

### 基于 Bi-LSTM 的期货市场关联交易行为检测方法

张源, 康乐, 宫朝辉, 张志鸿

#### 引用本文

张源, 康乐, 宫朝辉, 张志鸿. 基于 Bi-LSTM 的期货市场关联交易行为检测方法[J]. 计算机科学, 2022, 49(7): 31-39.

ZHANG Yuan, KANG Le, GONG Zhao-hui, ZHANG Zhi-hong. [Related Transaction Behavior Detection in Futures Market Based on Bi-LSTM](#) [J]. Computer Science, 2022, 49(7): 31-39.

---

#### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于 BERT 与 Bi-LSTM 融合注意力机制的中医病历文本的提取与自动分类](#)

Extraction and Automatic Classification of TCM Medical Records Based on Attention Mechanism of BERT and Bi-LSTM

计算机科学, 2020, 47(11A): 416-420. <https://doi.org/10.11896/jsjcx.200200020>

#### [面向国防科技领域的技术和术语识别方法研究](#)

Technology and Terminology Detection Oriented National Defense Science

计算机科学, 2019, 46(12): 231-236. <https://doi.org/10.11896/jsjcx.190300069>

#### [基于 DBN 深度学习的期货市场价格预测建模与决策](#)

Modeling and Decision-making of Futures Market Price Prediction with DBN

计算机科学, 2018, 45(6A): 75-78.

#### [基于双线性函数注意力 Bi-LSTM 模型的机器阅读理解](#)

Attention of Bilinear Function Based Bi-LSTM Model for Machine Reading Comprehension

计算机科学, 2017, 44(Z6): 92-96. <https://doi.org/10.11896/j.issn.1002-137X.2017.6A.019>

# 基于 Bi-LSTM 的期货市场关联交易行为检测方法

张源<sup>1</sup> 康乐<sup>2,3</sup> 宫朝辉<sup>3</sup> 张志鸿<sup>1</sup>

1 郑州大学信息工程学院 郑州 450001

2 清华大学计算机科学与技术系 北京 100084

3 郑州易盛信息金融创新实验室 郑州 450018

(iezyzhang@163.com)

**摘要** 随着期货市场的不断发展,其交易量屡创新高,但在海量交易的背后,一些交易者利用关联交易行为对市场进行操纵,扰乱了交易秩序,给市场监管和风险控制带来了严峻考验。因此,如何从海量交易中挖掘潜在关联交易行为成为维护期货市场公平交易的重要任务。针对该问题,提出了一种多特征信息融合的双向长短期记忆(Bi-LSTM)网络模型,从原始数据中提取交易时间、交易量、持仓变化、期货品种等多种维度的浅层特征信息,通过 Bi-LSTM 网络模型从时间序列上向前、向后两个方向的上下文关系学习深层特征,实现关联交易行为检测。针对浅层特征提取提出了一种基于交易行为的多粒度窗口特征提取方法,从日、小时、分钟、秒等级别捕捉账户间交易的关联性,从而解决了原始交易数据维度高、数据量大、关联性弱的问题。模型引入了 Dropout 策略,缓解了收敛速度慢和过拟合的问题。在郑州商品交易所真实数据上的实验结果表明,与一些传统的分类模型以及 RNN 和 LSTM 网络相比,所提方法在分类的准确率和召回率上有明显提升,同时,对特征中各个维度信息的消融实验证明了多特征融合方法和多粒度窗口策略的有效性。另外,抽取了两种期货品种的交易数据进行测试,结果表明所提模型具有良好的泛化能力。

**关键词:** 关联交易行为;多特征信息融合;多粒度特征提取;Bi-LSTM;期货市场

**中图分类号** TP301

## Related Transaction Behavior Detection in Futures Market Based on Bi-LSTM

ZHANG Yuan<sup>1</sup>, KANG Le<sup>2,3</sup>, GONG Zhao-hui<sup>3</sup> and ZHANG Zhi-hong<sup>1</sup>

1 School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China

2 Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

3 Finance Innovation Laboratory of Zhengzhou Esunny Information, Zhengzhou 450018, China

**Abstract** With the continuous development of the futures market, its transaction volume keeps increasing. Behind the massive transactions, some traders use related transaction behaviors to manipulate the futures market and disrupt the order of transactions, which has brought severe challenges to market supervision and risk control. How to mining potential related transaction behaviors from massive transaction data is an important task for maintaining fair transactions in the futures market. In response to this problem, this paper proposes a bidirectional long short-term memory(Bi-LSTM) network model with multi-feature information fusion, which extracts multiple dimensions of shallow feature information such as trading time, trading volume, position changes, and futures varieties from the original transaction data. The Bi-LSTM network model learns deep features from the contextual relationship in the forward and backward directions of the time series and realizes the detection of related transaction behavior. For shallow features extraction, a multi-granularity window feature extraction method based on transaction behavior is proposed, to captures the correlation of transactions between accounts from the levels of day, hour, minute, second, etc. It solves the problems of high data dimension, large amount of data, and weak correlation of original transaction data. The model introduces the Dropout strategy to alleviate problems of slow convergence and over-fitting. Experimental results on the real data of Zhengzhou Commodity Exchange show that the proposed method evidently improves the classification precision and recall compared with some traditional classification models and RNN and LSTM network. At the same time, the ablation experiment of each dimension information proves the effectiveness of the multi-feature fusion method and the multi-granularity window strategy. In addition, the transaction data of two futures varieties are extracted for testing, and the results show that the proposed model has good generalization ability.

到稿日期:2021-04-29 返修日期:2021-08-16

基金项目:河南省重大公益专项(201300210300)

This work was supported by the Major Public Welfare Project of Henan Province(201300210300).

通信作者:张志鸿(iezhzhang@zzu.edu.cn)

**Keywords** Related transaction behavior, Multi-feature information fusion, Multi-granularity feature extraction, Bi-LSTM, Futures market

## 1 引言

在金融市场,非正常交易行为给市场监管和风险控制带来了严峻的考验。这种迅速膨胀的非正常交易行为主要表现为:有经验的交易者形成关联交易群体,操纵某种商品的价格,迷惑其他投资者以达到个体利益最大化<sup>[1]</sup>。2019年11月16日,证监会修订了《期货交易所管理办法》<sup>[2]</sup>,明确要求期货交易所应当建立健全实际控制账户备案管理制度。但在实际操作中,依然有交易者不进行账户备案,企图通过实际控制账户进行违规交易或分仓,以逃避交易所对持仓限额和大户报告的监管。这部分账户通过虚构市场供求关系、控制期货价格来诱导其他投资者做出错误判断,并从中获得利益,给期货市场的公平性带来了隐患。因此,筛查并重点监控这部分账户的交易行为成为了交易所监察部门长期以来的重点工作。

期货市场的关联交易行为主要有以下两种表现形式:1)连续买卖和通谋买卖,即一人利用不同账户或多人串通连续买入某期货品种,制造行情上涨的假象来引诱其他投资者追涨,在价格抬升后卖出,等待价格回落后再次炒作,不断循环操作进而操纵期货价格;2)相对委托,即部分投资者经串通,分别扮演买方和卖方,按事先约定的期货品种、价格、数量,在相近的时间内买入和卖出,人为制造虚假交易,使某期货价格上涨或下跌,给其他投资者造成交易活跃的假象,串通者趁机获取利润。

