

一种基于深度 LSTM 和注意力机制的金融数据预测方法



刘 翀 杜军平

北京邮电大学计算机学院智能通信软件与多媒体北京市重点实验室 北京 100876

(Alen123456@163.com)

摘 要 随着互联网的迅速发展,金融市场每日产生了大量在线金融数据,如每日的交易次数以及交易的总金额等。近年来金融市场数据的动态预测成为了研究热点。金融市场数据量大,输入序列较多,且会随着时间发生变化。针对这些问题,文中提出了基于深度 LSTM 和注意力机制的金融数据预测模型。首先,该模型能处理复杂的金融市场数据输入,主要是多序列的输入;其次,该模型使用深度 LSTM 网络对金融数据进行建模,解决了数据间长依赖的问题,并能学习到更加复杂的市场动态特征;最后,该模型引入了注意力机制,使得不同时间的数据对预测的重要程度不同,预测更加精准。在真实的金融大数据集上的实验表明,所提模型在动态预测领域具有准确性高、稳定性好的特点。

关键词: 金融预测;深层 LSTM;注意力机制;序列模型

中图法分类号 TP391

Financial Data Prediction Method Based on Deep LSTM and Attention Mechanism

LIU Chong and DU Jun-ping

Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract With the rapid development of the Internet, financial markets generate a large amount of online financial data every day, such as the number of daily transactions and the total amount of transactions. The dynamic prediction of financial market data has become a research hotspot in recent years. However, the financial market has a large amount of data, many input sequences, and changes over time. Aiming at solving these problems, this paper proposes a financial data prediction model based on deep LSTM and attention mechanism. First, the model can handle complex financial market data which are mainly multi-sequence data. Second, the model uses deep LSTM networks to model financial data, solves the problem of long dependence between data, and can learn more complex market dynamic characteristics. Finally, the model introduces the attention mechanism, which makes the data of different time have different importance to the prediction and make the prediction more accurate. Experiments on real large data sets show that the proposed model has the characteristics of high accuracy and good stability in the field of dynamic prediction.

Keywords Financial forecasting, Deep LSTM, Attention mechanism, Sequence model

改革开放以来,传统的金融企业和互联网技术逐步进行交叉集成^[1],例如,原始实体银行和在线银行上,新的互联网财务模型已得到扩展,如在线财务管理、银行电子帐户、手机银行、P2P 模型、第三方支付平台、众筹等。此外,我国曾提出了实现绿色、创新和强大互联网国家的概念,这使互联网金融进入了快速发展时期^[2]。然而,在互联网金融飞速发展的过程中出现了很多问题,如政策滞后性以及监管力度不足等,这些问题可能会侵犯交易各方的权利。而且,互联网金融具有较强的主观性,容易产生潜在风险,给企业、个人乃至整个行业造成了巨大损失,从而阻碍了互联网金融的持续健康发展^[3]。因此,合理追踪互联网金融数据,并在一定算法下实现

动态数据的预测,得出对互联网金融影响更大的风险因素,并提出相应的改进建议,对互联网金融的风险预警和平台管理等具有十分重要的意义。传统的序列预测方法主要是回归模型,如自回归模型 ARIMA 等,然而这类方法对于复杂的金融数据等拟合能力较差。同时,基于深度学习的方法也逐渐应用于序列预测中,最为著名的是长短期记忆网络 LSTM 和门控 GRU 方法,这些方法基于时序模型,然而对于序列的学习依然存在不充分、无法有效获取长期依赖的问题。因此,本文提出了一种基于深度 LSTM 和注意力机制的金融数据预测方法(Financial Data prediction Based on Deep LSTM and Attention Mechanism, DLSTM-A)。其使用深度的长短期记忆

到稿日期:2020-07-08 返修日期:2020-09-03 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金项目(61902037,61532006,61772083,61802028);广西科技重大专项(桂科 AA18118054)

This work was supported by the National Natural Science Foundation of China(61902037,61532006,61772083,61802028) and Science and Technology Major Project of Guangxi (Guike AA18118054).

