计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190400072

基于节点演化分阶段优化的事件检测方法

富 坤 仇 倩 赵晓梦 高金辉
河北工业大学人工智能与数据科学学院 天津 300401
河北省大数据计算重点实验室 天津 300401

摘 要 链路预测技术是分析网络演化的有效方法,也为社会网络事件检测提供了一种新思路。当前采用链路预测进行事件检测的方法大多是从宏观的网络演化入手,也有少数结合节点演化的检测方法,但其稳定性不佳,对事件的敏感性也不够高,不能准确检测事件的发生。基于以上问题,提出了一种基于节点演化分阶段优化的事件检测方法(Node Evolution Staged Optimization,NESO_ED)。首先通过分阶段优化的方法加强事件检测的稳定性,并获取节点指标权重数组;然后根据不同阶段按不同规则选取节点的最佳相似性计算指标,使节点能更好地量化网络演化情况,以此提高事件检测的敏感性。此外,分析了网络演化过程中节点选取指标的变化情况,揭示了事件发生对节点演化产生的不同影响。基于真实社会网络 VAST 进行对比实验,结果显示 NESO_ED 方法在事件检测敏感性上比 LinkEvent 方法提高了 227%,比 NodeED 方法提高了 63%,NESO_ED 方法的稳定性也比 NodeED 方法提高了 66%,这表明 NESO_ED 方法能更加准确且稳定地进行事件检测。

关键词:事件检测;节点演化;链路预测;社会网络;动态网络
中图法分类号 TP181

Event Detection Method Based on Node Evolution Staged Optimization

FU Kun,QIU Qian,ZHAO Xiao-meng and GAO Jin-hui
School of Artificial Intelligence,Hebei University of Technology,Tianjin 300401,China
Hebei Province Key Laboratory of Big Data Calculation,Tianjin 300401,China

Abstract Link prediction technology is an effective method to analyze network evolution,it also provides a new idea to detect social events. Now,most of the event detections using link prediction start from the macroscopic overall network evolution. Although there are a few detection methods that combine node evolution,the stability of them are not good,and the sensitivity to the event are not high enough to accurately detect the occurrence of the event. Therefore,an event detection method based on node evolution staged optimization (NESO_ED) was proposed. Firstly,the stability of event detection is enhanced by a staged optimization method,and an array of node index weights is obtained. Then,according to different rules,the optimal similarity calculation index of the node is selected,so that the node can better quantify the network evolution and improve the sensitivity of event detection. In addition,the changes of indicators that selected by nodes in the process of network evolution were also analyzed. It reveals different effects of events on the evolution of nodes. On real social network VAST,the event detection sensibility of NESO_ED is increased by 227% compared with LinkEvent and 63% compared with NodeED. The stability of NESO_ED is also increased by 66% compared with NodeED,which shows that NESO_ED can detect events more accurately and stably.

Keywords Event detection,Node evolution,Link prediction,Social network,Dynamic network

1 引言

随着移动互联网技术的迅速发展与社交网络服务的繁荣,手机已经成为人们传播和获取各种信息的主要手段,使人们的生活和交流变得更加丰富多彩、便捷迅速,也使社会中发生的事件的影响力更大。事件检测是社交网络的重要内容,政府或企业可以通过对社交网络进行检测,来及时掌握网络上的敏感话题、舆情动态等,同时对恶性行为加以有效控制,对网络加以理性引导,从而建立和谐良好的网络社会。

在事件检测研究中,有适用于传统媒体的方法,如基于 LDA 主题模型方法[1]、基于突发词方法[2]等;有融合社交媒体特性的方法,如基于用户影响力方法[3]、结合地域分析方法[4]和情感分析方法[5-6]等;还有通过深度学习结合图片、视频等多模态形式进行事件检测的方法[7]。以上方法大多涉及文本数据,需要进行文本预处理,且过程繁琐、工作量大、效率低。另外,传统媒体的事件检测方法适用于规范书写的新闻

到稿日期:2019-04-11 返修日期:2019-08-13 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金青年科学基金(61806072)
This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China (61806072).
通信作者:富坤(fukun@hebut.edu.cn)

报道文章；社交媒体的事件检测方法适用于稀疏性、动态性、社会性的短文本[8-9]；利用深度学习的方法造价高、周期长。因此，这些方法都不利于快速地对社会网络进行事件检测。

在社会网络中利用网络演化进行事件检测的方法越来越受欢迎，有研究已证明通过链路预测可以对网络演化进行分析，链路预测和网络演化规律具有内在的一致性[10-11]。基于链路预测的事件检测方法，通过网络演化波动性来判断是否有事件发生，该方法主要分为两类：1）规定所有节点遵循相同的演化规律；2）不同节点可以遵循不同的演化规律。第一类方法依据 AUC[12]，Precision[13] 和 Rank Score[14] 3 种链路预测评价指标，选取使评价指标最好的相似性计算指标作为全部节点的计算指标，对宏观网络演化的波动性进行表示。第二类方法通过设计算法让每个节点选取最适合自己的相似性计算指标，从节点的微观角度对网络的波动性进行分析。

在真实社会网络中，节点间的演化规律往往不同，因此采用单一的相似性指标时，事件检测效果不如采用多种相似性指标，后者更符合社会网络的真实情况。在采用多种相似性指标的方法中，针对不同节点选取各自的最佳指标便是事件检测方法的关键。当前的一些研究都存在以下问题：首先，节点选取的最佳相似性指标不稳定，导致事件检测准确率不高，且对事件的敏感性不足，不能有效检测出事件的发生；其次，没有展现网络演化过程中每个阶段选取指标的数量变化情况，仅从总体数量占比进行了分析，不利于进一步的探索研究。

