



俄罗斯联邦科学和高等教育部 圣彼得堡彼得大帝理工大学计算机科
学与技术学院
软件工程研究生院。

该论文已被接受进行答辩

VSDPI主任

_____P.D.Drobintsev

" ____ " _____2021 г.

最后的资格认证工作

学士工作

基于对新闻文本的智能分析，预测股票趋势

在培训领域（专业） 09.03.04 - 软件工程

重点（简介） 09.03.04_01 – 开发和维护高质量软件的技术

由学生填写。3530904/70104

D.O. Mitusov

主管副教授
，技术科学
候选人

O.G.Maleev

顾问
在规范的控制下

E.G.Lokshina

圣彼得堡2021年

俄罗斯联邦科学和高等教育部 圣彼得堡彼得大帝理工大学计算机科
学与艺术学院
软件工程研究生院。

赞成

VSDPI主任

_____P.D.Drobintsev

"_____"_____20 ____ г.

要求

给学生Dmitry Olegovich Mitusov的毕业证书，3530904/70104

1. 主题: 基于新闻文本的智能分析预测股票趋势
2. 学生交出已完成工作的最后期限:
3. 工作的原始数据: Tensorflow文档、PyTorch文档
4. 工作内容 (待开发的问题清单):
 - 1.1 收集数据集
 - 1.1.1 新闻数据
 - 1.1.2 市场数据
 - 1.2 新闻分析模型的训练

圣彼得堡2021年

- 1.3 贸易的培训模式
- 1.4 数据流和存储
 - 1.4.1 新闻数据
 - 1.4.2 市场数据
 - 1.4.3 存储格式
- 2.1 描述与系统合作的算法
- 2.2 获得的结果

5. 图形材料清单（包括强制性图纸）：

-

6. 就业辅导员：

-

7. 转让日期 2021年1月

25日

ERC导师

O.G.Maleev

接受执行的转让

25.01.2021

学生

D.O. Mitusov

REFERAT

35页, 7个图, 4个表, 1个附录 关键字 字:

BERT、 发展 EUTH-

以下是一些最常见的语言类型, Q-learning, 深度学习
NIE

毕业论文的题目是 "基于新闻文本的智能分析的股票趋势
预测"。

本文重点讨论了股票市场交易的机器学习算法的实现。

在工作过程中已经解决的挑战:

1. 实施新闻分类算法, 用于
找到可能影响公司股价的消息。
2. 在后台执行交易的算法的实现
新市场是第一个在其战略中使用东北背景的市场。
3. 对分享雷亚尔结果的分析
算法不是一个简单的问题。

这项工作是利用自然语言处理方法和深度学习算法的支持完成的。所获得的解决方案可以在短时间内对日内交易的股票趋势做出预测。根据结果, 得出的结论是所实施的解决方案可以用于证券交易所的交易。

ABSTRACT

35页, 7个图, 4个表, 1个附录

关键词: Bert, 自然语言处理, Q-learning, 深度学习

习

最终合格作品的主题: "根据新闻文本的智力分析预测股票的趋势"。这项工作致力于实施机器学习算法来进行股票交易

。

在工作过程中解决的任务:

1. 实施新闻分类算法, 寻找能够影响公司股票价格的新闻。
2. 实施一种在证券交易所进行交易的算法, 在其策略中使用新闻背景。
3. 对所实施的算法的联合使用结果进行分析。

这项工作是利用自然语言处理技术和深度强化学习算法进行的。由此产生的解决方案可以预测股票在短时间内的趋势, 以进行离谱的交易。根据结果, 得出的结论是, 有可能将实施的解决方案用于证券交易所的交易。

目 录

简介.....	8
第1章：实施的系统描述.....	10
1.1. 收集数据集.....	10
1.1.1. 新闻数据	10
1.1.2. 市场数据.....	11
1.2. 新闻分析模型的训练.....	12
1.3. 贸易的培训模式.....	18
1.4. 数据流和存储.....	20
1.4.1. 新闻数据	20
1.4.2. 市场数据.....	22
1.4.3. 存储格式	22
第二章：系统操作.....	24
2.1. 描述与系统合作的算法.....	24
2.2. 获得的结果.....	27
总结	31
使用的资料来源清单.....	33
附件1.....	35

