

一种高效的基于教与学的社区发现算法

李佩茜, 冯少荣*

(厦门大学信息科学与技术学院 福建 厦门 361005)

摘要: 社区结构是复杂网络的重要特征,社区发现就是为了挖掘复杂网络中的社区结构.为了提高基于教与学的多目标社区发现算法(MODTLBO/D)的准确率,降低时间复杂度,提出了一种在多种群进化策略下的 MODTLBO/D (E-MODTLBO/D).在 E-MODTLBO/D 中,采用自适应学习因子加强在教学阶段的探索与搜索能力;在学习阶段,每个个体在各自的子种群内采用随机学习策略或者是改进的量子行为学习策略.在每次迭代更新后,子种群间进行信息交流,维持算法的多样性与避免早熟收敛.实验表明,E-MODTLBO/D 在时间复杂度与发现高质量的社区结构方面要优于 MODTLBO/D 等一些经典社区发现算法.

关键词: 社区发现; 教与学; 多目标; 多种群进化算法

中图分类号: TP 181

文献标志码: A

文章编号: 0438-0479(2018) 01-0104-08

现实世界有许多复杂网络体系,例如社交网络、生物网络、万维网等.在数学与计算机领域中,复杂网络被定义为图,包括一系列的节点与连接节点的边,节点代表个体,边代表个体与个体之间的联系.一般而言,社区是指复杂网络中联系较为紧密的子网络^[1],社区发现就是为了找到复杂网络的内部特征,便于更加了解复杂网络中的内在信息.社区发现问题通过发现节点之间的相似度或其他相关属性,对复杂网络进行社区划分.它可以被看作是一种最优化问题^[2].

基于教与学的最优化算法(TLBO)^[3]是一种种群智能优化算法,分为两个阶段:1) 学生向老师学习(称为教学阶段);2) 学生之间的相互学习(称为学习阶段).通过这两个阶段的共同作用提高学生成绩.TLBO 具有简单、易于理解、算法参数少、优化精度高和收敛速度快等优点,广泛运用于各个领域中的最优化问题^[4-7].Keesari 等^[8]用 TLBO 解决车间作业调度问题,Xia 等^[9]提出简化的 TLBO 解决拆分序列问题,Baykasoglu 等^[10]分别分析了 TLBO 在流水车间(FSSP)与车间作业调度(JSSP)的表现.与此同时,Dede^[11]把 TLBO 运用于离散化的桁架结构最优化问题,Li 等^[12]提出离散化 TLBO 解决流水车间调度问题.

从本质上来看,复杂网络中的社区发现问题是一种聚簇最优化问题.Chen 等^[13]提出了一种基于教与学的多目标社区发现算法(MODTLBO/D)来解决社区发现问题.该算法中,每个个体从邻居的平均值处进行学习,算法时间复杂度高,且在学习阶段,个体仅仅从邻居处进行学习,容易陷入局部最优.因此,本文中提出一种在多种群进化策略下的 MODTLBO/D (E-MODTLBO/D).采用多种群并进的进化策略,在不同种群中采用不同的进化规则.在教学阶段,采用自适应学习因子;在学习阶段,每个个体在各自的子种群内可以采用随机学习策略或者是改进的量子行为学习策略.在每次迭代更新后,子种群间进行信息交流,维持算法的多样性与避免早熟收敛.把改进后的算法带入真实网络中进行实验发现,本文中提出的 E-MODTLBO/D 算法在时间复杂度与发现高质量的社区结构方面比 MODTLBO/D 等算法表现更好.

1 社区发现

1.1 相关工作

近年来发现,很多现实世界的复杂系统可以由复

收稿日期: 2017-03-22 录用日期: 2017-11-30

基金项目: 国家社会科学基金重大项目(13&ZD148)

* 通信作者: shaorong@xmu.edu.cn

引文格式: 李佩茜,冯少荣.一种高效的基于教与学的社区发现算法[J].厦门大学学报(自然科学版),2018,57(1): 104-111.

Citation: LI P X, FENG S R. An Efficient multi-population community detection algorithm using teaching-learning-based optimization [J]. J Xiamen Univ Nat Sci, 2018, 57(1): 104-111. (in Chinese)



<http://jxmu.xmu.edu.cn>

杂网络来表示,发现复杂网络中的社区结构问题就是社区发现问题,该问题已经越来越受到人们的广泛关注。许多来自不同领域的算法被提出来解决社区发现问题,这些算法大致可以分为以下 3 类:

1) 基于图划分的算法:在这类算法中,节点通过预定义的方法被划分到不同社区,使得社区之间的连接尽可能少。Ezhilarasi 等^[14]提出一种基于图划分的算法,该算法定义社区之间连接与内在连接差异的评价函数,旨在最小化评价函数。Pothen 等^[15]利用拉普拉斯矩阵的特征进行社区划分,实现社区发现。但这两种方法都需要预先定义划分准则,最小化划分准则将会导致划分效果不佳。因此,又出现了一些其他准则,例如平均划分^[16]、比率划分^[17]、标准化划分^[18]等。基于图划分的算法存在需预定义社区大小等规则的缺点,这表明该种算法并不是理想的社区发现算法。

