

基于深度双向 LSTM 的股票推荐系统

曾 安¹ 聂文俊²

(广东工业大学计算机学院 广州 510006)¹ (广东省大数据分析处理重点实验室 广州 510006)²

摘 要 面对越来越复杂的数据环境,以经典统计学模型为主的股票预测模型在一定程度上已无法满足人们对预测准确性的要求。深度学习因具有较强的学习能力和抗干扰能力,已逐渐被应用于股票推荐中。但传统的股票推荐模型要么从未考虑时间因素,要么仅考虑时间上的单向关系。因此,文中提出了一种基于深度双向 LSTM 的神经网络预测模型。该模型充分利用了时间序列上向前、向后两个时间方向的上下文关系,解决了长时间序列上的梯度消失和梯度爆炸问题,能够学习到对时间有长期依赖性的信息。同时,该模型引入了 Dropout 策略,在一定程度上解决了深层网络模型带来的训练难、收敛速度慢和过拟合等问题。在 S&P500 数据集上的实验表明,基于深度双向 LSTM 的神经网络预测模型比现有预测模型在误差上降低了 2%~5%,使决定系数(r^2)提高了 10%。

关键词 推荐系统,股票预测,深度 RNN,双向 LSTM

中图法分类号 TP181 文献标识码 A DOI 10.11896/jsjcx.180901771

Stock Recommendation System Based on Deep Bidirectional LSTM

ZENG An¹ NIE Wen-jun²

(School of Computer, Guangdong University of Technology, Guangzhou 510006, China)¹

(Guangdong Key Laboratory of Big Data Analysis and Processing, Guangzhou 510006, China)²

Abstract With the diversity of applications scenarios and rapid growth of data, the stock prediction models based on classical statistical methods are unable to meet the requirements for high prediction accuracy. But traditional stock recommendation models either never consider the time factor or just consider the unidirectional relationship over time. However, existing stock recommendation models based on deep learning rarely consider the time factor. This paper proposed a deep bidirectional LSTM model for stock prediction, which makes full use of context relationship in the forward direction and backward direction of time series. The problem of vanishing gradient and exploding gradient are solved by introducing LSTM when dealing with long-term sequence. The proposed model can learn information which has long-term dependence on time. At the same time, dropout strategy is introduced to prevent over-fitting caused by deep network model and speed up the training. Experiments on S&P500 dataset show that the neural network prediction model based on the deep bidirectional LSTM outperforms the existing prediction models, the error is about 5% lower, and the coefficient of determination (r^2) is increased by 10%.

Keywords Recommendation system, Stock forecast, Deep recurrent neural networks, Bidirectional long short-term memory

1 引言

股票市场波动预测对投资者、投资组和经理、资产估值以及风险管理都很重要^[1]。风险投资在给投资者带来巨大回报的同时,也带来了相当大的损失^[2],为了减小损失和风险,股票预测与推荐逐渐成为热点。

从传统的统计角度出发,许多金融时间序列预测模型和方法被提出,以减小损失和风险。例如:Narendra 等^[3]将 ARMA 模型和 GARCH 模型应用于 NSE 印度股票市场数

据;Rosas-romero 等^[4]提出金融时间序列模型,来预测美国证券交易所上市的 20 家公司的未来价格趋势;Chen 等^[5]建立了基于灰色模型 GM(1,1)和灰色动态中性网的预测模型,并对 Dow Jones Cbn China 600 Inda 进行了预测;Huang 等^[6]将 ARX 预测模型与灰色系统理论和粗糙集理论相结合,建立了股票市场的自动预测和投资组合选择机制。这些模型虽然能直观地描述数据特征,但都受到了某些数据假设的限制,如正态性或自相关。然而,这些假设并不适用于复杂的现实世界数据^[2]。

到稿日期:2018-09-19 返修日期:2019-02-21 本文受国家自然科学基金项目(61772143, 61300107),广东省自然科学基金项目(S2012010010212),广州市科技计划项目(201601010034, 201505031501397),广东省大数据分析处理重点实验室开放基金项目(201801)资助。
曾 安(1978—),女,博士,CCF 会员,主要研究方向为数据挖掘, E-mail:437871335@qq.com(通信作者);聂文俊(1992—),男,硕士,主要研究方向为数据挖掘。

