

基于外汇舆情的人民币汇率波动预测研究

成舟¹ 余峰¹ 过弋^{2,3,4} 王志宏²

(中汇信息技术(上海)有限公司 上海 201203)¹ (华东理工大学信息科学与工程学院 上海 200237)²
(大数据流通与交易技术国家工程实验室 上海 200237)³
(石河子大学信息科学与技术学院 新疆 石河子 832003)⁴

摘要 舆情与金融市场波动之间的联系,对金融市场的监控、分析和异常发现有着重要的作用。外汇市场中,由于舆情的多样性和人民币汇率变化的复杂性,更好地量化舆情对汇率的影响对于实现人民币汇率的监测和分析有着重要的现实意义。首先对外汇舆情数据进行噪声过滤、分词等预处理,并基于汇率领域知识构建人民币汇率波动预测的特征,然后综合舆情的时效性和领域专家的知识设计了一种新的舆情对人民币汇率的影响力模型,并在此基础上实现了人民币汇率波动预测模型。实验结果表明,文中设计与实现的预测模型可以有效地对人民币汇率进行波动预测。

关键词 外汇市场,舆情监测,人民币汇率,汇率波动预测

中图分类号 TP181 文献标识码 A

Research on Volatility Forecasting of RMB Exchange Rate Based on Public Opinion

CHENG Zhou¹ YU Zheng¹ GUO Yi^{2,3,4} WANG Zhi-hong²

(R&D Dept. II, CFETS Information Technology (Shanghai) Co., Ltd, Shanghai 201203, China)¹

(School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China)²

(Business Intelligence and Visualization Research Center, National Engineering Laboratory for Big Data
Distribution and Exchange Technologies, Shanghai 200237, China)³

(School of Information Science and Technology, Shihezi University, Shihezi, Xinjiang 832003, China)⁴

Abstract Public opinion has an impact on financial volatility, which plays an essential role in monitoring, analysis and anomaly detection for financial market. Due to the diversity of public opinion and the complexity of RMB exchange rate, how to quantify the impact of public opinion in a better manner has an important industrial significance for realizing the monitoring and analysis of the RMB exchange rate. This paper firstly performed pre-processing for public news of the foreign market, such as noise filtering, word segmentation. Meanwhile, it constructed a series of features for volatility forecasting of RMB exchange rate based on the domain knowledge of foreign exchange rate. Moreover, a novel influence model was proposed to represent the relationship between public opinion and RMB exchange rate. Finally, the volatility forecasting model is realized for RMB exchange rate on the real dataset. The experimental results testify that the proposed method can effectively forecast the volatility of RMB exchange rate.

Keywords Foreign market, Public opinion monitoring, RMB exchange rate, Exchange rate volatility forecasting

1 引言

金融市场中,除了股票指数,汇率同样受到政府决策者和投资者的重点关注。外汇汇率(Foreign Exchange Rate)是将一个国家的货币折算为另一个国家货币时的换算比值,是两种货币之间的价值体现^[1]。相比于股票、现货等交易,外汇交易更具复杂性、多样性和不确定性,影响因素也更为众多。在更为自由和市场化了的人民币外汇市场中,网络舆情作为一种重要的信息载体,已成为外汇市场运行的重要风向标。在外汇市场中,民众通过互联网对自己所关心或与自身权益紧密相关的公共事件、社会现状等做出的主观反映,是民众多种态度和意见交互的综合表现,其信息量之大,影响范围之广,即时性之强,是交易员进行交易判断和操作的重要参考。

同时,各类舆情信息对市场的影响也是监管机构和交易中心重点关注的对象。作为境内外汇市场最为重要的货币对交易,人民币对美元交易是流动性最好、参与者最多、价格最具弹性、交易量最大的市场,互联网舆情也对人民币汇率价格的波动有着重要的影响。

传统的汇率预测研究^[2-4],更多地基于以数字形式存在的汇率历史数据进行分析,忽略了对非结构化的舆情信息的挖掘和分析。本文主要针对外汇市场舆情数据对外汇市场行情的影响,综合舆情时效性和领域专家知识提出了网络舆情对人民币汇率的影响力模型,实现了基于网络舆情预测在岸人民币价格的方法。对人民币汇率波动的预测分析,有利于市场监测,更能为监管层进行舆论引导提供坚实的理论基础和现实意义。

本文第2节介绍汇率预测的研究现状;第3节介绍本文模型的设计与实现,包括原始数据预处理、预测模型设计等;第4节给出实验结果与模型表现分析;最后总结全文,并展望该领域未来的研究前景。

2 相关工作

汇率预测是指对汇率未来变化趋势进行预先判断。由于汇率的异常变化会给各国货币政策和外汇管理带来严重干扰,扰乱正常经济秩序,甚至会激发货币危机,因此汇率预测已成为金融时间序列预测领域中的主要研究内容之一。

