

文章编号: 1001-0920(2012)07-0983-08

## 基于遗传算法的范畴化风险决策知识重组方法

王庆全

(东北财经大学 数学与数量经济学院, 辽宁 大连 116025)

**摘要:** 知识重组是风险管理中辅助决策者获得决策信息的重要环节, 它将应急问题情境具有强关联的文本性决策知识提供给决策者. 在范畴化结构的基础上, 提出一种基于遗传算法面向语义关联的知识重组方法, 为风险决策知识快速获取所需的高维语义信息处理提供了新的途径. 该算法从遗传编码、算子、收敛控制和参数选择 4 个方面来实现进化过程. 仿真实验结果表明, 在强制收敛的条件下, 该方法可获得较好的重组可靠性.

**关键词:** 风险决策; 知识重组; 遗传算法; 范畴知识

**中图分类号:** F062.3

**文献标识码:** A

## GAs based categorical knowledge reorganization for risk decision-making

WANG Qing-quan

(School of Mathematics and Quantitative Economics, Dongbei University of Finance and Economics, Dalian 116025, China. E-mail: dlwqq@hotmail.com)

**Abstract:** Knowledge reorganization is one of the key links in supporting decision-makers to acquire the information of decision-making in risk management. This process provides decision makers with textual knowledge which has strong correlation with decision-making problem situations. This paper proposes a knowledge reorganization method based on categorical structure semantic correlations based on genetic algorithm. This method provides a new way for the high-dimensional semantic information processing which are necessary for the quick knowledge acquisition of risk decision-making. The process of genetic evolution is achieved through four aspects including genetic coding, operators, convergence control and parameter selection. Simulation experiment results show that this method can obtain better reliability of reorganization under forced convergence conditions.

**Key words:** risk decision-making support; knowledge reorganization; genetic algorithms; categorical knowledge

### 1 引言

面对各类地震、疫病流行、金融危机等突发公共事件, 各级政府和应急控制中心的决策者需要在短时间内, 针对发生的突发事件的问题情境作出高效的决策和指挥. 而这一过程是与各级政府制定的应急相关法律、法规、预案和经验总结紧密相关的. 随着突发事件的复杂化, 这类应急文档的数量不断增多, 根据特定突发事件问题情境来获得决策背景知识愈发困难. 因此, 利用现有的计算机和网络技术, 快速而有效地向决策者提供解决特定问题所必须的决策背景知识是提高应急反应决策水平的重要保障. 目前, 对决策背景知识的人为筛查和判断往往显得力不从心. 根据复杂的应急问题情境, 对背景知识进行面向语义的重组<sup>[1]</sup>是实现有效辅助决策的重要环节.

在决策系统中, 知识不同形态具有一定结构, 如表示复杂语义关系的范畴化知识结构<sup>[2]</sup>, 用以避免知识处理过程中过度的信息缺失和提高决策知识重组过程中的可靠性. 应急决策中的问题情景和背景知识都可以表示成这种特定结构的知识. 知识重组即是解决这两种知识结构间匹配计量<sup>[3]</sup>的组合优化问题. 其主要原因在于: 对于问题情境和背景知识往往不具有相同的顺序和内容, 也就是说, 当面对决策需求的问题情境时, 并不可能按照预案、法律法规内条款进行描述.

由于面向语义的决策知识处理难以找到有效的导数信息来帮助求得最优解, 本文运用遗传算法来完成这个重组过程. 遗传算法(GA)是模拟遗传选择和自然淘汰的生物进化过程的全局优化搜索算法, 它具

收稿日期: 2011-01-11; 修回日期: 2011-08-16.

基金项目: 国家自然科学基金项目(70901016, 91024003).

作者简介: 王庆全(1974-), 男, 讲师, 博士, 从事风险分析计量、智能计算方法的研究.

有较强的鲁棒性, 适于并行处理<sup>[4]</sup>. 因此, 解决知识重组过程中遗传算法的运用需要解决 3 方面的问题: 1) 范畴知识结构上有效的遗传编码; 2) 个体适应度的计量问题; 3) 进化过程的终止判定条件.

## 2 范畴化知识重组

### 2.1 范畴化知识结构

范畴论是一个年轻的数学分支, 它着重于将不同数学领域的结构化概念, 通过统一的方式进行表示. 范畴论早在 20 世纪 90 年代初期被引入了计算机科学领域<sup>[6]</sup>, 这个理论同时也被看作知识科学和知识工程的数学基础. 同时, 范畴论被引入了知识工程, 如本体工程中的知识表示<sup>[6]</sup>, 元本体的描述, 元本体结构 (ACO)<sup>[7]</sup>, 领域分析<sup>[8]</sup>, 本体合并<sup>[9]</sup>, 本体融合<sup>[10]</sup>, 语义网中复杂数据结构表示<sup>[11]</sup>等相关领域.

从定型范畴论的视角, 一个知识片断就是一个定型范畴<sup>[12]</sup>, 它由一类对象、一类态射、一类态射定型和一类定型组合规则构成, 并具有两点间多边的网状的拓扑结构, 它能通过对范畴化结构间的相似度计算, 实现有效的知识匹配<sup>[3]</sup>. 这种在范畴化知识结构上的匹配, 更多地体现在其特有的拓扑结构特性上.

范畴知识结构的本质是对具有强语义关系的背景知识的特征描述, 并且这一过程尽量避免关系过度约简所带来的信息缺失<sup>[13]</sup>.