从机器学习的角度来看,机器学习的自学习能力强,同时对噪声数据的鲁棒性和容错性较强,从而可以逼近非线性关系,对非线性数据进行处理。因此,机器学习被逐渐应用于股票指数的预测。Wang 等^[7]将决策树(DT)算法和支持向量机(SVM)模型混合,首先利用 DT 算法过滤大部分噪声数据,然后利用 SVM 处理第二阶段的训练数据,从而预测期货价格走势。Chen 等^[8]建立特征加权的支持向量机(SVM)和 K 近邻算法,对上海和深圳股票市场指数进行预测,该模型在短期、中期和长期都具有较好的预测能力。Wang^[9]为了提高股票价格的预测能力,提出了一种基于小波神经网络的股票价格预测方法。Nayak 等^[10]利用人工化学反应优化(ACRO)算法训练多层感知机(MLP)来预测股票市场指数。Chatzis 等^[11]利用分类树、支持向量机、随机森林、神经网络、极端梯度提升和深度神经网络等一系列技术的优点,分析国际股票市场危机事件。目前,机器学习虽然在图像识别、语音识别、情绪分类、预测数据等诸多领域得到广泛应用并且收获了不错的成效,但是在金融行业的应用与研究还不够深入,也并未考虑到股票波动和时间维度的紧密联系。

循环神经网络(Recurrent Neural Networks,RNN)是一种具有时序性的神经网络,其隐藏层之间的互连结构反映了时间序列之间的相互影响关系,但是 RNN 存在梯度下降快、无法收敛到最优解等问题^[12]。双向 LSTM 模型^[13-15]通过添加门机制解决了 RNN 网络的梯度问题,并且利用时间序列向前、向后两个时间方向上的上下文关系,使得预测准确度较单向 RNN 网络有所提高。RNN 和 LSTM 模型主要应用于语音识别、文本处理等方面,近年来国内外学者也开始在股票市场预测方面使用循环神经网络。Sun^[16]分别利用 BP 神经网络、RNN 和 LSTM 模型对美股和上证指数进行预测,实验得出 LSTM 模型的预测精度最高。Fischer 等^[17]将 LSTM 网络应用于预测 S&P500 在 1992 年至 2015 年期间的样本外方向移动,并与随机森林(RAF)、深度神经网络(DNN)和 Logistic 回归分类器(LOG)进行比较,发现其优于无记忆分类方法。

本文针对现有的股票预测模型未考虑时间维度的问题,提出了一种深度双向 LSTM 的神经网络预测模型,并对股票指数进行预测,通过预测结果进行股票推荐。该模型能学习并利用向前、向后两个时间方向的上下文关系来进行预测,同时引入了 Dropout^[18]策略,以进一步提高模型的泛化能力。

2 相关工作

2.1 双向循环神经网络(BRNN)

在其他神经网络模型中,层与层之间是全连接的,而每层之间的节点是无连接的,这使得每个时刻的信息处理都是独立的。而循环神经网络 RNN 在隐藏层中增加了节点的互连,隐藏层的输入不仅包括输入层的输出,还包括上一时刻隐藏层的输出,因此网络模型记忆前面的信息并将其应用于处理当前输出数据的计算。双向循环神经网络 BRNN 能够同时利用序列中的历史信息和未来信息。将序列信息分两个方向输入模型中,使用两个隐藏层保存两个方向的输入信息,并将隐藏层相应的输出连接到相同的输出层。BRNN 网络的展开结构示意图如图 1 所示。

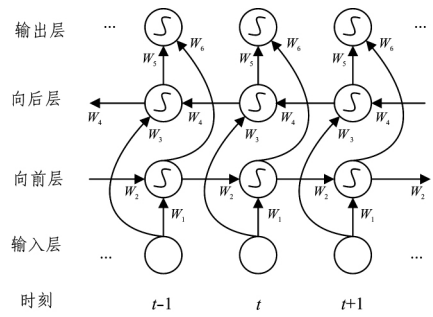


图 1 BRNN 的展开结构示意图
Fig. 1 Expansion structure of BRNN

2.2 LSTM 长短时记忆模型

长短时记忆模型(Long Short-Term Memory,LSTM)^[19]是循环神经网络的变体。尽管在理论上,RNN 可以处理任何长距离依赖问题,但实际上,其由于梯度消失、爆炸等问题很难实现。对此,LSTM 通过引入门机制和记忆单元提供了解决方案,即用 LSTM 单元代替 RNN 中的隐藏层神经单元。LSTM 单元结构如图 2 所示。