综上所述，本文提出了一种基于节点演化分阶段优化的事件检测方法 NESO_ED，旨在加强节点选取指标的准确性，提高事件检测的稳定性和敏感性。该方法包括节点相似性指标权重优化算法 ED_PSO、节点最佳相似性指标选取算法 ED_Sim 和检测网络波动的 ED_D 算法。本文的主要工作如下：

（1）将网络演化阶段划分为两个阶段，根据不同阶段采用不同的优化函数，将粒子群优化算法（Particle Swarm Optimization，PSO）引入节点相似性指标权重优化算法 ED_PSO 中，从而优化参数选择，以更加稳定地获取节点的相似性指标权重数组。

（2）根据网络演化不同阶段的特点设置不同节点选取指标的规则，使提出的节点最佳指标选取算法 ED_Sim 能更好地量化网络演化情况。

（3）展示网络演化过程中每阶段节点选取最佳指标的变化情况，以验证事件的发生对节点演化产生了不同的影响。

本文第 2 节介绍相关领域的研究工作；第 3 节详细介绍了 NESO_ED 方法；第 4 节在真实的社会网络数据集中进行对比实验，观察 NESO_ED 方法的表现；最后总结全文并进行展望。

2　相关工作

在已有的社会网络事件检测方法中，有只针对特定的邮件网络而设计的方法[15]、引入模糊概念模拟不确定性关系进行事件检测的方法[16]，以及通过网络演化进行事件检测的方法，其中最常见的网络演化分析方法是直接建立网络演化模型。

在直接建立网络演化模型的方法中，模型的构建依据准确的演化机制。最经典的模型是 W-S 小世界模型[17] 和 B-A 无标度模型[18]，其分别依据三元闭包[19] 和优先链接[18] 机制构建。然而，在真实的社会网络中，直接建立网络演化模型面临以下问题：1）在真实网络中很难准确预测网络演化机制，进而难以构建模型，如具有高聚类系数的网络不一定由聚类机制驱动，也可能是其他机制的副产品，如优先链接机制；2）社会网络统计特征众多，很难选择合适的统计量进行模型比较[20-21]，如聚类系数的值通常取决于网络的规模，即与小规模网络相比，大规模网络通常具有较小的聚类系数；3）大多数复杂的网络演化模型倾向于关注全局网络，虽然 Wang 等[22] 提出了一种考虑节点间信息传播和微观特征的网络演化模型，但该模型只能从一个小型的人工构建的初始网络演化而来，无法对现实网络进行演化分析。

网络演化分析的另一种方法是基于链路预测的方法，其中基于相似性的链路预测方法是目前运用最广泛的方法[23]，该方法快速、准确，并且需要的信息量少，计算量也小。

Kleinberg 等[24] 系统地提出了链路预测问题，并对比了多种相似性指标（共同邻居[19]、Jaccard 系数[25]、优先链接[18] 等）在链路预测中的表现。Liu 等[10] 率先提出了基于链路预测的网络演化分析方法，在推测中国城市航空网络演化机制的例子中，得到了与直接建立网络演化模型一致的结论，推动了网络演化分析的研究。Wu 等[26] 提出了一种基于事件的社会网络演化分析框架，首先发现网络演化中的重大事件，再基于事件对网络分段，利用图近似算法对网络演化进行分析。Zhang 等[27] 引入链路预测和似然分析两种方法来测量网络演化机制，该研究表明可以定量地对网络的多种演化机制进行测量，并且其提供了统一、有效和可扩展的测量方法。Hu 等提出了两种方法，一种是基于链路预测的事件检测方法[28]，考虑了单一指标下的网络演化分析方法；另一种是面向节点演化波动的事件检测方法[29]，其基于节点演化的差异性，采用各节点的最佳相似性指标提出了网络相似性计算算法，进而构建了事件检测算法。

Wu 等的方法和 Hu 等的第一种方法倾向于宏观整体网络的演化分析，没有考虑微观节点的演化；Zhang 等的方法虽然考虑了微观节点的演化，但忽略了不同指标的度量差异带来的影响；Hu 等的第二种方法因真实网络中节点演化具有多变性而存在节点选取指标不稳定、欠准确的问题，也没有具体分析演化过程中节点选取指标的变化情况，事件检测性能欠佳。

综上所述，直接建立网络演化模型的方法需要精确的网络演化机制，并且模型的建立都是基于宏观整体网络的，没有深入分析节点产生的影响。基于链路预测的方法虽然有面向节点演化的方法，但节点演化机制的多样性，使得事件检测的准确性和敏感性并不好，导致其不能准确高效地进行事件检测。本文在先前工作的基础上，考虑了网络内部节点的演化机制，提出了一种通过加强节点指标选择稳定性来检测网络异常的方法，进而提高了事件检测的稳定性和敏感性。

3　NESO_ED 方法

3.1　相关概念及定义

社会网络的研究通常涉及图论知识，利用图论能够使表

达简单化。表1列出了本文用到的符号及其定义。

在社会网络上进行链路预测，一般都是在同一张图上预测不同的、未连接的节点在未来被链接的可能性大小，其关键步骤便是计算这两个节点的相似性，通过节点间的相似性来判断两个节点链接的可能性。

