

# 基于投资者情感倾向的 P2P 市场成交量预测模型

张 帅<sup>1</sup> 傅湘玲<sup>1</sup> 后 羿<sup>2</sup>

(北京邮电大学软件学院 北京 100876)<sup>1</sup>

(中国华融资产管理股份有限公司上海自贸试验区分公司 上海 200002)<sup>2</sup>

**摘 要** 当前对 Peer-to-Peer 市场成交量的研究多种多样,但是常见方法只考虑了将投资者信息和市场信息作为特征,未考虑投资者情感变化与市场的关系。研究显示投资者的情感会对投资者的决策和行为产生深刻的影响。为此,以金融理论为基础,文中提出了基于投资者情感倾向预测 P2P 市场成交量的方法。首先以网贷之家的文本评论数据为研究对象,利用 TextCNN 模型对文本进行情感分类,得出情感倾向变化的时间序列,达到度量投资者情感变化趋势的目的;然后,通过格兰杰因果检验和皮尔逊相关系数验证投资者情感时间序列与成交量指数之间的关系;最终使用基于长短期记忆网络的预测模型预测 Peer-to-Peer 市场的成交量。实验结果表明,将情感特征加入到成交量预测模型能显著提高模型的预测能力。

**关键词** P2P 借贷,情感分类,自然语言处理,深度学习

**中图分类号** TP301 **文献标识码** A

## Prediction Model of P2P Trading Volume Based on Investor Sentiment

ZHANG Shuai<sup>1</sup> FU Xiang-ling<sup>1</sup> HOU Yi<sup>2</sup>

(School of Software, Beijing University of Posts and Telecommunications, Beijing 100876, China)<sup>1</sup>

(China Huarong Asset Management Co. Ltd., Shanghai Pilot Free Trade Zone Branch, Shanghai 200002, China)<sup>2</sup>

**Abstract** There are many kinds of studies on the trading volume of Peer-to-Peer market. However, the common methods only take investor and market information as characteristics, and donot consider the relationship between investor sentiment changes and the market. The research shows that investors' sentiments have a profound impact on their investment decisions and behaviors. Therefore, according to the financial theory, this paper proposed a method to predict the trading volume of Peer-to-Peer market based on investor's sentimental tendency. Firstly, the comments of Wang-DaiZhiJia is taken as the research object and applied TextCNN model for sentiment classification. The time series of sentiment tendency is obtained, so as to achieve the purpose of measuring the trend of investor sentiment. Secondly, it verifies the relationship between investor's emotion time series and trading volume index through Granger causality test and Pearson correlation coefficient. Finally, a predictive model based on long short term memory network is employed to predict the trading volume of the Peer-to-Peer market. The experimental results show that by adding sentimental features to the trading volume prediction model, the predictive ability of the model is improved significantly.

**Keywords** Peer-to-Peer lending, Sentiment classification, Natural language processing, Deep learning

## 1 引言

P2P 网络借贷是指贷款人与投资人之间通过网络借贷平台而不是传统银行等金融机构产生的借贷行为<sup>[1]</sup>。它是一种新兴的金融信贷模式,具有借贷双方直接交易、无中介,非专家参与,借贷金额小、期限短、无抵押等特点<sup>[2]</sup>。中国在 2007 年成立了第一家网贷平台,自 2012 年开始蓬勃发展,2017 年中国国内 P2P 理财成交额达到 28048.39 亿元,比 2016 年增长 35.9%,网贷行业历史累计成交量突破 6 万亿元。同时,由于准入门槛较低,中国 P2P 市场竞争异常激烈。截至 2017 年 12 月底,正常运营的平台达到 1931 家,但由于停业、经营路

线、欺诈等问题而关闭的问题平台数达到 645 个。造成这些问题的原因多种多样,通常是由于企业低估市场风险和盲目扩张。尽管近年来出台的相关政策减少了这类问题的发生,但是仍无法避免此类事件的发生。

P2P 市场成交量可以反映 P2P 市场的繁荣程度。如果能够预测市场的成交量,就能够把握市场变化规律,为政府监管和企业经营提供理论依据。而 P2P 成交量的变化是由投资者决策造成的。我们从心理学研究中得知,除了信息之外,情感在人类决策中起着重要的作用<sup>[3]</sup>。Chi 等<sup>[4]</sup>发现情感变化对股票收益有至关重要的影响。情感变化与股票市场的波动有关,那么类似的情感变化也与 P2P 市场有关吗?因此,

本文受国家自然科学基金(91546121),国家重点研发计划(2017YFB0803300),国家社会科学基金的重大研究课题(16ZDA055)资助。

张 帅(1993-),男,硕士,主要研究方向为自然语言处理、机器学习;傅湘玲(1975-),女,博士,副教授,主要研究方向为文本挖掘、情感分析等,E-mail:fuxiangling@bupt.edu.cn(通信作者);后 羿(1989-),男,硕士,工程师,主要研究方向为金融大数据。

我们提出问题:投资者的情感变化是否可用于预测 P2P 市场成交量?

