

风险调度中引入知识迁移的细菌觅食强化学习优化算法

韩传家, 张孝顺, 余 涛, 瞿凯平

(华南理工大学电力学院, 广东省广州市 510640)

摘要: 针对电力系统运行过程中负荷及故障的不确定性,在经济调度中引入风险评估原理,并提出了一种全新的基于知识迁移的细菌觅食强化学习优化算法。该算法将细菌觅食算法的寻优模式与Q学习算法的试错迭代机制结合,利用多主体协同合作来更新共有的知识矩阵,并以基于知识延伸的维度缩减方式避免了“维数灾难”。在预学习获得最优知识矩阵后,利用知识迁移加速在线学习进程。IEEE RTS-79测试系统的仿真结果表明:所提算法在保证获得高质量最优解的同时,寻优速度可达经典智能算法的9~20倍,适合求解大规模复杂电网的风险调度快速优化。

关键词: 知识迁移; 细菌觅食; 强化学习; 风险调度

0 引言

为更好地权衡系统安全性与经济效益,增强调度操作抵御运行风险的水平,国内外学者已经在发电优化中引入电力系统的风险理论,对风险调度进行了大量研究。目前,实际电力系统中多采用确定性“N-1”原则指导安全约束经济调度^[1],然而确定性准则下的结果偏保守。为此,McCalley教授等提出了基于概率风险的最优潮流模型与算法^[2-3],在考虑运行风险的机组调度领域做了一些开创性研究。文献[4]提出了一种多阶段相协调的调度方法,但其基于直流潮流的风险评估无法针对节点电压风险做出后果评价。文献[5]采用改进多目标差分进化算法对基于风险的多目标发电调度进行了优化。因此,本文所采用风险调度模型对基态以及预想故障下风险指标进行定量评估,考虑一天之内不同时间断面间的内在联系,通过挖掘断面之间的相似性实现风险调度的快速求解。

风险调度问题是一个基于潮流的复杂混合非线性规划问题,处理该问题的方法主要有经典优化算法和启发式人工智能算法两大类^[6]。梯度下降法^[7]、非线性规划法^[8]、内点法^[9]和牛顿法^[10]等经典优化算法求解风险调度时存在全局收敛性差、要求精确数学模型等缺点,虽然应用Gurobi和Cplex等成熟数学求解器可以实现对风险调度的快速求

解,但其算法应用设计较为繁琐。另一方面,遗传算法(genetic algorithm, GA)^[11]、量子遗传算法(quantum genetic algorithm, QGA)^[12]、人工蜂群算法(artificial bee colony, ABC)^[13]、粒子群优化(particle swarm optimization, PSO)^[14]等智能算法对具体数学模型依赖程度低,应用更为简便,只需要给出相应的优化变量、目标函数及约束的反馈,即可进行优化求解,且其对离散性、非线性优化问题的适应力强,在电力系统各类优化问题上已有较为成熟的应用^[15]。然而,这类智能算法对相似任务的优化是孤立进行的,不能有效保存过去任务的经验和知识,缺乏自学习能力,每次新任务执行时需重新初始化,导致寻优效率较低,难以适应大规模复杂风险调度的快速优化求解。

将迁移学习技术与启发式人工智能算法相结合是解决上述问题的有效途径。传统的机器学习认为不同任务间是相互孤立的,然而实际中这些任务经常是彼此关联的。迁移学习可将辅助领域中所学习到的知识或策略应用到相似但不相同的目标领域中进行学习,其本质是利用任务之间的联系,复用已有经验以加速新任务的学习速度^[16-17]。

因此,本文将细菌觅食优化(bacteria foraging optimization, BFO)算法与Q学习算法相结合,在此基础上提出了基于知识迁移的细菌觅食强化学习优化(transfer bacteria foraging optimization, TBFO)算法^[18-20]。为验证所提算法的寻优性能,本文在RTS-79可靠性测试平台上对风险调度问题进行了仿真验证。

收稿日期: 2016-06-19; 修回日期: 2016-11-10。

上网日期: 2017-01-10。

国家重点基础研究发展计划(973计划)资助项目(2013CB228205);国家自然科学基金资助项目(51477055)。

1 风险调度数学模型

1.1 运行风险指标计算

对电力系统运行过程中面临的不确定性因素进行可能性与严重性的综合评估的过程即为风险评估^[21]。其数学表达式可求解如下：

$$R(X_i) = \sum P_r(E_i) S_{ev}(E_i) \quad (1)$$

式中： X_i 为系统当前状态； E_i 为随机故障 i ； $P_r(E_i)$ 和 $S_{ev}(E_i)$ 分别为故障 i 出现的概率和故障后果的严重度指标。

统计规律表明，交流输电线路 i 在一定时间间隔 Δt 内发生停运的概率服从泊松分布，在 Δt 内线路的累积故障概率为^[22]：

$$P_r(F_i) = 1 - \exp(-\lambda_i \Delta t) \quad (2)$$

式中： F_i 表示输电线路 i 停运； $P_r(F_i)$ 为事件 F_i 出现的概率； λ_i 为线路故障率。

若系统中有 m 条输电线路，在 t 时刻发生多重故障，输电线路 i 停运概率为^[23]：

$$\rho_i = P_r(F_i) \prod_{j \in U, j \neq i} (1 - P_r(F_j)) \quad (3)$$

式中： U 为 t 时刻系统中全部无故障线路集。

输电线路停运可能造成的不良影响主要包括输电线路潮流越限和节点电压幅值越限，通常用线性函数描述故障后果，然而，线性化的风险指标不能充分体现高概率轻微故障和低概率严重故障这两类故障的区别。因此，本文采用非线性效用函数度量输电线路停运后果的严重程度^[5]。

故障后线路 i 潮流越限量的计算方法为：

$$\omega_{L_i} = \begin{cases} L_i - L_0 & L_i > L_0 \\ 0 & L_i \leq L_0 \end{cases} \quad (4)$$

式中： L_i 为线路 i 实际传输功率与线路功率传输上限之比； L_0 为设定阈值，本文取值为 0.9。

线路 i 的潮流越限严重度指标定义为：

$$S_{ev}(\omega_{L_i}) = \frac{\exp(A(\omega_{L_i}) + B) - 1}{C} \quad (5)$$

式中： A 、 B 和 C 均为正数。

同理，故障后节点 i 电压幅值越限量的计算方法为：

$$\omega_{V_i} = \begin{cases} U_{i \min} - U_i & 0 < U_i < U_{i \min} \\ 0 & U_{i \min} \leq U_i \leq U_{i \max} \\ U_i - U_{i \max} & U_i > U_{i \max} \end{cases} \quad (6)$$