简介

相关性和实际意义

预测股票市场走势是一个众所周知的问题。影响股票价格的因素之一是新闻。社会媒体是新闻的主要来源之一，也反映了公众对当前事件的看法。

本文涉及的是日内交易，这是有风险的。在日内交易中

损失可能是计划的5至10倍。[1]但同时，如果选择了正确的策略

它可以赚取大量的利润。因此，找到一种能将风险降到最低的方法是很重要的。在本文中，我们考虑了一种方法，即首先寻找那些股票在不久的将来可能会升值的公司。

问题的科学发展程度

这项工作涉及神经网络的两个应用--用于文本（新闻）内涵分类和股票交易。

对于第一个案例，有一篇类似领域的出版物--Twitter情绪预测股市[2]。该论文调查了从Twitter用户的信息中得出的公众的集体情绪状态（快乐、平静、焦虑）是否与工业品的价值相关。

道琼斯指数。他们使用神经模糊系统来进行预测。他们的结

果显示、

推特上的公众情绪状况与道琼斯工业指数密切相关。

第二种情况在《多DQN》一文中讨论：

用于股票市场预测的深度Q-learning代理的集合"[3] 用相同的数据训练多次的Q-learning代理来实现目标。前-

日内交易的实验结果表明，与传统的买入并持有的策略相比，性能更好。

这项工作结合了这两种方法来创造一种需要交易者进行最低限度控制的股票交易系统。

工作目的

本文的目的是实现股票市场交易的机器学习算法及其可能的应用。

工作目标

1. 实施新闻分类算法，寻找能影响公司股价的新闻。
2. 在后台执行交易的算法的实现
新市场是唯一一个在其战略中使用怀旧背景的市场。
3. 寻找和设置新闻来源和
股票市场数据
4. 对共享使用已实施算法的结果进行分析。

第1章 实施的系统描述

1.1. 收集数据集

1.1.1. 新闻数据

收集了金融新闻数据来教授模型。

来源：汤森路透 - 包含信息

关于已发布的新闻文章/警报，如文章细节、情绪和其他评论。
该数据是为2016年收集的，总数量为952396。

要使用的最有用的信息是

分类任务将是：

- `time(datetime64 [ns, UTC])` - UTC格式的时间戳，显示数据可用的时间（秒精度）。
- `headline(object)` - 新闻头条
- `(int8)` - 新闻的类型（1：警报，3：文章）。对于手头的任务来说，重要的是要准确地对警报进行分类
- `sentimentClass(int8)` - 表示该新闻项目与资产有关的普遍情绪等级。（-1: 负面, 0: 中性, 1: 正面）

1.1.2. 市场数据

已使用的历史市场数据--价格（以美元计），与其对应的日期和时间相比的数量。涵盖的时间段是从2020年3月到2021年3月。该区间已被选定在5分钟内，因为该模型必须对价格变化作出快速反应。

收集时使用了tiingo的API。表1显示了\$TSLA股票的一个数据例子。

表1.为股票\$TSLA收集数据的例子

日期	关闭	量
2020-03-27 12:10:00-04:00	500.665	479
2020-03-27 12:15:00-04:00	502.3	710
2020-03-27 12:20:00-04:00	502.36	103
2020-03-27 12:25:00-04:00	501.33	300
2020-03-27 12:30:00-04:00	501.735	484
2020-03-27 12:35:00-04:00	503.455	2755
2020-03-27 12:40:00-04:00	505.46	800
2020-03-27 12:45:00-04:00	504.97	1668

1.2. 新闻分析模型的训练

使用的库：Tensorflow, transformers(huggingface), sklearn, pandas, gdown

该模型是在Kaggle Kernels平台上用NVidia K80 GPU（包含GPU）训练的。

在运动学习过程中，训练的结果是加速了12.5倍。[4]

要导入一个训练数据集和使用了gdown + pandas库。

标题 "将作为模型的输入，而 "情感类 "和 "紧迫性 "将作为输出。

数据分为训练用数据--80%和验证用数据--20%。

BERT用于分类 - 一个双向的

该模型基于一个变压器架构，用一种基于注意力力学更快的方法取代了顺序递归神经网络。该模型还在两个无师自通的学习任务中进行了预训练：预测隐藏的单词和预测下一句话[5]。

使用一个预先训练好的BERT模型，经过

BERT模型通常在大量的文本上进行预训练，然后针对特定的任务进行微调，如情绪分类、意图检测、问题回答等。BERT模型通常在大量的文本上进行预训练，然后针对特定的任务进行调整。

