

金融时间序列数据预测: 文献回顾与展望

闫洪举^{1 2}

(1. 中国农业银行 博士后科研工作站 北京 100005; 2. 北京大学 经济学院 北京 100871)

摘要: 金融时间序列数据具有非线性、非平稳、高噪声等复杂特征,且随着移动互联网、人工智能的快速发展,海量结构化与非结构化数据不断产生,数据间的关联模式日益复杂。在此背景之下,构建科学合理的金融时间序列数据预测模型,充分挖掘金融时间序列数据隐含的重要信息至关重要。为此,在梳理金融时间序列数据预测的计量经济学方法、机器学习算法的基础上,着重分析深度学习应用于金融时间序列数据预测的理论基础与实证应用的相关文献,以期为大数据与人工智能背景下的金融时间序列数据预测以及多学科交叉融合研究提供相关借鉴。

关键词: 时间序列数据; 计量经济学; 机器学习; 深度学习

中图分类号: F832.29 **文献标识码:** A **文章编号:** 2095-0098(2021)03-0033-09

一、引言

金融市场是一个复杂的、进化的、非线性的动态系统(Abu - Mostafa 和 Atiya, 1996)^[1],金融时间序列数据通常具有非平稳性、非结构化、高噪声、不确定性大、关系复杂等特点(Hall, 1994)^[2],金融时间序列数据预测的理论方法与实际应用一直是学者关注的焦点。学者提出了诸如自回归模型、移动平均模型、广义自回归条件异方差类模型等经典的平稳时间序列数据模型,构建了诸如基于数据驱动建模的向量自回归等非平稳时间序列模型,将诸如模糊预测方法、灰色预测理论、演化算法、机器学习等人工智能算法应用于金融时间序列数据预测。然而,随着移动互联网、大数据等技术的快速发展,海量金融数据不断产生,数据间的关联模式日益复杂,诸如 ARIMA、VAR、模糊预测方法、灰色预测理论等金融时间序列数据预测模型,难以实现对大数据、人工智能背景下的金融时间序列数据预测。而深度学习具有较强的特征提取能力,在处理复杂数据方面极具优势,近年来在诸如语音处理、图像识别等领域已取得巨大成功。将深度学习应用于金融时间序列数据预测,不仅有助于充分挖掘时间序列数据的重要信息,提高对金融市场的认知能力,而且有助于实现多学科的交叉融合研究。

基于此,在梳理金融时间序列数据预测的计量经济学方法、机器学习算法基础之上,进一步梳理深度学习应用于金融时间序列数据预测的相关研究,分析其理论基础与实证应用,剖析深度学习应用于金融时间序列数据预测的优势与已有研究的不足,并提出其未来在金融时间序列数据预测领域可能的研究方向。

二、基于计量经济学模型的金融时间序列数据预测

Yule(1927) 构建了用以预测太阳黑子数的自回归(AR) 模型^[3],这也成为计量经济分析中较早的时间序列分析模型。其基本原理是将被解释变量的滞后项作为解释变量,对于 p 阶自回归模型 $AR(p)$ 的定义如下:

$$y_t = c + \sum_{i=1}^p \alpha_i y_{t-i} + \varepsilon_t \quad (1)$$

收稿日期: 2020-04-08

作者简介: 闫洪举(1987 -), 山东乐陵人, 博士, 研究方向为金融经济学。

其中 $\alpha_i (i=1, 2, \dots, p)$ 为参数, ε_t 为白噪声序列。

其后,学者进一步提出移动平均(MA)模型用以分析时间序列数据。与AR模型略有不同,MA模型利用白噪声序列的滞后项来预测被解释变量。对于 q 阶移动平均模型 $MA(q)$ 的表达式为:

$$y_t = \mu + \sum_{i=0}^q \theta_i \varepsilon_{t-i} \quad (2)$$

AR模型与MA模型并无本质区别,二者在一定条件下可以相互转化。Box和Jenkins将二者结合,构建更一般的包含 p 阶自回归模型和 q 阶移动平均模型的移动平均模型(ARMA)来描述时间序列数据。ARMA(p, q)的表达式为:

$$y_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \mu_t + \sum_{i=1}^q \theta_i \mu_{t-i} \quad (3)$$

