

模糊物元分析在经验库算法中的性能优化

Performance Optimization of Experience-Database Based on Fuzzy Matter-Element Analysis

李金龙 黄 英 (浙江传媒学院 电子信息学院 浙江 杭州 310018)

摘要: 经验往往是决定任务成败的关键因素。算法在原有经验库算法基础之上引入特征因子和模糊物元分析过程, 首先对发生事件或任务作归一化代码处理, 接着搜索存在的解决方案及组合解决方案并将其存入有序解集, 然后将有序解集中各解决方案的特征因子作为输入进行处理, 形成最优解决方案, 最后更新有序解集中各解决方案的动态特征因子。实验表明, 优化算法增加了处理时间, 但有效提升了解的优劣度, 改进了经验库的使用性能, 更好地满足了实际的需要。

关键词: 经验库 特征因

1 引言

在人们生活的各个领域, 经验往往是决定任务成败的关键因素。领域专家根据积累的经验知识, 针对各自领域所发生的各种问题, 能够给出正确的、恰当的、合理的解决方案, 例如: 医学家能够根据病症对症下药; 电子、机械专家能够根据所发生的电子、机械故障而提出合理的处理方案; 考古学家能够根据现有的文物推断出几千万年前发生的事实等等。计算机技术发展至今, 有没有可能综合各领域的发展成果通过人与计算机的协作来实现一种专家系统, 用于模拟人类来解决各领域中的实际问题呢?

多年来, 许多领域的科学家一直致力于使用计算机技术模拟人工智能来解决发生在各个领域较为复杂的问题。然而, 受限于计算机技术发展水平及生物、医学等领域对生物智能的研究水平, 使用计算机完全模拟生物智能来解决问题仍然是非常困难的。人工智能提供了一条可行的路径, 通向于解决机器智能和生物智能之间的差距。

本文基于原有经验库算法基础上, 引入特征因子及模糊物元分析过程。算法首先在原有经验库算法的

基础上, 增加了若干特征因子, 再通过模糊物元分析选择解集中的最优解, 从而提供给用户响应所发生事件对应的最佳解决方案。从本文结论来看, 特征因子和模糊物元分析过程的加入, 可以显著改善经验库的使用性能。在经验库的使用过程中, 我们还可以通过修正特征因子来不断提升经验库的可用性。

2 经验库问题描述

不同的领域, 经验的定义是不同的, 下面针对广播电视发射机智能监控系统(简称 BMS)给出具体的问题描述。

BMS 是通过对广播电视发射机进行实时监控与处理分析, 从而保障广电部门无故障播出的一套智能监控系统。BMS 采样量包括模拟量与状态量, 对发射机的动作处理主要通过控制量实现的。经验库中的经验来自于发射机实际运行过程中所发生的事件(模拟量与状态量的组合)及其解决方案(通过控制量及人工提示操作), 它的建立可由领域专家手工增加和 BMS 运行过程中按照算法自动增加。我们对于某类发射机某型号的单个或多个具体的模拟量、状态量越限(报

警)或出错定义一条经验,当隶属于该类型的发射机出现了经验所定义的事件时,系统将给出具体解决方案,经验库算法工作流程如图 1 所示。

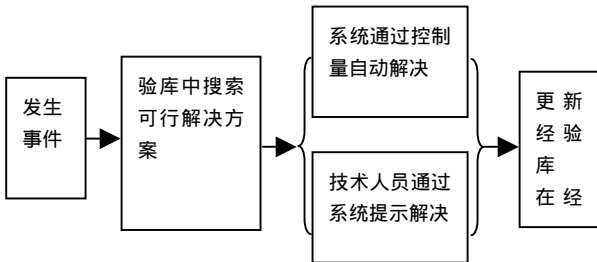


图 1 经验库原理框图

在搜索经验库中可行解决方案时,由于经验库中的解决方案是动态增长和修改的,也是不确定的(即经验库在某个时刻对某一具体事件还未形成最佳匹配解决),因此存在这样一个事实:匹配度高的解决方案未必就是发生事件的最佳解决方案,而匹配度低的解决方案可能刚好适用于该事件的解决。考虑到这一事实,本算法在进行解决方案搜索时,不将匹配度作为唯一的选择指标,我们还需考虑解决方案使用概率、检索代价、专家对解决方案的推荐度及模糊物元分析度等特征因子。

3 算法设计思想

3.1 算法说明

初始经验库的形成一般是由领域专家完成的。领域专家根据某一事件(e),给出一个或多个解决方案(S1, S2...),并且对每个解决方案给出推荐度(R1, R2...)。如图 2-1 所示,经验库在使用过程中,解决方案和推荐度将被动态修正或增加。

