

**本科毕业设计（论文）开题报告**

|  |  |
| --- | --- |
| **论文题目** | **基于确定学习的量化投资系统** |

|  |  |
| --- | --- |
| **学 院：** | 自动化科学与工程学院 |
| **专 业：** | 自动化 |
| **学生姓名：** | 何哲宇 |
| **学生学号：** | 201436580399 |
| **指导教师：** | 董训德 |

|  |  |  |
| --- | --- | --- |
| 1. **课题背景及意义（含国内外研究现状综述）**  1.量化投资 量化投资，即通过数量化的方式及计算机程序化的方式发出买卖指令。相较于传统投资方式，具有系统性，纪律性的特点，依赖稳定的数学模型，极大程度减少人为因素造成的不稳定，从而达到优化交易策略的目的。而其核心便在于建模。  量化交易最早可以追溯到拉斯维加斯的赌桌上，MIT教授Edward Thorp早期在21点的赌桌上运用自己的系统赢得了不菲财富。再到1982年，美国富国集团成立资产轮动基金，是当时世界上第一个将量化投资理念直接应用的大型机构，但在当时投资效果并不理想也不为太多人接受。尽管如此，仍然有一部分人看到了在信息产业飞速发展的前提下，量化投资的巨大应用潜能 [1]。  到了二十世纪末，更复杂的模型被构想出来满足投资者们的要求，于是机器学习便被应用到这一领域来。其中Andrew Bradley在他的文章中便展示了基础的机器学习算法在评估股票的ROC曲线上的巨大潜能 [2]，例如最近邻算法（KNN），感知机，多层感知机等等。  近年来随着深度学习与神经网络的飞速发展，量化交易策略正逐渐走向一个更加明朗的未来，伴随着自动化，智能化的趋势弱化传统投资人的角色 [3]。由于本质上，量化研究更多的归属于数据科学范畴，所以量化投资的模型与派别也主要是根据不同的数据模型产生，这也是其科学性的证明。它的稳定性和鲁棒性随着回归分析与分类器在投资管理上的大放异彩被学界认可 [4]。虽然不及华尔街对于量化投资与量化分析的热捧，但国内的公司近年来也开始拥抱量化投资的概念，其中以长信基金公司为首的企业对量化投资的理解走在了前端，但值得一提的是，量化交易在我国依旧不能得到广泛推广，其原因在于从交易所到经纪商，交易系统大都没有完全支持程序化交易，高频数据库也不完善（一些券商与交易所直联，可以实现程序化交易，但也多限于自营，机构客户很难获得这些）。这些都导致量化交易很难被大规模执行。所以从目前的形势来看，当下量化交易在中国的发展依旧是不成熟的。 2.确定学习 确定学习 [5] 作为模式识别的一个子类，具有对非线性系统敏感的优良特性，所以运用它作为量化投资的基础是值得探索的。可以大胆预测的是，一家值得投资的优秀的公司从长远来看一定是处于一个相对稳定的状态，所以在较短时间内，其一些特征参数应该维持在一定范围内，而表现差劲的公司的相关参数则应处于波动或是处于较低一级的区间内。在这个基础上，选择稳定的，更直接关系到公司经营情况的参数就尤为重要。  确定学习其基本要素包括：1）使用径向基函数神经网络；2）对周期（或者回归）状态轨迹满足部分持续激励条件；3）在周期（或回归）轨迹的领域内实现对非线性系统动态的局部准确神经网络逼近（局部准确建模）；4）所学的知识以时不变且空间分布的方式表达、以常值神经网络权值的方式存储，并可在动态环境下用于动态模式的快速检测。[6]  到目前为止，确定学习已经应用在心电图识别上，通过使用病例数据训练，得到的模型对于病人的心脏健康状况可以做出值得参考的判断 [7]。 3.基于确定学习的量化投资策略 由此来看，理论上来说只要挑选合适的数据进行训练，由确定学习得到的模型是有能力对于一个未知非线性系统（一只股票在一定时间内的公开财务数据）进行识别的，因为股票及其母公司的财务数据是符合一定经济规律的，在这前提下其不会受到太多无关变量的干扰，从而可以作为一个重要判据来进行投资与否。这样的判据其实已用于传统的基于专家个人能力的投资模式，可以遇见的是，倘若应用程序化的方法，会使得人为干扰的因素进一步减弱，实现极大程度的做出符合经济规律与投资逻辑的决策。  总的来说这个领域尚未开垦完全，需要更多的探索与发现，但数据化研究的观念还是逐渐被业界所接受。许多私募公司纷纷开发了自己的量化与半量化产品 [8]。  基于确定学习在心电图识别上的表现，可以如法炮制。只要取得合适的财务数据作为训练数据便可作为界定上市公司运营情况的依据从而实现指导投资的目的，这便是本课题的研究的初衷 | | |
| 1. **课题研究主要内容及研究基础**   课题的主要内容分为两大步骤，一部分是量化模型的建立，一部分是编程实现基于模型数据的决策。  已知的量化模型有很多种，例如Rodrigo Naranjo的基于k线模糊建模方法 [9]。  本课题设想的建模主要依赖于Greenblatt J Tobias所著的《一本可以打败市场的小书》 [10]，其中提到的EPS（市盈率），P/B（市净率）可以作为优质股票的判断依据。但原书中提到的投资策略仅仅是将eps和p/b的排名作为直接的判据，然后对他们采用等权重的处理办法。这样的决策方式可以看出是十分的简单的，不难想到单纯依赖eps和p/b的排名公式即stock(value)=rank(eps)+rank(p/b) 可能还有进步的空间。  课题主要是为了探讨将两者结合后，通过确定学习将优劣股票通过EPS与P/B的数值来区分出来从而得出更好的投资策略。目前的设想是通过结合EPS和P/B数据作为一个数据对，然后通过得到限定时间间隔内的一系列数据对可以得到这段时间的曲线，然后借由该股票在这一段时间内的曲线变化通过确定学习来训练模型，从而使得训练出来的模型达到区分优质股票和较低劣股票的目的 | | |
