

MAS 环境中一种基于反馈可信度的多维信誉计算方法*

霍星¹, 张阳洋¹, 景永俊^{2,3}, 邵堃²



¹(合肥工业大学 数学学院, 安徽 合肥 230009)

²(合肥工业大学 计算机与信息学院, 安徽 合肥 230009)

³(北方民族大学 计算机科学与工程学院, 宁夏 银川 750021)

通讯作者: 邵堃, Email: shaokun@hfut.edu.cn

摘要: 在分布式体系结构的 MAS(multi-agent system)中, Agent 之间通过彼此的交互, 协调完成共同的任務, 但是由于没有中心化的管理权威可以依赖, 导致对网络中 Agent 信誉信息进行判断存在一定的困难. 传统的基于评价反馈的信誉评估方法存在反馈评价属性信息利用不足以及缺少确保反馈评价信息可信的可行机制等问题, 为此, 提出一种综合的信誉计算方法. 该方法针对个别用户提交的恶意评价, 采用 CUSUM(cumulative sum)控制图理论对服务评价中的数据进行过滤; 利用信息熵的方法对不同维度的评价数据进行整合; 使用改进的 PageRank 算法对评价影响力进行度量, 最终得到融合反馈评价真值与评价影响力的综合信誉. 仿真结果表明, 该方法在提高信誉计算收敛性和准确性、抵抗恶意攻击行为等方面表现出较好的效果.

关键词: 信誉计算; 累积和; 多维度反馈评价; 个体影响力

中图法分类号: TP18

中文引用格式: 霍星, 张阳洋, 景永俊, 邵堃. MAS 环境中一种基于反馈可信度的多维信誉计算方法. 软件学报, 2020, 31(2): 374-394. <http://www.jos.org.cn/1000-9825/5619.htm>

英文引用格式: Huo X, Zhang YY, Jing YJ, Shao K. Multidimensional reputation calculation method based on feedback reliability in MAS environment. Ruan Jian Xue Bao/Journal of Software, 2020, 31(2): 374-394 (in Chinese). <http://www.jos.org.cn/1000-9825/5619.htm>

Multidimensional Reputation Calculation Method Based on Feedback Reliability in MAS Environment

HUO Xing¹, ZHANG Yang-Yang¹, JING Yong-Jun^{2,3}, SHAO Kun²

¹(School of Mathematics, Hefei University of Technology, Hefei 230009, China)

²(School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230009, China)

³(School of Computer Science & Engineering, North Minzu University, Yinchuan 750021, China)

Abstract: In MAS with distributed architecture, Agents are employed to achieve task objectives through mutual coordination and collaboration. Since there is no centralized management authority to rely on, it is difficult to judge the reputation information of Agents. Beyond that, traditional reputation evaluation methods based on evaluation and feedback have some problems, such as the insufficient usage of feedback evaluation and the absence of feasible mechanism for credible feedback evaluation information, etc. To solve these problems, a comprehensive reputation calculation method is proposed in this study. Concerning about malicious evaluation submitted by individual users, the proposed algorithm first filters the service evaluation data by the CUSUM (cumulative sum) control chart theory, then integrates different dimension of evaluation data using information entropy method, after that, uses PageRank algorithm to measure

* 基金项目: 国家自然科学基金(61572167, 61502136); 科技部国际合作项目(2015DFA11450)

Foundation item: National Natural Science Foundation of China (61572167, 61502136); International Cooperation Project of Ministry of Science and Technology (2015DFA11450)

收稿时间: 2018-01-02; 修改时间: 2018-04-16; 采用时间: 2018-06-21; jos 在线出版时间: 2019-01-21

CNKI 网络优先出版: 2019-01-22 13:49:10, <http://kns.cnki.net/kcms/detail/11.2560.TP.20190122.1348.013.html>

the influence of individual. Finally, it gives the comprehensive reputation model incorporating feedback evaluation of truth value and individual influence. Simulation results show that the proposed method performs well in improving reputation computation accuracy and convergence as well as the resisting of malicious attacks.

Key words: reputation computing; cumulative sum; multi-dimensional feedback evaluation; individual influence

信任是一种依赖关系,是人类解决复杂社会问题的有效机制之一,受到多个研究领域的高度重视^[1].由于信任具有主观性和易变性,并受到主观和客观等多种因素的影响,因而很难定量地描述和分析其演化的过程^[2-4].信誉是一种依附在人与人之间、单位之间和商品交易之间形成的相互信任的生产关系和社会关系^[5],是关于人或物的特征、立场的大众观点.在 MAS(multi-agent system)中,被评估 Agent 的信誉是指系统中其他 Agent 依据其自身与被评估 Agent 的过往交互所提供的对被评估 Agent 行为能力认可情况的集合.信誉作为一种用于支持信任评价的机制^[6],具有相对的客观性、稳定性和可靠性.因而,针对信誉的研究具有更高的现实意义和应用价值^[7].信誉研究的关键是针对社会个体的日常行为数据进行收集和分析,并建立相关的信誉模型.信誉计算方法作为信誉研究中一种重要的数据分析方法,综合考虑网络环境中影响信誉的各个因素,对评价客体的信誉给予数字化度量^[2].信誉作为用户选择服务的一个重要指标,能够保证服务提供者所发布的服务信息的准确性、客观性和可靠性^[8,9].特别是在开放动态的 MAS 环境中,各 Agent 是自治独立的,数据是分散存储和处理的,每一 Agent 都拥有一定的信息处理和解决问题能力,Agent 之间通过交互,协调合作完成共同的任务.由于提供相同服务的 Agent 越来越多,而不同服务之间又存在利益化的竞争关系,因此,研究准确、可靠和客观的 Agent 信誉度评估方法具有重要的理论和现实意义^[10].

目前,如何通过 Agent 的服务反馈信息获得准确、客观和可靠的信誉度计算方法,已得到相关学者的着重关注.这些研究使用不同的方法和工具提高了信誉评估算法的准确性、有效性和客观性,并在决策分析^[11]、推荐系统^[12-14]、交易系统^[15]、P2P 网络系统^[16]中得到大量的运用,但仍存在反馈评价属性信息的利用不足以及缺少确保反馈评价信息可信的可行机制等问题,具体如下.

- (1) 首先,在进行信誉评估的过程中,由于反馈评价的结果影响交互客体的信誉,在关系到被评价对象现实利益时,并不能确保所有 Agent 提供的评价信息都是诚实、公正的,造成了存在虚假的反馈评价数据^[14,17].本文将这些用来提高目标个体的信誉值(推举攻击)或者降低其信誉值(打压攻击)的虚假反馈行为定义为恶意反馈.为了确保信誉评估算法的准确性,必须识别并剔除这些恶意反馈.
- (2) 其次,虽然通过多维度评价属性划分的方式对信誉度进行度量已成共识^[18,19],但对于不同属性的评价信息进行整合缺乏有效手段,不同属性的评价指标数据未得到有效利用.
- (3) 最后,缺少对给出信誉评价的 Agent 所具有的影响力做出区分的具体方法.在实际情况中,由于 Agent 分布范围广泛,其知识水平、专业能力和个体素质不尽相同.对于特定的交互过程,拥有丰富专业背景和对交互过程较为熟悉的 Agent 所提供的反馈结果对于信誉评价的影响力更大.因此,为了提高信誉评估结果的可靠性,需要考虑不同 Agent 的反馈影响力的差异.

针对以上 3 个问题,本文提出了一种在 MAS 环境下的综合信誉计算方法 CRCM(comprehensive reputation calculation mechanism).本文提出的 CRCM 方法包含反馈检测、反馈信息整合、反馈影响力评估这 3 个组成部分,对经过这 3 个部分分析后得出的数据进行综合计算,得出最终的 Agent 信誉评估结果.仿真实验表明,与已有的信誉计算方法相比,本文所使用的 CRCM 方法更加准确、有效;同时,对虚假交互、共谋攻击等恶意行为表现出一定的抵抗能力.

本文第 1 节分析与本文研究相关的主要研究工作.第 2 节对提出的信誉计算方法实现框架做出介绍,并对 Agent 交互行为做出形式化说明.第 3 节给出具体的综合信誉计算方法,包括反馈检测、反馈评价信息整合、反馈影响力评估等,并对本文所提出的 CRCM 方法进行详细论述.第 4 节描述仿真实验及结果分析.第 5 节给出结论及未来工作展望.

1 相关工作

当前,常见的信誉计算机制主要分为两类^[7]:非数值型的信誉系统和数值型的信誉系统.非数值型的信誉系统依据权威信用评定机构对不同的实体进行不同等级的认证.类似于 Better Business Bureau Online(BBB),通过对不同企业对行业准则的遵守情况和对于客户投诉的处理情况,为不同的企业做出信用等级的评定,并对一些信用较低的企业给出警告,但并不提供任何数值型的信誉结果.非数值型的信誉系统需要有公认的权威机构进行信用评定,具有较高的公信力,但实际审理周期长、过程复杂,难以适用于复杂多变的网络环境.数值型的信誉系统因为具有结构简单、适应性广、动态性强的特点,被大量地运用于网络环境中对 Agent 信誉进行计算.传统的网络环境下的信誉计算往往采用简单的数值模型,如 eBay、淘宝通过对不同个体的所有评级结果逐步累积来得到个体信誉^[20,21],这种信誉计算方法快捷、简便,但实体很容易通过策略性的方式,如多次交互,提供不真实评级结果等来改变信誉计算结果.据此,相关研究人员提出了改进的模型:

Josang 等人^[22]建立了一种基于贝叶斯概率的信誉系统.该系统适用于交互结果可用二值化(好,坏)来描述的应用场景,利用统计学的方法更新 Beta 概率密度函数,通过组合新产生的计算结果与先验信誉结果,得到后验信誉值.该方法为信誉计算的合理性提供了理论基础,但在实际应用中实现困难,没有考虑信誉评估中可能出现的虚假信息等情况.为此,Teacy 等人^[23]提出一种基于 Beta reputation 的评价信息过滤方法.该方法依据评价信息与个体直接经验之间的差距,对评价信息进行调整.但该方法需要对历史评价数据按照时间轴顺序划分为若干时间窗,每次信誉更新只使用 1 个时间窗内的数据,对于反馈信息利用不足.

Zacharia 等人^[24]提出一种基于统计学的 Sporas 模型对 Agent 的全局信誉进行计算,在计算中引入评价者可信度这一因素,将评价者的信誉视为评价反馈可信度,考虑评价者的可信度对信誉值的影响.Sporas 模型中信誉的评价是单维度的,只考虑交互结果,每次评价的结果在 0.1~1 之间,并不对具体的交互内容如交互响应时间、交互开销等做细分.同时,由于采用了迭代的方式计算信誉,该方法对于团伙之间通过相互评价的团伙攻击行为抵抗能力不足.

针对单维度评价过程中评价粒度较粗^[25,26]、只针对交互本身做出单一评价的问题,Griffiths 等人^[27]从交互所包含的内容属性出发,分析一次交互过程中,不同 Agent 对于交互成败、交互时间、交互开销、交互质量这 4 个维度的满意程度,对不同的维度确定不同的指标权重,加权计算得到对此次交互的最终评价结果.由于缺少对数据可信度的分析,该方法的计算结果与实际信誉之间仍然存在偏离.

