

DOI: 10.13382/j.jemi.B1902793

基于子模优化的边界域处理社团发现算法*

杨雪洁^{1,2} 曹风云^{1,3} 陈洁² 赵姝² 张燕平²

(1. 合肥师范学院 计算机学院 合肥 230061; 2. 安徽大学 计算机科学与技术学院 合肥 230601;
3. 合肥师范学院 电子信息系统仿真设计安徽省重点实验室 合肥 230061)

摘要:使用聚类粒化方法求取非重叠社团结构时,经常会出现重叠区域。三支决策模型将两个存在重叠的社团的左边社团中非重叠部分定义为正域,右边社团中非重叠部分定义为负域,而两个社团的重叠部分定义为边界域。为了获得更好的社团性能,须将边界域中的节点进行二次划分。子模优化在机器学习中有广泛的应用,如果目标函数具有子模性,则存在一个简单的贪心算法能在多项式时间内以常数因子逼近问题的最优解。将子模优化思想引入社团重叠区域节点的处理,提出一种基于子模优化的边界域处理社团发现算法(SO-CDA)。定义设备选址函数进行子模优化,重叠节点的划分可以转化为子模函数最大化问题,在7个真实网络上的实验结果表明,SO-CDA能够有效地进行社团划分,性能更加稳定。

关键词:子模函数;三支决策;复杂网络;社团发现

中图分类号: TP181; TN99

文献标识码: A

国家标准学科分类代码: 520.20

Community detection algorithm for boundary region processing based on submodular optimization

Yang Xuejie^{1,2} Cao Fengyun^{1,3} Chen Jie² Zhao Shu² Zhang Yanping²

(1. School of Computer, Hefei Normal University, Hefei 230061, China; 2. School of Computer Science and Technology, Anhui University, Hefei 230601, China; 3. Anhui Province Key Laboratory of Simulation and Design for Electronic Information System, Hefei 230061, China)

Abstract: Overlapping regions often occur when non-overlapping community structure is obtained by clustering granulation method. The nodes in the non-overlapping parts of the community of the left side between two communities with overlapping parts were defined as positive regions. Then, the nodes on its right are denoted as the negative region, and nodes in the overlapping parts are denoted as the boundary region. In order to achieve better community structure, it is necessary to divide the nodes in the boundary region into non-overlapping parts. Submodular optimization is widely used in machine learning, If the objective function has sub-modularity, it exists a simple greedy algorithm which can approximate the optimal solution of the problem with constant factor in polynomial time. In this paper, submodular optimization is introduced into the processing of nodes in overlapping communities. and a community detection algorithm (SO-CDA) for boundary region processing based on submodular optimization is proposed. The device location function is defined for submodular optimization, the partition of overlapping nodes is converted to the maximization of submodular function. The experimental results on seven real networks show that SO-CDA can effectively divide communities and achieve more stable performance.

Keywords: submodular function; three-way decision; complex networks; community detection

收稿日期: 2019-11-28 Received Date: 2019-11-28

* 基金项目: 国家自然科学基金(61673020, 61602003)、安徽省自然科学基金(1708085QF156)、安徽省高校优秀青年人才项目(gxyq2019068)、电子信息系统仿真设计安徽省重点实验室开放基金(2019ZDSYSZY06)资助项目

0 引言

随着对社会网络、生物网络等复杂网络研究的不断深入,人们发现这些实际网络中都存在社团结构,即整个网络由若干个社团构成。在同一社团中的节点联系较为紧密,不同社团间的节点联系较为稀疏^[1]。研究网络中的社团结构,有助于对整个网络模块、功能的研究,方便更加准确理解复杂网络的拓扑结构和形成机制。

近年来,研究者从不同角度提出了很多社团发现算法,如标签传播模型^[2]、谱分析法^[3]、非负矩阵分解^[4]、层次聚类^[5-9]等。层次聚类的主要思想是依据节点间的相似程度对节点进行合并或分解。其中合并的层次聚类方法主要是寻求社团“最中心”的节点,不断将最相似的两个节点合并到一起,从而不断扩展社团规模。依据定义相似度的不同,产生了很多不同的算法。如 Newman 快速算法^[5]每次都合并能产生最大模块度的两个社团, CNM^[6]算法采用堆数据结构计算和更新模块度, Louvain^[7]算法基于多层次优化模块度,优化的目标是最大化整个图属性结构的模块度,可以发现层次性的社区结构。