| 1. **研究（或调研）方案和思路（技术路线）** 2. 首先获得NASDAQ公开的排名前列的以及排名中段的以及排名靠后的公司的EPS和P/B的数据（目前采用的是第三方API获取JSON数据） 3. 接着将数据重新封装成元组的格式，如{apple company (eps1,p/b1) (eps2,p/b2)…}，然后再传至确定学习的RBF神经网络中进行训练 4. 通过训练后得到的模块具有在给定一只股票的eps ，p/b数据的情况下，判断它是否可以被归类为值得投资的优秀股票 5. 目前取到的数据是离爬取数据时间最近的四个quarter公布的eps与p/b数据，这样对每一只股票就可以得到四个二维坐标，作为RBF神经网络的输入，来进行训练。用NASDAQ上的排名将前100，100-200，201-300，作为优质股票，中段股票，较低劣股票的分类。提取他们的（eps，p/b）数据 6. 代码管理采用GitHub，使用python3来进行原始数据获取的编程， | | |
| 1. **论文框架结构** 2. 综述研究现状   1.1 国外研究现状  1.2 国内研究现状   1. 算法的陈述，使用的平台，编程语言说明    1. 确定学习算法    2. Eps（市盈率），p/b（市净率）的介绍与重新建模    3. Python 与MATLAB 以及 IEX 股票数据 API的介绍 2. 数据建模的过程（包括训练数据的选取与确定学习训练的过程）    1. 原始数据处理    2. 训练过程    3. 检验模型（运用公开数据集） 3. 总结实验结果，提出问题 4. 致谢 | | |
| 1. **参考文献**   [1] “CAPITAL INVESTMENT APPRAISAL TECHNIQUES: A SURVEY OF CURRENT USAGE - Sangster - 1993 - Journal of Business Finance & Accounting - Wiley Online Library.” [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1468-5957.1993.tb00258.x/full. [Accessed: 07-Mar-2018].  [2] A. P. Bradley, “The use of the area under the ROC curve in the evaluation of machine learning algorithms,” *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, Jul. 1997.  [3] “Optimisation and quantitative investment management | SpringerLink.” [Online]. Available: https://link.springer.com/article/10.1057/palgrave.jam.2240205. [Accessed: 07-Mar-2018].  [4] G. Bassett, G. Gerber, and P. Rocco, “Robust Strategies for Quantitative Investment Management,” in *Theory and Applications of Recent Robust Methods*, Birkhäuser, Basel, 2004, pp. 27–37.  [5] C. Wang, D. J. Hill, and G. Chen, “Deterministic learning of nonlinear dynamical systems,” in *Proceedings of the 2003 IEEE International Symposium on Intelligent Control*, 2003, pp. 87–92.  [6] C. Wang and D. J. Hill, “Deterministic Learning and Rapid Dynamical Pattern Recognition,” *IEEE Trans. Neural Netw.*, vol. 18, no. 3, pp. 617–630, May 2007.  [7] X. Dong, C. Wang, and W. Si, “ECG beat classification via deterministic learning,” *Neurocomputing*, vol. 240, pp. 1–12, May 2017.  [8] “中国量化投资人才现状\_搜狐财经\_搜狐网.” [Online]. Available: http://www.sohu.com/a/132145930\_494939. [Accessed: 07-Mar-2018].  [9] R. Naranjo, J. Arroyo, and M. Santos, “Fuzzy modeling of stock trading with fuzzy candlesticks,” *Expert Syst. Appl.*, vol. 93, pp. 15–27, Mar. 2018.  [10] J. Greenblatt, *The Little Book That Still Beats the Market*. John Wiley & Sons, 2010. | | |
| 1. **工作进度安排** | | |
| **序号** | **设计（论文）各阶段任务** | **时间安排** |
| 1 | 获得封装好的训练数据， | 3.3-4.1 |
| 2 | 完成模型的框架工作与参数的调节 | 4.2-4.30 |
| 3 | 论文撰写 | 5.1-5.30 |
| 4 | 初稿提交 | 5.31 |
|  |  |  |
|  |  |  |
|  |  |  |