式中： U_i 为节点 i 的电压幅值； $U_{i \max}$ 和 $U_{i \min}$ 分别为节点 i 电压幅值的上限和下限。

节点 i 的电压幅值越限严重度指标为：

$$S_{ev}(\omega_{V_i}) = \frac{\exp(A(\omega_{V_i}) + B) - 1}{C} \quad (7)$$

综合线路潮流越限风险指标和节点电压幅值越限风险指标可得到全局性的电力系统运行风险指标 I_R ，其定义如下：

$$I_R = \mu_1 R_L + \mu_2 R_V \quad (8)$$

式中： R_L 为系统全部线路潮流越限风险指标之和； R_V 为系统全部节点电压幅值越限风险指标之和； μ_1 和 μ_2 为两者相应的权重， $\mu_1 + \mu_2 = 1$ 。

1.2 风险调度的目标函数及约束条件

本文所述风险调度优化的目的是在满足运行过程中各变量约束条件的基础上，尽可能减少机组发电的燃料成本和电力系统的运行风险。为降低优化的难度，本文利用罚函数法将带约束的风险调度优化问题转化为无约束优化问题^[24]，并采用线性加权法来实现不同量纲的经济性和安全性目标的统一优化，因此，经处理后的风险调度模型可表示为：

$$\begin{cases} \min(\omega_1 f_1(\mathbf{x}) + \omega_2 f_2(\mathbf{x}) + f_3(\mathbf{x})) \\ f_1(\mathbf{x}) = \frac{F_C(\mathbf{x})}{C_1} \\ f_2(\mathbf{x}) = \frac{I_R(\mathbf{x})}{C_2} \\ f_3(\mathbf{x}) = MC_V(\mathbf{x}) \end{cases} \quad (9)$$

式中： f_1 、 f_2 、 f_3 分别为规格化后的目标函数 F_C 、 I_R 、 C_V ； F_C 为系统的发电机组燃料成本； C_V 为基态下（未发生系统故障）系统总约束的违反程度； ω_1 和 ω_2 为权重系数， $\omega_1 \in [0, 1]$ ， $\omega_2 \in [0, 1]$ ， $\omega_1 + \omega_2 = 1$ ； $\mathbf{x} = [\mathbf{V}, \boldsymbol{\theta}, \mathbf{P}_G, \mathbf{Q}_G]^T$ ，其中 \mathbf{V} 、 $\boldsymbol{\theta}$ 、 \mathbf{P}_G 、 \mathbf{Q}_G 分别对应电网各节点电压值、各节点相角、发电机有功出力 and 无功出力； C_1 和 C_2 为规格化处理的基准值，其引入可以实现不同量纲目标函数 F_C 和 I_R 的直接叠加； M 为罚因子， $M > 0$ ，其引入可加大对越限方案的惩罚程度，将原有约束问题转化为无约束问题，从而降低优化的难度。

其中，系统燃料成本为^[25]：

$$F_C(\mathbf{x}) = \sum_{i \in S_G} (\alpha_i + \beta_i P_{G_i} + \gamma_i P_{G_i}^2) \quad (10)$$

式中： S_G 为发电机集合； P_{G_i} 为发电机 i 的有功出力； α_i 、 β_i 、 γ_i 为发电机 i 的燃料成本系数。

此外，系统总约束的违反程度为：

$$\begin{aligned} C_V(\mathbf{x}) = & \sum_{j=1}^q \max(0, g_j(\mathbf{x})) = \\ & \max(0, P_{G_s} - P_{G_{s \max}}, P_{G_{s \min}} - P_{G_s}) + \\ & \sum_{i \in S_G} \max(0, Q_{G_i} - Q_{G_{i \max}}, Q_{G_{i \min}} - Q_{G_i}) + \\ & \sum_{i \in S_D} \max(0, U_i - U_{i \max}, U_{i \min} - U_i) + \\ & \sum_{i \in S_L} \max(0, |T_i| - T_{i \max}) \end{aligned} \quad (11)$$

式中: $P_{G_{smax}}$ 和 $P_{G_{smin}}$ 分别为平衡机组的有功出力 P_{G_s} 的上限和下限; $Q_{G_{imax}}$ 和 $Q_{G_{imin}}$ 分别为发电机 i 的无功出力 Q_{G_i} 的上限和下限; $|T_i|$ 和 T_{imax} 分别为支路 i 传输的视在功率和容量极限; S_D 和 S_L 分别为负荷母线集合和支路集合。

由潮流方程表示的等式约束为^[26]:

$$\begin{cases} P_{G_i} - P_{D_i} = U_i \sum_{j \in i} U_j (G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij}) \\ Q_{G_i} - Q_{D_i} = U_i \sum_{j \in i} U_j (G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij}) \end{cases} \quad (12)$$

式中: P_{D_i} 和 Q_{D_i} 分别为节点 i 的有功负荷和无功负荷; θ_{ij} 为节点 i, j 的电压相角差; G_{ij} 和 B_{ij} 分别为支路 $i-j$ 的电导和电纳。

不等式约束包含状态变量约束和控制变量约束,即

$$\begin{cases} P_{G_{imin}} \leq P_{G_i} \leq P_{G_{imax}} & i \in S_G \\ Q_{G_{imin}} \leq Q_{G_i} \leq Q_{G_{imax}} & i \in S_G \\ U_{imin} \leq U_i \leq U_{imax} & i \in S_D \\ |T_i| \leq T_{imin} & i \in S_L \end{cases} \quad (13)$$

2 TBFO 算法原理

2.1 知识矩阵

在 Q 学习算法中, Q 矩阵中的元素 $Q(s, a)$ 反映了在系统当前状态 s 下选择动作 a 所得累积奖励值的期望,矩阵记录了智能体把状态映射到动作这一过程的知识。因此,本文将 Q 矩阵定义为细菌群的知识矩阵(见附录 A 图 A1)。TBFO 算法中,细菌群从知识矩阵中获得针对特定环境状态的动作策略,并利用从多次重复试验中获得的反馈信息更新原有知识,形成对特定状态的固有反应,以使细菌群觅食过程中累积的能量值达到最大。

标准 Q 算法中, Q 矩阵是一大小为 $|S| \times |A|$ 的lookup表格,各变量间是独立的,可行解的选择是一种并行式选择。若求解的问题有 k 个变量 x ,每个变量有动作 N_i 个,则总的动作集合数为 $|A| = N_1 N_2 \cdots N_{k-1} N_k$ 。随着控制变量数目的增加,算法的动作空间将以指数倍极速增长,发生“维数灾难”。

