doi:10.11835/j.issn.1000-582X.2020.11.011

以太坊庞氏骗局的类型分析与识别方法

喻文强,张艳梅,李梓宇,牛 娃
(中央财经大学 信息学院,北京 100081)

摘要:随着区块链投资领域投资者的增多,隐藏在智能合约中的庞氏骗局的影响也愈发恶劣。目前虽然有一些研究人员已经开始关注区块链上的庞氏骗局问题,但大部分还是停留在检测的层面上。将在现有的以太坊庞氏骗局检测方法的基础上进行进一步的研究,提出一种新颖的以太坊庞氏骗局类型识别方法。该方法基于智能合约的源代码和交易记录,通过分析提取关键词,将关键词与待测合约的源代码进行匹配,再结合交易记录的逻辑,进行二次分析,从而判断该合约属于哪一种骗局类型。在以太坊真实数据集上的实验表明:该方法的分类结果与人工分类的结果相比,分类准确率可以达到 80%。研究有助于研究人员和投资者更加深入的了解以太坊智能合约庞氏骗局的本质。

关键词:区块链;以太坊;庞氏骗局;骗局识别;骗局分类
中图分类号:TP391.1 **文献标志码**:A **文章编号**:1000-582X(2020)11-111-10

Study on type analysis and identification of Ethereum Ponzi scheme

YU Wenqiang, ZHANG Yanmei, LI Ziyu, NIU Wa
(School of Information, Central University of Finance and Economics, Beijing 100081, P. R. China)

Abstract: As the number of investors in the blockchain investment field increases, the impact of Ponzi schemes hidden in smart contracts becomes worse. Although some researchers have begun to pay attention to the Ponzi scheme in the blockchain, most of them remain at the level of detection. This paper will conduct further research on the basis of the existing Ethereum Ponzi scheme detection method, and propose a novel Ethereum Ponzi scheme type identification method. The method is based on the source code and transaction record of the smart contract. By analyzing the extracted keywords, we match the keywords with the source code of the contract to be tested, then combine the logic of the transaction record, and performe a secondary analysis to determine which type of scam the contract belongs to. Experiments on the real dataset of Ethereum show that the classification accuracy of the method can reach 80% compared with the results of manual classification. This study will help researchers and investors better understand the nature of ethereum smart contract ponzi scheme.

Keywords: blockchain; Ethereum; Ponzi scheme; scam identification; scam classification

---

收稿日期:2020-07-19
基金项目:国家自然科学基金资助项目(61602536, 61773415,61672104);北京市社会科学基金重点资助项目(16YJA001)。
Supported by National Natural Science Foundation of China(61602536 61773415,61672104) and Beijing Social Science Foundation(16YJA001).
作者简介:喻文强(1996—),男,硕士研究生,主要从事数据挖掘和区块链数据分析方向研究。
通信作者:张艳梅(1976—),女,博士,副教授,主要从事数据挖掘、商务智能和服务计算方向研究,(E-mail)jlzym0309@sina.com。

近几年区块链技术发展迅猛,越来越多的投资者将目光放到了这个新的领域上面。以太坊[1]是一个基于开源的区块链分布式平台,被誉为区块链2.0。在以太坊上可以部署一系列的智能合约。智能合约的本质是一段可以实现特定功能的代码[2],它是相互不信任的参与者之间的协议,当满足协议预设条件的时候,就会自动触发执行。智能合约一旦执行,无法人工终止,不依赖于任何中心机构。目前智能合约已经在各个领域得到了广泛运用[3-5]。

庞氏骗局[6]是一种传统的投资骗局,其典型的特征就是用新进场的投资者提供的资金来向现有的投资者支付所谓的回报。在以太坊智能合约中,庞氏骗局又有了一些新的特性[7]。最明显的就是它基于区块链的匿名性,研究人员无法知道合约发起者的真实身份,不能对其信用信息进行关联,只能通过以太坊官网公开的信息进行分析。此外,它的代码是公开的、不可变的,并且是自动强制执行的,所以投资者会对它产生一种信任感,降低防范意识,正因为如此,智能合约上的庞氏骗局也层出不穷,许多投资者因为看不懂智能合约的源代码而错误地投资了庞氏骗局,最终损失惨重。此外,由于区块链的匿名性和难以溯源的特点,受骗资金基本不可能被成功找回,只能让投资者蒙受损失。许多区块链庞氏骗局都被爆出获得巨额利润[8-10],单单2013年到2014年之间,比特币[11]的庞氏骗局就获得了700万美元的非法利益[12]。因此,加强对区块链市场的监管很有必要[13]。