目前已有许多汇率预测模型被提出,并在汇率预测领域取得了一定的成效。其中,典型的是基于统计学的预测模型,如回归模型(Autoregressive, AR)^[5]、自回归移动平均模型(Autoregressive Moving Average, ARMA)^[6]以及广义自回归异方差模型(Generalized Autoregressive Conditional Heteroscedasticity, GARCH)^[7-8]等。其中,杨梦昕等^[6]利用自相关、偏相关系数图等判断原序列是非平稳时间序列以及一阶差分后的序列是平稳时间序列,从而结合SIC等指标选择最优的ARIMA(1,1,2)模型进行汇率预测,取得了较好的预测效果。朱可飞^[8]通过将小波与GARCH模型结合,获得了比单独使用GRACH模型更精确的汇率预测结果,且小波结合的CARCH模型更具广泛性。但是,传统的统计学模型依赖于某些理想的前提假设,使其自身具有一定的局限性,因此越来越多的学者尝试引入机器学习方法来解决汇率预测问题。Shin等^[9]通过使用遗传算法进行小波变换的阈值选择,并为神经网络提供丰富的特征,从而将遗传算法、小波变换和神经网络3种方法相结合来预测韩元和美元之间的汇率;Cao等^[10]使用支持向量机(SVM)方法对美元和英镑的汇率进行预测,取得了不错的预测效果;Panda等^[11]和Majih等^[12]则分别提出了使用神经网络模型进行汇率预测,前者实现了卢布和美元间的汇率预测,后者通过简化神经网络预测模型的复杂度,在3种货币(卢布、英镑和日元)和美元间的汇率预测都表现出了更加优异的预测性能;王晴等^[13]为了提高汇率预测的精确度,对传统的KNN算法进行了改进,提出了基于加权KNN的汇率预测模型,在人民币对美元的汇率预测效果上得到了进一步的提升。从上述研究中可以发现,现有关于汇率预测的模型算法研究更多的是对以数字化形式存在的汇率历史数据进行分析,忽略了对非结构化的外汇舆情信息的挖掘和分析。

而随着信息技术的迅猛发展,网络平台已成为市场参与者获取信息的重要渠道。利用人工智能技术分析和挖掘网络舆情,从而对市场行情进行初步预判,已受到越来越多的学者和业界的关注。Wutrich等^[14]认为财经新闻中的文章内容除了传达市场表现外,还包含造成这种现象存在的潜在原因,他们基于KNN聚类算法和神经网络算法分析新闻舆情数据,并对全球主要股市指数的当天收盘价进行预测。Lavrenko等^[15]利用自然语言模型,结合新闻文字中的特点,分析出最可能影响未来股价趋势的新闻信息,并设计出一个预测股价趋势的系统。赵丽丽等^[16]将金融财经新闻内容转化为股价影响因子,并利用多元回归分析网络舆情对中国股市的影响。金雪军等^[17]采集了东方财富论坛上一年内的580万条帖子,通过分析投资者在平台上发表的意见,建立了看涨指数与趋

同指数。研究结果发现,看涨指数与股票收益率呈正向相关,且看涨指数对第二天的收益率具有预测作用;意见趋同指数对成交量有显著影响,意见趋同程度越低,股票的交易量越大。由此可见,网络舆情数据对金融市场行情的预测具有一定的参考价值,然而现有研究中对外汇舆情数据进行自动分析挖掘从而实现人民币汇率预测的研究还较少;且目前来说,外汇舆情对汇率行情的影响主要依靠监测人员对一段时间内的市场舆情进行人工分析来判断,这就需要舆情监测人员有较强的专业知识来分析各种舆情信息的影响力,并综合所有种类舆情信息进行市场行情预判。这种人工收集和主观判断的方式很容易滞后和遗漏重要的市场信息,从而导致对市场行情做出不准确、不客观、不全面的分析和判断。本文研究主要面向外汇舆情数据,探究其对人民币汇率的影响。为了更好地对外汇舆情数据进行表示和量化,本文首先基于汇率领域知识构建了一组人民币汇率波动预测特征,然后综合舆情时效性和领域专家知识提出了舆情对人民币汇率的影响力模型,最后利用机器学习算法实现了人民币汇率波动预测模型。

3 汇率波动预测模型的设计与实现

本文面向外汇舆情数据,探究其对人民币汇率的波动影响。整体研究框架如图1所示,主要包括以下4个阶段:1)外汇数据采集与预处理,包括外汇舆情数据和外汇行情数据;2)人民币汇率波动预测特征构建及表示;3)人民币汇率波动预测模型设计;4)模型评估与验证。