在对数据进行预处理后，加载一个BERT配置对象，它管理着模型和标记器。然后是标记器，它将在以后的脚本中用于将输入文本转换为BERT标记，然后将其增强和截断到最大可能的长度。然后，模型本身被加载BERT。

所使用的模型版本是以BERT为基础的，因为警报类的所有新闻都是用大写字母写的，而文章类的所有新闻都是用小写字母写的，在训练模型时不应该考虑到这一点。

BERT-base-uncased使用30,522个单词的字典。标记化过程包括将输入文本分割成字典中可用的标记列表。为了处理字典中没有的词，BERT使用一种叫做BPE（逐对编码）的技术。符号化框架的WordPiece tokens）。在这种方法中，单词外面的词汇被逐渐分解成子词，然后用一组子词来代表这个词。在这项任务中，标记的最大长度被当作68。

在变形金刚库中有几个不同的BERT分类模型。最主要的是一个简单地称为 "BertModel"（PyTorch）的模型或 "TFBertModel"（TensorFlow）。这个模型将被使用。

变形金刚库还附带了一个用于序列分类的预构建的 BERT模型、 称为 TFBertForSequenceClassification。在其中，加载一个干净的 BERT模型后，你只需向它添加滤波层和密集层。在双序列分类的情况下只是增加了两个致密层而不是一个。在图1中显示了所使用的模型的总体情况。

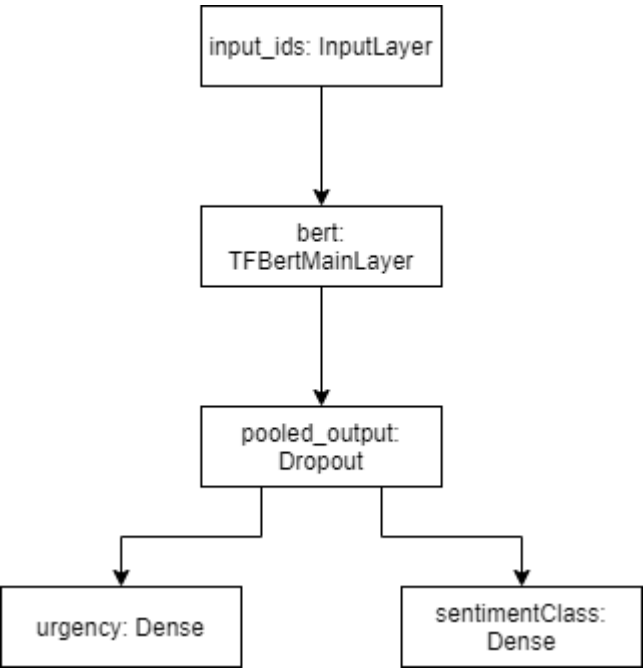


图1.该模型的总体方案

图2是对该模型更详细的描述。

Model: "BERT_model19"

Layer (type)	Output Shape	Param #	Connected to
input_ids (InputLayer)	[(None, 68)]	0	
bert (TFBertMainLayer)	TFBaseModelOutputwit	109482240	input_ids[0][0]
pooled_output (Dropout)	(None, 768)	0	bert[0][1]
sentimentClass (Dense)	(None, 3)	2307	pooled_output[0][0]
urgency (Dense)	(None, 3)	2307	pooled_output[0][0]

Total params: 109,486,854
 Trainable params: 109,486,854
 Non-trainable params: 0

图2. 所用模型的详细描述

然后剩下的就是编译模型并在测试数据上进行测试。

设置损失函数和度量：

```
loss = {'urgency': CategoricalCrossentropy(from_logits =
True), 'sentimentClass': CategoricalCrossentropy(from_logits)
= True)}
metric = {'urgency': CategoricalAccuracy('accuracy'), 'senti-
mentClass': CategoricalAccuracy('accuracy') }
```

这里的主要变化是使用了一个分类交叉熵损失函数，而不是通常用于多类分类的函数。分类交叉熵损失函数允许模型为标签分配独立的概率。

学习周期与原始BERT实现中提供的周期相同。[6] epoch和batch_size的值分别为6和32。

同样的优化器被用于微调、
BERT最初用它来训练："自适应时刻"

