

基于长短期记忆网络的汽车用户消费趋势预测

付羽佳 田梦

中汽数据（天津）有限公司

摘要：长短期记忆网络在诸多领域展现出强大的预测能力。对汽车用户消费趋势进行预测，可助力企业精准决策。利用长短期记忆网络对汽车用户消费数据进行深度挖掘与分析，能捕捉复杂时间序列特征。经模型训练与优化，实现对汽车用户消费趋势的有效预测，为汽车行业制定营销策略、规划产品提供科学依据，提升市场竞争力。

关键词：长短期记忆网络；汽车用户；消费趋势预测

引言：随着汽车市场竞争日益激烈，准确把握用户消费趋势成为企业致胜关键。传统预测方法在处理复杂时间序列数据时存在局限。长短期记忆网络作为一种特殊的循环神经网络，能有效解决梯度消失或爆炸问题，在时间序列预测方面表现出色。将其应用于汽车用户消费趋势预测，具有重要的理论和现实意义。

1 长短期记忆网络基础

1.1 长短期记忆网络原理

长短期记忆网络（LSTM）作为循环神经网络（RNN）的改进型网络，核心是解决传统 RNN 在处理长序列数据时易出现的梯度消失或梯度爆炸问题。其原理通过引入记忆单元和门控机制实现，记忆单元能长期存储序列信息，而输入门、遗忘门和输出门分别控制信息的输入、遗忘和输出。输入门决定哪些新信息能进入记忆单元，遗忘门判断记忆单元中哪些旧信息该被丢弃，输出门则筛选记忆单元中哪些信息需输出到下一层。通过这样的机制，LSTM 可有效捕捉长序列数据中的时序依赖关系，精准提取序列中的关键信息，为后续的趋势预测提供可靠的数据处理基础。

1.2 长短期记忆网络结构

长短期记忆网络结构由多个重复的 LSTM 单元组成，每个单元包含记忆单元、输入门、遗忘门和输出门四大核心部分。记忆单元是单元的核心，如同信息的“仓库”，负责持续存储序列数据中的关键信息；输入门由 sigmoid 函数和 tanh 函数构成，sigmoid 函数输出 0 到 1 之间的数值，决定新信息的纳入比例，tanh 函数生成候选信息供输入门筛选；遗忘门同样采用 sigmoid 函数，通过输出值控制记忆单元中历史信息的保留或遗忘；输出门结合 sigmoid 函数和 tanh 函数，先经 sigmoid 函数筛选记忆单元信息，再通过 tanh 函数将信息压缩后输出。这些部分协同工作，使 LSTM 单元能有序处理时序数据，各单元按序列顺序连

接，形成完整的 LSTM 网络结构，实现对长序列数据的有效处理。

1.3 长短期记忆网络优势

相较于传统的时序预测模型和其他神经网络模型，长短期记忆网络具有显著优势。首先，在处理长序列数据时，传统模型易丢失早期关键信息，而 LSTM 通过门控机制和记忆单元，能长期保留序列中的重要信息，准确捕捉长时序依赖关系，避免梯度消失或爆炸问题，提升对复杂时序数据的处理能力。其次，LSTM 对数据的适应性强，无论是线性还是非线性的时序数据，都能通过自身的网络结构和参数调整，精准提取数据中的潜在规律和特征，无需对数据进行复杂的预处理转换。此外，在多变量时序预测场景中，LSTM 可同时处理多个相关变量数据，综合分析各变量对预测目标的影响，提高预测结果的准确性和全面性，这一优势使其在汽车用户消费趋势预测等复杂场景中应用广泛。

2 汽车用户消费数据

2.1 数据类型与来源

汽车用户消费数据类型丰富，涵盖用户基本信息、消费行为信息、产品相关信息等多个维度。用户基本信息包括用户的年龄、性别、地域、收入水平、职业等，这些信息能反映用户的消费能力和潜在消费偏好；消费行为信息包含用户的购车时间、购车类型、购车频次、保养维修记录、配件购买情况等，可直观体现用户的消费习惯和消费规律；产品相关信息涉及汽车的车型、配置、价格、性能参数、品牌等，有助于分析产品特性与用户消费选择的关联。数据来源同样广泛，主要包括汽车销售企业的内部销售系统，记录用户购车及后续服务数据；汽车电商平台的用户浏览、下单、评价数据；汽车行业调研机构开展的用户消费调研数据，以及社交媒体中用户关于汽车消费的讨论、分享