关联交易行为的表现形式最终体现在交易账户上,其交易行为存在一定的关联性,因此关联交易行为检测的重点是对账户交易数据的检测分析。在交易量、交易时间、持仓变化、期货品种上有高度相似性的行为即为关联交易行为,其交易的账户为关联账户。为了躲避监管,这种目的相似的交易并不总是同时进行,交易者会有意增加关联交易的时间间隔,以迷惑监管者。

近年来,有学者和监管机构对这部分关联账户的交易行为进行了研究,提出了一些检测识别的方法。这些方法大体上可以分为基于聚类的方法、基于皮尔逊相关系数的方法、基于词向量表达的方法和基于耦合隐马尔可夫的方法,但这些方法主要集中于传统机器学习方法和数据挖掘方法。近些年,深度学习技术快速发展,并在股价预测<sup>[3]</sup>、时间序列预测、文本分析、计算机视觉<sup>[4]</sup>等领域都取得了突破性的成果,成为学术研究和工业应用的主流,但较少涉及针对期货市场的研究。

针对上述问题,本文提出了一种多特征信息融合<sup>[5]</sup>的双向长短期记忆网络(Bidirectional Long Short-Term Memory, Bi-LSTM)模型,该模型使用一种多粒度窗口的特征提取方法,从原始交易数据中提取交易时间、交易量、持仓变化、期货品种等多种维度的浅层特征信息,代替原始交易数据作为模型的输入。与原始交易数据相比,加工后的浅层特征数据维度更低,具有更明显的关联性。通过模型在浅层特征中学习更深层的特征,进而挖掘潜在的关联交易行为。该模型利用向前、向后两个时间方向的上下文关系来进行预测,同时引入

了Dropout策略<sup>[6]</sup>来缓解过拟合,以进一步提高模型的泛化能力。

本文的主要贡献包括:

(1)提出了一种基于交易行为的多粒度窗口特征提取方法,将两个账户的交易数据融合为一条特殊数据,从日、小时、分钟、秒等级别捕捉账户间交易的关联性,从而解决了原始交易数据维度高、数据量大、关联性弱的问题。

(2)提出了一种多特征信息融合的Bi-LSTM网络模型,利用Bi-LSTM对时间序列数据强大的上下文表征能力来融合多种浅层特征数据,并将其作为神经网络的输入,提高了关联交易行为检测的准确率和召回率。

(3)提出的多粒度窗口特征提取方法和多特征信息融合的Bi-LSTM网络模型在郑州商品交易所的真实数据上进行了实验,结果表明其对关联交易行为具有良好的检测能力。

## 2 相关工作

近年来,期货交易所利用终端信息一致、开户信息一致等其他辅助信息对未报备的疑似关联账户进行筛查监控。在实际工作中,部分不愿意报备的账户往往选择差异化的开户地、联系方式、交易地点和终端设备等来规避期货交易所的监管。

目前,针对期货市场的关联交易行为的研究较少,而较多的是针对股票市场的关联交易行为研究。近年来,有学者从不同角度对账户之间的交易行为进行了研究,为了区分正常交易中的非正常交易行为,Franke等<sup>[7]</sup>提出基于Hermitian邻接矩阵的聚类方法来对关联账户的交易行为和市场进行建模,以此来找到关联账户之间的循环交易(组内成员之间进行的大量交易)。然而,这项研究是在一个实验性的股票市场上进行的。在此基础上,Palshikar等<sup>[8]</sup>提出了一种基于Dempster-Shafer理论的新型图聚类方法,他们用合成的交易数据构建股票流程图来表示交易者之间的交易关系,并寻找产生循环交易的关联账户群体。Cao等<sup>[9]</sup>认为关联交易行为是由一群隐藏的账户通过精心安排交易价格、交易量和交易时间,并相互合作,同时进行买入和卖出活动产生的,他们提出了一种基于耦合隐马尔可夫的方法来描述群体成员之间的交易行为,并对股票数据上的异常交易行为进行了检测。针对如何在海量的交易数据下快速发现账户间的异常交易行为这一问题,Zhang等<sup>[10]</sup>提出了一种将交易数据的稀疏相似性矩阵映射到十字链表的聚类方法,并用真实数据做了实证研究,结果表明该算法减少了扫描数据库的次数并降低了传统算法的时间复杂度和空间复杂度。Zhai等<sup>[11]</sup>首先提出了一种基于交易数据的k-means聚类算法,该算法根据交易行为对账户进行分组;然后又提出了一种基于动态规划的算法对关联账户进行精细识别。其提出的两种算法利用纽约证券交易所和芝加哥商品交易所的真实数据进行了实验评估。Huang等<sup>[12]</sup>利用交易量变化的序列,评估了潜在的序列相似性度量,并选择适合行为模式聚类的距离,提出了一种用于关联交易账户识别的行为模式聚类算法。Wang等<sup>[13]</sup>提出了一种标准化的规整交易订单表达向量,对一定时间窗内的每一对标准化的

订单向量计算相关系数,基于该系数搭建账户的关系带权图;他们还设计了一种迭代算法对图中可疑关联账户进行检测,用上海期货交易所的真实交易数据进行了实验评估。但该方法只考虑了短时间内的交易记录,忽略了日间交易行为的相似性,导致检测结果存在偏差。

金融科技企业对关联交易检测也进行了一些研究。1)上海期货交易所:Yu等<sup>[14]</sup>提出了一种计算账户之间的皮尔逊相关系数的方法,其结合深度优先的搜索算法进行关联账户组的识别,但是在数据量增加时,该方法的时间复杂度也会增加;2)大连商品交易所:Ma等<sup>[15]</sup>从原始特征、数据变换及复合特征3个维度来表征账户的交易行为,并提出了两种不同的检测方法,分别是基于皮尔逊相关系数的识别方法和基于Word2Vec词向量的识别方法,但并未基于真实数据进行实验分析;3)上海金融期货信息技术有限公司:Miao等<sup>[16]</sup>使用数据分析处理方法和有监督、无监督的人工智能算法模型配合应用,通过账户的交易数据和基本信息构建账户的特征向量,并计算向量间的相似度,单交易日内使用图分析或聚类算法进行检测,跨交易日使用多子图融合分析或基于频繁项集挖掘的方法检测关联交易;4)支付宝(杭州)信息技术有限公司:Hu<sup>[17]</sup>利用账户交易数据的属性对待识别的交易数据进行聚类,当不同交易数据具有相同聚类属性值时说明交易存在关联,并通过多次迭代聚类,检测关联交易。

金融企业和普通学者对关联交易检测的研究都是基于传统的数据挖掘方法和机器学习方法。本文分析了关联账户之间的交易行为,深度挖掘账户日间交易行为的相似性和短时间内交易的关联性,并从深度学习方面着手,提出了基于多特征信息融合的Bi-LSTM分类模型<sup>[18-19]</sup>,该模型是对关联交易检测方法的一种探索。