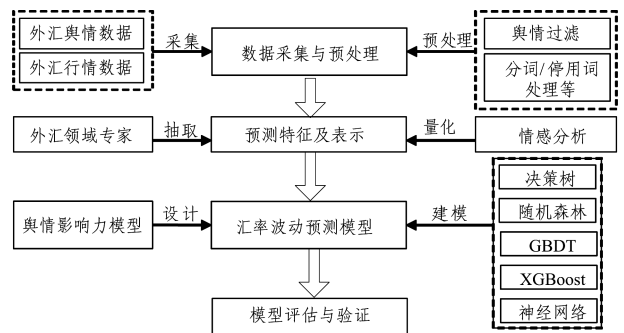


图1 人民币汇率波动预测模型的整体研究框架

3.1 外汇数据采集与预处理

本文的外汇数据主要包括外汇舆情数据和外汇行情数据。其中,外汇舆情数据是指有关外汇交易、外汇政策、外汇行情的新闻快讯类信息,主要用于人民币汇率波动预测特征的表示以及模型的训练、验证和评估;外汇行情数据是指外汇平台上的行情价格(本文只关注在岸人民币的行情),主要用于训练数据中汇率波动的评测标准。

为保证采集数据的整齐性和可用性,需要对外汇舆情数据做以下预处理。

(1)舆情过滤:主要是指过滤外汇舆情数据中所包含的图片、外链等非有效文本数据;

(2)来源删除:删除外汇舆情数据末尾的资讯来源,如“(彭博)”“(新华社)”等;

(3)分句、分词和停用词处理:对每一个时间点的舆情数据,使用自然语言处理工具进行分句、分词,并删除停用词,如“的”“进而”等。

3.2 预测特征构建及表示

原始舆情文本数据需要转换成模型训练和测试可用的数

值化特征数据,本节将重点介绍文中人民币汇率波动预测过程中使用的所有特征及其表示方式。

根据领域专家知识,影响在岸人民币价格涨跌的关键因素包括金融市场、宏观经济、央行政策、宏观政治、黑天鹅事件^[18]以及其他相关的舆情信息。为了选取有效特征,本文首先基于外汇专家领域知识获得了影响在岸人民币价格的舆情特征词,然后基于分类、分级原则对舆情特征词进行组织,如表 1 所列。基于舆情特征词的分类特点,对每个特征词都指定了唯一标识,标识使用 4 位大写英文字母,每个低层(第 n 层)特征都可与前 $n-1$ 位标注相同的特征组合成为较高级别特征,以用于特征合并。格式化后的基础特征记录共计 89 条,如表 2 所列。

表 1 外汇舆情特征层次分类示例

一级	二级	三级	四级
金融市场数据 (AA)	金融	贵金属 (AAA)	黄金、白银
		原油 (AAB)	布伦特原油、WTI 原油、上期所原油期货
		股市 (AAC)	标普 500、纳斯达克、道琼斯、恒生、沪指、深成指
		国债 (AAD)	美国 10 年期国债、中国 10 年期国债
		汇市 (AAE)	美元指数、欧元/美元、美元/日元、英镑/美元、离岸人民币、在岸人民币、人民币中间价
舆情数据 (A)	宏观	G7 国家相关经济数据 (ABA)	美国 GDP、美国季调后非农就业人口、美国当周初请失业金人数、美国 ADP 就业人数、美国 Markit 制造业 PMI 终值、美国 ISM 制造业 PMI、美国耐用品订单月率、美国 PCE 物价指数、美国核心 PPI、欧元区 CPI、欧元区 PPI、欧元区零售销售月率、欧元区 GDP、欧元区贸易帐
		国内宏观经济数据 (ABB)	财新制造业 PMI、中国外汇储备、中国贸易帐、中国 CPI、中国 PPI
		离岸人民币存款利率报价 (ACA)	—
离岸人民币数据 (AC)	CNH Hibor 利率 (ACB)	—	—

表 2 部分基础特征示例

特征标识	特征名称
AAAA	黄金
AABB	WTI 原油
AABD	OPEC
AACA	标普 500

经过上述处理,可以获得与人民币汇率波动预测相关的特征集合。在进行特征表示之前,我们首先基于舆情特征词对初始舆情数据进行了过滤,即根据舆情数据预处理的结果,将每条新闻数据的分析结果集与特征关键词列表进行比对(考虑同义词和缩写词的影响),删除与在岸人民币汇率预测不相关的新闻舆情数据。对于外汇舆情特征的表示,本文对特征词所在文本进行情感分析,并采用情感极性对特征进行量化。情感分析指的是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程^[19]。由于不同领域情感的表达具有差异性,通用的情感分析工具并不适用于外汇领域,在分析外汇市场舆情的情感时,需要构建外汇市场特有的情感分析工具。本文基于人工整理的外汇市场情感词典计算特征的情感值,并使用情感评价三元组模型(评价对象,情感词,程度词)来标定每个特征关键词,从而实现对外汇市场舆情的情感

