基于方面的情感分析及其在新闻驱动的股票预测中的应用

（申请清华大学工学硕士学位论文）

培 养 单 位：计 算 机 科 学 与 技 术 系
学　　科：计 算 机 科 学 与 技 术
研　究　生：潘　寅　旭
指 导 教 师：宋　斌　恒　副　教　授

二〇一九年五月

Aspect-based Sentiment Analysis for News-driven Stock Prediction

Thesis Submitted to

Tsinghua University

in partial fulfillment of the requirement

for the professional degree of

Master of Engineering

by

Pan Yinxu

(Computer Science and Technology)

Thesis Supervisor : Associate Professor Song Binheng

May, 2019

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：（1）已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；(2) 为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容。

本人保证遵守上述规定。

（保密的论文在解密后应遵守此规定）

作者签名：_____　　导师签名：_____

日　　期：_____　　日　　期：_____

摘 要

股票预测一直是学术界和商业界共同的研究热点之一。新闻事件能够影响交易员的决定，而股票价格的变动会被交易员的决定影响，因此，新闻事件是可以影响股票市场的。之前的研究将文本挖掘的技术应用到股票预测当中，通过向量或结构化实体的方式来表示事件，忽视了新闻文本中的大量细节；同时，之前研究中用到的数据集大多是不公开的。我们不再从新闻文本中提取事件表示，而是直接将新闻文本作为输入，为每支股票创建一个向量化的表示，并将股票预测转化为一个基于方面的情感分析问题，通过预测新闻对某支股票的影响，来预测股票价格的变动。我们构建了一个名为 senti-stock 的股票预测数据集。它包含约三万个样本，每个样本都是由来自路透社的新闻和来自雅虎财经的股票历史数据构建的。我们提出一个名为 MSRA 的模型，它包括情感分析模块和股票关系模块。情感分析模块利用多尺度卷积和股票系数，多尺度卷积可以提取不同粒度的句子特征；股票相关系数是通过股票向量计算得到的，反映了股票相关的信息。股票关系模块利用了 attention 机制，学习股票之间的相关关系。与其他基于方面的情感分析模型相比，我们的模型在 senti-stock 数据集上取得了最好的效果。短期（天）、中期（周）、长期（月）的股票预测准确率分别达到了 69.20%，62.73%，68.26%。

本文的贡献点主要有：

- 我们利用路透社的金融新闻和雅虎财经的股票历史数据，创建了新闻驱动的股票预测数据集 senti-stock；
- 我们提出了股票向量的概念，并将新闻驱动的股票预测转化为基于方面的情感分析问题；
- 提出了基于 transformer 和多尺度卷积的模型，并在多个开放数据集上取得了最佳的实验效果；
- 提出了 MSRA 模型（使用 attention 机制的多股票关系模型，Multi-Stock Relation model using Attention mechanism），并在 senti-stock 数据集上取得了最佳的效果。

关键词：股票预测；基于方面的情感分析；多核卷积；Transformer

Abstract

The stock market prediction has always been the hotspot research in academia and business. As news events affect human decisions and the volatility of stock prices is influenced by human trading, it is reasonable to say that news events can influence the stock market. Previous studies have applied text mining techniques to predict stock prices by using structured entities or dense vectors to represent news events. And most datasets used in previous studies are not publicly available. We take the news text as input, create an embedding vector for each stock, and regard news-driven stock prediction as an aspect-based sentiment analysis problem. We try to predict stock prices by evaluating the impact of news on stocks. We create a stock prediction dataset named senti-stock. The dataset consists of 30 thousand samples, which is made up of online financial news (from Reuters) and stock history price (from Yahoo Finance). We propose a model named MSRA (Multi-Stock Relation model using Attention mechanism), which consists of sentiment analysis module and stock relation module. The sentiment analysis module is based on multi-scale convolution neural network with stock coefficient. The multi-scale CNN can extract sentence features in different granularities. The stock coefficient, which is computed using the stock embedding is used to adjust the sentence features. The stock relation module uses attention mechanism to learn the relations between stocks. Compared with other aspect-based sentiment analysis models, our model gets the best performance on the senti-stock dataset. The accuracy of short-term (day), middle-term (week) and long-term (month) stock prediction is 69.20%, 62.73%, and 68.26%, respectively.

The main contributions of this paper are:

- We create a dataset named senti-stock using Reuters news and Yahoo Finance history data;
- We convert news-driven stock prediction to an aspect-based sentiment analysis problem by creating a vecotor for each stockd;
- We propose a model using transformer and multi-scale CNN and get the best performance on Restaurant, Laptop and Twitter dataset;
- We propose MSRA (Multi-Stock Relation model using Attention mechanism) and get the best performance on senti-stock dataset.

Key words: Stock Prediction; Aspect-based Sentiment Analysis; Multi-scale CNN; Transformer

目 录

主要符号对照表

CNN	卷积神经网络 (Convolution Neural Network)
RNN	循环神经网络 (Recurrent Neural Network)
LSTM	长短时记忆网络 (Long Short Term Memory Network)
GRU	Gated Recurrent Unit
ABSA	基于方面的情感分析 (Aspect-based Sentiment Analysis)
ACSA	对某一方面的情感分析 (Aspect-Category Sentiment Analysis)
ATSA	对某一目标的情感分析 (Aspect-Term Sentiment Analysis)
SVM	支持向量机 (Support Vector Machine)
Bag of Words	词袋模型
TFIDF	词频与逆向文档频率 (Term Frequency Inverse Document Frequency)
MSRA	使用 attention 的多股票关系模型 (Multi-Stock Relation Model using Attention mechanism)

第 1 章　介绍

1.1　研究内容

股票（stock）是股份公司发行的所有权凭证，是股份公司为筹集资金而发行给各个股东作为持股凭证并借以取得股息和红利的一种有价证券。每股股票都代表股东对企业拥有一个基本单位的所有权。每家上市公司都会发行股票。股票趋势预测一直是股票投资所关注的热点之一，优秀的股票预测可以带来巨大的收益。因此，股票预测一直是学术界和金融界共同的研究兴趣。

大量经验性的研究表明，股票在某种程度上是可以被预测的。行业研究员通过基本面研究，由股票公司的运行情况、相关政策、新闻等，股票的估值进行预测；量化分析师通过技术分析，运用数学、统计的方法，从股票历史交易数据中，寻找规律，进行股票趋势预测。而在本文中，我们专注于解决新闻驱动的股票预测问题。新闻事件能够影响交易员的决定，交易员的交易影响股票的价格。因此，我们可以认为，新闻事件是可以影响股票价格的。

随着自然语言处理技术的发展，使用人工智能的方法处理新闻文本，提取相关信息成为了可能。我们提出为每个股票创建一个相应的向量，从而将新闻驱动的股票预测转化为了基于方面的情感分析 (其目标是提取句子对某一方面或目标的情感极性)。我们假设股票的涨跌仅受前一个时间段的新闻的影响，通过预测新闻（文本）对股票（目标）的影响（情感极性），来预测股票的涨跌。

1.2　研究背景

近期的研究工作开始应用文本挖掘技术分析新闻对股票市场的影响。研究者使用 OpinionFinder[1] 和 Google-Profile of Mood States (GPOMS) 工具，来研究推特 (Twitter) 情绪是否会影响道琼指数 (Dow Jones Industrial Average, DJIA)[2]。他们利用上述两款工具处理推特数据，获取了情绪的时间序列，并以此预测道琼指数的每日涨跌，获得了 87.6% 的准确率。这一研究表明，网络内容的情感信息与股票市场的变动是相关的。之后，有研究者提出了结构化的事件表示方法[3]。他们提出使用信息抽取的方法来得到事件的表示，在不使用人力的情况下，从大规模的公开新闻中得到事件的结构化表示，即一个包括主语、谓语、动作和时间的四元组。他们利用线性和非线性的模型，探索新闻事件和股票市场之间的隐含的复杂关系。他们预测标普 500 的变化的准确率达到了 60%，预测单支股票的准确率超

过了 70%。为进一步提升事件驱动的股票预测的准确率，优化事件的表示形式，又有研究者用深度学习的方式来解决这一问题[4]。首先，他们从新闻文本中提取事件，用一个向量表示事件，并用一个神经网络对其进行训练；然后，用一个深度卷积网络来预测股票的短期和长期变化。相比于之前的方法，他们将预测标普 500 和单支股票涨跌的准确率提升了将近 6%。

上述研究工作主要存在两个问题：首先，他们利用词袋模型、结构化四元组、稠密向量的方式来表示事件，随着自然语言处理技术的发展，这些表示方式已经不是新闻事件的最优化表示方式了；然后，这些工作中所用到的数据集大都是不公开的，影响了进一步的后续研究。

1.3　本文贡献

为解决第一个问题，我们利用自然语言处理的技术，直接将新闻文本作为事件输入。同时，我们提出为每一个股票创建一个向量表示，即股票向量的概念。据我们所知，这是第一次提出股票向量的概念。这样，我们将新闻驱动的股票预测转化为预测新闻事件对某支股票的影响。这一问题在自然语言处理中，被称为基于方面的情感分析，其目标是预测一段文本（新闻文本）对某一特定方面（股票）或目标的影响（涨跌）。比如，“微软收购诺基亚手机部门”这一新闻对微软来说是积极的，将会导致微软的股票在下一个时间段内上涨；对诺基亚来说是消极的，将导致诺基亚的股票在下一时间段内下跌。

为解决第二个问题，我们创建了一个新闻驱动的股票预测的数据集。我们从互联网的公开数据中抓取原始数据。利用来自于路透社的新闻数据，以及来自于雅虎财经的股票历史数据，我们构建了名为 senti-stock 的数据集。在这里，我们假设股票价格的涨跌仅与股票在前一个时间段（前一天、周、月）的相关新闻有关，忽略其他的影响因素。我们将新闻文本以及与这条新闻相关的股票作为输入，将股票在下一个时间段（日、周、月）内的变动情况作为输出标签。具体地，如果股票在下一个时间段内（日、周、月）的涨幅超过 1%，则将输出标签设为 +1，如果股票在下一个时间段（日、周、月）的跌幅超过 1%，则将输出标签设为 −1，其他的输出标签设为 0。

为解决基于方面的情感分析问题，并将其应用于股票预测，我们提出了自己的模型。

首先，为解决 LSTM 的无法并行化的缺陷和 CNN 不擅长解决长期依赖的问题，我们提出使用 transformer[5] 和多尺度卷积的模型来解决基于方面的情感分析问题。Transformer 的结构由谷歌提出，它完全抛弃了 LSTM 和 CNN，仅使用 self-attention

就在自然语言处理的诸多应用中取得了优秀的结果，可以处理长期依赖问题，并且便于并行化。多尺度卷积利用了全部单词的表示，从不同的粒度提取了特征。我们的这一模型，在三个公开数据集 Restaurant、Laptop、Twitter 上取得了最佳的结果。

然后，为了在 senti-stock 这一数据集上取得更好的结果，我们提出了基于 attention 机制的股票关系模型 MSRA（使用 attention 机制的多股票关系模型, Multi-Stock Relation model using Attention mechanism）。MSRA 包括了情感分析模块和股票关系模块。情感分析模块使用多尺度卷积来学习新闻文本的特征。同时，为了考虑股票的相关信息，我们引入了股票系数。股票系数由股票向量通过一个全连接层学习得到，它引入了股票的相关信息。另外，股票之间的相互关系也对股票的变化有着巨大的影响。我们引入了 attention 机制，利用 attention 机制学习得到股票之间的相互关系。MSRA 在 senti-stock 数据集上取得了最佳的结果。

表 1.1　基于方面的情感分析主要研究内容

任务	应用场景	数据集	模型
ATSA	评论分析	Restaurant	**Transformer&MCNN**
		Laptop	
	社交媒体分析	Twitter	
ACSA	**股票预测**	**senti-stock**	**MSRA**

表 1.1显示了基于方面的情感分析的主要研究内容。基于方面的情感分析（Aspect-based Sentiment Analysis）旨在提取句子对某一方面或目标的情感极性，又可以细分为对某一目标的情感分析 (Aspect-Term Sentiment Analysis, ATSA) 和对某一方面的情感分析（Aspect-Category Sentiment Analysis, ACSA）。二者都是一种更细化的情感分析任务，不同的是，对某一目标的情感分析希望提取句子对某一目标的情感极性，这里的目标必须是在句子中出现的一个子串；而对某一方面的情感分析希望提取句子对某一方面的情感极性，这里的方面，并不一定出现在句子当中，而是与句子内容相关的某一方面。比如，“比萨很好吃！”这一句子，对目标“比萨”的情感极性是积极的，对方面“味道”的情感是积极的，“比萨”是出现在句子中的一个子串，而“味道”是与句子内容相关的某一方面，并不一定在句子中出现。在对某一目标的情感分析中，我们提出了使用 transformer 和多尺度卷积的模型；我们还将对某一方面的情感分析应用于股票预测，构建了 senti-stock 数据集，提出了 MSRA 基于 attention 机制的多股票关系模型。除此之外，我们还调研了大量的文献，总结了新闻驱动的股票预测和基于方面的情感分析之前的研究工作，并复现了多个之前基于方面的情感分析的模型，并将这些模型的实现开源。

表格中加粗的部分是我们的主要工作。

本文的贡献点主要有：

- 我们利用路透社的金融新闻和雅虎财经的股票历史数据，创建了新闻驱动的股票预测数据集；
- 我们提出了股票向量的概念，并将新闻驱动的股票预测转化为基于方面的情感分析问题；
- 提出了基于 transformer 和多尺度卷积的模型，并在多个开放数据集上取得了最佳的实验效果；
- 提出了 MSRA 模型（使用 attention 机制的多股票关系模型，Multi-Stock Relation model using Attention mechanism），并在 senti-stock 数据集上取得了最佳的效果。

