**承 诺 书**

我们仔细阅读了第十二届华中地区大学生数学建模邀请赛的竞赛细则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的, 如果引用别人的成果或其他公开的资料

（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们的参赛报名号为：89

参赛队员 (签名)

队员 1：许可

队员 2：李晨鸣

队员 3：符玉梵

湖北省工业与应用数学学会第十二届华中地区大学生数学建模邀请赛组委会

**编号专用页**

选择的题号： B

参赛的编号： 89

（以下内容参赛队伍不需要填写）

竞赛评阅编号：

基于长短期记忆网络（LSTM）的库存补单及销量预测模型

**【摘要】**

本文主要解决了商家商品补单以及销量预测的问题，并提出一个合理的，实用性强的补单策略。

针对问题一，本文应用长短期记忆网络（long short term memory简称LSTM）对商品销量进行预测。首先对所有商品的所有日期的销售数据使用滑动窗口进行取样，然后通过对销量分布的分析，划分销量范围的分类，并构建标签集。使用3层LSTM加两层全连接层的深度学习神经网络结构模型对数据进行训练，最后在测试集上进行预测。预测结果表明，模型不仅具有较高的准确性，而且对商品销量的偶然大幅度变化具有一定的泛化能力。解决了问题一中预测未来五天销售量的问题。

针对问题二，我们用已给出的数据，统计所有商品的上新量和延期比的分布情况，并画出分布饼状图来直观显示。同时，我们根据已知数据，求出上新量和延期比的分布范围以及置信空间。根据上述数据，我们分析得出上新量大都在5000以下，延期比大都在30%以下。

针对问题三，我们用已给出的数据，统计各个季度的商品的上新量和延期比的分布情况，并画出分布饼状图来直观显示。同时，我们根据已知数据，求出各个季度的商品的上新量和延期比的分布范围以及置信空间。根据上述数据，我们分析得到春季商品的上新量和延期比都要高于其他几个季度，夏季和冬季的相当，秋季的最低。故可以在春季多补单，夏季和冬季正常补单，秋季少补单。

针对问题四，我们是根据问题1和问题3的分析得到的结果来解决的。首先，我们把补单分为3部分，上新前的补单，上新当日的补单和上新一段时间后正常的补单。其中，上新前的补单我们是由问题三得到的各个季度的上新量和延期比推出的；上新当日的补单我们是通过当日的销售量和之后10天的销售量的关系以及问题三得到的延期比综合推测出的；上新一段时间后正常的补单我们是根据问题一建立的模型和问题三得到的延期比综合推测出的。这三步结合构成了我们的补单策略。

我们所建的模型具有良好的实用性，我们提出的补单策略可以作为各大商店的补单策略。

目录

[1 问题重述](#_Toc31488_WPSOffice_Level1) [1](#_Toc31488_WPSOffice_Level1)

[1.1 问题背景](#_Toc1410_WPSOffice_Level2) [1](#_Toc1410_WPSOffice_Level2)

[1.2 问题提出](#_Toc30703_WPSOffice_Level2) [1](#_Toc30703_WPSOffice_Level2)

[2 模型假设](#_Toc1410_WPSOffice_Level1) [2](#_Toc1410_WPSOffice_Level1)

[3 符号说明](#_Toc30703_WPSOffice_Level1) [2](#_Toc30703_WPSOffice_Level1)

[4 模型建立与求解](#_Toc1963_WPSOffice_Level1) [3](#_Toc1963_WPSOffice_Level1)

[4.1问题1的分析与建模](#_Toc1963_WPSOffice_Level2) [3](#_Toc1963_WPSOffice_Level2)

[4.1.1长短时记忆神经网络LSTM](#_Toc1410_WPSOffice_Level3) [3](#_Toc1410_WPSOffice_Level3)

[4.1.2数据的预处理](#_Toc30703_WPSOffice_Level3) [6](#_Toc30703_WPSOffice_Level3)

[4.1.3模型的建立](#_Toc1963_WPSOffice_Level3) [7](#_Toc1963_WPSOffice_Level3)

[4.1.4 结果分析](#_Toc9821_WPSOffice_Level3) [8](#_Toc9821_WPSOffice_Level3)

[4.1.5 不足与改进](#_Toc6090_WPSOffice_Level3) [9](#_Toc6090_WPSOffice_Level3)

[4.2问题2的模型建立与求解](#_Toc9821_WPSOffice_Level2) [9](#_Toc9821_WPSOffice_Level2)

[4.2.1 问题的分析](#_Toc30057_WPSOffice_Level3) [9](#_Toc30057_WPSOffice_Level3)

[4.2.2 置信区间的概念](#_Toc14222_WPSOffice_Level3) [9](#_Toc14222_WPSOffice_Level3)

[4.2.3 模型的建立](#_Toc19973_WPSOffice_Level3) [10](#_Toc19973_WPSOffice_Level3)

[4.2.4 模型的求解](#_Toc27161_WPSOffice_Level3) [10](#_Toc27161_WPSOffice_Level3)

[4.2.5分布情况的分析](#_Toc21732_WPSOffice_Level3) [12](#_Toc21732_WPSOffice_Level3)