当前对以太坊庞氏骗局的研究主要是对其特性的研究以及如何对庞氏骗局进行快速高效的检测识别。有学者[14]从论坛的帖子里分析庞氏骗局的特性以及对投资者的影响,也有学者[15]通过对比分析待测智能合约与被手动标记为庞氏骗局智能合约的智能合约组之间的字节码相似度来判别待测智能合约是否为庞氏骗局,还有学者[16]从交易记录出发,分析提取了多项用户特征,同时结合了操作码特征,在收集到的智能合约数据中进行监督训练之后,发现可以有效识别以太坊上的庞氏骗局。但目前大多数研究都只是停留在骗局检测的层面,在文献[15]中虽然提出将以太坊中庞氏骗局按源代码逻辑分为四个类别,但是也未进行深入的研究分析。笔者对庞氏骗局的类型进行深入分析,进一步揭示各种类型以太坊庞氏骗局的特征和规律。

主要贡献是:

1)采用案例分析法对每个具体的类别进行案例分析,以便找出每一类的显著特征;

2)依据智能合约源代码和交易记录的特征结合关键词提取方法和交易记录分析来实现骗局类型的自动识别。

1 相关工作

相关工作主要包含两类,第一类是对区块链庞氏骗局生态的研究,第二类是对区块链庞氏骗局的检测技术研究。下面对这2个方面的相关工作进行详细阐述。

1.1 区块链庞氏骗局生态特征研究

Marie Vasek 和 Tyler Moore[14]通过分析比特币论坛中庞氏骗局的广告帖的内容对基于比特币的庞氏骗局的生态进行分析,发现大多数受骗人在诈骗广告帖发布的前五天内会发帖投诉。然后用Cox比例风险模型对庞氏骗局存活周期数据进行分析,发现只发一次的庞氏骗局的骗子与受骗人的交流越多,骗局存活周期越长,而受骗人和发布多次骗局的骗子发布的帖子越多,骗局的存活周期越短。Massimo Bartoletti 等人[17]从比特币论坛上面手动搜索高收益投资项目的广告,然后访问这些广告的网站并寻找它们的比特币地址。由于之前对区块链数据集的收集大都是采用人工或者半自动下载数据集[18-19],他们开发了一个爬虫程序来对 blockchain.info/tags(一个允许用户标记比特币地址的网站)上的比特币地址链接的网站进行自动解析,并且根据它们页面中包含的与庞氏骗局相关的词汇的个数对它们进行排序。很多情况下,庞氏骗局会使用多个地址,为了检索这些地址,使用“多输入”的启发式算法对集合中的地址进行聚类。同时,Massimo Bartoletti 等人[15]作为第一批研究基于以太坊智能合约的庞氏骗局的研究者,他们从 etherchain.org 网站上搜集了811份公开源代码的智能合约,并将带有源代码的庞氏骗局按照代码的逻辑类型分成了4个类别,投资者可以通过观察并判断合约源代码的逻辑类型来判断该合约是否是庞氏骗局。此外,他们还通过列数据、分析案例的形式介绍了以太坊庞氏骗局的常用欺诈方式和经济影响。

1.2　区块链庞氏骗局检测技术研究

针对比特币，Massimo Bartoletti 等人[17]设计一组与庞氏骗局分类相关的特征（合约活跃天数，最大日交易额，用户数，用户平均投资金额等），然后使用 F-score、AUC 等度量指标来评估不同的监督学习分类算法模型的效果，最后发现随机森林算法的检测效果最好，可以成功找到 32 个庞氏骗局中的 31 个。针对以太坊智能合约，Massimo Bartoletti 等人[15]提出可以通过使用 NLD[20]（normalized Levenshtein distance）来计算待测合约与已收集庞氏骗局合约组之间的最大相似度来判断待测合约是否是庞氏骗局。文献[16]中从智能合约的交易数据中提取出交易特征，再结合智能合约的操作码中提取出的操作码频率特征，使用 XGBoost 二分类模型训练这些数据特征，来判断待测合约是否为庞氏骗局。Massimo Bartoletti 等人[15]将检测出的以太坊庞氏骗局合约分为 4 个类型，但是并未对这 4 个类型进行更加深入的研究与分析。所以，对识别出来的庞氏骗局合约进行自动分类，更深入地探析每种类型的特征和规律。

