# 基于LSTM对车型月销量预测——以秦PLUS为例

尽管在现有的研究中已经对多种车型的销量已经有了一些研究，但是还存在精准度方面的问题。为了提升精准度，本论文旨在利用长短期记忆网络（LSTM）对车型月销量进行更加精准的预测。本章以秦PLUS为案例进行探讨。通过对LSTM模型的建立和调整，研究者试图提高对车型销量的准确预测能力。该研究意在利用深度学习技术，特别是LSTM模型，来更好地理解和预测汽车销售趋势。通过该章节的研究，我们期望能够为汽车行业提供更为准确的销量预测方法，以帮助制定更有效的市场策略[1–6]。

## 问题分析

汽车销售行业一直以来都是一个充满挑战和变化的领域。随着消费者需求、市场竞争和技术发展的不断演变，汽车制造商和销售商面临着诸多问题和挑战。本论文的问题分析将聚焦于以五个方面：销量预测准确性问题、深度学习在销量预测中的应用问题、案例分析的一般性问题、研究方法的可靠性和有效性问题、市场策略制定与决策支持问题。

在汽车销售领域，准确预测车型的销量对于制定市场策略、生产计划和库存管理至关重要。然而，传统的销量预测方法往往受制于数据复杂性和模型局限性，难以有效捕捉市场变化和消费者行为的细微差异。因此，如何提高销量预测的准确性成为当前亟待解决的问题之一。

随着深度学习技术的发展，特别是长短期记忆网络（LSTM）等模型的出现，越来越多的研究开始探索将深度学习应用于销量预测中。然而，在实际应用过程中，如何克服数据稀缺、模型过拟合、超参数选择等问题，以及如何将深度学习模型与传统方法结合，提高预测效果，仍然具有挑战性。

本论文选择了秦PLUS作为案例进行研究，然而，这种特定车型的销售情况可能受到多种因素的影响，如市场趋势、竞争对手、政策环境等。因此，如何将秦PLUS的销售数据与其他因素进行综合分析，并推广到整个汽车销售市场，以获得更加普适和可靠的结论，是需要深入思考和解决的问题。

本论文采用了LSTM模型进行销量预测，然而，模型的可靠性和有效性直接影响到研究结论的可信度。因此，在选择和建立模型时，如何充分考虑数据特点、模型参数和评估指标，以及如何进行充分的实证分析和灵敏度检验，是需要认真对待的问题。

最终，汽车制造商和销售商希望通过销量预测研究得到有效的市场策略和决策支持。因此，如何将研究结果转化为实际操作建议，以及如何将销量预测与市场营销、供应链管理等方面有机结合，实现企业的长期发展目标，是需要深入探讨和研究的问题[7–14]。

## 模型结构

本节将介绍适用于增程汽车销量预测的长短期记忆网络（LSTM）模型。LSTM是一种常用于序列数据建模的深度学习模型，其特点在于能够有效地捕捉时间序列数据中的长期依赖关系，适用于销售数据这类具有时间序列特征的数据[2,15,16,12]。

### 适用于增程汽车销量预测的LSTM模型

在增程汽车销量预测领域，长短期记忆（LSTM）神经网络模型已被广泛应用，并且取得了显著的成果。LSTM是一种循环神经网络（RNN）的变体，专门设计用于处理时间序列数据，并能够捕捉长期依赖关系，这使得它在预测未来销量时具有优势。

LSTM模型的核心在于其对序列数据的记忆和遗忘机制。相比传统的RNN，LSTM通过三个门（输入门、遗忘门和输出门）来控制信息的流动，从而能够更好地处理长序列，并且在面对梯度消失或梯度爆炸等问题时表现更为稳定。

