原书 GitHub存储库网址（原书的代码文件与本书赠送的代码文件一致）

https://github.com/PacktPublishing/Machine-Learning-for-Algorithmic-Trading-Second-Edition

在GitHub存储库中的本章的README文件中可以找到其他资源和引用的链接

P9

数据提供商 Yipit 赞助的网站 alternativedata.org 中专门罗列了各类数据提供商，截至 2018 年 12 月，该网站已经有超过 375 家另类数据提供商。

P31

AlgoSeek 分钟采样基于证券信息处理器（SIP）提供的数据，该处理器负责管理证券买卖汇总记录带

源数据。 你可以在 [https://www.algoseek.com/samples/ 查看该文档。](https://www.algoseek.com/samples/查看该文档。)

P34

**sp\_url = 'https://en.wikipedia.org/wiki/List\_of\_S%26P\_500\_companies'**

P38

彭博（Bloomberg）和汤森路透（Thomson Reuters）一直是领先的数据提供商，在价值 285 亿美

元的金融数据市场上的总份额超过 55%。**其他规模较小的竞争对手和后来者包括 FactSet、**

**money.net、Quandl、Trading Economics 和 Barchart 等。**

P40

**SEC\_URL = 'https://www.sec.gov/files/dera/data/financial-statement-and-notes-**

**data-sets/'**

P52

另类数据提供商的市场相当分散，这也反映了这一新行业的快速增长。摩根大通（J. P. Morgan）列

出了 500 多家专业数据公司，AlternativeData.org 则列出了 300 多家。数据提供商扮演的角色也各不相同，有咨询公司、集成商，也有的是提供技术解决方案。卖方支持各种格式的数据传送，如原始数据、半处理数据甚至是从一个或多个数据源提取的某种形式的信号。

P55

**url = https://www.opentable.com/new-york-restaurant-listings**

P57

**url = https://www.opentable.com/new-york-restaurant-listings**

P58

**start\_urls = ['https://www.opentable.com/new-york-restaurant-listings']**

**P59**

**SA\_URL = 'https://seekingalpha.com/'**

**P101**

scikit-learn (http://scikit-learn.org/stable/modules/generated/sklearn.covariance.

LedoitWolf.html)

P170

Marcos Lopez de Prado(http://www.quantresearch.info/) 发表了大量关于回测风险以及如何检测或避免风险的文章，包括回测–过度拟合（backtest-overfitting）在线模拟器 (http://datagrid.lbl.gov/backtest/, *Bailey, et al. 2015*)

P180

**已知问题**

Zipline需要国债曲线和依赖于IEX API的标准普尔500指数收益作为基准（https：//github.com/quantopian/zipline/issues/2480），IEX API需要注册才能获得密钥。当然，我们也可以简单地给Zipline打个补丁，绕过这个问题从美联储（Federal Reserve）那里下载数据，GitHub描述了如何做到这一点。再或者，zipline/resources/market\_data/SPY\_benchmark.csv提供SPY返回值，您可以将其移动到.zipline文件夹中，如果您没有自行更改位置该文件夹通常位于主目录中。

另外，只能通过Interactive Brokers运行Quantopian并不提供完全支持的实时交易系统（https：//github.com/zipline-live/zipline）。

P266

LightGBM 模型接收大量的参数（参见https://lightgbm.readthedocs.io/ and the next chapter）。当然我们只是为了启用随机森林算法，因此只需要定义boosting\_type，设置 bagging\_freq 为正数，objective 为 regression，代码如下：

P433

仅使用少量训练集，我们就可以达到很高的准确率。然而，最先进的性能达到的错误率只有 1.02%（参见https://benchmarks.ai/svhn），要想更接近，最重要的步骤就是增加训练数据量。

P456

针对循环神经网络的 TensorFlow 2 指南强调了以下事实：GPU 支持仅在对大多数长短期记忆层设置

使用默认值时可用（https://www.tensorflow.org/guide/keras/rnn）。

P511

强化学习问题有几个元素，这些元素将它们与我们到目前为止所讨论的机器学习区别开来。下面两个部分概述了通过学习自动化决策的策略来定义和解决强化学习问题所需的关键特性。我们将使用这种表示法，并大致遵循强化学习:导论(Sutton和Barto 2018)和David Silver的伦敦大学学院RL课程(https://www.davidsilver.uk/teaching/)，建议在本章范围内的简短总结之外进行进一步研究。