

环境污染物分子结构与毒性关系专家系统*

全英

白乃彬

(北京工业大学计算机科学系, 100022) (中国科学院生态环境研究中心, 100085)

摘要 介绍了用于环境污染物结构毒性关系研究的软件, 做为软件性能的检验, 也进行了62种氮杂环芳胺及其致变活性关系的实例研究. 发现: 专家系统应用近于自然语言的规则形式易于表达特殊分子结构对毒性有明显激活或抑活效应的知识, 而此种作用是难于用数值描述的. 专家系统不失为环境污染物构效关系的良好模式之一.

关键词 专家系统, 氮杂环芳胺, 致变活性

分类号 TP319: O625.637

在以前的工作中^[1-3], 曾用专家系统方法研究了多环硝基芳烃结构与其致变活性关系. 发现专家系统应用近于自然语言的规则形式易于表达特殊分子结构对毒性有明显激活或抑活效应的知识, 而此种作用是难于用数值描述的. 本文将介绍用于环境污染物结构毒性关系研究的软件, 做为软件性能的检验, 也进行了62种氮杂环芳胺及其致变活性关系的实例研究.

1 专家系统用于 QSAR 研究的背景

环境污染物构效关系的一般表达式为:

$$A = F(S_1, S_2, \dots, S_n)$$

式中, 因变量 A 为活性描述符, 自变量 S 为结构描述符, F 为构效关系函数表达式. 由于分子结构的复杂性, 造成分子结构, 特别是对活性有潜在影响的分子结构知识表达具有复杂性和多种形式; 对于生物活性 A 和诸 S 变量间的函数形式 F 一般为多元非线性形式, 多元非线性形式知识表达也有几种可选择的模式.

1.1 A 和 S 皆可用数值表达

当每一种自变量 S 与 A 或 LOGA 皆呈线性关系, 且诸自变量间彼此线性无关, 那么 F 可简单表达为线性组合形式. 若所论系列化合物数目较少, 结构彼此变化不大时, 一般用多元线性回归方法生成构效关系模式.

如果有一个以上自变量 S 与 A 或 LOGA 存在非线性关系, 但自变量间彼此线性无关, 那么可对此类自变量做 Taylor 函数展开, 取二次项, 直至多次项近似, 于是 F 将被表达为含自变量多次项的线性组合形式, 这就是著名的 Hansch 表达式.

如果部分或全部自变量彼此有关, 且与因变量呈非线性关系, 那么 F 应表达为多元非线性

* 国家自然科学基金资助项目.

收稿日期: 1994-07-14.

性函数. 带有S形变换函数的三层反向传播神经网络系统是目前较为理想的模式^[4,5].

1.2 S可用数值表达; A用等级表达

由于生物活性数据具有较大离散性, 同一化合物不同文献来源数据彼此差异可能较大, 难以取舍. 我们认为: 与其精确不如准确, 可采用传统预防医学方法将数值转化为3个等级(阴性、可能阳性和阳性)或五个等级(-、+、++、+++、++++), 以增加其可比性. 聚类分析、判别分析和模式识别等分类方法常用来构筑QSAR模式^[6,7].

1.3 S用语言表达; A用等级表达

在环境污染物结构毒性关系实例研究中发现: 某些特殊子结构或分子片段对生物活性有明显激活作用; 而某些特殊子结构或分子片段, 或者在生成致变前兆体的分子反应部位制造空间障碍, 或者由于它的存在破坏了共轭大分子的平面性, 从而具有明显的抑活作用. 这些子结构或分子片段很难用数值表达, 也就不可能用前述的数值计算模式进行QSAR研究. 有幸的是, 以规则集为基础的专家系统方法擅长于用人类自然语言描述分子结构, 人类自然语言词汇又十分丰富, 几乎没有一种自然现象不能为其表达的, 故而它有利于描述难于数值化的分子结构描述符和预报活性等级较多的情况. 另一方面, 基于规则的专家系统应用符号处理方法很容易实现数值计算难于进行的多元非线性函数关联. 因此专家系统不失为建立环境污染物构效关系的良好模式.

