



Top Quant
CHRD前海智库
CHINA RESEARCH & DEVELOPMENT FRONTIER CENTER
以高质量研究 助力国家决策



零起点 TensorFlow 与量化交易

何海群 著



Win Or Out

一本让你轻松了解神经网络与量化分析的实战教程
图书+开发平台+成套的教学案例，系统讲解、逐步深入



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>



零起点 TensorFlow 与量化交易

何海群 著

电子工业出版社

Publishing House of Electronics Industry

北京•BEIJING

内 容 简 介

Python 量化回溯、TensorFlow、PyTorch、MXNet 深度学习平台以及神经网络模型，都是近年来兴起的前沿科技项目，相关理论、平台、工具目前尚处于摸索阶段。

TensorFlow 是近年来影响最大的神经网络、深度学习平台，本书从入门者的角度，对 TensorFlow 进行了介绍，书中通过大量的实际案例，让初学者快速掌握神经网络和金融量化分析的基本编程，为进一步学习奠定扎实的基础。

本书中的案例、程序以教学为主，且进行了高度简化，以便读者能够快速理解相关内容，用最短的时间了解 Python 量化回溯的整个流程，以及数据分析、机器学习、神经网络的应用。

本书仅仅作作为入门课程，具体的实盘策略，有待广大读者通过进一步深入学习 TensorFlow、PyTorch 等新一代深度学习平台来获得。最重要的是，广大的一线实盘操作人员需要结合专业的金融操盘经验，与各种神经网络模型融会贯通，构建更加符合金融量化实际应用的神经网络模型，从而获得更好的投资回报收益。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

零起点 TensorFlow 与量化交易 / 何海群著. —北京：电子工业出版社，2018.4

（金融科技丛书）

ISBN 978-7-121-33584-6

I. ①零… II. ①何… III. ①人工智能—算法—研究 IV. ①TP18

中国版本图书馆 CIP 数据核字（2018）第 019875 号

策划编辑：黄爱萍

责任编辑：葛 娜

印 刷：北京京科印刷有限公司

装 订：三河市良远印务有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编：100036

开 本：787×980 1/16 印张：27.5 字数：506 千字

版 次：2018 年 4 月第 1 版

印 次：2018 年 4 月第 1 次印刷

定 价：99.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：（010）51260888-819，faq@phei.com.cn。

推 荐 序

AlphaGo 与柯洁的黑白大战，因为对阵的一方是中国顶级围棋高手柯洁，所以引起国人的高度关注。利用百度搜索引擎输入 AlphaGo，一度可以得出 7000 多万条搜索结果，这远远高于其他热门词条。

事实上，AlphaGo 只是 Google 拥有的两套人工智能系统中的一套。它是 Google 2014 年收购的 DeepMind 的人工智能系统，专注于棋赛开发。Google 的另外一套人工智能系统就是本书介绍的 TensorFlow 系统。

在 TensorFlow 等人工智能系统出现之前，计算机所做的事情往往是简单重复的。计算机会按照人类编好的既定程序，简单重复、按部就班地运行，没有超越人类事先为其设定的思维边界。

计算机与人类的大脑相比，根本的区别在于不具备学习和创新能力。

计算机顶多也就是记忆的信息多，重复计算的速度快，不受情绪的影响等。但是，在 TensorFlow 等人工智能系统出现之后，计算机所做的事情除简单重复运行之外，更重要的是其具备了一定的自我学习和创新能力。

TensorFlow 等人工智能系统使得计算机在一定程度上能够自主学习，自我提高，总结过去的经验，汲取以往的教训，具备一定的创新性。这一点在 AlphaGo 与柯洁对垒的 3 场棋局的结果中不难看出。

这正是以 AlphaGo 和 TensorFlow 为代表的人工智能系统区别于以往任何计算机技术的关键所在，也是 TensorFlow 被称为互联网以来唯一的“黑科技”项目的原因。

具备了一定的自我学习和创造能力的人工智能系统的出现，将对经济系统的各个领域产生重大影响。笔者有着超过 20 年境内外金融行业从业经历，将从一个侧面分享人工智能对金融领域的影响。