### 2.2 知识重组过程

针对上面的 3 个问题, 可通过对范畴定型的运用来解决遗传算法的编码问题; 运用两个知识片断间的相似度计量个体进化过程中的适应值; 运用知识片断的语义融合度计量种群适应度<sup>[14]</sup>, 进而进行遗传进化终止条件的判断. 它根据这样 1 个假设, 即在知识重组的过程中, 获得的背景知识片断的语义关联度是随着背景知识的相关性增大而增大的.

知识重组的过程主要解决的问题是相对固定的背景知识和相对变化的问题情境描述之间重组的计量问题, 是一个自动聚类过程. 这个过程主要体现在: 突发事件发生之前, 文本性背景知识可以被描述成若干个具有独立主题的范畴化知识片断; 而在突发事件发生之后, 其文本性的问题情境描述也可以用同样的结构进行表示. 根据知识片断匹配的计量方法, 对供需双方进行基于概念的预匹配, 其中每一情境知识片断即是一个个体, 个体中的每一个定型和超定型即是一个基因, 初始问题描述知识片断中的定型和超定型的集合便构成了基因空间, 即基因池, 它是范畴化概念模型中的定型和超定型集合. 进化的过程依据两个适应性进行: 1) 依据个体同其关联的背景知识片断相似度计量的个体适应度; 2) 通过面向语义融合统

计距离的群体适应性. 经过若干代的进化过程后, 最终生成的是满意解集.

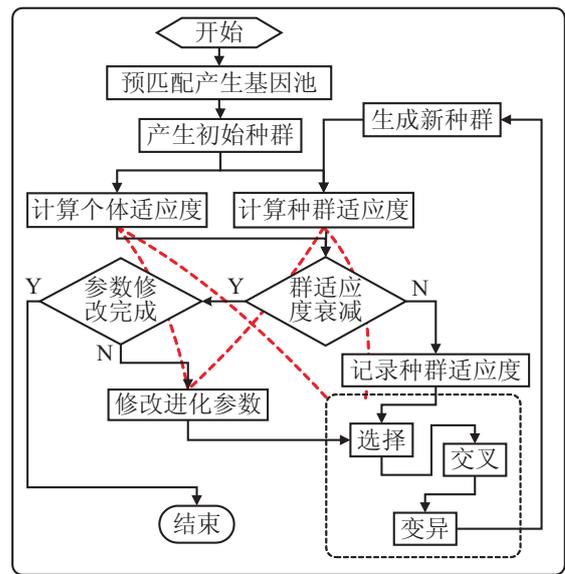


图 1 遗传算法的流程

图 1 给出了遗传算法实现知识重组的流程, 其中预匹配过程是在应急问题的范畴化描述后, 针对涉及的概念之间快速匹配符合一定阈值的决策背景知识, 形成一定数量的、与问题描述涉及概念相关的背景知识, 生成基因池 (GP) 和预匹配结果集.

若应急问题根据不同的情形被分散地描述成  $m$  个知识片断, 在预匹配的条件下, 有  $n$  个背景知识片断, 一般情况下会有  $m \ll n$ , 这也正是知识重组过程需要解决的问题. 经过预匹配, 即可认定本文所需要的知识片断存于其中, 知识需要对应急问题知识片断 (个体) 进行必要的基因 (定型) 调整来迅速发现这些背景知识片断. 预匹配的过程只是为缩小搜索范围的一个粗略的知识获取方式.

遗传进化过程中个体适应度和种群适应度是控制整个遗传过程的重要计量方式, 它们不仅会影响遗传的选择、交叉和变异的过程, 而且还能够影响进化过程选择、交叉和变异概率的调整, 加快算法的收敛速度. 另外, 种群的变化并不是个体的完全的替换, 而是在原有个体的基础上作进一步的调整, 以使得算法能够将个体中的基因进行优化组合.

## 3 知识重组的遗传进化

### 3.1 范畴化的遗传编码

目前, 遗传编码大致分为两派, 一派是基于模式定理, 用尽量少的符号来编码; 另一派是以数值优化精度为准, 使用一个基因一个参数的方法. Bosworth 等<sup>[15]</sup>是后一派的开创者. 许多学者<sup>[16]</sup>发现, 采用大符号集编码的遗传算法比采用二进制编码的遗传算法的性能好. Antonisse<sup>[17]</sup>从理论上证明了大符号集编码

的设计可提供更多的模式. 高维语义信息编码的处理需要采用大符号集的编码方式, 并且在处理特定问题时, 为了缩小处理的信息域, 引入了基因池(基因空间)的概念. 基因池<sup>[18]</sup>是指一个族群内所有基因的总和, 也就是说遗传过程中的基因类型包含于基因池中.

实现基于遗传算法的语义关联, 首要的环节是如何给复杂的语义信息进行编码, 并且尽可能地符合遗传编码的特性, 如完备性、封闭性、紧致性、可扩展性、多重性、个体可塑性、模块性、冗余性与非冗余性和复杂性<sup>[19]</sup>. 对于语义处理的遗传编码目前尚没有有效的方法, 其难度主要表现为两个方面: 1) 语义知识难以进行遗传编码; 2) 语义处理的量化问题难以解决, 难以实现编码基础上的进化过程. 然而, 定型范畴论给知识的语义化处理提供了强有力的理论基础, 使得范畴结构基础上的知识重组问题的解决成为可能.