### 3 数据分析与特征提取

本文采用郑州商品交易所历史上标记的关联账户之间的交易数据进行统计分析。通过分析历史上确认的关联账户之间的交易行为,从中发现其相似性,并总结归纳出潜在规律。在此基础上,提出了一种基于交易行为的多粒度窗口的特征提取方法,并且把两个账户的交易数据融合成一条特殊数据,从而解决了原始交易行为数据维度高、数据量大、关联性弱的问题。

#### 3.1 历史数据分析

为了从关联账户的交易行为中寻找潜在关系,本文将市场上账户的交易行为划分成两个最基本的单元,分别是盘中交易与盘后持仓。其中,盘中交易可分为4个子类,分别是买开仓、卖开仓、买平仓、卖平仓;盘后持仓可分为买持仓与卖持仓。从这两个维度对交易行为进行分析,数据统计时间为每日收盘结算后,即按日统计。

为了能更直观地观察账户交易行为的变化,本文以时间为轴线,统计了指定期货品种上关联账户的交易行为数据,而据柱状图的不断变化反映了交易行为的变化。如图1所示,图中截取了关联账户7天的交易数据,每个柱体表示一个账户每日收盘时的交易情况,上半部分表示交易数据,其中,红色表示买开仓(Buy Open Volume,BOV)、绿色表示卖开仓(Sell Open Volume,SOV)、橙色表示买平仓(Buy Liqui-

date Volume,BLV)、蓝色表示卖平仓(Sell Liquidate Volume,SLV),0轴上方为买交易行为,下方为卖交易行为。图中下半部分表示持仓数据,0轴上方为买持仓量,下方为卖持仓量,红色表示买持仓量(Buy Holding Volume,BHV),绿色表示卖持仓量(Sell Holding Volume,SHV)。

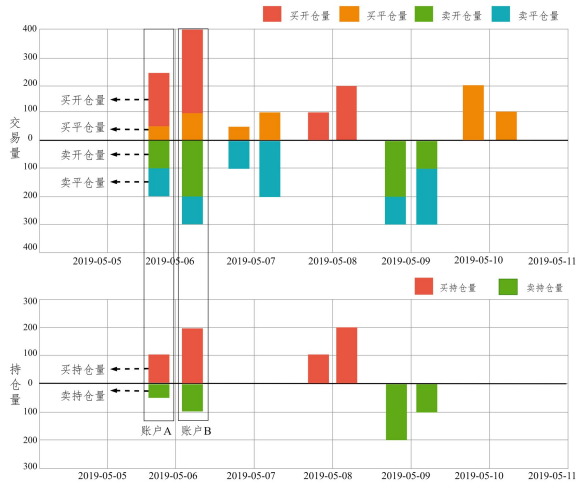


图1 交易行为分析(电子版为彩图)

Fig. 1 Analysis of transaction behaviors

根据两个账户每天的交易行为变化情况,可以分析出一些关联信息。从图1中可以看出,2019年5月6日账户A和账户B都进行了买开仓、买平仓、卖开仓、卖平仓操作,并且交易量都相近或成比例,当日都有买持仓和卖持仓;5月7日同步进行了平仓,持仓量同时转化为0;5月8日同步买开仓;5月9日同步卖平仓和卖开仓,持仓由买持仓转换为卖持仓;5月10日同步买平仓,持仓量再次同时转换为0。可以看出,账户A和账户B的交易时间非常同步,交易量相近或成比例,持仓变化高度一致。

接下来从盘中交易数据和盘后持仓数据两方面对一些关联账户的典型交易行为进行分析。

(1)账户盘后不留仓,日内交易结构高度一致。图2截取了两个账户在同种期货合约上7天的交易数据。可以看出,两个账户在每个交易日都同步进行开仓,并在当日结算前同步进行平仓,每个交易日结束时都没有持仓数据,盘中交易数据高度相似。

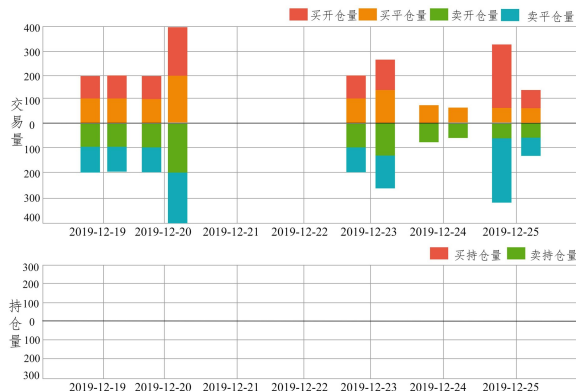


图2 盘中交易分析(电子版为彩图)

Fig. 2 Intraday trading analysis

(2)账户多个交易日无任何交易数据,但长时间维持相同持仓方向。图3截取了两个账户在同种期货合约上8天的

交易数据。可以看出,两个账户在 2019 年 1 月 2 日同时卖开仓后,一直到 1 月 7 日期间没有任何交易数据,连续 6 天保持相同方向的持仓,在 1 月 8 日同步买平仓,持仓由卖持仓转换为 0,盘后持仓数据高度相似。

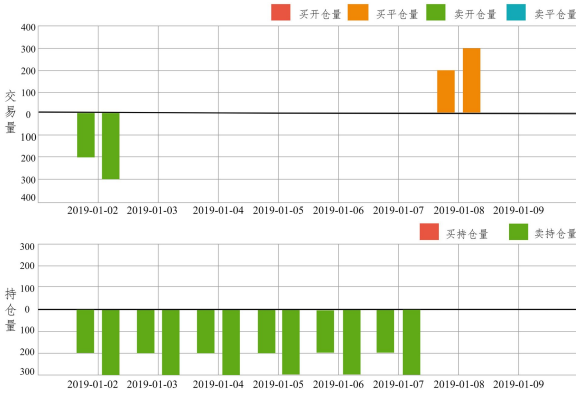


图 3 盘后持仓分析(电子版为彩图)  
Fig. 3 Analysis of after-hours position

(3)账户持仓变化同增同减,开平仓节奏高度一致。图 4 截取了两个账户在同种期货合约上 9 天的交易数据。可以看出,两个账户在 9 天内的持仓变化表现出同增同减的趋势,买持仓、卖持仓同天切换,开平仓的交易方向一致。在 2019 年 8 月 5 日同步买平仓和卖平仓,持仓量同步转换为 0;8 月 6 日同步卖开仓,持仓方向由买持仓切换为卖持仓;8 月 7 日同步由卖持仓转换为买持仓。