当利用链路预测进行事件检测时，对不同时刻网络的状态变化进行分析，需要计算不同时刻图的相似性，因此需要计算不同时刻图上同一节点的相似性。不同时刻图上的节点可能稍有不同，上一时刻存在的节点在下一时刻未必存在，反之亦然。为消除此类节点对网络演化描述的不良影响，在每一时刻的网络快照中加入一个与图中全部节点相连的虚拟节点[28]，再进行后续图的相似性计算。引入虚拟节点后的计算公式如表2所列。表2中，$\Gamma(v_i^t)$表示在 t 时刻与节点 i 直接连接的节点集合；$k(v_i^t)$表示在 t 时刻节点 i 的度数。

表1 符号及定义

Table 1 Symbols and definitions

符号	定义
$g^t=(V^t,E^t)$	g^t 是 t 时刻的网络快照，由 V^t 和 E^t 组成，V^t 表示 t 时刻网络的节点集，E^t 表示 t 时刻网络的边集
G	G 是由连续时间段的网络快照组成的图序列，即 $G=(g^1,g^2,g^3,\cdots,g^t,\cdots,g^n)$
v_i^t	v_i^t 表示 t 时刻的节点 i
$s(v_i^t,v_i^{t+1})$	$s(v_i^t,v_i^{t+1})$ 表示节点 i 在 t 和 $t+1$ 时刻的相似性
$S(g^t,g^{t+1})$	$S(g^t,g^{t+1})$ 表示 t 时刻的网络快照 g^t 和 $t+1$ 时刻的网络快照 g^{t+1} 的相似性
$\hat{D}(g^{t+1}\parallel g^t)$	$\hat{D}(g^{t+1}\parallel g^t)$ 表示 t 时刻的网络快照 g^t 到 $t+1$ 时刻的网络快照 g^{t+1} 的波动性
$vD(v_i^{t+1}\parallel v_i^t)$	$vD(v_i^{t+1}\parallel v_i^t)$ 表示 t 到 $t+1$ 时刻节点 i 的波动性
$Bindex_{(t,t+1)}(i)$	$Bindex_{(t,t+1)}(i)$ 表示节点 i 在[t,$t+1$]时段的最佳相似性计算指标

表2 引入虚拟节点后的节点相似性计算指标

Table 2 Node similarity calculation index after introducing virtual node

名称	节点相似性计算公式	名称	节点相似性计算公式
共同邻居指标(CNS)	$\lvert\Gamma(v_i^t)\cap\Gamma(v_i^{t+1})\rvert+1$	Jaccard指标(JAS)	$\dfrac{\lvert\Gamma(v_i^t)\cap\Gamma(v_i^{t+1})\rvert+1}{\lvert\Gamma(v_i^t)\cup\Gamma(v_i^{t+1})\rvert+1}$
优先链接指标(PAS)	$(k(v_i^t)+1)*(k(v_i^{t+1})+1)$	Sorenson指标(SOS)	$\dfrac{2(\lvert\Gamma(v_i^t)\cap\Gamma(v_i^{t+1})\rvert+1)}{k(v_i^t)+k(v_i^{t+1})+2}$
Adamic-Adar指标(AAS)	$\sum_{z\in\Gamma(v_i^t)\cap\Gamma(v_i^{t+1})}\dfrac{1}{\frac{k(v_i^t)+k(v_i^{t+1})+2}{2}}$	大度节点有利指标(HPIS)	$\dfrac{\lvert\Gamma(v_i^t)\cap\Gamma(v_i^{t+1})\rvert+1}{\min\{(k(v_i^t)+1),(k(v_i^{t+1})+1)\}}$
Salton指标(SAS)	$\dfrac{\lvert\Gamma(v_i^t)\cap\Gamma(v_i^{t+1})\rvert+1}{\sqrt{(k(v_i^t)+1)*(k(v_i^{t+1})+1)}}$	LNH-I指标(LNHS)	$\dfrac{\lvert\Gamma(v_i^t)\cap\Gamma(v_i^{t+1})\rvert+1}{(k(v_i^t)+1)*(k(v_i^{t+1})+1)}$

3.2 整体框架

本文设计的NESO_ED方法采用链路预测的方式来进行网络演化分析，计算不同时刻相同节点的相似性，从而获得相邻时间片网络的相似性和网络演化的波动性，通过阈值判断是否有事件发生。NESO_ED方法主要包括ED_PSO算法、ED_Sim算法和ED_D算法。NESO_ED方法的输入是社会网络的图序列，输出是社会网络事件发生的时段，其具体步骤如下：

(1)输入社会网络数据集的图序列；

(2)执行ED_PSO算法，获得每阶段中表2所列的8个指标的权重；

(3)执行ED_Sim算法，获得每阶段每个节点的最佳相似性指标；

(4)执行ED_D算法，计算图相似性和波动值，针对波动值设定阈值，当波动值大于阈值时意味着此阶段有事件发生；

(5)输出社会网络中有事件发生的时段。

3.3 NESO_ED算法

NESO_ED算法按时间顺序将网络演化划分为不同时段，第一个时段到第二个时段的网络演化称为初始阶段，其余时段更替称为后续阶段；再通过分阶段优化方法加强节点选取指标的准确性，以提高事件检测的稳定性和敏感性。

ED_PSO算法的目的是获取每阶段8种指标的权重，并将其作为该阶段每个节点的权重。在量化网络演化的波动性中，每个节点要选取各自的最佳指标。从表2可以看出，不同的相似性计算指标具有不同的度量单位，需要平衡各指标之间的度量差异，因此构建相似性权重数组 *params* 对指标进行加权调整[29]。由8种指标组成的相似性计算指标集合为SimList=[CNS,PAS,AAS,SAS,JAS,SOS,HPIS,LNHS]，与之相对应的是相似性指标权重数组 *params*。