通信作者:杜军平(junpingdu@126.com)

网络对市场的动态数据序列(如日交易量和次数等)进行建模,并使用注意力机制得到序列对不同时刻预测的影响因素,在一定程度上解决了金融数据长时间的依赖和影响问题,并利用深度网络对序列特征进行了进一步学习,使之能应用于复杂序列的学习。本文使用互联网金融平台得到的大数据对所提出的 DLSTM-A 算法进行实验和验证,结果表明本文方法对于互联网金融数据预测具有有效性和稳定性。

1 相关工作

互联网金融数据预测是近年来的研究热点。长短期记忆网络能对序列进行编码和解码,通过隐藏层参数在序列之间进行传递得到历史信息,并通过门控等方法解决长期依赖导致的梯度消失问题,同时注意力机制的引入也使得预测的关注点不再均衡,使得预测更有针对性。

1.1 长短期记忆网络

Greff 等^[4]针对语音识别、手写识别和和弦音乐建模任务,对 8 个 LSTM 变体进行了大规模分析,使用随机搜索分别优化了每个任务的所有 LSTM 变体的超参数,并使用 Variance 分析框架评估了不同部件的重要性,结果表明这些变体都不能显著改善标准 LSTM 体系结构,而遗忘门和输出激活功能是其最关键的组件。Merity 等^[5]考虑了词级语言建模的特定问题,研究了规范化和优化基于 LSTM 的模型的策略,并提出了权重下降的 LSTM,对隐藏的权重使用 Drop-Connect 作为递归正则化的形式。Zhao 等^[6]针对短期交通预测问题,提出了一种基于 LSTM 网络的流量预测模型,并通过由许多存储单元组成的二维网络来考虑交通系统中的时空相关性。Karim 等^[7]使用长短期记忆递归神经网络子模块增强全卷积网络,提出了长短期记忆全连接卷积网络(Long Short Term Memory Fully Convolutional Network, LSTM-FCN)来进行时间序列分类,并显著增强了全卷积网络的性能,其对数据集只需进行最少的预处理,并且能够可视化决策过程。Fu 等^[8]使用长短期记忆和门控递归单元神经网络方法来预测短期交通流量,并且通过实验证明了基于递归神经网络的深度学习方法的性能优于自回归综合移动平均值模型。

由于 LSTM 网络能很好地学习序列中时间的相关性,对时序序列预测有较高适用性,因此本文采用 LSTM 网络作为基础组件。然而现有方法对金融数据的特征提取仍有不完全的问题,因此本文使用深度网络在一定程度上缓解此类问题。

1.2 注意力机制

Choi 等^[9]提出了一种细粒度注意力机制,使得翻译中上下文向量的每个维度将获得单独的注意力得分。在以 En-De 和 En-Fi 翻译为任务的实验中,细粒度的注意力方法提高了 BLEU 分数的翻译质量,并揭示了细粒度的注意机制如何利用上下文向量的相关性。Tilk 等^[10]提出了一种具有注意力机制的双向递归神经网络模型,用于未分段文本中的标点恢复,该模型可以在方向和直接注意两个方面利用较长的上下文,从而使其在很大程度上超越英语和爱沙尼亚数据集上的现有技术水平。Wang 等^[11]提出了 CLVSA 模型,这是一种由随机递归网络、序列到序列结构、自我注意、相互注意机制,以及卷积 LSTM 单元组成的混合模型,该模型用于捕获原始金融交易数据中的潜在变化特征。Chen 等^[12]为了提高金融

时间序列预测的准确性,提出了一种由经验模式分解和基于注意力的长期短期记忆网络组成的混合模型。该模型可以有效地将金融时间序列分解为多个级别的固有模式函数,并将这些函数输入 LSTM-ATTE 中进行预测。其通过股票市场指数的线性回归分析验证了所提模型的预测性能。Jiang 等^[13]提出了一种神经网络架构 GABi-LSTM 来解决来自金融微博和新闻的细粒度金融目标相关情绪分析。该模型采用门控机制自适应地集成字符级和词级嵌入来进行单词表示,然后提出了一种基于注意力的 Bi-LSTM 组件,将与目标相关的信息嵌入到句子表示中,最后使用线性回归层来预测情感得分。

注意力机制能有效捕捉数据的动态变化特征,使得相关性分析能更加准确,而时间序列通过注意力机制后能快速捕捉特定范围数据的动态变化特征,因此本文将注意力机制引入时序数据预测中。

2 DLSTM-A 方法

2.1 DLSTM-A 方法的模型框架

本文提出的 DLSTM-A 方法由 1 层输入层对所有时刻的序列进行输入,并依次纵向通过 3 层 LSTM 堆叠构建循环神经网络进行输出,其中还包括同层 LSTM 的隐藏状态传递。本文方法在最后一层设计注意力机制,根据不同时刻对结果求得的重要性来计算所有权值的相加量,然后根据权值得到最终的输出。整个框架如图 1 所示。