图 4 同向交易行为分析(电子版为彩图)  
Fig. 4 Analysis of codirectional transaction behaviors

(4)账户持仓变化一致,但交易方向和持仓方向相反。图 5 截取了两个账户在同种期货合约上 8 天的交易数据。

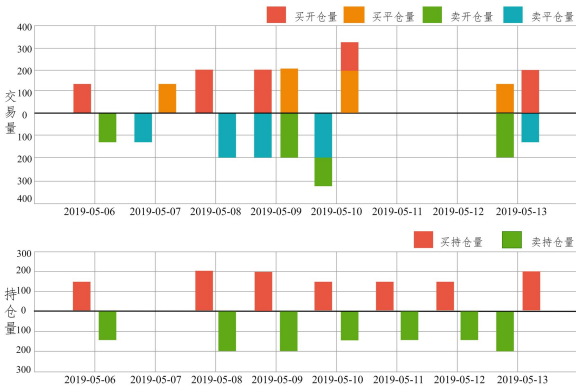


图 5 反向交易行为分析(电子版为彩图)  
Fig. 5 Analysis of reverse transaction behaviors

可以看出,两个账户在各个交易日内的持仓方向相反,但持仓变化一致,每个交易日开平仓的交易方向相反。2019 年 5 月 6 日账户 A 进行买开仓,账户 B 进行卖开仓,持仓分别为买持仓和卖持仓;5 月 7 日分别卖平仓和买平仓,持仓同步转换为 0;5 月 11 日和 5 月 12 日两账户都未进行交易,买持仓和卖持仓保持不变。

通过以上统计分析可以看出,无论在何种交易行为中,关联账户为了获取相应利益,通常都会在交易行为上表现出相似性,在持仓变化上显示联动性,主要表现在交易同种期货、交易时间相近,交易量相近或成比例,持仓变化相关。这说明捕捉账户之间交易行为和持仓变化的潜在关系就是挖掘关联交易行为的关键。

### 3.2 特征提取方法

由于要对账户之间的关联交易行为进行分类检测,因此如何将数据送入模型训练成为了一个问题。针对此问题,提出了一种基于交易行为的多粒度窗口的特征提取方法,如图 6 所示,将两个账户的交易数据用特征提取方法转换为一条特征数据,从日、小时、分钟、秒等级别捕捉账户间交易的关联性。提取的特征包括交易时间特征(Trading Time Features, TTF)、交易量特征(Trading Volume Features, TVF)、持仓特征(Position Features, PF)、期货品种特征(Futures Varieties Features, FVF)。其中最主要的是交易时间,其通过划分多粒度的时间窗口,逐步缩小窗口大小,从不同的粒度提取交易时间特征。

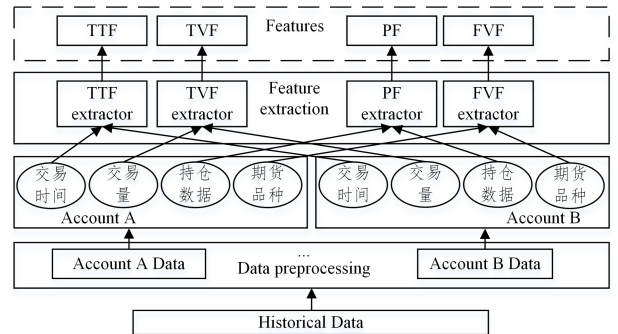


图 6 特征提取方法  
Fig. 6 Feature extraction method

#### 3.2.1 交易时间特征

交易数据共有 4 个方向,分别是买开仓、买平仓、卖开仓、卖平仓。在账户 A 和账户 B 的原始数据中,以日、时、分、3 秒为不同粒度划分时间窗口,统计每个窗口下两个账户的交易情况。

用  $i$  表示时间窗口粒度,  $i \in (day, hour, minute, second)$ , 其中  $day$  表示时间窗口以日为粒度划分,共划分了  $n$  天;  $hour$  表示时间窗口以小时为粒度划分,共划分了  $m$  个小时;  $minute$  表示时间窗口以分钟为粒度划分,共划分了  $a$  个分钟;  $second$  表示时间窗口以 3 秒为粒度划分,共划分了  $b$  个 3 秒。

用  $j$  表示交易的方向,  $j \in (bo, bl, so, sl)$ 。其中,  $bo$  表示买开仓,  $bl$  表示买平仓,  $so$  表示卖开仓,  $sl$  表示卖平仓。

#### (1)同向交易特征提取

同向交易即两个账户在一个时间窗口下具有相同方向的交易。特征提取方法如式(1)所示,在不同粒度的窗口下,记录账户在哪些窗口有交易,在哪些窗口同时有交易。

$$F = \frac{\text{count}(A_{ij} \cap B_{ij})}{\text{count}(A_{ij} \cup B_{ij})} \quad (1)$$

其中,  $A_{ij}, B_{ij}$  分别表示账户 A 和账户 B 在不同窗口粒度下、不同交易方向上有交易的窗口集合;  $A_{ij} \cap B_{ij}$  表示账户 A 和账户 B 同时有同方向交易的窗口集合;  $A_{ij} \cup B_{ij}$  表示账户 A 或账户 B 有交易的窗口集合;  $\text{count}$  用于统计集合的元素个数;  $F \in [0, 1]$ , 表示在不同粒度窗口下, 账户 A 和账户 B 在不同交易方向上有同向交易的概率。

### (2) 反向交易特征提取

反向交易就是一个账户进行买操作, 另一个账户进行卖操作。特征提取方法如式(2)所示, 在不同粒度的窗口下, 记录有交易记录的窗口和两个账户有反方向交易的窗口。

$$C = \frac{\text{count}((A_{ij_1} \cap B_{ij_2}) \cup (A_{ij_2} \cap B_{ij_1}))}{\text{count}(A_{ij} \cup B_{ij})} \quad (2)$$

其中,  $j_1$  表示买方向,  $j_1 \in (bo, bl)$ ;  $j_2$  表示卖方向,  $j_2 \in (so, sl)$ ;  $A_{ij_1}, B_{ij_1}$  分别表示账户 A、账户 B 在不同粒度窗口下, 有买方向交易的窗口集合;  $A_{ij_2}, B_{ij_2}$  分别表示账户 A、账户 B 在不同粒度窗口下, 有卖方向交易的窗口集合;  $A_{ij}, B_{ij}$  分别表示账户 A、账户 B 在不同粒度窗口下, 有交易的窗口集合;  $C \in [0, 1]$  表示在不同粒度窗口下, 账户 A 和账户 B 有反向交易的概率。

### 3.2.2 交易量特征

账户在不同交易方向上都有对应的交易量, 以日为单位, 在不同交易方向上, 统计账户 A 每天的交易量  $X_j, X_j \in (x_1, x_2, x_3, \dots, x_n)$ , 账户 B 每天的交易量  $Y_j, Y_j \in (y_1, y_2, y_3, \dots, y_n)$ , 特征提取方法如式(3)所示, 计算两个账户在  $n$  个交易日上交易量的协方差, 并进行标准化。

$$\rho = \left| \frac{\text{Cov}(X_j, Y_j)}{\sqrt{D(X_j)D(Y_j)}} \right| \quad (3)$$

其中, 时间粒度按日划分, 共  $n$  天;  $j \in (bo, bl, so, sl)$ ;  $\text{Cov}(X_j, Y_j)$  表示两个账户在  $n$  个交易日不同交易方向上交易量的协方差;  $D(X_j), D(Y_j)$  分别表示账户 A、账户 B 在  $n$  个交易日不同交易方向上交易量的方差;  $\rho$  表示账户 A 和账户 B 在不同交易方向上交易量的相关系数。

