

运用BP神经网络构建比特币市场预测模型

李 靖(副教授)

【摘要】 本文针对比特币市场发展的随机性、突变性特点,采用BP神经网络建立了比特币市场预测模型。选取2010年8月18日至2016年3月2日的数据作为样本,构建了多个预测模型,分别运用不同时间段的数据对单日 and 一周的比特币市场进行预测。结果表明,越近的数据越有利于准确预测比特币的价格,越是短期预测越能较好地拟合比特币发展趋势。

【关键词】 比特币; BP神经网络; 市场预测

【中图分类号】 F820;F830.9

【文献标识码】 A

【文章编号】 1004-0994(2016)21-0033-4

比特币是一种P2P形式的数字货币,点对点传输意味着一个去中心化的支付系统,即任何在线支付都不需要通过第三方金融机构,而是直接由一方发起并支付给另外一方。比特币构建的是一种公共交易平台和记账平台,全民记账,用户自治。自2009年问世以来,比特币去中心化、匿名、数量有限、全球流通等特性引起了国际关注。比特币的汇率先是一路上涨,再逐渐回落。期间,大量用户涌入比特币投资市场,比特币的日交易量最高接近600万元,用户面临着巨大的比特币投资风险,比特币市场成为民众和学者关注的热点。

为了帮助用户认清比特币发展趋势,进行理性投资,规避投资风险,同时为政府及相关部门政策制定提供量化指标及参考,本文拟将人工智能技术引入比特币市场研究,以对比特币市场进行预测。

一、研究综述

目前,国内外很多学者对比特币市场尤其是价格进行了深入研究。Luisanna Cocco等(2014)搜集了2012年1月1日至2014年4月10日的比特币交易数据,采用定性 with 定量结合的研究方法,构建基于代理的比特币市场模型,模型基于帕累托的比特币初始分配策略,模拟了投机和非投机两类用户现实订单交易、比特币增量产生以及用户不断增多的过程。研究结果表明:比特币价格随机性强,存在聚类波动,用户的交易行为与累计收益分布有相关性,投机者尤为明显。

Martins Buchholz(2012)对2010年6月至2012年3月的比特币市场相关数据进行收集和分析,构建了ARCH/GARCH模型,研究指出比特币价格存在泡沫。

刘力臻、王庆龙(2015)基于2013年4月2日至2014年4月2日的比特币市场交易数据构建回归模型,指出用户交易

行为存在羊群效应,用户投机行为和比特币价格相互影响,市场结构对比特币价格有影响。

David Garcia等(2014)基于时间序列数据构建了向量自回归模型,对比特币用户、价格、搜索量和社会媒体之间的相互影响进行了研究,结果表明:四者之间形成了社会 and 用户两个反馈回路。

Jermain Kaminski等(2014)结合推特网上2013年11月至2014年3月的161200条数据对比特币价格、交易量和用户的推特情绪的相关性进行了研究。Ladislav Kristoufek(2013)研究指出,搜索引擎和比特币价格双向影响,投机行为影响比特币价格,并且比特币价格存在泡沫。

Devavrat Shah等(2014)基于OKCOIN交易平台2014年2月至7月的交易数据,以10秒为时间点分别建立以30分钟、60分钟、120分钟为间隔的时间序列,运用贝叶斯回归模型研究了比特币未来价格的平均变化,并用于实际检验。研究结果表明:价格变化呈现头肩形和三角形,但是要想扩大使用规模需要再深入研究。梁秋等(2015)利用小波分析方法对比特币市场价格进行了预测。

综上所述,学者们对比特币价格及相关影响因素进行了一些研究,他们多数使用回归和相关性分析方法,针对某些特定的影响因素对比特币市场进行研究,但针对比特币价格的预测研究较少。从他们的研究也可以看出,比特币市场是一个庞大而复杂的系统,价格变化受到多种因素的影响,具有随机性、突变性、非线性等特点,如果用传统的方法很难对比特币价格作出较为准确的预测。而BP神经网络可以对复杂的非线性系统进行训练和学习,建立数据之间的复杂函数关系,并在此基础上实现预测功能。

【基金项目】 国家社会科学基金青年项目“比特币的发展及其风险研究”(项目编号:14CJY069)