[4.3问题3的模型建立与求解](#_Toc6090_WPSOffice_Level2) [12](#_Toc6090_WPSOffice_Level2)

[4.3.1 问题的分析](#_Toc32545_WPSOffice_Level3) [12](#_Toc32545_WPSOffice_Level3)

[4.3.2 模型的求解](#_Toc2432_WPSOffice_Level3) [12](#_Toc2432_WPSOffice_Level3)

[4.3.3 分布情况的分析](#_Toc13388_WPSOffice_Level3) [17](#_Toc13388_WPSOffice_Level3)

[4.4问题4的模型建立与求解](#_Toc30057_WPSOffice_Level2) [18](#_Toc30057_WPSOffice_Level2)

[4.4.1 问题的分析](#_Toc26270_WPSOffice_Level3) [18](#_Toc26270_WPSOffice_Level3)

[4.4.2 模型的建立与求解](#_Toc17980_WPSOffice_Level3) [18](#_Toc17980_WPSOffice_Level3)

[5 参考文献](#_Toc9821_WPSOffice_Level1) [19](#_Toc9821_WPSOffice_Level1)

# 1 问题重述

## 1.1 问题背景

商家原来已有订单，在此基础上，买家又增加新的需求量，追加订单，就是补单。有时补单也特指因品质异常不能满足客户要求而导致的补单。但是补单对于许多电商来说也是一个十分烦恼的问题，其问题主要表现在以下两个方面。

补单的第一个问题是对现金流的占用。在每年双11时的货物成本对于许多电商来说那就是几千万的货值，一但补单没预测好，几千万的资金占用，对于大多数的电商而言都是不小的压力，所以合理的补单预测，尽量减少流动资金的占用对于体量大的公司尤为重要。

补单的第二大问题就是库存问题。一个商品能卖多少实际在你的店铺里面是有个上限的，根据单个商品的价格，点击，流量，转化率，活动表现等，最终体现在销售数据上，而过于乐观的补单将造成大量的库存，库存的积压带来的不仅是仓储成本的增加，更为严重的是要考虑清仓问题，清库存的方式无非是加大营销力度，以更低的成本清理货物，这时候不仅资金的流转变慢，而且低价的清仓对于品牌而已也会拉低品牌溢价，而过于悲观的补单则会导致货不够卖而造成资源的浪费和利润的流失，这又是企业不愿意看到的，所以科学合理的补单预测尤为重要。

## 1.2 问题提出

（1）请根据附件一中的销量数据建立一个销量预测的数学模型，要求至少能够比较精确地预测未来五天的货物销量大小；

（2）请根据附件二分析货物的上新量和延期比的分布情况，并给出分布范围及置信区间；

（3）请根据附件二中的数据进一步分析各个季度货物的上新量和延期比的分布情况，并给出分布范围及置信区间；

（4）请根据上述分析结果，制定合理的补单策略，写出具体操作流程。

# 2 模型假设

·假设所有商品从上新日开始后的销量变化符合相同的数学模型；

·假设刚上新的商品，在上新日前几天不在预测范围内，但上新前几日的销量数据可作为历史销量数据用来预测后面的销量；

·假设日期销量的值缺失意味着在该天未售出该商品，即销售量为0；

·假设题目2和题目3中的上新量指总实收量。

# 3 符号说明

|  |  |
| --- | --- |
| 符号 | 含义 |
|  | x的权值矩阵（x可以为i、f、a、o） |
|  | x的偏移量（x可以为i、f、a、o） |
|  | Sigmoid函数 |
| tanh | tanh函数 |
|  | t时间步i的值 |
|  | t时间步f的值 |
|  | t时间步o的值 |
|  | t时间步a的值 |
|  | t时间步的细胞状态 |
|  | t时间步的隐藏状态 |
|  | t时间步的预测输出值 |
|  | 样本的平均值。 |
| SE | 平均值的标准误差 |
|  | 样本的标准差 |
| n | 样本的总数 |
|  | 样本总体 |
| X1,X2... | 取自样本总体的样本 |
|  | 总体的待估参数。 |

# 4 模型建立与求解

## 4.1问题1的分析与建模

结合实际情况与问题一分析，商家预测未来的销量，肯定是通过之前的销量数据来预测的，所以在问题一中，我们将以前十天的销量数据为历史数据，预测未来5天的销量。我们将该问题定义为时间序列预测问题，由于时间序列分析法对于短，近期预测比较显著，但延伸到更远的将来，会出现很大的局限性。因此，我们建立模型跟据前十天的销量历史数据预测未来五天的销量的总和。近两年，深度学习在时间序列预测的领域有突破性进展，在此，我们将使用长短时记忆神经网络（long-short term memory简称LSTM）建立销量预测模型

### 4.1.1长短时记忆神经网络LSTM

循环神经网络（Recurrent Neural Network）是一种节点定向连接成环的人工神经网络。这种网络的内部状态可以展示动态时序行为。不同于传统神经网络的是，RNN可以利用它内部的记忆来处理任意时序的输入序列，这让它可以更容易处理时间序列预测问题。但是普通的RNN存在一定的弊端，由于原始RNN模型也是使用BP算法进行参数的训练，梯度消失和梯度爆炸的问题不仅得不到解决，反而在内部的迭代过程中得到了放大。

