

基于神经网络的股票预测模型

乔若羽

(中国科学技术大学 统计与金融系, 安徽 合肥 230026)

摘要: 针对股票市场的特征提取困难、预测精度较低等问题, 本文基于深度学习算法, 构建了一系列用于股票市场预测的神经网络模型, 包括基于多层感知机(MLP)、卷积神经网络(CNN)、递归神经网络(RNN)、长短期记忆网络(LSTM)和门控神经元(GRU)的模型。针对RNN、LSTM和GRU无法充分利用所参考的时间维度的信息, 引入注意力机制(Attention Mechanism)给各时间维度的信息赋予不同权重, 区分不同信息对预测的重要程度, 从而提升递归网络模型的性能。上述模型均基于股票数据进行了优化, 基于上证指数对各类模型进行了充分的对比实验, 探索了模型中重要变量对性能的影响, 旨在为基于神经网络的股票预测模型给出具体的优化方向。

关键词: 股票预测; 深度学习; 神经网络; 注意力机制

中图分类号: F830.91 **文章标识码:** A **文章编号:** 1007-3221(2019)10-0132-09 **doi:** 10.12005/orms.2019.0233

Stock Prediction Model Based on Neural Network

QIAO Ruoyu

(Department of Statistics and Finance, University of Science and Technology of China, Hefei 230026, China)

Abstract: To address the issues corresponding to the difficult feature extraction and low prediction accuracy in the stock market, this paper proposes a series of neural network-based models for stock market forecasting based on deep learning, which refers to multi-layer perceptron, convolutional neural network, recurrent neural network, long-and-short-term memory network and gated recurrent unit, respectively. In consideration of the case that RNN, LSTM and GRU models can not make full use of the referred information in time dimension, this paper introduces the attention mechanism to give different weights to the information of each time dimension which can distinguish the importance of different information to the prediction, and then improve the performance of the recursive network model. All the models are optimized with respect to the stock data. Extensive experiments on the Shanghai Stock Index demonstrate the effectiveness of the proposed models. In addition, the comparisons between different models further explore the effect of the important parameters in the models. The purpose of this paper is to give the optimization direction of the stock forecasting models based on neural network.

Key words: stock prediction; deep learning; neural network; attention mechanism

0 引言

股票市场是宏观经济形势的晴雨表和风向标, 股票市场的走向可在一定程度上反映当前的经济形势, 对股票市场的准确预测有助于国家及时调整经济政策, 维护市场稳定, 因此对股票市场的准确预测意义重大。

但是股票预测的准确性受到多种因素的限制。由于股票数据是随机游走的金融时间序列, Fama

论证了金融时间序列的预测难点在于其存在的高噪音^[1]。若用统计学模型预测金融时间序列, 必须对数据进行预处理, 再输入到模型中, 而这样会破坏数据的完整性和真实性。其次, 股票数据是非线性且非平稳的, Singh 和 Srivastava 认为股票数据的非线性和非平稳性导致传统的多元回归以及线性回归模型应用受限, 只有更高级的模型才可以准确描述此类非线性的金融时间序列^[2]。最后股票市场受到多种复杂且不确定的因素影响, 如长期受到宏观政策的影响, 短期受到市场预期影响, 也

收稿日期: 2018-08-10

作者简介: 乔若羽(1995-), 男, 安徽合肥人, 硕士研究生, 研究方向: 金融工程。

会受到一些突发事件或国际事件的干扰,如 2018 年中美贸易战对中国股市的影响巨大。

国内外的学者尝试过多种方法预测股票市场, Chulho Jung 和 Roy Boyd 在 1996 年采用向量自回归(VAR) 模型、误差修正模型(ECM) 和卡尔曼滤波模型(KFM) 对英国股市进行预测^[3]。

Wolfgang Bessler 和 Peter Lückoff 利用贝叶斯向量自回归模(Bvar) 模型预测德国一些大型公司的投资组合回报,但是预测效果不佳^[4]。

Adebiyi 等人基于纽约证券交易所和尼日利亚证券交易所的股票数据尝试运用 ARIMA 模型预测股票价格,结果显示该模型具有短期预测潜力^[5]。但由于股票数据是非线性非平稳的金融时间序列,并且影响股票价格的因素繁多复杂,传统的统计方法和计量模型需要对输入数据进行预处理且数据量不能过多,导致预测效果并不理想。

近年来,深度学习和神经网络崭露头角,在目标检测^[6]、图像分类^[7]、图像分割^[8]等多个领域取得了瞩目的成就。神经网络预测解决了传统预测方法难以解决的问题,为股票预测带来了曙光^[9]。神经网络是对人脑抽象和具象能力的人工模拟,具有强大的拟合与表达能力^[10]。首先神经网络具有非线性映射能力,数学理论可证明神经网络可以无限逼近所有的非线性连续函数;其次神经网络具有自学习能力,神经网络在训练时会自动调节,找出输入与输出之间的内在关系;最后神经网络具有一定的容错能力,即使部分神经元受损也不会对结果产生较大的影响。神经网络的这些能力可有效地解决股票数据的不确定、非线性、非平稳等问题,多位学者的研究结果均显示神经网络模型预测股票价格的能力要优于传统计量模型,从而验证了神经网络在股票市场预测中的可行性。