□ 改革与发展

根据Kolmogorov原理,三层BP(Back Propagation)网络可以在任意希望的精度上建立任意的连续函数,本文拟基于三层BP神经网络构建比特币市场的时间序列预测模型,以期对比特币的市场发展及其趋势进行预测。考虑到比特币市场随机性强,因此本文主要针对比特币市场的短期发展趋势进行预测。

二、BP神经网络

BP(Back Propagation)神经网络又称为误差反向传播网络,该网络由输入层、中间层和输出层组成。中间层位于输入层和输出层之间,不与外界建立直接联系,亦称为隐含层,它可以是单层也可以是多层。各层之间的神经元全连接,同层内的神经元无连接。BP神经网络结构如图1所示:

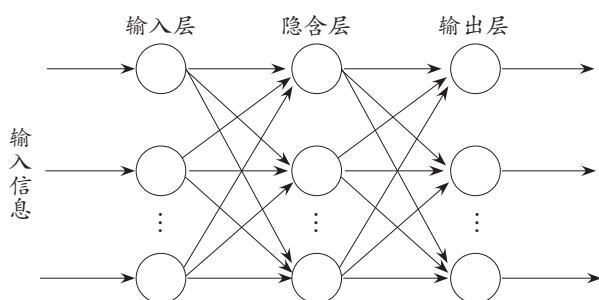


图1 三层BP神经网络结构

BP神经网络采用指导的方式进行学习和训练,学习和训练的过程主要由工作信号正向传播和误差信号反向传播两部分组成。在正向传播时,输入信号从输入层经由隐含层逐层处理,在输出层输出信号。如果输出层得到的信号误差超过预期,就将误差信号由输出端逐层反向向前传播,反向传播主要是不断调整网络的权值和阈值,使网络的误差平方和最小。通过这种原理,BP神经网络学习和存储各个变量之间的复杂非线性映射关系,从而实现预测。

三、比特币价格预测模型建立

(一)用比特币前一天的市场数据预测当日比特币市场

1. 参数确定及数据采集。根据现有的研究成果,比特币的交易市场、用户的心理和行为、搜索引擎、资本进入、外部政策等都会影响到比特币的价格,所以可以考虑将这些参数作为比特币价格预测的自变量。但是这些参数对市场的影响最终体现在比特币交易方面,比特币交易是所有影响因素的直接体现。因此,本模型(模型1)中选择比特币每一天交易市场中的相关数据作为参数,把第二天比特币的价格和交易量作为预测参数。

通过编写程序,抓取相关网站(www.blockchain.info; http://sourceforge.net/projects/bitcoin/files/stats/)2009年1月1日至2016年3月2日的所有数据。按照时间序列排列数据,选取比特币价格非零的数据作为模型数据源,即2010年8月18日以后的1774条数据。

具体变量名称及假设如表1所示。

表1 比特币市场预测自变量

变量名称	假设
比特币价格	前期价格对比特币后期市场有影响
全球比特币钱包累计下载量	钱包下载量越多,说明用户越多,而用户数量(人气)对比特币市场及价格有影响
比特币每日增量	比特币的供求量与市场及价格有关系
比特币每日交易宗数	交易量是衡量比特币市场发展的重要指标
比特币每天的估计交易量	
独立的比特币活跃地址	
非热门地址交易宗数	
交易不包括长链10宗数	
交易不包括长链100宗数	
交易不包括长链1000宗数	少数人企图操纵比特币交易量,这可能导致比特币价格波动
交易不包括长链10000宗数	

2. 数据预处理。本文采用Mapminmax函数对原始数据样本中的输入和输出进行归一化处理,使处理以后的数据映射在对应的区间上,默认区间是[-1,1],计算公式为:

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} (y_{\max} - y_{\min}) + y_{\min}$$

该函数归一化是按行进行处理的,上式中x是待处理的数据, x_{\max} 、 x_{\min} 分别是x所在矩阵所在行的最大值和最小值, y_{\max} 、 y_{\min} 分别是指定区间的最大值和最小值,默认值分别为1和-1。BP神经网络训练结束以后,得到的结果要经过反归一化处理,还原为正常值。

3. 数据建模。

(1)初始化参数的设定。本模型学习率设定为0.01,最大训练次数设定为1000,目标误差设定为0.0001,性能函数为MSE,训练函数为trainlm函数,其他采用默认值。