LSTM的出现解决了上述所说的问题。LSTM的结构与RNN相似，都为链式结构。不同的是，RNN内部的每个记忆单元使用的是单一网络层（见图4-1），而LSTM对RNN的每个记忆单元都增加了四个门控结构见图（4-2）。[4][5][7]

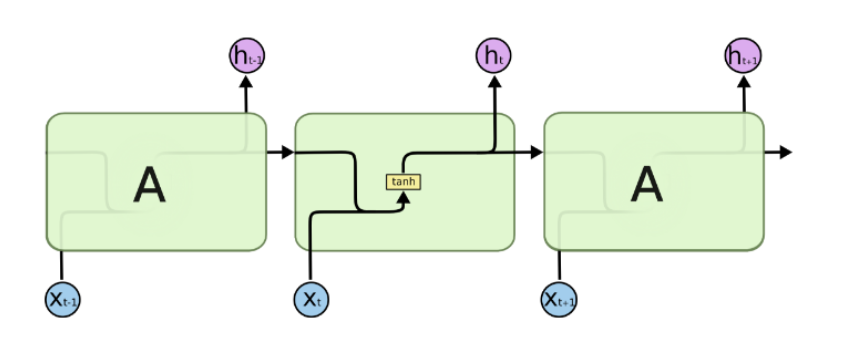


图 4‑1 RNN结构图

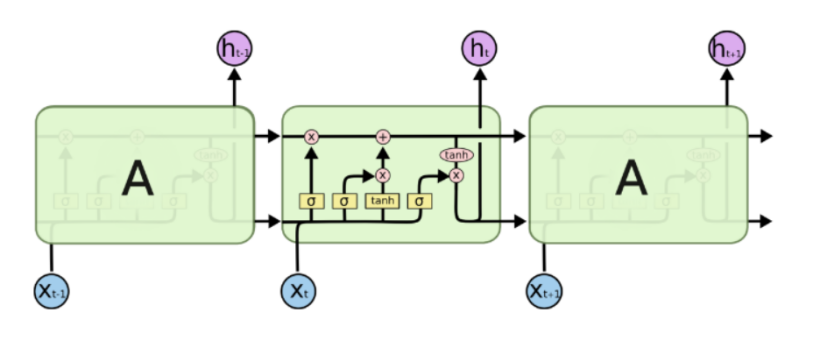


图 4‑2 LSTM结构图

从图4-2中可以看出，在每个序列索引位置t时刻向前传播的除了和RNN一样的隐藏状态，还多了另一个隐藏状态（以下记为细胞状态），如图4-2中的长横线。如图4-3所示：

