

参与式感知系统中用户累积行为信誉计算^①



张晓滨, 张嘉诚

(西安工程大学 计算机科学学院, 西安 710048)

通讯作者: 张嘉诚, E-mail: 834550537@qq.com

摘要: 参与式感知系统中, 由于感知数据质量可能受参与者影响, 提出了基于用户累积行为的信誉计算模型以帮助选择可信赖用户. 针对感知环境中用户群体的广泛性及核心用户的不确定性, 该模型采用 OPTICS 聚类算法定义用户场景并划分行为数据集, 建立用户累积行为信誉计算模型, 同时引入时间戳标记信息抛弃部分旧行为以更新用户信誉. 实验表明, 该信誉模型能够结合新旧行为较好地计算并调整用户信誉, 在感知环境用户信誉评价中具有良好的应用前景.

关键词: 参与式感知; 场景定义; 累积行为; 用户可信度; 信誉计算模型

引用格式: 张晓滨, 张嘉诚. 参与式感知系统中用户累积行为信誉计算. 计算机系统应用, 2021, 30(2): 154-159. <http://www.c-s-a.org.cn/1003-3254/7698.html>

Reputation Calculation Model Based on Cumulative Behavior of Users in Participatory Sensing System

ZHANG Xiao-Bin, ZHANG Jia-Cheng

(School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: In a participatory sensing system, since the quality of the perceived data may be affected by the participants, a reputation calculation model based on the cumulative behavior of users is proposed to help select the trustworthy users. According to the extensiveness of the user groups and the uncertainty of the core users in the perceived environment, this model uses the OPTICS clustering algorithm to define the user scenarios and divide the behavioral data set. Furthermore, it introduces time stamps to label information and discard some old behaviors, thus updating the user reputation. The experimental results show that the proposed reputation model can combine old and new behaviors to calculate and adjust the user reputation well, displaying a good application prospect with respect to the evaluation of user reputation in the perceived environment.

Key words: participatory sensing; scenario definition; cumulative behavior; user credibility; reputation calculation model

共享经济和移动技术的广泛普及推动了参与式感知的发展, 大量基于智能手机的参与式感知系统被人们重视并应用到现实生活中^[1], 如任务分配^[2]、数据转发^[3]、在线社区^[4]、昆虫监视^[5]、空气污染监测^[6]等. 由于感知数据是由普通参与者自愿收集并共享, 因而

错误操作或恶意贡献者很大程度上会降低数据质量. 如何提高感知数据质量及数据使用率越来越引起学者重视和研究.

在感知数据质量方面, 部分学者主要针对低质量数据问题展开研究, 文献 [7] 建议使用信誉管理对所收

① 基金项目: 西安工程大学研究生创新基金 (chx2019051); 陕西省自然科学基金 (2015JQ5157)

Foundation item: Graduate Student Innovation Program of Xi'an Polytechnic University (chx2019051); Natural Science Foundation of Shaanxi Province (2015JQ5157)

收稿时间: 2020-04-25; 修改时间: 2020-05-21, 2020-05-26; 采用时间: 2020-06-01; csa 在线出版时间: 2021-01-27

集的数据进行分类,以向数据分析员提供有用信息;文献[8]在处理低质量数据的同时考虑私人数据的机密性,提出了线性盲回归建模方式,该模型在不了解本地私有数据的情况下既可以收集参与者感知数据又可以有效地建立全局线性回归模型的最佳估计,以帮助检测感知数据;文献[9]提出基于累积声誉模型提高数据质量,该模型通过对参与者历史感知数据质量及参与频率的积累,识别并减少不良影响以获得较准确的感应结果;文献[10,11]等基于低质量数据研究进行了深入探索,以便能够很好地区分高质量与低质量数据.但上述研究并没有考虑将数据质量与影响数据质量的参与者关联,以减少恶意贡献者或数据篡改者对数据真实性的影响.部分学者针对可能影响数据质量的参与者展开研究,文献[12]以参与者行为为研究起点,从时间、空间两个方面分析感知用户的行为及感知活动空间,但未分析行为因素与参与者自身或感知数据质量之间的联系;文献[13]通过考虑参与者的时空可用性解决参与者选择问题,以选择有效的参与者继而提高数据质量;文献[14]主要针对影响参与者贡献度的激励机制和隐私问题提出了一种通用的隐私感知激励方案,以减少用户对隐私泄露的担忧,提高用户主动参与性.上述研究虽对数据质量或影响数据质量的参与者因素展开了研究,但在进行参与者研究时,也可以考虑通过对用户可信度的分析提高对信赖参与者的选择,从而提高数据质量.