Pedrycz 等人^[28]对评价信息的挖掘进行了相应的研究,通过对评价信息进行过滤,将主要评价意见进行融合,提出了一种新的信誉可视化方法.但对于如何动态地处理多维评价信息中不同维度的评价信息之间的相互关系,并将这些不同维度评价信息合理、有效地整合为节点信誉未做出有效的说明.

Huynh 等人^[29]从信誉评价来源的角度进行考虑,将信誉分为 IT(interaction trust)、WR(witness reputation)、RT(role-based trust)和 CR(certificate reputation)这 4 个组成部分,基于这 4 个部分进行归一化加权平均,得到最终 Agent 的全局信誉.针对 Huynh 等人提出的模型使用固定参数进行信誉计算,难以适应网络环境动态变化的问题,You 等人^[30]提出了一种适用于网络环境的自适应的信誉模型.该模型将信誉分为 DR(direct reputation)和 WR 两个部分,在每次交互完成之后,对 DR 和 WR 的值进行更新,并对两者在融合计算中的权重进行重新分配得到新的信誉结果.该模型在一定程度上遏制了恶意评价对于信誉评估的影响,因为恶意攻击者必须同时对几种不同类型的信誉数据进行恶意攻击才能对最终的信誉评估结果产生影响,提高了恶意行为的成本.但这种方式对于已经产生的恶意行为缺乏有效的检测机制.

Su 等人^[31]提出了一种能够识别群托攻击的评估方法.群托攻击是指系统中多个用户联合起来抬高或打压某个个体,该辨别方法的关键在于识别出一些一起进行了很多评价活动且评价的结果十分相似(通常为异常评分)的用户群.Chirita 等人^[32]提出的探测恶意评分的方法与攻击类型无关.该方法通过计算新的评分数据与已有的评分数据的观点相似度、近邻行为相似度和平均意见等的偏离程度,根据历史经验确定一个概率函数,依据设定阈值,用户评分会被分类为正常评分或者恶意评分.在 MovieLens 数据库上评估这种探测方法的结果表明,这种方法能够准确地探测出过去只收到数量很少且相对低的评分的用户所受到的推举攻击.这两种方法都具

有较高的实用价值,但忽略了数据可信度和数据来源可信度之间存在的差异,没有将二者进行综合考虑;同时,采用经验确定阈值的方法在 MAS 环境中仍然存在动态适应性不足的问题。

通过相关文献分析,定量地计算个体的信誉需要构建一套可测量的指标,以适应不同的应用背景。同时,信誉的计算模型还需要符合信誉自身所具有的一些特性,如增长困难、毁坏容易、逐渐趋于稳定^[33,34]等特性。本文通过对反馈信息进行合理过滤,结合信誉的组成(评价反馈内容)和评价影响力(Agent 在网络中的链接关系分析)两个方面的内容,对 Agent 的全局信誉进行综合计算,在一定程度上克服了信誉模型中反馈评价属性信息利用不充分以及反馈评价信息可信度难以判断的不足。

2 基于反馈可信度的多维信誉计算方法

通过相关工作的分析,在不同的应用环境下,信誉计算模型存在一定的差异。本文构建的信誉计算方法在没有中心化管理权威的支持下也可以对 MAS 环境中的 Agent 全局信誉进行有效评估^[7]。

本文中,任意 Agent 都由一组关键字进行识别,Agent 通过关键字来检索得到其他 Agent 的信誉信息。Agent 通过检索得到的陌生 Agent 的信誉信息包括综合信誉计算值和 Agent 的每一次交互所得到的反馈评价结果。

本文设定新加入网络的 Agent 的初始信誉值为 0。在每一次交互完成后,Agent 的信誉依据评价反馈和 Agent 在网络中的链接关系进行更新,如果两个 Agent 在一段时间内进行了多次交互,模型只将最近的 a 次交互列入计算中。网络中任意节点的信誉值都大于一个新加入节点的初始值。随着交互次数的增加,信誉逐渐收敛,在单次的交互活动中,高信誉 Agent 的信誉变化量比低信誉 Agent 要小。

信誉计算模型所依据的数据基础分为两部分:一部分来源于不同 Agent 之间交互完成后的反馈评价,这部分内容具有主观属性;另一部分来源于对 Agent 在网络中链接结构的分析,这部分信息通过信誉系统自动获取,具有客观属性。CRCM 两部分数据信息采用综合计算机制得到最终的信誉计算结果。

2.1 网络行为模型

MAS 环境具有复杂的结构特性和未知的生成机制,进而表现出一定的随机性、模糊性和不可预测性^[35]。同时,网络异构性、复杂性、动态性和交互行为的主观性导致准确地度量网络节点间的交互关系较为困难,而信誉计算的关键在于合理地量化影响信誉计算的各个因素。为了精确地对信誉计算机制进行描述,本文使用网络拓扑图刻画 Agent 之间的交互链接行为。依照网络中 Agent 之间的交互关系建立网络拓扑图,图中的边可以表示 Agent 之间存在交互行为,并对网络行为做以下形式化说明。

交互行为 b_{ij} 表示在网络中,节点 v_i 与 v_j 在 T 时间内的交互行为,可以表现为交易或数据的传输。通过节点与节点之间的交互行为,在某一时间段内构造基于交互行为的网络拓扑图 $G=(V,E)$,如图 1 所示。

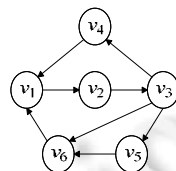


Fig.1 An example of interactive topology diagram

图 1 交互拓扑图示例

图 1 中的节点表示网络中的 Agent,节点间的边表示直接的交互行为。设 V 为网络中的节点集合 $V=\{v_1, v_2, \dots, v_n\}$, 设 B_{ij}^T 为交互时间序列集合,则在 T 时间内,节点 v_i 与 v_j 之间的 n 次交互过程时间序列集合 $B_{ij}^T = \{b_{ij}^1, b_{ij}^2, \dots, b_{ij}^n\}$, T_k 表示第 k 次交易发生的时间。设任意 v_i 与 v_j 在交互完成后提供的 m 项反馈评价指标分别为 $X_1(v_i, v_j)$, $X_2(v_i, v_j), \dots, X_m(v_i, v_j)$, 则反馈评价集合表示为 $X = \{X_1, X_2, \dots, X_m\}$, 其中,每一个元素 $X_m(v_i, v_j) \in [0, 1]$ ($m=1, 2, \dots, 5$) 为一个

评价属性.设 ω_m 表示第 m 个评价属性相对于其他评价属性的重要程度,且 ω_m 满足 $\omega_m \in [0,1], \sum_{m=1}^5 \omega_m = 1$,则 ω_m 为 $X_m(v_i, v_j) (m=1,2,\dots,5)$ 的评价权重.为了方便信誉计算,设该图为有向图,图中节点之间的指向关系定义如下.

若 v_i 选择与 v_j 进行交互,则在图中有一条由 v_i 指向 v_j 的边,在交互结束后, v_i 给予 v_j 交互评价.图中的边不仅表示存在交互行为,还包括一定的交互信息.例如,图1中节点 v_1 与 v_2 仅有一条有向边相连接,但 v_1 与 v_2 可能进行了多次交互.

定义1(节点反馈评价 $E(\text{evaluation})$). E_{ij} 表示节点 v_i 与节点 v_j 交互完成后所进行的评价活动,表现为 v_i 对于 v_j 行为能力的认可情况.

定义2(节点影响力 $Pr(\text{prestige})$). 节点 v_i 的影响力 Pr_i 由 v_i 的入链数量和出链数量进行链接分析计算得到.节点 v_i 影响力值越高,表示节点越具有权威性, v_i 在交互后进行的反馈评价具有较高的应用价值.节点 v_i 的影响力与 v_i 发生交互次数无关,只与 v_i 在网络中与其他节点的链接情况有关.

定义3(节点综合评价值 $SE(\text{systematic evaluation})$). 节点 v_i 与 v_j 交互活动后所得的节点综合评价值 SE_{ij} 由 E_{ij} 与 Pr_i 进行复合运算得到.

定义4(节点综合信誉 $SR(\text{systematic reputation})$). 节点 v_i 的综合信誉 SR_i 由与 v_i 进行了交互的节点集 $V = \{v_1, v_2, \dots, v_k\}$ 中所有节点的综合反馈评价累积得到.例如在图1中,节点 v_1 的信誉由 SE_{41} 与 SE_{61} 累加得到.

2.2 信誉度计算方法

本文提出的信誉度评估方法首先对恶意反馈进行检测与清除;其次,从反馈来源的角度对不同 Agent 提供反馈的影响力做出划分;同时,对合理的反馈评价结果进行整合.在此过程中,本文考虑了不同 Agent 的评价能力具有差异性以及信誉累积过程中的信誉评价随信誉增长单次评价效用逐渐降低等情况,得到较为完善的适用于 MAS 环境的信誉计算机制.信誉综合计算具体实现框架如图2所示.

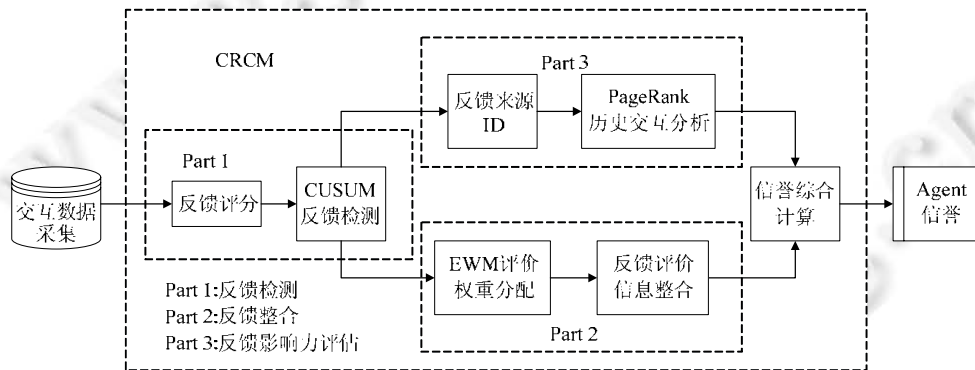


Fig.2 Framework of comprehensive reputation calculation

图2 综合信誉计算框架

本文提出的 CRCM 架构包含 3 个部分^[7].

- 第 1 部分为反馈检测,主要用于对评价数据中的恶意评价进行检测.由于恶意评价在评价数据集中较为分散,本文基于 CUSUM(cumulative sum)算法,提出针对反馈评价信息进行预处理的 Q-CUSUM 算法用于检测恶意反馈.该方法的基本思想是,通过检测时间序列上反馈评价值的变化情况对评分数据集中的恶意评分进行去除,确保评分数据的真实性.
- 第 2 部分为反馈信息整合,该部分将信誉评价反馈的内容做 5 个不同维度的划分,之后采用基于熵权的 EWM(entropy weight method)方法来计算不同维度指标所对应评价的权重.通过对不同维度的评价信息赋予不同的权重,使评价结果反映出不同评价指标之间的联系,并结合效用函数得到评价反馈的整合结果.