遗传编码是在范畴知识结构的基础上进行的, 具体思想为: 一个个体包含一个单倍染色体, 一个染色体是由问题描述片断构成的单链结构, 每一个染色体是由一系列问题描述的范畴化定型(或超定型)排列而成, 它所代表的是原有描述问题情境的知识片断中所提取的二元关系(或多元关系)的抽象描述. 应急问题具有重组特性, 因而采用变长度的染色体编码, 并且其长度是受交叉算子和变异算子的直接影响. 一个染色体长度  $L = f(i, t)$ , 表示第  $i$  个个体在第  $t$  代的染色体长度. 建立编码空间为

$$S^L = \{I_1, I_2, \dots, I_K\}, k = 1, 2, \dots, K, K = 2^L. \quad (1)$$

$$I_{ii} = (\text{dom}(t_{ii}), \text{cod}(t_{ii}), \text{val}(t_{ii}))^n,$$

$$I_i = \{I_{i1}, I_{i2}, \dots, I_i, \dots, I_{iL}\}.$$

其中: 个体向量  $I_i$  为一个问题情境的范畴化描述, 称为第  $i$  个个体对应的染色体;  $\text{dom}(t_{ii}), \text{cod}(t_{ii}), \text{val}(t_{ii})$  分别表示定型  $t_{ii}$  在概念模型中的域、共域和态射取值,  $n$  表示该定型为几元的定型或超定型, 当  $n > 1$  时, 它表示的是一个超定型.

编码空间中的定型  $t_{ii} (1 \leq i \leq K, 1 \leq j \leq L)$  是由这些定型和超定型的非重复子集所构成的定型的集合, 称之为重组过程中的基因池 GP, 记为

$$GP = \sum_{i=1}^{K'} \sum_{j=1}^{L'} t_{ij}, K' < K, L' < L. \quad (2)$$

将染色体的编码空间转化成问题空间的译码函数, 即

$$\Gamma : \{t_{i1}, t_{i2}, \dots, t_{iL} | t_{ij} \in P\}^L \rightarrow \{e_{i1}, e_{i2}, \dots, e_{iL}, \dots, e_{iL}\}, \quad (3)$$

其中  $e_i$  为一个染色体, 是对应急问题描述的一个知识

片断. 染色体的改变是通过遗传算法过程中的基于定型基因的互操作(交叉、变异算子)实现的, 这个编码过程是对应急问题表示的一个特征提取过程.

### 3.2 适应度测定

遗传算法的求解过程通常是: 首先产生一定数量的染色体, 形成一个种群; 再根据个体适应度函数评价各染色体的优劣, 通过对染色体的选择、交叉及变异的遗传操作, 获得新的个体, 构成新的种群; 如此反复进行, 通过一定代数的迭代后, 得到的最优染色体个体的集合, 即是求得问题的最优解或较优解集. 因为个体适应度是种群中个体生存机会选择的唯一确定指标, 所以适应度函数的形式直接决定着群体的进化行为和算法的收敛性.

知识重组过程的遗传进化是以问题情境为导向的, 因而, 其适应度的评价函数需要从两个角度进行测量, 一个是个体染色体针对背景知识的匹配度测定; 另一个是面向整个问题情境描述的背景知识片断的适应度测定. 前者的测量主要通过基于最大公约子范畴的相似度进行度量, 而后者则需要对关联整个应急决策问题情境的背景知识片断进行语义融合度的测量, 它是建立在这样一个假设的基础上, 即所提取的背景知识本身针对当前的情境问题具有较强的相关度.

**定义 1** 个体染色体适应度测定. 设有个体  $I$ , 包含  $L$  个定型基因  $\{t_{i1}, t_{i2}, \dots, t_{iL}\}$ , 其个体染色体适应度  $\alpha_L(I)$  为

$$\alpha_L(I) = S \left[ x \left( \sum_{i=1}^L t_i \right), x \left( \sum_{i=1}^L q_i \right) \right]; \quad (4)$$

$$S(A, B) = \begin{cases} 1, & d(A, B) = 0; \\ 1 - \frac{d(A, B)}{\sigma}, & d(A, B) < \sigma; \\ 0, & d(A, B) > \sigma. \end{cases}$$

其中:  $\sigma$  是指定的两个定型范畴之间的最大距离阈值;  $x \left( \sum_{i=1}^L t_i \right), x \left( \sum_{i=1}^L q_i \right)$  分别表示一个单链染色体的表现型及其预匹配的背景知识片断, 其匹配的方式是依据知识片断间匹配的计量方法实现的.

**定义 2** 种群适应度测定. 设第  $t$  代种群中包含  $n$  个个体, 表示为  $I_i, 1 \leq i \leq n$ , 每个个体包含  $L = f(i)$  个变长的染色体, 则整个群体的染色体适应度  $\beta_n(I_i)$  为其背景知识中提取知识片断的语义融合度, 记为

$$\beta_n(I_i) = S \left( \sum I_i, a_i \right), \quad (5)$$

其中  $a_i$  表示未被匹配的背景知识中的定型和超定型, 它体现了基于目前的应急问题情境提取的背景知识

片断的相关性大小.

因为面向知识重组的遗传算法需要进行个体的优化而不是替换,所以个体适应度和种群适应度需要对遗传进化过程有必要的约束,使其成为一个优化的过程.在进化过程中,经过交叉和变异的个体需要与上一代的个体进行比对,如果个体适应性优于上一代,则执行交叉算子;否则放弃,选择其他的个体进行交叉.种群适应度也需要与上一代的种群适应度进行比对,如果优于上一代,则此次遗传进化成功;否则调整遗传进化参数,重新进行此次进化过程.因此,个体适应度和种群适应度在进化过程中可以实现遗传算法的启发式过程,尤其是种群适应度的引入可以改变传统遗传算法固定进化代数的限制,作为整个遗传进化过程的终止条件,如果没有种群适应度的约束,则交叉和变异算子也会失去效力.