ED_PSO通过PSO来获取每阶段指标的权重，PSO算法是Kennedy等于1995年提出的一种群智能优化算法，源于对鸟群捕食行为的研究[30]。它从随机解出发，通过迭代寻找最优解，并使用适应度来评价解的优劣。该算法结构简单、易于实现、收敛速度快，最主要的是不需要借助问题的特征信息，只需要利用少数参数自适应调节适应度函数来获取最优解，非常适合利用链路预测进行社会网络事件检测的研究。

假设在PSO算法中，最大迭代次数为 $g_{\max}$，种群中共有 m 个粒子，每个粒子代表寻优空间中一个潜在的解，通过适应度来评价解的优劣。初始状态时每个粒子都携带一个随机生成的权重数组 *params*，粒子在迭代优化中通过自身和群体的历史最优权重来更新当前的速度和权重，不断逼近整个粒子群中最优的权重数组。v_k^g 和 $params_k^g$ 分别是粒子 k 在第 g 次迭代结束时的速度和权重，其计算公式如下：

$$v_k^g=wv_k^{g-1}+c_1r_1(params_{kb}-params_k^{g-1})+c_2r_2(params_G-params_k^{g-1}) \tag{1}$$

$$params_k^g=params_k^{g-1}+v_k^g \tag{2}$$

其中，$params_{kb}$ 是粒子 k 当前的最优权重数组，$params_G$ 是当前种群中的最优权重数组。式(1)主要由3部分构成：1)v_k^{g-1}

表示粒子先前的速度，w 是惯性权重因子，可调节先前速度对当前速度的影响；2）$params_{kb}-params_k^{g-1}$ 表示粒子的自我认知，是粒子 k 的当前权重与自己最好权重之间的距离；$params_G-params_k^{g-1}$ 表示社会经验，是粒子 k 的当前权重与群体最好权重之间的距离。c_1 和 c_2 分别是个体和全局学习因子，用于控制个体和全局最佳权重对速度更新的影响；r_1 和 r_2 是服从[0,1]均匀分布的随机数，这两个参数使算法具有不确定性。

由于 NESO_ED 方法采用分阶段的优化方法，因此适应度函数也是分段函数。定义整体优化指标(Overall Optimization Index，OOI)，并将其作为适应度函数，*OOI* 计算的是网络演化过程中当前阶段包含的所有指标的波动值与之前阶段的平均波动值之差和平均波动值的比值，该指标在当前阶段的优化包含了所有节点波动值的计算，属于整体层面，其计算式如式(3)所示：

$$OOI=\begin{cases} D(g^{t+1}\parallel g^t), & t=1 \\ \dfrac{D(g^{t+1}\parallel g^t)-\dfrac{\sum_{h=1}^{t-1}\widehat{D}(g^{h+1}\parallel g^h)}{t-1}}{\dfrac{\sum_{h=1}^{t-1}\widehat{D}(g^{h+1}\parallel g^h)}{t-1}}, & t>1 \end{cases} \tag{3}$$

在 *OOI* 中，$t(t>0)$是社会网络数据集的时间片。当 $t=1$ 时，*OOI* 计算的是网络演化初始阶段的一个绝对波动值，在图的波动性中，值越大代表相邻时间片的网络结构越不同，发生事件的可能性就越大。当 $t>1$ 时，式(3)表示网络演化的后续阶段，可以看出 *OOI* 计算的是一个相对波动值，同样是值越大发生事件的可能性就越大，且在后续阶段的优化中，*OOI* 采用了上一阶段的网络演化波动值$\widehat{D}(g^{h+1}\parallel g^h)$来优化当前阶段，未采用包含全部指标的网络波动值 $D(g^{h+1}\parallel g^h)$，这样不仅减小了计算量，而且由于$\widehat{D}(g^{h+1}\parallel g^h)$采用节点的最佳指标来计算网络波动值，因此会使下一阶段权重数组的优化更加稳定。$D(g^{t+1}\parallel g^t)$实际上是 $S'(g^t,g^{t+1})$的倒数，而$\widehat{D}(g^{h+1}\parallel g^h)$是$S(g^h,g^{h+1})$的倒数。$S'(g^t,g^{t+1})$和 $S(g^h,g^{h+1})$的计算公式如式(4)、式(5)所示，其中 $U_{t,t+1}=g^t\cup g^{t+1}$。

$$S'(g^t,g^{t+1})=\sum_{i\in U_{t,t+1}}\sum_{index\in SimList}params_{(t,t+1)}[index]\times s_{index}(v_i^t,v_i^{t+1})\times\frac{1}{|g^t\cup g^{t+1}|} \tag{4}$$

$$S(g^h,g^{h+1})=\sum_{i\in U_{h,h+1}}params_{(h,h+1)}[Bindex_{(h,h+1)}(i)]\times s_{Bindex_{(h,h+1)}(i)}(v_i^h,v_i^{h+1})\times\frac{1}{|g^h\cup g^{h+1}|} \tag{5}$$

ED_PSO 每获取一阶段的指标权重后，就通过 ED_Sim 算法来选取该阶段节点的最佳相似性指标。类似网络的波动性定义，本文定义了节点的波动性，节点的波动性指某相似性指标下的值与对应权重之积的倒数，如式(6)所示：

$$vD(v_i^{t+1}\parallel v_i^t)=\frac{1}{params_{(t,t+1)}[index]*s_{index}(v_i^t,v_i^{t+1})} \tag{6}$$

