基于动量粒子群优化的社会网络分析

马瑞新[1]，　邓贵仕[2]，　闫兆法[1]，　史哲文[1]

(1. 大连理工大学 软件学院　辽宁 大连 116620；2. 大连理工大学 管理与经济学部　辽宁 大连 116024)

摘要：针对社会网络分析中的社区发现问题，在原有的粒子群优化算法的基础上，提出了一种基于动量粒子群优化算法，并且将此算法应用于社会网络分析中的社区发现研究中，提出了一种自适应社区发现方法. 利用 Newman 提出的模块度作为适应度函数，在优化过程中自动获取社区数目，在 Karate 网络上的实验结果表明，所提出的算法能够有效地进行社区预测，并且获得了较高的预测精度.

关键词：社会网络分析；动量粒子群优化；社区发现

中图分类号：TP 311　　**文献标志码**：A　　**文章编号**：1671－6841(2011)02－0038－05

0　引言

社会网络分析(Social Network Analysis)，是用来分析社会网络的. 英国人类学家 Brown 于 1954 年首次提出了“社会网络”的概念，至今已有 50 多年的发展. 对于社会网络中的社区发现的研究，最早起源于对社会学的研究，其理论知识涉及到图论、模式识别[1-2]等. 传统的图分割方法总是假设网络是可以分解的，待划分的子图个数由用户指定. 但要对大规模的真实网络进行社区划分，社区发现方法必须解答以下两个方面的问题：一是所在网络中是否包含社区结构；二是如何有效发现网络内在的社区结构.

迄今为止，人们已经提出很多的社区发现方法，其基本思想大都是根据某个节点内聚性度量，递归地对网络进行分裂或合并，最终把网络分解为嵌套的社区层次结构. 比较典型的代表方法有：基于边介数的社区发现方法[3]和模块度优化方法[4]等.

正是基于社会网络中社区发现问题，应用粒子群优化算法来对其进行社会网络分析的研究. 粒子群优化算法一直是近年来的一个研究热点，应用极其广泛. 在原有算法的基础上，作者提出了一种基于动量粒子群优化算法，对于社会网络问题的分析取得了良好的性能，并且利用此算法成功解决了社区发现问题，进而利用这种社区发现算法经过系统的实现为用户提供更好的个性化服务.

1　动量粒子群算法

社会网络中人与人之间的各种联系就形成了一个群体，在这个群体中的成员拥有着不同的特点，为了发掘他们的关系，人类受到社会系统、物理系统、生物系统等运行机制启发，建立和发展起一个个研究工具和手段来解决和攻克研究过程汇总遇到的困难. 典型的有遗传算法(GA)，人工神经网络(ANN)，进化规划(EP)，蚁群优化算法(ACO)等，这些算法已经被广泛应用并有成功的实例. 在计算智能(CI)领域中有基于群体智能(SI)的蚁群优化算法和粒子群优化算法. 前者是对蚂蚁群落搜索食物行为的模拟，已经成功运用在很多组合优化问题上；后者就是粒子群优化算法，也是作者将重点介绍和应用的一种优化算法.

粒子群优化算法(Particle Swarm Optimization，PSO)是由 Eberhart 和 Kennedy[5]于 1995 年提出的一种基于群体的演化算法，其思想来源于人工生命和演化计算理论. 对鸟群飞行的研究发现，鸟仅仅是追踪它有限数量的邻居，但最终的整体结果是整个鸟群好像在一个中心的控制之下，即复杂的全局行为是由简单规则的相互作用引起的. PSO 算法就是从这种模型中得到启示而产生的，并用于解决优化问题. 另外，人们

收稿日期：2010-01-01
基金项目：中央高校基本科研业务费专项资金资助项目.

作者简介：马瑞新(1975－)，男，讲师，博士，主要从事电子商务、群智能及社区挖掘研究，E-mail：teacher_mrx@126.com.

通常是以自己及他人的经验作为决策的依据，这就构成了 PSO 的一个基本概念[6].

