

物联网搜索中主客观融合的动态信誉评估

张会兵¹, 李超², 胡晓丽¹, 周娅¹

(1. 桂林电子科技大学 广西可信软件重点实验室, 广西 桂林 541004; 2. 中国科学院 信息工程研究所, 北京 100093)

摘 要: 与传统的互联网搜索相比, 物联网搜索更加强调数据质量。为了提高数据质量, 物联网搜索中心需要依据数据质量为提供者支付相应报酬以激励其持续提供符合质量需求的数据。这就使如何评估数据提供者的数据质量及其信誉成为物联网搜索中的一个基础问题。为此, 引入动态信誉机制来综合评估数据提供者的可信性, 为数据选择、收集提供依据。首先, 提出了主观意愿及客观质量评价计算方法, 并设计了交互行为贴现及信誉衰减机制; 然后, 提出了融合主客观要素的动态信誉计算模型, 并基于信号传递机制进行博弈分析, 以获取交互次数、贴现率、支付价格及数据成本之间约束关系。实验证明该模型能够较好地反映数据质量变化情况, 为数据选择提供依据, 并满足物联网搜索的实时性及动态性需求。

关键词: 物联网搜索; 信誉; 数据质量; 主客观因素

中图分类号: TP309.2

文献标识码: A

Fusing subjective and objective factors: a dynamic approach to evaluating reputation for IoT search

ZHANG Hui-bing¹, LI Chao², HU Xiao-li¹, ZHOU Ya¹

(1. Guangxi Key Lab of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China;

2. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China)

Abstract: Compared with the traditional Internet search, IoT search data center needs a higher data quality. In order to effectively motivate data owners to continuously provide higher quality data, the center needs pay the corresponding rewards to them according to the quality of data (QoD). As a result, QoD and the trustworthy evaluation become a basic problem for the development of IoT search. To address this problem and support data selection of IoT center, a dynamic reputation model was proposed to comprehensively assess the reputation of data owners. In detail, first, an approach to assessing the subjective and objective quality was proposed and mechanism of interactive discount and reputation attenuation was designed. Then, fusing subjective and objective factors, a novel dynamic reputation evaluation scheme was presented. Last, in order to get constraint relation of interaction times, discount, payment price, and data cost, an economic analysis based on signaling game was conducted. Experiment results show that the proposed approach can reflect dynamic change of QoD effectively, and provide the basis for data selection. It also can be adapted to real-time and dynamic of IoT search.

Key words: IoT search; reputation; data quality; subjective and objective factors

1 引言

随着物联网、传感器、移动互联网技术的发展

融合及应用普及, 网络空间中智能终端节点数量急剧增加。据 IDTechEX 预测, 未来 10 年全球智能终端数量将突破 10 万亿台。其中, 我国物联网终端

收稿日期: 2015-09-22; 修回日期: 2015-12-14

基金项目: 国家自然科学基金资助项目(61562015,61363005); 广西省自然科学基金资助项目(2014GXNSFDA118036, 2013GXNSFAA019324); 广西高等学校高水平创新团队及卓越学者计划基金资助项目; 广西高校科研重点基金资助项目(ZD2014049)

Foundation Items: The National Natural Science Foundation of China (61562015, 61363005); The Natural Science Foundation of Guangxi Province(2014GXNSFDA118036,2013GXNSFAA019324); The High Level of Innovation Team of Colleges and Universities in Guangxi and Outstanding Scholars Program Funding; Education Department of Guangxi (ZD2014049)

数量将达到 100~1 000 亿，这些种类繁多的终端节点将物体、人和信息充分互联在一起，并为社会资源、物理资源、信息资源的深度融合及智能应用奠定了基础^[1]。在智慧城市、智慧旅游、环境监测、健康管理等领域有着广泛应用。面对由如此庞大终端节点、海量实时异构数据和智能服务组成的物联网空间，物联网搜索为快速准确地发现用户所需的数据和服务提供了重要途径。

传统网络搜索(如百度、Google 等)的数据主要来源于公开共享且免费的 Web 页面。这些页面数据往往是无状态的，一旦获取就可以长期使用，并且数据质量受环境及其他人为因素影响很小。与传统的基于网页搜索系统不同，物联网搜索系统中的数据有很大一部分来自于各种智能终端、传感器的监测感知信息。这些信息与特定时空要素相关的有状态信息，反映了被监测对象的状态、行为等；同时，这些数据的实时监测、生成与传输往往需要相应的费用，数据提供者为了提供比较高的数据质量往往需要更多的投入以提升监测设备的精度、网络传输的带宽、节点的供电能力等。所以，数据拥有者因安全、隐私、成本等原因往往不愿无偿分享相关数据。因此，为了得到符合质量要求的数据，物联网搜索中心通常会采取一些措施，如用支付报酬方式来激励数据提供者提供高质量的数据。这样就带来了一个基本问题：数据提供者提供的数据质量存在差别，且存在着部分数据提供者为了获取超额的报酬出现不诚实行为。在这种情况下，物联网搜索中心如何评估、选择合适的数据提供者并支付合理报酬成为一个需要解决的基本问题。通常地，基于信誉的信任评估机制是解决这类问题的有效途径。

