西北大学信息科学与技术学院

本科毕业设计开题报告/答辩登记表

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 学生学号 | |  | | 姓名 |  | 年级 |  | | | |
| 专业 | |  | | | | | | | | |
| 论文（设计）题 目 | | 时序数据下降波动状态识别方法研究 | | | | | | | | |
| 指导教师  姓 名 | |  | | 专业技术职务 | |  | 开题报告日期 | | |  |
| 企业导师  姓 名 | |  | 文献综述成绩 | | |  | | 开题报告成绩 |  | |
| 答辩小组成员（姓名，职称）： | | | | | | | | | | | |
| 答辩小组组长签字： 年 月 日 | | | | | | | | | | | |
| **开 题 报 告 内 容** | | | | | | | | | | |
| 选题来源 | 1．教师指定（√）2．教师课题（ ）3．创新基金项目（ ）4．自选（ ） | | | | | | | | | |
| 设计选题的背景与意义、理论与实证准备、拟解决的问题、研究（设计）方法与技术路线 | 1. **选题的背景与意义**   随着物联网和 5G 技术的不断发展，社会开始进入互感、互联的大数据应用时代。人们可以通过安装在物体上的传感器来实时收集“信息”（即时间序列数据，简称时序数据），从而监控物体现有状态和预测物体未来状态。时间序列数据指的是同一对象在不同时间点上收集到的一串有序数据，比如，电商公司每个月的销售总数量、某城市每天PM2.5 的变化值和某道路每小时的交通流量等等。它广泛存在于工业、医疗和金融等各个领域中。  通常时间序列数据中蕴含着丰富的信息，挖掘分析这些信息能够帮助人们理解现象预测未来。比如，在股市中，分析股票历史以来的每日收盘价格组成的时间序列数据来预测未来时间的股价。  通常,对于时间序列数据的分析主要包括时间序列预测以及时间序列分类，而本文我们主要研究时间序列预测的问题。  时间序列预测是通过分析历史的时间序列数据，找到时间序列数据所表现出来的发展趋势，从而预测下一刻时间或以后一段时间可能的取值。传统的时间序列预测方法仅仅依据历史值进行线性加权预测，难以对序列数据的非线性进行建模，导致预测精度不高。传统的机器学习方法通过构建时间序列特征对未来值进行预测，虽然其具有优秀的非线性建模能力，但未考虑序列数据的时间依赖关系和变量之间的相关性，从而难以达到高精度预测。而深度学习不仅具有强大的非线性建模能力，同时循环神经网络及其变形能有效对序列的长期依赖关系进行建模，自注意力机制能有效捕获序列数据之间的相关性，对于提高预测精度具有重要的作用。  综上所述，深度学习能学习数据隐含信息的抽象特征表示，同时能有效对序列变量之间的相关性以及数据长期依赖关系进行建模，从而可以更好的理解和表达数据的隐含信息，使得预测精度和分类准确率都有极大的提高。因此，本文利用深度学习方法对时间序列数据进行预测具有重要的研究意义。   1. **理论与实证准备**   机器学习指计算机通过分析和学习大量已有数据，从而拥有预测判断和做出最佳决策的能力。其代表算法有深度学习、人工神经网络、决策树、增强算法等。它使用算法来解析数据，从中学习，然后对真实世界中的事件做出决策和预测。  BP(back propagation)神经网络是1986年由Rumelhart和McClelland为首的科学家提出的概念，是一种按照误差逆向传播算法训练的[多层前馈神经网络](https://baike.baidu.com/item/%E5%A4%9A%E5%B1%82%E5%89%8D%E9%A6%88%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C/10435655)，是应用最广泛的神经网络模型之一。  递归神经网络（RNN）是深度神经网络（DNN）的一大研究分支，而长短期记忆神经网络（LSTM）则是一种新型的具有记忆块机制的改进递归神经网络，LSTM同时具有网络模型学习与记忆功能。  LSTM 作为循环神经网络的一种特殊形式，首先由 Hochreiter 等在 1997 年提出［19］，如今已广泛应用于时间序列数据的预测。经典 LSTM 是在原始 RNN 隐藏层中增加一个传输的单元状态（Cell state），并由三个门控单元（Gate）控制，即输入门 It、遗忘门 Ft 和输出门 Ot，结构如图 1 所示    图1  文献[1] 针对现有股价预测模型较复杂导致实际应用性不强的问题,构建一个计算简便可广泛应用的时序权重均值模型,并选取沪深A股市场207支股票的243个交易日收盘价对该模型与常用的算数平均数法的预测精度进行对比。  文献[2]利用1961-2020年黄河流域河南段的逐月气象数据,计算不同时间尺度的标准化降水蒸散指数（SPEI）,在Copula熵的基础上根据Hampel准则选择干旱驱动因子,构建多变量长短时记忆（LSTM）神经网络预测模型.  文献[3] 提出了一种基于Softmax函数的注意力模块，并将其应用在LSTM的输入前，使模型可以根据输入数据中的时间和空间信息，自主地生成带有权重的词义向量，并赋予输入序列时间和空间注意力权重值.注意力机制增强了LSTM模型对时间序列的处理能力  文献[4] 提出一种基于长短期记忆网络（long short-term memory，LSTM）与一维卷积神经网络（1-dimensional convolutional neural network， 1DCNN）的目标轨迹预测方法。  