AR、MA、ARMA等模型适用于平稳时间序列数据,在处理非平稳时间序列数据方面存在一定局限性。然而,已有研究发现大多数金融时间序列数据通常具有非平稳特征(Mikosch & Stărică 2004)^[4]。为此,Box & Jenkins进一步提出单整移动平均(ARIMA)模型以处理非平稳性时间序列数据。ARIMA模型建模较为简单,且能有效处理非平稳金融时间序列数据,故成为金融时间序列数据预测领域一种常用的计量经济模型。例如,池启水(2007)采用Box-Jenkins的ARIMA预测中国石油消费增长量^[5],认为与结构性因果模型、AR、MA、ARMA等模型相比,ARIMA更适用于石油消费量的非平稳时间序列数据,具有更好的预测效果。Ediger & Akar(2007)采用ARIMA模型和季节调整的ARIMA模型预测能源需求量^[6],其结果表明采用ARIMA模型预测能源需求更具可靠性。

Box-Jenkins的ARIMA模型,其基本原理是用差分剔除时间序列数据中的趋势项和周期项,然后对平稳项用ARMA模型进行分析与预测,这种方法最大的缺点是丢掉了金融时间序列数据中最重要的信息,即趋势项和周期项(谷政和张维 2013)^[7],也无法捕捉金融时间序列数据的非线性模式(Pai & Lin 2004)^[8]。

此外,Engle(1982)提出了时间序列数据分析的自回归条件异方差模型(ARCH)^[9],Bollerslev(1986)在ARCH模型的基础之上^[10],进一步对误差的方差建立自回归模型,构建了广义自回归条件异方差模型(GARCH)。其后,GARCH类模型在金融时间序列数据分析中有着较为广泛的应用,惠晓峰等(2003)基于GARCH模型预测人民币对美元的汇率^[11],其实证结果一方面说明了人民币汇率存在的显著的自相关和异方差特性,同时证实了GARCH模型在预测人民币汇率方面具有很好的预测效果。刘国旗(2000)采用二次GARCH模型(QGARCH)和Glosten提出的GSR等非线性模型预测我国股市波动性^[12],基于上证指数的实证结果发现,相较于EGRACH、GSR、标准GARCH等GARCH类模型,QGARCH模型在预测上证指数波动性方面预测效果最好。于志军和杨善林(2013)构建了误差校正与GARCH模型相结合的金融时间序列数据预测模型^[13],即采用GARCH模型对时间序列数据进行初步预测,然后利用回归分析预测GARCH模型的残差序列,最终集成得到时间序列数据的预测值,基于上证指数的实证预测发现,该模型预测精度明显优于未校正的预测精度。

综上所述,金融时间序列数据预测的计量经济学方法大都依赖于经济理论建模,具有明显的经济意义,建模思路较为简单,可以实现对小样本数据的预测分析,在金融时间序列数据预测以及其他金融经济问题分析中也有着广泛应用。然而,传统时间序列预测方法诸如ARMA、GARCH等计量经济学模型通过假定具体的模型来刻画金融时间序列数据。但复杂且含有噪声的真实时间序列数据无法通过含有参数的解析方程来反映,因为时间序列数据的动力学方程或者太过复杂,或者未知(Taylor 2009)^[14]。此外,金融时间序列数据的序列依赖特征是金融时间序列数据预测中不可或缺的一部分,ARMA、GARCH等模型虽然可以考虑金融时间序列数据的序列相关特征,但是序列相关长度的确定或者需要通过复杂的计量检验来确定,或者是难以基于所研究问题实现序列依赖长度的自动识别。因此,传统的计量经济模型对于具有复杂特征的金融时间序列数据预测存在一定局限性。

三、基于机器学习算法的金融时间序列数据预测

机器学习探究的是如何通过计算的手段,利用学习经验改善模型自身的性能,其研究的主要内容是从输入的数据中产生“模型”的算法,挖掘出输入数据之间的关系,即“学习算法”(周志华,2016)^[15]。国内外学者的相关研究已经证实诸如支持向量机(SVM)、BP神经网络等机器学习算法对金融时间序列数据的预测精度优于传统的ARIMA、GARCH等计量模型(Ture & Kurt, 2006^[16];刘海玥和白艳萍,2011^[17])。为此,进一步梳理机器学习算法在金融时间序列数据预测领域的应用。

人工神经网络是由若干称之为神经元或节点的简单处理单元相互连接而成,每个神经元接收来自其他神经元或外部的所有信息,通过激活函数转换为输出信号给其他神经元或外部输出。人工神经网络是一种非常强大的非参数化工具,被广泛应用于信号处理、模式识别等诸多领域,可以有效处理模型参数之间的非线性关系,与其他计量经济学方法相比,它不需定义模型的方程形式,因此无需假设变量之间的函数关系(Kristjanpoller & Minutolo, 2015)^[18]。