随着研究的深入,人们发现在社团实际划分时经常出现重叠区域,即一个节点被多个社团包含,重叠节点不易划分的情况,实际上,将重叠节点划分到正确的社团中更有助于发现社团内存在的规律,并预测网络的行为和功能^[10-11]。因此,实现非重叠的社团划分是非常有必要的。

现有大多数社团划分算法在处理重叠区域时都是基于传统的二支决策^[12-13]思想,即依据现有的信息直接做出接受或拒绝决策,实际上重叠区域节点经常是因为现有信息量不精确或不完整无法决定其归属,才会出现在重叠区域。三支决策理论将传统的正域、负域的二支决策语义拓展为正域、边界域和负域的三支决策^[14]语义,将暂时无法处理的节点(重叠节点)放在边界域中,暂时不处理,等待新信息后再进行二次划分。因此基于层次聚类的社团划分算法的研究重点在于边界域的处理问题。

子模优化理论是机器学习、数据挖掘领域的研究热点,如果目标函数具有子模性,则基于简单的寻优策略,如贪心算法,就可以在多项式时间内获得高精度的解。它已被应用在传感器放置、文本摘要^[15]、特征选择^[16]、数据子集选择^[17]和多标签主动学习^[18]等多领域。Wu 等^[18]使用基于子模函数的多标签主动学习方法来选择实例标签对,将多标签主动学习问题转换为子模函数最大化问题。Wei 等^[19]使用基于约束子模函数最大化的数据,解决了训练语音识别系统时需要大量训练数据的

问题。Jin 等^[20]使用子模函数对神经网络模型结构进行优化,将网络结构优化转化为子集选择,相关实验证明了优化后的模型具有更高的识别率、速度更快。

本文将子模优化引入社团边界域处理中,提出基于子模优化的边界域处理社团发现算法(SO-CDA)。该算法通过初始粒化后,选择合适的粒化准则进行聚类粒化,选择 EQ 值最大层作为最优层形成重叠的社团划分结果。依据社团重叠部分(边界域)中节点与已明确划分的社团(正域、负域)之间的紧密关系构造子模函数,依据子模函数最大化选择合适子集(节点)加入到正域或负域,获取接近全局最优的非重叠社团划分。

1 相关工作

1.1 聚类粒化

聚类粒化主要是依据样本间的相似性,选择合适的粒化原则进行聚类,形成划分结果。假设无向网络 $G = (V, E)$, 这里 $V = \{v_1, v_2, \dots, v_n\}$ 为图中节点的集合, $E = \{(v_i, v_j) \mid v_i, v_j \in V\}$ 为图中边的集合。

文献[21]将聚类粒化思想引入社团划分,将每个节点 v_i 视为基本粒 $Gr(v_i)$, 通过初始粒化操作 $Gran(G, Gr)$ 构造以每个节点为中心的粒,形成初始粒度空间 Gr^1 , 再利用聚类粒化操作 $EXGran(Gr^m, \lambda)$ 对第 m 层粒集合 Gr^m 进行聚类粒化,若粒度空间 Gr^m 层中任意两个粒的粒化系数 $f(Gr_i^m, Gr_j^m) \geq \lambda$ 时,对这两个粒进行聚类粒化形成新的粒 Gr^{m+1} , 否则就不进行聚类粒化,重复粒化操作,最终形成多层次的社团划分结果。粒 $Gr(v_i)$ 、初始粒化 $Gran(G, Gr)$ 、聚类粒化 $EXGran(Gr^m, \lambda)$ 相关定义如下。

定义 1 粒 $Gr(v_i)$, 简称为 Gr_i 。以节点 $v_i \in V$ 为中心的 $Gr(v_i)$ 满足以下条件^[21]:

- 1) $v_i \in Gr(v_i)$;
- 2) $\forall v_j \in V$, 若 v_i 与 v_j 有边直接相连, 则 $v_j \in Gr(v_i)$ 。