### 3.2.3 持仓特征

持仓指账户进行交易后的状态, 它的变化可以间接反映用户交易的过程, 共有两个持仓方向, 分别是买持仓和卖持仓。持仓数据的统计时间为每日收盘结算后, 以日为粒度划分时间窗口。

用  $k$  表示持仓方向,  $k \in (bh, sh)$ ,  $bh$  表示买持仓,  $sh$  表示卖持仓。

#### (1) 持仓变化特征

在账户 A、账户 B 持仓量日间变化的过程中提取持仓变化特征, 将每一天的持仓量与前一天相比, 统计账户 A 和账户 B 在不同方向上持仓量同时增加的次数  $v_1$ , 同时减少的次数  $v_2$ , 同时保持不变且同时大于 0 的次数  $v_3$ ; 统计账户 A 或账户 B 持仓量增加的次数  $e_1$ , 减少的次数  $e_2$ , 保持不变且大于 0 的次数  $e_3$ 。提取方法如式(4)~式(6)所示。

$$V = \begin{cases} v_1(k), & H_{(i+1)}^k(A) > H_i^k(A) \\ & \text{且 } H_{(i+1)}^k(B) > H_i^k(B) \\ v_2(k), & H_{(i+1)}^k(A) < H_i^k(A) \\ & \text{且 } H_{(i+1)}^k(B) < H_i^k(B) \\ v_3(k), & H_i^k(A) = H_{(i+1)}^k(A) > 0 \\ & \text{且 } H_i^k(B) = H_{(i+1)}^k(B) > 0 \end{cases} \quad (4)$$

$$E = \begin{cases} e_1(k), & H_{(i+1)}^k(A) > H_i^k(A) \\ & \text{或 } H_{(i+1)}^k(B) > H_i^k(B) \\ e_2(k), & H_{(i+1)}^k(A) < H_i^k(A) \\ & \text{或 } H_{(i+1)}^k(B) < H_i^k(B) \\ e_3(k), & H_i^k(A) = H_{(i+1)}^k(A) > 0 \\ & \text{或 } H_i^k(B) = H_{(i+1)}^k(B) > 0 \end{cases} \quad (5)$$

$$I = \frac{v_1(k) + v_2(k) + v_3(k)}{e_1(k) + e_2(k) + e_3(k)} \quad (6)$$

其中,  $i$  表示时间,  $i \in (1, 2, 3, \dots, n)$ , 共  $n$  天;  $H$  表示账户每天的持仓量;  $H_i^k(A), H_i^k(B)$  分别表示账户 A、账户 B 在不同方向上每天的持仓量;  $V$  表示账户 A 和账户 B 持仓量同时变化的次数;  $E$  表示账户 A 或账户 B 持仓量变化的次数,  $I \in [0, 1]$  表示两个账户在不同方向上持仓量同时变化的概率。

#### (2) 同向持仓特征

同向持仓是看两个账户在时间窗内是否有相同方向的持仓, 提取方法与交易时间的提取方法相似, 以日为粒度划分时间窗口, 提取方法如式(7)所示, 记录两个账户每个时间窗内的持仓情况。

$$T = \frac{\text{count}(A_k \cap B_k)}{\text{count}(A_k \cup B_k)} \quad (7)$$

其中,  $A_k, B_k$  分别表示账户 A、账户 B 在买持仓或卖持仓方向有持仓的窗口集合;  $A_k \cap B_k$  表示账户 A 和账户 B 在买持仓或卖持仓方向同时有持仓的窗口集合;  $A_k \cup B_k$  表示账户 A 或账户 B 在买持仓或卖持仓方向有持仓的窗口集合;  $\text{count}$  用于统计集合的元素个数;  $T \in [0, 1]$  表示账户 A、账户 B 在不同方向上有同向持仓的概率。

#### (3) 反向持仓特征

反向持仓指一个账户有买持仓量, 另一个账户有卖持仓量。它是反向交易的体现, 特征提取方法如式(8)所示, 记录两个账户每个时间窗内的持仓情况。

$$M = \frac{\text{count}((A_{bh} \cap B_{sh}) \cup (A_{sh} \cap B_{bh}))}{\text{count}(A_k \cup B_k)} \quad (8)$$

其中,  $A_{bh}, B_{bh}$  分别表示账户 A、账户 B 有买持仓的窗口集合;  $A_{sh}, B_{sh}$  分别表示账户 A、账户 B 有卖持仓的窗口集合;  $A_k, B_k$  分别表示账户 A、账户 B 有持仓的窗口集合;  $\text{count}$  用于统计集合的元素个数;  $M \in [0, 1]$  表示账户 A 和账户 B 有反向持仓的概率。

### 3.2.4 期货品种特征

期货品种的特征提取方法如式(9)所示,  $P_A$  表示账户 A 交易的期货品种,  $P_B$  表示账户 B 交易的期货品种,  $P$  表示两个账户的期货品种特征, 两个账户交易的期货品种相同时, 特征为 1; 期货品种不同时, 特征为 0。

$$P = \begin{cases} 1, & P_A = P_B \\ 0, & P_A \neq P_B \end{cases} \quad (9)$$

## 4 Bi-LSTM 网络模型

本文提出的多特征信息融合的双向长短期记忆网络模型结构如图 7、图 8 所示, 该模型分为输入层、双向 LSTM 部分、全连接部分(Fully Connected Layer, FC)和输出层。输入层将 3 种浅层特征数据按时间串联成一个多通道的输入数据, 双向 LSTM 部分由 4 层双向 LSTM 网络组成, 全连接部分由 3 层全连接组合构成。其中, 中间层 LSTM 每个时刻的隐藏

状态全部输出,最后一层 LSTM 仅返回最后一个时刻的输出,减少深层特征个数,降低参数量。在 4 个串联的双向 LSTM 网络后,使用 3 层全连接网络对深层特征进行分类。

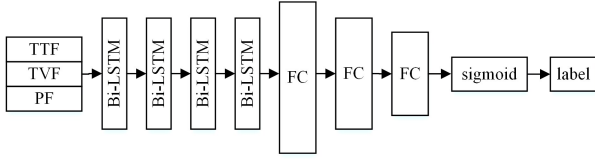


图 7 本文提出的模型结构

Fig. 7 Structure of the proposed model

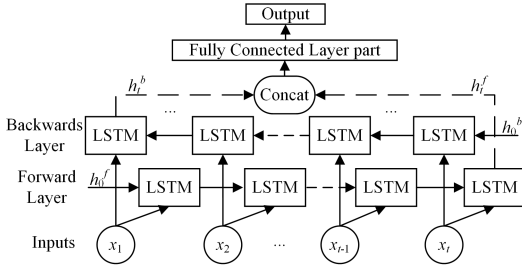


图 8 Bi-LSTM 网络结构

Fig. 8 Structure of Bi-LSTM network

账户在每个时刻都有可能进行交易,检测过程中的每个时刻不仅依赖过去的信息,未来的信息对其也有影响。因此,本文采用双向 LSTM 网络,避免了 RNN 在长时间序列上的梯度消失和梯度爆炸问题。该模型还可以利用时间序列向前、向后两个时间方向的上下文关系,学习对时间有长期依赖性的信息。为了深层次地挖掘序列数据的特征,模型将 4 个双向 LSTM 网络叠加起来,并且添加了全连接层,全连接层具有良好的非线性映射能力,能对双向 LSTM 网络输出的非线性特征进行加权处理。随着网络层的增加,模型的训练难度也会增大且容易出现收敛速度慢和过拟合的问题,因此在双向 LSTM 部分和全连接部分引入了 Dropout 策略来缓解这些问题,即在模型训练时,通过预先设置的概率停止神经元的输出,每次的模型训练只有部分数据特征参与,从而防止模型过多地学习训练集的数据特征。

