# 摘要

长短期记忆（LSTM）网络是一种最先进的序列学习技术。它们较少被应用于金融时间序列的预测，但本质上适合这一领域。 我们部署了LSTM网络来预测1992年至2015年标普500指数成分股的样本外方向性变动。在交易成本之前，LSTM网络的日回报率为0.46%，夏普比率为5.8，我们发现LSTM网络的表现优于无记忆分类方法，即ran-dom森林（RAF）、深度神经网（DNN）和逻辑回归分类器（LOG）。从1992年到2009年，相对于一般市场的超额收益是非常明显的，但到2010年，超额收益似乎已经被套利走了，LSTM的盈利能力在交易成本之后在零左右波动。我们进一步揭示了盈利能力的来源，从而揭示了人工神经网络工程的黑箱。具体来说，我们发现被选中交易的股票有一个共同的模式–它们表现出高波动性和短期反转的回报特征。利用这些发现，我们能够正式确定一个基于规则的短期反转策略，在交易成本之前收益率为0.23%。进一步的回归分析显示，与三个基准模型相比，LSTM的收益率对常见的系统性风险来源的暴露程度也很低。

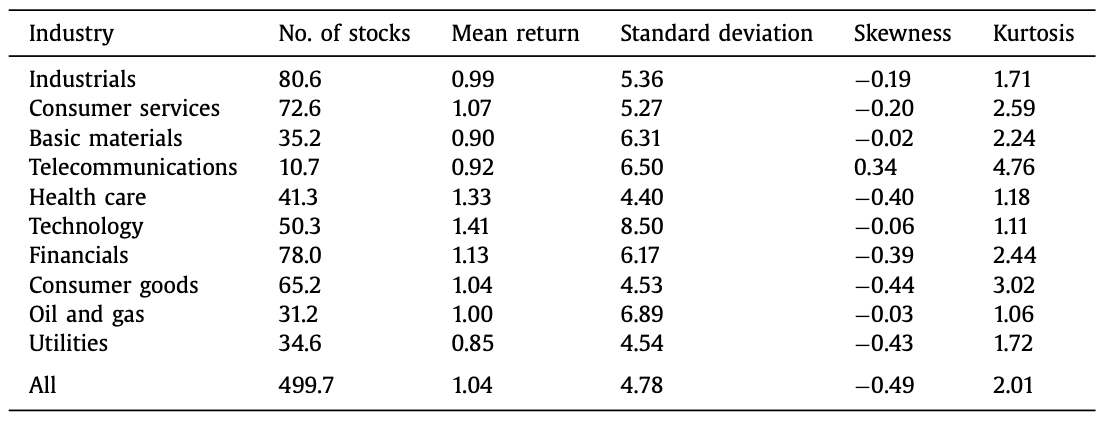
介绍 金融时间序列的预测任务是出了名的困难，主要是由高程度的噪音和普遍接受的半强形式的市场效率（Fama, 1970）所驱动。 然而，有大量众所周知的资本市场异常现象，与市场效率的概念形成鲜明对比。例如，Jacobs（2015）或Green，Hand和Zhang（2013）提供了包括100多个这样的资本市场异常现象的调查，这些异常现象有效地依靠回报预测信号来超越市场的表现。然而，用于建立这些回报预测信号，（特征）和未来回报（目标）之间关系的金融模型，通常是透明的，无法捕捉复杂的非线性依赖关系。 在过去的几年里，已经有初步证据表明，机器学习技术能够识别金融市场数据中的（非线性）结构，见Huck（2009，2010）。Takeuchi和Lee（2013），Moritz和Zimmermann（2014），Dixon，Klabjan和Bang（2015），以及Atsalakis和Valavanis（2009）以及Sermpinis，Theofilatos，Karathana- sopoulos，Georgopoulos和Dunis（2013）中进一步参考。具体来说，为了便于比较，我们对Krauss、Do和Huck（2017）最近在同一数据样本上的工作进行了扩展。作者使用深度学习、随机森林、梯度增强树和不同的en-sembles作为1992年至2015年所有标普500指数成分股的预测方法。一个关键的发现是，在交易成本之前，深度神经网络每天的回报率为0.33%，低于梯度增强树的0.37%和随机森林的0.43%。后一个事实令人惊讶，因为深度学习已经 “极大地提高了语音识别、视觉物体识别、物体检测和许多其他领域的先进水平”（LeCun, Bengio, and Hinton, 2015 , p. 436）。乍一看，我们会期望在时间序列预测领域有类似的改进。然而，Krauss等人（2017 ，第695页）指出，“神经网络是出了名的难以训练”，而且 “很可能在参数空间有配置，以进一步提高深度学习的性能”。 在本文中，我们主要关注深度学习，并进一步探索其在大规模时间序列预测问题上的潜力。在这方面，我们对文献做出了三个贡献。