为此,本文提出了一种知识延伸的维度缩减方法,将知识矩阵 Q 划分为多个子知识矩阵 Q_i ,与各变量一一对应(见附录 A 图 A2)。变量间通过知识矩阵联系起来,相邻矩阵中的元素为相关知识,也就是说 x_i 的动作空间 A_i 即为 x_{i+1} 的状态空间 S_{i+1} 。只有先确定了变量 x_i 的动作,才能基于其选择结果选择 x_{i+1} 的动作,从而在相关知识间形成了一种链式的延伸,实现了对知识矩阵的分解降维。

2.2 知识迁移

迁移学习是指利用已有知识帮助学习新知识的过程,如果状态空间 S 与动作空间 A 保持不变,则可将源任务的最优知识矩阵作为目标的初始知识矩阵,通过这种方式利用先前所学到的知识,实现知识的迁移^[17]。旧知识矩阵包含了与新任务之间的共性,但不可避免地存在一些无效知识,若不加以处理将导致学习效率的降低(即“负迁移”现象)。为此,在知识迁移的过程中,TBFO 更加注重对源任务和目标任务间有效知识的提取和相似性的挖掘。

TBFO 需要在预学习阶段执行一系列的源任务以获取最优知识矩阵,并从中挖掘出初始知识,为将来相关的新任务做好准备。如图 1 所示,来自源任务的相关初始知识将用于在线优化中,根据源任务与新任务间的相似性,源任务 Q_S 的最优知识矩阵将通过线性组合迁移为新任务 Q_N 的初始知识矩阵。

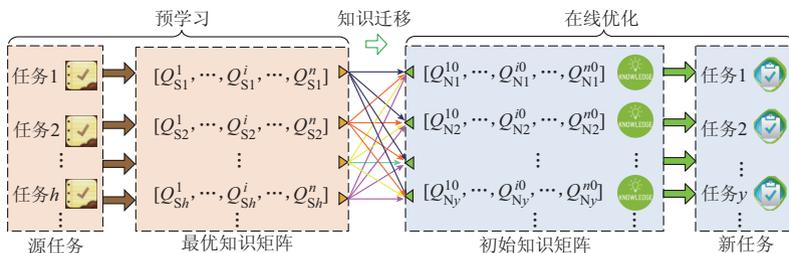


图 1 知识迁移
Fig.1 Knowledge transfer

2.3 知识获取

标准 BFO 算法结构中包含 3 层嵌套循环,由外向内依次为迁徙性操作循环、复制性操作循环和趋向性操作循环^[19]。主要的搜索任务由趋向性操作承担,使得算法可以执行细致的局部寻优,复制性操

作按照细菌所在位置能量高低排序,复制精英个体,淘汰劣解,迁徙行为按概率的随机移动则丰富了种群的多样性。BFO 算法利用细菌群不同操作的协调配合对解空间进行概率搜索,期望直接获得优化任务的最优解。

与标准 BFO 算法纯随机搜索方式不同, TBFO 算法中全部细菌将根据初始知识矩阵对觅食区域进行搜索, 并将所得奖励反馈到知识矩阵。如图 2 所示, 按照正在执行的操作, TBFO 将细菌划分为趋向和迁徙两种状态。在算法单次迭代循环中, 将两种状态分别赋予一定比例的细菌个体, 两组细菌执行完各操作后, 计算并排序全部个体的能量值以评估其对环境的适应程度, 进入复制性操作。为增强菌群的多样性, 本文算法在复制性操作中引入了交叉的过程, 其交叉方式如下:

$$\theta_{i+S/2}(j, k, l) = r\theta_i(j, k, l) + (1-r)\theta_{i+S/2}(j, k, l) \quad (14)$$

式中: S 为细菌个体数; $i \in [1, S/2]$; r 为 $[0, 1]$ 内的随机数。

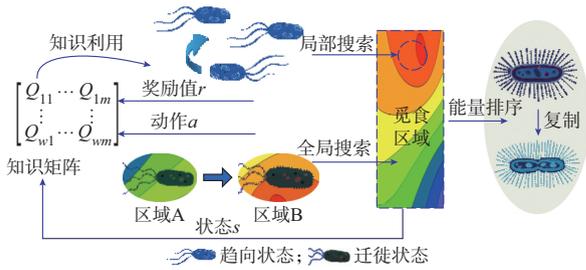


图 2 TBFO 算法的知识获取

Fig.2 Knowledge learning of TBFO algorithm

新一轮迭代循环中, TBFO 依据上次迭代中能量值高低对细菌状态进行再分配, 能量值较大的细菌保持所在区域不变并进行趋向性操作, 而能量值较低的细菌被重置为迁徙状态。

Q 学习算法由单主体进行 Q 矩阵更新, 一次只能更新矩阵中的一个元素, 收敛速度慢。而 TBFO 算法将菌群作为多主体对知识矩阵协同更新, 全部细菌共享一个知识矩阵, 单次迭代中可同时更新多个知识元素, 大大加快了寻优的效率。每次试错探索后, TBFO 算法会对各主体进行奖励值评估。引入菌群协同后, 子知识矩阵 Q_i 更新方式如下^[27]:

$$\rho_{ij}^{(k)} = R(s_{ij}^{(k)}, s_{ij}^{(k+1)}, a_{ij}^{(k)}) + \mu \max_{a_i \in A_i} Q_i^{(k)}(s_i^{(k+1)}, a) - Q_i^{(k)}(s_{ij}^{(k)}, a_{ij}^{(k)}) \quad (15)$$

$$Q_i^{(k+1)}(s_{ij}^{(k)}, a_{ij}^{(k)}) = Q_i^{(k+1)}(s_{ij}^{(k)}, a_{ij}^{(k)}) + \sigma \rho_{ij}^{(k)} s_{ij}^{(k)} \quad (16)$$

式中: 下标 i 代表第 i 个子知识矩阵, j 表示第 j 个细菌; $R(s_{ij}^{(k)}, s_{ij}^{(k+1)}, a_{ij}^{(k)})$ 表示第 k 次迭代在状态 $s_{ij}^{(k)}$ 下选择动作 $a_{ij}^{(k)}$ 当前状态转移为 $s_{ij}^{(k+1)}$ 时得到的奖励函数值; σ 为学习因子; μ 为折扣因子。

2.4 动作策略

菌群进行探索学习时, 全部个体都面临着如何执行动作选择的问题。TBFO 算法中包含搜索和利

用两种倾向的博弈: 动作选择倾向于随机搜索增加了全局收敛的可能性, 但对知识矩阵的利用率低, 算法寻优效率低; 倾向于知识利用可加快迭代收敛, 但增大了局部收敛的概率。为此, 本文结合了 BFO 算法的随机搜索模式与基于概率空间的动作选择策略, 提出了一种新的动作策略。