#### 4.1 LSTM

LSTM 是循环神经网络的变体。理论上,RNN 可以处理任何长距离依赖问题,但其梯度消失、梯度爆炸等问题导致 RNN 仅有短期记忆,无法实现信息的长期保存。因此,LSTM 通过添加内部的门控机制和记忆单元来维持信息的长期保存,从而解决了信息的长期依赖问题。LSTM 单元结构如图 9 所示。

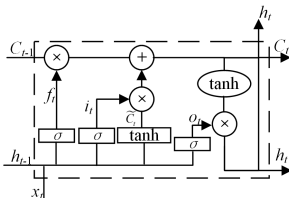


图 9 LSTM 单元结构

Fig. 9 Structure of LSTM cell

LSTM 单元包括输入门  $i_t$ 、遗忘门  $f_t$  以及输出门  $o_t$ , LSTM 隐藏层的输入和输出向量分别为  $x_t$  和  $h_t$ ,记忆单元为

$C_t$ 。输入门用于控制当前输入数据  $x_t$  流入  $C_t$  的数量,即保存到  $C_t$  中的输入信息的数量;遗忘门控制信息的保留及遗忘,并以特定的方式避免当梯度随时间反向传播时引发的梯度消失问题,即上一时刻信息  $C_{t-1}$  对当前时刻  $C_t$  的影响;输出门可以控制记忆单元  $C_t$  对当前输出值  $h_t$  的影响,即在  $t$  时刻控制记忆单元  $C_t$  的输出信息。输出  $h_t$  不仅受  $C_t$  和  $x_t$  的影响,还受  $h_{t-1}$  的影响;记忆单元  $C_t$  不仅受  $h_{t-1}$  和  $x_t$  的影响,还受  $C_{t-1}$  的影响。

在  $t$  时刻,各门状态的数学表达式如下:

$$\tilde{C}_t = \tanh(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c) \quad (10)$$

$$i_t = \sigma(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i) \quad (11)$$

$$f_t = \sigma(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f) \quad (12)$$

$$o_t = \sigma(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o) \quad (13)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t \quad (14)$$

$$h_t = o_t \otimes \tanh C_t \quad (15)$$

其中, $i_t$  是输入信息和过去信息经过输入门的中间计算结果; $f_t$  是输入信息和过去信息经过遗忘门的中间计算结果; $o_t$  是输入信息和过去信息经过输出门的中间计算结果; $\tilde{C}_t$  是输入信息和过去信息经过激活函数  $\tanh$  所得的中间结果; $\sigma$  是 sigmoid 激活函数; $\otimes$  表示元素间的点积,逐点相乘; $W$  和  $U$  是神经网络的权重矩阵; $b$  是神经网络的偏执向量。

LSTM 的门控机制可以捕捉长距离过去信息,因此为了同时获得上下文信息,本文采用双向 LSTM。Bi-LSTM 中的隐藏状态  $h_t$  如式(16)所示。

$$\vec{h}_t = \vec{h}_t \oplus \overleftarrow{h}_t \quad (16)$$

其中, $\vec{h}_t$  和  $\overleftarrow{h}_t$  分别是  $t$  时刻 LSTM 中输入信息的向前和向后的隐藏状态, $\oplus$  表示拼接操作。

#### 4.2 损失函数和优化算法

损失函数(Loss Function)用来评估模型预测值与实际值不一致的程度,损失函数值越小,模型的性能越好。本文采用二元交叉熵损失函数,其计算方法如式(17)所示:

$$L(y, \tilde{y}) = - \sum_{i=1}^n (y_i \log \tilde{y}_i + (1 - y_i) \log(1 - \tilde{y}_i)) \quad (17)$$

其中, $y_i$  表示数据的真实标签, $\tilde{y}_i$  表示模型的预测值, $n$  表示样本的数量。

在模型训练时,需要用优化算法来调整网络的权重参数,减少损失函数的值,从而优化模型性能。Adam (Adaptive Moment Estimation) 算法适用于大数据和高维度的空间,对内存的需求较小,每次迭代参数的学习步长都有一个确定的范围,不会因为梯度较大而学习步长过大,参数更新比较稳定。与其他自适应学习率算法相比,Adam 算法收敛速度更快,学习效果更好。因此本文使用 Adam 优化算法。

## 5 实验结果与分析

### 5.1 数据集与实验环境

#### (1) 数据集

本文采用郑州商品交易所 2019 年 5 月 1 日—5 月 31 日的苹果期货合约 AP910 的交易数据。该期货合约 5 月份约有 16 万账户进行过交易,其中被交易所历史上标记为关联账户(有关联交易行为)的有 9525 对,共 3078 个账户,这部分账户将作为正例。模型的作用是检测对市场价格和公平性有

影响的交易行为,但由于交易量小的账户对市场没有风险,因此下文从交易日在 10 天以上的非关联账户群体中随机选取了 4000 个账户,将这 7078 个账户随机组合出 9600 对非关联账户,作为反例。

实验数据指定了期货品种,故选取交易时间、交易量、持仓量等作为基础特征数据,共 400 多万条。对 19125 对账户的数据,通过提出的多粒度窗口的特征提取方法对其进行浅层特征提取,将结果作为模型的输入,按 3:1:1 的比例划分训练集、验证集、测试集。

## (2) 实验环境

实验采用 Windows10 操作系统,模型基于 pytorch1.8.0 深度学习框架、python3.8 解释器、集成开发环境 pycharm profession2020.1。

## 5.2 评价指标

本文所提模型为二分类模型,预测结果中有关联交易行为用正例表示,无关联交易行为用负例表示,被正确分类的正例和负例分别被称为真正例(True Positive, TP)和真负例(True Negative, TN)。若将负例样本预测为正例,则称其为假正例(False Positive, FP);若将正例样本预测为负例,则称其为假负例(False Negative, FN)。模型的评估指标有准确率(Precision, P)、召回率(Recall, R)和 F1 值(F1-score, F1)。计算方法如式(18)~式(20)所示:

$$P = \frac{TP}{TP + FP} \quad (18)$$

$$R = \frac{TP}{TP + FN} \quad (19)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (20)$$