* 首先，我们专注于长短时记忆（LSTM）网络，这是用于序列学习任务的最先进的深度学习架构之一，例如手写识别、语音识别或时间序列预测（Graves等人，2009；Graves，Mohamed，& Hinton，2013；Hochreiter & Schmidhuber，1997；Schmidhuber，2015）。令人惊讶的是，据我们所知，之前还没有人尝试将LSTM网络部署在一个大型的、流动性强的、无幸存者偏差的股票世界中，以评估其在大规模金融市场预测任务中的表现。 部分应用，如Xiong, Nichols, and Shen (2015)，主要是预测标准普尔500指数的波动性，预测小样本的外汇汇率（Giles, Lawrence, & Tsoi, 2001），或评估纳入特定公司新闻的影响（Siah and Myers (2016)）。我们填补了这一空白，将LSTM网络应用于1992年至2015年的所有标普500指数成分股。在此，我们对数据预处理以及LSTM网络在金融时间序列预测任务中的开发、训练和部署提供了深入指导。最后但并非最不重要的是，我们将我们的发现与文献中选定的基准进行了对比–随机森林（表现最好的基准）、标准深度神经网络（显示LSTM架构的增值）和标准逻辑回归（建立基线）。LSTM网络以每天0.46%的统计学和生态学意义上的回报胜过无记忆方法–相比之下，RAF为0.43%，标准的DNN为0.32%，逻辑回归为0.26%。这种相关优势在预测能力方面也是如此，Diebold-Mariano测试证实LSTM网络的预测能力优于应用基准。我们的发现在很大程度上对微观结构的影响是稳健的。 具体来说，当我们在成交量加权平均价格（VWAPs）而不是收盘价上实施LSTM策略时，我们看到盈利能力有所下降，但结果在统计和经济上仍然是显著的。对于周转率较低的每周执行也是如此–即使在信号发出后引入了一天的等待规则。只是到了2010年，LSTM的优势似乎已经被套利掉了，LSTM的盈利能力在交易成本之后在零左右波动，而RAF的盈利能力则变成了负值。
  + 第二，我们旨在揭示人工神经网络的黑匣子–从而揭开盈利的来源。一般来说，我们发现被选中交易的股票表现出高波动性，低于平均水平的势头，在交易前的最后几天有极端的方向性运动，并且在近期内有扭转这些极端运动的趋势。
  + 第三，我们将后半部分的研究结果综合为一个简化的、基于规则的交易策略，旨在抓住LSTM在选择赢家和输家股票时的模式的精髓。一个买入短期极值输家和卖出短期极值赢家的策略，在扣除交易成本之前，每天的回报率为0.23%–因此只有LSTM回报的50%左右。对系统性风险因素的回归分析揭示了在交易成本之前LSTM剩余的0.42%的阿尔法，与基准模型相比，对常见的系统性风险来源的暴露普遍较低。 本文的其余部分组织如下。 第2节简要介绍了数据样本、软件包和硬件。第3节深入讨论了我们的方法，即训练和交易集的生成，输入序列的构建，模型结构和训练，以及预测和交易步骤。第4节介绍了结果，并根据现有文献讨论了我们最相关的发现。最后，第5节是结论。

# 数据集，软件，硬件

## 2.1 数据集

对于实证应用，我们使用来自汤森路透的标普500指数成分股。为了消除幸存者偏差，我们首先从汤森路透公司获得1989年12月至2015年9月的标普500指数的所有月末成分股名单。我们将这些名单合并成一个二进制矩阵，表明该股票在随后的一个月里是否是指数成分股。因此，我们能够在1989年12月至2015年9月期间的任何特定时间点上近似地再现标普500指数。在第二步，对于在该时间段内曾经是指数成分股的所有股票，我们下载了从1990年1月到2015年10月的每日总回报指数。回报指数是累积分红的价格，并考虑到所有相关的公司行动和股票分割，使其成为回报计算的最充分的指标。根据Clegg和Krauss（2018），我们在表1中报告了按行业部门划分的平均年龄汇总统计数据。它们是基于每个行业的等权投资组合，每月生成，并限制在标准普尔500指数的成分中。

**表1** 从1990年1月到2015年10月，标准普尔500指数成分股的平均每月汇总统计数据，按行业划分。它们是基于全球行业分类标准代码所定义的每个行业的等权投资组合，按月形成，并限于标普500指数成分股。月度回报和标准偏差以百分比表示。 

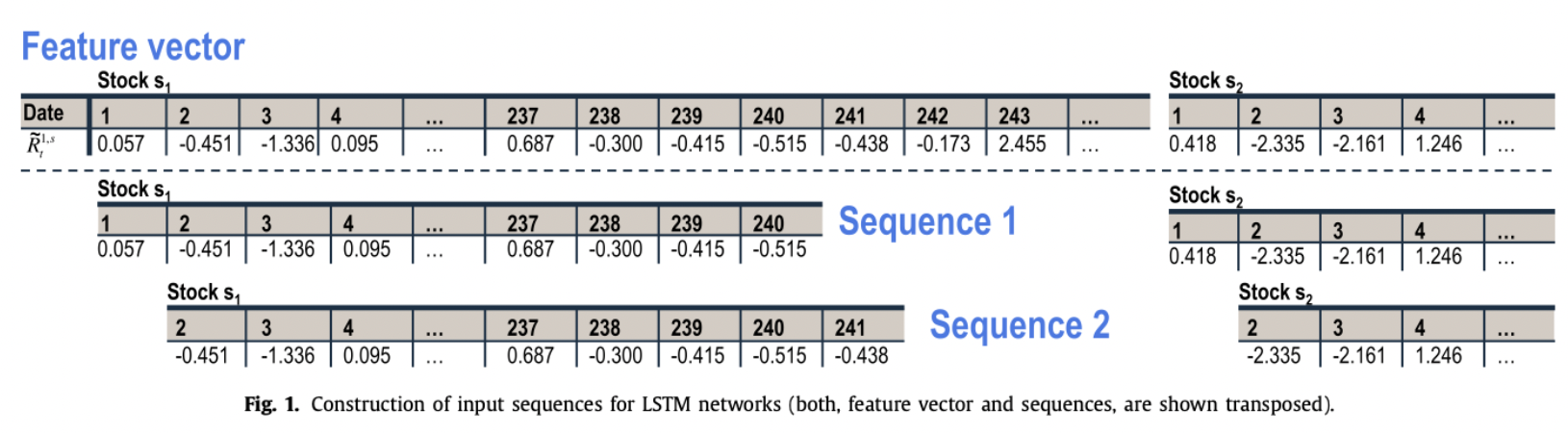
## 2.2 软件和硬件

数据准备和处理完全在Python 3.5（Python软件基金会，2016）中进行，依赖于numpy（Van Der Walt, Colbert, & Varoquaux, 2011）和pandas（McKinney, 2010）软件包。我们的深度学习LSTM网络是在谷歌TensorFlow的基础上用keras（Chollet，2016）开发的，TensorFlow是谷歌的一款软件。在异质系统上进行大规模机器学习的强大库（Abadi等人，2015）。此外，我们利用sci-kit learn（Pedregosa等人，2011年）建立了随机森林和逻辑回归模型，利用H2O（H2O，2016年）建立了标准的深度网。对于性能评估，我们使用R，一种用于统计计算的编程语言（R Core Team, 2016）和Peterson和Carl（2014）的Per- formanceAnalytics包。LSTM网络是在NVIDIA GPU上训练的，所有其他模型都是在CPU集群上训练的。