2　庞氏骗局合约案例分析

对 130 个庞氏骗局合约的源代码进行分析之后，确定了和 Massimo Bartoletti 等人[15]相同的 4 个庞氏骗局类别，并对每一种庞氏骗局类型进行深入案例分析，以便找出其特征，下面一一举例介绍：

2.1　基于树结构的庞氏骗局

基于树的金字塔类型使用树结构来记录用户的地址。每个投资者都有一个邀请者，树的根部除外，因为他是合约的所有者。投资者的钱会在他的祖先之间进行分配。如图 1 所示，投资者要想加入该合约，他必须投入一些钱并指明其邀请者，如果投资金额太少，或用户已经存在，或者邀请方案不存在（第 3～4 行），则拒绝该投资者进入合约，若全部符合条件，则将该投资者的信息插入树结构中（第 8 行），该投资者投资金额由其祖先共享，每层减半（第 13 行）。

```
1   function enter(address inviter) public {
2     uint amount = msg.value;
3     if ((amount < contribution) || (Tree[msg.sender].inviter !=
4     0x0) || (Tree[inviter].inviter == 0x0)) {
5       msg.sender.send(msg.value);
6       return;
7     }
8     addParticipant(msg.sender, inviter);
9     address next = inviter;
10     uint rest = amount;
11    uint level = 1;
12    while ( (next != top) && (level < 7) ){
13      uint toSend = rest/2;
14      next.send(toSend);
15      Tree[next].totalPayout += toSend;
16      rest -= toSend;
17      next = Tree[next].inviter;
18      level++;}
19    next.send(rest);
20    Tree[next].totalPayout += rest;
21  }
```

图 1　Etheramid 核心功能源代码

Fig. 1　The core function source codes of Etheramid

投资者的主要回款来源是发展下线，然后获得下线投资额一定比例的回款收益。这一类庞氏骗局的生存周期较长，因为新进入的投资者总会源源不断地发展下线，以此来保证自己的收益。分析 Etheramid 的合约交易记录之后发现该合约涉及到了共 98 个账户，其中有 20 个账户投资回报率大于 1（占总账户数目的 20.41%），合约生命周期为 27 d 左右，影响生命周期的因素即旧的投资者是否能够动员新的投资者进入，并且新投资者如果能够邀请更多的人，那么这个投资者也可以获利。

2.2　基于数组结构的庞氏骗局

基于数组的金字塔骗局按到达顺序向用户回款。一般来说，这类骗局承诺将投资乘以预先指定的因子。当用户从后来加入该计划的用户那里筹集到足够的资金时，她就可以赎回成倍的投资。如图 2 所示，为了加

入这个合约，用户投资 msg.value，从而触发函数（第 1 行），该合约要求用户投资最小数额要大于 500 芬妮（第 2 行），如果这个投资者的投资数额小于 500 芬妮，他会被拒绝进入该合约，否则，这位投资者的地址会被加入 array 中。投资者进入后，合约更新 balance（第 8 行）。判断 balance 如果足够支付在数组中等待被回款的投资者，那么合约将给与该投资者投资额的 2 倍（第 14 行）。之后，合约试图支付数组中下一个等待回款的投资者，直到 balance 不满足判断条件。

```
1   function enter() {
2       if (msg.value > 500 finney) {
3       uint Amount=msg.value;
4       Total_Players=depositors.length+1;
5       depositors.length += 1;
6       depositors[depositors.length-1].EtherAddress = msg.sender;
7       depositors[depositors.length-1].Amount = Amount;
8       Balance += Amount;
9       Total_Deposited+=Amount;
10      uint payout;
11      uint nr=0;
12       while (Balance > depositors[nr].Amount * 200/100 &&
          nr<Total_Players)
13      {
14      payout = depositors[nr].Amount *200/100;
15      depositors[nr].EtherAddress.send(payout);
16      Balance -= depositors[nr].Amount *200/100;
17      Total_Paid_Out += depositors[nr].Amount *200/100;
18      }
19  }
```

