

结合特征学习的粒子群求解极小碰集方法

刘娟^{1,2}, 欧阳彤^{1,2}, 王艺源^{1,2}, 张立明^{1,2}

(1. 吉林大学计算机科学与技术学院, 吉林长春 130012; 2. 符号计算与知识工程教育部重点实验室, 吉林长春 130012)

摘要: 基于模型诊断中的极小碰集问题是人工智能领域的一个重要课题, 现实中很多实际问题都可以转化为极小碰集问题, 如老师与课程问题, 极小覆盖集问题等. 通过对极小碰集问题特征的研究, 本文结合粒子群优化求解极小碰集的算法提出了一个新的算法, 来指导极小碰集的求解; 引入学习机制, 减少极小碰集求解中对无解空间的搜索; 加入翻转策略, 来加速极小碰集有解空间中的求解. 实验结果表明本文提出的算法在求解极小碰集问题上的效率有明显提高.

关键词: 极小碰集; 特征; 学习机制; 翻转策略

中图分类号: TN911.23

文献标识码: A

文章编号: 0372-2112 (2015)05-0841-05

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2015.05.002

Computing Minimal Hitting Sets with Particle Swarm Optimization Combined Characteristics Learning

LIU Juan^{1,2}, OUYANG Dan-tong^{1,2}, WANG Yi-yuan^{1,2}, ZHANG Li-ming^{1,2}

(1. College of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China;

2. Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Changchun, Jilin 130012, China)

Abstract: In model-based diagnosis, minimal hitting sets problem is an important topic of artificial intelligence, and a lot of practical problems can be converted into it, such as teachers and curricula problem, minimum covering sets problem, etc. By studying the characteristics of minimal hitting sets, combining with the original PSO algorithm of computing minimal hitting sets, this paper proposes a new algorithm to guide the minimum hitting sets computation: introducing a learning mechanism to cut down some search of the no-solution space; adding a flipping strategy to accelerate some solving of solution space. Experimental results show a significant improvement of this new algorithm in computing minimal hitting sets.

Key words: minimal hitting sets; characteristics; learning mechanism; flipping strategy

1 引言

基于模型诊断 (Model-Based Diagnosis, MBD) 问题是基本的极小碰集求解问题, 通常先求出所有极小冲突集合, 再通过极小冲突集合簇求解极小碰集, 即系统的极小诊断. 现实和理论中的很多问题在某种程度上都可以归结为极小碰集问题. 如智能规划问题^[1], 实时多故障诊断中的动态碰集问题^[2], 不可满足核求解问题^[3]等.

人工智能领域专家 Reiter 于 1987 年首次给出计算碰集的算法 HS-TREE^[4]. 后来, 碰集问题被证明是 NP-Complete 问题, 其求解效率的提高显得尤为重要. 许多计算极小碰集的算法被相继提出, 如 BHS-TREE^[5,6], Boolean^[7], DMD-TREE^[8], 分布式方法^[9]等. 以上方法大

都基于树或图, 数据结构复杂且计算量大, 不适合求解大规模的碰集问题. 除了这些算法, 还有使用遗传算法^[10], 粒子群算法^[11]来求解极小碰集. 碰集问题的求解呈现出一种更加复杂化的发展趋势, 使用完备算法求解极小碰集的复杂度将会随着问题复杂性的增大而大幅增加, 因此一种能够更好地平衡复杂度与完备性的算法至关重要.

群智能算法是对群体觅食过程的模拟, 最具代表性的算法有粒子群算法, 蚁群算法和人工蜂群算法, 许多源于蜂群和蚁群模型设计的算法已越来越多地应用于解决实际问题^[12-14]. 粒子群优化 (Particle Swarm Optimization, PSO) 算法相对于其它群智能算法具有参数简洁、易于实现、精度高、收敛快等优点, 并被广泛应用于

不同的领域.如求解流水线调度问题^[15,16],电磁优化问题^[17]等.