为提高感知用户可信度分析的真实性与可靠性,本研究提出了基于用户累积行为的信誉计算模型.首先,该模型在考虑实现感知环境累积行为信誉计算时参与者数量庞大、分布广泛的前提下,采用 OPTICS 算法在降低对输入参数敏感度的同时对行为数据集进行聚类处理,实现场景定义;其次,考虑到实际应用中感知用户行为的多样性,为避免用户单一行为的伪装对信誉判断的错误影响,提出融合多行为的累积行为信誉计算模型;最后,考虑用户行为对信誉判断的影响度可能随时间增加而降低,通过引入时间戳标记抛弃部分旧行为实现信誉更新.

1 场景定义

在实际感知环境中,由于参与用户数量庞大、分布广泛.因此,在进行信誉计算时如何高效、精确地选取所需用户成为感知信誉研究中需要解决的首要问题.

通常,用户分类或场景定义通过发现用户之间的潜在联系解决分布广泛的问题,帮助选择具备某些特征或符合条件的用户.学者们针对场景分类问题展开了一系列研究并提出多种可用方法^[15-20],但针对感知环境中的用户复杂性、行为多样性以及核心用户不确定性等特殊问题,本文使用 OPTICS 聚类算法对用户进行场景分类,该算法可以降低数据集对输入参数的敏感度,帮助获取不同邻域半径的簇集,有助于选择合适的用户及行为集进行累积信誉计算.

OPTICS 算法在对行为数据集进行聚类时,首先遍历样本集 X ,寻找所有符合其 ϵ -邻域内样本点数大于邻域最小点数 $MinPts$ 的样本点,并将这些样本点加入核心对象集合.其次,随意选择一个未处理的核心对象 x_i ,通过对 x_i 与其 ϵ -邻域内未访问的样本点 (x_j) 间的可达距离排序得到有序列表决策图.最后,通过决策图可以得到 ϵ 取不同值时的行为数据聚类结果.(可达距离是指 x_i 的核心距离和 x_i 与 x_j 欧几里得距离间的最大值,核心距离是指使 x_i 成为核心对象的最小距离).算法伪代码如下如算法 1.

算法 1. OPTICS 算法

输入: 样本集 X , 邻域半径 ϵ , ϵ -邻域最小点数 $MinPts$, 种子集合 $Seeds$

输出: 有序样本点

1. 初始化核心对象集合 $\Omega=0$;
2. 遍历 X 的元素,如果是核心对象,则将其加入到核心对象集合 Ω 中;
3. while $X \neq 0$ do
4. for Ω 中所有核心对象,随机选择一个未处理的核心对象 x_i 并标记为已处理,同时加入有序列表 p ;
5. 计算 x_i 与 ϵ -邻域中所有 x_j 的可达距离,并将 x_j 插入到根据可达距离排序的种子集合 $Seeds$ 中;
6. if $Seeds=0$
 执行步骤 4;
 else 选择 $Seeds$ 中可达距离最近的点 $seed$, 标记为已处理同时将其压入有序列表 p ;
7. if $seed$ 不是核心对象
 执行步骤 6;
 else 将 $seed$ 的所有邻居点加入到种子集合,重新计算可达距离,执行步骤 6;
8. end
9. end

2 累积行为信誉模型

2.1 累积行为信誉计算

采用 OPTICS 算法得到不同场景下的行为数据集,定

义上述获得的场景集 $Beha = \{b_{i1}, b_{i2}, \dots, b_{ik}\} (i = 1, 2, \dots, k$ 表示共有 k 个场景), 每一场景下的行为集 $UserBeha = \{b_{11}^1, \dots, b_{ix}^m\} (m = 1, 2, \dots, m$ 表示第 m 个行为) 中可以得到 x 用户的 m 个行为. 对比同一场景中同一行为下不同用户的执行结果, 通过每一个结果的出现率和偏差率给出对该行为真实性与可信度判断的初步分值, 并存入行为分值集 $S = \{s_{11}^1, \dots, s_{ix}^m\}$ 中, 使用稳健平均值迭代算法^[10] 计算每一行为分值在信誉计算中所占权重 w_m , 所有行为的权重集可以表示为 $W = \{w_1, w_2, \dots, w_m\}$. 具体计算方法如下:

$$\bar{w} = \sum_{i=1}^m w_i s_i \quad (1)$$

$$v_i = \frac{s_i - \bar{w}}{\sum_{i=1}^m (s_i - \bar{w})^2}, \quad i = 1, 2, \dots, m \quad (2)$$

$$w_i = \frac{\frac{1}{v_i + \varepsilon}}{\sum_{i=1}^m \frac{1}{v_i + \varepsilon}}, \quad i = 1, 2, \dots, m \quad (3)$$

