与超市中的日常必需品不同,它并非必需品,因此不会像超市中出售的产品那样频繁地被购买,鉴于此,为了生成关联规则,最小支持度可以更低。经过试错法后,每种产品类型的最低支持值设为 10,如果 500 笔交易中某种类型产品有 10 笔以上的交易,则案例公司认为该类型产品很受欢迎。考虑到最小支持度的决策较严格并且依赖于一系列试错程序,所以应对选择的值进行定期评估以保证获得决策的质量^[32]。通过关联规则挖掘找出流行产品之间的关系,最小规则置信度设置为 75%,表 2 列出了从其中一个集群中获得的关联规则。如果顾客购买了某件产品(设为条件产品),则以置信度为比例购买另一件产品(设为结果产品),根据顾客对条件产品的需求,案例公司可以更好地估计对结果产品的需求。考虑到顾客希望在产品下单后能够快速交货,需要在顾客下订单之前将相关产品运送到最近的分配中心,可以建立更有

效的预期运输计划。

在下一部分中,使用遗传算法生成先行发货的配送解决方案,同时考虑预测规则的置信度,该解决方案可以确定从特定源点到特定集群应该装运产品的数量。

(二)先行发货的修正遗传算法

根据参数范围随机生成 10 组数据,如表 3 所示,每组将由遗传算法分别确定运输成本、运输距离和预测规则置信度的近似最优值,之后,每组将同时考虑 $\alpha = 0.5$, $\beta = 0.1$ 和 $\gamma = 0.4$ 权重的所有 3 个因素。如上文所述,因为 T 既不会影响顾客,也不会影响总运输成本,在总体目标中不那么重要,所以 β 很小。将每个因子下每一组运行 10 次,其中:种群大小 = 50,交叉率 = 0.7,突变率 = 0.02。记录每 10 次运行的平均结果,并总结在表 4 中。例如,在第 2 ~ 4 栏中,应用遗传算法来使运输成本最小化,而记录运输成本 CC 、运输距离 TC 和预测规则置信度 RC ,并作比较。

表 2 关联规则

条件产品 (A)	结果产品 (C)	A 的支 持度	C 的支 持度	A 和 C 的 支持度	置信 度/%
1MG270 & AM0021 W	AA074P	13	24	11	84.615
2AB109 W & MM7512 K	MM0561 -7	15	26	13	86.667
3MM0561 -7 & MM7512 K	AB109 W	15	26	13	86.667
4AM0021 W & AA074P	MG270	14	31	11	78.572

表 3 模拟配送网络中的参数

参数	取值范围
产品种类数量	3 ~ 10
集群数量(DC / 配送中心数量)	1 ~ 2
源点数量	3 ~ 5
每单位运输成本	10 ~ 150
运输时间	30 ~ 120
服务水平	= 50%, 置信度 ≥ 0.8
	= 20%, 置信度 ≥ 0.6
	= 0, 其他
配送中心容量	6 000
每个配送中心的各类产品的需求量	100 ~ 1 000
每个源点的各类产品的可用需求	1 000 ~ 3 000