其中,  $P$  表示正确预测正例的样本数与预测结果为正样本数的比例;  $R$  表示正确预测正例的样本数与实际正样本数的比例;  $F1$  值是准确率和召回率的调和平均值。

## 5.3 实验结果

### 5.3.1 Bi-LSTM 模型调参结果

本文采用网格搜索法进行主要超参数调节,获取到的模型的最优超参数集合如表 1 所列,包括 LSTM 隐层的单元数(Hidden Size),LSTM 的层数(Num Layers),批次大小(Batch Size),学习率(Learning Rate),LSTM 层的 Dropout 值(LSTM Dropout),全连接层的 Dropout 值(FC Dropout),全连接层数(FC Layers)以及每一层的维度。

表 1 Bi-LSTM 模型的超参数

Table 1 Hyperparameters of Bi-LSTM model

| parameter names | parameter values |
|-----------------|------------------|
| hidden size     | 128              |
| num layers      | 4                |
| batch size      | 64               |
| learning rate   | 0.0006           |
| LSTM dropout    | 0.8              |
| FC dropout      | 0.5              |
| FC layers       | 3                |
| FC1             | (512,128)        |
| FC2             | (128,64)         |
| FC3             | (64,1)           |

为了更好地学习数据间的关系,在神经网络中的数据集上训练一轮是不够的,往往需要训练多次才能学习出数据中的潜在关系。为此本文选择训练 100 轮,训练一轮用验证集

测试一次,训练集和验证集的损失函数变化曲线如图 10 所示,验证集精确(Accuracy)变化曲线如图 11 所示。可以看出,随着训练轮次 epoch 的增加,精确率也相应提高,后期精确率提升的幅度明显减小,逐渐趋于平稳。模型在训练 60 轮时损失函数开始收敛,此时模型达到最优性能,保存模型内部参数。

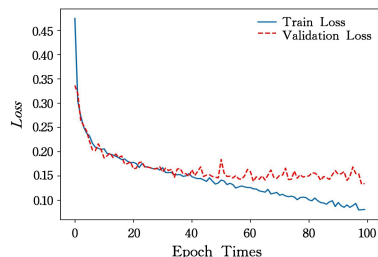


图 10 损失函数曲线

Fig. 10 Loss function curve

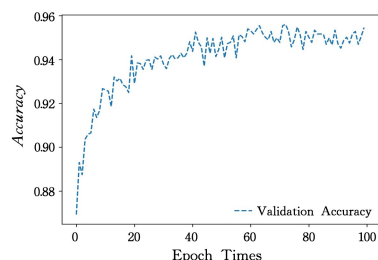


图 11 验证集精确率曲线

Fig. 11 Accuracy curve of validation set

### 5.3.2 多特征信息融合对实验结果的影响

本文采用的多粒度特征提取方法分别从交易时间、交易量、持仓变化方面进行特征提取,为了验证提取特征的有效性,本文比较了浅层特征单个或多个组合的分类结果,将 20 维的交易时间特征 TTF、4 维的交易量特征 TVF、5 维的持仓数据特征 PF 进行组合,测试了不同情况下的实验结果,使用表 1 中的超参数来记录每次测试集的评价指标。如表 2 所列,交易时间特征提供了单通道特征的最好分类结果,在测试集上的准确率和召回率达到了 91.90% 和 82.30%。当融合两类特征数据时,都获得了比单种特征更好的分类结果,其中 TTF 与 PF 的组合提供了双通道特征的最好分类结果,准确率和召回率达到了 93.51% 和 92.08%。当我们融合 TTF, TVF, FP 特征时,分类准确率达到 95.39%,召回率达到了 94.24%,优于单种特征数据和两种特征数据组合的分类准确率和召回率。实验结果表明,融合多个特征数据可以提高对关联交易行为分类的准确率和召回率。

表 2 不同特征融合的实验结果

Table 2 Experimental results of different features fusion strategies

(单位: %)

| Feature fusion |     |    | Evaluation Metrics |        |       |
|----------------|-----|----|--------------------|--------|-------|
| TTF            | TVF | PF | Precision          | Recall | F1    |
| ✓              |     |    | 91.90              | 82.30  | 86.84 |
|                | ✓   |    | 72.18              | 67.15  | 69.28 |
|                |     | ✓  | 88.31              | 71.37  | 78.94 |
| ✓              | ✓   |    | 91.34              | 83.70  | 87.36 |
| ✓              |     | ✓  | 93.51              | 90.70  | 92.08 |
|                | ✓   | ✓  | 86.33              | 82.62  | 84.43 |
| ✓              | ✓   | ✓  | 95.39              | 94.24  | 94.81 |

### 5.3.3 多粒度特征提取对实验结果的影响

多粒度特征提取主要体现在交易时间的特征提取上,其用不同的粒度划分时间窗口,分别以日、时、分、3秒为粒度划分,逐渐缩小窗口的粒度,在不同的粒度上,提取交易时间的浅层特征。为了验证多粒度特征提取方法的有效性,本文从20维的交易时间特征 TTF 入手,将不同粒度的特征进行组合,测试了不同情况下的实验结果。使用表 1 中的超参数,进行实验并记录每次实验结果。如表 3 所列,秒粒度特征提供了单粒度实验的最好分类结果,在测试集上的准确率和召回率分别达到了 69.00% 和 62.70%。当在秒粒度特征上组合其他粒度特征时,都获得了比单粒度更好的结果,其中秒粒度与分粒度的特征组合得到了双粒度特征实验的最好分类结果,准确率和召回率分别达到了 79.42% 和 76.22%。当在秒粒度、分粒度和小时粒度的特征组合分类效果最好,其分类的准确率和召回率分别达到了 85.62% 和 81.14%。实验结果表明,每种粒度的特征对分类结果都有不同程度的影响,粒度越小对分类结果的影响越大,秒粒度特征对实验结果的影响最大,融合 4 种粒度特征后的效果最好,分类的准确率和召回率分别达到了 91.90% 和 82.30%。

表 3 多粒度特征提取的影响

Table 3 Impact of multi-granularity feature extraction

| Multi-granularity Features |   |   |   | Evaluation Metrics |        |       |
|----------------------------|---|---|---|--------------------|--------|-------|
| S                          | M | H | D | Precision          | Recall | F1    |
| √                          |   |   |   | 69.00              | 62.70  | 65.70 |
|                            | √ |   |   | 62.29              | 61.95  | 62.12 |
|                            |   | √ |   | 61.45              | 59.81  | 60.62 |
|                            |   |   | √ | 56.89              | 47.58  | 51.82 |
| √                          | √ |   |   | 79.42              | 76.22  | 77.79 |
| √                          |   | √ |   | 76.95              | 74.97  | 75.95 |
| √                          |   |   | √ | 74.30              | 67.93  | 70.97 |
| √                          | √ | √ |   | 85.62              | 81.14  | 83.32 |
| √                          | √ |   | √ | 83.70              | 80.58  | 82.11 |
| √                          | √ | √ | √ | 91.90              | 82.30  | 86.84 |