研究中发现: 由于分子结构的复杂性和毒性数据的离散性, 或学习样本集合数目较少, 有时将某化合物硬性派往某类欠妥, 不如去讨论该样本归属所有类的隶属度. 基于规则的不精确推理专家系统, 分类时注意到亦此亦彼, 推理过程中算出未知样本派往所有可能类的概率, 从而得到既灵活而又反映客观实际的样本归类. 在环境科学中不乏此类分类的应用实例, 如环境质量评价等.

为了适应不同的用户课题要求, 构筑了两个独立的专家系统, 然后用一个统一的人机界面和以前已经完成的神经网络系统、模式识别系统连接起来形成一个整体的环境污染物构效关系模式.

2 精确推理和不精确推理专家系统^[8]

基于规则的精确推理专家系统中, 领域知识被表达为产生式规则, 即IF(前提)THAN(结论)表达式, 它特别适用于具有必然性的因果关系和逻辑关系知识结构. 本系统包含了正向推理和反向推理两种推理机制, 后者适用于前题条件多而可选择的结论少的情况, 运行速度快一些; 前者则相反. 系统载入规则继而运行时, 用户应不断回答系统所提示的问题, 直至给出结论为止. 反向推理仅需询问和回答部分前提, 而正向推理则要逐一询问各前提成立与否, 然后推理.

精确推理专家系统由四个功能不同的模块组成:

头文件: 定义了有关的数据类型和常量, 从而确定了知识库的描述结构.

主控模块: 定义了知识库, 负载全部领域知识; 提供用户使用的正、反推理机选择菜单; 提供知识库分支选择菜单, 根据用户选择, 依据知识库的不同分支进行推理, 由于知识可能包括多方面, 故知识库为树型结构, 各方面能独立分支, 便于推理, 缩短运行时间.

推理机：包括正向推理和反向推理两个独立平行部分。通过与用户交互式回答得到有关前提，诸如化合物的分子结构描述符，通过规则进行推理，导出结论。

数据库：为推理机的支持模块。存放推理过程用到的有关前提及推理过程中产生的中间信息。实现对数据库的种种操作：信息添加及信息是否匹配的检查等。

基于规则的不精确专家系统适用于不具有明确因果关系的知识结构。领域知识虽也表达为 IF THEN 形式，但前提不一定是肯定的。还可以多个彼此相互独立的前提支持同一结论。在某些环境问题划分等级时，诸如环境质量评价，化学品有无毒性等，等级之间往往是界限不清的，给出未知样本在所有可能类中的概率可以更好模拟人的思维。例如，对于预报氮杂环芳胺的对鼠伤寒沙门式菌的致变活性时，从分子结构来讲，要考虑端环抑活子结构、端环增活子结构、母体苯环骈接数及氮杂环上氮原子取代甲基数等四个要素，即取论域 U ，

$U = \{u_1(\text{抑活结构}), u_2(\text{增活结构}), u_3(\text{母体环数}), u_4(\text{氮甲基数})\}$ 评定毒性分为 3 个等级，即取评语集为 V

$$V = \{v_1(\text{阳性}), v_2(\text{可疑}), v_3(\text{阴性})\}$$

给出 U 和 V 中各要素的权值(统计学中的概率)，既而根据这些权值进行组合运算，导出各个结论成立的概率，而不是象精确推理那样仅给出各个结论中的某一个结论。

不精确推理专家系统由 3 个功能不同的模块组成：

主控模块：提供主要菜单和 2 级菜单，用户用以控制程序流。

知识库模块：它实现主菜单中 5 种功能用两种用户可选择的方法建立知识库，即建立结论组及出现概率；建立前提组；建立结论与前提间的关系(同时给出前提分类数)；将知识存入知识库；或用已有知识文件建立知识库。

推理模块：要求用户先把前提分类，例如本文所选氮杂环芳胺案例研究中，前提可分为抑活结构数、增活结构数、母体环数及氮原子上甲基数四类，依其特性每一类又可分为几个等级。在系统运行时，根据提示用户依次输入各类的某一个前提(顺序无关紧要)，系统便会给出样本归属于各个结论的概率。