Zhang 等人研究发现神经网络模型预测非线性时间序列数据的准确度远远高于 ARIMA 模型^[11]。Bayogly 和 Bahadir 以不同的标准对贝叶斯估计和神经网络模型的预测效果进行比较,结果表明神经网络模型的预测效果更好^[12]。刘海玥和白艳萍建立了 AR 模型、RBF 和 GRNN 神经网络模型,滚动预测上证指数的开盘价、收盘价、最高价和最低价并与实际价格对比和分析误差,论证了三种模型的有效性,但是 AR 模型相对预测不稳定, RBF 和 GRNN 网络训练速度很快且 GRNN 表现出了更优秀的效果^[13]。

本文基于深度学习算法,构建了一系列用于股票市场预测的神经网络模型,包括基于多层感知机

(Multi-Layer perceptrons ,MLP)^[14]、卷积神经网络(Convolution Neural Network ,CNN)、递归神经网络(Recurrent Neural Network ,RNN)^[15]、长短期记忆网络(Long Short-Term Memory ,LSTM)^[16]和门控神经单元(Gated Recurrent Unit ,GRU)^[17]的模型。针对 RNN、LSTM 和 GRU 等网络模型难以充分利用时间维度信息的问题,本文引入注意力机制(Attention Mechanism)^[18]来区分不同时间维度信息对预测的重要程度,进而提高了模型的预测精度。本文基于上证指数数据进行了充分的实验,论证了所构建模型的有效性,并通过调控模型中的重要参数进行了充分的对比实验,探究不同的架构对性能的影响。本文旨在为基于神经网络的股票预测模型给出合理的优化方向。

1 基于神经网络的多种模型构建

针对金融数据的格式和需要预测的指标,本文使用神经网络中的不同模块构建了多个模型进行比较分析。模块包括全连接层(Fully-Connected Layer)、CNN、RNN 以及 RNN 的两种变体网络: LSTM 和 GRU。此外,本文考虑将注意力机制引入 RNN、LSTM 和 GRU 中,进一步提升网络模型的学习能力。由于需要处理的股票数据本身维度较低,本文的模型中未使用 Batch Normalization(BN)^[19]、LayerNormalization(LN)^[20]等标准化手段来防止梯度消失、梯度爆炸问题和加快训练速度。

1.1 基于多层感知机的网络模型

多层感知机(Multi-Layer perceptrons ,MLP) 由多层的全连接层构成。层与层之间使用的非线性激活函数通常为 Sigmoid、Tanh、线性整流函数(Rectified Linear Unit ,ReLU) 及 ReLU 的一系列变种如(LeakyReLU ,PReLU ,CReLU 等等)。全连接层的运算本质是矩阵运算,第 i 层全连接层的运算模式可表示为:

$$x_{i+1} = W_i x_i + b_i \quad (1)$$

其中, W_i 和 b_i 分别表示第 i 层全连接层的权值和偏置项。

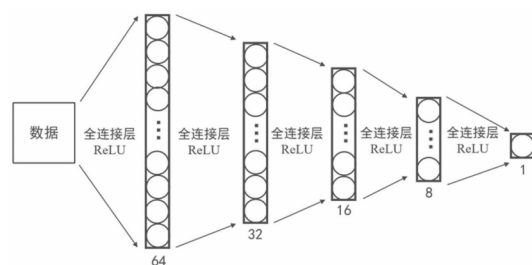


图 1 基于多层感知机的网络模型

本节构建了含有五层全连接层的网络模型,如图1所示,每层输出元素逐渐递减,分别为64、32、16、8、1。全连接层由于参数量过大,且输出与输入的所有元素均有关系,很容易导致过拟合的发生。因此模型中每层之间加入了 ReLU 函数作为非线性激活单元来防止过拟合,增加非线性表达能力,具体表达形式如下:

$$\text{ReLU}(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (2)$$

Sigmoid 函数和 Tanh 函数作为非线性激活函数在全连接层和卷积层中表现较差,收敛速度慢且易导致梯度消失现象。这两种函数被广泛应用于递归神经网络中,具体表现形式为:

$$\text{Sigmoid}(x) = \frac{1}{1 + \exp(-x)} \quad (3)$$

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (4)$$

1.2 基于 CNN 的网络模型

卷积神经网络 CNN 多用于图像的各种处理任务,如目标检测、图像分类、图像分割等等,并展示了远远超出传统方法的精度。由于图片数据多具有局部相关性,某个像素点的值通常和它临近像素点均有关联,CNN 能够很好地挖掘局部特征相关性,相比于全连接层对全局特征的关注,能够更好地提取出纹理信息和边缘信息。

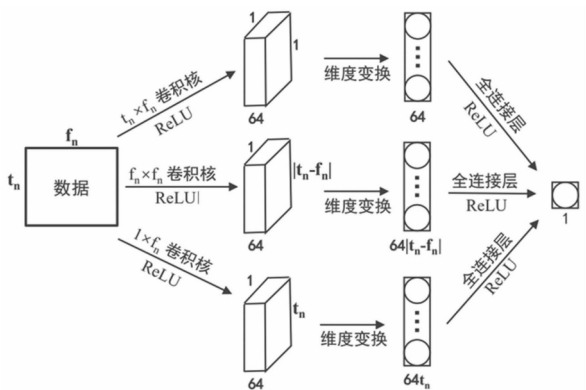


图2 基于 CNN 的网络模型

本节针对需要处理的时间序列金融数据构建了基于 CNN 的网络模型,如图2所示。考虑到金融数据除了具有时间维度之外,在每个具体时间下也有一系列描述特征的数据,称为特征维度。该模型中,本文使用 t_n 表示预测时需要参考的时间维度数, f_n 表示每一时间维度下的特征维度数。