### 5.3.4 实验结果对比分析

为了验证模型的性能,将其和传统的分类模型以及 RNN 和 LSTM 网络进行对比,采用的传统分类模型有逻辑斯蒂回归(Logistic Regression, LR)、支持向量机(Support Vector Machine, SVM)、决策树(Decision Tree, CART)、LightGBM 以及 XGboost。

实验的数据采用相同的多粒度窗口的特征提取方法,输入特征的维度保持一致。对所有的模型进行调优后,对比每个模型测试集的准确率、召回率和 F1 值。如表 4 所列,神经网络模型和传统的分类模型的性能各有优劣,其中 LightGBM 和 XGboost 的性能优于 RNN,并且和 LSTM 网络的性能相近,而本文提出的 Bi-LSTM 模型在准确率、召回率以及 F1 值上都优于传统分类的模型以及 RNN 和 LSTM 网络,分别高出 1%~5%,2%~14% 和 1%~10%。

表 4 模型评价指标对比

Table 4 Comparison of model evaluation metrics

(单位:%)

| Model    | Evaluation Metrics |        |       |
|----------|--------------------|--------|-------|
|          | Precision          | Recall | F1    |
| LR       | 89.57              | 79.42  | 84.19 |
| SVM      | 92.52              | 86.49  | 89.40 |
| CART     | 89.62              | 89.75  | 89.69 |
| LightGBM | 93.91              | 87.39  | 90.53 |
| XGboost  | 93.89              | 91.99  | 92.93 |
| RNN      | 90.83              | 89.68  | 90.25 |
| LSTM     | 92.97              | 91.30  | 92.13 |
| Bi-LSTM  | 95.39              | 94.24  | 94.81 |

为了验证模型的泛化能力,另外抽取了 2019 年 2 月份玻璃期货合约 FG905 和 5 月份动力煤期货合约 ZC909 的交易数据。使用本文提出的多粒度窗口特征提取方法提取特征,对训练好的模型进行测试,结果如表 5 所列。在玻璃期货 FG905 的数据上,所提模型分类的准确率和召回率分别达到了 92.37% 和 91.48%;在动力煤期货 ZC909 的数据上分别达到了 94.01% 和 92.20%。实验结果表明,所提多特征信息融合的 Bi-LSTM 网络模型具有良好的泛化能力。

表 5 泛化能力实验结果

Table 5 Experiment results of generalization ability

(单位:%)

| Futures varieties | Evaluation Metrics |        |       |
|-------------------|--------------------|--------|-------|
|                   | Precision          | Recall | F1    |
| FG905             | 92.37              | 91.48  | 91.92 |
| ZC909             | 94.01              | 92.20  | 93.10 |

**结束语** 关联交易行为检测一直是期货交易所的重点工作,本文提出的基于交易行为的多粒度窗口特征提取方法将两个账户的数据融合为一条特殊数据,在原始数据中提取浅层特征。通过所提 Bi-LSTM 网络模型,从浅层特征中利用向前向后两个方向的上下文关系深度挖掘账户间的关联关系,并在模型训练中引入了 Dropout 策略,缓解了模型训练时的过拟合问题。在真实的数据上的实验结果表明,本文提出的多信息特征融合的 Bi-LSTM 网络模型在分类的性能上要优于传统的分类模型以及 RNN 和 LSTM 网络,其中准确率和召回率分别提高了 1%~5%,2%~14%。

本文对关联交易行为的研究还处于初级阶段,接下来研究工作的重点是继续挖掘关联交易行为的浅层特征,并在现阶段工作基础上检测多个账户之间的关联交易行为。

## 参考文献

- [1] HSU C H, LEE H C. Insider trading and information revelation with the introduction of futures markets[J]. *Economic Modelling*, 2014, 43:173-182.
- [2] MENG K. The China Securities Regulatory Commission revise the management method of Futures Exchange [EB/OL]. (2019-11-16) [2019-11-16]. [http://epaper.zqrb.cn/html/2019-11/16/content\\_530969.htm](http://epaper.zqrb.cn/html/2019-11/16/content_530969.htm).
- [3] CHEN C, ZHAO L, BIAN J, et al. Investment behaviors can tell

- what inside: Exploring stock intrinsic properties for stock trend prediction[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2019: 2376-2384.
- [4] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2818-2826.
- [5] ZHU Z K, LI J B. Multi-Feature Information Fusion LSTM-RNN Detection for OSA[J]. Journal of Computer Research and Development, 2020, 57(12): 2547-2555.
- [6] SRIVASTAVA N, HINTON G, KRIZHE A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [7] FRANKE M, HOSER B. On the analysis of irregular stock market trading behavior[M]//Data Analysis, Machine Learning and Applications. Berlin: Springer, 2008: 355-362.
- [8] PALSHIKAR G K, APTE M. Collusion set detection using graph clustering[J]. Data Mining & Knowledge Discovery, 2008, 16(2): 135-164.
- [9] CAO H B, YU M O, PHILIP S, et al. Detecting abnormal coupled sequences and sequence changes in group-based manipulative trading behaviors[C]//Association for Computing Machinery. 2010: 85-94.
- [10] ZHANG X J, ZHANG T L, ZHANG Z H. a clustering detection method of related transaction behavior for futures market[J]. Computer Applications and Software, 2016(12): 275-278.
- [11] ZHAI J, CAO Y, YAO Y, et al. Coarse and fine identification of collusive clique in financial market[J]. Expert Systems with Applications, 2017, 69: 225-238.
- [12] HUANG B, LIU Z, CHEN J, et al. Behavior pattern clustering in blockchain networks[J]. Multimedia Tools & Application, 2017, 76(19): 20099-20110.
- [13] WANG J, ZHOU S, GUAN J. Detecting potential collusive cliques in futures markets based on trading behaviors from real data[J]. Neurocomputing, 2012, 92: 44-53.
- [14] YU X P. Related account group identification technology in futures market[D]. Shanghai: Shanghai JiaoTong University, 2013.
- [15] MA X F, FENG Z. identification method of suspected related account based on transaction behavior[J]. The Foreland of Trading Technology, 2017(12): 52-56.
- [16] MIAO Z C, SHI G W, LU J D, et al. An Automatic Identification System for Related Behaviors of Financial Derivatives Investors: CN 109636608 A[P]. 2019-04-16.
- [17] HU L. Associated account identification method and electronic equipment; CN 110879816 A[P]. 2020-03-13.
- [18] SUNNY M, MASWOOD M, ALHARBI A G. Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model[C]//2020 2nd Novel Intelligent and Leading Emerging Sciences Conference(NILES). 2020: 87-92.
- [19] XIE Y, LIANG R Y, LIANG Z Y, et al. Speech Emotion Classification Using Attention-Based LSTM[J]. IEEE/ACM Transactions on Audio, Speech and Language Processing(TASLP), 2019, 27(11): 1675-1685.



**ZHANG Yuan**, born in 1996, postgraduate, is a student member of China Computer Federation. His main research interests include data mining and financial big data.



**ZHANG Zhi-hong**, born in 1965, Ph.D., professor, is a senior member of China Computer Federation. His main research interests include financial big data, blockchain, and distributed computing.

(责任编辑:喻黎)