其中, w_i 初始值为 $\frac{1}{m}$, ε 是一个用来改进算法的较小数值^[21], 稳健平均值求权重算法如算法 2.

算法 2. 稳健平均值求权重

输入: 行为集 $UserBeha = \{b_{11}^1, \dots, b_{ix}^m\}$, 行为分值 $S = \{s_{11}^1, \dots, s_{ix}^m\}$
 输出: 行为权重 $W = \{w_1, w_2, \dots, w_m\}$

1. for $i = 1$ to m
2. 初始化 $w_i = \frac{1}{m}$;
3. while $w_i^m - w_i^{m-1}$ 收敛 do
4. 式 (1)~式 (3) 计算 \bar{w}, w_i ;
5. end
6. end

通过算法 2 得到每一行为分值在信誉计算时所占的权重, 再使用式 (4) 计算用户信誉.

$$Rep = \sum_{i=1}^m w_i s_i \quad (4)$$

2.2 信誉的更新

用户行为的时效性主要体现在其在信誉计算中的影响度可能随时间的增加而降低, 因而本文在进行累积行为计算时引入时间戳标记信息, 减少旧行为对信誉的影响并及时更新用户信誉.

研究中将时间轴划分为若干个长度为 t_0 的时间窗口, 定义每一行为的初始时间为 t_{ini} , 计算信誉的当前时间为 t_{end} , $|t_{end} - t_{ini}|$ 时间段内存在 j 个时间窗口且用户在这一时间段内行为 $i (i = 1, 2, \dots, m)$ 的发生频数为 k , 则基于时间因素更新用户信誉的计算函数如下:

$$Rep_{new} = \frac{\sum_{g=1}^j \left(e^{\frac{1}{\lambda g}} \times Rep_g \right)}{\sum_{i=1}^m k_i}, \quad j = \frac{|t_{end} - t_{ini}|}{t_0} \quad (5)$$

其中, Rep_g 表示第 g 个时间窗口的用户信誉. $e^{\frac{1}{\lambda g}}$ 为时间衰减函数, λ 为可主观设定的时间调节系数^[21], g 越大距离当前信誉计算时间间隔越长, 时间衰减函数值越小, 则该时间窗口的信誉值在用户信誉更新时所占比重越小.

通过对累积行为信誉计算模型中场景定义、信誉计算、信誉更新的描述, 得到累积行为信誉算法如算法 3.

算法 3. 累积行为信誉算法

输入: 样本集 X , 邻域半径 ε, ε -邻域最小点数 $MinPts$, 种子集合 $Seeds$, 行为集 $UserBeha = \{b_{11}^1, \dots, b_{ix}^m\}$, 行为分值 $S = \{s_{11}^1, \dots, s_{ix}^m\}$
 输出: 用户信誉值 $Rep-n = \{Rep_{new1}, Rep_{new2}, \dots, Rep_{newn}\}$

1. 执行算法 1, 场景定义;
2. if 新行为 $g \in UserBeha$
 执行步骤 1;
3. else 执行算法 2;
4. if 经历 t_0 时长
5. 式 (5) 更新用户信誉;
6. end

3 实验分析

实验采用 SNAP 公布的 Bitcoin Alpha trust weighted signed network 数据集^[22,23]. 该数据集信息来源 Alpha 平台下 5 种不同网络环境下比特币交易的用户信誉信息, 包含 3783 个用户节点, 24 186 条用户交互边, 每条边的权值从 -10 到 10 不等, 其中积极数据集占总数据集的 93%. 本文选取任意一个网络下的用户交互信息进行实验, 实验先对这些数据进行统一处理, 形成包含源用户 ID、目标用户 ID、源用户对目标用户的行为分值和用户初始信誉 4 个字段的数据集.