定义 2 初始粒化 $Gran(G, Gr)$ 。按照粒 Gr_i 定义为每一个节点为中心构造初始粒集合^[21] $Gr = \{Gr_i\}$ 。

定义 3 聚类粒化 $EXGran(Gr^m, \lambda)$ 。对第 m 层粒集合 Gr^m 进行聚类粒化,得到新的粒化集合 Gr^{m+1} , 满足粒化系数 $f(Gr_i^m, Gr_j^m) \geq \lambda$, 则进行聚类粒化^[21]:

- 1) $Gr'_i = Gr_i \cup Gr_j$;
- 2) 从 Gr^m 中删除 Gr_i 和 Gr_j ;
- 3) $Gr^{m+1} = Gr^m + \{Gr'_i\}$ 。

粒化系数 $f(Gr_i^m, Gr_j^m)$ 主要依据样本的相似程度,参数 λ 主要控制粒化合成的粗细,依据不同的相似度量方法和粒化程度,可以设置相应的粒化系数 f 和参数 λ 。

本文 SO-CDA 算法中的粒化系数 $f(Gr_i^m, Gr_j^m)$ 主要考虑粒 Gr_i^m 与 Gr_j^m 的相似性, 如式(1)所示。

$$f(Gr_i^m, Gr_j^m) = \frac{|Gr_i^m \cap Gr_j^m|}{\min(|Gr_i^m|, |Gr_j^m|)} Gr_i^m, Gr_j^m \in Gr^m \quad (1)$$

式中: $|Gr_i^m \cap Gr_j^m|$ 表示粒 Gr_i^m 与粒 Gr_j^m 中相同元素的个数; $\min(|Gr_i^m|, |Gr_j^m|)$ 表示粒 Gr_i^m 与粒 Gr_j^m 中较少元素的个数。

参数 λ 设置为当前 Gr^m 层中所有任意两个粒 $f(Gr_i^m, Gr_j^m)$ 值求和之后的平均值, 以保证在每一层中都能依据当前层粒集合特点进行聚类粒化。

1.2 重叠社团的三个域

聚类粒化后形成了多层次的重叠社团, 为获得更合理的非重叠社团划分结果, 需要选择合适粒度的重叠社团结构, 扩展模块度值 EQ 是评价重叠社团划分结果的主要指标, EQ 值越大, 说明社团结构越强, 反之结构越弱, 因此本文算法选择所有粒度空间中 $EQ^{[22]}$ 最大的层作为最终获得的重叠社团结构(最优层), EQ 计算公式如下:

$$EQ = \frac{1}{2m} \sum_{i,j} \frac{1}{O_i O_j} (A_{ij} - \frac{d_i d_j}{2m}) \delta(C_i, C_j) \quad (2)$$

式中: m 表示网络中边的个数; O_i 表示节点 i 属于相应社团的数量; A_{ij} 为邻接矩阵; d_i 为节点 i 的度。如果节点 i 和节点 j 在同一个社团, 则 $\delta(C_i, C_j)$ 值为 1, 否则为 0。

使用三支决策思想处理 EQ 最大层的重叠社团结构, 对任意两个存在重叠的社团做出如下划分。

- 1) 正域 (POS), 两个存在重叠区域的社团中左边社团非重叠区域。
- 2) 负域 (NEG), 两个存在重叠区域的社团中右边社团非重叠区域。
- 3) 边界域 (BND), 两个社团的重叠部分。

其中, 正域与负域为相对的概念, 也可将左边界定义为负域, 右边定义为正域。本文为叙述方便, 将左边称为正域, 右边称为负域。由此社团划分最终关键在于从边界域节点集合中选择合适的节点子集加入到正域或负域中。

2 基于子模优化的边界域处理社团发现算法

2.1 子模优化

基于三支决策理论, 对基于聚类粒化获得的最优层重叠社团结构做出正域、负域和边界域的区分, 对于社团划分, 边界域的存在是暂时的。随着对边界域节点与正域、负域节点间关系认识的深入, 边界域节点最终将被二

