

# 基于多种群量子粒子群优化的属性约简<sup>①</sup>

李三波

(丽水职业技术学院 机电信息分院, 丽水 323000)

**摘要:** 现代工业发展要求迅速、可靠地实现故障诊断。针对粒子群约简算法易陷入局部最优等问题, 提出了一种多种群量子粒子群优化算法 (MIQPSO)。该算法对量子粒子群算法进行分群, 并通过接种疫苗, 指导粒子朝更优化方向进化, 提高了量子粒子群的收敛速度和寻优能力。利用 UCI 相关数据集, 通过对 Hu 算法、粒子群算法、量子粒子群算法、多种群量子粒子群算法的粗糙集属性约简验证, 结果表明, 基于多种群量子粒子群优化的约简算法具有良好的约简效果。

**关键词:** 粒子群算法; 多种群量子粒子群优化; 粗糙集; 属性约简

## Attribute Reduction Based on Quantum-Behaved Particle Swarm Optimization with Multi-Swarm Algorithm

LI San-Bo

(School of Mechanical Electronic and Information Engineering, Lishui Vocational & Technical College, Lishui 323000, China)

**Abstract:** Requirements of modern industrial development rapidly and reliably achieve the fault diagnosis. Against particle swarm algorithm for the reduction and other issues so easy to fall into local optimum problem, this paper aims to present the MIQPSO Algorithm. The quantum particle swarm algorithm for clustering by the MIQPSO Algorithm, and through vaccination, to guide the direction of the particle evolution towards more optimized, improve the convergence rates and optimization searching ability of the quantum particle swarm. The use of UCI data sets, and by Hu algorithm, particle swarm optimization, quantum particle swarm optimization, multi-species quantum particle swarm algorithm for rough set attribute reduction verification, the results show that the algorithm based on the quantum particle swarm optimization has good reduction effect on the reduction.

**Key words:** particle swarm optimization; quantum-behaved particle swarm optimization; rough set; attribute reduction

### 1 引言

属性约简是粗糙集理论研究的一个重要方面, 人们总是希望在保持系统分类能力不变的条件下, 删除其中冗余甚至是不相关的属性, 找到一组包含最少属性个数的约简, 即最小属性约简, 并已证明找出决策表的最小约简是 NP 问题<sup>[1]</sup>。尽管目前已经有许多较为成熟的属性约简算法, 但是探索一种快速、有效的启发式搜索算法仍是属性约简的一个重要趋势。Hu 算法<sup>[2]</sup>是传统的属性约简算法的代表之

一, 但它所求出的属性约简并不一定是最小约简。粒子群算法通过模拟生物种群社会信息交流的过程来对约简进行迭代搜索, 最后得到最小约简或者近似最小约简, 可得到较好的寻优性能, 但易陷入局部最优。变异粒子群算法、自适应调整粒子群算法<sup>[3]</sup>、免疫粒子群算法<sup>[4]</sup>等算法在一定程度上可避免 PSO 陷入局部最优, 但也存在一些不足之处。

针对 PSO 易陷入局部最优问题, 本文依据多种群分群的改进策略, 提出一种多种群量子粒子群优

① 基金项目: 浙江省高校学科带头人资助项目(2007-209); 学校培育基金(LZYA201003)

收稿时间: 2011-07-30; 收到修改稿时间: 2011-09-07

化(MQPSO) 算法, 通过引入扩张变异和粒子扰动策略<sup>[5]</sup>, 迫使粒子群摆脱局部极小点。为把多种群量子粒子群优化应用于属性约简, 验证多种群量子粒子群的有效性, 本文分别采用 Hu 算法、粒子群约简算法、量子粒子群约简算法和多种群量子粒子群约简算法对 UCI 数据进行实验, 并且比较分析实验结果。结果表明, 基于多种群量子粒子群优化的约简算法在收敛速度和寻优能力都取得了更好的效果。

## 2 粗糙集基本理论

粗糙集理论是波兰数学家 Z. Pawlak 于 1982 年提出的一种分析不完整、不确定数据的理论, 其主要思想是在保持分类能力不变的前提下, 通过知识约简, 导出问题的决策或分类规则。

