区块链 DPoS 共识机制改进研究

张星星,何利文
(南京邮电大学,江苏 南京 210003)

摘　要:区块链共识算法可用于增强物联网安全性,提高网络节点之间的协作效率。委托权益证明(Delegated Proof of Stake,DPoS)可以同时满足低成本和高效率的要求,提高节点协作的服务质量。然而,协作节点的恶意攻击、自私行为和投票积极性不高都会影响 DPoS 的共识过程。针对这些挑战,该文对 DPoS 共识机制进行了改进,为了提高节点的投票积极性,提出了一种信任值模型,根据节点行为将信用评价指标分为“交易情况”“性能”“信用级别”三个一级指标以及对应的二级指标,并采用动态分配二级指标权重的方法对节点信任值进行计算,从而使选出的节点更加可信。同时,针对恶意节点以及自私行为,提出了一种基于高斯混合模型的异常节点剔除算法,对投票数据进行划分,计算其混合高斯概率密度值,并设定阈值,将低于阈值的节点剔除,从而识别并剔除异常数据。相对于传统的 DPoS,改进后的 DPoS 节点出块速率以及异常节点剔除率都有显著提升。

关键词:区块链;DPoS 共识机制;信任值模型;高斯混合模型;共识节点
中图分类号:TP399　　**文献标识码**:A　　**文章编号**:1673-629X(2023)09-0078-05
doi:10.3969/j.issn.1673-629X.2023.09.012

Research on Improvement of Blockchain DPoS Consensus Mechanism

ZHANG Xing-xing,HE Li-wen
(Nanjing University of Posts and Telecommunications,Nanjing 210003,China)

Abstract:The blockchain consensus algorithm can be used to enhance the security of the Internet of Things and improve the cooperation efficiency between network nodes. Delegated Proof of Stake (DPoS) can meet the requirements of low cost and high efficiency at the same time,and improve the service quality of node collaboration. However,malicious attacks,selfish behavior and low voting enthusiasm of cooperative nodes will affect the consensus process of DPoS. In response to these challenges,we improve the DPoS consensus mechanism. In order to improve the voting enthusiasm of nodes,a trust value model is proposed. According to the behavior of nodes,credit evaluation indicators are divided into three first-level indicators:“transaction status”“performance”“credit level” and the corresponding second-level indicators. The method of dynamically allocating the weight of the second-level indicators is used to calculate the trust value of nodes,which makes the selected node more reliable. At the same time,aiming at malicious nodes and selfish behaviors,an outlier removal algorithm based on Gaussian mixture model is proposed,which divides the voting data,calculates its Gaussian mixture probability density value,and sets a threshold to remove nodes below the threshold,so as to identify and remove outlier data. Compared with the traditional DPoS,the improved DPoS has significantly improves node block rate and exception node rejection rate.

Key words:blockchain;DPoS consensus mechanism;trust value model;Gaussian mixture model;consensus node

0　引　言

2008 年 11 月 1 日,中本聪发表的《比特币白皮书:一种点对点的电子现金系统》[1] 标志着比特币的诞生,同时也将区块链展示在世人面前。区块链本质上是一种数据结构,由封装数据信息的块按照时间顺序链接。区块链使用非对称加密的分布式账本来确保区块链中数据的安全性和可靠性。随着区块链技术的不断发展,它将对金融、教育改革、物联网、人工智能等产生深远影响[2]。共识算法作为区块链的核心部分,它的效率直接决定了区块链的性能。共识算法解决的是区块链数据记录的合法性与数据存储的同一性,它的漏洞很容易被不法分子利用并对网络产生巨大的破坏作用。因此,研究更加高效安全的共识算法对推动区块链技术的广泛运用有着重要且深远的现实意义。

随着区块链技术的发展、应用场景和协议条件选择的不同,提出了许多不同的共识算法,如用于区块链中的共识算法 PoW、PoS、DPoS,用于非拜占庭网络的 Paxos 算法和 Raft 算法、解决拜占庭问题的 BFT 和