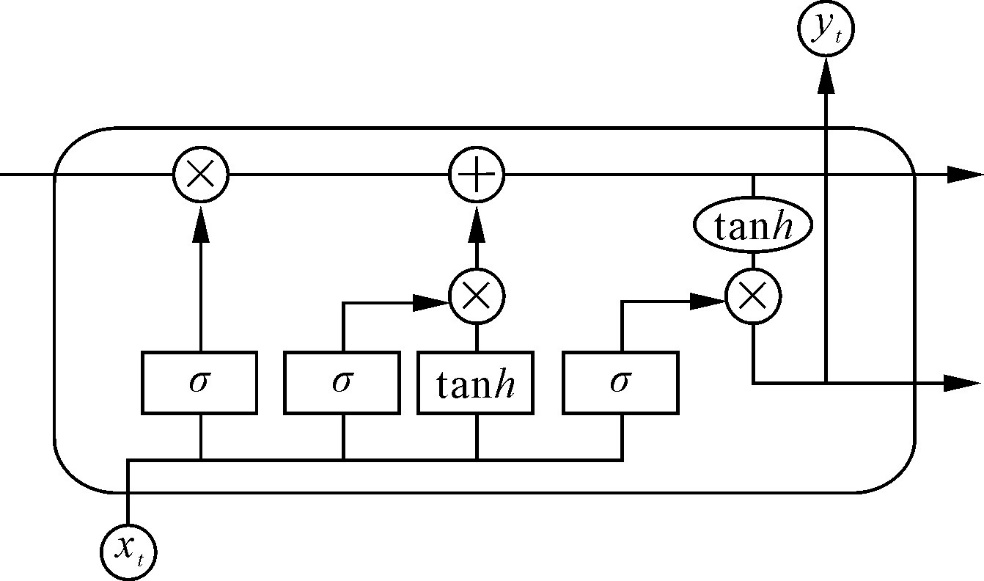


图 4.1 适用于车型销量预测的LSTM模型示意图

输入门（Input Gate）: 控制新信息的输入程度。

遗忘门（Forget Gate）: 控制过去信息的遗忘程度。

输出门（Output Gate）: 控制输出的信息量。

这些门结构使得LSTM能够在训练过程中选择性地记住或遗忘先前的信息，从而更好地捕捉序列中的长期依赖关系。

### 损失函数

在深度学习中，损失函数（也称为目标函数或成本函数）是用来度量模型预测结果与真实标签之间差异的函数。在汽车销量预测的LSTM模型中，常用的损失函数包括均方误差（Mean Squared Error，MSE）和平均绝对误差（Mean Absolute Error，MAE）。以下将详细介绍这两种算法：

均方误差（Mean Squared Error，MSE）：

均方误差是最常见的回归问题损失函数之一。它计算预测值与真实值之间的差异的平方，并对所有样本进行求和后取均值。数学表达式为：

其中，n是样本数量，是第个样本的真实销量， 是模型预测的销量。MSE越小表示模型预测结果与真实值的拟合程度越好。

平均绝对误差（Mean Absolute Error，MAE）：

平均绝对误差是另一种常用的回归问题损失函数，它计算预测值与真实值之间的绝对差异，并对所有样本进行求和后取均值。数学表达式为：

与MSE不同，MAE是使用绝对值来度量误差，因此它对异常值（Outliers）不敏感。MAE越小表示模型的预测结果与真实值的整体偏差越小。

在汽车销量预测的LSTM模型中，将使用实验数据比较两种损失函数的优劣，根据具体情况选择适合的损失函数。通常希望模型更关注与真实值的接近程度，并且对异常值不敏感，可以选择MAE作为损失函数；而如果希望更注重预测值与真实值之间的差异，可以选择MSE作为损失函数。特殊情况下，这两种损失函数可以结合使用，以综合评估模型的性能。

### 优化器

为了提高模型的预测准确性和泛化能力，我们还将进行模型的参数调优和训练过程的优化。在参数调优过程中，我们将选择合适的网络结构、激活函数、学习率等参数，以及进行正则化和防止过拟合的处理。在训练过程的优化中，我们将采用适当的优化算法，例如随机梯度下降（SGD）或Adam优化算法，来加速模型收敛并提高训练效率。

在深度学习中，优化器是用来更新模型参数以最小化损失函数的算法。在汽车销量预测的LSTM模型中，常用的优化器包括随机梯度下降（Stochastic Gradient Descent，SGD）和Adam优化器。

1. 随机梯度下降（Stochastic Gradient Descent，SGD）：

SGD是最基本的优化算法之一，在每个训练批次中都会随机选择一部分数据来计算梯度并更新模型参数。其更新规则如下：

其中， 是第 轮迭代后的模型参数， 是损失函数， 是学习率， 是损失函数关于模型参数的梯度。SGD的优点是简单易实现，但在参数更新时可能存在收敛速度慢、易陷入局部最优等问题。