(Adam)。学习率取为 $5e-5$ ，正如原文所建议的那样[5]。

一旦模型被训练出来，就可以在数据上对其进行评估验证，看其表现如何：

损失: 0.3628

感情类损失: 0.3288

迫切性_损失: 0.0341

SentimentClass_accuracy: 0.8031

迫切性_准确性: 0.9874

事实证明，该模型在识别新闻的紧迫性方面相当出色，但在分析情绪时却可能出错。但鉴于此：

(A) 情感类别--3，而不是变体

新闻的紧迫性；

B) 金融新闻往往没有任何明显的感情色彩；

B) 一条新闻可能同时涉及几家公司，每家公司的
情绪都可能不同；

结果令人满意，该模型可以应用于Twitter新闻分析。

表2显示了在Twitter（股票代码：AAPL）数据集上的一个工作实例。

表2.关于Twitter数据集的工作实例，（股票代码：AAPL）。

	日期	时间	=	感情类	鸣叫
0	20.03.2021	18:56:54	2	2	在过去20年中，苹果的股票有许多粗糙的补丁。它很难坚持到最后...。但我总是强调同样的观点，拥有它不要交易它，正是因为它做得有多好!\$AAPL
1	20.03.2021	17:21:59	2	0	BTW @TMobile现在也在收集和出售你的数据，你需要选择退出。不容易搞清楚。为什么我必须选择退出。应该是，我必须选择加入。这些设备需要由用户控制 @tim_cook - 够了，这种间谍活动。\$aapl \$tmus
2	20.03.2021	17:20:01	2	0	我们需要一个设置来关闭所有来自iPhone上的应用程序的数据收集 @tim_cook - 只是重新发布它，苹果。这是我的手机。它是我的信息，我应该能够控制它而不被应用程序所欺骗。删掉Facebook
4	20.03.2021	4:00:32	0	1	马克-扎克伯格对苹果(\$AAPL)即将推出的iOS系统更新所带来的隐私变化有了改观。 https://t.co/kR8VCziQ3a

在这个数据集中，有1000个2020-2021年的警报。对于紧迫性等级，得到的警报数量是

– 539.对于情绪类，细分如下：

- 561 - 中立
- 245 - 正面
- 195 - 负面

1.3. 贸易的培训模式

由于在这项工作中只使用了一种策略--上涨游戏，所以只需要了解一般的趋势--购买股票后短时间内的价格上涨就可以了。因此，没有必要使用大量公司的数据，只需使用高度波动的股票的历史数据即可。

实现了一个股票交易的模型，使用深度强化学习，特别是深度Q-学习进行训练。

强化学习是一系列的方法机器学习，它允许创建代理，通过与环境的互动，从环境中学习，因为他们通过试验和错误学习最佳行为路线。这里的重要思想是，这种方法可以应用于任何现实世界的任务，这些任务可以被传统地描述为马尔科夫过程。

在Q-learning中，基于从环境中收到的输入，代理人形成了一个效用函数Q，它随后允许代理人不再随机选择行为策略，而是考虑到以前与环境互动的经验。Q-learning的优点之一是

- 即它能够比较预期效用

可用的行动，而不形成环境的模型。

由于标准Q网（DQN）存在一些统计问题，而这些问题在DDQN中得到了解决，因此采用了双深度Q网（DDQN）：

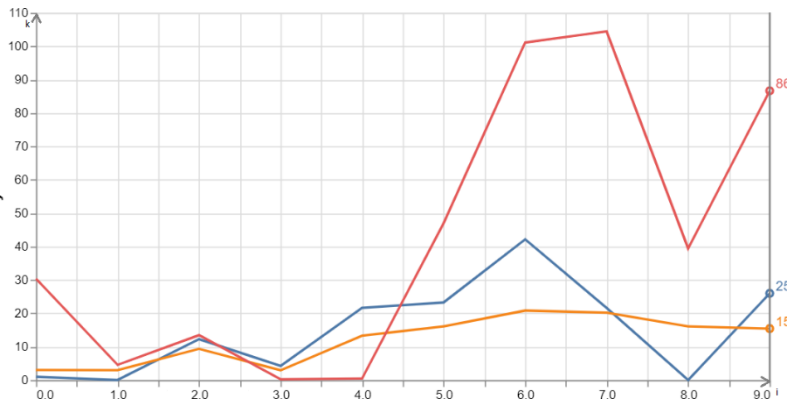
- 1) DQN近似地表示出以下数值的集合是高度相互关联的；
- 2) DQN往往是过度乐观的。它将高估停留在一个状态的时间，尽管这只是由于统计误差；[7]。