类似于整体优化指标 *OOI*，本文定义了节点优化指标(Node Optimization Index，NOI)，ED_Sim 算法根据指标 *NOI* 的值选取节点的最佳指标。*NOI* 是网络演化过程中节点 i 在当前阶段的波动值与之前阶段的平均波动值之差和平均波动值的比值，是对单个节点波动值的操作，属于微观层面，其计算式如式(7)所示：

$$NOI=\begin{cases} vD(v_i^{t+1}\parallel v_i^t), & t=1 \\ \dfrac{vD(v_i^{t+1}\parallel v_i^t)-\dfrac{\sum_{h=1}^{t-1}vD(v_i^{h+1}\parallel v_i^h)}{t-1}}{\dfrac{\sum_{h=1}^{t}vD(v_i^{h+1}\parallel v_i^h)}{t}}, & t>1 \end{cases} \tag{7}$$

在 NOI 中，$t(t>0)$是社会网络数据集的时间片。当 $t=1$ 时，*NOI* 计算的是网络演化初始阶段中同一个节点的波动值，是一个绝对值，值越大对整体网络的波动性影响就越大，事件越容易被检测。当 $t>1$ 时，*NOI* 计算的是一个相对值，同样是值越大，对整体的波动性影响越大，事件越容易被检测。

在以上两种算法的执行过程中，ED_PSO 算法根据 *OOI* 来确定各阶段的指标权重数组，ED_Sim 算法根据 *NOI* 来选择每阶段每个节点的最佳相似指标。式(3)为 ED_PSO 算法的适应度函数，$t=1$ 时的公式为网络演化初始阶段的适应度函数，$t>1$ 时的公式为后续阶段的适应度函数。由前文的分析可知，*OOI* 的优化是面向网络演化过程中某一阶段网络宏观的、整体的波动性值，值越大代表社会网络结构的变化越大，事件在此阶段发生的可能性就越大，因此通过最大化适应度函数来计算每个阶段的权重数组。ED_Sim 算法在选取节点的最佳相似性指标时，初始阶段的节点根据式(7)中 $t=1$ 条件下的公式进行选择，后续阶段的节点根据式(7)中 $t>1$ 条件下的公式进行选择，并且可以选择使 *NOI* 值最大或最小的指标，因此组合排列后的选择方案如表 3 所列。

表 3　ED_PSO 和 ED_Sim 的 4 种方案

Table 3　Four options for ED_PSO and ED_Sim

	ED_PSO	ED_Sim	方案
初始阶段	最大化	最大化	①
	最大化	最小化	②
后续阶段	最大化	最大化	③
	最大化	最小化	④

默认假设在初始阶段没有事件发生，在后续阶段均有事件可能发生。*NOI* 指标在网络演化的初始阶段是一个绝对的波动值，而在后续阶段是一个相对的波动值，并且 *NOI* 优化的是网络演化过程中各阶段节点的波动性值，属于微观角度，其值越大对整体的网络波动性的影响就越大，事件也就越容易被检测。在初始阶段没有事件发生，且 *NOI* 值作为一个绝对值比相对值大，如果最大化 *NOI* 值将不利于后续阶段的检测，因为后续阶段 *OOI* 的优化过程采用了先前阶段的结果，所以在初始阶段选择使 *NOI* 值最小的指标作为该节点的最佳相似性指标；而在后续阶段可能会发生事件，*NOI* 作为一个相对值比绝对值小，选择最大化 *NOI* 指标更利于网络波动性的表示。因此，在优化过程中，本文选择了方案②和方案③。ED_PSO 算法和 ED_Sim 算法的具体步骤如算法 1 所示。

算法 1 ED_PSO 算法和 ED_Sim 算法

输入：社会网络图序列 $G=(g^1,g^2,g^3,\cdots,g^t,\cdots,g^n)$

输出：社会网络演化各阶段的权重数组及每个节点的最佳相似性指标

1. 构建网络演化过程中节点相似性计算指标集合 SimList
2. For t=1 to n−1：
  If t=1：
    For g=1 to g_{max}：
      For k=1 to m：
        If fit($params_k^g$)>fit($params_{kb}$) then $params_{kb}=params_k^g$
        If fit($params_{kb}$)>fit($params_G$) then $params_G=params_{kb}$
        执行式(1)和式(2)，并限制速度和权重的范围
      End
    Return $params_{(t,t+1)}$
    End
  $Bindex_{(t,t+1)}(i)=argmin\{NOI\}$
  $Ind\_List[i][t]=Bindex_{(t,t+1)}(i)$
  Else：
    For g=1 to g_{max}：
      For k=1 to m：
        If fit($params_k^g$)>fit($params_{kb}$) then $params_{kb}=params_k^g$
        If fit($params_{kb}$)>fit($params_G$) then $params_G=params_{kb}$
        执行式(1)和式(2)，并限制速度和权重的范围
      End
    Return $params_{(t,t+1)}$
    End
  $Bindex_{(t,t+1)}(i)=argmax\{NOI\}$
  $Ind\_List[i][t]=Bindex_{(t,t+1)}(i)$
3. End
4. Return Ind_List

事件检测的最后一步便是执行 ED_D 算法，该算法首先计算节点的加权相似性，即节点相似性值与相应权重的乘积；再利用节点的加权相似性来计算相邻时间片的图的相似性和波动值，即计算 $S(g^h,g^{h+1})$和$\widehat{D}(g^{h+1}\parallel g^h)$；最后通过将波动值与实验设定的阈值进行比较，来判断是否检测到事件发生，当大于阈值时表明此阶段有事件发生，否则没有事件发生。

4 实验分析

4.1 实验环境与数据

实验环境为 Intel(R) Core(TM) i5-4210U CPU @1.70 GHz，4.0GB 内存，操作系统为 Microsoft Windows 7，编程语言为 Python3.7。

