0 引言

销售是企业对供应链和运营进行优化时必要的一步(Rao, 1985)。KATI S（2018年2月）[1]研究表明，需求因素对未来产生影响的前提下进行[2].准确的销售预测可以推动产业链的发展，是企业运营的重要组成部分。相反，当企业不能科学的做出销售预测和发展趋势的判断时[3]，往往会导致公司的供应链和运营部门进入混乱或者停滞状态(mindykim, 2019)。

事实上，时间序列预测不仅在工业尤为重要，而且在学术界亦是如此。用Adhikari和Agrawal (2013)的话来说，”研究人员做了很多努力来开发有效的模型,并想以此来提高时间序列的预测精度“。然而，不同

的销售渠道和销售场景有者不同的销售时间序列预测方法，这也让很多研究人员没有办法建立出合适的模型来应用于特定场景中。尤其是在多任务、长时间预测时，因使用不适用于给定时间序列的方法进行预测可能会导致企业季度或年度战略的失败。因此，不论是管理者还是学术研究人员，必须谨慎选择适合当前情况的预测方法(Purthan et al)。

在过去的几十年里，研究人员和企业经营者通常会采用例如HW、神经网络、ARIMA等方法进行相关的销售时间序列预测，在少量数据和特定的预测任务背景下，类似的预测模型能够让预测误差在可接受的范围内。Cranage和Andrew(1992)使用79个月的餐馆销售数据，将计量经济学方法与时间序列方法(SARIMA和HW)进行比较。他们的结果表明，SARIMA方法在初始期和七个月的预测期都有更好的表现。然而，这项研究只针对一家餐厅，并没有对多任务时间序列预测任务进行总结。Makatjane和Moroke(2016)使用南非19年的月度汽车销量数据，发现HW方法相对于SARIMA方法具有更强的预测能力。Udom(2014)使用移动平均法、HW法和SARIMA法对塑料行业中单个分销商的五种不同产品的销售额进行了预测，并以MAPE为评价指标，发现SARIMA方法对销售额的预测效果最好。Frank等人(2003)采用单季节指数平滑法(SSES)，HW和ANNs方法用于女装销售预测。他们发现神经网络显示出令人满意的拟合优度统计量(R2)，但HW方法在比较实际销售和预测销售方面表现得更好。但神经网络对序列数据的前一个时间步长没有记忆，其不能在较长的时间范围内捕获时间序列的特征，例RNN有时会出现梯度消失的问题。

虽然以上文献均采用销售数据进行预测，但得出的结果并不一致。在对不同的时间序列进行建模时，各种方法能够取得性能通常会和数据中的时间维度特征相关，而当机器学习在预测任务上通常会展现比传统模型更高的预测性能时，更多的人选择用此类方法来进行时间序列数据，例如：LSTM (Hochreiter和Schmidhuber, 1997)、XGBoost以及Prophet等。为了验证LSTM网络是否能够取得比传统模型更好的预测效果，Weytjens等人(2019)还将SARIMA和Prophet与ANN和LSTM方法进行比较，以预测现金流。在他们的工作中，他们引入了一种新的绩效衡量方法，即利益机会成本(IOC)。研究利用IOC和MSE作为成本函数进行模型的评价，最终得出结论:LSTM是预测现金流的最佳方法。而在多任务预测中，Yu等(2018)使用LSTM方法预测销售。他们分析了66种产品，包括45周的数据。研究结果表明LSTM为66种产品中的17种提供了准确的预测，而并非全部。考虑到这些数据几乎没有季节性这一特点，单用一个LSTM网络进行多任务的销售预测会面临时间序列长度不够的情形，追其本质，是单个LSTM网络并不能稳定地识别序列数据中的时序特征。

以上的研究工作虽然为销售预测提供了离线预测场景下的可靠方法，但随着云计算、分布式数据库存储技术的发展，企业会要求模型能够进行实时、在线地预测。并且，随着企业推广能力的增强，企业的营业点会分布在各个范围内。这就要求模型要能够进行在线的多任务预测，这使得模型要能够高效地识别出时序数据中的深层次特征，即数据的时空相关性。为适应这一场景，研究设计了一种基于CNN-LSTM网络的在线多任务销售预测模型。此外，本文还用此模型对比了传统的经济预测模型、统计机器学习和单个LSTM模型。目的为在线多任务预测提供理论支持。

1 数据处理

将数据集划分为训练集和测试集，为了得到准确的预测结果和清晰的展示，商品销售数据按照时间序列，将商店和商品类别分类整理。采用特殊值填充法对测试集进行缺失值处理，将所有的销售额用“0”值填充。为了保持训练集和测试集的同步性，让训练集和测试集中的列名和顺序保持一致。

将日期按照one-hot进行日、周、月编码处理，以便确保我们的网络模型可以跟好的处理时间序列数据，对季节性的的数据更加敏感。预测方法选择在线单步预测，因此数据的输入采用追加方式将数据在训练的过程中不断地输入到模型中来进行模型调节。由于神经网络的数据输入会造成一定的抖动，所以采用滑动窗口算法来消除这种抖动，窗口大小=∆N，步长=n，时间间隔=∆T。最后，使用min-max标准化(Min-max normalization)方法对数据集中的销售额进项归一化处理。



3.滑动窗口算法

TODO：背景 原因

3.1算法流程



4.在线预测算法

在线预测与传统的机器学习方法不同，传统的机器学习方法是离线的，通过批量的样本输入来训练模型，模型更新周期较长，实时效果反馈效果差，在小规模数据集时会有一定效果，但是，当大量数据输入模型时会无法实时响应和反馈模型。

首先，将预处理好的销售数据进行特征的提取，并初始化模型参数的权重ω，根据滑动窗口算法选取训练样本实例，使训练样本持续到来，从而减少过多样本带来的阻塞。第二步，使用实时训练的模型预测出当前结果，做出相应的决策。第三步，将预测值和实测值进行比较，依据损失函数得出误差(loss)。第四步，利用损失函数的梯度更新当前的决策，反向传播更新模型的权重，直到损失值很小或者不再下降，当损失函数是连续凸函数时，在线梯度下降可以达到最优的遗憾上界。

|  |
| --- |
| **算法**：在线预测 |
| **Input**：权重ω，变量xi∈D预处理的数据集 |
| **Output**：实时预测结果 |
| **Begin：**  **Repeat**：：  **For** *x∈D={x1,x2,x3…,xn} do*  提取x特征值  **End for**  滑动窗口算法选取输入变量  CNN-LSTM-Model = model.fit(x) //数据输入到CNN-LSTM模型训练  predicted = CNN-LSTM-Model.predict(y) //预测结果  Loss = loss(predicted，true) //计算损失函数  ω = ω + F(Loss) //在线预测更新参数  调节模型  **Until**：loss损失值很小或者不再下降  **End function** |



5 模型



6 实验结果

［1］KATI S,TEEMU L,PETRI S，et al．Forecasting sales in industrial services [J］．Journal of Service Management，2018，29（2）：277－300．

[2]周小溪,徐行,孟剑飞,苏旭中.服装销售预测方法研究进展[J].针织工业,2020(03):68-72.

[3]江雪.基于时间序列模型的商业企业销售量预测方法研究[J].创新科技,2015(07):41-44.