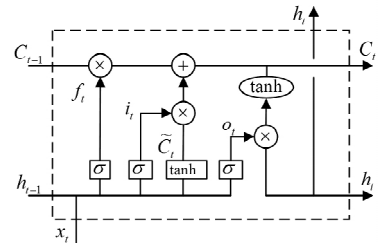


图 2 LSTM 模型
Fig. 2 LSTM model

LSTM 中保存的历史信息受输入门、遗忘门和输出门控制。在 t 时刻,各门状态的数学表达式如下:

$$\tilde{c}^{(t)} = \tanh(W_{xc}x^{(t-1)} + W_{hc}h^{(t-1)} + b_c) \tag{1}$$

$$i^{(t)} = \sigma(W_{xi}x^{(t-1)} + W_{hi}h^{(t-1)} + W_{ci}c^{(t-1)} + b_i) \tag{2}$$

$$f^{(t)} = \sigma(W_{xf}x^{(t-1)} + W_{hf}h^{(t-1)} + W_{cf}c^{(t-1)} + b_f) \tag{3}$$

$$c^{(t)} = f^{(t)} \otimes c^{(t-1)} + i^{(t)} \otimes \tilde{c}^{(t)} \tag{4}$$

$$o^{(t)} = \sigma(W_{xo}x^{(t-1)} + W_{ho}h^{(t-1)} + W_{co}c^{(t-1)} + b_o) \tag{5}$$

$$h^{(t)} = o^{(t)} \otimes \tanh(c^{(t)}) \tag{6}$$

其中, $x^{(t)}$ 是 t 时刻的输入数据; $h^{(t)}$ 是 t 时刻 LSTM 单元的 outputs 状态值; $\tilde{c}^{(t)}$ 是 t 时刻记忆单元的候选值; $i^{(t)}$ 是输入门 t 时刻的状态值; $f^{(t)}$ 是遗忘门 t 时刻的状态值; $o^{(t)}$ 是输出门 t 时刻的状态值; W 为对应的权值; b 是对应的偏置参数; \otimes 表示元素间的点积,逐点相乘。记忆单元的状态值由输入门和遗忘门共同调节。

LSTM 的门机制使得模型可以捕捉长距离历史信息。为了同时获取上下文信息,本文采用双向 LSTM,因此 BLSTM 中的隐藏状态 $h^{(t)}$ 可表示为:

$$h^{(t)} = \vec{h}^{(t)} \parallel \overleftarrow{h}^{(t)} \tag{7}$$

其中, $\vec{h}^{(t)}$ 和 $\overleftarrow{h}^{(t)}$ 分别是 t 时刻 LSTM 中输入数据的前向和后向的隐藏状态, \parallel 表示整合操作。

2.3 Dropout

Dropout 的数学形式表达式为:

$$y=f(W \cdot d(x))$$

$$d(x)=\begin{cases} mask \star x, & \text{训练阶段} \\ (1-p)x, & \text{其他} \end{cases} \quad (8)$$

其中, p 为 Dropout 率, $mask$ 为以 $1-p$ 为概率的贝努力分布生成的二值向量。

从式(8)可以看出, Dropout 与 L1 和 L2 范式正则化不同, Dropout 并不会修改代价函数, 而是修改深度网络本身。Dropout 随机“删除”网络中的一些隐藏神经元, 保持输入输出神经元不变。这样, 对于一个网络而言, Dropout 便是用相同的数据训练了多个不同的神经网络, 产生了多个不同程度的拟合状态。但这些网络共用一个损失函数, 相当于对神经网络本身进行了优化, 求取了所有状态的平均值; 同时, 减少了神经元之间的相互协同关系, 增加了网络的鲁棒性。

3 模型构建

本文提出的深度双向 LSTM 模型 (Deep Bidirectional Long Short-Term Memory, DBLSTM) 的结构如图 3 所示。该模型分为输入层、双向 LSTM 部分、全连接层部分、输出层。双向 LSTM 部分由多层双向 LSTM 组成, 全连接层部分由多层全连接层组合。

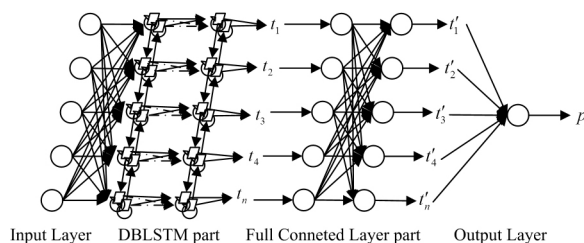


图 3 深度双向 LSTM 模型 (DBLSTM)

Fig. 3 Deep bidirectional LSTM model (DBLSTM)