基于能量值排序, 将菌群中的优势个体置于趋向状态, 仍承担局部搜索的任务。设 $\theta_i(j, k, l)$ 为细菌个体 i 在第 l 代迁徙操作、第 k 代复制操作和第 j 代趋向操作后的位置, 其趋向行为可表示为^[19]:

$$\theta_i(j+1, k, l) = \theta_i(j, k, l) + C^{(k)}(i) \frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}} \quad (17)$$

式中: $C^{(k)}(i)$ 为单次游动步长; Δ 为游动后确定的方向上的单位向量。

在标准 BFO 算法中, 细菌的游动步长为一定固定值, 游动步长过小会影响算法的全局搜索速率, 而过大的游动步长将影响优化后期局部搜索的精度, 为此, 本文引入了非线性的惯性步长 $C^{(k)}(i)$, 随着优化的进行, 逐步缩小, 其更新方式如下:

$$C^{(k)}(i) = C_{\text{start}}(i) -$$

$$(C_{\text{start}}(i) - C_{\text{end}}(i)) \left[\frac{2k}{T_{\text{max}}} - \left(\frac{k}{T_{\text{max}}} \right)^2 \right] \quad (18)$$

式中: $C^{(k)}(i)$ 为第 k 次循环中细菌 i 的游动步长; C_{start} 为初始游动步长; C_{end} 为最终游动步长; T_{max} 为最大循环步数。

标准 BFO 算法中, 迁徙性操作是无选择性的纯随机行为, 这使得菌群中的精英个体面临被淘汰的风险。TBFO 算法中, 处于迁徙状态的细菌在知识矩阵的指导下进行探索学习。对于特定状态, 知识元素越大, 执行对应觅食行为细菌获得的能量值越高。因此, 知识矩阵在更新知识的过程中已保留了精英细菌个体信息。对处于迁徙状态的细菌, 当其满足迁徙概率 P_{ed} 时, 细菌按照动作概率矩阵进行轮盘选择; 否则细菌按照最大知识元素对应的动作迁徙(贪婪策略), 即

$$a_{ij}^{(k+1)} = \begin{cases} \arg \max_{a_i \in A_i} Q_i^{(k+1)}(s_{ij}^{(k+1)}, a_i) & r \geq P_{\text{ed}} \\ a_s & \text{其他} \end{cases} \quad (19)$$

式中: r 为一随机数; P_{ed} 为迁徙概率。当满足迁徙条件时, 细菌按照动作概率矩阵 P_i 执行伪随机轮盘选择 a_s 。

P_i 的更新方式如下:

$$\begin{cases} \mathbf{e}_i(s_i, a_i) = \frac{1}{\mathbf{Q}_i(s_i, a_i) - \xi \max_{a' \in A_i} \mathbf{Q}_i(s_i, a')} \\ \mathbf{P}_i(s_i, a_i) = \frac{\mathbf{e}_i(s_i, a_i)}{\sum_{a' \in A_i} \mathbf{e}_i(s_i, a')} \end{cases} \quad (20)$$

式中: ξ 为差别系数,用于放大子知识矩阵 \mathbf{Q}_i 的差异; \mathbf{e}_i 为过渡矩阵。

3 风险调度求解

3.1 算法结构优化

本文所述风险调度模型不同于一般的交流最优潮流问题,模型的目标函数涉及系统运行风险指标的计算,为得到这一指标,使用人工智能算法求解时需进行基态及所有故障状态下的交流潮流计算。设预想故障集中包涵 N_p 个故障,则优化过程中进行潮流计算的次数将是常规交流最优潮流问题的 N_p+1 倍,因此,风险调度的求解时间将远大于常规交流最优潮流问题。若使用标准BFO算法,设 M_e, M_r 和 M_q 分别表示迁徙、复制和趋向行为的操作数,最大游动次数为 N_s ,则潮流计算次数可多达 $M_e M_r N_q N_s (N_p+1)$ 次,使得求解过程极为缓慢。如2.3节所述, TBFO通过对算法寻优模式的改进,去除了原算法的嵌套循环,提高了算法的效率。

3.2 状态与动作设计

在本文风险调度求解中,选取PV节点处的发电机有功出力 P_G 为控制变量,动作变量空间 \mathbf{A} ,即 $[\mathbf{A}_{PG1}, \mathbf{A}_{PG2}, \dots, \mathbf{A}_{PGi}]$ 与控制变量空间是一一对应的, i 为PV节点上机组总数。

如前文所述,前一个变量的动作空间即为下一个变量的状态空间。与各变量状态-动作空间对应的子知识矩阵分别为 $\mathbf{Q}_{PG1}, \mathbf{Q}_{PG2}, \dots, \mathbf{Q}_{PGi}$ 。

3.3 奖励函数设计

上文提及的风险调度数学模型中,更小的目标函数代表着更好的优化效果。而在TBFO算法中,立即奖励值反映了优化的方向,菌群通过迭代优化知识矩阵获得最优策略,以期望获得最大的累积奖励函数值。根据式(9)定义的目标函数,本文的奖励函数设计如下:

$$R_{ij} = \frac{1}{\omega_1 \left(\frac{F_C}{C_1} \right) + \omega_2 \left(\frac{I_R}{C_2} \right) + MC_V} \quad (21)$$

由式(21)可以得出如下结论:①仿真试验表明,不同断面下的机组燃料成本基本波动范围为21 135~39 402美元,而风险指标基本波动范围为 $2.42d^{-5} \sim 5.2d^{-5}$,其中 d 为指标基,故将 C_1 和 C_2 分别设置为 10^4 美元和 10^{-5} 。因此, F_C 和 I_R 可转

化为统一的无量纲值 f_1 和 f_2 ,分别在 $[2.1, 4]$ 与 $[2.4, 5.2]$ 之间波动;②一般来说,罚因子 M 的值不宜选得过小或过大, M 过小时,则罚函数的极小点远离约束问题的最优解,计算效率很差; M 过大时,则给罚函数的极小化增加计算上的困难^[28]。仿真表明: C_V 相比于经过规格化处理后的机组燃料成本与风险指标,在数值上已足够大,故本文选择 $M=1$ 。