# 3 方法

我们的方法包括五个步骤。首先，我们将原始数据分为研究期，由训练集（用于样本内训练）和交易集（用于样本外预测）组成。第二，我们讨论训练和预测所需的特征空间和目标。第三，我们对LSTM网络进行了深入讨论。第四，我们简要介绍了随机森林、深度网和我们应用的逻辑回归模型。第五，我们设计了交易方法。本节的其余部分遵循上述的五步逻辑。 ## 3.1 训练和交易集的生成 按照Krauss等人（2017）的说法，我们将 “研究期”定义为训练-交易集，包括750天的训练期（约三年）和250天的交易期（约一年）。我们把从1990年到2015年的整个数据集分成23个研究期，交易期不重叠。 换句话说，研究期是由1000天组成的滚动块。训练是在前750天的滚动窗口中进行的，也就是说，通过在整个区块中滚动输入240天的回视（相当于最大特征长度，见第3.2节），然后提前一天进行预测。在过去的250天里，用训练好的参数进行交易，完全脱离了样本。然后后，轮胎系统向前滚动250天–导致总共有23个不重叠的交易期。 让表示在研究期i的训练期最后一天属于标准普尔500指数的股票数量，所以非常接近500。 对于训练集，我们考虑所有股票的历史。有些股票展示了完整的750天的训练历史，有些只是这个时间段的一部分，例如，当他们在较晚阶段上市时。交易集也是由所有只股票组成。如果一个成分股在交易期的某一天之后没有表现出价格数据，那么它直到那一天才会被考虑用于交易。 ## 3.2 特征和目标的生成 ### 3.2.1 特征 - LSTM网络的返回序列 让被定义为股票s在时间t的价格过程，，是股票s在m个时期内的简单回报，即

对于LSTM网络，我们首先计算每一天(m = 1)的回报率，并将它们堆积在一个大的特征向量V的维度为，其中表示研究期的天数。然后，我们通过减去平均数（）并除以从训练集得到的标准差（）来返回标准化后的值:

LSTM网络需要输入特征序列进行训练，即连续时间点的特征值。我们使用标准化的单日回报率作为唯一的特征。我们选择序列长度为240，因此包含了大约一个交易年度的信息。因此，我们按照以下方式生成重叠的240个连续标准化单日回报率R的序列，如果对于每个股票s和每个t≥240在研究期间内。对于第一支股票S1，序列包括标准化的单日回报率。第二个序列包括，依此类推。图1提供了一个示例。总体上，每个研究期间包含了大约380,000个这样的序列，其中大约255,000个用于样本内训练，大约125,000个用于样本外预测。 

### 3.2.2 目标

为了便于比较，我们效仿Takeuchi和Lee（2013），定义一个二元分类问题，即每个股票s和日期t的响应变量可以有两个不同的值。为了定义这两个类，我们将所有股票s在t+1期的单期回报按升序排列，并将其切分为两个同等大小的类。如果股票s的单期回报率小于t+1时期所有股票的横截面回报率中位数，那么0类就实现了。同样地，如果s的单期回报率大于或等于截面中位数，那么第1类就实现了。 LSTM网络的描述遵循Graves（2013）、olah（2015）和Chollet（2016）。在Karpathy（2015）和Britz（2015）中可以找到有价值的额外介绍，提供逐步的图形演练和代码片段。

LSTM网络属于循环神经网络（RNNs）类，即神经网络的”内部神经元连接拓扑至少包含一个循环”（Medsker, 2000，p. 82）。它们由Hochreiter和Schmidhuber（1997）提出，并在接下来的几年中得到进一步完善，例如Gers, Schmidhuber和Cummins（2000）和Graves和Schmidhuber（2005）等。LSTM网络专门设计用于学习长期依赖关系，并能够克服RNNs先前固有的问题，即梯度消失和梯度爆炸（Sak, Senior, & Beaufays, 2014）。

LSTM网络由输入层、一个或多个隐藏层和输出层组成。输入层的神经元数量等于解释变量（特征空间）的数量。输出层的神经元数量反映了输出空间，即在我们的案例中，两个神经元表示股票在t+1时期是否超过横向中位数。LSTM网络的主要特征包含在隐藏层中，由所谓的记忆单元组成。每个记忆单元都有三个维护并调整其单元状态的门：遗忘门（），输入门（）和输出门（）。记忆单元的结构如图2所示。

在每个时间步t，三个门都会接收输入（输入序列的一个元素）以及前一时间步的记忆单元输出。此时，门充当滤波器，每个滤波器都有不同的功能：

• 遗忘门定义从单元状态中移除的信息。 • 输入门指定添加到单元状态的信息。 • 输出门指定从单元状态中用作输出的信息。

### 3.4 基准模型 - 随机森林、深度神经网络和逻辑回归

为了对LSTM进行基准测试，我们选择随机森林，即一种强大而高性能的机器学习方法，一个标准的深度神经网络，即用于展示LSTM的优势，以及逻辑回归，即一个标准的分类器作为基线。请注意，随机森林、标准深度网络以及无记忆方法的特征生成遵循Krauss等人（2017年）中概述的规格。具体来说，我们使用累计收益Rm.s作为特征，其中m∈{61，…，20}U {40,60,…,2401}，见公式(1)，并且使用与第3.2.2节中定义的相同的目标。对于逻辑回归模型，我们按照公式(2)对收益进行标准化。在下文中，我们简要概述了如何校准基准测试方法。