BP 神经网络未考虑序列时间因素, 单向 RNN 网络也只能向一个方向学习时间序列的数据特征。但是在时间序列中, 对于某一时间点的数据来说, 除了过去时刻的数据与其相关联外, 未来时刻的数据对其也有影响。例如, 对于序列 abcdef 和 abcba, 如果只已知序列 abc, 是无法准确预测后面的序列是 def 还是 cba。因此, 本文运用双向 LSTM 网络, 其具有以下优势: 1) 可以回避 RNN 在长时间序列上的梯度消失和梯度爆炸问题; 2) 可以学习对时间有长期依赖性的信息; 3) 可以利用时间序列向前、向后两个时间方向的上下文关系。为了实现数据的深层次特征挖掘, 将多个双向 LSTM 层叠加起来, 通过多层神经网络结构来学习时间序列的深层特征。除了添加双向 LSTM 层以外, 还可以添加全连接层 (Fully Connected Layers, FC)。全连接层具有良好的非线性映射能力, 能对双向 LSTM 输出的非线性特征进行加权处理, 即对这些非线性特征进行组合。这一过程实质上就是在一个向量空间中学习一个 (非线性) 方程, 以简化的权重方式学习到这些非线性的组合特征, 类似于统计学中的主成分分析 (PCA) 与数学期望。但是, 随着网络层的增加, 模型的训练难度增加、收敛速度变慢, 容易出现过拟合等问题, 因此使用 Dropout 策略来解决这些问题。直观来说, Dropout 的原理就是在训练网络时, 以预先设置的概率停止神经元的输出, 部分神

经单元的“罢工”意味着每次的网络训练只有一部分数据特征参与, 从而防止网络过多地学习训练集的数据特征, 以达到防止过拟合的目的。

3.1 算法描述

Step 1 数据预处理。将数据 $X=\{x_1, x_2, \dots, x_N\}$ 归一化处理。

Step 2 初始化神经元的细胞状态 $C^{(0)}$ 和 $h^{(0)}$, 将预处理后的数据输入第一层双向 LSTM 神经元。

Step 3 按照式(1)一式(3)和式(5), 分别计算出当前神经元的记忆候选值 $\tilde{C}^{(t)}$ 、输入门状态 $i^{(t)}$ 、遗忘门状态 $f^{(t)}$ 和输出门状态 $o^{(t)}$ 。

Step 4 根据式(4)计算当前神经元的记忆状态值 $C^{(t)}$ 。

Step 5 根据式(6), 计算当前神经元的输出值 $h^{(t)}$ 。

Step 6 保留 $C^{(t)}$ 和 $h^{(t)}$, 并将其使用到下一时刻的 LSTM 神经元计算中。

Step 7 重复 Step 3—Step 6, 直到向前层、向后层的 LSTM 神经元均学习完全部的时间序列。根据式(7)和正、反向神经元最后的输出 $\vec{h}^{(n)}$ 和 $\overleftarrow{h}^{(n)}$, 计算 $h^{(n)}$, 进入下一层双向 LSTM。

Step 8 重复以上步骤, 直到最后一层双向 LSTM, 输出非线性数据特征 t 。

Step 9 进入多层 FC 部分, 对非线性特征进行加权计算, 从而得到隐含层的输出 t' 。

Step 10 将 t' 输入输出层, 从而得到预测值 p 。

Step 11 计算损失函数值, 不断更新网络参数, 直到迭代结束。

4 实验验证

4.1 数据集

数据集来自加州大学欧文机器学习库 (UC Irvine Machine Learning Repository), 其收集了 Google Finance API 中 2017 年 4 月到 2017 年 8 月的 S&P 500 指数, 每天正常交易的时间为 6.5 h, 每周正常交易的天数为 5, 总计 500 支证券 41266 min 的证券指数。在实验阶段, 按照文献[20]中的实验策略, 使用每支证券中前 2100 min 的证券指数, 将选取的数据按 2:1 的比例分为训练集和测试集。

4.2 性能评价

本文选取 MAE, MSE, RMSE, Explained Variance Score, r^2 这 5 项数据作为预测性能的评价标准。

平均绝对误差 (Mean Absolute Error, MAE) 是绝对误差的平均值, 能更好地反映预测值误差的实际情况。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - p_i|$$

均方误差 (Mean Squared Error, MSE) 是指参数估计值与参数真值之差平方的期望值, 可以评价数据的变化程度。MSE 值越小, 说明预测模型具有更好的精确度。

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2$$

均方根误差(Root Mean Squared Error, RMSE)是均方误差的算术平方根。

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^n(y_i-p_i)^2}$$

解释方差评分(Explained Variance Score, EVS)是指数学模型对于给定的数据集,其预测数据方差与原始数据方差的比值。解释方差的最大值是 1,代表预测值是最优的,解释方差评分越小代表预测结果越差。解释方差评分的计算公式如下:

$$explained\ variance(y,p)=1-\frac{Var(y-p)}{Var(y)}$$

决定系数(coefficient of determination) r^2 是预测值和观测值之间的线性相关性, r^2 值的范围在 0 到 1 之间, r^2 值为 1 表示完全相关。

$$r^2=\frac{(\sum_{i=1}^n(y_i-\bar{y})(p_i-\bar{p}))^2}{\sum_{i=1}^n(y_i-\bar{y})^2\sum_{i=1}^n(p_i-\bar{p})^2}$$