- 第3部分为反馈影响力评估,此部分的工作与第2部分反馈信息整合并行进行.反馈影响力评估通过对MAS环境中Agent的交互记录建立档案,分析Agent在交互网络中的交互链接关系,得到与评价内容无关的Agent影响力;然后,对所得的影响力结果做适度划分并赋予不同的权重,增强信誉评估结果的客观性.

最终通过对第2部分反馈信息整合与第3部分Agent影响力分析得到的结果进行综合计算,得到综合信誉评估结果.

3 信誉综合计算

3.1 反馈检测

3.1.1 恶意反馈检测方法

当前,针对评价的恶意反馈分为两种^[7].

- 第1种通过提交极端的恶意评价 EME(extreme malicious evaluation)(提供在评价域范围内的最高值或最低值)来进行恶意攻击.这类攻击方法很容易通过阈值过滤的方式进行检测.
- 第2种恶意反馈行为通过多次提供与已有评分记录相似的反馈数据混入评分结构中,达到推举攻击或者打压攻击的目的.由于此种恶意行为数据特征不明显,在评分数据中较为分散,因而检测困难.

本文将这两种恶意行为统一记为 ME(malicious evaluation).文中定义的恶意反馈会同时对本文第3.2节中5个不同属性的信誉评价指标进行恶意攻击,为了降低检测开销,选取反馈评价样本中最直观地表现出交互情况的服务质量评价数据进行监测.将检测出的异常反馈视为 ME,并将该反馈记录从评分数据集中去除.

一般来说,均值极差图和均值标准差图对于恶意数据检测有较好的效果.这种效力基于:随着待检测样本数量的增加,样本均值的分布趋近于正态分布.样本数越大,均值图对均值显著偏移越敏感.但是在实际情况中,Agent之间的交互次数是有限的,因而无法通过提高子组样本量的方法来提高检测效果.在这种可获得的数据很少的情况下,需要采用更加灵敏的方式对恶意数据进行检测.

本文所提出的恶意反馈检测办法利用评价反馈自身的统计信息来识别恶意评价.从统计方法学来说,对ME行为的检测就是研究在时间序列上的评分结果是否在统计分布规律上保持一致的问题:如果不一致,则以最小的延迟找到分布规律发生变化的点^[36],将这些点视为恶意评分点,并从评价数据中去除.本文采用连续时间序列检测方式监控恶意反馈行为,常用的序列检测方法包括CUSUM控制图(cumulative sum control chart)、Shewhart控制图、指数加权滑动平均控制图EWMA(exponentially weighted moving average)和自适应阈值算法(adaptive threshold algorithm).其中,CUSUM控制图和自适应阈值算法适用于检测变化值比较小的序列,文献^[37]中,对两种算法进行了比较分析.由于CUSUM算法具有较高的检测效率^[38],能够适应恶意反馈检测中需要同时对多个Agent的不同评分结果进行监测的特点,为此,本文基于CUSUM算法,提出针对反馈评价信息进行预处理的Q-CUSUM算法.

3.1.2 针对反馈评价异常检测的改进CUSUM算法(Q-CUSUM)

CUSUM算法分为参数化和非参数化两种^[39].参数化的CUSUM算法是指:设观察到的一个随机序列为 X_1, X_2, \dots, X_n ,随机变量 x_1, x_2, \dots, x_n 相互独立,且样本序列中元素取值有限. x_1, x_2, \dots, x_{k-1} 概率密度函数为 $f_0(\cdot)$,随机变量 x_k, x_{k+1}, \dots, x_n 概率密度函数为 $f_1(\cdot)$, $k \in [1, n]$.定义对数似然比累积和 $S_n = \sum_{i=1}^n s_i$,其中, $s_i = \ln f_1(\cdot) / f_0(\cdot)$,通过计算当前时间的对数似然比累积和与正常状态的对数似然比累积和最小值之差得到统计变量 $g_n = S_n - \min_{0 \leq j \leq n} S_j$, $I(\cdot)$ 可用公式(1)中的迭代形式表示,其中, $[a]^+ = \max[a, 0]$:

$$\begin{cases} g_0 = 0 \\ g_n = \left[g_{n-1} + \ln \frac{f_1(\cdot)}{f_0(\cdot)} \right]^+ \end{cases} \quad (1)$$

CUSUM 算法的决策函数为 $d_n=d(g_n)=I(g_n \geq \delta)$, $I(\cdot)$ 为真时取 1, 反之取 0. 当统计变量超过某一阈值 δ 时, 认定发生了数据异常.

在实际运用中, 由于难以对参数进行估计, 特别是获取评分结果的概率分布比较困难, 为此, 相关学者提出了非参数化的 CUSUM 算法. 该算法适用于分析序列的统计分布无法预知的情况.

本文所采用的 Q-CUSUM 算法由两个部分组成: 极端评价处理和基于非参数 CUSUM 算法的数据处理. 前者主要对 EME 进行过滤, 避免 EME 在 CUSUM 检测中造成统计量大量累计, 影响后者对评分数据中不明显的异常评分进行过滤的精度. 整个 Q-CUSUM 算法流程如图 3 所示.

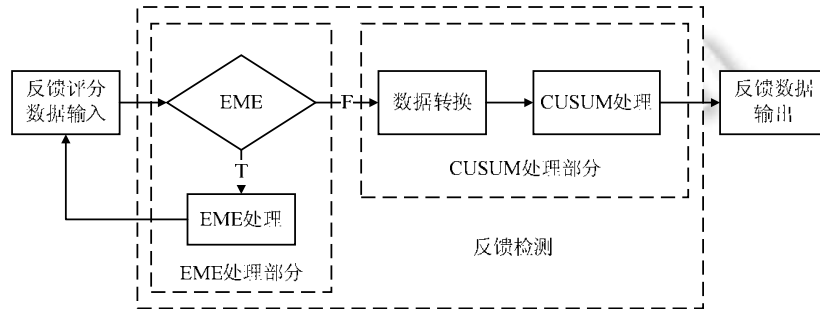


Fig.3 A flowchart of the Q-CUSUM algorithm

图3 Q-CUSUM 算法流程图

极端评价处理基于正态分布的基本理论, 将超出 $\mu \pm 3\sigma$ 范围的评分结果视为极端恶意评价, 其中, μ 为评分均值, σ 为评分标准差, $\sigma = \sqrt{\sum_{i=1}^n (x_i - \mu)^2 / (n-1)}$, 并将 EME 从评价数据中去除.

对于在评分结构 $\mu \pm 3\sigma$ 范围内的 ME 行为, 采用基于非参数的 CUSUM 算法来进行检测. 将评价结果的偏移累积起来, 达到放大的效果, 从而提高检测过程对于评价偏移的灵敏度. 由于评分结果既有可能增加, 也有可能减少, 为了降低算法的处理开销, 采用双边的 CUSUM 算法的迭代形式用于检测 ME 行为. 定义 CUSUM 控制图如公式(2)所示.

$$\begin{cases} \text{上累积和: } \begin{cases} g_0^+ = 0 \\ g_n^+ = \max(0, g_{n-1}^+ + x_n - (\mu_0 + \beta_s)) \end{cases} \\ \text{下累积和: } \begin{cases} g_0^- = 0 \\ g_n^- = \min(0, g_{n-1}^- - x_n + (\mu_0 - \beta_i)) \end{cases} \end{cases} \quad (2)$$

其中, g_n^+ 是用来监测在 CUSUM 控制图上超过上限的数据点, g_n^- 监测在 CUSUM 控制图上超过下限的数据点, 上(下)限用来识别超过过程理想值 μ_0 (变点出现前的序列均值) 的变化. β_s 为上参考值, β_i 下参考值, 分别表示非参数 CUSUM 算法的上、下限. $\beta_i = D_i \times \sigma / 2$, $\beta_s = U_s \times \sigma / 2$, 其中, U_s, D_i 分别表示检测到的过程均值的向上或向下偏移, 以标注差为单位, 通常在 $0.5\sigma \sim 1.5\sigma$ 之间. 决策区间 $H = h\sigma$, 其作用与控制限一样. h 为决策参数, 通常定义为 5.

CUSUM 处理过程分 3 个步骤^[7].

- 第 1 步, 在服务质量评分序列未发生突变之前, CUSUM 统计量 g_n 是一个在 0 附近随机波动的变量.
- 第 2 步, 若评分发生突变, 当发生正向偏移, 即该评分记录高于正常水平时, g_n^+ 会不断累积增大; 当发生反向偏移, 即该评分结果低于正常水平时, g_n^- 不断累加.
- 第 3 步, 当 g_n 累积量超过设定的阈值时, 就可以认为已经发生了统计量突变, 引发统计量突变的评分记录将被视为 ME, 需要在评分数据集中去除该评价记录. 在 CUSUM 检测部分完成后, 评分数据集中的剩余数据将用于对网络中的 Agent 信誉进行评估.

3.2 反馈整合

由于 MAS 中提供相同服务的 Agent 越来越多,在 Agent 之间的交互过程中,除了考虑功能性的需求,也对非功能性的服务质量 QoS(quality of service)提出了一定的需求^[40,41]。为此,本文在计算信誉时,将信誉评价反馈的具体内容做 5 个不同维度的划分,采用熵权的方法,利用系统采集到的不同维度指标数据的差异性来计算各评价指标权重,对 Agent 提交的不同维度的评价信息进行整合,结合效用函数得到最终的评价反馈整合结果。本文采用的反馈评价整合方式如图 4 所示。

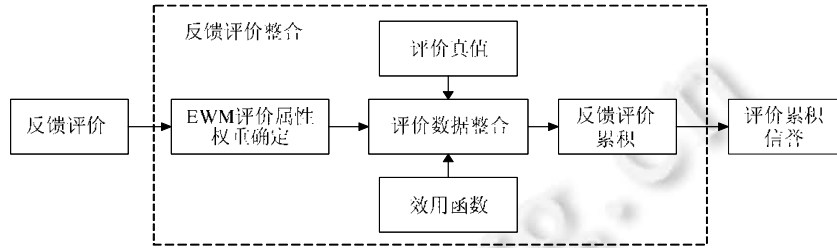


Fig.4 Framework of feedback evaluation processing
图 4 反馈评价处理框架

依据第 2.1 节的介绍,对交互行为和交互结果做以下形式化说明。

假设某 Agent P 共提供 n 次交互服务,其中,成功交互的次数用 n_s 表示,第 k 次发出交互请求的时间为 t_R^k ,交互开始时间为 t_S^k ,交互结束时间为 t_E^k ,即交互响应时间跨度为 $\Delta t^k = t_S^k - t_R^k$,交互时间跨度为 $\Delta t^k = t_E^k - t_S^k$,第 k 次交互的交互开销为 Co_k 。得到反馈评价信誉(reputation,简称 RE)的计算方法如下。

(1) 将信誉评价定义为五元组:

$$Evaluation=(Service\ Result,Service\ Quality,Service\ Response\ Time,Service\ Time,Service\ Cost).$$