数据等，这些来源共同构成了全面的汽车用户消费数据集。

2.2 数据预处理方法

汽车用户消费数据在采集过程中，常存在缺失值、异常值、重复值等问题，需通过科学的预处理方法提升数据质量。对于缺失值，根据数据类型和缺失比例采用不同处理方式，若缺失比例低，对数值型数据采用均值、中位数填充，对分类数据采用众数填充；若缺失比例高，结合业务逻辑判断是否删除该字段或采用插值法、模型预测法进行填充。针对异常值，通过绘制箱线图、计算 Z-score 值等方法识别，再依据业务场景判断是数据采集错误还是真实异常数据，错误数据予以修正，真实异常数据则根据其对分析的影响程度，选择保留、删除或进行数据转换。对于重复值，通过数据比对去重，删除完全重复或核心字段重复的数据。此外，还需对数据进行标准化或归一化处理，消除不同量级数据对模型训练的影响，将分类数据通过独热编码、标签编码等方式转换为数值型数据，为后续模型训练提供规范、高质量的数据。

2.3 数据特征分析

汽车用户消费数据特征分析是挖掘数据潜在价值、为模型训练提供有效输入的关键环节。首先从单特征分析入手，对用户年龄、收入水平等数值型特征，通过计算均值、方差、四分位数等统计指标，结合直方图、密度图等可视化方式，了解特征的分布情况，判断是否存在偏态分布等异常；对用户性别、购车品牌等分类特征，通过计算各类别的频次、占比，绘制条形图，分析不同类别的分布差异。其次进行多特征关联分析，探究不同特征之间的相互关系，例如分析用户年龄与购车类型的关联，判断不同年龄段用户对轿车、SUV 等车型的偏好差异；分析用户收入水平与购车价格区间的关系，了解消费能力与消费选择的对应规律。同时，还需识别对消费趋势预测具有重要影响的关键特征，通过相关性分析、特征重要性评估等方法，筛选出与用户消费趋势紧密相关的特征，减少冗余特征对模型的干扰，提升模型训练效率和预测准确性。

3 基于长短期记忆网络的预测模型

3.1 模型构建过程

基于长短期记忆网络的汽车用户消费趋势预测模型构建，需遵循严谨的流程。首先明确预测目标，确定是预测用户未来的购车需求、购车类型偏好，还是

消费金额变化趋势等，不同目标对应不同的模型设计思路。接着进行数据准备，将预处理后的汽车用户消费数据按时间序列顺序整理，划分训练集、验证集和测试集，通常采用时间切片的方式将时序数据转换为模型可接受的输入格式，确定合适的时间步长，使模型能基于历史多步数据预测未来趋势。然后设计网络结构，根据预测目标和数据特征，确定 LSTM 层的数量、每层的神经元个数，同时设置输入层和输出层，输入层维度与数据特征维度匹配，输出层根据预测任务（分类或回归）选择相应的激活函数和输出维度。最后，选择合适的优化器、损失函数，配置模型训练的迭代次数、批次大小等超参数，完成模型的初步构建。

3.2 模型训练与优化

模型训练阶段，将整理好的训练集数据输入构建好的 LSTM 模型，按照设定的超参数开始训练。训练过程中，模型通过前向传播计算预测值，再通过损失函数计算预测值与真实值的误差，然后利用反向传播算法，根据误差调整网络中的权重和偏置参数，不断降低损失值，使模型逐步学习到数据中的时序规律。在训练过程中，需实时监控训练集和验证集的损失变化，若验证集损失不再下降甚至上升，说明模型可能出现过拟合，此时需采取相应的优化措施。优化方法包括增加训练数据量、引入 dropout 层减少神经元间的过度依赖、调整超参数（如减小学习率、增加批次大小、调整 LSTM 层神经元个数）、采用早停策略在验证集损失最优时停止训练等。通过反复的训练与优化，不断提升模型的泛化能力和预测准确性，确保模型在测试集上也能保持良好的性能。

3.3 模型评估指标

为全面、客观地评价基于长短期记忆网络的汽车用户消费趋势预测模型性能，需采用多种合适的评估指标。对于回归类预测任务（如消费金额预测），常用的评估指标包括平均绝对误差（MAE）、均方误差（MSE）和决定系数（ R^2 ）。MAE 计算预测值与真实值的绝对误差平均值，能直观反映预测的平均偏差；MSE 通过计算误差的平方平均值，放大较大误差的影响，更能体现模型对极端值的预测能力； R^2 衡量模型对数据变异的解释能力， R^2 越接近 1，说明模型的预测效果越好。对于分类类预测任务（如购车类型预测），则采用准确率、精确率、召回率和 F1 分数等指标。准确率反映模型整体预测正确的比例；精确率衡量预