### 3.3 遗传编码进化过程

设种群初始化为遗传进化过程中的第1步,它一般是通过个体染色体适应度随机选取.在知识重组的过程中,遗传的编码是以应急问题描述的单链结构,所以初始种群即为应急问题描述的初始状态,种群的数量也是应急问题情境描述的知识片断的数量.初始种群的种群适应性测度的计量是在预匹配结果集中选取预匹配值最大的相同数量匹配的背景知识片断.

遗传算子主要是在编码生成的初始种群的基础上,通过基因的选择、交叉和变异实现遗传进化,寻得最优解.其中选择算子是最简单的,其主要操作是选择个体适应度最高的个体,不对其进行任何交叉和变异操作,直接传至下一代.选择算子对于群体多样性具有严格单调减少的作用.经过选择运算后个体的基因保持不变,同时它对基因池也不造成影响,在知识重组过程中,选择的概率应当大于交叉概率和变异概率,只有这样才能够淘汰劣质个体,确保整个种群品质的提高.另外,过低的选择概率会使问题的求解出现早熟现象.

交叉算子是模仿自然界有性繁殖的基因重组过程,而使得种群的品质得以提高,可以支持生成更多优良基因的新个体.在知识重组过程中,交叉算子所解决的问题是在应急问题情境描述中,因对背景知识组织形式的未知而造成的描述顺序上的混乱.

设个体染色体  $I_i = \{I_{i1}, I_{i2}, \dots, I_{ik}, \dots, I_{iL}\}$ ,  $I_j = \{I_{j1}, I_{j2}, \dots, I_{js}, \dots, I_{jL'}\}$ . 假设第  $t$  代分别第  $k$  点和第  $s$  点进行交叉,则交叉后所产生的  $t+1$  代的两个新个体为

$$\begin{aligned} I_i^{t+1} &= M(I_i^t - I_{ik}, I_{is}), \\ I_j^{t+1} &= M(I_j^t - I_{js}, I_{jk}). \end{aligned} \quad (6)$$

知识融合函数  $M(x, y)$  会出现以下4种情况:

**假设 1** 满足  $r(I_{ik}) = r(I_{js}) = 0$ , 且  $S_M(M(I_i^t - I_{ik}, I_{js})) = 0$ , 即定型的阶数相同, 而融合度等于 0.

**假设 2** 满足  $r(I_{ik}) = r(I_{js}) = 0$ , 且  $S_M(M(I_i^t - I_{ik}, I_{js})) > 0$ , 即定型的阶数相同, 而融合度大于 0.

**假设 3** 满足  $r(I_{ik}) = r(I_{js}) \neq 0$ , 且  $S_M(M(I_i^t - I_{ik}, I_{js})) = 0$ , 即定型的阶数不同, 而融合度等于 0.

**假设 4** 满足  $r(I_{ik}) = r(I_{js}) \neq 0$ , 且  $S_M(M(I_i^t - I_{ik}, I_{js})) > 0$ , 即定型的阶数不同, 而融合度大于 0.

显然,对于以上4种交叉结果,假设1和假设3是在交叉后新的基因型与要交叉的染色体无法融合,这2种交叉可以被视为失败的交叉算子,不予考虑.符合假设2和假设3的算子为可行的交叉过程.当满足假设2时,染色体中的基因数量保持不变,染色体中的态射数量也保持不变.但满足假设4时,染色体的基因数量保持不变,但染色体中的态射数量将会随之改变.更为重要的是,在完成一个成功的交叉后,由于融合的过程中需要有新增的态射添加,基因池中基因的类型和数量也将随之改变.

交叉过程实质上是一个定型  $t$  与一个知识片断  $K$  进行融合的过程,期间会从知识库中以推理的形式提取出相应的定型补充到个体染色体和基因池中.除了假设2和假设4应当满足外,还需要符合以下条件:

$$\begin{cases} \Delta t_i \geq 1, \\ S(M((I_i^t - I_{ik}), I_{js}), e_i) - S(I_i^t, e_i) \geq 0, \\ \Delta t_j \geq 1, \\ S(M((I_j^t - I_{js}), I_{ik}), e_j) - S(I_j^t, e_j) \geq 0. \end{cases} \quad (7)$$

其中:  $\Delta t_i$  和  $\Delta t_j$  分别表示个体在交叉后新增的定型数量,  $e_i$  和  $e_j$  分别表示合并前  $I_i^t$  和  $I_j^t$  匹配的背景知识片断,并且基因池的数量也会增加为  $|GP| + \Delta t_i + \Delta t_j$ .

在遗传进化过程中,变异算子主要用来保持种群的多样性.在知识重组过程中,需重新定义变异算子,因为变异算子是决策知识重组结果优劣的关键环节.在应急问题情境描述过程中,可能出现问题情境描述的知识片断与背景知识片断的数量不一致,或者小部分的语义信息是共用的语义信息.根据知识重组和面向语义关联的这一特点,将变异算法定义为3种形式.