极性类别(积极、消极、中性)的判断。基于此,对过滤后的舆情数据逐一进行情感分析,得出每条新闻针对 89 个特征关键词的情感极性分类结果及情感值。其中,情感极性分类为正向的情感值为正值,情感极性分类为负向的情感值为负值,中性的舆情情感值为 0。例如,对于舆情数据“【离岸人民币盘中跌破 6.70 元创近半年新低】离岸人民币(CNH)兑美元盘中跳水,从 6.6600 元水平附近跌至 6.7021 元,10 多分钟内下跌逾 400 点;目前跌幅有所收窄,报 6.6878 元。离岸人民币(CNY)兑美元盘中也一并跳水,10 多分钟内下跌约 100 点,一度至 6.6539 元。”计算得到的特征向量为: $[0, 0 \cdots 0, -0.3876, 0 \cdots 0, 0]$ 。其中,特征向量中的每个值对应一个特征,值为 0 的特征表示本条舆情与该值对应的特征无关或影响为中性,不为 0 的特征值表示对应特征在舆情中表现的情感值。“-0.3876”对应的特征是“离岸人民币”,负值表示了该舆情的情感分析结果为消极,与舆情实际内容“离岸人民币(CNH)兑美元盘中跳水”相符。

3.3 人民币汇率波动预测

本节将重点介绍外汇舆情对人民币汇率的影响力模型的设计与实现,并基于此模型给出人民币汇率波动的定义。

3.3.1 外汇舆情影响力模型

为了确定舆情发生后对在岸人民币价格的影响随着时间的变动,本文结合舆情时效性和领域专家知识设计了外汇舆情影响汇率的时间模型。

假设时间段(T_j 到 T_k)内的舆情对 T_k 点汇率的影响为 D_{jk} 。一般可知,距离目标价格时刻点越近的舆情对价格变动的的影响越大。另外,由于外汇舆情对外汇行情的影响具有一定的时滞性,即交易人员从看到外汇舆情到对行情产生影响存在一定的时间差,因此本文设计了如下外汇舆情影响力模型:

$$f(t) = \begin{cases} \frac{1}{T_p}t, & 0 \leq t \leq T_p \\ \frac{T_N}{T_N - T_p} - \frac{1}{T_N - T_p}t, & T_p \leq t \leq T_N \end{cases} \quad (1)$$

其中, t 表示时间段内的时刻点, T_p 表示舆情对价格影响达到最大的时刻点, T_N 表示舆情对价格影响趋于 0 的时刻点,舆情影响力 $f(t)$ 与时间 t 的关系如图 2 所示。

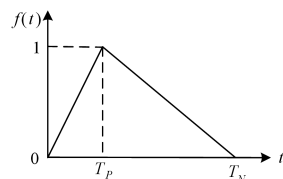


图 2 舆情影响力与时间的关系

由式(1)可知, D_{jk} 是 $f(t)$ 在 t 上的积分($t \in [j, k)$),即:

$$D_{jk} = \int_j^k f(t) \quad (2)$$

对同一时间间隔内的多条舆情,需要对其影响进行汇总,以确定对汇率的综合影响。本文采用对多条舆情的影响进行累加的方式对各条舆情对应特征向量的各列求和,以获得最终的影响力特征向量,即 $\sum_j^k D_{jk}$ 。

3.3.2 人民币汇率波动定义

预测 T_k 时刻的汇率,需要知道 T_j 时刻的汇率和 T_j 到

T_k 这段时间内的舆情影响汇率变动的总和 $\sum_j^k D_{jk}$, 即:

$$P_k = P_j + g(\sum_j^k D_{jk}) \quad (3)$$

其中, P_j 和 P_k 分别为 T_j 和 T_k 时刻的汇率, $g(D)$ 是舆情影响汇率变动的模型, 将通过机器学习算法训练确定, 模型的选择和参数的优化将在实验部分具体讨论。结合式(3), 在时间段 $[T_j, T_k]$ 内, $g(\sum_j^k D_{jk})$ 表示舆情影响汇率波动的值, 定义如下:

$P_k - P_j > 0$, 表示人民币汇率上涨;

$P_k - P_j < 0$, 表示人民币汇率下跌;

$P_k - P_j = 0$, 表示人民币汇率无变化。

4 实验及结果

4.1 实验数据

本文采用的数据集包括外汇舆情数据集和外汇行情数据集两部分。其中, 外汇舆情数据集主要来源于华尔街见闻的外汇快讯, 采集了从 2016 年 6 月至 2017 年 12 月的实时快讯, 共计 49279 条记录。根据 2.2 小节的预测特征过滤筛选后的有效舆情数据共计 22058 条。华尔街见闻是中国排名第一的财经新闻发布平台, 会快速推送影响金融市场的重大新闻、全球财经资讯和市场观点。外汇舆情数据集的部分示例如表 3 所列。