随机决策森林的第一个算法是由Ho（1995）提出的，后来由Breiman（2001）扩展。简单地说，随机森林由许多不同的训练数据的自举样本组成的深度但不相关的决策树组成。随机森林算法使用了两个关键技术 - 随机特征选择以解相关树和bagging，以在不同的自举样本上构建它们。算法相当简单：对于委员会中的B个树中的每个树，从训练数据中抽取一个自举样本。在自举样本上开发决策树。在每个分割中，仅m个p特征的子集可用作潜在分割标准。当达到最大深度J时，生长停止。最终输出是B个树的委员会，并且进行多数投票分类。我们将树的数量B设置为1000，并将最大深度设置为J=20，允许大量高阶交互作用。对于分类，将随机特征选择保留为默认值m=Vp，请参见Pedregosa等人（2011）。

最终输出是由B个树组成的委员会，并且分类是通过多数投票来进行的。我们将树的数量B设置为1000，最大深度设置为J=20，允许大量高阶交互作用。对于分类，将随机特征选择保留为默认值m=Vp，请参见Pedregosa等人（2011）。

随机森林：我们使用随机森林作为基准有两个令人信服的原因。首先，它是一种先进的机器学习模型，几乎不需要调整，并且通常可以提供良好的结果。其次，在Krauss等人（2017年）中，随机森林在这种配置下是最好的单个技术，并且是Moritz和Zimmermann（2014年）的首选方法-一种在月度股票市场数据上的大规模机器学习应用。因此，随机森林作为任何创新型机器学习模型的强大基准。

深度神经网络：我们部署一个标准的DNN来展示LSTM网络的相对优势。具体来说，我们使用一个31个输入神经元的前馈神经网络，第一层有31个神经元，第二层有10个神经元，第三个隐藏层有5个神经元，输出层有2个神经元。激活函数是具有两个通道的maxout函数，遵循Goodfellow、Warde-Farley、Mirza、Courville和Bengio（2013年）的方法，在输出层使用softmax函数。我们将Dropout设置为0.5，使用收缩为o.00001的L1正则化-有关详细信息，请参见Krauss等人（2017年）。

逻辑回归：作为基线模型，我们还部署了逻辑回归。有关我们实现的详细信息，请参见sci-kit learn的文档（Pedregosa等人，2011）及其中的参考文献。最优的L2正则化是在0.0001和10,000之间的对数尺度上的100个选择中通过对相应的训练集进行5折交叉验证来确定的，使用L-BFGS来找到最优解，同时将最大迭代次数限制为100。逻辑回归作为基线，以便我们可以推导出相对于标准分类器，更加复杂和计算密集的LSTM网络的增量价值。

### 预测，排名，交易

针对所有模型，我们预测每只股票s在t+1时期相对于横截面中位数的表现，即超越或低于中位数的概率p，仅利用时间t之前的信息。然后，我们将每个时期t+1的所有股票按照这个概率值的降序排名。

排名最前面的股票是被认为在t+1期表现最佳的最被低估的股票。因此，我们选择买入每个排名前k只股票和卖出排名后k只股票，组成一个由2k只股票组成的多空组合，具体可见Huck (2009, 2010)的论文。

# 结果

我们的研究分为三个阶段。首先，我们按照Avellaneda和Lee（2010）的方法，分析5个基点每半轮的交易成本前后的收益，并对比ISTM网络与随机森林、深度神经网络和逻辑回归的表现。其次，我们在排名前k和后k的股票中提取共同的模式，揭示盈利的来源。第三，我们基于这些发现开发了一个简化的交易策略，并展示我们可以通过捕捉最明显的模式来实现LSTM的部分性能，使用一个透明的交易规则。

## 4. 表现回顾

4.1.1. 概述

首先，我们分析由2k股票组成的投资组合的特点，即我们做多的前k个股票和做空的后k个股票。我们选择k∈（10，50，100，150，200}，并在日均回报、年化标准差、年化夏普比率和准确性（在扣除交易成本之前）等方面比较新型LSTM与其他方法的表现。

我们发现以下趋势。无论投资组合大小k如何，LSTM相对于其他方法都表现出有利的特点。具体来说，在扣除交易成本之前的日回报率为0.46%，而RAF为0.43%，DNN为0.32%，LOG为0.26%（k=10时）。对于较大的投资组合大小，LSTM也实现了最高的日均回报，除了k=200时，它与RAF并列。

在标准差（一种风险度量）方面，LSTM与RAF处于相似水平，k=10时略低，投资组合规模增加时略高。在所有k级别上，LSTM和RAF的标准差都比DNN和逻辑回归低得多。夏普比率，即单位风险回报，在k=100时为LSTM最高，对于更大的投资组合，当RAF的较低标准差抵消了LSTM的较高回报时，略低于RAF。准确性，即正确分类的股份，是一个重要的机器学习指标。我们看到LSTM在k=10的投资组合中具有明显的优势，在k=100时具有轻微的优势，在规模增加时与RAF持平。

我们将后续分析集中在k=10的多空投资组合上。

4.1.2. 预测准确性详解

所采用的机器学习方法的关键任务是准确预测一只股票是否超过其横向中位数。在这一段中，我们将LSTM的预测准确性与其他方法以及随机猜测进行对比。此外，我们将LSTM的财务表现与100,000个随机生成的多空投资组合进行比较。

