

工程与应用

用于专家系统规则库的冗余校验方法研究

安 莉, 王建林

AN Li, WANG Jian-lin

北京化工大学 信息科学与技术学院, 北京 100029

School of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

E-mail: wangjl@mail.buct.edu.cn

AN Li, WANG Jian-lin. Research of detecting redundancy method for rule-based in expert system. *Computer Engineering and Applications*, 2008, 44(34): 191-193.

Abstract: Production rules is a knowledge representation applied in numerous fields presently. In expert system for confirming structure of biomass soft-sensor hybrid model in fermentation process, when adding new rule into rule-base, exist of redundancy affects reasoning efficiency and reasoning veracity. A detecting redundancy method for rule-based in the expert system is introduced in this paper. Distinguish of redundancy, disposal of redundancy and realization of redundancy verification are presented. Experiment shows that, the detecting redundancy method can judge whether rule added is redundant according to inputting conditions and existent rules, and can debase complexity of model when avoiding effect of redundancy to reasoning efficiency, which favors to optimize structure of hybrid model.

Key words: production rules; redundancy verification; soft-sensor models; complexity of model

摘 要: 产生式规则是目前应用较多的一种知识表示方法。在用于确定发酵过程生物量软测量混合模型结构的专家系统中, 当向产生式规则知识库添加新的规则时, 冗余的存在会影响推理的效率以及推理的准确性。提出了一种用于该专家系统规则库的冗余校验方法, 给出了冗余规则的判别、冗余规则的处理以及冗余校验的实现方法。实验结果表明, 该冗余校验方法可以根据输入条件和已有规则, 判断出新添加的规则是否冗余, 并在消除冗余对推理效率影响的同时, 降低模型复杂度, 有利于优化混合模型的结构。

关键词: 产生式规则; 冗余校验; 软测量模型; 模型复杂度

DOI: 10.3778/j.issn.1002-8331.2008.34.059 文章编号: 1002-8331(2008)34-0191-03 文献标识码: A 中图分类号: TP391

1 前言

产生式知识表示又称为产生式规则表示法, 是目前应用较多的一种知识表示方法, 它用接近于人类思维特点的形式获取和表达知识, 直观且便于推理。但是, 产生式规则之间相互作用的限制可能导致降低推理效率, 需要解决一致性和完整性的问题。近年来, 国内外学者对上述问题进行了一系列的研究^[1-3], J G Schmolze 等^[4]在文中给出了一种在产生式规则中探测冗余的算法, 但是其算法复杂不适合规则较多的大系统, 而且也不能解决规则间冲突, 即一致性的问题; 薛冬娟等^[5]在文中鉴于知识规则的代表量非常大的特点, 提出了一种称为分类产生式规则的知识表示方法, 有效地解决了知识组合爆炸和搜索匹配代价高的问题, 但不具备自学习功能, 不能添加新知识; 孙伟等^[6]在文中提出了一种基于有向超图的规则库表达方法, 并在此基础上给出了基于可达矩阵的冗余和环路检测算法, 然而适用于确定发酵过程生物量软测量混合模型结构的专家系统知识库

的冗余校验方法还有待于进一步研究。

针对上述问题, 在用于确定发酵过程生物量软测量混合模型结构的专家系统中, 为了避免在向产生式规则知识库添加新知识时出现冗余, 影响推理效率以及推理的准确性, 本文提出了一种用于该专家系统规则库的冗余校验方法。

2 冗余校验的方法

2.1 冗余规则的判别

用于确定发酵过程生物量软测量混合模型结构的专家系统的知识库主要以模型类知识为主, 采用产生式规则表达, 称为“模型规则知识库”。一条规则即为一个机理模型, 规则的前提条件为变量或参数的集合, 如式(1)所示:

$$V_k = \{r_1, r_2, r_3, \dots, r_k\} \quad (1)$$

规则的结论部分为机理模型表达式, 如式(2)所示:

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.20476007, No.20676013)。

作者简介: 安莉(1981-), 女, 博士生, 主要研究领域为微生物发酵过程智能检测; 王建林(1965-), 通讯作者, 男, 工学博士, 教授, 博士生导师, 主要研究领域为复杂过程智能检测与信息处理。

收稿日期: 2008-05-22 修回日期: 2008-08-05

$$f_k[r_1, r_2, r_3, \dots, r_{i_k}] = 0 \quad (2)$$

式中, i_k 为第 k 条规则中的变量个数。

发酵过程生物量软测量混合模型结构的确定过程就是通过对已知事实的分析, 获得那些是待估变量或参数的信息, 进而使用模型类规则进行推理, 获得与待估变量或参数相关的机理模型, 最后根据可用机理模型的组合确定混合模型结构。

冗余规则的存在会影响模型类规则的推理效率, 甚至影响推理结果的准确性, 所以在添加新的规则时, 需要首先对其进行冗余校验, 判别其是否冗余。

当机理模型个数=变量总个数时, 说明机理模型的组合有解; 当机理模型个数<变量总个数时, 说明存在需要辨识的未知变量; 当机理模型个数>变量总个数时, 说明存在规则冗余。

2.2 冗余规则的处理

在用于确定发酵过程生物量软测量混合模型结构的专家系统的模型规则知识库中, 对于冗余规则的处理, 分为放弃添加和替换两种操作。

为了优化专家系统推理得出的混合模型结构, 定义了模型复杂度 C , 其表达式为:

$$C = R + S \quad (3)$$

式中, R, S 分别为混合模型中未知变量个数与机理模型个数。这里未知变量为找不到可用的机理模型需要建立辨识模型的部分。

为了优化混合模型结构, 降低混合模型的复杂度, 需要在删除冗余规则的同时尽量降低模型规则的复杂程度, 因此还需要比较所有相关规则(包括待添加规则)中各个相关模型的变量个数 $i_k, k=1, 2, \dots, p+1$, 若 $i_{p+1} > i_k, k=1, 2, \dots, p$, 即待添加规则的变量个数最大, 放弃添加; 否则添加新规则, 并删除原规则库中变量个数最大的模型, 即执行替换操作。

2.3 冗余校验的实现

在向模型类规则库添加新的规则时, 首先需要根据待添加规则前提条件中的变量在模型规则库中找到与之相关的所有规则, 定义规则的个数(即模型个数)为 p 。

定义所有相关规则(包括待添加规则)中的变量总个数为 m , 其计算表达式如式(4)所示:

$$m = \sum_{k=1}^{p+1} i_k - \sum_{k=1}^p n_k \quad (4)$$

式中, n_k 为第 k 条规则中的变量与待添加规则中的变量存在重复的个数。

定义所有相关机理模型总个数为 j , 即找到的所有规则(包括待添加规则)总数, 如式(5)所示:

$$j = p + 1 \quad (5)$$

然后比较相关规则中的变量总个数 m 和相关机理模型总个数 j , 如式(6)所示:

$$\begin{cases} m \geq j & \text{没有冗余} \\ m < j & \text{存在冗余} \end{cases} \quad (6)$$

若 m 大于等于 j , 说明没有冗余, 可以添加新规则; 若 m 小于 j , 说明添加新规则后会出现冗余规则, 此时还需要比较所有相关规则(包括待添加规则)中各个相关模型的变量个数 $i_k, k=1, 2, \dots, p+1$, 若 $i_{p+1} > i_k, k=1, 2, \dots, p$, 即待添加规则的变量个数最大, 放弃添加; 否则添加新规则, 并删除原规则库中变量个数最大的模型, 即执行替换操作。