图 2　CrystalDoubler 核心功能源代码

Fig. 2　The core function source codes of Crystal Doubler

在这一合约中，投资者有可能得到数额为投资的两倍的回款，但是，只有排名靠前的投资者可以得到，当资金池没有后续资金来源的时候，资金链就会断裂，排在后面的投资者就会血本无归。分析 CrystalDoubler 的合约交易记录之后发现该合约涉及到了共 4 个账户，只有第一个账户收到了回款，但该账户共收到 5 次回款，每次回款额为该账户投资额的 2 倍。经分析发现，源代码回款函数中少写了一个'nr＋＋'，这样数组就不会往下遍历，只有数组中第一个投资者能收到回款。如表 1 所示，分析另一个数组结构类型庞氏骗局合约 LuckyDoubler 发现该合约涉及 5 个账户，前面几个账户都收到了回款，且金额为投资额的 1.25 倍，只有最后 2 个账户未收到回款。该类合约的回款速度比较慢，投资热度下降比较快，生命周期较短，收益高、风险高，前期投机获利明显。

表 1　LuckyDoubler 交易记录统计

Table 1　The Trasaction statistics of Lucky Doubler

交易账户	投资次数/次	投资总金额/以太	回款次数/次	回款总金额/以太	投资回报率
0x 61 964	4	4	19	4.55	1.137 5
0xFe0A3	5	5	5	6.25	1.25
0x 68 483	3	3	1	1.25	0.416 666 667
0x710F7	1	1	0	0	0
0x07e2E	1	1	0	0	0

2.3　瀑布结构的庞氏骗局

瀑布类型旁氏骗局从第一个投资开始，将每个新加入的投资分配给已经加入的投资者。只要有足够的钱，每个投资者都能得到固定比例的投资。如图 3 所示，合约将收到的金额的 50％给合约的所有者（第 4～8 行），剩下的回款给之前的一些投资者。如果余额足够支付数组中的第一个投资者，则合约将该投资者原始

投资以一固定比例发送给该投资者(第 9～11 行)。然后，合约向数组中的下一个投资者回款，以此类推，直到余额不足够支付为止。

```
1   function invest() payable
2       accreditedInvestor()
3       {
4           fee = amount/2;
5           balances[owner] += fee
6           uint dividend = msg.value;
7           uint fee = ownerFee(dividend);
8           dividend -= fee;
9           for (uint i = 0; i < investors.length; i++) {
10                  balances[investors[i]] += dividend *
                        invested[investors[i]] / total;
11          }
12          if (invested[msg.sender] == 0) {
13              investors.push(msg.sender);
14              invested[msg.sender] = msg.value;
15          } else {
16              invested[msg.sender] += msg.value;
17          }
18          total += msg.value;
19          LogInvestment(msg.sender, msg.value);
```

图 3　PonzICO 核心功能源代码

Fig. 3　The core function soure codes of ponzICO

分析 PonzICO 合约的交易记录之后发现该合约共涉及 27 个交易账户，生命周期为 280 d 左右，影响生命周期因素为是否有新的投资者进入，该合约中 27 个投资者仅有 7 个投资者(分布在前 6 d)获利(25.93%)，可发现这类合约投资者相对更加追求低风险低收益。

2.4　转移权限庞氏骗局

转移权限庞氏骗局只存储最后一个投资者的地址，若有投资者想要加入，他必须偿还最后一个投资者的投资和固定利息。根据这个规则，每个投资者每次应支付的金额将增加。如图 4 所示，投资者要加入该合约，须向合约输入 startingAmount。合约将这笔金额转交给前投资者(第 10 行)，记录新投资者的地址(第 11 行)，金额变为之前的两倍(第 12 行)。

```
1   function() payable {
2       if(round == 1) {
3           if(msg.value != startingAmount) {
4               throw;
5           }
6       } else {
7           checkAmount(msg.value);
8           lastDepositor.send(msg.value);
9       }
10      lastDepositorAmount = msg.value;
11      lastDepositor = msg.sender;
12      nextAmount = msg.value * 2;
13      increaseRound();
14  }
```

图 4　PonziScheme 核心功能源代码

Fig. 4　The core function source code of PonziScheme

分析 PonziScheme 合约的交易记录之后发现该合约共涉及 10 个交易账户，只有最后投资的投资者没有收到回款。四类合约中，此类合约极容易崩溃，因为倍数增长的速度远快于前面三类合约。

对于投资者来说，投资者仅能根据合约描述或者交易记录来判断是否投资。合约描述错综复杂，难以有效判断。从交易记录来看，第四类转移权限庞氏骗局与部分第二类基于数组的庞氏骗局交易记录较少，难以有效区分二者；第一类基于树结构的庞氏骗局与第三类瀑布结构庞氏骗局投资与回款记录较多且较为复杂，分析二者中每个用户回款与投资记录比可以发现比值都小于 1，不利于判断。仅从交易记录难以判断不同的