BP神经网络是人工神经网络中较为常用的神经网络之一,其框架图可以用图1表示。

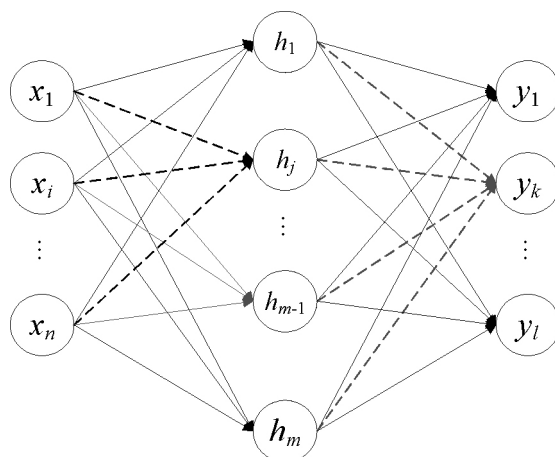


图1 BP神经网络示意图

在图1中 $x_1, \dots, x_i, \dots, x_n$ 为输入层的输入特征向量(对应计量经济学中的解释变量) $h_1, \dots, h_j, \dots, h_m$ 为隐藏层的隐藏单元 $y_1, \dots, y_k, \dots, y_l$ 为输出层的输出单元(对应计量经济学中的被解释变量)。

对于隐藏层的第 j 个隐藏单元有:

$$h_j = g_h(w_1x_1 + \dots + w_ix_i + \dots + w_nx_n + b_{hj}) \quad (4)$$

对于输出层的第 k 个输出单元有:

$$y_k = g_o(v_1h_1 + \dots + v_jh_j + \dots + v_mh_m + b_{ok}) \quad (5)$$

其中 $w_1, \dots, w_i, \dots, w_n$ 为连接输入层与隐藏层的第 j 个隐藏单元的权重 $v_1, \dots, v_j, \dots, v_m$ 为连接隐藏层与输出层的第 k 个输出单元的权重 b_{hj} 与 b_{ok} 为偏置项。 g_h 与 g_o 分别为激活函数,将激活函数设定为非线性激活函数时,可以实现对加权数据的非线性转换。

式(4)说明隐藏层的任一神经元是通过输入特征向量的线性加权和非线性转换得到,式(5)说明输出层的任一神经元是通过隐藏层的所有神经元线性加权和非线性转换得到。可以通过下式概括BP神经网络中输入层与输出层(即解释变量与被解释变量)的关系:

$$y_k = f(x_1, \dots, x_i, \dots, x_n) + b \quad (6)$$

其中 f 为非线性函数,反映神经网络的结构(如隐藏层数量、隐藏单元数量)和不同层之间的连接权重等。因此,该模型可以刻画特征向量 $x_1, \dots, x_i, \dots, x_n$ 对输出金融时间序列数据 y_k 的非线性影响。

人工神经网络是一种非参数化的非线性模型,与诸如ARIMA、GARCH等参数计量经济模型相比,具有以下优势:首先,作为非线性模型的神经网络在选择输入输出关系方面提供了更多的可能性,非线性模型的

预测能力强于线性模型(Lendasse 2000)^[19]。其次,人工神经网络没有标准的结构方程,更容易适应金融市场的变化,故更适用于金融时间序列数据的预测(Guresen et al. 2011)^[20]。最后,神经网络是基于数据驱动的建模思路,是非参数化的弱模型,与大多数参数化的统计、计量模型相比,神经网络更容易避免模型误设问题,因而在学习金融时间序列数据的动力学方面更强大(Tay & Cao 2000)^[21]。例如,李春伟和张骏(2006)采用 BP 神经网络对股票价格序列数据进行中期预测^[22],其实证结果证实了 BP 神经网络在股价预测中的有效性。Wang et al. (2011) 指出基于线性和正态分布假设的计量模型无法预测具有复杂特征的股票价格等金融时间序列数据^[23],而不存在限制性假设的 BP 神经网络等人工智能算法更具适用性,基于上证指数的实证结果也证实了 BP 神经网络预测金融时间序列数据的有效性。

在金融时间序列数据预测中,另外一种常用的机器学习算法是支持向量机(SVM)。基于结构风险最小化的独特原理,通过最小化泛化误差的上界, SVM 更容易避免过拟合问题,提高对样本外数据的预测能力,同时 SVM 的解是通过线性约束的二次规划求得,故是唯一的全局最优解,而神经网络容易陷入局部最优解(Huang et al. 2015)^[24]。股票市场数据具有高噪声和复杂维度等特点,而人工神经网络对噪声数据的预测经常表现出不一致性, SVM 更适用于金融时间序列数据预测(Kim 2003)^[25]。徐国祥和杨振建(2011)认为 BP 神经网络容易陷入局部最优解^[26],继而构建了基于主成分分析、遗传算法以及 SVM 的金融时间序列数据预测模型,实证结果证实了该模型对金融时间序列数据预测的有效性。Bustos et al. (2017) 比较了 SVM 与人工神经网络在预测股票市场价格指数方面的区别^[27],实证结果发现 SVM 具有更好的预测结果。