2) 基于分层的算法:这类算法的核心思想是运用相似度量不同节点对之间的相似程度来解决社区发现问题。分层算法包括凝聚分层算法和分裂分层算法。典型的凝聚分层算法是标签传播算法(LPA)^[19]。典型的分裂分层算法是由 Girvan 与 Newman 提出的 GN 算法^[20]。这些方法自然生成了网络的社区划分,但是它们需要一个停止的规则。

3) 基于模块度的算法:此类算法在社区发现问题中被广泛运用。模块度 Q 是由 Newman 等^[1]提出的用来评价社区划分质量的评价函数。Fast Newman (FN) 算法^[21]是一个经典的模块度算法,该算法从一些独立的节点开始,用贪心策略将原始图的连接信息反复加入到图中,获得最大可能的模块度增长。Blondel 等^[22]提出一种快速多步贪心策略(BGLL)来实现模块度最优,找到最优社区划分。除此之外,有一些单目标进化算法^[23-27]被用来最优化模块度 Q 。由于单目标算法存在分辨率限制等缺陷,还有许多多目标进化算法被提出,其中包括 Shi 等^[28]提出的多目标社区发现算法(MOCD)、Pizzuti 等^[29]提出的多目标遗传算法(MOGA-Net)、Gong 等^[30]提出的基于分解的多目标遗传算法(MOEA/D-Net)。与此同时,Gong 等^[31]还提出一种基于分解的多目标离散粒子群智能算法来解决社区发现问题。除此之外,还有一些多目标进化算法解决社区发现问题^[32-36],它们都是同时最优化两个互相矛盾的目标函数实现社区发现。

1.2 社区定义

一个复杂网络可以抽象为一个图 $G=(V,E)$,其中 V 为节点的集合, E 为连接节点的边的集合。在复杂网络中的社区是指具有密切关系的节点的集合^[37],也

就是说在社区内部节点连接紧密,而在社区之间节点连接稀疏。复杂网络中的社区发现问题是一个聚类问题,旨在把图 G 划分成若干子图 $G_i(i=1,2,\dots,k)$,满足 $\cup_{1 \leq i \leq k} G_i = G$ 且 $\cap_{1 \leq i \leq k} G_i = \phi$ 。

1.3 评价指标

为了找到更好的社区划分与评价社区划分的质量,必须采用恰当的质量评价指标。本文中采用标准互换信息(normalized mutual information, NMI)与模块度 Q 这两个最常见的评价指标。近些年来,这两个评价指标被广泛运用于社区发现问题。

由 Danon 等^[38]提出的 NMI 是用于评价所得社区划分与真实网络社区划分相似度的评价指标,给出一个网络的两种划分 p_1 与 p_2 , C 为混淆矩阵, C_{ij} 代表 p_1 中属于社区 i 的节点也属于 p_2 中属于社区 j 的节点的数目总和, $NMI(p_1, p_2)$ 为:

$$NMI(p_1, p_2) = \frac{-2 \sum_{i=1}^{C_p} \sum_{j=1}^{C_p} \log(C_{ij}N/C_iC_j)}{\sum_{i=1}^{C_p} C_i \log(C_i/N) + \sum_{j=1}^{C_p} C_j \log(C_j/N)}$$

其中, C_p (C_p) 表示 p_1 (p_2) 中的社区数目, C_i (C_j) 表示混淆矩阵 C 的第 i 行(第 j 列)的元素和, N 指节点总数。如果 $p_1=p_2$, 则 $NMI(p_1, p_2)=1$, 如果 p_1 与 p_2 完全不同, 则 $NMI(p_1, p_2)=0$ 。NMI 值越大代表 p_1 与 p_2 的相似度越高。

由 Newman 等^[1]提出的模块度 Q 是一个非常受欢迎的用来评价社区划分质量的评价函数。假设 K 为复杂网络中的社区划分数目, 则模块度 Q 值定义如下:

$$Q = \sum_{s=1}^K \left[\frac{l_s}{m} - \left(\frac{d_s}{2m} \right)^2 \right],$$

其中 m 表示网络中的边总数, l_s 与 d_s 分别表示社区 S 中的边总数与节点总数。当 Q 值越接近 1, 表示网络的社区划分质量越好。在实际例子中, Q 值一般为 0.3 ~ 0.7。

2 预备知识

在 E-MODTLBO/D 中,由于学习规则的改变,本文中需要重新定义 Chen 等^[13]提到的一些变量。

定义 1 设 $Y=(y_1, y_2, \dots, y_N)$, $X=(x_1, x_2, \dots, x_N)$, 则定义函数 $Y=S(X)$ 如下:

$$y_i = \begin{cases} 1, & \text{当 } \text{rand}(0, 1) < \text{sigmoid}(x_j), \\ 0, & \text{其他,} \end{cases}$$

<http://jxmu.xmu.edu.cn>

其中 $j = 1, 2, \dots, N$ sigmoid 函数^[39] 定义如下:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$$

定义2 假设 $Y = (y_1, y_2, \dots, y_N)$, $X = (x_1, x_2, \dots, x_N)$, $Z = (z_1, z_2, \dots, z_N)$, 则定义函数 $Z = XY$ 如下:

$$z_j = \begin{cases} x_j, & y_j = 0, \\ S_{\text{Best}_j}, & \text{其他}, \end{cases}$$

其中 S_{Best_j} 是代表个体所有邻居中第 j 个节点的最多标签数.