本文对投资者情感与 P2P 市场的关系做了研究,并利用获得的信息预测了 P2P 市场成交量的变化。

本文的主要贡献如下:

(1)构建情感分类数据集并验证情感变化与 P2P 市场成交量具有相关性。

(2)提出了一种基于长短期记忆网络的 P2P 市场成交量预测模型,来验证投资者情感能否影响 P2P 市场并预测 P2P 市场成交量。

本文第 2 节讨论了相关工作;第 3 节对问题进行定义,并概述本文框架;第 4 节介绍了模型使用的数据,包括如何获得、预处理和分析,提出了情感分类模型及分类结果;第 5 节和第 6 节分介绍了 P2P 市场成交量预测模型和实验结果分析;最后,总结全文。

## 2 相关工作

目前,许多学者做了以网络文本相关的研究。Oliveira 等<sup>[5]</sup>使用从微博中提取的情感和注意力指标(采用大型 Twitter 数据集)以及调查指数(AAII 和 II,USMC 和 Sentix)来预测股市行为。Ding 等<sup>[6]</sup>发现 Facebook 的“like”数量会影响票房表现。Si 等<sup>[7]</sup>提出了一种基于 Twitter 情感主题的技术来预测股票市场。黄润鹏等<sup>[8]</sup>也利用微博数据预测股市。Chi 等<sup>[9]</sup>发现投资者的情感变化会对中国股票收益产生相对较强的影响。Ranco 等<sup>[10]</sup>发现 Twitter 情绪会影响股票价格回报。

对 P2P 的研究目前主要集中在风险控制方面。Everett<sup>[11]</sup>发现私人贷款机构比公共贷款机构的套牢问题更为严重,而人际关系缓解了道德风险问题。李焰等<sup>[2]</sup>发现在 P2P 借贷关系中没有抵押机制的情况下,自愿性信息披露对投资人的决策会产生影响。Guo 等<sup>[12]</sup>建立了一种新的信用风险评估模型来评估和研究 P2P 信用风险。Ge 等<sup>[13]</sup>使用借款人自我披露的社交媒体信息来预测及防止借款人违约的发生。

P2P 市场基于网络信息,用户根据从互联网获得的信息做出投资决策。因此,使用用户评论分析用户情感变化和 P2P 市场的关系是非常有用的。近年来,中国的 P2P 市场蓬勃发展,每天都在互联网上产生大量文本数据,这有利于我们的研究。

本文研究采用格兰杰因果检验验证了投资者情感变化对 P2P 市场成交量的影响,并建立了用于预测 P2P 市场成交量的深度学习模型,对政府和企业具有重要的现实意义和参考价值。

## 3 研究框架

### 3.1 问题定义

针对第 2 节提出的问题,本文做出如下假设:情感变化的时间序列与 P2P 市场相关并可以用来预测 P2P 市场成交量。

为了验证这一假设,我们建立了情绪分类模型和 P2P 市场成交量预测模型。在基于 TextCNN 的情感分类模型中,

用标注过的文本向量作为输入,用情感倾向作为输出。

在 P2P 成交量预测中,我们用每 7 天的成交量、情感趋势等作为输入,并设置对照组来比较预测结果。表 1 中定义了本文中用到的符号。

表 1 符号定义

符号	含义
ND	每天的负面情感评论的百分比时间序列
TVI	每日成交量指数
VD	每日成交量指数的时间序列
L	表示长短期记忆网络的输入
Accuracy	准确率,表示对于给定的测试集,模型分类正确的样本占全部样本的百分比
PCC	皮尔逊相关系数
MSE	真实值和预测值的均方误差