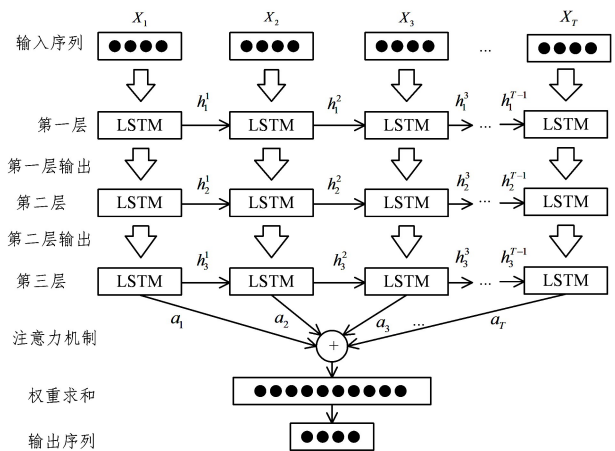


图 1 DLSTM-A 方法的总体流程图

Fig. 1 General flowchart of proposed DLSTM-A method

2.2 LSTM 构件介绍

首先,输入层中的输入序列由 $X = (x_1, x_2, \dots, x_t)$ 构成,序列长度为 T , x_i 代表时间 i 模型的输入。 x_i 是由市场交易量以及交易次数所构成的嵌入向量,由不同类别的项目构成,如文娱类或者科技类的项目,按照一定顺序进行嵌入操作得到。

接着是第一层的 LSTM,第一层之间主要传递隐藏状态 h_i^1 ,其中 i 代表第 i 个时刻的序列。在这一层之间共享第一层 LSTM 网络的参数 W_1 和 b_1 。然后是第二层的 LSTM,与第一层类似,第二层的 LSTM 接收第一层得到的输出激活值,用 a_i^1 表示, i 代表第 i 个时刻的序列。第三层的 LSTM 与第二层类似。

同一层 LSTM 之间除了需要传递隐藏状态 a_m^m 外,还需

要传递记忆状态 c'_m 。而记忆状态组成之一就是记忆状态的候选值,某时刻 t 的第 m 层 LSTM 的记忆状态候选值 \tilde{c}'_m 如式(1)所示:

$$\tilde{c}'_m = \tanh(\mathbf{W}_m^c [a_m^{t-1}, a'_{m-1}] + b_m^c) \quad (1)$$

其中, a_m^{t-1} 表示从上个时刻同一层得到的激活值, a'_{m-1} 表示由同一时刻上一层 LSTM 得到的激活值。即 \tilde{c}'_m 接收两个输入,为使模型统一,在第一层输入加入由 0 向量组成的激活值 a'_0 ,并且为了计算方便对第 m 层的 LSTM 参数 \mathbf{W}_m^c 进行堆叠处理,直接与 a_m^{t-1} 及 a'_{m-1} 拼接构成的向量进行乘积操作。

同时,对于第 m 层 LSTM 而言,还需要记录 3 个门控操作,即记忆门 Γ_m^c 和遗忘门 Γ_m^f 以及输出门 Γ_m^o ,这 3 个门控决定了该 LSTM 单元的真实记忆值 c'_m 以及激活值 a'_m 。第 m 层 LSTM 的记忆门 Γ_m^c 和遗忘门 Γ_m^f 用式(2)和式(3)求得。

$$\Gamma_m^c = \sigma(\mathbf{W}_m^c [a_m^{t-1}, a'_{m-1}] + b_m^c) \quad (2)$$

$$\Gamma_m^f = \sigma(\mathbf{W}_m^f [a_m^{t-1}, a'_{m-1}] + b_m^f) \quad (3)$$

接着可以求得第 m 层第 t 时刻的记忆值 c'_m ,而每个时刻 t 中 c'_m 有两种组成方式,一种是由当前时刻产生的候选值,另一种由前一时刻的记忆值构成,而这两者分别通过记忆门 Γ_m^c 和遗忘门 Γ_m^f 控制,计算式如式(4)所示:

$$c'_m = \Gamma_m^c \tilde{c}'_m + \Gamma_m^f c_m^{t-1} \quad (4)$$

可以根据输出门 Γ_m^o 求得第 m 层第 t 时刻的隐藏状态 a'_m ,计算式分别如式(5)和式(6)所示:

$$\Gamma_m^o = \sigma(\mathbf{W}_m^o [a_m^{t-1}, a'_{m-1}] + b_m^o) \quad (5)$$

$$a'_m = \Gamma_m^o c'_m \quad (6)$$

最后第 t 时刻的输出序列 y^t 由最后一层激活值得到,计算式如式(7)所示:

$$y^t = \text{softmax}(a'_M) \quad (7)$$

其中, M 为 LSTM 的总层数,本文的 DLSTM-A 模型中 M 取值为 3。

2.3 注意力机制的引入

当输入序列通过三层 LSTM 输出后,可以得到每个时刻互联网金融市场的状态,同时也能得到最后输出的隐藏状态和记忆值。从经验考虑,不同时刻的信息对当前时刻的状态预测所产生的比重不同。太过久远的信息对现在时刻的状态影响较小,而相近的时间影响较大。因此,模型需要对不同时刻产生的信息的重要性进行评估,从而引入了注意力机制。