2. Adam优化器：

Adam是一种自适应学习率的优化算法，它能够根据每个参数的梯度自动调整学习率。Adam算法结合了动量法和RMSProp算法的思想，在每个时间步都会计算一个估计的一阶矩和二阶矩，并使用它们来调整学习率。其更新规则如下：

其中，和 分别是第 轮迭代时的一阶矩和二阶矩估计，和 是衰减率， 是学习率， 是一个小常数用于数值稳定性。Adam优化器具有快速收敛、对超参数不敏感等优点，因此在实践中被广泛应用。

在汽车销量预测的LSTM模型中，根据具体情实验况选择适合的优化器。SGD算法简单易用；而Adam算法自适应性更强，适合大规模数据集或需要更快收敛的情况。

## 实验结果和分析

### 实验设置

软硬件设置：

本研究的实验环境和配置如下：

- 操作系统：Windows

- Python版本：3.8.18

- Numpy版本：1.24.3

- FastReID版本：1.3

- PyTorch版本：1.7.1

- CUDA版本：12.0，运行时11.0

- GPU配置：单个NVIDIA GeForce RTX 4060 Laptop GPU，支持CUDA

其他关键库版本：Pillow 10.2.0，OpenCV (cv2) 4.9.0，torchvision 0.8.2

PyTorch构建信息包括：

- C++版本：199711

- 使用Intel(R) Math Kernel Library Version 2020.0.0

- 支持AVX2指令集

- 使用CuDNN 8.0.4

- 使用Magma 2.5.4进行矩阵计算

- 编译设置包括使用MKL、OpenMP、CUDA等

以上配置确保了实验的重复性和结果的准确性，同时也体现了本研究在实验设置上的细致考虑。

实验参数设置：统一学习率为0.001，epoch为100，

### 模型收敛情况

下图展示了模型在训练过程中损失函数的收敛情况：

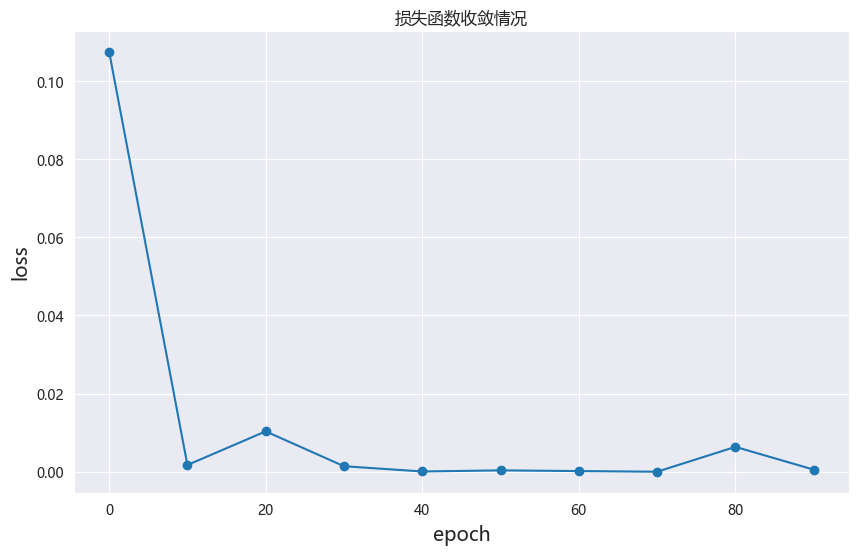


图 4.2 损失函数随着模型训练收敛情况

可以看出模型的收敛情况比较好，说明模型正确学习到了销量随着时间的变化趋势。

### 对比实验

通过大量的对比实验，我们加入

表 4.1大量的对比实验表明本文模型的优越性

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **模型配置** | **MAE** | **MSE** | **RMSE** |  |
| LSTM+Adam+MAE | 0.1215 | 0.0425 | 0.2063 | 0.8084 |
| LSTM+Adam+MSE | 0.0511 | 0.0091 | 0.0952 | 0.9592 |
| LSTM+SGD+MSE | 0.3795 | 0.2160 | 0.4648 | 0.0271 |
| LSTM+SGD+ L1 | 0.3558 | 0.2217 | 0.4708 | 0.0017 |
| LSTM+SGD+ SmoothL1 | 0.3838 | 0.2185 | 0.4674 | 0.0162 |
| **LSTM+Adam+ SmoothL1** | **0.0303** | **0.0016** | **0.0405** | **0.9908** |

