

用于复杂系统环境的遗传分类器系统*

赵宏 汪浩 邹雯

(国防科技大学系统工程与数学系 长沙 410073)

摘要 设计了一个针对复杂系统分析的遗传分类器系统,该系统不仅利用了遗传算法强大的搜索能力和便于从大量历史数据中发现规律的优点进行分类器获取,而且还具有如下两个主要特点:①它能够适用于复杂的、缺少完整描述且训练信号带有噪声的系统环境;②它是一个白箱系统,所获取分类器的置信度与统计结果相符,从而支持定性、定量相结合的综合推理模式。

关键词 遗传算法,复杂系统,分类器 推理

分类号 O224

Genetic Classifier System for Complex System Environment

Zhao Hong Wang Hao Zuo Wen

(Department of Systems Engineering, NUDT, Changsha, 410073)

Abstract In this paper, a genetic classifier system for complex system analysis is proposed. The system not only makes use of the remarkable genetic algorithms searching ability and the ability to find laws from a great amount of historical data to acquire classifiers, but also possesses the following two characteristics: ①It is suitable for such complex systems which have noise and lack complete description ②The credit of each classifier acquired by this system is identical to statistical results, so it can support meta-synthesis reasoning method.

Key words genetic algorithms, complex systems, classifier, reasoning

现代研究表明,社会经济系统(尤指证券、期货等风险投资系统)作为复杂的开放大系统,其非线性自组织特性很具体地表现在:时间序列数据具有明显的类随机、自相

* 1996年4月15日收稿

似、时间高度相关等混沌特征。运用混沌动力学的分析方法得出的分形数维、Lyapunov 指数等特征量更是可以清楚地表明这一点。因此，对于该类系统行为的分析，传统的基于稳定趋势或基于随机过程等线性模型方法明显是不适合的，而混沌动力学模型就目前的成就，多用于研究复杂系统的结构相变。对于系统行为的分析尚缺乏良策。

随着认知科学和技术的发展，人工神经网络、示例机器学习等现代人工智能技术已越来越多地运用于复杂系统分析领域。但是，人工神经网络和示例机器学习方法在应用中也有着一些局限性，主要表现在：①容错性仅体现在判别过程中，而在学习过程中，当训练样本带有噪声或有异常特例地，它们却无能为力；②它们都不具有统计特性，对于判别的置信度很难与背景知识联系在一起，即不支持定性、定量相结合的综合决策模式。鉴于上述分析，本文提出了一个用于复杂系统环境的遗传分类器系统。

1 遗传算法

遗传算法是一种以自然选择和遗传理论为基础，将生物进化过程中适者生存规则与同一群体中的个体与个体间的随机信息交换机制相结合的搜索算法。它的主要优点在于：(1) 对于搜索空间的性质和先验知识没有任何特殊要求，而能够搜索出最优解，这是其它方法无法做到的；(2) 对于搜索过程中的解变量对客观环境的适应值，是根据具体问题进行设计评价的。这样，一方面遗传算法在应用实际问题时具有充分的灵活性，同时对于所搜索到的最优解（适应值最大的个体），也易于用背景知识进行解释。

遗传算法实现过程如图 1 示出，详细原理和说明可参见文献 [1]。

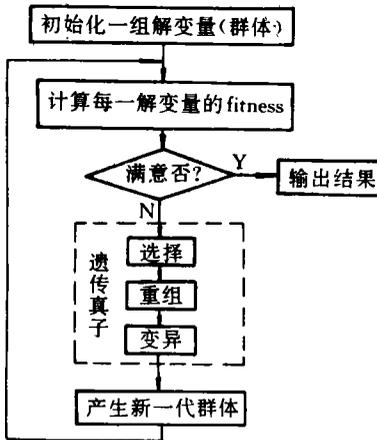


图 1 遗传算法实现过程图

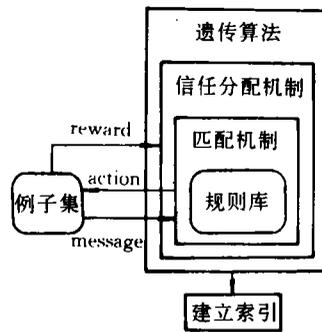


图 2 分类器组织结构图

2 遗传分类器系统

本系统的组织结构由图 2 示出，系统的目的是要从已知的例子集中，找到一些与之相容的分类器，而这些分类器将用于对未来可能发生事件的分类判别。

例子集（用 $[E]$ 表示）由一系列例子组成。每个例子都表示一次实际发生过的因果

事件，事件由前因（先兆）和结构构成。为了便于遗传操作的实现，将事件的前因部分（用〈message〉表示）和结果部分（用〈action〉表示）都编码成二进制串的形式，串的长度对所有的例子统一限定。这样，例子可形式化描述为：

$$\text{Example} ::= \langle \text{message} \rangle : \langle \text{action} \rangle$$

其中， $\langle \text{message} \rangle ::= \{0, 1\}^L$ ， $\langle \text{action} \rangle ::= \{0, 1\}^M$ 。

规则库由一系列分类器组成。分类器的形式是如下的产生式：

$$\text{IF } \langle \text{condition} \rangle \quad \text{THEN } \langle \text{action} \rangle$$

其中， $\langle \text{condition} \rangle ::= \{0, 1, \#\}^L$ ， $\#$ 是0和1的通配符， $\langle \text{action} \rangle$ 同上。

本系统的执行主要通过以下三个算法实现：

算法1（主流程）：

(1) 对 $[E]$ 进行冲突处理。将 $[E]$ 中的例子进行两两匹配，对于那种只有〈message〉部分匹配，而〈action〉部分不匹配的例子作为冲突例子，将它们全部从 $[E]$ 取出，组成新的冲突例子集 $[C]$ 。