粗糙集理论研究的对象是知识表达系统, 因此, 要在实际中应用粗糙集理论, 就首先要研究的对象表示成粗糙集理论所能研究的形式, 也就是知识表达系统。

设信息系统  $S = \{U, A, V, f\}$ ; 其中  $U = \{x_1, x_2, \dots, x_n\}$  为对象的非空有限集合,  $A = \{a_1, a_2, \dots, a_n\}$  称为论域;  $V_a$  为属性  $a \in A$  的非空有限集合,  $V = \bigcup_{a \in A} V_a$  是属性的值域, 是所有属性值域的并集;  $f: U \times A \rightarrow A$  是一个信息函数, 它为每个对象的每个属性赋予一个信息值, 即  $\forall a \in A, x \in U, f(x, a) \in V_a$ 。

知识表达系统的数据能够通过关系表的形式进行表示, 我们要研究的对象对应为关系表的行, 关系表的列则对应着对象的属性, 对象的信息是通过指定对象的各个属性值来表达确定的。

我们可以看出, 每一个属性对应着一个等价关系, 一个表我们可以看做是定义的一族等价关系, 也就是知识库, 因此, 知识约简也就可以转化为属性约简。

决策表是知识表达系统中一类特殊而重要的形式, 多数决策问题都可以通过决策表的形式来表达, 这个工具在决策应用中起到了非常重要的作用。

参照知识表达系统, 决策表可以定义如下:

在一完备知识表达系统  $S = \{U, A, V, f\}$  中, 如果属性集  $A$  被分为条件属性集合  $C$  和决策属性集合  $D$ , 即  $A = C \cup D$  且  $C \cap D = \emptyset$ , 则称

$S = \{U, C \cup D, V, f\}$  是决策信息系统, 这种具有条件属性和决策属性的知识表达系统也被称为决策表。

不可分辨关系: 设  $S = \{U, C \cup D, V, f\}$  是一个决策信息系统,  $P \subseteq C$ , 不可分辨关系  $ind(P)$  为:  $ind(P) = \{(x, y) \in U \times U \mid \forall p \in P, f(x, p) = f(y, p)\}$ 。

上近似、下近似、边界域: 对于决策信息系统  $S = \{U, C \cup D, V, f\}, X \subseteq U, P \subseteq C$  为一定义在  $U$  上的等价关系, 则  $X$  关于  $ind(P)$  的上下近似集分别为:

$$\bar{P}X = \{x \in U \mid [x]_P \cap X \neq \emptyset\}$$

$$PX = \{x \in U \mid [x]_P \subseteq X\}$$

$bn_P(X) = \bar{P}X - PX$  称为集合  $X$  的边界域, 也就是论域的不确定域

如果边界区  $bn_P$  为空, 则称  $X$  关于  $P$  是清晰的; 反之, 则集合  $X$  关于  $P$  是粗糙的。

属性约简, 就是在保持知识库分类能力不变的条件下, 删除其中冗余甚至是不相关的知识。知识约简中包含两个最基本概念: 约简和核。

设  $R$  是一个等价关系,  $r \in R$ , 如果  $ind(R) = ind(R - \{r\})$ , 则称  $r$  为  $R$  中不必要的,  $r$  在  $R$  中是可被约去的知识, 否则称  $r$  为  $R$  中必要的。

如果每一个  $r \in R$  都为  $R$  中必要的, 则称  $R$  为独立; 否则称  $R$  为依赖的。

设  $P \subseteq R$ , 如果  $P$  是独立的, 且  $ind(P) = ind(R)$ , 则称  $P$  为  $R$  的一个约简。显然,  $R$  可以有多种约简,  $R$  中所有不可约去的关系称为核, 由它构成的集合称为  $R$  的核集, 记为  $Core(R)$ 。

## 3 多种群量子粒子群算法优化的属性约简

### 3.1 粒子群算法

粒子群算法是群智能算法的一个重要分支, 和遗传算法相类似, 这也是一种基于迭代优化算法, 粒子的初始化大部分为一组随机的解, 此外可以通过一定的先验信息, 同时给出一些特定分布的初始粒子来加速收敛速度。

对于一个具体的问题, 首先根据实际情况设置一个合理的粒子总数  $m$ , 以及每个粒子的维度  $d$ 。在迭代过程中  $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$  和  $v_i = (v_{i1}, v_{i2}, \dots, v_{id})$  分别表示了粒子的当前位置以及当前飞行速度  $p_i = (p_{i1}, p_{i2}, \dots, p_{id})$ , 为粒子  $i$  所经历的当前最好位