1.4　组织结构

本文研究基于方面的情感分析及其在新闻驱动的股票预测中的应用[①]。第一章介绍了全文的主要内容；第二章介绍新闻驱动的股票预测和基于方面的情感分析的相关工作；第三章介绍我们所用到的数据集，包括开放数据集 Reataurant、Laptop、Twitter，以及我们所创建的新闻驱动的股票预测数据集 senti-stock；第四章，我们介绍用于对目标情感分析的模型，使用了 transformer 结构和多尺度卷积；第五章，我们介绍用于新闻驱动的股票预测模型，提出股票向量的概念，使用了多尺度卷积，利用股票系数来调整句子特征，并利用 attention 机制学习股票之间的相关关系；第六章介绍了我们的实验设置以及实验结果，我们的模型在公开数据集和股票预测数据集上分别取得了最佳的结果；最后，我们在第七章总结本文的内容。

① 由于字母符号数量限制，本文中各章节的字母符号可能表示不同的含义，相互不冲突。

第 2 章　相关工作

股票预测一直是股票投资所关注的一个重点，成功的预测可以为投资者创造巨大的收益。通过人工智能的方法，对股票涨跌进行预测，成为了计算机与金融的交叉领域的研究热点之一。在本章中，我们将介绍股票市场相关知识、新闻事件驱动的股票市场预测和基于方面的情感分析的相关工作。

2.1　股票市场相关知识

股票能否被预测一直是一个值得讨论的话题，在这些讨论中，有效市场假说是非常重要的一个理论。有效市场假说（Efficient Market Hypothesis，简称 EMH）是由尤金法玛于 1970 年整理提出的[6]。这一研究起源于路易斯巴舍利，他使用随机过程的方法，研究股票价格变化的随机性和布郎运动，他认为，过去、现在和未来的事件都反映在股票市场的价格当中，股价遵循公平游戏模型。尤金法玛总结了前人的理论和实证，提出了有效市场假说，包括以下三个要点。

- 市场中每个人都是理性的，市场中每支股票都处于这些理性人的监视之下，他们每天对股票价格进行分析和预测，并谨慎地在风险与收益之间进行取舍；
- 股票价格反映了这些理性人的供求平衡，买方等于卖方；
- 股票的价格反映了全部的可获取信息，当有新的消息出现时，股票价格会迅速地调整到合理的价位。

有效市场假说意味着“天下没有免费的午餐”，人们无法在有效市场中获得超额收益。然后有效市场假说也不一定完全正确，不是每个交易者都是完全的理性人，信息也并不一定在每个时刻都发生效果。有效市场假说被越来越多地证明不符合现实。

有效市场假说面临许多理论挑战：

- 投资者并非完全理性的；
- 投资者不仅偶然偏离理性，而是经常以同样的方式偏离理性；
- 套利者不会完全消除非理性投资者的错误对价格的影响。

随着人工智能技术的发展，之前的研究表明，股票价格在某种程度上是可以被预测的。如图2.1显示了两条金融新闻。在苹果前 CEO 乔布斯去世后，苹果的股票价格开始下跌；同时，在谷歌的收入报告公布后，由于收入不理想，谷歌的价格开始下跌。从中可以看出，新闻事件对股票市场有着重要的影响。

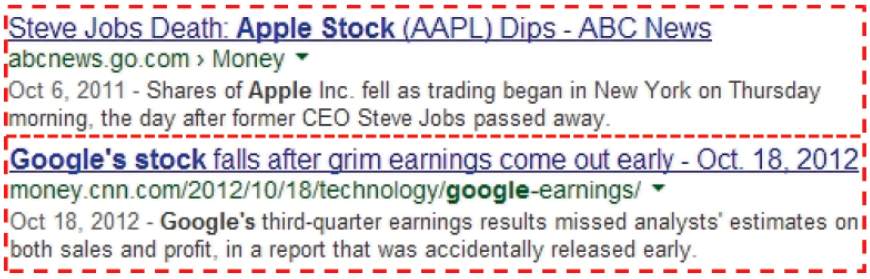


图 2.1　苹果公司和谷歌公司相关新闻两则

2.2　新闻驱动的股票预测

股票市场能否被预测一直都是一个值得讨论的问题。大量经验性的研究表明，股价是可以被预测的[2–4,7]。新闻事件能够影响交易员的决定，而交易员的交易会影响股票的价格。因此，新闻事件可以影响股票价格的变动。

2.2.1　推特情绪预测指数涨跌

依照行为金融学的观点，情绪可以影响个体的行为和决定。为验证这一规律是否能扩展到大规模群体上，研究者研究了大规模的推特情绪状态与道琼指数（Dow Jones Industrial Average, DJIA）的相关性[2]。他们使用两种情绪分析工具来分析每天推特的文本内容：OpinionFinder[1] 和 GPOMS（Google-Profile of Mood States）[8]。

2.2.1.1　OpinionFinder

OpinionFinder 是一个公开的软件工具①，用于提取句子级的主观情绪极性（积极或消极）。

预处理阶段，利用 Standford 提供的词性标注工具，对输入进行句子分割和词性标注；然后从中选取主观的能反应情绪的词语；然后基于朴素贝叶斯方法和情感词库，对文本进行分类。

2.2.1.2　GPOMS

GPOMS 是原作者自己提出的一款情感分析工具，它使用了 Profile of Mood States（POMS-bi），一个由严格审查的心理学仪器提供的词库。为了使这一词库可以应用于推特情绪的分析，作者对词库进行了扩充。推特内容的情感被细分为了六类，包括冷静（Calm），警惕（Alert），肯定（Sure），重要（Vital），和善（Kind）和高兴（Happy）。

① http://mpqa.cs.pitt.edu/opinionfinder/opinionfinder_2/

利用上述两种工具，他们将推特情感分为六类，包括冷静（Calm），警惕（Alert），肯定（Sure），重要（Vital），和善（Kind）和高兴（Happy）。然后，他们利用格兰杰因果分析（Granger Causality Analysis）和自组织模糊神经网络（Self-Organizing Fuzzy Neural Network），来验证利用上述两种工具得到的情感时间序列是可以用来预测道琼指数的收盘价。他们预测道琼指数收盘价的涨跌的准确率达到了 87.6%。

这一研究表明，网络信息的情感极性是可以用来预测股票市场的变动的。

2.2.2　结构化的事件表示

为了解决新闻事件驱动的股票市场预测，研究者提出事件的结构化表示[3]。以往的新闻事件驱动的股票预测，往往使用浅层特征，比如词袋特征（Bag of Words），命名实体，名词等，无法获取实体关系信息。

他们提出结构化的事件表示，使用一个四元组 (O_1, P, O_2, T) 来表示事件，其中 P 表示动作（Action），O_1 表示动作的发起者（Actor），O_2 表示动作的作用者（Object），T 表示动作发生的时间（Time）。比如，“2013 年 9 月 3 日，微软同意以 7.2 亿美元的价格收购诺基亚的手机部门”这一新闻事件可以表示为（Actor= 微软,Action= 收购,Object= 诺基亚手机部门,Time=2013 年 9 月 3 日）。

具体地，研究者们使用依存关系分析器提取句子的结构，选取动词 P（Action），然后将其左侧最近的名词作为动作的发起者 O_1，将其右侧最近的名词作为动作的作用者 O_2。下面我们简单介绍词袋特征和结构化的事件表示方法。

2.2.2.1　词袋模型

在这里，使用经典的 TFIDF（Term Frequency–Inverse Document Frequency，词频与逆向文档频率）值作为词袋特征。TFIDF 被用来评估某一字词对于一个文件集或一个语料库的重要程度。字词的重要性随着它在文件中出现的次数成正比增加，但同时会随着它在语料库中出现的频率成反比下降。

2.2.2.2　结构化事件表示

前面已经提到用一个四元组 (O_1, P, O_2, T) 来表示事件。为了解决数据稀疏性的问题，使用退化的特征，即使用元素的合并作为事件的表示，即 $(O_1, P, O_2, O_1 + P, P + O_2, O_1 + P + O_2)$。具体地，对事件四元组 (Microsoft,buy,Nokia's mobile phone business) 可以表示为 (#arg1=Microsoft,#arg2=buy,#arg3=Nokia's mobile phone,#arg4=Microsoft buy,#arg5=buy Nokia's mobile phone business,#arg6=Microsoft buy Nokia's mobile phone business)。然后，为每个字符串生成一个特定的向量表示，

这几个字符串拼接起来作为事件的表示。因为动作、动作发起者、动作作用者的数据非常稀疏，这里去掉时态、复数等特征，并将动作转换为动作的类别，以减少数据的稀疏性。

同时，这里还使用了线性和非线性的模型来解决新闻事件驱动的股票预测问题。线性模型选取的是 SVM（Supported Vector Machine，支持向量机）模型，非线性模型选取的是三层神经网络。

实验结果显示，预测标普 500 的准确率达到了 60%，单支股票预测的准确率超过 70%。

2.2.3　深度学习与新闻驱动的股票预测

前面所提到的结构化的事件表示方法，使用的是单词向量的拼接。为改进这一方法，研究者提出用一个神经网络来学习事件的表示。图 2.2显示了提取事件特征所使用的网络结构。

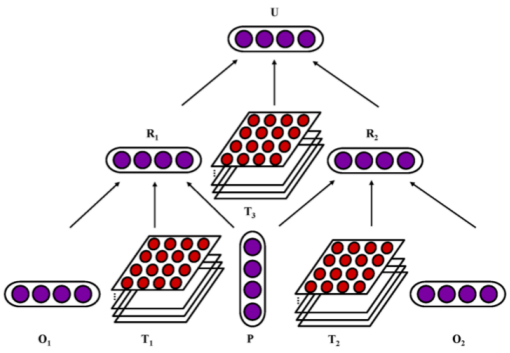


图 2.2　Neural Tensor Network 结构

这里用到的事件表示方法，与前面提到的四元组结构化表示一致。不同的是，不再使用退化的特征，而是使用一个神经网络 Neural Tensor Network 来训练学习得到事件的向量化表示。具体地，Actor、Action、Object 均被表示为一个向量 O_1、P、O_2。然后，利用矩阵 T_1 融合 Actor O_1 和 Action P 得到 R_1，利用矩阵 T_2 融合 Action P 和 Object O_2 得到 R_2，利用矩阵 T_3 融合 R_1 和 R_2 得到事件最终的向量表示。

模型上，这里用到了卷积神经网络。图 2.3显示了模型的结构。上面得到的事件的向量化表示被用来当作短期事件的表示，然后利用 CNN 和 max pooling 层得到中期和长期的事件表示。将短期、中期、长期的事件表示拼接起来，并通过两层神经网络预测得到最终的预测结果。

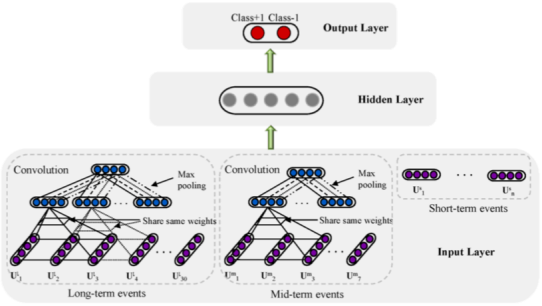


图 2.3　基于 CNN 的预测模型结构

相比于之前的方法，利用神经网络学习得到的事件表示，可以将标普 500 的预测准确率提升将近 6%。

这些新闻事件驱动的股票预测表明，新闻事件与股票的变动是相关的，新闻事件可以用来作为股票预测的依据。但是，无论是利用 OpinionFinder 等工具的方法，还是结构化、向量化的表示，都忽视了新闻文本中的一些细节信息。随着自然语言处理技术的发展和词向量的提出，直接将新闻文本作为事件的表示，将股票预测转化为文本分类成为了更为合适的处理方式。

2.3　基于方面的情感分析

基于方面的情感分析旨在提取某个句子对某一方面或目标的情感极性[9]。比如，“这家餐厅的菜很好吃，但服务太差了”对“食物”方面的情感是积极的，对“服务”方面的情感是消极的。基于方面的情感分析又细分为对某一方面的情感分析和对某一目标的情感分析。对某一方面的情感分析（Aspect-Category Sentiment Analysis）中，方面可能是一个抽象的概念，并不一定在句子当中出现；而对某一目标的情感分析（Aspect-Term Sentiment Analysis or Target-oriented Sentiment Analysis）中，目标一定是句子中的某一个词。随着自然语言处理技术的发展，研究者们提出了许多深度学习的模型来解决这一问题。

2.3.1　ATAE-LSTM

ATAE-LSTM（Attention-based LSTM with Aspect Embedding）[10] 是一个结合了 LSTM[11] 和 attention 机制的模型。与直接使用 LSTM 模型相比，图 2.4显示了 ATAE-LSTM 模型的结构。ATAE-LSTM 使用了 attention 机制，当输入不同的 aspect（方面）时，会对句子的不同部分有不同的侧重。因为 aspect（方面）起着非常关键的作用，ATAE-LSTM 模型用两种方式使用了 aspect 的信息：一个是在计算 attention 的加权和时，把 aspect 的信息和句子的隐式信息表示拼接起来；另一

