



兰州大学

本科毕业论文（设计）

论文题目（中文） 基于机器学习的连锁零售店需求预测和库存决策研究

论文题目（英文） Research on demand forecast and inventory decision of chain retail stores based on machine learning

学生姓名 王畅

指导教师 洪兆富（教授）

学 院 管理学院

专 业 管理学基地班（信息管理与信息系统）

年 级 2017 级

兰州大学教务处

诚信责任书

本人郑重声明：本人所提交的毕业论文（设计），是在导师的指导下独立进行研究所取得的成果。毕业论文（设计）中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，不包含任何其他个人或集体已经发表或在网上发表的论文。

本声明的法律责任由本人承担。

论文作者签名： 王畅

日期： 2021.5.29

关于毕业论文（设计）使用授权的声明

本人在导师指导下所完成的论文及相关的职务作品，知识产权归属兰州大学。本人完全了解兰州大学有关保存、使用毕业论文的规定，同意学校保存或向国家有关部门或机构送交论文的纸质版和电子版，允许论文被查阅和借阅；本人授权兰州大学可以将本毕业论文的全部或部分内 容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本毕业论文。本人离校后发表、使用毕业论文或与该论文直接相关的学术论文或成果时，第一署名单位仍然为兰州大学。

本毕业论文研究内容：

可以公开

不宜公开，已在学位办公室办理保密申请，解密后适用本授权书。

（请在以上选项内选择其中一项打“√”）

论文作者签名： 王畅

导师签名： 张世高

日期： 2021.5.29

日期： 2021.5.28

基于机器学习的连锁零售店需求预测和库存决策研究

中文摘要

连锁零售店作为分销商的一种形式，具有商品种类固定、顾客流量大但忠诚度不高、需求不稳定等特点。近年来由于线上销售的兴起和突发新冠疫情的影响，使得连锁零售店的生存环境进一步恶化。在如此多的不稳定因素下，如何准确地预测各类商品的需求，进而帮助连锁零售店做出合理的采购库存决策，是一个急需解决的问题。本文采用随机森林回归（RFR）、K近邻回归（KNN）、支持向量回归（SVR）以及极端随机树回归（ETR）四种机器学习算法对验证集的商品需求进行预测，并采用五折交叉验证的方法来评估模型，选择极端随机树回归（ETR）作为最优的预测模型并对测试集的商品需求进行预测。随后基于极端随机树回归的需求预测结果，构建考虑订货提前期和服务质量的两种库存决策模型，采用整数线性规划算法求解订货点和订货量，并对比分析了动态调整库存策略的成本优势。最后对服务水平 α 进行灵敏度分析，发现总成本会随着服务水平的上升而增加，且动态调整库存决策模型的总成本总低于静态库存决策模型的总成本，进一步验证了动态调整库存决策模型对于指导连锁零售店进行库存决策的优势。

关键词：连锁零售店，机器学习，需求预测，库存决策

Research on demand forecast and inventory decision of chain retail stores based on machine learning

Abstract

As a form of distributors, chain retail stores have the characteristics of fixed product types, large customer flow but low loyalty and unstable demand. In recent years, due to the rise of online sales and the sudden impact of the COVID-19, the living environment of chain retail stores has further deteriorated. With so many unstable factors, how to accurately predict the demand for various commodities, and then help chain retail stores to make reasonable purchasing and inventory decisions, is an urgent problem to be solved. This article uses four machine learning algorithms: Random Forest Regression (RFR), K-Nearest Neighbor Regression (KNN), Support Vector Regression (SVR), and Extra Tree Regression (ETR) to predict the commodity demand of the validation set. The five-fold-cross-validation method is used to evaluate every model, and Extra Tree Regression (ETR) is selected as the optimal prediction model to predict the commodity demand of the test set. Subsequently, based on the demand forecast results of Extra Tree Regression, two inventory decision-making models considering order lead time and service quality were constructed. The integer linear programming algorithm is used to solve the order point and order quantity, and the cost advantage of dynamic-adjusting inventory strategy is compared and analyzed. Finally, a sensitivity analysis of service level is carried out, and it is found that the total cost will increase with the increase of service level, and the total cost of dynamic-adjusting inventory decision model is always lower than the total cost of the static inventory decision model, further verify the advantages of the dynamic-adjustment inventory decision model for guiding the chain retail stores to make inventory decisions.

Key words: chain retail stores, machine learning, demand forecast, inventory decision

目 录

第一章 绪 论	1
1.1 研究背景与意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	1
1.2 国内外研究现状.....	2
1.2.1 需求预测.....	2
1.2.2 库存决策.....	3
1.3 研究内容.....	4
第二章 需求预测和库存决策理论基础	6
2.1 需求预测相关理论.....	6
2.1.1 连锁零售店需求预测.....	6
2.1.2 机器学习预测模型.....	6
2.1.3 模型性能评估指标.....	8
2.2 库存管理理论.....	9
2.2.1 库存管理的作用.....	9
2.2.2 确定需求库存模型.....	9
2.2.3 随机需求库存模型.....	10
第三章 基于机器学习的连锁零售店需求预测	11
3.1 数据收集.....	11
3.2 数据预处理.....	11
3.2.1 数据清洗.....	11
3.2.2 特征构建.....	12
3.3 模型预测结果分析.....	13
3.3.1 模型超参数.....	14
3.3.2 模型性能评估.....	14

3.3.3 极端随机树回归预测结果.....	15
第四章 基于需求预测的连锁零售店库存决策模型.....	18
4.1 静态库存决策模型.....	18
4.1.1 模型变量说明.....	18
4.1.2 模型建立.....	19
4.1.3 模型求解.....	21
4.2 动态调整库存决策模型.....	22
4.2.1 模型变量补充说明.....	22
4.2.2 模型建立.....	22
4.3 实验结果分析.....	24
4.3.1 静态库存策略数值实验.....	24
4.3.2 动态调整库存策略数值实验及对比分析.....	25
4.3.3 服务水平灵敏度分析.....	26
第五章 总结与展望.....	27
5.1 总结.....	27
5.2 展望.....	27
参考文献.....	29
致 谢.....	30

第一章 绪论

1.1 研究背景与意义

1.1.1 研究背景

自中国加入世贸组织以来，国内连锁零售业迅速发展。但自2008年以来，由于电子商务的逐渐兴起、在线零售商店的竞争加剧以及国内外的复杂局面，社会消费品零售总额增长率持续趋缓，从2008年的22.7%降到2019年的8%，这表明行业逐渐趋于饱和，业内竞争日益激烈。此外2020年突发的新冠肺炎疫情，导致线下购物遭遇爆冷状况，实体零售业一度陷入崩溃状态，超级市场、连锁店、便利店、零售店等多种业态遭受巨大冲击，消费者需求不确定性增加，市场环境日益复杂，行业生存、竞争压力逐渐增大。连锁零售业想要在这次大考面前立于不败之地，就要找到企业运营质量和运营成本的平衡点，建设以市场环境变化为导向的供应链体系，提升企业竞争力，而这一决策过程中一个极其重要的环节便是有效地进行库存管理^[1]。

库存的直接目的是满足需求，因此需求预测是库存管理的基础和前提，并发挥着重要的作用，特别是在零售企业^[2]。零售商依靠需求预测来做出营销、生产、库存和财务方面的商业决策。不准确的预测可能会导致缺货，从而导致销量和客户的损失；或过量库存，从而导致企业资金周转困难以及产品过期等库存成本问题。因此，连锁零售业想要维持合理的库存水平，在保证运营效率的同时使得库存成本最小化、收益最大化，就要准确的进行需求预测。

在大数据呈井喷式爆发的背景下，无论是线上还是线下的企业都掌握着大量的用户消费数据，且数据类型也越来越多样化。传统统计学方法已经不能够对多元数据之间的相关性进行很好地呈现和拟合，因此近几年越来越多的研究基于包含历史销售数据在内的各种多元化数据，比如产品类型、地区类型、价格变动等，建立预测精度高、泛化能力强、大数据处理效率高的预测模型，从而为合理的库存决策提供可靠保障。

机器学习模型相比传统的预测模型有着更为出色的处理复杂庞大数据的能力，近年来受到越来越多学者的关注。这些技术已经被广泛的应用于各个研究领域的预测中，并且展示出了较传统预测算法更为强大、精确的预测能力。因此针对连锁零售店海量用户消费数据，采用机器学习方法进行零售需求预测，基于预测结果制订合理的库存策略，为变量复杂、数据庞大的连锁零售店提供更加精准的需求预测和更有效的库存管理。

1.1.2 研究意义

1、理论意义

本文探究了基于机器学习的连锁零售店需求预测与库存决策问题，是大数据技术在库

存管理中的应用,能够降低库存成本,提高运营效率,对连锁零售企业的发展具有指导意义。针对连锁零售行业变量复杂、数据庞大的特点,通过选择有别于传统预测方法的几种泛化能力强、大数据处理效率高的机器学习模型,大大提高了预测精度。这对需求预测模型的优化具有借鉴意义。根据预测结果,设计不同的库存决策方案,这对库存决策是否合理具有重要指导意义。

2、实际意义

本文的需求预测结果服务于库存决策模型的构建。采用机器学习方法,刻画了包含历史销售数据在内的各种多元化的数据,分析未来不同商品的需求趋势,为连锁零售店采购计划和库存管理提供了依据。准确的需求预测能够降低连锁零售店库存成本。同时,连锁零售店把客户需求信息共享给供应商,也能够帮助供应商进行合理生产、采购,降低供应商库存成本,从而达成共赢。对连锁零售业销售数据的充分挖掘,有利于企业在线上零售兴起和疫情冲击等复杂市场环境中站稳脚步,快速灵活地响应市场需求和局势的变化。此外,通过对海量销售数据的分析,可以让连锁零售店更好的掌握消费者的需求特性,从而制定相应的服务策略,提升行业竞争力。

1.2 国内外研究现状

1.2.1 需求预测

本文涉及的需求预测是针对连锁零售店商品的需求预测,但实际生活中,需求预测应用范围特别广泛,包括能源需求预测、交通需求预测、物流需求预测等。不同应用领域在进行需求预测时选取的特征和影响因素不同,采取的预测方法也不尽相同。例如 Ma 等介绍了一种空调系统能耗预测方法,考虑了天气、是否工作日、室外温度等特征,建立了基于组合权重选择的空调系统能耗预测模型,最后通过算例验证了该方法的有效性^[3]。Li 结合神经网络和模糊系统的优点,对动态模糊神经网络(D-FNN)交通流进行优化改进,用于交通流时间序列预测,并对比分析了该方法相较于神经网络方法能够有效地提高预测精度^[4]。李顺等采用灰色关联算法,分析某港口物流需求同当地经济状况的相关度,建立港口物流需求预测特征集,提出一种基于遗传算法的极限梯度提升树优化模型^[5]。