在证券交易所的交易任务中，Q-learning被用来详见下文：

- 1) 在某个时间点上（一个插曲），代理人在一个时间窗口中保存了当前的股价表示法
- 2) 代理人选择并执行行动--购买、出售或持有。
- 3) 代理人观察随后的状态，获得一个薪酬信号，衡量投资组合的差异
- 4) 代理人根据计算出的损失梯度来调整其参数。

在算法的开始，买入固定数量的股票或购买固定数量的股票。在结束时，所有未售出的股票被售出。

集数被选择为100。每隔10集，所产生的模型被保存。此后，该模型
这些模型的性能已经在各种测试数据上进行了测试。结果显



示在图中（图3）。

图3. 不同测试数据上的模型验证结果

横轴是模型编号，纵轴是投资组合的变化。

根据结果，可以得出结论，经过70次训练的模型
_50_20，是日内交易上涨策略的模型、
是最有效的。

1.4. 数据流和存储

1.4.1. 新闻数据

从Twitter检索实时数据的能力仅由官方Twitter API支持，其免费版本在数量上有限制。

请求，并在时间上。所以对于这项任务，我们使用了库twint，它有一个用Python编写的Twitter解析工具。Twint允许你在不使用Twitter API的情况下从Twitter配置文件中检索通信。

多处理被用来加快数据收集的速度。采用这种方法，平均速度约为每分钟600次。

工作选择：

- 含有"\$<name_tyker>"或公司名称的信息来自所有用户--信息数量最多，有大量的垃圾邮件和不必要的信息。
- 含有"\$<name_tyker>"或公司名称的帖子，来自经过验证的用户--大大限制了用户群，没有垃圾邮件，但并非所有帖子都是新闻。
- 含有"\$<name_ticker>"或公司名称并被所有用户标记为 "新闻 "的帖子--帖子的数量被大大限制，几乎没有垃圾邮件或不必要的信息，但并非所有属于新闻的帖子都被标记为 "新闻"，一些信息可能被省略。
- 含有"\$<name_tyker>"或公司名称的信息，并被验证用户标记为 "新闻"--每条信息都有价值，但信息量很小。

由于方案1中不必要的信息过多，方案4中缺乏信息，决定使用方案3。

纳斯达克的股票列表在大约7分钟内通过、如果你在搜索中也包括完整的公司名称。这是运行该脚本的最小频率。但是，在搜索选项2、3、4中，这样的频率下，发现的新闻的平均数量小于1。因此，每10-15分钟进行一次搜索是比较理想的。在这段时间内，新闻将继续保持相关性[8]。

1.4.2. 市场数据

使用的库：Alpaca-trading-api，web-socket，sqlite3。

Alpaca Data API是用来获取当前股票价格的信息。在与新闻工作时收到的股票代码是输入的。每次价格更新时，都会收到一条消息，并通过函数on_message，将收到的信息与类型、价格和数量写入数据库。

1.4.3. 存储格式

得到的数据被存储在一个sqlite数据库中。图4显示了数据库的可视化。4显示了该数据库的可视化情况。

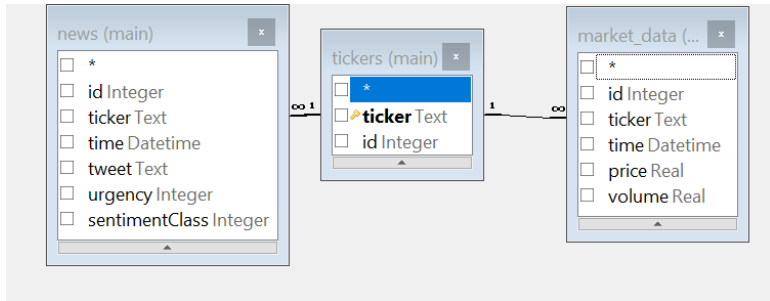


图4：数据库存储格式

股票列表 - 股票表：

- "滴答"--滴答

新闻和它的类 - 表新闻

- "滴答"--滴答
- "时间" - 当前的日期和时间
- "推特" - 新闻
- "紧迫性"是新闻
- "SentimentClass" - 此刻的成交量 市场数据

- market_data表：

- "滴答"--滴答
- "时间" - 当前的日期和时间
- "价格" - 到目前为止的价格
- "音量" - 此刻的音量

当系统被验证后，该类型的数据将从数据库中删除。

第2章 系统操作

2.1. 描述与系统合作的算法

交钥匙解决方案是一个用于分析交易策略的应用程序包。完成培训任务后，你可以开始测试系统。该系统的总体方案如图5所示。5.

