

结合投资者情绪运用LSTM-GAN模型 预测股票

张家宏

贵州大学经济学院, 贵州 贵阳

收稿日期: 2025年4月14日; 录用日期: 2025年4月28日; 发布日期: 2025年6月4日

摘要

本研究运用股票市场历史数据以及投资者情绪数据, 通过实证方法和对比分析方法, 探讨了长短期记忆网络模型和对抗生成网络模型以及二者结合在股票收盘价预测中的效果。通过机器学习模型对这些数据进行分析, 实验结果表明, 混合模型相比于单独的长短期记忆网络模型和对抗生成网络模型误差更低, 效果更好, 并且模型中加入投资者情绪后, 显著提高了股票预测的准确性和鲁棒性。

关键词

长短期记忆网络模型, 对抗生成网络模型, 投资者情绪

Combining Investor Sentiment, Using LSTM-GAN Model to Forecast Stocks

Jiahong Zhang

School of Economics, Guizhou University, Guiyang Guizhou

Received: Apr. 14th, 2025; accepted: Apr. 28th, 2025; published: Jun. 4th, 2025

Abstract

This study utilizes historical data of the stock market and investor sentiment data. Through empirical methods and comparative analysis methods, it explores the effects of the long short-term memory network model and the adversarial generative network model, as well as their combination, in the prediction of stock closing prices. These data were analyzed through the machine learning model. The experimental results show that the hybrid model has lower errors and better effects compared

with the individual long short-term memory network model and the adversarial generative network model. Moreover, after adding investor sentiment to the model, the accuracy and robustness of stock prediction were significantly improved.

Keywords

Long Short-Term Memory Network Model, Antagonistic Generative Network Model, Investor Sentiment

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言与文献综述

在当今快速变化的经济环境中，股票市场的波动性和不确定性越来越激烈，股票市场的变化对投资者、企业和政策制定者都具有深远的影响。准确预测股票市场的走势对于资本的有效配置、风险管理以及投资决策至关重要。然而，股票市场的预测充满挑战，因为本身受到多种复杂因素的影响，例如宏观经济状况、投资者情绪、政治变动和技术发展等。

在股票市场预测的研究历程中，方法和技术经历了从简单到复杂、从传统到现代的演变。过去对股票市场的预测方法许多基于股票市场的基本面分析和技术分析，例如，Alexander (1961) [1]指出格雷厄姆和大卫·多德在 1934 年提出，预测某公司的股票走势可以通过分析该公司的基本财务状况、或者行业地位等基本信息；张国文(1992) [2]指出，通过结合公司的财务信息和业务前景进行说明，探究短期内的股票价格波动情况，用来寻找买卖股票的时机选择。早期对于股票的预测过于简单，很多一些影响股票的因素并未考虑完全。自 20 世纪 80 年代起，随着计算机技术的进步，许多投资者开始运用机器学习技术方法预测股票，最初的应用包括神经网络、支持向量机等，张健(1997) [3]运用神经网络中 BP 模型预测股票。潘晓明等(2009) [4]通过支持向量来优化神经网络模型的预测，这些方法能够处理非线性关系，提高了预测的精度。进入 21 世纪后，深度学习技术的兴起为股票市场预测带来了新的突破，张晶华等(2019) [5]利用深度学习支持向量机方法，结合数学上全概率知识，对上证指数数据进行预测分析，并得出其预测的效果更好，提高了对上证指数预测的鲁棒性和准确性。

在当前机器学习领域，长短期记忆网络和生成对抗网络因其独特的优势而被广泛应用于各类数据处理与预测任务中。长短期记忆网络在捕获时间序列数据中的长期依赖关系方面展现出更强的建模能力，这一特性使其在股票价格预测、自然语言处理等时序建模任务中获得广泛应用，齐太威等(2024) [6]设计改进了一种基于长短期记忆网络模型预测股票的策略，改进模型预测获得了更加精准的预测数据，并且这种策略可行性更高，但股票价格受市场噪声和非平稳性影响显著，而长短期记忆网络模型的梯度传播机制可能导致过度拟合，影响预测结果不稳定。因此，随着深度学习的发展，许多学者为了克服单一模型使用的缺点，开始将各种机器学习模型结合构造出一个新的混合股票预测模型，方便结合各种模型的优势，例如崔笑宁等(2024) [7]结合了时间卷积与长短时记忆网络模型，两者结合集成两个模型的优势，权衡了特征提取时局部和长期的分配问题，优化了以往的单一的股票预测模型，但是该方法未考虑市场不同阶段对短期或者长期信号的依赖差异，可能导致信息冗余或冲突，未建立动态权重分配机制；韩莹等(2023) [8]构建了宽度学习系统(BLS)与深度学习的混合架构，创新性地采用互补集成经验模态分解对非平稳股票序列进行自适应降噪。经数据集验证，该模型在 RMSE、MAE 指标上分别有所降低，有效缓解