针对上述问题，本文提出了主客观评价融合的动态信誉计算模型：将数据提供者的行为动机、方式态度等建模成信誉计算的主观要素；将数据提供者提供的数据质量建模成信誉计算的客观要素；并在时间域上对主客观要素进行有机融合。然后，将数据提供者的信誉作为其能提供的数据质量信号，基于信号传递模型分析了信誉形成过程中的 2 个关键要素：博弈次数与支付(贴现)之间的约束关系。

2 相关工作

在美国国家自然科学基金委的资助下，Ostermaier 最早对物联网搜索进行研究并开发了

Dyser 系统^[2]。随后物联网搜索研究受到越来越广泛的关注，如 Chen^[3]、Miliaraki^[4]、Shen^[5]及 Ebrahimi^[6]等。这些研究或系统有一个基本假设：数据提供者主动、诚实地向搜索中心提供符合质量要求的数据。因此，在这些工作中没有涉及到数据提供者动机、能力等信誉情况的研究。据我们所知，目前还没有针对物联网搜索中数据提供者信誉值计算模型的研究。因此，本文主要从具有类似特征的服务计算及社会化商务领域中信任管理与信誉计算的角度来介绍一些典型研究成果。

Su^[7]提出了一个基于优先级的信誉计算模型，该模型从 4 个方面来建模服务提供者的信誉度：服务提供者的服务经验、服务请求者与数据提供者关于服务质量属性优先级分布的相似度、候选服务对服务需求的适应性以及由第三方提供的评价信息。这 4 个属性共同决定了最终的信誉值。Zhang 等^[8]将对服务的各种评价信息转换为对信誉计算的支持证据，基于证据理论来计算服务信誉值。Adali 根据社交网络中用户的通信行为方式建立了相应的行为信誉计算模型：综合会话信任和传播信任来计算行为可信度，该模型的前提假设是有规律的通信行为比随机通信行为有更高可信度。其中，会话信任用来表示用户之间的会话时间和会话频率，频率越高、时间越长表明他们之间信任度越高；传播信任是指信息的传播度，信息从一个用户传到其他用户的越多说明对信息源用户的信任度越高^[9]。Nepal 基于社会资本(social capital)、交互关系、社会信任之间的基本关系，建立了融合声望信任(popularity trust)和参与信任(engagement trust)的信任模型，其目标是提升社区的整体信任度。前者指用户被社区其他用户接受或认可的程度，它从社区角度来评估特定用户的信誉度；后者指用户参与社区的程度，它从特定用户的角度来评估社区信誉度^[10]。文献[11]提出了基于图的综合信誉模型，该模型通过融合用户行为和社会关系来降低直接交互信息稀疏问题，并通过关键因子识别技术来降低行为动态变化及评价主观性带来的影响。文献[12]通过利用潜在因子来发现用户之间的潜在相似度以解决信誉值的冷启动问题。针对网络交易信任问题的动态性、匿名性、虚拟性等特点，文献[13]提出了基于多 Agent 系统的网络交易动态信任计算模型与信誉系统。模型包括用户时域的信任度、信誉反馈评分的加权平均计算及社区贡献加权，建立了事前

开展防范, 事中进行协调, 事后给予惩罚三位一体的信誉约束机制。文献[14]提出了由交互经验计算用户态度, 以及利用交互时间序列计算用户行为模式变化的方法, 进而利用态度、交互经验、行为模式 3 种信息综合建立新的信任评估模型。文献[15]给出了合作认证中基于效用的合作决策机制。

服务计算中的信誉问题主要关注的是用户对服务执行情况的评价, 而社会化商务中的信誉问题主要从用户行为及用户间的拓扑入手研究评价信息的可信性。而物联网搜索中的信誉更加关注数据提供者的行为模式、数据及数据的状态信息等。针对这一特点, 本文引入主观意愿和客观质量评价机制, 给出了相应的评价方法及信誉计算模型。

3 信誉计算模型

3.1 物联网搜索基本交互模型

物联网搜索系统的基本构成如图 1 所示, 共有 $m+n+1$ 个实体(机构、个人或智能设备等)参与交互: 1 个物联网搜索中心 C 、 m 个用户(u_1, u_2, \dots, u_m)、 n 个数据提供者(p_1, p_2, \dots, p_n)。当用户向物联网搜索中心 C 提出搜索请求时, 物联网搜索中心 C 向若干潜在的数据提供者(依据其以往的信誉值来选择)发出数据请求。每个接收到请求的数据提供者如果决定提供数据的话就返回当前能够提供的数据质量信息 $AQ(q_1, q_2, \dots, q_i)$ (假定值越小越好), 搜索中心依据收到的数据质量信息及该数据提供者的信誉值 R 来选择相应的数据提供者(一个或多个), 然后被选中的数据提供者将相应的数据发送给搜索中心 C , 搜索中心 C 根据收到数据的实际质量 $RQ(q_1, q_2, \dots, q_i)$ 来支付相应的报酬并更新数据提供者的信誉值 R 。最后, 搜索中心将相应的数据返回给请求用户。本

文将这样一次数据交互所对应的特定时间记为 t_i 。通常地, 物联网搜索中心需要与数据提供者多次交互才能获取到一个完整的数据序列, 这里将包含多次交互的某一特定时间段记为 T_i , 并由 $T_0, T_1, \dots, T_i, T_{i+1}, \dots$ 构成长期完整的交互时域 T 。