C 语言是一种过程式语言，采用 C 语言设计的程序有代码短、质符高和运行速度快的优点。虽然在人工智能程序设计中不如 LISP 和 PROLOG 那样简捷，但是它能充分发挥机器的硬件特性，系统要和涉及数值计算的神经网络系统及涉及图形操作的人机界面系统组装为一个统一的人工智能系统，应用 C 语言就比 LISP 和 PROLOG 要方便得多。本专家系统应用 TURBO C 语言自行编程，系统源文件含 C 语句约 1600 条；在 PC/AT 286 计算机上实现，系统执行文件大小 110KB。

3 62 种杂环芳胺结构与致变活性关系专家系统案例研究

近年发现，烹调肉、鱼类食品过程中可产生百余种氮杂环芳胺，经鉴定：有相当部分显示较强致变活性，且为致癌剂。对该类化合物进行构效研究有助于对目前由于含量甚微而尚未测定出的多种杂环芳胺的致变性能作出科学预报。

将文献 [9 ~ 12] 所载 62 种氮杂环芳胺经 S_9 活化的 TA98 和 TA1538 菌株变活性进行

分级：若活性小于 1 Rev/nmol，则该化合物为非致变物，(-)；若大于 2.4 Rev/nmol，则该化合物为致变物，(+); 若在 1~2.4 之间，则为可疑致变物，(+/-)。选择此标准的医学依据是：对于某化合物，测试菌种的自然回变率为至少为阴性对照物自然回变率两倍以上，才能较为充分断定其为致变物。

根据专家经验和多次结构描述符选择尝试，发现端环所含以下分子碎片为增活结构： $\text{NH}_2-\text{C}=\text{C}-\text{C}=\text{C}$ 或 $\text{NH}_2-\text{C}=\text{N}-\text{C}=\text{C}$ ；而端环所含以下分子碎片为抑活结构： $\text{NH}_2-\text{C}=\text{C}-\text{C}=\text{N}-\text{C}$ 或 $\text{NH}_2-\text{C}-\text{C}-\text{N}=\text{C}$ 或 $\text{NH}_2-\text{C}-\text{NH}-\text{C}=\text{C}$ 或 $\text{NH}_2-\text{C}=\text{C}-\text{C}=\text{N}-\text{C}-\text{N}-\text{C}$ (C. 为两个芳环并接处碳原子)。显然，在某一化合物中是否存在这些分子碎片，用人类自然语言一下子就能说清，可在以符号处理为特点的专家系统框架内进行分类；但很难用传统方法将这句话数值化，继而建立数值模式通过数值计算而进行分类。此外，芳环氮原子取代甲基数和母体共轭环数的增加也有次要增活作用。62 种氮杂环芳胺结构与致变活性描述符皆列于附表。