时序数据中的相位偏移问题。然而,研究同时揭示:当序列长度超过一定时间步时,长短期记忆网络模型仍存在梯度弥散率,导致长期记忆能力下降;而 BLS 受限于浅层网络结构,其特征表征损失未能有效补偿深度模型的退化现象;任君等(2019) [9]提出将长短期记忆网络与 Lasso 相结合,新的混合模型相对于其他组合的模型,该预测模型在股票市场应用中展现出显著的超额收益潜力与风险抵御效能,其特有的成本吸收机制使模型在较高交易费用环境下仍保持策略有效性,但是新模型依赖历史统计规律,对极端事件缺乏生成式建模能力。

生成对抗网络作为前沿深度学习框架,借力生成器与判别器的双网络博弈机制,其合成数据的信噪比优于传统方法,为突破金融时序数据稀疏性瓶颈提供了创新路径,马甜等(2022) [10]基于生成对抗网络构建的深度学习架构,相较于传统线性模型及常规深度学习模型,在因子投资收益预测的准确性及策略稳健性维度实现了系统性优化;刘玉玲等(2022) [11]提出的生成对抗网络耦合情感分析框架,通过动态反馈机制实现股价预测的在线校准,其误差收敛性较传统时序模型提升显著;王静(2020) [12]提出了一种融合多元经验模态分解的对抗生成网络架构,生成器通过 MEMD 与注意力机制的双重优化实现特征提取,同时采用卷积神经网络作为判别器进行时序模式识别,最终使沪深 300 指数预测的均方误差较传统模型有了显著降低;严冬梅和李斌(2022) [13]提出的 SAR-GAN 生成对抗框架,通过自注意力机制与残差网络的深度耦合,构建了对抗模型的双模块协同架构(自注意力层提取全局时序依赖,残差层捕获局部波动模式),该模型在风险中性测度下有效提升股价预测效果,为量化投资策略的鲁棒性优化提供可微分的决策拓扑空间。尽管生成对抗网络模型在股票预测中展现出了对抗去噪的能力,并能适应市场周期性规律,有效应对相关预测难题,但目前仍有一些学者未能充分认识到投资者情绪对股价的影响,即便部分学者已经意识到这种影响,现有预测架构普遍存在对过拟合风险的评估盲区,其根源在于未能充分考量投资者情绪因子与市场波动间的非对称传导机制,该因子通过认知偏差放大与信息扩散滞后等路径,与基本面指标形成高维度非线性关联,导致传统正则化方法在防范模型退化时产生显著性错误,投资者情绪反映了投资者在接收市场信息时的心理状态和情绪反应,包括乐观、悲观、恐慌等,这些情绪不仅映射了他们对市场未来走势的预期,还会实际影响他们的投资决策,进而对股票价格产生显著影响[14]。

鉴于此,本研究旨在构建一个更为强大和精确的预测模型,创新性地将生成对抗网络与长短期记忆网络相结合,用于股票价格预测,在深入探索 GAN 和 LSTM 的深度学习应用过程中,进一步引入了投资者情绪这一关键因素,期望能够提供一个更为全面和深入的市场分析工具,从而提高预测的准确性和可靠性,通过这种结合,不仅能够利用 GAN 生成高质量的合成数据,还能借助 LSTM 捕捉时间序列中的长期依赖关系,并同时考虑投资者情绪对股价的影响,为投资者和金融机构提供更科学的决策支持。

2. 理论分析与实证假设

本研究方法体系的核心建模框架由以下三部分构成:(1)长短期记忆网络,长短期记忆网络作为循环神经网络的优化架构,通过门控机制调控信息流,有效抑制长序列建模中的梯度不稳定现象;(2)生成对抗网络依托生成器与判别器的对抗性训练范式,在特征空间构建过程中实现数据分布的动态博弈均衡;(3)基于文本挖掘的投资者情绪量化分析,捕获市场参与者的非理性预期偏差。通过设计情绪因子与两类深度学习模型的嵌套耦合机制,将情绪向量分别嵌入 LSTM 和 GAN 的组合预测模型之中,形成融合市场微观结构与行为金融特征的多尺度预测系统,此举在理论上可增强模型对市场风格转换的适应能力。