3 改进的 E-MODTLBO/D

3.1 离散化个体表示

本算法采用整数值编码规则^[13], 个体向量代表社区划分结果. 假设一个节点的社区值与另一个节点的社区值相同, 代表着两个节点属于同一社区, 反之则属于不同社区. 假设网络共有 C 个社区, N 个节点, 则第 i 个个体向量 X_i 定义如下:

$$X_i = (x_{i1}, x_{i2}, \dots, x_{iN}),$$

其中 $x_{ij} (1 \leq j \leq N, 1 \leq i \leq p)$ 代表第 i 个个体的第 j 个节点所属的社区值, p 为种群大小. 若 $x_{ij} = x_{ik}$, 则第 i 个个体的第 j 个节点与第 k 个节点属于同一社区内. 除此之外, 本算法采用 Gong 等^[40] 提出的一种启发式算法获得初始个体向量.

3.2 多种群进化策略

本算法采用一种改进的多种群进化策略^[47]. 种群内所有个体以相同概率被均等地分到各个子种群内, 不同子种群内部进化规则不同, 以维持算法的探索能力与提高算法多样性. 在每一轮迭代中, 每个个体通过教学阶段与学习阶段在各个子种群内进行学习更新, 子种群间进行信息交流, 把种群内部表现最优的个体传给其他子种群, 维持种群的多样性, 避免陷入局部最优. 多种群进化策略不仅可以提高算法的多样性, 还可以降低时间复杂度.

3.3 自适应教学因子

在 Chen 等^[13] 提出的 MODTLBO/D 中, 教学因子 (T_{TF}) 决定平均值的改变程度与平均值对生成新解的影响程度, 其值随机取为 1 或 2, 分别反映学员什么也没学到, 或者是学到老师的全部知识两种情况. 但在实际教学中, 情况随着学习时间的推移有所不同. 在教学阶段的前期, 学员与教师之间的知识存在较大差距, 这是由于此时学员对所学知识掌握很少, 所以学员学习效率较高, 学到的知识多而且快; 而在教学阶段的

后期, 经过一段时间的学习, 学员掌握的知识越来越多, 与教师之间的差距越来越小, 学习效率会显著下降, 学到的知识少而且慢.

在优化算法中, T_{TF} 值小, 则代表算法探索能力强, 但搜索能力弱; 反之则代表算法探索能力弱, 但搜索能力强. 考虑到此问题, 本算法参考 2015 年 Yue 等^[41] 提出的自适应教学因子, 改进本算法的教学因子, 使得 T_{TF} 随着迭代次数的增加而呈线性递减, 表达式如下:

$$T_{TF}(t) = \frac{1}{2} \left[(T_{TF\max} - T_{TF\min}) \left(\frac{t_{\max} - t_i}{t_{\max}} \right)^2 + (T_{TF\max} - T_{TF\min}) \left(\frac{t_{\max} - t_i}{t_{\max}} \right) \right] + T_{TF\min},$$

其中 $T_{TF\max}$ 与 $T_{TF\min}$ 分别表示 T_{TF} 的最大值与最小值, t_{\max} 表示算法的迭代总次数, t 表示当前迭代次数. 从上式可以看出, 在搜索前期, T_{TF} 值大; 在搜索后期, T_{TF} 值小. 这表明在算法搜索前期, 学员学习效率高, 学到的知识多且快, T_{TF} 值大有利于增强算法的全局探索能力, 加快算法的收敛速度. 随着迭代不断进行, 学习效率下降, 学到的知识少且慢, 此时 T_{TF} 值小有利于增强算法的局部探索能力, 使算法搜索逐步向最优解靠拢, 获得更高精度的解.

3.4 E-MODTLBO/D 教学阶段

在实际教学中, 教师给学生传播知识提高全班的知识水平, 帮助学生得到较好的分数, 因此全班的平均成绩也可提高, 个体基于全班平均成绩与教师进行更新. 本算法把每个个体分配到不同种群内, 个体从种群内最优个体 (S_{Teacher_i}) 与平均值 S_{Mean_i} 学习更新.

S_{Teacher_i} 与 S_{Mean_i} 定义如下:

$$S_{\text{Teacher}_i} = B_{\text{Best}}(X_{i1}, X_{i2}, \dots, X_{ik}, \dots, X_{iN}),$$

$$S_{\text{Mean}_i} = \left(\sum_{k=1}^K X_{ik1}, \sum_{k=1}^K X_{ik2}, \dots, \sum_{k=1}^K X_{ikD} \right),$$

其中 $X_{i1}, X_{i2}, \dots, X_{ik}, \dots, X_{iN}$ 代表个体 X_i 所在种群的个体集合, 且 $X_{ik} = (x_{ik1}, x_{ik2}, \dots, x_{ikD})$.