### 模型可解释性

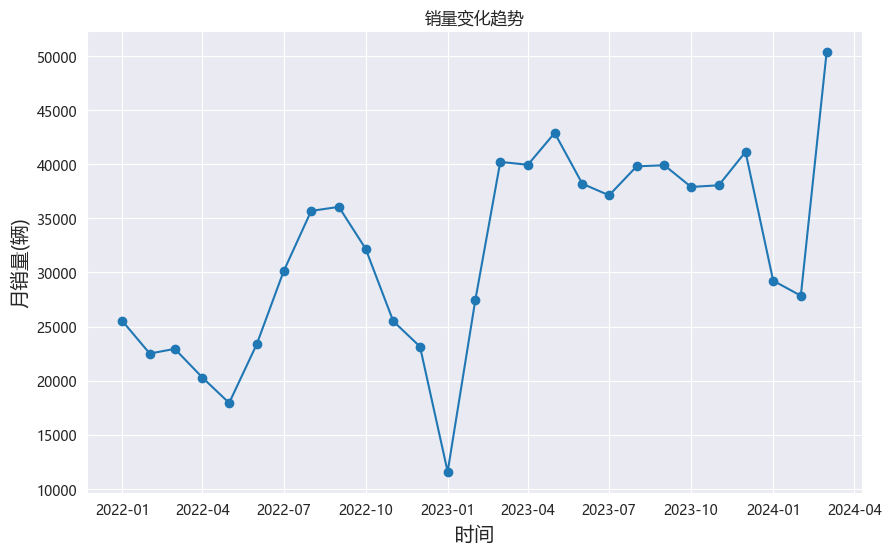
首先对输入数据进行可视化，如下图所示，根据网络公开信息（不同信源的数据可能不同，本文采用的是车主之家的数据），采集到某车型的销售量，可以看出，销售量具有一定的时间波动性和时间周期性：

图 4.3 车型销量随时间变化示意图

下图展示了预测4月销量的结果：

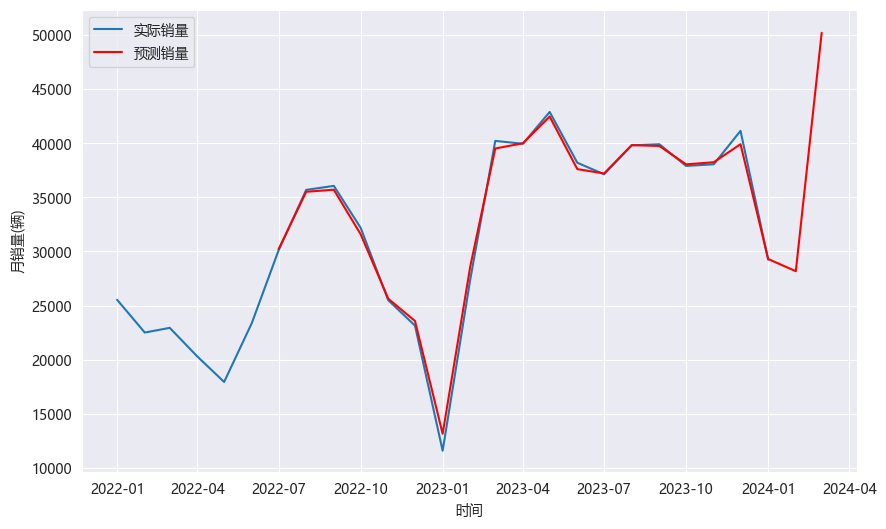


图 4.4 模型预测效果可视化

红色线是使用LSTM模型结合之前的销售数据对四月销量的预测，销量为模型最后预测的50530辆，结合所公开的最新消息，比亚迪秦PLUS四月销量为51170辆[17]，模型具有很强的参考价值。