表 4 优化结果摘要

运行	运输成本最优(C)			运输时间最优(T)			规则置信度最优(R)			所有因素最优(O)		
	成本 (C-C)	时间 (T-C)	规则置信 度(R-C)	成本 (C-T)	时间 (T-T)	规则置信 度(R-T)	成本 (C-R)	时间 (T-R)	规则置信 度(R-R)	成本 (C-O)	时间 (T-O)	规则置信 度(R-O)
1	23 421.8	424.6	0.72	329 211.2	337.6	0.75	530 536.9	554.9	0.78	24 063.8	358.7	0.75
2	38 987.4	313.3	0.75	175 315.1	264.5	0.77	277 720.7	594.6	0.80	39 407.0	297.0	0.77
3	24 142.3	342.7	0.74	83 762.1	287.9	0.76	121 856.1	407.3	0.81	24 416.8	307.7	0.78
4	41 215.0	370.2	0.75	100 019.2	289.5	0.80	156 694.3	541.2	0.83	41 819.0	310.9	0.80
5	31 652.5	277.9	0.75	39 690.2	234.8	0.78	44 868.8	309.2	0.80	31 848.0	241.1	0.78
6	25 855.2	468.1	0.73	125 534.3	450.8	0.73	162 244.3	703.5	0.76	26 018.8	463.7	0.73
7	7 151.2	345.6	0.67	54 958.6	221.9	0.68	78 637.1	570.6	0.69	7 950.4	319.8	0.68
8	13 142.4	566.4	0.64	27 282.5	394.3	0.66	43 743.6	602.8	0.70	13 468.6	439.4	0.66
9	58 910.0	422.1	0.73	147 345.6	416.0	0.73	177 826.4	726.5	0.75	58 910.0	416.0	0.73
10	197 868.8	384.8	0.71	387 503.7	297.3	0.75	415 519.3	530.4	0.79	19 911.2	324.3	0.76

图 4 ~ 图 6 是比较结果,证明了优化一个单一因素会牺牲其他因素。如图 4 所示,运输时间的最小化导致配送网络中的运输成本增高。如图 6 所示,其他因素的最小化可能会产生相对较低的置信度值。因此,如果根据先行发货的预测结果提供产品,则配送计划将与基于其他因素(例如运输成本和运输时间)的计划大不相同。然而,考虑到所有因素的重要性,C-O、T-O 和 R-O 的值表明每个因素的效能越来越接近最优值 C、T 和 R。在图 4 中,这种改进更为明显,C-O 线更接近 C-C 线,并且比 C-T 和 C-R 线好得多,这是由于赋予 C 较大的权重($\alpha = 0.5$)。另外,线 T-O 是图 5 中最靠近线 T-T 的线,线 R-O 是图 6 中最靠近线 R-R 的线,证明了提出用遗传算法在处理先行发货多种因素时的优化可靠性。

四、结语

本文是基于以下 3 种现象而产生思考的。第一,电子商务的增长使顾客可以通过不同渠道购物,每秒产生大量数据,那些未能设置分析大数据业务部门的企业将面临大数据革命中的生存挑战。第二,在全渠道零售商业中,竞争越来越激烈。由于网上提供了很多产品信息,顾客在商店中定位商品,但同时在线搜索更多信息,并有可能在竞争对手的网站上购买,这种现象越来越普遍,因此了解顾客购买行为并积极响