需求预测研究的主要关注点为预测方法。传统需求预测方法主要是运用统计学的原理进行预测,主要有:移动平均法、指数平滑法、ARIMA 预测法、回归分析法等。赵志翔对 B 公司防护商品的管理现状进行分析,并在此基础上采用移动平均、加权移动平均等方法预测产品需求^[6]。赵军和王晓针考虑需求独立分布、随机不确定等因素影响,将指数平滑法运用于物流配送中心的库存预测^[7]。Huber 等提出了决策支持系统的效用,利用多元 ARIMA 来支持工业化烘焙连锁店的日常运营决策,同时考虑了汇总预测的日内销售模式^[8]。Shi 等采用多元线性回归的方式对跨境时装电商企业 A 的产品需求进行预测,并采用机器学习的方法进行预测优劣对比^[9]。

但近年来由于大数据时代的到来和计算机硬件性能的提升,企业的运营数据逐渐趋于复杂化和多样化,传统的预测方法已不能准确地表现出各类型数据之间复杂、非线性的关系,从而预测性能也会大打折扣。这使得学者们开始探究新的预测方法,这时基于数据挖掘和机器学习的预测方法成为研究热点。

1.2.2 库存决策

库存决策的目的就是通过决策补货点和补货量来控制库存成本,即确定最优的补货点和补货量,使得库存总成本最小。常见的库存订货方法有定期订货和定量订货。定期订货是按照固定的周期对库存进行核查,当库存量不足时进行订货。定量订货是在确定订货点和订货量的前提下,如果库存水平达到订货点时立即进行订货,且订货量为一确定值。经济补货批量(EOQ)模型是定量订货最常见的模型。罗兵等研究了产品价格随外界需求变化的企业库存决策,建立了关于易腐产品基于价格和销售量的库存决策模型^[10]。孙士雅等分析了允许缺货且存在提前期补货条件下的企业缺货成本,提出了缺货成本细分理论(缺货成本和销售机会损失成本),构建了缺货部分延迟的模型^[11]。然而在大多数情况下市场的需求具有很大的不确定性,不确定性随机需求情况下的库存策略依赖需求的准确预测,基于预测的结果进行库存决策。

对于随机需求下的库存决策问题,大多研究是通过获取需求分布规律来制订最优库存决策模型。徐贤浩等研究了易逝产品的生命周期规律,改进 BASS 模型,得到易逝产品的需求分布特征,进而制定了更加合理的订货策略^[12]。李卓群等通过对不确定需求因素的分析,构建一种不受不确定需求干扰的库存系统动态模型,并通过仿真实验得出了此模型下的最优订货策略,具有一定实践意义^[13]。Huang 等结合不确定性理论和竞合博弈的思想,在需求不确定的情况下,同时考虑生产成本和采购成本,提出一种合理的订货和定价策略^[14]。Qin 等将不确定性理论引入订货模型中,得到了企业在外部需求不确定性的情况下的最优订货量^[15]。

需求预测是库存决策的基础和依托。近年来学者们探究、设计了适用不同场景的需求预测方法,并对比分析得到最优预测模型,依此优化企业库存决策,制订最佳库存策略。罗晓萌等在考虑价格影响因素的前提下,采用时间序列的方法依据历史需求对未来需求进行预测,并根据需求预测情况,提出 (s,S) 订货策略^[16]。李长春采用不同的机器学习模型来进行产品需求预测,通过对比分析得到预测能力最好的模型——随机森林,基于预测结果,创建报童模型,得到产品的最优订货量^[17]。Roberto 等研究了汽车备件的库存问题,采用了不同的需求预测方式并通过实验仿真确定了最优库存策略^[18]。Rong 研究了多阶段订货、库存决策问题,建立了多阶段库存马尔可夫链预测模型,求得最优订货点和订货量^[19]。

综上所述,库存决策的研究大多与需求联系在一起,而在实际情况下市场需求大多为不确定的。不确定性需求需要采用合理的预测方法得到其需求分布情况,从而更好的进行库存决策。而大多数基于需求预测的库存决策研究都是单周期订货模型,忽略了订货提前

期，这与实际情况是不相符的。

1.3 研究内容

根据以上研究现状的分析，可以看到需求预测和库存决策的研究还存在以下不足：

(1) 连锁零售店商品种类繁多，且该销售模式受到线上销售和疫情的冲击，需求极不稳定，采购、库存风险较大，运用经典预测方法预测效果不好，而机器学习预测方法可以解决这一问题。但运用机器学习方法预测连锁零售店需求并且进行对比分析得到最优预测模型的研究不多。

(2) 在需求预测指导库存决策的研究中，预测方法单一，与实际需求情况不相符。针对连锁零售店采用机器学习方法预测需求，指导企业库存决策的研究较少。确定订货点和订货量后根据每期已实现的需求进行动态调整的库存决策研究不多。

基于以上的问题，本文研究内容如下：

(1) 针对连锁零售店历史销售数据，采用随机森林回归 (RFR)、K 近邻回归 (KNN)、支持向量回归 (SVR) 以及极端随机树回归 (ETR)，根据样本量大小，采用五折交叉验证的方法来评估每一种模型的泛化能力和预测精度，选出最优的预测模型。

(2) 基于极端随机树回归的需求预测结果，构建考虑订货提前期和服务质量的动态调整库存决策模型，动态调整库存决策模型是基于静态库存决策模型优化的结果。最后在数值实验中，通过改变服务水平对比分析两种库存决策模型所产生的库存成本的高低，从而进一步验证了动态调整库存决策模型对于指导连锁零售店进行库存决策的优势。本文结构如下：

第一章为本文的绪论。本章首先介绍了连锁零售店需求预测及库存决策的研究背景和研究意义，接着回顾了国内外相关领域的研究现状，通过对比分析以往研究的不足，引出本文的研究内容和论文结构。

第二章为需求预测和库存决策理论基础。本章介绍了需求预测理论和库存基础理论，引出了本文用到的四种机器学习方法以及性能指标，为后续需求预测模型和库存决策模型的建立提供理论基础。

第三章为基于机器学习的连锁零售店需求预测。本章对某连锁零售店的历史销售数据进行了数据清洗、特征转换、特征筛选，数据标准化以及训练集、测试集划分等预处理工作，然后通过五折交叉验证的方法划分训练集和验证集，并对比分析了四种机器学习模型的泛化能力和预测精度，选出最优预测模型。

第四章为基于最优需求预测的库存决策模型。本章基于最优预测模型的需求预测结果，构建了静态库存决策模型与动态调整库存决策模型，通过改变服务水平对比分析两种库存决策模型所产生的库存成本的高低，从而进一步验证动态调整库存决策模型对于指导连锁零售店进行库存决策的优势。

最后一章为结论和展望。本章根据上文的研究结果提出了本文的主要研究结论，并阐述了研究展望。

第二章 需求预测和库存决策理论基础

2.1 需求预测相关理论

2.1.1 连锁零售店需求预测

连锁零售店作为分销商的一种形式，具有商品种类固定、顾客流量大但忠诚度不高、需求不稳定等特点。近年来由于线上销售的兴起和突发新冠疫情的影响，使得连锁零售店的生存环境进一步恶化，行业竞争压力进一步提升。在如此多的不稳定因素下，如何根据大量的历史流量数据准确地预测各类商品的需求，进而帮助连锁零售店做出合理的采购库存决策，是一个急需解决的问题。

需求预测不仅会影响连锁零售店，而且对连锁零售店所在的整个供应链上下游都会产生很大影响。整个供应链上下游包含了制造商、供应商、零售商、最终消费者。需求信息由供应链中最下游的最终消费者产生，通过供应量向上游层层传递，在此过程中由于噪声和其他因素的干扰会导致需求逐渐失真，当到达最上游的时候，企业获得的需求信息可能与真实的需求存在很大差异，因此不利于企业做出下一步的战略决策。连锁零售店作为供应链中最接近最终消费者的一环，能够获取大量及时、准确的需求动态信息，在连接供应链上下游各环节中起到了桥梁纽带作用。因此，连锁零售店的准确需求预测可以帮助上游环节做出合理的生产、分配决策，同时也有利于连锁零售店自身的采购和库存决策，从而达成共赢。本文的研究主要是从供应链角度出发，以连锁零售店自身的库存决策为目标，准确地进行需求预测，从而提升连锁零售企业在供应链环境中的竞争力。接下来介绍下本文用到的机器学习预测模型和性能指标。

2.1.2 机器学习预测模型

1、随机森林（Random Forest, RF）

随机森林在近十年的机器学习应用中广受学者欢迎，由于其具有很强的泛化性能、可扩展性和易用性。简单说，可以把随机森林看成是多个决策树的集合，其中每棵决策树通过随机有放回从数据集中抽取多个样本来构建。若为分类问题，则以基尼杂质、熵等作为杂质度量目标、以最大化信息增益原则分裂节点，最后通过多数票机制确定标签的分类；若为回归问题，则以均方误差作为杂质度量目标，以最大化信息增益原则分裂节点，最后通过取多个决策树预测均值的方式得到最终预测结果。本研究是连续型需求预测问题，因此采用随机森林回归（Random Forest Regressor）取均值的方法得到预测结果。

随机森林使用了决策树 CART（Classification And Regression Tree）作为弱学习器，但随机森林不是对多个决策树的简单集合。在生成每棵树的时候，每个树选取的特征都仅仅是随机选出的少数特征，一般为 \sqrt{m} （ m 为总特征数），而决策树是选取所有特征进行建模。此外随机森林在构建每棵决策树的时候随机选择 n 个可替换样本，而决策树则选取全

部样本进行建模。由于随机选择特征集和样本集，这使得随机森林有着很好的泛化能力和较低的方差，不容易陷入过拟合。随机森林的其他优点还包括对于离群值不敏感，可并行化处理，不需要太多的参数调优，唯一需要测试的参数是决策树的个数。