置,  $p_{gbest} = p_g = (p_{g1}, p_{g2}, \dots, p_{gd})$  则是整个粒子群中所有粒子经历的最优位置。

在 PSO 算法中, 粒子对位置和速度的更新如式(1)所示:

$$\begin{cases} v_{id}^{k+1} = v_{id}^k + c_1 \text{rand}() (p_{id} - x_{id}^k) + c_2 \text{rand}() (p_{gbest} - x_{id}^k) \\ x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \end{cases} \quad (1)$$

式中,  $c_1$  和  $c_2$  是两个正常数, 称为加速因子, 表现了社会信息对粒子的影响程度  $v_{id}^k$ 。从中我们也可以看出, 粒子更新后的移动方向由原有的速度  $p_{id} - x_{id}^k$ , 经历的最佳距离以及群体经历的最佳距离  $p_{gbest} - x_{id}^k$  共同决定。

由于在 PSO 算法中, 粒子搜索区域局限于全局最优点和局部最优附近, 使种群容易陷入局部最优。

### 3.2 量子粒子群算法

QPSO 算法对 PSO 算法进化搜索策略做了根本改变, 在搜索能力上优于传统 PSO 算法。量子粒子群算法能够克服一般粒子群算法在收敛性能上的缺陷这主要是由于量子系统特有的优点决定的。

1) 量子系统是一个复杂的非线性系统, 它符合状态重叠原理, 因此, 比一般的线性系统和低价非线性系统具有更多状态。

2) 量子系统是一个非确定性系统, 所以粒子的运动没有一条确定的轨迹, 而是通过一定的概率出现在整个可行解空间中的任意一个位置<sup>[4,6]</sup>。所以 QPSO 算法比基本 PSO 算法具有更广泛的搜索空间, 更容易跳出局部最优。

QPSO 模型中的量子即为粒子, 波函数  $\varphi(x, t)$  能够对粒子的状态进行描述和确定, 而粒子在空间某个点出现的概率密度函数可以由薛定谔方程求解得到, 最后, 通过 Monte Carlo 随机模拟得到粒子的位置方程为:

$$X(t) = P \pm \frac{L}{2} \ln\left(\frac{1}{u}\right)$$

其中  $L$  值由式  $L(t+1) = 2\beta |mbest - X(t)|$  确定。

具体的模拟过程可以表示如下:

$$\begin{cases} x_i^{t+1} = p_i + \beta |mbest - x_i^t| \ln\left(\frac{1}{u}\right) & \text{if } r \geq 0.5 \\ x_i^{t+1} = p_i - \beta |mbest - x_i^t| \ln\left(\frac{1}{u}\right) & \text{if } r < 0.5 \end{cases} \quad (2)$$

$$p_{id} = cp_{id} + (1-c)p_{gd} \quad (3)$$

式中,  $x_i^t$  为第  $i$  个粒子在  $t$  时刻的位置,  $\beta$  称为收缩扩张因子, 主要用来控制算法的收敛速度, 目前算法中  $\beta$  一般通过如下的方法在迭代过程中动态调节。

$$\beta = \frac{(1-0.5) \times (\text{Maxiter} - t)}{\text{Maxiter}} + 0.5 \quad (4)$$

其中  $u$ ,  $r$  和  $c$  均为  $(0, 1)$  间的随机数。

$mbest$  称为群体中最优中心, 是种群中所有粒子局部最优位置的中心, 可表示为

$$mbest = \frac{1}{m} \sum_{i=1}^m p_i = \left( \frac{1}{m} \sum_{i=1}^m p_{i1}, \frac{1}{m} \sum_{i=1}^m p_{i2}, \dots, \frac{1}{m} \sum_{i=1}^m p_{id} \right) \quad (5)$$

### 3.3 多种群量子粒子群算法

对粒子群进行多种群分群, 可以扩大粒子的寻优空间, 各个子种群的粒子之间不但具有独立性的一面, 同时也具有合作性的一面。每个子种群中的粒子都按照各自种群中的规则进行独立的计算; 另一方面, 子种群之间能够进行信息的交流, 每个粒子群中个体的速度更新, 不仅应受该子种群中所有粒子个体经过的最好位置  $p_g$  的影响, 还受所有子种群中的最优位置  $p_g$  的影响。