**定义 3** 染色体断裂.染色体断裂是将在知识片断中连接度最弱的点分为2个个体,它所解决的问题是问题知识片断描述中包含2个以上的背景知识片断.

设  $I_i^t = (I_{i1}, I_{i2}, \dots, I_{ik}, I_{i,k+1}, \dots, I_{iL})$  为一个染色体,如在基因连接处发生断裂,且设断裂处的对象为  $a$ ,若满足如下条件:

$$\left\{ \begin{array}{l} a \in \{\text{dom}(I_{ik}), \text{cod}(I_{ik})\}, \\ a \in \{\text{dom}(I_{i,k+1}), \text{cod}(I_{i,k+1})\}, \\ \Delta t_{i(1)} = 0, \\ \Delta t_{i(2)} = 0, \\ S(I_{i(1)}^t, e_{i(1)}) - S(I_i^t, e_i) \geq 0, \\ S(I_{i(2)}^t, e_{i(2)}) - S(I_i^t, e_i) \geq 0, \\ r(t(a)) = 1. \end{array} \right. \quad (8)$$

其中:  $\Delta t_{i(1)} = 0$  和  $\Delta t_{i(2)} = 0$  分别表示断裂后态射定型的改变量;  $e_{i(1)}$  和  $e_{i(2)}$  分别表示重新经过匹配过程得到的背景知识片断;  $t(a)$  为包含概念  $a$  的定型. 则该变异算子操作成功, 得到的结果为

$$\begin{aligned} I_{i(1)}^{t+1} &= (I_{i1}, I_{i2}, \dots, I_{ik}), \\ I_{i(2)}^{t+1} &= (I_{i,k+1}, \dots, I_{iL}); \end{aligned} \quad (9)$$

否则为失败, 并且基因池的数量不会增加.

**定义 4** 染色体合并. 染色体合并是将具有相同定型的两个个体进行语义融合, 合并成为一个个体, 并且合并相同的对象. 它所解决的问题是两个问题知识片断描述中包含的背景信息被包含在同一个背景知识片断中.

设个体染色体  $I_i^t = (I_{i1}, I_{i2}, \dots, I_{ik}, \dots, I_{iL})$ ,  $I_i^t = (I_{i1}, I_{i2}, \dots, I_{ik'}, \dots, I_{iL'})$ , 如存在基因  $I_{ik} = I_{j'k'}$ , 并且满足如下条件:

$$\left\{ \begin{array}{l} S(I_i^t, e_i) - S(I_j^t, e_j) \geq \frac{2}{|I_i^t|}, \\ \Delta t_g \geq 0, \\ S(I_i^t, e_i) - S(I_g^t, e_g) \geq 0, \\ S(I_j^t, e_j) - S(I_g^t, e_g) \geq 0. \end{array} \right. \quad (10)$$

其中:  $\Delta t_g$  表示合并个体态射定型的改变量;  $e_i$  和  $e_j$  分别表示合并前  $I_i^t$  和  $I_j^t$  匹配的背景知识片断;  $e_g$  表示重新经过匹配过程得来的背景知识片断;  $|I_i^t|$  表示  $I_i^t$  中的态射数量. 则该变异算子操作成功, 得到的结果为

$$I_g^{t+1} = (I_{i1}, I_{i2}, \dots, I_{ik}, \dots, I_{i,L+L'}); \quad (11)$$

否则失败, 基因池数量增加为  $|GP| + \Delta t_g$ .

**定义 5** 定型基因增补. 基因增补是指在选定的个体上从基因池中选择添加具有相同对象的定型基因, 即新增基因的域或共域的对象属于该个体, 并且采用单点变异的形式. 它所解决的问题是部分小知识点语义信息的移位. 变异操作会影响染色体的长度.

设  $I_i^t = (I_{i1}, I_{i2}, \dots, I_{ik}, \dots, I_{ik'}, \dots, I_{iL})$  为一个染色体, 第  $t$  代在基因  $I_{ik}$  和  $I_{ik'}$  之间的对象  $a$  产生变异, 并且若满足如下条件:

$$\left\{ \begin{array}{l} a \in \{\text{dom}(I_{ik}), \text{cod}(I_{ik})\}; \\ a \in \{\text{dom}(I_{i,k+1}), \text{cod}(I_{i,k+1})\}; \\ S(I_i^t, e_i) - S(M(I_i^t, t(a)), e_i') \geq 0. \end{array} \right. \quad (12)$$

其中:  $e_i'$  表示重新匹配的背景知识片断;  $t(a)$  为包含概念  $a$  基因池中的定型基因. 则该变异算子操作成功, 得到的结果为

$$I_i^{t+1} = (I_{i1}, \dots, I_{ik}, \dots, I_{ik'}, \dots, I_{iL}, t(a)); \quad (13)$$

否则失败, 并且基因池的数量保持不变.

对这 3 种基因突变方式采用顺序选择的方式, 即先选择染色体断裂, 再选择染色体合并, 最后选择定型基因增补的方式, 若有一个符合条件则成功运行变异算子.