本文给出一种结合结构特征学习的粒子群求解极小碰集的方法,来减少碰集求解中对无解空间的搜索;同时加入翻转策略来加速碰集有解空间中的求解.

2 基于 PSO 的极小碰集求解方法

PSO 中优化问题的解被看作“粒子”,每个粒子有一个速度决定它们运动的距离和方向,所有粒子依据当前所得到的最优粒子在解空间中进行搜索.本节详细介绍文献[11]中基于 PSO 优化算法求解极小碰集,下面给出几个相关定义.

定义 1^[3] 冲突集

集合 $C = \{c_1, c_2, \dots, c_k\} \subseteq \text{COMPS}$ 为系统 (SD, COMPS, OBS) 的一个冲突集,若 C 使得 $\text{SD} \cup \text{OBS} \cup \{\sim \text{AB}(c_1), \sim \text{AB}(c_2), \dots, \sim \text{AB}(c_n)\}$ 不一致.若该冲突集的任意真子集都不是一个冲突集,那么它是一个极小冲突集.

定义 2^[4] 碰集

设 $F = \{C_1, C_2, \dots, C_n\}$ 是一个冲突集合簇,称 H 是该冲突集合簇的一个碰集,当且仅当 H 满足:

(1) $H \subseteq \bigcup C_i (i = 1, 2, \dots, n)$; (2) $\forall C_i \in F, H \cap C_i$

$\neq \varphi$.

称 F 的一个碰集是极小碰集当且仅当它的任意真子集都不是 F 的碰集.

定义 3^[11] 精英集

在整个搜索过程中会产生很多碰集,把每次迭代产生的极小碰集放入一个集合中,这个集合就称为精英集.并且对于精英集中的任一碰集,它的超集都不在精英集中.

在基于 PSO 求解极小碰集问题中,速度被赋予新的更新方式,公式被改变为更适应碰集问题性质的如下公式:

$$V_i^{k+1} = \omega V_i^k + c_1 * r_1 * (Pe - X_i^k) \quad (1)$$

$$+ c_2 * r_2 * (\mathbf{0} - X_i^k)$$

$$X_i^{k+1} = X_i^k + V_i^{k+1} \quad (2)$$

其中, Pe 为精英集中的某个粒子, $\mathbf{0}$ 表示 D 维零向量.

根据碰集的定义,给出适应度函数 $f(X) = h/c$, h 表示冲突集合簇中与当前粒子有交的集合数, c 表示冲突集合簇中的集合数.结合精英集的定义,当 $f(X) = 1$ 时,所求的集合为碰集.在第一次迭代时,精英集被初始化为只含有 $(\mathbf{1}, \mathbf{1}, \dots, \mathbf{1})$ 的集合.

在标准的粒子群算法中,速度使粒子从一个位置运动到另一个位置,而在碰集问题中,速度中的每一维的分量只能从 $\mathbf{0}$ 变化为 $\mathbf{1}$, 或从 $\mathbf{1}$ 变化为 $\mathbf{0}$, 因此把速度定

义为向量的每一维取“1”的概率.公式(1)给出了粒子速度的更新公式, $c_1 * r_1 * (Pe - X_i^k)$ 包含了两层意思,首先对于 Pe 中存在, X_i^k 中不存在的分量其选择概率增加 $c_1 * r_1$; 其次,对于 X_i^k 中存在而 Pe 中不存在的分量其选择概率减少 $c_1 * r_1$.这就使得新产生的个体向优秀个体 Pe 靠近,同时也实现了对 Pe 的子集的搜索.在选择 Pe 时,若选择到的碰集本身就是一个极小碰集,那么就只能起到搜索其它碰集的作用,而无法去除精英集中的超集.因此在 Pe 的选择上更倾向于选择包含 1 的个数较多的个体. $c_2 * r_2 * (\mathbf{0} - X_i^k)$ 部分减少 X_i^k 中已存在分量的选择概率,也是加速搜索极小碰集的一种策略.根据式(1)完成速度的更新后所得到的的结果不一定在 $[0, 1]$ 区间内,因此给出规范式(3)来调整速度使其符合规范.其中 $v_{i,m}$ 是 V_i 的最大分量,当 $v_{i,m} > 1$ 时执行式(3).