收稿日期:2022-12-19　　**修回日期**:2023-04-20
作者简介:张星星(1998-),女,硕士研究生,通信作者,研究方向为信息安全;何利文(1968-),男,博士,教授,研究方向为网络、信息安全、云计算大数据分析与应用。

PBFT 算法等[3]。

PoW 共识算法具有简单的验证和安全性,一个 nonce 的验证只需要两次 SHA-256 操作,此验证模式可防止节点被伪 nonce 欺骗。此外 PoW 具有“51% 攻击”的容错率[4],只有当攻击者拥有整个计算资源的 51% 以上时,它才有可能修改已上链的区块链信息,这是不现实的。但与此同时 PoW 也带来了过度的资源浪费,并且需要确认交易的时间过长。PoS 共识算法是为了弥补 PoW 算法的不足而产生的,PoS 的核心思想是“节点的权益越大,更容易获得记账权”[5]。PoS 算法是在一个有限的空间里进行共识,不需要消耗过多的外部算力和资源,因此可以有效地弥补 PoW 的劣势,并且能够在一定程度上缩短达成共识的时间,提升系统运行性能[6]。虽然在一定程度上减少了系统的挖矿时间,但本质上还是需要挖矿,依然会造成算力浪费。

Dan Larimer 设计并提出了委托权益证明机制(Delegated Proof of Stake,DPoS)[7],它是 PoS 的一种演化版本。在 DPoS 算法中,所有的节点都可以通过投票来选举代理节点,被选举出的代理节点按照一定的规则负责区块生成及验证。如果代理节点出现问题,例如没有在规定的时间内产生区块,那么它就会失去代理权。相比于 PoS 算法,DPoS 减少了参与验证区块的节点数量,提升了区块确认速率[8],同时也降低了能耗,区块链系统的性能得到了进一步的提升。但 DPoS 也存在一些缺点,总结来说有以下四点:

(1)投票积极性不高,大多数节点只是持股,从来不参与投票。

(2)垄断性高,只有持币的人才能参与区块验证。

(3)没有对错误节点进行快速剔除,不仅影响代理节点投票结果,还增加了投票周期,耗费资源[9]。

(4)恶意节点贿赂投票节点导致“腐败攻击”,破坏整个系统。

针对上述缺点,国内外一些学者对 DPoS 算法进行了改进。针对 DPoS 中没有生成块故障的处理方案的问题,Tan C 和 Xiong L[10]将块生成的节点故障行为记录为下一次选择见证节点的投票数的计算因子,以降低恶意节点再次被选为见证节点的概率,但并没有考虑到大规模并发问题,用于商用时还需考虑吞吐量问题。Chen Y 和 Liu F[11]考虑时间动态因素的声誉模型,构建了基于声誉的投票机制和奖惩激励机制,实现了对节点的声誉激励,并设计了一种新的计票方法,提高了选举效率。何帅等[12]针对 DPoS 共识机制存在恶意节点相互勾结以及权益分配不合理的两大问题,引入 RBF 神经网络模型,计算综合信任值,使得通过综合信誉值选举出的节点更加权威可信;同时,加入基于动态博弈的信誉激励机制,利用沙普利值对节点权益进行合理划分,使得节点的权益得到了分散,增强了“去中心化”程度。但与神经网络算法的结合增加了整体算法的空间复杂度,反而降低了运行效率。

该文提出一种信任评估模型,使用信任评估模型计算每个节点的信任值,根据节点行为将信用评价指标分为“交易情况”“性能”“信用级别”三个一级指标,每个一级指标下划分若干二级指标,动态分配二级指标权重,将二级指标根据量化函数进行量化后,根据对应的权重对节点信任值进行计算并排序,从而使选出的节点更加可信。同时,针对恶意节点以及自私行为,提出了一种基于高斯混合模型的异常节点剔除算法,首先对投票数据情况进行划分,根据高斯混合模型计算其混合高斯概率密度值,并设定阈值,将计算出的混合高斯概率密度值低于阈值的节点剔除,从而达到在节点的恶意攻击和自私行为中识别并剔除异常数据的目的。最后通过仿真实验得出结论。