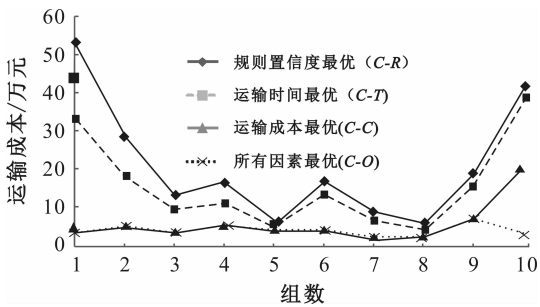


图 4 优化运输成本时各种因素的比较结果

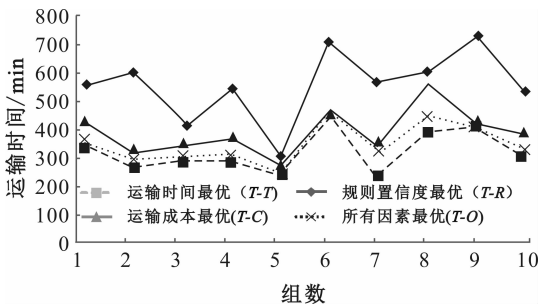


图 5 优化运输时间时不同因素的比较结果

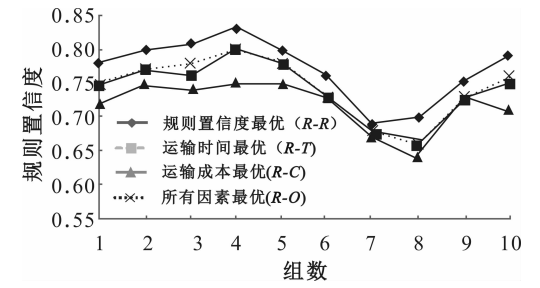


图 6 优化预测规则置信度时不同因素的比较结果

应市场非常重要。第三,顾客在下订单后希望快速交付产品,因此企业需要通过先行发货以增强顾客购物体验,也要求企业预测顾客的购买情况。

本文基于以上问题,提出遗传算法和基于聚类的关联规则挖掘相结合的优化模型。首先利用云计算来管理从所有有效渠道生成的大数据,应用基于集群的关联规则挖掘来分析大数据并发现顾客隐含的购买模式,求出规则,帮助预测未来购买趋势。然后根据这些规则,考虑多种因素,利用遗传算法选择先行发货的近似最优解决方案。本文贡献主要表现在两方面,一方面在研究配送网络时,不仅考虑运输成本、时间等运输自身存在的影响因素,还考虑到顾客购买产品的情况,通过对顾客购买产品的预测,提高了先行发货的准确性;另一方面,通过对遗传算法的引入,丰富了大数据分析在营销渠道领域的应用,解决了快速交付产品带来的问题。实验结果给了实践者一个启示,即在处理先行发货问题时,优化单一因素将会牺牲其他因素。因此,我们应该同时考虑所有因素,这样才能控制配送网络多方面的整体效能,使得其接近最优状态。

本文也有一些局限。一个局限在于文中的聚类是根据地理标准预先定义的,即需求点是根据它们的距离聚类。另一个局限在于模型中做出的假设。文中假设所有计划物品都立即发货,而且无论数量多少,每次运输都有固定的时间。这种假设使得模型变得相对简单,并使运输时间与发货数量独立。因此,需要在未来的研究中建立没有上述假设的模型,使得运输时间与数量相关,这可以更好地模拟实际情况。

参考文献:

- [1] 申维娜. 零售业全渠道商业模式选择——基于移动电子商务的视角[J]. 商业时代, 2016(15): 85-87.
- [2] 王坤, 相峰. “新零售”的理论架构与研究范式[J]. 中国流通经济, 2018, 32(1): 3-11.
- [3] Wang G, Gunasekaran A, Ngai E W T, et al. Big data analytics in logistics and supply chain management: certain investigations for research and applications[J]. International Journal of Production Economics, 2016, 176: 98-110.
- [4] 郑淑蓉. 零售业大数据: 形成、应用及启示[J]. 理论探索, 2014(2): 90-94.
- [5] 田歆, 汪寿阳, 鄂尔江, 等. 零售大数据与商业智能系统的设计、实现与应用[J]. 系统工程理论与实践, 2017, 37(5): 1282-1293.
- [6] Banker S. Amazon and anticipatory shipping: a dubious patent? [EB/OL]. (2014-01-24) [2020-01-11]. <https://www.forbes.com/sites/stevebanker/2014/01/24/amazon-and-anticipatory-shipping-a-dubious-patent>.
- [7] Zhao F, Wu D, Liang L, et al. Lateral inventory transshipment problem in online-to-offline supply chain[J]. International Journal of Production Research, 2016, 54(7): 1951-1963.
- [8] 李飞. 全渠道零售的含义、成因及对策——再论迎接中国多渠道零售革命风暴[J]. 北京工商大学学报(社会科学版), 2013, 28(2): 1-11.
- [9] Fairchild A M. Extending the network: defining product delivery partnering preferences for omni-channel commerce[J]. Procedia Technology, 2014, 16: 447-451.
- [10] 董志刚, 徐庆, 马骋. 电子商务环境下双渠道供应链的制造商分销渠道选择[J]. 系统工程, 2015(6): 26-33.
- [11] 王晓锋. 重构零售: 新零售时代企业生存法则与经营实践[M]. 杭州: 浙江大学出版社, 2018.
- [12] 王晓锋, 张永强, 吴笑一. 零售4.0时代[M]. 北京: 中信出版社, 2015.
- [13] 马慧敏. 移动互联网时代我国零售企业全渠道模式的应用[J]. 中国流通经济, 2017, 31(4): 10-16.
- [14] Herhausen D, Binder J, Schoegel M, et al. Integrating bricks with clicks: retailer-level and channel-level outcomes of online-offline channel integration[J]. Journal of Retailing, 2015, 91(2): 309-325.
- [15] Erevelles S, Fukawa N, Swayne L. Big data consumer analytics and the transformation of marketing[J]. Journal of Business Research, 2016, 69(2): 897-904.
- [16] Fan S, Lau R Y K, Zhao J L. Demystifying big data analytics for business intelligence through the lens of marketing mix[J]. Big Data Research, 2015, 2(1): 28-32.
- [17] 褚洪洋, 柴跃廷, 刘义. 基于层次分裂算法的价格指数序列聚类[J]. 清华大学学报(自然科学版), 2015, 55(11): 1178-1183.