为验证累积行为及时间戳概念能够较准确的计算并更新信誉值, 首先使用 OPTICS 算法对随机选取数据集中的 500 条行为数据进行聚类处理, 得到有序列

表决策图和可视化分类图,结果如图1所示。

图1(a)图为有序列表决策图,图1(b)为可视化聚类结果图。从决策图可以看出实验数据集被分为5个类别(决策图的每一个凹槽代表1个类别)。从这5个类别的行为数据集中随机选取a、b、c三组进行以下两组对比实验。

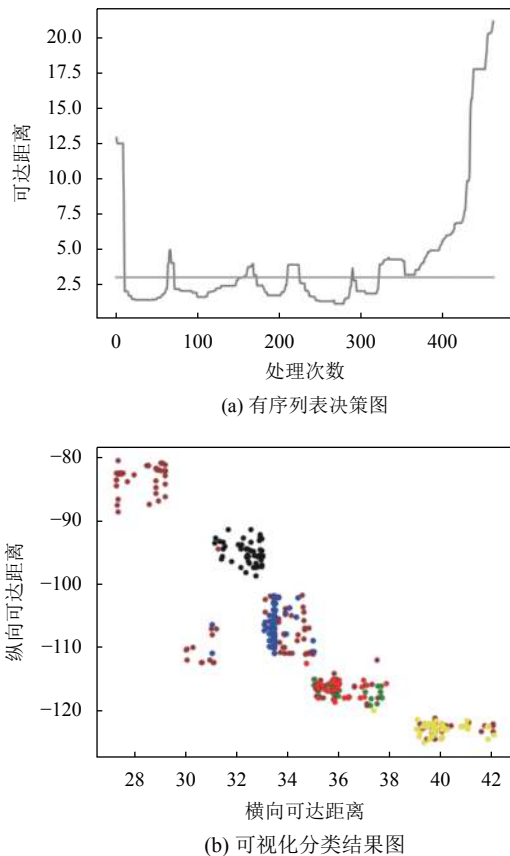


图1 场景分类结果图

(1) 对比单一行为、累积行为计算的用户信誉及用户的真实信誉。

从图2信誉计算对比图可以看出,使用累积行为计算用户信誉更贴近真实信誉。从图2(a)可以看出,a组数据中仅有7.36%的单一行为信誉比累积行为信誉更贴近真实信誉,相反,有92.64%的累积行为信誉更贴近真实信誉;从图2(b)可以看出,b组数据中有10.27%的单一行为信誉比累积行为信誉更贴近真实信誉,3.63%的单一行为信誉与累积行为信誉相等,但有86.1%的累积行为信誉更贴近真实信誉;从图2(c)可以看出,c组数据中仅有2.46%的单一行为信誉比累积行为信誉更贴近真实值,而有97.54%的累积行为信誉更贴近真实值。

(2) 对比未引入时间戳的用户信誉及引入时间戳的信誉更新值。

从图3信誉更新对比图可以看出,引入时间戳概念对用户信誉进行更新,能够减少旧行为的不恰当影响,得到较为准确的信誉值。从图3(a)可以看出,在a组数据中仅有2.46%的更新信誉值与未更新信誉值相等,而有62.38%的更新信誉值更贴近真实信誉值且有36.69%的更新信誉值与真实信誉值相等;从图3(b)可以看出,在b组数据中仅有2.46%的未更新信誉值更贴近真实信誉,但65.36%的更新信誉值更贴近真实信誉且有33.78%的更新信誉值与真实信誉相等;从图3(c)可以看出,在c组数据中所有更新信誉值的效果均优于未更新信誉值,且有30.27%的更新信誉值与真实信誉相等。