骗局,需要我们做更进一步的分析与处理。对于研究人员来说,仅从源代码逻辑上来自动判断庞氏骗局合约类别是很难实现的,因为源代码还没有实现规范化,很多合约源代码的变量都是随意设置的,所以需要结合具体的真实交易记录才能更有效的判断庞氏骗局合约的具体类型。

3 基于关键词提取和交易记录分析的以太坊庞氏骗局分类

使用的骗局分类方法流程如图5所示,首先通过爬虫爬取庞氏骗局合约的源代码数据,然后根据源代码的具体回款逻辑,将合约按照源代码人工标记为4个类别。然后对合约的源代码回款函数部分的关键词进行提取,用正则表达式进行归纳整理,再对合约的交易记录进行分析。最后将归纳好的正则表达式对待测庞氏骗局合约进行匹配,同时使用交易记录逻辑对待测合约的类别进行判断,判定成功且则输出为该类,并输出分类结果。

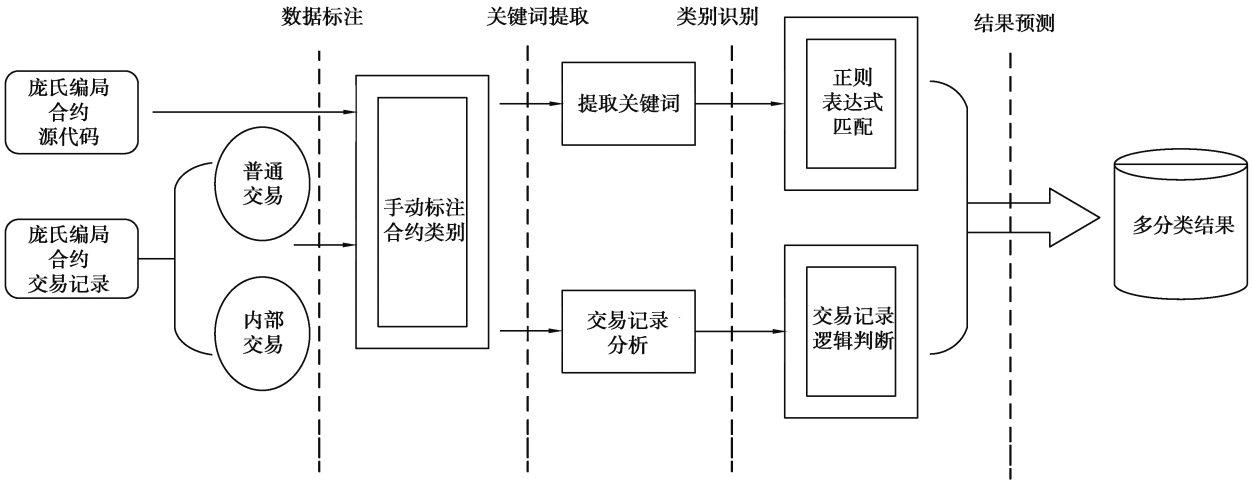

图5 骗局分类方法流程图

Fig. 5 The flow chart of ponzi scheme classification

3.1 源代码关键词选取

按照上述4种分类标准对收集到的庞氏骗局合约进行标记,之后对其中的部分庞氏骗局合约的回款功能相关代码进行分析,得出不同类型的合约,其实现该功能的关键字是不同的。为了区分树结构类型和其他类型,选择了'tree'、'Tree'、'Node'、'node'、'parent'的关键词集合,出现频率统计如表2所示,从表2可以看出这个关键词集合的区分度明显,可以作为关键词使用。同理,在区分转移权限类型和剩下两种类型合约的时候,选择了关键词'last',区分度如表3所示,效果明显。最后选择了'200','150'等大于100的数字类型的关键词来区分数组结构类型和瀑布结构类型,区分效果如表4所示,效果不是特别明显,但仍然具有不小的区分度。

