

基于两层混合遗传算法的IT外包进度风险控制

王雷震^{1,2}, 朱锦文¹, 卢福强^{1,2}, 汪定伟¹
(1. 东北大学 信息科学与工程学院, 辽宁 沈阳 110819; 2. 东北大学秦皇岛分校 管理学院, 河北 秦皇岛 066004)

摘 要: 针对 IT(information technology)外包项目的两层进度风险控制优化问题,设计了两层混合遗传算法. 该算法是在传统遗传算法中引入模拟退火和自适应机制,并结合优化问题的两层特点而设计的,能够克服传统遗传算法易于早熟、局部搜索能力较差的弱点. 在算例分析中,首先分析了两层数学模型在 IT 外包项目进度风险控制中的管理意义,进而将两层混合遗传算法的仿真结果与两层粒子群优化算法和传统遗传算法的仿真结果进行比较,验证了改进算法的效率和有效性.

关 键 词: IT 外包;进度风险;混合算法;遗传算法;模拟退火

中图分类号: N 945 **文献标志码:** A **文章编号:** 1005-3026(2019)02-0164-06

IT Outsourcing Schedule Risk Control Based on Two-Level Hybrid Genetic Algorithm

WANG Lei-zhen^{1,2}, ZHU Jin-wen¹, LU Fu-qiang^{1,2}, WANG Ding-wei¹
(1. School of Information Science & Engineering, Northeastern University, Shenyang 110819, China; 2. School of Management, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China. Corresponding author: LU Fu-qiang, E-mail: 19217238@qq.com)

Abstract: Focusing on the optimization problem of schedule risk control in information technology(IT)outsourcing project, a two-level hybrid genetic algorithm(TLHGA)is proposed. The TLHGA incorporates simulated annealing, adaptive mechanism and the two-level feature of optimization problem to improve the traditional genetic algorithm(TGA), which could overcome the shortcomings of TGA such as early mature and weak local searching ability. In the experimental analyses, the management meanings of the two-level mathematical model in IT outsourcing schedule risk control is analyzed. Next, the simulation results of TLHGA are compared with the TGA and two-level particle swarm optimization algorithm, which verifies the rationality and effectiveness of the improved algorithm.

Key words: IT outsourcing; schedule risk control; hybrid algorithm; genetic algorithm; simulated annealing

在 IT 外包项目管理中,能否按时完工一直是委托方和承包商的关注重点,工期控制的成功与否在一定程度上决定着项目的成败^[1-2]. 在项目进度控制方面,项目管理者常用的方法和技术包括 Gantt 图、关键路径法、基于挣值(earned value)的技术、基于里程碑的方法^[3-5]等. 在风险分析方面,有等风险图法、模糊分析法、风险报酬法和蒙特卡罗模拟法等多种方法^[6-7]. 本文采用的 IT 外包进度风险控制的结构化分布式决策模型^[1],不仅着眼于项目本身作业过程的进度风险的评估和控制,还考虑了 IT 外包项目在执行过程中所涉及到的多方交互合作的实际情况.

用遗传算法求解项目进度相关的例子有很多. 周方明等将主成分分析(PCA)、遗传算法(GA)与 BP 神经网络相结合,提出基于 PCA-GA-BP 的工程项目工期风险预测模型,并证明其对实际工程具有指导意义^[6]; Dao 等曾在项目日程安排的优化上使用遗传算法,为该类问题提

供了一种新的思路^[7].

IT外包项目中进度风险控制是一个分布式优化问题,具有两层复杂结构,且决策变量存在离散和连续的混合情况,因此,该优化问题的求解难度较高.本文根据问题特点对传统遗传算法(traditional genetic algorithm, TGA)进行改进.首先,设计两层遗传算法以分解问题的解空间,降低求解难度;然后,为了避免TGA易于早熟和局部搜索能力不强的问题,引入模拟退火和自适应机制,设计了TLHGA(two-level hybrid genetic algorithm),以期为本优化问题提供一种有效的求解手段.

1 IT外包项目进度风险控制问题

一个委托方按照开发流程将项目进行分解,委托给多个承包商并使其完成相应的目标^[1].模型分为上、下两层,上层的决策者为委托方,由委托方将风险控制的预算资金分配给各承包商;下层的决策者为承包商,有 M 个承包商接受资金分配并对所承包的子项目中的 N 个活动选择相应的进度风险控制措施.以承包商 i 为例,将其得到的风险控制投入资金 x_i 分配至其负责子项目的关键路径上的 N 个活动中.