2.1. 生成式对抗网络

生成对抗网络(Generative Adversarial Networks, 简称 GAN)采用生成-判别双模块的对抗训练范式,通过纳什均衡驱动的参数优化机制,使生成器在策略博弈中逐步逼近真实数据分布。

2.1.1. 生成器和判别器的简介

生成对抗网络的结构如图 1 所示，其核心由生成器与判别器组成双模块架构。在训练过程中，这两个模块形成动态博弈，生成器基于真实数据分布特征合成仿真样本，旨在提升生成内容的逼真度以误导判别器的判断；而判别器则通过持续优化自身的鉴别能力，力图准确区分输入数据源自主样本库还是生成器产物。这种对抗性训练机制促使双方性能同步提升，经过多轮迭代后，系统将收敛至纳什均衡状态，此时生成器产生的样本已具备与真实数据相当的统计特性，使得判别器难以有效辨识其来源。

2.1.2. 训练过程

如图 2，生成对抗网络的训练过程中，黑色曲线表征原始数据概率密度，绿色曲线对应生成样本分布形态，蓝色折线反映判别器输出置信度。在训练周期初始化阶段，系统同步更新生成网络与判别网络的参数，生成器通过参数优化策略最大限度地缩小生成样本与真实样本间的区别，使判别器对合成样本的判别置信度趋近于对真实数据的判定阈值；而判别器则采用梯度反向传播算法强化其决策边界，增强对两类样本的区分能力。通过迭代训练驱动模型向纳什均衡逼近，直至生成分布与真实分布在统计维度达到不可区分状态。具体来说，生成器从噪声数据中随机采样生成一组数据，同时从真实数据中随机抽取一组数据，将这两种数据随机输入判别器，判别器会根据输入的数据判断输出数据为真实的概率，根据这些计算出来的概率，运用反向传播计算每个部分具体需要修正的参数，依据修改的误差，交替更新其相关模型参数，通过训练过程中的优化，生成器能够生产出和真实数据特别相似的数据，判别器不能依据最大化的模型效果的参数识别其真假。

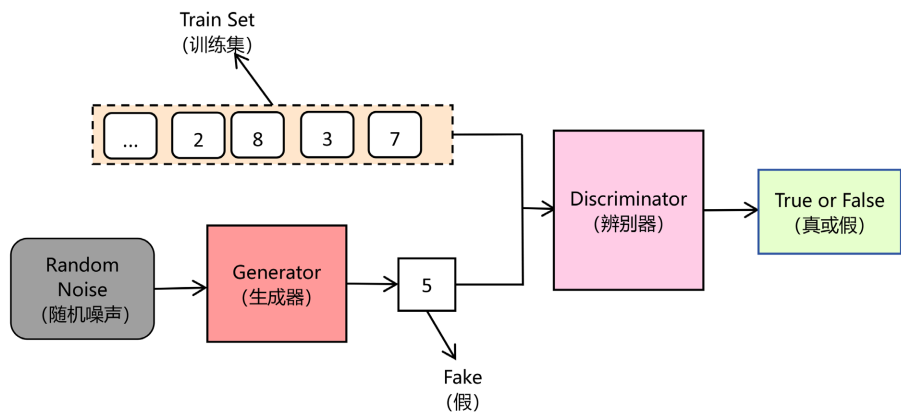


Figure 1. Antagonistically generating network model diagrams
图 1. 对抗生成网络模型图

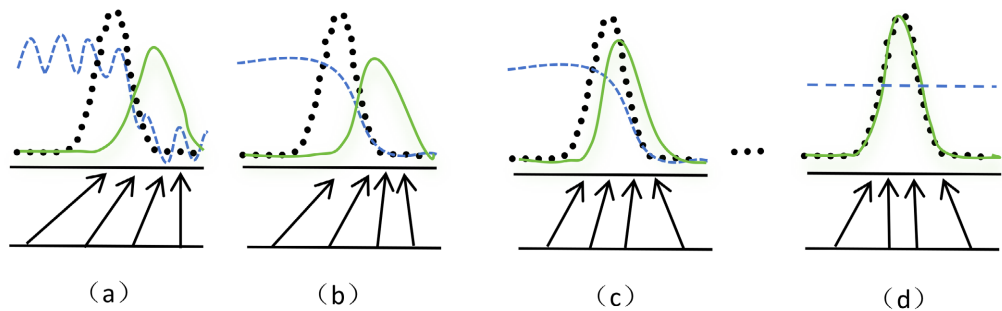


Figure 2. GAN model training process
图 2. GAN 模型训练过程

2.2. 长短期记忆网络

长短期记忆网络(Long Short-Term Memory, 简称 LSTM)模型是一种用于训练具有长久关联的序列数据模型, 也属于递归神经网络模型一类, 根据自身模型的性能特点能有效地解决以往循环神经网络中出现的梯度消失以及梯度爆炸的问题, 能够学习到时间序列数据中较长久的信息。LSTM 有输入门、输出门和遗忘门三种独特的门控机制, 可以有效将数据流入和输出之间的特征提取, 有效地存储长久的数据特征。