其中,Service Result 表示交互结果,包含交互成功或失败两方面;Service Quality 表示交互质量;Service Response Time 表示交互响应时间;Service Time 表示交互时间跨度;Service Cost 表示交互开销。

通过五元组将服务抽象为具体的 5 个指标,设指标数据集合 $D(data),D_k=\{d_{k1},d_{k2},d_{k3},d_{k4},d_{k5}\},d_{kj}$ 表示 P 所提供服务的第 j 维度在第 k 次交互中的指标数据。并将针对这 5 个评价指标的反馈评价形式化表示为 $X_k=\{x_{k1},dx_{k2},x_{k3},x_{k4},x_{k5}\},X_k$ 表示针对 Agent P 所提供的第 k 次交互服务得到的反馈评价结果,针对 Agent P 所提供的第 k 次交互服务中第 j 维度的评价记为 x_{kj} ,其中, $\{0 \leq x_{kj} \leq 1 | j=2,3,4,5\},x_{k1} \in \{0,1\}$,即交易成功时, $x_{k1}=1$;交易失败时, $x_{k1}=0$ 。 $d_{k1}=x_{k1},d_{k2}=x_{k2},d_{k3}=\Delta t^k,d_{k4}=\Delta t^k,d_{k5}=Co_k$ 。

(2) 各指标权重采用熵权的方法来确定。某个评价指标数据的信息熵越大,则表明该指标数据的离散程度越大,针对该指标的评价数据就越重要,其所对应的评价权重也越大;相反,某个指标数据的信息熵越小,则表明该指标值的离散程度越小,所能提供的信息量就越少,针对该指标的评价所分配的权重应该越小。通过熵权的方法来确定评价向量的权重将评价指标数据与反馈评价数据进行了有效的联系,同时考虑了交互属性对于信誉的影响。

依据评价指标向量建立包含 n 个评价、5 个评价指标的评价指标数据矩阵 $D=(d_{kj})_{n \times 5}$,计算第 k 个项目的第 j 个指标值的比重 Z_{kj} :

$$Z_{kj} = d_{kj} / \sum_{k=1}^n d_{kj} \tag{3}$$

计算第 j 个指标的熵值 e_j :

$$e_j = -l \sum_{k=1}^n Z_{kj} \ln Z_{kj}, \text{ 其中, } l = (\ln m)^{-1} \tag{4}$$

计算对应的第 j 个指标的熵权:

$$\omega_j = \frac{1 - e_j}{\sum_{j=1}^5 (1 - e_j)}, \text{ 且 } \sum_{j=1}^5 \omega_j = 1 \quad (5)$$

Agent P 所提供的第 k 次服务所得反馈评价计算结果可用公式(6)表示.

$$E(P_k) = \sum_{j=1}^5 \omega_j x_{kj}, \quad k = 1, 2, 3, \dots, n \quad (6)$$

评价反馈的结果表现了 Agent 对于预期交互结果的满意程度.该评价结果取决于 Agent 的经验水平,因此不同 Agent 由于自身喜好、专业性和对于服务要求的区别,对于同一次交互可能给出完全不同的评价结果.一个好的信誉机制需要采用合理化的度量方式使主观的评价内容转化为可信赖的数值结果,在此基础上,通过一定的计算机刻画画出 Agent 的信誉.同时,信誉机制应该对 Agent 的不良行为起到约束作用^[7].综合考虑评价结果对于信誉的影响,对反馈评价的结果做如下处理.

定义 5(反馈评价真值 TE(truth-evaluation)). TE_k 表示 Agent i 对于 Agent j 的评价结果与节点 i 在交互前预期之间的差值,节点 i 交互之前的预期基于节点 j 在与 i 交互之前所进行的 $k-1$ 次交互过程中所累积的信誉. SR_{\max} 表示当前整个网络中节点信誉所能达到的最大值.

$$TE_k = E_k - \frac{SR_{k-1}}{SR_{\max}} \quad (7)$$

通过反馈评价真值,提高了评价结果的二值化程度.只有当评价结果高于预期时,才会促成信誉的增长;当评价结果低于预期时,会造成信誉的损失.节点 SR 值越大,想要继续获得 SR 值的生长就需要提供更优质的服务,这就对高信誉 Agent 的行为起到了约束作用;相反,在 SR 值较小时, SR_k/SR_{\max} 较小, TE_k 与 E_k 近乎相等,此时信誉对 Agent 行为的约束较小,对 Agent 信誉此时的增长起到促进作用.

据此得到反馈评价累积通过公式(8)进行计算.

$$E(P) = \sum_{k=1}^N TE(P_k) = \sum_{k=1}^N \left[E(P_k) - \frac{SR_{k-1}}{SR_{\max}} \right] = \sum_{k=1}^N \left[\sum_{j=1}^5 \omega_j x_{kj} - \frac{SR_{k-1}}{SR_{\max}} \right] \quad (8)$$

考虑到信誉有其特定的现实意义,拥有较高信誉的 Agent 在每一次更新过程中信誉值的变化会更小.为此,本文中引入效用函数 UF(utility function) $\varphi(\cdot)$ 对评价有效性进行描述.具体的 $\varphi(\cdot)$ 形式如公式(9)所示.

$$\varphi(k) = 1 - (1 - \alpha) \frac{1}{1 + e^{\left(\frac{SR_{k-1}}{1 - SR_{\max}} \right)}}, \quad \alpha \in (0, 1) \quad (9)$$

其中, $\varphi(k)$ 表示 Agent P 所提供的第 k 次服务所得反馈评价的效用; α 表示反馈结果的合理性,体现不同的反馈结果与实际交互结果之间的联系.实际交互结果是指在评价者有充分专业认知背景并对交互内容做出公正、合理评价的理想结果,最能反映交互的真实情况.由于服务用户分布范围广泛,知识水平、专业能力和个人素质不尽相同,无法保证每个用户在交互完成后都能进行合理的反馈评价.拥有丰富专业知识,对服务内容熟悉的用户的反馈评价结果往往比缺乏相关知识储备,对服务内容不熟悉的用户提供的反馈结果更具有合理性^[42]. α 越大,表示评价结果越接近于实际交互结果. α 值可以依据用户过往的评价经历通过公式(10)计算获得.

$$\alpha = \frac{CEN_j}{SEN_j}, \quad \alpha \in (0, 1) \quad (10)$$

其中, CEN_j 表示节点 j 做出的合理评价的数量(合理评价是指满足 $\frac{|E_{ij}^k - SR_{ij}^k/k|}{SR_{ij}^k/k} < \delta$, E^k 表示节点 j 对节点 i 进行第 k 次反馈评价, SR_{ij}^k 表示节点 j 在对节点 i 进行第 k 次反馈评价后,节点 i 的信誉, δ 表示反馈合理的门限,一般取较小值), SEN_j 表示节点 j 做出的评价的总数量. $\varphi(\cdot)$ 的变化趋势如图 5 所示.

图 5 中,从上到下分别为 α 取 0.9, 0.8, 0.7, 0.6, 0.5 时, φ 函数的变化情况.随着交互次数的不断增加,单次交互的评价结果对信誉累积的影响力不断下降.

在引入效用函数后的反馈评价信誉累积结果可用公式(11)表示,其中, $Re(P)$ 表示 Agent P 提供 n 次交互所得到的反馈评价累积的信誉.

$$Re(P) = \sum_{k=1}^n \left(\varphi(R_k) \sum_{j=1}^5 \omega_j x_{kj} - \frac{SR_{k-1}}{SR_{max}} \right), \alpha \in (0,1) \quad (11)$$

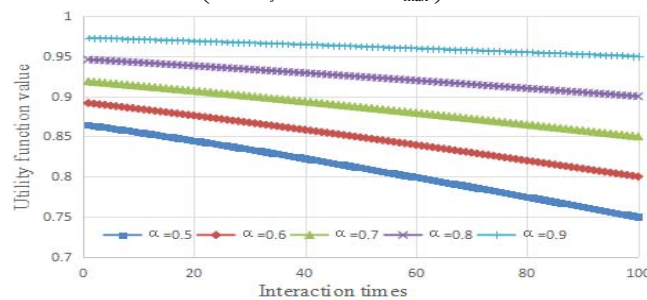


Fig.5 Trends of evaluation effectiveness changes
图 5 评价有效性变化趋势

3.3 反馈影响力评估

3.3.1 PageRank 算法应用分析

反馈影响力评估通过对不同 Agent 交互产生的链接关系进行分析,得出对应于该 Agent 的反馈影响力结果.这种交互分析基于这样一种现实假设:与具有高影响力的 Agent 进行交互活动,能够在团体中更快地获得更高的认可.本文对于此种假设模型提出基于具有同样特性的改进 PageRank 算法.

PageRank 算法起源于搜索引擎排序,其核心思想基于“从许多优质的网页链接过来的网页必定还是优质网页”的回归关系,来判定网页的重要性.经典的 PageRank 算法如公式(12)所示.

$$P(A) = (1-d) + d \sum_{i=1}^n \frac{P(T_i)}{C(T_i)} \quad (12)$$

其中,

- $P(A)$ 表示网页 A 的 PageRank 值.
- d 是阻尼系数,通常设定 $d=0.85$,表示从网页 A 出发继续往下浏览的概率. $(1-d)$ 表示浏览网页过程中随机跳转到一个新的网页的概率,如果把真实的网络转化为转移矩阵,这将是一个极其稀疏的矩阵,极度稀疏矩阵的迭代相乘会使迭代向量变得不平滑,即一些节点具有很大的 PageRank 值,而另一些节点的 PageRank 值很小,甚至为 $0.d$ 的选用使迭代向量变得平滑,同时克服了 Spider Traps 节点的影响.
- $P(T_i)$ 表示链入网页 A 的网页 T_i 的 PageRank 值.
- $C(T_i)$ 表示网页 T_i 链出网页的数量.

通过 PageRank 算法得到网页 A 在当前网络中的影响力值.

由于通过 PageRank 算法得到的网页影响力具有客观性,因此,PageRank 算法也常用来对社会网络中的图结构进行分析.网页和网页之间可以通过链接关系传递影响力,这在 MAS 环境下也是成立的,但在 MAS 环境中,情况会更加复杂,主要表现在以下两个方面.

- (1) Agent 属性与网页不同.网页的主题属性一般是固定的,通常只有 1 种,易于确定网页主题归属.但 Agent 的属性较多,理解为 Agent 可以选择进行多种不同类型的交互.
- (2) 一个网页在诞生之后,其主题属性基本保持不变(网页的 PageRank 值会由于链接关系的变化发生改变),但 Agent 的属性会随着时间的推移发生一定的改变,即交互类型发生改变.

基于以上两个方面,直接采用 PageRank 算法对 Agent 影响力进行全量计算(即所有 Agent 均纳入计算)是不合理的.为了对 MAS 环境中的 Agent 影响力进行计算,本文对 PageRank 算法做出了改进,克服了以上两个方面存在的问题,提出了 Pr-PageRank 算法^[7].

3.3.2 基于 Topic-Sensitive PageRank 的 Agent 影响力计算(Pr-PageRank)

Topic-Sensitive PageRank^[43]参考 ODP(open directory project)网站,定义了 16 个大的主题类别,通过离线计算得出一个 PageRank 向量集合.该集合中的每一个向量与某一主题相关,即计算某个页面关于不同主题的得分.主要分为两个阶段:主题相关的 PageRank 向量集合的计算和在线查询时主题的确.