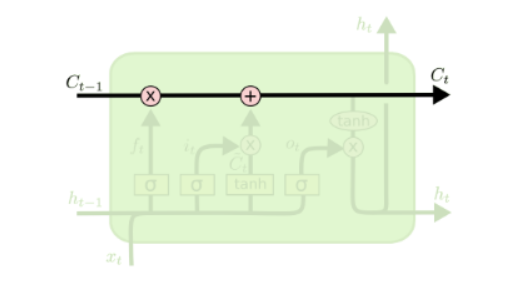


图 4-3 LSTM细胞状态

除了细胞状态之外，LSTM还多了三个门控结构，我们称之为遗忘门，输入门和输出门。

遗忘门（forget gate）是控制记忆单元是否遗忘上一层的隐藏细胞状态。遗忘门子结构如图4-4所示：

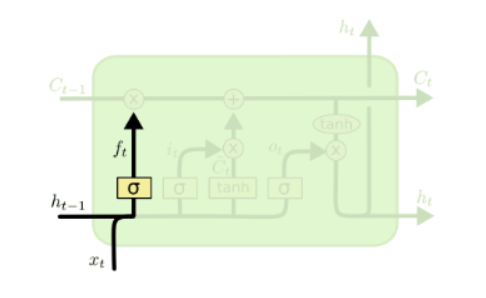


图 4-4 LSTM遗忘门

遗忘门输入之前的隐藏状态，和当前序列数据，通过激活函数sigmoid输出上一层细胞状态的遗忘概率。

输入门（input gate）负责处理当前序列位置的输入，它的结构如图4-5：

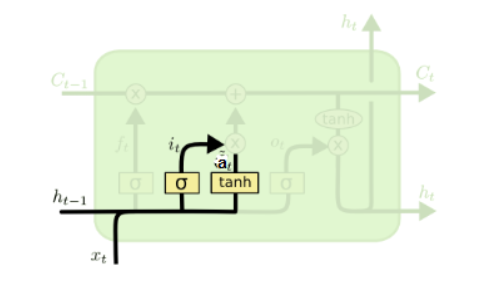


图 4-5 LSTM输入门

输入门输入之前的隐藏状态，和当前序列数据，分别通过激活函数sigmoid和tanh得到和，再相乘，得到结果用于更新细胞状态。

状态更新结构如图4-6：

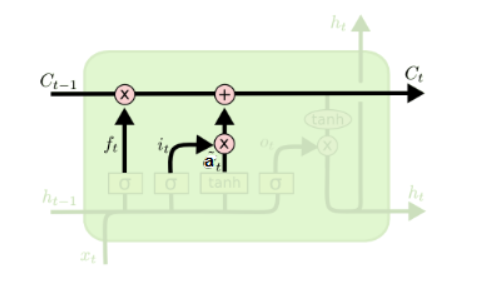


图 4-6 LSTM状态更新

输入之前的细胞状态和遗忘门输出结果以及输入门的输出结果和来更新细胞状态。

输出门结构如图4-7：

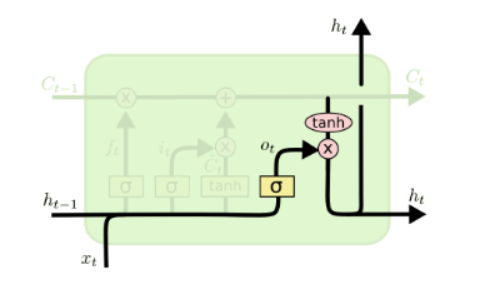


图 4-7 LSTM输出门

输出门用来更新隐藏状态，分为两个部分，第一部分输入之前的隐藏状态和当前序列输入通过激活函数sigmoid得到，第二部分输入状态更新的输出通过激活函数tanh得到tanh()，两部分结果做Hadmard积[6]得到当前隐藏状态。

最后更新当前序列引索预测输出。

LSTM的前向传播算法总结为：[7]

·更新遗忘门输出：

·更新输入门两部分输出：

·更新细胞状态 ：

·更新输出门输出：

·更新当前序列索引预测输出：

LSTM的后向算法通过链式法则可对参数梯度一一求解。

### 4.1.2数据的预处理

在对附件一中商品销售量进行分析的过程中，我们发现很多日期的销量数据缺失如图4-8。

图 4-8 数据样例

结合本题的情景我们对数据进行了如下的预处理：

·对于销量缺省的日期，我们对当天的销售量进行填0处理。

·数据中存在同一商品同一天内有两条记录的情况，这种数据属于因记录错误导致的异常数据，将其去除。

·对于每一个商品，我们选择长度为15的窗口对每一商品的所有日期的销量数据连续滑动100次进行取样，每个窗口内的前10个数据为销量历史数据，每个窗口的最5个数据的和为预测数据的标签。用该方法对每个商品取出100个样本，总计32900个样本。

·因为训练的数据及其不均匀，我们按照销量值在“0、1、2、3、4、5、6、7-8、9-10、11-12、13-17、18-30、31-45、46-60、60-100、101-200、201-300、301-400、401-500、500+”不同区间内进行统计（统计结果如表4-1所示），分别对每个区间内的数据分别处理为“0、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19”。