3.4 迁移设计

TBFO算法迁移效率的关键是如何挖掘源任务与新任务之间的相似性。在电力系统风险调度中,风险调度的求解依赖于获得系统的潮流分布。由于短期内系统的网络拓扑和运行方式不会发生显著性改变,因而该问题求解主要取决于系统的负荷需求。据此,本文将有功功率偏差定义为源任务和新任务间的相似性,并将有功需求由小到大划分成多个负荷断面 $[P_{Ds1}, P_{Ds2}), [P_{Ds2}, P_{Ds3}), \dots, [P_{Dsi-1}, P_{Dsi}), \dots, [P_{Dsn-1}, P_{Dsn})$,其中 P_{Dsi} 为风险调度源任务中第 i 个断面的负荷需求,并有 $P_{Ds1} < P_{Ds2} < P_{Dsi} < P_{Dsn-1} < P_{Dsn}$ 。为避免迁移受到无效知识对新任务学习质量和速率产生负干扰,学习过程中应尽量利用与新任务相似度高的知识,本文仅使用最接近新任务负荷需求的两个源任务断面信息进行迁移。假设新任务 x 的有功需求是 P_{Dx}, P_{Dj} 和 P_{Dk} 为源任务中与任务 x 最接近的两个断面负荷,且满足 $P_{Dj} < P_{Dx} < P_{Dk}$,则两个源任务对迁移学习的贡献系数 η_1 和 η_2 可由式(22)计算,其中 $\eta_1 + \eta_2 = 1$ 。

$$\begin{cases} \eta_1 = \frac{P_{Dx} - P_{Dj}}{P_{Dk} - P_{Dj}} \\ \eta_2 = \frac{P_{Dk} - P_{Dx}}{P_{Dk} - P_{Dj}} \end{cases} \quad (22)$$

利用线性迁移方式,可以得到新任务 x 的知识矩阵为:

$$\mathbf{Q}_i^x = \eta_1 \mathbf{Q}_i^j + \eta_2 \mathbf{Q}_i^k \quad (23)$$

式中: $\mathbf{Q}_i^x, \mathbf{Q}_i^j$ 和 \mathbf{Q}_i^k 分别为新任务 x 、源任务 j 和源任务 k 中对应于变量 i 的子知识矩阵。

综上,迁移学习步骤如下:①以适当的间隔从日负荷曲线上选择多个断面作为源任务学习的内容;②对源任务进行预学习并将优化信息储存在知识矩阵中;③根据新任务的有功功率选取源任务中与新任务最为接近的两个断面,计算贡献系数并按照线性迁移策略得到新任务的初始知识矩阵;④利用知识矩阵的初始策略进行新任务的在线优化(风险调度的求解流程见附录A图A3)。

3.5 参数设置

TBFO算法中,对算法效果影响较大的参数主

要有种群数量 S 、迁徙概率 P_{ed} 、学习因子 σ 、折扣因子 μ ^[29]。各参数取值及其对寻优结果影响的原理如下。

种群数量 S :代表进行搜索的细菌个体的数目。较大种群规模可以提高接近最优解的机会,但也将耗费更多的计算时间。经过大量仿真调试,文中源任务学习阶段与新任务学习阶段 S 的最优取值分别为 200 和 64。

迁徙概率 P_{ed} :体现菌群进行迁徙性操作时对贪婪策略与随机选择的权衡。 P_{ed} 越大,迁徙性行为中执行随机轮盘选择的可能性越大。 P_{ed} 的取值范围为 $[0,1]$,在本文的源任务学习阶段与新任务学习阶段其取值分别为 200 和 64。

学习因子 σ :影响菌群寻找食物时从觅食区域获取知识的速度。 σ 越大,算法自学习的速度越快,但算法容易陷入局部最优解。 σ 的取值范围为 $(0,1)$,在本文的源任务学习阶段与新任务学习阶段其取值分别为 0.1 和 0.99。

折扣因子 μ :反映了更新知识矩阵过程中对历史奖励值的折扣程度, μ 越大,表明对立即奖励值越重视,即对历史奖励值折扣越大。 μ 的取值范围为 $(0,1)$,在本文的源任务学习阶段与新任务学习阶段其取值分别为 0.9 和 0.99。

4 仿真算例

本文仿真计算是在 CPU 为 Intel Xeon E5-2670、主频为 2.3 GHz、内存 64 GB 的 AMAX 服务器上运行,算法中的潮流计算基于 MATLAB R2014a 平台上的 Matpower 4.0 工具箱实现。为验证 TBFO 算法的效果,算例中还引入了 BFO,GA,QGA,PSO,ABC,Q 学习算法与之进行对照。对于本文提出的风险调度优化问题,其存在可行解或不可行解。当智能算法的个体搜索到不可行解时,将导致潮流不收敛或变量越界,此时 f_3 将大于 0,其适应度函数相比其他个体将显得更大,从而迫使算法对该个体进行淘汰或改进。当个体搜索到较优的可行解时,将产生正常的潮流结果,并满足所有运行约束,此时 f_3 等于 0,其适应度函数相比其他个体将显得更小,从而引导算法其他个体对其进行趋近或交叉。当算法在所有个体结束完设定的迭代步数后,其最优个体能保证潮流收敛,同时满足所有运行约束,且获得最低的适应度函数值。因此,对于每种智能算法来说,只要其种群规模和迭代最大步数设置得足够大,其最后收敛的结果必定是可行解,优化的结果是有意义的。

4.1 仿真模型

本文将 IEEE RTS-79 测试系统作为风险调度的仿真对象^[30]。选取系统基准容量为 100 MVA,系统中共有 24 条母线节点、34 条传输线/变压器和 32 台发电机(见附录 A 图 A4)。在全部 10 个发电机节点中,将单机容量最大的发电机节点 21 定为全系统的平衡节点,其余 9 个节点为 PV 节点。文中选取 PV 节点上的 31 台机组有功出力为控制变量,并将同一节点上有相同燃料成本系数的机组划为一个控制变量^[31]。

为测试算法对不同负荷水平进行优化的适应性,文中对 RTS-79 算例进行 96 个断面的风险调度优化仿真。本文选取了典型日负荷曲线(见附录 A 图 A5),并按照时序每隔 15 min 划分一个断面,得到断面 1 至断面 96。

与相邻断面时间间隔相对应,设置故障计算时间 t_c 为 15 min。本文所设预想故障集包含 5 个故障率最高的单重线路故障和 2 个具有相关停运模式的 $N-2$ 线路故障(L_{13} 与 L_{15} 、 L_{29} 与 L_{30})。由式(3)可计算得到各故障发生的概率。

4.2 对源任务的预学习

TBFO 算法在进行在线优化前,需要选取合适样本作为源任务进行预学习,菌群通过对源任务的随机搜索和试探获取知识策略从而形成初始知识矩阵。RTS-79 测试系统 96 个断面的有功功率分布为 1 685~2 850 MW。本文以 55 MW 的跨度等间距取样 21 个负荷断面,按照有功功率由小到大的顺序分别为 16,21,24,5,26,2,1,31,32,94,56,51,36,88,40,42,68,70,80,79,77。