长短期记忆网络通过其门控系统的微分架构实现对金融时序信号的建模与分析。该模型的动态特征提取能力源于以下三个核心门控单元的协同作用:

$$I_t = \sigma(W_{xi}X_t + W_{hi}H_{t-1} + b_i) \quad (1)$$

$$F_t = \sigma(W_{xf}X_t + W_{hf}H_{t-1} + b_f) \quad (2)$$

$$O_t = \sigma(W_{xo}X_t + W_{ho}H_{t-1} + b_o) \quad (3)$$

$$\tilde{C}_t = \tanh(W_{xc}X_t + W_{hc}H_{t-1} + b_c) \quad (4)$$

$$C_t = F_t \cdot C_{t-1} + I_t \cdot \tilde{C}_t \quad (5)$$

$$H_t = O_t \tanh(C_t) \quad (6)$$

其中, I_t 为该模型的输入门, 用于确定当前的输入, F_t 为遗忘门用于确定需要保留多少消息, O_t 为输出门用于确定数据的输出, 为候选记忆单元用于生成新的候选值向量, 这些候选值也将被写入到记忆单元中, C_t 为记忆单元, 用于存储和传递长期信息, H_t 为该模型的隐藏状态。如图 3 所示, 在 LSTM 的运行结构中, 最上面的一条线是记忆单元 C_t , 贯穿着整个网络的运行过程中, 而黄色矩阵、粉色圈圈和黑色箭头分别表示神经网络层、运算符和向量传输。在 LSTM 的运作过程中, 不仅隐藏状态 h_t 会随时间变化, 记忆单元 C_t 也会不断更新, 这样变化也有助于适应新的输入信息和时间序列的变化。

LSTM 模型系统选取上一期的向量 $h_{(t-1)}$ 和当前数据 X_t , 通过遗忘门得到一个 0 到 1 之间的值 f_t , 旨在根据当前的输入决定保留多少之前的记忆存储 $C_{(t-1)}$, 并通过单元状态得到当前时刻的临时状态 C_t ; 结合 sigmoid 层进行数据筛选, 并将记忆单元送入 tanh 层; 最后将筛选出来的数据与 tanh 层输出的数据相融合, 从而得到 LSTM 模型的输出结果。

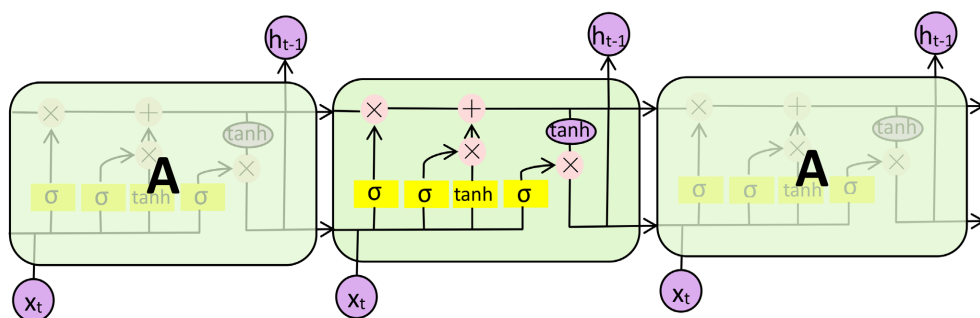


Figure 3. Long short-term memory network model diagram

图 3. 长短期记忆网络模型图

2.3. 投资者情绪

投资者情绪测度体系在行为金融学研究中存在经典分类框架, 其中第一类为直接观测法, 该方法主

要依托结构化问卷工具与深度访谈技术，通过构建多维李克特量表采集市场参与者的风险偏好、投资预期与趋势判断。例如 Lee 等(2002) [15]通过咨询投资者推荐购买的股票或对市场上的预测信息衡量投资者情绪，Brown 和 Cliff (2004) [16]通过调查问卷的方式收集投资者对市场上的预期用来衡量投资者情绪，然而，该方法容易让投资者受多种因素的影响，投资者会选择非内心的答案进行回答；二是指标替代法，该方法是寻找市场上常用的具有代表性的指标组合衡量投资者情绪，如崔文星等(2025) [17]依据六个具有代表性的指标来衡量投资者情绪，该方法虽然在一定程度上克服了投资者主观性的问题，但容易出现时滞性问题，即相关指标可能无法及时准确地反映当前投资者情绪，导致衡量结果存在偏差；三是文本分析法，通过财经新闻网站、社交媒体等平台利用 python 抓取相关帖子进行投资者情绪指标构建，如张卫国和丘启君(2024) [18]通过 python 抓取投资者对股票或者市场看法的帖子，将抓取到的帖子采用词典法对文本内容进行情感分析，运用投资者对股票市场的看涨和看跌的数量计算出投资者情绪指标。