### 3.2 研究框架图

本文具体研究方案如图 1 所示。

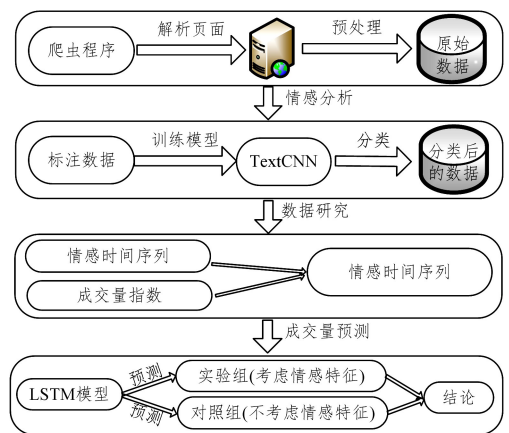


图 1 研究方案

本文方案的具体步骤如下。

1)数据采集和预处理:利用 Python 语言编写爬虫程序来抓取网贷之家所有 P2P 平台的评论数据和每日成交量指数(TVI),并将采集的原始数据进行清洗和分词处理,形成原始数据集。

2)情感分类:从原始数据集数据中选取部分数据进行正负两种情感的人工标注。使用标注数据训练 TextCNN 模型。使用训练好的模型将数据进行情感分类。

3)数据分析:将分类好的情感数据集按天计算每日负情感的评论占当日全部评论的比例。按照日期顺序生成时间序列。利用格兰杰因果关系检验情感变化对 P2P 市场的影响。

4)基于情感倾向的 P2P 市场成交量预测:将负情感时间序列作为特征预测成交量,并与对照组比较。

## 4 数据获取和情感分类

### 4.1 数据获取和预处理

P2P 平台的数据来源一般分为两类:新浪微博和百度贴吧等社交媒体平台,以及网贷资讯网站。我们在研究中发现前者有如下特点:

1)用户量大,用户群体涉及范围广,内容良莠不齐,含有大量无关内容。

2)微博或帖子的行文不规范,涉及的网络用语较多,并且

往往带有大量的表情、符号、链接和图片等,不利于分析数据。

3) 这些平台的数据存在大量的广告贴和炒作帖,并且文字包含相当多的无关内容。

P2P 资讯网站的大多数用户是投资者和借款人,他们会在咨询平台上评价和分享关于网贷企业的信息。基于以上因素我们选择网站作为数据来源。通过比较网贷天眼、网贷之家和贷罗盘等多家网站,最终选择了网贷之家<sup>1)</sup>的评论数据和日常公布的行业数据(成交量)作为研究对象。网贷之家是中国首家权威的 P2P 资讯平台,也是最大的网络借贷行业门户网站之一。它收集的网贷平台数据较为全面,并且数据更新及时,方便研究。它评级排名指标也成为投资者衡量平台综合实力的重要指标,影响着用户投资决策。

我们编写了 Python 爬虫程序来分析网页并爬取数据,最终共获得从 2013 年 2 月到 2018 年 4 月关于 6086 家平台的 212964 条评论数据。此外采集 2016 年 1 月 1 日至 2018 年 4 月 8 日期间每日的成交指数,将其记作 TVI。TVI 是反映 P2P 市场成交量整体变化的指标,是每日成分平台成交量的算术平均值。成分平台由专家选择的 20 家网贷平台组成,选择的原则为影响力大、代表性强、多样化和持续稳定。TVI 的具体公式如下:

$$TVI = \frac{\sum \text{成交量}}{n \times 10000} \quad (1)$$

其中,  $n=20$ 。

在爬取到全部数据后,对文本数据进行预处理,去除特殊字符和表情,然后对文本进行分词。分词后的数据集有 211664 条评论。分词工具使用的是百度云提供的词法分析的接口。百度云是百度提供的公有云平台,在国内非常受欢迎,对 NLP 任务提供了丰富的接口支持。在分词之后,获得了大小为 7971 的字典。