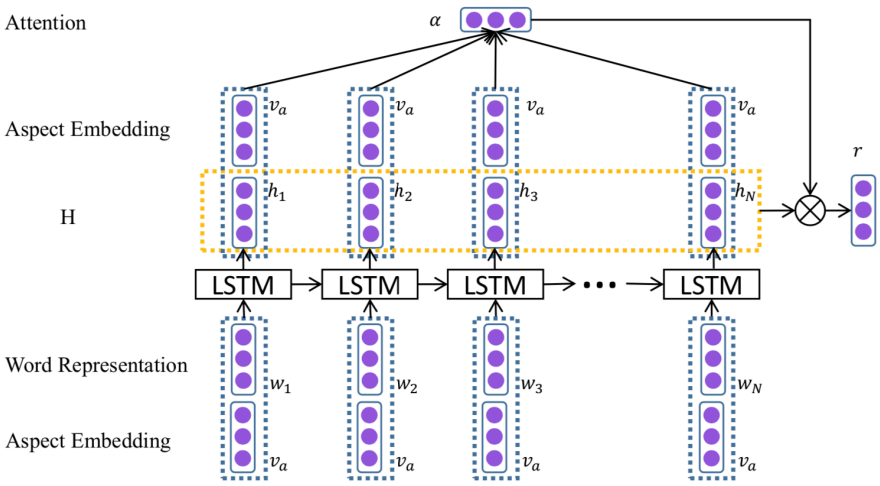


图 2.4　ATAE-LSTM 模型结构

个是把 aspect 表示和输入单词向量拼接起来。ATAE-LSTM 模型既可以处理对某一方面的情感分析（Aspect-Category Sentiment Analysis），也可以处理对某一目标的情感分析（Aspect-Term Sentiment Analysis or Target-oriented Sentiment Analysis）。ATAE-LSTM 相比于 LSTM 模型，考虑了 aspect 的信息，引入了 attention 的机制。

2.3.2　IAN

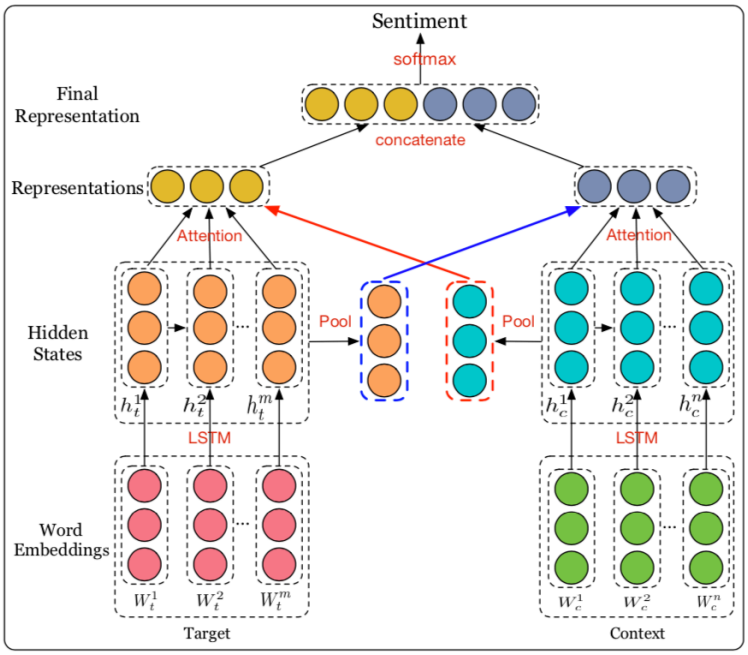


图 2.5　IAN 模型结构

在对某一目标的情感分析（Aspect-Term Sentiment Analysis or Target-oriented

Sentiment Analysis）中，目标（target）可能由多个词组成。在针对不同的句子时，目标中的单词也可能发挥着不同的重要性。因此，IAN（Interactive Attention Networks）[12] 提出目标的表示和句子的表示都需要通过交互式地学习得到。图 2.5显示了 IAN 模型的结构。

首先，IAN 使用 LSTM 处理句子和目标，得到了句子和目标的隐式向量表示。然后，通过均值 pooling 的方式，得到句子和目标的表示，并以此计算 attention 的权重，对句子和目标中的隐式表示进行加权求和。最后，将句子和目标的加权和结果拼接，作为整个输入的表示。IAN 相比于 ATAE-LSTM，使用了交互式的 attention 机制，同样考虑了目标中不同单词的不同重要性。

2.3.3　BILSTM-ATT-G

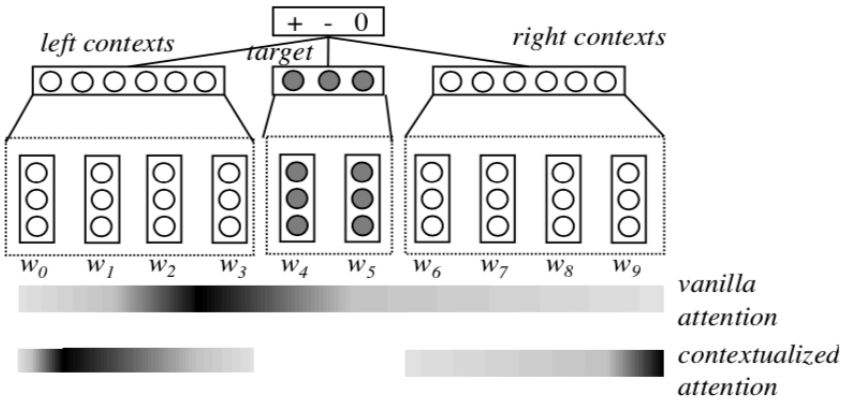


图 2.6　BILSTM-ATT-G 模型结构

在处理对某一目标的情感分析（Aspect-Term Sentiment Analysis or Target-oriented Sentiment Analysis）时，BILSTM-ATT-G（BiLSTM with Attentional Gates）[13] 除了考虑不同单词的重要性，还考虑了句子不同部分的重要性。图 2.6显示了 BILSTM-ATT-G 的结构。它用目标将句子分成了左边、右边和目标三部分。首先，它使用与上面提到的简单 attention 模型来处理句子的左边部分、右边部分和目标；然后，利用 attention 模型的输出和最后一个隐层表示来计算一个门，并通过这个门对不同的部分的结果加权求和。这一方法对模型的效果有巨大的提升。从中可以看出，句子中不同的部分有着不同的重要性，gate 方法可以很好地学习到这种重要性。

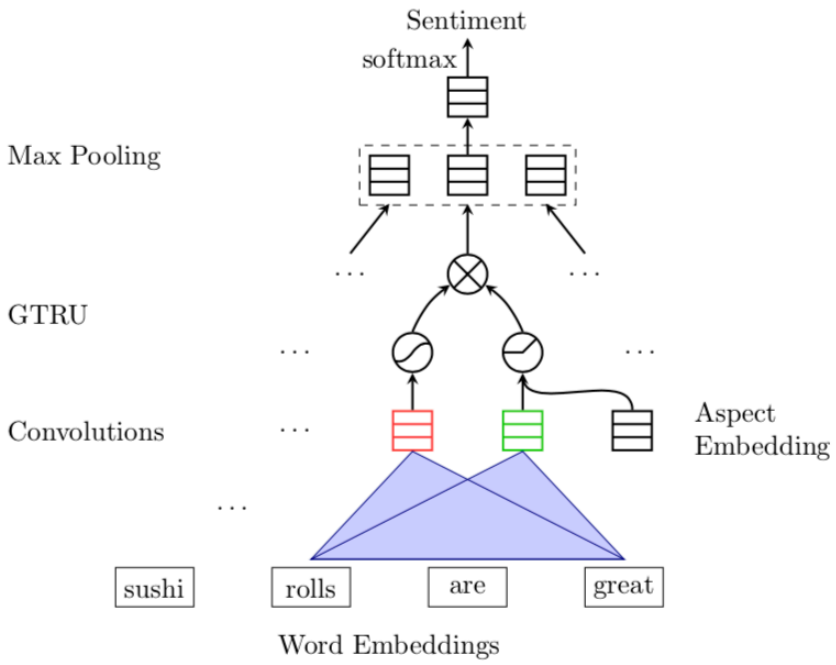


图 2.7　GCAE 模型结构

2.3.4　GCAE

在前面所提到的模型中，LSTM 等循环结构被用来处理单词的向量，得到句子的隐式表示。由于 LSTM 难以并行化，且当句子长度较长时训练困难，GCAE（Gated Convolutional Network with Aspect Embedding）开始尝试用 CNN[14,15] 解决处理句子。图 2.7显示了 GCAE 模型的结构。首先，GCAE 用一维卷积处理句子输入，得到了句子的表示；然后对句子的单词和 target 的单词进行卷积，得到控制门；将二者相乘，得到了输入的表示。为了从不同粒度处理句子的特征，GCAE 用到了多尺度的卷积，即使用了多个不同卷积核大小的卷积，并将它们的结果拼接起来。

2.3.5　Memnet

除了 attention 机制和 gate 方法，memory network[16] 也被应用到基于方面的情感分析中来。Memory network 是一项机器学习技术，在问答系统中取得了优秀的表现[16,17]。图 2.8显示了 Memnet 模型的结构。Memnet 是一个多层结构。首先，它把单词的向量表示当作记忆单元，把目标单词的均值当作初始 query；每层是一个 attention 的结构，上一层的结果作为下一层的 query；各层结果之和作为最终的输出。Memnet 利用了多层 attention 结构，它很好地解决了长期依赖的问题。但是，Memnet 也存在一些问题：它直接使用单词的表示作为记忆单元，使得难以处理词组的信息；attention 方法对单词的顺序不敏感，这样忽视了单词的时序信息。

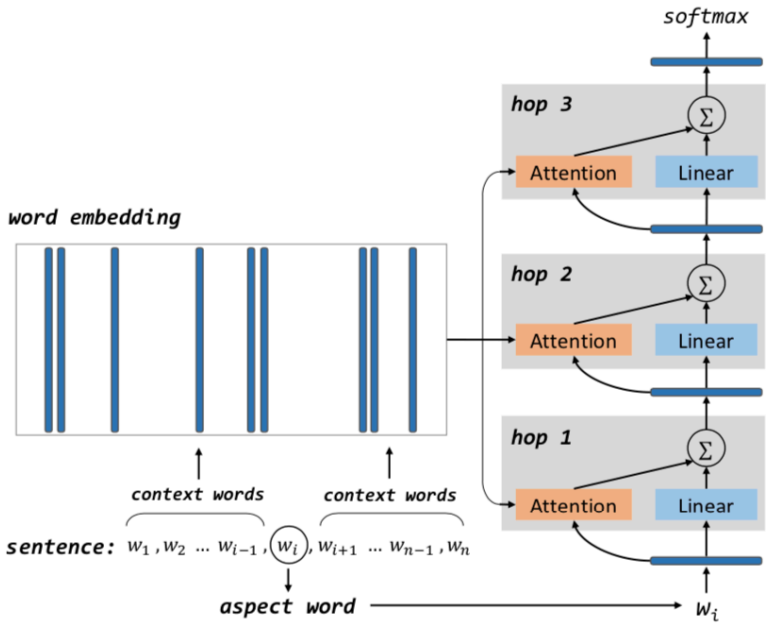


图 2.8　Memnet 模型结构

Memnet 在处理需要推断、否定和比较的问题时遇到了困难。

2.3.6　RAM

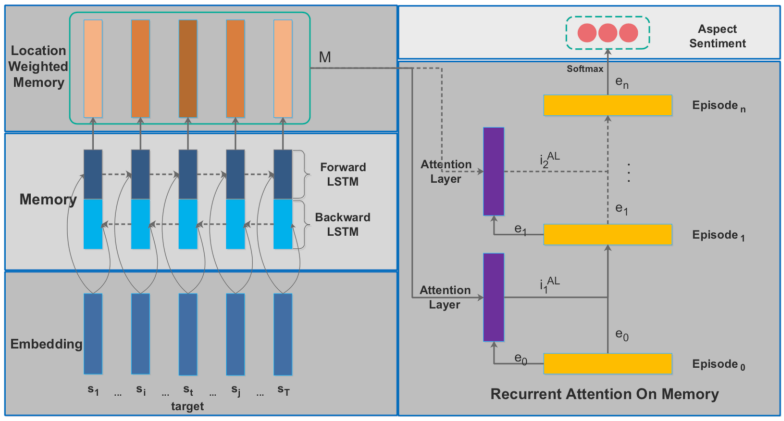
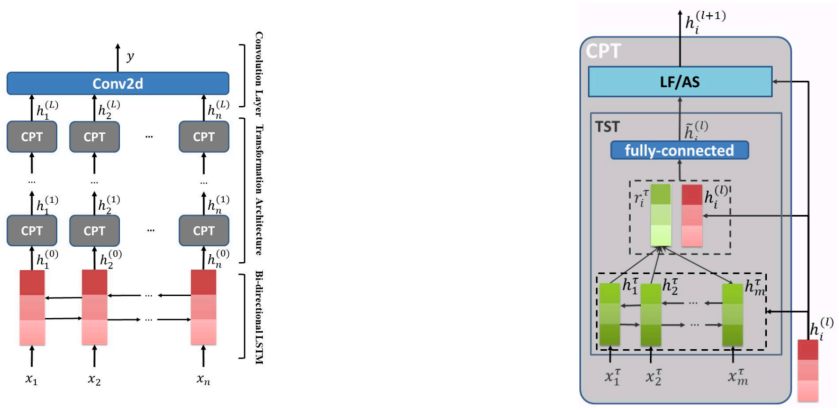


图 2.9　RAM 模型结构

针对 Memnet 的一些缺陷，RAM（Recurrent Attention Network）[18]。图 2.9显示了 RAM 模型的结构。为了解决 Memnet 不善处理词组的问题，它使用双向 LSTM 来处理句子单词，得到了结合上下文的句子表示，并以此作为记忆单元；为了使模型具备一定的推断、否定和比较的能力，它不再简单地将上一层的输出作为下一层的 query，而是利用一个[19] 单元，得到下一层的 query；同时，为了弥补 Memnet

