

基于拟态计算的社会网络划分算法

成平广

(重庆第二师范学院数学与信息工程系 重庆 400067)

摘要 信息社会中, 社会网络结构的形成与演变是一个动态过程, 拟态计算(Mimic Computing, MC)是根据资源、任务、安全、服务和时效性等不同约束条件, 动态适应用户不同的应用需求, 改进计算模式的一种计算架构, 是当前适应动态多变网络环境的一种有效方法。基于拟态计算, 应用当前研究前沿的传染病 SIR 模型, 提出了基于拟态计算的社会网络划分算法(Mimic Community Clustering, MCC)。应用该方法进行网络社区划分, 并用真实数据来验证模型的可行性和有效性。

关键词 拟态计算, 社区划分

中图法分类号 TP301.6

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2015.8.029

Mimic Community Clustering of Society Networks

CHENG Ping-guang

(Department of Mathematics and Information Engineering, Chongqing University of Education, Chongqing 400067, China)

Abstract The formation and evolution of social network are a dynamic process. Mimic computing can dynamically meet the users' needs based on resources, tasks, safety, effectiveness and service, so it is a useful method to adapt to the dynamic social network. The paper proposed a mimic community clustering(MCC) based on traditional SIR model and mimic computing. The results of experiment performed on real data show that the method has high value.

Keywords Mimic computing, Community clustering

1 研究背景

拟态计算是以实现高效能计算为目的, 在任务处理的全过程中通过感知自变量动态地选择或生成应用问题的最佳计算结构集合。计算结构的函数化是拟态计算的本质。拟态计算与虚拟计算的区别在于, 拟态计算是物理和逻辑结构的同时变换, 而虚拟计算仅仅是逻辑结构的变换。

拟态计算机是一种根据资源、任务、安全、服务和时效性等的不同, 变换成尽可能合适的结构形态, 从而得到高的处理效能, 并且还可以基于认知而改变结构的计算机。

拟态计算以多维重构函数化体系结构与动态多变量运行机制为核心, 对高性能计算在多个典型领域的应用、结构和效能关系进行了深入分析, 引入了“应用决定结构, 结构决定效能”的理念。邬江兴^[1]提出了基于多维重构函数化结构与动态多变量运行机制的拟态计算体系——拟态计算。相比于传统信息系统的静态性、相似性和确定性, 拟态计算系统具有非持续性、非相似性和非确定性等基本属性; 而拟态计算机的结构动态可变, 软硬件结合计算, 针对用户不同的应用需求, 拟态计算机可通过改变自身结构提高效能。

随着信息和网络技术的发展, 世界上社会网络的数量正在以前所未有的速度增长。社会网络为人们提供了一种新的交流方式、工作方式、购物方式, 甚至全新的生活方式, 这种新

的社会组织形式越来越显著地影响着传统社区中的每一个人, 同时社会网络是动态多变的, 拟态计算应用于社会网络是应对大数据的社会网络划分研究的方向^[2]。社会网络即虚拟社区, 又被称为虚拟社群、网络社群或电子社群等, 是互联网使用者经过互动后产生的一种社会群体。1993 年社会网络研究先驱莱茵戈德(Rheingold)率先提出“虚拟社区”的概念: 一群主要藉由计算机网络彼此沟通的人们, 彼此有某种程度的认识, 分享某种程度的知识和信息, 在很大程度上如同对待朋友般彼此关怀, 从而形成的团体^[3]。

信息社会网络结构的形成与演变是一个动态过程, 基于动态模型的社区发现方法主要是通过对该过程的动态特征建模来实现社区结构的发现, 目前主要有 3 类, 即随机漫步模型、旋转模型、同步机制。随机漫步模型算法设计的基本原则是: 若网络具有很强的社区结构性, 则随机漫步者多数时间是停留在社区内部边上。旋转模型是通过网络节点与 Potts 自旋系统的旋转体对应将信息网络中的一个社区映射到一个处于基态或哈密顿极小态的磁畴。同步现象广泛存在于社会体系和自然界中, 若将网络节点一一映射到不同初相的振荡器中, 则可利用同步机制的原理进行社区发现与探索的研究。文献[4]在 Kuramoto 模型的基础上给出 t 时刻的节点相关度矩阵 $\rho_{ij}(t) = \cos(\theta_i(t) - \theta_j(t))$, 对此矩阵二值化得到动态关联矩阵, 对矩阵进行谱分析可得到社区数。

到稿日期: 2014-10-15 返修日期: 2015-01-12 本文受重庆市教委自然科学基金项目(KJ1401405), 重庆市教委人文社科项目(12sksz65), 重庆第二师范学院教学改革重点项目(JG2014106)资助。