从整个金融业的历史沿革来看，这大致经历了 4 个阶段：纯人工阶段、单机电脑阶段、互联网（含移动互联网）阶段和人工智能阶段。

随着每个阶段的渐次演进，提供金融服务一方的人力成本投入在逐渐减少，提供金融服务的效率在提高；对于接受金融服务的一方来说，金融服务的可获得性，以及便捷程度在逐渐增加，金融服务越来越围绕着人进行，以人为中心的全方位的社会经济服务体系正在形成。

在金融服务体系中，银行服务、证券服务、保险服务等内部界限开始变得模糊，金融服务与其他非金融的社会经济服务之间的界限开始变得不清。

特别是金融业进入人工智能阶段之后，人工智能系统将接受金融服务一方的身份特征数据、交易数据和行为数据等大数据，进行实时分析和动态跟踪，以远低于人工成本的成本，为每个人建立一个基于生命周期的综合金融模型，对每个人未来的金融行为进行预测，自动为他们提供账户资金管理、货币兑换、证券买卖、保险购买、购房购车计划、旅行休闲、子女教育、养老规划等方面的金融建议和授权代理操作，并将模型预测结果与实际情况相比对，自主学习和修正模型，以便更加贴合接受金融服务一方的真实金融意图，使得人工智能模型的预测建议和人的实际金融行为无限接近。

由此人类将从日常繁杂的各种金融交易中解放出来，投身到更需要自己或自己更感兴趣的方面。

展望未来，人工智能的应用前景无限美好；探寻当下，人工智能在世界各地的各行各业方兴未艾。

千里之行，始于足下。何海群先生的《零起点 TensorFlow 与量化交易》是有志

于人工智能领域的 IT 人士的一块敲门砖和铺路石。

祝愿人工智能在华夏大地生根发芽，开花结果。

梁 忠

梁忠：中国人民大学财政金融系博士，曾任里昂证券 CLSA 分析员；瑞银证券 UBSS 董事，财富管理中国研究部主管；瑞士信贷（香港）有限公司中国研究部董事；瑞信方正证券执行董事，研究部主管，具有 20 年国际顶级金融机构从业经历。

前言

感谢梁忠先生在百忙之中为本书撰写序言。以 TensorFlow 为代表的神经网络，被视为自互联网以来唯一的“黑科技”，无远弗届，无分行业领域，对社会各界从上至下带来彻底的颠覆与革命。

梁忠先生作为非 IT 领域的学者、专家，从第三方角度，冷静地观察这场数字革命，同时向更多的大众介绍这场革命的火花，推动行业变革，功莫大焉。

随着类似于 Titanic 数据集案例、梵高画风等一系列，基于 TensorFlow 等神经网络、深度学习项目的不断涌现，未来的各个学科都会结合人工智能（AI），进行新的学术重组。

“Python 量化三部曲”

“Python 量化三部曲”包括：

- 《零起点 Python 大数据与量化交易》（入门课程）
- 《零起点机器学习与量化交易》（重点分析 SKLearn）
- 《零起点 TensorFlow 与量化交易》（重点分析 TensorFlow）

此外，还有几部补充作品：

- 《零起点 Python 足彩大数据与机器学习实盘分析》
- 《零起点 Python 机器学习快速入门》

- 《零起点 TensorFlow 快速入门》
- 《MXNet 神经网络与量化交易》
- 《Plotly 可视化数据分析》

本书是《零起点 TensorFlow 快速入门》的后续之作，原本是 TopQuant.vip 极宽量化培训课程高级班的教学课件，为了节省篇幅，删除了 Python 基础教程，以及 SKLearn、TensorFlow 等机器学习方面的入门内容。没有经验的读者，建议先阅读《零起点 Python 机器学习快速入门》《零起点 TensorFlow 快速入门》，再开始本书的学习，这样会收到事半功倍的效果。

本书是目前较好的 TensorFlow 神经网络与量化分析入门教程：

- 无需任何理论基础，全程采用 MBA 案例模式，懂 Excel 就可看懂本书。
- 独创的逆向式课件模式，结合 TensorBoard 可视化系统，案例、图表优先，层层剖析。
- 系统介绍 TensorFlow 在金融量化领域的具体应用，提供多组配套案例。
- 全套神经网络股票趋势预测、股票价格预测案例源码。
- TDS 金融数据集的创建与使用。
- 三位一体的课件模式：图书+开发平台+成套的教学案例，系统讲解，逐步深入。