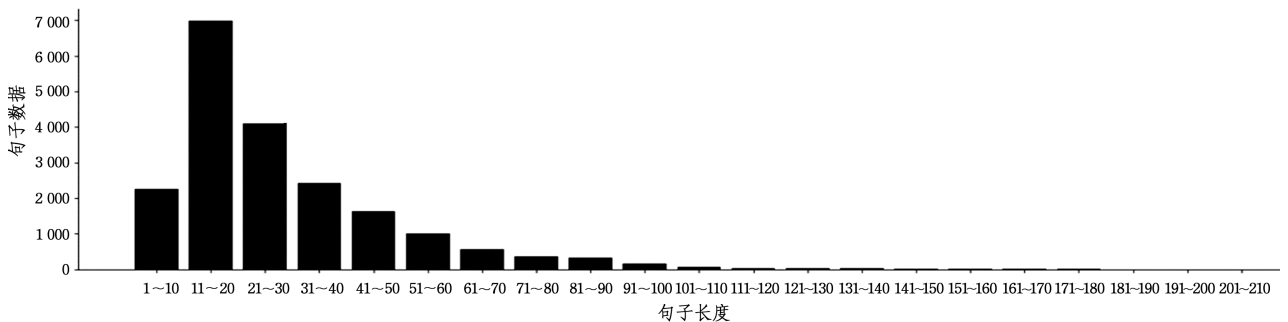


图 2 句子长度分布

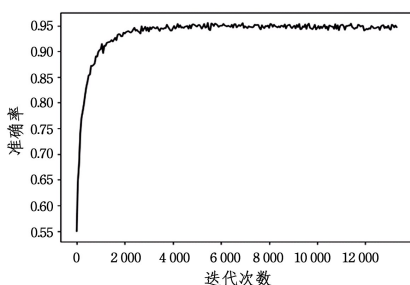


图 3 模型准确率

## 4.2 情感分类

### 4.2.1 标注数据

本文将评论分为 Negative 和 Positive 两类。将有负情感倾向的评论标注为 Negative,负情感倾向的类型包括对 P2P 企业服务、网站、客服的投诉的评论,对 P2P 企业业务、利息质疑或者不满意的评论,认为企业存在诈骗风险的评论等,将其余的数据标注为 Positive。

本文从分词后的数据集中随机选择了 20000 条评论进行标注。数据集分布如表 2 所列。

表 2 标注结果

Positive 数量	Negative 数量
13268	6732

### 4.2.2 TextCNN 模型

本文使用 TextCNN 模型实现情感分类。TextCNN 是利用卷积神经网络对文本进行分类的算法,由多个卷积核组成的卷积层、池化层和全连接层拼接而成。它对文本有非常出色分类效果,这个在以往的实验中得到验证<sup>[14]</sup>。

我们使用 90% 的标注数据集作为训练集,句子长度的分布如图 2 所示,所有句子的最大长度小于 220。因此,我们在模型中选择 220 作为句子的长度,长度不足的补 0。另外词向量为 100,我们通过 Word2Vec 预训练词向量,从而提高了最终模型的准确率。准确率表示对于给定的测试集中,模型分类正确的样本占全部样本的百分比。准确率的定义如式(2)所示。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

其中,  $TP$  和  $TN$  分别表示被正确分类的正类和负类;  $FP$  和  $FN$  分别表示被错误分类的正类和负类。参数调整后,模型的准确率如图 3 所示,准确率、召回率和 F1 值如表 3 所列。

表 3 模型的准确率、召回率和 F1 值

Precision	Recall	F1
0.943	0.911	0.927

情感分类结果如表 4 所列。

表 4 情感分类结果

Negative 评论数量	Positive 评论数量
82011	129653

### 4.2.3 情感和 TVI 的相关性检验

实验将投资者情感只分为正负两类,每日两类情感占比

<sup>1)</sup> <https://www.wdzt.com>

的和为 1,因此任选一类情感研究都能反映情感与 P2P 市场的关系,所以本文选择负情感作为研究对象。我们计算每日负面评论数量的百分比,并按日期顺序生成时间序列,记作