本文基于客观性和时效性原则，选择第三类方法来研究投资者情绪，以期更准确地反映投资者情绪对股票价格的影响，借鉴 Antweiler 和 Frank [19]对投资者情绪的指标构建，构建其投资者情绪指标如下。

$$B_t = \frac{M_t^{pos} - M_t^{neg}}{M_t^{pos} + M_t^{neg}} \quad (7)$$

其中 B_t 为投资者情绪指标， M 为股吧帖子， pos 代表看涨的帖子， neg 代表看跌的帖子， t 是时间指标， B_t 介于-1 到 1 之间，当指标偏向 1 时，则表示投资者心里面向好发展，反之，当指标偏向-1 时，投资者心里偏向不好。

2.4. LSTM-GAN 模型

通过将长短期记忆网络模型和生成对抗网络结构起来创造一个新的预测模型，既能发挥长短期记忆网络的优势，也能利用 GAN 网络的优势，通过结合两种模型的优点，更好的预测股票的发展。除此之外，本文的一个创新点是运用投资者情绪，将其作为一个预测股票的特征输入，投资者情绪能反映市场上经济的某种状态，加入的元素对于股票的预测更加精准。

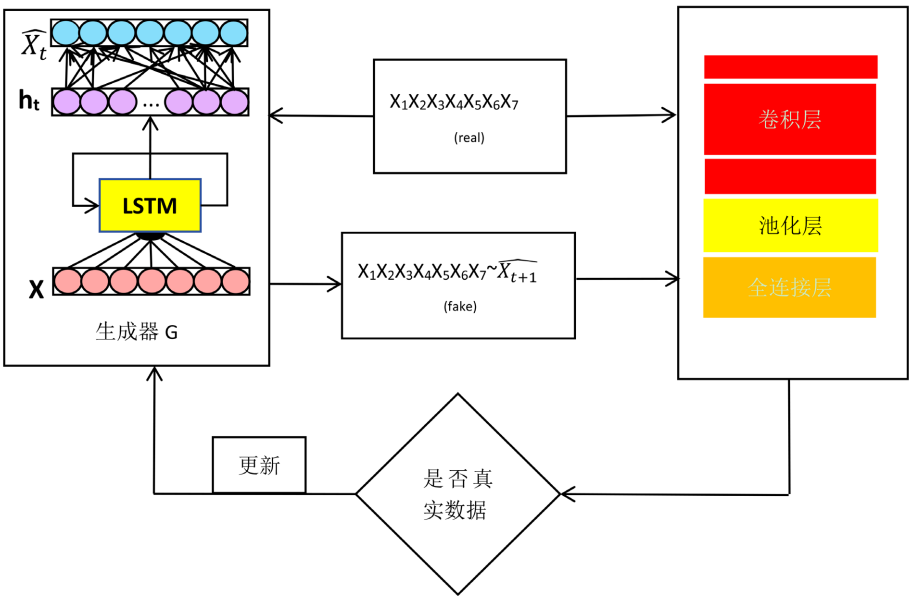


Figure 4. LSTM-GAN model
图 4. LSTM-GAN 模型

如图 4, LSTM-GAN 模型依然包含两个部分, 生成器和鉴别器, 生成器运用 LSTM 设计处理数据, 本文选择最高价、最低价、开盘价、收盘价、成交量、换手率和投资者情绪这 7 个指标预测股票的收盘价 $\widehat{X}_{t+1}$ 。在这一过程中, 首先运用 LSTM 对这七个指标进行数据提取 h_t , 将提取到的数据特征运用一个全连接层预测其下一期的股票收盘价 $\widehat{X}_{t+1}$ 。