设 $E = \{e_1, e_2, \dots\}$ 为事件的集合。

$e = (S, R)$, 其中 e 为某一事件(如前所述,在 BMS 中,它由若干模拟量和状态量的组合而成), $S = \{S_1, S_2, \dots\}$ 为解决方案的集合, $R = \{R_1, R_2, \dots\}$ 为解决方案所对应领域专家推荐度集合。

发生事件与所匹配事件和事件与解决方案有如下关系:

设经验库中存储事件 m 个,解决方案共 n 个,事件匹配解决方案上限 k 个,则发生事件与所匹配事件

和事件与解决方案对应关系如表 1、表 2 所示:

表 1 事件匹配度

发生事件	匹配事件 1	匹配事件 2	...	匹配事件 m
1	P_{11}	P_{12}	...	P_{1k}
2	P_{21}	P_{22}	...	P_{2k}
...

表 1 中, $P_{ij} \in [0, 1]$ ($i=1, 2, \dots; k \geq j \geq 1$)

表 2 事件与解决方案对应关系

事件	解决方案 1	解决方案 2	...	解决方案 k
1	X_{11}	X_{12}	...	X_{1k}
2	X_{21}	X_{22}	...	X_{2k}
...
m	X_{m1}	X_{m2}	...	X_{mk}

表 2 中, $X_{ij} \in [0, n]$, $m \geq i \geq 1, k \geq j \geq 1$, 向量 $(X_{11}, X_{12}, \dots, X_{1k})$ 、 $(X_{21}, X_{22}, \dots, X_{2k})$, ..., $(X_{m1}, X_{m2}, \dots, X_{mk})$ 具有邻域特征。对每个事件,相应解决方案小于 K 时,以 0 填充(表示不存在)。根据表 2,我们总结事件与解决方案特征如下:

(1)事件与解决方案之间是多对多的关系。即对于某一事件,解决方案有 n 种($n >= 0$);同样,一个解决方案也可应用于多个事件。

(2)对于某一事件,其解决方案的分布具有邻域特征。即相对于某一事件,其最优的解决方案分布通常情况集中在一块区域。当我们搜索到一个解决方案时,意味着,最优解决方案可能位于该解决方案所在邻域的附近或是其本身。

(3)事件之间具有包含性,某一事件可能是其它 n 个事件的组合;同样,解决方案也具有包含性,某一解决方案可能是其它 n 个解决方案的组合($n >= 1$)。

3.2 经验库搜索算法的选择

GA 具有很好的总体搜索能力,但其局部寻优能力较差。相对于很多应用问题而言,存在许多专门针对该问题的知识型启发算法,其局部寻优能力较高。对于经验库中经验的搜索,模拟退火算法能够很好地调节种群的个体性能且局部寻优能力强,因此,经验库

的搜索，我们使用 GA 与模拟退火混合算法^[1,5]。

3.3 原经验库算法的设计

经验库中，经验的存储我们使用数据库来实现，某类发射机某型号的单个或多个模拟量和状态量的组合与解决方案及推荐度定义为一条记录。初始经验库形成后，我们需要对数据库中每条记录的各字段参数作归一化代码处理，为后面的经验库搜索做准备。经验库使用时，算法将所发生事件 e 进行编码，使用混合 GA^[1,5]搜索经验库中的最优解决方案。搜索步骤如下：

- 第一步：搜索已存解决方案；
- 第二步：搜索组合解决方案；
- 第三步：确定最优解决方案。

根据 3.1 (见 3.4.1) 我们可知，这里搜索到的解决方案可能不只一个，当存在多个时，我们根据推荐度的高低，来决定哪个是最优解决方案。值得注意的是：

(1) 搜索出来的解决方案并不一定是领域专家给出的解决方案。如图 1，一种情况是，解决方案是由算法动态添加或修正的；另一种情况是，这个解决方案是由使用人员手工添加或修正的。

(2) 搜索出来的解决方案有时不仅仅对应单个事件。初始化时，领域专家将这个解决方案赋予了多个事件。

3.4 特征因子及模糊物元分析的引入

基于原有经验库算法^[2]，我们引入特征因子及模糊物元分析过程。虽然 GA 具有很好的全局寻优能力且模拟退火具有较强的局部寻优能力，但是由于事件与解决方案之间关系的复杂性 (见 3.1)，原有经验库算法仍然很容易陷入局部最优解^[3]，特征因子的引入是为了增加解的判断指标，模糊物元分析的引入是评判相应邻域中各解决方案优劣度。

3.4.1 优化算法说明

在原经验库算法基础之上，我们增加若干特征因子：使用概率、检索代价、模糊物元分析度、事件匹配度等。其中，解决方案所对应的使用概率、检索代价、模糊物元分析度等特征因子均初始化为同一数值。在经验库的使用过程中，这些数值将被动态修正。并定义有序解集，搜索过程中，根据适应度函数，选择