首先，我们采用Diebold和Mariano（1995）（DM）测试来评估方法i的预测准确性是否低于方法j的预测准确性的零假设，其中i, j∈（LSTM, RAF, DNN, LOG）且i≠j。对于每种方法的每个预测，如果k=10的投资组合中的个别股票被正确分类，我们分配一个0，否则分配一个1，并将这个分类错误的向量作为DM测试的输入。因此，我们总共考虑了k=10的投资组合中在总共5750个交易日内的5750 x 2 x k = 115,000个个别股票的预测。结果如表2的面板A所示。在第一行，对于LSTM预测低于RAF、DNN或LOG预测的零假设，我们分别得到p值0.0143、0.0037和0.0000。如果我们以5%的显著性水平进行测试，并对三个比较应用Bonferroni校正，调整后的显著性水平为1.67%，我们仍然可以拒绝LSTM预测比RAF、DNN或LOG预测准确性低的个别零假设。因此，可以假设LSTM预测优于其他被考虑方法的预测。类似地，我们可以拒绝RAF预测低于LOG预测的零假设以及DNN预测低于LOG预测的零假设。换句话说，复杂的机器学习方法的预测都优于标准逻辑回归分类器的预测。显然，前者能够捕捉到我们的金融时间序列数据中无法通过标准逻辑回归提取的复杂依赖关系。然而，从DM测试矩阵中，我们不能推断出RAF预测优于DNN预测，反之亦然——两种方法似乎具有相似的预测准确性。我们的关键发现是，尽管LSTM网络的计算成本明显更高，但在预测准确性方面，它是首选方法。 其次，我们使用Pesaran-Timmermann（PT）检验来评估每种预测方法的预测和响应是否独立分布的零假设。我们发现p值为零，小数点后面有四位数，这表明可以在任何合理的显著水平下拒绝零假设。换句话说，我们采用的每种机器学习方法都展现出具有统计学意义的预测准确性。

第三，我们提供了LSTM网络在随机情况下获得这些结果的统计估计。对于k=10，我们考虑了总共5750×10×2=115,000个优势和劣势股票，其中54.3％被正确分类。如果LSTM网络的真实准确率确实为50％，我们可以将正确分类的股票数量X建模为二项分布中的“成功”数量，即X ~ B（n = 115,000，p = 0.5，q = 0.5）。对于如此大的n，我们可以将x近似为正态分布，即X ~ N（μ = np，σ² = npq）。现在，我们可以轻松计算出在LSTM网络的真实准确率为50％的情况下，实现超过54.3％准确率的概率。我们使用Maechler（2016）的Rmpfr R包评估多精度浮点数并计算一个概率，即一个随机分类器仅通过偶然机会就能像LSTM一样表现的概率为2.7742e-187。

最后，我们按照Malkiel（2007）将猴子把飞镖扔在《华尔街日报》股票页面上的方式，评估了随机抽取的10万个投资组合的财务表现。对于每个交易日，我们随机抽取10个股票用于多头组合和10个股票用于空头组合，且不重复。所有这些组合在5750天内可以解释为由一只猴子选择的组合。然后，我们计算组合在这5750天内的平均每日回报率，以评估猴子的表现。100,000次复制的结果，即100,000只不同的猴子的结果，如图4所示。如预期的那样，在交易成本之前，我们看到平均每日回报率为零。更重要的是，即使是表现最好的“猴子”，其平均每日回报率为0.05％，也远远不及图3所示的应用模型的结果。 4.1.3. 关于财务表现的详细信息 表3提供了LSTM与基准之间的财务表现的见解，包括交易成本前后的情况。回报特征：在表3的A面板中，我们看到LSTM展现出有利的回报特征。交易成本前的平均回报为0.46％，交易成本后为0.26％，Newey-West t统计值在交易成本前为16.9336，在交易成本后为9.5792，与临界值1.9600（5％显著性水平）相比，统计显著。中位数只略小于平均回报，四分位数以及最小值和最大值表明结果并非由异常值引起。在交易成本后，正回报的占比高达55.74％，对于一个多空组合来说，这是一个惊人的高值。第二好的模型是RAF，交易成本后平均回报为0.23％，尽管标准差略高（0.0209 LSTM vs. 0.0215 RAF）。DNN排名第三，交易成本后每日平均回报为0.12％，仍然是显著的，与逻辑回归相比。最简单的模型实现了每日0.06％的平均回报，这已不再与零显著不同（Newey-West t统计值为1.6666，与临界值1.9600相比）。需要注意的是，与文献相比，LSTM表现强劲。Krauss等人（2017）的组合，由深度网络，梯度提升树和随机森林组成，在相同的时间框架，数据集和交易成本后，平均每天回报为0.25％。换句话说，单一的LSTM模型实现了比一个完整的组合略强的表现。

风险特征：在表3的B面板中，我们观察到风险特征方面的混合画面。就每日风险价值（VaR）而言，LSTM在RAF之后排名第二，其1％的VaR为-5.45％，而RAF为-4.95％。最冒险的策略来自于逻辑回归模型，在所有情况中1％的情况下超过-7.66％的损失，比一般市场上的买入持有投资风险高出两倍多。然而，LSTM的最低最大回撤为52.33％，相比其他模型和一般市场而言是最低的。

年化风险回报指标：在表3的C面板中，我们分析了年化基础上的风险回报指标。我们看到，LSTM在交易成本后实现了最高的年化回报率82.29％，相比RAF（67.87％），DNN（24.60％），LOG（7.11％）和一般市场（9.25％）更高。年化标准差处于第二低水平，为33.23％，相比所有基准都更低。夏普比率通过标准差缩放超额回报，因此可以被解释为金融中的信噪比，或者风险单位回报。我们看到，LSTM实现了最高水平的2.34，RAF排名第二，为1.87，而所有其他方法的夏普比率均远低于1.0。