### 3.4 遗传参数调整的适应性策略

由于在遗传进化过程中, 固定的交叉和变异概率会对进化过程产生影响, 交叉概率  $P_c$  越大, 群体中新问题描述知识片断的机会增多, 原有的应急问题被破坏的可能性也随之增大. 同样, 变异概率  $P_m$  越大, 整个应急问题描述可以保持更高的多样性, 同时遗传进化趋近于随机搜索. 因此这里考虑交叉概率  $P_c$  和变异概率  $P_m$  参与的参数调整的适应性策略. 变异概率  $P_s$  表示选择概率.  $P_s^{\max}, P_c^{\max}, P_m^{\max}$  分别表示 3 个概率的最大可能取值,  $P_s^{\min}, P_c^{\min}, P_m^{\min}$  分别表示 3 个概率的最小可能取值, 并且可以有

$$P_s = 1 - P_c - P_m. \quad (14)$$

将  $P_c, P_m$  群体的收敛性和个体的适应性相联系, 使得  $P_c$  和  $P_m$  在不同代数的遗传进化过程中采用不同的值. 因而, 有参数  $P_c$  调整适应性策略为

$$P_c = \gamma_c \cdot \left( \frac{(\alpha_{\max} - \bar{\alpha})(P_c^{\max} - P_c^{\min})}{\alpha_{\max} - \alpha_{\min}} \right). \quad (15)$$

其中:  $\alpha_{\max}, \alpha_{\min}$  分别表示进化到当前代数的个体最大和最小的适应度值;  $\bar{\alpha}$  为该个体的平均适应度值;  $\gamma_c$  为调整权系数.

显然, 随着种群中个体的适应度的提高, 其交叉概率应随之降低. 因遗传算法被用于语义处理, 并且应急问题背景的描述相对完备, 所以其变异适应度是随着种群适应度进行衡量的, 即种群适应度越高, 其相应的变异概率应该越低. 参数  $P_m$  调整的适应性策略为

$$P_m = \gamma_m \cdot \left( \frac{\beta(P_c^{\max} - P_c^{\min})}{\beta_0 - \beta} \right). \quad (16)$$

其中:  $\beta$  为进化到当前代数的种群适应度,  $\beta_0$  为初始种群的种群适应度,  $\gamma_m$  为调整权系数. 设立  $\gamma_c$  和  $\gamma_m$  取值的目的在于, 进行参数  $P_m$  和  $P_c$  的调整策略时, 实际上压缩了它们的取值范围, 所以需要按照一定比例进行放大.

## 4 知识重组的仿真实验

### 4.1 应急背景知识的实验环境

实验样本取自国家关于突发卫生公共事件的法律、法规和预案11篇,包括《重大动物疫情应急条例》、《传染病防治法》、《传染病防治法实施办法》、《突发公共卫生事件应急条例》、《食品卫生法》、《进出口动植物检疫法实施条例》、《动物防疫法》、《国境卫生检疫法》、《进出口动植物检疫法》、《植物检疫条例》、《国境卫生检疫法实施细则》。文本性的知识片断划分是按照政策文件中的条款或自然段落进行划分,因为独立的自然段和条款通常具有相对独立的应急领域相关主题。按照段落和条款,将这11篇文档划分为1440个背景知识片断和包含在内的1394个领域概念(不包括同义词标注)。

### 4.2 应急问题情境仿真

实验过程首先根据假定的应急决策问题描述和解决问题所需的背景知识,通过文本形式描述某地区突发的公共卫生事件,根据事件发生发展的时间顺序,将该描述生成范畴化结构的知识片断,选取一个由应急问题描述的知识片断,将其进行范畴化的知识表示处理,表示成具有范畴结构的情境知识片断。

模拟出这样一个问题情境:“某县发生重大动物疫情,可能部分被感染动物已经装载至出境口岸,并且已经发现有人感染疑似感染病例,需要及时隔离处理”。

面对这样的灾情,哪些部门需要采取预防措施,避免灾情扩大,需要采用什么样的现场预防措施,其他哪些相关部门应该做好配合工作,哪个单位负责疑似病例,如何处理。兽医主管部门是否需要同检疫机关联系,医疗机构对被感染人怎样处理,以及由谁来采集病料,哪个部门来实施封锁,哪个部门负责现场工作,上报给政府的哪个部门,运输工具营运单位应该怎样做?

这个问题情境描述可分为2个部分:1)对应急问题环境的描述(前一个自然段),即对突发事件中待解决的问题的陈述,在什么地点、什么时间发生了什么样的突发事件。2)对所需要获得背景知识的需求性描述,即对突发事件中参与部门、履行职责、涉及任务、处理方法等方面的需求描述。例如,哪些部门需要在什么样的条件下完成什么样的任务,当某个部门遇到何种情况时,应当采取什么样的措施。在问题描述的过程中,这两种应急问题情境的描述虽然有着紧密的联系,但是也有着明显的区别,主要体现在:前者是由发生的事件确定的,具有一定的描述上的不确定性;而后者是对应急决策背景知识的需求,主要通过

应急文档中背景知识的描述方法来体现,具有一定进一步抽象的空间。从应急问题的一般性归纳角度,面向背景知识的需求性描述可以从问题情境描述中提取出来,抽象成一般性的描述框架,以支持建立起针对不同突发事件决策知识获取的一般性应急问题模式。本文主要通过对原始问题描述和一般性应急问题抽象的两种形式的比较,考察知识重组仿真的重组过程。

### 4.3 重组实验设计

根据概念的相似度计算,并且取预匹配的概念数量阈值为2,根据两个问题情境描述的术语获取了101个背景知识片断(重复了4个),它们分散在7个应急文档中,通过范畴化的知识表示形式,将这两个问题情境表示成16个范畴结构的知识片断。

表1给出了知识重组的遗传进化过程开始之前的各项实验数据。最大最小的选择、交叉编译的概率都取 $[0, 0.9]$ ,交叉和变异概率先取其最大最小值的平均值,其调整系数是根据知识片断间匹配的取值范围以及交叉和变异概率的映射获得的。