金融数据首先被并行地输入三个不同的卷积层中,卷积核大小分别为 $1 \times f_n$ 、 $f_n \times f_n$ 和 $t_n \times f_n$,对应的特征图(feature maps)数量均为64。三个卷积层的输出维度分别为 $64 \times 1 \times 1$ 、 $64 \times 1 \times |t_n - f_n|$ 和

$64 \times 1 \times t_n$ 。卷积核尺寸这样设置的原因在于,金融数据同一时间维度下的所有特征均用来描述该数据的特性,且与图片数据的区别在于,这些特征具有较少的局部相关性。如果使用多用于图像任务的方形尺寸卷积核如 3×3 、 5×5 等,只能提取较弱的局部信息,而且会部分丢失同一时间维度下的所有特征之间互相的关联性。考虑到这一点,本文构建了三种不同的卷积核,分别用来提取不同尺度下的特征。 $1 \times f_n$ 、 $f_n \times f_n$ 和 $t_n \times f_n$ 的卷积核分别用于提取每一时间维度、 f_n 个时间维度和所有参考的时间维度下所有特征之间的关联信息,分别对应于从小尺度到大尺度信息的提取。卷积层的输出经过 ReLU 函数后,经过维度变换转化为一维向量,以便于输入到下一层全连接层。三个不同尺寸的全连接层分别用于处理三个不同维度的向量,输出的值相加之后得到最终的预测值。

通常来说,多尺度的特征提取能够提升网络的性能,同时针对金融数据的特性,本文构建了一个较为适合处理时间序列数据的基于 CNN 的模型,并在第2章实证分析中和方形尺寸卷积核的 CNN 模型进行了对比。此外,本文也通过增加卷积层的数量探究模型性能的变化,并在实证分析中进行了详细论证。

1.3 基于 RNN、LSTM 和 GRU 的递归网络模型

递归神经网络是一类由各单元链式连接而成的有向循环神经网络。通过各单元的连接,递归神经网络可以储存神经元当前时刻的输入和前一时刻的输出关系,在处理时间序列数据上优于其他类型的神经网络。

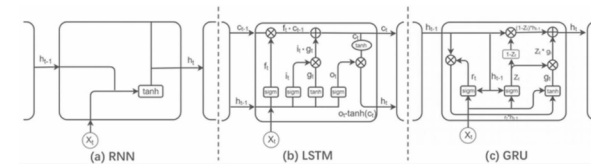


图3 RNN、LSTM 和 GRU 的单元

最早提出的 RNN 模型单元如图3(a)所示,每一个 RNN 单元仅包含全连接层和一个 tanh 激活函数,用来学习当前时刻输入和输出的函数关系。 t 时刻 RNN 单元的具体计算模式如下:

$$h_t = \text{tanh}(w_{hx}x_t + w_{hh}h_{t-1} + b_h) \quad (5)$$

$$o_t = g(w_{(t+1)h}h_t + b_o) \quad (6)$$

其中 w 表示从各类输入到输出的全连接层权值,在不同的 RNN 单元内共享权值。 x_t 、 h_t 和 o_t 分别表示输入序列、隐藏层序列和最终输出序列。 b_h 和 b_o 分别表示两次运算中的偏置项。 g 函数根据任务的不同可以选取不同的函数,例如在分类任务中

一般选择 softmax 函数用于输出最终分类结果。

RNN 对于短期记忆效果很好,但是结构较为简单,难以胜任长期记忆任务,易导致结果失真,运算效率低下。LSTM 通过复杂的设计实现了长时记忆的任务,在时序处理任务中效果显著。LSTM 单元的具体结构如图 3(b) 所示, t 时刻的具体计算模式为:

$$\begin{aligned} f_t &= \text{sigm}(w_{fx}x_t + w_{fh}h_{t-1} + b_f) \\ i_t &= \text{sigm}(w_{ix}x_t + w_{ih}h_{t-1} + b_i) \\ g_t &= \tanh(w_{gx}x_t + w_{gh}h_{t-1} + b_g) \\ c_t &= f_t \times c_{t-1} + i_t \times g_t \\ o_t &= \text{sigm}(w_{ox}x_t + w_{oh}h_{t-1} + b_o) \end{aligned} \quad (7)$$

其中 f_t 表示遗忘门,决定有多少来自上一层的信息会被记录。 i_t 表示输入门,决定有多少输入的信息会被使用。 g_t 表示一个产生备选信息的来源,随时用于更新新的单元状态 c_t 。由当前的单元状态 c_t 和中间输出 o_t 共同决定最终输出 h_t ,并输入到下一时刻 LSTM 单元中。LSTM 通过构建输入门、遗忘门和输出门来实现信息的长时传递,保证之前的信息能够一直参与网络训练。

基于 LSTM 有一些变种结构,GRU 结构是其中一种,如图 3(c) 所示, t 时刻的具体计算模式为:

$$\begin{aligned} r_t &= \text{sigm}(w_{rx}x_t + w_{rh}h_{t-1} + b_r) \\ z_t &= \text{sigm}(w_{zx}x_t + w_{zh}h_{t-1} + b_z) \\ g_t &= \tanh(w_{gx}x_t + w_{gh}h_{t-1} + b_g) \\ h_t &= (1 - z_t) \times h_{t-1} + z_t \times g_t \end{aligned} \quad (8)$$