次划分, 本文将子模优化理论引入边界域节点处理, 完成非重叠社团的划分。

子模函数定义如下, 设 V 为有限数据集, 2^V 为 V 中的元素组成的子集集合, 记集合函数 $f: 2^V \rightarrow R$, 使得任意 $S \subseteq V$ 对应一个实数, 同时对任意 $S \subseteq R \subseteq V$ 且 $s \in V \setminus R$, 式(3)都成立, 则称集合函数 $f(\cdot)$ 为子模函数。

$$f(R \cup \{s\}) - f(R) \leq f(S \cup \{s\}) - f(S) \quad (3)$$

子模函数具有效益递减属性, 在 S 中增加一个元素所增加的收益要小于等于在 S 的子集中增加一个元素所增加的收益。通俗的说就是你把所有商品看成一个集合, 随着你所拥有的物品数量的增加, 那么你获得剩下物品所得到的满足程度越来越小。从寻优角度看, 如果目标函数具有子模性, 则存在一个简单的贪心算法能在多项式时间内以常数因子 $(1-1/e)$ 逼近问题的最优解^[23]。

设备选址函数是经典的子模函数模型, 该模型表明, 如果可以定义两样本间的紧密程度 w_{ij} , 就可以为选定的样本集构造该子模函数。本文选择设备选址模型来构造子模函数, V 为正域和负域中的节点集合, S 为需加入正域或负域的节点集合, 属于边界域, 子模函数定义为:

$$f(s) = \sum_{i \in V} \max_{j \in S} w_{ij} = \max_{i \in V, j \in S} \sum w_{ij} \quad (4)$$

模块度 $Q^{[24]}$ 可以衡量社团内部连接紧密程度, 计算公式如下:

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \delta(C_i, C_j) = \sum_{i,j} \frac{1}{2m} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \delta(C_i, C_j) \quad (5)$$

式中: m 表示网络中边的个数; A_{ij} 为邻接矩阵; d_i 为节点 i 的度。如果节点 i 和节点 j 在同一个社团, 则 $\delta(C_i, C_j)$ 值为 1, 否则为 0。

Q 值计算公式与子模函数公式形式具有一致性, 将 w_{ij} 选取为 $\frac{1}{2m} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \delta(C_i, C_j)$, $f(s)$ 可以转换为:

$$f(s) = \max_{i \in V, j \in S} Q \quad (6)$$

由此重叠节点的划分可以转化为子模函数最大化问题。即通过选择每一个边界域节点最大 Q 值的划分, 累加后形成最终 Q 值最大的划分, 如下所示:

$$s^* \in \operatorname{argmax}_{s \in V \setminus S} f(S \cup \{s\}) - f(S) \quad (7)$$

依据构造的子模函数进行的社团划分, 通过“贪心”选择当前最大 Q 值节点划入相应的正域或负域, 由于子模函数在贪心算法中能以常数因子 $(1-1/e)$ 逼近问题的最优解, 这为基于子模优化的社团划分获取接近全局最优的非重叠划分提供了理论保证。

2.2 算法描述

基于子模优化的边界域处理社团发现算法通过初始粒化、聚类粒化后, 选择 EQ 值最大的层作为最优层形成

重叠的社团划分。依据边界域中节点与已明确划分的正域、负域之间的紧密关系构造子模函数,依据求解子模函数最大化选择合适节点加入到正域或负域,获取接近全局最优的非重叠社团划分。

算法 1 基于子模优化的边界域处理社团发现算法 SO-CDA

输入:一个无向无权网络 $G = (V, E)$

输出:无重叠的社团结构 $POS(G), NEG(G)$

- 1) 初始化网络,依据式(1)进行聚类粒化操作,形成多层重叠社团结构
- 2) 依据式(2)以 EQ 最大值所在层为最优层,形成三域: $POS(G), NEG(G), BND(G)$
- 3) $\forall i \in BND(G)$, 依据式(6)构造子模函数 $f(s)$
- 4) 求解式(7),选择当前最大 Q 值节点划入相应的 $POS(G)$ 或 $NEG(G)$
- 5) 对 $\forall i \in BND(G)$ 重复步骤 3)、4),直至 $BND(G)$ 中所有节点划分完成
- 6) 输出 $POS(G), NEG(G)$

3 实验

为验证算法的有效性,将 SO-CDA 算法应用于真实网络,同时和经典模块度优化算法 CNM 算法、Louvain 算法、较新算法 MEAs-SN、MODBSA/D 算法、MDSTA 算法进行比较。