1　DPoS 共识机制的改进

1.1　系统模型

文中系统模型如图 1 所示。首先通过信任评估模型对节点信任值进行排序,利用基于高斯混合模型剔除算法剔除节点的异常投票,将最终投票数与信任值结合选出排名为前 2TN 的节点为候选节点,排名为前 TN 的节点为见证节点,见证节点轮流产生区块并验证区块信息,然后开始下一轮出块权力竞争,n 轮区块产生之后开始下一轮见证节点以及候选节点的竞选。

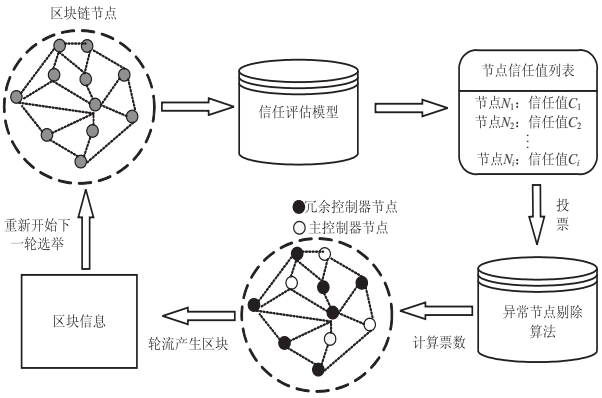

图 1　系统模型

1.2　信任值模型

基于 DPoS 共识算法的区块链节点主要包含三种类型:普通节点、见证人节点以及候选节点。普通节点在共识过程中负责投票选出见证节点,而见证节点负责产生区块。当见证节点没有在规定时间内产生新的区块或者产生了无效块,那么 DPoS 会跳过这个节点并由下一个候选节点负责区块的产生。为了使选出的见证节点更加安全可信,构建了信任值模型,结合每个

节点的行为对节点做出评估得出信任值并加入到选举机制当中。如表 1 所示,根据节点行为将信用评价指标分为交易情况、性能、信用级别三个一级指标。节点的交易情况由二级指标交易总量 TN_i(Total Number of transactions)和货币流动性 CL_i(Currency Liquidity)来衡量。节点性能由网络延迟 NL_i(Network Latency)、节点下线次数 OTN_i(Offline Times of Nodes)和节点活跃度 NA_i(Node Activeness)来衡量。信用级别由有效区块数 EP_i(Effective block ratio)和上一轮节点信任值 C_i(Credit degree of the lastround of nodes)来衡量。

表 1 节点信用评价指标

一级指标	二级指标(IN_i)	权重(W_i)
交易情况	交易总量(TN_i)	t_1
	货币流动性(CL_i)	t_2
性能	网络延迟(NL_i)	p_1
	节点下线次数(OTN_i)	p_2
	节点活跃度(NA_i)	p_3
信用级别	有效区块数(EP_i)	0.2
	上一轮节点信任值(C_{si})	0.2

在节点信用评价指标中,有些二级指标的属性是离散的,如交易总量、网络延迟、节点离线次数等。这些值的波动幅度较大且不同属性不同范围的维度容易影响实验,所以在利用这些指标计算节点信任值之前,需要对离散的指标进行量化。该文采用 min-max 标准化方法将所有指标量化在[0,1]之间,量化函数如下:

$$\chi^* = \frac{\chi - \mathrm{MIN}}{\mathrm{MAX} - \mathrm{MIN}}$$

其中,MAX 为样本数据的最大值,MIN 为样本数据的最小值,χ 为原始数据,χ^* 为量化后的值。

在数据量化后,根据选取指标计算信任值,信任值的计算方法如下:

$$C_i = \sum_{i=1}^{n} IN_i * W_i$$

其中,交易情况和性能下对应的二级指标所占权重为 t_1、t_2、p_1、p_2、p_3,其中:

$$t_1 + t_2 = 0.25$$

$$p_1 + p_2 + p_3 = 0.35$$

可根据系统当前状态动态调整各个指标所对应的权重,从而激励整个系统保持高效的运作。如当前系统节点活跃度较低,可适当增大节点活跃度(NA_i)的权重 p_3,减小网络延迟(NL_i)、节点下线次数(OTN_i)的权重 p_1、p_2 来激励节点提升活跃度。

1.3 基于高斯混合模型的异常节点剔除算法

在 DPoS 共识机制节点投票过程中,包含以下行为的节点定义为异常节点:

(1)节点投给信任值较低的节点;

(2)节点不愿意消耗自己的资源,放弃投票;

(3)节点采取欺骗行为以防止系统识别异常行为。

针对投票过程中的节点异常行为,提出一种基于高斯混合模型(Gaussian Mixture Model,GMM)的异常节点剔除算法。该算法首先对投票情况数据进行划分,得到 M 个聚类中心[13]。然后根据期望值最大(Expectation Maximization,EM)估计方法计算 GMM 参数的初始值[14],并构建 GMM。并根据构建的 GMM 计算投票数据的混合高斯概率值,将投票数据的准确率与召回率之比设置为检测异常数据的阈值。然后计算该高斯分布的概率密度,与异常数据阈值进行比较。如果概率密度小于阈值,则可以将投票数据识别为异常数据。

1.3.1 高斯混合模型

高斯混合模型本质上是通过多个不同的高斯分布拟合而成的样本空间分布情况模型[15]。

设 x 为一个 N 维的特征向量,其高斯混合模型是由 M 个分量混合而成。则 x 的混合高斯密度为:

$$P(x) = \sum_{i=1}^{M} \alpha_i P_i(x)$$

其中,α_i 为第 i 个高斯分量的加权系数[16],且 $0 < \alpha_i < 1$,$\sum_{i=1}^{M} \alpha_i = 1$,$P_i(x)$ 为高斯混合模型中第 i 个高斯分布的高斯密度函数,即:

$$P_i(x) = \frac{1}{(2\pi)^{N/2} \left|\sum_i\right|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu_i)^{\mathrm{T}} \sum_i^{-1} (x-\mu_i)\right\}$$

式中,μ_i 为第 i 个高斯分量的均值,$\sum_i$ 为第 i 个高斯分量的协方差。

GMM 模型训练阶段,使用 EM 算法以最大化似然函数的方式求解模型最佳参数,即每个高斯分量的加权系数 α_i、均值 μ_i 和协方差 $\sum_i$,直至模型收敛。

1.3.2 贝叶斯信息准则

BIC(贝叶斯信息准则)是 Schwar 在 1978 年提出的一种统计模型的适用性度量方法,现已广泛用于时间序列和线性回归的模型识别中。BIC 奖励模型准确性并惩罚模型复杂性,模型复杂度越高,所得结果越大。最小检测值则为模型最佳聚类数目[17]。

$$\mathrm{BIC} = x\ln(n) - 2\ln(L)$$

其中,x 为模型参数,n 为样本数量,L 为似然函数。

1.3.3 算法流程

Step1:根据信誉值模型以及投票情况计算出最终

得票情况并加入数据集。

可将节点 N_i 的最终投票得分表示为:

$$v_i = C_i * \text{votes}_i + \text{coin} * \text{coinTime} * \text{Weight}$$

其中, C_i 为节点信任值, votes_i 为节点对应的票数。coin 为节点代币数量, coinTime 为币龄, Weight 为代币权重,将计算得到的 v_i 加入到数据集中并将数据集表示为 $V = \{v_1, v_2, \cdots, v_n\}$ 。