4.3 网络参数的确定

4.3.1 时间窗口长度

在三隐层 LSTM 网络和五隐层 LSTM 网络上同时进行实验,实验测试变量为时间窗口长度 W ,实验数据是某支证

券前 2 100 min 的证券指数,实验 5 次后求得 train Score 和 test Score 的平均值,通过比较得出时间窗口长度的最佳参数,结果如图 4 所示。

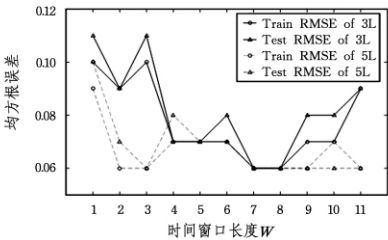


图 4 不同 W 值下的 RMSE 值
Fig. 4 RMSE values with different W

由图 4 可知,三隐层网络在 $W=7$ 或 $W=8$ 时达到最优;五隐层网络在 $W=3,7,8,9$ 时达到最优。 $W=7$ 和 $W=8$ 时的预测结果如图 5 所示,不难看出 $W=8$ 时五隐层网络在数据末尾处较为准确地预测到了证券指数的突然增长。因此,时间窗口长度 W 取 8 时为最佳的调优参数。此外,五隐层网络普遍比三隐层网络的 RMSE 低,说明深层网络能适应更多不同长度的时间窗口。

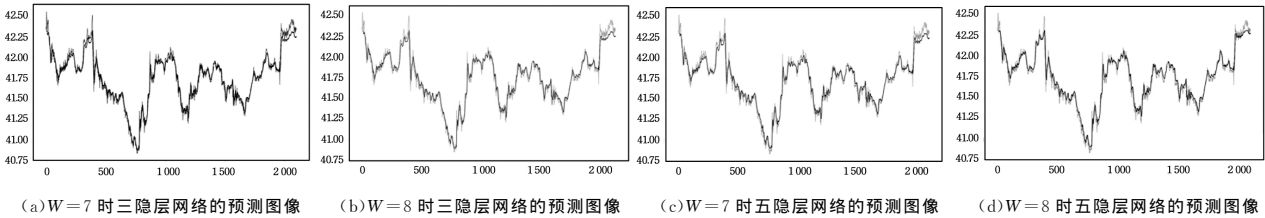


图 5 预测结果
Fig. 5 Prediction results

4.3.2 Dropout

在 RNN 网络模型中,Dropout 可以作为网络层添加到其他网络层之后。现以五隐层 LSTM 网络为例,在不同隐含层之后添加 Dropout,找出添加 Dropout 的最佳位置。五隐层 LSTM 网络的结构如图 6 所示,其中 1—4 号箭头是添加 Dropout 的位置。具体方案如下:方案 1 不添加 Dropout;方案 2 在 LSTM 层和全连接层之间(3 号位置)添加 Dropout;方案 3 在每个 LSTM 层后(1—3 号位置)都添加 Dropout;方案 4 在所有隐含层之间(1—4 号位置)都添加 Dropout。

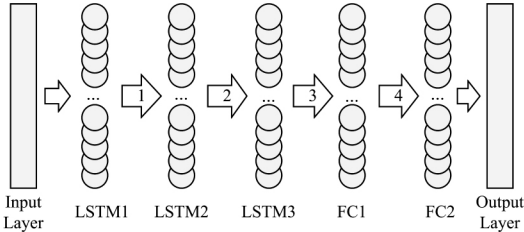


图 6 五隐层 LSTM 模型的结构
Fig. 6 Structure of five hidden layers of LSTM model

1) 泛化能力

抽取 10 支股票,以此进行泛化能力实验,每支股票数据实验 5 次,计算 5 次结果的平均值。其中,9 支股票的前后涨幅情况相似,而第 10 支股票的前后涨幅情况有所不同,表现出了明显不同的数据特征。在 9 支普通股票中,4 种方案的

表现比较相近,差异不大。而在第 10 支股票中(见表 1),方案 2 保持了相对较高的 r^2 ;方案 1 和方案 3 的 r^2 值偏低,未超过 0.5;而方案 4 则出现了负值。结果表明,对于数据特征有明显改变的数据集,方案 2 可以很快地学习到数据特征的变化趋势。