## 本章小结

本章旨在利用长短期记忆网络（LSTM）对车型月销量进行预测，以秦PLUS为案例进行探讨。通过建立和调整LSTM模型，研究者试图提高对车型销量的准确预测能力。本研究致力于利用深度学习技术，特别是LSTM模型，更好地理解和预测汽车销售趋势，为汽车行业提供更为准确的销量预测方法，以帮助制定更有效的市场策略。

在研究过程中，我们首先对当前汽车销售行业面临的问题进行了分析，包括销量预测准确性、深度学习应用、案例分析的一般性、研究方法的可靠性和有效性、以及市场策略制定与决策支持等方面。随后，我们介绍了适用于增程汽车销量预测的LSTM模型的原理和结构，以及损失函数和优化器的选择。

通过实验设置和模型收敛情况的分析，我们验证了模型在训练过程中的有效性和收敛性。对比实验结果显示，不同配置下的模型在MAE、MSE、RMSE和等指标上存在差异，进一步验证了模型的性能。

最后，我们还对模型的可解释性进行了探讨，通过对输入数据的可视化和预测结果的展示，验证了模型对销量的预测能力，为汽车行业提供了有力的决策支持。

综上所述，本章的研究成果为汽车销售行业提供了一种基于深度学习的销量预测方法，并且在实证分析中取得了良好的效果，具有重要的理论和实践意义。

# 参考文献

[1] 陈瑜. 全球道路燃料需求2027年达峰[J]. 中国石油企业, 2023(7): 46.

[2] 陈尚林. 基于LSTM神经网络和百度指数的新能源汽车销量预测[D]. 湖北大学, 2024.

[3] 李媛媛. 基于改进Bass模型的汽车销量预测方法研究[D]. 辽宁大学, 2024.

[4] 林琳. 全生命周期汽车销量预测方案研究[J]. 汽车工业研究, 2023(3): 44-48.

[5] 蓝晨卉, 李卓, 任俊杰, 等. 基于季节性因素影响的汽车销量预测模型研究及应用[J]. 时代汽车, 2023(14): 163-166.

[6] 段昊江, 吴冰. 多变量乘用车销量预测模型研究[J]. 汽车文摘, 2023(12): 55-62.

[7] 邹婷婷. 基于灰色Lotka-Volterra模型的中国新能源汽车销量预测及其环境影响研究[D]. 江南大学, 2024.

[8] 余明洋, 沈斌. 多变国际形势下新能源汽车销量分析——基于突发因素的复合预测模型[J]. 中国商论, 2023(12): 164-168.

[9] 杨鸿渐. 大数据处理技术在汽车产销协同的应用综述[J]. 汽车工业研究, 2024(1): 42-46.

[10] 杨光. 车企数字化破局之路[N]. 中国信息化周报, 2023-04-24: 008.

[11] 许妙遥. 我国新能源汽车的销量预测及发展路径研究[D]. 山东工商学院, 2024.

[12] 辛宇涛. 基于G-SARIMA-BP-LSTM的新能源汽车销量预测研究[D]. 山西大学, 2024.

[13] 吴静, 黄安, 谢涵鹏, 等. 基于机器学习势函数的材料力热性质多尺度模拟研究进展[J]. 硅酸盐学报, 2023: 1-13.

[14] 王书田, 林岩, 朱国庆, 等. 基于在线评论和网络搜索的汽车销量预测研究[J]. 东北大学学报(自然科学版), 2023, 44(5): 752-760.

[15] 冯非凡, 武雪玲, 牛瑞卿, 等. 一种V/S和LSTM结合的滑坡变形分析方法[J]. 武汉大学学报(信息科学版), 2019, 44(05): 784-790.

[16] 卢志平, 玉晓晶, 陆成裕. 融合VMD和LSTM模型的新能源汽车销量预测方法[J]. 武汉理工大学学报(信息与管理工程版), 2023, 45(4): 546-551.

[17] 东方网. 快马加鞭，比亚迪24年4月销量31.2万辆，继续遥遥领先[EB]. (2024-05-02)[2024-05-10]. https://finance.sina.com.cn/jjxw/2024-05-02/doc-inatvexc6987913.shtml.