本书采用独创的黑箱模式、MBA 案例教学机制，结合大量的经典案例，介绍 TensorFlow 系统和常用的深度学习算法、神经网络模型，以及它们在量化分析当中的具体应用。

进一步学习

读者如有兴趣可以进一步学习“Python 量化三部曲”的内容，以及《零起点 Python 足彩大数据与机器学习实盘分析》。

机器学习、人工智能、金融量化，它们的基本原理是相通的，本质上都是数据分析。对于“Python 量化三部曲”的读者而言，本书也有很大的价值，特别是对于

第一部入门课程的读者。

Python 量化回溯与 TensorFlow、PyTorch、MXNet 等神经网络深度学习平台，都是近年来兴起的科技前沿领域，有关的理论、平台、工具目前还处于摸索阶段。“Python 量化三部曲”图书和 TopQuant.vip 极宽智能量化系统，只是在这些领域的起步阶段，作为入门教程，抛砖引玉。

本书中的案例、程序以教学为主，进行了很多简化，以便大家能够快速理解相关内容，用最短的时间，了解 Python 量化回溯的整个流程，以及数据分析、机器学习、神经网络在这些领域的应用操作技巧。

神经网络、深度学习在量化实盘当中的应用，是目前全世界都在研究的顶尖课题，当前尚未有很好的模型与应用案例。

本书仅仅作为入门课程，具体的实盘策略，有待广大读者通过进一步深入学习 TensorFlow、PyTorch、MXNet 等新一代深度学习平台来获得。

最重要的是，还有待广大的一线实盘操作人员结合专业的金融操盘经验，与各种神经网络模型融会贯通，构建更加符合金融量化实际应用的神经网络模型，从而获得更好的投资回报。

网络资源

为避免版本冲突，建议本书的读者下载 zwPython 2018m1 版本的软件和最新版本的《零起点 TensorFlow 与量化交易》配套课件程序，作为配套学习课件。配套程序的下载地址是 <http://www.broadview.com.cn/33584>。

使用其他 Python 运行环境如 Linux、Mac 平台的读者，请尽量使用 Python 3.5 和 TensorFlow 1.1 版本，并自行安装所需的其它模块库。

此外，需要注意的是，读者在运行书中案例时得到的结果，可能与本书略有差别，甚至多次运行同一个案例的结果也会有所差异，这属于正常情况。因为 TensorFlow 等深度学习系统，内部都使用了随机数作为种子数，用于系统变量初始

化等操作，每次分析的起点或者中间参数都有所不同。

本书的案例程序已经做过优化处理，不需要 GPU 显卡，全部支持单 CPU 平台。不过，为了提高运行效率，笔者建议尽量使用 NVIDIA 公司最新一代的 GPU 显卡。

目前是大数据、人工智能+时代，在这样的时代，计算力=生产力。

与本书相关的网络资源如下。

- 网站: <http://www.TopQuant.vip> <http://www.ziwan.com>。
- 网盘地址: <http://pan.baidu.com/s/1jIg944u>。
- 极宽量化 QQ 群: 总群, 124134140; QQ 2 群, 650924099; QQ 3 群, 450853713。
- 技术 Blog: <http://blog.sina.com.cn/zbrow>。
- 字王 Git 项目总览: <https://github.com/ziwan-com/>, 包括: 字王 4k 云字库、zwPython、zwpy_lst。