由于物联网搜索中心数据获取的特殊性, 为了经济有效地获得数量和质量都能够满足用户需求的数据, 在该模型中引入信誉机制来评估数据提供者的数据质量(即数据可信性、真实度), 并通过支付相应的报酬来激励数据提供者积极提供真实、可信的高质量数据。从本质上讲, 这是通过采用信誉机制来保证数据提供者与物联网搜索中心数据交互的顺利进行。

3.2 信誉计算中的关键因素

由图 1 可以看出, 这里的信誉机制属于典型的集中式信誉计算模型: 物联网搜索中心对每个数据提供者分别计算其信誉值并存储在本地, 当需要时依据该信誉值来选择相应的数据提供者, 并根据数据提供者最新的数据质量来更新其信誉值, 为下次选择提供依据。为得到准确的信誉值, 本文从数据提供者的主观态度意愿、提供数据的客观真实质量、以及在时间频度域上的演变这 3 个维度来综合刻画、建模信誉形成机制, 可以表示为: $R = TW(A, T) = TW(F(O, S), T)$, 函数 TW 表示在时间域上随着数据交互的信誉值变化情况。其中, 搜索中心 C 与数据提供者的每次交互 A_i 用函数 $F(O, S)$ 来表示, O 代表客观评价函数, S 代表主观评价函数。

1) 主观意愿函数

与基于网页数据的传统搜索系统不同的是, 物联网搜索中的数据获取有其诸多特殊性。首先, 在物联网搜索中往往关注的是客观物理实体及其实时状态信息, 需要实时监测、传输。出于安全、隐

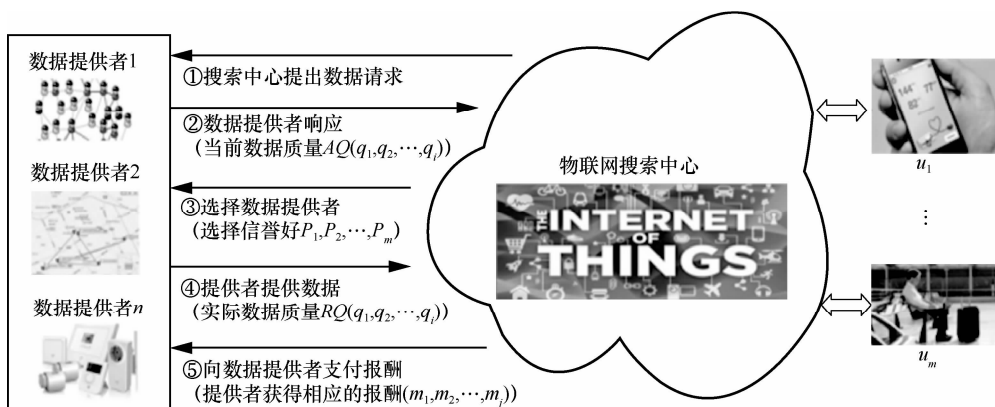


图 1 物联网搜索系统基本模型

私考虑或实体的计算调度、通信能力等原因导致提供的数据出现滞后，使数据的效用降低，甚至变成无用数据，这种情形称之为在交互时间 t_i 的数据响应实时性因素 rt 。将 rt 表示为

$$rt = \begin{cases} 1 - \frac{t_e - t_r}{\tau}, & t_e - t_r \leq \tau \\ 0, & t_e - t_r > \tau \end{cases} \quad (1)$$

其中， t_e 表示数据响应时间， t_r 表示数据请求时间， τ 表示对该次数据请求能够允许的最大响应延时间。

其次，当众多的数据提供者存在利益竞争或某些数据提供者为了获取超额不正当利益时，往往虚报能够提供的数据质量信息 $AQ(q_{a1}, q_{a2}, \dots, q_{ai})$ ，在这种情形下称实际提供的数据质量 $RQ_k(q_{r1}, q_{r2}, \dots, q_{ri})$ 与 $AQ(q_{a1}, q_{a2}, \dots, q_{ai})$ 之间的偏差为数据质量声明真实性 tr ，偏差越大表示真实性越低。这里用数据质量向量 AQ 与 RQ_k 的余弦相似度来表示

$$tr = \cos(AQ, RQ_k) \quad (2)$$

第三，物联网搜索中的数据往往表现为时空、状态、轨迹、行为模式等数据类型，这些数据类型往往需要在一个相对完整的时间内进行连续监测响应，数据提供者是否愿意提供持续性的响应是保障数据完整性及其有效性的关键，并且数据的完整性与其有效性是非线性关系，随着数据缺失比例的增加其有效性加速降低。本文将这种关系表示成

$$ef = \frac{1}{\sqrt{1 + \mu^2 \left(\frac{t_{ee} - t_{es}}{t_{re} - t_{rs}} + \frac{1}{\frac{t_{ee} - t_{es}}{t_{re} - t_{rs}}} \right)^2}} \quad (3)$$

其中， μ 表示有效性因子，它决定了数据的完整性与有效性之间的映射关系， μ 值越大对完整性要求越高； t_{rs} 、 t_{re} 表示搜索中心要求开始和结束的时间， t_{es} 、 t_{ee} 表示实际数据交互开始和结束的时间，并且 $0 < \frac{t_{ee} - t_{es}}{t_{re} - t_{rs}} \leq 1$ 。