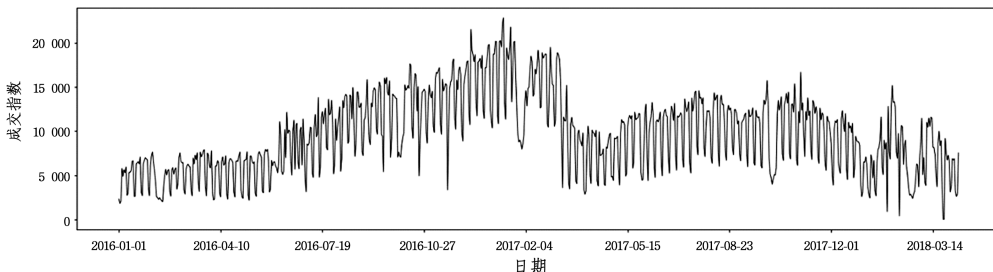


图 4 每日成交量趋势图

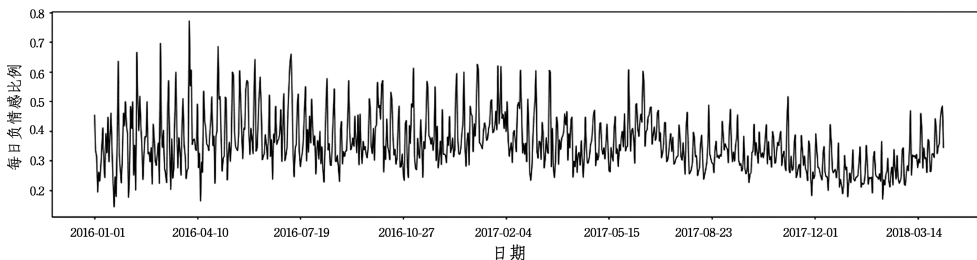


图 5 每日负情感占比的趋势图

从数据中可以看出,成交量指数呈现周期性波动。因此,在成交量预测模型中,需要考虑周期性特征。

为了比较  $ND$  和  $VD$ ,使用  $z$ -score 方法来归一化实验数据。 $\bar{X}$  是  $X$  的均值, $\sigma$  是  $X$  的标准差。计算  $Z$  值的公式如下:

$$Z = \frac{X - \bar{X}}{\sigma} \quad (3)$$

处理后的数据符合标准正态分布,平均值为 0,标准偏差为 1。数据标准化后我们发现, $VD$  和  $ND$  的 Pearson 系数为  $-0.2097$ , $p$  值小于 1%。这意味着成交量随着负面情感的比例的增加而下降,这是符合我们的直觉的。当然,这种关系需要经过严格的检验。

#### 4.2.4 格兰杰因果检验

格兰杰因果关系检验是一种检验一组时间序列是否可被用来预测另一组时间序列的假设检定<sup>[15]</sup>。如果时间序列  $X$  是时间序列  $Y$  的格兰杰原因,那么就说明统计意义上它们具有因果关系。

格兰杰因果检验的前提是时间序列的平稳性,因此首先需要对本实验的时间序列进行单位根检验。对时间序列单位根的检验就是对时间序列平稳性的检验。本实验使用 R 语言进行计算,单位根检验应用的方法是 Augmented Dickey-Fuller test(ADF),原假设是该时间序列有单位根。计算结果如表 5 所列。

表 5 单位根检验的结果

名称	$VD$	$ND$
T 值	-10.8384	-16.9337

T 值均小于 0.01 的可能性下的临界值为  $-3.96$ 。因此  $VD$  和  $ND$  没有单位根,均是平稳时间序列。然后对它们做格兰杰因果关系检验。在格兰杰因果检验中, $p$  值是统计模型的概率。表 6 列出格兰杰因果检验的结果。

$ND$ 。将  $TVI$  的时间序列记作  $VD$ 。

$VD$  的趋势变化如图 4 所示,相同时间跨度的  $ND$  趋势如图 5 所示。

表 6 格兰杰因果检验的结果

滞后天数	1	2	3	4	5	6	7
$p$ 值	$3.916 \times 10^{-9}$	$2.261 \times 10^{-16}$	$1.227 \times 10^{-13}$	$5.207 \times 10^{-14}$	$4.827 \times 10^{-12}$	$9.172 \times 10^{-7}$	$2.206 \times 10^{-8}$

$p$  值均小于 1%,拒绝原假设,表明负情感变化和成交量的变化在统计上存在因果关系,这也证明可以使用情感变化预测成交量。

## 5 基于情感时间序列的 P2P 成交量预测模型

长短期记忆模型(LSTM)由 Hochreiter 和 Schmidhuber 在 1997 年提出<sup>[16]</sup>,并于 2000 年由 Gers 的团队进行改进<sup>[17]</sup>。LSTM 适合于处理和预测时间序列事件。基于 LSTM 的模型可以做文档摘要、图像识别、股票预测等。本文基于 LSTM 建立深度学习模型,利用情感倾向时间序列预测  $TVI$ 。

首先,我们观察到  $VD$  呈周期性波动。这是因为网贷平台通常在工作日发标(指一个借款人或投资人发出贷款请求或投资请求,即创建一个借款项目或投资项目的行为),在周末发标较少,所以成交量指数会呈现周期性波动。本文将成交量指数的时间特征表示为 7 维的 one-hot 向量并加入到预测模型中(如第 0 维为 1 表示当前是周一,第一维为 1 表示当前是周二,以此类推),记作  $WeekDay$ 。为了衡量情感变化对  $TVI$  的影响,我们设立了对照组进行实验。由于  $VD$  的波动周期为 7 天,因此使用每 7 天的  $VD$  和  $WeekDay$  作为输入。 $t$  表示想要预测的那一天, $V_{D_{t-n}}$  和  $WeekDay_{t-n}$  分别代表  $t$  的  $n$  天前的  $VD$  和  $WeekDay$ 。对照组 1 被记作  $L_1$ :

$$L_1 = \{(VD_{t-1}, WeekDay_{t-1}), (VD_{t-2}, WeekDay_{t-2}), \dots, (VD_{t-7}, WeekDay_{t-7})\}$$