其中 r_t 表示重置门,决定上一时刻的状态对信息更新的贡献比例。 z_t 表示更新门,决定新的状态 g_t 更新到最终输出 h_t 的比例。和 LSTM 相比,GRU 的特征在于缺少了产生备选信息的来源,同时不再限制当前数据的输入贡献比例。GRU 的结构相较于 LSTM 更简单,但在一些任务处理中表现更加出色。

本节在递归神经网络的前端和末端分别加入了一个全连接层用于改变数据输出的维度,并控制递归神经网络中网络单元的叠加数量和单元内部神经单元的数量,以此构建出多个模型,并在第 2 章实证分析中详细论述了不同参数对预测性能的影响。

1.4 基于注意力机制的递归网络模型

上述递归神经网络处理时间序列数据的过程是一个马尔科夫过程, t 时刻的输出只和 $t-1$ 时刻有关。通常状况下,递归神经网络的条件分布函数满足等式:

$$P\{X(t) \leq x | X(t_n) = x_n, \dots, X(t_1) = x_1\} =$$

$$P\{X(t) \leq x | X(t_n) = x_n\} \quad (9)$$

虽然递归神经网络可以传递 $t-1$ 时刻之前的信息,并将其输送到 t 时刻的信息中,但是这种隐式的更早信息的表示并不能很好地被网络提取。尤其当输入序列很长的时候,模型难以学到合理的向量表示。注意力机制通过保留递归网络中每个时间次序下的输出,训练一个模型对这些输出进行选择性的学习,并将输出序列与之进行关联。

注意力机制的描述如图 4 所示,具体的计算模式如下:

$$\begin{aligned} u_t &= v^T \tanh(w \times h_t + b_u) \\ a_t &= \frac{\exp(u_t)}{\sum_{k=1}^T \exp(u_k)} \\ c_t &= \sum_{k=1}^T a_k^k u_t^k \end{aligned} \quad (10)$$

其中, v^T 和 w 均为需要学习的权值, a^t 是通过 softmax 分类函数对 u_t 进行归一化来计算概率,通过求得概率 a^t 和 u^t 的加权求和得到最终的输出 c^t 。注意力机制会对每一时序的输出生成一个概率,通过概率来决定每一时刻的输出对最终输出所占有的贡献。注意力机制本质上是一种相似度的度量,当前的输入和目标状态越相似,那么当前输入所占的权重就会越大。通过加入注意力机制,可以打破传统递归神经网络的马尔科夫链,将输出和所有参考的时序数据直接关联,获得更加准确的输出结果。

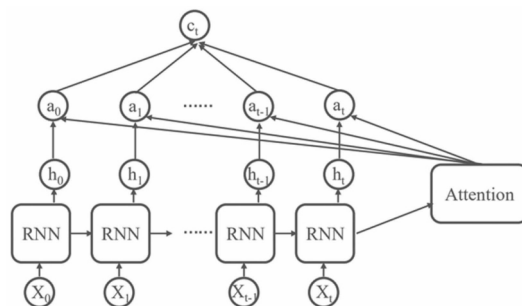


图4 注意力机制

本节将注意力机制加入 1.3 节的递归网络模型中,构建了新型的预测模型,并在第 2 章实证分析中详细论述了该机制对性能的影响。

2 实证分析

2.1 样本选取和实验设置

本文的实验数据为 1991 年 7 月 1 日至 2019 年 4 月 1 日的上证指数,有效数据共计 6771 条,数据来源于 wind。本实验使用前 6650 条数据作为

训练集, 剩余 121 条数据作为测试集。数据指标包含开盘价、收盘价、最高价、最低价、当日成交量、涨跌幅和换手率共计七项指标, 对应于模型中每一时间下特征维度。本文使用当天所有七项指标来预测明日收盘价。在模型训练中, 本文采取过去 N 天的所有数据来预测第 $N+1$ 天的数据, 并使用均方根误差 (Root Mean Squared Error, RMSE) 作为损失函数使其输出结果不断逼近真实数据:

$$L(x, \theta) = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (11)$$

其中 x 表示输入的 N 天数据, θ 表示网络中需要学习的参数, y_i 和 \hat{y}_i 分别表示第 $N+1$ 天的真实数据和预测值。

本实验选取 Tensorflow 作为实验平台, 实验环境为配置 GTX 1080Ti 显卡和 32G 内存的电脑。模

型初始学习率为 0.0007, 共计学习 2000 个轮次 (一个轮次代表训练集中的所有样本全部参与训练一次)。数据首先进行预处理, 减去每一特征维度下的均值再除以标准差, 得到标准化的数据, 然后将数据输入模型中进行训练。

2.2 基本模型验证与对比

本节比较基于 MLP 的模型、基于 CNN 的模型, 基于 RNN、LSTM 和 GRU 的模型和引入注意力机制的上述三种模型, 共计八种神经网络模型。本节中基于 CNN 的模型为图 2 所示的基本模型, 六种递归神经网络模型均只采用了一个递归网络单元, 单元内部神经元数量设为 10。所使用的注意力机制中神经元数量设置为 10。本节中实验使用前 20 天的数据来预测第 21 天的数据, 即 $N=20$ 。图 5 为各模型在测试集 121 条数据样本上的预测曲线图。