在以上 3 个要素中， rt 、 tr 用来表征每一次数据交互的效果，而 ef 用来表征一个数据交互序列的效果。因此，用主观意愿函数 $S_i(rt, tr)$ 来刻画数据提供者实际提供数据时的主观因素， ef 对信誉值的影响后面再讨论。其中， α 、 β 是调节系数。

$$S_i(rt, tr) = \alpha \cdot rt - \beta \cdot tr \quad (4)$$

2) 客观评价函数

当数据提供者向物联网搜索中心提供数据时，由于设备精度、运行状态、环境状况等原因使发送到搜索中心的实际数据质量处于动态变化之中，这种变化可以从多个维度来刻画。首先，不同性能的设备受环境影响不同，其实，实际监测精度或准确度会随着环境变化出现波动。波动范围的大小反映了该监测设备数据信息的稳定性、可信性，用函数 $flu = \cos(st(pre_1, pre_2, \dots, pre_i), mo_k(pre_1, pre_2, \dots, pre_i))$ 来表示波动性，其中 $st(pre_1, pre_2, \dots, pre_i)$ 表示设备的标准精度， $mo_i(pre_1, pre_2, \dots, pre_i)$ 表示设备在 t_i 时刻交互时的实际监测精度(受环境等因素影响会出现波动)。

其次，物联网搜索中心在发出数据请求时对相应数据的质量会有一个需求预期 $de(q_1, q_2, \dots, q_i)$ ，但数据提供者的实际监测值 $mo_i(q_1, q_2, \dots, q_i)$ 往往与该要求存在一定偏差。偏差的大小反映了满足需求的程度，用函数 $dev(de(q_1, q_2, \dots, q_i), mo_i(q_1, q_2, \dots, q_i))$ 表示偏差值。这里采用 2 个向量之间的欧氏距离来表示偏差大小，并采用如下公式

$$dev(de, mo_i) = \begin{cases} \sqrt{(de - mo_i)(de - mo_i)^T} & (|de| \geq |mo_i|) \\ -\sqrt{(de - mo_i)(de - mo_i)^T} & (|de| < |mo_i|) \end{cases} \quad (5)$$

最后，当物联网搜索中心向潜在的数据提供者发出数据请求时，假定有 $n(n > 0)$ 个数据提供者愿意提供数据，其最终提供的数据质量分别为 $mo_1(q_1, q_2, \dots, q_i)$ ， $mo_2(q_1, q_2, \dots, q_i)$ ， \dots ， $mo_n(q_1, q_2, \dots, q_i)$ 。由前面的讨论可知，由于监测设备本身的性能差异，这 n 个潜在数据提供者的数据监测质量会存在差异，这种差异性反映了数据提供者之间的相对可信性及数据可用性。这里选取数据质量的平均值作为数据提供者的基准，数据提供者 P_i 与该均值之间的距离 dd_i 反映了该数据的可用性(数据质量向量的模越小越好)。

$$dd_i = \begin{cases} \sqrt{(ma - mo_i)(ma - mo_i)^T} & (|ma| \geq |mo_i|) \\ -\sqrt{(ma - mo_i)(ma - mo_i)^T} & (|ma| < |mo_i|) \end{cases} \quad (6)$$

其中， $ma = \frac{1}{n} \sum_{k=1}^n (mo_1 + mo_2 + \dots + mo_n)$ 。

综合以上 3 个要素，这里用函数 $O(flu, dev, dd)$ 来刻画数据提供者的数据质量，其中， λ 、 φ 、 δ 是调节系数。

$$O_i(flu, dev, dd) = \varphi dev + \delta dd_i - \lambda flu \quad (7)$$

3) 基于时间频度域的衰减机制

物联网搜索中心与数据提供者的每次交互都会产生相应的主客观评价价值，随着交互次数的增加，经过长期积累逐步建立起对每个数据提供者的总体评价，这里称之为信誉值。并且随着时间推移及交互次数的变化，该信誉值呈现动态变化。一方面，早期的信誉值随时间推移动态衰减，时间越久远衰减越多并且呈现加速衰减趋势，由此借鉴巴特沃斯滤波器设计思想，引入演变衰减函数 $A(t_k)$

$$A(t_k) = \frac{1}{1 + \left(\frac{m(t - t_k)}{t - t_0} \right)^n} \quad (8)$$

其中， $t_k \in [0, t]$ 表示最近一次交互时间， t_0 表示初始时间； t 表示当前时间；另一方面，信誉值也会随着新的交互行为动态改变，越是近期的交互行为越能真实反映数据提供者的可信度，信誉的积累在时间域上不具有简单的累加性。这里，通过引入时间贴现函数 $\Gamma(t_i, t)$ 来刻画数据交互行为在时间域上的影响权重。这里定义

$$\Gamma(t_i, t) = \rho^{\sigma(t-t_i)}, 0 < \rho < 1 \quad (9)$$

其中， t_i 表示在时间段 $[t-1, t]$ 内数据提供者 i 与 C 交互的具体时间，并通过函数 $\sigma(t-t_i)$ 将 $t-t_i$ 映射到 $[0, L]$ 离散空间中的整数，以减缓衰减过程。