PSO 求解优化问题时，问题的解对应于搜索空间中一只鸟的位置，称这些鸟为“粒子”(particle) 或“主体”(agent). 每个粒子都有自己的位置和速度(决定飞行的方向和距离)，还有一个由被优化函数决定的适应值. 令 PSO 初始化为一群随机粒子，在每一次迭代中，粒子通过跟踪两个“极值”来更新自己：第一个是粒子本身所找到的最好解，叫做个体极值点(用 *pbest* 表示其位置)，全局版 PSO 中的另一个极值点是整个种群目前找到的最好解，称为全局极值点(用 *gbest* 表示其位置)，而局部版 PSO 不用整个种群而是用其中一部分作为粒子的邻居，所有邻居中的最好解就是局部极值点(用 *lbest* 表示其位置).

假设 m 个粒子组成一个种群，在 D 维的空间$[X^d_{\min},X^d_{\max}]^D$ 中搜索. 第 i 个粒子在第 t 代的位置为 $X_i(t)=(x_{i1},x_{i2},\cdots,x_{iD})$，速度为 $V_i(t)=(v_{i1},v_{i2},\cdots,v_{iD})$. 粒子本身目前所找到的最优解 *pbest* 为 $P_i(t)=(p_{i1},p_{i2},\cdots,p_{iD})$，整个种群目前找到的最优解 *gbest* 为 $P_g(t)=(p_{g1},p_{g2},\cdots,p_{gD})$. 根据追随最好解原理，粒子在下一代的速度和位置按式(1)、式(2)计算，跟随 *pbest* 和 *gbest* 运动：

$$v_{id}(t+1)=v_{id}(t)+c_1\cdot\mathrm{rand}()\cdot(p_{id}(t)-x_{id}(t))+c_2\cdot\mathrm{rand}()\cdot(p_{gd}(t)-x_{id}(t)),\tag{1}$$

$$x_{id}(t+1)=(1-mc)\cdot x_{id}(t)+mc\cdot v_{id}(t),\tag{2}$$

其中，mc 为动量因子$(0<mc<1)$；c_1,c_2 为学习因子；rand()为$[0,1]$之间的随机数. 粒子速度的每一维向量都限制在$[V^d_{\min},V^d_{\max}]$，如比 $V^d_{\min}$ 小，设置为 $V^d_{\min}$，如比 $V^d_{\max}$ 大，设置为 $V^d_{\max}$. $V^d_{\min},V^d_{\max}$ 设为对应决策变量取值的上下限.

另外，该算法中还使用速度松弛迭代策略：如果当前代粒子的适应度好于前一代粒子的适应度，那么下一代粒子的速度保持不变，否则按照速度更新式(1)对速度进行更新. 在算法中速度看作引导粒子的虚拟位置，而不再是粒子的运动步长，速度与位置的关系组成了导师—学生模型.

动量 PSO 的流程可以描述为：

Step1　初始化　初始搜索点的位置 X^0_i 及其速度 V^0_i 通常是在允许的范围内随机产生的，每个粒子的 *pbest* 坐标设置为其当前位置，且计算出其相应的个体极值，而全局极值就是个体极值中最好的，记录该最好值的粒子序号，并将 *gbest* 设置为该最好粒子的当前位置.

Step2　评价每一个粒子　计算粒子的适应度值，如果好于该粒子当前的个体极值，则将 *pbest* 设置为该粒子的位置，且更新个体极值. 如果所有粒子的个体极值中最好的好于当前的全局极值，则将 *gbest* 设置为该粒子的位置，记录该粒子的序号，且更新全局极值.

Step3　粒子的更新　用式(1)和式(2)对每一个粒子的速度和位置进行更新.

Step4　检验是否符合结束条件　如果当前的迭代次数达到了预先设定的最大次数(或达到最小错误要求)，则停止迭代，输出最优解，否则转到 Step2.