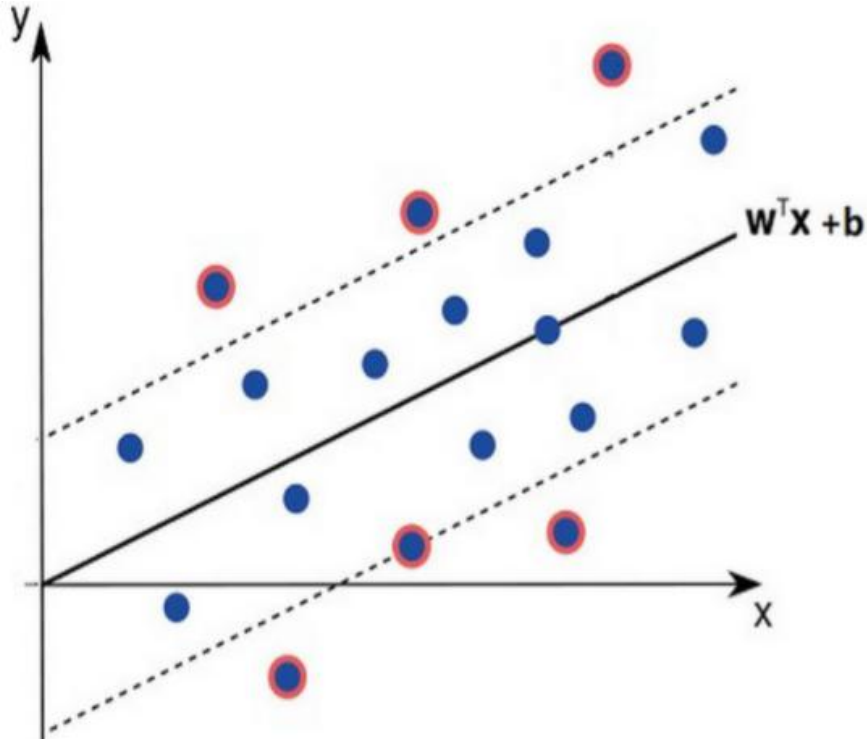
2、K 近邻 (K-Nearest Neighbor, KNN)

KNN 算法是一种非参数机器学习模型，通过记忆训练过的样本来进行后续的预测工作，既可以进行分类预测，又可以进行回归预测。对于测试样本，基于样本距离度量找出训练集中 K 个与其最接近的训练样本，然后基于这 K 个训练样本的信息来进行预测。对于回归预测，可使用平均或加权平均法把 K 个训练样本的平均值或按照距离度量的加权平均值作为回归预测结果。其中距离度量标准常用的有闵可夫斯基距离、欧氏距离以及曼哈顿距离。本研究选择 k 近邻回归 (K-Neighbors Regressor) 的方法来进行需求预测。

3、支持向量机 (Support Vector Machine, SVM)

支持向量机 (SVM) 是一种由传统感知机延伸的机器学习方法，支持向量机算法优化的目标是寻找分离决策边界和其样本集之间的最大距离。该机器学习方法应用范围较广的一个原因是可以通过核函数来处理非线性问题，对于低维空间中线性不可分的样本集，通过映射函数，把原始特征投影到高维空间进行处理。在高维空间中，特征变得线性可分，通过构建决策模型使得正负超平面之间的距离最大化，并且尽量保持支持向量距离最大化。这就是 SVM 的分类原理。而本文的研究问题是连续型的需求预测，所以我们通过采用 SVM 的延伸模型支持向量回归 (Support Vector Regression) 来进行建模。

SVR 的原理和 SVM 相同，区别在于构建模型时，SVR 在线性函数的两侧引入了一个“隔断带”，该算法的原理是寻找一个超平面以最小化该平面与所有支持向量的距离。当样本落入“隔断带”之内时被认为预测准确，不计算损失，当样本落入“隔断带”之外时，计入损失函数，最终通过最小化损失函数来优化模型，提升预测能力。其构建原理如图 2.1 所示。

图 2.1 支持向量回归建模原理¹

4、极端随机树（Extra Tree）

极端随机树的原理和随机森林非常相似，都是由多棵决策树构成。它们的主要区别是：随机森林的原理是 **Bagging** 模型，即样本是有放回的随机抽取，而极端随机树使用的是所有样本，仅每一棵树的特征是随机选取的；随机森林是在每一个随机选择样本的决策树中得到最佳分枝属性，而极端随机树是完全随机的得到分枝属性。因为节点随机分裂，所以某些情况下比随机森林得到的结果更接近实际情况。最后，尽管所有决策树采用同一个样本集，但由于随机选择最佳分枝属性，所以每棵树仍会得到完全不同的预测结果。对于本研究，极端随机树同样采用均方误差作为杂质度量目标，采用极端随机树回归（**Extra Tree Regressor**）取均值的方法得到预测结果。

2.1.3 模型性能评估指标

机器学习预测模型需要在测试集上评估其预测性能，这时评估指标可以用来表示模型预测的效果。评估指标是机器学习中不可或缺的一环，不同的预测算法有着不同的评估指标，同一种预测算法也有着不同的评估指标，每个指标从不同的角度对模型进行评估，侧重点不同。本研究的需求预测为连续型回归问题，所以以下介绍要用到的回归模型评估指标：

R^2 ：即预测值解释了实际值方差的多大比例，反映了模型的拟合好坏程度。 R^2 越大，表明预测值对于实际值的拟合程度好。公式如下：

¹ 图片来源：https://blog.csdn.net/qq_41909317/article/details/88542892

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (1)$$

MAE: 即实际值与预测值的平均绝对偏差, 该值小, 表明预测越接近真实值。公式如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

RMSE: 即实际值与预测值的平均误差平方和的算术平方根, 该值越小, 表明训练的模型误差越低。公式如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

2.2 库存管理理论

2.2.1 库存管理的作用

库存是企业对任何有经济价值的物品的闲置与储备, 它对企业维持正常的生产运作至关重要。而库存管理是企业对库存进行计划、协调、控制等一系列行为, 它与企业总体成本、利润联系密切。库存管理中最关键的是库存控制, 即在合适的时机做出合适的库存决策, 确定合理的订货点以及订货量, 在保证的服务质量约束下以最小的成本满足顾客需求。库存管理的作用主要有:

- (1) 调节企业生产经营, 控制库存成本, 加速资金周转。
- (2) 平稳生产过程, 使企业做出合理生产决策满足顾客需求。
- (3) 掌握库存动态信息, 适时, 适量做出订货决策, 避免积压或缺货。

2.2.2 确定需求库存模型

确定需求库存模型中最常用的是基本经济补货批量 (EOQ), 通过求解最小库存成本得到每次的最优订货量。该模型存在一些前提假设:

- 1) 无限补充能力, 在补货点一次性到货。
- 2) 不允许缺货, 库存到达零时立即补货。
- 3) 无补货提前期, 订货点和补货点一致。
- 4) 市场需求恒定。
- 5) 无补货数量折扣。

总成本=持货成本+补货成本

$$TC(Q) = \frac{1}{2}Qh + \frac{Ds}{Q} \quad (4)$$

TC : 总成本

Q : 单次补货量

h : 单位产品持有成本

s : 单次补货成本

D : 总需求量

把公式 (2.4) 对 Q 求一阶导并令其等于零, 得到单次最优订货量 Q^* 。

$$Q^* = \sqrt{\frac{2Ds}{h}} \quad (5)$$

2.2.3 随机需求库存模型

随机需求下的库存模型的主要决策变量为库存点和库存量, 具体库存策略有: (1) (r, Q) 策略, 该策略持续跟踪库存水平, 一旦发现库存水平低于 r 时立即进行补货, 且补货量为 Q 。(2) (s, S) 策略, 依照一定的周期盘点库存, 即在盘点时如果库存水平超过 s , 不进行补货; 如果库存水平不超过 s , 就将库存水平补充到 S 。(3) (t, S) 策略, 该策略为固定周期 t 的库存盘点, 即在盘点时如果库存水平低于 S , 则立即进行补货, 并且将库存水平补至 S 。

随机需求下的库存模型同时还包括了静态库存模型和动态库存模型。静态库存模型, 即根据产品的需求分布, 在计划期开始时就确定整个计划期内的订货点和订货量。但随着每个周期需求的实现, 确定好的库存点和库存量不能调整, 可能会造成库存成本的上升。而动态库存模型是根据需求的变化和每期需求的满足情况, 对订货点和订货量进行调整。但由于需求的不确定性, 往往不能制定出一定时期内的补货方案, 造成补货成本上升, 管理困难等问题。

本文基于机器学习需求预测结果, 在商品随机需求的情况下, 考虑了更贴合企业运营实际的订货提前期和服务水平约束等条件, 构建了静态库存决策模型和动态调整库存决策模型。

第三章 基于机器学习的连锁零售店需求预测

3.1 数据收集

本文中使用的数据来自于大数据竞赛平台 Kaggle 上的大型连锁零售企业 A 的历史销售数据¹。企业 A 在印度 Maharashtra, Telangana and Kerala 三个州设有连锁零售店。这些连锁零售店出售种类固定的多种商品，主要包括快速消费品（fast moving consumer goods），饮料和食物（drinks and food）以及其他类商品（others），共 50 种商品。各地共有 10 家连锁零售店，其中 Maharashtra 有 4 家连锁零售店，Telangana 和 Kerala 分别有 3 家连锁零售店。每家零售店还设 6 个不同的销售区域，每个区域售卖固定种类的商品，每种商品的价格会随着时间以及需求的变化做出微小的调整。连锁零售企业 A 为方便统一化管理，要求每个州的每个店铺中的商品数量、商品种类、价格以及销售区域数量都是完全相同的。

目前为止，每一种商品的需求都是由零售店经理根据历史销售情况和自己的经验判断进行预测。但由于不同地区复杂多变的市场环境，依据经验进行预测的准确性与实际需求会有很大差别。为了避免这一问题，连锁零售店急需一个科学准确的预测模型来对各种商品的需求进行预测，从而为合理的库存决策提供可靠保障。

3.2 数据预处理

3.2.1 数据清洗

该竞赛数据为连锁零售企业 A 从 2012 年 1 月到 2014 年 2 月的历史销售数据，共 395000 条，分为商品基本信息、时间信息、销售信息三个数据集。其中商品基本信息数据集包括店铺编号、商品编号、周数以及商品价格；时间信息数据集包括日期和周数；销售信息数据集包括日期、商品编号、商品种类、店铺编号、销售区域、州名以及商品日销量。本文采用 excel、python 数据分析包 numpy、pandas 等，通过保留三个数据集中不同的特征，删除重复的特征，初步构建了包含商品所有信息的历史销售数据，如表 3.1 所示。

表 3.1 历史销售数据样本

日期	商品编号	销售区域编号	商品种类	店铺编号	州	销量	周数	价格
2012/1/1	74	11	others	111	Maharashtra	0	49	2.94
2012/1/1	337	11	others	111	Maharashtra	1	49	7.44
2012/1/1	423	12	others	111	Maharashtra	0	49	0.97
2012/1/1	432	12	others	111	Maharashtra	0	49	4.97
2012/1/1	1753	31	drinks and food	111	Maharashtra	1	49	1.97

通过对上述数值型特征进行描述性统计，如表 3.2 所示。可以看到样本数据质量较好，分布比较均匀，不存在离群值、缺失值，不需要进行删除、插补等操作。

表 3.2 描述性统计

¹ 数据来源：<https://www.kaggle.com/mragpavank/predicting-the-sales-of-products-of-a-retail-chain>