表2 树结构类型关键词区分度

Table 2 The discrimination of tree structure type keyword

庞氏骗局类型	中位数	平均数
数组结构类型	0	0
瀑布结构类型	0	0.01
树结构类型	18.57	14.5
转移权限类型	0	0

表3 转移权限类型关键词区分度

Table 3 The discrimination of transfer authority type keyword

庞氏骗局类型	中位数	平均数
数组结构类型	0	0.32
瀑布结构类型	0	0.45
转移权限类型	11.0	11.0

表 4　数组结构类型关键词区分度

Table 4　The discrimination of array structure type keyword

庞氏骗局类型	中位数	平均数
数组结构类型	0	0.76
瀑布结构类型	0	0.14

3.2　交易记录逻辑分析

按照案例分析结果对智能合约的交易记录进行进一步的分析，发现如下逻辑，在树结构和瀑布结构的庞氏骗局中，一次投资可能会导致多次回款，因为树结构庞氏骗局中触发了回款条件之后会对该节点的所有父节点进行回款，二瀑布结构的庞氏骗局则是对之前所有的投资者进行回款。在数组结构的庞氏骗局中，每次回款的金额必定比该笔投资的金额大，在转移权限庞氏骗局中，后面的投资金额必定大于前面的投资金额。结合这几个逻辑，可通过看合约交易记录来判定出数组和转移权限庞氏骗局，其他 2 种仍然需要根据源代码来区分。

3.3　基于关键词匹配和交易记录的庞氏诈骗分类模型

由于机器学习需要先提取特征，而当前的智能合约源代码大多都不是很规范，变量和函数的命名都没有一个统一的标准。因此它的代码含义特征很难被提取。但交易记录是具有一定规范和逻辑的，不同类别的合约的交易会存在一定的区别。将源代码关键词提取和交易记录逻辑分析 2 种方法结合，具体算法如下：

输入：关键词集合 K，待测合约源代码 D，待测合约投资记录 V，待测合约回款记录 G

输出：庞氏骗局合约类别

1.　if 树结构类型的关键词集和 K_1 in D
2.　return 数组结构类型；
3.　for i in V['value']
4.　计算每一次的投资额，若每次的投资额都大于上一次则 flag 1 = 1
5.　if 转移权限类型关键词集合 K_2 in D or flag1 == 1
6.　return 转移权限类型
7.　for i in V['value']
8.　　for j in G['value']
9.　若每次投资的回款额都大于投资额，则 flag2 = 1
10.　if 数组结构类型关键词 K_3 in D or flag2 == 1
11.　return 数组结构类型
12.　else return 瀑布结构类型

上述算法步骤如下：先通过源代码匹配的方法判断待测合约是否为树结构的庞氏骗局，是则返回树结构类型(第 1～2 行)，然后计算待测合约的每一次的投资额，若每次投资额都大于前一次的投资额且关键词匹配成功，则返回转移权限类型(第 3～6 行)，最后计算待测合约每次的回款额和投资额，若每次投资的回款额都大于投资额，且相应的关键字匹配成功，则返回数组结构类型，否则返回瀑布结构类型(第 7～12 行)。

4　实　验

主要是对庞氏骗局的分类效果进行评估。由于目前没有相关分类方法，仅对提出的方法进行对比分析。实验将回答以下问题。

4.1　实验环境

实验数据集：数据集是文献[16]中的数据，这份数据包含 130 个庞氏骗局(包含 14 个树结构庞氏骗局、

81个数组结构庞氏骗局、33个瀑布结构庞氏骗局和2个转移权限庞氏骗局)。然后编写爬虫在etherscan.io上下载了这些合约对应的源代码和交易记录数据。

实验环境:Windows10系统,使用Spyder平台Python语言。

4.2 评价标准

实验选了Precision(查准率)、Recall(召回率)以及F-score(F值)作为评价标准。其中,Precision是所有被判定为某类别的合约中真正为该类别合约所占的比例,Recall是被检测到的合约类别数量在总的该类别合约数量中所占的比例,F-score是一个综合了Precision值和Recall值的调和指标。其求解公式如下

$$\text{Precision}=\frac{tp}{tp+fp}, \tag{1}$$

$$\text{Recall}=\frac{tp}{tp+fn}, \tag{2}$$

$$\text{F}-\text{score}=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}, \tag{3}$$

其中,tp为庞氏骗局类别判定正确的数量,fp为合约被误判为其他类别的数量,fn为其他类别合约被判定为该类别合约的数量。

4.3 分类效果评价

使用了两种方法来分别进行实验,第一种方法是只用关键词匹配,第二种方法是关键词匹配和交易记录分析结合。第一种方法的实验结果热力图如图7所示,第二种方法的实验结果热力图如图8所示,其中横纵坐标的0到3分别按顺序对应提到的4个类别。