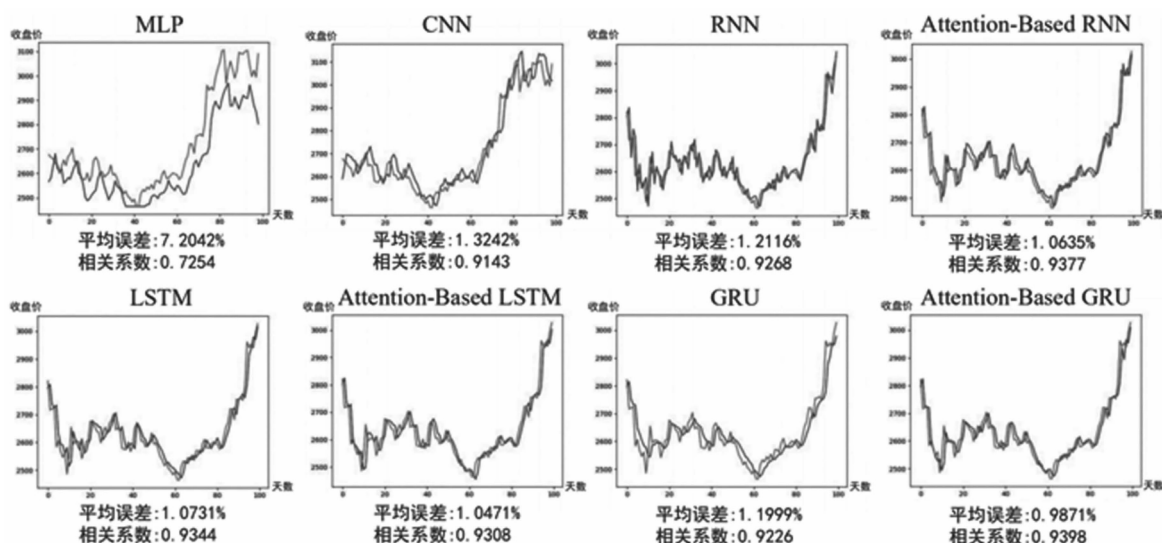


图5 八种模型下的预测曲线

预测曲线图的横坐标为测试集 121 天的天数, 纵坐标为当天的收盘价格, 其中浅色的曲线为真实数据, 深色的曲线为预测数据。本文使用平均误差和相关系数作为衡量模型性能的指标, 分别用来预测整体偏差情况以及预测趋势是否跟实际趋势一致。其中, 平均误差计算方式为:

$$Err(x) = \frac{1}{N_t} \sum_{i=1}^{N_t} |x - x'| / x \quad (12)$$

其中 x 和 x' 表示真实值和预测值, N_t 表示测试集中数据的数量, 在本实验中 $N_t = 121$, 相关系数计算方式为:

$$Corr(x, x') = \frac{Cov(x, x')}{\sqrt{Var(x) Var(x')}} \quad (13)$$

其中 $Cov(x, x')$ 表示真实值和预测值的协方差, $Var(x)$ 和 $Var(x')$ 分别表示真实值和预测值的方差。

可以看到, 基于 MLP 的模型因为很难针对股票数据进行优化, 在预测任务中表现最差; 而基于 CNN 的模型针对股票数据优化之后, 表现出远优于 MLP 的性能, 并能够一定程度逼近递归神经网络的性能。在所构造的六种递归神经网络模型中, 基于注意力机制的 GRU (Attention-Based GRU) 模型表现出了最佳的性能, 而普通的 RNN 结构表现最差。可以看到, 在第二行的四条曲线图中, 深色和浅色的线条几乎完全重合, 这表示模型的预测非常准确。需要强调的是, 加入注意力机制的模型和对应的基础模型相比, 均有不同程度性能提升, 这说明注意力机制确实有助于提高预测任务的模型精度。

2.3 探究基于 CNN 的模型性能与结构的关系

本节通过调整基于 CNN 模型的结构参数, 探究这些重要的结构参数对模型性能的影响。首先

将图 2 中的卷积核替换为图像任务中经常使用的方形卷积核,分别使用了 2×2 、 3×3 、 4×4 、 5×5 的卷积核,模型的预测曲线如图 7 第一行所示。可以看到, 2×2 的卷积核表现最差,平均误差高达 1.6357%, 远远高于其他三个卷积核。这是由于尺寸过小的卷积核对数据的全局信息提取更加困难,

尤其是全局相关性显著的股票数据。 5×5 的卷积核表现出了最佳的效果,可能是由于尺寸较大的卷积核能够更丰富地提取到各特征维度之间的关联信息。然而,上述的所有模型性能均无法和优化后的 CNN 模型相比,这证明了该针对数据形式进行的卷积核优化方式的有效性。