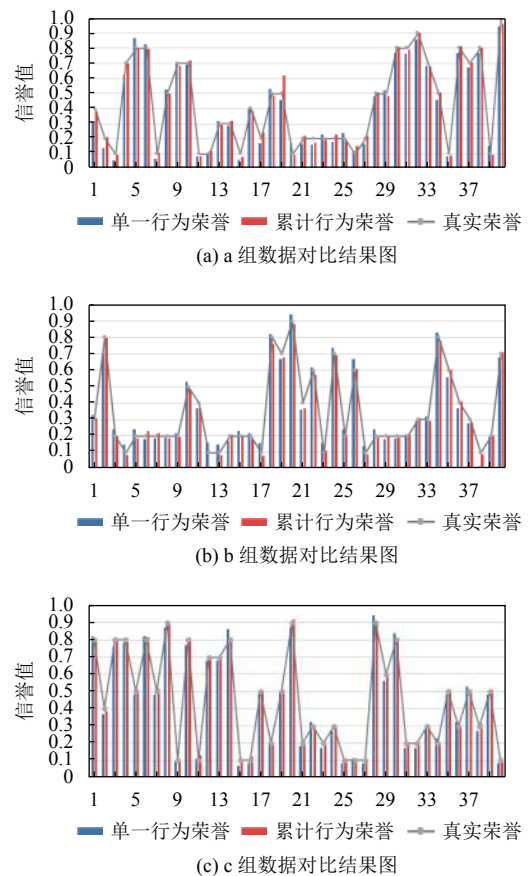


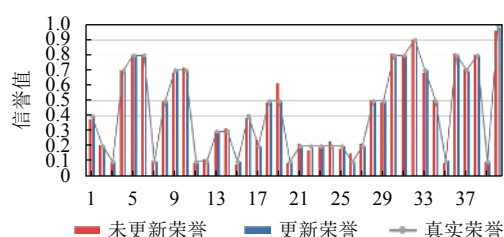
图2 信誉计算对比图

4 总结

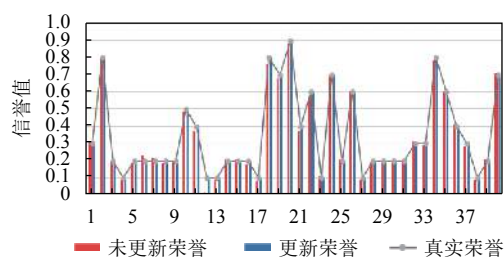
本文针对感知用户信誉计算过程中行为多样性的影响和信誉更新的问题,提出基于累积行为的信誉计

算模型. 该模型结合 OPTICS 算法对用户行为场景定义, 基于用户分类结果再使用稳健平均值迭代算法调整每一行在累积信誉计算中的比重, 最后引入时间戳标记剔除较久发生的行为, 实现用户信誉更新.

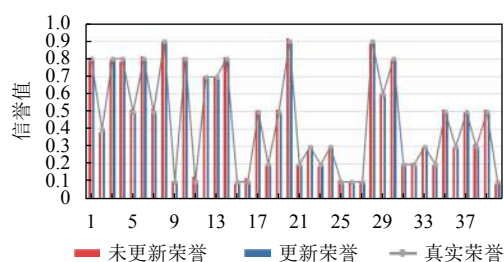
该信誉计算模型能够考虑到用户行为的多样性和时效性, 对用户信誉有一个较为准确的判断, 使感知用户的信誉计算更加具有客观性, 符合实际生活规律, 有助于感知环境中的参与者选择. 后期可考虑根据所更新的用户信誉进行可信赖社区探索.



(a) a组数据对比结果图



(b) b组数据对比结果图



(c) c组数据对比结果图

图3 信誉更新对比图

参考文献

- Guo B, Chen C, Zhang DQ, *et al.* Mobile crowd sensing and computing: When participatory sensing meets participatory social media. *IEEE Communications Magazine*, 2016, 54(2): 131–137. [doi: 10.1109/MCOM.2016.7402272]
- He SB, Shin DH, Zhang JS, *et al.* Toward optimal allocation of location dependent tasks in crowdsensing. *Proceedings of 2014 IEEE Conference on Computer Communications*. Toronto, ON, Canada. 2014. 745–753.
- Zhou H, Chen JM, Zhao HY, *et al.* On exploiting contact patterns for data forwarding in duty-cycle opportunistic mobile networks. *IEEE Transactions on Vehicular Technology*, 2013, 62(9): 4629–4642. [doi: 10.1109/TVT.2013.2267236]
- Guo B, Wang Z, Yu ZW, *et al.* Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm. *ACM Computing Surveys*, 2015, 48(1): 7.
- Malek R, Tattoni C, Ciolli M, *et al.* Coupling traditional monitoring and citizen science to disentangle the invasion of *Halyomorpha halys*. *ISPRS International Journal of Geo-Information*, 2018, 7(5): 171. [doi: 10.3390/ijgi7050171]
- Méndez D, Pérez AJ, Labrador MA, *et al.* P-sense: A participatory sensing system for air pollution monitoring and control. *IEEE International Conference on Pervasive Computing and Communications Workshops*. Seattle, WA, USA. 2011. 344–347.
- Yang HF, Zhang JL, Roe P. Using reputation management in participatory sensing for data classification. *Procedia Computer Science*, 2011, 5: 190–197. [doi: 10.1016/j.procs.2011.07.026]
- Chang S, Zhu HZ, Zhang W, *et al.* Pure: Blind regression modeling for low quality data with participatory sensing. *IEEE Transactions on Parallel and Distributed Systems*, 2016, 27(4): 1199–1211. [doi: 10.1109/TPDS.2015.2427805]
- Yu RY, Liu R, Wang XW, *et al.* Improving data quality with an accumulated reputation model in participatory sensing systems. *Sensors*, 2014, 14(3): 5573–5594. [doi: 10.3390/s140305573]
- Chou CT, Ignjatovic A, Hu W. Efficient computation of robust average of compressive sensing data in wireless sensor networks in the presence of sensor faults. *IEEE Transactions on Parallel and Distributed Systems*, 2013, 24(8): 1525–1534. [doi: 10.1109/TPDS.2012.260]
- Kounadi O, Resch B. A geoprivacy by design guideline for research campaigns that use participatory sensing data. *Journal of Empirical Research on Human Research Ethics*, 2018, 13(3): 203–222. [doi: 10.1177/1556264618759877]
- Khoi NM, Casteleyn S. Analyzing spatial and temporal user behavior in participatory sensing. *ISPRS International Journal of Geo-Information*, 2018, 7(9): 344. [doi: 10.3390/ijgi7090344]
- Abdel Hamid S, Takahara G, Hassanein HS. On the recruitment of smart vehicles for urban sensing. *Proceedings of 2013 IEEE Global Communications Conference*. Atlanta, GA, USA. 2013. 36–41.

