第23卷 第5期
2013年5月

计算机技术与发展
COMPUTER TECHNOLOGY AND DEVELOPMENT

Vol. 23 No. 5
May. 2013

二元群智能算法求解组卷问题研究

程美英[1],钱 乾[1,2]

(1. 安徽商贸职业技术学院 电子信息工程系,安徽 芜湖 241002;
2. 安徽工程大学 计算机与信息学院,安徽 芜湖 241000)

摘 要:二元蚁群优化算法(BACO-CA)及二元粒子群优化算法(BPSO-CA)作为基于概率的随机搜索智能算法,二者在寻优机理上有着显著的不同。以大规模组合优化问题组卷问题为例,通过设置算法中的参数,探讨二元蚁群优化算法和二元粒子群优化算法求解组卷问题性能的优劣。仿真实验表明,二元蚁群优化算法和二元粒子群优化算法虽然均能在多项式时间内完成组卷问题的求解,但二元粒子群优化算法在求解组卷问题时较二元蚁群优化算法具有更好的时间性能,能在较短的时间收敛到全局最优解。

关键词:二元蚁群算法;二元粒子群算法;组卷问题;时间性能对比分析

中图分类号:TP301.6 **文献标识码**:A **文章编号**:1673-629X(2013)05-0079-04

doi:10.3969/j.issn.1673-629X.2013.05.020

Research on Binary Swarm Intelligence Algorithm for Test Paper Problem

CHENG Mei-ying[1], QIAN Qian[1,2]

(1. Electronic Information Engineering Department, Anhui Business College of Vocational Technology, Wuhu 241002, China;
2. College of Computer and Information, Anhui Polytechnic University, Wuhu 241000, China)

Abstract: As a random search algorithm based on probability, the binary ant colony algorithm (BACO-CA) and the binary particle swarm optimization (BPSO-CA) has the different optimization mechanism. Take the test paper for example, by setting the parameters of the algorithm, the performance of the BACO-CA and the BPSO-CA for the test paper problem was discussed. Experimental results show that the two algorithm can finish the group problem solving in polynomial time, but the BPSO-CA has the good performance, and can solve the test paper problem in polynomial time, and converge to the global optimal solution in a short time.

Key words: binary ant colony algotithm; binary particle swarm optimization; test paper problem; time performance analysis

0 引 言

近年来,群集智能(Swarm Intelligence)成为众多学者研究的热点之一。群是指某种交互作用的组织或Agent之结构的集合。在群集智能计算研究中,群的个体组织包括蚂蚁、白蚁、蜜蜂、鱼群以及鸟群等。在这些群体中,个体在结构上很简单,而它们的集体行为却变得很复杂,如:在一个蚁群中,每只蚂蚁个体只能执行一组很简单的任务中的一项,而在整体上蚂蚁的动作和行为却能确保建造最佳的蚁巢结构、保护蚁后和幼蚁、清净蚁巢、发现最好的食物源以及优化攻击策略等全局任务的实现等。人们从群智能的生物进化机理中受到启发,提出了解决优化问题的新方法。目前群集智能研究领域主要有两种算法:蚁群优化算法(Ant Colony Optimization,即ACO)[1]和粒子群优化算法(Particle Swarm Optimization,即PSO)[2],二者在数据分类、模式识别、电信QoS管理、生物系统建模、机器人控制等领域得到较好的应用。

蚁群优化算法(ACO)和粒子群优化算法(PSO)这两种基于概率的随机搜索算法在算法结构、研究内容、方法以及步骤上有较大的相似性:即从一组初始解出发,计算适应值;根据某种规则,产生下一组解,如此重复直到满足迭代次数或某种收敛条件! ACO与PSO虽然依靠概率搜索,需要较多的评价函数,但与梯度方法和传统的演化算法相比较而言,其优点还是相当显著的。

ACO与PSO虽然有众多的相似性,但由于二者寻优机理的不同导致二者的算法流程也有着显著的差别(蚁群优化算法的灵感来自于对蚂蚁群体食物采集过程的模拟,而粒子群优化算法则是模拟鸟群寻找食

收稿日期:2012-08-28;**修回日期**:2012-11-30

基金项目:安徽省教育自然科学基金重点项目(KJ2007A046);安徽省教育自然科学研究项目(KJ2011Z131);安徽商贸职业技术学院院级科研项目(KY20100624)