附表 杂环芳胺分子结构与对鼠伤寒沙门氏菌致变活性关系专家系统分类结果

No.	化合物	分子结构编码				活性编码								
		氮 基 数	母 体 环 数	增 活 结 构	抑 活 结 构	TA98 或 TA1538 (Rev / nmol)	实 验 等 级	预 报 等 级	预报等级					
								+	+/-	0				
1	2-氨基-9H-吡啶并[2,3-b]吡啶	0	1	2	0	>24	+	+	1.0	0.0	0.0			
2	2-氨基-3-甲基-9H-吡啶并[2,3-b]吡啶	0	1	2	0	>24	+	+	1.0	0.0	0.0			
3	3-氨基-4-甲基-5H-吡啶并[4,3-b]吡啶	0	1	2	0	>24	+	+	1.0	0.0	0.0			
4	3-氨基-5H-吡啶并[4,3-b]吡啶	0	1	2	0	>24	+	+	1.0	0.0	0.0			
5	3-氨基-1-甲基-5H-吡啶并[4,3-b]吡啶	0	1	2	0	>24	+	+	1.0	0.0	0.0			
6	3-氨基-1,4-二甲基-5H-吡啶并[4,3-b]吡啶	0	1	2	0	>24	+	+	1.0	0.0	0.0			
7	3-氨基-9H-吡啶并[3,4-b]吡啶	0	1	1	1	<1	-	-	0.14	0.14	0.71			
8	3-氨基-1-甲基-9H-吡啶并[3,4-b]吡啶	0	1	1	1	<1	-	-	0.14	0.14	0.71			
9	2-氨基-3H-苯并[1,2-d]咪唑	0	1	1	1	<1	-	-	0.14	0.14	0.71			
10	2-氨基-1-甲基苯并[1,2-d]咪唑	1	1	0	1	<1	-	-	0.0	1.0	0.0			
11	2-氨基-5-甲基-3H-苯并[1,2-d]咪唑	0	1	1	1	<1	-	-	0.14	0.14	0.71			
12	2-氨基-1,5-二甲苯并[1,2-d]咪唑	1	1	1	0	>24	+	+	0.83	0.17	0.0			
13	2-氨基-3H-5,6-二甲苯并[2,3-d]咪唑	0	1	1	1	<1	-	-	0.14	0.14	0.71			
14	2-氨基-1,5,6-三甲基苯并[2,3-d]咪唑	1	1	1	0	>24	+	+	0.83	0.17	0.0			
15	2-氨基-3H-萘并[3,2-d]咪唑	0	2	1	1	>24	+	+	0.91	0.09	0.0			
16	2-氨基-1-甲基-萘并[3,2-d]咪唑	1	2	1	0	>24	+	-	1.0	0.0	0.0			
17	2-氨基-3H-萘并[2,1-d]咪唑	0	2	0	1	<1	-	-	0.0	0.0	1.0			
18	2-氨基-1-甲基萘并[2,1-d]咪唑	1	2	1	0	>24	+	+	1.0	0.0	0.0			
19	2-氨基-3-甲基萘并[2,1-d]咪唑	1	2	0	0	>24	+	+	1.0	0.0	0.0			
20	2-氨基-二吡啶并[1,2-a:3',2'-d]咪唑	0	2	1	1	>24	+	+	0.91	0.09	0.0			
21	2-氨基-3-甲基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	1	1	1-2.4	+/-	+	0.91	0.09	0.0			
22	2-氨基-4-甲基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	1	1	>24	+	+	0.91	0.09	0.0			
23	2-氨基-6-甲基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	1	1	>24	+	+	0.91	0.09	0.0			
24	2-氨基-7-甲基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	1	1	>24	+	+	0.91	0.09	0.0			