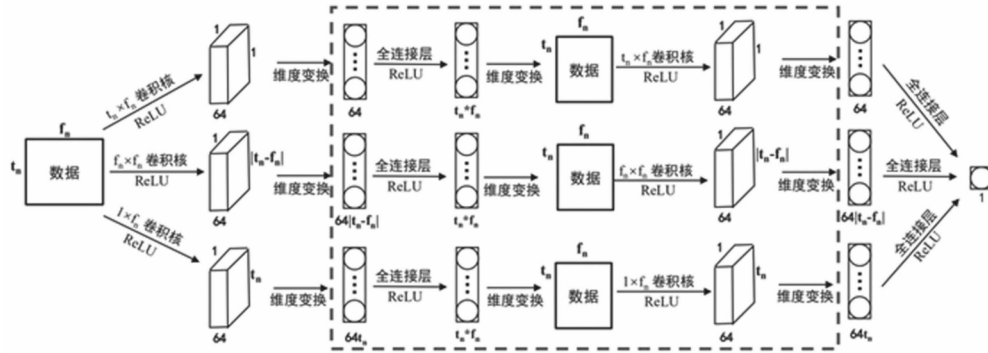


图 6 加一层卷积块的 CNN 模型

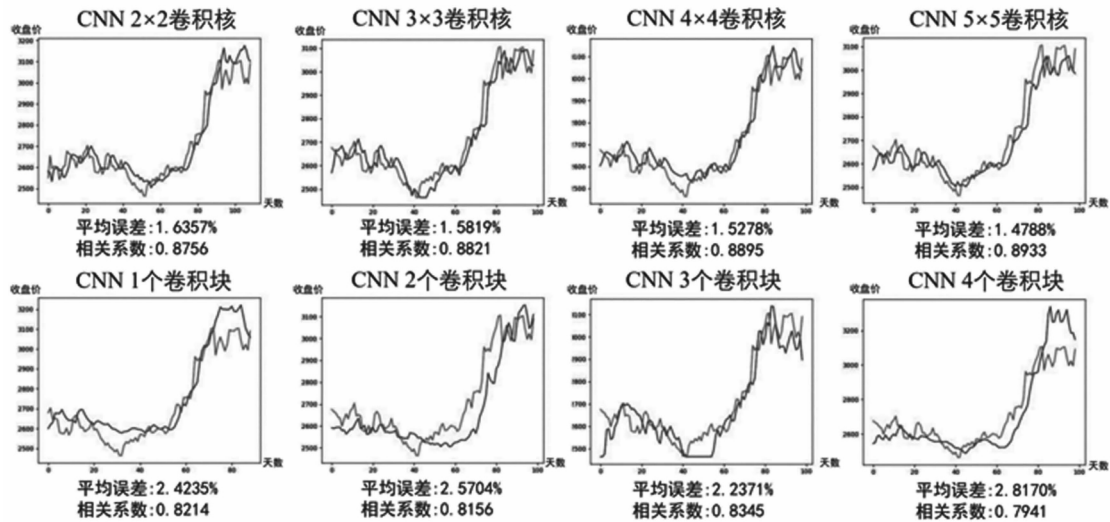


图 7 八种 CNN 模型下的预测曲线

同时,本节针对模型中卷积层的数量对性能的影响也进行了评估。首先构建一个如图 6 的虚线框内所示的卷积块,所构建卷积块包含三条通路上独立的全连接层和卷积层。在卷积块中,首先将卷积层的输出使用全连接层改变维度为 $t_n \times f_n$,然后变换其维度为原始数据的维度,之后将其输入到优化后的卷积核中。图 6 所示模型比图 2 的模型仅多出一个卷积块,在首尾网络层的中间无缝接入。本节对比了加入 1、2、3、4 个卷积块的模型对性能的影响,如图 7 第二行所示。可以看到,随着卷积块的增多,模型性能呈现出下降的趋势。其原因主要有两点:一方面是股票的数据结构较为简单,复杂程度远远不及图像数据,因此使用较为简单的网络结构已经可以逼近模型的上限,更复杂的网络只会导致严重的过拟合,另一方面 CNN 并不适合处理时间序列任务。

2.4 探究递归神经网络模型性能与结构的关系

本节通过修改六种递归网络模型的结构,分别探究递归神经网络中神经元数量和递归神经网络堆叠的数量对模型性能的影响。考虑到需要对比的模型数量过多,本节不再使用如节 2.2、2.3 所示的曲线图来展示模型性能,而是利用统计直方图来直观地展示平均误差的变化趋势。直方图的横坐标表示各类不同的模型,纵坐标表示平均误差。此外,本节使用了表格展示所有递归模型的相关系数。

首先,本节探讨了递归神经网络中神经元数量对模型性能的影响。针对节 2.2 中的基础模型,实验将神经单元中神经元的数量分别改变为 30、50、100、200,相当于增加了单元中全连接层的维度。实验共统计了 6 种模型在不同神经元数量下的表现性能,如图 8 所示。

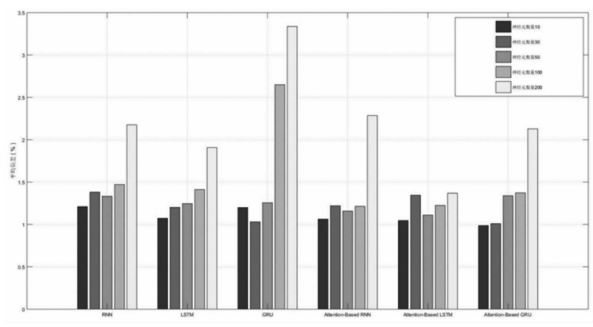


图8 神经元数量对性能的影响

可以看到,在 RNN、LSTM、GRU 三种模型中,整体趋势表现为性能随着神经元数量的增加而降低,节 2.2 的基础模型反而表现出了最佳性能。这是由于所用股票数据的特征维度只有 7,在神经元数量远远大于特征维度时,网络很容易过拟合,而导致在测试集上性能变差。在神经元数量与特征维度接近或相当的时候,网络的表达能力刚好和数据集契合,此时才能达到网络的最优性能。