作者简介:程美英(1983-),女,硕士,助教,主要研究方向为智能计算。

物的过程),其最大不同是“单个个体如何根据某种规则产生下一组解”。如蚁群优化算法通过调整外激素(即信息素),根据信息素的大小得到转移概率产生下一代解;而粒子群算法是通过更新粒子的速度和位置来产生下一组解。ACO 与 PSO 作为比较成熟的优化算法,目前主要致力于算法的应用研究,而对算法内部机理的探讨以及二者之间性能特点的对比研究还不是十分充分,文献[3~5]中虽然提到 ACO 与 PSO 的区别,但仅仅局限于理论层面。

在实际应用中,一些易于表述但难于求解的 NP 难组合优化问题(如:排课问题、车间作业调度、大规模集成电路综合布线、组卷问题、集装箱装载等)始终是研究的一个热点。穷举搜索这种自然但近乎幼稚的算法在问题规模 n 较大时,算法的运行时间将以指数增长,人工智能学家们为了使“时间-问题大小”曲线的变化尽可能缓慢的增长或根据实际需要能得到一个人们可以接受的时间而一直不懈的努力。对于一个实际的优化问题,要具体选用一种优化方法来解决,而在实际应用中,几乎每一种优化算法经过适当的变换都可应用到各种优化问题。因而,对于一个具体的优化问题,可以选取任意一种优化算法来解决。但是问题是:究竟选择哪一种优化算法效果更好、且程序的运行时间更短呢?对于这个问题,无论在理论上还是在实际应用中,大都采取回避的态度。

文献[6~8]将二进制编码分别引入到蚁群算法及粒子群算法中,提出了二元蚁群算法(BACO-CA)及二元粒子群算法(BPSO-CA),BACO-CA 和 BPSO-CA 虽然均能在多项式时间内解决二元离散的 NP 难问题[9~11],但是实际应用中,二者有着明显的优劣。文中将以大规模的组合优化问题——组卷问题为例,通过设置 BACO-CA 与 BPSO-CA 中的参数进行大量实验,对二者求解问题的性能进行对比分析,探讨二者求解特定问题的优劣,这就为后续 NP 难组合优化问题选择合适的算法以及算法的融合提供了很好的参考依据。

1 二元群智能算法基本模型描述

1.1 二元蚁群算法基本模型描述(BACO-CA)

假设在一条直线上均匀分布着 L 个细胞(见图1),蚂蚁种群的规模为 m;细胞状态集合 $Q \in \{0,1\}$,$\forall t$,蚂蚁 k,第 $i(i=1,2,\cdots,L)$ 个细胞状态0上分布的信息素为:$\tau_{i0}^{k}(t)$,第 i 个细胞状态1上分布的信息素为:$\tau_{i1}^{k}(t)$,P_p 为预先设定的概率,$P_{i0}^{k}(t)=\dfrac{\tau_{i0}^{k}}{\tau_{i0}^{k}+\tau_{i1}^{k}}$,任意时刻,若 $P_{i0}^{k}(t) \leqslant P_p$,则蚂蚁选择状态0(即 $Q=0$),反之选择状态1(即 $Q=1$)。

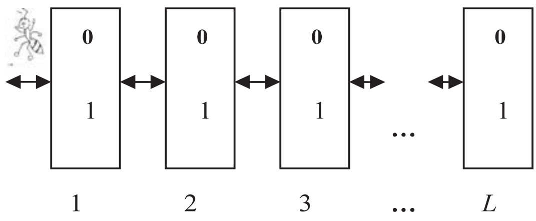

图1 二元蚁群算法细胞自动机模型

经过 t 时刻,当所有蚂蚁遍历完细胞阵列之后,细胞内部的信息素按如下公式(1)~(2)进行调整。同时为了避免信息素在较短的时间内集中在某一条特定的路径上,在信息素更新过程中引入 MAX-MIN 原则,即每一次迭代之后,只有在本次迭代中取得最优的细胞阵列上的信息素进行更新调整:

$$\tau_{i0}(t+1)=\rho \cdot \tau_{i0}(t)+\Delta\tau \tag{1}$$

$$\tau_{i1}(t+1)=\rho \cdot \tau_{i1}(t)+\Delta\tau \tag{2}$$

公式(1)、(2)中的 ρ 是信息素挥发系数,$\Delta\tau=\dfrac{1}{f(\mathrm{best})}$,$f(\mathrm{best})$ 是每一代的最优解或全局最优解,并将信息素设置上、下界,若信息素更新之后大于 $\tau_{\max}$,则为 $\tau_{\max}$,反之为 $\tau_{\min}$。