$$v_{i,j} = v_{i,j} / v_{i,m} (j = 0, 1, 2, \dots, D - 1) \quad (3)$$

通过以上过程得到的每个分量的选择概率为 $[0, 1]$ 区间内的实数,不能直接得出选择或不选择某个分量,必须将每个概率映射到 0 或 1,其方法是产生一个 $(0, 1)$ 之间的随机数,比较每个分量的选择概率与随机数的大小,若大于,则该分量的对应编码为 1, 否则为 0.映射公式为:

$$x_{i,j} = \begin{cases} 1, \text{rand}(0, 1) < v_{i,j} \\ 0, \text{rand}(0, 1) > v_{i,j} \end{cases} \quad (4)$$

上述内容详细介绍了使用粒子群优化算法求解极小碰集的相关概念和公式,下面给出利用 PSO 求解极小碰集的优化算法 OA (Original Algorithm).

Algorithm OA

Input: $F = \{C_1, C_2, \dots, C_n\}, \text{maxStep}$

Output: Minimal HS

```

1 begin
2 initialize all  $V_0$ , get  $P_0$  according (4);
3 initialize a set  $E$ , only  $(\mathbf{1}, \mathbf{1}, \dots, \mathbf{1})$  included;
4 while  $\text{step} < \text{maxStep}$  do
5    $E = \{E \cup P_i\}$  if  $f(P_i) = 1$ 
6   select  $Pe$  from  $E$  such that  $Pe$  has the largest number of 1, breaking ties randomly;
7   update  $V_{i+1}$  according to (1) and get  $P_{i+1}$  according (4);
8 end
```

3 结合特征学习的 PSO 求解极小碰集算法

上一节中已经详细介绍了使用粒子群优化算法求解极小碰集的具体过程,算法的主要操作在于公式(1),该公式通过 Pe 和 $\mathbf{0}$ 两个向量的调节来生成新粒子,粒子对问题的整个解空间都要进行搜索,无论某些

空间中是否存在解。

本文提出一种结合特征学习求解碰集的方法,通过结合特征学习机制指导后续碰集求解问题,有效地减少对无解空间的搜索,同时又采取翻转策略加速对有解空间的搜索.其主要思想是在 PSO 算法中的 $maxStep$ (其中 $maxStep$ 表示本文算法中所设置的最大迭代次数)次迭代中,每隔一定的迭代次数,对碰集长度进行一个学习过程.学习过程结束,得到一个最大碰集长度和一个最小碰集长度;在接下来的搜索过程中,将它们作为两个衡量标准来对每次迭代求到的个体进行过滤,进而避免对无解空间的搜索,同时通过在有解空间中加入翻转策略加速有解空间的求解。

例如, $maxStep = 500$, 在 $step = 1 - 10$ 时,执行一个特征学习过程,接下来的 $step = 11 - 100$ 阶段,执行结合特征信息的求解过程; $step = 101 - 110$ 及 $step = 111 - 200$ 次迭代中,继续执行特征学习过程和结合特征信息的求解过程,直到 $step = maxStep$, 算法结束。

本文在上一节中所介绍的 OA 算法基础上,加入了特征学习方法,结合特征学习指导的后续求解方法和加速求解策略方法,进而得到一个新的 LCA (Learning with Characteristics Algorithm) 算法,其基本描述如下。