从财务角度来看，我们有两个关键发现。首先，LSTM在回报特征和风险回报指标方面明显优于RAF。因此，我们能够证明选择LSTM网络-这些网络天生适合于时间序列预测任务-优于浅树基模型以及标准的深度学习模型。其次，我们证明了标准的逻辑回归无法从特征空间中捕捉到与LSTM相同水平的信息，即使我们执行样本内交叉验证以找到最佳的正则化值。

4.1.4. LSTM盈利能力随时间的关键回顾 在图5中，我们展示了策略的时间表现，即从1993年1月到2015年10月。我们关注最具竞争力的技术，即LSTM和随机森林。 1993/01-2000/12:这些早期的时期特点是表现强劲- LSTM在平均每日回报、夏普比率和准确性方面优于RAF，几乎在所有年份都是如此。每天1美元平均投资的累计支付额达到了LSTM超过11美元，RAF超过8美元。在考虑这种超额表现时，重要的是要注意，LSTM网络是在1997年引入的，并且只有在2000年代末GPU计算机的出现之后才能够被实际应用。因此，在90年代实现异常高的回报很可能是由于LSTM在那个时候要么是未知的，要么对大多数市场专业人士来说是完全不可行的。类似的论点也适用于随机森林。 2001/01-2009/12:第二个时期对应于温和时期。LSTM仍然能够在所有年份的交易成本后产生正回报，尽管与90年代相比要低得多。当考虑累计支付时，我们看到LSTM与随机森林相比的超额表现一直持续到金融危机。这些基于树的方法的一个关键优势是它们对噪声和异常值的稳健性-这在这样波动性大的时期表现出来。RAF在夏普比率高达6的情况下实现了异常高的回报和一致的准确性值。因此，LSTM的1美元投资总支付额为4美元，RAF的总支付额为5.6美元-RAF的大部分支付额是在金融危机期间实现的。 这段时间的温和阶段似乎是由这些策略在行业专业人士中逐渐扩散，从而逐渐侵蚀盈利能力所导致的。然而，对于RAF而言，2008/2009年的全球金融危机则构成了一个例外-RAF的盈利能力有了强烈的回升。根据文献，这些利润可能由两个因素驱动。首先，在金融动荡时期，认为投资者有可能“把树木看不见森林”（Jacobs和Weber，2015，第75页）-从而创造出相对价值套利机会，参见Clegg和Krauss（2018）。其次，在高波动时期，套利限制也非常高，因此很难捕捉到这样的相对价值套利机会。具体来说，对于难以借入的股票，卖空成本可能会上升-参见Gregoriou（2012），Engelberg，Reed和Ringgenberg（2017），或者在更严重的情况下，卖空可能会被完全禁止。但是，长期的投资也会受到影响，例如当扩大价差和降低流动性限制回报时。 第三个时期对应于恶化的时期。随机森林失去了优势，并且在平均每天1美元的投资基础上毁掉了1美元以上的价值。相比之下，LSTM在几乎所有年份中实现了更高的准确性得分，并且在交易成本之后能够使资本大致保持不变。与文献比较时，值得注意的是，Krauss等人（2017）提出的组合策略在2010年以后也出现了巨大的回撤-而单一模型的LSTM回报在零附近波动。尽管如此，所提出的机器学习方法的优势似乎已被套利消除。这种效应可以在许多针对定量策略的学术研究论文中观察到-例如Bogomolov（2013），Rad，Low和Faff（2016），Green，Hand和Zhang（2017），Clegg和Krauss（2018），等等。 由于LSTM识别结构的能力，我们接下来分析在k = 10组合中某些行业的潜在偏好。具体来说，我们考虑该行业在k = 10组合中的份额与该行业在当时标普500指数中的份额之间的差异。正值表示该行业被LSTM网络超配，负值表示该行业被低配。 图6展示了我们对最有趣的行业-石油和天然气、技术、金融和其他行业的发现。首先，我们看到在90年代末，技术股的重量有显著的超配，对应于不断增长的互联网泡沫和其崩溃。其次，我们观察到金融股在2008/2009年上升-对应于全球金融危机。第三，自2014年以来，石油和天然气股票的权重增加-随着最近的石油供过于求和原油价格的显著下跌而下降。 注意：附录提供了更详细的分类，显示每个GICS行业在前k和后k组合中相对超配。有趣的是，每个行业的超配充分捕捉了主要市场发展趋势。我们假设这种行为是由于不断增加的波动水平所驱动的，并在以下部分进一步阐述这一点。 在本节中，我们旨在揭示前十和后十股票中的共同模式，以破解人工神经网络等机器学习方法的黑匣子。首先，我们进行了一项非常简单但有效的分析。对于每一天，我们提取前十和后十股票的所有240天回报序列。然后，将所有5750 x 10个顶部和同样数量的下跌序列堆叠在一起。为更好地表示，我们将每个序列的240个回报累积到一个回报指数中，从t-240天的水平开始，然后平均这些回报指数序列。因此，我们获得了两个概括序列，包含前十和后十股票的模式。

结果如图7所示，并与所有股票（均值）的横截面行为进行对比。我们看到，前十和后十股票都表现出Jegadeesh和Titman（1993）意义下低于均值的动量，即从t-240日到t-10日表现不佳，与横截面相比。从t-9日到t（注意：预测是在t日进行的），前十股票开始以加速度崩盘，平均损失了之前230天获得的约50％。相反，后十股票在交易前的过去10天中表现出相反的模式，并呈现出越来越高的回报。看到ISTM提取了这样强烈的下跌股票和前十股票之间的共性，让人信服。