不带有时序信息的问题，它引入了位置权重，使得与目标更近的单词权重更大。

2.3.7　TNet



(a) 整体结构　　(b) CPT 模块

图 2.10　TNet 模型

TNet（Transformation Network）[20] 集成了众多的技术，在多个公开数据集上取得了最佳的结果。图 2.10左图显示了 TNet 模型的结构。首先，它使用双向 LSTM 处理句子单词，得到结合上下文的表示，并以此作为记忆单元；然后，利用一个多层结构处理记忆单元；最后，利用二维卷积得到最终的句子表示。具体地，在每一层中，使用一个名为 CPT（Context-Preserving Transformation）的结构，其结构如图 2.10右图所示。在 CPT 结构中，对每个句子中的单词，使用 attention 方法求得与句子单词相关的目标表示，将其与上一层的记忆单元通过一个单层映射得到下一层的记忆单元。同时，TNet 还使用了更加复杂的位置权重。

第 3 章　数据

在本章中，我们将介绍本文中使用的数据集，包括一些公开数据集如 Restaurant，Laptop，Twitter 以及我们所构建的新闻驱动的股票预测数据集 senti-stock。

3.1　SemEval 2014 Task 4

SemEval 2014 Task 4 是一个基于方面的情感分析数据集，是在基于方面的情感分析领域应用最为广泛的数据集。基于方面的情感分析旨在提取句子中对某一方面或目标的情感极性。这一数据集由人工标注的餐厅 Restaurant 和笔记本电脑 Laptop 的评论组成。数据集分为训练集和测试集两部分。从表中的数据可以看出，在训练集和测试集中，积极、消极、中性的标签分布并不均衡，尤其是笔记本电脑 Laptop 评论数据集。同时，训练数据的量并不是很大。这些特点给这一数据集增加了难度。Restaurant 和 Laptop 数据集都是一个比较正式的数据，句子都比较完整，语法较规范。

表 3.1　SemEval 2014 Task 4 数据集统计信息

Dataset	Positive	Negative	Neutral
Laptop-Train	994	870	464
Laptop-Test	341	128	169
Restaurant-Train	2164	807	637
Restaurant-Test	728	196	196

表 3.1显示了 Restaurant 和 Laptop 数据集的统计信息。

3.2　Twitter

Twitter[21] 是一个由推特内容构成的数据集。这个数据集针对于对某一目标的情感分析问题，是由人工标注的。表 3.2显示了该数据集的统计信息。推特 Twitter 数据集分为训练集和测试集两部分，从表3.2中可以看出，推特 Twitter 数据集中训练集和测试集的标签分布相对来说还是比较均匀的。但推特数据集中的句子多比较口语化，涉及很多口语化的表达，句式不规范，甚至包含 emoji 表情等内容，这些都给这一数据集增加了难度。

表 3.2　Twitter 数据集统计信息

Dataset	Positive	Negative	Neutral
Twitter-Train	1567	1563	3127
Twitter-Test	174	174	346

3.3　senti-stock

在之前的研究中用到的股票预测的数据集大都是不公开的。为此，我们构建了新闻驱动的股票预测数据集 senti-stock 并将其公开。本章将详细介绍我们如何收集数据、处理数据以及数据如何获取。图 3.1显示了数据收集和处理的流程。

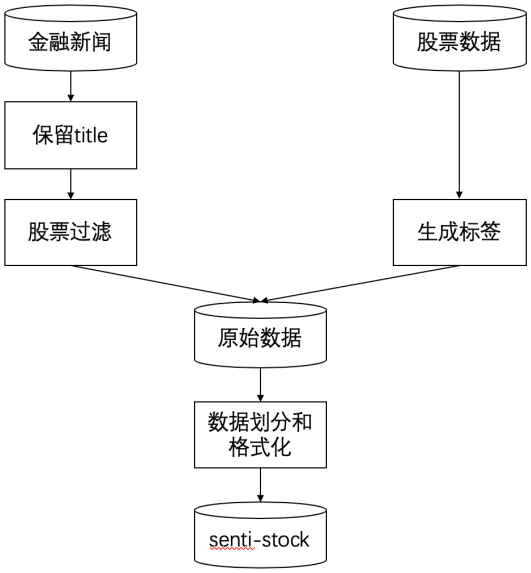


图 3.1　senti-stock 数据收集和处理流程图

3.3.1　收集数据

我们选择美国股票市场作为研究对象。美国股市是指包含纽约证券交易所 (New York Stock Exchange) 及纳斯达克证券市场 (Nasdaq Stock Market) 上市的股票。道琼斯工业股票指数、纳斯达克指数和标准普尔 500 指数三大股指代表着美国股市的兴衰。美国股票市场于 1811 年建立，至今已经有两百年的历史。十八世纪，美国股票市场得到了初步发展；十八世纪末到十九世纪初，美国股市进一步发展，但市场操纵和内幕交易情况严重；十九世纪中期以前，美国股票市场进入规范化的发展时期；十九世纪中期至今，机构投资发展迅速，美国股票市场进入现代投资时代。因为其历史悠久，美国股票市场目前已经处于一个比较规范化的发展阶段。纽约证券交易所有超过三千支股票，包括一些历史悠久的大型企业，股

份总值达到七兆亿美元；纳斯达克证券交易所是一个虚拟交易所，虽然历史较短，但有超过五千支股票，股票公司多为小型新公司。美国市场具有以下特点：

- 规模大、市场成熟、运作规范、股价稳定；
- 证券市场管理严格、规范；
- 美国允许外国股份公司在美国证券市场发行股票并进行交易。

美国股票市场交易时间是周一到周五，美股没有单日股票涨跌幅限制，实行 T+0 交易制度，当天买入的股票可以当天卖出。正因为美国股票市场具有上述特点，美国股票市场比中国股票市场更适合用来作为研究对象。

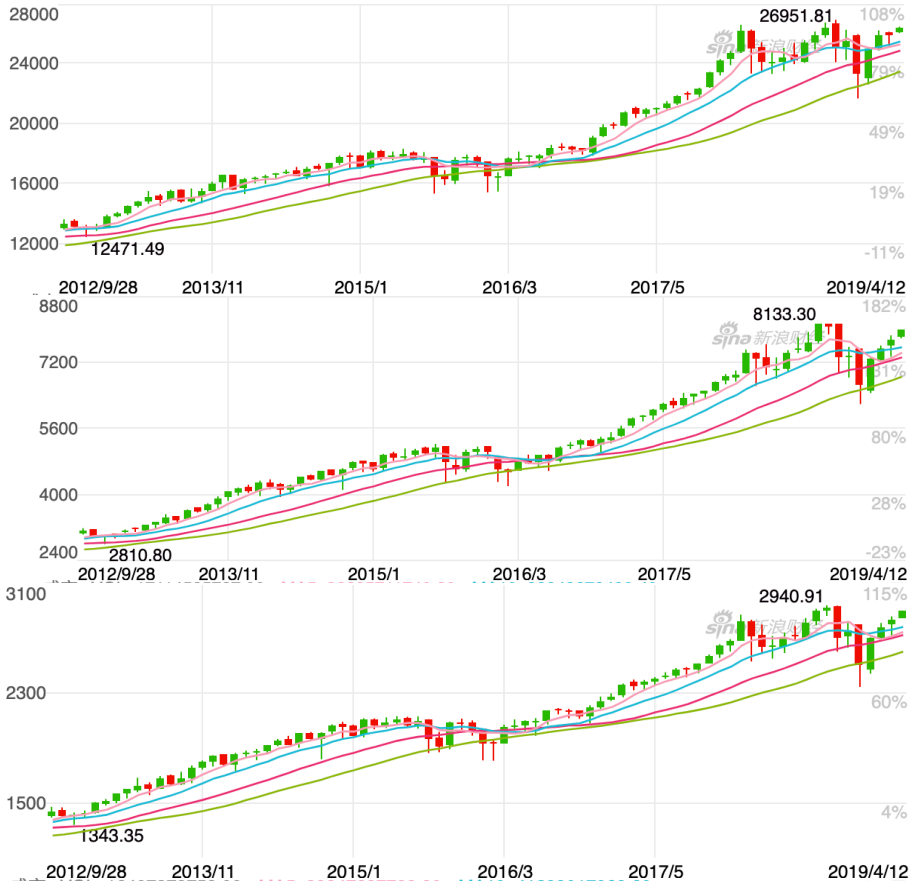


图 3.2　美国市场三大指数：道琼指数、纳斯达克指数、标普 500 的月均线，图片来自新浪财经

我们主要从路透社收集新闻数据，从雅虎财经收集股票历史价格数据。

路透社（Reuters）是世界上最早创办的通讯社之一，也是目前英国最大的通讯社和西方四大通讯社之一。路透社是世界前三大的多媒体新闻通讯社，提供各类新闻和金融数据，在 128 个国家运行。路透提供新闻报导给报刊、电视台等各式媒体，并向来以迅速、准确享誉国际。雅虎财经提供金融新闻、数据、评论等内容，包括股票历史数据、金融新闻报道和原创内容。雅虎财经提供的股票历史数

表 3.3　新闻样例

键	值
news_type	topStory
symbol	AMD
name	Advanced Micro Devices Inc
date	20161020
title	AMD revenue beats on demand for chips used in gaming consoles
content	Chipmaker Advanced Micro Devices Inc reported a better-than-expected 23.2 percent increase in quarterly revenue helped by higher demand for graphics chips used in gaming consoles.

据以准确及时而著称，并且雅虎财经提供了方便使用的数据接口。

为了从路透社网站和雅虎财经网站上抓取数据，我们使用了 Scrapy 框架。Scrapy[①]是一个为了爬取网站数据，提取结构性数据而编写的应用框架。它可以应用在包括数据挖掘，信息处理或存储历史数据等一系列的程序中。其最初是为了页面抓取 (更确切来说, 网络抓取) 所设计的，也可以应用在获取 API 所返回的数据 (例如 Amazon Associates Web Services) 或者通用的网络爬虫。我们所使用的爬取数据的源码可以在 github 上获得[②]。

我们从路透社的网站上抓取了从 2015 年 1 月 1 日到 2018 年 1 月 1 日的美国市场的股票新闻数据。图 3.2显示了美国市场三大指数（道琼指数、纳斯达克指数、标普 500 指数）在近几年的变化波动趋势。可以看到，在 2015 年到 2018 年期间，美国股市整体上呈现稳定增长的趋势。我们得到了关于 1297 支股票的共计 121096 条新闻数据。这些新闻全部以 JSON 的格式保存。表 3.3显示了我们一个新闻的样例。其中，news_type 指的是新闻的类型，包括"topStory" 和"normal" 两种。"symbol" 是股票的代码。"name" 是股票的全称。"date" 是新闻的日期，格式为"%Y%m%d"。"title" 和"content" 分别表示新闻的标题和内容。我们使用一个 python 包 fix-yahoo-finance🄵[③]来获取每支股票的历史价格数据。

```
1    from pandas_datareader import data as pdr
2    import fix_yahoo_finance as yf
3    yf.pdr_override()
4
```

① https://pypi.org/project/Scrapy/
② https://www.github.com/Cppowboy/stock_data_crawler
③ https://pypi.org/project/fix-yahoo-finance/

```
5        data = pdr.get_data_yahoo("SPY", start="2017-01-01"
             , end="2017-04-30")
6        data = pdr.get_data_yahoo(["SPY", "IWM"], start="
             2017-01-01", end="2017-04-30")
```

上面给出的 fix-yahoo-finance 的示例代码, 我们利用这一工具得到了历史数据。在这里，我们假设某支股票的涨跌公与前一天、周、月的相关新闻有关，不受其他因素的影响。然后从短期（天）、中期（周）、长期（月）三个时间间隔去生成输出的标签[3]。三种不同的时间间隔的标签构成了 Short、Middle、Long 三个不同的数据集。如果股票价格在下一天、周、月的涨幅超过 1%，则将标签设为 +1；如果股票价格在下一天、周、月的跌幅超过 1%，则将标签设为 −1；其他股票的标签设为 0。

3.3.2　处理数据

之前的研究表明，新闻的标题更有助于预测新闻事件的影响，而新闻的内容往往包含太多不相关的信息。因此，我们只使用新闻的标题[3]。图3.3显示了全部

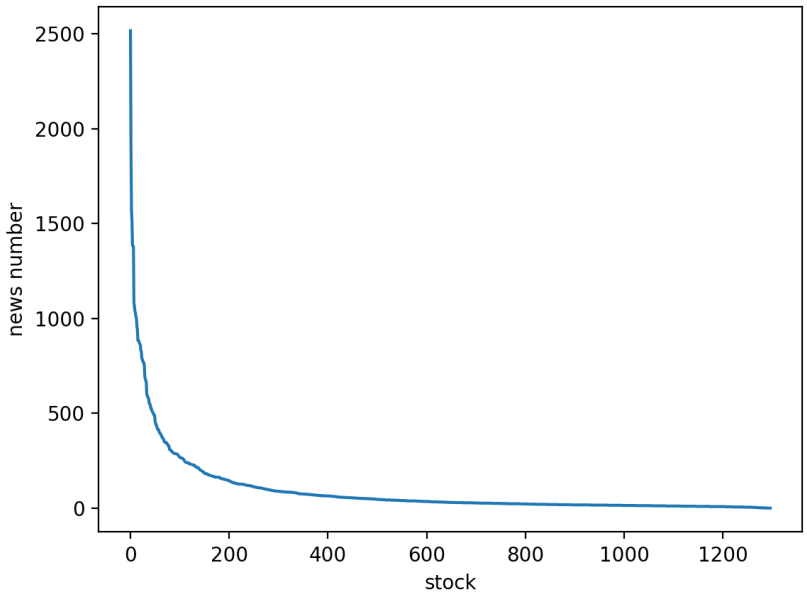


图 3.3　1297 支股票的新闻数量分布

股票的新闻数量分布。可以看到，大部分股票的相关新闻数量非常少，但有些比较知名的股票公司，其相关的新闻非常多。近期的研究[3] 显示，知名股票的预测准

确率要高于不那么知名的股票，因为与知名股票相关的新闻事件数量更高。不知名股票的日常新闻非常少，以致于难以提取足够的信息来预测股票的变动。因此，我们对所有的股票进行简单的过滤，只保留新闻数量超过五百的股票，得到 46 支股票。这 46 支股票的新闻数量分布如图3.4所示, 这 46 支股票均是在美国市场中比较有知名度的股票。