在教学阶段, 教师的任务是提高全班的成绩, 通过对最优个体的学习来提高整体水平, 因此, 对社区发现问题重新定义对第 i 个个体 X_i 的更新规则:

$$\text{new}X_i = X_i \oplus \text{Difference}_i,$$

$$\text{Difference}_i = S(\text{rand}(0, 1) \times (S_{\text{Teacher}_i} \oplus (T_{TF}(t) \times S_{\text{Mean}_i}))),$$

其中, $\text{rand}(0, 1)$ 是 0 与 1 之间的随机数, $T_{TF}(t)$ 是前文 3.3 节提到的自适应教学因子, (S_{Teacher_i}) 与 S_{Mean_i} 分别是个体 X_i 所在种群的最优个体与平均值, \oplus 是异或算子.

3.5 E-MODTLBO/D 学习阶段

在学习阶段,学生随机向种群内的其他学生学习知识,学生从其他学生那里学习他们还未掌握的新知识。在这一阶段,个体通过以下规则更新:

$$\text{new}X_i = X_i \theta \text{ Difference}_i,$$

$$\text{Difference}_i = S(X_i \oplus X_j),$$

其中 \oplus 是异或算子, X_j 代表 X_i 所在种群内部的任意个体,这种策略被称为随机学习策略。

为了提高算法的表现效果,本算法提出一种改进的量子行为学习策略^[42]。在学习阶段,每个个体在各自的子种群内可以采用随机学习策略或者是改进的量子行为学习策略。给出学习概率 p_l , 决定个体在学习阶段采用的学习策略。随机产生一个介于 0 与 1 之间的随机数 r , 当 $r < p_l$ 时,采用传统的随机学习策略;否则采用改进的量子行为学习策略。事实上, p_l 值越大则算法中的个体多样性越好, p_l 值越小则算法的收敛速度越快。

假设产生的新个体 $\text{new}X_i = (z_1, z_2, \dots, z_N)$, 改进的量子行为学习策略规则如下:

$$z_j = \begin{cases} S_{\text{Teacher}_j}, & \text{当 } \text{rand}(0, 1) < \text{sigmoid}(\text{Temp}X_j), \\ G_{\text{Teacher}_j}, & \text{其他}. \end{cases}$$

其中 S_{Teacher} 与 G_{Teacher} 分别表示个体 X_i 所在种群内的最优个体与全局最优个体,且 $\text{Temp}X$ 定义如下:

$$\text{Temp}X = \beta \times (X \oplus S_{\text{Mean}_i}) \times (-\log(u)),$$

其中 μ 为介于 0 与 1 之间的随机值, S_{Mean_i} 是个体 X_j 所在种群内的平均值。在传统的学习阶段,个体只是在种群中随机选择个体进行学习,而本算法中,个体有两种不同的学习策略。在一定概率下,个体采用传统的随机学习策略;但在一定概率下,个体通过向全局最优个体与种群内最优个体学习知识,使得个体可以在其他个体、全局最优个体与种群内最优个体的共同帮助下学习,使学习效率明显提高,加快算法收敛速度。

3.6 变异规则

当算法中存在完全相同的两个解,则必须修改它们以避免陷入局部最优,维持种群多样性。采用 Gong 等^[40]提出的基于邻居的变异算子,对每个个体随机生成一个介于 0 与 1 之间的数,如果小于变异概率 p_m , 则执行变异算子。这里需要指出每次教学阶段或者学习阶段完成后,都需进行一次变异操作。

3.7 算法框架

3.7.1 基于分解的多目标优化算法(MOEA/D)

MOEA/D^[43]是基于传统的多目标最优化策略的

算法,该算法把多目标优化问题分解为一系列单目标子问题,同时最优化这些子问题,每个子问题通过其邻居子问题进行优化,而子问题的邻居关系通过子问题之间的距离与聚合多目标权重向量定义。不同于传统的帕累托分级方法,该算法通过实现每个子问题的最优化来获得最优解。

在 MOEA/D 中,有许多方法可以把多目标最优化算法分解为一系列子问题,其中一种是通过构造聚合函数实现分解。至今为止,有许多方法可以构造聚合函数,其中最受欢迎的是权重求和法与切比雪夫法。本算法采用切比雪夫法,因为所研究的问题中两个目标函数是不连续的,所以不能简单判断帕累托前沿(PF)是不是凹的。如果 PF 是凸的,权重求和法的效果就会不佳。切比雪夫法把问题分割为一些具有单目标子问题^[43],最优化问题可以表示为

$$\text{ming}(x | \lambda, z^*) = \max_{1 \leq i \leq m} \lambda_i |f_i(x) - z_i^*|,$$

其中权重向量 $\lambda = (\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_m)$ 且 $\lambda_i \geq 0, i = 1, 2, \dots, m$ 且 $\sum_{i=1}^m \lambda_i = 1$, z^* 代表参考点, m 是目标函数的数目。