实验中使用真实的社会网络 VAST，VAST[31]是一个拥有 400 人通话记录的数据集，来自一个开放的竞赛项目 IEEE VAST 2008，它包含连续 10 天的通话记录，并且在这 10 天中该网络发生了唯一一次事件，属于单一事件检测，非常有利于社会网络演化的分析。

在现实情况中，VAST 社会网络在第 8 天发生事件，第 8 天与无事件的第 7 天和第 9 天的网络结构是不同的，因此在 VAST 网络演化的第 7 天到第 8 天即第 7 阶段，以及第 8 天到第 9 天即第 8 阶段均有可能引起波动值的突变。由式(3)也可以看出，如果第 7 阶段的波动值突变，由于计算第 8 阶段时利用了第 7 阶段的波动值，因此会很难再突变，而第 7 阶段没有突变时第 8 阶段是可以突变的，因此只要第 7 阶段和第 8 阶段某一阶段的波动值突变，就表示事件检测成功。

4.2 实验参数调整

在 NESO_ED 方法中，需要对 ED_PSO 算法的参数进行调节，ED_PSO 算法的本质是粒子群算法，由 3.3 节可知，PSO 算法的可调参数有 m，g_{max}，w，c_1，c_2，r_1 和 r_2。本文参考文献[29,32]的参数设置。这些参数都是经过十几个基准函数测试后，能够普遍契合各个优化过程的参数。设 m 为 100，g_{max}为 1000，$w=0.8$，$c_1=c_2=2$，$r_1=0.6$，$r_2=0.3$。实验结果如图 1 所示，可以看出波峰在第 1 阶段或第 2 阶段产生，在之后的阶段很平滑，事件检测结果不准确。

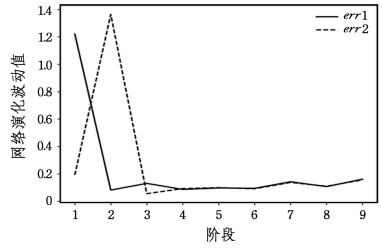

图 1 事件检测方法产生的干扰图

Fig. 1 Interference graph generated by event detection method

图 1 所示的结果可从 *OOI* 指标作为适应度函数和优化时的参数设置两方面来解释。在 ED_PSO 优化过程中，将 *OOI* 作为适应度函数，初始阶段的波动值是绝对值，该值过大会使后续阶段的适应度函数偏小，导致所求的波动值也偏小，其结果如图 1 中的 err1 所示。为了减少此种情况的发生，本文采用表 3 中的方案②，但由于 PSO 本身的随机性，该情况依旧可能发生。

另外，在上述实验中 PSO 参数在初始阶段和其余阶段的设置一样。但 w 是惯性权重，值较大时算法有较强的全局探索能力，值较小时有较强的局部搜索能力，初始阶段默认没有事件发生，网络结构稳定，不需要较强的全局探索能力，而后续阶段可能发生事件，则粒子需要较强的全局探索能力，因此两个阶段中 w 参数不应设置相同的值。此外，实验在发生单一事件的社会网络上运行，不存在连续事件的相互影响，即网络结构在大部分时刻较为稳定，因此对该网络进行粒子群优化时更侧重于自我认知而非社会认知，也就是说粒子在学习过程中更倾向于自身的经验学习，对社会经验的依赖度不高，因此学习因子 c_1，c_2 的值也应不同。

基于上述问题，本文经过百次以上的实验发现，对参数进行以下设置时，事件检测效果最稳定：r_1 和 r_2 保持不变，粒子群数目设置为 100，迭代次数为 100[33]，以减少计算量；对于 ED_PSO，在 $t=1$ 时，设置参数 $w=0.3$，$c_1=3$，$c_2=1$；在 $t>1$ 时，设置参数为 $w=0.8$，$c_1=3$，$c_2=1$。

4.3 实验结果与分析

在 VAST 社会网络数据集上，将本文方法与 LinkEvent 方法[28]和 NodeED 方法[29]进行对比分析。LinkEvent 方法基于固定单一指标 PAS 进行网络演化分析，NodeED 方法采用最佳指标进行网络演化分析，后者具有随机性。实验对各方法的稳定性和敏感性进行对比，稳定性(Event Detection Stability，STA)是根据优化算法的随机性来设定的，指成功检

测事件的次数与运行总次数的比值，该比值越大表示事件检测算法的稳定性越好；敏感性（Event Detection Sensitivity，SEN）是在网络演化波动序列中事件发生时的波动值和无事件时平均波动值之差除以无事件时平均波动值的值，敏感值越大表示方法对事件检测的效果越好。为了更好地展现实验结果，将几种方法所得到的网络演化波动值进行归一化处理，结果如图 2 所示。

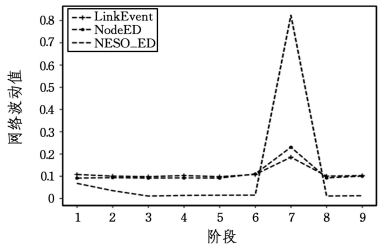

图 2　VAST 中 3 种方法的网络演化波动图

Fig. 2　Network evolution fluctuation of three methods in VAST

由图 2 可以看出，每种方法在第 7 阶段的波动值都变大，其余阶段的波动值都很小，这说明第 7 阶段有事件发生，而事实证明 VAST 社会网络在第 7 阶段确实发生了事件。从图 2 也可以直观地看出 NESO_ED 方法的效果是最显著的。针对 3 种方法运行相同的次数，计算敏感值 SEN 和稳定值 STA 的平均值，结果如表 4 所列。