表 1 神经元数量对相关系数的影响

	RNN	LSTM	GRU	A-RNN	A-LSTM	A-GRU
神经元数量 10	0.9268	0.9344	0.9226	0.9377	0.9308	0.9398
神经元数量 30	0.8781	0.9238	0.9346	0.9081	0.9047	0.9389
神经元数量 50	0.8765	0.9192	0.9184	0.9246	0.9341	0.9198
神经元数量 100	0.8048	0.9005	0.7825	0.9206	0.9372	0.9013
神经元数量 200	0.8902	0.8356	0.6524	0.8895	0.9009	0.6973

表 1 中展示的相关系数也表现出了和平均误差相似的变化趋势,其中 A-表示基于注意力机制的模型。值得强调的是,GRU 模型在高神经元数量下性能表现明显差于另外两类神经网络,而 LSTM 模型展现出了最稳定的性能。

而在基于注意力机制的三种模型中,可以看到,其性能变化趋势较基础模型非常稳定,这是由注意力机制所带来的增益,该机制可以充分地利用所参考的时间维度下的信息,其预测过程不是基础的递归神经网络模型中简单的马尔科夫过程,故而对复杂的网络有更强的适应能力。值得注意的是,基于注意力机制的 GRU 模型在高神经元数量下表现出了和基础模型相似的高预测误差,而 LSTM 模型展示出和基础模型类似的鲁棒性。

关于递归神经单元堆叠的数量对性能的影响,本节也进行了大量的对比实验。实验使用神经元数量为 10 的递归神经单元,针对六种递归网络模型分别堆叠了 2、3、4、5 个神经单元,模型性能展示如图 9 所示。

可以看到,在 RNN、LSTM、GRU 三种基础递归

网络模型中,堆叠两到三个神经单元取得了最优性能,之后随着神经单元的增多,整体呈现性能变差的趋势。在基于注意力机制的三个模型中,堆叠一到两个神经单元可以取得最优性能,但是在单元数量增多后,预测性能依然会变差。值得一提的是,三种基于注意力机制的模型整体性能和稳定性优于基础模型,这佐证了图 8 中得出的结论:注意力机制可以有效改善过拟合问题,提高模型的性能和稳定性。表 2 展示了以上各模型的相关系数,其变化趋势整体和平均误差相似。

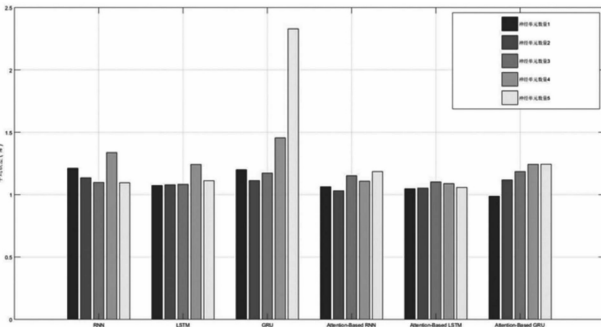


图9 递归神经单元数量对性能的影响

表 2 递归神经单元数量对相关系数的影响

	RNN	LSTM	GRU	A-RNN	A-LSTM	A-GRU
神经元数量 1	0.9268	0.9344	0.9226	0.9377	0.9308	0.9398
神经元数量 2	0.9324	0.9350	0.9317	0.9391	0.9388	0.9335
神经元数量 3	0.9284	0.9299	0.9247	0.9301	0.9400	0.9318
神经元数量 4	0.8861	0.9257	0.8900	0.9313	0.9373	0.9201
神经元数量 5	0.9298	0.9273	0.6671	0.9298	0.9378	0.9212

对比图 8 的实验结果 2 块神经元数量为 10 的递归神经单元和 1 块神经元数量为 20 的神经单元在性能表现上整体呈现出相反的改变趋势 这说明改变堆叠的递归神经单元数量和改变神经单元中神经元数量对性能的影响是不一致的。同时可以发现 ,即使堆叠神经单元数量增多导致网络过拟合和性能变差 ,其性能变差趋势依然缓和于改变单元内神经元数量。该实验对比结果表明 ,对网络模型需求更复杂的结构和更强大的拟合能力时 ,堆叠神经单元数量的优先级高于增加单元内神经元的数量。

2.5 探究网络模型性能与参考时间维度 N 的关系

本节通过修改参考时间维度的长度来探索该因素对模型性能的影响。如节 2.2 描述 ,模型使用前 20 天的数据预测第 21 天的数据 2.3、2.4 节中的模型遵循这一设置 ,取 $N = 20$ 。在本节中 ,实验修改 N 的数值分别为 10、30、40 ,并针对节 2.2 中的八种网络模型进行了性能评估 ,结果如图 10 所示。因基于 MLP 的模型性能表现过差 ,对直方图

的绘制会产生影响 ,因此图 10 并没有绘制基于 MLP 的模型的结果。

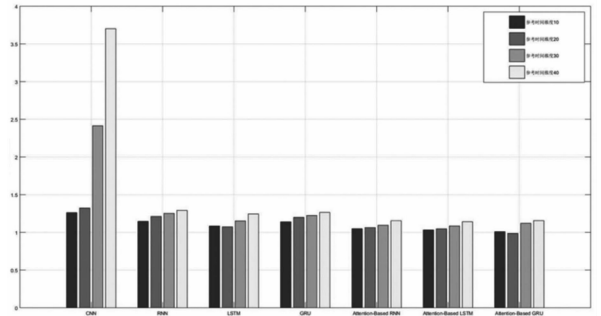


图 10 参考时间维度对性能的影响

表 3 展示了各时间维度下模型的相关系数变化。可以看到 ,随着参考时间维度的增加 ,模型性能呈下降趋势。这是由于过早的时间数据对当天数据的影响已经微乎其微 ,若依然将其作为模型输入 ,反而会对预测结果产生负面影响。值得注意的是 ,基于 CNN 的模型也表现出了这一趋势 ,这说明针对股票数据进行的 CNN 模型的优化也对时间维度具有敏感性。