在跳出局部最优方面, 多种群的搜索策略远远优于单个种群: 只要  $n$  个种群没有任何两个陷入相同的位置, 那么种群陷入局部最优的概率将小于单个种群的  $1/n^2$ 。但是, 多个种群的并行搜索也将耗费比单种群更多的时间<sup>[5]</sup>。为了能够更好地跳出局部最优, 还可以引入了扩张变异和扰动的思想。当粒子处于当前种群全局最优位置时容易陷入局部最优, 进行适当的扩张变异能使其更容易跳出局部最优; 扰动的思想主要原理是当算法每迭代到一定次数时, 对不动粒子或者不再更新全局最优的种群进行扰动, 促进其对未知的空间进行探索。这些思想具有一定的优越性, 能让算法的收敛速度更快, 稳定性能更强。

在多种群粒子群的分群策略上, 许多学者给出了创新性的思路。在李爱国提出的多粒子群协同优化算法中, 粒子群被分成两个层次: 下层部分的几个种群通过独立搜索扩大了搜索的空间; 上层部分的粒子群通过追随当前全局最优来达到加速全局收敛的目的<sup>[6]</sup>。陈国初等提出的两群微粒群优化算法中, 通过两个搜索空间相同但飞行方向相反的粒子

群,进行反向搜索,同时两个种群的粒子能够以一定的概率进行交换<sup>[7]</sup>。上述的两种算法在一定程度上都能够上扩大了粒子的搜索空间,此外,还有一些效果比较不错的分群方法和进化策略<sup>[8]</sup>。

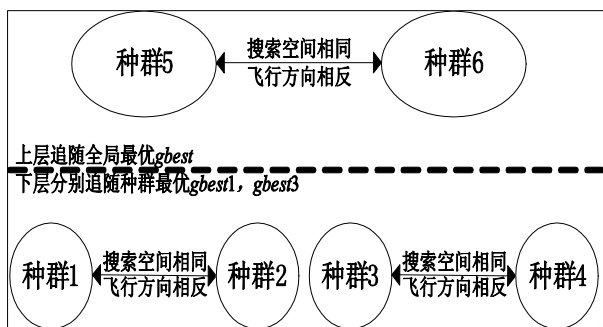


图1 多种群分群示意图

受到这些启发,本文提出了一种多种群粒子群算法。在该算法中,主要分为两个层次,如图1所示:下层的1号种群和2号种群,3号种群和4号种群分别为两对搜索空间相同但是飞行方向相反的种群;上层的5号种群追随全局最优,6号种群则是朝着和5号种群相反的方向搜索。这样不但利用了下层4个种群的独立搜索策略来保证寻优过程可以在较大范围进行,又可以利用上层粒子群追逐当前全局最优点来保证算法的收敛性,兼顾了优化过程的精度和效率,在优化结果和收敛特性上均具有优势,有效地避免了收敛过程中早熟引起的陷入局部最优。

### 3.4 基于多种群粒子群优化的属性约简

在多种群方案的具体设计上,既利用下层群的独立搜索保证寻优过程可以在较大范围进行,又可以利用上层粒子群追逐当前全局最优点来保证算法的收敛性,兼顾了优化过程的精度和效率,为了防止错过全局最优,还设置了向群体最优和全局最优收敛的过程中同时还向相反的方向搜索。MQPSO的属性约简算法流程如下:

**Step1:** 初始化6个粒子群,给定m个粒子的初始位置与初始速度;设置最大迭代次数  $Maxiter$ ,  $t=1$ ,初始化历史最优、种群最优和全局最优;

**Step2:** 通过适应函数,计算每个粒子的适应值  $fitness$ ,更新历史最优、种群最优和全局最优;

**Step3:** 按不同要求分别对不同种群的粒子更新位置;

**Step4:** 编码转换;

**Step5:** 判断是否要进行粒子变异或者种群变异;

**Step6:**  $t=t+1$ ,如果,  $t < Maxiter$  或未达到预定最优值,则转 Step2;

**Step7:** 输出全局最优粒子,即为所求的最优属性约简。

## 4 实验结果及分析

为了验证算法的有效性和正确性,本文采用UCI机器学习数据库中的标准数据集作为测试数据,UCI数据库是加州大学欧文分校(University of CaliforniaIrvine)提出的用于机器学习的数据库,这个数据库目前共有187个数据集,UCI数据集是一个常用的标准测试数据集。这些数据来自 <http://archive.ics.uci.edu/ml/>,并设计了类比实验,比较Hu算法、PSO、QPSO、MQPSO算法在属性约简方面的效果。本文选取了Sponge、Zoo、Lymphography、Wine、Soybean\_large和Vote等5个数据集。