合适的解存入解集。解集变动如下：

$S = \{S_1, S_2, \dots, S_n\}$, $n \in [0, M]$, S 为有序解集, S_n 为一个解及其相应特征的无序集合, M 为最大解集个数。初始时, $n = 0$, 随着混合 GA 搜索的进度, n 会逐步增长, 直至 M 。当 $n = M$, 搜索到另一个更好的解时, 我们将从 S 中删除一个解集。其中, 每添加或替代一个解, 算法会对 S 按解的优劣程度进行排序 (一般 M 较小, 排序在算法中所花费代价可以忽略不计)。

3.4.2 优化算法设计原理

优化后的经验库中每条经验是由事件、解决方案及其解决方案若干特征所构成的。这些特征主要包括：专家推荐度 (r)、经验使用概率 (e)、检索代价 (c)、模糊物元分析度 (f) 等。在初始经验库中，客观特征 e、c、f 等为一常量。

初始经验库构造完毕后，首先我们需要对事件作归一化代码处理，以便于使用混合 GA 进行最优解决方案的搜索。接下来需要提供一些实例对经验库进行搜索训练，以完成对 e、c、f 等参量的初始修正工作^[4]。基于事件与解决方案的特征，我们使用 GA 对经验库中的事件进行搜索，使用模拟退火策略来调节种群的个体性能^[5]。搜索过程中，算法保留较优个体，存入有序解集。搜索完毕后，算法将有序解集中的元素所对应解决方案的客观特征 e、c 进行相应修正。

搜索出来的有序解集，根据解决方案的特征，通过模糊物元分析与处理，最后给出事件解决方案优劣次序并修正各解决方案的模糊物元分析度。

已知搜索模型^[6]:

$$\begin{cases} \min E(X) = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j + \sum_{i=1}^n b_i x_i + C \\ \text{或} \\ t(X) > t_{max} \end{cases} \quad (1)$$

其中, $X = (x_1, x_2, \dots, x_n) \in \{-1, 1\}^n$, $W = (w_{ij})_{n \times n}$ 是一对称、对角元非负的实矩阵, $b = (b_1, b_2, \dots, b_n)^T \in R^n$ 是实向量, C 是常数; $t(x)$ 表示已耗费时间, t_{max} 表示所设定的最大搜索时间。

模糊物元处理模型^[7,8]:

表 3 经验库算法特征因子

解决方案(s)	适应度(m)	专家推荐度(r)	使用概率(e)	检索代价(c)	模糊物元分析度(f)
s1	m1	r1	e1	c1	f1
s2	m2	r2	e2	c2	f2
...

设 $R = (M, C, X)$, M 表示事物, C 表示特征, X 表示量值。记隶属函数为 $u(x)$, $u(x) \in [0, 1]$ 。

如表 3 所示, 特征 m 、 r 、 e 越大越优。则其变换公式为:

$$u_{ji} = X_{ji} / \max X_{ji}; \tag{2}$$

特征 c 越小越优。则其变换公式为:

$$u_{ji} = \min X_{ji} / X_{ji}; \tag{3}$$

特征 f 接近某个常数越优。则其变换公式为:

$$u_{ji} = \min(X_{ji}, u_0) / \max(X_{ji}, u_0). \tag{4}$$

上面 3 个公式中, u_{ji} 表示第 j 个解决方案第 i 项特征的从优隶属度; $\max X_{ji}$ 是各特征中每项特征的最大值; $\min X_{ji}$ 是各特征中每项特征的最小值; $\min(X_{ji}, u_0)$ 表示在 X_{ji} 和 u_0 中取最小的; $\max(X_{ji}, u_0)$ 表示在 X_{ji} 和 u_0 中取最大的 ($j=1, 2, \dots, m; i=1, 2, \dots, n$)。

(1) 将表 3-3 数据通过公式 (3.2), (3.3) (3.4) 转换为关联系数矩阵, 矩阵中每个元素的值设为 ξ_{ij} , 则 $x_i(t+1) \in [0, 1]$ 。

(2) 以 R_w 表示每个解决方案各特征的权重复合物元, 以 R_k 表示由 m 个关联度所组成的关联度复合模糊物元 (关联度向量), 采用加权平均集中处理, 则有 $R_k = R_w * R_\xi$, 这里运算符 “*” 可以有五种情况^[8]。根据 BMS 中所使用的权重复合物元, 我们使用

$$R_k = \left\{ \begin{matrix} M_j & \dots & M_m \\ k_1 = \sum_{i=1}^n w_i \xi_{1i} & \dots & \end{matrix} \right\} \tag{5}$$