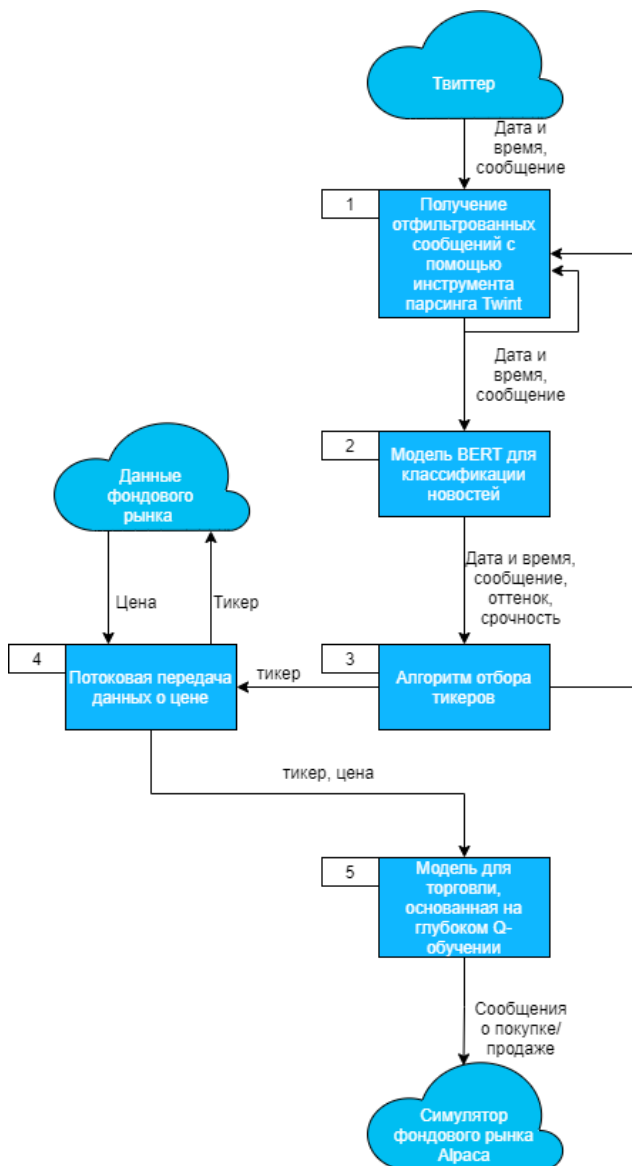


图5. 已实施的系统的一般操作方案

整个项目分为几个脚本，用一个bat文件启动。对操作程序进行更详细的描述：

1 - 运行搜索过去15分钟内具有指定过滤器的信息，输入在数据库中发现的结果，如果是具有指定过滤器的信息已被发现。如果在给定的过滤器中没有发现任何信息，那么15分钟后将重新开始搜索。

2 - 一旦信息搜索完成，脚本就以运行中的BERT模型开始，初步

在所收集的新闻集上进行训练。得到的结果是情感类和新闻紧迫性类。这些结果被输入一个数据库。

3- 使用股票选择算法进行搜索。你-

选择有关其各自公司的新闻具有积极内涵和紧迫性等级--通知--的股票。具有积极内涵的新闻数量必须大于具有消极内涵的新闻（中性新闻不考虑在内）。如果找到一个以上符合要求的股票，则选择具有最高数量的正面和中性新闻的股票。如果没有找到股票，将出现15分钟的等待期，然后重新开始新闻搜索。

4 - 当找到一个合适的股票代码时，股票价格数据就会使用Alpaca Data API进行流转。

收到的信息开始被输入数据库。

5 - 来自数据库的信息被送入模型输入

基于深度Q-learning的证券交易所交易模型。该模型的输出是在特定时间采取的行动，以获得最高利润。可能的行动有：买入、卖出、持有。工作结束的时间是北美东部时间0:00。数据库中关于这个股票的所有条目都被删除。

可调整的参数：

- 1) 新闻搜索的开始时间，即脚本的开始时间
- 2) 消息的过滤器（选项 - 新闻过滤器激活，验证用户过滤器激活，两个过滤器都激活）。
- 3) 交易模式的选择（适合股票波动性较大的公司的模式和适合股票波动性较小的公司的模式）。