表 3 参考时间维度对相关系数的影响

	CNN	RNN	LSTM	GRU	A-RNN	A-LSTM	A-GRU
参考时间维度 10	0.9254	0.9647	0.9677	0.9624	0.9684	0.9676	0.9665
参考时间维度 20	0.9143	0.9268	0.9344	0.9226	0.9377	0.9308	0.9398
参考时间维度 30	0.7224	0.8328	0.8433	0.8388	0.8416	0.8483	0.8547
参考时间维度 40	0.5532	0.8029	0.8460	0.8280	0.8514	0.8511	0.8372

3 结论

本文选取了近 28 年上证指数的数据 ,分别构造了基于 MLP、CNN、RNN、LSTM、GRU 的网络模型进行数据的预测。同时 ,为了克服基于 RNN、LSTM、GRU 等递归网络模型的局限性 ,本文引入了注意力机制来增加模型性能和防止过拟合。针对构造的八种网络模型 ,本文通过修改模型参数 ,进行了大量的对比实验 ,旨在探索基于神经网络的股票预测模型合理的优化方向。实验结果表明 ,MLP 和 CNN 并不适合处理时间序列任务 ,其性能和递归网络模型差距较大。在需要对递归网络模型进行复杂设计以适应庞大数据集时 ,优先增加其单元数量而不是增加单元内神经元数量 ,可以有效防止过拟合且增加网络性能。而注意力机制通过打破基础递归网络模型预测时的马尔科夫过程 ,直

接利用更早时期的数据 ,在提升网络性能的同时可以有效防止过拟合带来的性能下降 ,在时间序列预测任务中表现优异。

参考文献:

[1] Fama Eugene F. Random walks in stock market prices [J]. Financial analysts journal ,1995 ,51(1) : 75-80.

[2] Singh R , Srivastava S. Stock prediction using deep learning [J]. Multimedia Tools and Applications ,2017 ,76(18) : 18569-18584.

[3] Chulho Jung , Roy Boyd. Forecasting UK stock prices [J]. Applied Financial Economics ,1996 ,6: 279-286.

[4] Bessler W , Lückoff P. Predicting stock returns with bayesian vector autoregressive models [M]. Data Analysis , Machine Learning and Applications. Springer , Berlin , Heidelberg ,2008. 499-506.

[5] Ayodele Adebisi A , Aderemi O. Adewumi Charles K. Ayo. Stock price prediction using the ARIMA model [C]. 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. Cambridge:

- IEEE, 2014. 106-112.
- [6] Ren S, He K, Girshick R, et al. Faster r-cnn: towards real-time object detection with region proposal networks [C]. Advances in neural information processing systems. Montreal: NIPS, 2015. 91-99.
- [7] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]. Advances in neural information processing systems. Harrahs and Harveys: NIPS, 2012. 1097-1105.
- [8] Milletari F, Navab N, Ahmadi S A. V-net: fully convolutional neural networks for volumetric medical image segmentation [C]. 2016 Fourth International Conference on 3D Vision (3DV). Stanford: IEEE, 2016. 565-571.
- [9] 刘豹, 胡代平. 神经网络在预测中的一些应用研究 [J]. 系统工程学报, 1999, 14(4): 338-343.
- [10] 孙瑞奇. 基于 LSTM 神经网络的美股股指价格趋势预测模型的研究 [D]. 北京: 对外经贸大学, 2015. 1-61.
- [11] Peter Zhang G. Time series forecasting using a hybrid ARIMA and neural network model [J]. Neurocomputing, 2003, (50): 159-175.
- [12] Ozbayoglu Murat A, Ismet Bahadir. Comparon of bayesian estimation and neural network model in stock market trading [J]. Intelligent Engineering Systems through Artificial Neural Networks, 2011, (20): 74-81.
- [13] 刘海玥, 白艳萍. 时间序列模型和神经网络模型在股票预测中的分析 [J]. 数学的实践与认识, 2011, 41(4): 14-19.
- [14] Pal S K, Mitra S. Multilayer perceptron, fuzzy sets, and classification [J]. IEEE Transactions on neural networks, 1992, 3(5): 683-697.
- [15] Mikolov T, Karafiát M, Burget L, et al. Recurrent neural network based language model [C]. Eleventh annual conference of the international speech communication association. Chiba: ISCA, 2010. 1045-1048.
- [16] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural computation, 1997, 9(8): 1735-1780.
- [17] Chung J, Gulcehre C, Cho K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J]. arXiv preprint arXiv: 1412. 3555, 2014.
- [18] Chorowski J K, Bahdanau D, Serdyuk D, et al. Attention-based models for speech recognition [C]. Advances in neural information processing systems. Montreal: NIPS, 2015. 577-585.
- [19] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [J]. arXiv preprint arXiv: 1502. 03167, 2015.
- [20] Lei Ba J, Kiros J R, Hinton G E. Layer normalization [J]. arXiv preprint arXiv: 1607. 06450, 2016.