此外,学者对 K 近邻、决策树等机器学习算法在金融时间序列数据预测上的效果也做了一定探讨。Ahmed et al. (2010) 比较了多层感知机、K 近邻、分类与回归树、支持向量回归、高斯过程等不同机器学习算法对金融时间序列数据的预测能力^[28],发现多层感知机和高斯过程对金融时间序列数据表现出了更好的预测效果。Kumar & Thenmozhi(2006) 认为机器学习算法在建模时不需考虑数据的分布特征^[29],只需关注在给定特征向量的情况下如何提高预测精度,同时, SVM 和随机森林的解可能是全局最优的,而 BP 神经网络的解容易陷入局部最优,并且 SVM 和随机森林具有更好的泛化能力,更容易避免过拟合问题,其实证结果也证实了 SVM 对于金融时间序列数据的预测精度高于其他机器学习算法。

近年来,也有学者提出将 ARIMA、GARCH 等计量经济学模型与神经网络、SVM 等机器学习算法相结合,以预测金融时间序列数据。Bildirici & Ersin(2009) 指出人工神经网络(ANN) 在处理非线性数据方面存在较大优势^[30],故提出将 GARCH 类模型和人工神经网络(ANN) 相结合构建 ANN-GARCH 模型,以预测股票收益率,其实证结果发现,相较于单独的 GARCH 类模型, ANN-GARCH 模型具有更高的预测精度。Kumar & Thenmozhi(2014) 将线性的 ARIMA 模型与非线性的 SVM、神经网络、随机森林等机器学习算法结合预测股指收益率^[31],发现 ARIMA-SVM 预测精度最高。熊志斌(2011) 构建了基于 ARIMA-神经网络的金融时间序列数据预测模型^[32],即将时间序列分解为线性主体和非线性残差两部分,进一步利用 ARIMA 预测线性部分,利用神经网络预测残差部分,最后集成两部分的预测效果,基于人民币汇率的预测结果表明,集成模型的预测精度高于 ARIMA、神经网络等单个模型的预测精度。张贵生和张信东(2016) 认为不同金融市场存在显著的联动关系^[33],应充分考虑不同金融市场信息的相互影响,故提出 SVM-GARCH 模型预测股票价格,进而同时纳入时间序列数据的非线性复杂结构和异方差效应,其实证表明该模型预测精度优于 ARMA-GARCH 模型。

综上所述,与 ARIMA、GARCH 等计量模型不同,诸如 BP 神经网络、支持向量机、随机森林等机器学习算法更大程度上是一种非参数化、非线性、数据驱动的模式,可以有效处理金融时间序列数据的非线性、非平稳等特征,更容易挖掘金融时间序列数据间的隐含模式,实现对金融时间序列数据间未知关系的认知。因此,将机器学习算法应用于金融时间序列数据预测更具优势。然而,机器学习算法在处理金融时间序列数据序列依赖特征方面存在一定缺陷性,为将时间序列数据的序列相关特征纳入预测模型,需要通过人为设定输入特征的方式来实现,这无疑存在一定的主观性和局限性。此外,机器学习是一种浅层次学习算法,对于具有高维且具有复杂特征的金融时间序列数据,难以实现对其复杂特征以及重要信息的挖掘。因此,诸如 BP 神

神经网络、支持向量机等机器学习算法对于大数据以及人工智能背景下的金融时间序列数据预测存在的问题。

四、基于深度学习算法的金融时间序列数据预测

学者从理论与实证角度探究了机器学习在金融时间序列数据预测领域的可行性与有效性。然而,SVM、BP神经网络等浅层次机器学习算法,在处理复杂高维度的数据方面存在较大局限性,且存在维数灾难与无效的特征表示等诸多问题(Bengio & LeCun, 2007)^[34]。深度学习是具有多层级表示的表示学习方法,通过叠加多层简单但非线性的模块获得,每个模块将每层的表示(从输入层开始)转换为下一层更抽象的表示,只要有足够多的这种转换,就可以学习非常复杂的函数(Lecun et al., 2015)^[35]。在提高样本内数据预测精度的同时,更容易缓解过拟合问题(Heaton et al., 2016)^[36]。这将非常有助于挖掘金融时间序列数据的内在复杂特征。为此,本文进一步从理论与实证角度分析深度学习在金融时间序列数据预测领域的应用。