表 1 第 10 支股票的实验结果

Table 1 Experimental results of 10th security					
	MAE	MSE	RMSE	EVS	r^2
方案 1	0.17	0.05	0.23	0.69	0.44
方案 2	0.12	0.02	0.16	0.88	0.74
方案 3	0.17	0.05	0.23	0.70	0.45
方案 4	0.24	0.10	0.32	0.44	-0.09

从 10 支股票的各项平均值来看(见表 2),在各项误差值上 4 种方案均表现良好。其中方案 1 的各项误差值都是最低的,分别为 0.12,0.08,0.17,其次是方案 2 和方案 3。在解释方差评分和 r^2 值上,方案 2 表现最佳,解释方差评分为 0.87, r^2 为 0.82,其次是方案 1,再次是方案 3。整体来看,虽然方案 2 的误差较方案 1 稍大,但是在解释方差评分和 r^2 值上却比方案 1 好,这也证明了添加 Dropout 可以增加网络模型的泛化能力。方案 3 的效果没有方案 2 好,与方案 1 相比有略微提升。而方案 4 因为在全连接层之间也添加了 Dropout,导致网络效果反而下降很多。

表 2 4 种方案 5 项标准的平均值
Table 2 Average values of five standards in four projects

	MAE	MSE	RMSE	EVS	r^2
方案 1	0.116	0.080	0.165	0.847	0.809
方案 2	0.124	0.102	0.174	0.857	0.822
方案 3	0.129	0.105	0.183	0.841	0.794
方案 4	0.143	0.126	0.202	0.811	0.735

2)收敛性

4 种方案在某一支股票上的损失函数图像如图 7 所示。不难看出,前 3 种方案的收敛速度很快,在 100~150 步之间便达到了收敛;而方案 4 则要在 500 步左右才能收敛,收敛速度相对较慢。

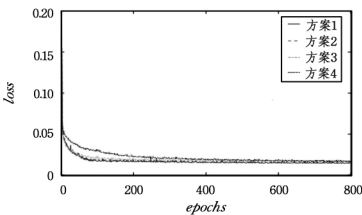


图 7 损失函数图像
Fig. 7 Loss function figure

4 种方案在泛化能力和收敛速度上的对比证明了添加 Dropout 可以加强模型的泛化能力,同时也可发现,Dropout 的添加数目和添加位置对于泛化能力的提升程度起着决定性作用。对于本文提出的模型,最佳方案为在 RNN 网络部分和全连接网部分之间添加 Dropout,其次是在所有 RNN 层后都添加 Dropout。

4.3.3 各 RNN 改进网络的比较

设计方案 1—方案 3,分别使用双向 LSTM、单向 LSTM 以及 GRU 这 3 种网络与全连接网组合,每种方案分别在 30 支股票数据上进行实验,计算出平均值。在这 30 支股票数据中,方案 1(双向 LSTM)只有一支股票的 r^2 值低于 0,而方案 2(单向 LSTM)有两支股票的 r^2 值低于 0,方案 3(GRU)则有 4 支股票的 r^2 值低于 0。30 支股票预测结果的 5 项评价标准的平均值如表 3 所列,从结果来看,方案 1 在各项上都优于其他方案,其次是方案 2。实验表明,双向 LSTM 网络比单向 LSTM 网络、GRU 网络的泛化能力更强,更能适应一些异常数据集,学习到不规律的数据变化特征。

表 3 5 项标准的平均值
Table 3 Average value of five standards

	MAE	MSE	RMSE	EVS	r^2
方案 1	0.075	0.041	0.114	0.891	0.861
方案 2	0.087	0.052	0.129	0.870	0.789
方案 3	0.105	0.0677	0.151	0.860	0.710

本节从时间窗口长度 W 、Dropout 的添加位置和参数,以及 RNN 变形网络的优劣这 3 个方面进行了大量的对比实验。实验结果表明:时间窗口长度 W 为 8 时,在不同深度的网络结构中都能使模型达到最优,并且深度越深的网络对时间窗口长度 W 的要求越低;对于 Dropout 的添加位置,本文提出的模型的最优方案是在 RNN 网络层之后添加 Dropout,在全连接层之间添加会导致收敛速度变慢,并且实验证明适

当地添加 Dropout 能在一定程度上提高模型的泛化能力;双向 LSTM 网络在预测方面的性能优于单向 LSTM 网络和 GRU 网络。最后,本文模型的最优参数与结构为:1)时间窗口长度为 8;2)添加 Dropout,添加位置在 BLSTM 部分与全连接层之间;3)使用多层双向 LSTM 网络。