图 4-9 销量数据统计直方图

·对所有的训练样本做min-max标准化，将数据映射到[0,1]上。

·将所有训练样本分为三个部分，前27000个样本和标签作为训练集，27000-30000样本和标签作为验证集，剩下的样本和标签作为测试集。

### 4.1.3模型的建立

我们选取27000个样本，每个样本的10天的销量作为10个LSTM网络的时间步，用来推测销量在时间序列上的关系。

建立模型结构图如图4-9所示：

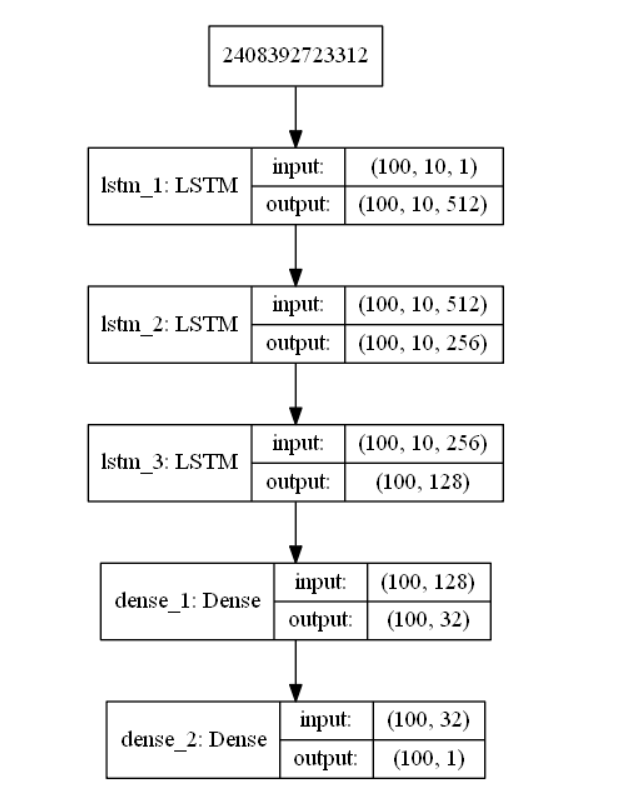


图 4-10 模型结构图

首先用三层LSTM层学习序列的特征。接着模型加入全连接层进一步学习之前学习到的序列的特征，最后保留最有用的特征。

我们使用mse损失函数。优化器使用Adam算法，采用批量梯度下降，每一批训练数据为100个。经过反复调参，最优训练结果如图4-9所示。

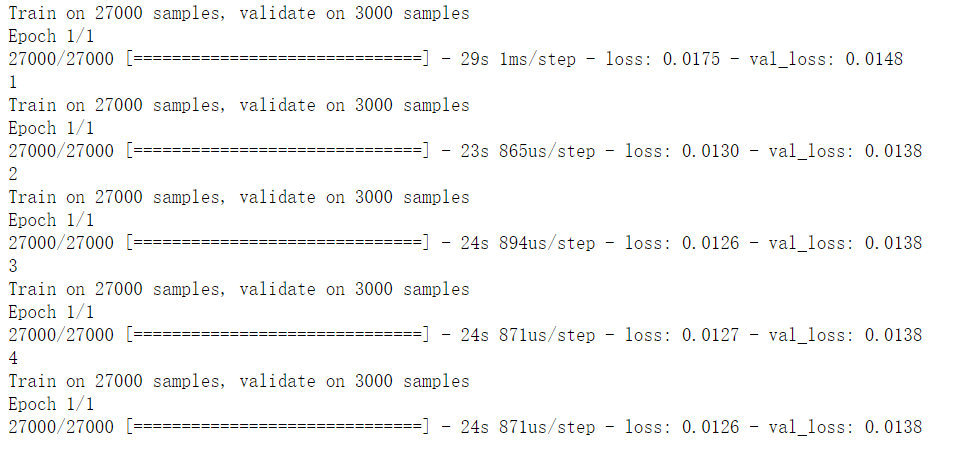


图 4-11 模型训练情况图

在测试集上验证模型的结果，并随机选取了6个商品的共计300个样本画图，如图4-10所示：

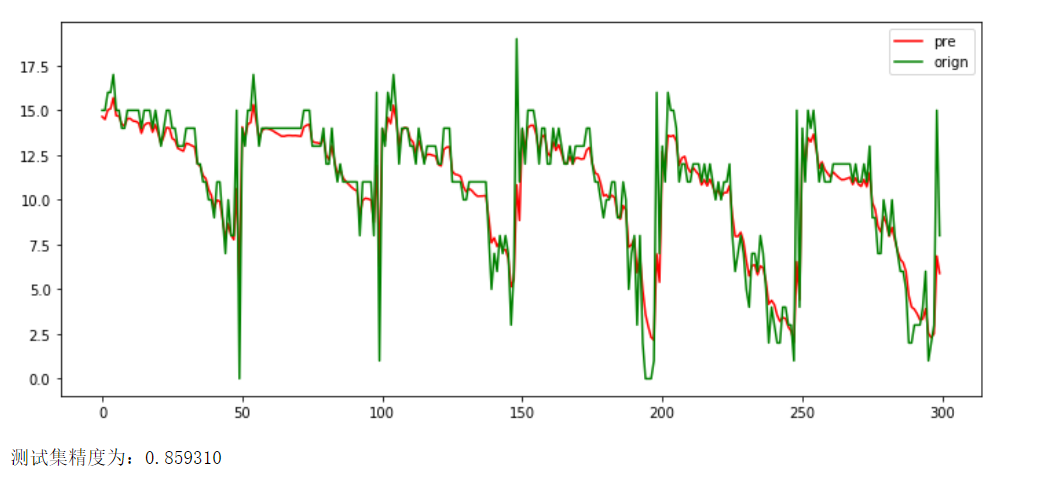


图 4-12 测试情况图

红色线为预测结果，绿色线为原始结果。测试集精度为0.859310

### 4.1.4 结果分析

Loss基本上在第一个epoch就已经收敛了，后三个epoch训练的模型存在略微的过拟合，因此我们用第二个epoch结束后的模型参数用做最终的模型参数，并在测试集上测试。结果发现，对销量高的类如17,18,19的预测结果比较差，考虑原因可能是因为高类样本点较少，模型的学习效果一般。模型还存在的一点问题是，商品销量存在一定的突变情况，即可能从一个正常值突然在第二天骤降或者骤升，模型对这种情况的预测效果并不理想，并且会在一定程度上影响后面的预测。但正常情况下，模型的预测效果还是非常可观的。

### 4.1.5 不足与改进

该模型对销量趋势的预测比较精确，对于大部分正常销量值的预测比较精准，但本模型只考虑了序列特征，只进行了简单地LSTM和全连接层的堆叠，但是销量可能还与其他因素有关，比如距离上新日的时间，商品上新的季节等等。可以再对数据选取可用的外部特征，与本模型使用的特征构建一个复合模型。

同时，对数据的采样和筛选处理存在一定的问题，保留一些离群点，例如一个商品在一段时间内销量很低，但是突然有一天销量变得特别高。我们在训练模型的时候保留了这些离群点是为了让模型具有更好的泛化能力，但实验结果发现，这些离群点的保留好像影响了模型预测的准确性。因此，在增强模型泛化能力和提高模型预测准确度上还值得权衡。

最后，整个模型最后的输出销量特征是一个1维结果。实际上，我们在之前对数据的预处理上把销售量分为了20个类，因此如果最后把模型的输出变为20维的向量，相当于把销量的特征增加到了20维，销量之间的信息关联将会被保留的更多，再使用交叉熵验证筛选出概率最高的类，即销量情况，准确度会更高。