在鉴别器中, 本研究采用卷积神经网络模型进行设计, 通过损失函数计算损失值, 运用梯度下降算法调节整体权重。鉴别器的网络架构选择卷积层、池化层和全连接层三个核心组件, 在特征提取阶段, 卷积层通过滑动窗口机制捕捉股价时序数据中的局部波动特征; 池化层随后对特征图进行下采样处理, 在保留关键信息的同时有效抑制过拟合风险, 并通过特征聚合实现数据维度压缩; 最终借助全连接层对经多层抽象的特征表示进行非线性组合和高层特征映射, 经分类器处理后输出表征股价趋势判断的概率分布值。

$V(D,G)$ 表示新模型的训练过程, $G(x)$ 是生成器, $D(x)$ 是鉴别器, 目标函数如下:

$$\frac{\min}{D} \frac{\max}{G} vD, G = E_{x \sim P_{data}(x)} [\log D(x_i)] + E_{z \sim P_z(x)} [\log 1 - D(G(z))] \quad (8)$$

在公式(8)中, $D(x)$ 作为判别器, 用于鉴别输入的数据是真实数据的概率; 而 $1 - D(G(z))$ 则表示判别器识别生成器生成的数据 $G(z)$ 判定为假样本的概率。生成器旨在尽可能去欺骗判别器的判断, 而判别器则努力提高自身的判断正确效率。通过双方之间的交替更新, 生成器和判别器的权重也会在对抗过程中不断优化, 直至达到动态平衡。

3. 实证分析

3.1. 数据说明

本文选取上证 50 指数作为研究目标, 相关股票数据来自 Tushare 数据库, 时间范围为 2020 年 10 月 1 日至 2024 年 9 月 30 日。数据指标涉及收盘价、开盘价、最高价、最低价和成交量等。此外, 本文通过财经新闻网站和社交媒体平台, 利用 Python 抓取相关帖子, 并采用词典法分析出看涨和看跌帖子的数量。

3.2. 数据预处理

对于数据的不同维度的状态, 学者采用了归一化处理, 即将所有数据统一到一个标准范围内。其计算公式为:

$$Z_j = \frac{X_j - \mu}{\sigma} \quad (9)$$

式子(9)中, X_j 表示数据样本, μ 为数据样本的平均值, σ 为数据样本的标准差, 用于表示数据样本的离散程度。

3.3. 数据集划分

本研究按 1:4 比例划分数据集, 训练集用于优化模型参数至最优状态, 测试集则验证模型泛化性能。该划分方式在保证充分训练的同时, 通过数据隔离有效防止过拟合。如图 5 所示。

3.4. 参数设置

GAN-LSTM 模型中, LSTM 充当生成器, CNN 充当判别器。LSTM 作为生成器, 其网络结构由 3 层组成, 每层包含 128 个隐藏单元, 旨在有效捕捉时间序列数据中的长期依赖性。该生成器的学习率设定为 0.001, 训练周期(epochs)为 60, 批量大小(batch_size)为 64, 时间窗口设置为 5, 以适应短期预测的需求。

求。判别器选择卷积神经网络模型，由三层卷积层组成，分别配置 32、64、128 个神经元节点，学习率同样设定为 0.001，以稳定学习过程。最大池化策略被用于降低过拟合风险，而输出层采用 Sigmoid 激活函数，实现单步预测的二分类任务。

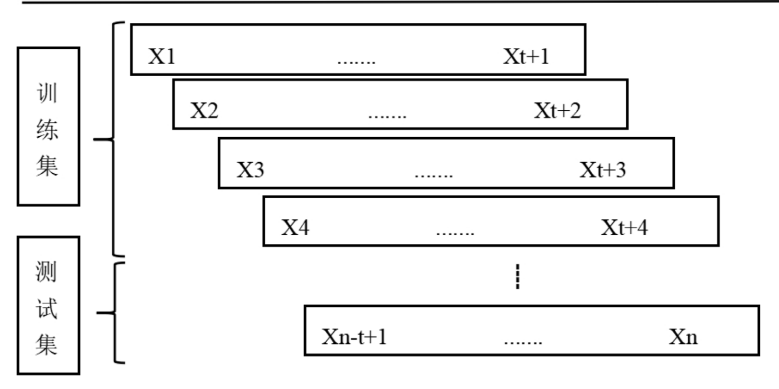


Figure 5. Data partitioning
图 5. 数据划分

3.5. 评价指标

3.5.1. 均方误差(MSE)

均方误差(MSE)是旨在衡量数据的真实值和预测值之间的偏差状况，计算公式为

$$MSE = \frac{1}{N} \sum_i^n (y_i - \hat{y})^2 \quad (10)$$

通过平方误差的平均值来量化模型的预测偏差，数值越大表示模型效果越差，预测的精确度越小，反之，效果越好，精确度越大。MSE 对误差的大小非常敏感，因此在数据中存在极端值时，其结果可能会受到较大影响。