在本文方法中,Pr-PageRank 算法将网络中的 Agent 交互行为限定为常见的 3 类:文件传输服务、交易服务、推荐服务(可以将这 3 种交互行为之外的交互作为新的分类,并按照类似方法进行计算,在此忽略).按照 Agent 的历史交互经历,计算交互过程对于这 3 类行为的归属感.结合 Topic-Sensitive PageRank 算法得出 Agent 对于这 3 类服务的 PageRank 值.考虑 Agent 影响力的冷启动问题,对所得结果做进一步划分,最终得到 Agent 对应于不同服务的影响力值.具体计算过程如下^[7].

(1) 交互行为归属感 MF(membership function)计算.依据第 2 节对模型的说明,在该模型中,所有 Agent 的评价信息可以进行准确定位,同时能够对所有的交互类型进行准确的识别.对交互行为归属度的计算基于 Agent 所进行某一类别的活动越多,Agent 对于该类别的服务越熟悉,归属感就越高的假设.得到归属感计算公式(13).

$$s_j^p = \frac{cit_j^p}{tit_j^p}, j=1,2,3 \quad (13)$$

其中, s_j^p 表示 Agent P 在类别 j 中的归属感, cit_j^p 表示 Agent P 在分类 j 中所进行的交互次数, tit_j^p 表示 Agent P 所进行的总的交互次数.

(2) 通过公式(14),计算 Agent P 在分类 j 下的 PageRank 值.

$$P_j(P) = (1-d)s_j^p + d \sum_{i=1}^n \frac{P_j(T_i)}{C(T_i)} \quad (14)$$

其中, $P_j(T_i)$ 表示在分类 j 下选择与 Agent P 进行交互的 Agent T_i 的 PageRank 值, $C(T_i)$ 表示 Agent T_i 选择与其他 Agent 进行交互的总次数.

(3) 由于 PageRank 级别并不是线性增长的,如 PageRank 4 级比 PageRank 3 级仅仅高了 1 级,但在影响力上却高出 6 倍~7 倍.对于新加入网络中的 Agent,由于缺少交互机会,其 PageRank 值会很低.直接采用 PageRank 对节点影响力进行评估不利于网络中新 Agent 建立信誉.基于此,本文采用一种较为简单的基于正态分布的划分方式对 MAS 环境中节点的 PageRank 值进行划分.划分方式如公式(15)所示.

$$Pr = \begin{cases} 0.2, & P < \mu - 3\sigma \\ 0.3, & \mu - 3\sigma \leq P < \mu - 2\sigma \\ 0.4, & \mu - 2\sigma \leq P < \mu - \sigma \\ 0.5, & \mu - \sigma \leq P < \mu + \sigma \\ 0.7, & \mu + \sigma \leq P < \mu + 2\sigma \\ 0.9, & \mu + 2\sigma \leq P < \mu + 3\sigma \\ 1, & P \geq \mu + 3\sigma \end{cases} \quad (15)$$

其中, $\mu = \frac{\sum_{i=1}^n P_i}{n}$, $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - \mu)^2}$, n 表示 Agent 的总数, P 表示通过 PageRank 算法计算得到的值, Pr 表示按照正态等级划分后的影响力评估结果.

依据公式(15)中的划分方式,一个新加入网络中的 Agent 所提供的评价反馈信息影响力为 0.2.在网络中, P 值最高的 Agent 的影响力为 1,本文中将其所提供的评价反馈信息视为完全可信.

通过这种链接分析方法所得到的节点影响力,考虑了整个网络中节点之间相互影响的过程,节点间的链接影响力可自动分析获得.该方法具有以下优点:一方面,由于反馈评价可能在伪造记录的帮助下被操控,将影响力纳入信誉计算的过程,不仅有助于提高信誉评估的准确度,同时也增加了恶意评价行为的成本,因为攻击者必须同时构造出一个基于其自身的链接网络来增强其评价的可信度,在一定程度上削弱了恶意评价的影响;另一

方面,由于 PageRank 值本身与交互双方的交互次数无关,这种对网络中不同 Agent 链接关系的分析能够有效抵御信誉评估中的共谋攻击 CA(collusion attack)行为(两个 Agent 之间通过互相给予大量优良评价,从而达到共同提高信誉评估结果的目的).

3.4 综合信誉计算过程

通过第 3.2 节和第 3.3 节的介绍,对信誉计算过程两个主要的数据来源,即反馈评价信息和 Agent 交互链接关系分别进行了处理,得到了两类精确量化的数值结果.结合信誉自身演化规律,得到综合信誉计算方法^[7],如公式(16)所示.

$$SR_k = SR_{k-1} + Pr_k \left(\varphi(R_k)E_k - \frac{SR_{k-1}}{SR_{max}} \right) = \sum_{k=1}^n Pr_k \left[\varphi(R_k) \sum_{j=1}^5 \omega_j x_{kj} - \frac{SR_{k-1}}{SR_{max}} \right] \quad (16)$$

其中, SR_k 表示经过 n 次交互后,Agent P 的综合信誉值; SR_{k-1} 表示在进行第 k 次交互之前,Agent P 的综合信誉值; $\varphi(R_{k-1})$ 是选用的效用函数; Pr_k 表示与 Agent P 进行第 k 次交互的节点在网络中的影响力值.

在以往的模型中,人们往往无法区分一个新加入网络的 Agent 是一个全新的 Agent 还是进行了恶意行为后重新进入网络的 Agent,因此,Agent 本身就对陌生 Agent 保持谨慎的态度.同时,开放式网络中匿名的特性,为 Agent 转换身份重新进入网络提供了便利条件,这在一定程度上加大了新 Agent 在网络中建立自己信誉的难度,不利于一个可信网络的建设.本文提出的综合信誉计算机制由于综合了反馈信息和 Agent 个体的影响力分析,当新个体由于反馈信息的不足难以在网络中建立自己的信誉时,依托该个体的影响力因素,同样有机会实现信誉的较快增长.通过对公式(16)的分析,可以得出本文提出的信誉计算机制具有如下特性.

- (1) 现有网络中的任一 Agent 的信誉值都大于一个新加入的 Agent,提高了 Agent 转换身份重新加入网络建立信誉的成本,使恶意 Agent 转换身份重新进入网络获得相同的信誉需要更大的交互开销,有效地抑制了通过不断注册新账户进行匿名攻击的恶意行为.
- (2) 本文提出的综合计算过程结合了对节点链接情况的分析,使一个新节点可以通过与拥有高影响力的 Agent 之间的交互获得更高的信誉增长,缩短了信誉增长所需的时间,有利于网络中 Agent 建立自身的信誉,对可信网络的建设起促进作用.

有关以上算法性质,本文提供了相关证明,见证明 1、证明 2.

证明 1:任意 Agent 的信誉值均大于新加入的 Agent 的信誉.

$$\begin{aligned} SR_k &= SR_{k-1} + Pr_k \left(\varphi(R_k)E_k - \frac{SR_{k-1}}{SR_{max}} \right) \\ &> SR_{k-1} - Pr_k \frac{SR_{k-1}}{SR_{max}}, \text{ since } Pr_k \varphi(R_k)E_k > 0 \\ &> SR_{k-1} - \frac{SR_{k-1}}{SR_{max}}, \text{ since } 0.2 \leq Pr_k \leq 1 \\ &> 0, \text{ since } SR_{max} > SR_{k-1}. \end{aligned} \quad \square$$

证明 2:在进行了充分的交互过程之后,节点累积的信誉值存在上界.

设信誉值的上界为 $U(U>1)$,令 $SR_{k-1}=U-\Delta(SR)$, $\Delta(SR)>0$ 表示在进行第 k 次交互之前,Agent 信誉值与上界的距离:

$$\begin{aligned} SR_k &= SR_{k-1} + Pr_k \left(\varphi(R_{k-1})E_k - \frac{SR_{k-1}}{SR_{max}} \right) \\ &= U - \Delta(SR) + Pr_k \left(\varphi(R_k)E_k - \frac{U - \Delta(SR)}{U} \right) \\ &\leq U - \Delta(SR) + Pr_k \left(1 - \frac{U - \Delta(SR)}{U} \right), \text{ since } \varphi(R_k)E_k \leq 1 \end{aligned}$$

$$\begin{aligned} &\leq U - \Delta(SR) + \frac{\Delta(SR)}{U}, \text{ since } 0.2 \leq Pr_k \leq 1 \\ &\leq U, \text{ since } \Delta(SR) \leq U. \end{aligned}$$

即 Agent 信誉值存在上界. □

3.5 相关实现算法

本节给出本文中模型的实现算法伪代码,包括第 3.1 节中的 Q-CUSUM 反馈评价过滤算法、第 3.2 节中的 EI 反馈评价整合算法、第 3.3 节中的 Pr-PageRank 反馈影响力评估算法以及 CRCM 模型总体实现算法.

算法 1. Q-CUSUM 算法.

输入: X_data /*反馈评分数据*/

输出: Y_data /*已过过滤评分数据*/

```

1: for all  $X_m \in X\_data$  /*对所有 Agent 获得的反馈评价记录进行检测*/
2:   While  $EME(X_m) == \text{TRUE}$ 
3:      $EME(X_m)$  /*清除包含极端评价的反馈评价记录*/
4:      $X \leftarrow \text{transform}(X_m)$  /*选取服务质量属性数据分别计算上累积和,下累积和*/
5:      $Y\_data \leftarrow Q\_CUSUM(X)$  /*清除反馈数据中恶意评价*/
6:   end for
7: return  $Y\_data$ 

```

算法 2. 反馈评价整合 EI 算法.

输入: $flag_data, Y_data$ /*输入评价指标数据集,反馈评价数据集*/

输出: rep_data /*反馈评价数据整合结果*/

```

1: for all  $D_k \in flag\_data$  do
2:   for  $j=0$  to 5 do
3:      $\omega_j \leftarrow EWM(D_k)$  /*计算各评价指标权重*/
4:      $E \leftarrow (\omega, D_k)$  /*计算反馈评价价值*/
5:      $TE \leftarrow E - (SR_{k-1} / SR_{max})$  /*计算反馈评价真值*/
6:      $u \leftarrow UF(SR_{k-1}, SR_k)$  /*计算反馈评价效用*/
7:      $rep\_data \leftarrow (u, TE)$  /*公式(11)反馈数据整合结果*/
8:   end for
9: return  $rep\_data$ 

```

算法 3. Pr-PageRank 算法.