## 4.2问题2的模型建立与求解

### 4.2.1 问题的分析

上新量：商品上新当日的销售量，即

上新量=上新日销量

延期比：商品上新时，延期商品和实收商品的比值，即



所以该问题和下一问题的实质是上新日销量和延期比。

### 4.2.2 置信区间的概念

设是总体的待估参数，是取自总体的样本，对给定值，若统计量和满足



则称随机区间为的置信水平为1-的双侧置信区间。和分别称为置信下限和置信上限。[1]

### 4.2.3 模型的建立

构建模型中，我们先假设样本均值遵循正态分布，这个假设可以得到中心极限定理的支持。该定理告诉我们：样本的平均值约等于总体的平均值，不管总体是什么分布，任意一个总体的样本平均值都会围绕在整体的平均值周围，并且呈正态分布。这是群体平均置信区间估计的基础。

为了估计样本平均值的范围，我们定义平均值的标准误差如下：



其中，是样本的标准差，是样本的总数。

在本题中，我们选择了90%，95%和99%三个值作为置信区间的置信水平，这三个置信水平的临界值分别为1.64，1.96和2.32。数学上的具体表示为，其中，为样本的平均值：







### 4.2.4 模型的求解

#### 4.2.4.1 上新量的分布范围及置信区间

（1）分布范围

按照0-5000、5000-10000、10000-15000、15000以上四个区间得到上新量的分布情况如图4-13所示，总共有333种商品。

图 4-13 上新量的分布饼图

其中最小值为294，最大值为25930，故分布范围为294-25930。

（2）置信区间

计算出,,,，置信区间为:

90% confidence inverval: (3850.5661476666055, 4724.4308493303915)

95% confidence inverval: (3765.311054821358, 4809.685942175639)

99% confidence inverval: (3669.3990753704547, 4905.597921626542)

#### 4.2.4.2 延期比的分布范围及置信区间

（1）分布范围

按照每十个百分点为一个区间将商品的延期比进行划分，结果如图4-14所示

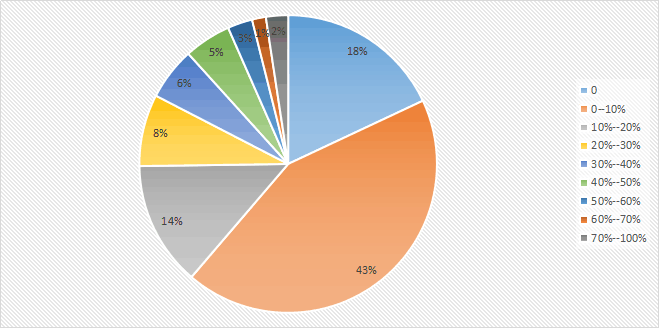


图 4-14 延期比的分布饼图

其中最小值为0%，最大值98.6%，故分布范围为0-98.6%。

（2）置信区间

计算出,,,，置信区间为：

90% confidence inverval: (0.1228940187731031, 0.15664351876443441)

95% confidence inverval: (0.11960138462760736, 0.15993615290993016)

99% confidence inverval: (0.11589717121392466, 0.16364036632361284)

4.2.5分布情况的分析

根据2.4.1中的分析，我们可以知道大部分货物的上新量分布在0-5000，占比达到了80%，一部分货物的上新量在5000-10000，只有少量货物的上新量为10000-15000，而上新量在15000以上的货物基本没有，虽然货物的最大上新量达到了25930，但只有一例这种情况。这说明货物上新量分布除特殊情况外，较为集中。所以可以以0-5000的上新量作为第4题的补单依据。

根据2.4.2中的分析，我们可以知道，大部分货物还是比较准时的，延期比在10%以下的商品占了61%，延期比在30%以下的占83%。因此可以以此为依据在第4题适当考虑延期比对补单的影响。

## 4.3问题3的模型建立与求解

4.3.1 问题的分析

题目与上一题要求类似，只是按照季度不同：春（3月-5月）、夏（6月-8月）、秋（9月-11月）、冬（12月-2月）将所有商品划分成四个部分，对每个季度的商品分别分析其上新量和延期比。

4.3.2 模型的求解

4.3.2.1 四个季度的上新量

（1）春季（3月-5月）

分布范围：如图4-15所示

图 4-15 春季上新量的分布饼图

其中最小值为522，最大值为25930，故分布范围为522-25930。

置信区间：

90% confidence inverval: (3971.3935120332126, 5210.446487966788)

95% confidence inverval: (3850.510294868961, 5331.329705131039)

99% confidence inverval: (3714.5166755591786, 5467.323324440821)

（2）夏季（6月-8月）

分布范围：如图4-16所示

图 4-16 夏季上新量的分布范围饼图

其中最小值为288，最大值为14500，故分布范围为288-14500。

置信区间：

90% confidence inverval: (2815.1476412740067, 3503.3253316989667)

95% confidence inverval: (2748.008354403279, 3570.4646185696943)

99% confidence inverval: (2672.4766566737103, 3645.996316299263)

（3）秋季（9月）

分布范围：如图4-17所示

图 4-17 秋季上新量的分布范围饼图

其中最小值为517，最大值为13004，分布范围为517-13004。

置信区间：

90% confidence inverval: (1929.3035078803314, 3256.6358860590626)