本文将第三层输出的隐藏层的每个时刻状态作为隐藏状态集合 $A = (a_M^t, a_M^{t-1}, \dots, a_M^1)$,其中 a_M^t 表示第 t 时刻最后一层 LSTM 的隐藏层状态,并将 A 作为注意力机制的输入,可以计算出 t 时刻的输出序列需要关注 t' 时刻隐藏状态的程度 e'_t 。 e'_t 可以通过搭建一层简单的神经网络来计算,网络参数是 \mathbf{W}_h 和 b_h ,如式(8)所示:

$$e'_t = \tanh(\mathbf{W}_h [a_M^{t-1}, a_M^t] + b_h) \quad (8)$$

其中, a_M^{t-1} 表示上一个时刻序列的隐藏状态, a_M^t 表示 t' 时刻的隐藏状态。同样采取拼接的方式将 a_M^{t-1} 和 a_M^t 进行拼接,以方便矩阵计算。

得到 e'_t 后就可以计算出 t 时刻的输出序列关注 t' 时刻隐藏状态的权重 α'_t ,计算式如式(9)所示,即求 t' 时刻的关注程度占所有时刻的关注程度的比重。

$$\alpha'_t = \frac{\exp(e'_t)}{\sum_{i=1}^T \exp(e'_i)} \quad (9)$$

因此,求得 t 时刻的总权重因子 r_t 如式(10)所示:

$$r_t = \sum_{i=1}^T \alpha'_i a'_M \quad (10)$$

最终求得 t 时刻的输出值如式(11)所示:

$$y_t = \text{LSTM}(r_t, a_{t-1}, c_{t-1}) \quad (11)$$

2.4 模型的正则化处理

由于采用深度 LSTM 网络,训练参数较多,如果不做任何处理会导致过拟合,使得模型的泛化能力降低。因此,本文使用随机 Dropout 方法来关闭深度 LSTM 网络中的某些神经元,这种方法可以防止神经元的共同适应,每个神经元都有一定几率被随机抹去。每次训练时,整个网络都只有一个子部分受到训练,而其他部分保持不变。此外,本文将 Dropout 应用于 LSTM 的循环连接,以忽略某些序列带来的影响。

此外,本文对所有的网络参数都加上了 L2 正则化,在训练时将网络参数的值的平方添加到损失函数中防止网络参数过大。同时,本文在训练时会随时记录和验证损失函数的值,当满足某一条件后提前结束训练,防止过拟合。

3 实验结果与分析

实验使用采集自 Prosper 借贷平台的数据,收集并分析了从 2013 年 1 月 1 日到 2017 年 12 月 31 日 5 年期间每一天借贷人的投资记录,共计 1826 天的数据。数据包含每日的交易金额总数、每日交易次数以及每日违约总次数。实验将这 3 种记录作为每日的序列输入模型,并以 10 天为时间窗口根据近 10 天的历史数据预测未来 10 天的数据。

3.1 实验数据集

Prosper 提供的数据集主要包含 1826 天的日交易金额、日交易次数和日违约次数。实验将 70% 的数据作为训练集,30% 的数据作为测试集对性能进行测试。由于数据集中存在少量序列数据缺失的问题,为了保证序列在时间上的完整性,对缺失的数据采用三次样条插值,以尽可能准确地还原缺失的原始数据,从而能用 DLSTM-A 模型进行处理。三次样条插值代价低、计算速度快,具有较好的收敛性以及低阶可导性质,利于模型训练,同时其还原数据效果比二次和线性插值好。同时对单天的输入序列进行嵌入操作,把日交易金额、日交易次数、日违约次数 3 部分信息嵌入成一维向量进行处理,使模型能识别输入信息,并采用归一化处理,将值压缩到 $[0, 1]$ 区间,避免评价范围不同导致评估不准确的问题。

3.2 不同方法的实验对比

本文主要使用时序预测中普遍使用的均方根误差 (RMSE) 以及平均绝对误差 (MAE) 作为模型的评价指标。它们是对预测值和真实值的实际差异大小进行描述,当这两个值越低,说明预测值与真实值越接近,结果也就越好。RMSE 和 MAE 的计算式如下:

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t' - y_t)^2} \quad (12)$$

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |y_t' - y_t| \quad (13)$$

其中, y_t' 表示第 t 天的预测值, y_t 表示第 t 天的真实值。

首先,以 10 天的时间窗口为观测数据的窗口,预测未来 1—10 天的交易总金额,并以 RMSE 和 MAE 为指标衡量预

测值与真实值的误差, 所得结果分别如表 1 和表 2 所列。可以看出, 提出的 DLSTM-A 方法在 Prosper 数据集上预测未

来 10 天交易总金额的效应在 RMSE 和 MAE 指标上优于其他方法。

表 1 序列预测方法对未来 10 天总金额预测的 RMSE 值

Table 1 RMSE value of serial prediction method for total transaction amount forecast in next 10 days

	第 1 天	第 2 天	第 3 天	第 4 天	第 5 天	第 6 天	第 7 天	第 8 天	第 9 天	第 10 天
ARIMA	0.0636	0.0640	0.0633	0.0646	0.0655	0.0664	0.0664	0.0680	0.0688	0.0701
GRU	0.0612	0.0631	0.0621	0.0636	0.0646	0.0638	0.0642	0.0649	0.0665	0.0687
LSTM	0.0604	0.0623	0.0612	0.0616	0.0637	0.0633	0.0646	0.0643	0.0652	0.0665
MALSTM	0.0591	0.0584	0.0592	0.0599	0.0600	0.0600	0.0621	0.0616	0.0618	0.0629
OURS	0.0583	0.0583	0.0590	0.0588	0.0590	0.0598	0.0613	0.0605	0.0613	0.0604

表 2 序列预测方法对未来 10 天交易总金额预测的 MAE 值

Table 2 MAE value of serial prediction method for total transaction amount forecast in next 10 days

	第 1 天	第 2 天	第 3 天	第 4 天	第 5 天	第 6 天	第 7 天	第 8 天	第 9 天	第 10 天
ARIMA	0.0454	0.0461	0.0467	0.0463	0.0495	0.0481	0.0498	0.0507	0.0525	0.0531
GRU	0.0435	0.0449	0.0454	0.0457	0.0481	0.0474	0.0491	0.0487	0.0509	0.0522
LSTM	0.0438	0.0433	0.0445	0.0446	0.0454	0.0473	0.0480	0.0476	0.0497	0.0510
MALSTM	0.0422	0.0423	0.0427	0.0427	0.0442	0.0439	0.0451	0.0458	0.0443	0.0458
OURS	0.0414	0.0410	0.0411	0.0416	0.0426	0.0439	0.0444	0.0436	0.0450	0.0454

首先, 对于 ARIMA 方法^[14]而言, 因为它仅使用了回归模型, 并不考虑其他相关变量带来的时间变化, 同时也没有考虑历史数据带来的影响, 因此效果较差, 误差远高于其他方法。其次是基本的 GRU^[15]和 LSTM^[16]模型, 这两种模型都是基本的 RNN 网络, 其中 GRU 网络仅使用遗忘门和输出门, 相对于 LSTM 模型少了一个计算步骤, 因此在预测性能上稍弱于 LSTM 模型, 但是由于传递了隐藏状态, 考虑了时间对预测结果的影响, 因此预测性能优于一般的拟合方法。而 LSTM 模型相对于 MALSTM 和本文模型少了注意力机制, 导致忽略了序列某时刻对其他时刻的影响是不同的事实, 训练时对关键特征的抓取效果稍弱, 因此最终预测偏差较加入了注意力机制的方法更大。MALSTM 方法^[17]主要考虑了多序列输入以及注

意力机制, 并在序列预测上取得了较好的效果。而本文方法由于使用了深度的 LSTM 网络, 并使用了正则化手段避免模型过拟合, 因此预测性能总体上稍优于 MALSTM 方法。

总的来看, 基于卷积神经网络的方法比使用拟合的序列预测方法在 MAE 指标上至少要提高 5% 左右, 特别是对长时间的预测, 效果更加明显。具有注意力机制的方法和深度的网络所预测的结果效果更好。

然后, 以 10 天的时间窗口为观测数据的窗口, 预测未来 1—10 天的交易总交易次数, 同样以 RMSE 和 MAE 为指标衡量预测值与真实值的误差, 所得到的结果如表 3 和表 4 所列。最后预测 10 天后违约次数的变化, 同样以 RMSE 和 MAE 为评价指标, 得到的结果如表 5 和表 6 所列。

表 3 序列预测方法对未来 10 天总交易次数预测的 RMSE 值

Table 3 RMSE value of serial prediction method for total transaction count forecast in next 10 days