Step2:利用贝叶斯信息准则对数据集进行预处理,得到高斯混合模型最佳聚类数目。

Step3:设置初始参数

$$\lambda^{(0)} = \{\alpha_i^{(0)}, \mu_i^{(0)}, \sum\nolimits_i^{(0)}, i = 1, 2, \cdots, M\}$$

Step4:根据样本集和初始参数计算随机变量的期望 $Q(\lambda; \lambda^{(0)})$ 。计算公式如下:

$$Q(\lambda; \lambda^{(0)}) = \sum_{i=1}^{M} \sum_{j=1}^{J} \log(\alpha_i P(v_i \mid \lambda_i))$$

$$\frac{P_i(v_j \mid \lambda_i^{(0)}) \alpha_i}{\sum_{i=1}^{M} P_i(v_j \mid \lambda_i^{(0)}) \alpha_i}$$

Step5:计算更新后的加权系数 α_i^{new} ,均值 μ_i^{new} 和协方差 $\sum\nolimits_i^{\text{new}}$ 。

$$\alpha_i^{\text{new}} = \frac{1}{J} \sum_{i=1}^{J} P(i \mid v_j, \lambda^{(0)})$$

$$\mu_i^{\text{new}} = \frac{\sum_{j=1}^{J} v_j P(i \mid v_j \lambda^{(0)})}{\sum_{j=1}^{J} P(i \mid v_j, \lambda^{(0)})}$$

$$\sum\nolimits_i^{\text{new}} = \frac{\sum_{j=1}^{J} P(i \mid v_j, \lambda^{(0)})(v_j - \mu_i^{\text{new}})(v_j - \mu_i^{\text{new}})}{\sum_{j=1}^{J} P(i \mid v_j, \lambda^{(0)})}$$

Step6:根据更新后的参数计算概率密度函数,即:

$$P_i(v_i) = \sum_{i=1}^{M} \alpha_i^{\text{new}} \frac{1}{(2\pi)^{N/2} \left| \sum\nolimits_i^{\text{new}} \right|^{1/2}} \exp\{-\frac{1}{2}$$

$$(x - \mu_i^{\text{new}})^{\mathrm{T}} \sum\nolimits_i^{-1} (x - \mu_i^{\text{new}})\}$$

Step7:计算出训练数据的异常检测阈值 P ,其中 P 为投票数据的正确率与召回率之比,并将概率密度低于阈值的节点剔除,并将该节点本轮信任值置为 0。

2　实验结果与分析

2.1　实验环境

为了验证改进前后 DPoS 共识算法的有效性,在相同的环境下对 DPoS 算法改进前后的效率以及安全性进行分析。该文使用 python 语言对信任值模型以及基于高斯混合模型的异常节点剔除算法进行构建,并模拟 DPoS 共识算法,构建了 301 个模拟区块链交易的节点集群,实验设置见证(出块)节点个数 101 个,最后,通过 MATLAB R2020a 对最终的实验数据进行可视化对比评价。

2.2　模型训练

利用 EOS 项目的历史数据集 XBlock-EOS 对系统模型进行训练并保存了训练模型。在实验仿真过程中,选取了 5 000 个节点的历史数据进行输入,并对系统模型进行了 50 次训练,同时对每次训练的模型选取 500 个节点进行测试,其中包括 40% 的异常节点,并分析整个测试集综合投票得分前 60% 的恶意节点所占比例,从而得出整个模型剔除异常节点的比率。实验结果如图 2 所示。经过实验结果对比分析可得,随着训练集节点数量的增加,异常节点剔除率也在不断增加。当训练集节点达到 2 000 时,异常节点剔除率高达 93.5% 。当节点进一步增加时,异常节点剔除率基本维持在 93% 左右。对比传统的 DPoS 共识算法,该算法提高了共识节点的安全性,并且能够很好地剔除错误节点,降低错误节点成为共识节点的概率。