输入: V_data, B_data /*节点集合,节点历史交互记录*/

输出: p_data /*节点 Pr 值*/

```

1: for all  $V \in V\_data$  do
2:   for all  $V\_data \in B\_data$  do
3:      $s \leftarrow MF(V\_data)$  /*计算节点 V 归属感*/
4:     repeat
5:        $P_{k+1} \leftarrow (1-d)s + dA^T P_k$ 
6:        $k = k+1$ 
7:     until  $\|P_{k+1} - P_k\|_1 < \varepsilon$ 
8:   return  $P_{k+1}$  /*计算节点在不同交互类别中的影响力*/
9: switch the value of  $P$  do /*公式(15)对所得 PageRank 值进行划分*/
10:  case 1  $P < \mu - 3\sigma$ 

```

```

11:         output: 0.2
12:     case 2  $\mu-3\sigma \leq P < \mu-2\sigma$ 
13:         output: 0.3
14:     case 3  $\mu-2\sigma \leq P < \mu-\sigma$ 
15:         output: 0.4
16:     case 4  $\mu-\sigma \leq P < \mu+\sigma$ 
17:         output: 0.5
18:     case 5  $\mu+\sigma \leq P < \mu+2\sigma$ 
19:         output: 0.7
20:     case 6  $\mu+2\sigma \leq P < \mu+3\sigma$ 
21:         output: 0.9
22:     otherwise
23:         output: 1
24:     endsw
25: endsw
26: P_data ← P
27: return P_data

```

算法 4. CRCM 算法.

输入: $X_data, flag_data, V_data, B_data$ /*反馈评分数据集,评价指标数据集,节点集合,节点历史交互集合*/

输出: SR_data /*Agent 综合信誉值*/

```

1: for all  $X_m \in X\_data$ 
2:   for all  $D_k \in flag\_data$  do
3:     for all  $V \in V\_data$  do
4:       for all  $V\_data \in B\_data$  do
5:          $Y\_data \leftarrow Q\_CUSUM(X\_data)$  /*Q-CUSUM 处理*/
6:          $REP\_data \leftarrow EI(Y\_data, flag\_data)$  /*EI 反馈数据整合*/
7:          $P\_data \leftarrow Pr\_PageRank(V\_data, b\_data)$  /*Pr-PageRank 反馈影响力评估*/
8:       end for
9:     end for
10:   end for
11: end for
12:  $SR\_data \leftarrow SR(rep\_data, P\_data)$  /*公式(16)Agent 信誉综合计算结果*/
13: return  $SR\_data$ 

```

4 仿真实验及结果分析

为了验证整合多维度反馈评价数据和节点影响力数据对提高信誉计算结果的鲁棒性、稳定性和有效性的作用,本文设计了一系列仿真模拟实验.仿真模拟实验是目前采用较为广泛的信任模型评估方法,通过模拟信任在具体的应用场景和实体之间的交互行为,达到评估信任模型实际效果的目的^[7].

本文中,为了使仿真实验更贴近真实环境,实验中的反馈指标数据采用 QWS 真实数据集.该数据集包括 2 507 条 Web 服务信息,每条信息包含 13 个具体属性,关于该数据集的具体信息详见文献[44].实验中假设所有 Agent 均从交互结果、交互质量、交互响应时间、交互时间跨度、交互开销这 5 个维度对目标 Agent 给出交

互评价.为了方便说明,实验中将 Agent 转化为抽象节点,节点信息包含节点编号、节点类别、节点历史交互对象等信息.各节点拥有独立存储空间用来存储交互数据.

具体实验内容包括 CUSUM 算法检测能力测试(考察 Q-CUSUM 算法对于恶意行为识别的有效性)、信誉度的迭代收敛性分析(考察综合信誉计算的收敛性)、信誉计算的有效性(考察综合信誉计算对于恶意攻击的有效性和鲁棒性).为了验证本文中的模型实际计算效果,本文通过预先设定的几项评价指标——恶意反馈检测效果、收敛速度、评估准确率和信誉评估误差度,将本文的模型与已有的 Sporas 模型进行了对比.

- 恶意反馈检测率 MDR(malicious evaluation detection rate)反映模型对于恶意反馈行为的检测能力, $MDR=D(P)/\alpha N(P)$,其中, $D(P)$ 表示检测到针对 Agent P 的反馈数据中属于仿真设定的恶意反馈的数量, α 表示仿真中设定的 ME 所占比例.为了防止检测过程中将正常反馈误识别为恶意反馈,设定检测成功率 SDR(successful detection rate)来表示检测的准确率, $SDR=D(P)/S(P)$,其中, $S(P)$ 表示检测出的针对 Agent P 的恶意反馈总数.
- 收敛速度表示信誉评估算法中信誉聚合的速度,通过计算收敛所需迭代次数 CIN(convergence iteration number)来表示.本文中,通过对不同算法基于仿真数据的信誉评估结果演化趋势进行分析来比较不同计算方法的收敛能力.收敛速度越快,表示该信誉计算方法能够使信誉较快的达到稳定状态,这种稳定的状态有利于整个信誉系统平稳运行.
- 评估的准确率 SER(successful estimate rate),即计算正确的评估结果在所有评估结果中的比例.计算正确是指满足 $(SR_i - SR'_i)/SR_i \leq \varepsilon$,其中, SR_i 和 SR'_i 分别表示节点 i 的信誉度和评估信誉度.实验中, ε 取 0.1.SER 直观地表示出不同计算方法的应用效果.
- 信誉误差度 RAE(reputation aggregation error), $RAE = \sqrt{\frac{1}{N} \sum_{i=1}^N (SR_i - SR'_i)^2}$, N 为网络环境中节点总数.

RAE 越小,表示信誉评估模型的精度越高,也意味着对于恶意行为的抵抗能力越强.

仿真实验基于 NetLogo 实现,NetLogo 能够允许使用者对多个独立 Agent 下达运行指令,适用于研究随时间变化的复杂系统.本文通过 NetLogo 实现了一个拥有 50 个独立 Agent 的 MAS 环境.硬件平台配置为 CORE i7 3.6GHz,内存 8GB.

4.1 仿真实验

4.1.1 Q-CUSUM 控制图检测效果

该实验主要针对 Q-CUSUM 控制图对于恶意反馈的检测效果进行检验,检测过程分为 EME 行为检测和 CUSUM 处理两个部分.实验中的反馈数据来自于 QWS 数据集中一组 Web 服务的反馈评级结果,该数据集中对服务的评级区间为 1~5.考虑到 QWS 数据集中含有 Reliability ($Reliability \in [0,1]$)属性,实验中对评分数据经过公式(17)映射到[0,1],所得反馈评价数据先采用 EME 行为检测方法进行过滤,再对经过 EME 行为检测的数据进行 CUSUM 处理:

$$evaluation = \begin{cases} 0.2([class \cdot Reliability] + 1), & class \cdot Reliability - [class \cdot Reliability] \geq 0.5 \\ 0.2([class \cdot Reliability]), & class \cdot Reliability - [class \cdot Reliability] < 0.5 \end{cases} \quad (17)$$

其中, $class$ 表示评价等级, $evaluation$ 表示评价等级映射到[0,1]的结果.从图 6 中可以看出,在 $[\mu - 3\sigma, \mu + 3\sigma]$ 范围之内包含了大部分的评价数据,将超出此范围的评价数据视为 EME.EME 评分结果明显异于正常的评分结构,通过检测将 EME 评分记录从反馈数据中去除.经过 EME 行为检测的数据转入到 CUSUM 处理部分.CUSUM 处理主要对评分结果与正常评分结构相近的评价数据进行过滤.从图 7 中可以看出,通过 CUSUM 控制图得到的数据将评分结果中均值的变化趋势进行了放大,对不同评价过程中的偏移量进行累计.在恶意反馈对于评价均值未造成明显影响的情况下,CUSUM 控制图中 g_n^+ , g_n^- 表现出了明显的波动,并对于恶意节点进行了准确的定位,将超出控制限 $H[-5\sigma, 5\sigma]$ 的评价数据视为 ME 行为.CUSUM 处理部分所检测出的恶意评分与正常评分十分相似,但累积量 g_n 出现了明显的变化.

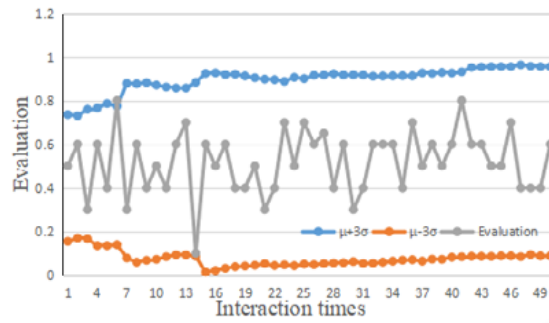


Fig.6 Detection results of EME

图 6 极端恶意行为检测结果

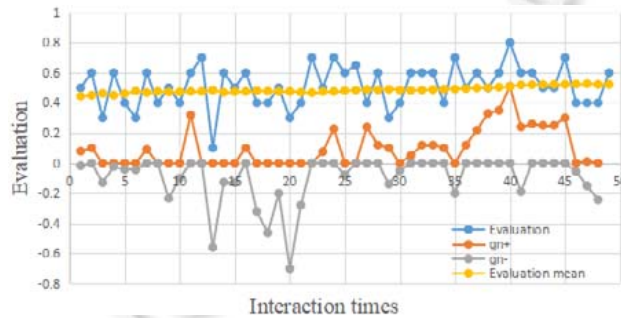


Fig.7 Detection results of CUSUM

图 7 CUSUM 控制图检测结果

针对恶意行为,Q-CUSUM 包含的两个部分都发挥了作用.为了进一步说明 Q-CUSUM 的检测效果,本文设计模拟实验,通过提高恶意反馈在整个反馈结果中的比例来表现 Q-CUSUM 在不同情况下的检测能力.为了便于比较,本文对文献[45]中同样采用基于统计数据本身的过滤方式(K-Means)进行了仿真,得到的实验对比结果如图 8、图 9 所示.

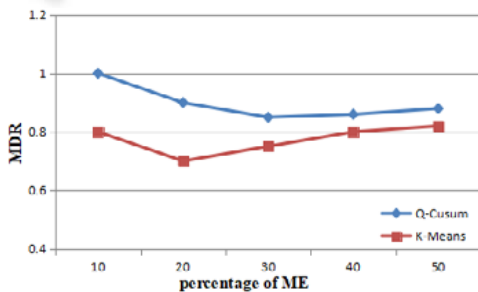


Fig.8 Comparison of MDR

图 8 反馈检测率对比

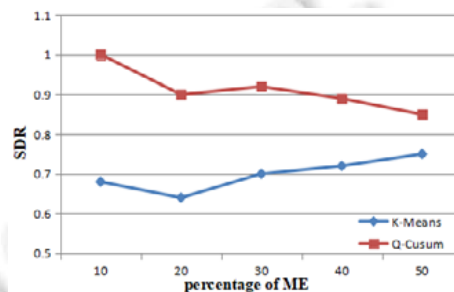


Fig.9 Comparison of SDR

图 9 检测成功率对比

从图 8 中可以看出,随着 ME 在整个反馈中的比例从 10%不断增加到 50%,两种方式对于恶意反馈的检测都有较好的检测效果;在 ME 比例达到 50%时,本文所采用的过滤方法检测率依然保持在 85%以上.由于 K-Means 方法采用聚类的方式识别恶意反馈和正常反馈,并未从时间序列角度考虑数据前后的相关性,因而在 ME 比例较低时,对于 ME 的检测能力和识别成功率明显低于本文所采用的方法.即在 ME 占比较低时,K-Means 方法未能充分识别出恶意反馈;同时,又误将正常反馈分类为恶意反馈.K-Means 方法识别成功率随着 ME 比例的

增加,反馈数据中数据差异性提高,检测率也有所提高.而本文中所采用的方法在保持稳定检测率的同时,表现出较高的检测成功率,能够对反馈评价中的 ME 行为进行有效的监控.