3.3 综合信誉计算模型

综合上面对数据交互过程中主客观要素及时间权重因子的讨论，在时间序列 T 的给定时段 $[t-1, t]$ 内数据提供者 p 的信誉值计算模型为

$$R_t = \begin{cases} R_{t-1} A(t_k) + ef \left(\sum_{t_i \in [t-1, t]} (S_i(rt, tr) + O_i(flu, dev, dd_i)) \Gamma(t_i, t) \pi_i(dev/pay) \right) & \text{(有数据交互)} \\ R_{t-1} A(t_k) & \text{(无数据交互)} \end{cases} \quad (10)$$

其中， R_{t-1} 是最近的前一时域数据提供者的信誉值； $A(t_k)$ 是信誉值随时间变化的衰减因子；

$$ef = \left(\frac{1}{\sqrt{1 + u^2 \left(\frac{t_{ee} - t_{es}}{t_{re} - t_{rs}} + \frac{1}{\frac{t_{ee} - t_{es}}{t_{re} - t_{rs}}} \right)^2}} \right)$$

表示在时域 $[t-1, t]$

内数据序列的完整性对其信誉值的影响因子。 $S_i(rt, tr)$ 和 $O_i(flu, dev, dd_i)$ 分别是每次数据交互的主客观评价函数； t_i 表示在时域 $[t-1, t]$ 内数据提供者与物

联网搜索中心 C 每次交互的时间； $\Gamma(t_i, t)$ 表示在 t_i 时刻交互行为对当前信誉值的时间贴现函数； $\pi_i(dev/pay)$ 表示在 t_i 时刻交互行为的性价比， pay 表示物联网搜索中心为当前数据交互的支付额度。为计算方便，假设当数据质量超过搜索中心的需求标准时 $\pi_i(dev/pay) = 1.2 (dev > 0)$ ；当数据质量刚好满足搜索中心的需求标准时 $\pi_i(dev/pay) = 1 (dev = 0)$ ；当数据质量低于搜索中心的需求标准时 $\pi_i(dev/pay) = 0.8 (dev < 0)$ 。

4 基于信号传递模型的博弈分析

在物联网搜索系统中，数据提供者 p_i 和搜索中心 C 关于提供者能提供的真实数据质量信息是不对称的，并且数据提供者 p_i 知道自己能提供的真实质量，而搜索中心 C 在获得并使用该数据之前并不了解真实的质量信息。他们之间的交互是典型的不完全信息动态博弈，为保障数据交互的正常进行，通常以信誉为信号来传递质量信息，以达到有效均衡。

通过上面的讨论，知道数据提供者 p_i 的信誉值是在和搜索中心 C 多次重复博弈过程中形成的，它能很好地向搜索中心传递 p_i 的数据质量真实性信息。搜索中心 C 依据该信誉值来选择合适的 p_i 并支付相应报酬。这里基于信号传递博弈模型来分析数据提供者 p_i 与搜索中心 C 之间动态博弈中的利益关系及影响信誉值形成的关键要素。

假定数据提供者 (p_1, p_2, \dots, p_n) 集合中数据质量为低 (LQ) 的概率为 λ 、为高 (HQ) 的概率为 $1-\lambda$ 。同时，低质量的提供者其数据提供成本为 $C(LQ)$ ，高质量的数据提供者其数据提供成本为 $C(HQ)$ 。相应地，假定在信息对称情况下物联网搜索中心为高质量提供者支付的报酬为 $P(HQ)$ ，低质量数据提供者支付的报酬为 $P(LQ)$ 。进一步假定，高质量数据提供者提供真实的高质量数据付出的基础成本为 $C(B)$ ，低质量的数据提供者假冒高质量数据提供者付出的综合成本(基础成本、节点之间的合作压力等)为 $C(F)$ ，并且 $C(B) = K C(F) (K > 1 \text{ 为比例系数})$ 。

这里将一次完整的数据交互、支付过程视作一次博弈。经过 n 次博弈之后，搜索中心 C 就能够建立起数据提供者的信誉值，并且该信誉值能够真实传递数据质量信息。换句话说，搜索中心 C 能够依据该信誉值来区分高质量的数据提供者和低质量的数据提供者。由此，对于低质量数据提供者而言，其只可能在 $1 \sim n$ 次的交互中假冒高质量的数据提供

者，并获得相应的支付价格 $P(HQ)$ ， n 次以后只能以 $P(LQ)$ 的支付价格进行交互；而高质量的数据提供者可以一直以支付价格 $P(HQ)$ 与搜索中心交互。假定共同的贴现率为 ξ ，数据提供者重复交互净现值为 $U(HF, I, Q)$ ，其中， HF 表示诚实 H 或假冒 F ， I 表示一次交互， Q 表示质量高低。在数据交互博弈的过程中，达到分离均衡的 2 个必要条件是：①低质量的数据提供者不愿付出成本 $C(F)$ 来冒充高质量的数据提供者；②高质量的数据提供者愿意付出成本 $C(B)$ 来提供高质量的数据。从经济利益来看，对于低质量的数据提供者来说，只有满足 $U(H, I, LQ) > U(F, I, LQ)$ 时，条件①成立；对于高质量的数据提供者来说，只有满足 $U(H, I, LQ) < U(H, I, HQ)$ 时，条件②成立。即 $U(F, I, LQ) < U(H, I, LQ) < U(H, I, HQ)$ 时信誉机制才能发挥作用。