4.4 对比实验

现将深度双向 LSTM 模型与文献[20~22]中的模型进行比较。模型 1 为本文提出的 4 隐层双向 LSTM 模型,隐含层由三层双向 LSTM 加一层全连接网组成,每层双向 LSTM 含有 32 个神经单元,全连接网含有 64 个神经单元,在最后一层双向 LSTM 层后添加系数为 0.1 的 Dropout。模型 2 为文献[20]中提及的 CNN 模型,模型结构由两层卷积层(卷积核个数为 64)、一层最大池化层、两层卷积层(卷积核个数为 128)、一层平均池化层和全连接层构成,Dropout 系数为 0.5,卷积层激活函数均为 relu 函数,全连接层激活函数为 sigmoid 函数。模型 3 为文献[21]中提及的包含 3 个隐含层的 MLP 神经网络,其隐含层神经单元个数均为 64。模型 4 为文献[22]提出的一种弹性网正则化 LSTM 模型,采用正则化方法修改 LSTM 模型的目标函数,两层单向 LSTM 组成隐含层,L1 和 L2 范数的系数均为 0.01。除网络模型以外,其他参数、环境保持一致。时间窗口 W 的长度为 8,全连接层激活函数为 sigmoid 函数,使用 $\text{adam}(lr=0.001)$ 作为优化器, batch size 为 40,使用 30 组数据,每组数据迭代 1000 步,实验 5 次并取 5 次结果的平均值。

由图 8 可知,在 30 组数据上,所有模型的 r^2 值都集中在 0.8 到 1.0 之间,同时每个模型都有奇异值出现。模型 1 只有一个小于 0 的奇异值,除奇异值以外的其他 r^2 值中,最小值为 0.86,中位数为 0.97,主要集中在 0.92~0.98;模型 2 的奇异值个数是 5(部分重合),其中小于 0 的奇异值有 3 个,除奇异值以外的 r^2 值中,最小为 0.44,中位数为 0.90,主要集中在 0.79~0.96;模型 3 的奇异值个数为 3,其中小于 0 的奇异值有 2 个,最小值为 0.73,中位数为 0.94,主要集中在 0.90~0.98;模型 4 的奇异值个数为 2,其中一个小于 0,除奇异值以外的其他 r^2 值最小为 0.62,中位数为 0.94,主要集中在 0.87~0.97。从 r^2 值的分布可以看出,模型 1 相比其他模型,奇异值的个数更少,说明其泛化能力更好;在模型 1 中,另外 29 个非奇异的 r^2 值相对其他模型来说分布更加集中,说明模型 1 的稳定性更好,并能对 29 组数据进行较为准确的拟合。

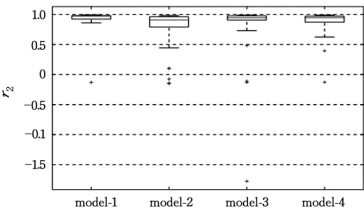


图 8 r^2 值箱型图
Fig. 8 Box-plot of r^2

如表 4 所列,结合所有评价标准的平均值和方差,模型 1 的 5 项评价标准值都是最优的,并且方差最小,这说明模型 1 的预测准确度、泛化能力以及模型的稳定性都是最优的。

表 4 5 项标准的平均值与方差
Table 4 Average value and variance of five standards

算法		MAE	MSE	RMSE	EVS	r^2
模型 1	AVE	0.060	0.030	0.100	0.900	0.890
	VER	0.007	0.008	0.021	0.031	0.041
模型 2	AVE	0.119	0.082	0.172	0.814	0.716
	VER	0.024	0.044	0.055	0.043	0.137
模型 3	AVE	0.110	0.090	0.151	0.850	0.741
	VER	0.040	0.098	0.071	0.057	0.311
模型 4	AVE	0.088	0.049	0.129	0.869	0.838
	VER	0.013	0.018	0.034	0.034	0.051

从以上实验结果可看出,本文提出的引入 Dropout 的深度双向 LSTM 网络对于股市变化的预测更加准确,误差更小,且模型更稳定。相较其他模型,所提模型的各项误差降低了 2%~5%, r^2 提高了 5%~10%。因此,在证券指数预测上,引入 Dropout 的深度双向 LSTM 网络能更合理地拟合证券指数走势,能够对未来证券指数进行预测。

结束语 本文提出了引入 Dropout 的深度双向 LSTM 模型,并将其应用于证券指数的预测任务中。该模型从前、后两个方向同时学习证券指数的变化特征,并运用深层的神经网络抽象特征,通过全连接网组合非线性特征,最后输出预测结果。在找到最佳时间窗口长度和 Dropout 添加位置的情况下,本文模型与其他文献提出的预测模型相比,各项误差降低了 2%~5%, r^2 提高了 5%~10%,较好地完成了预测任务;此外,在实验过程中发现 BatchNormalization 并不适用于本文模型。虽然范数正则化、Dropout、BatchNormalization 都能防止过拟合,但由于模型与数据特征的特殊性,导致不同的防过拟合策略带来的收益并不相同。

参 考 文 献

[1] DEGIANNAKIS S, FILIS G, HASSANI H. Forecasting global stock market implied volatility indices[J]. Journal of Empirical Finance, 2018, 46: 111-129.