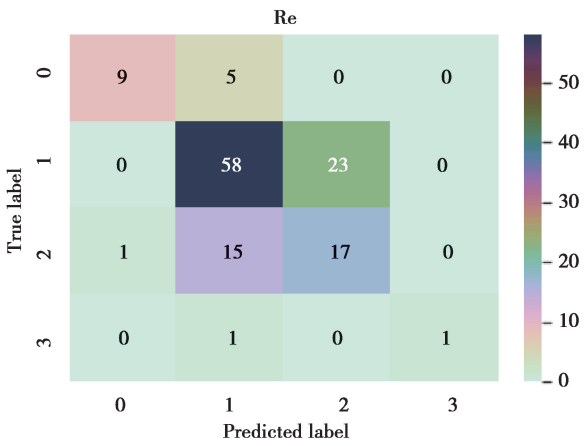

图6 关键词匹配热力图

Fig. 6 The thermodynamic diagram of keywords mathing

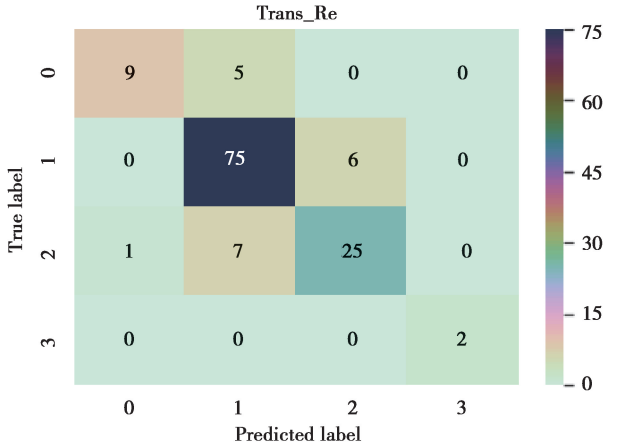

图7 关键词匹配加交易记录分析热力图

Fig. 7 The thermodynamic diagram of keyword mathing combind with transaction analysis

从图6和图7可以看出组合方法在除树结构外的其他三类中好的改善,这是因为关键词提取方法中不能提取到所有能够判别庞氏骗局类别的关键词。此外,从图6中可以看出,数组结构的庞氏骗局和瀑布结构的庞氏骗局的关键词有大量交集,即很多关键词在这两类合约中都会出现,这就导致这两类合约的区分度不够,经常会误判。而在图7中的误判率就明显下降许多,这是因为在交易记录中,数组结构的庞氏骗局和瀑布结构的庞氏骗局有2个很明显的区分点,就是瀑布结构的庞氏骗局经常是一次触发多次回款,且回款金额通常都是小于投资金额,而数组结构的庞氏骗局一次只触发一次或者少数几次回款,并且回款金额必定大于投资金额。

关键词匹配方法的综合指标分析如表5所示。

表 5　关键词匹配效果评价

Table 5　The effect evaluation of keyword matching

类别	准确率	召回率	F 值
树结构	0.9	0.64	0.75
数组结构	0.73	0.71	0.72
瀑布结构	0.43	0.52	0.47
转移权限	1.0	0.5	0.67

关键词匹配加交易记录分析方法的综合指标分析如表 6 所示。

表 6　关键词匹配加交易记录分析效果评价

Table 6　The effect evalution of keyword matching combined with transaction recored analysis

类别	准确率	召回率	F 值
树结构	0.9	0.64	0.75
数组结构	0.86	0.92	0.89
瀑布结构	0.80	0.76	0.78
转移权限	1.0	1.0	1.0

由表 5 和表 6 可以看出，在数组结构和瀑布结构的类别判定中，无论是准确率还是召回率，组合方法都比关键词匹配方法有一个很大的提升。但最终的判定结果还不是很高，需要提取更多的关键词来继续完善。

总之，选用的组合方法的效果对关键词匹配方法有一个明显的改进，但总体的 F 值还不是很好，只能作为一个参考，让投资者对庞氏骗局有一个更深入的了解，以避免冲动投机，减少不必要的损失。

5　结　论

通过对以太坊庞氏骗局智能合约的实例分析，从智能合约源代码和智能合约交易记录 2 个层面出发来讨论四类骗局的区别以及特征。同时，经过实验，发现通过关键词匹配加交易记录分析的方法可以对庞氏骗局合约进行一个有效的分类，可以更加深入的了解庞氏骗局合约。后续，考虑在如下几个方面进行深入研究：第一，扩充数据集，优化分类效果。第二，在该分类的基础上对智能合约的投资风险进行评估，进而揭示以太坊庞氏骗局投资风险的本来面目，更好的引起投资者的警惕和防范。

参考文献：

[1] Buterin V. A next-generation smart contract and decentralized application platform[J]. Etherum, 2014: 1-36.