Lecun & Bengio(1995)指出标准的全连接前馈神经网络忽略了输入特征向量的拓扑结构^[37],即忽略了图片、语音、时间序列等数据所存在的高度的时间或空间局部相关性,而局部相关性是识别空间或时间对象之前提取和组合局部特征的一个重要原因。卷积神经网络(CNN)通过限制隐藏单元的局部感知来提取局部特征。卷积神经网络包括卷积运算、池化层运算等步骤,极大地降低CNN中需要训练的诸如神经网络权重等参数的数量,进而降低神经网络的复杂程度,使得神经网络的学习和训练过程更加简单。同时,在各种信号和信息处理任务中性能优于标准的全连接神经网络(周飞燕等,2017)^[38]。因此,学者探究了CNN在金融时间序列数据预测领域的适用性。Tsantekidis et al.(2017)将卷积神经网络应用于股票价格预测^[39],其实证结果表明该方法的预测效果优于BP神经网络、支持向量机等模型的预测效果。Sezer & Ozbayoglu(2018)采用CNN构建算法交易^[40],以识别买、卖和持有时机,基于股票和ETF的实证结果表明该策略的表现优于买入、持有策略和其他算法交易策略。林杰和龚正(2017)指出期货价格具有显著的非线性特征^[41],故提出将BP神经网络和CNN应用于期货价格预测,其实证结果发现两种模型都可以取得较好的预测效果,但整体而言,BP神经网络的预测效果优于CNN。

CNN可以有效处理数据的空间关联特征,但是难以捕捉金融时间序列数据的序列相关特征,故在金融时间序列数据预测领域仍存在一定的局限性。而深度学习的循环神经网络(RNN)则纳入了序列相关性,在学习过程中当前状态会包含先前时间序列的所有历史信息,克服机器学习算法难以处理金融时间序列数据序列相关特征以及学习能力差的缺陷,RNN在金融时间序列数据预测领域也取得广泛的应用。Giles et al.(2001)指出时间序列数据具有高噪声、非平稳、非线性等复杂特征^[42],而BP神经网络在处理高噪声、小样本时间序列数据时存在一定的局限性,RNN应用于时间序列数据预测则可以考虑时间序列数据的序列依赖关系,故RNN更适用于金融时间序列数据预测。Hsieh et al.(2011)指出RNN的结构比传统人工神经网络的结构更简单^[43],更适用于金融时间序列数据预测,故提出采用RNN预测股票价格,其实证结果证实RNN的预测效果优于BP神经网络。Rather et al.(2015)探究了非线性模型RNN以及线性模型^[44],诸如ARMA、指数平滑模型等方法对股票收益率的预测效果,实证结果表明RNN的预测效果优于线性模型。

然而,标准RNN结构可能存在梯度消失或者爆炸问题,对于序列长期依赖关系的学习存在一定困难(Hochreiter et al., 2001)^[45]。而包含记忆模块的长短期记忆(LSTM)神经网络,很好地解决了序列的长期依赖问题,且在诸如顺序文本翻译(Sutskever et al., 2014)^[46]等时序数据应用领域取得重大成功。LSTM神经网络是RNN的一种形式,可以有效处理金融时间序列数据的序列相关问题,在金融时间序列数据预测领域,LSTM神经网络逐渐取代RNN模型。Bao et al.(2017)指出^[47],相比于传统的人工神经网络、SVM等算法,LSTM神经网络可以通过提取稳健的特征以描述更加复杂的真实数据,同时捕捉金融时间序列数据的序列相关信息,故将LSTM神经网络应用于金融时间序列数据预测,其实证结果也证实了LSTM神经网络优于其他预测模型,可以获得更加精确的预测结果,具有更好的盈利能力。Fischer & Krauss(2018)认为^[48],LSTM神经网络可以从含有大量噪声的金融时间序列数据中提取重要的信息,其实证结果表明,与随机森林、Logis-

tic 回归等模型相比,LSTM 神经网络预测效果更好。Kim & Won(2018)提出将 LSTM 神经网络和 GARCH 类模型结合构建股票价格波动率预测模型^[49],即将 GRACH 类模型预测的波动率作为 LSTM 神经网络的输入特征向量,实证结果表明该混合模型比单个的 GARCH 类模型和 LSTM 神经网络模型更好地预测了股价波动率。谢琪等(2019)认为^[50],LSTM 神经网络可以有效金融时间序列数据中长时间的信息,适用于金融时间序列数据的分析与预测,故采用 LSTM 神经网络的集成学习预测股票价格,其实证结果表明该方法的预测精度高于传统的神经网络。杨青和王晨蔚(2019)指出^[51],LSTM 神经网络具有更强的泛化能力,在股票指数预测方面,比支持向量回归、多层感知机、ARIMA 等模型更具优势。