问题定义:子项目 i 的进度风险由风险暴露(risk exposure)来定义,即子项目出现拖期的概率乘以损失的程度(拖期的时间):

$$\text{Risk}_{ij} = P(T_i > T_i^0) \times (T_i - T_i^0) = P(\bar{A}_i) \times (T_i - T_i^0). \quad (1)$$

模型假设:

1) 委托方在与各承包商签订的合同中规定各合同总价款均为 U_i ,若各子项目发生延期,则需赔款 $V_i, i = 1, 2, \dots, M, M$ 为承包商数量; x_i 为承包商 i 的预算资金; B_{\max} 为总的预算资金的最大上限值; R_{\max} 是各承包商负责的子项目的进度风险水平的最大上限值; Benefit_i 表示承包商 i 所获得的收益; $P(A_i)$ 表示承包商 i 项目的完成概率, $P(\bar{A}_i)$ 表示承包商 i 的拖期概率.

2) 子项目 i 关键路径上的每一个活动 j 的直接费用为 $c_{ij} (j = 1, \dots, N_i)$,子项目计划工期为 T_i^0 ,子项目实际工期为 T_i .子项目 i 间接费用一般随着总工期的增加而增长,假设单位时间的间接费用为 $C_i, t_{ij}(y_{ij})$ 是子项目 i 每个活动 j 的拖期时间, $\text{cost}_{ij}(y_{ij})$ 表示每个活动 j 上所投入的资金.

3) 承包商 i 的进度风险水平是 $\text{Risk}_i(x_i)$.

4) 子项目之间在工序上是串行的.

上层模型:

$$\min \sum_{i=1}^M \text{Risk}_i(x_i); \quad (2)$$

$$\text{s. t.} \quad \sum_{i=1}^M x_i \leq B_{\max}, \quad (3)$$

$$\text{Risk}_i(x_i)/T_i^0 \leq R_{\max}, \quad (4)$$

$$x_i \geq 0, i = 1, 2, \dots, M. \quad (5)$$

上层模型以项目整体的进度风险为目标函数,即式(2).式(3)表示对总预算资金投入的限制,式(4)指各承包商负责的子项目的进度风险不能超过给定的水平 R_{\max} .上层模型的最优解为 $(x_1^*, x_2^*, \dots, x_M^*)$.

下层模型:

$$\max \text{Benefit}_i = P(A_i)U_i + x_i - P(\bar{A}_i)V_i - \sum_{j=1}^{N_i} c_{ij} - C_i \sum_{j=1}^{N_i} t_{ij}(y_{ij}) - \sum_{j=1}^{N_i} \text{cost}_{ij}(y_{ij}); \quad (6)$$

$$\text{s. t.} \quad \text{Benefit}_i/U_i \geq r, \quad (7)$$

$$\text{Risk}_i(y_{ij})/T_i^0 \leq R_{\max}, \quad (8)$$

$$\sum_{j=1}^{N_i} \text{cost}_{ij}(y_{ij}) \leq x_i^*, \quad (9)$$

$$y_{ij} \in [0, Q], y_{ij} \in \mathbf{Z}, j = 1, 2, \dots, N_i. \quad (10)$$

下层模型以承包商 i 总收益最大化为目标函数,即式(6).式(7)表示承包商的净收益不小于其可接受的最低收益水平;式(8)指项目进度风险不大于预期的风险水平;约束(9)表示风险控制费用在委托方风险投入的可承受范围内;式(10)表示备选的风险控制措施,其中表示子项目 i 中活动 j 备选的风险控制措施的数量,若 $y_{ij} = 0$,则表示不对活动 j 采取任何风险控制措施, Q 为最大可选择的风险控制措施.第 i 个下层模型的最优解为 $(y_{i1}^*, y_{i2}^*, \dots, y_{iN_i}^*)$.

