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 | | 2019117160 | | 姓名 | 詹超 | | 年级 | 2019级 | | | |
| 专业 | | 计算机科学与技术 | | | | | | | | | |
| 论文（设计）题 目 | | 基于强化学习的时序大数据分析方法研究与开发 | | | | | | | | | |
| 指导教师  姓 名 | | 秦宝生 | | 专业技术职务 | | |  | 开题报告日期 | | |  |
| 企业导师  姓 名 | | 吴江 | 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**   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。其代表算法有专家系统、深度学习、人工神经网络、决策树、强化学习等。它根据算法来解析数据，进行学习，进而用于对真实世界做出预测和决策，随着发展被广泛应用于医疗健康、无人驾驶、游戏娱乐、AI智慧助手等领域[1]。  强化学习是一种交互式学习，其基本思想是智能体学习如何将环境状态映射到动作，目的是在交互过程中获取最大累积奖赏值，允许从经验中学习，侧重于学习解决问题的策略，是一种典型的奖惩式学习。强化学习通常分为基于值函数和基于策略的强化学习，基于值函数方法主要为Q-learning算法，基于策略方法主要为循环强化学习算法[5]。  Watkins(1992)最早提出了广为人知的Q-learning算法[7]，算法的主要思想是根据处理过程中新状态下新动作的价值，不断迭代更新Q函数的状态-动作值，实现函数最大化，从而提升预测准确度和实现优化。Q-learning算法具体流程如下：    Q-learning算法采用一个Q-tabel来记录每个状态下的动作值，当状态空间或动作空间较大时，需要的存储空间也会较大。如果状态空间或动作空间连续，则该算法无法使用。因此，Q-learning算法只能用于解决离散低维状态空间和动作空间类问题。  为了解决状态空间或动作空间连续，Q-learning算法无法使用的问题，之后的学者将将深度学习与强化学习相结合，综合深度学习的环境感知和特征提取能力和强化学习动态决策领域的能力[13]，可以为复杂系统提供感知决策的自主学习能力，这种学习方法也被称为深度强化学习，主要包含深度Q网络（DQN）算法及其改进算法，深度确定性策略梯度（DDPG）、信赖域策略优化（TRPO）等算法[14]。  Mnih最早对强化学习Q-learning算法结合深度神经网络进行了改进[15]，通过引入单独的Q函数网络，即用神经网络前向传播的输出替代原本Q-learning算法Q表中存储的状态-动作值解决了传统Q-Learing中面对过于复杂的状态和动作时难以维护Q表逐个进行更新的问题。  随着近年来信息产业的高速发展，以及计算机处理能力的不断提升，数字化已经进入越来越多的领域，同时为金融市场利用计算机进行交易也带来了更多可能性。在此基础上，众多学者尝试将计算机理论研究内容应用于量化投资领域，利用深度学习、强化学习和深度强化学习的算法、理论为金融领域的海量时序数据进行数据分析、交易策略构建和应用成为研究热点，  量化投资具有科学性、纪律性、系统性、及时性和准确性等特点，对于具有高维、非线性、高噪声、随机性的金融时序数据分析[5]，强化学习可以通过不断“试错”的学习方式与环境进行交互，通过多个适合的决策来达到期望值目标，同时利用环境反馈的奖励值与环境进行交互，具有在线学习、自动探索和自动持续学习产生决策的能力，符合量化投资特点，深度强化学习利用神经网络结构获得更强的环境感知和表征能力，可以从高维数据中学习控制策略，进而利用强化学习方法做出决策并调整，为构建量化投资策略提供了可行的方法。  本文利用强化学习方法设计股票趋势预测模型，构建基于Q-Learning算法的模型，构建基于深度神经网络的DQN算法趋势预测模型，提高模型预测准确率和决策速度。通过趋势预测模型，实现构建量化投资策略，进而提高回报率，降低投资风险。   1. **理论与实证准备。**   Q-learning算法采用一个Q-tabel来记录每个状态下的动作值，当状态空间或动作空间较大时，需要的存储空间也会较大。如果状态空间或动作空间连续，则该算法无法使用。因此，Q-learning算法只能用于解决离散低维状态空间和动作空间类问题。DQN算法的核心就是用一个人工神经网络来代替Q-tabel，即动作价值函数。网络的输入为状态信息，输出为每个动作的价值，因此DQN算法可以用来解决连续状态空间和离散动作空间问题，无法解决连续动作空间类问题。  Watkins（1992）[7]提出了QLearning算法，算法的主要思想是根据处理过程中新状态下新动作的价值，不断迭代更新Q函数的状态-动作值，实现函数最大化，从而提升预测准确度和实现优化。  Lee（2007）[8]将多个Q-Learning智能体结合，让多智能体分工合作参与股票价格预测和交易选择，在韩国的KOSPI200股指上进行自动交易测试，结果得到较好的收益率，在风险管理也取得不错效果。  