	第 1 天	第 2 天	第 3 天	第 4 天	第 5 天	第 6 天	第 7 天	第 8 天	第 9 天	第 10 天
ARIMA	0.0829	0.0812	0.0809	0.0779	0.0764	0.0765	0.0788	0.0805	0.0812	0.0837
GRU	0.0803	0.0797	0.0805	0.0780	0.0774	0.0757	0.0775	0.0777	0.0803	0.0805
LSTM	0.0819	0.0787	0.0784	0.0776	0.0771	0.0764	0.0759	0.0799	0.0807	0.0807
MALSTM	0.0778	0.0778	0.0763	0.0758	0.0726	0.0736	0.0766	0.0765	0.0755	0.0763
OURS	0.0721	0.0691	0.0685	0.0655	0.0674	0.0677	0.0668	0.0669	0.0688	0.0688

表 4 序列预测方法对未来 10 天总交易次数预测的 MAE 值

Table 4 MAE value of serial prediction method for total transaction count forecast in next 10 days

	第 1 天	第 2 天	第 3 天	第 4 天	第 5 天	第 6 天	第 7 天	第 8 天	第 9 天	第 10 天
ARIMA	0.0583	0.0570	0.0557	0.0553	0.0540	0.0542	0.0555	0.0562	0.0570	0.0580
GRU	0.0588	0.0573	0.0558	0.0553	0.0537	0.0542	0.0556	0.0561	0.0582	0.0586
LSTM	0.0587	0.0576	0.0553	0.0543	0.0528	0.0533	0.0546	0.0569	0.0583	0.0589
MALSTM	0.0544	0.0544	0.0530	0.0516	0.0516	0.0518	0.0530	0.0529	0.0539	0.0538
OURS	0.0543	0.0537	0.0522	0.0512	0.0501	0.0512	0.0523	0.0514	0.0529	0.0528

表 5 序列预测方法对未来 10 天违约次数预测的 RMSE 值

Table 5 RMSE value of serial prediction method for number of defaults forecast in next 10 days

	第 1 天	第 2 天	第 3 天	第 4 天	第 5 天	第 6 天	第 7 天	第 8 天	第 9 天	第 10 天
ARIMA	0.0240	0.0236	0.0233	0.0247	0.0253	0.0269	0.0264	0.0283	0.0284	0.0305
GRU	0.0207	0.0234	0.0223	0.0234	0.0244	0.0239	0.0246	0.0244	0.0268	0.0287
LSTM	0.0203	0.0220	0.0207	0.0219	0.0239	0.0235	0.0248	0.0248	0.0252	0.0264
MALSTM	0.0195	0.0185	0.0189	0.0201	0.0198	0.0204	0.0220	0.0220	0.0222	0.0231
OURS	0.0188	0.0188	0.0193	0.0190	0.0188	0.0193	0.0208	0.0206	0.0217	0.0206

表 6 序列预测方法对未来 10 天违约次数预测的 MAE 值

Table 6 MAE value of serial prediction method for number of defaults forecast in next 10 days

	第 1 天	第 2 天	第 3 天	第 4 天	第 5 天	第 6 天	第 7 天	第 8 天	第 9 天	第 10 天
ARIMA	0.0254	0.0241	0.0236	0.0249	0.0236	0.0279	0.0254	0.0209	0.0293	0.0304
GRU	0.0221	0.0222	0.0232	0.0194	0.0212	0.0178	0.0251	0.0230	0.0299	0.0296
LSTM	0.0187	0.0135	0.0218	0.0157	0.0190	0.0262	0.0241	0.0203	0.0296	0.0264
MALSTM	0.0149	0.0204	0.0161	0.0138	0.0242	0.0175	0.0186	0.0164	0.0204	0.0230
OURS	0.0133	0.0200	0.0142	0.0119	0.0152	0.0188	0.0240	0.0138	0.0193	0.0165

下面根据以上的数据分析源数据变化程度对模型的稳定性的影响。用户每日的交易次数属于变化较大的数据,因此预测难度高于预测的总金额和违约次数。从实验结果来看,使用了循环神经网络和注意力机制的模型效果较好,预测的准确性随着时间的增加不会显著降低。同样的情况适用于每日违约次数的预测。由于违约次数在短时间内较为稳定,难以快速变化,因此预测难度小于预测交易金额和次数, RMSE 和 MAE 值随着时间推移增长较为缓慢。在日期较远时交易次数预测的 RMSE 和 MAE 值出现了小幅下降,推测原因是交易次数的平均 RMSE 和 MAE 值都较高,整体波动范围较大,短时间内不能体现整体趋势,后几天恰好处于波动的下降区间,整体上升趋势需要在更长远的时间才能得以体现。