表1 遗传进化初始参数

序号	名称	值
1	预匹配阈值	2
2	预匹配结果集数量	101
3	基因池数量	108
4	初始化种群数量	16
5	交叉概率调整系数	1.439
6	变异概率调整系数	1.502
7	交叉概率调整步长	0.05
8	变异概率调整步长	0.05
9	最小选择概率	0.1
10	最小交叉概率	0.1
11	最小变异概率	0.1
12	最大选择概率	0.9
13	最大交叉概率	0.9
14	最大变异概率	0.9

### 4.4 重组实验结果

给出遗传进化过程的求解性能,横坐标表示进化的代数,纵坐标表示个体平均适应度和种群适应度,都为 $[0, 1]$ 间的取值。整个遗传进化过程共发生33代后终止,终止时的个体平均适应度和种群适应度分别为0.1622和0.1808。

遗传进化到33代时终止,此时的种群数量为3,基因池数量为142个,并且获得了种群数量不发生变化后的最后5代匹配最优的前3个(可人为设定)背景知识片断。

图2给出了遗传进化结束时的3个个体(情境知识片断)。将问题情境分为问题情境A,问题情境B和问题情境C。给出了最后5代匹配最优的背景知识片断(不重复)与重组后的问题情境的匹配度。

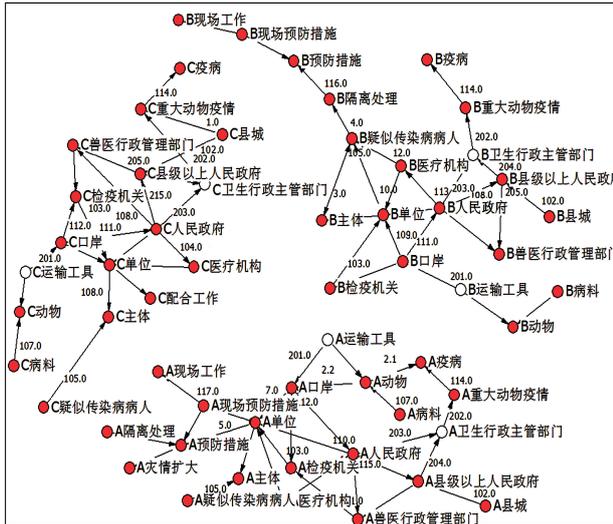


图 2 重组后的问题情境知识片断

表 2 给出了问题情境 A, B 和 C 与这 9 个选取出的背景知识片断之间的语义匹配结果. 面向遗传进化过程的知识重组将应急问题进行有效的补充和组合优化, 进而提高决策知识的语义关联的可靠性, 它将应急决策知识供需匹配的一个无序的状态优化成具有可靠语义关联的有序状态.

表 2 实验结果匹配度

背景知识片断	问题情境	匹配度
1	C	0.1622
2	B	0.0976
3	B	0.0476
4	B	0.1064
5	A	0.1316
6	A	0.1556
7	C	0.1591
8	C	0.1389
9	A	0.1190

### 5 结 论

本文在知识范畴化表示和语义关联计量方法研究的基础上, 提出了一种基于遗传算法的知识重组方法, 它利用范畴知识结构中的定型和超定型来实现遗传编码, 运用应急决策问题情境知识片断建立遗传进化的初始种群, 通过个体与背景知识片断的匹配相似度计量值进行个体适应度的测定, 并运用知识片断语义融合度来计量种群适应度, 以确定种群进化的约束条件和遗传进化过程的终止条件, 并且根据个体适应度和种群适应度的变化来实时地调整进化参数. 经过对突发公共卫生事件样本数据的实验分析, 范畴化知识结构上的面向语义关联的知识重组方法能够较好地优化应急问题情境的描述, 以进行可靠的决策知识关联匹配. 该方法通过范畴结构进行知识重组的遗传编码, 一定程度上保证了知识重组遗传进化过程中的处理精度. 以定型为最小单位, 执行交叉和变异操作

能够较好地保证语义关联的质量和完整性. 通过基于匹配的个体适应度计量和基于语义融合的种群适应度计量使得整个遗传进化过程有一个较好的收敛速度和优化性能. 用于知识重组的遗传算法具有较强的空间搜索能力, 比单纯的语义关联计量匹配具有更高的可靠性.

### 参考文献(References)

- [1] Rong L L. A method of managing the knowledge in government documents for quick response[J]. Int J of Knowledge and Systems Sciences, 2005, 2(1): 67-73.
- [2] 王庆全, 荣莉莉, 于凯. 一种基于范畴论的应急决策概念建模方法[J]. 情报学报, 2009, 28(6): 929-938. (Wang Q Q, Rong L L, Yu K. A category theory based conceptual modeling method for emergency decision-making[J]. J of the China Society for Scientific and Technical Information, 2009, 28(6): 929-938.)
- [3] 王庆全, 荣莉莉, 于凯. 基于最大公约子范畴的应急知识匹配方法研究[J]. 控制与决策, 2009, 24(7): 990-995. (Wang Q Q, Rong L L, Yu K. Emergency decision-making knowledge matching based on greatest common subcategory[J]. Control and Decision, 2009, 24(7): 990-995.)
- [4] 李敏强, 寇纪淞, 林丹, 等. 遗传算法的基本理论与应用[M]. 北京: 科学出版社, 2002. (Li M Q, Kou J S, Lin D, et al. The basis theories and applications of genetic algorithms[M]. Beijing: Science Press, 2002.)
- [5] Walters R F C. Categories and computer science[M]. London: Cambridge University Press, 1992.
- [6] Oles F J. An application of lattice theory to knowledge representation[J]. Theoretical Computer Science, 2000, 249(1): 163-196.
- [7] Herre H, Heller B, Burek P, Hoehndorf, et al. General formal ontology(GFO): A foundational ontology integrating objects and processes[M]. Leipzig: University of Leipzig, 2006: 1-81.
- [8] Poli R. Ontological methodology[J]. Int J of Human-Computer Studies, 2002, 56(6): 639-664.
- [9] Krotzsch M, Hitzler P, Ehrig M, et al. Category theory in ontology research: Concrete gain from an abstract approach[R]. Germany: University of Karlsruhe, 2005: 1-6.
- [10] Hitzler P, Krotzsch M, Ehrig M, et al. What is ontology merging — A category-theoretical perspective using pushouts[C]. American Association for Artificial Intelligence Workshop. Menlo Park: AAAI Press, 2005: 104-108.