- [18] 范生万,刘放. 商品精确营销中聚类分析与关联规则分析应用研究[J]. 华东经济管理,2017,31(5):182-184.
- [19] 张玲玲,张笑,崔怡雯. 基于聚类方法的百度搜索指数关键词优化及客流量预测研究[J]. 管理评论,2018,30(8):126-137.
- [20] Ren P, Liao Z, Ge P. Real-time forecast of tourists distribution based on the improved K-means method[J]. Mathematical Problems in Engineering, 2014(4):1-10
- [21] Chung W, Tseng T L. Discovering business intelligence from online product reviews: a rule-induction framework[J]. Expert Systems with Applications, 2012, 39(15):11870-11879.
- [22] 戴稳胜,袁卫,谢邦昌. 顾客喜欢的下一个产品是什么——以数据挖掘发现交叉销售机会[J]. 中国统计,2005(8):42-43.
- [23] 张德鹏,黄雪玲. 基于关联规则的 PB 商品营销组合[J]. 工业工程,2007(5):50-52,57.
- [24] 何黎,何跃,霍叶青. 微博用户特征分析和核心用户挖掘[J]. 情报理论与实践,2011,34(11):121-125.
- [25] 郭燕,陈国华,陈之昶. “互联网+”背景下传统零售业转型的思考[J]. 经济问题,2016(11):71-74.
- [26] 窦润亮,郭均鹏,田祥龙,等. 面向客户个性化需求的交互式遗传算法[J]. 管理科学学报,2016,(1):24-34.
- [27] Lin C, Choy K L, Ho G T S, et al. A genetic algorithm-based optimization model for supporting green transportation operations[J]. Expert Systems with Applications an International Journal, 2014, 41(7):3284-3296.
- [28] Ko Y D, Jang Y J, Lee M S. The optimal economic design of the wireless powered intelligent transportation system using genetic algorithm considering nonlinear cost function[J]. Computers & Industrial Engineering, 2015, 89(11):67-79.
- [29] Dib O, Manier M A, Moalic L, et al. Combining VNS with genetic algorithm to solve the one-to-one routing issue in road networks[J]. Computers & Operations Research, 2017, 78:420-430.
- [30] 赵泉午,赵军平,林娅. 基于 O2O 的大型零售企业城市配送网络优化研究[J]. 中国管理科学,2017,25(9):159-167.
- [31] Rimal B P, Jukan A, Katsaros D, et al. Architectural requirements for cloud computing systems: an enterprise cloud approach[J]. Journal of Grid Computing, 2011, 9(1):3-26.
- [32] Lee C K H, Ho G T S, Choy K L, et al. A RFID-based recursive process mining system for quality assurance in the garment industry[J]. International Journal of Production Research, 2014, 52(14):4216-4238.