2 两层混合遗传算法设计

IT外包项目进度风险控制问题是一个两层的、混合数学规划问题.当子项目和措施的数量较多时,该问题的解空间将迅速增大,导致其是一个NP难问题,因此,选择GA作为本问题的求解方法.本文根据问题特点对TGA进行了改进.首先,设计两层GA使问题的解空间得以分解,降低求解难度.然后,为了避免TGA易于早熟和局部搜索能力不强的问题^[5-7],引入模拟退火和自适应机制,对上层模型采用模拟退火算法,对下层模型采用自适应遗传算法,设计了TLHGA,以期为本优化问题提供一种有效的求解手段.

2.1 模拟退火机制

对于上层种群中的最优个体,实施模拟退火

机制. 模拟退火具有较强的局部搜索能力,易于跳出局部最优解,能弥补 TGA 易于早熟的缺陷. 由于本文上层的自变量是具体的资金额,且有一定的限制. 因此,模拟退火算法中产生函数的机制为:随机产生 a , 将其与自变量相加或相减,使最终结果符合限制条件. 模拟退火机制的具体操作过程为:①利用产生函数,产生新个体;②判断新个体的适应值是否减小(求最小值),若适应值减小,则用新个体取代原个体,否则按照接受准则判断是否接受新个体;③在降温机制和每个温度条件一定的交换次数限制下,重复步骤①和步骤②. 接受准则采用 Boltzmann 概率分布^[7],公式为

$$h(\Delta E, T) = \frac{1}{1 + \exp[\Delta E/T]} \quad (11)$$

式中: E 为温度 T 时的风险水平, ΔE 为其增量. 降温机制是指温度按 $T = T \cdot T_R$ 变化, T 为初始温度, T_R 为温度变化率.

2.2 自适应机制

遗传算法的核心部分在于选择、交叉和变异,而算法中的交叉和变异过程受到交叉概率 P_c 与变异概率 P_m 影响,所以实际上, P_c 与 P_m 的选择是影响遗传算法性能的关键^[7]. 自适应机制使得

P_c 与 P_m 的值在程序运行过程中能够随着当代种群适应值的改变而改变.

1) 将单点交叉与自适应交叉相结合. 自适应交叉率的公式为

$$P_c = \begin{cases} P_{c1} - \frac{P_{c1} - P_{c2}}{f_m - f_{avg}} \cdot (f' - f_{avg}), & f' \geq f_{avg}; \\ P_{c1}, & f' < f_{avg}. \end{cases} \quad (12)$$

2) 将基本位变异与自适应变异相结合. 自适应变异率的公式为

$$P_m = \begin{cases} P_{m1} - \frac{P_{m1} - P_{m2}}{f_m - f_{avg}} \cdot (f_{avg} - f), & f \geq f_{avg}; \\ P_{m1}, & f < f_{avg}. \end{cases} \quad (13)$$

在式(12)和式(13)中, P_{c1} 和 P_{m1} 表示种群取平均适应值时的交叉率和变异率, P_{c2} 和 P_{m2} 表示种群取最大适应值时的交叉率和变异率. f_m 表示当次迭代时种群的最大适应值, f_{avg} 表示当次迭代时种群的平均适应值, f' 表示要交叉的两个个体中较大的适应值, f 表示要变异个体的适应值.