(2)隐含层节点数的确定。最常用的隐含层节点数的公式为: $j = \sqrt{n + m} + a$ 。其中,j是隐含层节点数,n是输入层节点数,m是输出层节点数,a是1~10之间的常数。本模型中输入层变量为11个,对应n=11;输出变量为2个,对应m=2。因此,根据公式计算得到隐含层节点数为4~13之间的常数。分别在网络其他参数不变的情况下,对不同的隐含层节点进行试验,并比较其均方误差,结果如表2所示:

表2 不同隐含层节点的均方误差

隐含层节点数	MSE	隐含层节点数	MSE
4	0.003882	9	0.0016988
5	0.0013192	10	0.0013232
6	0.0019868	11	0.0026267
7	0.0029922	12	0.0027171
8	0.0028504	13	0.0018012

当隐含层节点数为10时均方误差最小,所以将隐含层节点数设定为10。

(3)模型训练和仿真结果。将收集到的数据集分为两部分进行模型研究:第一部分,取前2013组数据,即2010年8月18日到2016年2月20日的数据作为训练集,通过训练和学习这些数据得到各变量之间的复杂映射关系;第二部分,把剩余10组数据,即2015年2月21日到2015年3月2日的数据作为仿真数据集,用学习和训练后得到的复杂映射关系来预测后面10组数据,预测值和实际值对比,得到相应的误差。预测和误差结果如图2~图5所示。