1.2 二元粒子群算法基本模型描述(BPSO-CA)

假设在一条直线上分布着 N 个细胞(见图2),第 i 个细胞对应粒子的第 i 维空间,将 N 维空间中全局最优粒子的位置 p_g 存储在图2所示的细胞中。细胞状态集合为{0,1}。任一细胞下一时刻的状态取值由粒子自身携带的信息及存储在细胞中的全局最优粒子的信息根据1和0的概率分布来决定(如公式3~5所示):

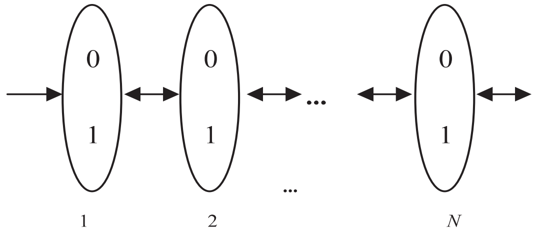

图2 二元粒子群算法细胞自动机模型

$$v_{ij}(t+1)=v_{ij}(t)+c_1 r_1(\overline{p_{ij}}(t)-\overline{x_{ij}}(t))+c_2 r_2(\overline{p_g}(t)-\overline{x_{ij}}(t)) \tag{3}$$

$$\overline{x_{ij}}(t+1)=\overline{x_{ij}}(t)+v_{ij}(t+1) \tag{4}$$

$$X_{ij}(t+1)=\begin{cases}0 & \mathrm{sign}(\overline{x_{ij}}(t+1)) \leqslant 0.5 \\ 1 & \mathrm{sign}(\overline{x_{ij}}(t+1)) > 0.5\end{cases} \tag{5}$$

其中,$\mathrm{sign}(x)=1/(1+\mathrm{e}^{-x})$ 为模糊函数,c_1,c_2 为加速系数,r_1,r_2 为相互独立的随机函数。任意时刻,当所有粒子完成对细胞阵列的遍历之后,更新调整存储在细胞中的全局最优粒子的位置得到 $\overline{p_g}(t+1)$ 及种群中粒子的个体最优位置 $\overline{p_{ij}}(t+1)$。

2　二元群智能算法求解组卷问题

2.1　组卷问题数学模型描述

智能组卷指计算机按照用户输入的要求,自动抽取满足用户需求的一套试卷的过程。试卷中的每一道题目,由 n 维向量来描述($a_1,a_2,a_3,a_4,a_5,a_6,\cdots,a_n$),其中 a_1 表示题型、a_2 为题分、a_3 为题目难度、a_4 为教学要求、a_5 为知识单元、a_6 为估时…, 决定一份试卷,就决定一个 $m\times n$ 矩阵 $\boldsymbol{S}_g=\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$,其中 m 是试卷所含的题目数,则组卷的数学模型可描述如下[9]:

目标函数 $\min f=\sum_{i=1}^{n} w_i e_i$(其中 $i=1,2,\cdots,n$,即所求试卷与目标试卷的误差越小越好),且总分满足 $\sum_{i=1}^{m} a_{i2}=P_1$,试卷难度 $P_2=\sum_{i=1}^{m} a_{i2}a_{i3}/P_1$,$P_K=\sum_{i=1}^{m} C_{1i}a_{i2}$,K 为教学要求,即教学约束。教学要求 K 取值{识记、理解、综合、应用}等,所占分数由用户在初始组卷时给出。

其中 $C_{1i}=\begin{cases}1,当\ a_{i4}=k\ 时\\0,当\ a_{i4}\neq k\ 时\end{cases}\cdots$,同理,知识单元分值分布、全卷估时、区分度、题型、期望值等约束和上面类似,均由用户在初始组卷时给出。