综上所述,学者证实了诸如 RNN、LSTM 神经网络以及 CNN 等深度学习算法应用于金融时间序列数据预测的可行性与有效性。同时,在实证预测效果上,深度学习的预测效果优于 BP 神经网络、SVM 等机器学习算法。与机器学习算相比,深度学习不仅具有更强大的层次特征提取能力,且诸如 LSTM 神经网络等深度学习算法可以很好地刻画金融时间序列数据的序列相关特征。因此,相较于机器学习算法以及计量经济学模型等金融时间序列数据预测模型,深度学习更能有效挖掘金融时间序列数据所包含的重要信息,在金融时间序列数据预测领域更具优势。

五、结论

本文认为,科学有效的金融时间序列数据预测模型应具备以下特征:

第一,具有强大且稳健的特征学习能力,能够充分捕捉金融时间序列数据的高噪声、非线性、非平稳等复杂特征,以提取影响金融时间序列数据变动的重要影响信息。

第二,可有效反映金融时间序列影响因素的非线性动态交互关系,进而将金融时间序列数据间的非线性交互关系纳入预测模型。

第三,具有较强的泛化能力,进而实现对样本外数据的精确预测,以提高预测模型在现实中的应用能力。

第四,应考虑金融时间序列数据的长短期序列依赖特征,实现基于输入数据自动识别序列依赖长度。

因此,相较于计量经济学模型、机器学习等时间序列数据预测方法,在图像识别、无人驾驶、自然语言等诸多人工智能领域取得巨大成功的深度学习算法,更适用于大数据以及人工智能背景下的金融时间序列数据预测。其主要原因如下:

第一,深度学习具有强大的特征学习能力,可实现对金融时间序列数据复杂特征的挖掘,这可以克服传统计量模型对于具有复杂特征的金融时间序列数据难以建模的缺陷,克服浅层次机器学习算法难以学习高维复杂数据特征的缺陷。

第二,诸如 LSTM 神经网络等深度学习算法可以有效纳入时间序列数据的序列依赖特征,实现对时间序列数据长短期依关系的动态识别,进而克服其他预测模型难以纳入金融时间序列数据序列相关特征或者难以实现序列依赖长度难以自动识别的缺陷。

第三,深度学习具有更好的泛化能力,可有效提高对样本外时间序列数据的预测能力,继而增强预测模型在实际问题中的适用性。

第四,深度学习可实现对海量及非结构化金融数据的建模,提高对文本、图片等非结构化数据信息的挖掘能力,这将克服计量经济模型、机器学习算法难以挖掘非结构化数据、难以实现对海量数据挖掘的缺陷。将深度学习应用于金融时间序列数据预测,对于提高预测精度,实现大数据、人工智能背景下的金融时间序列数据预测和多学科交叉融合研究,拓展现有的金融时间序列研究方法,具有重要的理论和现实意义。

参考文献:

- [1] Abu - Mostafa Y S ,Atiya A F. Introduction to financial forecasting [J]. Applied Intelligence ,1996(3) : 205 - 213.
- [2] Hall J W. Adaptive Selection of US Stocks with Neural Nets [J]. Trading on the Edge: Neural ,Genetic ,and Fuzzy Systems for Chaotic Financial Markets. New York: Wiley ,1994:45 - 65.