4.1.2 对单个节点的信誉评估

对网络中单个节点的信誉度计算进行模拟,重点验证不同方法对于评价数据的敏感性和收敛效果的区别,不考虑数据的可信因素,即取 $Pr=1, \alpha=0.8$.

实验采用 3 种不同的方式对于单个服务的信誉度评估算法进行比较^[7].即对不同维度的评价信息,分别采用平均值算法(AVE)^[21]、基于熵权的信誉算法(EWM)和本文的引入效用函数后的信誉计算方式(TE)对信息的整合效果作对比.实验中,将 QWS 数据集中包含同一交互内容的 service 划分为同一类,选取该类别中 100 次交互反馈结果作为评价指标数据,据此计算得到针对不同评价指标的指标权重.由于不同 Agent 对于交互要求的苛刻程度不同,本文采用与文献[43]相同的反馈评价结果服从正态分布的假设.

图 10 给出了 3 种方法的评价信息整合对比效果.实验结果表明,采用本文提出的方法扩大了评分数据的离散化程度,相对于平均值算法和基于熵权的算法,反馈数据方差分别提高了 33%和 24%,使反馈评价数据呈现出明显的好坏区别.从图 11 中可以看出,考虑了信誉社会属性的基于效用反馈评价的信誉度评估方法 CIN 明显低于采用平均值的信誉评估方式,达到了较好的收敛效果.

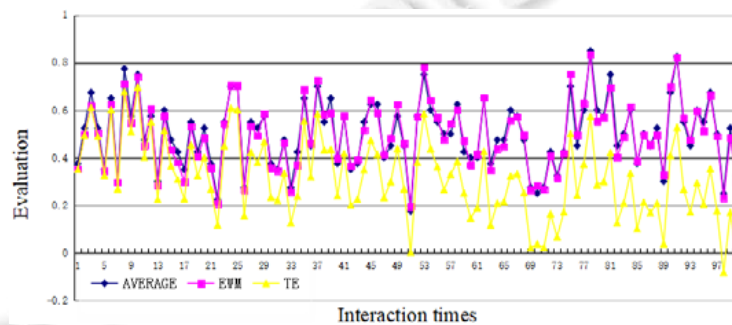


Fig.10 Comparison of single evaluation

图 10 单次评价对比

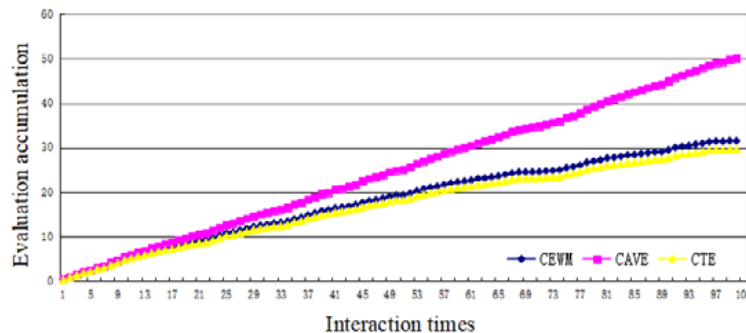


Fig.11 Comparison of evaluation cumulative

图 11 评价累积对比

4.1.3 对多个节点的信誉评估

对网络中的多个节点信誉进行仿真模拟,重点在于验证引入个体影响力后对于综合评估算法 CRCM 的修正作用,以及对于共谋攻击 CA 行为的抵抗能力.作为参照对本文模型与 Sporas 模型计算性能进行比较,表 1 为部分实验参数.

Table 1 Parameter settings

表 1 参数设置

参数设置	CRCM	Sporas
实体总个数	50	50
模拟运行次数	500	500
纳入计算的交互次数 a	3	1
交互归属感 s	1	无

实验采用 NetLogo 模拟了一个拥有 50 个对象的交互系统,假设能够对系统中的所有个体成功定位.每个对象任意选择与其他个体进行至少 1 次交互活动,共进行 500 次的交互行为,并在交互活动之后进行反馈评价(这里的反馈值是指经过过滤并结合效用函数计算得到的反馈评价结果).将 50 个个体分为 A,B 两类,其中,20 个 A 类个体只与同类个体进行交互,进行了 200 次交互活动;另外 30 个 B 类个体获得剩余 60%的交互机会,并随机选择交互对象,据此构建出交互网络^[7].

实验中不断增加 CA 在整个评价反馈中所占的百分比 $percentage=new/original,new$ 表示 A 类个体新获得的交互机会, $original$ 表示初始状态所有节点交互总次数为 500.依据第 2 节中对模型的前提设定,实验中对 CRCM 设置 $a=3$,即两个个体之间互相给与交互评价时,反馈数据只取前 3 次列入信誉计算过程中.通过两种方式,Sporas 模型^[24]和本文中的 CRMC 模型对个体信誉评估的 SER 和 RAE 进行比较.

从图 12 可以看出,本文所采用的 CRCM 算法随着 CA 类反馈评价的比例不断升高,其 SER 有一定的下降;当 CA 行为所占比例达到 50%时,即 A 类个体新获得 250 次交互机会时,CRCM 算法对应的 SER 仍可达到 83%.与此对应,采用 Sporas 信誉计算评估方法由于将评价者的信誉作为评价可信度,因此对 CA 行为的抵抗能力较差,其准确度已经下降至 68%.为了进一步说明 CRCM 算法对于抵抗 ME 行为的有效性,将 RAE 纳入考量.如图 13 所示,随着 CA 类反馈比例不断升高,CRCM 的误差也明显小于 Sporas 模型.该实验说明,在面临共谋攻击时,由于加入了 PageRank,降低了 CA 行为对于信誉评估的影响,使得本文提出的 CRCM 计算方法能够更好地表现出信誉的真实情况.

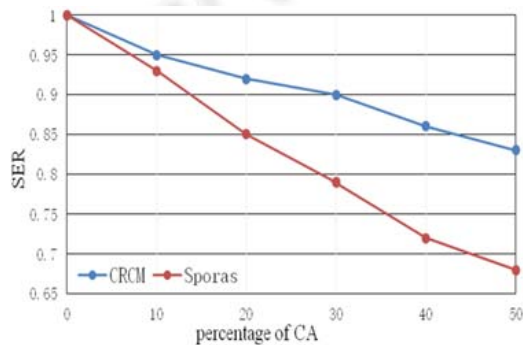


Fig.12 Comparison of SER
图 12 评估准确度对比

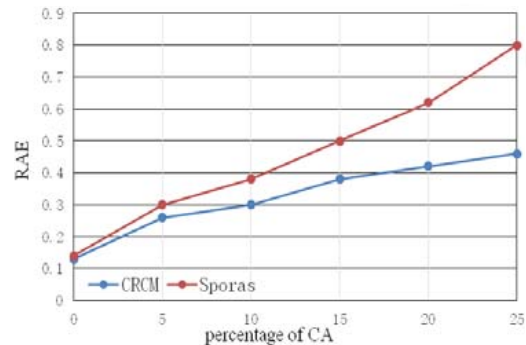


Fig.13 Comparison of RAE
图 13 评估误差对比

4.2 结果讨论

如第 4.1 节实验结果所示,本文提出的信誉评估算法的 3 个组成部分在确保反馈信息可信的同时,能够充分利用反馈属性信息^[7],在进行信誉评估之前通过改进的 CUSUM 算法对评分数据集进行预处理,对反馈数据进行连续监测,与其他过滤方式相比动态适应性更强.实验中,该方法检测正确率和成功率在恶意攻击下能够保持相对稳定;采用基于熵权的方法对过滤后的反馈数据进行了有效的整合,在引入评价真值及效用函数后,增强了评价反馈的现实基础,提高了信誉的收敛速度,这使得信誉能够较快达到稳定状态,在动态变化的网络中提高信誉计算的动态适应性;运用改进的 PageRank 算法,从数据来源角度对 Agent 影响力进行刻画.实验中没有单独对不同 Agent 影响力计算方法与其他算法进行比较,而是将该算法融入信誉综合计算的过程中,着重考察融合算

法对于共谋攻击的抵抗能力.实验结果表明,融合后的算法对于共谋攻击表现出较强的抵抗能力.经过实验分析,CRCM算法性能良好的原因在于,基于链接分析的PageRank方法具有一定的客观性,这与主观的反馈评价相互补充,从而使这两部分综合计算得到的信誉获得了较好的评估准确性,对于ME行为和CA行为表现出较高的抵抗能力.综上所述,本文提出的方法解决了当前信誉研究存在的关键性问题,表现出较好的性能.

5 结论及未来工作展望

本文提出了一种MAS环境下的信誉计算方法CRCM,从反馈数据以及反馈数据来源两个方面考虑,在对反馈数据进行合理过滤的基础上,采用基于多属性融合的节点反馈评价方法,并综合考虑节点在网络中的链接关系分析方法对网络中节点的信誉进行综合评估计算.实验结果表明,该信誉综合评估算法对于抵抗恶意攻击和加快迭代计算收敛性上有较明显的作用.

本文的信誉计算方法的基础来源于反馈评价数据.在现实生活中,由于缺乏有效的鼓励机制,所能得到的有效反馈信息数量并不充足.部分个体出于安全性的考量,通常会避免进行评分较低的反馈,这对信誉评估的准确性造成了一定影响.如何提高诚信个体进行评价活动的积极性,同时对恶意个体进行惩罚,是有待解决的另外一个关键问题.同时,如何结合信誉系统或应用程序中所收集到的隐式评分对用户信誉行为进行检测,需要进一步研究.