续附表 杂环芳胺分子结构与对鼠伤寒沙门氏菌致变活性关系专家系统分类结果

No.	化合物	分子结构编码			活性编码						
		氮 甲 基 数	母 体 环 数	增 活 结 构	抑 活 结 构	TA98或 TA1538 (Rev/ n mol)	实 验 等 级	预 报 等 级	预报等级		
									+	+/-	0
25	2-氨基-8-甲基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	1	1	>24	+	+	0.91	0.09	0.0
26	2-氨基-9-甲基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	1	1	>24	+	+	0.91	0.09	0.0
27	3-氨基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	2	0	>24	+	+	1.0	0.0	0.0
28	3-氨基-4-甲基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	2	0	>24	+	+	1.0	0.0	0.0
29	3-氨基-4,6-二甲基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	2	0	>24	+	+	1.0	0.0	0.0
30	3-氨基-4,7-二甲基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	2	0	>24	+	+	1.0	0.0	0.0
31	4-氨基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	1	2	<1	-	-	0.0	0.25	0.75
32	4-氨基-2-甲基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	1	2	<1	-	-	0.0	0.25	0.75
33	4-氨基-7-甲基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	1	2	1-24	+/-	-	0.0	0.25	0.75
34	4-氨基-2,7-二甲基二吡啶并[1,2-a:3',2'-d]咪唑	0	2	1	2	<1	-	-	0.0	0.25	0.75
35	2-氨基-3H-咪唑并[4,5-b]吡啶	0	1	1	1	1-24	+/-	-	0.14	0.14	0.71
36	2-氨基-3-甲基咪唑并[4,5-b]吡啶	1	1	1	1	1-24	+/-	+/-	0.0	1.0	0.0
37	2-氨基-1,6-二甲基咪唑并[4,5-b]吡啶	1	1	1	0	1-24	+/-	+	0.83	0.17	0.0
38	2-氨基-3H-5,7-二甲基咪唑并[4,5-b]吡啶	0	1	1	1	>24	+	-	0.14	0.14	0.71
39	2-氨基-3,5,7-三甲基咪唑并[4,5-b]吡啶	1	1	1	0	>24	+	+	0.83	0.17	0.0
40	2-氨基-5-苯基吡啶	0	1	2	0	>24	+	+	1.0	0.0	0.0
41	2-氨基-1-甲基-6-苯基咪唑并[4,5-b]吡啶	1	1	1	0	>24	+	+	0.83	0.17	0.0
42	2-氨基-3-甲基-6-苯基咪唑并[4,5-b]吡啶	1	1	1	0	>24	+	+	0.83	0.17	0.0
43	2-氨基-3H-咪唑并[4,5-f]喹啉	0	2	1	1	>24	+	+	0.91	0.09	0.0
44	2-氨基-1-甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
45	2-氨基-3-甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
46	2-氨基-3H-4-甲基咪唑并[4,5-f]喹啉	0	2	1	1	>24	+	+	0.91	0.09	0.0
47	2-氨基-3H-5-甲基咪唑并[4,5-f]喹啉	0	2	1	1	>24	+	+	0.91	0.09	0.0
48	2-氨基-1,4-二甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
49	2-氨基-1,5-二甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
50	2-氨基-3,4-二甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
51	2-氨基-3,5-二甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
52	2-氨基-3-甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
53	2-氨基-3,4-甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
54	2-氨基-3,7-二甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
55	2-氨基-3,8-二甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
56	2-氨基-3,4,7-三甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
57	2-氨基-3,4,8-三甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
58	2-氨基-3,5,7-三甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
59	2-氨基-3,5,8-三甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
60	2-氨基-3,7,8-三甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
61	2-氨基-3,4,7,8-四甲基咪唑并[4,5-f]喹啉	1	2	1	0	>24	+	+	1.0	0.0	0.0
62	3-氨基三唑	0	0	1	1	<1	-	-	0.0	1.0	0.0

根据图1所示推理网络, 按抑活结构、增活结构、母体环数和氮原子甲基数四个结构描述符层次写出七条产生式规则, 载入知识库. 启动专家系统后, 要求用户一一回答欲进行预报活性分类的相应结构描述符的值, 最终专家系统给出该化合物的归类. 62种氮杂环芳胺所讨论的致变活性专家系统预报结果亦列入附表. 结果表明: 62个样本错分类仅5个, 正确分类率达到0.92, 满足统计完好标准.

应用不精确推理专家系统, 对62个样本致变活性分类概率结果亦列于附表. 由附表可知正确分类率与精确推理专家系统结果类似, 但具有分类灵活的特点. 以上表明: 用附表所示四项分子结构指标可以进行未知化合物是否为致变物的先期预报; 专家系统在表达难以用数值表达的特殊子结构的激活或抑活作用, 及表达复杂的、非线性的环境污染物致变活性行为等两方面, 可成为良好QSAR模式.

一般认为^[12], 在氮杂环芳胺致变历程中氨基起重作用. 在生物体内芳胺代谢氧化为羟胺, 继而脱水形成活性很大的仲胺正离子, 它能作用于细胞的DNA, 成为致变作用的前兆物. 据此, 仲胺正离子稳定性应是影响致变活性大小的关键. 有趣的是, 对于增活分子结构: $\text{NH}_2-\text{C}=\text{C}-\text{C}=\text{C}$ 或 $\text{NH}_2-\text{C}=\text{N}-\text{C}=\text{C}$ 具有共轭烯烃结构, 由于 π 电子的共轭效应, 有助于仲胺正离子的稳定性; 对于抑活分子结构: $\text{NH}_2-\text{C}=\text{N}-\text{C}-\text{NH}-$ 或 $\text{NH}_2-\text{C}-\text{C}-\text{N}-$ 或 $\text{NH}_2-\text{C}-\text{NH}-\text{C}=\text{C}=\text{N}-\text{C}-\text{N}-$ (C. 为两个芳环并接处碳原子) 则不具有共轭烯烃结构或仅具有短共轭烯烃结构, 由于 π 电子共轭效应的中断, 而无助于仲胺正离子的稳定. 杂环芳胺的母体部分一般为共轭芳环结构, 环数越多, π 电子共轭效应越大, 应越有助于仲胺正离子的稳定. 甲基具有推电子作用, 可能稍利于仲胺正离子的稳定. 以上讨论可做为本文所建立的杂环芳胺构效关系的定性解释.