Algorithm LCA

Input: $F = \{C_1, C_2, \dots, C_n\}, maxStep$

Output: Minimal HS

```

1 begin
2 initialize all  $V_0$ , get  $P_0$  according (4);
3 initialize a set  $E$ , only  $(1, 1, \dots, 1)$  included;
4  $min = 999, max = 0$ ;
5 while  $step < maxStep$  do
6  $E = \{E \cup P_i\}$  if  $f(P_i) = 1$ 
7 if  $step = fixedStep$ 
8   if  $(length(P_{min}) < min)$  //  $P_{min} \in E$ 
9      $min = length(P_{min})$ ;
10    else if  $(length(P_{max}) > max)$  //  $P_{max} \in E$ 
11       $max = length(P_{max})$ 
12 else
13   if  $(length(P_i) < min)$ 
14     flip 0 - > 1 for  $m$  dimension,  $m = min - length(P_i)$ ;
15   if  $(length(P_i) > max \ \&\& \ f(P_i) < 1)$ 
16     Initialize  $P_i$ ;
17   select  $P_e$  from  $E$  such that  $P_e$  has the maximum number of 1, breaking
ties randomly;
18   update  $V_{i+1}$  according to (1) and get  $P_{i+1}$  according (4);
19 end

```

在 LCA 中, 7~16 为本文中提出的策略, 在 OA 基础上加以改进. 7~11 是特征学习过程, 其中 $fixedStep$ 表示执行特征学习过程的迭代. P_{min} 和 P_{max} 分别为精英

集 E 中长度最小和最大的碰集, 在此过程中, 记录并更新已求得碰集的最大和最小长度, 用来指导后续碰集的求解. 在特征学习过程中, 每一次迭代都记录下所求得碰集的最大和最小长度 (这里所说的碰集长度是指碰集中含有的“1”的个数), 若下次迭代产生的碰集的最小长度小于当前的最小长度, 则更新这个最小长度, 同理, 若下次迭代产生的最大长度大于当前的最大长度, 也进行更新; 若下次迭代产生的碰集长度介于最小长度和最大长度之间, 不需要更新。

算法 12~16 行则是利用之前学习到的信息指导后续碰集的求解, 有效避免了对无解空间的搜索, 同时加速了对有解空间中解的搜索. 在此过程中: (1) 每次迭代求得的粒子的长度若大于最大长度且该粒子不是碰集, 则直接舍弃这些解, 再重新初始化相同数量的解以保证群体规模恒定. 通过这个策略, 其长度大于最大长度且非碰集的粒子就不需要再去记录, 那些不含有极小碰集的搜索空间也无需浪费时间进行搜索; (2) 若求得的碰集的长度小于最小长度, 则对当前粒子中的“0”进行随机翻转为“1”, 翻转的个数为“最小长度-该粒子的长度”, 这样不仅可以减少对小于最小长度的搜索空间的搜索, 还可以加速对有解空间的搜索。

4 实验

本文算法运行于 Linux (Dell Dimension C521, AMD Athlon(tm) 64 X2 Dual Core Processor 3600+, 1.90 GHz, 2.00GB RAM) 系统下, 各测试用例为随机生成的变量出现概率为 0.1~0.9 的冲突集合簇, 每个用例测试 20 次取其平均值作为最终结果. 本文中对两种规模的冲突集问题进行了测试, 分别为 $v30s30$ 、 $v40s40$, 其中 v 代表变量, s 代表冲突集, 则 $v30s30$ 即代表含有 30 变量 30 个冲突集的测试用例. 图 1 和图 2 分别给出了 $v30s30$ 和 $v40s40$ 两种规模的测试用例在两种算法上的碰集求解数量的对比。