95% confidence inverval: (1799.8076661067967, 3386.1317278325973)

99% confidence inverval: (1654.1248441115704, 3531.8145498278236)

（4）冬季（1月）

分布范围：如图4-18所示

图 4-18冬季上新量的分布范围饼图

其中最小值为294，最大值为9148，故分布范围为294-9148。

置信区间：

90% confidence inverval: (2157.447438609919, 3566.1821910197104)

95% confidence inverval: (2020.0099017894518, 3703.619727840178)

99% confidence inverval: (1865.392672866426, 3858.2369567632036)

4.3.2.2 四个季度的延期比

（1）春季（3月-5月）

分布范围：如图4-19所示

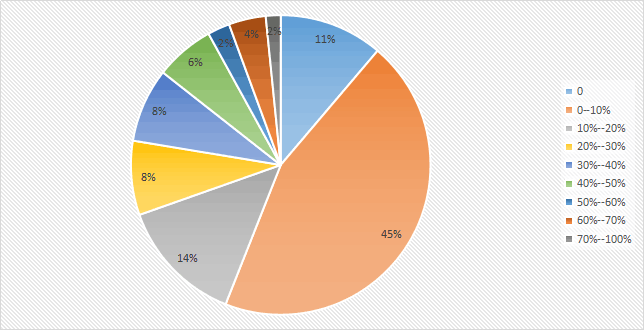


图 4-19 春季延期比的分布饼图

其中最小值为0%，最大值71.1%，故分布范围为0-71.1%。

置信区间：

90% confidence inverval: (0.13287775684450695, 0.18909024315549303)

95% confidence inverval: (0.12739361183855707, 0.1945743881614429)

99% confidence inverval: (0.1212239487068635, 0.2007440512931365)

（2）夏季（6月-8月）

分布范围：如图4-20所示

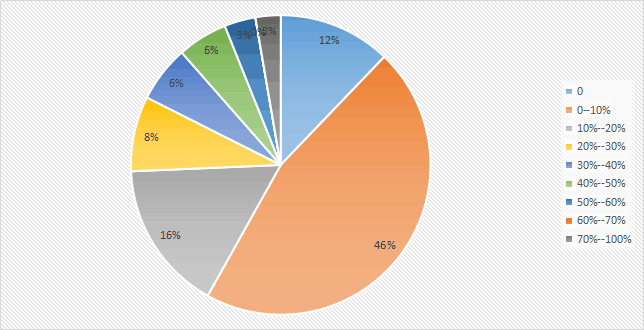


图 4-20 夏季延期比的分布饼图

其中最小值为0%，最大值72.3%，故分布范围为0-72.3%。

置信区间：

90% confidence inverval: (0.11909048363518626, 0.16708519204048938)

95% confidence inverval: (0.11440807305905912, 0.17176760261661653)

99% confidence inverval: (0.1091403611609161, 0.17703531451475954)

（3）秋季（9月）

分布范围：如图4-21所示

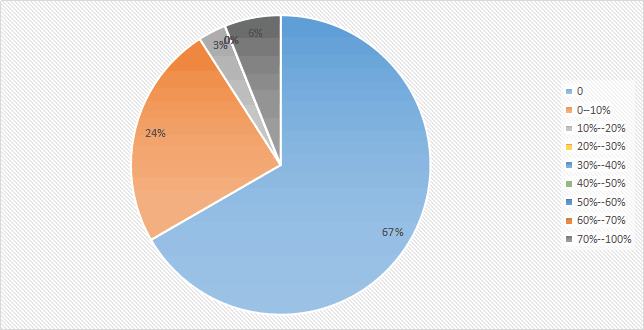


图 4-21 秋季延期比的分布饼图

其中最小值为0%，最大值98.6%，故分布范围为0-98.6%。

置信区间：

90% confidence inverval: (0.0039070492293824255, 0.13469901137667817)

95% confidence inverval: (-0.008853142199622036, 0.14745920280568264)

99% confidence inverval: (-0.023208357557252063, 0.16181441816331266)

（4）冬季（1月）

分布范围：如图4-22所示

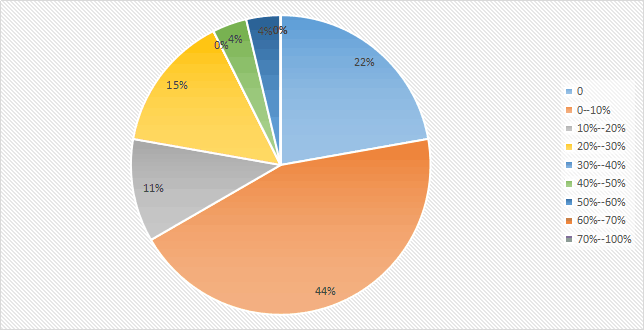


图 4-22 冬季延期比的分布饼图

其中最小值为0%，最大值51.7%，故分布范围为0-51.7%。

置信区间：

90% confidence inverval: (0.1228940187731031, 0.15664351876443441)

95% confidence inverval: (0.11960138462760736, 0.15993615290993016)

99% confidence inverval: (0.11589717121392466, 0.16364036632361284)