4.3. 盈利来源 简化交易策略：在本节中，我们基于之前的分析构建了一个简化的交易策略。从图7和图8中，我们可以看到最显著的特征是在交易前几天收益的暴跌或急剧上升。类似地，Krauss等人（2017）发现最后的收益是他们的机器学习模型中最重要的变量，Moritz和Zimmermann（2014）也是如此。为了简单起见，我们基于这个最显著的观点，松散地遵循短期反转异常的两位创造者Jegadeesh（1990）和Lehmann（1990）。具体而言，我们做多前k只在交易日前5天累积收益最负的股票，做空前k只在交易日前5天累积收益最正的股票 - 均等权重。对于这个非常简单但透明的策略，我们发现在交易成本之前的平均日收益率为0.23％，在交易成本之后为0.03％。因此，我们发现仅仅利用最突出的特征，即过去5天的累积收益，仅能产生LSTM策略在交易成本之前约50％的收益。在交易成本之后，短期反转策略失去了优势，与LSTM相反。因此，我们可以谨慎地得出结论，其他特征包含了额外的解释力量，形成更微妙的模式，不直接从图8的汇总视角中识别出来。基于这个发现，我们接下来将评估LSTM回报对系统风险的共同来源的暴露度。 对系统性风险来源的暴露： 正如在投资组合分析中通常进行的那样，与Krauss等人（2017）类似，我们评估多空投资组合对常见的系统性风险来源的暴露。我们使用Fama和French（1996）中的Fama-French三因素模型（FF3），并通过三个额外的因素来加强。 因此，它包括一个市场因子，一个衡量小市值股票减去大市值股票（SMB）的因子，以及一个衡量高市值股票减去低市值股票（HML）的因子。此外，我们还包括一个动量因子（Carhart，1997），一个短期反转因子（Gatev等人，2006）和一个长期反转因子。这些因子模型所使用的所有数据都是从Ken-neth French的网站上获得的。10 其基本思想是衡量对横向价格动态的额外暴露。我们的研究结果在表4中进行了总结。首先，我们观察到LSTM回报率在所有模型中表现出最低的决定系数0.0330。换句话说–每日LSTM原始收益的0.46%，只有一小部分可以被这种常见的系统性风险来源所解释。事实上，每天仍有0.42%的阿尔法（拦截）。相比之下，RAF的决定系数较高，为0.1338，每天只有0.34%的阿尔法再主（与0.43%的原始回报相比）。即使R2很低（在金融研究中经常会出现这种情况–见Gatev等人，2006）–我们仍然可以将因子载荷与我们的预期进行比较。LSTM基本上是市场中立的，这意味着多头和空头组合（我们在其中的投资额相等）表现出类似的β值。相比之下，其他策略在长线上似乎比短线上更倾向于高贝塔的股票。在所有的模型中，SMB因子通常是不显著的，接近于零。对HML的暴露在不同的策略中是不同的。 LSTM偏向于魅力股（负值和显著负载），而RAF倾向于选择价值股（正值和显著负载）；DNN和简单逻辑回归没有任何暴露。最有趣的是捕捉截面价格动态的因素，因为它们在很大程度上与我们所反馈的特征相一致。我们看到，所有的模型对动量都有正面的和统计上显著的暴露的短期反转因子–这意味着较短期的横截面价格动态被捕捉到。有趣的是，与LSTM相比，这些基于价格的因子负荷在对数回归中更占优势。负荷的增加可能是由于较简单的模型在结构上无法从提供的特征中提取高阶模式–鉴于我们没有提供任何互动。最后，长期恢复因子对LSTM来说是不明显的，而对其他策略来说有轻微的负值，尽管符号水平较低。我们认为这个因素的解释力普遍较低，因为我们要解决的是一个很短的交易期限。 简而言之，LSTM的回报几乎不能用系统风险源来解释–而且比我们部署的其他机器学习模型的回报要少得多