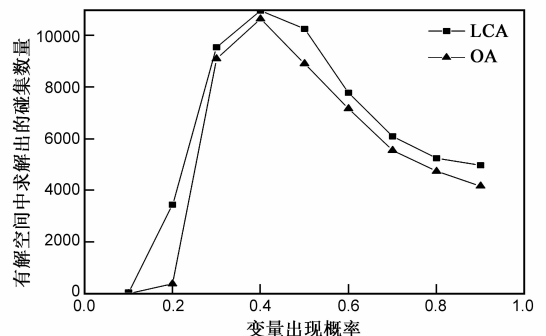


图1 两种算法在 $v30s30$ 上的碰集求解数量对比

从图 1 和图 2 可以看出 LCA 在各个规模、各个出现概率上所求得的碰集的数量都大于 OA. 碰集数量在变

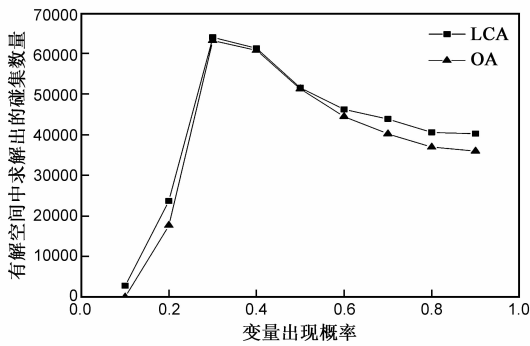


图2 两种算法在v40s40上的碰集求解数量对比

量的出现概率为0.3~0.5时最多,在这些概率下,LCA相比于OA在求解出的碰集数量的提高率上有明显差异.如,变量出现概率为0.5时,在v40s40规模下效率仅提高了0.5%,而在v30s30规模下效率提高了15.1%.当出现概率小于0.3或是大于0.5的时候,求得的碰集数量提高率也逐渐增大.

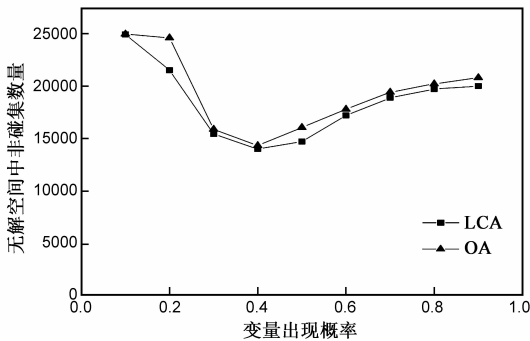


图3 两种算法在v30s30上搜索无解空间中的非碰集数量对比

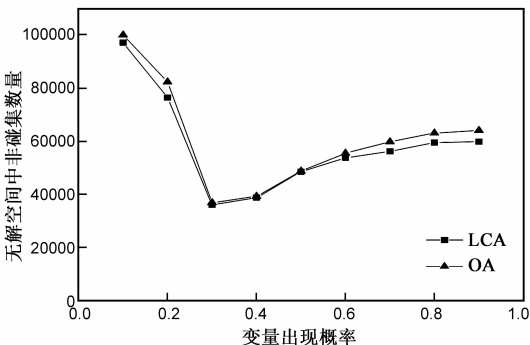


图4 两种算法在v40s40上搜索无解空间中的非碰集数量对比

图3和图4描述OA、LCA两种算法在整个求解过程中搜索的无解空间中的非碰集个体的数量对比.两条折线中间的部分即表示LCA中通过学习策略所避免搜索的不含有碰集的那部分搜索空间.在学习过程中,当个体长度大于最大碰集长度且非碰集时,舍弃该个体重新初始化一个新的个体,避免了对一部分无解空间的搜索.

LCA算法相对于OA算法削减了对部分无解空间的搜索,并通过翻转策略加速求解,但由于LCA求得的碰集较多,且碰集在极小化时复杂度较高,导致LCA的总求解时间略大于OA.表1给出了v40s40规模下求得的碰集数目与时间,其数据为20次试验的平均值.特别值得指出的是在各规模下,当变量的出现概率为0.1时,OA几乎求不出任何的碰集,而LCA可以求出较多数目的碰集.碰集极小化的复杂度高于求解碰集的复杂度,求解碰集只需要根据其适应度函数的值判断某个个体是否为碰集,而碰集的极小化则要将新产生的碰集与之前产生的所有极小碰集进行比较,判断是否为超集.并且越处于极小碰集求解的后期,碰集的极小化复杂度越高.