所有权重参与运算, 关联度包括所有因素的共同作用。

(3) 关联度处理

$K^* = \max(k_1, k_2, \dots, k_m)$, K^* 所对应的就是最符合的解决方案。

整个算法步骤如下:

初始化工作。设定有序解集; 对事件进行编码, 作归一化处理, 使初始种群 $X(0) \in \{-1, 1\}^N$, 指定交叉概率 P_c , 变异概率 P_m , 温度冷却系数 D 、退火初始温度 T 及搜索终止准则。设 $n=0$ 并定义适应度度量 $J(X) = -E(X)$, 设定 t_{max} 的值。

根据适应度度量, 使用 GA 算法在解决方案空间 (静态存储解决方案和动态组合解决方案) 里搜索。

对有序解集进行处理, 使用择优添加原则 (3.2, 3.3, 3.4), 更新有序解集。

若连续进化 k 次而有序解集无显著变化, 则修改退火温度。即令 $T = CT$ ($C \in (0, 1)$), 转步骤 2)。

判断搜索终止准则 (3.1), 输出有序解集; 否则, 转步骤 2)。

将有序解集作为输入, 通过模糊物元处理 (3.5), 输出最优个体所对应的最优解决方案, 更新有序解集所对应特征因子 (表 3)。

4 算法设计思想

为了验证优化算法的有效性, 我们从原经验库中选取了一批历史发生事件进行测试, 比较了原有经验库算法和优化算法的性能区别。

表 4 历史经验库数据

发射机类型	总事件数 (m)	解决方案 (n)	匹配解决方案上限 (k)
调幅发射机	1782	2876	4
调频发射机	1017	1322	3
电视发射机	837	1070	3
...

(注意: 表 4 中解决方案是指静态存储的解决方案, 而搜索过程中所用到的组合解决方案数是根据发生事件动态变化的。)

这里, 我们分别选取了某型号调频发射机, 调幅发射机和电视发射机三类发射机作为实验测试对象 (说明: 这里型号与品牌有关, 品牌从略)。某型号调幅发射机选取了 109 个事件, 某型号调频发射机选取了 68 个事件, 某型号电视发射机选取了 30 个事件。测试事件我们用文本文件作为输入, 运算后, 程序内

部进行计时,算法给出的每个最优解决方案,我们由领域专家给出解的优劣度,使用加权平均计算出平均优劣度。

表 5 实验结果表

发射机类型	测试事件数	经验库算法		优化算法	
		总耗费(小时)	优劣度(%)	总耗费(小时)	优劣度(%)
调幅	109	2.22	61.17	2.64	84.52
调频	68	1.45	71.18	1.67	80.24
电视	30	0.73	81.24	0.81	87.05

硬件测试环境: DELL PowerEdge 2950 (标准配置)

从表 5 中,我们可以算出,原经验库算法一个事件的平均处理时间约为 73.32~87.6 秒,最优解决方案的平均优劣度为 71.20,而优化的算法平均处理时间约为 87.20~97.2 秒,最优解决方案的平均优劣度为 83.94。实验表明,特征因子和模糊物元分析过程的加入增加了处理时间,但优劣度有了很好的提升,能更好地满足实际的需要。

5 结束语

经验库算法中的解集具有邻域特征,即解集主要分布在解空间里的一块集中的区域,提升优劣度的主

要代价耗费在这块集中区域的选取上。本文基于原有经验库算法,引入了特征因子和模糊物元分析过程,能以较少代价获取较高的最优解优劣度的提升,进一步提高经验库的适用性。

参考文献

- 1 Chakraborty U K,Dastidar D G. Using reliability analysis to estimate the number of generations to convergence in genetic algorithms. Information Processing Letters, 1993,2:73-110.
- 2 Leung K S,Duan Q,Xu Z B , Kwong C K. A new model of simulated evolutionary computation: convergence analysis & specifications. IEEE Trans. On Evolutionary Computation 1999,5(1):1-15.
- 3 Salhi S,Quen N M. A hybrid algorithm for identifying global and local minima when optimizing functions with many minima. European Journal of Operational Research 2004,155(1):51-67.
- 4 刘智,端木京顺,等.基于熵权多目标决策方案评估方法研究,数学的实践与认识,2005,35(10):116-118.
- 5 潘全科,朱剑英.基于进化算法和模拟退火算法的混合算法.机械工程学报,2005,41(6):224-227.
- 6 徐宗本.计算智能-模拟进化计算.北京:高等教育出版社,2004:50-60.
- 7 杨纶标,高英仪.模糊数学原理及应用(第3版).广州:华南理工大学出版社.2003:37-46,108-111.
- 8 李鸿吉.模糊数学基础及实用算法.北京:科学出版社,2005:352-386.