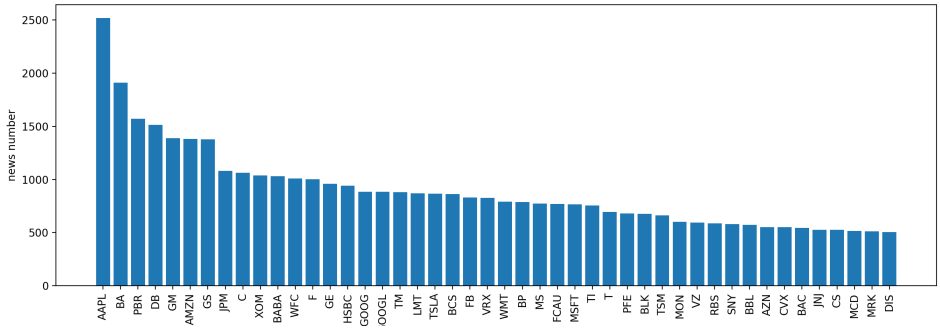


图 3.4　新闻数超过五百的股票新闻数分布

表 3.4　数据集的统计信息

数据集	全部数据	多标签数据	困难数据
Short-Train	25496	3623	33
Short-Test	6374	287	8
Middle-Train	30637	4368	176
Middle-Test	7660	353	17
Long-Train	18324	2510	157
Long-Test	4581	218	12

在完成上述处理后，我们将数据按照 4:1 的比例划分为训练集和测试集。表3.4显示了数据集的一些统计信息。其中，多标签数据指的是同一条新闻有多支相关的股票；困难数据指的是同一条新闻有多支相关股票且这些股票的涨跌情况不同，即同一条新闻对不同的股票有着不同的影响。困难数据是多标签数据的一个子集。

表3.5显示了数据集中各类标签的分布情况。在短期标签的数据集中，大部分标签的类别都是 0。这是因为在很短的时间（一天）内，股票的价格往往不会出现较大的变动。中期标签的数据集比另外两个数据集要更加均衡。长期数据集中，大部分标签的类别都是 +1，这是因为在 2015 年到 2018 年之间，股市整体上呈现稳定上升的趋势。在这三种标签的数据集中，短期数据集和和长期数据集各类标签

表 3.5　数据集中标签的分布

	+1	0	−1
Short-Train	5483	141915	5098
Short-Test	1388	3640	1346
Middle-Train	11716	8743	10178
Middle-Test	2936	2207	2517
Long-Train	9109	2470	6745
Long-Test	2194	653	1734

分布相对比较不均匀，因此相对容易，而中期数据集相对困难一些。

3.3.3　公开数据

我们的数据集 senti-stock 将会以 xml 的形式存储，其存储格式与 SemEval 2014 数据集格式一致[9]，数据集可以从 github 上下载①。

① https://www.github.com/Cppowboy/senti-stock

第 4 章　基于 transformer 和多尺度卷积的模型

为了更好地解决对某一目标的情感分析问题（Aspect-Term Sentiment Analysis，Target-oriented Sentiment Analysis），我们提出了使用 transformer 和多尺度卷积的模型。本章中，首先我们给出问题的形式化描述；然后，我们介绍 transformer 的结构；最后，介绍多尺度卷积的部分和目标函数。

4.1　问题描述

问题的输入是一个句子目标对 (w, w^{τ})，其中目标 $w^{\tau} = \{w_1^{\tau}, w\tau_2 ... w\tau_m\}$ 是句子 $w = \{w_1, w_2 ... w_n\}$ 的子串，m 是目标的长度，n 是句子的长度。对某一目标的情感分析 Aspect-Term Sentiment Analysis，Target-oriented Sentiment Analysis）的目标是提取句子所表达的对目标的情感极性 $y \in \{P, N, O\}$，P，N，O 分别表示积极、消极和中立。比如，句子“great food but the service was dreadful!”对目标“food”的情感极性是积极的，对目标“service”的情感极性是消极的。

4.2　模型介绍

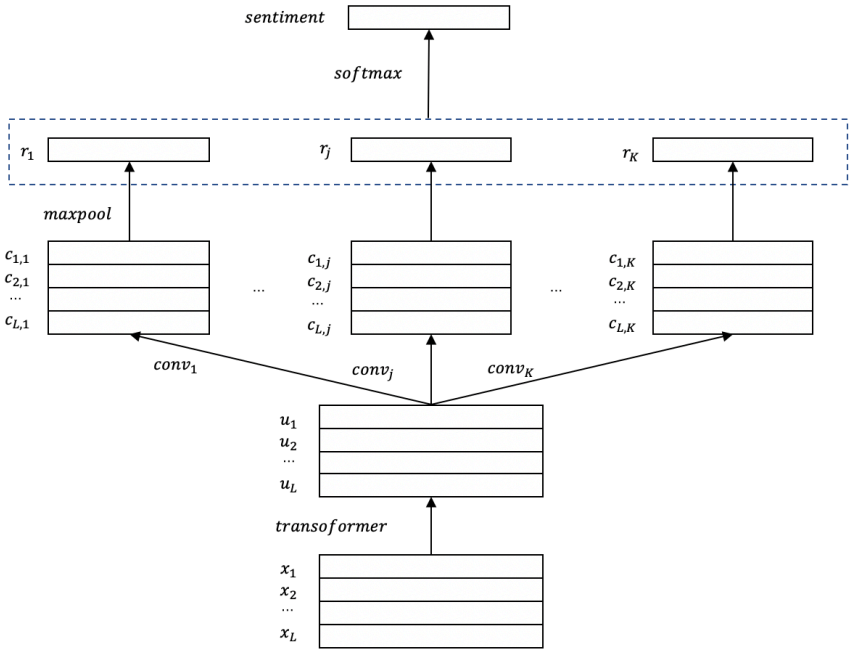


图 4.1　基于 transformer 和多尺度卷积的模型

图 4.1显示了我们所提出的模型的主要结构。我们将目标和句子拼接起来，并用多层 transformer 结构来处理；多尺度卷积可以学习变长 n-gram 特征。

4.3　Transformer 结构

循环网络结构 RNN 和卷积神经网络 CNN 经常被用来将一个变长序列映射成一个固定长度的序列。循环网络结构 RNN 一般沿着序列的时间维度，逐个地生成隐层状态。循环网络 RNN 天然地具有时序性，这也使得循环网络结构在训练过程中难以被并行。当序列长度较长时，RNN 往往难以训练。

卷积神经网络 CNN 通过局部感受野，获取输入序列的局部特征。但是，由于卷积的局部性，导致单层卷积难以学习长期的依赖关系；多层卷积可以弥补这一问题，但长期依赖的路径变长。

Transformer[5] 不依赖循环网络结构和卷积神经网络，只通过 self-attention 就将输入序列映射到固定长度的输出序列。在本小节中，我们将简要地介绍 transformer 的结构。

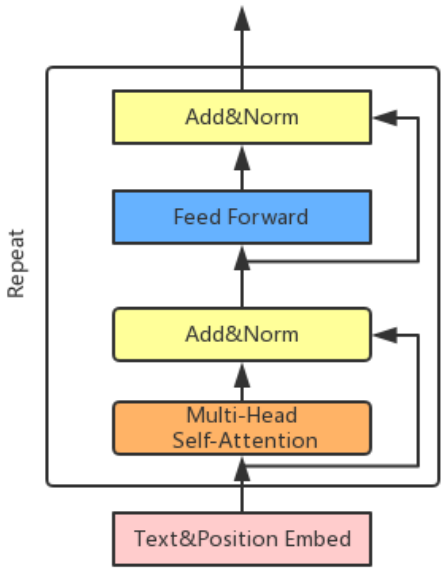


图 4.2　Transformer 结构

图 4.2显示了 transfomer 的结构。Transformer 是一个多层结构，每一层包括 self-attention 和前向网络。

4.3.1　点积 attention

点积 attention 与常见的 attention 机制不同，它的 Attention 权重是由点积求得的。对于 query $Q \in R^{T_q \times d_k}$，key $K \in R^{T_v \times d_k}$ 和 value$V \in R^{T_v \times d_v}$，attention 的输出是

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{4-1}$$

其中，T_q 是 query 的长度，T_v 是 key 和 value 的长度，d_k 是 query 和 key 的维度，d_v 是 value 的维度。$\sqrt{d_k}$ 用来地权值进行缩放，保证了权值的稳定性。

4.3.2　多头 attention

Transformer 中并没有使用单一的 attention 机制，而是用不同的线性映射将 query，key，value 映射 h 次。多次 attention 机制可以看作是单一 attention 的一种 ensemble 模型。多个映射将 query, key, value 映射到了不同的子空间, 多头 attention 侧重于不同的值。

$$\begin{aligned} MultiHead(Q, K, V) &= concat(head_1, head_2 ... head_h)W^O \\ head_i &= Attention(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \tag{4-2}$$

其中用到的映射都是不包含偏置的线性映射，其中 $W_i^Q \in R^{d_{model} \times d_k}$，$W_i^K \in R^{d_{model} \times d_k}$，$W_i^V \in R^{d_{model} \times d_v}$ 是需要学习的参数，d_{model} 是单词向量的维度。Self-attention 是 Q、K、V 相同的多头 attention。

$$SelfAttention(V) = MultiHead(V, V, V) \tag{4-3}$$

4.3.3　前向网络

在 transformer 的每一层中，除了用到了 attention 机制，还使用了一个两层全连接网络，使用 relu 激活函数。

$$FFN(x) = relu(xW_1 + b_1)W_2 + b_2 \tag{4-4}$$

4.3.4　词向量和 softmax

与其他序列模型类似，我们将输入单词和输出单词映射为一个需要学习的向量，向量的维度是 d_{model}。模型的最后，使用一个带有 softmax 激活函数的线性层

得到预测结果。

4.3.5　位置编码

因为多头 attention 机制和前向网络没有卷积层和循环结构，不能处理输入序列的时序信息，我们使用了位置编码来弥补这一缺点。这里用到的位置编码是在训练过程中学习得到的，而不是利用提前定义好的[22]。

4.4　语言模型和预训练

在本小节，我们都语言模型和使用 transformer 的语言模型做简要的介绍。一个统计语言模型通常构建为句子 w 的概率分布 $P(w)$，这里 $P(w)$ 实际上反映的是 w 作为一个句子出现的概率。对于一个由 n 个词按顺序构成的句子 $w = \{w_1, w_2 ... w_n\}$，$P(w)$ 实际上求解的是字符串的联合概率，利用贝叶斯公式，链式分解如下：

$$P(w) = P(w_1, w_2 ... w_n) = P(w_1)P(w_2|w_1)P(w_3|w_1, w_2)...P(w_n|w_1, w_2 ... w_{n-1}) \tag{4-5}$$

从上面可以看出，语言模型可以看成，在给定了前一个词的情况下，求下一个词出现的条件概率。在这里，我们使用的是基于 transformer 的语言模型。设 W_e 是词向量表示，W_p 是位置编码，h_l 表示第 l 层的中间结果。初始句子表示为

$$h_0 = UW_e + W_p \tag{4-6}$$

我们用 $transformer\_block$ 表示 transfomer 的每一层。

$$h_l = transformer\_block(h_{l-1}) \tag{4-7}$$

则最终的概率分布为

$$P(w) = softmax(h_n W_e^T) \tag{4-8}$$

语言模型的训练只需要句子本身，因此其训练数据非常的充足。之前的研究表明，使用在大规模语料库上预训练的语言模型的参数，可以大大提升分类任务的效果。我们使用 OpenAI GPT[22] 提供的预训练的语言模型，将语言模型的输出用作特征。OpenAI GPT 是在 BooksCorpus 数据集上训练的，它包括七千本未发表的书籍，内

容包括了冒险、奇幻、史诗类作品。另一个数据集是 1B Word Benchmark[23]，它与 BookCorpus 有类似的大小。在这个语料库上训练该语言模型，直到达到了 18.4 的困惑度（perplexity）。我们用预训练得到的参数去初始化 transformer 的参数，之后在对某一目标的情感分析任务中去调优这些参数。