CNM^[6]为基于模块度优化的分层聚类快速贪心算法;Louvain^[7]为基于模块度优化的社团挖掘算法,能快速处理大规模网络,并得到层次化社团结构;MEAs-SN^[25]为基于相似性的多目标进化算法;MODBSA/D^[26]为具有分解的多目标离散回溯搜索优化算法;MDSTA^[27]为基于模块化的离散状态转移社团发现算法。

上述对比算法的参数依据对应论文中的推荐设置进行设置,其中 MEAs-SN 算法 popsize and gen 设置为 100, MODBSA/D 算法 popsize 设置为 100, neighbourhood size 设置为 40,测试问题的最大迭代次数设为 200, MDSTA 算法考虑全局搜索能力和计算时间,将控制状态转换(SE)和 popsize 设置为 40 和 20, stagnate_number(终止条件)设置为 10。

本文 SO-CDA 算法聚类粒化中参数 λ 设置为当前 Gr^m 层中所有任意两个粒 $f(Gr_i^m, Gr_j^m)$ 值求和之后的平均值,依据每一层特点确定相应的 λ 值,以保证在每一层中都能依据当前层粒集合特点进行聚类粒化。实验环境为 Intel i7-4790 3.6 GHz CPU、内存 8G、MATLAB r2014b 和 Windows10 操作系统。

3.1 实验数据

实验选取了 7 种真实网络数据集,空手道网络(zachary's karate club)、宽吻海豚网络(dolphin social network)、美国大学足球队网络(american college football)、政治书籍网络(books about US politics)、电子邮件网络(the email network)、科学家合作网络(the netscience network)、电网(power grid),数据集可在(<http://www-personal.umich.edu/~mejn/netdata>) Newman 数据集中获取。具体信息如表 1 所示。

表 1 实验数据集信息

Table 1 Information of data sets

Network	节点数	边数
Karate	34	78
Dolphins	62	159
Football	115	613
Polbooks	105	441
Email	1 133	5 451
Netscience	1 589	2 742
Powergrid	4 941	6 594

3.2 评价指标

目前已有很多指标^[28]可以评价复杂网络的社团划分质量,在这些指标中,模块度 Q ^[24]是复杂网络社团发现使用最广泛的方法。模块度 Q 可以衡量社团内部边密度与社团之间边密度的比值,模块度值越大,说明社团结构越好。

归一化互信息 NMI ^[24]指标可以用来估计具有已知分区的真实社团结构与社团划分结果之间的相似性。高 NMI 值反映了划分的社团结构与真实社团结构非常相似,若 NMI 值为 1,两者结构完全相同,若 NMI 值为 0,两者完全不同。

对于大规模网络,其实际社团结构是未知的,因此常用模块度 Q 值来评价社团划分质量。对于已知实际分区的真实网络,可以用 NMI 值来衡量。

3.3 实验结果及分析

Dolphins 网络是近年来验证社团发现算法的一个重要基准网络,该网络是 David 等在 1994~2001 年的 7 年时间内对 62 只宽吻海豚群落进行观察而构建的。网络中的节点代表海豚,节点之间的边代表海豚对之间的关联,该网络主要描述了 62 只海豚的社会关系,其真实网络结构已知,实际应包含两个族群。应用 SO-CDA 算法后该网络也被划分为 2 个社团,对比真实结构,只有一个节点划分错误,如图 1 所示。

对于很多小规模真实网络,其真实社团结构已知,为衡量算法优劣,可以直接对比其 NMI 值,表 2 对比了 SO-CDA 算法与其他算法的 NMI 值。

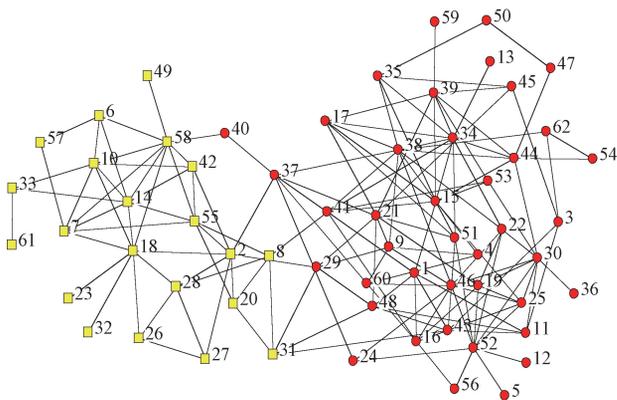


图 1 SO-CDA 划分的 Dolphins 社团结构
Fig. 1 Dolphin community structure divided by SO-CDA algorithm