	商品编号	销售区域编号	店铺编号	销量	周数	价格
样本量	395000	395000	395000	395000	395000	395000
均值	1509.96	24.46	211.2	1.228919	105.0709	4.987644
标准差	809.7995	6.337862549	91.16129	3.595266	32.57875	3.874444
第一分位数	926	21	113	0	77	2.68
中位数	1325	22	221.5	0	105	3.98
第三分位数	1753	31	331	1	133	6.48
最小值	74	11	111	0	49	0.05
最大值	3021	33	333	293	161	44.36
缺失值	0	0	0	0	0	0

由上图可以看出，日销量的中位数为 0，表明每个商铺的大多数商品的日销量为 0。如果以日销量为单位进行需求预测，在商品不可分割的情况下（商品需求数量不能为小数），得到的预测结果大多为 0 或 1，不能够很好地反应实际需求分布情况，从而会产生较大的误差。因此考虑以周作为销量计数周期，来进行需求预测。

本文把 2012 整年、2013 整年和 2014 年前两个月的销售数据以周为时间窗口进行划分，舍去 2012 年的最后两天和 2013 年的最后一天，以及 2014 年二月的后三天。然后以周为单位把各个店铺相同商品的日销量进行加总，得到 2012 年 52 周销售数据、2013 年 52 周销售数据和 2014 年前 8 周的销售数据，并分别把周数编码为 1-52、1-52 和 1-8，得到初步的机器学习样本数据集。

3.2.2 特征构建

在完成数据清洗工作后，还需要对数据进行特征构建，得到适合的特征子集才能让机器学习算法更好地拟合训练数据。本文将依次完成以下特征构建工作：特征转换、特征筛选、数据标准化以及通过 k 折交叉验证划分训练集、验证集和测试集。

1、特征转换

由表 3.1 可得商品种类和州为字符串类型，但机器学习算法只能处理数值型数据，因此把上述两个特征转换为 category（分类）数据类型，然后采用 Label Encoder 对 category 数据类型进行编码。最终得到以下编码方式：drinks and food = 0, fast moving consumer goods = 1, others = 2; Kerala = 0, Maharashtra = 1, Telangana = 2。

2、特征筛选

通过特征筛选剔除无关特征，选择适合机器学习算法的特征子集，提高模型效率。对于日期特征，由于需求预测的时间窗是以周为单位，为无关特征，予以剔除。对于周数 id 特征，由于该编码从 49 开始升序编码到 166，不同年份的周次数值不能对应，为无关特征，予以剔除，并保留每年 1-52 周的编码方式。此外对于商品种类和州两个数据类型为字符串的冗余特征，予以剔除，并保留 0-1-2 的编码方式。

3、划分训练集、验证集和测试集

在构建机器学习模型的过程中，评估模型在未见过的数据上的预测效果是对模型泛化

性能的有效评估。对于传统的数据集划分方法，一般把数据集划分成独立的训练集和测试集，前者用于训练模型，后者用来评估模型的泛化性能。然而在参数调整过程中，机器学习模型反复使用相同的测试集评估模型泛化性能，它将会成为训练数据的一部分，这样很容易导致模型过拟合。

因此本文考虑采用 k 折交叉验证的方法把数据集分裂为训练集、验证集和测试集。训练集用于拟合不同的模型，验证集用来验证模型对于未见过的数据的泛化性能，测试集用来评估模型的预测性能。通过得到模型在每一个子集上的性能评估结果 E_i 来计算模型的平均性能评估结果 E ，进而提高模型的泛化性能，避免过拟合。k 折交叉验证的原理图如下：

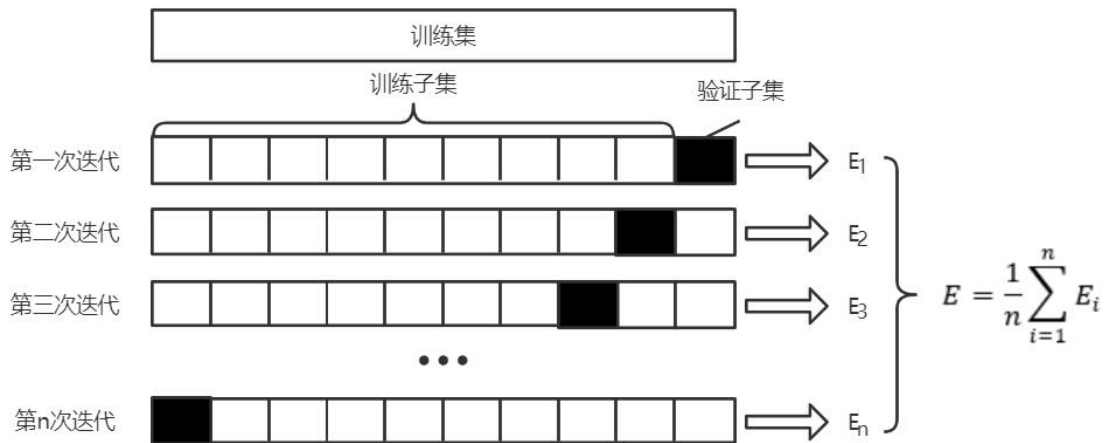


图 3.1 k 折交叉验证原理

本文首先把 2014 年 1-8 周的销售数据作为测试集，然后再对剩余的样本数据进行训练集和验证集的划分。考虑到样本量大小和各特征的类别占比状况，决定以每个店铺的每个销售区域作为分组，在每一个分组中采用 5 折交叉验证的方法划分训练集和验证集，保证训练集和验证集中能够包含所有商品，避免了完全随机划分训练集和验证集而导致的类不平衡问题，保证了模型的泛化性能，提高了模型对于未见过的数据的预测精度。

4、数据标准化

在原始数据样本集中，每个特征取值的范围不同，数值相差较大的特征会使机器学习模型训练时产生较大误差，影响预测精度。因此采用标准化的方式对训练集、验证集和测试集的特征进行处理，在保留了离群值的有用信息情况下把数据调整到有限的值域，进而提高算法预测性能。转换公式如下：

$$x_{std}^{(i)} = \frac{x^{(i)} - \mu_x}{\sigma_x} \quad (6)$$

其中 μ_x 为某个特征的样本均值， σ_x 为其标准差。

3.3 模型预测结果分析

本文以 R^2 、MAE 和 RMSE 作为模型性能评价指标，通过 5 折交叉验证来比较随机森

林回归 (RFR)、K 近邻回归 (KNN)、支持向量回归 (SVR) 和极端随机树回归 (ETM) 四种机器学习模型的泛化性能和预测精度, 选出最优的预测模型。最后基于最优预测模型对 2014 年 1-8 周的销量进行预测, 并随机选取不同商品类型的三种商品进行预测结果可视化分析。

3.3.1 模型超参数

本文利用 python 的 Sklearn 模块来对随机森林回归 (RFR)、K 近邻回归 (KNN)、支持向量回归 (SVR) 和极端随机树回归 (ETM) 四种机器学习模型进行 5 折交叉验证。表 3.3 为本文所使用模型的部分超参数 (以下所有机器学习模型的超参数组合都经过多次调优, 这里假设都可以使得对应模型在所有超参数组合中表现最优)。

表 3.3 本文所使用模型的部分超参数

机器学习模型	超参数名	超参数值
随机森林回归	n_jobs	-1
	criterion	mse
	min_samples_leaf	1
	min_samples_split	2
	n_estimators	1000
K 近邻回归	n_jobs	-1
	n_neighbors	5
	weights	uniform
	metric	minkowski
支持向量回归	p	2
	kernel	rbf
	gamma	0.1
极端随机树回归	C	100
	n_jobs	-1
	criterion	mse
	min_samples_leaf	2
	min_samples_split	2
	n_estimators	1000

3.3.2 模型性能评估

本文以 R^2 、MAE 和 RMSE 作为以上四种机器学习模型性能评价指标, 通过对 5 折交叉验证的每次性能评估结果求均值的方法得到模型最终的评估结果。性能评估结果如表 3.4 所示。

表 3.4 5 折交叉验证模型性能评估

机器学习模型	迭代次数/性能均值	R^2	MAE	RMSE
随机森林回归	1	0.2086	6.6353	15.2764
	2	0.6161	5.4323	13.8526
	3	0.2983	6.0311	15.217

	4	0.3867	5.7341	13.7946
	5	0.4092	5.4749	13.6733
	性能均值	0.3838	5.8616	14.3628
K 近邻回归	1	0.3585	6.5935	13.755
	2	0.532	6.0996	15.2951
	3	0.3563	6.4187	14.5747
	4	0.3951	6.3246	13.7002
	5	0.4598	5.7013	13.0744
	性能均值	0.4203	6.2275	14.0799
支持向量回归	1	0.3148	6.4219	14.2162
	2	0.2683	6.8726	19.1255
	3	0.2773	6.5199	15.4431
	4	0.3953	5.6779	13.6979
	5	0.3066	5.9433	14.8124
	性能均值	0.3125	6.2871	15.459
极端随机树回归	1	0.2937	6.3438	14.4327
	2	0.6392	5.2703	13.4307
	3	0.377	5.5795	14.3391
	4	0.49	5.3107	12.5788
	5	0.4937	5.1241	12.6574
	性能均值	0.4587	5.5257	13.4877

由上图可知，极端随机树回归的 R^2 为 0.4587 大于另外三种预测模型，表明极端随机树回归的模型解释力最优。同样极端随机树的 MAE、RMSE 分别为 5.5257 和 13.4877 小于另外三种预测模型，表明极端随机树回归的预测精度最高。综合可得，极端随机树回归相较于其他三种模型在未见过的数据上的预测精度和泛化能力较好，因此选择极端随机树回归进行测试集的需求预测，并把预测结果作为指导库存决策的依据。

3.3.3 极端随机树回归预测结果

本文把极端随机树回归作为最优预测模型，在保证模型超参数不变的情况下对 2014 年 1-8 周的需求进行预测，得到 R^2 为 0.3411，MAE 为 6.0088，RMSE 为 20.6638。

为更加直观的展现极端随机树回归的预测效果，我们随机选取不同州、不同店铺以及不同种类的三种商品，并分别命名为商品一、商品二和商品三，然后对其进行预测需求和实际需求的可视化对比。三种商品的基本信息如表 3.5 所示。

表 3.5 可视化商品基本信息

商品	产品编号	店铺编号	产品类型	州
商品一	74	113	others	Maharashtra
商品二	659	112	fast moving consumer goods	Telangana
商品三	2853	333	drinks and food	Kerala