3.5.2. 平均绝对误差(MAE)

平均绝对误差(MAE)通过计算模型预测值与实际观测值之间绝对偏差的算术均值，量化预测结果的准确性。其计算公式为

$$MAE = \frac{1}{N} \sum_i^n |y_i - \hat{y}| \quad (11)$$

MAE 对所有误差的权重相同，不受极端值影响较大，能够提供一个稳健的误差度量。平均绝对误差的数值越小，表示模型的预测结果与实际产生的数值的偏差越小，模型效果越好，反正的较差，适用于对误差大小较为敏感的应用场景。

3.5.3. 均方根误差(RMSE)

均方根误差衡量数据的预测值和实际值之间偏差状况的平方根，也是平方误差的开方，该指标与均方误差相似，对较大的误差给予更大的权重，但其量纲与原始数据相同，因此更易于解释。RMSE 的计算公式为：

$$RMSE = \sqrt{\frac{1}{N} \sum_i^n (y_i - \hat{y})^2} \quad (12)$$

RMSE 与原始数据的单位相同，因此指标结果更直观、更易于理解，数值越小，说明模型的预测性能越好。

3.6. 模型性能与比较分析

本研究分别采用单独的 LSTM 模型和 GAN 模型进行分析，并将这两个模型与投资者情绪结合作为 LSTM-GAN 模型一部分进行预测和对比分析。结果表 1 显示，单独的 LSTM 模型在预测股票价格时，其均方误差、平均绝对误差和均方根误差的值为 0.05646、0.22149、0.23761，误差非常小，但是并不是预测效果最好。尽管 LSTM 模型能够利用其长短期记忆能力有效捕捉价格的时间序列特征，但其主要依赖于历史价格数据，并未能充分考虑市场情绪等外部因素的影响。在市场情绪波动较大或出现突发新闻事件时，这可能导致预测结果的准确性下降。而单独的 GAN 模型在生成的股票价格数据方面表现出色，三者的值为 0.16249、0.33877 和 0.40310，GAN 模型在直接预测股票价格时，由于缺乏对时间序列数据的长期依赖性分析，预测出来的结果不够稳定，容易受到生成数据质量的影响，因此其误差指标也并非最优。相比之下，基于情绪指标的 LSTM-GAN 模型在 MSE、MAE 和 RMSE 上均有显著改善。该模型结合了 LSTM 的时间序列分析能力和 GAN 的数据生成能力，并引入情绪指标作为额外的输入信息。通过综合考虑历史价格走势和市场情绪等因素，该模型能够更全面地捕捉市场动态，从而在预测误差上相对更小，显示出更高的预测准确性。

Table 1. Error of model prediction
表 1. 模型预测的误差

模型	MSE	MAE	RMSE
LSTM	0.05646	0.22149	0.23761
GAN	0.16249	0.33877	0.40310
LSTM-GAN	0.05590	0.20256	0.23434

3.7. 不同特征预测股票结果

在股票价格预测任务中，特征工程策略的差异性会导致显著的预测偏差差异。从表 2 中可以看出，当建模过程中仅纳入收盘价单变量特征时，该新型预测模型展现出的误差均较小，均方误差达 0.05662，平均绝对误控制在 0.20358，均方根误差为 0.23794，特别值得注意的是，这三项核心评估指标在保持单位一致性的同时，均处于较低数值区间，充分验证了模型在单一特征条件下的预测稳定性。若综合考虑历史价格数据，这些误差指标分别是 0.20816、0.39419 和 0.45625，同时当将投资者情绪作为特征加入模型时，误差指标进一步进化至 0.05590、0.20256 和 0.23434。这表明，投资者情绪的加入显著改善了模型的预测效果，使得预测误差大幅减小。因此，可以得出结论，投资者情绪是一个重要的预测因素，其对股票价格预测具有显著的正向影响。

Table 2. Prediction errors of different characteristics
表 2. 不同特征的预测误差

特征	MSE	MAE	RMSE
收盘价	0.05662	0.20358	0.23794
历史数据	0.20816	0.39439	0.45625
历史数据 + 投资者情绪	0.05590	0.20256	0.23434

4. 主要结论与建议

本研究通过融合生成对抗网络模型和长短期记忆网络模型进行预测。其结果表明，二者结合对于股票的预测的准确性相比单一的模型较好，同时，加上投资者情绪对股票价格进行了预测，组合模型显著提升了模型的预测性能，降低了预测误差。投资者情绪在股票价格形成过程中扮演着重要角色，对于股票价格预测具有重要的指导作用。