成平广(1975-), 男, 硕士, 副教授, 主要研究方向为计算机应用、信息安全, E-mail: cadou@163.com。

2 拟态计算的社会网络动态划分算法

当下,已经进入“自媒体传播”的“全民传播时代”。20世纪美国著名传播学者李普曼提出,大众媒介既反映现实,也建构现实,传播内容是一种营造“拟态环境”的活动,通过对信息或事件进行筛选、加工和重构后向人们展示的环境,不仅制约着读者的认知与行为,而且还对客观的现实环境产生影响^[5,6]。基于以上研究,本文提出了基于拟态计算的社区动态划分算法。

本文利用经典的传染病动力学模型来进行社区动态描述;针对信息流动过程的特殊性,由于社会网络中每个节点只能随机接触其拓扑相邻的节点,接触的节点数量由相邻的节点数量决定,因此信息仅能从传播者集合向它们相邻的节点传播。用 $\Theta(v, t)$ 表示 t 时刻社会网络中节点 v 与相邻节点的接触在 $S(t)$ 中的比例, $I_\theta(t)$ 表示已将信息传播了 θ 个时间片的节点所构成的集合。则动态社区变化可以表示为

$$I(t+1) = I(t) + S(t) \sum_{v \in \bigcup_{\theta=0}^t I_\theta} \Theta(v, t) \cdot \beta_{v,t}$$

其中, $\beta_{v,t}$ 表示 t 时刻节点 v 与相邻节点的接触率。

节点 v 与相邻节点的接触率 $\beta_{v,t}$ 可以表示为

$$\beta_{v,t} = \frac{1}{S(t)\Theta(v, t) \cdot \Delta t_{v,t}}$$

在实际应用时, $\Theta(v, t)$ 可根据具体问题进行不同定义。本文在后续的介绍和应用中,将 $\Theta(v, t)$ 定义为 t 时刻节点 v 相邻的节点中接收者占有所有接收者的比例,因此, $S(t) \cdot \Theta(v, t)$ 为 t 时刻节点 v 相邻节点中的接收者数量。

基于上述模型,应用拟态计算的理论将一个社区的状态用一个三元组矩阵 $\Omega = (X, \sigma, Sp)$ 来进行描述。

$$X = \begin{pmatrix} X(t_1) \\ X(t_2) \\ \vdots \\ X(t_l) \end{pmatrix} = \begin{pmatrix} x_{t_1}^1 & x_{t_1}^2 & \cdots & x_{t_1}^n \\ x_{t_2}^1 & x_{t_2}^2 & \cdots & x_{t_2}^n \\ \vdots & \vdots & \vdots & \vdots \\ x_{t_l}^1 & x_{t_l}^2 & \cdots & x_{t_l}^n \end{pmatrix}$$

σ 表示拟态转换, $\sigma: \Omega(t_i) \rightarrow \Omega(t_{i+1})$ 。

x 为网络中各个社区的状态,设定社会网络 $G(V, E)$ 中 W_{vu} 表示节点 v 到 u 的边的权重, W 表示社会网络中所有边的权重之和, S_v^{out} 为从节点 v 流出的权重值之和, S_u^{in} 为从节点 u 流入的权重值之和, G 的社区划分结果 x 为:

$$x = \{x_1, x_2, \dots, x_i \cap x_j = \emptyset (i \neq j), \bigcup_{x_i \in X} x_j = V\}$$

则社区密度 $D(x)$ 表示为

$$D(x) = \sum_{i=1}^m d(G_i) = \sum_{i=1}^m \frac{W_{vu} - S_v^{out} S_u^{in}}{|W_i|}$$

社区划分的实时度量 $Sp = (Sp_1, Sp_2, \dots, Sp_l)$, Sp_i 为社区动态电导度。电导率^[7]用来测量结点聚类的质量,表示聚类外的边数与聚类内的边数的比率,可作为评价发现的社区好坏的一个标准。形式如下:

$$\phi(Sp) = \frac{c_{Sp}}{\min(\text{Vol}(S_p), \text{Vol}(V \setminus S_p))}$$

式中, $c_{Sp} = |\{(u, v): u \in G_i, v \notin G_i\}|$ 表示聚类外的边数, $\text{Vol}(S_p) = \sum_{u \in S_p} d(u)$ 表示聚类 G_i 中结点的度的总和。电导率越低,结点集越偏向于社区。