1、商品一需求可视化对比

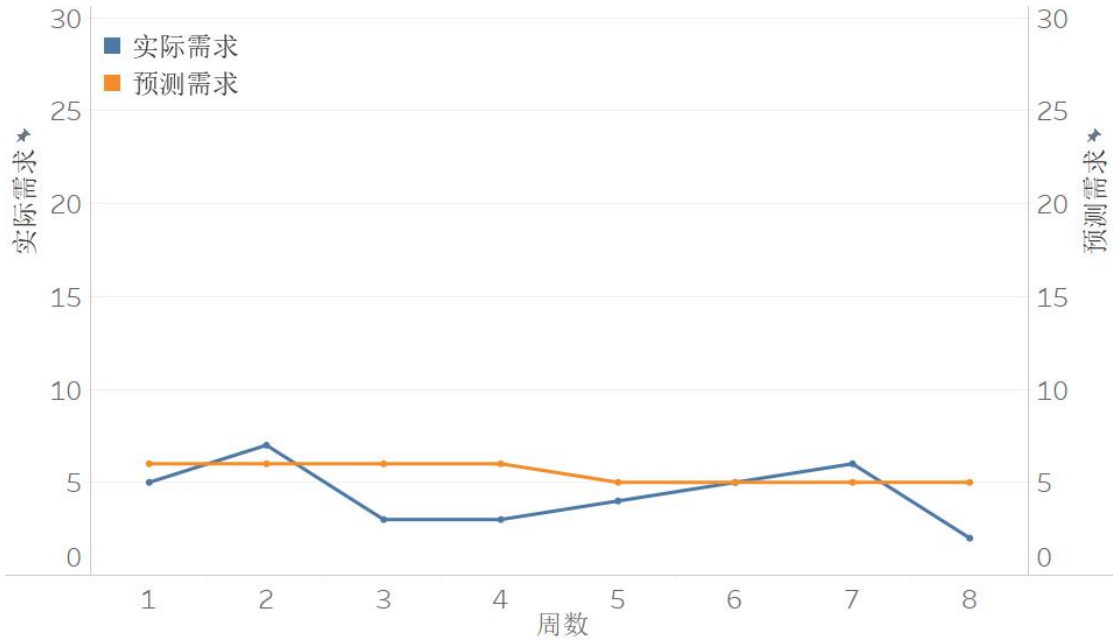


图 3.2 商品一预测需求和实际需求可视化对比

图 3.2 是 2014 年 1-8 周商品一 (other 类商品) 的预测需求和实际需求的可视化对比。由上图可以看出, 该商品的周需求量较小且需求波动较小。商品一第 3、4、8 周的预测需求与实际需求偏离最大, 偏离值为 3; 第 6 周的偏离最小, 偏离为 0, 总体预测效果较好。

2、商品二需求可视化对比



图 3.3 商品二预测需求和实际需求可视化对比

图 3.3 是 2014 年 1-8 周商品二 (fast moving consumer goods 类商品) 的预测需求和实际需求的可视化对比。由上图可以看出, 该商品的周需求量较大且需求波动较小。商品二第 6 周的预测需求与实际需求偏离最大, 偏离值为 5; 第 5 周的偏离最小, 偏离为 1, 总

体预测效果较好。

3、商品三需求可视化对比

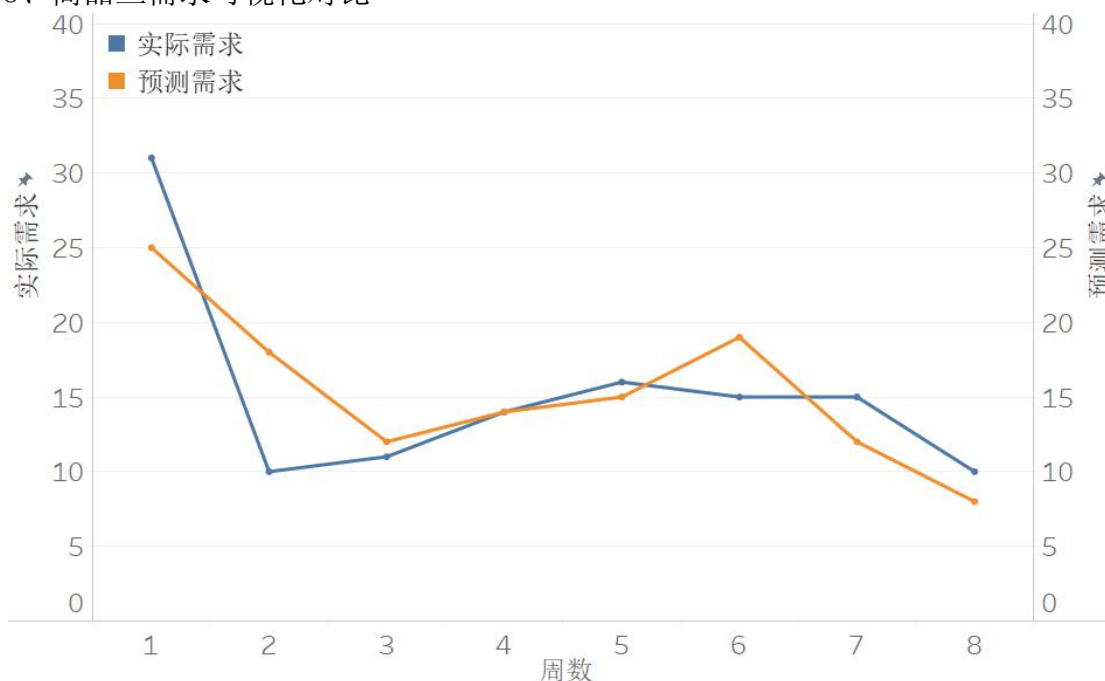


图 3.4 商品三预测需求和实际需求可视化对比

图 3.4 是 2014 年 1-8 周商品三（drinks and food 类商品）的预测需求和实际需求的可视化对比。由上图可以看出，该类商品的周需求量很大且需求波动较大。商品三第 2 周的预测需求与实际需求偏离最大，偏离值为 8；第 4 周的偏离最小，偏离为 0，总体预测效果较好。

总结来看，极端随机树回归的总体预测效果基本能反映出实际需求变化规律，预测精度较高，但是预测需求和实际需求仍存在或大或小的误差，预测值不能完全替代实际值。因此直接将预测值作为连锁零售店的订货量显然是不合理的，这会导致库存成本增加。所以本文需要构建更加符合实际的库存决策模型，以极端随机树回归的预测值作为建模依据，得到最优的订货点和订货量，以最大程度降低库存成本。

第四章 基于需求预测的连锁零售店库存决策模型

本章基于极端随机树回归的需求预测结果，构建静态库存决策模型与动态调整库存决策模型，然后通过改变服务水平对比分析两种库存决策模型所产生的库存成本的高低，选择适合连锁零售店的库存策略，进而降低连锁零售店的库存成本，提高服务质量。

4.1 静态库存决策模型

静态库存决策模型在计划期开始前就确定整个计划期的订货点和订货量，并且整个计划期内保持不变。在静态库存决策模型中，每个周期的需求会直接影响订货点和订货量的确定，进而影响库存决策成本。本章中整个计划期的商品需求通过极端随机树回归预测得出，并作为一个输入变量指导库存决策。

4.1.1 模型变量说明

本章的连锁零售店库存决策模型是基于极端随机树回归的需求预测结果建立的，以预测值为输入变量，建立静态库存决策模型与动态调整库存决策模型，确定合理的订货点和订货量。模型变量如下：

表 4.1 模型变量

变量符号	变量说明
T	指计划期，本文以 2014 年 1-8 周的预测需求作为库存决策依据，故 T 为 8 周
d_t	第 t 周的需求， $t=1, 2, 3, \dots, 8$ ，其中每周需求由极端随机树回归预测得到
$g(\cdot)$	d_t 的密度函数
X_t	第 t 周订货的决策变量，当值取 0 表示第 t 周不订货；当值取 1 表示第 t 周订货
Q_t	第 t 周的订货量，订货数量为正整数
I_t	第 t 周的期末库存水平
I_0	计划期期初库存水平
l	订货提前期
a	服务水平
k	固定订货成本，即每次订货产生的启动成本
v	可变订货成本，与订货数量呈正比关系
h	单位产品库存持有成本
$D(t)$	1-8 周总需求
$F_{D(t)}(\cdot)$	$D(t)$ 的累积分布函数
$F_{D(t)}^{-1}(\cdot)$	$F_{D(t)}(\cdot)$ 的反函数
$S_{D(t)}$	$D(t)$ 的标准差

基本决策思路：若 $t-1$ 周有订货，则订货会在第 t 周到达，在满足第 t 周的需求后，剩余的货物计入第 t 周期末库存水平，并根据库存水平计算库存成本。随后根据第 t 周末的库存量以及第 t 周后的需求预测做出下一次的订货决策。（其中整个决策过程要满足一定的服务水平）。

4.1.2 模型建立

静态库存决策模型中的总成本包括订货成本和库存持有成本，其中订货成本是由固定成本和可变成本组成，库存持有成本与每期期末的库存水平呈正相关关系。

此外整个订货决策需要满足一定的服务水平 a ，且有服务水平约束的库存决策模型一般要求本周的期末库存水平大于或等于下一周的需求，因此这里不考虑缺货成本。基于以上条件，静态库存模型 A 可以表示如下：

目标函数：

$$\text{Min } E\{TC\} = \sum_{t=1}^T (vQ_t + kX_t + hE\{I_t\}) \quad (7)$$

约束条件：

$$P\{I_t \geq d_{t+1}\} \geq a \quad (8)$$

$$E\{I_t\} = E\{I_{t-1}\} + Q_t - E\{d_t\} \quad (9)$$

$$Q_t \leq N * X_t \quad (10)$$

$$Q_t \geq 0 \quad (11)$$

$$E\{I_t\} \geq 0 \quad (12)$$

$$X \in \{0,1\} \quad (13)$$

其中 N 为一个无限大的正整数。

静态库存决策模型的目标函数是在计划期 $[1, T]$ 周内使订货成本和库存持有成本之和最小。因为有订货提前期，计划期初订购的货物将在 l 周后到达，则第 1 周到第 l 周的库存持有成本不在计划期的决策之内，所以 1 周到 l 周的库存持有成本不计入总成本。同理， $[T-l+1, T]$ 周的订货在计划期内不能送达，所以 $T-l+1$ 周到 T 周时的订货成本不计入总成本。则目标函数可以变为：

$$\text{Min } E\{TC\} = \sum_{t=1}^{T-l} (vQ_t + kX_t + hE\{I_{t+l}\}) \quad (14)$$