表 4　3 种方法的 SEN 和 STA

Table 4　SEN and STA values of three methods

	LinkEvent	NodeED	NESO_ED
SEN	15.73	31.52	51.47
STA	100%	30%左右	50%～60%

由表 4 可以看出，NESO_ED 方法的效果更佳，事件敏感值 SEN 相较 LinkEvent 方法提高了 227%，相较 NodeED 方法提高了 63%。另外，在稳定性比较中，由于 LinkEvent 方法是基于固定的单一指标进行网络演化波动性分析的，事件检测的敏感性值较低，但该方法没有随机性，稳定性最高。而 NESO_ED 方法的稳定性相较 NodeED 方法提高了 66%，说明 NESO_ED 方法在节点选择最佳指标时更加稳定、准确，能更好地表示网络演化的波动性，因此事件检测的效果更好。

在实验中保存将事件检测正确的数据，计算整个网络演化过程中所有节点的最佳相似性计算指标占比，结果如图 3 所示。

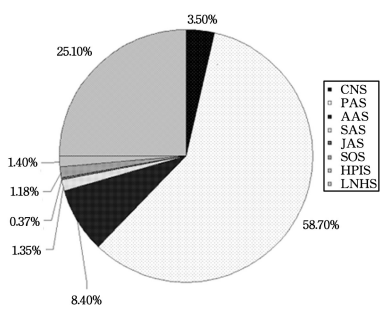

图 3　全部节点最佳相似性计算指标的占比

Fig. 3　Ratio of nodes' optimal similarity index

由图 3 可以明显看出，在所有指标中 PAS 指标占比最大，其次是 LNHS，AAS 和 CNS 指标，剩余指标占比非常小。这说明 VAST 网络中绝大部分节点按照优先链接的方式进行演化，此结果也解释了采用单一指标的 LinkEvent 方法选择 PAS 指标的原因。由于从图 3 中并不能发现网络演化过程中节点选取最佳指标的变化情况，因此本文用图 4 给出了网络演化中每阶段节点选择最佳指标的数量变化情况，其中横坐标是网络演化阶段，纵坐标是所有节点选择最佳指标的数量。

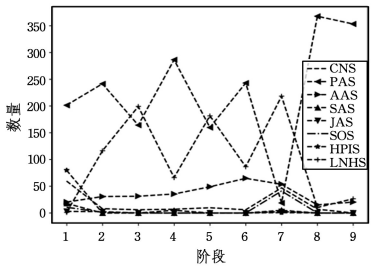

图 4　每阶段节点的最佳指标波动图

Fig. 4　Fluctuation of nodes' optimal index in each stage

从图 4 可以看出，在第 1 阶段各指标数目相对分散，指标数目值相比其他阶段更大，这验证了 3.3 节关于 ED_PSO 和 ED_Sim 的分析。同时也能看出，PAS 的数目最多，其次是 LNHS，该结果与图 3 所示结果吻合。此外，在演化过程中 PAS 和 LNHS 的数量此起彼伏，在平稳阶段的起伏程度不算太大，但是在事件发生的阶段起伏变化剧烈，尤其是 PAS 的数量在事件发生阶段急剧减少。同时，在没有事件发生的阶段中，CNS，SAS，JAS 和 SOS 的数量都很少甚至为零，而在事件发生阶段其数量却有了提升，其中 CNS 和 SOS 的提升最为明显。上述内容说明事件的发生对节点的演化产生了不同的影响，部分节点改变了原有的演化机制，选择的演化方向也大有不同，这间接证明了节点演化的多样性。总而言之，在平稳阶段节点多选择 PAS，LNHS 和 AAS 指标，其余相似性指标数量很少，但在事件发生阶段其余指标数量有所提升，表明了基于节点演化的事件检测方法的有效性。

结束语　针对当前社会网络中面向节点演化的事件检测方法存在准确性欠佳和敏感度不高的问题，本文提出了一种基于节点演化分阶段优化的事件检测方法。该方法将网络演化阶段划分为两段，针对不同的演化阶段采用不同的优化方法，使每阶段获取的指标权重数组稳定性增强，同时采用不同规则选取节点的最佳相似性指标，使节点的演化表示更佳。本文在 VAST 社会网络数据集上对提出的算法进行验证，并将其与几种方法进行对比，实验结果表明 NESO_ED 方法不仅在敏感性上有了明显提高，而且能够准确且快速地检测事件的发生。本文根据实验结果分析了节点在网络演化的每阶段节点最佳指标的选取情况，发现指标选取在网络平稳阶段和波动段有明显的不同，充分反映了事件发生对节点演化产生了不同的影响，从而导致网络结构发生改变，证明了基于节点演化分阶段优化的事件检测方法的有效性。

未来的研究工作主要围绕以下几个方面：1）本文方法在发生单一独立事件的社会网络中表现得非常好，在多事件发生的社会网络事件检测中有待进一步研究；2）探究相似性指标与网络演化的对应关系，并考虑引入链路预测的其他方法来弥补基于相似性链路预测的不足。

参 考 文 献

[1] SHI L, DU J P, LIANG M Y. Social network burst topic discovery based on RNN and topic model[J]. Journal on Communications, 2018, 39(4): 189- 198.

[2] JIE F, XIE F, LI L, et al. Latent event-related burst detection in social networks[J]. Acta Automatica Sinica, 2018, 44(4): 730-742.

[3] HAN Z M, CHEN Y, LIU W, et al. Research on node influence analysis in social network [J]. Journal of Software, 2017, 28(1): 84-104.