表1 v40s40下两种算法求得碰集数量与求解时间对比

变量出现概率	OA		LCA		比率(LCA/OA)	
	碰集数量	求解时间(s)	碰集数量	求解时间(s)	数量	时间
0.1	1	2.172	2819.8	5.666	2819.8	2.609
0.2	17703	104.1	23678.6	196.834	1.338	1.89
0.3	63168.4	1405.558	64033.8	1445.748	1.014	0.028
0.4	60787.8	1411.13	61297.8	1419.822	1.008	1.006
0.5	51289.4	1085.496	51550.6	1049.816	1.005	1.009
0.6	44487.2	887.604	46242.2	928.356	1.039	1.045
0.7	40180.8	765.462	43877.8	847.662	1.092	1.107
0.8	36968.8	659.138	40573	772.434	1.097	1.172
0.9	35986.2	678.594	40236.4	769.426	1.118	1.174

本文还通过实验测试给出特征学习过程中迭代次数间隔对算法的影响.选取出现概率为0.4、0.5和0.6,规模为v30s30的测试用例,迭代次数间隔参数分别选取50、100和200.每个概率随机生成10个测试用例并测试10次,取平均值作为测试用例的最终值.从图5可以看出,随着迭代次数间隔参数从50增加到200,求得的碰集个数也随之有较小幅度增加.在求解的特征学习过程中,特征学习次数较多时,则对求解空间限定较严格,进而可能缩减部分可能解;特征学习次数较少时,则对求解空间限定相对宽松,进而可得到更多的解.

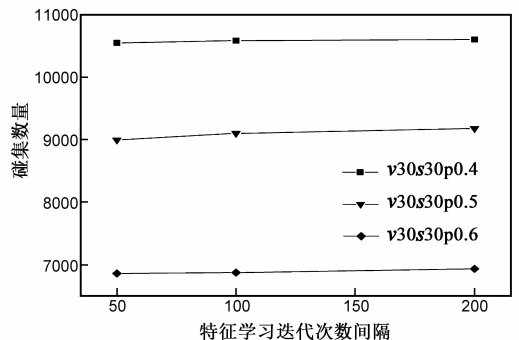


图5 v30s30特征学习迭代次数间隔为50、100、200时求解出极小碰集数量对比

5 结论

本文对基于 PSO 的算法进行改进,加入了特征学习、结合特征学习指导的后续求解和加速求解策略。本文提出的算法不仅加速了对有解空间的搜索,同时削减了对无解空间的搜索,有效提高了碰集的求解数量。相比于基于树或图的碰集求解算法只能求解小规模的问题,本文提出的算法可以求解大规模的碰集问题,编程容易,更具有实际意义。

参考文献

- [1] Bonet B, Helmer M. Strengthening landmark heuristics via hitting sets[A]. 2010 19th European Conference on Artificial Intelligence[C]. Lisbon: Frontiers in Artificial Intelligence and Application, 2010. 329 – 334.
- [2] Kodali A, Singh S, Pattipati K. Dynamic set-covering for real-time multiple fault diagnosis with delayed test outcomes[J]. Systems, Man, and Cybernetics: Systems, IEEE Transactions on, 2013, 43(3): 547 – 562.
- [3] Malik S. Design debugging using boolean satisfiability[R]. Invited Talk, First International SAT/SMT Solver Summer School 2011, MIT, June 2011.
- [4] Reiter R. A theory of diagnosis from first principles[J]. Artificial intelligence, 1987, 32(1): 57 – 95.
- [5] Pill I, Quaritsch T. And yet another variant of reiters complete on-the-fly hitting set algorithm [A]. The 24th International Workshop on Principles of Diagnosis [C]. Jerusalem; online, 2013. 210 – 215.
- [6] 姜云飞, 林笠. 用对分 HS-树计算最小碰集[J]. 软件学报, 2002, 13(12): 2267 – 2274.
- [7] Pill I, Quaritsch T. Optimizations for the boolean approach to computing minimal hitting sets[A]. 2012 20th European Conference on Artificial Intelligence [C]. Montpellier: Frontiers in Artificial Intelligence and Application, 2012. 648 – 653.
- [8] 张立明, 欧阳丹彤, 曾海林. 基于动态极大度的极小碰集求解方法[J]. 计算机研究与发展, 2011, 48(2): 209 – 215. Zhang Li-ming, Ouyang Dan-tong, Zeng Hai-lin. Computing the minimal hitting sets based on dynamic maximum degree[J]. Journal of Computer Research and Development, 2011, 48(2): 209 – 215. (in Chinese)
- [9] Zhao Xiangfu, Ouyang Dantong. A distributed strategy for deriving minimal hitting-sets[A]. The 24th International Workshop on Principles of Diagnosis[C]. Jerusalem; online, 2013. 33 – 38.
- [10] Li L, Yunfei J. Computing minimal hitting sets with genetic algorithm[J]. Algorithmica, 2002, 31(2): 95 – 106.
- [11] 张楠, 孙吉贵, 等. 求极小碰集的遗传算法[J]. 广西师范大学学报: 自然科学版, 2006, 24(4): 62 – 65.
- [12] Xiangtao Li, et al. Hybrid differential evolution with artificial bee colony and its application for design of a reconfigurable