第 $t+l$ 周的期末库存水平可以用该周之前的订货量和需求量表示，即：

$$I_{t+l} = I_0 + \sum_{i=1}^t Q_i - \sum_{i=1}^{t+l} d_i \quad (15)$$

此时服务水平约束变为：

$$P\{I_{t+l} \geq d_{t+l+1}\} \geq a, t \in \{1, 2, 3, \dots, T-l\} \quad (16)$$

将公式 (15) 带入上式得到与订货量和需求量有关的服务水平约束：

$$P\left\{I_0 + \sum_{i=1}^t Q_i \geq \sum_{i=1}^{t+l+1} d_i\right\} \geq a \quad (17)$$

令 $D(t) = d_1 + d_2 + d_3 + \dots + d_t$ ，其中 $D(t)$ 为第 1 周到第 t 周的总需求且为随机变量。 $F_{D(t)}(\cdot)$ 为第 1 周到第 t 周总需求的累积分布函数。则上式可以变形为：

$$P\left\{I_0 + \sum_{i=1}^t Q_i \geq D(t+l+1)\right\} \geq a \quad (18)$$

由累积分布函数的性质得：

$$F_{D(t+l+1)}\left(I_0 + \sum_{i=1}^t Q_i\right) \geq a \quad (19)$$

通过反函数 $F_{D(t+l+1)}^{-1}$ ，可将上式变形为：

$$\sum_{i=1}^t Q_i \geq F_{D(t+l+1)}^{-1}(a) - I_0 \quad (20)$$

令 $F_{D(t+l+1)}^{-1}(a) = \mu$ ，则 $F_{D(t+l+1)}(\mu) = a$ ，由累积分布函数的性质得， $D(t+l+1) \leq \mu$ 的概率为 a 。其中 μ 表示在满足服务水平 a 下的第1周到第 $t+l+1$ 周的广义需求。因此，静态库存决策模型A最终可以写成：

目标函数：

$$\text{Min } E\{TC\} = \sum_{t=1}^{T-1} (vQ_t + kX_t + hE\{I_{t+l}\}) \quad (21)$$

约束条件：

$$E\{I_{t+l}\} = I_0 + \sum_{i=1}^t Q_i - \sum_{i=1}^{t+l} E\{d_i\} \quad (22)$$

$$\sum_{i=1}^t Q_i \geq F_{D(t+l+1)}^{-1}(a) - I_0 \quad (23)$$

$$Q_t \leq N * X_t \quad (24)$$

$$Q_t \geq 0 \quad (25)$$

$$E\{I_{t+l}\} \geq 0 \quad (26)$$

$$X \in \{0,1\} \quad (27)$$

其中 N 为一个无限大的正整数。

本章假设每周的需求服从正态分布，那么就可以通过需求预测值和标准差计算 $F_{D(t)}^{-1}(a)$ 。因为每周的需求服从正态分布且相互独立，所以第1周到第 t 周的总需求 $D(t) = \sum_{i=1}^t d_i$ 也服从正态分布，令 $E\{d_i\}$ 表示第 i 周的需求预测值， S_{d_i} 表示第 i 周需求的标准差，则第1周到第 t 周的总需求预测值和总需求标准差分别为： $E\{D(t)\} = \sum_{i=1}^t E\{d_i\}$ 和 $S_{D(t)} = (\sum_{i=1}^t S_{d_i}^2)^{1/2}$ 。当满足服务水平 a 时，正态分布的 Z 值为： $Z = (\mu - E\{D(t)\})/S_{D(t)}$ 。那么 $F_{D(t)}^{-1}(a)$ 可以表示为：

$$F_{D(t)}^{-1}(a) = \sum_{i=1}^t E\{d_i\} + Z * \left(\sum_{i=1}^t S_{d_i}^2\right)^{\frac{1}{2}} \quad (28)$$

这时就可以通过需求预测值、需求标准差以及服务水平 a 的 Z 值求解广义需求值。

4.1.3 模型求解

可以通过启发式算法求解最终的静态库存决策模型 A。可以把随机需求 $E\{d_i\}$ 通过 $F_{D(t)}^{-1}(a)$ 来表示。令 $\Delta d(t) = F_{D(t)}^{-1}(a) - \sum_{i=1}^t E\{d_i\}$ ，其中 $\Delta d(t)$ 表示 1 到 t 周的需求波动，且由公式 (28) 得， $\Delta d(t) = Z * (\sum_{i=1}^t S_{d_i}^2)^{1/2}$ 为一个确定值。

则模型 A 可以变为 A_0 ：

目标函数：

$$\text{Min } E\{TC\} = \sum_{t=1}^{T-l} [vQ_t + kX_t + h(E\{I_{t+l}\} - \Delta d(t+l))] \quad (29)$$

约束条件：

$$E\{I_{t+l}\} - \Delta d(t+l) = I_0 + \sum_{i=1}^t Q_i - F_{D(t+l)}^{-1}(a) \quad (30)$$

$$\sum_{i=1}^t Q_i \geq F_{D(t+l+1)}^{-1}(a) - I_0 \quad (31)$$

$$Q_t \leq N * X_t \quad (32)$$

$$Q_t \geq 0 \quad (33)$$

$$X \in \{0,1\} \quad (34)$$

其中 N 为一个无限大的正整数。

模型 A_0 与模型 A 相比，总成本多了一项确定值 $\sum_{i=1}^{T-l} -\Delta d(t+l)$ ，因此不影响关于订货点和订货量的决策结果。此外模型 A_0 与确定性需求下的库存决策模型有很大的相似性，约束条件中的项存在对应关系。确定性需求下的库存决策模型如下：

目标函数：

$$\text{Min } E\{TC\} = \sum_{t=1}^{T-l} (vQ_t + kX_t + hI_{t+l}) \quad (35)$$

约束条件：

$$I_{t+l} = I_0 + \sum_{i=1}^t Q_i - \sum_{i=1}^{t+l} d_i \quad (36)$$

$$Q_t \leq N * X_t \quad (37)$$

$$Q_t \geq 0 \quad (38)$$

$$I_{t+l} \geq d_{t+l+1} \quad (39)$$

$$X \in \{0,1\} \quad (40)$$

其中 N 为一个无限大的正整数。

确定性需求下的库存决策模型中的 I_{t+l} 的相当于静态库存决策模型 A_0 中的 $E\{I_{t+l}\} - \Delta d(t+l)$ ；确定性需求下的库存决策模型中的 $\sum_{i=1}^{t+l} d_i$ 相当于静态库存决策模型 A_0 中的 $F_{D(t+l)}^{-1}(a)$ ；确定性需求下的库存决策模型中的 d_{t+l+1} 相当于静态库存决策模型 A_0 中的 $F_{D(t+l+1)}^{-1}(a) - F_{D(t+l)}^{-1}(a)$ 。由于两种模型的高度相似性，可以采用求解确定性需求下的库存

决策模型的启发式算法来求解模型 A_0 ，如整数线性规划。

4.2 动态调整库存决策模型

动态调整库存决策模型的决策过程可以分为两阶段。第一阶段是确定订货点，该决策过程与静态库存决策模型相似。第二阶段根据已经实现的需求调整后期决策的订货量，并保持第一阶段决策的订货时间点不变。

4.2.1 模型变量补充说明

动态调整库存策略先是在计划期开始前确定订货点，然后根据计划期内需求的实现状况调整每次订货点的计划订货量。动态调整库存模型要补充以下变量：

表 4.2 模型变量补充

变量符号	变量说明
n	第 1 周到第 $T-1$ 周的订货次数
T_i	订货点, T_i 不超过 $T-1$
B	订货点集合, $B=\{T_1, T_2, T_3, \dots, T_m\}$
Q_{T_i}	T_i 点的订货量
M_{T_i}	根据 $[T_{i-1}, T_i)$ 内实现的需求调整的订货量

4.2.2 模型建立

第一阶段中的订货点可以通过静态库存决策模型求出。接下来根据订货点和订货次数建立动态调整库存策略第二阶段模型。由于订货量要根据已满足的需求进行调整，因此需要先确定调整量。已满足需求与调整量的关系如下：

$$\begin{aligned}
 Q_{T_1} &= M_{T_1}, \\
 Q_{T_2} &= M_{T_2} + d_{T_1} + d_{T_1+1} + \dots + d_{T_2-1}, \\
 &\dots \\
 Q_{T_n} &= M_{T_n} + d_{T_{n-1}} + d_{T_{n-1}+1} + \dots + d_{T_n-1}, \\
 X_t &= 0, \text{ 当 } t \text{ 不属于集合 } B
 \end{aligned}
 \tag{41}$$

对上式进行合并可得：

$$Q_{T_i} = M_{T_i} + \sum_{j=T_{i-1}}^{T_i-1} d_j
 \tag{42}$$

对每次的订货量进行求和可得：

$$\sum_{j=1}^i Q_{T_j} = \sum_{j=1}^i M_{T_j} + \sum_{k=1}^{T_i-1} d_k
 \tag{43}$$

因为对于 $\forall T_i \in B$ ，订货量 Q_t 不能为负，由公式 (42) 可得：

$$M_{T_i} \geq - \sum_{j=T_{i-1}}^{T_i-1} d_j \quad (44)$$

当 $\forall T_i \in B$ 时, $X_t = 0$ 。那么第 $T_{i+1} + l - 1$ 周的期末库存水平为:

$$I_{T_{i+1}+l-1} = I_0 + \sum_{j=1}^i Q_{T_j} - \sum_{k=1}^{T_{i+1}+l-1} d_k \quad (45)$$

把公式(43)带入后可得:

$$I_{T_{i+1}+l-1} = I_0 + \sum_{j=1}^i M_{T_j} - \sum_{k=T_i}^{T_{i+1}+l-1} d_k \quad (46)$$

则服务水平约束条件可以写成:

$$P \{ I_0 + \sum_{j=1}^i M_{T_j} - \sum_{k=T_i}^{T_{i+1}+l-1} d_k \geq d_{T_{i+1}+l} \} \geq a \quad (47)$$

最终通过累积分布函数的性质得到服务水平与调整订货量的约束关系:

$$\sum_{j=1}^i M_{T_j} \geq F_{d_{T_i}+d_{T_{i+1}}+\dots+d_{T_{i+1}+l}}^{-1}(a) - I_0, i \in \{1, 2, \dots, n\} \quad (48)$$

其中 $F_{d_{T_i}+d_{T_{i+1}}+\dots+d_{T_{i+1}+l}}^{-1}(a)$ 为服务质量 a 下第 T_i 周到第 $T_{i+1} + l$ 周的广义需求。由公式(28)可得:

$$F_{d_{T_i}+d_{T_{i+1}}+\dots+d_{T_{i+1}+l}}^{-1}(a) = \sum_{j=T_i}^{T_{i+1}+l} E\{d_j\} + Z * \left(\sum_{j=T_i}^{T_{i+1}+l} S_{d_j}^2 \right)^{\frac{1}{2}} \quad (49)$$