- [3] Yule G U. VII. On a Method of Investigating Periodicities Disturbed Series ,with Special Reference to Wolfer's Sunspot Numbers [J]. Phil. Trans. R. Soc. Lond. A ,1927(636 – 646) : 267 – 298.
- [4] Mikosch T ,Stărică C. Nonstationarities in Financial Time Series ,The Long – range Dependence ,and The IGARCH Effects [J]. Review of Economics and Statistics 2004(1) : 378 – 390.
- [5] 池启水. 中国石油消费量增长趋势分析——基于 ARIMA 模型的预测与分析 [J]. 资源科学 2007(5) : 69 – 73.
- [6] Ediger V. Ş. ,Akar S. Arima Forecasting of Primary Energy Demand by Fuel in Turkey [J]. Energy Policy , 2007(3) : 1701 – 1708.
- [7] 谷政 张维. 基于 WAVELET – GARCH 组合方法的中国保险深度分析 [J]. 江西科学 2013(3) : 403 – 408.
- [8] Pai P F ,Lin C S. A Hybrid ARIMA and Support Vector Machines Model in Stock Price Forecasting [J]. Omega 2005(6) : 497 – 505.
- [9] Robert F. Engle. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom inflation [J]. Econometrica ,1982(4) : 987 – 1007.
- [10] Bollerslev T. Generalized Autoregressive Conditional Heteroskedasticity [J]. Journal of Econometrics ,1986 (3) : 307 – 327.
- [11] 惠晓峰 柳鸿生 胡伟 等. 基于时间序列 GARCH 模型的人民币汇率预测 [J]. 金融研究 2003(5) : 99 – 105.
- [12] 刘国旗. 非线性 GARCH 模型在中国股市波动预测中的应用研究 [J]. 统计研究 2000(1) : 49 – 52.
- [13] 于志军 杨善林. 基于误差校正的 GARCH 股票价格预测模型 [J]. 中国管理科学 2013(S1) : 341 – 345.
- [14] Taylor G W. Composable ,Distributed – State Models for High – Dimensional Time Series [M]. University of Toronto 2009.
- [15] 周志华. 机器学习 [M]. 北京: 清华大学出版社 2016.
- [16] Ture M ,Kurt I. Comparison of Four Different Time Series Methods to Forecast Hepatitis A Virus Infection [J]. Expert Systems with Applications 2006(1) : 41 – 46.
- [17] 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(4) : 14 – 19.
- [18] Kristjanpoller W ,Minutolo M C. Gold Price Volatility: A Forecasting Approach Using the Artificial Neural Network – GARCH model [J]. Expert Systems with Applications 2015(20) : 7245 – 7251.
- [19] Lendasse A ,de Bodt E ,Wertz V ,et al. Non – linear Financial Time Series Forecasting – Application to the Bel 20 Stock Market Index [J]. European Journal of Economic and Social Systems 2000(1) : 81 – 91.
- [20] Guresen E ,Kayakutlu G ,Daim T U. Using Artificial Neural Network Models in Stock Market Index Prediction [J]. Expert Systems with Applications 2011(8) : 10389 – 10397.
- [21] Tay F E H ,Cao L. Application of Support Vector Machines in Financial time Series Forecasting [J]. Omega , 2001(4) : 309 – 317.
- [22] 李春伟 张骏. 基于神经网络的股票中期预测 [J]. 计算机工程与科学 2006(5) : 115 – 117.
- [23] Wang J Z ,Wang J J ,Zhang Z G ,et al. Forecasting Stock Indices with Back Propagation Neural Network [J]. Expert Systems with Applications 2011(11) : 14346 – 14355.
- [24] Huang W ,Nakamori Y ,Wang S Y. Forecasting Stock Market Movement Direction with Support Vector Machine [J]. Computers & Operations Research 2005(10) : 2513 – 2522.
- [25] Kim K. Financial Time Series Forecasting Using Support Vector Machines [J]. Neurocomputing 2003(1 – 2) : 307 – 319.
- [26] 徐国祥 杨振建. PCA – GA – SVM 模型的构建及应用研究——沪深 300 指数预测精度实证分析 [J]. 数量经济技术经济研究 2011(2) : 135 – 147.