另外,为了验证将  $WeekDay$  加入到模型的必要性,设置了对照组 2,其被记作  $L_2$ :

$$L_2 = \{(VD_{t-1}), (VD_{t-2}), \dots, (VD_{t-7})\}$$

实验组加入情感变化特征, $ND_{t-n}$  表示  $t$  的  $n$  天前的负情

感比例,实验组输入记作  $L_3$ :

$$L_3 = \{(VD_{t-1}, ND_{t-1}, WeekDay_{t-1}), (VD_{t-2}, ND_{t-2}, WeekDay_{t-2}), \dots, (VD_{t-7}, ND_{t-7}, WeekDay_{t-7})\}$$

LSTM模型如图6所示,其中  $I$  和  $O$  表示 LSTM 每个时间点的输入和输出,OUTPUT 为最终输出。

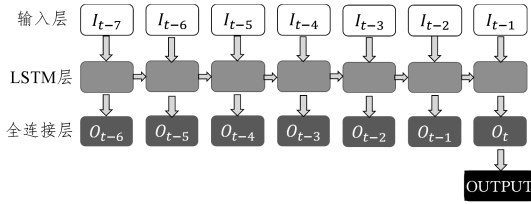


图6 LSTM模型的结构

## 6 实验结果

### 6.1 参数调整和结果分析

最初,本文的训练模型没有考虑  $WeekDay$  特征。在实验组中,成交指数的实际值和模型的预测值的皮尔逊系数仅为 0.70。添加  $WeekDay$  特征后,两者的拟合效果更显著。我们

使用随机梯度下降(SGD)和 Adam 作为优化函数。实验结果表明 Adam 可以达到更好的拟合结果。最后,我们选择将 0.001 作为学习率。表7列出模型的超参数。

表7 P2P成交量预测模型的超参数

参数	值	描述
N epoch	800	使用训练集全部样本训练的次数
Cell Size	10	模型隐藏层状态变量的维度
Learning Rate	0.001	学习步长
Time Steps	7	LSTM 时间窗口大小
Batch Size	20	一次迭代使用的样本数量

因为获得的 TVI 的时间跨度在 2016 年 1 月 1 日至 2018 年 4 月 8 日之间,所以我们也截取了同样的时间跨度  $ND$ (在此时间跨度的评论数为 184 203),因此按照每 7 天作为一个样本的方式处理后共获得了 819 个样本。本文使用前 600 组样本作为训练数据集,最后 219 组样本作为验证数据集。

我们使用训练数据集进行模型训练。模型训练完成后,使用验证数据集进行预测。对照组的结果如图7和图8所示,实验组的预测结果如图9所示。其中实线表示成交量实际值,虚线表示模型预测值。