为了对比 TBFO 与 BFO 算法菌群的寻优效果,在源任务样本中也对标准 BFO 算法的效果进行了验证(见附录 A 图 A6)。以断面 1 为例,TBFO 进行预学习的速度优于标准 BFO 算法,所需收敛时间仅为其 36%,且收敛值比其低 33%左右。

为保证在线优化期间迁移学习的全局收敛性,在预学习阶段 TBFO 的参数设置倾向于大种群、高迭代次数,不强调对算法寻优速度的追求,但 BFO 算法的寻优效率仍逊于 TBFO 算法,说明其嵌套循环结构不适用于风险调度等复杂问题的求解。

由于 BFO 算法趋向、复制、迁徙三种寻优行为缺乏朝向精英细菌个体的聚集过程,这导致其在求解高维优化问题时盲目性较强,难以寻找到全局最优解。而 TBFO 算法融合了贪婪策略和菌群多状态随机探索,可更充分地利用觅食区域的能量信息,优化所得目标函数值远小于 BFO 算法。

4.3 新任务的迁移学习

TBFO 通过预学习形成了源任务学习样本的最优动作策略,并将其存储在知识矩阵中。算法从源任务样本中选取与新任务最为接近的两个样本,并将其知识矩阵线性加权作为新任务的初始知识矩阵,实现由源任务到新任务的知识迁移。以断面 4 为例,由负荷曲线可知其负荷为 1 887 MW,与其有功功率最接近的源任务样本是断面 5 和断面 26,其有功功率分别为 1 858 MW 和 1 916 MW,根据式(22)可计算两断面对新任务的贡献系数并加权得到断面 4 的初始知识矩阵。新任务中其他断面的优化也按照同样方式进行。

图 3 为断面 4 在线学习过程中目标函数值的收敛情况。从图中可以发现,TBFO 执行断面 4 的优化时,仅需 46 s 即可完成收敛,明显短于其余人工智能算法,这里也体现出强化学习的迁移技术在速度方面的显著优势。对于机组经济调度问题,在线优化的时间尺度约为 15 min,当在调度问题中融入风险指标的计算时,若采用包含事故范围更广的预想故障集,其他算法的优化速度将难以满足速度要求,而 TBFO 算法依旧可胜任快速滚动风险调度任务。

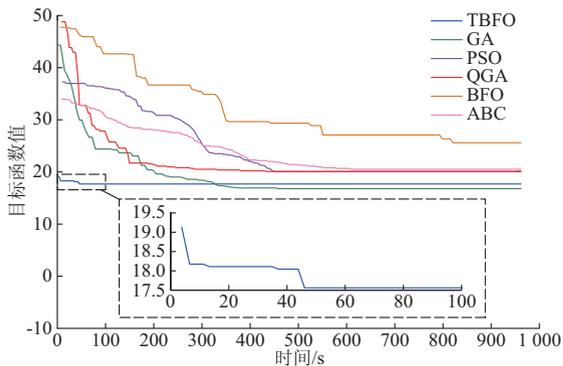


图 3 断面 4 各算法在线优化过程
Fig.3 Online optimization of each algorithm in scenario 4

对于 Q 学习算法,其寻优过程需完全遍历马尔可夫过程。本文算例中将出现“维数灾难”使得 Q 学习算法无法收敛。这验证了 TBFO 算法采用知识延伸的方式缩小解空间的有效性(TBFO 算法日优化曲线见附录 A 图 A7)。

图 4 是一天 24 h 内 7 种算法的目标函数优化结果,可以看出,本文所提算法可顺利完成全部断面的优化。在 RTS-79 测试系统中,搜索得到的目标函数收敛曲线仅略高于 GA 算法而低于其余算法,这体现了 TBFO 具有较好的全局收敛性能。

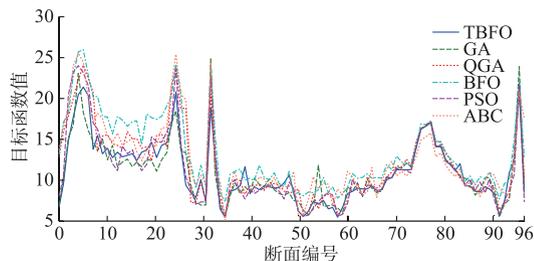


图 4 96 断面目标函数收敛值
Fig.4 Convergence results of objective function by each algorithm of ninety-six load sections

因智能算法的求解结果具有随机性,本文对上述 7 种算法各进行了 10 次仿真以进一步探讨 TBFO 算法应用于风险调度问题的优化效果(见附录 A 图 A8)。表 1 统计了各算法运行 10 次的平均结果,表中机组燃料成本、可靠性指标、目标函数和计算时间均为 96 个断面之和(箱型图分析见附录 A 图 A9 和图 A10;算法运行 10 次目标函数收敛的稳定性数据见附录 A 表 A1)。

此外,算例中还利用最优化数值分析软件 GAMS 进行风险调度模型的求解,GAMS 具有经典优化方法求解迅速、收敛稳定的优点。然而,由于系统含有大量强非线性以及导数不连续的目标及约束(包含基态约束和全部故障状态的约束),求解效果并不理想。

表 1 算法 96 断面优化结果 10 次运行平均值

Table 1 Average optimization results of ninety-six sections by each algorithm in ten runs

算法	计算时间/s	收敛时间/s	机组燃料成本/美元	可靠性指标	目标函数
BFO	100 409.90	1 045.93	2 877 187.63	4.283×10^{-3}	1 277.52
GA	45 590.41	474.90	2 832 148.91	4.346×10^{-3}	1 041.36
QGA	47 462.28	494.40	2 835 208.40	4.338×10^{-3}	1 147.77
ABC	49 824.74	519.01	2 864 487.85	4.305×10^{-3}	1 200.89
PSO	73 122.14	761.69	2 865 308.35	4.273×10^{-3}	1 086.82
Q 学习		动作空间过大,产生“维数灾难”,算法无法收敛			
TBFO	4 904.30	51.08	2 839 588.33	4.279×10^{-3}	1 066.52

5 结语

本文提出了一种基于知识迁移的细菌觅食强化学习快速风险调度优化算法,该算法不依赖于求解问题的数学模型,可以求解含多极值、不连续、多约束的凸或非凸优化问题,具有自学习及存储知识的能力,能够将历史优化任务的有效信息转化到值函数中,从而实现快速的在线优化。算法设计简便,在较短的时间内,可以获得较高质量的解,实用性强。