表 3 外汇舆情数据集示例

时间	外汇舆情内容
2016/06/30 15:14:56	【离岸人民币盘中跌破 6.70 元创近半年新低】离岸人民币(CNH)兑美元盘中跳水, 从 6.6600 元水平附近跌至 6.7021 元, 10 多分钟内下跌逾 400 点, 目前跌幅有所收窄, 报 6.6878 元。离岸人民币(CNY)兑美元盘中也一并跳水, 10 多分钟内下跌约 100 点, 一度至 6.6539 元。
2016/06/30 15:23:22	新浪援引外媒称, 中国央行愿意让人民币汇率在今年降至 1 美元兑 6.8 元。
2016/06/30 17:07:58	BBH 高级副总裁 Marc Chandler: 英镑的走势和基本面没有关系, 反弹的原因是空头平仓以及机构投资者在月末和季末调整仓位。
2016/06/30 18:10:37	MBM 研究团队首席研究员 Richard Cox: 瑞士央行如市场预期维持了其扩张性的货币政策, 理事会主席 Thomas Jordan 也承认负利率政策还将维持一段时间。
2016/06/30 18:56:03	前伦敦市长、退欧派代表人物 Boris Johnson 称不会参加保守党领导人及首相竞选。该消息过后, 英镑/美元扩大涨幅至近 0.4%, 触及 1.3478。

外汇行情数据集来源于中国外汇交易中心发布的行情数据。为保证两个数据集的对应关系, 共抽取 2016 年 6 月至 2017 年 12 月的 USD/CNY 的交易行情数据, 共计 241792 条, 每条数据只包含时间和汇率, 格式化后的行情数据示例如表 4 所列。

表 4 历史行情数据示例

时间	汇率(USD/CNY)
2016/06/30 17:21:49	6.6470
2016/06/30 17:22:07	6.6472
2016/06/30 17:24:39	6.6460
2016/06/30 17:25:45	6.6459
2016/06/30 17:27:37	6.6453

本文获得的外汇舆情数据集和人民币对美元的外汇行情数据集的详细统计如表 5 所列。

表 5 数据集总览

数据集统计指标	本文数据集情况
数据集时间范围	2016.06—2017.12
有效行情记录数	241792
有效舆情记录数	22058

4.2 实验对比

4.2.1 评价标准

为保证实验的公平性, 所有模型使用相同的训练集和测试集, 并针对人民币汇率波动预测结果使用准确率(Acc)评价模型的性能。准确率表示预测正确的结果占样本总数的比值, 计算公式为:

$$Acc = \frac{\text{预测汇率波动正确的样本数}}{\text{样本总数}} \quad (4)$$

4.2.2 实验设计与结果分析

在模型训练过程中, 为了减少因汇率随机波动对训练模型的影响, 本文在时间维度上根据不同的采样频率对行情数据进行处理, 选取多个时间间隔进行采样, 采样时间间隔包括: 1 min, 5 min, 10 min, 15 min, 20 min, 30 min, 45 min, 60 min, 120 min, 240 min, 420 min, 840 min。

对于给定的数据集, 本文使用分类算法构建外汇舆情影响人民币汇率涨跌的预测模型。将数据集划分为训练集和测试集, 其中 80% 的数据作为训练集, 20% 的数据作为测试集。本文共使用 5 种模型进行人民币汇率波动预测, 分别是决策树、随机森林、GBDT、XGBoost 和神经网络。

5 种算法在相同的数据集上进行训练, 得到价格预测模型, 并通过模型进行测试, 评估得出模型的性能。本文在使用神经网络算法构建模型时, 使用 2 层隐藏层, 第一层网络含有 5 个神经元, 第二层网络使用 2 个神经元, 采用梯度下降法训练模型; 在使用其他几种算法构建模型时, 选用 80 个弱分类器, 学习率为 0.1, CART 分类树的最大深度为 3, 目标函数为 Logistics 损失函数。

5 种模型在同一测试集上对不同采样时间间隔的模型进行评估, 预测准确率如表 6 所列。

表 6 5 种模型的预测实验结果

间隔	神经网络	XGBoost	GBDT	决策树	随机森林
1	0.4896	0.4889	0.4891	0.4733	0.4790
5	0.5015	0.5036	0.4994	0.4879	0.4875
10	0.5170	0.5160	0.5038	0.4900	0.4925
15	0.4836	0.4887	0.5034	0.4894	0.4849
20	0.5201	0.5216	0.5231	0.5162	0.4824
20	0.5174	0.5124	0.5025	0.5204	0.4919
45	0.5452	0.5217	0.5056	0.5006	0.4932
60	0.5526	0.5336	0.5307	0.5015	0.4985
120	0.5733	0.5492	0.5286	0.5055	0.4989
240	0.5564	0.5451	0.5414	0.4586	0.4774
420	0.5667	0.5524	0.4857	0.5333	0.4957
840	0.4593	0.4889	0.4741	0.4963	0.4940