antenna array with discrete phase shifters [J]. IET Microwaves, Antennas & Propagation, 2012, 6(14): 1573 – 1582.

- [13] 刘全, 等. 一种动态挥发率和启发式修正的蚁群优化算法[J]. 计算机研究与发展, 2012, 49(3): 620 – 627.
- [14] 刘全, 王晓燕, 傅启明, 等. 双精英协同进化遗传算法[J]. 软件学报, 2012, 23(4): 765 – 775. Liu Quan, Wang Xiao-yan, Fu Qi-ming, et al. Double elite co-evolutionary genetic algorithm[J]. Journal of Software, 2012, 23(4): 765 – 775. (in Chinese)
- [15] 张长胜, 孙吉贵, 欧阳丹彤. 一种自适应离散粒子群算法及其应用研究[J]. 电子学报, 2009, 37(2): 299 – 304. Zhang Chang-sheng, Sun Ji-gui, Ouyang Dan-tong. A self-adaptive discrete particle swarm optimization algorithm [J]. Acta Electronica Sinica, 2009, 37(2): 299 – 304. (in Chinese)
- [16] 田野, 刘大有. 求解流水线车间调度问题的混合粒子群算法[J]. 电子学报, 2011, 39(5): 1087 – 1093. Tian Ye, Liu Da-you. A hybrid particle swarm optimization method for flow shop scheduling problem[J]. Acta Electronica Sinica, 2011, 39(5): 1087-1093. (in Chinese)
- [17] 张波, 薛正辉, 任武, 等. 采用粒子群算法的频率选择表面优化设计[J]. 电子学报, 2013, 41(3): 603 – 608. Zhang Bo, Xue Zheng-hui, Ren Wu, et al. Particle swarm optimization of frequency selective surface [J]. Acta Electronica Sinica, 2013, 41(3): 603 – 608. (in Chinese)

作者简介



刘娟 女, 1989 年 11 月出生, 山东菏泽人. 2008 年毕业于长春理工大学软件开发与测试专业, 现为吉林大学计算机科学与技术学院硕士研究生, 研究方向为基于模型诊断.

E-mail: liujuanfree@126.com



欧阳丹彤 女, 1968 年 6 月出生, 吉林长春人. 教授、博士生导师、中国计算机学会理论计算机科学、人工智能与模式识别专业委员会委员, 现为吉林大学符号计算与知识工程教育部重点实验室主任, 主要研究方向为基于模型诊断、模型验证和自动定理证明.

E-mail: ouyangdantong@163.com

张立明 (通讯作者) 男, 1980 年 9 月出生, 吉林榆树人. 吉林大学博士后. 主要研究方向为基于模型诊断, 模型验证和自动定理证明.

E-mail: limingzhang@jlu.edu.cn