2.2. 获得的结果

这些结果来自于历史数据。

为此，首先收集了新闻（表3）。

表3：收集的新闻，Nvidia

	date	tweet	urgency	sentimentClass
0	27.05.2021	NVIDIA Virtual GPU and @VMware empower enterprises with graphics-rich, secure, and cost-effective VDIs. Read this @Forbes article to learn why #vGPU solutions are essential for "anywhere workspaces" across industries. #VDI #vWS #vPC https://t.co/CzJBQuY5aY https://t.co/HvcnKdy8qC	2	2
1	27.05.2021	Nvidia made \$5.66 billion in revenue last quarter, despite RTX 30-Series shortages. https://t.co/q6X06c7yIp https://t.co/t2p2l6NOx6	0	1
2	27.05.2021	Nvidia anticipates RTX 30-Series graphics card shortages will continue until the "second half of the year." https://t.co/Mbqua7Ofyx https://t.co/MiqAegaZLt	0	0
3	27.05.2021	Nvidia teases GeForce RTX 3080 Ti announcement for May 31st https://t.co/lRbN4pG4M0 https://t.co/wWqmYhOq4q	0	1
4	27.05.2021	Nvidia gaming revenue spikes by 106%, data center by 79% https://t.co/xjocHanDDp by @DanielHowley https://t.co/SqBPkwBBia	0	2
5	27.05.2021	Nvidia forecast beats expectations but crypto mining's role remains unclear https://t.co/BNrowthXxe https://t.co/QJqU9kdTG8	0	1

该模型并不总是能正确检测出新闻的颜色，因此需要进一步完善。

在正面色调的新闻比负面色调的新闻多的情况下，该模型被测试为股票交易。在这种情况下，具有积极内涵的新闻更多，所以对模型进行了测试，并绘制了图表（图6）。

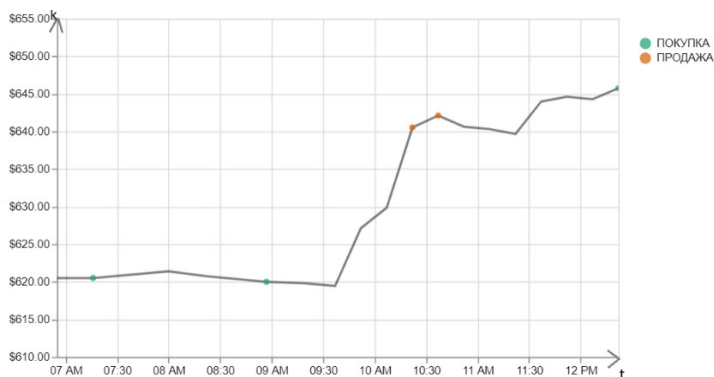


图6.Nvidia的系统结果

横轴是以美元计的每股价格。横轴是以美元计的股价，纵轴是时间。

所涉公司是Nvidia，日期是2021年5月28日。

交易机会之间有一个15分钟的窗口。每次只买入和卖出一只股票。结果是获利为\$42.15.42.15美元。

所述的第二个案例是一个例子、当新闻具有较长期的相关性时。公司的股票保持趋势的时间更长。在这个例子中，新闻被收集和分类超过两天。结果显示在表4中。

表4. 收集的新闻，DocuSign

	date	tweet	urgency	sentimentClass
0	4.6.21	\$DOCU is 'in a sweet spot to continue to receive sig	2	2
1	4.6.21	Premarket Movers \$CLB +9.6% \$CLNE +7.7% \$DOCU	0	1
2	4.6.21	More investors sign up for DocuSign shares, while C	0	2
3	3.6.21	\$DOCU Q1 FY22 International revenue increased 8	0	2
4	3.6.21	\$DOCU reported record dollar net retention of 125	0	0
5	3.6.21	Wowza DocuSign with another heater of a quarter	2	2
6	3.6.21	The most accurate \$DOCU estimate is from @sPAC	2	1
7	3.6.21	\$DOCU beats the Estimize EPS Consensus by 8c and	0	2
8	3.6.21	\$DOCU reports FQ1 earnings of 44c EPS and \$469.1	0	1
9	3.6.21	\$DOCU reports first quarter results for FY22 with 5	0	2
10	3.6.21	\$DOCU * Q1 Adj. EPS \$0.44 Beats \$0.28 Estimate	0	1
11	3.6.21	\$DOCU Q1 Earnings: EPS: \$0.44 vs. \$0.27 Expected	0	1
12	3.6.21	Tune into a #live webcast of \$DOCU Q1 FY22 earni	2	0
13	3.6.21	\$DOCU reports after the close, Estimize Consensus	0	0

这些交易也是在两天的时间内进行的。第二种情况的图表显示在图7中。7.