2.3 TLHGA 流程图

TLHGA 算法流程图如图 1 所示,其中图 1b 是图 1a 中的虚线部分,代表下层结构. 在算法中,

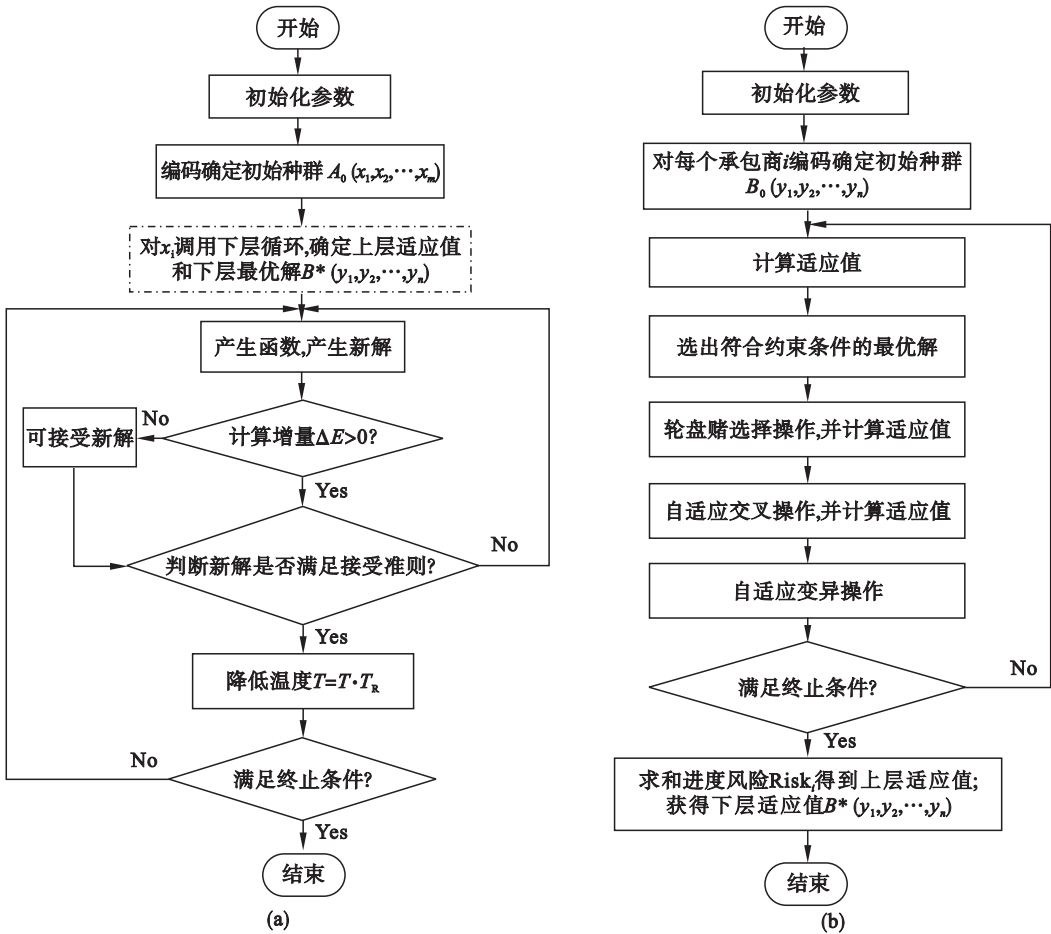


图 1 TLHGA 流程图
Fig. 1 Flow chart of TLHGA
(a)—TLHGA 流程图; (b)—下层结构.

上层搜索的终止由初始温度和每个温度下的迭代次数确定,下层的终止由迭代次数确定.

3 算例及分析

本节以国内某 IT 外包服务公司为背景,设计仿真实例来验证算法的有效性. 算例考虑有 1 个委托方和 M 个承包商的情况, $M=1,2,\cdots,5$;这 5 种情况总的预算资金上限分别为 $B_{\max}=900,1\,800,2\,700,3\,600,4\,500$,进度风险水平约束为 $R_{\max}=0.05$. 各承包商与委托方通过签订合同开展合作,合同中的总价款 $U_i=10\,000$,项目延期的赔偿款 $V_i=3\,000,i=1,2,\cdots,5$. 承包商的利润率 $r=0.2$. 假定每个项目由 10 个活动组成 ($N_i=10$),且活动在工序上是串行的. 各活动的费用 $c_{ij}=650,i=1,2,\cdots,5,j=1,2,\cdots,10$. 各活动的进度风险发生概率 $p_{ij}=\{0.93,0.87,0.79,0.79,0.75,0.73,0.71,0.67,0.65\}$. 为了防止进度风险发生致使项目失败,对于每个活动,可以从 4 个 ($Q=4$) 风险控制措施中选择一个用于风险防范. 在风险措施的控制下,项目在执行过程中的工期会有相应的缩减,由 t_{ij} 变为 $t_{ij}(y_{ij})$. 在本算例中风险控制措施编号越大,则活动工期缩减得越多. 其中项目 i 的计划工期 $T_i^0=1\,000$,实际工期由式 (14) 表示:

$$t_{ij}(y_{ij})=120\exp(-\mu_{ij}y_{ij}). \tag{14}$$

参数 $\mu_{ij}=\{0.02,0.025,0.03,0.035,0.04,0.045,0.05,0.055,0.06,0.065\}$,不同取值表示相同风险控制措施对不同活动工期缩减作用的不同效果. 从式 (14) 可以看出,各活动未进行风险控制的实际工期为 120. 执行风险控制措施需要一定的费用,费用的函数由式 (15) 表示:

$$\text{cost}_{ij}(y_{ij})=100[1-\exp(-\omega_{ij}y_{ij})]. \tag{15}$$

其中, $\omega_{ij}=\{0.55,0.6,0.65,0.7,0.75,0.8,0.85,0.9,0.95,1\}$.