2.2　二元蚁群算法求解组卷问题

下面具体给出二元蚁群算法求解组卷问题伪代码描述:

input: An objective function $f(x)=\sum_{i=1}^{n} w_i e_i$
/ * $f(x)$ 组卷的目标函数 * /
initialize pheromone distribution T;　　/ * 初始化路径上的信息素 * /
while (stop condition not met) do
for $i=1$ to m do / * m 为种群中蚂蚁的个数 * /
$v_i\leftarrow$empty binary string
/ * v_i 用来存放蚂蚁遍历的细胞阵列 * /
for $j=1$ to $LC\cdot VN$ do　　/ * 蚂蚁遍历细胞阵列的过程, $LC*VN$ 为题库中总的试题个数 * /
choose a value $c\in\{0,1\}$ for a_j
/ * 蚂蚁根据路径上的信息素大小在{0,1}之间进行选择, a_j 为细胞阵列中的第 j 位 * /
append $a_j=c$ to v_i
end for
end for
encode and evaluate all the colony size solution
$v_i, i\in\{1\cdots m\}$　/ * 将每个蚂蚁所抽出的试题与用户输入的要求进行对比 * /
perform local pheromone update
/ * 信息更新(挥发) * /
$v_{\rm Ibest}\leftarrow$iteration best solution
$v_{\rm Gbest}\leftarrow$best of $v_{\rm Ibest}$ and previous global best $v_{\rm Gbest}$
perform global pheromone update
/ * 全局信息素更新 * /
end while
output: Best solution found $v_{\rm Gbest}$
/ * 输出最佳试卷 * /

2.3　二元粒子群算法求解组卷问题

下面具体给出二元粒子群算法求解组卷问题伪代码描述:

input: An objective function $f(x)=\sum_{i=1}^{n} w_i e_i$
/ * $f(x)$ 组卷的目标函数 * /
initializeinformation of the particle;　　/ * 初始粒子自身的信息 * /
while (stop condition not met) do
for $i=1$ to m do / * m 为种群中粒子的个数 * /
$v_i\leftarrow$empty binary string
/ * v_i 用来存放粒子遍历的细胞阵列 * /
for $j=1$ to $LC\cdot VN$　do　/ * 粒子遍历细胞阵列的过程, $LC*VN$ 为题库中总的试题个数 * /
choose a value $c\in\{0,1\}$ for a_j
/ * 粒子自身携带的信息并感知存储在细胞中的信息按概率公式选择状态 0/1, a_j 为细胞阵列中的第 j 位 * /
append $a_j=c$ to v_i
end for
end for
解码
$v_i, i\in\{1\cdots m\}$　/ * 将每个粒子所抽出的试题与用户输入的要求进行对比 * /
更新全局最优粒子位置和个体粒子最优位置
end while
output: Best solution
/ * 输出最佳试卷 * /

3　二元群智能算法求解组卷问题性能分析

3.1　二元群智能算法求解组卷问题时间复杂度分析

假设 N 为算法的迭代次数,m 为种群规模,$LC\cdot VN$ 为题库的规模(即细胞阵列的长度为 $LC\cdot VN$),二元网络上的任一个细胞代表题库中等待被选择的试题,则二元蚁群算法(BACO-CA)求解组卷问题的时间复杂度为:$O(N\cdot m\cdot LC\cdot VN)$。

同理二元粒子群算法(BPSO-CA)求解组卷问题的时间复杂度为 $O(N\cdot \mathrm{popsize}\cdot LC\cdot VN)$。其中 popsize 为粒子群种群规模。

由此分析看出,对于 NP 难的组合优化问题——组卷问题,二元蚁群算法和二元粒子群算法均能在多项式时间内完成问题的求解。

3.2 二元群智能算法求解组卷问题具体时间性能分析

算法执行时间需通过依据该算法编制的程序在计算机上运行时所消耗的时间来度量。而度量一个程序的执行时间通常有两种方法:事前统计法和事后统计法。在这里将使用事后统计方法中的 clock/CLOCKS_PER_SEC,并且文中的所有程序均在 3.0GHz、3.25GB 内存的计算机上运行。

在本实验中,题库规模为 500(其中题型为:填空题、选择题、判断题、改错题、编程题,每种题型各 100 道,总共 500 道题存在题库中),卷面总分为 100,时间 100 分钟,试卷难度 0.6,教学要求分值分布为:识记 10 分、理解 35 分、综合 35 分、应用 20 分,知识单元共 10 个,其相应的知识分值分布为:10 分、4 分、10 分、10 分、14 分、8 分、12 分、8 分、10 分、14 分,试卷中各种题型要求的个数分别为:填空题 10 道、选择题 10 道,编程题 2 道,改错题 5 道,判断题 5 道。假设种群规模均为 30;$\rho=0.95$,$P_p=0.5$,加速常数 $c_1=c_2=1.8$;粒子速度 $v_{ij}\in[-5,5]$,粒子位置 $\overline{x_{ij}}\in[-7.5,7.5]$,个体最优位置 $\overline{p_{ij}}\in[-7.5,7.5]$,每隔 10 代运行 10 次,求解平均最优值及平均运行时间(以毫秒 ms 为计量单位),如表 1、表 2 所示:

表 1 运行 BACO-CA 求解组卷问题实验结果

运行代数	平均最优值	平均运行时间(ms)
10	95.32724	963.9
20	75.93128	1920.5
30	72.10536	2878.1
40	66.52532	3828.6
50	59.91498	4822.1
60	62.74202	5748.2
70	57.32556	6803.1
80	54.73114	7696.3
90	52.50224	8564.4
100	45.8079	14729.7
200	33.5432	200325.9
300	19.5547	373837.2

在图 3 中,实曲线表示运行 BACO 算法求解组卷问题平均最优值随平均运行时间变化趋势;虚曲线表示运行 BPSO 算法求解组卷问题平均最优值随平均运行时间变化趋势;由图 3 可以得出:BPSO 能以较快的速度收敛到全局最优解,在求解大规模组卷问题时比 BACO 具有更好的时间性能。

表 2 运行 BPSO-CA 求解组卷问题实验结果

运行代数	平均最优值	平均运行时间(ms)
10	58.1906	4156
20	54.131	8390
30	40.156	12109
40	36.1794	16093
50	32.1778	19860
60	32.1604	24078
70	29.2298	27594
80	28.0616	31422
90	27.2102	35250
100	24.2174	38828
200	16.108	77469
300	14.1574	117094

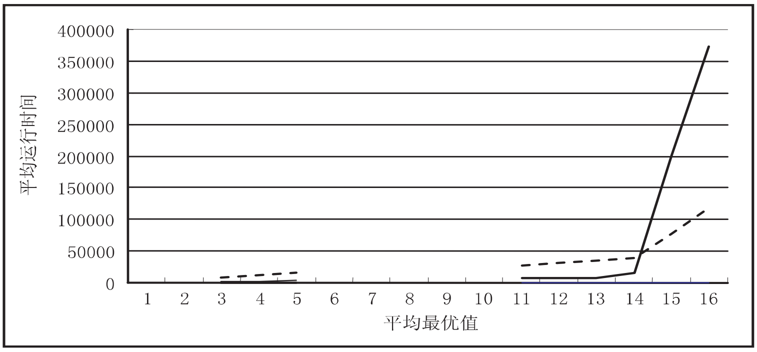

图 3 BACO-CA 与 BPSO-CA 求解组卷问题时间性能对比图

4 结束语

二元蚁群算法及二元粒子群算法作为人工生命中一种,二者均通过 Agent 之间的局部作用而“涌现”出复杂的现象。BACO-CA 及 BPSO-CA 虽然均能在多项式时间完成组合优化问题——组卷问题的求解[12,13],求解效果优于文献[14],但在实际应用中,BPSO-CA 具有更好的时间性能,这也为后续其他优化问题的解决选择合适的算法提供了一定的参考依据。

参考文献:

[1] Colorni A,Dorigo M,Maffioli F,et al. Heuristics from nature for hard combinatorial problem[J]. International Transactions in Operational Research,1996,3(1):1-21.

[2] Kennedy J,Eberhart R C. Particle Swarm Optimization[C]//Proc. of IEEE International Conference on Neural Networks. Perth,Australia:[s. n.],1995:1942-1948.

[3] 杨劲秋. 智能优化算法评价模型研究[D]. 杭州:浙江大学,2012.

[4] 冯春时. 群智能优化算法及其应用[D]. 合肥:中国科学技术大学,2009.

(下转第 85 页)

图 4　提升小波变换去噪图像

图 5 形态学去噪图像

图 6　融合图像

4.3　实验处理数据

表 1 为实验处理数据。

表 1　实验处理数据

	均值	标准差	平滑指数	信噪比	均方误差
含噪图	179.59	103.19	1.74	13.83	51.29
软阈值去噪	177.82	91.29	1.94	15.58	45.38
形态学去噪	167.95	78.86	2.13	20.48	39.27
文中算法	172.90	73.68	2.34	19.61	36.69