4.4. 对市场微观结构的稳健性

上一小节的一个关键结果是，基于LSTM的股票选择与典型的逆向投资策略有相似之处。众所周知，微观结构效应，特别是买卖反弹，会给这种逆向投资策略带来向上的偏差–见Conrad和Kaul（1989），Jegadeesh（1990），Jegadeesh和Titman（1995），Gatev等人（2006）。其根本原因很简单。LSTM策略倾向于买入最近的Iosers，卖出最近的赢家。在名列前茅的K股中，赢家的报价更有可能是卖出报价，而在落榜的K股中，输家的报价更有可能是买入报价–类似于Gatev等人（2006）中所提出的变通投资策略的道理。因此，在输家的情况下，我们会以买入价买入，而在赢家的情况下以卖出价卖出。 鉴于这些考虑，有必要对我们的研究结果的稳健性进行分析。 首先，较长的交易期限通常对米–结构性影响更为稳健。例如，Lehmann(1990)在他的短线研究中采用了这种方法。在他的机器学习策略中，我们引入了LSTM策略的每周变体–有效地弥补了低营业额（每周与每日）之间的差距。 我们引入了LSTM策略的每周变体–有效地弥补了较低营业额（每周与每日）之间的差距，并为LSTM保留了足够数量的训练实例，以成功地从数据中提取结构。后者排除了每月的变体，而没有明显改变研究设计。具体来说，我们设计的LSTM完全类似于第3节中概述的基石，但是以5天而不是1天为期限。12此外，我们以5天的预测期限实现了五个重叠的LSTM策略端口，以避免由起始日期引入的偏差（Jegadeesh & Titman, 1993）。 因此，每个投资组合在一个交易日内相互抵消，所有投资组合的收益率为平均数。请注意，我们每天对所有五个投资组合的投资资本进行市场对市场的记录，所以我们能够为了可比性而显示出每日的回报。 第二，按成交量加权平均价格（VwAPs）执行更现实，更不容易受到买卖反弹的影响（例如，见Lee, Chan, Faff, & Kalev, 2003）。因此，我们使用Quantouote提供的1998年1月至2015年10月所有标普500指数股票的分档交易数据来创建VWAPs，我们将其用于特征工程、目标创建和回测–见第3节。在23个持续时间为17分钟的仓位中，我们得到23个VWAPs。我们使用第22个VWAP作为创建特征的锚点，使用第23个VWAP来执行我们的交易，并完全按照第3.2节来创建特征。 第三，Gatev等人（2006）的一天等待规则是对买卖反弹的适当补救措施。具体来说，我们在信号产生后延迟一整个交易日来执行。这个规则只对周策略有意义。如果是日线策略，延迟覆盖整个预测范围，使预测无效。 第四，为了完整起见，我们在本文中使用了每半圈5个基点的交易成本–对于美国大盘股来说是一个相当稳健的数值。例如，Jha（2016）假设在类似的时间范围内，美国股票宇宙中最大的500只股票只有2个基点。 简而言之，我们对LSTM策略进行了以下稳健性检查。模型Mi是基线，显示了前几节所讨论的标准LSTM策略的结果，即每天的营业额和收盘时的执行，但限制在1998年至2015年的QuantQuote数据集的时间范围内。 模型M显示了在VAP而不是收盘价上执行该策略时的效果。模型Mg-Ms是基线策略的每周变量。因此，Ms按收盘价执行，Ma按VWAP执行，Ms按收盘价执行，有一天的等待规则。表5显示了这些变体的结果–每半圈5个基点的交易成本之前和之后。 在交易成本之后，基线LSTM策略M，在1998年至2015年较短的时间范围内，每天的平均收益为0.20%。在模型M的情况下，以VWAPs而不是以收盘价执行，会导致每天0.16个百分点的恶化–t统计学上仍然显著，而且经济上有意义，每年有43%的回报。即Ms（收盘时执行）为0.11%，M4（VWAP时执行）为0.09%，M5（收盘时执行，等待一天）为0.0%–所有这些都具有统计学和经济学上的意义，t统计数字高于3，年化收益率高于20%。表的其余部分与表3完全相似，显示了LSTM策略在稳健性检查后，相对于一般市场的表现。 然而，在最近几年，如图9所示，我们看到重新转为平缓，所以LSTM的优势似乎已经被仲裁掉了。尽管这样，在很长一段时间内，即从1998年到2009年，模型的表现是稳健的，特别是在金融危机前后，收益率上升的高峰。 # 5结论 　 在本文中，我们将长短期记忆网络应用于标普500指数的大规模金融市场预测任务，从1992年12月到2015年10月。通过我们的工作，我们对文献做出了三个关键的贡献：第一个贡献是关注LSTM网络在金融时间序列预测任务中的大规模实证应用。我们提供了一个深入的指导，密切关注整个数据科学价值链。具体来说，我们构建了一个适当的预测任务，以240天回报序列的形式推导出合理的特征，在预处理过程中对特征进行标准化以促进模型训练，讨论合适的ISTM架构和训练算法，并根据预测结果推导出交易策略，与现有文献一致。我们将LSTM的结果与随机森林、标准深度网以及简单的逻辑回归进行比较。我们发现LSTM这个方法本身就适合这个领域，它以非常明显的优势击败了标准的深度网和逻辑回归。 在大多数情况下–除了全球金融危机–随机森林的表现也超过了它。我们发现，在交易成本之前，每天有0.46%的统计学和经济学意义上的回报，这对半强势的市场道德水平提出了明确的挑战，并表明在2010年之前，深度学习可以成为这一领域的有效预测建模技术。考虑到市场摩擦，这些发现在很大程度上是稳健的，因为在对VWAPs而不是收盘价执行每日LSTM策略时，以及在信号产生后有一天等待期的每周LSTM策略运行时，盈利能力仍然具有统计学和经济学意义。我们的机器学习方法的效率提高了。截至2010年，市场在我们部署的机器学习方法方面表现出效率的提高，LSTM的盈利能力在零左右波动，而RAF的盈利能力则严格地降到了负值领域。 此外，本文所概述的关于LSTM网络的概念和经验方面的内容超出了纯粹的金融市场应用，而是作为其他研究人员的指南，希望将这种有效的方法部署到具有大量训练数据的其他时间序列预测任务中。 第二，我们分解了黑箱 “LSTM”，并揭开了为盈利交易而选择的股票的共同模式。我们发现ISTM组合由具有低于平均水平的动量、强烈的短期反转特征、高波动性和β值的股票组成。所有这些发现在某种程度上与现有的资本市场异常现象有关。令人印象深刻的是，其中一些是由LSTM从240个标准化的原始回报序列中独立提取的。这有待于未来的重新搜索，以确定LSTM神经网络工程从金融市场数据中学习到的更多、更微妙的模式，并验证这些模式在更精细、基于规则的交易策略中的盈利潜力。 第三，基于LSTM组合的共同模式，我们设计了一个简化的基于规则的交易策略。具体来说，我们做空短期赢家，买入短期输家，并持仓一天–就像LSTM应用中一样。通过这种透明的、简化的策略，在交易成本之前，我们每天取得0.23%的回报–大约是LSTM回报的50%。对系统风险的常见来源的进一步回归分析显示，在交易成本之前，LSTM的剩余阿尔法为0.42%，与我们部署的其他模型相比，风险暴露普遍较低。 总的来说，我们已经成功证明了LSTM网络能够有效地从嘈杂的金融时间序列数据中提取有意义的信息。与随机森林、标准深度网和逻辑回归相比，它是预测准确性和扣除交易成本后的每日重新转折方面的首选方法。事实证明，深度学习–以ISTM网络的形式–似乎也构成了这个领域的一个进步。