3.7.2 目标函数

采用 Gong 等^[30]提出的两个目标函数 NRA 与 RC, 目标是最小化这两个函数。定义一个无向图 $G = (V, E)$, 定义邻接矩阵 A , 一种划分方法 $S = (V_1, V_2, \dots, V_m)$, V_i 是社区 G_i 的节点集合 $i = 1, 2, \dots, m$ 。定义 $L(V_1, V_2) =$

$$\sum_{i \in V_1, j \in V_2} A_{ij}, L(V_1, V_2) = \sum_{i \in V_1, j \in V_2} A_{ij}.$$

$$\begin{cases} \text{min}f_1 = \text{NRA} = - \sum_{i=1}^m \frac{L(V_i) V_i}{|V_i|}, \\ \text{min}f_2 = \text{RC} = - \sum_{i=1}^m \frac{L(V_i) V_i}{|V_i|}. \end{cases}$$

一般而言, NRA 可看作是社区内部连接的密度, RC 可看作是社区之间连接的密度。最小化 NRA 趋向于把网络切割成很多的小型社区,而最小化 RC 趋向于把网络切割成大型社区,这两个目标函数有着权衡对方对于社区数目增加或者减少的趋势的潜力。通过权衡最小化这两个目标函数,可以获得高质量的网络划分。

3.7.3 E-MODTLBO/D 伪代码

本文中采用 MOEA/D 的主要框架,加入多种群进化策略,在每个种群内部,个体通过种群内部信息进行更新进化。因此 N_{Teacher_i} 与 N_{Mean_i} 代表第 i 个个体所在种群内部最优值与平均值。E-MODTLBO/D 的伪代码如下:

<http://jxmu.xmu.edu.cn>

算法: E-MODTLBO/D

输入: 初始化个体 X , 个体数目 N , 迭代次数 gen_{max} , 变异概率 p_m , 学习概率 p_l , 种群数目 S_p , 种群内个体数目 S_N , 权重向量 W , 初始化最优参考点 z^* ;

输出: 最优个体

Begin

for $t = 1: gen_{max}$

把个体 X 随机分入不同种群内, 更新全局最优个体与种群内最优个体;

for 每个种群 do

通过教学阶段与学习阶段更新每个个体 X_i , 计算个体与全局最优参考点的距离 z^* ;

保留每个种群内表现最优的 S_N 个个体;

更新全局最优参考点 z^* ;

end for

把种群内最优个体传给其他种群, 并保留前 S_N 个个体在种群内;

end for

End

4 实验结果

4.1 参数设置

所有实验都在 2.3 GHz CPU, 4 GB RAM 的处理器上运行, 操作系统为 Windows7, 软件版本为 MATLAB 8.3. 为了减少系统误差, 每个函数都被单独运行了 20 次. 对于所有的方法, 种群大小 popsize 设置为 100, 更新迭代次数 $g_{gen_{max}}$ 设置为 100, 变异概率 $p_m = 0.06$, 种群数目设置为 20, 种群内个体数目为 5. 为了维持算法多样性与收敛性的平衡, 设置学习概率 $p_l = 0.5$.

4.2 真实网络

采用 Zachary 空手道俱乐部网络、Bottlenose 海豚网络、美国大学足球队网络和 Krebs 美国政治书网络这 4 个真实的复杂网络对本算法进行验证.

空手道俱乐部网络由 Zachary 提出^[44], 通过观察在一段时期内空手道俱乐部 34 名成员的活动所获得. 在研究期间, 俱乐部管理者与俱乐部开发者之间存在分歧, 这个分歧最终导致开发者离开并且成立新的俱乐部; 与此同时, 开发者带走原俱乐部近 1/2 成员, 网络自然分成了两个大小相近的社区.

Bottlenose 海豚网络由 62 只海豚构成^[45], 海豚们住在新西兰的一个神奇峡湾中, 它们的行为活动被

Lusseau 等学者所观察. 学者们花了 7 年时间观察它们的行为活动, 定义如果有两只海豚发生固定且频繁的协作行为, 则认为这两只海豚之间有联系. 这个网络中的海豚自然的分成 2 组, 共有 159 组联系.

美国大学足球队网络来自于对美国大学足球队员的观察^[20], 网络代表在 Fall 2000 常规赛赛季中不同球队之间的比赛场次, 网络中的节点代表足球队, 而边则代表两个足球队会进行比赛. 比赛队伍会被分到几个不同的联盟中去, 每队平均进行 4 场联盟间比赛与 7 场联盟内比赛, 该网络包含 115 个节点与 616 条边.

Krebs 的政治书网络中, 节点代表 105 本购自亚马逊的关于美国政治的书, 边代表经常由同一个买家买的书之间具有关联关系. 不同的书按不同的政治取向被分类^[46].

4.3 实验结果分析

为了评价本算法在复杂网络的社区发现问题中的表现, 把本算法与其他一些经典的算法进行比较.

首先, 与两个经典的单目标算法: FM 算法^[21] (采用贪心策略实现模块度 Q 最优), BGLL 算法进行比较, 结果如表 1 所示.