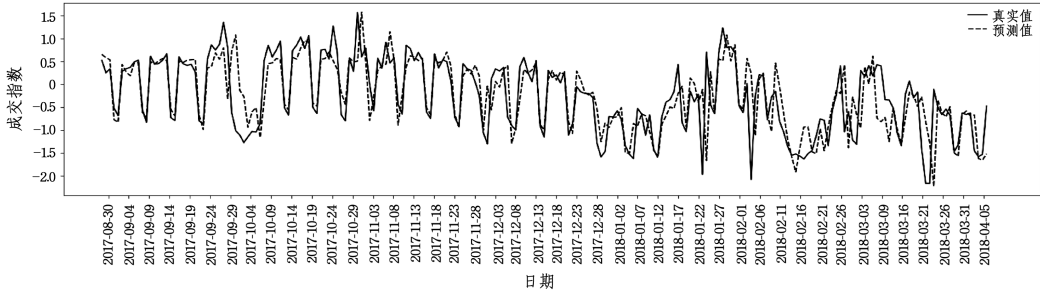


图7 对照组1的实验结果

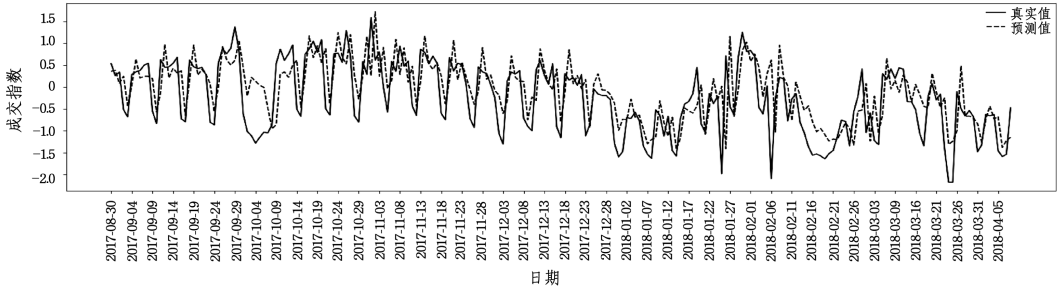


图8 对照组2的实验结果

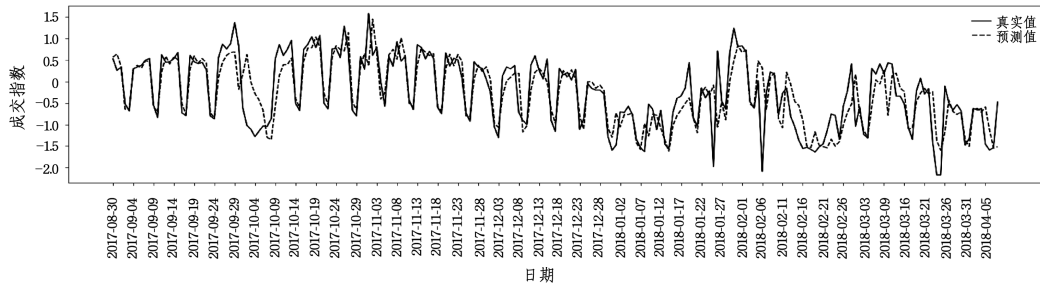


图9 实验组的实验结果

我们使用 Pearson 相关系数和均方误差(MSE)来衡量预测效果。Pearson 相关系数(PCC)是两个变量  $X$  和  $Y$  之间的线性相关性的度量。PCC 由式(4)定义,其中  $COV$  是协方差,  $\sigma$  是标准偏差。式(5)是 MSE 的定义,其中  $\hat{Y}$  是  $n$  个预测的向量,并且  $Y$  是被预测的变量的观测值的向量。计算

结果如表8所列。

$$PCC = \frac{COV(X, Y)}{\sigma_X \sigma_Y} \quad (4)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (5)$$

表8 预测值和真实值的 Pearson 相关系数和均方误差的结果

组名	皮尔逊相关系数	均方误差
$L_1$	0.7355	0.3184
$L_2$	0.6575	0.4141
$L_3$	0.7994	0.2390

比较  $L_1$  和  $L_2$  可以看出,预测模型在加入 *WeekDay* 特征后获得了更好的预测结果。这意味着考虑模型周期性变化的这个特征是正确的。此外,我们发现模型加入 *ND* 后,  $L_3$  与  $L_1$  相比,预测值与实际值之间的相关性增加了 8.62%,而 *MSE* 减少了 21.7%。因此,加入 *ND* 特征后提高了预测 P2P 市场交易量的准确性,表明投资者情感变化可以影响和预测 P2P 市场成交量,验证了本文的假设和工作的可行性。

## 6.2 与其他模型的比较

为了验证模型的有效性和准确性,将预测模型与以下几个常见模型进行性能对比。本文分别尝试了深度神经网络(DNN)、线性回归、随机森林和 SVR。

DNN: DNN 是一种深度神经网络算法,是可以使用反向传播算法进行训练的判别模型,近几年在工业界和学术界广受关注。

线性回归(Linear regression): 线性回归是监督学习的一种,使用被称为线性回归方程的最小二乘函数对一个或多个独立变量与因变量之间的关系进行回归分析。

随机森林(Random forest): 随机森林是利用多棵树对样本进行训练并预测的一种分类器。它是集合学习的组合分类算法,可用于分类和回归。

SVR: SVM 是一种常见判别方法,是一种有监督模型,而支持向量回归(SVR)是支持向量机(SVM)的回归版本,主要是通过高维空间中构造线性决策函数来实现回归。

各模型的比较结果如表 9 所列。

表9 实验结果

模型	皮尔逊相关系数	均方误差
LSTM	0.7994	0.2390
DNN	0.6315	0.4236
Linear regression	0.7058	0.3302
Random forest	0.6447	0.4054
SVR	0.7113	0.3286