对于 $U(H, I, LQ) > U(F, I, LQ)$ ，由于

$$U(H, I, LQ) = \frac{P(LQ) - C(LQ)}{1 - \xi} \quad (11)$$

$$U(F, I, LQ) = \frac{(1 - \xi^n)[P(HQ) - C(LQ)]}{1 - \xi} + \frac{(\xi^n)[P(LQ) - C(LQ)]}{1 - \xi} - C(F) \quad (12)$$

由此得到

$$C(F) > \frac{(1 - \xi^n)[P(HQ) - P(LQ)]}{1 - \xi} \quad (13)$$

对于 $U(H, I, LQ) < U(H, I, HQ)$ ，由于

$$U(H, I, LQ) = \frac{P(LQ) - C(LQ)}{1 - \xi} \quad (14)$$

$$U(H, I, HQ) = \frac{P(HQ) - C(HQ)}{1 - \xi} - C(B) \quad (15)$$

由此得到

$$C(B) < \frac{[P(HQ) - C(HQ)]}{1 - \xi} - \frac{P(LQ) - C(LQ)}{1 - \xi} \quad (16)$$

结合式(13)、式(16)及 $C(B) = KC(F)$ ，可以得到

$$\xi > \sqrt[n]{\frac{1}{k} \left(\frac{P(HQ) - P(LQ)}{C(HQ) - C(LQ)} + k - 1 \right)} \quad (17)$$

进而得到

$$n < \log_{\xi} \frac{1}{k} \left(\frac{P(HQ) - P(LQ)}{C(HQ) - C(LQ)} + k - 1 \right) \quad (18)$$

由此得到数据提供者与搜索中心的博弈次数 n 与贴现率 ξ 之间的约束关系，并进一步将 n 、 ξ 与成本、数据质量与支付价格统一到信誉管理模型中。

5 数据测试与分析

这里以实时收集某路段的交通、环境信息为目标，数据提供者实体是在该路段行驶的车辆。以此为例如来验证信誉算法的性能：每天下班时间（17:40~18:00 之间），物联网搜索中每隔 2 min 发出一次信息请求，要求车辆在接收到请求之后的 $\tau = 30$ s 内返回当前所处的位置(L)、时间(T)、速度(V)、环境噪音(N)、PM2.5(P)等信息。这里选择 L 、 N 、 P 这三类信息的数据精度及其响应时间的延时、响应时间段的完整性来表征数据提供者的数据质量。

5.1 信誉计算模型的区分度

首先根据经验设定相关参数值：交互时间贴现因子 $\rho = 0.9$ ；离散空间 $L = 4$ ；衰减函数 $A(t_k)$ 中 $m = 1.5$ 、 $n = 2$ ；交互序列有效性因子 $\mu = 6.4$ 。 $\alpha = 1$ 、 $\beta = 1.5$ ； $\lambda = 2$ 、 $\varphi = 3$ 、 $\delta = 0.5$ 。按照表 1 所列的标准生成测试数据来验证信誉计算模型的有效性。其中， st 、 de 、 AQ 用数据质量向量 (L, N, P) 表示： P_{HQ} 、 P_{MQ} 、 P_{LQ} 分别代表高质量、中等质量及低质量的数据提供者。在表 1 中，如“10~14:90%”这样的数据表示数值区间在 10~14，所占整个数据的概率为 90%。

表 1 主要参数设置

项目	高 P_{HQ}	中 P_{MQ}	低 P_{LQ}
L	10~14:90%	13~15:80%	13~19:70%
	14~15:10%	15~20:20%	12~13:30%
生成测试数据质量区间及其分布	N	4~5:80%	5~6:70%
		5~8:20%	4~5:30%
P	4~8:100%	6~7:80%	8~9:75%
		7~11:20%	5~8:25%
实时性 rt	0.9~1:90%	0.8~1:80%	0.7~1:25%
	0~0.9:10%	0~0.8:20%	0~0.7:75%
完整性 ef	0.9~1:100%	0.9~1:80%	0.6~1:30%
		0~0.9:20%	0~0.6:70%
信誉初始值	1	1	1
标准精度 st	(10,3,4)	(13,4,6)	(12,4,5)
声明精度 AQ	(12,4,6)	(14,5,7)	(13,4,6)
需求预期 de		(15,5,8)	

根据表 1 的参数设置要求，随着交互次数增加/时间推移的信誉值变化情况如图 2 所示。由图可以看出所提信誉计算模型具有较好的区分度：无论是从单一交互序列，还是从多次交互来看，信誉值都

能够较好地反映数据质量变化情况，并且具有较好灵敏度。

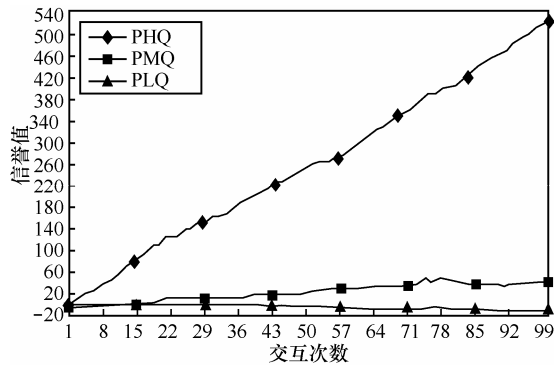


图 2 3 种不同质量提供者的信誉值变化

5.2 信誉计算模型中影响因子分析

以表 1 中数据质量为中等的数据提供者 P_{MQ} 为例，测试实时性 rt 、完整性 ef 、时间贴现函数 ρ 以及信誉值衰减函数中 m 、 n 在不同取值情况下的信誉值的变化情况，分析信誉计算模型对各个因子的敏感性。