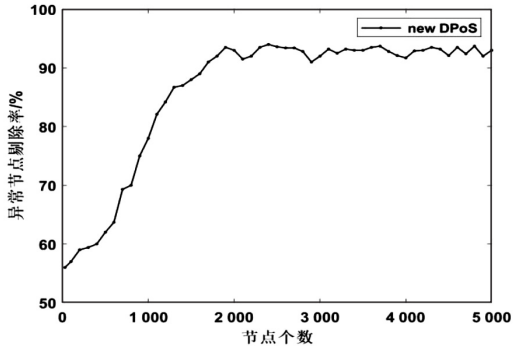

图 2　异常节点剔除率

2.3　算法验证

通过信任值模型以及高斯混合模型剔除异常节点选出的共识节点,如果在生成区块的过程中产生错误区块或者没有在规定时间内产出区块,那么在下一轮的信任值计算中,由于有效区块数较低,相应的得到的信任值也会相对较低,如果其他节点投票给这些信任值较低的节点,经高斯混合模型剔除模型计算后,投票节点便会失去成为共识节点的资格,并且其自身信任值也会被清零,相应的在下一轮的信任值计算中也会得到一个极低的信任值。图 3 为在 60% 异常节点比率下,改进前后 DPoS 共识算法共识节点集合中恶意节点所占比例。经过实验结果对比分析可得,改进后的 DPoS 共识机制随着共识次数的增加,异常节点在共识节点中的比例逐渐减小,当达到 20 次共识时,异常节点占比接近于 0。而传统的 DPoS 共识机制,因为没有异常节点剔除的过程,随着共识次数增加,异常节点在共识节点中的占比并没有明显减少。

传统的 DPoS 共识机制对出块时间没有限制,出块时间大致在 4 s 左右,改进后的 DPoS 共识机制将有效区块数作为信任值的一项计算标准,所以选出的共

识节点的出块时间相对于改进前也有一定提高。图 4 为改进前后 DPoS 共识机制出块时间对比。当区块个数增加到 1 000 时,改进后的算法出块时间降至 2 s 左右,比改进前的 DPoS 共识机制节约了 50% 左右的时间,能够更好地应对日益增长的交易量。

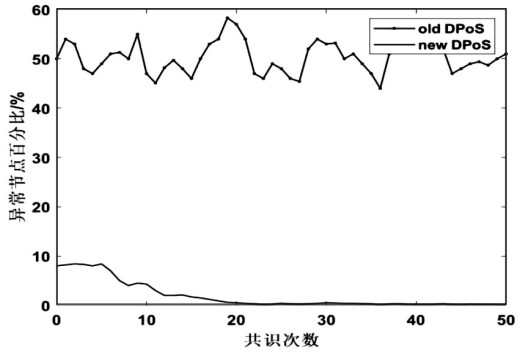

图 3 改进前后 DPoS 异常节点占比

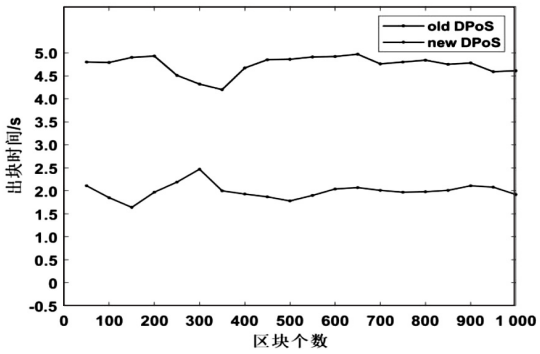

图 4 改进前后 DPoS 出块时间

3 结束语

传统的 DPoS 共识机制相较于 POS、POW 算法具有更低的能耗以及更高的安全性[18],但仍然存在投票积极性不高、节点恶意投票等缺点。针对上述缺点,该文设计了一种信任值模型,对节点活跃度、节点性能以及产出有效区块数等进行综合评分,同时构建高斯混合模型将恶意节点进行剔除,有效提高了共识节点安全性,缩短了节点出块时间以满足日益增大的交易量。在后续的工作中,计划优化基于高斯混合模型的异常节点剔除算法的时间复杂度,并将提出的系统综合模型应用到更广泛的实际应用场景中,从而创造更大的价值。

参考文献:

[1] NAKAMOTO S. Bitcoin:a peer-to-peer electronic cash system[EB/OL]. (2009). https://bitcoin. org/bitcoin. pdf.