总体来说,通过对用户交易金额、交易次数以及违约次数的预测,最终发现使用注意力机制的模型效果显著优于其他方法,说明了本文提出的 DLSTM-A 模型的优越性。

3.3 LSTM-A 方法的结构调优

DLSTM-A 方法在传统的 RNN 网络的基础上主要引入了 LSTM 这类新的 RNN 网络,并在传统单层的网络上加入了深层网络来提高模型的学习能力,并引入了注意力机制来学习更加有效的特征。本节通过对比方法来对 DLSTM-A 模型的各个部分进行调优。

采用控制变量的方法进行实验,即在原本 DLSTM-A 模型的基础上仅改变某一部分,来对比改变后的模型和原本模型的效果。此时采用 RMSE 值作为评价指标,以 10 天为观察窗口,对比这两个模型对未来 1-10 天的总交易金额、交易次数以及违约次数的预测情况。

首先讨论 RNN 衍生模型的选择,本文选择使用 LSTM 模型作为基础。而另外一种常见的衍生模型是 GRU 模型,本文仅改变神经网络的构造,来得到预测结果的 RMSE 值,结果如图 2 所示。

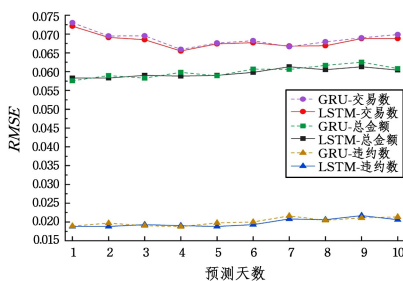


图 2 RNN 框架选择对 RMSE 值的影响

Fig. 2 Impact of RNN framework selection on RMSE values

从图 2 可以看出,在交易总次数的预测结果上,使用深度 LSTM 网络的性能明显优于 GRU,在总金额的预测上 LSTM 网络稍高于 GRU,而在一些稍简单的任务中,尤其是违约次

数的评估上, LSTM 和 GRU 网络没有明显的优劣。因此, LSTM 网络总体上稍优于 GRU 网络。在训练时间上 GRU 网络由于少一个控制门,正向传播和反向传播速度稍快,训练时间较短。因此,实际使用情况可以根据对精度或者时间的要求来选择 LSTM 网络或是 GRU 网络。

下面讨论网络深度对 DLSTM-A 模型的影响。本文采用三层 LSTM 堆叠的方式来构建深度网络,因此将其与两层的 LSTM 网络和单层的 LSTM 网络做对比,最终得到的结果如图 3 所示。可以得出,总体上网络的深度越深,预测效果就越好。但是从提升的幅度看,从单层 LSTM 网络变成两层 LSTM 网络的效果提升明显,而从两层 LSTM 网络到三层 LSTM 网络的效果提升不明显。因为单层 LSTM 网络对于序列的特征学习不充分,使用深层网络能学习到深度特征,从而加大了预测准确率,三层的深度网络容易受到过拟合的干扰,导致效果提升达到瓶颈,同时三层的 LSTM 训练速度较慢,不适合大规模训练。当网络层数增加时,训练时间呈几何倍数增长。受限于硬件条件,本文不进行更高层网络的对比实验。

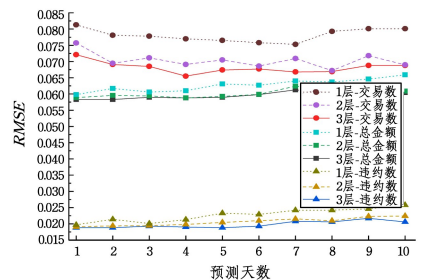


图 3 网络深度对 RMSE 值的影响

Fig. 3 Impact of network depth on RMSE values

最后讨论注意力机制对 DLSTM-A 模型的影响。由于使用了注意力机制,模型能在某时刻对之前的序列进行重点关注,从而使学习效率加快。引入和不引入注意力机制的对比实验如图 4 所示。

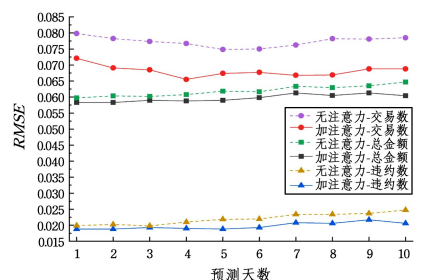


图 4 注意力机制对 RMSE 值的影响

Fig. 4 Impact of attention mechanism on RMSE values

从图 4 可以得出,引入注意力机制后对序列预测结果的

准确性有很大的提升,主要体现在预测的稳定性上,对未来的预测误差没有随着时间增加而快速上升。基于注意力机制能更好解决长序列的依赖问题,因此在进行信息处理时更加高效和准确。