References:

- [1] Jøsang A, Ismail R, Boyd C. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 2007, 43(2):618–644.
- [2] Sherchan W, Nepal S, Paris C. A survey of trust in social networks. *ACM Computing Surveys*, 2013,45(4):1–33.
- [3] Golbeck J. Trust on the World Wide Web: A survey. *Foundations and Trends in Web Science*, 2006,1(2):131–197.
- [4] Ruan YF, Durresi A. A survey of trust management for online social communities—Trust modeling, trust inference and attacks. *Knowledge-Based Systems*, 2016,106:150–163.
- [5] Hoffman K, Zage D, NitaRotaru C. A survey of attack and defense techniques for reputation systems. *ACM Computing Surveys*, 2009,42(1):1–31.
- [6] Wang YN, Singh MP. Trust representation and aggregation in a distributed Agent system. In: *Proc. of the National Conf. on Artificial Intelligence*. 2006. 1425–1430.
- [7] Zhang YY. Research on reputation-based trust mechanism in open network environment [MS. Thesis]. Hefei: Hefei University of Technology, 2019 (in Chinese with English abstract).
- [8] Alrifai M, Dolog P, Balke WT, Nejdl W. Distributed management of concurrent Web service transactions. *IEEE Trans. on Services Computing*, 2009,2(4):289–302.
- [9] Jiang ZY, Han JH, Wang Z. Optimization model for dynamic QoS-aware Web services selection and composition. *Chinese Journal of Computers*, 2009,32(5):1014–1025 (in Chinese with English abstract).
- [10] Wang SG, Sun QB, Yang FC. Reputation evaluation approach in Web service selection. *Ruan Jian Xue Bao/Journal of Software*, 2012,23(6):1350–1367 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4051.htm> [doi: 10.3724/SP.J.1001.2012.04051]
- [11] Shen HW, Shao K, Zhang YY, Huo X, Liu ZT. Trust decision model based on naive Bayesian. *Journal of Chinese Computer Systems*, 2018,39(2):275–279 (in Chinese with English abstract).
- [12] Shambour Q, Lu J. A hybrid trust-enhanced collaborative filtering recommendation approach for personalized government-to-business e-services. *Int'l Journal of Intelligent Systems*, 2011,26(9):814–843.
- [13] Fang H, Zhang J, Sensoy M, Magnenat-Thalmann N. Reputation mechanism for e-commerce in virtual reality environments. *Electronic Commerce Research & Application*, 2014,13(6):409–422.
- [14] Zhu ML, Jin Z. Approach for evaluating the trustworthiness of service agent. *Ruan Jian Xue Bao/Journal of Software*, 2011,22(11):2593–2609 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3921.htm> [doi: 10.3724/SP.J.1001.2011.03921]
- [15] Massa P, Avesani P. Controversial users demand local trust metrics: an experimental study on epinions.com community. In: *Proc. of the 20th National Conf. on Artificial Intelligence*. 2005. 121–126.

- [16] Kamvar SD, Schlosser MT, Garcia-Molina H. The eigen-trust algorithm for reputation management in P2P networks. In: Proc. of the 12th Int'l Conf. on World Wide Web. ACM, 2003. 640–651.
- [17] You J, Shangguan JL, Xu SK, Li QM, Wang YH. Distributed dynamic trust management model based on trust reliability. Ruan Jian Xue Bao/Journal of Software, 2017,28(9):2354–2369 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5180.htm> [doi: 10.13328/j.cnki.jos.005180]
- [18] Gan ZB, Ding Q, Li K, Xiao GQ. Reputation-based multi-dimensional trust algorithm. Ruan Jian Xue Bao/Journal of Software, 2011,22(10):2401–2411 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3909.htm> [doi: 10.3724/SP.J.1001.2011.03909]
- [19] Duma C, Shahmehri N, Caronni G. Dynamic trust metrics for peer-to-peer systems. In: Proc. of the Int'l Workshop on Database and Expert Systems Applications. Washington: IEEE Computer Society Press, 2005. 776–781.
- [20] Fan Y, Ju JD, Xiao M. Reputation premium and reputation management: Evidence from the largest e-commerce platform in China. Int'l Journal of Industrial Organization, 2016,46(6):63–76.
- [21] Cabral L, Hortacsu A. The dynamic of seller reputation: Evidence from eBay. Journal of Industrial Economics, 2010,58(1):54–78.
- [22] Jøsang A, Ismail R. The beta reputation system. In: Proc. of the 15th Bled Electronic Commerce Conf. 2002.
- [23] Teacy WTL, Patel J, Jennings NR, Luck M. TRAVOS: Trust reputation in the context of inaccurate information sources. Autonomous Agents and Multi-Agent Systems, 2006,12(2):183–198.
- [24] Zacharia G, Moukas A, Maes P. Collaborative reputation mechanisms for electronic marketplaces. Decision Support Systems, 2000, 29(4):371–388.
- [25] Dan C, Lam SK, Albert I, Konstan JA, Riedl J. Is seeing believing? How recommender interfaces affect user's opinions. In: Proc. of the Sigchi Conf. on Human Factors in Computing Systems, Vol.5. 2003. 585–592.
- [26] Goldberg K, Roeder T, Gupta D, Perkins C. Eigentaste: A constant time collaborative filtering algorithm. Information Retrieval, 2001,4(2):133–151.
- [27] Griffiths N. Task delegation using experience based multi-dimensional trust. In: Proc. of the Int'l Joint Conf. on Autonomous Agents and Multi-agent Systems. 2005. 89–496.
- [28] Yan Z, Jing XY, Pedrycz W. Fusing and mining opinions for reputation generation. Information Fusion, 2017,36:172–184.
- [29] Huynh TD, Jennings NR, Shadbolt NR. An integrated trust and reputation model for open multi-agent systems. Autonomous Agents and Multi-Agent Systems, 2006,13(2):119–154.
- [30] You L, Sikora R. An adaptive evaluation mechanism for online traders. European Journal of Operational Research, 2011,214(3): 739–748.
- [31] Su XF, Zeng HJ, Chen Z. Finding group shilling in recommendation system. In: Proc. of the Special Internet Tracks and Posters of 14th Int'l Conf. on World Wide Web. 2005. 960–961.
- [32] Chirita PA, Nejd W, Zamfir C. Preventing shilling attack in online recommender system. In: Proc. of the 7th Annual ACM Int'l Workshop on Web Information and Data Management. 2005. 80–89.
- [33] Yu B, Singh MP. A social mechanism of reputation management in electronic communities. In: Proc. of the 4th Int'l Workshop on Cooperative Information Agents. 2000. 154–165.
- [34] Shao K, Luo F, Mei NX, Liu ZT. Normal distribution based dynamical recommendation trust model. Ruan Jian Xue Bao/Journal of Software, 2012,23(12):3130–3148 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4204.htm> [doi: 10.3724/SP.J.1001.2012.04204]
- [35] Zhang SB, Xu CX. Study on the trust evaluation approach based on cloud model. Chinese Journal of Computers, 2013,36(2): 422–431 (in Chinese with English abstract).
- [36] Brodsky B. Non-parametric Methods in Change-point Problems. Kluwer Academic Publishers, 1993.
- [37] Siris VA, Papagalou F. Application of anomaly detection algorithms for detecting SYN flooding attacks. Computer Communications, 2006,29(9):1433–1442.
- [38] Sun ZX, Tang YW, Cheng Y. Router anomaly traffic detection based on modified-CUSUM algorithms. Ruan Jian Xue Bao/Journal of Software, 2005,16(12):2117–2123 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/16/2117.htm> [doi: 10.1360/jos162117]
- [39] Chen W, He YX, Peng WL. A light weight detection method against DDoS attack. Chinese Journal of Computers, 2006,29(8): 1392–1400 (in Chinese with English abstract).

- [40] Limam N, Boutaba R. Assessing software service quality and trustworthiness at selection time. *IEEE Trans. on Software Engineering*, 2010,36(4):559–574.
- [41] Wang Y, Vassileva J. A review on trust and reputation for Web service selection. In: *Proc. of the Int'l Conf. on Distributed Computing Systems Workshops*. 2007. 25–32.
- [42] Li Y, Zhou MH, Li RC, Cao DG, Mei H. Service selection approach considering the trustworthiness of QoS data. *Ruan Jian Xue Bao/Journal of Software*, 2008,19(10):2620–2627 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/19/2620.htm> [doi: 10.3724/SP.J.1001.2008.02620]
- [43] Haveliwala TH. Topic-sensitive PageRank: A context-sensitive ranking algorithm for Web search. *IEEE Trans. on Knowledge and Data Engineering*, 2003,15(4):784–796.
- [44] Al-Masri E, Mahmoud QH. QoS based discovery and ranking of Web service. In: *Proc. of the IEEE 16th Int'l Conf. on Computer Communications and Networks (ICCCN)*. 2007. 529–534.
- [45] Dakhel GM, Mahdavi M. A new collaborative filtering algorithm using K-means clustering and neighbor's voting. In: *Proc. of the Int'l Conf. on Hybrid Intelligent Systems*. 2012. 179–184.

附中文参考文献:

- [7] 张阳洋. 开放网络环境下基于信誉的信任机制研究[硕士学位论文]. 合肥: 合肥工业大学, 2019.
- [9] 蒋哲远, 韩江洪, 王钊. 动态的 QoS 感知 Web 服务选择和组合优化模型. *计算机学报*, 2009,32(5):1014–1025.
- [10] 王尚广, 孙其博, 杨放春. Web 服务选择中信誉度评估方法. *软件学报*, 2012,23(6):1350–1367. <http://www.jos.org.cn/1000-9825/4051.htm> [doi: 10.3724/SP.J.1001.2012.04051]
- [11] 沈宏伟, 邵堃, 张阳洋, 霍星, 刘宗田. 基于朴素贝叶斯的信任决策模型. *小型微信计算机系统*, 2018,39(2):275–279.
- [14] 朱曼玲, 金芝. 一种服务 Agent 的可信性评估方法. *软件学报*, 2011,22(11):2593–2609. <http://www.jos.org.cn/1000-9825/3921.htm> [doi: 10.3724/SP.J.1001.2011.03921]
- [17] 游静, 上官经伦, 徐守坤, 李千目, 王印海. 考虑信任可靠度的分布式动态信任管理模型. *软件学报*, 2017,28(9):2354–2369. <http://www.jos.org.cn/1000-9825/5180.htm> [doi: 10.13328/j.cnki.jos.005180]
- [18] 甘早斌, 丁倩, 李开, 肖国强. 基于声誉的多维度信任计算算法. *软件学报*, 2011,22(10):2401–2411. <http://www.jos.org.cn/1000-9825/3909.htm> [doi: 10.3724/SP.J.1001.2011.03909]
- [34] 邵堃, 罗飞, 梅袅雄, 刘宗田. 一种正态分布下的动态推荐信任模型. *软件学报*, 2012,23(12):3130–3148. <http://www.jos.org.cn/1000-9825/4204.htm> [doi: 10.3724/SP.J.1001.2012.04204]
- [35] 张仕斌, 许春香. 基于云模型的信任评估方法研究. *计算机学报*, 2013,36(2):422–431.
- [38] 孙知信, 唐益慰, 程媛. 基于改进 CUSUM 算法的路由器异常流量检测. *软件学报*, 2005,16(12):2117–2123. <http://www.jos.org.cn/1000-9825/16/2117.htm> [doi: 10.1360/jos162117]
- [39] 陈伟, 何炎祥, 彭文灵. 一种轻量级的拒绝服务攻击检测办法. *计算机学报*, 2006,29(8):1392–1400.
- [42] 李研, 周明辉, 李瑞超, 曹东刚, 梅宏. 一种考虑 QoS 数据可信性的服务选择方法. *软件学报*, 2008,19(10):2620–2627. <http://www.jos.org.cn/1000-9825/19/2620.htm> [doi: 10.3724/SP.J.1001.2008.02620]



霍星(1979—),女,浙江余姚人,博士,副教授,CCF 专业会员,主要研究领域为软件理论方法,非线性计算方法.



景永俊(1977—),男,博士,高级工程师,CCF 专业会员,主要研究领域为软件工程,社会计算,机器学习.



张阳洋(1993—),男,硕士,主要研究领域为可信计算,数据挖掘.



邵堃(1967—),男,博士,副教授,CCF 高级会员,主要研究领域为软件需求分析与建模.