与本书相关的程序和数据下载，请浏览网站: TopQuant.vip 极宽量化社区，在网站的“下载中心”有最新的程序和数据下载地址。

本书在 TopQuant.vip 极宽量化网站设有专栏，若对本书、人工智能和机器学习有任何建议，请在网站专栏或 QQ 群留言，我们会在第一时间进行反馈和答复。

TopQuant 极宽量化网站“资源中心”的网址：

<http://www.topquant.vip/?p=56>

<http://ziwan.com/>

致谢

本书的出版要特别感谢电子工业出版社的黄爱萍和葛娜编辑，感谢她们在选题策划和稿件整理方面做出的大量工作。

同时，在本书创作过程中，极宽开源量化团队和培训班的全体成员，提出了很多宝贵的意见，并对部分课件程序做了中文注解。

特别是吴娜、余勤、邢梦来、孙励、王硕几位成员，为极宽开源量化文库和

TopQuant 极宽量化开源软件编写了文档，并在团队成员管理方面做了大量工作，为他们的付出表示感谢。

何海群（字王）

TopQuant.vip 极宽量化开源组 · 创始人

2018 年 2 月 14 日

轻松注册成为博文视点社区用户（www.broadview.com.cn），扫码直达本书页面。

- **下载资源：**本书所提供的示例代码及资源文件均可在 [下载资源](#) 处下载。
- **提交勘误：**您对书中内容的修改意见可在 [提交勘误](#) 处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- **交流互动：**在页面下方 [读者评论](#) 处留下您的疑问或观点，与我们和其他读者一同学习交流。

页面入口：<http://www.broadview.com.cn/33584>。



目 录

第 1 章 TensorFlow 概述	1
1.1 TensorFlow 要点概括	2
1.2 TensorFlow 简化接口	2
1.3 Keras 简介	3
1.4 运行环境模块的安装	4
1.4.1 CUDA 运行环境的安装	4
案例 1-1: 重点模块版本测试	5
案例 1-2: GPU 开发环境测试	8
1.4.2 GPU 平台运行结果	9
第 2 章 无数据不量化 (上)	12
2.1 金融数据源	13
2.1.1 TopDat 金融数据集	14
2.1.2 量化分析与试错成本	15
2.2 OHLC 金融数据格式	16
案例 2-1: 金融数据格式	17
2.3 K 线图	18
案例 2-2: 绘制金融数据 K 线图	19
2.4 Tick 数据格式	22
案例 2-3: Tick 数据格式	23

2.4.1	Tick 数据与分时数据转换	25
	案例 2-4: 分时数据	25
2.4.2	resample 函数	26
2.4.3	分时数据	26
2.5	离线金融数据集	29
	案例 2-5: TopDat 金融数据集的日线数据	29
	案例 2-6: TopDat 金融数据集的 Tick 数据	31
2.6	TopDown 金融数据下载	33
	案例 2-7: 更新单一 A 股日线数据	34
	案例 2-8: 批量更新 A 股日线数据	37
2.6.1	Tick 数据与分时数据	40
	案例 2-9: 更新单一 A 股分时数据	40
	案例 2-10: 批量更新分时数据	43
2.6.2	Tick 数据与实时数据	45
	案例 2-11: 更新单一实时数据	45
	案例 2-12: 更新全部实时数据	48
第 3 章	无数据不量化 (下)	51
3.1	均值优先	51
	案例 3-1: 均值计算与价格曲线图	52
3.2	多因子策略和泛因子策略	54
3.2.1	多因子策略	54
3.2.2	泛因子策略	55
	案例 3-2: 均线因子	55
3.3	“25 日神定律”	59
	案例 3-3: 时间因子	61
	案例 3-4: 分时时间因子	63
3.4	TA-Lib 金融指标	66
3.5	TQ 智能量化回溯系统	70
3.6	全内存计算	70
	案例 3-5: 增强版指数索引	71

案例 3-6: AI 版索引数据库	73
3.7 股票池	77
案例 3-7: 股票池的使用	77
3.8 TQ_bar 全局变量类	81
案例 3-8: TQ_bar 初始化	82
案例 3-9: TQ 版本日线数据	85
3.9 大盘指数	87
案例 3-10: 指数日线数据	88
案例 3-11: TQ 版本指数 K 线图	89
案例 3-12: 个股和指数曲线对照图	92
3.10 TDS 金融数据集	96
案例 3-13: TDS 衍生数据	98
案例 3-14: TDS 金融数据集的制作	102
案例 3-15: TDS 金融数据集 2.0	105
案例 3-16: 读取 TDS 金融数据集	108
第 4 章 人工智能与趋势预测	112
4.1 TFLearn 简化接口	112
4.2 人工智能与统计关联度分析	113
4.3 关联分析函数 corr	113
4.3.1 Pearson 相关系数	114
4.3.2 Spearman 相关系数	114
4.3.3 Kendall 相关系数	115
4.4 open (开盘价) 关联性分析	115
案例 4-1: open 关联性分析	115
4.5 数值预测与趋势预测	118
4.5.1 数值预测	119
4.5.2 趋势预测	120
案例 4-2: ROC 计算	120
案例 4-3: ROC 与交易数据分类	123
4.6 $n+1$ 大盘指数预测	128