[2] Wood G. Ethereum: a secure decentralised generalised transaction[J]. Ethereum Project Yellow Paper, 2014: 1-32.

[3] Christidis K, Devetsikiotis M. Blockchains and smart contracts for the Internet of Things[J]. IEEE Access, 2016, 4: 2292-2303.

[4] Juels A, Kosba A, Shi E. The Ring of Gyges: investigating the future of criminal smart contracts[C]// CCS '16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York, USA: ACM, 2016: 283-295.

[5] Norta A. Creation of smart-contracting collaborations for decentralized autonomous organizations[M]. Cham: Springer International Publishing, 2015: 3-17.

[6] Artzrouni M. The mathematics of ponzi schemes[J]. Mathematical Social Sciences, 2009, 58(2): 190-201.

[7] Bartoletti M, Carta S, Cimoli T, et al. Dissecting ponzi schemes on ethereum: identification, analysis, and impact[J/OL]. P2P Financial Systems International Workshop (P2PFISY 2017), 2019[2020-09-29]. https://arxiv.org/abs/1703.03779.

[8] Higgins S. SEC seizes assets from alleged altcoin pyra-mid scheme[EB/OL]. https://www.coindesk.com/sec-seizesalleged-altcoin-pyramid-scheme.

[9] Keirns G. 'Gemcoin' ponzi scheme operator hit with $74 million judgment[EB/OL]. https://bitcoinwiki.co/gemcoinponzi-scheme-operator-hit-with-74-million-judgment/.

[10] Morris D Z. The rise of cryptocurrency ponzischemes[EB/OL].https://www.theatlantic.com/technology/archive/2017/05/cryptocurrency-ponzi-schemes/528624/.

[11] Bonneau J, Miller A, Clark J, et al. SoK: research perspectives and challenges for bitcoin and cryptocurrencies[C]// 2015 IEEE Symposium on Security and Privacy. Piscataway, NJ: IEEE, 2015: 104-121.

[12] Vasek M, Moore T. There's no free lunch, even using bitcoin: Tracking the popularity and profits of virtual currency scams[C]// Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Berlin, Germany: Springer, 2015, 8975: 44-61.

[13] Elwell C K, Murphy M M, Seitzinger M V. Bitcoin: questions, answers, and analysis of legal issues[J/OL]. Virtual Currencies: Regulatory and Tax Compliance Issues.2014[2020-09-29]. https://fas.org/sgp/crs/misc/R43339.pdf.

[14] Vasek M, Moore T. Analyzing the bitcoin ponzi scheme ecosystem[J]. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019, 10958: 101-112.

[15] Bartoletti M, Carta S, Cimoli T, et al. Dissecting Ponzi schemes on Ethereum: Identification, analysis, and impact[J]. Future Generation Computer Systems, 2020, 102: 259-277.

[16] Chen W, Zheng Z, Cui J, et al. Detecting ponzi schemes on ethereum: towards healthier blockchain technology[C]// WWW '18: Proceedings of the 2018 World Wide Web Conference. New York, USA: ACM Press, 2018: 1409-1418.

[17] Bartoletti M, Pes B, Serusi S. Data mining for detecting bitcoin ponzi schemes[C]// 2018 Crypto Valley Conference on Blockchain Technology (CVCBT). Piscataway, NJ: IEEE, 2018: 75-84.

[18] Brenig C, Accorsi R, Möller G. Economic analysis of cryptocurrency backed money laundering[J/OL]. 23rd European Conference on Information Systems, 2015[2020-09-29]. http://aisel.aisnet.org/ecis2015_cr/20

[19] Moser M, Bohme R, Breuker D. An inquiry into money laundering tools in the Bitcoin ecosystem[C]// 2013 APWG eCrime Researchers Summit. Piscataway, NJ: IEEE, 2013:1-14.

[20] Li Y J, Liu B. A normalized levenshtein distance metric[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1091-1095.

(编辑 侯 湘)