1) 实时性因素分析

在其他参数不变的情况下，数据响应时间 t_e 分别取值 6、12、18 和 24 时的信誉值变化情况如图 3 所示。由图可以看出，随着响应延时的增加信誉值明显降低，对响应时间具有较好的敏感性。

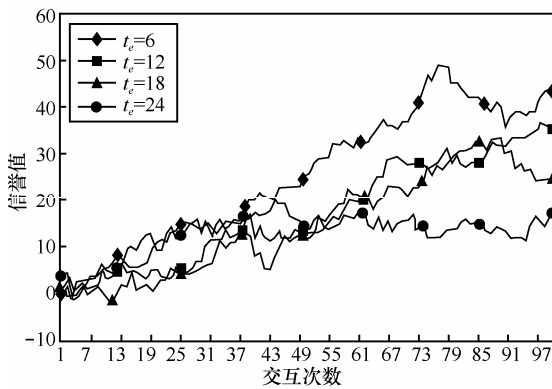


图 3 不同响应时间延时对信誉值的影响

2) 响应数据的完整性因素分析

响应数据的完整性对信誉值的影响表现在 2 个方面：有效性因子 μ 对信誉值的影响以及数据序列完整性对信誉值的影响。图 4 是 μ 分别取 3.2、6.4 及 9.6 这 3 种情况下信誉值的变化情况；图 5 是在 $\mu=6.4$ 时，数据的完整性分别是 50% 的概率为 50%，70% 的概率为 80%，以及 90% 的概率为 80% 时的信誉值变化情况。

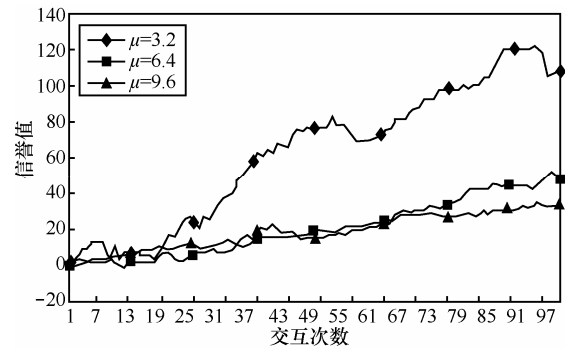


图 4 μ 的取值对信誉值的影响

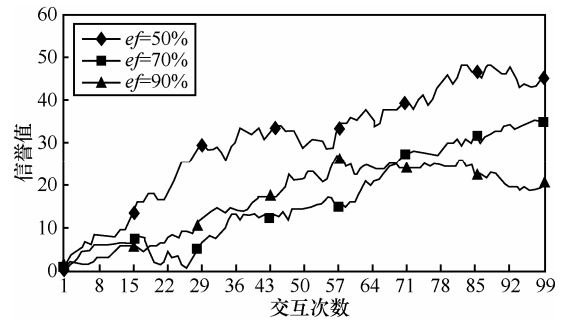


图 5 数据完整性对信誉值的影响

由图 4 可以看出不同取值对信誉值有较大影响，且随着值的变化相对影响在减小。由图 5 可以看出，完整性对信誉值的影响不大，特别是随着完整性比例的增加，影响逐渐减小。

3) 时间贴现因子对信誉值的影响

在其他参数不变的情况下，时间贴现因子 ρ 取值为 0.9、0.6 及 0.3 时信誉值的变化情况如图 6 所示。可以看出具有较好的区分度。

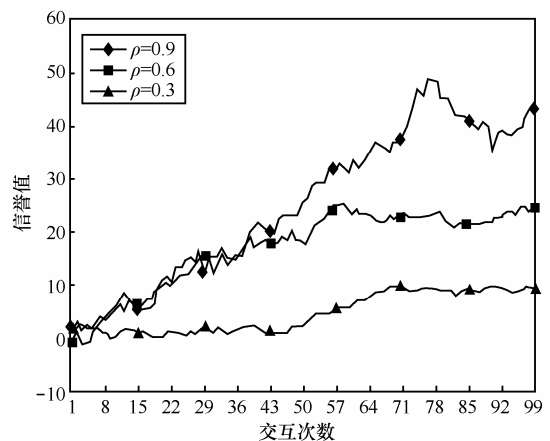
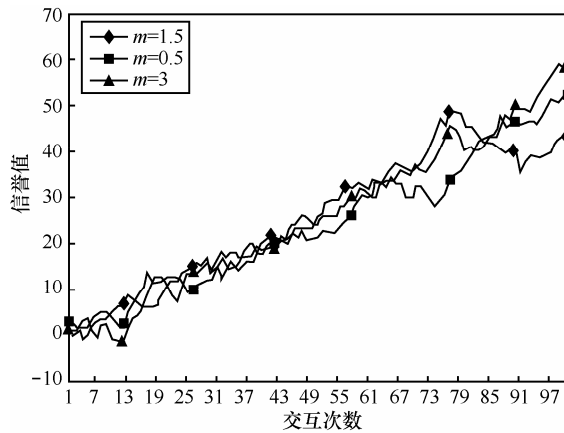