实验环境如下:CPU主频为1.73GHz,内存为512MB,操作系统为Windows XP,运行平台为Matlab 7.2。参数设置为:粒子数  $m=30$ ,其中MQPSO算法把粒子分为6个种群,每个种群包含5个粒子,适应度函数中  $k_1=0.8$ ,  $k_2=0.2$ ,最大迭代次数  $Maxiter=100$ 。实验结果如表1所示。

数据集Sponge原有属性数45个,共有记录76个,各种算法对其属性约简效果如表2所示,为了更好地表现出迭代的过程中,图2给出了具体的进化过程适应度曲线。

为了比较PSO、QPSO、MQPSO三个约简算法的收敛速度和寻优能力,本文对三个算法的进化过程进行详细分析。

数据集Lymphography原有属性数19个,共有记录148个,各种算法对其属性约简效果如表3所示,为了更好地表现出迭代的过程中,图3给出了具体的进化过程适应度曲线。

表 1 属性简约结果

数据集	属性	记录数	算法	约简属性数	运行时间/ms
Sponge	45	76	Hu 算法	8	93.355
			PSO	12	178.487
			QPSO	12	187.596
			MQPSO	8	200.178
Zoo	17	101	Hu 算法	8	17.922
			PSO	8	273.538
			QPSO	10	295.608
			MQPSO	8	303.013
Lymphography	19	148	Hu 算法	6	25.945
			PSO	6	589.673
			QPSO	6	629.761
			MQPSO	6	643.665
Wime	14	178	Hu 算法	5	33.357
			PSO	5	871.369
			QPSO	6	900.763
			MQPSO	5	914.456
Sogbeanbarge	36	307	Hu 算法	10	366.088
			PSO	12	2675.304
			QPSO	10	3175.218
			MQPSO	10	3203.289
Vote	17	435	Hu 算法	9	192.698
			PSO	10	5334.526
			QPSO	9	5528.956
			MQPSO	9	5588.197

表 2 数据集 Sponge 属性约简结果

Sponge	属性数	达优率	约简属性数	最小约简属性数	最快收敛步数	平均收敛步数	运行时间
Hu 算法	45	/	8	8	/	/	93.355
PSO		/	12		54	82	187.596
QPSO		70	8		42	71	200.178
MQPSO		90	8		21	54	337.554

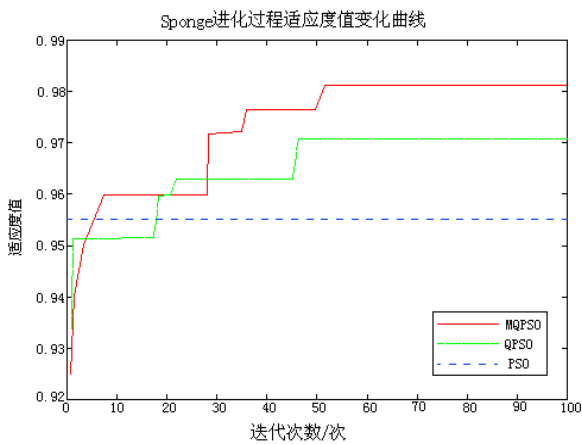


图 2 Sponge 进化过程适应度值变化曲线

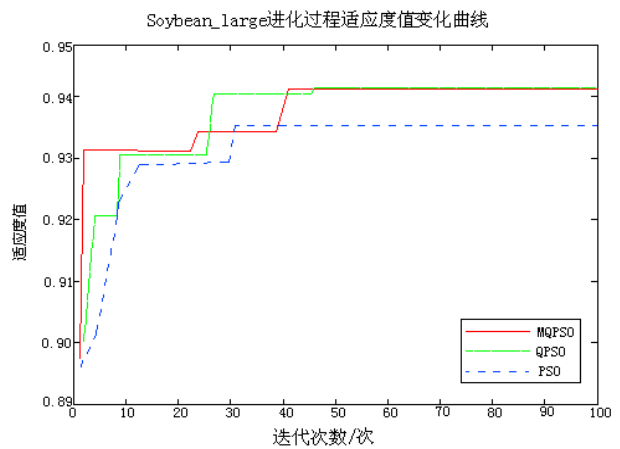


图 3 Soybean\_large 进化过程适应度值变化曲线

表3 数据集 Lymphography 属性约简结果

Lymphography	属性数	达优率(%)	约简属性数	最小约简属性数	最快收敛数	平均收敛数	运行时间
Hu 算法	19	/	6	6	/	/	25.945
PSO		70	6		33	95	629.761
QPSO		90	6		27	60	643.665
MQPSO		100	6		11	41	656.244