4.3.3 分布情况的分析

由3.2.1的图像分析可知，上新量与季度有关。据我们观察，春季的货物的上新量相对其他几个季度最高，上新量在5000以上的商品占了31%；冬季和夏季的货物的上新量相似，上新量在5000以上的商品都占10%左右，上新量在10000以上的商品占3%；秋季的货物的上新量最低，虽然上新量在5000的商品占了11%，但所有商品的上新量都在10000以下。可以说是离春季越近货物的上新量越高。据我们分析，这应该为春季人们的购买欲望最强所导致。

由3.2.2的图像分析可知，延期比也与季度有关。据我们观察，春季货物的延期比最高；夏季和冬季货物的延期比中等，也可以说是较正常；秋季货物的延期比最低。据我们分析这应该和上新量有关。因为春季上新量最大，导致货物更容易延期，而秋季上新量最小，导致货物不那么容易延期。所以使得延期比与季度有关。

## 4.4问题4的模型建立与求解

### 4.4.1 问题的分析

在我们的分析中，具体补单步骤应该分为3部分：

·上新前的补单。

·上新当日的补单。

·上新一段时间后正常的补单。

其中，上新前的补单，应该由该季节的上新量来决定；上新当日的补单是由当日的销售量决定；上新一段时间后正常的补单由我们第一题中建立的模型来决定。

### 4.4.2 模型的建立与求解

#### 4.4.2.1 上新前的补单

上新前的补单在我们的理解里主要为应对上新10天内的商品销售。为计算某个商品的上新前补单量，我们要首先根据上新时间把商品归为某个季节的商品。如上新时间为12-2月，则为冬季商品。同理3-5月为春季商品，6-8月为夏季商品，9-11月为秋季商品。

然后根据不同季节的商品，根据第三题的分析结果，可知当季商品的上新量的大致范围。就可以根据这个范围进行上新前的补单。如春季商品的上新量的大致范围为3714-5467，夏季商品的上新量的大致范围为2672-3645，秋季商品的大致范围为1654-3531，冬季商品的上新量大致范围为1865-3858。所以我们估计春季商品的上新量为4590，夏季商品的上新量为3158，秋季商品的上新量为2592，冬季商品的上新量为2861。

我们根据销售数据分析，上新10天内的销售量大概为上新量的115%。所以春季商品的上新前补单量为5278，夏季商品的上新前补单量为3631，秋季商品的上新前补单量为2980，冬季商品的上新前补单量为3290。

当然，这些数据没有考虑延期比，根据第三题的分析结果，可知当季商品的延期比的大致范围。如春季商品的延期比的大致范围为0.12-0.2，夏季商品的延期比的大致范围为0.11-0.18，秋季商品的大致范围为0-0.16，冬季商品的延期比大致范围为0.12-0.16。所以我们估计春季商品的延期比为0.16，夏季商品的延期比为0.14，秋季商品的延期比为0.08，冬季商品的延期比为0.14。

所以最后春季商品的上新前补单量为6122，夏季商品的上新前补单量为4139，秋季商品的上新前补单量为3218，冬季商品的上新前补单量为3751。

#### 4.4.2.2 上新当日的补单

根据4.2.1的分析结果，可知根据实际第一天的销售量，乘以1.15，就可以得到大致上新后10天的销售量，再乘以1加上估计的该商品的延期比，得到估计的10天内需要的货物量。减去上新前的补单量，就可以得到实际的上新当日的补单量。当然，如果估计的需要的货物量小于上新前的补单量，则上新当日的补单量为0。

#### 4.4.2.3 上新一段时间后正常的补单

我们假设10天进行一次补单。第一次正常补单是在上新10天后。我们根据这10天的数据，根据第一题中的模型，推测出之后10天的销售量。10天销售量相加得到这次需要的补单量。如果现有的存货量大于这次需要的补单量，则不进行补单；如果小于，则按补单的季节，将需要的补单量乘以1加上补单季节的延期比得到最终的补单量，然后按这个补单量进行补单。之后的补单重复这一过程。

# 5 参考文献

1. yeluping666，概率论与数理统计\_7.3置信区间，<https://wenku.baidu.com/view/7f34d80d5fbfc77da369b1a1.html?from=search，2019.5.4>
2. coderpai，Python金融系列第四篇：置信区间和假设检验，<https://blog.csdn.net/CoderPai/article/details/82978556?tdsourcetag=s_pcqq_aiomsg，2019.5.4> 。
3. 百度百科RNN <https://baike.baidu.com/item/RNN/5707183?fr=aladdin>
4. LSTM结构详解 <https://blog.csdn.net/zhangbaoanhadoop/article/details/81952284>
5. 袁祥枫，代银兴 基于LSTM的期货微观市场的趋势预测模 北京邮电大学经济管理学院；北京邮电大学远程与继续教育处
6. 百度百科Hadmard积 <https://baike.baidu.com/item/%E5%93%88%E8%BE%BE%E7%8E%9B%E7%A7%AF/18894493>
7. LSTM模型的前向反向传播算法 <https://www.cnblogs.com/pinard/p/6519110.html>
8. Adam算法Kingma D, Ba J. Adam: A Method for Stochastic Optimization[J]. Computer Science, 2014