(2) 初始化分类器群体。将 $[E]$ 中的例子作为初始分类器，直接存入规则集。

(3) 重复调度信任分配机制（算法2）。修改规则库中每一个分类器的适应值 fitness。

(4) 调度遗传算法（算法3）。生成新一代分类器群体，存入规则集。

(5) 返回（3），直至分类器群体达到满意。

(6) 生成冲突分类器。冲突分类器由 $[C]$ 中的例子直接生成，即〈condition〉部分由〈message〉给出，〈action〉部分不变，fitness由算法2得出。

(7) 给规则库中保留的分类器冲突分类器建立与之相容例子之间的索引，形成类型判别分类器。

算法2（信任分配机制）

(1) 定义分类器的 fitness 是一个二元组：

$$\text{fitness} ::= (\text{fit1}, \text{fit2}) \quad \text{其中, fit1, fit2 为非负整数。}$$

(2) 初始化分类器的适应值。

$$\text{fitness} \leftarrow 0, \text{即: fit1} \leftarrow 0, \text{fit2} \leftarrow 0$$

(3) 从 $[E]$ 中取出一个例子，将其与该分类器进行匹配，若完全匹配，则 $\text{fit1} \leftarrow \text{fit1} + 1$ ；若只有〈condition〉部分匹配，而〈action〉部分匹配失败，则 $\text{fit2} \leftarrow \text{fit2} + 1$ ；若〈condition〉部分匹配失败，则 $\text{fitness} \leftarrow \text{fitness}$ 。

(4) 重复步骤（3），直到将 $[E]$ 中的例子取完。

算法3（分类器系统的遗传算子。遗传算子的运行就是将分类器逐步净化，最终获取到一组与客观环境最相容的分类器）。

(1) 选择。

1.1) 对于规则库中的每一分类器，若其对应的 $\text{fit1} > 0$ ，且 $\text{fit2} = 0$ ，则保留；否则淘汰。

1.2) 将保留下来的分类器两两匹配。设 $R1, R2$ 为两个保留下来的分类器，若 $R1 \supseteq R2$ ，且 $\text{fit1}(R1) = \text{fit1}(R2)$ ，则保留 $R2$ ，淘汰 $R1$ ；若 $R1 \supseteq R2$ ，且 $\text{fit1}(R1) > \text{fit}(R2)$ ，则保留 $R1$ ，淘汰 $R2$ 。

(2)重组。被选中的两个分类器,随机交换彼此位串中若干比特的值,来产生新的分类器,如:

$$\begin{array}{l} R1 = a_1a_2a_3 | a_4a_5a_6 | a_7a_8a_9 \\ R2 = b_1b_2b_3 | b_4b_5b_6 | b_7b_8b_9 \end{array} \xrightarrow{\text{重组}} \begin{array}{l} R1 = a_1a_2a_3 | b_4b_5b_6 | a_7a_8a_9 \\ R2 = b_1b_2b_3 | a_4a_5a_6 | b_7b_8b_9 \end{array}$$

(3)变异。随机地改变所选中分类器位串中若干比特的值。对于〈condition〉部分将 1 变成 0 或 #, 0 变成 1 或 #, # 变成 0 或 1; 对于〈action〉部分将 1 变成 0, 0 变成 1, 从而产生新的分类器。

这里有两点需要说明:①在重组和变异算子中,分类器位串中的若干比特一般在分类器位串长度的 10% 以内随机选取;②由重组或变异算子所产生的新的分类器,可能会破坏原有的适应性,这在遗传算法中是十分正常的现象。若新分类器的适应值超过原分类器,则保留新分类器,否则淘汰。规则库中分类器群体的适应性是由选择算子来保证的。

运用本系统获取的规则集对事件的未知结果进行分类判决时,首先输入前因描述〈message〉,再将其与规则集中的每个分类器的〈condition〉部分进行匹配。若没有分类器能够与之匹配,则本系统拒绝对此事件作出判决;若匹配成功,则考察该分类器的 fitness 是否满足。一般说来,若 fit1 较 fit2 大得多,则认为满意,这时将〈action〉部分输出,可直接作为分类判决结果;否则需根据索引检索所有与该分类器相对应的例子。这时不仅要考察例子〈message〉部分的背景描述,还要综合许多〈message〉之外的各种信息进一步分析,运用基于事例的推理技术,得出定性、定量相结合的综合评判结果。

3 结束语

本文设计并实现了一个用于复杂系统环境的遗传分类器系统。该系统较同类的人工神经网络模型和示例机器学习模型在鲁棒性、判别精度和灵活性等方面有较大的改善,而且特别适合于结构复杂并带有噪声的系统环境。

参 考 文 献

- 1 Edgar E. Chaos and order in the capital markets. John Wiley & Sons Inc, 1992
- 2 孙广振,王劲松. 深圳股市混沌现象的辨别及其讨论. 数量经济技术经济研究, 1994(1)
- 3 Goldberg D E. Genetic algorithms in search, optimization & machine learning. MA: Addison wesley, 1989
- 4 Alastair D. McAulay and Jae chan Oh. Improving learning of genetic rule-based classifier systems. IEEE Trans. Syst. Man Cybern., 1994, 24(1): 152~159
- 5 Raed A Z. Neurocontrollers trained with rules extracted by a genetic assisted. IJCNN, 1995, 6(4): 859~875
- 6 Zhongzhi shi. Principles of machine learning. International academic press, 1992, 243~272
- 7 陈文伟. 决策支持系统及其开发. 北京:清华大学出版社, 1994, 152~217

(责任编辑 石少平)