3.1 算法参数设置

算法上层需要确定的参数是模拟退火的初始温度 T 和温度变化率 T_{R} . 以 $M=1$ 为例,对 $T=60,65,70,75,80,85$ 和 $T_{\text{R}}=0.4,0.5,0.6$ 的情况进行组合,分别设计 18 组算例,每组运行 30 次. 以项目进度风险水平的最小值和平均值为评判指标,通过实验得到参数 $T=60$ 和 $T_{\text{R}}=0.6$. 其他参数包括:种群规模取 10,温度的下限 T_{\min} 取 4. 算法下层自适应遗传算法中 $P_{\text{c1}}=0.9,P_{\text{c2}}=0.6,P_{\text{m1}}=0.1,P_{\text{m2}}=0.01$. 种群规模取 30,迭代次数为 60.

3.2 仿真结果及其分析

使用 TLHGA 对模型进行求解,得到的最优仿真结果见表 1,其中 pr 是上层最优适应值, x 是委托方为承包商分配的资金,prisks 是承包商的进度风险值, p 为所选的进度风险措施编号,Benefit 是承包商的收益,tsum 项目工期缩短量.

表 1 最优仿真结果
Table 1 Optimal simulation results

M	pr	x	prisks	p	Benefit	tsum
1	35.23	899	35.23	4144444433	2 450.42	34.58
		849	40.74	4043444443	2 450.31	35.42
2	80.63	884	39.89	3044444444	2 485.72	27.73
		880	38.59	4404444444	2 475.24	27.69
3	123.55	892	44.12	4434440444	2 454.21	48.95
		885	40.84	4404444244	2 459.44	38.78
4	167.68	896	41.98	2444144444	2 462.33	31.63
		888	40.20	4044443444	2 485.47	30.53
		861	44.15	3444444044	2 466.46	40.00
		865	41.35	4444044443	2 463.64	38.03
5	217.82	854	39.05	4044444344	2 450.98	30.93
		826	45.72	4014444444	2 462.77	35.51
		836	43.06	3304444443	2 447.77	38.84
		862	43.66	4344444404	2 466.89	42.42
		875	46.33	4042443433	2 493.74	50.14

以 $M = 2$ 的结果为例来说明仿真结果的含义. 委托方将项目委托给 2 个承包商, 委托方分配给承包商 1 和 2 的资金分别为 849 和 884. 在风险控制措施的作用下, 整个项目的进度风险水平为 80.63, 其中承包商 1 和承包商 2 的进度风险水平分别控制在 40.74 和 39.89. 在进行风险控制前各子项目的实际工期均为 1 200, 采取相应的风险控制措施后, 承包商 1 和承包商 2 的子项目的工期分别缩短为 35.42 和 27.73. 与此同时, 承包商 1 和承包商 2 分别获得风险控制收益为 2 450.31 和 2 485.72.

为了验证算法的有效性, 将 TLHGA 的仿真结果与 TLPSO^[1] 和 TGA 的结果进行比较, 结果见图 2, 其中 prBest 是最优值、prAVG 是最优值的

平均值、prVAR 是最优值的方差、CPUTimeAVG 是平均 CPU 时间. 由图 2a 可以看出, 随着问题规模的增加, TGA 的整体图线在 TLPSO 的下方, TLHGA 的整体图线在 TGA 的下方, 显然, TGA 比文献[1]中使用的 TLPSO 算法从结果上看更适合该问题, 而 TLHGA 得到了比 TGA 更好的最优解. 图 2b 描述的是仿真结果得到的目标值的平均值, 可以看出与图 2a 中有相似的结论. 由图 2c 可以看出, TGA 和 TLHGA 表现出来的稳定性不及 TLPSO. 在图 2d 中, 随着问题规模的不断增加, TLPSO 和 TGA 的平均 CPU 时间都有明显的增大, 但 TLHGA 的 CPU 时间并没有随着问题规模的增加而明显增大, 表明该算法的寻优效率较高.