本小节中使用的符号是为了解释语言模型的相关问题，与其他章节的符号无关，并不冲突。

4.5　多尺度卷积

我们把对某一目标的情感分析问题（Aspect-Term Sentiment Analysis，Target-oriented Sentiment Analysis）转化为一个序列对分类问题。具体地，我们把句子 $w = \{w_1, w_2 ... w_n\}$ 和目标 $w^\tau = \{w_1^\tau, w_2^\tau ... w_m^\tau\}$ 拼接起来，在开头添加一个 *start* 记号，在结尾添加一个 *end* 记号，并在句子和目标之间添加一个 *del* 记号。拼接得到的序列记为 X，$X = \{start, w_1, w_2 ... w_n, del, w_1^\tau, w_2^\tau ... w_m^\tau, end\} = \{x_1, x_2 ... x_L\}$，其中 $L = m + n + 3$。将 X 输入到一个多层 transformer 结构当中，得到每个单词的表示。在这里，为了简洁，我们用下面的方式表示多层 transformer 结构。

$$u_1, u_2 ... u_L = transformer(x_1, x_2 ... x_L)t \tag{4-9}$$

一般情况下，*start* 记号的表示 u_1 被用来当作整个输入的表示。这种方式忽略了其它单词的表示，而这些信息可能会对预测正确的情感极性有着重要的作用。为了解决这一问题，我们使用多尺度的卷积来处理所有单词的表示。在句子中，有许多不同长度的短语。多尺度的卷积可以提取不同长度短语的特征。我们使用了多个不同卷积核大小的卷积层，并在序列的两端补零使不同卷积的输出大小保持一致。K 是卷积的数量。

$$c_{i,j} = tanh(u_{i:i+k_j} * W_j + b_j) \tag{4-10}$$

其中，W_j 和 b_j 是第 j 个卷积的参数，k_j 是第 j 个卷积核的大小。我们使用 max pooling 层来选取最重要的特征。

$$r_j = max(c_{1,j}, c_{2,j} ... c_{L,j}) \tag{4-11}$$

把 max pooling 的输出结果拼接起来，得到句子的最终表示。

$$r = concat(r_1, r_2 ... r_K) \tag{4-12}$$

最后，使用带有 softmax 激活函数和全连接层得到预测结果。

$$\hat{y} = softmax(Wr + b) \tag{4-13}$$

其中 W 和 b 是全连接层的参数。

4.6　目标函数

训练过程中，我们的目标是最小化损失函数，损失函数是语言模型的损失函数和分类损失函数的和。

$$Loss = Loss_{lm} + Loss_{clf} \tag{4-14}$$

语言模型的损失函数为

$$Loss_{lm} = \sum_i \sum_n \hat{x_{i,n}} log(x_{i,n}) \tag{4-15}$$

其中，x_i 是语言模型的第 i 个输出，n 是单词的序号。分类损失函数是真实值 y 和预测值 $\hat{y}$ 的交叉熵。

$$Loss_{clf} = \sum_i \sum_j \hat{y_{i,j}} log(y_{i,j}) \tag{4-16}$$

其中，i 是样本的序号，j 是输出类别的序号。

第 5 章　基于 attention 机制的多股票关系模型

为了解决新闻驱动的股票预测问题，我们把它转化为针对某一方面的情感分析问题（Aspect-Category Sentiment Analysis）。我们提出了 MSRA（Multi-Stock Relation using Attention mechanism，使用 attention 的多股票关系模型）模型，该模型包括情感分析模块和股票关系模块。在本章中，我们首先给出总是的形式化描述，并简要介绍模型的结构，然后分别介绍情感分析模块和股票关系模块。

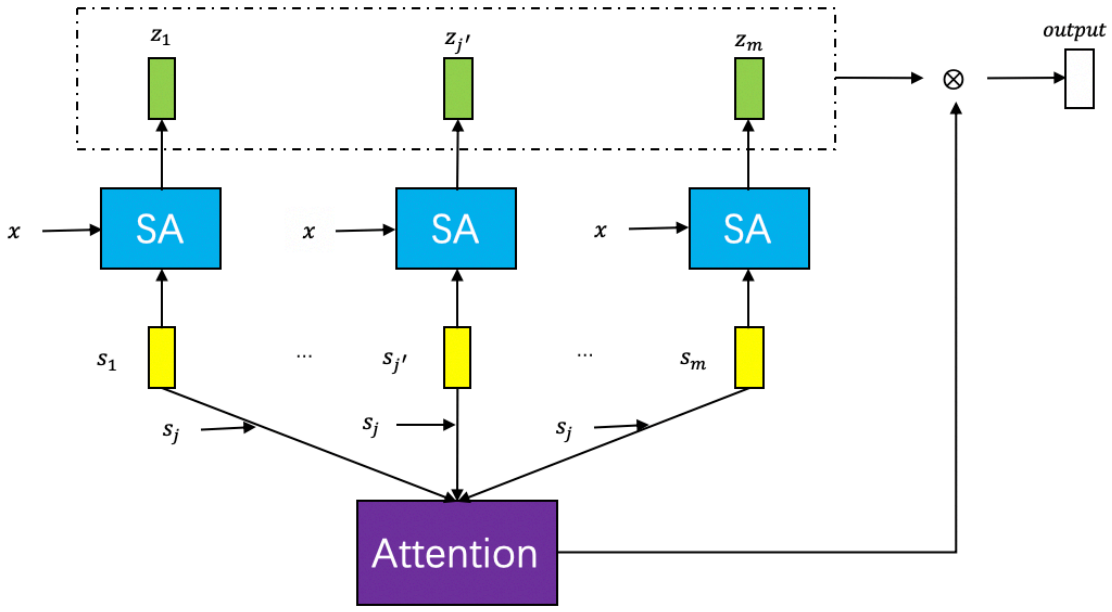


图 5.1　MSRA 模型结构

5.1　问题描述

问题的输入包括一支股票 s 以及一条与这支股票相关的新闻 $w = \{w_1, w_2 ... w_n\}$，其中 n 是新闻句子的长度。句子相应的词向量表示是 $x = \{x_1, x_2 ... x_n\}$。我们使用股票向量 s_j 来表示第 j 支股票，其中 $j = 1, 2 ... m$，m 是所有股票的数量。问题的输出是股票价格在下一个时间段内将会如何变化 $y \in \{+1, 0, -1\}$，其中，$+1$ 表示股票的价格将会在下一天、周、月之后上涨，-1 表示股票的价格将会在下一天、周、月之后下跌，其它的为 0。$C = 3$ 是输出标签的数量。比如，“微软收购诺基亚的手机部门”将会使微软 (MSFT) 的股票价格上涨，诺基亚 (NOK) 的股票价格下跌。

5.2　模型结构

图 5.1展示了模型的基本结构。我们使用情感分析模块提取与股票 j 相关的新闻 w 的情感表示；使用点积 attention 来学习股票之间的相关关系。所有股票的情感表示的加权和被用来作为输入的最终表示。最后使用带有 softmax 激活函数的全连接得到输出结果。

5.3　股票向量

在预测股票价格变化时，股票相关的信息有着重要的作用。同一条新闻，对不同的股票可能会产生完全不同的影响。为了更好地利用股票相关的信息，我们提出了股票向量的概念。我们为全部的股票构建一个词表，并为每支股票创建一个向量表示。向量 $s_i \in R^{d_s}$ 表示的是第 i 支股票的向量表示，其中 d_s 表示的是股票向量的维度。$S \in R^{d_s \times |S|}$ 由全部的股票向量组成。据我们所知，这是第一次提出股票向量的概念。

5.4　情感分析模块

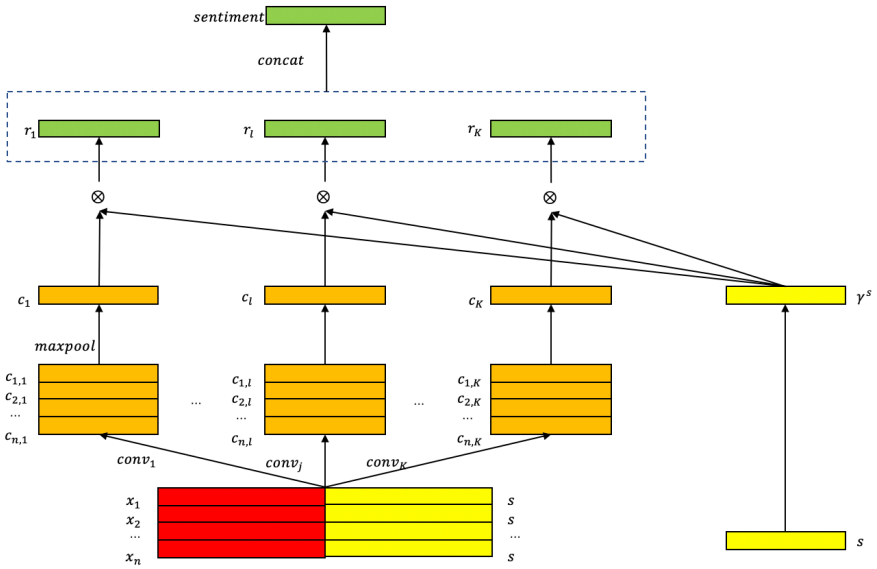


图 5.2　情感分析模块

如图 5.2所示，情感分析模块使用了带有股票系数的多尺度卷积。

5.4.1　多尺度卷积

我们利用多尺度卷积来处理词向量。自然语言处理中常使用 LSTM 来处理句子序列，LSTM 沿着时间维度，逐次计算下一个时间的隐层状态，从而将变长的序列转化为了定长输出。然后，因为 LSTM 的时序性，使得它难以并行化。CNN 最初用于图片分类，CNN 通过局部感受野，能够提取局部的特征。多层 CNN 在图像相关领域取得了非常优秀的表现，然而，CNN 在自然语言处理领域一直没有得到很好的应用。直到 TextCNN[24] 提出，CNN 才开始在自然语言处理中有所表现。相比于 LSTM，CNN 使用卷积核处理序列中局部的序列，更易于并行化。为了更好地结合股票相关信息，我们将句子的词向量 x_i 与股票向量 s 拼接起来。

$$u_i = concat(x_i, s) \tag{5-1}$$

卷积操作结合相邻的若干个单词的信息，相当于不同长度短语的特征。多尺度卷积使用了多个不同卷积核大小的卷积层，这样，可以提取不同长度的短语的特征。

$$c_{i,l} = relu(u_{i:i+k_l} * W_l + b_l) \tag{5-2}$$

其中，W_l 和 b_l 是第 l 个卷积的参数，k_l 是第 l 个卷积的卷积核大小，$l = 1, 2...K$，K 是不同尺度卷积的数量。$relu$ 激活函数的定义如下。

$$relu(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{5-3}$$

Max-pooling 层选取值最高的特征，其思想在于把最重要的特征选取出来。同时，max-pooling 层可以将变长的序列转化为固定的长度，方便后续的处理。

$$c_l = max(c_{1,l}, c_{2,l}...cn, l) \tag{5-4}$$

之前的研究表明[24]，多层 CNN 在自然语言处理中表现并不突出，却使得模型更深，更加难以训练，因此，我们这里只使用了单层卷积，并未使用多层的 CNN 结构。

5.4.2　股票系数

不同的股票具有不同的特性，有些股票更倾向于上涨，有些股票更倾向于下跌。为了更好地利用股票的特性，我们计算股票系数，它反映了股票的信息。

$$\gamma^s = tanh(W_s x^s + b_s) \tag{5-5}$$

其中, W_s 和 b_s 是需要学习的参数。$tanh$ 函数的定义如下：

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{5-6}$$

我们用按元素的乘法将股票系数和句子表示结合起来。这里使用按元素的乘法，股票系数起到了对原有特征进行缩放的作用，从而使得本来的特征带有了股票的相关信息。

$$r_l = c_l \otimes \gamma^s \tag{5-7}$$

全部卷积的结果拼接起来，得到句子的最终表示。

$$r = concat(r_1, r_2 ... r_K) \tag{5-8}$$

5.5　股票关系模块

股票公司之间的竞争、合作、上游、下游关系，可能会影响股票价格的变动。比如，“亚马逊向中国及其他七个国家的卖家提供贷款”对亚马逊（AMZN）来说是一条积极的新闻，它将会使得亚马逊的股票价格上升；同时，它也会对亚马逊在中国的竞争者阿里巴巴（BABA）造成消极的影响，使得其股票价格下跌。我们使用点积 attention 来学习股票之间的相互关系。为了简便，我们使用下面的公式来表示情感分析模块的输出。

$$z_{j'} = SA(x, s_{j'}) \tag{5-9}$$

其中，$j' = 1, 2...m$ 是股票的序号。Attention 的权重通过下面的公式求得。我们把所有股票的情感表示的加权和作为输入的最终表示。

$$\alpha_{j'} = \frac{s_j^T s_{j'}}{\sum_{j'=1}^{m} s_j^T s_{j'}} z \quad = \sum_{j'=1}^{m} \alpha_{j'} z_{j'} \tag{5-10}$$

5.6 目标函数

最后，由全连接层和 softmax 激活函数得到最终的预测结果。

$$\hat{y} = softmax(Wz + b) \tag{5-11}$$

其中，W 和 b 是需要学习的参数。目标函数是实际值 y 和预测值 $\hat{y}$ 之间的交叉熵。

$$loss = \sum_{p}^{N} \sum_{c}^{C} \hat{y_{i,j}} log(y_{i,j}) \tag{5-12}$$

其中，p 是样本序号，c 是类别序号，N 是样本数量，C 是类别数量。

第 6 章　实验

在本章中，我们分别介绍对某一目标的情感分析（Aspect-Term Sentiment Analysis or Target-oriented Sentiment Analysis）和新闻驱动的股票预测的实验设置及实验结果。

6.1　对某一目标的情感分析

在本小节，我们介绍对某一目标的情感分析（Aspect-Term Sentiment Analysis or Target-oriented Sentiment Analysis）的实验。这里，我们使用的是基于 transformer 和多尺度卷积的模型。

6.1.1　实验设置

我们在 Restaurant、Laptop、Twitter 三个公开数据集上进行了实验。这三个数据集均在第 3中有所介绍。我们沿用了之前工作中数据预处理的方法[20,25]，我们除去了一些冲突的标签，所有单词均转化为小写，不去除任何的停止词、符号、数字。我们使用 nltk①工具提供的分词工具对全部的句子进行分词。所有的句子都使用“PAD”字符补齐成最大长度。

我们使用准确率和 macro-averaged F1 score 作为评价指标。对每一类来说，精确率 $P = \frac{TP}{TP+FN}$，召回率 $R = \frac{TP}{TP+FP}$，F1 score 由 $\frac{2PR}{P+R}$ 求得。Macro-averaged F1 score 是所有类别的 F1 score 的均值[25]。

我们与下列模型进行了对比:

- Majority 把训练集中出现频率最高的类别作为预测类别;
- SVM 是支持向量机模型，它使用的是人为构建的 n-gram 特征、解析特征和语义特征[26];
- AE-LSTM 是一个将目标和句子拼接起来作为输入的 LSTM 模型;
- ATAE-LSTM 对 AE-LSTM 进行了扩展，利用 attention 机制选取最重要的单词[10];
- IAN 利用交互式地方式，分别使用 attention 方法计算句子和目标的表示[12];
- BILSTM-ATT-G 使用控制门来控制目标左边、右边部分的重要性[13];
- GCAE 使用卷积神经网络和控制门，得到了更高的准确率，更容易并行化[27];

① http://www.nltk.org/

- Memnet 将词向量当作记忆单元，使用多层 attention 方法来得到最终表示，为克服 attention 机制不能获取时序信息的缺点，它还使用了位置权重[28]；
- RAM 对 Memnet 做出了改进，它使用双向 LSTM 的结果作为记忆单元，使用 GRU 来生成下一层的表示，同时使用了与 Memnet 不同的位置权重[18]；
- TNet 提出生成与目标相关的句子表示，同时结合上下文信息。[20].