3 实验验证

在实验中,为了验证所提算法的高效性,采用了网络下载

的2组真实数据:某论坛关于大众汽车召回事件的讨论情况和某校园论坛关于学生评价教师活动的讨论情况。使用兰德指数(RI)^[8]和调整兰德指数(ARI)^[9]作为所评价指标。采集的数据有发帖数、回帖数、转帖数、评价数等,对事件在话题活跃期间,以小时为周期,计算回帖率、转帖率、发帖质量、回帖质量、帖子热度、周期活动率等指标。聚类的结果如表1所列。

表1 数据集描述

Dataset	Records-number	Clusters-number
车友网	115	4
校园学评教	236	5

使用 Matlab 实现本文描述模型,并与 spin model 模型和 Kuramoto 模型在同一数据集上做对比实验,结果如表2和表3所列。

表2 社区划分的 RI 指标

Dataset	spin model	Kuramoto	SNMF-Kmeans
车友网	0.9123	0.9265	0.9457
校园学评教	0.8536	0.8386	0.8852

表3 社区划分的 ARI 指标

Dataset	spin model	Kuramoto	SNMF-Kmeans
车友网	0.5773	0.8656	0.9234
校园学评教	0.6122	0.6236	0.6534

通过社区划分指标可以看出,用拟态模型进行社区划分的结果比经典的两个动态划分模型要好。但是受采集热点事件数据节点偏少、问题社会影响力有限、用于实验的计算机不适合做大规模数据处理等约束条件的限制,进一步实验需要在后续的工作中来完成。

实验数据经过社区划分模型处理后,可视化的结果如图1和图2所示。

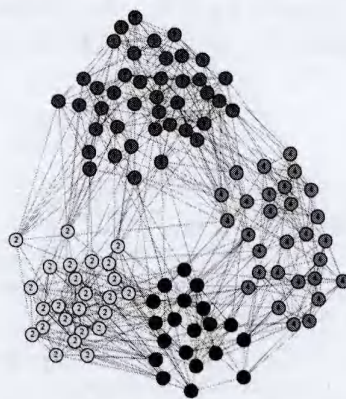


图1 车友网对大众汽车召回事件的评价

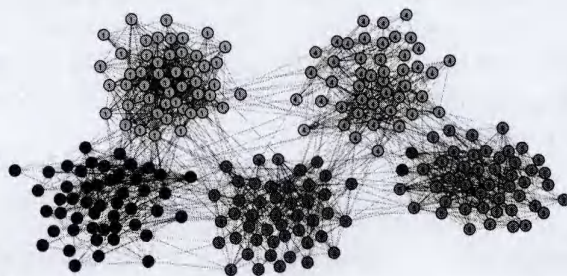


图2 校园网对评教活动的评价

(下转第156页)

- for Humanity(TIC-STH). IEEE, 2009; 147-152
- [3] Alhazmi O H, Malaiya Y K. Evaluating disaster recovery plans using the cloud[C]//Reliability and Maintainability Symposium (RAMS), 2013 Proceedings-Annual. IEEE, 2013; 1-6
- [4] 陈敏刚, 董军, 张丽亮, 等. AHP 和模糊综合评判在灾难恢复能力评估中的应用[J]. 计算机工程, 2006, 32(18): 135-137, 140
Chen Min-gang, Dong Jun, Zhang Li-liang, et al. AHP and Fuzzy Comprehensive Evaluation Application in Disaster Recovery Ability Evaluation [J]. Computer Engineering, 2006, 32(18): 135-137, 140
- [5] 李佳, 徐达, 王晓磊. 基于风险分析的信息系统灾难恢复能力评估[C]//中国通信学会. 虚拟运营与云计算——第十八届全国青年通信学术年会论文集(上册). 2013; 6
Li Jia, Xu Da, Wang Xiao-lei. Information System Disaster Recovery Capability Evaluation Based on Risk Analysis[C]//China Institute of Communications. Virtual Operation and Cloud Computing—The Eighteenth Annual Conference Proceedings National Youth Communication(Volume). 2013; 6
- [6] Zhang G, Yang Y, Mao X. Disaster recovery evaluation PROC model framework based on information flow[C]//2011 International Conference on Computer Science and Network Technology(ICCSNT). IEEE, 2011; 1841-1845
- [7] 秦丽, 李兵. 一种基于云模型的不确定性数据的建模与分类方法[J]. 计算机科学, 2014, 41(8): 233-240
Qin Li, Li Bing. Novel Method of Uncertain Data Modeling and Classification Based on Cloud Model [J]. Computer Science, 2014, 41(8): 233-240
- [8] 吕辉军, 王晔, 李德毅, 等. 逆向云在定性评价中的应用[J]. 计算机学报, 2003, 26(8): 1009-1014
Lv Hui-jun, Wang Ye, Li De-yi, et al. The Application of Backward Cloud in Qualitative Evaluation [J]. Chinese Journal of Computers, 2003, 26(8): 1009-1014
- [9] 刘常昱, 冯芒, 戴晓军, 等. 基于云 X 信息的逆向云新算法[J]. 系统仿真学报, 2004, 16(11): 2417-2420
Liu Chang-yu, Feng Mang, Dai Xiao-jun, et al. A New Algorithm of Backward Cloud [J]. Journal of System Simulation, 2004, 16(11): 2417-2420
- [10] 于少伟, 史忠科. 基于正态分布区间数的逆向云新算法[J]. 系统工程理论与实践, 2011, 31(10): 2021-2026
Yu Shao-wei, Shi Zhong-ke. New Algorithm of Backward Cloud Based on Normal Interval Number [J]. Systems Engineering—Theory & Practice, 2011, 31(10): 2021-2026
- [11] 杨义先, 姚文斌, 陈钊. 信息系统灾备技术综论[J]. 北京邮电大学学报, 2010, 33(2): 1-6
Yang Yi-xian, Yao Wen-bin, Chen Zhao. Review of Disaster Backup and Recovery Technology of Information System [J]. Journal of Beijing University of Posts and Telecommunications, 2010, 33(2): 1-6
- [12] 李军, 李继光, 姚建刚, 等. 属性识别和 G1-熵权法在电能质量评价中的应用[J]. 电网技术, 2009, 33(14): 67-73
Li Jun, Li Ji-guang, Yao Jian-gang, et al. Application of Attribute Recognition and G1-Entropy Method in Evaluation of Power Quality [J]. Power System Technology, 2009, 33(14): 67-73