表 1 4 个真实网络的模块度 Q 值
Tab. 1 The modularity values obtained on four real-world networks

单目标算法	Zachary 空手道 俱乐部	Battlenose 海豚	美国大学 足球队	Krebs 美国 政治书
FM	0.380 79	0.494 80	0.581 40	0.512 36
BGLL	0.421 08	0.518 79	0.604 38	0.518 64
E-MODTLBO/D	0.419 79	0.519 04	0.604 47	0.525 43

从表 1 可以看出: 在 Zachary 空手道俱乐部网络中, BGLL 算法表现最优, 可以获得最大的模块度 Q 值为 0.421 08; 而对于另外 3 个网络, E-MODTLBO/D 算法表现最优, 比其他两个单目标算法表现都要好.

其次, 把 E-MODTLBO/D 与一些多目标进化算法进行比较, 分别为 MOCD^[28]、MOGA-Net^[29]、MOEA/D-Net^[30]、MODPSO^[31] 和 MODTLBO/D^[13], 所得结果如表 2 所示.

从表 2 可以看出: 从模块度 Q 值的角度比较, 在 Zachary 空手道俱乐部网络中, E-MODTLBO/D 与 MOEA/D-Net、MODPSO、MODTLBO/D 表现同样好; 但在另外 3 个网络中, E-MODTLBO/D 比其他算法表现

<http://jxmu.xmu.edu.cn>

表 2 4 个真实网络的模块度 Q 值与标准互换信息 NMI

Tab. 2 Results obtained on four real-world networks

算法	Q				NMI			
	Zachary 空手道俱乐部	Battlenose 海豚	美国大学足球队	Krebs 美国政治书	Zachary 空手道俱乐部	Battlenose 海豚	美国大学足球队	Krebs 美国政治书
MOCD	0.418 76	0.521 95	0.595 86	0.525 74	0.837 18	1.000 00	0.893 32	0.594 98
MOGA-Net	0.414 94	0.514 17	0.526 48	0.517 42	0.837 18	1.000 00	0.801 16	0.594 19
MOEA/D-Net	0.419 79	0.521 18	0.603 98	0.526 55	1.000 00	1.000 00	0.926 79	0.596 42
MODPSO	0.419 79	0.526 05	0.604 55	0.526 62	1.000 00	1.000 00	0.928 85	0.615 83
MODTLBO/D	0.419 79	0.522 01	0.604 55	0.526 72	1.000 00	1.000 00	0.926 68	0.617 27
E-MODTLBO/D	0.419 79	0.526 80	0.604 57	0.526 94	1.000 00	1.000 00	0.924 20	0.677 10

更优.从 NMI 这个评价指标来看,在 Zachary 空手道俱乐部网络中,E-MODTLBO/D 与 MOEA/D-Net、MODPSO、MODTLBO/D 表现同样好;在 Battlenose 海豚网络中,所有算法的 NMI 值都可以达到 1;在美国大学足球队网络中,MODPSO 表现最优,NMI 值为 0.928 85,本算法的 NMI 值为 0.924 20.在 Krebs 美国政治书网络中,本算法表现最优,可以获得最大的 NMI 值,为 0.677 10,远大于其他算法.

从与这些算法比较中更可以看出,本算法在解决社区发现问题时是很有潜力的.在 NMI 值上来看,它可以在 Zachary 空手道俱乐部网络、Bottlenose 海豚网络与 Krebs 美国政治书网络中都表现优于其他算法;从 Q 值来看,它在 4 个网络中的表现都优于其他算法.

5 结 论

本文中提出了一种高效的在多种群进化策略下的 E-MODTLBO/D.不同于 MODTLBO/D,本算法采用多种群进化策略来降低时间复杂度并且增加算法的多样性.种群内所有个体以相同概率被均等地分到各个子种群内,不同子种群内部进化规则不同,以维持算法的探索能力.每个个体在教学阶段与学习阶段在各个种群内进行学习更新,每次迭代后,种群间进行信息交流,把种群内部表现最优的个体传给其他种群,以维持种群的多样性,避免陷入局部最优.在教学阶段,每个个体从种群内学习,同时采用自适应教学因子.在学习阶段,每个个体在各自的子种群内采用随机学习策略或者是改进的量子行为学习策略,以保证算法的收敛性与多样性.在每代更新完成后,子种群之间的信息交流可以维持算法的多样性与避免早熟收敛.

为了验证提出该算法的有效性,把该算法运用于不同的真实网络中,并与其他算法进行比较.实验结果表明,该算法在时间复杂度与发现高质量的社区结构方面要优于 MODTLBO/D 等一些经典社区发现算法,在解决复杂网络中的社区发现问题方面高效且有广泛应用前景.

为了让该算法更加高效,未来的工作将致力于研究基于特殊问题的更新策略与如何维持多样性与收敛性的方法.进而将离散性基于教与学最优化的方法运用于限制性的、动态的复杂网络中,这有助于解决各种各样的真实复杂网络的最优化问题.