2　基于动量粒子群优化的社区发现算法

利用动量粒子群优化算法来解决社会网络分析中的社区发现问题，用模块度函数作为适应度函数，在优化过程中自动获取社区数目，通过实验测试，具有较好的精度.

理论指出，对于一个社区结构比较明显的网络，假设社区数目为 m，则矩阵 $\boldsymbol{N}$ 有 $k=m-1$ 个非常接近 1 的非平凡特征值，而其他的特征值都与 1 有明显的差距. 文本算法把 Capocci 算法得到的前 k 个非平凡特征向量 $\boldsymbol{V}_p=(v_{p1},v_{p2},\cdots,v_{ps})(p=1,\cdots,k)$ 作为输入，利用粒子群算法挖掘社区结构. 记所有特征向量的转置构成的矩阵为 $\boldsymbol{T}$.

$$\boldsymbol{T}=\begin{bmatrix}t_1\\t_2\\\vdots\\t_s\end{bmatrix}=\begin{bmatrix}t_{11}&t_{12}&\cdots&t_{1k}\\t_{21}&t_{22}&\cdots&t_{2k}\\\vdots&\vdots&\ddots&\vdots\\t_{s1}&t_{s2}&\cdots&t_{sk}\end{bmatrix},\tag{3}$$

其中，$\boldsymbol{T}$ 为一个 s 行 k 列的矩阵，每一列对应了一个特征向量，每一行 t_i 与原始网络中每个节点相对应.

但在实际的网络结构中，特征值的分布特征并不明显，m 的取值跟具体的网络结构相关，随着 Internet

的发展,网络越来越复杂,m 的可能取值范围也非常大,因此,采用经验的方法很难得到确切的社区数目,而试探的方法会占用大量的运算时间. 作者选取第一个为负的特征值之前的 r 个非平凡特征值,以及其对应的 r 个非平凡特征向量,也就是预测网络有 $r+1$ 个社区. 但这么选择的 r 一定大于等于 $k=m-1$. 在此基础上,将 m 的选取融入编码结构中,在优化的过程中动态发现.

2.1 粒子编码方案

对网络的各个节点进行编码的方案忽略了节点之间的相关性,不仅会增大问题的维度,而且需要复杂的后期处理消除孤立点[7]. 因此,本文算法仅对社区中心以及社区数进行编码,粒子编码结构见表 1.

表 1 粒子编码方案

Tab. 1 The particles encoding scheme

中心存在序列				中心位置序列			
$flag_1$	$flag_2$	…	$flag_{r+1}$	$center_1$	$center_2$	…	$center_{r+1}$

表 1 中 r 是第一个负值特征值之前对应的非平凡特征值个数,也就是作为优化输入的数据的维数. $flag_i$ 表征了第 i 个中心是否有效,取值在[0,1]之间,当 $flag_i \leqslant 0.5$ 时,对应的第 i 个中心是无效的,否则第 i 个中心是社区的有效中心.

$center_i$ 是 $flag_i$ 对应的社区中心,鉴于粒子群优化算法(PSO)在处理连续问题时优越的优化能力以及各个节点对应的 t_i 中各个元素的相关性,编码结构中 $center_i$ 仅映射在第一个非平凡特征向量元素的取值空间中. 假定 $a=\min(t_{11},t_{21},\cdots,t_{s1})$,$b=\max(t_{11},t_{21},\cdots,t_{s1})$,则 $center_i \in [a,b]$.

对每个粒子的编码间接映射了一种社区划分方案,在将粒子转换成具体的社区划分时,先寻找与有效的 $center_i$ 最接近的粒子,作为社区的中心.

2.2 粒子适应度定义

对于社区划分问题,每个粒子代表一种社区划分方案. 为了判定网络的划分是否优越,利用 Newman[8] 提出的模块度函数作为社区预测的评价标准.