(上接第 137 页)

从社区划分可视化的结果来看, 社区划分的区分度比较明显, 社区内节点之间的联系明显大于社区之间节点的联系。这说明了本文利用的社区划分评价指标的有效性。

结束语 动态社会网络是社交媒体发展的必然趋势, 本文引入拟态计算的概念来描述动态多变的社会网络, 利用传染病模型来描述信息传播的过程并加入到社区划分的模型中。为验证模型的有效性, 利用下载的真实数据, 与 spin model 和 Kuramoto 模型进行对比实验。实验表明 MCC 模型对动态社区划分有明显的提高。

参 考 文 献

- [1] 郭江兴. 拟态计算与拟态安全防御的原意和愿景[J]. 电信科学, 2014, 30(7): 2-7
Wu Jiang-xing. Meaning and Vision of Mimic Computing and Mimic Security Defense [J]. Telecommunications Science, 2014, 30(7): 2-7
- [2] 崔可嘉. 微博及其构建的拟态环境研究[J]. 新闻知识, 2014(10): 73-74
Cui Ke-jia. Study on micro-blog pseudo environment and its construction [J]. News Research, 2014(10): 73-74
- [3] 贺小光, 兰讽. 网络社区研究综述——从信息交流到知识共享[J]. 情报科学, 2011, 29(8): 1268-1271
He Xiao-guang, Lan Feng. A Review on Network Community—Information Communication to Knowledge Sharing [J]. Information Science, 2011, 29(8): 1268-1271
- [4] Hu Y, Li M, Zhang P, et al. Community detection by signaling on complex networks [J]. Phys Rev E, 2008, 78(1): 016115
- [5] 李姣姣. 微博拟态环境构建分析[J]. 青年记者, 2012(33): 82-83
Li Jiao-jiao. Analysis of the construction of micro-blog mimicry environment [J]. Youth Journalist, 2012(33): 82-83
- [6] 苏娜. 新媒体拟态环境对大学生网络舆情的影响及应对[J]. 江苏高教, 2014(3): 123-125
Su Na. New media pseudo environment influence on College Students' network public opinion and the Countermeasures [J]. Jiangsu Higher Education, 2014(3): 123-125
- [7] Leskovec J, Lang K J, Mahoney M W. Empirical comparison of Algorithms for network community Detection [C]//Proceedings of the International Conference on World Wide Web. Raleigh, NC, USA, 2010: 631-640
- [8] Rand, William M. Objective criteria for the evaluation of clustering methods [J]. Journal of the American Statistical association, 1971, 66(336): 846-850
- [9] Wagner S, Wagner D. Comparing clusterings: an overview [R]. Universität Karlsruhe, Fakultät für Informatik, 2007
- [10] 吉福生, 王燕燕, 徐明玉, 等. 社会化机会网络中节点归属位置感知的路由机制[J]. 重庆邮电大学学报(自然科学版), 2015, 27(3): 404-410
Ji Fu-sheng, Wang Yang-yan, Xu Ming-yu, et al. Node home location aware routing mechanism for social opportunistic networks [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2015, 27(3): 404-410