[4] ZHONG Z M, GUAN Y, LI C H, et al. Localized Top-k burst event detection in microblog[J]. Chinese Journal of Computers, 2018, 41(7): 1504-1516.

[5] ZHANG L M, JIA Y, ZHOU B, et al. Online burst event detection based on emotions[J]. Chinese Journal of Computers, 2013, 36(8): 1659-1667.

[6] FEI S D, YANG Y Z, LIU P Y, et al. Method of burst events detection based on sentiment filter[J]. Journal of Computer Applications, 2015, 35(5): 1320-1323.

[7] XIONG Y, ZHANG Y F, FENG S, et al. Event detection and tracking in microblog stream based on multimodal feature deep fusion[J/OL]. Control and Decision. [2019-05-30]. http://kns.cnki.net/KCMS/detail/21.1124.TP.20180416.0932.015.html.

[8] WANG B Y, WU Z Y, SHEN S B, et al. Survey on event detection research in social media[J]. Computer Technology and Development, 2018, 28(9): 105-111.

[9] LIN C, LIN C, LI J, et al. Generating event storylines from microblogs[C]// Proceedings of the 21st ACM international conference on information and knowledge management (CIKM'12). Maui, Hawaii, USA, 2012: 175-184.

[10] LIU H K, LV L Y, ZHOU T. Uncovering the network evolution mechanism by link prediction [J]. Scientia Sinica (Physica, Mechanica & Astronomica), 2011, 41(7): 816-823.

[11] WANG H, HU W B, QIU Z Y, et al. Nodes' evolution diversity and link prediction in social networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(10): 2263-2274.

[12] HANLEY J A, MCNEIL B J. The meaning and use of the area under a receiver oSENating characteristic (ROC) curve[J]. Radiology, 1982, 143(1): 29-36.

[13] HERLOCKER J L, KONSTAN J A, TERVEEN L G, et al. Evaluating collaborative filtering recommender systems[J]. ACM Transactions on Information Systems, 2004, 22(1): 5-53.

[14] ZHOU T, REN J, MEDO M, et al. Bipartite network projection and personal recommendation[J]. Physical Review E, 2007, 76(4): 70-80.

[15] LI Q G, SHI J Q, QIN Z G, et al. Mining user behavior patterns for detection in email networks[J]. Chinese Journal of Computers, 2014, 37(5): 1135-1146.

[16] YANG L M, ZHANG W, CHEN Y F, et al. Time-series prediction based on global fuzzy measure in social networks[J]. Front Inform Technol Electron Eng, 2015, 16(10): 805-816.

[17] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. Nature, 1998, 393(6684): 440-442.

[18] BARABASI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.

[19] RAPOPORT A. Spread of information through a population with socio-structural bias: I. assumption of transitivity[J]. The Bulletin of Mathematical Biophysics, 1953, 15(4): 523-533.

[20] XIONG C, CHEN Y F, CANG J Y. Event-based node influence analysis in social network evolution [J]. Computer Science, 2016, 43(S1): 404-409.

[21] WU X D, LI Y, LI L. Influence analysis of online social networks [J]. Chinese Journal of Computers, 2014, 37(4): 735-752.

[22] WANG Y, CUI J H, ZHANG T, et al. A complex network evolution model based on microscopic characteristic of nodes[C]// Proceedings of IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C). Lisbon, Portugal: IEEE Press, 2018: 388-393.

[23] LV L Y. Link prediction on complex network [J]. Journal of University of Electronic Science and Technology of China, 2010, 39(5): 651-661.

[24] LIBEN-NOWELL D, KLEINBERG J. The link-prediction problem for social networks[J]. Journal of the American for Information Science and Technology, 2007, 58(7): 1019-1031.

[25] JACCARD P. Etude comparative de la distribution florale dans une portion des Alpes et du Jura [J]. Impr. Corbaz, 1901, 37(139): 547-579.

[26] WU B, WANG B, YANG S Q. Framework for tracking the event-based evolution in social networks[J]. Journal of Software, 2011, 22(7): 1488-1502.

[27] ZHANG Q M, XU X K, ZHU Y X, et al. Measuring multiple evolution mechanisms of complex networks[J]. Scientific Reports, 2015, 5: 10350.

[28] HU W B, PENG C, LIANG H L, et al. Event detection method based on link prediction for social network evolution[J]. Journal of software, 2015, 26(9): 2339-2355.

[29] WANG H, HU W B, QIU Z Y, et al. An event detection method for social networks based on evolution fluctuations of nodes[J]. IEEE Access, 2018, 6: 12351-12359.

[30] KENEDY J, EBERHART R. Particle swarm optimization[C]// Proceedings of IEEE International Conference on Neural Network. Perth, Australia: IEEE Press. 1995: 1942-1948.

[31] GRINSTEIN G, PLAISANT C, LASKOWSKI S, et al. VAST 2008 Challenge: Introducing mini-challenges[C]// Proceedings of IEEE Symposium on Visual Analytics Science and Technology. Columbus, OH, USA: IEEE Press, 2008: 195-196.

[32] WANG D F, MENG L. Performance analysis and parameter selection of PSO algorithm[J]. Acta Automatica Sinica, 2016, 42(10): 1552-1561.

[33] CAI Q, GONG M, MA L, et al. Greedy discrete particle swarm optimization for large-scale social network clustering[J]. Information Sciences, 2015, 316(41): 503-516.

FU Kun, born in 1979, Ph.D, associate professor. Her main research interests include social network analysis, community detection and reconfigurable calculation.