对于通过粒子结构得到的网络社区划分方式,假定得到的网络一共有 n 个社区,定义一个 $n\times n$ 的对称矩阵 $\boldsymbol{E}=(e_{ij})$,其中,元素 e_{ij} 表示网络中连接 i 社区和 j 社区的边在所有边中所占的比例,当 $i=j$ 时,表示网络中 i 社区内部的边占所有边的比例,这里所说的边是指原始网络当中的,不是被算法已经破坏的网络,而是利用完整的原始网络来计算的.

设矩阵 $\boldsymbol{E}$ 中对角线各元素之和为 $Trace(\boldsymbol{E})=\sum_{i=1}^{n} e_{ii}$,记录了网络里所有社区内部的边占所有边的比例. 设矩阵 $\boldsymbol{E}$ 中每行元素之和为 $a_i=\sum_{j=1}^{n} e_{ij}$,记录了跟社区 i 中节点相连的边占所有边的比例. 以此为基础,定义了模块度函数,作为网络社区的划分评价标准.

$$Q(y)=\sum_{i=1}^{n}(e_{ii}-a_i^2)=Trace(\boldsymbol{E})-\|\boldsymbol{E}^2\|, \tag{4}$$

其中,$\|\boldsymbol{E}^2\|$ 记录了矩阵 a^2 中所有元素之和;$Q(y)$ 表征了网络中连接两个同种节点的边的比例,减去任意连接这两个节点的边的比例的期望值. Q 越接近于 1,社区结构越明显,通过这个衡量标准就可以建立起社区划分质量的全局度量函数.

利用该模块度函数作为适应度函数,将社区发现问题转化为最优化问题,从而借助粒子群算法求解.

2.3 算法流程

整体算法流程如图 1 所示.

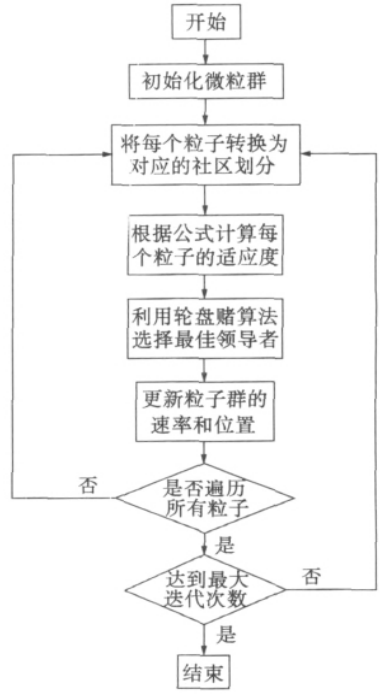

图 1 算法流程

Fig. 1 The algorithm flow

3　社会网络实验测试

采用 Karate 网络对所提出的算法尝试进行社区划分，Karate 网络包含 34 个节点，78 条边. 首先利用谱分法寻找标准矩阵的非平凡特征值，得到前 3 个特征值为正，分别为 0.867 2，0.704 4，0.613 0，于是取前 3 个非平凡特征向量作为待处理数据.

网络的最大可能社团数为 4，粒子群的编码纬度为 8. 种群数目设为 20，最大迭代次数为 100，从 0.9 到 0.4 变换，$c_1=c_2=0.05$，搜索空间前 4 维在[0，1]之间，后 4 维的取值在第一个非平凡特征向量的最小、最大值之间，即[−0.206 0，0.343 1].

算法运行 30 次，最终划分为 4 个社区，得到的模块度均为 0.419 6，如图 2 所示.