- [11] Colomb R M, Dampney C N G. An approach to ontology for institutional facts in the semantic web[J]. *Information and Software Technology*, 2005, 47(12): 775-783.
- [12] Lu R Q. Towards a mathematical theory of knowledge[J]. *J of Computer Science and Technology*, 2005, 20(6): 751-757.
- [13] Brazhnik O. Databases and the geometry of knowledge[J]. *Data and Knowledge Engineering*, 2007, 61(2): 207-227.
- [14] 王庆全, 荣莉莉, 于凯. 应急决策知识发现的推理方法研究[J]. *运筹与管理*, 2010, 1(7): 23-30.  
(Wang Q Q, Rong L L. A knowledge reasoning method for emergency decision-making knowledge discovery[J]. *Operation and Management*[J]. 2010, 1(7): 23-30.)
- [15] Bosworth J, Foo N, Zeigler B P. Comparison of genetic algorithms with conjugate gradient methods[R]. Michigan: The University of Michigan, 1972.
- [16] Liepins G E, Vose M D. Representational issues in genetic optimization[J]. *J of Experimental and Theoretical Artificial Intelligence*, 1990, 2(2): 101-115.
- [17] Antonisse J. A new interpretation of schema notation that overturns the binary encoding constraint[C]. *The 3rd Int Conf on Genetic Algorithms*. San Mateo: Morgan Kaufmann, 1989: 86-91.
- [18] 李航, 寇纪淞, 李敏强. 参数搜索空间规模对遗传算法稳定性的影响[J]. *系统工程学报*, 2007, 22(2): 162-169.  
(Li H, Kou J S, Li M Q. Influence of solution space scale on stability of GA[J]. *J of Systems engineering*, 2007, 22(2): 162-169.)
- [19] 王小平, 曹立明. 遗传算法-理论、应用与软件实现[M]. 西安: 西安交通大学出版社, 2002.  
(Wang X P, Cao L M. *Genetic algorithms-theory, application and software implement*[M]. Xi'an: Sian Jiaotong University Press, 2002.)

(上接第982页)

- [24] Gan X, Sethi S P, Yan H. Channel coordination with a risk-neutral supplier and a downside-risk-averse retailer[J]. *Production & Operations Management*, 2005, 14(1): 80-89.
- [25] 于春云, 赵希南, 彭艳东, 等. 具有风险规避者和偏爱者加盟的供应链优化与协调模型[J]. *系统工程*, 2007, 25(1): 13-20.  
(Yu C Y, Zhao X N, Peng Y D, et al. Supply chains optimization and coordination mechanisms mode with a risk-averse supplier(or retailer) and a risk taking retailer(or supplier)[J]. *Systems Engineering*, 2007, 25(1): 13-20.)
- [26] 王虹, 周晶. Loss-averse 零售商参与的供应链协调机制研究[J]. *统计与决策*, 2009, 279(3): 180-182.  
(Wang H, Zhou J. Supply chain coordination mechanism with a loss-averse retailer[J]. *Statistics and Decision*, 2009, 279(3): 180-182.)
- [27] 刘珩, 潘景铭, 唐小我. 基于损失厌恶型零售商的易逝品供应链价格补贴契约研究[J]. *控制与决策*, 2010, 25(8): 1149-1154.  
(Liu H, Pan J M, Tang X W. Research on perishable product supply chain markdown money contract with a loss-averse retailer[J]. *Control and Decision*, 2010, 25(8): 1149-1154.)
- [28] Shi K, Xiao T. Coordination of a supply chain with a loss-averse retailer under two types of contracts[J]. *Int J of Information and Decision Sciences*, 2008, 1(1): 5-25.
- [29] Wang C X, Webster S. Channel coordination for a supply chain with a risk-neutral manufacturer and a loss-averse retailer[J]. *Decision Sciences*, 2007, 38(3): 361-389.
- [30] 庞庆华. 供应商具有损失厌恶的收益共享契约模型研究[J]. *统计与决策*, 2009, 299(23): 44-46.  
(Pang Q H. Research on revenue-sharing contract with a loss-averse supplier[J]. *Statistics and Decision*, 2009, 299(23): 44-46.)
- [31] 林志炳, 蔡晨, 许保光. 损失厌恶下的供应链收益共享契约研究[J]. *管理科学学报*, 2010, 13(18): 33-41.  
(Lin Z B, Cai C, Xu B G. Research on revenue-sharing contract of supply chain with a loss-averse supplier and retailer[J]. *J of Management Sciences in China*, 2010, 13(18): 33-41.)