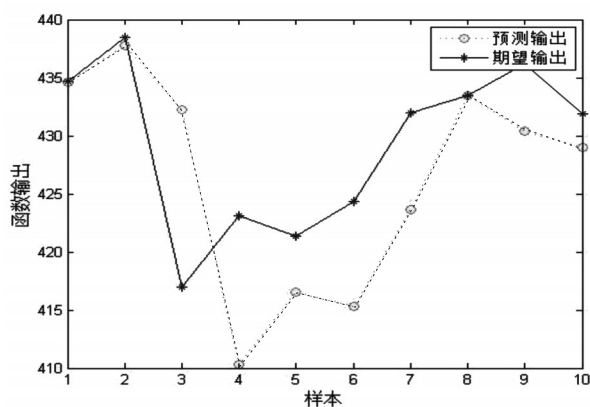


图2 用比特币前一天数据预测当日比特币价格

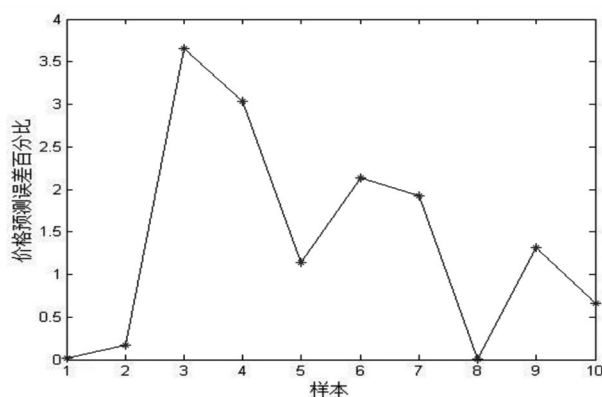


图3 用比特币前一天数据预测当日比特币价格误差

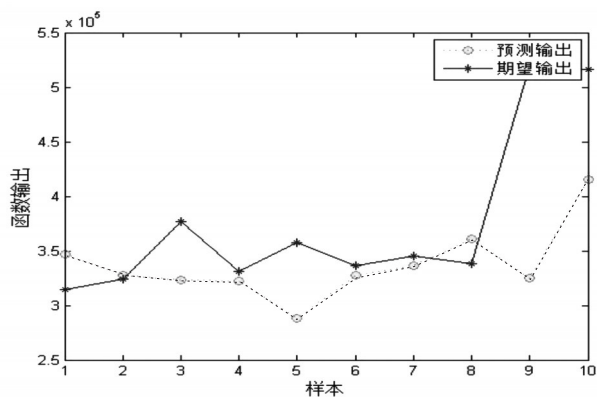


图4 比特币交易量预测

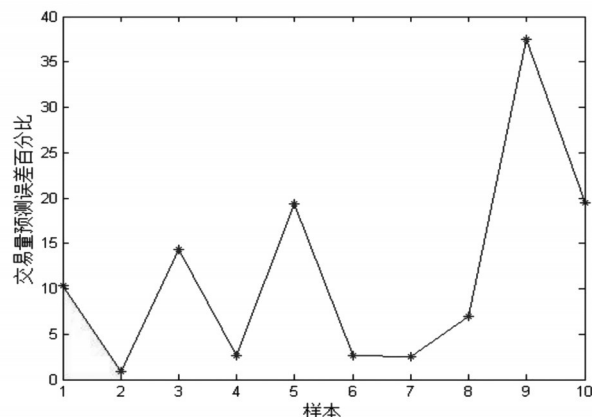


图5 比特币交易量误差百分比

从图中可以看出,比特币价格误差范围在3.8%以内,交易量最大误差超过35%。由于比特币交易量预测的误差太大,以下模型中去掉对比特币交易量的预测,而仅仅对比特币的价格进行预测。

(二)用比特币前一周的市场数据预测当日比特币市场

与模型1对比,本模型(模型2)的不同之处有三点:第一,输入值是数据样本中输出变量上一周7天的值;第二,输出值仅仅为次日的价格,而没有交易量;第三,数据样本为2017个。其他预测参数和过程类似。比特币价格预测误差控制在5%以下,预测结果如图6、图7所示:

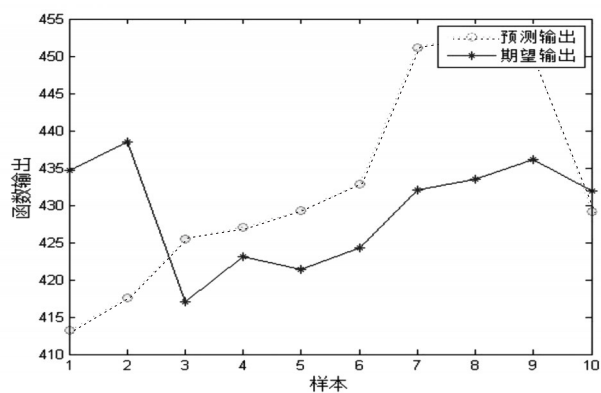


图6 用比特币前一周数据预测当日比特币价格

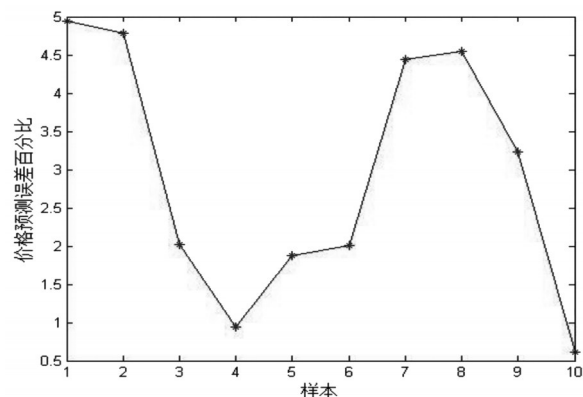


图7 用比特币前一周数据预测当日比特币价格误差

□ 改革与发展

(三) 用比特币前一个月的市场数据预测当日比特币市场

与模型1对比,本模型(模型3)的不同之处有三点:第一,输入值是数据样本中输出变量前30天的值;第二,输出值仅仅为次日的价格,而没有交易量;第三,数据样本为1993个。其他预测参数和过程类似。比特币价格预测误差控制在10%以下,如图8所示:

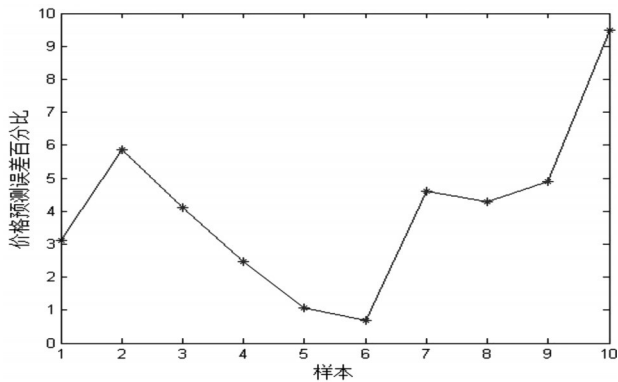


图8 用比特币前一个月数据预测当日比特币价格误差

以上三个模型预测结果表明,用前一日的比特币市场价格数据对比特币当日市场价格进行预测的误差最小,表明越是临近的数据对于比特币当日价格的预测越精确,也反映出比特币价格市场具有较大的突变性和随机性。但是比特币交易市场预测效果不理想。