在确定的订货点情况下, 由于计划订购量 Q_{T_i} 和调整量 M_{T_i} 存在一定的数量关系, 因此动态调整库存决策模型第二阶段考虑以 M_{T_i} 为决策变量进行建模, 模型如下:

目标函数:

$$\begin{aligned} \text{Min } E\{TC\} = & nk + v \left(\sum_{i=1}^n M_{T_i} + \sum_{i=1}^{T_n-1} E\{d_i\} \right) + h[(T-l)I_0 + \sum_{i=1}^n (T-l-T_i+1)M_{T_i} \\ & + \sum_{i=1}^n (T-l-T_i+1) \sum_{j=T_{i-1}}^{T_i-1} d_j \\ & - \sum_{i=1}^n \sum_{j=T_i}^{T_{i+1}-1} (T-l-j+1)E\{d_{j+l}\}] \end{aligned} \quad (50)$$

约束条件:

$$M_{T_i} \geq - \sum_{j=T_{i-1}}^{T_i-1} d_j, i \in \{1, 2, \dots, n\} \quad (51)$$

$$\sum_{j=1}^i M_{T_j} \geq F_{d_{T_i}+d_{T_{i+1}}+\dots+d_{T_{i+1}+l}}^{-1}(a) - I_0, i \in \{1, 2, \dots, n\} \quad (52)$$

由目标函数可得，总成本随着 M_{T_i} 的增加而增加，且为线性相关。所以当 M_{T_i} 取最小值时，有最小总成本。此时 M_{T_i} 要大于以上两个约束的最大值，即：

$$\begin{aligned}
 M_{T_1} &= \max \left(- \sum_{j=1}^{T_1-1} d_j, F_{d_{T_1+d_{T_1+1}+\dots+d_{T_2+l}}}(a) - I_0 \right) \\
 M_{T_2} &= \max \left(- \sum_{j=T_1}^{T_2-1} d_j, F_{d_{T_2+d_{T_2+1}+\dots+d_{T_3+l}}}(a) - I_0 - M_{T_1} \right) \\
 &\quad \dots\dots \\
 M_{T_n} &= \max \left(- \sum_{j=T_{n-1}}^{T_n-1} d_j, F_{d_{T_n+d_{T_n+1}+\dots+d_T}}(a) - I_0 - \sum_{i=1}^{n-1} M_{T_i} \right)
 \end{aligned} \tag{53}$$

再根据公式（42），可以求出每个订货点的最终订货量 Q_{T_i} 。

4.3 实验结果分析

本节选取商品三作为实验对象，基于极端随机树回归的需求预测结果，求出商品三在静态库存决策模型和动态调整库存决策模型下的最佳订货点和订货量，并计算 1-8 周内的总成本和平均库存水平。最后对服务水平 a 进行灵敏度分析，得出服务水平对两种库存策略总成本的影响，并对比分析两种库存决策模型所产生的库存成本的高低。

4.3.1 静态库存策略数值实验

在静态库存决策模型中，商品三每周的需求由极端随机树回归预测得出且假设服从正态分布，标准差由 2012、2013 和 2014 三年前 1-8 周的历史销售数据得出，如表 4.3 所示。此外为完成数值实验，本节对以下参数进行赋值： $I_0=50$ ， $k=5$ ， $v=2$ ， $h=0.2$ ， $a=95\%$ ， $l=2$ 。

表 4.3 计划期内商品三需求预测值和标准差

T	1	2	3	4	5	6	7	8
$E\{d_t\}$	25	18	12	14	15	19	12	8
S_{d_t}	3	7	3	1	3	2	5	7

因为 $F_{D(t)}^{-1}(a)$ 是求解静态库存决策模型最优订货点和订货量的关键，所以根据公式(28)先求出 $F_{D(t)}^{-1}(a)$ 的值，结果如表 4.4 所示。

表 4.4 商品三的 $F_{D(t)}^{-1}(a)$ 值

T	1	2	3	4	5	6	7	8
$F_{D(t)}^{-1}(a)$	32	61	74	88	104	124	139	152

根据商品三的各时期的 $F_{D(t)}^{-1}(a)$ 值，采用 Matlab 软件的 intlinprog 函数编写整数线性规划算法进行求解。

输入： $I_0=50$ ， $k=5$ ， $v=2$ ， $h=0.2$ ， $a=95\%$ ， $l=2$

- 1: 分配整数变量地址给订货决策变量
 - 2: 用订货量和每周预测需求替代目标函数的库存变量
 - 3: 展开广义需求分布与订货量的关系并作为整数线性规划的不等式约束
 - 4: 把订货量的下界定义为 0, 上界定义为 1000 (足够大的正整数); 订货决策变量下界定义为 0, 上界定义为 1
 - 5: 调用 `intlinprog` 函数求解
- 输出: 最优订货点、最优订货量以及最小总成本

通过以上方式进行求解得到商品三的订货点集合 $B = \{1,3,4\}$, 即在第 1 周、第 3 周和第 4 周初订货。具体的订购量、库存量及库存成本如表 4.5 所示。

表 4.5 商品三静态库存策略及库存成本

T	1	2	3	4	5	6	7	8
第 t 周订货量	48		20	26				
第 t 周初库存量	50	25	55	43	49	60	41	29
第 t 周末库存量	25	7	43	29	34	41	29	21
周平均库存量				29				
次平均订货量				31				
总成本				242				

由上表可以看出商品三 8 周内共订货 3 次, 分别为第 1 周、第 3 周和第 4 周。其中第 1 周的订货量为 48, 第 3 周的订货量为 20, 第 4 周的订货量为 26。8 周内的周平均库存量为 29, 次平均订货量为 31, 总成本为 242 (由于商品的不可分割性, 故在计算时采用四舍五入的方式对小数进行取整)。

4.3.2 动态调整库存策略数值实验及对比分析

动态调整库存决策模型的第一阶段求解方法同样采用整数线性规划, 且每周的需求预测值和 $F_{D(t)}^{-1}(a)$ 值与静态库存决策模型相同, 故订货点的决策不发生变化。而第二阶段是对第一阶段确定的订货点的对应订货量进行决策, 其中订货量根据已实现的需求进行调整, 所以两个模型的订货数量会有所不同。动态调整库存决策模型的第二阶段求解方式在上一节的最后已经给出。商品三的动态调整库存决策模型的具体的订购量、库存量及库存成本如表 4.6 所示。

表 4.6 商品三动态调整库存策略及库存成本

T	1	2	3	4	5	6	7	8
第 t 周订货量	48		7	25				
第 t 周初库存量	50	25	55	43	36	46	27	15
第 t 周末库存量	25	7	43	29	21	27	15	7
周平均库存量				22				
次平均订货量				27				
总成本				203				

可以看出动态调整库存策略与静态库存策略的订货点相同, 分别为第 1 周、第 3 周和

第 4 周。其中第 1 周的订货量为 48，第 3 周的订货量为 7，第 4 周的订货量为 25。8 周内的周平均库存量为 22，次平均订货量为 27，总成本为 203（由于商品的不可分割性，故在计算时采用四舍五入的方式对小数进行取整）。

根据两种库存决策模型的计算结果可得，对于商品三，动态调整库存策略模型下的周平均库存量比静态库存策略模型下的周平均库存量低 9，次平均订货量低 4，总成本低 52。这表明动态调整库存决策模型相较于静态库存决策模型更能够降低连锁零售库存成本。

4.3.3 服务水平灵敏度分析

服务水平 α 影响着模型中的 $F_{D(t)}^{-1}(\alpha)$ 值，进而影响订货点和订货量。以上数值实验的 α 为 95%，具有一定的片面性，因此接下来尝试取多组不同的 α 值来比较静态库存策略模型和动态调整库存策略模型的总成本。在其他参数保持不变的情况下， α 分别取 85%、90%、95%、99%（其对应的正态分布 Z 值分别为 1.04、1.29、1.65、2.33），可以得到服务水平 α 对两种模型总成本的影响。如图 4.2 所示。

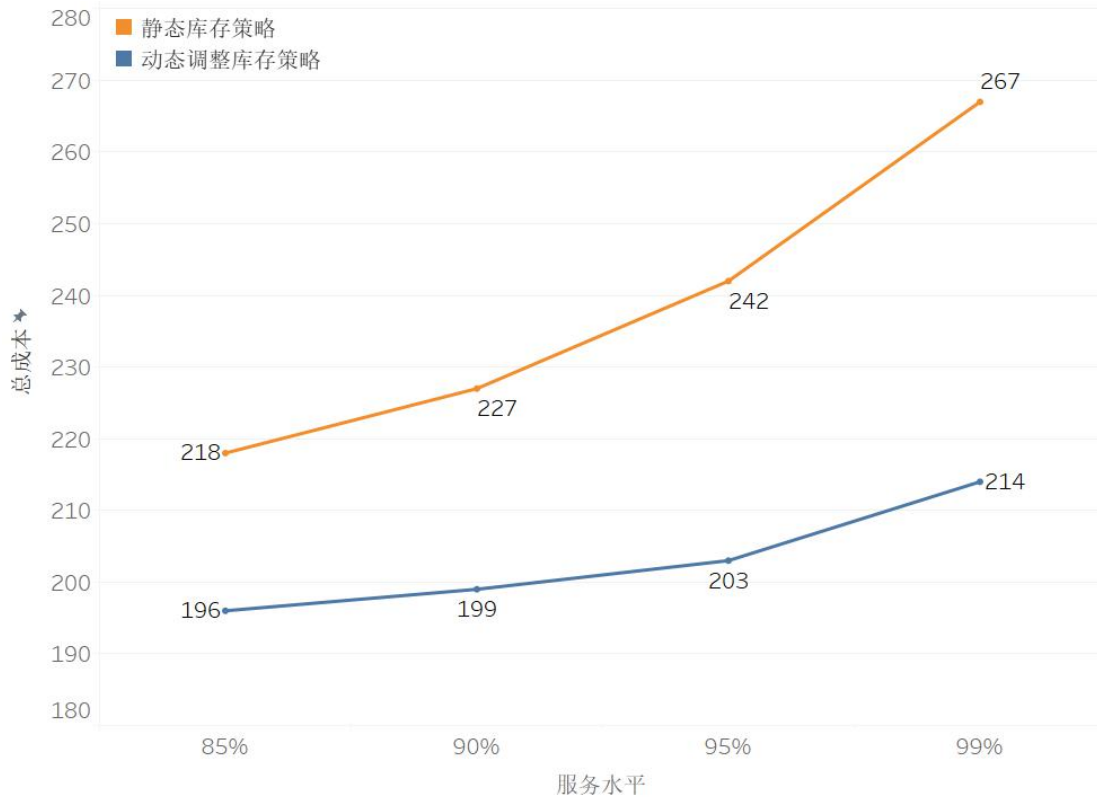


图 4.1 商品三两种库存策略下服务水平对总成本的影响

根据图 4.2 可以看出，两种库存策略下的总成本都随着服务质量的提高而增加，当服务水平越接近 99%，由于正态分布的函数的“长尾”特征， Z 值增加速度越快，对应的两种库存策略产生的总成本增幅越大，并且动态调整库存策略的总成本总比静态库存策略的总成本低。这说明无论服务水平如何改变，动态调整库存策略相较于静态库存策略在指导连锁零售店进行库存决策时更具优势，能够更大程度地降低连锁零售店库存成本。