限于篇幅, 本文仅画出两个数据集 sponge 和 Soybean\_large 进化过程适应度值的变化曲线, 对其他数据集并没有列出。

表1、表2、表3实验结果表明, MQPSO 约简算法比 Hu 算法、PSO、QPSO 约简算法取得了更好的结果, 表现在 MQPSO 约简算法取得了包含更少属性个数的约简结果, MQPSO 算法具有更好的收敛速度和寻优能力。图2和图3的曲线图是迭代次数与适应度值之间的关系图, 适应度值越大说明约简结果越优。从两图中可以看出 MQPSO 约简算法比 QPSO 和 PSO 约简算法具有更好的收敛速度和寻优能力。

从运行时间的角度来看, Hu 算法所用的时间最少, MQPSO 算法所用的时间最多。这是由于 Hu 算法采用了属性重要性优先选择的贪心策略, 这种策略能够帮助算法很快得到结果, 但是得到的结果往往不是最小的属性约简; 另外, MQPSO 算法是在 QPSO 的基础上, 通过引入免疫算法, 再进行分群搜索的策略, 加上几种算法的迭代次数相同, 所以 MQPSO 算法需要耗费的时间最多。

## 5 结语

本文把多种群分群的思想运用于量子粒子群算法, 提出了基于多种群免疫量子粒子群的粗糙集属性约简算法, 实验表明, 该算法在收敛速度和寻优能力

都得到了改进, 这种结合是可行的, 并能得到较好的效果。

## 参考文献

- 1 Wong SKM, Ziarko W. On optimal decision rules indecision tables. Bulletin of Polish Academy of Science, 1985, 33(11-12): 693-696.
- 2 Hu XH, Cercone N. Learning in relational database: a rough set approach. Computational Intelligence, 1995, 11(2): 323-337.
- 3 Yang XM, Yuan JS, Yuan JY, et al. A modified particle swarm optimizer with dynamic adaptation. Applied Mathematics and Computation, 2007, 189: 1205-1213.
- 4 廖建坤, 叶东毅. 基于免疫粒子群优化的最小属性约简算法. 计算机应用, 2007, 27(3): 550-555.
- 5 张洪波. 多种群粒子群分层进化优化算法. 中国科技信息, 2010, 8: 40-42.
- 6 李爱国. 多粒子群协同优化算法. 复旦大学学报, 2004, 43(5): 923-925.
- 7 陈国初, 俞金寿. 两群微粒群优化算法及其应用. 控制理论与应用, 2007, 24(2): 294-298.
- 8 许珂, 刘栋. 多粒子群协同进化算法. 计算机工程与应用, 2009, 45(3): 51-54.

(上接第114页)

## 参考文献

- 1 范玉顺. workflow管理技术基础. 北京: 清华大学出版社, 2001.
- 2 周东华. 容错控制理论及其应用. 自动化学报, 2000, 26(6): 788-797.
- 3 张成, 廖建新, 朱晓民. 一种基于增量贝叶斯疑似度的事件驱动故障定位算法. 电子与信息学报, 2009, 31(6): 1501-1504.
- 4 周建涛, 史美林, 叶新铭. 柔性 workflow 技术研究的现状与趋势. 计算机集成制造学院, 2005, 11(11): 1502-1510.
- 5 王松磊, 陈永生. 基于消息队列的独立通信层的设计与应用. 计算机系统应用, 2011, 20(12): 12-22.
- 6 董云卫, 郝克刚. 一种乐观嵌套 workflow 事务模型. 计算机科学, 2005, 32(8): 90-93.
- 7 朱锐, 郭长国, 王怀民. 一种基于补偿代价的长事务调度算法. 软件学报, 2009, (3): 744-753.
- 8 杨林, 王晶, 沈奇威, 朱晓民. 基于 OSGI 的移动广告平台订单系统. 计算机系统应用, 2011, 20(20): 32-35.