我们使用 pytorch①框架实现了这些模型中除 IAN 之外的模型，并使它们的实验结果尽可能地与原论文相似。每个模型都是独立训练的。我们使用的是 GloVe[29] 词向量来初始化我们的词向量，并在训练的过程中调整词向量的值。在我们的模型中，我们使用的模型参数设置与 OpenAI GPT[22] 一致。具体地，transformer 的层数是 12，多头 attention 的 head 数是 12，词向量的维度设为 768，中间层的维度设为了 3072。我们首先加载 OpenAI GPT[22] 的预训练参数，然后与后续结构一起进行调优。我们使用了五个不同的卷积核，卷积核大小从 1 到 5。卷积的输出通道数设为 100。我们使用 Adam[30] 优化器，学习率设为 6.25e-5。模型在 20 轮训练内就得到了最好的效果。

6.1.2　实验结果

表 6.1　对某一目标的情感分析实验结果（%）。带“*”的实验结果是从原文中获取的。

Models	Restaurant		Laptop		Twitter	
	ACC	Macro-F1	ACC	Macro-F1	ACC	Macro-F1
Majority	65.00	-	53.45	-	50.00	22.22
SVM	80.89	-	72.10	-	63.40	63.30
LSTM	76.70	63.57	69.28	63.30	66.04	63.46
ATAE-LSTM	77.23	63.73	69.44	63.46	71.24	69.19
IAN	78.60*	-	72.10*	-	-	-
BILSTM-ATT-G	79.20	67.07	71.32	64.88	71.68	70.37
GCAE	78.12	62.50	70.38	64.02	72.40	70.89
MemNet	77.86	64.47	68.18	62.46	69.80	66.86
RAM	78.30	65.42	71.63	66.73	71.24	68.75
TNet	78.39	65.37	73.98	68.64	72.11	70.01
Ours	84.20	76.35	78.21	73.31	72.98	71.40

表 6.1显示了对某一目标的情感分析的实验结果。我们的模型在 Restaurant、Laptop、Twitter 数据集上均取得了最佳的实验结果。我们在 Restaurant 数据集上取

① https://pytorch.org

得了 84.20% 的准确率（提升 5.81%），在 Laptop 数据集上取得了 78.21% 的准确率（提升 4.23%），在推特数据集上取得了 72.98% 的准确率（提升 0.87%）。

在所有的神经网络模型中，LSTM 模型效果最差。ATAE-LSTM 考虑到了目标并使用了 attention 方法，效果有所提升。IAN 使用了句子对目标的 attention 和目标对句子的 attention，更进一步提升了实验效果。BILSTM-ATT-G 和 RAM 在 Restaurant 和 Laptop 上的实验效果较好，但在 Twitter 数据集上效果提升并没有那么大，可见 LSTM 不擅长处理 Twitter 数据中大量的口语化文本。TNet 使用了 CNN 和 LSTM，并在三个数据集上都取得了不错的效果。

不同于这些模型，我们的模型使用 transformer 提取句子特征，更易于处理长期依赖关系，易于并行。我们的模型利用多尺度卷积，从不同粒度提取特征。我们的模型在三个数据集上均取得了最佳的效果。

6.1.3　预训练

为了证明 transformer 预训练的效果，我们进行了对比实验：一组使用预训练的 transformer 参数，另一组完全随机初始化参数，并从头开始进行训练。表 6.2显

表 6.2　预训练的作用

Models	Restaurant		Laptop	
	ACC	Marco-F1	ACC	Marco-F1
w/o pre-training	69.20	48.16	64.89	59.25
w/ pre-training	84.20	76.35	78.21	73.31

示了对比实验的实验结果。当不使用预训练的参数时，实验效果大大下降。在大规模语料下的预训练，使得 transformer 可以获取句子的语义信息。

6.1.4　多尺度卷积

为证明多尺度卷积的重要性，我们将模型中的多尺度卷积去掉，对比了简化后的模型与完整模型的实验效果。表 6.3显示了对比实验的结果。多尺度卷积能够

表 6.3　多尺度卷积的作用

Models	Restaurant		Laptop	
	ACC	Marco-F1	ACC	Marco-F1
w/o cnn	83.39	74.40	77.43	72.42
w/ cnn	84.20	76.35	78.21	73.31

利用全部单词的表示信息，提取不同长度的短语的表示，并选取最为重要的信息。多尺度卷积将 Restaurant 数据集上的准确率提升约 0.81%，将 Laptop 数据集上的准确率提升约 0.7%。

6.1.5　样例分析

表 6.4　结果示例。输入目标用中括号标记，正确输出标签以标的形式给出

Sentence	ATAE-LSTM	GCAE	Ours
[Coffee]$_P$ is a better deal than overpriced sandwiches.	P	$O^{\times}$	P
But make sure you have enough room on your credit card as the [bill]$_P$ will leave a big dent in your wallet.	P	$O^{\times}$	P
Aww, it 's okay... You have a [PSP]$_P$. :D That 's good already.	$O^{\times}$	P	P
I hate my [iPod]$_N$! It's dead! dead dead dead! ! ! Someone wanna fix it for me?	$O^{\times}$	N	N
I have never had a bad [meal]$_P$ (or bad service) at pigalle.	$N^{\times}$	$N^{\times}$	P
The [staff]$_N$ should be a bit more friendly.	$P^{\times}$	$P^{\times}$	N
It's a basic pizza joint, not much to look at, but the [pizza]$_P$ is what I go for.	$N^{\times}$	$O^{\times}$	P

表 6.4显示了一些预测的例子。输入目标用中括号标记，正确输出标签以下标的形式给出。其中 P、N、O 分别表示积极、消极和中立。比如，在第一个句子中，对 “coffee” 的情感极性是积极的。我们的模型比 ATAE-LSTM 和 GCAE 的预测结果要好。前两行中，句子的句式比较正式，LSTM 的模型更擅长解决此类问题。后两行中，句子比较口语化，句子结构比较零散，CNN 更擅长解决此类问题。Transformer 的模型经过了大规模语料库的预训练，在处理这两种情况时都有不错的表现。另外，我们的模型具备一定的推理和比较的能力，第五、六行显示了相应的例子。最后一行，表示情感的词是 “what I go for”，因为使用了多尺度卷积，我们的模型可以提取到短语的信息。

6.2　新闻驱动的股票预测

在本小节中，我们介绍新闻驱动的股票预测相关实验。我们使用的是 MSRA 模型（Multi-Stock Relation model using Attention mechanism，使用 attention 机制的多股票关系模型）。

6.2.1　实验设置

我们在前面提到的 senti-stock 数据集上进行实验。所有单词都转化为小写，不移除任何停止词、符号或数字。我们使用 nltk[①]工具提供的分词工具对全部的句子进行分词。所有句子用"PAD" 符号补齐到最大长度。我们以准确率和 F1 平均值为主要的评价指标。对每一类而言，精确率 $P = \frac{TP}{TP+FN}$，召回率 $R = \frac{TP}{TP+FP}$，F1 值由 $\frac{2PR}{P+R}$ 求得。其中，TP、TN、FN、FP 分别表示真正例、真负例、假负例、假正例的个数。最终的 F1 平均值是各个类的 F1 值的均值。相比于准确率，F1 值更能评价在不均衡数据集下预测结果的好坏。

在前面提到了基于方面的情感分析模型中，有一些模型要求目标必须出现在句子当中，或目标不止包含有一个单词，因此不能应用到新闻驱动的股票预测这一场景当中。因此，我们的模型只与下列模型进行了对比：

- Majority 将训练集中大多数的标签作为测试集的预测标签。
- LSTM 利用简单的 LSTM 来处理新闻句子信息，不考虑股票的相关信息。
- ATAE-LSTM 在 LSTM 的基础上，引入了 attention 机制，结合了股票信息和新闻句子信息。
- GCAE 使用 CNN 来处理新闻句子，通过 CNN 计算得到了 gate 来调整预测的结果。

我们使用 pytorch[②]复现了上述模型。每个模型都是独立训练的。每个模型都使用 GloVe[31] 预训练的词向来初始化词向量，词向量的维度是 300。CNN 的卷积核大小从 3 到 5，卷积的输出通道数设为 100。我们使用 AdaGrad[32] 优化器，学习率设为 0.01。模型的实现是开源的[③]。

6.2.2　实验结果

表 6.5显示了在 senti-stock 数据集上的主要实验结果。LSTM 在所有神经网络模型中效果最差，因为它只考虑了新闻句子的信息，没有考虑股票的相关信息。ATAE-LSTM 引入了 attention 机制，考虑了股票相关信息，使得它相比 LSTM 有

① http://www.nltk.org/
② https://pytorch.org
③ https://www.github.com/Cppowboy/icann2019

表 6.5　senti-stock 数据集上的实验结果

Models	Short		Middle		Long	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
Majority	0.5710	0.2423	0.3832	0.1847	0.4789	0.2158
LSTM	0.6316	0.5542	0.5655	0.5594	0.6162	0.5641
ATAE-LSTM	0.6401	0.5688	0.5687	0.5663	0.6241	0.5791
GCAE	0.6828	0.6072	0.6097	0.6047	0.6630	0.6161
MSRA	**0.6920**	**0.6198**	**0.6273**	**0.6237**	**0.6826**	**0.6387**

所提升。GCAE 使用了 CNN 来处理新闻标题这种比较简洁的文本，取得了更好的效果。我们所提出的模型在三种不同的时间间隔上的预测效果最好。

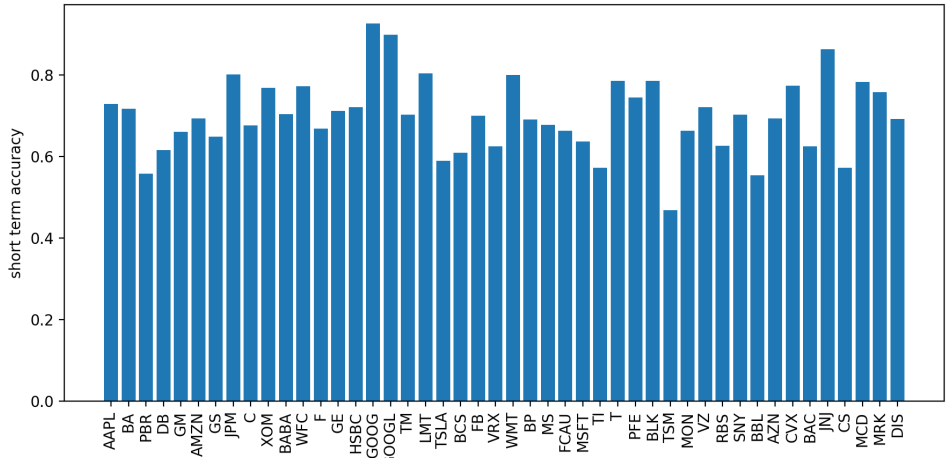


图 6.1　各股票短期预测准确率

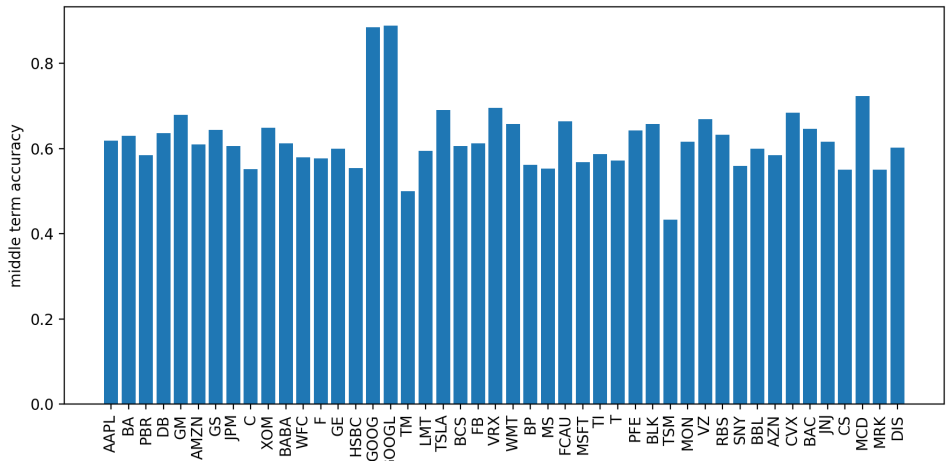


图 6.2　各股票中期预测准确率

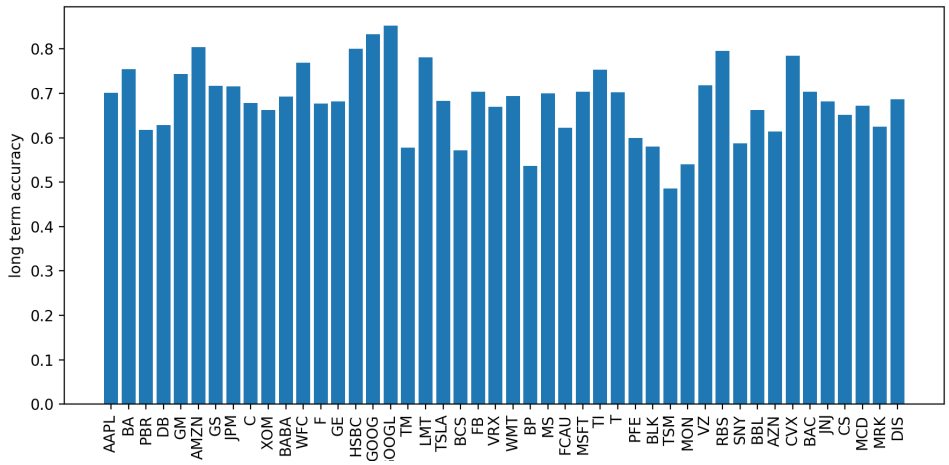


图 6.3　各股票长期预测准确率

图6.1、图6.2、图6.3分别显示了各支股票的短期、中期、长期的预测准确率。短期预测和长期预测的准确率明显比中期预测要高，大部分股票的短期和长期预测准确率均超过了百分之五十。而中期预测的准确率相对较低，大部分股票的中期预测准确率超过了百分之四十。谷歌（GOOG）在短期、中期和长期的预测准确率均超过了百分之八十，说明谷歌的股票价格与其相关新闻关系非常密切。