参考文献:

- [1] NEWMAN M E J, GIRVAN M M. Finding and evaluating community structure in networks [J]. Phys Rev E, 2004, 69 (2): 99-106.
- [2] NEWMAN M E J. Fast algorithm for detecting community structure in networks [J]. Phys Rev E, 2004, 69 (6): 066133.
- [3] RAO R V, SAVSANI V J, VAKHARIA D P. Teaching-learning-based optimization: an optimization method for continuous non-linear large scale problems [J]. Information Sciences, 2012, 183(1): 1-15.
- [4] YU K, WANG X, WANG Z. Constrained optimization based on improved teaching-learning-based optimization algorithm [J]. Information Sciences, 2016, 352/353: 61-78.
- [5] GHASEMI M, GHAVIDEL S, GITIZADEH M, et al. An improved teaching-learning-based optimization algorithm using Levy mutation strategy for non-smooth strategy for non-smooth power flow [J]. International Journal of Electrical Power & Energy Systems, 2015, 65: 375-384.
- [6] BASU M. Teaching-learning-based optimization algorithm for multi-area economic dispatch [J]. Energy, 2014, 68: 21-28.

<http://jxmu.xmu.edu.cn>

- [7] PATEL V K ,SAVSANI V J. A multi-objective improved teaching-learning-based optimization algorithm (MO-ITLBO) [J]. *Information Sciences* 2014 ,357: 182-200.
- [8] KEESARI H S ,RAO R V. Optimization of job shop scheduling problems using teaching-learning-based optimization algorithm [J]. *Opsearch* 2013 ,51(4) : 545-561.
- [9] XIA K ,GAO L ,LI W D ,et al. Disassembly sequence planning using a simplified teaching-learning-based optimization algorithm [J]. *Advanced Engineering Informatics* , 2014 ,28: 518-527.
- [10] BAYKASOGLU A ,HAMZADAYI A ,KOSE S Y. Testing the performance of teaching-learning-based optimization (TLBO) algorithm on combinatorial problems: Flow shop and job shop scheduling cases [J]. *Information Sciences* , 2014 ,276(C) : 204-218.
- [11] DEDE T. Application of teaching-learning-based-optimization algorithm for the discrete optimization of truss structures [J]. *Ksce Journal of Civil Engineering* ,2014 ,18(6) : 1759-1767.
- [12] LI J ,PAN Q ,MAO K. A discrete teaching-learning-based optimization algorithm for realistic flow shop rescheduling problems [J]. *Engineering Applications of Artificial Intelligence* 2015 ,37: 279-292.
- [13] CHEN D ,ZOU F ,LU R Q , et al. Multi-objective optimization of community detection using discrete teaching-learning-based optimization & with decomposition [J]. *Information Sciences* 2016 ,369: 402-418.
- [14] EZHILARASI G A ,SWARUP K S. Network decomposition using Kernighan-Lin strategy aided harmony search algorithm [J]. *Swarm & Evolutionary Computation* ,2012 (7) : 1-6.
- [15] POTHEN A ,SIMON H ,LIU K P. Partitioning sparse matrices with eigenvectors of graphs [J]. *Siam Journal on Matrix Analysis & Applications* ,1990 ,11(3) : 430-452.
- [16] FIEDLER M. A property of eigenvectors of non-negative symmetric matrices and its applications to graph theory [J]. *Czechoslovak Mathematical Journal* ,1975 ,25: 619-637.
- [17] WEI Y C ,CHENG C K. Ration cut partitioning for hierarchical designs [J]. *IEEE Transactions on Computer-Aided Design* ,1991 ,10(7) : 911-921.
- [18] HE L ,ZHANG H. Iterative ensemble normalized cuts [J]. *Pattern Recognition* 2015 ,52: 274-286.
- [19] ZHANG X ,FEI S ,SONG C , et al. Label propagation algorithm based on local cycles for community detection [J]. *International Journal of Modern Physics B* ,2015 ,29(5) : 1550029.
- [20] GIRVAN M ,NEWMAN M E J. Community structure in social and biological networks [J]. *Proceedings of the National Academy of Sciences of the United States of America* 2002 ,99(12) : 7821-7826.
- [21] CLAUSET A ,NEWMAN M E J ,MOORE C. Finding community structure in very large networks [J]. *Physical Review E* 2004 ,70(2) : 066111.
- [22] BLONDEL V D ,GUILAUME J L ,LAMBIOTTE R ,et al. Fast unfolding of communities in large networks [J]. *Journal of Statistical Mechanics: Theory and Experiment* ,2008 ,10: P10008.
- [23] CRUZ J D ,BOTHOREL C ,POULET F. Entropy based community detection in augmented social networks [C]// *International Conference on Computational aspects of Social Networks*. Salamanca: IEEE ,2011: 163-168.
- [24] HASSAN E A ,HAFAZ A I ,HASSANIEN A E , et al. Community detection algorithm based on artificial fish swarm optimization [J]. *Advances in Intelligent Systems & Computing* 2014 ,323: 509-521.
- [25] KARIMI-MAID A M ,FATHIAN M ,AMIRI B. A hybrid artificial immune network for detecting communities in complex networks [J]. *Computing* 2015 ,97(5) : 483-507.
- [26] LI Y ,LIU J ,LIU C. A comparative analysis of evolutionary and memetic algorithms for community detection from signed social networks [J]. *Soft Computing* ,2014 ,18(2) : 329-348.
- [27] LI Z ,LIU J. A multi-agent genetic algorithm for community detection in complex networks [J]. *Physica A Statistical Mechanics & Its Applications* 2016 ,449: 336-347.
- [28] SHI C ,YAN Z ,CAI Y , et al. Multi-objective community detection in complex networks [J]. *Applied Soft Computing* , 2012 ,12(2) : 850-859.
- [29] PIZZUTI C. A multi-objective genetic algorithm to find communities in complex networks [J]. *IEEE Transactions on Evolutionary Computation* 2012 ,6(3) : 418-430.
- [30] GONG M G ,MA L ,ZHANG Q F , et al. Community detection in networks by using multi-objective evolutionary algorithm with decomposition [J]. *Physica A Statistical Mechanics & Its Application* 2012 ,391: 4050-4060.
- [31] GONG M G ,CAI Q ,CHEN X , et al. Complex network clustering by multi-objective discrete particle swarm optimization based on decomposition [J]. *IEEE Transactions on Evolutionary Computation* 2014 ,18(1) : 82-97.
- [32] AMIRI B ,HOSSAIN L ,CRAWFORD J W , et al. Community detection in complex networks: multi-objective enhanced firefly algorithm [J]. *Knowledge-Based Systems* ,2013 ,46: 1-11.
- [33] CAI Q ,GONG M G ,MA L , et al. Greedy discrete particle swarm optimization for large-scale social network clustering [J]. *Information Sciences* 2015 ,316(41) : 503-516.