LSTM 是循环神经网络,能够在网络中传递先时刻的信息,并且所有时刻共享权重,减少了参数并降低了模型的复杂性,因此其更加适合做时间序列的预测。DNN, SVR 等模型无法记录时间序列之间的关系。LSTM 在预测时间序列方面比 DNN, SVR 和其他模型更有优势。表 9 中的结果表明, LSTM 比其他模型有更好的预测效果。这表明使用基于 LSTM 的成交量预测模型是正确的。

**结束语** P2P 市场的分析和研究是一个非常广泛的应用课题。将投资者情感变化用于研究成交量的变化是一个不错的选择,也为政府和企业提供了重要的参考依据。

本文研究了投资者情感变化与 P2P 市场成交量的相关性,并使用深度学习预测 P2P 平台成交量,实验证明在加入情感特征后能显著提升模型的预测精度。

本文的不足与研究展望:

1) 本文使用 TextCNN 作为情感分类的模型,虽然经过模型调优,但是 Precision 和 Recall 的值仍有优化空间。同时,本实验只把情感分为正负两类,而人的情感是多维的,可

以进一步对情感详细分类。

2) 受数据集规模限制,成交量预测模型虽然在加入情感指标后,预测值更加接近实际值,但对峰值和谷值的变化仍不能精确预测。P2P 市场成交量的变化与诸多因素有关,为了进一步提高模型的准确率,可以考虑把更多的特征融入预测模型。

## 参 考 文 献

- [1] BACHMANN A, BECKER A, BUERCKNER D, et al. Online peer-to-peer lending-A literature review[J]. Journal of Internet Banking & Commerce, 2011, 16(2), Special section p1.
- [2] 李焰, 高弋君, 李珍妮, 等. 借款人描述性信息对投资人决策的影响——基于 P2P 网络借贷平台的分析[J]. 经济研究, 2014 (S1): 143-155.
- [3] BOLLEN J, MAO H. Twitter Mood as a Stock Market Predictor [J]. Computer, 2011, 44(10): 91-94.
- [4] CHI L X, ZHANG G S, ZHUANG X T, et al. A Study on Investor Sentiment Index and Chinese Stock Market Based on the Extended Kalman Filter Method [J]. Journal of Industrial Engineering & Engineering Management, 2012(3): 122-128.
- [5] OLIVEIRA N, CORTEZ P, AREAL N. The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices[J]. Expert Systems with Applications, 2017, 73: 125-144.
- [6] DING C, CHENG H K, DUAN Y, et al. The power of the "like" button: The impact of social media on box office [J]. Decision Support Systems, 2017, 94: 77-84.
- [7] SI J, MUKHERJEE A, LIU B, et al. Exploiting Topic based Twitter Sentiment for Stock Prediction [C] // ACL. 2013: 1779-1787.
- [8] 黄润鹏, 左文明, 毕凌燕. 基于微博情绪信息的股票市场预测 [J]. 管理工程学报, 2015, 29(1): 47-52.
- [9] CHI L X, ZHUANG X T, SONG D L. Investor sentiment in the Chinese stock market: an empirical analysis [J]. Applied Economics Letters, 2012, 19(4): 345-348.
- [10] RANCO G, ALEKSOVSKI D, CALDARELLI G, et al. The Effects of Twitter Sentiment on Stock Price Returns [J]. PLOS ONE, 2015, 10(9): e0138441.
- [11] EVERETT C R. Group Membership, Relationship Banking and Loan Default Risk: The Case of Online Social Lending [J]. Social Science Electronic Publishing, 2015, 7(2): 15-54.
- [12] GUO Y, ZHOU W, LUO C, et al. Instance-based credit risk assessment for investment decisions in P2P lending [J]. European Journal of Operational Research, 2015, 249(2): 417-426.
- [13] GE R, FENG J, GU B, et al. Predicting and Deterring Default with Social Media Information in Peer-to-Peer Lending [J]. Journal of Management Information Systems, 2017, 34(2): 401-424.
- [14] KIM Y. Convolutional Neural Networks for Sentence Classification [J]. arXiv: 1408. 5882.
- [15] GRANGER C W J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods [J]. Econometrica, 1969, 37(3): 424-438.
- [16] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [17] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to Forget: Continual Prediction with LSTM [M]. Istituto Dalle Molle Di Studi Sull'Intelligenza Artificiale, 1999.