- 14 Li QH, Cao GH. Providing privacy-aware incentives in mobile sensing systems. *IEEE Transactions on Mobile Computing*, 2016, 15(6): 1485–1498. [doi: [10.1109/TMC.2015.2465375](https://doi.org/10.1109/TMC.2015.2465375)]
- 15 叶舒雁, 张未展, 齐天亮, 等. 一种基于传感器与用户行为数据分析的移动学习场景感知分类方法. *计算机研究与发展*, 2016, 52(12): 2721–2728. [doi: [10.7544/issn1000-1239.2016.20160633](https://doi.org/10.7544/issn1000-1239.2016.20160633)]
- 16 林一, 刘越, 王涌天, 等. 一种增强现实分场景推送情景感知服务的方法. *软件学报*, 2016, 27(8): 2115–2134. [doi: [10.13328/j.cnki.jos.004846](https://doi.org/10.13328/j.cnki.jos.004846)]
- 17 阮锦佳, 罗丹, 罗海勇. 基于支持向量机多分类器的室内外场景感知算法. *计算机应用*, 2015, 35(11): 3135–3138, 3145. [doi: [10.11772/j.issn.1001-9081.2015.11.3135](https://doi.org/10.11772/j.issn.1001-9081.2015.11.3135)]
- 18 Du L, Ling HB. Dynamic scene classification using redundant spatial scenelets. *IEEE Transactions on Cybernetics*, 2016, 46(9): 2156–2165. [doi: [10.1109/TCYB.2015.2466692](https://doi.org/10.1109/TCYB.2015.2466692)]
- 19 Liu WX, Lau RWH, Manocha D. Robust individual and holistic features for crowd scene classification. *Pattern Recognition*, 2016, 58: 110–120. [doi: [10.1016/j.patcog.2016.03.031](https://doi.org/10.1016/j.patcog.2016.03.031)]
- 20 Zhang XY, Du SH, Zhang Y. Semantic and spatial co-occurrence analysis on object pairs for urban scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, 11(8): 2630–2643. [doi: [10.1109/JSTARS.2018.2854159](https://doi.org/10.1109/JSTARS.2018.2854159)]
- 21 赵治国, 谭邦, 夏石莹, 等. 基于时间序列的P2P网络综合信任模型研究. *计算机工程与应用*, 2017, 53(15): 127–131, 265.
- 22 Kumar S, Spezzano F, Subrahmanian VS, *et al.* Edge weight prediction in weighted signed networks. *Proceedings of the 16th IEEE International Conference on Data Mining (ICDM)*. Barcelona, Spain. 2016. 221–230.
- 23 Kumar S, Hooi B, Makhija D, *et al.* REV2: Fraudulent user prediction in rating platforms. *Proceedings of the 11th ACM International Conference on Web Search and Data Mining (WSDM)*. Marina Del Rey, CA, USA. 2018. 333–341.