冗余校验的流程如图 1 所示。

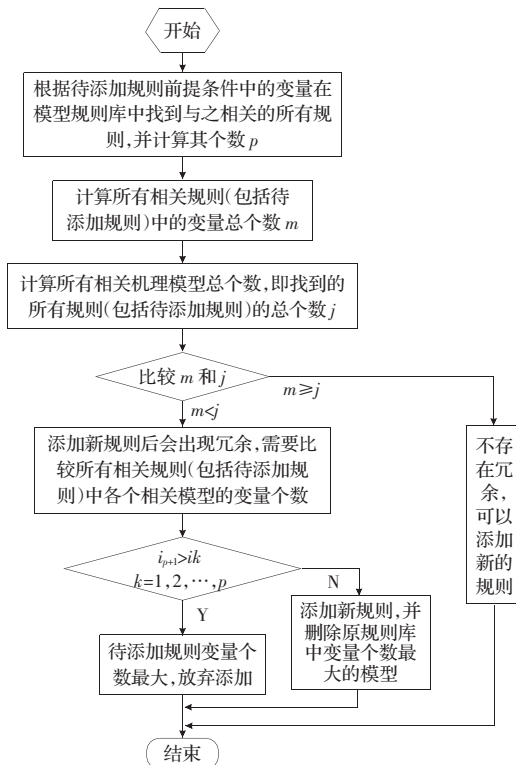


图 1 冗余校验流程

3 实验与讨论

为了对系统进行测试, 从用于确定发酵过程生物量软测量混合模型结构的专家系统知识库的模型类规则库中, 小规模地选取了与待估变量有关的几条规则进行实验。

待添加规则:

规则的前提条件: 变量为 $Sv1, Sv2, Sv4$; 参数为 $Sp1$; 常数型参数为 $Sp2$ 。

规则的结论: 模型编号为 $M6$, 模型中变量的个数: 4 个。

规则库中已有的相关规则如表 1 所示。

表 1 规则库中已有的相关规则

规则条件			规则结论
变量	参数	常数型参数	模型号
$Sv1, Sv2$	$Sp1$	$Sp2$	$M1$
$Sv1, Sv2, Sv3$		$Sp2$	$M2$
$Sv1, Sv3$	$Sp1$	$Sp2$	$M3$
$Sv1, Sv3, Sv4$		$Sp2$	$M4$
$Sv4$	$Sp1$	$Sp2$	$M5$

将上述规则表示成数据库表, 规则条件表、规则结论表和规则表分别如表 2~表 4 所示。

表 2 规则条件表

Rule_Cond ID	Condition	PmTrust
Cond1	$Sv1$	
Cond2	$Sv2$	
Cond3	$Sv3$	
Cond4	$Sv4$	
Cond5	$Sp1$	0
Cond6	$Sp2$	1

表 3 规则结论表

Rule_Conclu ID	$M(i)$
Conclu1	$M1$
Conclu2	$M2$
Conclu3	$M3$
Conclu4	$M4$
Conclu5	$M5$
Conclu6	$M6$

其中, PmTrust 表示参数可靠性, 参数的可靠性定义为该参数是否可辨识, 取值“0”该参数进化为变量, 取值“1”该参数为



图2 规则添加界面



图3 规则添加界面



图4 操作结果界面

表4 规则表

Rule ID	Rule_Cond ID	Rule_Conclu ID
Rule1	Cond1	Conclu1
Rule2	Cond2	Conclu1
Rule3	Cond5	Conclu1
Rule5	Cond1	Conclu2
Rule6	Cond2	Conclu2
Rule7	Cond3	Conclu2
Rule8	Cond1	Conclu3
Rule9	Cond3	Conclu3
Rule10	Cond5	Conclu3
Rule11	Cond1	Conclu4
Rule12	Cond3	Conclu4
Rule13	Cond4	Conclu4
Rule14	Cond4	Conclu5
Rule15	Cond5	Conclu5

常数,推理时不考虑其影响。

实验时,首先将待添加规则的前提条件:变量为 $Sv1, Sv2, Sv4$;参数为 $Sp1$;常数型参数为 $Sp2$ (不考虑),输入规则添加界面,每输入一个条件,根据规则表中多条件对应多结论的设计,会给出与其相关的规则结论,如图2所示。