[2] ZHANG W Y, ZHANG S X, ZHANG S, et al. A multi-factor and high-order stock forecast model based on Type-2 FTS using cuckoo search and self-adaptive harmony search[J]. Neurocomputing, 2017, 240: 13-24.

[3] NARENDRA BABU C, REDDY B E. Prediction of selected Indian stock using a partitioning-interpolation based ARIMA-GARCH model[J]. Applied Computing and Informatics, 2015, 11(2): 130-143.

[4] ROSAS-ROMERO R, DÍAZ-TORRES A, ETCHEVERRY G. Forecasting of stock return prices with sparse representation of financial time series over redundant dictionaries[J]. Expert Systems With Applications, 2016, 57: 37-48.

[5] CHEN Y T, TAO L, LI H B. The Prediction of Dow Jones Cbn China 600 Inde Based on “Most Updated” GM(1,1) Model and Grey Dynamic Neural Network[C]// 2010 Third International Conference on Information and Computing. Wuxi: IEEE, 2010: 56-59.

[6] HUANG K Y, JANE C J. A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories[J]. Expert Systems With Applications, 2008, 36(3): 5387-5392.

[7] WANG D X, LIU X, WANG M D. A DT-SVM Strategy for Stock Futures Prediction with Big Data[C]// Computational Science and Engineering. Sydney, NSW, Australia: IEEE, 2013: 1005-1012.

[8] CHEN Y J, HAO Y T. A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction[J]. Expert Systems with Applications, 2017, 80: 340-355.

[9] WANG P G. Research on Stock Price Prediction Based on BP Wavelet Neural Network with Mexico Hat Wavelet Basis[C]// Proceedings of 1st International Conference on Education, Economics and Management Research. Advances in Social Science, Education and Humanities Research (ICEEMR 2017). Singapore Management University, 2017, 95: 4.

[10] NAYAK S C, MISRA B B, BEHERA H S. Artificial chemical reaction optimization of neural networks for efficient prediction of stock market indices[J]. Ain Shams Engineering Journal, 2015, 8(3): 371-390.

[11] CHATZIS S P, SIAKOULIS V, PETROPOULOS A, et al. Forecasting stock market crisis events using deep and statistical machine learning techniques[J]. Expert Systems with Applications, 2018, 112: 353-371.

[12] GILES C L, KUHN G M, WILLIAMS R J. Dynamic recurrent neural networks: Theory and applications [J]. IEEE Trans. Neural Networks, 1994, 15: 153-156.

[13] SCHUSTER M, PALIWAL K K. Bidirectional Recurrent Neural Networks [J]. IEEE Transactions on Signal Processing, 1997, 45: 11.

[14] HOCHREITER S, SCHMIDHUBER J. Long Short-term Memory [J]. Neural computation, 1997, 9(8): 1735-1780.

[15] GRAVES A, SCHMIDHUBER J. Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures [J]. Neural Networks, 2005, 18(5): 602-610.

[16] SUN R Q. Research on the price trend forecasting model of us stock index based on LSTM neural network[D]. Capital University of Economics and Business, 2015.

[17] FISCHER T, KRAUSS C. Deep learning with long ? short-term memory networks for financial market predictions[J]. European Journal of Operational Research, 2018, 270(2): 654-669.

[18] SRIVASTAVA N, HINTON G, KRIZHEVSKY A. Dropout: A Simple Way to Prevent Neural Networks from Overfitting [J]. Journal of Machine Learning Research, 2014(15): 1929-1958.

[19] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8).

[20] HIRANSHA M, GOPALAKRISHNAN E A, VIJAY KRISHNA Menon, et al. NSE Stock Market Prediction Using Deep-Learning Models [J]. Procedia Computer Science, 2018, 132: 1351-1362.

[21] YETIS Y, KAPLAN H, JAMSHIDI M. Stock market prediction by using artificial neural network[C]// World Automation Congress (WAC). Waikoloa, HI, USA: IEEE, 2014: 718-722.

[22] REN J, WANG J H, WANG C M, et al. Stock Index Forecast Based on Regularized LSTM Model[J]. Computer Application and Software, 2018, 35(4): 44-48, 108.