本文主要创新点及下一步的研究方向归纳如下。

1)改进了BFO算法的结构和搜索方式,将其与传统Q学习算法的试错迭代机制相结合,通过菌群作为主体实现群智能优化,加速了知识矩阵的形成。

2)利用知识延伸对高维度知识矩阵进行维度缩减,明显降低求解难度,有效解决了“维数灾难”问题。

3)以系统负荷有功偏差定义源任务与新任务间的相似性,通过迁移学习极大地提高了在线学习的速度,首次实现了风险调度问题的在线优化。当问题规模进一步扩大,TBFO仍能保证较快的求解速度。

4)本文以线性加权的方式将模型中的多目标优化问题转化为单目标求解,因而研究基于多目标迁移学习的风险调度优化算法将是下一步的研究方向。此外,本文假设源任务与新任务仅在有功功率方面存在差异,降低了迁移学习的难度。实际调度运行中系统网络拓扑、机组组合、故障类型及后果都可能发生较大变化,对于这些情况,若采用文中所提方法,迁移效果将受到影响,为此,下一步将对具有强差异性的源任务与新任务之间相关性的挖掘进行深入研究。

5)目前,本文提出的风险调度优化模型只是基于静态的经济调度模型,暂不考虑含多个时段耦合约束的动态经济调度模式。因此,在下一步研究工作中将把本文的单断面风险调度模型扩展到含多时段耦合约束的动态风险调度模型。由于优化变量的增加,知识矩阵规模将随之增大。同时,相似性的评估计算将发生改变:对于单断面风险调度模型,相似性的评估只是基于不同断面的有功功率需求偏差来实现计算,但对于含多时段耦合约束的动态风险调度模型来说,相似性的评估则要基于不同优化周期的负荷曲线偏差来实现计算,例如对于考虑一天的动态风险调度,其任务相似性可按照不同日负荷曲线的偏差来评估计算。

附录见本刊网络版(<http://www.aeps-info.com/aeps/ch/index.aspx>)。

参考文献

- [1] 朱继忠,徐国禹.电力系统 $N-1$ 安全有功经济调度[J].重庆大学学报(自然科学版),1992,15(2):105-109.
ZHU Jizhong, XU Guoyu. The economic dispatch of real power with security [J]. Journal of Chongqing University (Nature Science), 1992, 15(2): 105-109.
- [2] FU W, MCCALLEY J D. Risk based optimal power flow[C]// Porto Power Tech Conference, September 10-13, 2001, Porto, Portugal: 1-6.
- [3] LI Y, MCCALLEY J D. Risk-based optimal power flow and system operation state[C]// Power and Energy Society General Meeting, July 29-30, 2009, Calgary, Canada: 1-6.
- [4] 文云峰,江宇飞,沈策,等.多阶段协调的风险调度模型及算法研究[J].中国电机工程学报,2014,34(1):153-160.
WEN Yunfeng, JIANG Yufei, SHEN Ce, et al. Research on the model and algorithm of multistage coordinated risk-based dispatch[J]. Proceedings of the CSEE, 2014, 34(1): 153-160.
- [5] 邱威,张建华,刘念,等.计及运行风险的多目标发电优化调度[J].中国电机工程学报,2012,32(22):64-72.
QIU Wei, ZHANG Jianhua, LIU Nian, et al. Multi-objective optimal generation dispatch with consideration of operation risk [J]. Proceedings of the CSEE, 2012, 32(22): 64-72.
- [6] WANG Q, YANG A, WEN F, et al. Risk-based security-constrained economic dispatch in power systems[J]. Journal of Modern Power Systems and Clean Energy, 2013, 1(2): 142-149.
- [7] 裴喜平,郝晓弘,陈伟,等.基于梯度下降法的单相幅锁相环优化设计[J].电力系统自动化,2014,38(2):115-121. DOI: 10.7500/AEPS20130425011.
PEI Xiping, HAO Xiaohong, CHEN Wei, et al. Design principle and parameter calculation for distribution network low voltage anti-islanding devices[J]. Automation of Electric Power Systems, 2014, 38(2): 115-121. DOI: 10.7500/AEPS20130425011.
- [8] 李芳.线性规划法最优潮流的实用化研究[D].北京:中国电力科学研究院,2003.
- [9] 吴阿琴,韦化,白晓清.基于半定规划的(0,1)-经济调度[J].电力系统及其自动化学报,2008,20(2):121-125.
WU Aqin, WEI Hua, BAI Xiaqing. (0,1)-economic dispatch problem based on semidefinite programming[J]. Proceedings of the CSU-EPSCA, 2008, 20(2): 121-125.
- [10] 王永刚,柳焯.牛顿法优化潮流的实用化研究[J].电力系统保护与控制,1999,27(5):6-8.
WANG Yonggang, LIU Zhuo. The research on the practical strategy of the optimal load flow[J]. Power System Protection and Control, 1999, 27(5): 6-8.
- [11] 任博强,蒋传文,彭鸣鸿,等.基于改进遗传算法的含风电场的电力系统短期经济调度及其风险管理[J].现代电力,2010,27(1):76-80.
REN Boqiang, JIANG Chuanwen, PENG Minghong, et al. Short-term economic scheduling model including wind park based on improved genetic algorithm and risk management[J]. Modern Electric Power, 2010, 27(1): 76-80.