4 结论

专家系统应用近于自然语言的规则形式易于表达特殊分子结构对毒性有明显激活或抑活效应的知识, 而此种作用是难于用数值描述的. 专家系统不失为环境污染物构效关系的良好模式之一.

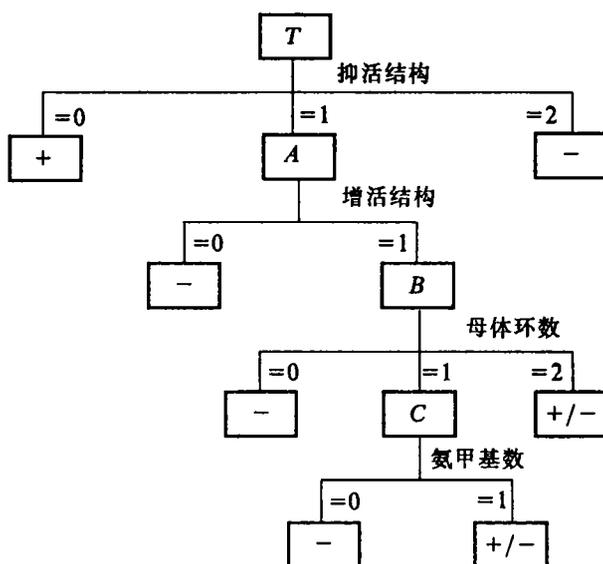


图1 推理网络示意图

参 考 文 献

- 1 白乃彬, 杜敏. 有毒有机物环境行为和生态毒理论文集. 北京: 中国科学技术出版社, 1990, 252 ~ 259
- 2 Bai Naibin, Du Min. *Journal of Environmental Sciences(China)*. 1992, 4(4): 106 ~ 112
- 3 Bai Naibin, Du Min. *Toxicological and Environmental Chemistry*. 1994, 38: 1 ~ 7
- 4 包约翰. 自适应模式识别与神经网络. 北京: 科学出版社, 1992, 117 ~ 125
- 5 Blank T B. *Analytical Chemistry*. 1993, 65(21): 3 081 ~ 3 089
- 6 白乃彬, 王忠诚. 硝基多环芳烃结构对鼠伤寒沙门氏菌致变性能关系研究. *环境化学*, 1989, 8(6): 15
- 7 白乃彬. 80种硝基多环芳烃结构与其对鼠伤寒沙门氏菌致变性能关系研究. *环境化学*, 1991, 10(1): 52
- 8 Castillo E, Alvarez E. *Expert systems: uncertainty and learning*. New Yonk: Elsevier applied science, 34 ~ 96
- 9 Nagao M. *Mutagenicity testing in environmental pollution control*. Chichester, Ellis Horwood, 69 ~ 85
- 10 Hatch F T. *Environmental and molecular mutagenesis*. 1991, 17: 4 ~ 19
- 11 Smith C. *Mutation research*. 1992, 279: 61 ~ 73
- 12 Ying Pingzhang, Klopman G, Rosenkrans H S. *Enviornmental and molecular mutagenesis*. 1993, 21: 100 ~ 115

A Study of the QSAR (Quantitative Structural Activity Relationship) on Enviromental Pollution Using an Expert System

Quan Ying

(Department of Computer Science, Beijing Polytechnic University, 100022)

Bai Naibin

(Research Center for Eco-environmental Sciences, Chinese Academy of Sciences, 100085)

Abstract In this paper, an expert software system for the study of the QSAR on environmental pollution is introduced. The case study of the QSAR of 62 heterocyclic amines has been carried out. To test the functionality of the software, the results of the study show that, using logical rules, the expert system is able to determine the enhanced or inhibited effects of particular molecular structures; however, those effects are very difficult to determine as numerical values. Thus, in some cases, the expert system is a better QSAR model than the traditional numerical value models.

Keywords expert system, heterocyclic amines, QSAR (Quantitative Structural Activity Relationship)