表 2 真实网络中不同算法的 NMI 值比较
Table 2 Comparison of NMI by different algorithms on real-world networks

Network	CNM	Louvain	MEAs-SN	MDSTA	SO-CDA
karate	0. 7	0. 59	0. 782	0. 687 3	0. 837
Dolphins	0. 56	0. 55	0. 437	0. 587 2	0. 889
football	0. 7	0. 89	0. 924	0. 889 2	0. 864
polbooks	0. 53	0. 51	0. 449	0. 560 2	0. 579

Karate 网络一个由空手道俱乐部的 34 个成员在两年的时间里建立起来的关系网络,有 34 个节点和 78 条边,每个节点代表一个俱乐部成员,连边代表成员在俱乐部内或外有社会交往关系,该网络实际包含两个团体。

polbooks 网络描述的是 2004 年美国选举期间出版的有关政治的书的销售情况。网络共包含 105 个节点和 441 条边,节点表示线上亚马逊网站销售的关于政治的书籍,边代表其连接两端的书籍经常被同一买主一起购买。该网络被书籍在政治方面所蕴含的态度:自由、中性和保守 3 个方面分成了 3 个部分,实际对应 3 个社团。

football 网络是 2000 年秋季常规赛期间美国各大学之间的美式足球比赛网络,由 115 个顶点和 613 个边组成,节点代表每个大学的足球队,连边表示两个球队之间有比赛。该网络被分成 12 组,每个组有 8 到 12 个足球队,实际构成了 12 个社团结构。

karate、Dolphins、polbooks 节点数较少,且边数与节点比例较小,网络稀疏,本文 SO-CDA 算法在上述 3 个网络中获得了最高的 NMI 值,football 网络节点数虽少,但边数较多,网络较为稠密,网络结构特征不明显,MEAs-SN 算法从符号网络角度从正相似性和负相似性两方面进行社团划分,适合处理稠密网络,在 football 网络中达到最大值。MDSTA 算法需要使用状态变化算子进行个体进化形成精英群体,然后再进行群体进化,因此更适用于处

理节点数较多的大规模网络,在小规模网络中表现一般。CNM、Louvain 两种算法只基于模块度优化,存在分辨率局限,无法发现规模小于一定值的社团,模块度高的社团划分结果与真实社区结构经常差异很大。由此可见,相对于其他算法,SO-CDA 算法在发现网络真实社区结构上具有一定优势。

Email、Netscience 和 Powergrid 三个网络真实社团结构未知,规模较大,因此使用模块度 Q 评价划分结果。表 3 对比了 SO-CDA 算法与其他算法的 Q 值。

表 3 真实网络中不同算法的 Q 值比较
Table 3 Comparison of Q-value by different algorithms on real-world networks

Network	CNM	Louvain	MODBSA/D	MDSTA	SO-CDA
Email	0. 511 6	0. 541 2	0. 520 7	0. 566 6	0. 571
Netscience	0. 858 1	0. 934 6	0. 912 7	0. 959 7	0. 962
powergrid	0. 934 1	0. 775 6	0. 808 5	0. 934 5	0. 935

SO-CDA 算法依据子模函数最优化,可以获取接近全局最优的非重叠社团划分,在 Email、powergrid、Netscience 较大规模网络上均能获得最高的模块度值,MDSTA 算法模块度值稍差。CNM 算法在 powergrid 网络中基本接近最大值,但在 Netscience 网络中并无优势,Netscience 网络不连通,散点较多,网络较稀疏,Louvain 算法在 powergrid 网络中 Q 值最低,MODBSA/D 算法将社团发现问题转换为若干子问题,分别子问题进行优化,全局优化性能需要提高,因此在 3 个网络中优势不明显,由此可见,SO-CDA 算法在较大规模网络中能得到较好的社团划分结果,具有明显优势。