股票市场就像一个超级复杂的迷宫，里面有很多看不见的规则 and 变化。预测股票价格就像是在猜谜语，需要考虑很多东西，比如整个国家的经济情况、公司的具体情况、政府的政策变化，还有大家的情绪等。虽然这次研究的模型让预测更准了一些，但要找到一个能完美预测股票价格的模型，还有很长的路要走。因为股市一直在变，而且变化很难捉摸，所以用一个模型来完全搞懂它是很难的。

这项研究为未来股票价格预测提供了新思路，通过结合多种模型和因素来提升预测效果。未来的研究可以在现有基础上，探索更多影响因素，如宏观经济、公司财务和行业发展等，以构建更全面和准确的预测模型。此外，还可以引入先进的技术，如注意力机制和图神经网络，帮助更好地捕捉市场的复杂动态。

参考文献

- [1] Alexander, S.S. (1961) Price Movements in Speculative Markets: Trends or Random Walks. *Industrial Management Review of Mit*, 2, 7-26.
- [2] 张国文. 股票价格技术分析的若干指标简析[J]. 上海金融, 1992(7): 14-15.
- [3] 张健, 陈勇, 夏罡, 等. 人工神经网络之股票预测[J]. 计算机工程, 1997(2): 52-55.
- [4] 潘晓明, 吴建生. 基于支持向量机的进化神经网络集成股市模型[J]. 广西工学院学报, 2009, 20(2): 58-62+72.
- [5] 张晶华, 甘宇健. 基于深度学习支持向量机的上证指数预测[J]. 统计与决策, 2019, 35(2): 176-178.
- [6] 齐太威, 于文年. 基于 LSTM 的多指标股票预测[J]. 计算机与数字工程, 2024, 52(2): 337-342.
- [7] 崔笑宁, 苏丹华, 尚维. 基于互联网新闻和时间卷积长短时记忆神经网络的股票指数预测研究[J]. 管理评论, 2024, 36(7): 113-127.
- [8] 韩莹, 张栋, 孙凯强, 等. 结合长短时记忆网络和宽度学习的股票预测新模型研究[J]. 运筹与管理, 2023, 32(8): 187-192.
- [9] 任君, 王建华, 王传美, 等. 基于 ELSTM-L 模型的股票预测系统[J]. 统计与决策, 2019, 35(21): 160-164.
- [10] 马甜, 姜富伟, 唐国豪. 深度学习与中国股票市场因子投资——基于生成式对抗网络方法[J]. 经济学(季刊), 2022, 22(3): 819-842.
- [11] 刘玉玲, 赵国龙, 邹自然, 等. 基于情感分析和 GAN 的股票价格预测方法[J]. 湖南大学学报(自然科学版), 2022, 49(10): 111-118.
- [12] 王静. 基于多元经验模态分解生成对抗网络的金融时间序列预测[D]: [硕士学位论文]. 大连: 大连海事大学, 2020.
- [13] 严冬梅, 李斌. 基于生成式对抗神经网络的股票预测研究[J]. 计算机工程与应用, 2022, 58(13): 185-194.
- [14] 彭浩杰, 张宗强. 媒体情绪与股价崩盘风险——基于机器学习和文本分析的证据[J/OL]. 兰州财经大学学报, 1-19. <http://kns.cnki.net/kcms/detail/62.1213.F.20240319.1731.038.html>, 2025-01-07.
- [15] Lee, W.Y., Jiang, C.X. and Indro, D.C. (2002) Stock Market Volatility, Excess Returns, and the Role of Investor Sentiment. *Journal of Banking & Finance*, 26, 2277-2299. [https://doi.org/10.1016/s0378-4266\(01\)00202-3](https://doi.org/10.1016/s0378-4266(01)00202-3)
- [16] Brown, G.W. and Cliff, M.T. (2004) Investor Sentiment and the Near-Term Stock Market. *Journal of Empirical Finance*, 11, 1-27. <https://doi.org/10.1016/j.jempfin.2002.12.001>
- [17] 崔文星, 曾繁华, 刘晓君. 经济政策不确定性、投资者情绪与股票收益的关联效应研究[J]. 统计与决策, 2025, 41(4): 150-155.
- [18] 张卫国, 丘启君. 投资者情绪对基金业绩的影响研究——基于基金资金净流入视角[J]. 投资研究, 2024, 43(2): 54-74.
- [19] Antweiler, W. and Frank, M.Z. (2004) Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, 59, 1259-1294. <https://doi.org/10.1111/j.1540-6261.2004.00662.x>