Moody（1998）[9]提出了循环强化学习（Recurrent Reinforcement Learning，RRL）并将RRL算法应用于训练交易策略和投资组合管理当中，优化了模型的目标函数，并在标普500和部分美股中进行测试，、在交易次数少于传统的Q-learning策略的同时有着更高的回报。  Gold（2003）[10]提出将RRL模型中的多层神经网络替换为单层神经网络，并在25个外汇交易市场上进行自动交易测试，测试结果表明发现多层RRL的效果更差，说明了多层RRL没有显著提升，实验中学者们没有采用正则化和Dropout等方法，实验出现过拟合的现象。  Almahdi（2017）[11]提出了基于RRL模型的多目标函数自动交易算法，可以根据市场风格自适应选用交易函数，实验结果表明该模型显著由于单一目标函数RRL模型。  国内学者梁天新（2019）[12]通过对强化学习在量化交易策略的研究，认为经典RRL模型将持续发展,但是RRL基于循环的自适应框架不会改变，目标函数的采用会更加多样化,时序特征提取方式将更多选择深度学习模型  Deng（2016）[16]使用深度神经网络(DNN)、循环强化学习模型(RRL)和模糊学习(fuzzy learning)的交易策略算法(FRDNN)，应用模糊学习来减少数据中的不确定性，并使用DNN来利用数据的高维度非线性特征，最后利用RRL算法来进行交易行为选择，通过在中国的股票指数和期货市场上进行测试，FRDNN模型的收益显著高于RRL模型。  齐岳（2018）[17]建立了改进的深度确定性策略梯度DDPG算法应用于投资组合管理，对选取的16只中证100指数成分股作为风险资产进行实验。结果表明该算法的收益率明显高于其他对照组，验证了算法的有效性。  韩道岐（2020）[18]建立了基于双价值网络深度强化学习（DDQN）的智能股市交易模型，并利用中国股市1400多支股票的长周期数据进行多组对比实验，结果表明该模型具有发现市场价格变化，适应市场环境，做出正确操作策略的能力。  梁少文（2021）[19]将Q-Learning中不符合现实的最佳行为映射到相近的交易策略中，建立了LSTM-DNN混合模型降低数据维度，通过DQN算法来识别最佳交易操作，得到了比基准策略更好的收益率，缺点是未降低风险，同时处理大数据时对算力要求提高，对算法效率提出挑战。  王康（2021）[20]将双延迟深度确定性策略梯度算法（TD3）和策略梯度算法（PG）应用于中国市场龙头企业中成交量较大的5支股票和国债组成的资产包中。实验结果表明，该算法的回报率明显高于其他对照组，说明算法的有效性。   1. **拟解决的问题** 2. 现阶段我国金融市场时序数据分析与量化投资领域，深度学习与强化学习的理论与模型的应用已经展开，相对而言深度强化学习的应用很少，没有利用到深度学习的环境感知和强化学习的决策判断能力，对收益率和风险控制能力较差。 3. 基于深度强化学习DQN算法的模型大多都基于一个神经网络的算法，缺少基于多个不同神经网络和Q-learning算法的模型对照。 4. 所见文献多从单一数据维度开展研究工作，没有从多个数据维度去构建特征，研究视角单一，如训练数据只使用了日级数据，没有使用小时级数据，缺乏对照实验。 5. **研究方法与技术路线**   针对上述存在问题，本文会去解决上面三个问题，本系统会以多数据维度去分别构建特征作为深度强化学习模型的输入数据。采用基于MLP、CNN、LSTM的不同神经网络的DQN算法模型对数据进行分析预测，比对已证明有效的模型和在已存在模型基础上改变的模型。具体研究内容如下：  1）数据获取与预处理。获取每只股票的历史数据，在已经证明有效的特征构建基础上，再从多个视角去分析股票数据，并构建相应的特征，进行特征扩充。  2）分别构建基于MLP、CNN等不同神经网络的DQN算法股票数据分析模型。  3）将不同模型应用于多维度特征数据进行比对，得出实验结果  对于最终的预测结果，我将使用准确率和收益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符合回报预期和风险控制预期。 | | | | | | | | | | |
| 论文写作提纲 | （除题目外，具体到三级标题）：   1. 绪论    1. 研究背景及意义   1.1.1研究背景  1.1.2研究意义   * 1. 研究现状   2. 存在的问题   3. 研究内容及论文框架   1.4.1研究内容  1.4.2论文框架   * 1. 研究思路及研究方法   2. 本章小结  1. 深度强化学习理论基础   2.1深度神经网络  2.2马尔可夫决策过程  2.3 Q-learning算法   1. 强化学习的股票趋势预测模型研究   3.1数据处理与指标获取  3.1基于Q-learning算法的股票预测模型  3.2基于DQN算法的股票预测模型  3.3股票预测模型的实验结果与分析   1. ：总结   …  参考文献  致谢  参考文献：   1. 人工智能的发展与应用综述\_袁云佳 2. 