进而利用已经找到的规则结论,在规则表中寻找与其相关的规则条件如图3所示。

由图2和图3,计算可得变量总数 $m=5$,模型总数 $j=6$,比较 m 和 j ,得到 $m < j$,说明添加新规则后会出现冗余,为了降低模型的复杂度,需要再对待添加规则条件中的变量总个数 $i_6=4$ 与找出的各个相关模型中的变量个数 ($i_k, k=1, 2, \dots, 5$) 进行比较,如上图所示, $\max\{i_k, k=1, 2, \dots, 5\}=3$;得出 $i_6 > i_k, k=1, 2, \dots, 5$,不需替换,放弃添加,结束操作。此次操作结果如图4所示。

所提出的冗余校验方法能够较为快速地检测出规则库中

存在的冗余,与文献中所示的方法相比,该算法简单,计算量小。因此,该算法具有较高的计算速度,从而进一步提高了检测效率。

4 结束语

所提出的冗余校验方法,可以根据输入条件和已有规则,判断出新添加的规则是否冗余,并对其执行不同的操作,便于产生式规则知识库的管理,消除冗余对推理效率的影响,降低模型复杂度,有利于优化混合模型的结构,对发酵过程生物量软测量混合模型的建立具有重要意义。

参考文献:

- [1] Revault N. Model transformation based on production rules[J]. Electronic Notes in Theoretical Computer Science, 2003, 72(4): 1-14.
- [2] 陈明亮,李怀祖,施太和,等.分类产生式规则[J].计算机应用研究, 1999, 2: 9-12.
- [3] 邓超,郭茂祖,王亚东.一种基于产生式规则的不确定推理模板模型的研究[J].计算机工程与应用, 2003, 39(30): 57-61.
- [4] 刘承洋,黄志军,徐红贤.基于产生式规则知识系统的设计与实现[J].计算机与数字工程, 2000, 28(6): 30-32.
- [5] 叶亚齐,许梦国.面向对象方法构造基于规则专家系统[J].信息技术, 2004, 28(8): 69-71.
- [6] Schmolze J G, Snyder W. Detecting redundancy among production rules using term rewrite semantics[J]. Knowledge-based System, 1999, 12: 3-11.
- [7] 薛冬娟,张冬冬,张彦峰,等.农业专家系统中分类产生式规则的知识表示方法[J].河北农业大学学报, 2004, 27(3): 104-108.
- [8] 孙伟,郭莉,高天一,等.一种基于有向超图的规则库冗余及环路检测算法[J].大连理工大学学报, 2008, 48(1): 74-78.
- [6] 张晓玲,张培强,沈兰荪.基于信息失真测度的VQ及在高光谱图像无损压缩中的应用[J].遥感学报, 2004, 8(5): 414-418.
- [7] 郑巍.基于矢量量化的图像压缩技术研究[D].合肥:合肥工业大学, 2003-11.
- [8] Wagner M. Video coding with adaptive vector quantization and rate distortion optimization[EB/OL]. (2000). UML: <http://www.inf.uni-konstanz.de/cgip/bib/files/WaSa03.pdf>.
- [9] McVeigh J S. Efficient compression of arbitrary multi-view video signals[D]. Pittsburgh PA: Department of Electrical and Computer Engineering, Carnegie Mellon University, 1996-06.
- [10] 木梅春.图像压缩中矢量量化的技术研究[D].合肥:合肥工业大学, 2005-05.

(上接 184 页)

参考文献:

- [1] 张良培,张立福.高光谱遥感[M].武汉:武汉大学出版社, 2005.
- [2] 苏令华,李纲,衣同胜,等.一种稳健的高光谱图像压缩方法[J].光学精密工程, 2007, 15(10): 1609-1615.
- [3] 闫敬文,屈小波,陈嘉臻.分组 Karhun-Loeve 变换/整数小波变换高光谱影像准无损压缩新方法[J].光学学报, 2007(10).
- [4] 路岩,张晔,陈雨时.基于分类和压缩比联合测度的高光谱图像压缩方法研究[J].光电技术应用, 2007(2).
- [5] 王晋,张晓玲,柴焱,等.一种基于自适应预测的高光谱图像近无损压缩方法[J].计算机应用研究, 2007, 24(5): 305-307.