4.6.1	线性回归模型	128
案例 4-4:	上证指数 $n+1$ 的开盘价预测	129
案例 4-5:	预测数据评估	133
4.6.2	效果评估函数	136
4.6.3	常用的评测指标	138
4.7	$n+1$ 大盘指数趋势预测	139
案例 4-6:	涨跌趋势归一化分类	140
案例 4-7:	经典版涨跌趋势归一化分类	143
4.8	One-Hot	145
案例 4-8:	One-Hot 格式	146
4.9	DNN 模型	149
案例 4-9:	DNN 趋势预测	150
第 5 章	单层神经网络预测股价	156
5.1	Keras 简化接口	156
5.2	单层神经网络	158
案例 5-1:	单层神经网络模型	158
5.3	神经网络常用模块	168
案例 5-2:	可视化神经网络模型	170
案例 5-3:	模型读写	174
案例 5-4:	参数调优入门	177
第 6 章	MLP 与股价预测	182
6.1	MLP	182
案例 6-1:	MLP 价格预测模型	183
6.2	神经网络模型应用四大环节	189
案例 6-2:	MLP 模型评估	190
案例 6-3:	优化 MLP 价格预测模型	194
案例 6-4:	优化版 MLP 模型评估	197
第 7 章	RNN 与趋势预测	200
7.1	RNN	200

7.2	IRNN 与趋势预测	201
	案例 7-1: RNN 趋势预测模型	201
	案例 7-2: RNN 模型评估	209
	案例 7-3: RNN 趋势预测模型 2	211
	案例 7-4: RNN 模型 2 评估	214
第 8 章	LSTM 与量化分析	217
8.1	LSTM 模型	217
	8.1.1 数值预测	218
	案例 8-1: LSTM 价格预测模型	219
	案例 8-2: LSTM 价格预测模型评估	226
	8.1.2 趋势预测	230
	案例 8-3: LSTM 股价趋势预测模型	231
	案例 8-4: LSTM 趋势模型评估	239
8.2	LSTM 量化回溯分析	242
	8.2.1 构建模型	243
	案例 8-5: 构建模型	243
	8.2.2 数据整理	251
	案例 8-6: 数据整理	251
	8.2.3 回溯分析	262
	案例 8-7: 回溯分析	262
	8.2.4 专业回报分析	268
	案例 8-8: 量化交易回报分析	268
8.3	完整的 LSTM 量化分析程序	279
	案例 8-9: LSTM 量化分析程序	280
	8.3.1 数据整理	280
	8.3.2 量化回溯	284
	8.3.3 回报分析	285
	8.3.4 专业回报分析	288

第 9 章	日线数据回溯分析	293
9.1	数据整理	293
	案例 9-1: 数据更新	294
	案例 9-2: 数据整理	296
9.2	回溯分析	307
9.2.1	回溯主函数	307
9.2.2	交易信号	308
9.3	交易接口函数	309
	案例 9-3: 回溯分析	309
	案例 9-4: 多模式回溯分析	316
第 10 章	Tick 数据回溯分析	318
10.1	ffn 金融模块库	318
	案例 10-1: ffn 功能演示	318
	案例 10-2: 量化交易回报分析	330
	案例 10-3: 完整的量化分析程序	343
10.2	Tick 分时数据量化分析	357
	案例 10-4: Tick 分时量化分析程序	357
总结		371
附录 A	TensorFlow 1.1 函数接口变化	372
附录 B	神经网络常用算法模型	377
附录 C	机器学习常用算法模型	414