文献[5]提出一种基于时序数据对工作面设备进行故障预测的方法。  文献[6] 提出了一种基于时序集成森林的股票多类别预测算法。该算法分为两部分,首先是随机森林改进的LSTM进行股票的收盘价预测;其次,根据第二天预测的收盘价与前一天的收盘价对比得到涨跌信号;最后将LSTM提取的时序特征与涨跌信号输入集成森林进行股票的涨跌预测。  文献[7] 提出了一种人工经验与主成分分析相结合的长短期记忆网络方法 （AEPCA-LSTM）,利用运行过程中的监测时序数据对设备运行趋势进行预测。  文献[8] 对时间序列数据传统的预测和分类方法进行调研,同时调研深度学习的相关理论知识,主要包括卷积神经网络、循环神经网络和注意力机制,并重点研究基于深度学习的时序预测和分类方法。  文献[9] 基于Xgboost模型的特征生成方法研究。研究发现输入特征对股票指数预测模型的性能具有重大影响。现有特征选择与特征提取方法在利用基础数据信息方面存在着部分丢失与不充分的问题。  文献[10] 采用了一种基于经验模态分解（EMD）和长短期记忆网络（LSTM）模型的组合预测方法,对股指进行统计性描述,发现中国3个股指的波动具有明显区别,就这一特征对数据进行建模。  文献[11] 与单一结构的LSTM神经网络模型预测相比,本文提出的RF-LSTM组合模型预测的平均绝对误差（MAE）、均方误差（MSE）和均方根误差（RMSE）分别减小了13.11%,6.70%和12.54%。该组合模型可提高股票价格预测的准确性。  文献[12] 使用ARIMA模型抓取数据的线性特征,使用GRU模型抓取数据的非线性特征,选取合适的参数以及网络优化算法调整网络以获得更好结果,其后与其他模型的预测结果进行了对比.  文献[13] 提出一种基于改进型BP神经网络的多变量时间序列预测方法。该方法对多变量时序数据建模预测能力较强,能够在有效减少训练时间的前提下,提高数据预测的精度。  文献[14]利用小波分析对金融时间序列做多尺度分解、去噪,借助改进的粒子群算法对BP神经网络的隐层进行优化,并建立金融时间序列的分层预测模型。  文献[15] 设计了一种基于惯性权重改进花朵授粉算法（MFPA）和误差逆向传播（BP）神经网络结合的MFPA-BP短时交通流预测模型。  […]   1. **拟解决的问题**   1）如何抽象并研究时序数据的下降波动特征的指标。  2）标测试模型能否识别下降波动数据。  3）结果的拟合程度，并对优化前和优化后进行对比。  4）重复迭代优化模型参数，提高模型准确率。   1. **研究方法与技术路线**   针对上述存在问题，本文会去解决上面四个问题，本系统会以长短期神经记忆网络为基础对训练集进行训练。具体研究内容如下：   1. 数据获取与预处理。获取时序数据的预选数据集并进行预处理。剔除掉一些明显错误的特殊数据，并搭建LSTM神经网络架构。 2. 处理数据，并选择出训练集和测试集。 3. 构建一种基于BP神经网络的预测模型设计。 4. 构建一种新的以LSTM为主要基分类器的集成学习模型。 5. 构建新的集成学习模型，判断模型是否能识别出下降波动特征并与传统的集成学习模型对比，以验证新的集成学习模型的有效性。 6. 对于最终的预测结果，将使用准确率和召回率对预测的结果进行计算，根据计算的结果判定预测结果是否正确。 | | | | | | | | | |
| 论文写作提纲 | （除题目外，具体到三级标题）：  第一章 绪论  1.1研究背景及意义  1.2研究现状  1.3本文研究内容  1.4本章组织结构  第二章 时序数据获取与处理  2.1数据获取  2.2股票价格下降信号  2.3指标的计算和处理  2.4特征数据归一化  2.5本章小结  第三章 模型的构建以及对比  3.1基于BP（Back Propagation）神经网络的预测模型设计  3.1.1 BP神经网络模型介绍  3.1.2 BP神经网络模型构建  3.1.3 BP神经网络实验设置  3.2基于长短期神经记忆网络（LSTM）的预测模型设计  3.1.1 模型介绍  3.1.2 LSTM神经网络模型构建  3.1.3 LSTM神经网络实验设置  3.3 模型训练以及结果对比  3.4本章小结  第四章：总结与展望  4.1本文总结  4.2 未来展望  **参考文献：**  [1]赵庆国,孔祥月,刘莉明,杨龙倩.短期股票价格预测的时序权重均值模型构建[J].沈阳航空航天大学学报,2020,37(04):81-89.  [2]李艳玲,巩雅杰.基于驱动分析的LSTM干旱预测模型研究[J/OL].数学的实践与认识:1-11[2022-05-07].http://kns.cnki.net/kcms/detail/11.2018.O1.20220425.1140.030.html  [3]王迎飞,黄应平,肖敏,熊彪,周爽爽,靳专.基于注意力机制的LSTM长江汛期水位预测方法研究[J].三峡大学学报(自然科学版),2022,44(03):13-19.DOI:10.13393/j.cnki.issn.1672-948x.2022.03.003.  [4]宋波涛,许广亮.基于LSTM与1DCNN的导弹轨迹预测方法[J/OL].系统工程与电子技术:1-11[2022-05-07].http://kns.cnki.net/kcms/detail/11.2422.tn.20220421.1454.006.html  [5]郑磊.基于时序数据的工作面设备故障预测研究[J].工矿自动化,2021,47(08):90-95.DOI:10.13272/j.issn.1671-251x.17694.