6.2.3　模型简化实验

表 6.6　模型简化实验结果

Models	Short		Middle		Long	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
CNN	0.6498	0.5786	0.5813	0.5761	0.6490	0.5961
MCNN	0.6616	0.5852	0.5920	0.5871	0.6531	0.6005
MCSC	**0.6955**	**0.6208**	0.6236	0.6205	0.6765	0.6275
MSRA	0.6920	0.6198	**0.6273**	**0.6237**	**0.6826**	**0.6387**

为了证明 MSRA 模型各个部分的有效性，我们进行了模型简化实验。CNN 是一个基本的卷积神经网络模型，它只考虑新闻句子的输入，不考虑股票的信息。MCNN 使用了多尺度卷积，可以提取不同长度的短语的信息。MCSC 使用了股票系数，股票系数引入了股票相关信息。MSRA 是我们的完整模型，它使用了 attention 机制，学习股票之间的相关关系。从表 6.6可以看到，多尺度卷积和股票系数都可以明显地提升模型的性能。股票关系模型对中期和长期的预测有更大的作用。一种可能的解释是，股票关系模型反映了股票之间的相互关系，短期来看，这些关

表 6.7　新闻驱动的股票预测样例分析

news title	stock	ATAE-LSTM	GCAE	MSRA
Tesla delivers quarterly record of 25000 vehicles in first quarter	$TSLA_{+1}$	+1✓	-1✗	+1✓
Taiwan stocks hit over 18-mth highs TSMC up ahead of Q4 result	TSM_{+1}	+1✓	0✗	+1✓
AT&T to offer hulu to customers this year.	T_{+1}	0✗	+1✓	+1✓
Consumer reports says Tesla misunderstands 'positive' Model 3 rating	$TSLA_{-1}$	+1✗	-1✓	-1✓
Tesla has recalled 53 000 of its model s model x cars	$TSLA_{+1}$	-1✗	-1✗	-1✗

系可能作用不大，但股票之间的关系对股票长期的变动影响更大。

6.2.4　样例分析

为了进一步分析 MSRA 模型的特点，我们在表6.7中展示了新闻驱动的股票预测的样例，正确预测值以下标的形式给出。我们比较了以 ATAE-LSTM 为代表的 LSTM 模型和以 GCAE 为代表了 CNN 模型，以及我们所提出的股票关系模型。在前两行所展示的例子中，句子都是比较正式，LSTM 更擅长处理此类问题。在第三行的例子中，因为标题的简洁性，标题与普通句子的语法并不完全一致，CNN 能够很好地处理此类问题。第四行中，“misunderstands 'positive' rating”，LSTM 误将“positive”识别为积极的消息，CNN 模型可以将误解与积极联系成一个短语。另外，在最后一行的例子中，模型预测消息是消极的，从人的主观观察来看，这一结果是正确的，但是，它与股票实际的涨跌不同。因为股票的涨跌受到许多因素的影响，新闻并不是决定股票涨跌的唯一因素。为了进一步提升股票预测的准确率，我们可以考虑补充股票历史价格、技术指标等其他信息，这也是后续研究的方向之一。

6.2.5　风险提示

利用 MSRA 模型进行新闻驱动的股票预测是对历史经验的总结，存在失效的可能。股市有风险，投资须谨慎。

第 7 章　结论

本文研究了基于方面的情感分析问题及其在新闻驱动的股票预测中的应用。基于方面的情感分析旨在提取一个句子对某一方面或目标的情感极性。我们提出了使用 transformer 和多尺度卷积的模型来解决对某一目标的情感分析问题，并在公开数据集 Restaurant、Laptop、Twitter 上取得了最佳的结果。同时，我们将基于方面的情感应用于新闻驱动的股票预测。通过为每个股票创建一个向量化的表示，我们将新闻驱动的股票预测转化为对某一方面的情感分析问题。我们创建了名为 senti-stock 的数据集，这一数据集由来自于路透社的金融新闻和雅虎财经的股票历史数据构成，共包括三万条数据。我们提出 MSRA（多股票关系模型），使用带有股票系数的多尺度卷积提取情感信息，利用 attention 机制学习股票之间的相关关系。MSRA 模型在 senti-stock 数据集上取得了最佳的结果。

插图索引

表格索引

公式索引

参考文献

[1] Wilson T, Hoffmann P, Somasundaran S, et al. Opinionfinder: A system for subjectivity analysis[C]//HLT/EMNLP 2005, Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 6-8 October 2005, Vancouver, British Columbia, Canada. [S.l.: s.n.], 2005.

[2] Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market[J]. Journal of Computational Science, 2011, 2(1): 1–8.

[3] Ding X, Zhang Y, Liu T, et al. Using structured events to predict stock price movement: An empirical investigation[M]. [S.l.: s.n.], 2014: 1415–1425.

[4] Ding X, Zhang Y, Liu T, et al. Deep learning for event-driven stock prediction[M]. [S.l.: s.n.], 2015: 2327–2333.

[5] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[M/OL]//Guyon I, Luxburg U V, Bengio S, et al. Advances in Neural Information Processing Systems 30. Curran Associates, Inc., 2017: 5998–6008. http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf

[6] Fama E F. Efficient market hypothesis: A review of theory and empirical work[J]. Journal of Finance, 1970, 25(2).

[7] Ding X, Zhang Y, Liu T, et al. Knowledge-driven event embedding for stock prediction.[M]. [S.l.: s.n.], 2016: 2133–2142.

[8] Spielberger C D. Profile of mood states[J]. Professional Psychology, 1972(4): 387–388.

[9] Pontiki M, Galanis D, Pavlopoulos J, et al. Semeval-2014 task 4: Aspect based sentiment analysis[M]. [S.l.: s.n.], 2014: 27–35.

[10] Wang Y, Huang M, Zhu X, et al. Attention-based lstm for aspect-level sentiment classification [M]. [S.l.: s.n.], 2016: 606–615.

[11] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735–1780.

[12] Ma D, Li S, Zhang X, et al. Interactive attention networks for aspect-level sentiment classification[J]. international joint conference on artificial intelligence, 2017: 4068–4074.

[13] Liu J, Zhang Y. Attention modeling for targeted sentiment.: volume 2[M]. [S.l.: s.n.], 2017: 572–577.

[14] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. [S.l.: s.n.], 2012: 1097–1105.

[15] Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences[C]//ACL. [S.l.: s.n.], 2014.

[16] Weston J, Chopra S, Bordes A. Memory networks[J]. Eprint Arxiv, 2014.

[17] Sukhbaatar S, Szlam A, Weston J, et al. End-to-end memory networks[J]. Computer Science, 2015.

[18] Al-Smadi M, Qawasmeh O, Al-Ayyoub M, et al. Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of arabic hotels' reviews[J]. Journal of Computational Science, 2017: S1877750317305252.

[19] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078, 2014.

[20] Xin L, Bing L, Lam W, et al. Transformation networks for target-oriented sentiment classification[M]. [S.l.: s.n.], 2018.

[21] Dong L, Wei F, Tan C, et al. Adaptive recursive neural network for target-dependent twitter sentiment classification[C]//Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers): volume 2. [S.l.: s.n.], 2014: 49–54.

[22] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[J]. URL https://s3-us-west-2. amazonaws. com/openai-assets/research-covers/languageunsupervised/language understanding paper. pdf, 2018.

[23] Zhu Y, Kiros R, Zemel R, et al. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books[M]. [S.l.: s.n.], 2015.

[24] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.

[25] Tang D, Qin B, Feng X, et al. Effective lstms for target-dependent sentiment classification[J]. Computer Science, 2015.

[26] Kiritchenko S, Zhu X, Cherry C, et al. Nrc-canada-2014: Detecting aspects and sentiment in customer reviews[M]. [S.l.: s.n.], 2014: 437–442.

[27] Xue W, Li T. Aspect based sentiment analysis with gated convolutional networks[J]. meeting of the association for computational linguistics, 2018, 1: 2514–2523.

[28] Tang D, Bing Q, Liu T. Aspect level sentiment classification with deep memory network[M]. [S.l.: s.n.], 2016.

[29] Pennington J, Socher R, Manning C. Glove: Global vectors for word representation[C]// Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). [S.l.: s.n.], 2014: 1532–1543.

[30] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

[31] Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[M]. [S.l.: s.n.], 2014: 1532–1543.

[32] Duchi J C, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of Machine Learning Research, 2011, 12: 2121–2159.

致　谢

衷心感谢导师宋斌恒副教授对本人的精心指导。他们的言传身教将使我终生受益。感谢罗宁奇同学和沈彬同学对论文的帮助。感谢 ThuThesis 提供的论文模板。

声　明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签　名：_____ 日　期：_____

附录 A　本文中使用的股票列表

表 A.1　本文中使用的股票列表

symbol	name	symbol	name
AAPL	Apple Inc	BA	Boeing Company (The)
PBR	Petroleo Brasileiro SA- Petrobras	DB	Deutsche Bank AG
GM	General Motors Company	AMZN	Amazoncom Inc
GS	Goldman Sachs Group Inc (The)	JPM	J P Morgan Chase & Co
C	Citigroup Inc	XOM	Exxon Mobil Corporation
BABA	Alibaba Group Holding Limited	WFC	Wells Fargo & Company
F	Ford Motor Company	GE	General Electric Company
HSBC	HSBC Holdings plc	GOOG	Alphabet Inc
GOOGL	Alphabet Inc	TM	Toyota Motor Corp Ltd Ord
LMT	Lockheed Martin Corporation	TSLA	Tesla Inc
BCS	Barclays PLC	FB	Facebook Inc
VRX	Valeant Pharmaceuticals International Inc	WMT	Wal-Mart Stores Inc
BP	BP plc	MS	Morgan Stanley
FCAU	Fiat Chrysler Automobiles NV	MSFT	Microsoft Corporation
TI	Telecom Italia SPA	T	AT&T Inc
PFE	Pfizer Inc	BLK	BlackRock Inc
TSM	Taiwan Semiconductor Manufacturing Company Ltd	MON	Monsanto Company
VZ	Verizon Communications Inc	RBS	Royal Bank Scotland plc (The)
SNY	Sanofi	BBL	BHP Billiton plc
AZN	Astrazeneca PLC	CVX	Chevron Corporation
BAC	Bank of America Corporation	JNJ	Johnson & Johnson
CS	Credit Suisse Group	MCD	McDonald&s Corporation
MRK	Merck & Company Inc	DIS	Walt Disney Company (The)

个人简历、在学期间发表的学术论文与研究成果

个人简历

1993 年 5 月 7 日出生于河北省衡水市深县。

2012 年 9 月考入车辆工程系，2013 年 7 月转入计算机科学与技术系，2016 年 7 月本科毕业并获得工学学士学位。

2016 年 9 月免试进入清华大学计算机科学与技术系攻读工学学位至今。

学术论文

[1] Yinxu Pan, Binheng Song, et al. Transformer and Multi-scale CNN for Target-oriented Sentiment Classification[C]//Asia Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data(APWeb-WAIM) 2019.Chengdu,China.Springer,2019, (CCF-C 类会议, 已录用)

[2] Yinxu Pan, Binheng Song, et al. MSRA: Multi-Stock Relation model using Attention mechanism for News-driven Stock Prediction[C]//28th International Conference on Artificial Neural Networks (ICANN).Munich,German:Springer,2019, (CCF-C 类会议, 已录用)

[3] Ningqi Luo, Binheng Song, Yinxu Pan, et al. Bcmlp: Binary-connected multiplayer perceptrons[C]//The 25th International Conference on Neural Information Processing (ICONIP). Siem Reap, Cambodia: Springer, 2018:77-88. (CCF-C 类会议，EI 检索，索引号 20185206311203)

实践经历

[1] 2017 年 6 月至 2017 年 9 月，于富途证券公司担任量化平台开发工程师。

[2] 2017 年 10 月至 2018 年 3 月，于商汤科技公司担任算法实习。

[3] 2018 年 6 月至 2018 年 11 月，于蚂蚁金服（杭州）网络技术有限公司担任实习算法工程师。

[4] 2018 年 12 月至 2019 年 4 月，于苹果中国担任实习算法工程师。