从实验结果可以看出, 随着采样时间间隔的增大, 模型的准确率先逐渐上升并在达到峰值后下降, 符合一般认知影响规律。总的来说, 对于神经网络、XGBoost、GBDT、决策树和随机森林 5 种算法, 模型的预测准确率依次降低, 其中随机森林效果最差。这是因为在随机森林、GBDT 和 XGBoost 这 3 种基于决策树改进的模型中, 只有随机森林模型是采用有放回地思想构造多棵决策树, 在基于外汇舆情进行人民币汇率预测的过程中, 对外汇舆情数据有放回地构造多棵决策树, 但

是由于外汇舆情数据具有时效性,在当前场景下重复利用数据进行决策树构建会引入更多对汇率波动无效的噪声数据,因此随机森林模型的预测效果最差。相对来说,XGBoost 和 GBDT 两个模型的预测准确率都有所提高,最大值分别为 55.24%(420 min)和 54.14%(240 min)。5 种模型中,神经网络模型在采样的时间间隔为 120 min 时,模型准确率达到最大,为 57.33%,这是因为神经网络模型可以获取到更高维空间的特征,因此神经网络模型预测结果的整体准确率较高,可以更加有效地预测出人民币汇率的波动。

另外,本文对预先选取的特征进行了重要程度评估,如图 3 所示。其中,AAAA(黄金价格走势)特征的重要程度最高,BAAA(美联储相关舆情)、BAAB(欧洲央行相关舆情)、AAEE(离岸人民币走势)、AABD(OPEC 相关舆情)也呈现出较高的重要性,模型评判结果与基于专家领域知识的预期相符。

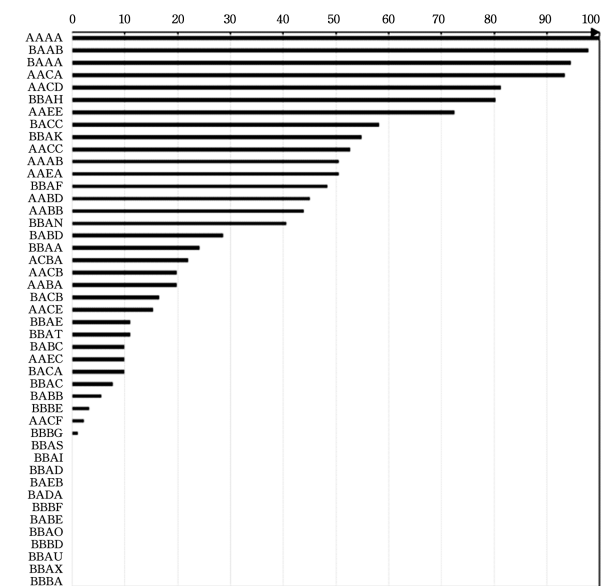


图 3 特征重要程度的对比

基于上述实验结果和分析,对模型的有效性可以得出以下结论。

(1)采用 5 种模型进行人民币汇率波动预测。实验结果表明,本文设计实现的基于外汇舆情的人民币汇率波动预测模型具有一定的可行性,其中神经网络算法在价格采样间隔为 120 min 时,准确率达到 57%。

(2)模型验证了外汇市场中对在岸人民币汇率走势影响较大的舆情要素,与实际市场监测中需要关注的要素基本符合,同时也说明本文提出的方法对外汇舆情数据建模预测分析具有一定的指导作用。

(3)对于黄金价格信息、OPEC 相关新闻(影响油价)等与外汇市场非直接因果关系的影响因素,模型可以揭示出其与汇市相关性较强且与汇市价格具有联动作用。

4.2.3 预测结果总结和分析

通过模型预测结果与在岸人民币实际汇率变动的比较,对汇率波动预测结果进行总结,如表 7 所列。

中国央行政策类的历史舆情数据较少,导致此类数据机器学习样本较少,无法准确预测。上述舆情信息在向市场发布后,人民币汇率向升值方向出现较大波动。而对于预测

模型,由于央行汇率政策发布频率较低,学习样本较为有限,模型预测结果显示为价格稳定,故不能进行准确预测。对于交易参与者来说,上述舆情信息的披露属于随机信息,难以进行直接分解,市场需要加深判断逻辑。