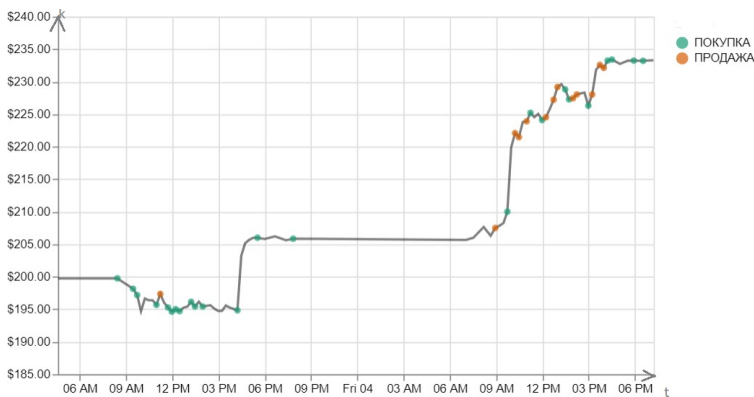


图7.DocuSign两天的系统运行结果

审议的公司是DocuSign, Inc.，日期是2021年6月3日到2021年6月4日。

交易机会之间有一个15分钟的窗口。每次只买入和卖出一只股票。结果是获利为\$343.28.费用为343.28美元。

在解决方案的标准操作中，在一天内进行交易的情况下，利润为18.45美元（附录1。利润为18.45美元（附录1）。

因此，经过进一步的可能的修改，创建的解决方案，可以实现资本的大幅增加。

其结果是一个用于分析基于历史数据的交易策略的应用程序包。

总结

在工作过程中，解决了以下问题
任务：

1. 实施了一种新闻分类算法，以寻找能够影响公司股价的新闻。
2. 一个证券交易所的交易算法已经实现，该算法采用了一种新的策略。
东部的背景。
3. 找到了新闻和股票市场数据的来源。
4. 的联合使用的结果。
已经实施的算法。

5. 创建了一个应用程序包，以分析基于历史数据的交易策略。

基于这些结果，我们可以得出结论，将新闻纳入考虑范围对于制定交易策略是有用的。然而，新闻选择和分类算法应该在未来得到改进，因为在某些情况下，当用于不需要交易者控制的交易策略时，其性能可能不够准确。

参考文献列表

1. 亚当-格莱姆斯技术分析的艺术与科学
/ Adam Grimes.Hoboken: John Wiley & Sons, 2012.- 463
c.
2. Bollen, J. Twitter情绪预测股市 / Bollen J., Mao H., Zeng
X.// 《计算科学》杂志。- 2011.- №2(1).- c.1-8
3. Salvatore Carta.Multi -DQN: An ensemble of Deep Q-
learning agents for stock market forecasting /
Salvatore Carta, Anselmo Ferreira, Alessandro Sebastian
Podda // Expert Systems with Applications.- 2021.-
№164.- c.1- 16
4. 用 gpu 运行 kaggle 内核 。 URL:
<https://www.kaggle.com/dansbecker/running-kaggle-kernels-with-a-gpu>
5. Jacob Devlin.BERT: Pre-training of Deep Bidirectional
Transformers for Language Understanding / Jacob Devlin,
Ming-Wei Chang, Kenton Lee, Kristina Toutanova //
2019.- c.1-16
6. BERT 资源库 。 URL: <https://github.com/google-research/bert>
7. 郝多-范-哈瑟尔特.Deep Reinforcement Learning with
Double Q-learning / Hado van Hasselt, Arthur Guez, Da-
vid Silver // Google DeepMind.- 2015.- c.1-13
8. Emanuele Tetia.扭曲与股票价格之间的关系.美国技术
行业的证据 / Emanuele Tetia, Maurizio Dallochio,
Alberto Aniasi /.

技术预测与社会变革.- 2019.-

№149

9. Hikmat Ullah Khan. Twitter趋势：实时数据的排名算法分析 / Hikmat Ullah Khan, Shumaila Nasir, Kishwar Nasim // Expert Systems with Applications.- 2021.- №164.
10. Tensorflow API 文档。URL: http://www.tensorflow.org/api_docs
11. Twitter 的 API 文档。URL: <https://developer.twitter.com/en/docs>
12. Twintproject. URL: <https://github.com/twintproject/twint>

附件1