通过上面的数据,可以看出采用基于提升小波与形态学去噪方法,峰值信噪比比软阈值去噪方法要大,标准差比软阈值去噪及形态学去噪方法要小,均值比软阈值去噪方法要大,均方误差比软阈值去噪及形态学去噪方法要小,平滑指数比软阈值去噪及形态学去噪方法要大,有效地去除了噪声。

从图 6 及实验数据中可以看出融合后的图像的去噪效果比较好。

5　结束语

在提升小波变换与数学形态学对含噪图像的处理中,不同的小波选择和不同的结构元素选择在去噪声能力方面是不一样的。

基于提升小波变换,研究了一种融合数学形态学的去噪方法——基于数学形态学的提升小波去噪法,有效地改善了图像质量,提高了峰值信噪比,使图像更加清晰。通过实验数据和分析说明,该方法能更加快速、有效地去除噪声。

参考文献:

[1] Daubechies I,Sweldens W. Factoring Wavelet Transforms into Lifting Steps[J]. Journal of Fourier Analysis and Application,1998,4(3):245-267.

[2] 季　忠,黄　捷,秦树人. 提升小波在齿轮箱故障诊断中的应用[J]. 震动、测试与诊断,2010,30(3):291-294.

[3] 袁　静,何正嘉,訾艳阳. 基于提升多小波的机电设备复合故障分离和提取[J]. 机械工程学报,2010,46(1):79-85.

[4] 周建萍,郑应平,王志萍. 基于提升小波的输电线路短路故障奇异点检测[J]. 上海电力学院学报,2009,25(4):329-331.

[5] 赵高鹏,薄煜明,刘　娣. 基于提升小波的红外和可见光图像融合方法[J]. 计算机工程与设计,2009,30(7):1697-1699.

[6] 黄宝贵,马春梅,卢振泰. 新的形态学图像降噪方法[J]. 计算机应用,2011,31(3):757-759.

[7] 夏开建,姚宇峰,钟　珊,等. 基于形态学小波变换的图像融合算法[J]. 计算机工程,2010,36(10):224-226.

[8] 杨　鹏. 改进的数学形态学小波图像融合算法[J]. 计算机仿真,2011,28(2):288-291.

[9] 王　新,黄兆云. 基于多结构元素的数学形态学图像边缘检测[J]. 计算机工程与应用,2008,44(7):89-90.

[10] 闫海霞,赵晓晖. 基于数学形态学的边缘检测方法[J]. 计算机应用研究,2008,25(11):3496-3497.

[11] 徐国保,王　骥,赵桂艳,等. 基于数学形态学的自适应边缘检测新算法[J]. 计算机应用,2009,29(4):997-999.

[12] 薛丽霞,李　涛,王佐成. 自适应的形态学边缘检测算法[J]. 计算机工程,2010,36(12):214-216.

[13] 何新英,王家忠,孙晨霞,等. 基于数学形态学和 Canny 算子的边缘提取方法[J]. 计算机应用,2008,28(2):477-483.

(上接第 82 页)

[5] 陈恩修. 离散群体智能算法的研究及应用[D]. 济南:山东师范大学,2009.

[6] 熊伟清,魏　平. 二进制蚁群进化算法[J]. 自动化学报,2007,33(3):259-264.

[7] 熊伟清,魏　平,赵杰煜. 传递信号的二元蚁群算法[J]. 模式识别与人工智能,2007,20(1):15-20.

[8] 程美英,熊伟清,严　彬,等. 求解多维 0/1 背包问题的二元粒子群算法[J]. 系统仿真学报,2009,18(9):5735-5743.

[9] 程美英,熊伟清,魏　平. 二元蚁群算法求解组卷问题[J]. 计算机应用研究,2008,25(9):2637-2639.

[10] 严　彬,熊伟清,程美英,等. 带拥塞控制的多种群二元蚁群算法[J]. 控制理论与应用,2009,26(4):387-394.

[11] 胡　钢,熊伟清,张　翔,等. 可控搜索偏向的二元蚁群算法[J]. 控制理论与应用,2011,28(8):1071-1080.

[12] 钱　乾,程美英. 二元蚁群优化算法研究综述[J]. 计算机应用研究,2012,29(4):1211-1215.

[13] 魏　平,熊伟清,魏　颖. 一种求解组卷问题的二元粒子群算法[J]. 计算机工程与应用,2009,45(30):80-83.

[14] 楼玉萍,金炳尧,骆红波. PBIL 进化算法在自动组卷系统中的应用[J]. 计算机技术与发展,2006,16(6):80-82.