第五章 总结与展望

5.1 总结

连锁零售店作为分销商的一种形式,具有商品种类固定、顾客流量大但忠诚度不高、需求不稳定等特点。近年来由于线上销售的兴起和突发新冠疫情的影响,使得连锁零售店的生存环境进一步恶化,行业竞争压力进一步提升,为了在复杂市场环境中站稳脚步,快速地响应市场需求和局势的变化,就要提高供应链效率进而彰显竞争优势。在供应链的各个环节中,库存环节至关重要,不合理的库存决策会提升连锁零售店运营成本,造成缺货或积压,使得资金周转困难,利润遭受损失。连锁零售店在库存决策之前,要了解各商品的需求状况,因此就要对需求进行预测。准确的需求预测能够为连锁零售店采购计划和库存管理提供依据,从而有效地降低连锁零售店库存成本。在大数据时代下,无论是线上还是线下的企业都掌握着大量的用户消费数据,且数据类型也越来越多样化。采用机器学习方法,充分挖掘、分析包含历史销售数据在内的各种多元化的数据,预测未来不同商品的需求趋势,帮助连锁零售店做出正确的采购计划和库存决策。本文通过机器学习方法预测连锁零售店的未来商品需求,并根据预测结果建立合理的库存决策模型,选择最适合连锁零售店的库存策略。本文主要的研究工作如下:

(1) 利用机器学习方法对连锁零售店的各商品未来需求进行预测。针对连锁零售店包含历史销售数据在内的各种多元化的数据,本文先后进行了数据清洗、特征转换、特征筛选,数据标准化以及训练集、验证集和测试集划分等预处理工作,然后采用随机森林回归(RFR)、K近邻回归(KNR)、支持向量回归(SVR)以及极端随机树回归(ETR)四种机器学习算法对验证集的商品需求进行预测,采用五折交叉验证的方法来评估每一种模型的泛化能力和预测精度,选出极端随机树回归(ETR)作为最优的预测模型。然后根据极端随机树回归(ETR)对测试集的商品需求进行预测,并把预测结果作为指导库存决策的输入变量。

(2) 基于极端随机树回归的需求预测结果,构建考虑订货提前期和服务质量的静态库存决策模型和动态调整库存决策模型,并根据模型的特征采用整数线性规划算法进行求解。在数值实验中,假设需求服从正态分布的基础上,设置模型参数,计算两种库存策略下的订货点和订货量,并对比分析两种模型对应的库存量和总成本。最后对服务水平 α 进行灵敏度分析,发现总成本会随着服务水平的上升而增加,且动态调整库存决策模型的总成本总低于静态库存决策模型的总成本,从而进一步验证动态调整库存决策模型对于指导连锁零售店进行库存决策的优势。

5.2 展望

由于自身专业水平的限制以及各种因素的影响,本文还有很多不足之处,望在后续研

究中有进一步的考虑和改进：

(1) 特征选择在机器学习预处理过程中特别重要，它决定着数据集的质量和值，进而影响着预测模型的精度和泛化能力。但本文中由于特征数量较少，过拟合现象不严重，因此在特征选择的过程中没有采用如以随机森林的方式对特征重要性进行排序，这样多少会对模型的预测精度和泛化能力产生负面的影响，因此在之后的研究中会尝试多种的特征选择方法，通过对比选出最适合数据集的特征集合。

(2) 本文中只选取了四种机器学习算法进行需求预测，这对依据大数据的运营决策是不够的。后续的研究中需要采用更多的机器学习算法进行预测，对比不同模型的预测精度和泛化能力，从而选出更加合适的模型对连锁零售店的库存决策进行指导。此外可以考虑集成学习和深度学习算法进一步的提高模型的性能。

(3) 本文构建的库存决策模型假设订货提前期 l 为固定值。但由于多种因素的不确定性，订货提前期会随着商品的变化而不同，在后续的研究中会考虑订货提前期在不同时期的变化，根据订货提前期的随机分布特征来建立库存决策模型。

参考文献

- [1] Dubelaar C, Chow G, Larson P D. Relationships between inventory, sales and service in a retail chain store operation[J]. International Journal of Physical Distribution & Logistics Management, 2001, 31(2):96-96.
- [2] Xiao, T., Qi, X. Price competition, cost and demand disruptions and coordination of a supply chain with one manufacturer and two competing retailers[J]. Omega, 2008, 36 (5),741-753.
- [3] Zhongjiao, Ma, Jialin, et al. Energy consumption prediction of air-conditioning systems in buildings by selecting similar days based on combined weights[J]. Energy and Buildings, 2017, 151(6):157-166.
- [4] Haitao. Research on prediction of traffic flow based on dynamic fuzzy neural networks.[J]. Neuraluting & Applications, 2016, 27(7):1969-1980.
- [5] 李顺, 李君, 吴鑫, 梅碧舟. 基于 GA-XGBoost 的宁波港物流需求预测 [J]. 浙江万里学院学报, 2021, 34(02) :71-77.
- [6] 赵志翔. B 公司防护货品需求预测与库存控制研究[D]. 中国科学院大学(中国科学院大学工程科学学院), 2020.
- [7] 赵军, 王晓. 基于数据挖掘的第三方物流中心库存需求预测模型 [J]. 物流技术, 2014, 33(03) :148-150+170.
- [8] Jakob Huber, Alexander Gossmann, Heiner Stuckenschmidt. Cluster-based hierarchical demand forecasting for perishable goods[J]. Expert Systems With Applications,2017, 76(Jun.):140-151.
- [9] Shi Y, Wang T, Alwan L C . Analytics for Cross-Border E-Commerce: Inventory Risk Management of an Online Fashion Retailer[J]. Decision Sciences, 2020, 51(6):1347-1376.
- [10] 罗兵, 于会强. 存货影响销售率且销售价格可变的 EOQ 模型 [J]. 重庆大学学报(自然科学版), 2002, 11(26) :100-103.
- [11] 孙士雅, 罗兵, 于会强. 部分短缺量拖后的非瞬时补货的 EOQ 模型 [J]. 重庆大学学报(自然科学版), 2002, 01(33) :134-137.
- [12] 徐贤浩, 廖丽平, 任英. BASS 预测模型与库存控制集成研究 [J]. 工业工程与管理, 2010, 15(03) :1-6.
- [13] 李卓群, 梁美婷. 不确定需求影响下动态供应链库存策略选择 [J]. 工业工程与管理, 2018, 23(04) :23-29.
- [14] Huang H, Ke H . Pricing decision problem for substitutable products based on uncertainty theory[J]. Journal of Intelligent Manufacturing, 2017, 28(3):503-514.
- [15] Zhongfeng Qin, Samarjit Kar. Single-period inventory problem under uncertain environment[J]. Applied Mathematics and Computation,2013,219(18):9630-9638
- [16] 罗晓萌, 李建斌, 胡鹏. 基于时间序列预测的电子商务库存优化策略 [J]. 系统工程, 2014, 32(06) :91-98.
- [17] 李长春. 大数据背景下的商品需求预测与分仓规划 [J]. 数学的实践与认识, 2017, 47(07) :70-79.
- [18] Rego, José Roberto do, Mesquita, Marco Aurélio de. Demand forecasting and inventory control: A simulation study on automotive spare parts[J]. International Journal of Production Economics, 2015, 161(mar.):1-16.
- [19] Ke Rong. Research on Multi-Stage Inventory Model by Markov Decision Process[J]. Physics Procedia,2012,33(none):1074-1077.

致 谢

时光清浅，大学四年惊鸿一瞥，但却终生难忘。回首过去，从接到写着“自强不息，独树一帜”的校训的录取通知书时的激动，到初见西北黄土高原的荒凉苍劲的忐忑，再到如今，足踏萃英，手揽黄河。四年时间，我从迷茫懵懂、不谙世事的纯朴少年变成了弘毅致远、坚定自信的兰大青年，兰大“自强不息，独树一帜”的校训深深的烙在心头，铭在骨里。四年里，我有幸结识了一位位博学广智、德高望重的老师，他们带领着我向知识的海洋不断探索，开阔了我的视野，深邃了我的思考，他们更以身体力行向我传达着兰大人的坚守和奋斗，向我展示着兰大人的风骨和情怀。幸甚之至，以后的三年我还能留在兰大继续攻读研究生。此时的我不在如当初的迷茫困惑，而是一腔热血为学术和科研。我抱着吾校虽瘦，必肥华夏的初心，抱着传承师道，扎根西北的信念，砥砺前行，勇攀高峰，做西北高原上一只坚韧拼搏的鹰。

在大学四年的，我有过“一点浩然气，千里快哉风”的自信和快意，也有过“今朝有酒今朝醉，明日愁来明日愁”的失落和彷徨。但是，我的导师洪兆富教授总是给予我最大程度的帮助和坚定不移的信任，帮助我从只会啃书本死知识的状态一步步走上了专研实践的学术之路，使我的未来方向不在迷茫。在毕业论文即将结尾的时候，我必须向洪兆富老师致以最由衷的感谢。

另外，再次，我要感谢我的几位同学，是他们在日常的学习以及生活中给了我精神上的支持，特别是游鸿宾、王玉宝、陈黎明、欧贻杰、张东阳和易嘉豪，四年同窗，厚爱无言，笺短情深，山海可平。

最后，谢谢论文评阅老师们的辛苦工作。衷心感谢我的家人、朋友、以及同学们，正是在他们的鼓励和支持下我才得以顺利完成此论文。

毕业论文(设计)成绩表

导师评语

论文之中机器的学习算法研究的透彻性
看和的识别与序序决策问题, 选题具有一定
的新颖性, 论证的严谨, 写作规范, 是一篇
优秀的本科毕业论文。

建议成绩

优秀

指导教师(签字)

沈明富

答辩委员会意见

答辩委员会负责人(签字)

成绩

学院(盖章)

年 月 日