[6]王平飞.基于时序集成森林的股票多类别预测研究[J].现代计算机,2021(18):169-175.  [7]杨柯,范世东.基于长短期记忆网络时序数据趋势预测及应用[J].推进术,2021,42(03):675-682.DOI:10.13675/j.cnki.tjjs.200394.  [8]谭振宁. 基于深度学习的时序预测和分类[D].华南理工大学,2020.DOI:10.27151/d.cnki.ghnlu.2020.000504.  [9]耿旭东. 基于机器学习的股票指数预测研究[D].河南大学,2019.  [10]刘铭,单玉莹.基于EMD-LSTM模型的股指收盘价预测[J].重庆理工大学学报(自然科学),2021,35(12):269-276.  [11]李辉,化金金,邹波蓉.基于RF-LSTM组合模型的股票价格预测[J].河南理工大学学报(自然科学版),2022,41(01):136-142.DOI:10.16186/j.cnki.1673-9787.2019100021.  [12]刘昌荣,黄珍,袁贝贝,李岚.基于深度学习的股价趋势预测研究[J].兰州文理学院学报(自然科学版),2021,35(06):56-61.DOI:10.13804/j.cnki.2095-6991.2021.06.012.  [13]陈建婷.基于改进型BP神经网络的多变量时序预测方法[J].电子技术与软件工程,2019(05):163-165.  [14][1]苗旭东,魏连鑫.基于小波和PSO-BP神经网络的金融时序预测[J].信息技术,2018(05):26-29.DOI:10.13274/j.cnki.hdzj.2018.05.007.  [16] Bisoi R, Dash P K. A hybrid evolutionary dynamic neural network for stock market trend analysis and  prediction using unscented Kalman filter[J]. Applied Soft Computing Journal, 2014, 19(6):41-56.  [17] Bhanja S, Das A. Impact of Data Normalization on Deep Neural Network for Time SeriesForecasting[J]. 2018.  [18] Gupta A, Chaudhary D K, Choudhury T. Stock Prediction Using Functional Link Artificial NeuralNetwork (FLANN)[C]// 2017 3rd International Conference on Computational Intelligence and Networks (CINE). IEEE, 2018.  [19] Nelson D M Q, Pereira A C M, Oliveira R A D. Stock market's price movement prediction with LSTM neural networks[C]// 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, 2017.  [20] Tan Y, Shi Y, Tang Q. [Lecture Notes in Computer Science] Data Mining and Big Data Volume 10943| Deep Stock Ranker: A LSTM Neural Network Model for Stock Selection[J]. 2018,10.1007/978-3-319-93803-5(Chapter 58):614-623. | | | | | | | | | |
| 工作步骤与时间安排 | 2021年9月11日-2021年9月25日：阅读相关文献.  2021年9月26日-2021年-12月25日：学习python相关语法，numpy，pandas的简单使用机器学习、深度学习等理论与技术。  2022年1月10日前完成开题报告并完成答辩。  2022年1月10日-2022年2月1日 获取时序数据的预选数据集，并进行数据的剔除以及训练集，验证集，测试集的分割。并撰写关于时序数据获取和处理部分的毕业论文。  2022年2月1日-2022年2月20日 编写代码，构建基于BP神经网络的预测模型，进行训练并调参优化。并撰写关于BP神经网络预测模型构建的毕业论文。  2022年2月21日-2022年3月10日 编写代码，构建LSTM神经网络模型，并进行训练并调参优化。并撰写关于长短期记忆神经网络构建的毕业论文。  2022年3月10日-2011年3月15 测试模型的有效性并完成修改，进而完成关于模型对比的毕业论文  2022年3月15日-2022年3月25日 完成论文初稿撰写  2022年3月25日-2022年4月15日 继续完善代码，并完成毕业论文的终稿。  2022年4月15日-2022年5月26日 提交毕业论文并完成毕业答辩 | | | | | | | | | |
| 开题答辩评语 | （从选题、理论与实证准备、研究（设计）方法、工作安排等方面给出评价，并提出指导意见）  指导教师签名：  年 月 日 | | | | | | | | | |

注：此表由学生填写后交指导教师签署意见，并交院系教务办保存，否则不得开题；此表将作为毕业设计最终评分的依据。