- [12] 黄小庆,杨夯,陈颀,等.基于LCC和量子遗传算法的电动汽车充电站优化规划[J].电力系统自动化,2015,39(17):176-182. DOI:10.7500/AEPS20150323009.
HUANG Xiaoqing, YANG Hang, CHEN Jie, et al. Optimal planning of electric vehicle charging stations based on life cycle cost and quantum genetic algorithm[J]. Automation of Electric Power Systems, 2015, 39(17): 176-182. DOI: 10.7500/AEPS20150323009.
- [13] ADARYNI M R, KARAMI A. Artificial bee colony algorithm for solving multi-objective optimal power flow problem[J]. International Journal of Electrical Power & Energy Systems, 2013, 53(1): 219-230.
- [14] 王晶,陈骏宇,蓝恺.基于实时电价的微网PSO最优潮流算法研究[J].电力系统保护与控制,2013,41(16):34-40.
WANG Jing, CHEN Junyu, LAN Kai. PSO optimal power flow algorithm for a microgrid based on spot power prices[J]. Power System Protection and Control, 2013, 41(16): 34-40.
- [15] 舒隽,张粒子,王雁凌,等.基于竞价的日发电计划混合智能优化算法[J].电力系统自动化,2002,26(21):34-38.
SHU Jun, ZHANG Lizi, WANG Yanling, et al. Bid-based daily generation scheduling using a mixed intelligence optimal algorithm[J]. Automation of Electric Power Systems, 2002, 26(21): 34-38.
- [16] 庄福振,罗平,何清,等.迁移学习研究进展[J].软件学报,2015,26(1):26-39.
ZHUANG Fuzhen, LUO Ping, HE Qing, et al. Survey on transfer learning research [J]. Journal of Software, 2015, 26(1): 26-39.
- [17] 王皓,高阳,陈兴国.强化学习中的迁移:方法和进展[J].电子学报,2008,36(增刊1):39-43.
WANG Hao, GAO Yang, CHEN Xingguo. Transfer of reinforcement learning: the state of the art [J]. Acta Electronica Sinica, 2008, 36(Supplement 1): 39-43.
- [18] LIU Y, PASSINO K M. Biomimicry of social foraging bacteria for distributed optimization: models, principles, and emergent behaviors[J]. Journal of Optimization Theory & Applications, 2002, 115(3): 603-628.
- [19] GAZI V, PASSINO K M. Bacteria foraging optimization[M]. Germany: Springer Berlin Heidelberg, 2011.
- [20] 张孝顺,郑理民,余涛.基于多步回溯 $Q(\lambda)$ 学习的电网多目标最优碳流算法[J].电力系统自动化,2014,38(17):118-123. DOI:10.7500/AEPS20140513010.
ZHANG Xiaoshun, ZHENG Limin, YU Tao. Multi-objective optimal carbon emission flow calculation of power grid based on multi-step $Q(\lambda)$ learning algorithm[J]. Automation of Electric Power Systems, 2014, 38(17): 118-123. DOI: 10.7500/AEPS20140513010.
- [21] NI M, MCCALLEY J D, VITTAL V, et al. Online risk-based security assessment [J]. IEEE Power Engineering Review, 2002, 22(11): 59-59.
- [22] 黄家栋,张富春,周庆捷.基于输电断面识别的电力系统连锁故障风险评估模型[J].电力系统保护与控制,2013,41(24):30-35.
HUANG Jiadong, ZHANG Fuchun, ZHOU Qingjie. Risk assessment model of cascading failures in power system based on identification of transmission section [J]. Power System Protection and Control, 2013, 41(24): 30-35.
- [23] 李文沅.电力系统风险评估:模型方法和应用(精)[M].北京:科学出版社,2006.
- [24] 俞国燕,郑时雄,刘桂雄,等.复杂工程问题全局优化算法研究[J].华南理工大学学报(自然科学版),2000,28(8):104-110.
YU Guoyan, ZHENG Shixiong, LIU Guixiong, et al. Study on global optimization algorithms for complex Engineering Problem[J]. Journal of South China University of Technology (Nature Science), 2000, 28(8): 104-110.
- [25] 徐豪,张孝顺,余涛.非理想通信网络条件下的经济调度鲁棒协同一致性算法[J].电力系统自动化,2016,40(14):15-24. DOI: 10.7500/AEPS20160112003.
XU Hao, ZHANG Xiaoshun, YU Tao. Robust collaborative consensus algorithm for economic dispatch under non-ideal communication network [J]. Automation of Electric Power Systems, 2016, 40(14): 15-24. DOI: 10.7500/AEPS20160112003.
- [26] YU T, LIU J, CHAN K W, et al. Distributed multi-step $Q(\lambda)$ learning for optimal power flow of large-scale power grids[J]. International Journal of Electrical Power & Energy Systems, 2012, 42(1): 614-620.
- [27] 余涛,周斌,陈家荣.基于 Q 学习的互联电网动态最优CPS控制[J].中国电机工程学报,2009,29(19):13-19.
YU Tao, ZHOU Bin, CHAN Kawing. Q -learning based dynamic optimal CPS control methodology for interconnected power systems[J]. Proceedings of the CSEE, 2009, 29(19): 13-19.
- [28] 黄文涛,邓长虹,汪志强,等.基于外点罚函数法的实时低频切泵策略[J].中国电机工程学报,2013,33(16):104-111.
HUANG Wentao, DENG Changhong, WANG Zhiqiang, et al. A real-time strategy for under frequency pump shedding of pumped power storage station based on external point function method[J]. Proceedings of the CSEE, 2013, 33(16): 104-111.
- [29] SUTTON R, BARTO A. Reinforcement learning: an introduction[M]. USA: MIT Press, 1998.
- [30] IEEE Reliability Test System Task Force. IEEE reliability test system [J]. IEEE Trans on Power Apparatus & Systems, 1979, 98: 2047-2054.
- [31] FOTUHI-FIRUZABAD M, BILLINTON R, ABORESHAIID S. Spinning reserve allocation using response health analysis [J]. IET Proceedings: Generation Transmission and Distribution, 1996, 143(4): 337-343.

韩传家(1992—),男,硕士,主要研究方向:电力系统风险调度与可靠性评估。E-mail: chuanjia71@126.com

张孝顺(1990—),男,通信作者,博士研究生,主要研究方向:电力系统优化运行与控制。E-mail: xszhang1990@sina.cn

余涛(1974—),男,博士,教授,主要研究方向:复杂电力系统的非线性控制理论和仿真。E-mail: taoyul@scut.edu.cn

(编辑 孔丽蓓)

(下转第 97 页 continued on page 97)

Optimization Algorithm of Reinforcement Learning Based Knowledge Transfer Bacteria Foraging for Risk Dispatch

HAN Chuanjia, ZHANG Xiaoshun, YU Tao, QU Kaiping

(School of Electric Power, South China University of Technology, Guangzhou 510640, China)

Abstract: Referring to the load and fault uncertainty during power system operation, the risk assessment theory is introduced into economic dispatch. Moreover, a new knowledge transfer bacteria foraging optimization (TBFO) algorithm is proposed for risk based economic dispatch, which is developed by combining bacteria foraging optimization (BFO) and the try-error mechanism of Q-learning. Besides, the knowledge matrix is updated by multiple agents with cooperative collaboration, in which the knowledge extension is adopted to handle the curse of dimension. After obtaining all the optimal knowledge matrices, the convergence of the online learning can be accelerated by knowledge transfer. The performance of TBFO has been fully tested for risk based economic dispatch on the IEEE RTS-79. The simulation demonstrates that the convergence rate of TBFO can be approximately 9 to 20 times faster than that of classical intelligent algorithm, while the quality of the obtained optimal solution can be guaranteed. Hence, it is suitable for fast risk based economic dispatch of large-scale and complex power grid.

This work is supported by National Basic Research Program of China (973 Program) (No. 2013CB228205) and National Natural Science Foundation of China (No. 51477055).

Key words: knowledge transfer; bacteria foraging; reinforcement learning; risk dispatch