综合表 2、3 可知,SO-CDA 算法不仅在较大规模网络获得较好的社团划分效果,也能较好的发现网络真实社区结构,对比其他算法,表现更加稳定。

4 结 论

本文将子模优化思想引入社团划分中,提出基于子模优化的边界域处理社团发现算法 SO-CDA,以划分出更加合理的非重叠社团。算法首先依据粒计算思想,将网络进行初始粒化、聚类粒化,选择 EQ 值最大层作为最优层形成重叠的社团划分结果,依据边界域中节点与正域、负域之间的紧密关系构造子模函数,依据求解子模函数最大化选择合适子集加入到正域或负域,获取接近全局最优的非重叠社团划分。本文选取了 7 个真实网络作为测试数据集与其他社团算法进行了对比,本文算法在实验网络中都能获取较高的 NMI 值和 Q 值,边界域的重叠节点得到了更加稳定的划分。下一步将考虑进一步改进初始重叠社团的获取方法。

参考文献

- [1] NEWMAN M E J. Detecting community structure in networks [J]. *The European Physical Journal B - Condensed Matter and Complex Systems*, 2004, 38(2) : 321-330.
- [2] SU C, JIA X, XIE X, et al. A new random-walk based label propagation community detection algorithm [C]. *Proceedings of the ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 2015: 137-140.
- [3] FANUEL M, ALAIZ C M, SUYKENS J A K. Magnetic eigenmaps for community detection in directed networks [J]. *Physical Review E*, 2017, 95(2-1) : 022302.
- [4] 李国朋,潘志松,姚清,等. 融合先验信息的非负矩阵分解社区发现算法 [J]. *模式识别与人工智能*, 2016, 29(7) : 608-615.
LI G P, PAN ZH S, YAO Q, et al. Nonnegative matrix factorization algorithm with prior information for community detection [J]. *Pattern Recognition and Artificial Intelligence*. 2016, 29(7) : 608-615.
- [5] NEWMAN M E J. Fast algorithm for detecting community structure in networks [J]. *Physical Review E*, 2004, 69(6) : 066133.
- [6] CLAUSET A, NEWMAN M E J, MOORE C. Finding community structure in very large networks [J]. *Physical Review E*, 2004, 70 (6) : 066111.
- [7] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of community hierarchies in large networks [J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10) : P10008.
- [8] 何选森,何帆. 基于单源点检测的欠定混合矩阵的聚类分析 [J]. *电子测量与仪器学报*, 2019, 33 (6) : 157-164.
HE X S, HE F. Clustering analysis of underdetermined mixing matrix based on single-source-point detection [J]. *Journal of Electronic Measurement and Instrumentation*, 2019, 33(6) : 157-164.
- [9] NEWMAN M E J. Communities, modules and large-scale structure in networks [J]. *Nature Physics*, 2012, 8 (1) : 25-31.
- [10] 谢斌,彭晨,张浩,等. 基于网络攻击节点可信度的电力系统状态估计 [J]. *仪器仪表学报*, 2018, 39(3) : 157-166.
XIE B, PENG CH, ZHANG H, et al. Power system state estimation based on network attack node credibility [J]. *Chinese Journal of Scientific Instrument*, 2018, 39(3) : 157-166.
- [11] 王震,黄如意,李霁蒲,等. 一种用于故障分类与预测的多任务特征共享神经网络 [J]. *仪器仪表学报*, 2019, 40(7) : 169-177.
WANG ZH, HUANG R Y, LI J P, et al. Multi-task feature sharing neural network used for fault diagnosis and prognosis [J]. *Chinese Journal of Scientific Instrument*, 40(7) : 169-177.
- [12] YAO Y Y. Three-way decisions with probabilistic rough sets [J]. *Information Sciences*, 2010, 180(3) : 341-353.
- [13] YAO Y Y. Two semantic issues in a probabilistic rough set model [J]. *Fundamentalia Informaticae*, 2011, 108(3-4) : 249-265.
- [14] YAO Y Y. An Outline of a theory of three-way decisions [C]. *Rough Sets and Current Trends in Computing*, Springer, 2012: 1-17.
- [15] FANG H Y, LU W M, WU F, et al. Topic aspect-oriented summarization via group selection [J]. *Neurocomputing*, 2015, 149: 1613-1619.
- [16] DAS A, DASGUPTA A, KUMAR R. Selecting diverse features via spectral regularization [C]. *Advances in Neural Information Processing Systems*, 2012: 1583-1591.
- [17] WEI K, IYER R K, BILMES J A. Submodularity in data subset selection and active learning [C]. *ICML*, 2015: 1954-1963.
- [18] WU K L, CAI D, HE X F. Multi-label Active Learning Based on Submodular Functions [J]. *Neurocomputing*, 2018, 313: 436-442.
- [19] WEI K, LIU Y Z, KIRCHHOFF K, et al. Submodular subset selection for large-scale speech training data [C]. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014: 3311-3315.
- [20] JIN J Q, YAN Z A, FU K, et al. Neural network architecture optimization through submodularity and supermodularity [J]. *Machine Learning*, 2016, arXiv:1609.00074.
- [21] 赵姝,柯望,陈洁,等. 基于聚类粒化的社团发现算法 [J]. *计算机应用*, 2014, 34(10) : 2812-2815.
ZHAO SH, KE W, CHEN J, et al. Community detection algorithm based on clustering granulation [J]. *Journal of Computer Applications*, 2014, 34(10) : 2812-2815.
- [22] SHEN H W, CHENG X Q, CAI K, et al. Detect Overlapping and hierarchical community structure in networks [J]. *Physica A: Statistical Mechanics & Its Applications*, 2008, 388(8) : 1706-1712.
- [23] JIANG Z L, DAVIS L S. Submodular salient region detection [J]. *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2013: 2043-2050.