- [27] Bustos O ,Pomares A ,Gonzalez E. A Comparison between SVM and Multilayer Perceptron in Predicting an Emerging Financial Market: Colombian Stock Market [C]//de Innovacion y Tendencias en Ingenieria(CONITI) 2017 Congreso Internacional. IEEE 2017:1 – 6.
- [28] Ahmed N K ,Atiya A F ,Gayar N E ,et al. An Empirical Comparison of Machinelearning Models for time Series Forecasting [J]. *Econometric Reviews* 2010(5 – 6) :594 – 621.
- [29] Kumar M ,Thenmozhi M. Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Forest [J]. *Social Science Electronic Publishinl* 2006(2) :87 – 102.
- [30] Bildirici M ,Ersin ÖÖ. Improving Forecasts of GARCH Family Models with the Artificial Neural Networks: An Application to the Daily Returns in Lstanbul Stock Exchange [J]. *Expert Systems with Applications* 2009(4) :7355 – 7362.
- [31] Kumar M ,Thenmozhi M. Forecasting Stock Index Returns Using ARIMA – SVM ,ARIMA – ANN ,and ARIMA – random forest hybrid models [J]. *International Journal of Banking ,Accounting and Finance* 2014(3) :284 – 308.
- [32] 熊志斌. ARIMA 融合神经网络的人民币汇率预测模型研究 [J]. *数量经济技术经济研究* 2011(6) :64 – 76.
- [33] 张贵生, 张信东. 基于近邻互信息的 SVM – GARCH 股票价格预测模型研究 [J]. *中国管理科学* 2016(9) :11 – 20.
- [34] Bengio Y ,LeCun Y. Scaling Learning Algorithms Towards AI [J]. *Large – scale Kernel Machines* 2007(5) :1 – 41.
- [35] Lecun Y ,Bengio Y ,Hinton G. Deep learning [J]. *Nature* 2015(3) :436 – 444.
- [36] Heaton J B ,Polson N G ,Witte J H. Deep Learning in Finance [J]. *arXiv preprint arXiv:1602.06561* 2016.
- [37] LeCun Y ,Bengio Y. Convolutional Networks for Images ,Speech ,and Time Series [J]. *The Handbook of Brain Theory and Neural Networks* ,1995(10) :1995.
- [38] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述 [J]. *计算机学报* 2017(6) :1229 – 1251.
- [39] Tsantekidis A ,Passalis N ,Tefas A ,et al. Forecasting Stock Prices From the Limit Order Book Using Convolutional Neural networks [C]//Business Informatics(CBI) 2017 IEEE 19th Conference on. IEEE 2017 ,1:7 – 12.
- [40] Sezer O B ,Ozbayoglu A M. Algorithmic Financial Trading with Deep Convolutional Neural Networks: Time Series to Image Conversion Approach [J]. *Applied Soft Computing* 2018(70) :525 – 538.
- [41] 林杰, 龚正. 基于人工神经网络的沪锌期货价格预测研究 [J]. *财经理论与实践* 2017(2) :54 – 57.
- [42] Giles C L ,Lawrence S ,Tsoi A C. Noisy Time Series Prediction Using Recurrent Neural Networks and Grammatical Inference [J]. *Machine Learning* 2001(1 – 2) :161 – 183.
- [43] Hsieh T J ,Hsiao H F ,Yeh W C. Forecasting Stock Markets Using Wavelet Transforms and Recurrent Neural Networks: An Integrated System Based on Artificial Bee Colony Algorithm [J]. *Applied Soft Computing* 2011(2) :2510 – 2525.
- [44] Rather A M ,Agarwal A ,Sastry V N. Recurrent Neural Network and a Hybrid Model for Prediction of Stock Returns [J]. *Expert Systems with Applications* 2015(6) :3234 – 3241.
- [45] Schmidhuber J. Gradient Flow in Recurrent Nets: the Difficulty of Learning Long – Term Dependencies [M]. New Jersey: Wiley – IEEE Press 2001.
- [46] Sutskever I ,Vinyals O ,Le Q V. Sequence to Sequence Learning with Neural Networks [C]//Advances in Neural Information Processing Systems. 2014:3104 – 3112.
- [47] Bao W ,Yue J ,Rao Y. A Deep Learning Framework for Financial Time Series Using Stacked Autoencoders and Long – short Term Memory [J]. *PIOS ONE* 2017(7) :e0180944.

- [48] Fischer T, Krauss C. Deep Learning with Long short-term Memory Networks for Financial Market Predictions [J]. European Journal of Operational Research 2018(2): 654 – 669.
- [49] Kim H Y, Won C H. Forecasting the Volatility of Stock Price Index: A Hybrid Model Integrating LSTM with Multiple GARCH-type Models [J]. Expert Systems with Applications 2018(103): 25 – 37.
- [50] 谢琪, 程耕国, 徐旭. 基于神经网络集成学习股票预测模型的研究 [J/OL]. 计算机工程与应用: 1 – 8 [2019 – 02 – 11]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20190123.1540.012.html>.
- [51] 杨青, 王晨蔚. 基于深度学习 LSTM 神经网络的全球股票指数预测研究 [J]. 统计研究 2019(3): 65 – 77.

Financial Time Series Prediction: A Literature Review and Prospect

YAN Hongju^{1 2}

(1. Postdoctoral Research Station of Agricultural Bank of China, Beijing 100005, China;

2. School of Economics, Peking University, Beijing 100871, China)

Abstract: Financial time series data has complex characteristics such as non-linear, non-stationary and high noise. With the rapid development of mobile internet and artificial intelligence, massive structured and unstructured data are constantly generated, and the relation among them is increasingly complex. So, it is very important to build a scientific and reasonable financial time series data prediction model and fully mine the important information hidden in the financial time series data. Compared with econometric model and machine learning, deep learning algorithm, which has achieved great success in image recognition, self-driving, natural language and many other fields of artificial intelligence, is more suitable for financial time series data prediction under the background of big data and artificial intelligence.

Key words: Time series data; Econometric model; Machine learning; Artificial intelligence

(责任编辑: 黎芳)