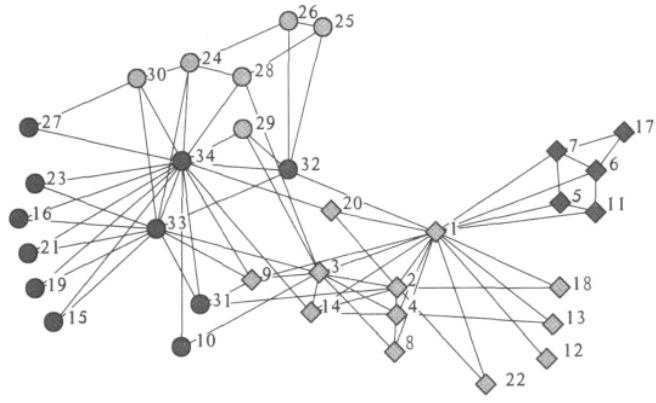

图 2　Karate 网络划分结果

Fig. 2　The Karate network division results

由图可知，把每个社区又分成了 2 个. 相比于 2 个社区的模块度 0.371 5，本算法获得了更大的模块度 0.419 6. 表 2 显示了与 GN[8]、快速 GN[9]、DA[5] 以及 Newman 算法[10] 的划分结果进行对比，此算法不仅取得了正确的划分，而且取得了较大的模块度，充分显示出了良好的性能.

表 2　Karate 网络划分结果比较

Tab. 2　The comparison of Karate network division results

算法	划分社区数	模块度
GN	3	0.401 0
快速 GN	2	0.381 0
DA	4	0.418 8
Newman 算法	4	0.418 8
本文方法	4	0.419 6

4　结论

提出了一种基于动量粒子群的自适应社区发现方法. 利用 Newman 提出的模块度作为适应度函数，利用粒子群算法优化聚类中心. 在对粒子编码时，将聚类数目隐含放入编码方案，可在优化的过程中自动发现社区数目. 在 Karate 网络上的实验证明，此算法能够有效地进行社区预测，获得了较高的预测精度. 通过与 GN、快速 GN、DA 等经典算法的分析比较，此方法在 3 个数据集上均取得了最大的模块适应度.

参考文献：

[1]　Scott J. Social Network Analysis：a Handbook[M]. 2nd ed. London：Sage Publications，2000：113-119.

[2]　West D B. Introduction to Graph Theory[M]. 2nd ed. Upper Saddle River：Prentice Hall，2001：87-92.

[3] Girvan M, Newman M E J. Community structure in social and biological networks[J]. The National Academy of Sciences of the United States of America, 2002, 99(12):7821-7826.

[4] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2):026113.

[5] Kennedy J, Eberhart R. Particle swarm optimization[C]//IEEE International Conference on Neural Networks. Australia, 1995:1942-1948.

[6] 杨维,李歧强. 粒子群优化算法综述[J]. 中国工程科学,2004,6(5):87-94.

[7] 段晓东,王存睿,刘向东,等. 基于粒子群算法的 Web 社区发现[J]. 计算机科学,2008,3(35):18-21.

[8] Newman M E J. Fast algorithm for detecting community structure in networks[J]. Physical Review E, 2004, 69(6):321-330.

[9] Radicchi F, Castellano C, Cecconi F, et al. Defining and identifying communities in networks[C]//Proceedings of the National Academy of Sciences. USA, 2004:2658-2663.

[10] Parsopoulos K, Vrahatis M. On the computation of all global minimizers through particle swarm optimization[J]. IEEE Transactions on Evolutionary Computation, 2004, 8(3):211-224.

Analysis of Social Networks Based on the Momentum Particle Swarm Optimization

MA Rui-xin[1], DENG Gui-shi[2], YAN Zhao-fa[1], SHI Ze-wen[1]

(1. *School of Software, Dalian University of Technology, Dalian* 116620, *China*;
2. *School of Management, Dalian University of Technology, Dalian* 116024, *China*)

Abstract: In terms of social network analysis, a new momentum particle swarm optimization algorithm based on the original thoughts of PSO was proposed. By this algorithm, the social network analysis was applied to solve community detection problems. An adaptive community discovery algorithm based on momentum particle swarm optimization was further proposed. By using Newman's modularity as fitness function, the number of communities in the optimization process was obtained. Experiments on Karate network showed that the algorithm could effectively predict the community and obtain perfect prediction accuracy.

Key words: social network analysis; momentum particle swarm optimization; community discovery