表 7 基于外汇舆情的人民币汇率波动预测结果分析

舆情类型	舆情样本示例	预测结果分析
发达市场金融舆情	1. 美元指数涨破 94 关口,创逾两周新高,日内涨约 0.6%。美元/加元涨至一个月高位 1.2778。美元/离岸人民币最高至 6.7012,日内累计涨逾百点。此前美国零售销售大幅好于预期。 2. 美国二季度实际 GDP 年化季环比初值 4.1%,创 2014 年三季度来最高增速,预期 4.2%,前值由 2%修正为 2.2%。	样本充足,预测准确率较高
中国央行政策类舆情	1. 外管局局长潘功胜表示“有信心保持人民币汇率在合理均衡水平上的基本稳定”;易纲行长接受采访时表示“将运用已有经验和充足的政策工具,发挥好宏观审慎政策的调节作用,保持人民币汇率在合理均衡水平上的基本稳定”。 2. 中国人民银行:将远期售汇业务的外汇风险准备金率上调为 20%。	样本较少,预测准确率稍低
突发类、新兴市场类舆情	1. 美国政府发布加征关税的商品清单,将对从中国进口的 500 亿美元商品征收 25% 关税。 2. 创业板创 2016 年 4 月以来最大周跌幅,进口博览会板块午后风云再起。	与近期时事相关,无历史信息参考,预测不理想

发达市场金融舆情信息预测成功率较高。发达市场新闻样本数较多,使得舆情判断和价格预测的结果更为准确。

突发类、新兴市场类舆情由于与近期事实相关,无历史记录参考,无关键词进行准确匹配,无法准确预测;并且在不同经济周期内,各个金融市场的联动关系也有所区别。机器学习在分析非外汇市场直接关联的舆情时,仍有巨大的学习空间。

结束语 本文面向外汇领域,探究外汇市场舆情对汇率行情的影响。具体来说,本文首先基于外汇专家领域知识构建了一系列影响人民币汇率波动的舆情特征词,然后综合舆情时效性和领域专家知识提出了网络舆情对人民币汇率的影响力模型,最后实现了基于网络舆情预测人民币汇率波动的方法。对人民币汇率波动的预测分析,不仅有利于市场监测,而且具有为监管层进行舆论引导的现实意义。在接下来的研究中,我们将加强数据采集的全面性和舆情特征词情感量化的领域针对性,进一步提高特征的表达能力和模型的可解释性,从而最终提高模型的预测准确率,更好地为市场监测和引导发挥作用。

参考文献

[1] 李小文. 汇率变化的经济影响分析[J]. 现代交际, 2011(4): 138-138.

[2] 谢赤,王丽平. 基于 MS-MIDAS 模型的人民币汇率预测研究 [EB/OL]. <http://www.paper.ed-u.cn/releasepaper/content/201706-153>.

[3] 孟建钊. 基于压缩感知的汇率预测与粗糙集加权的聚类研究 [D]. 广州:华南理工大学, 2017.

[4] 王晴,朱家明. KNN 算法在汇率预测中的应用及改进[J]. 兰州文理学院学报(自然科学版), 2017, 31(3): 27-31.

[5] 钱倩倩. 基于向量自回归模型的人民币汇率预测研究 [D]. 北京:北京外国语大学, 2015.

[6] YANG M X, GAO Z, FINANCE S O. The empirical analysis of

- RMB-USD exchange rate forecasting based on ARMA model [J]. *Journal of Science of Teachers College & University*, 2018 (4).
- [7] LUO X. The Research of the Fluctuation Rules of USD/RMB Exchange Rate Series Based on GARCH Model[J]. *Application of Statistics & Management*, 2009, 28(2): 295-300.
- [8] 朱可飞. 基于小波分析的人民币汇率预测方法研究[D]. 杭州: 浙江工商大学, 2014.
- [9] SHIN T, HAN I. Optimal Signal Multi-Resolution by Genetic Algorithms to Support Artificial Neural Network Models for Financial Forecasting[C]// *International Conference on Information Intelligence and Systems*, 1999. IEEE, 2000: 586-593.
- [10] CAO D Z, PANG S L, BAI Y H. Forecasting exchange rate using support vector machines[C]// *International Conference on Machine Learning and Cybernetics*. IEEE, 2005: 3448-34526.
- [11] HUANG W, LAI K K, NAKAMORI Y, et al. Forecasting foreign exchange rates with artificial neural networks: a review[J]. *International Journal of Information Technology & Decision Making*, 2004, 3(1): 145-165.
- [12] MAJHI R, PANDA G, SAHOO G. Efficient prediction of exchange rates with low complexity artificial neural network models[J]. *Expert Systems with Applications*, 2009, 36(1): 181-189.
- [13] 王晴, 朱家明. KNN算法在汇率预测中的应用及改进[J]. *兰州文理学院学报(自然科学版)*, 2017, 31(3): 27-31.
- [14] WUTHRICH B, PERMUNETILLEKE D, LEUNG S, et al. Daily prediction of major stock indices from textual www data[J]. *HKIE Transaction*, 1998, 5: 151-166.
- [15] LAVRENKO V, T SCHMILIM, LAWRIE D, et al. Mining of concurrent text and time series [C]// *KDD-2000 workshop on Text Mining*. 2000: 37-44.
- [16] 赵丽丽, 赵茜倩, 杨娟, 等. 财经新闻对中国股价影响的定量分析[J]. *山东大学学报*, 2012, 47: 70-75.
- [17] 金雪军, 祝宇, 杨晓兰. 网络媒体对股票市场的影响——以东方财富网股吧为例的实证研究[J]. *新闻与传播研究*, 2013(12): 36-51.
- [18] 佟瑞鹏, 谢贝贝, 安宇. 黑天鹅事件定义及分类的探讨[J]. *中国公共安全(学术版)*, 2017(2).
- [19] 赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. *软件学报*, 2010, 21(8): 1834-1848.