- [34] LIU C ,LIU J ,JIANG Z.A multi-objective evolutionary algorithm based on similarity for community detection from signed social networks [J].IEEE Transactions on Cybernetics 2014 44(12) : 2274-2287.
- [35] ZHAO Y ,LI S ,JIN F.Overlapping community detection in complex networks using multi-objective evolutionary algorithm [J]. Computational & Applied Mathematics , 2015 36(1) : 1-20.
- [36] ZHOU X ,LIU Y ,LI B ,et al.Multi-objective biogeography based optimization algorithm with decomposition for community detection in dynamic networks [J].Physica A: Statistical Mechanics & Its Applications 2015 436: 430-442.
- [37] RADICCHI F ,CASTELLANO ,CECCONI F ,et al.Defining and identifying communities in networks [J].Proceedings of the National Academy of Sciences of the United States of America 2004 101(9) : 2658-2663.
- [38] DANON L ,DAZ-GUILERA A ,DUCH J ,et al.Comparing community structure identification [J].Journal of Statistical Mechanics Theory & Experiment 2005 78: P09008.
- [39] KENNEDY J ,EBERHART R C.A discrete binary version of the particle swarm algorithm [J].Computational Cybernetics and Simulation 1997 5: 4104-4108.
- [40] GONG M G ,CAI Q ,LI Y ,et al.An improved memetic algorithm for community detection in complex networks [C]// 2012 IEEE Congress on Evolutionary Computation. Brisbane: IEEE 2012: 1-8.
- [41] YUE Z ,GAO Y.An improved teaching-learning-based optimization algorithm [J]. Journal of Lanzhou University of Technology 2015 41(6) : 99-103.
- [42] ZOU F ,WANG L ,HEI X ,et al.Teaching-learning-based optimization with dynamic group strategy for global optimization [J].Information Sciences 2014 273: 112-131.
- [43] ZHANG Q F ,LI H. MOEA/D: a multi-objective evolutionary algorithm based on decomposition [J]. IEEE Transactions on Evolutionary Computation ,2017 ,11 (6) : 712-731.
- [44] ZACHARY W W. An information flow model for conflict and fission in small groups [J].Journal of Anthropological Research ,1997 33(4) : 452-473.
- [45] LUSSEAU D ,SCHNEIDER K ,BOISSEAU O J ,et al.The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations [J].Behavioral Ecology and Sociobiology 2003 54(4) : 396-405.
- [46] NEWMAN M E J. Modularity and community structure in networks [J].Proceedings of the National Academy of Science 2006 103 (23) : 8577-8582.
- [47] ROSINC C ,BELEW R ,MORRIS G ,et al.New methods for competitive co-evolution [J]. Evolutionary Computation , 1997 5(1) : 1-29.

An Efficient Multi-population Community Detection Algorithm Using Teaching-learning-based Optimization

LI Peixi ,FENG Shaorong*

(School of Information Science and Engineering ,Xiamen University ,Xiamen 361005 ,China)

Abstract: Community structure is an important feature of complex networks and community detection aims at mining the community structure of complex networks. In order to improve the multi-objective optimization of community detection using discrete teaching-learning-based optimization with decomposition(MODTLBO/D) and decrease time complexity, we propose an efficient teaching-learning-based optimization algorithm combined with multi-population evolutionary strategy for community detection. In this study, we adopt adaptive learning factor in teacher phase to enhance the ability of exploration and search. In learner phase, each learner employs the random learning strategy or modified quantum-behaved learning strategy in corresponding subpopulation. After each generation, subpopulations exchange information to maintain the diversity and discourage premature convergence. The experiments results demonstrate that our proposed algorithm has an advantage of time complexity and is highly efficient at discovering quality community structure.

Key words: community detection; teaching-learning; multi-objective; multi-population evolutionary algorithm