测为正类的样本中实际为正类的比例；召回率表示实际为正类的样本中被模型正确预测的比例；F1 分数是精确率和召回率的调和平均数，综合两者的性能，避免单一指标的片面性。通过这些指标的综合分析，可全面判断模型的预测性能，为模型的改进和应用提供依据。

4 预测结果与应用

4.1 消费趋势预测结果

基于长短期记忆网络的汽车用户消费趋势预测结果，能清晰呈现未来一段时间内汽车用户的消费变化方向和规律。从用户购车需求来看，预测结果可明确不同群体用户（如不同年龄、地域、收入群体）未来购车需求的增减趋势，例如某一地域年轻用户群体的购车需求可能呈上升趋势，而另一地域中年用户群体的购车需求可能趋于稳定。在购车类型偏好方面，预测结果能反映用户对不同车型（如轿车、SUV、新能源汽车）、不同配置（如智能科技配置、安全配置）的偏好变化，比如未来用户对新能源汽车的偏好可能持续上升，对智能驾驶辅助配置的需求可能显著增加。此外，预测结果在多个关键维度上具有重要意义。它能够清晰体现用户消费金额的变化趋势，比如是逐渐上升、保持稳定还是出现下滑，让企业知晓消费者的购买力走向。同时，购车频次的动态也能被精准捕捉，了解消费者更换或新增车辆的频率规律。这些信息为汽车企业全面了解市场需求动态、精准把握用户消费方向提供了极为精准且有价值的参考依据。

4.2 对汽车企业的启示

预测结果为汽车企业的生产、销售、研发等各环节提供重要启示，助力企业制定科学合理的经营策略。在生产环节，企业可根据预测的用户购车类型偏好和需求规模，调整生产计划，优化产品生产结构，加大对用户偏好车型（如新能源汽车）的生产投入，减少滞销车型的产量，避免库存积压，提高生产效率和资源利用率。在销售环节，结合预测的不同群体用户消费趋势，针对不同用户群体制定差异化的营销策略，例如对偏好智能科技配置的年轻用户，重点宣传汽车的智能功能；对注重安全的家庭用户，突出汽车的安全配置优势，同时合理规划销售渠道和促销活动，提升产品的市场占有率。在研发环节，依据预测的用户消费需求变化，明确研发方向，加大对用户关注领域（如新能源技术、智能驾驶技术）的研发投入，推出

更符合用户未来需求的新产品，增强企业的市场竞争力和行业影响力。

4.3 预测的局限性与展望

基于长短期记忆网络的汽车用户消费趋势预测虽能提供有价值的参考，但仍存在一定局限性。一方面，预测结果受数据质量和数据范围影响较大，若数据存在缺失、偏差或数据覆盖范围有限，可能导致预测结果不准确；另一方面，市场环境中存在诸多不可预测的突发因素（如政策重大调整、经济形势突变、自然灾害等），这些因素难以在模型中完全量化和体现，可能使预测结果与实际情况产生偏差。未来，可从多方面完善预测体系，一是扩大数据来源，纳入更多维度的相关数据（如宏观经济数据、政策数据、交通数据等），提升数据的全面性和代表性；二是优化模型结构，将 LSTM 与其他先进的神经网络模型（如卷积神经网络、注意力机制模型）结合，构建混合神经网络模型，进一步提升模型的预测能力；三是加强对突发因素的研究，建立突发因素影响评估机制，提高预测结果的稳健性和可靠性，更好地为汽车行业的发展提供支持。

结束语：长短期记忆网络为汽车用户消费趋势预测提供了新途径。通过对相关数据的处理与模型构建，能得到较准确的预测结果。这有助于汽车企业了解市场动态、优化产品与服务。但预测仍存在一定局限性，未来需进一步改进模型、拓展数据，以提升预测精度，更好地服务汽车行业发展。

参考文献

- [1] 孙宏贤，徐兰．基于长短期记忆网络的高速公路车辆变道轨迹预测模型[J]．计算机测量与控制，2023, 31(12): 316-321.
- [2] 王明祥．基于长短期记忆神经网络的汽车主动悬架优化控制研究[D]．西安理工大学，2023.
- [3] 柏海舰，孙婷，丁恒，等．基于长短期记忆神经网络与注意力机制的智能汽车分车型跟驰模型[J]．汽车工程学报，2023, 13(06): 821-831.
- [4] 刘影．基于部分亲和域和长短期记忆网络的行人轨迹预测研究[D]．上海工程技术大学，2021.
- [5] 刘宏阳，杨林，李济霖．基于长短期记忆网络的电动汽车电池故障诊断[J]．机电一体化，2020, 26(Z1): 17-23+37.