(上接第 111 页)

参考文献

- [1] 孙丽妍. 如何培养小学生的语文朗读能力[J]. *语文建设*, 2018, 12: 97.
- [2] BERTIN-MAHIEUX T, ELLIS D P W. Large-scale cover song recognition using hashed chroma landmarks[C]// *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. NY, USA: IEEE, 2011: 117-120.
- [3] OORD A V D, DIELEMAN S, ZEN H G, et al. Wavenet: A generative model for raw audio[C]// *arXiv:1609.03499*. 2016.
- [4] EZZAT S, EL GAYAR N, GHANEM M M. Sentiment analysis of call centre audio conversations using text classification[J]. *International Journal of Computer Information Systems and Industrial Management Applications*, 2012, 4(1): 619-627.
- [5] 韩文静, 李海峰, 阮华斌, 等. 语音情感识别研究进展综述[J]. *软件学报*, 2014, 25(1): 37-50.
- [6] TRABELSI I, AYED D B. On the Use of Different Feature Extraction Methods for Linear and Non Linear kernels[J]. *Computer Science*, 2014.
- [7] PALO H K, MOHANTY M N, CHANDRA M. Computational Vision and Robotics[C]// *Advances in Intelligent Systems and Computing*. 2015: 63-70.
- [8] RODDY C. Emotion recognition in human-computer interaction [J]. *Signal Processing Magazine, IEEE*, 2001, 18(1): 32-80.
- [9] GEROSA M, LEE S, GIULIANI D, et al. Analyzing children's speech: An acoustic study of consonants and consonant-vowel transition[C]// *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2006 (ICASSP 2006). 2006: 1393-1396.
- [10] YILDIRIM S, NARAYANAN S, BOYD D, et al. Acoustic analysis of preschool children's speech[C]// *Proc. 15th ICPhS*. 2003: 949-952.
- [11] LI, RUSSELL M J. An analysis of the causes of increased error rates in children's speech recognition[C]// *Seventh International Conference on Spoken Language Processing*. 2002.
- [12] AFAVI S, NAJAFIAN M, HANANI A, et al. Speaker recognition for children's speech[J]. *arXiv:1609.07498*, 2016.
- [13] LI P C, SONG Y, MCLOUGHLIN I V, et al. An Attention Pooling Based Representation Learning Method for Speech Emotion Recognition[C]// *Interspeech*. 2018: 3087-3091
- [14] BADSHAH A M, AHMAD J, RAHIM N, et al. Speech Emotion Recognition from Spectrograms with Deep Convolutional Neural Network[C]// *International Conference on Platform Technology & Service*. IEEE, 2017.
- [15] ETIENNE C, FIDANZA G, PETROVSKII A, et al. CNN+LSTM Architecture for Speech Emotion Recognition with Data Augmentation[J]. *Computer Science*, 2018.
- [16] CUMMINS N, AMIRIPARIAN S, HAGERER G. An Image-based Deep Spectrum Feature Representation for the Recognition of Emotional Speech[C]// *ACM on Multimedia Conference*. 2017.
- [17] KANG J, ZHANG W Q, LIU J. Gated recurrent units based hybrid acoustic models for robust speech recognition[C]// *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)*. IEEE, 2016.
- [18] HUANG Y S, CHOU S Y, YANG Y H. Pop Music Highlighter: Marking the Emotion Keypoints[J]. *arXiv:1802.10495*, 2018.
- [19] MIRSAMADI S, BARSOUM E, ZHANG C. Automatic Speech Emotion Recognition Using Recurrent Neural Networks with Local Attention[C]// *ICASSP*. IEEE, 2017.
- [20] PASSALIS N, TEFAS A. Neural bag-of-features learning[J]. *Pattern Recognition*, 2017, 64: 277-294.
- [21] KAUSHIK L, SANGWAN A, HANSEN J H. Automatic audio sentiment extraction using keyword spotting[C]// *Sixteenth Annual Conference of the International Speech Communication Association*. 2015.
- [22] AMODEI D, ANUBHAI R, BATTENBERG E, et al. Deep Speech 2: End-to-End Speech Recognition in English and Mandarin[J]. *Computer Science*, 2015.