图 6 时间贴现因子对信誉值的影响

4) 衰减函数 $A(t_k)$ 中 m 的变化对信誉值的影响

图 7 给出了 m 在不同取值时的信誉值变化情况，由图可以看出 m 对信誉值的影响不大，原因在于测试中数据时间间隔比较小，从而使 m 的影响因素降低。

图 7 m 的取值对信誉值的影响

6 结束语

传统网络搜索(如百度、Google 等)的数据主要来源于公开、共享、免费的 Web 页面,并且数据质量受环境及其他因素影响很小。而在物联网搜索系统中,因安全、隐私、成本等原因,数据的拥有者往往不愿无偿分享相关数据,并且数据质量与智能节点、环境及拥有者的主观态度密切相关。这就要求物联网搜索中心综合平衡数据需求(如质量、数量、完整性等要素)与成本支出,以优化数据获取机制。基于这种需求,本文将信誉评估机制引入到数据交互过程中,提出了融合主客观要素的动态信誉计算机制,并基于信号传递模型分析了约束信誉形成的 2 个关键要素。通过数据测试证明了所提信誉机制具有较好的区分度,能够帮助搜索中心选择合适的数据提供者。

在本文的信誉计算模型中,默认数据提供者与物联网搜索中心的数据交互是多次重复的。然而,也存在着一些交互次数比较少的情况,下一步的工作中将借鉴协同过滤及节点间合作的思想来扩展信誉计算模型,使其能更好地适应物联网搜索的需求。

参考文献:

- [1] 方滨兴,刘克,吴曼青,等.大搜索技术白皮书[M].国家自然科学基金委,2015.
FANG B X, LIU K, WU M Q, et al. Big Search[M]. National Natural Science Foundation of China, 2015.
- [2] OSTERMAIER B, R'OMERY K, MATTERN F, et al. A real-time search engine for the Web of Things[A]. Internet of Things[C]. 2010. 1-8.
- [3] YUANYI CHEN, JINGYU ZHOU, MINYI GUO. A context-aware search system for Internet of Things based on hierarchical context model[J]. Telecommunication Systems, 2015(3): 1-15.

- [4] MILIARAKI I, BLANCO R, LALMAS M. From "selena Gomez" to "marlon brando": understanding explorative entity search[A]. 24th International World Wide Web Conference[C]. 2015. 765-775.
- [5] SHEN, H Y, LIU J W, CHEN K, et al. SCPS: a Social-aware distributed cyber-physical human-centric search engine[A]. IEEE Transactions on Computers, 2015, 64(2):518-532.
- [6] EBRAHIMI, MOHAMMAD; SHAFIEIBAVANI, at al. A new meta-heuristic approach for efficient search in the Internet of Things[A]. 2015 IEEE International Conference on Services Computing (SCC)[C]. 2015.264-270.
- [7] SU X, ZHANG M J, MU Y, et al. A robust trust model for service-oriented systems[J]. Journal of Computer and System Sciences, 2013, 79: 596-608.
- [8] ZHANG H I, SHOU Z Y, ZHANG J W, et al. Building reputation for service-oriented am I: modeling[J]. Algorithms, and Analysis Transactions of Electrical Engineering, 2014, 38(E1): 99-109.
- [9] ADALI S, ESCRIVA R, GOLDBERG M K. Measuring behavioral trust in social networks[A]. Proceedings of the IEEE International Conference on Intelligence and Security Informatics, 2010. 150-152.
- [10] NEPAL S, SHERCHAN W, PARIS C. STTrust: a trust model for social networks[A]. 2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2011. 841-846.
- [11] YAN S R, ZHENG X L, WANG Y, et al. A graph-based comprehensive reputation model: exploiting the social context of opinions to enhance trust in social commerce[J]. Information Sciences, 2015, 3(18): 51-72.
- [12] WANG Y J, YING G S, CAI Z P, et al. A trust-based probabilistic recommendation model for social networks[J]. Journal of Network and Computer Applications, 2015 (55):59-67.
- [13] JIANG W J, XU Y S, HONG G, et al. Dynamic trust calculation model and credit management mech-anism of online trading[J]. Sci Sin Inform, 2014,44(9):1084-1101.
- [14] MENG X J, MA J F, LU D, et al. Comprehensive trust evaluation model in social networks[J]. Journal on Communications, 2014, 35(12): 136-143.
- [15] GUO Y C, YIN L H, LIU L C, et al. Utility-based cooperative decision in cooperative authentication[A]. IEEE Conference on Computer Communications (INFOCOM)[C]. 2014.1006-1014.

作者简介:



张会兵(1976-),男,河南南阳人,博士,桂林电子科技大学讲师,主要研究方向为物联网、信任评估。

李超[通信作者](1981-),男,重庆人,博士,中国科学院信息工程研究所助理研究员,主要研究方向为物联网搜索技术、信息流安全、无线传感网络安全。E-mail:lichao@iie.ac.cn。

胡晓丽(1978-),女,江苏宿迁人,桂林电子科技大学讲师,主要研究方向为物联网安全。

周娅(1966-),女,湖北荆州人,桂林电子科技大学教授,主要研究方向为大数据。