(四) 用比特币前一周和前一个月市场数据预测未来一周比特币市场价格趋势

以上模型仅仅是预测比特币当日(一天)的价格,为了能够预测某个周期内比特币价格的波动,为用户提供价格波动预警,本研究尝试构建BP神经网络模型(模型4),对比特币未来某个周期内的价格变动进行预测。考虑到比特币价格波动较大,本研究将价格预测周期设定为7天。

分别利用比特币前一周和前一个月的市场数据对比特币未来7天的市场价格进行预测。模型构建过程和上述模型基本相同,得到价格百分比的误差如图9、10所示:

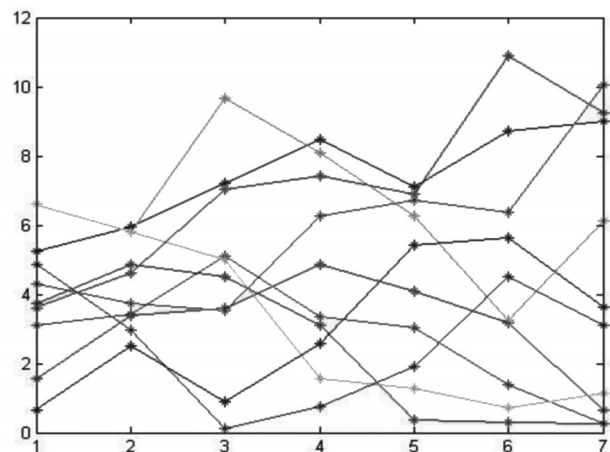


图9 用前一周数据预测未来7天比特币价格误差

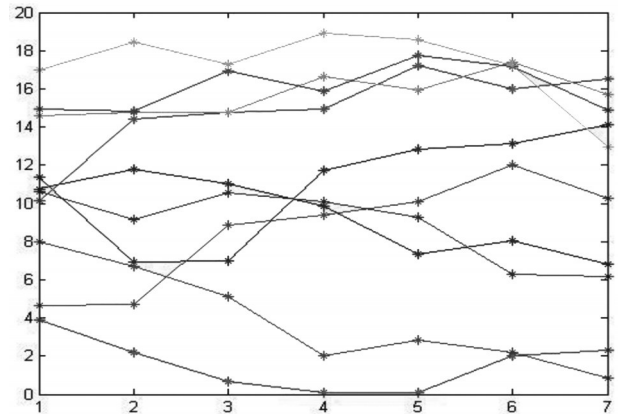


图10 用前一个月数据预测未来7天比特币价格误差

预测结果表明,用前一周数据对未来7天比特币价格进行预测的误差大部分能够控制在10%以下,用前一个月对未来7天比特币价格进行预测的误差范围相对较大,集中在20%以内。

四、结论

本文创新性地把人工智能技术引入比特币市场的研究,运用BP神经网络对比特币的短期市场构建了多个模型。从数据分析中可以得到:

在比特币价格预测方面,当日价格的整体预测效果较好,而且,越是临近的数据,对于比特币当日价格的预测越准确,误差可以控制到3.8%以下,说明比特币市场价格变化快,随机性强;未来一周价格的预测效果一般,同时通过对比表明,越是临近的数据对于价格预测越有效。总而言之,通过4个比特币价格的预测模型比较表明,比特币输入数据样本越新鲜,比特币价格预测误差就越小;比特币市场发展趋势的预测周期越长,预测误差越大。

在比特币交易市场的预测方面,误差较大,由于市场具有高度的随机性和突变性,交易量很难预测。

总之,比特币价格波动大,存在较高的市场风险,本文针对比特币价格的预测研究对于比特币市场风险防范有预警效应,对相关决策和管理部门能够提供相关措施的量化指标,对比特币企业和投资者有一定的参考和实用价值。未来笔者将进一步研究比特币价格波动模型,并开发基于人工智能的比特币价格预测系统,以更好地为比特币行业以及用户提供风险防范预警。

主要参考文献:

梁秋,王凡彬,余龙秀等.基于小波分析的比特币价格预测[J].内江科技,2015(5).

刘力臻,王庆龙.基于模仿传染模型的比特币羊群效应分析[J].北京邮电大学学报(社会科学版),2015(2).

作者单位:武汉商贸职业学院电商管理学院,武汉430205