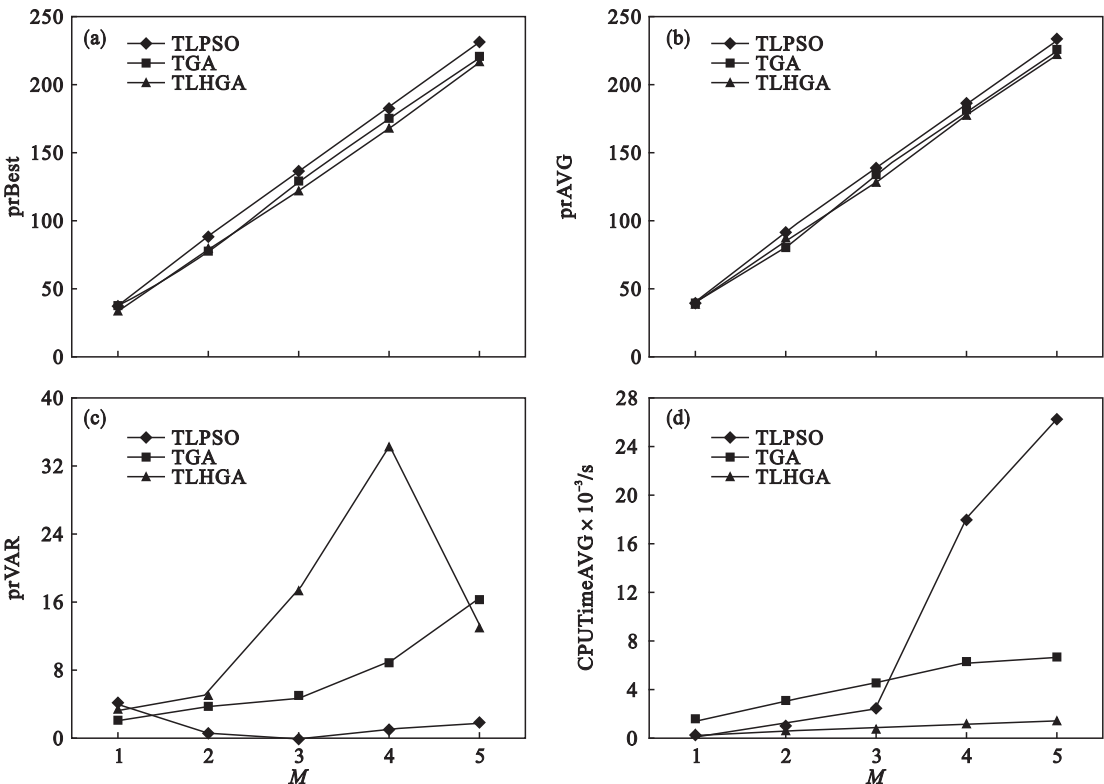


图 2 三种算法的性能对比

Fig. 2 Comparison of the performance of the three algorithms

(a)—最优值; (b)—平均最优值; (c)—最优值方差; (d)—平均 CPU 时间.

4 结 语

本文针对 IT 外包项目进度风险控制的两层分布式决策问题, 分别引入了模拟退火和自适应机制, 设计了两层混合遗传算法 (TLHGA). 在仿真实验中, 设计了五个数值算例, 分别采用 TGA 和 TLHGA 进行求解, 并将仿真结果与文献[1]中 TLPSO 的结果进行了对比. 结果表明, 在最优适

应值、平均适应值、方差和 CPU 时间等诸多方面, TLHGA 都表现不错, 尤其是在时间效率方面表现尤为突出.

参考文献:

[1] 卢福强, 毕华玲, 黄敏, 等. IT 外包进度风险控制的结构化分布式决策模型[J]. 系统工程, 2016, 34(7): 138 - 145.
(Lu Fu-qiang, Bi Hua-ling, Huang Min, et al. Study on CDDM model of IT outsourcing schedule risk control [J]. Systems Engineering, 2016, 34(7): 138 - 145.)

(下转第 179 页)