- [24] STEINHAEUSER K, CHAWLA N V. Identifying and evaluating community structure in complex networks[J]. Pattern Recognition Letters, 2010, 31(5):413-421.
- [25] LIU C, LIU J, JIANG Z. A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks[J]. IEEE Transactions on Cybernetics, 2014, 44(12): 2274-2287.
- [26] ZOU F, CHEN D B, LI S W, et al. Community detection in complex networks: Multi-objective discrete backtracking search optimization algorithm with decomposition [J]. Applied Soft Computing, 2017(53):285-295.
- [27] ZHOU X J, YANG K, XIE Y F, et al. A novel modularity-based discrete state transition algorithm for community detection in networks[J]. Neuro Computing, 2019, 334: 89-99.
- [28] 唐益明, 丰刚永, 任福继, 等. 面向结构复杂数据集的模糊聚类有效性指标[J]. 电子测量与仪器学报, 2018, 32(4): 119-127.
- TANG Y M, FENG G Y, REN F R, et al. Fuzzy clustering validity index facing data set with complexity structure [J]. Journal of Electronic Measurement and Instrumentation, 2018, 32(4): 119-127.

作者简介



杨雪洁, 2005年于安徽建筑大学获得学士学位, 2008年于安徽大学获得硕士学位, 现为合肥师范学院讲师, 主要研究方向为数据挖掘、机器学习。

E-mail: yxj1982_colour@163.com

Yang Xuejie received her B. Sc. degree from Anhui Jianzhu University in 2005, M. Sc. degree from Anhui University in 2008. Now she is a lecturer at Hefei Normal University. Her main research interests include data mining and machine learning.



曹风云, 2010年于滁州学院获得学士学位, 2013年于合肥工业大学获得硕士学位, 现为合肥师范学院讲师, 主要研究方向为计算机视觉、数字图像处理等。

E-mail: caofengyun11@126.com

Cao Fengyun received his B. Sc. degree from Chuzhou University in 2010, M. Sc. degree from Hefei University of Technology in 2013. Now he is a lecturer at Hefei Normal University. His main research interests include computer vision, digital image processing and so on.



陈洁(通信作者), 2003年于安徽师范大学获得学士学位, 2006年于安徽大学获得硕士学位, 2014年于安徽大学获得博士学位, 现为安徽大学副教授, 主要研究方向为机器学习、三支决策。

E-mail: chenjie200398@163.com

Chen Jie (Corresponding author) received her B. Sc. degree from Anhui Normal University in 2003, M. Sc. degree from Anhui University in 2006, Ph. D. degree from Anhui University in 2014. Now she is an associate professor at Anhui University. Her main research interests include machine learning and three-way decision.