结束语 本文对互联网金融大数据的动态序列预测方向进行研究,提出了基于深度 LSTM 网络和注意力机制的金融数据预测方法,对金融平台的日交易总金额、日交易总次数和违约次数做基本的预测,用于评估当前金融市场的风险值。本文方法基于深度 LSTM 网络,能学习到序列的深度特征,并对序列的隐藏状态和记忆状态进行同层以及跨层的传递,显著提升了模型的学习能力;同时引入了注意力机制,将输出层的不同时刻对当前时刻预测的结果的影响力做出了区分,使得网络能对需要关注的序列进行深入学习,提升了模型学习的效率,并在一定程度上解决了长依赖问题。本文还使用了 L2 正则项和随机 Dropout 手段来减轻深度网络带来的过拟合问题。最后,根据采集的实际金融市场数据进行实验,实验结果表明该模型在准确性和稳定性上高于其他对比算法;对模型内部构成进行调优和实验,结果表明引入注意力机制能显著提升模型的性能,使用 LSTM 网络和增加 LSTM 网络深度能在一定程度上改善模型的性能,但是可能会花费更多的计算资源和时间。因此,将来的研究可以从两个方面展开,一是研究模型训练的加速方案,二是使用更多种类的金融数据,如将金融市场相关的文本特征作为观测的输入序列,并观察对实验结果的影响。

参 考 文 献

- [1] ZHANG J, SUN Q. Research on Financing Cost of Small and Medium-Sized Enterprises by Internet Finance[J]. *Open Journal of Social Sciences*, 2017, 5(11):95.
- [2] LIN Y H, CHEN C F. Research on Enterprise Financial Risk Evaluation Based on Association Rules[J]. *Friends of Accounting*, 2017(1):32-35.
- [3] LIU J X, JIA X Y. A Multi-label Classification Algorithm Based on Association Rules Mining[J]. *Journal of Software*, 2017, 28(11):2865-2878.
- [4] GREFF K, SRIVASTAVA R K, KOUTNÍK J, et al. LSTM: A search space odyssey[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, 28(10):2222-2232.
- [5] MERITY S, KESKAR N S, SOCHER R. Regularizing and optimizing LSTM language models[J]. *arXiv*:1708.02182.
- [6] ZHAO Z, CHEN W, WU X, et al. LSTM network: a deep learning approach for short-term traffic forecast[J]. *IET Intelligent Transport Systems*, 2017, 11(2):68-75.
- [7] KARIM F, MAJUMDAR S, DARABI H, et al. LSTM fully convolutional networks for time series classification[J]. *IEEE Access*, 2017, 6:1662-1669.
- [8] FU R, ZHANG Z, LI L. Using LSTM and GRU neural network methods for traffic flow prediction[C]// 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE, 2016:324-328.
- [9] CHOI H, CHO K, BENGIO Y. Fine-grained attention mechanism for neural machine translation[J]. *Neurocomputing*, 2018, 284:171-176.
- [10] TILK O, ALUMÄE T. Bidirectional Recurrent Neural Network with Attention Mechanism for Punctuation Restoration[C]// *Interspeech*. 2016:3047-3051.
- [11] WANG J, SUN T, LIU B, et al. CLVSA: A convolutional LSTM based variational sequence-to-sequence model with attention for predicting trends of financial markets[C]// *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. AAAI Press, 2019:3705-3711.
- [12] CHEN L, CHI Y, GUAN Y, et al. A Hybrid Attention-Based EMD-LSTM Model for Financial Time Series Prediction[C]// 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE, 2019:113-118.
- [13] JIANG M, WANG J, LAN M, et al. An effective gated and attention-based neural network model for fine-grained financial target-dependent sentiment analysis[C]// *International Conference on Knowledge Science, Engineering and Management*. Springer, Cham, 2017:42-54.
- [14] CONTRERAS J, ESPINOLA R, NOGALES F J, et al. ARIMA models to predict next-day electricity prices[J]. *IEEE Transactions on Power Systems*, 2003, 18(3):1014-1020.
- [15] GAO Y, GLOWACKA D. Deep gate recurrent neural network [C]// *Asian Conference on Machine Learning*. 2016:350-365.
- [16] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8):1735-1780.
- [17] ZHAO H K, WU L K, LI H, et al. Predicting the Dynamics in Internet Finance Based on Deep Neural Network Structure[J]. *Journal of Computer Research and Development*, 2019, 56(8):1621-1631.



LIU Chong, born in 1995, postgraduate, is a member of China Computer Federation. His main research interests include nature language processing, computer vision and deep learning.



DU Jun-ping, born in 1963, Ph.D, professor, is a fellow of China Computer Federation and CAAI. Her main research interests include artificial intelligence, image processing and pattern recognition.