[2] YANG F, ZHOU W, WU Q, et al. Delegated proof of stake with downgrade:a secure and efficient blockchain consensus algorithm with downgrade mechanism [J]. IEEE Access, 2019,7:118541-118555.

[3] 谭 超. 基于 DPoS 算法的区块链共识机制优化[D]. 重庆:重庆邮电大学,2020.

[4] CHEN S,XIE M,LIU J,et al. Improvement of the DPoS consensus mechanism in blockchain based on PLTS[C]//2021 7th IEEE intl conference on big data security on cloud (BigDataSecurity),IEEE intl conference on high performance and smart computing (HPSC) and IEEE intl conference on intelligent data and security (IDS). NY:IEEE,2021:32-37.

[5] 胡 倩. 基于以太坊的区块链共识算法研究与实现[D]. 济南:齐鲁工业大学,2021.

[6] 沈 瑞. 区块链共识算法的研究与应用[D]. 南京:南京邮电大学,2021.

[7] LARIMER D. Delegated proof-of-stake white paper[EB/OL]. 2014. http://docs. Bitshares. org/bitshares/dpos. html.

[8] WEI Y, LIANG L, ZHOU B, et al. A modified blockchain DPoS consensus algorithm based on anomaly detection and reward-punishment[C]//2021 13th international conference on communication software and networks (ICCSN). Chongqing: IEEE,2021:283-288.

[9] 高 迎,谭学程. DPOS 共识机制的改进方案[J]. 计算机应用研究,2020,37(10):3086-3090.

[10] TAN C,XIONG L. DPoSB:delegated proof of stake with node's behavior and borda count[C]//2020 IEEE 5th information technology and mechatronics engineering conference (ITOEC). Chongqing:IEEE,2020:1429-1434.

[11] CHEN Y,LIU F. Improvement of DPoS consensus mechanism in collaborative governance of network public opinion[C]// 2021 4th international conference on advanced electronic materials, computers and software engineering (AEMCSE). Changsha:IEEE,2021:483-488.

[12] 何 帅,黄襄念,刘谦博,等. DPoS 区块链共识机制的改进研究[J]. 计算机应用研究,2021,38(12):3551-3557.

[13] AYIAD M M,LEITE H,MARTINS H. State estimation for hybrid VSC based HVDC/AC:unified bad data detection integrated with Gaussian mixture model[J]. IEEE Access,2021, 9:91730-91740.

[14] HOJJATINIA H,JAHANSHAHI M,SHEHNEPOOR S. Improving lifetime of wireless sensor networks based on nodes´ distribution using Gaussian mixture model in multi-mobile sink approach (Feb,10. 1007/s11235-021-00753-6,2021)[J]. Telecommunication Systems: Modeling, Analysis, Design and Management,2021(1):77.

[15] 朱壮壮,周治平. 高斯混合生成模型检测健康数据异常[J]. 计算机科学与探索,2022,16(5):1128-1135.

[16] 于冰洁. 基于高斯模型的异常检测算法[D]. 徐州:中国矿业大学,2017.

[17] 李文政,顾益军,闫红丽. 基于网络贝叶斯信息准则算法的社区数量预测研究[J]. 数据分析与知识发现,2020,4(4):72-82.

[18] 谭敏生,杨 杰,丁 琳,等. 区块链共识机制综述[J]. 计算机工程,2020,46(12):1-11.