朱俊璋. 基于深度学习的股票价格预测研究[D].哈尔滨工业大学,2019. 3. Zhang R, Yuan Z, Shao X. A New Combined CNN-RNN Model for Sector Stock Price Analysis[C]// 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC). IEEE Computer Society, 2018. 4. 张栗粽,王谨平,刘贵松,罗光春,卢国明.面向金融数据的神经网络时间序列预测模型[J].计算机应用研究,2018,35(9):2632-2637. 5. 程朝智. 基于深度学习的金融时间序列预测研究[D].电子科技大学,2021. 6. 刘全,翟建伟,章宗长,钟珊,周倩,章鹏,徐进.深度强化学习综述[J].计算机学报,2018,41(1):1-27. 7. Watkins C, Dayan P. Technical Note: Q-Learning[J]. Machine Learning, 1992, 8(3): 279-292. 8. Lee J W, Park J, Jangmin O, et al. A Multiagent Approach to Q-Learning for Daily Stock Trading[J]. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 2007, 37(6): 864-877. 9. Moody J, Saffell M. Reinforcement learning for trading[C]//Conference on Advances in Neural Information Processing Systems II. MIT Press, 1999. 917−923. 10. Gold C.FX trading via recurrent reinforcement learning[C]//IEEE International Conference on Computational Intelligence for Financial Engineering. IEEE, 2003. 363−370. 11. Almahdi S, Yang SY. An adaptive portfolio trading system:A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown[J]. Expert Systems with Applications, 2017, 87: 267−279. 12. 梁天新,杨小平,王良,韩镇远.基于强化学习的金融交易系统研究与发展[J].软件学报,2019,30(3):845-864. 13. 赵星宇,丁世飞.深度强化学习研究综述[J].计算机科学,2018,45(7):1-6. 14. 刘建伟,高峰,罗雄麟.基于值函数和策略梯度的深度强化学习综述[J].计算机学报,2019,42(6):1406-1438. 15. Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533. 16. Deng Y, Bao F, Kong Y, et al. Deep direct reinforcement learning for financial signal representation and trading[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(3): 653-664. 17. 齐岳,黄硕华.基于深度强化学习DDPG算法的投资组合管理[J].计算机与现代化,2018(5):93-99. 18. 韩道岐,张钧垚,周玉航,刘青.基于深度强化学习的股市操盘手模型研究[J].计算机工程与应用,2020,56(21):145-153. 19. 梁少文. 基于深度强化学习的中国股市投资组合交易策略[D].华中科技大学,2021. 20. 王康,白迪.基于深度强化学习的投资组合管理研究[J].现代计算机,2021(1):3-11. | | | | | | | | | | |
| 工作步骤与时间安排 | 2022年2月2日-2月7日：搜集强化学习相关文献，阅读学习并从中筛选参考文献  2023年2月8日-2月18日：学习python相关语法，学习numpy、pandas、matplotlib库的基本用法和强化学习、深度学习等理论与技术  2023年2月19日-2023年3月15日：完成开题报告  2023年3月16日-2023年4月15日：完成对Q-learning算法、DQN算法的研究，开发基于Q-learning算法、DQN算法的股票预测模型，并对模型进行测试，对实验结果进行分析，完成论文撰写。 | | | | | | | | | | |
| 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。