

文章编号: 1000-582X(2001)06-0064-03

# 基于粗集理论的蔬菜栽培病虫害预警技术<sup>\*</sup>

易正俊<sup>1</sup>, 黄翰敏<sup>2</sup>, 黄席越<sup>2</sup>

(1. 重庆大学 数理学院, 重庆 400044; 2. 重庆大学 自动化学院, 重庆 400044)

**摘要:** 笔者将粗集理论应用在蔬菜栽培的病虫害预警中, 构造出一类最简单的知识提取规则, 进而获得最快的信息融合算法, 为蔬菜栽培病虫害防治提供了一种优化预警技术。这对栽培知识的传播, 绿色蔬菜的生产、丰产丰收、稳定市场具有重要意义。

**关键词:** 粗集; 信息融合; 蔬菜栽培; 预警技术

**中图分类号:** TP 387

**文献标识码:** A

近年来, 为满足城乡蔬菜周年均衡供应, 以塑料大棚和温室为主的蔬菜生产基地纷纷建成, 进一步促进了蔬菜生产。由于蔬菜集中生产和保护地栽培, 生态环境有利于病虫害的发生, 给生产带来了巨大损失。如 1989 年全国主要蔬菜病虫害发生面积高达 7001.62 万亩, 占蔬菜播种面积的 74.21%, 经防治后挽回损失 942.54 万吨<sup>[1]</sup>。如果待植株生长过程已呈现出病虫害症状, 此时只能施用农药进行治疗。但农药可能在蔬菜体内残留, 被人吸入体内, 这是人们最不愿得到的结果。因此病虫害的预警就显得非常必要。目前病虫害预警大多根据病虫害发病的条件, 利用专家的经验知识, 建立一种条件特征库, 然后将获得的信息如种子处理方法, 栽种季节, 气温, 品种等特征与特征库进行匹配并得出预测结果和报警, 然后根据结果进行预防。但包含了一些冗余的条件属性及重复信息, 影响了融合速度和知识的传播, 且对于不完整和不精确的信息难于融合。为克服这些缺点, 笔者在本文中把粗集理论用于在蔬菜栽培的病虫害预警中, 形成最简单的知识提取规则, 为病虫害预警提供了一种快速的融合算法。

## 1 粗集理论介绍

粗集理论是由波兰数学家 Z. Pawlak 在 80 年代初提出的<sup>[2,3]</sup>, 是一种处理模糊和不精确性问题的新型数学工具。他针对边界线区域思想提出了粗集, 并把无法确认的个体都归属于边界线区域, 而这种边界线区域被定义为上近似集和下近似集的差集。粗集理论在处理有限元集合时, 通过对大量数据进行分析, 根据论域中的两个等价关系的依赖关系, 剔除相容信息, 抽取

潜在有价值的规则知识。其中, 简化和核是两个重要的概念<sup>[4]</sup>。下面给出这些基本概念的定义:

### 1.1 上近似集, 下近似集、正域的概念

定义 1: 设  $U$  是论域,  $R$  是  $U$  上不分明的关系,  $X \subset U$ , 则  $X$  的上近似集  $R_+(X)$  和下近似集  $R_-(X)$  为:

$$R_+(X) = \{x \in U, R(x) \cap X \neq \Phi\}$$

$$R_-(X) = \{x \in U, R(x) \subseteq X\}$$

其中  $\Phi$  为空集,  $X$  的边界  $BN_R(X)$  定义为:

$$BN_R(X) = R_+(X) - R_-(X)$$

$R_+(X)$  是对于知识  $R$ ,  $U$  中所有一定能归入  $X$  的元素集合,  $R_-(X)$  是对于知识  $R$ ,  $U$  中可能归入  $X$  的元素集合,  $B_{NR}(X)$  是对于知识  $R$  既不能归入  $X$  也不能归入  $\bar{X}$  的元素的集合。把  $pos_R(X) = R_+(X)$  定义为  $X$  的  $R$  正域。

### 1.2 核与简化概念

简化: 令  $R$  为一等价关系, 且  $r \in R$ , 当  $ind(R) = ind(R - \{r\})$  则称  $R$  可简化为  $R - \{r\}$  ( $ind(\cdot)$  表示不分明关系), 如果它不可以进一步简化, 则称它为  $R$  的一个简化, 记为  $red(R)$

核:  $R$  的所有简化的交集称为  $R$  的核, 记为  $core(R)$ , 即:

$$core(R) = \bigcap red(R)$$

核包含在所有的简化簇中, 核是不能消除的知识特征部分的集合。

令  $R$  和  $Q$  为  $U$  中的等价关系,  $Q$  的  $R$  正域为:

$$Pos_R(Q) = \bigcup R_+(X)$$

$$X \in Q$$

\* 收稿日期: 2001-05-16

基金项目: 本项目获得 863 资助(863-306-ZD05-01-A)

作者简介: 易正俊(1963-), 男, 重庆忠县人, 副教授、重庆大学博士生。主研方向: 信息融合。

$R$  和  $Q$  的依赖关系定义为：

$$r_R(Q) = \text{card}(\text{pos}_R(Q)) / \text{card}(U)$$

其中  $\text{card}(B)$  表示集合  $B$  的基数, 显然  $0 \leq r_R(Q) \leq 1$ , 利用  $R$  和  $Q$  的依赖关系  $r_R(Q)$ , 可以判定  $R$  和  $Q$  等价类的相容性, 当  $r_R(Q) = 1$  时, 表示  $R$  和  $Q$  是相容的, 而  $r_R(Q) \neq 1$  时,  $R$  和  $Q$  是不相容的。

### 1.3 粗集理论融合信息的一般步骤

用粗集理论进行信息融合的步骤为：

- 1) 将采集到的样本信息按条件属性和结论属性编制一张信息表。
- 2) 利用属性化简及核等概念去掉冗余的条件属性及重复信息, 得出简化信息表。
- 3) 求出核值表。
- 4) 由核值表求出信息表的简化形式。
- 5) 汇总对应的最小规则, 得出最快融合算法。

## 2 病虫害预警设置方法

保护地蔬菜栽培时一般是选择经济价值较高的蔬菜进行栽培, 不同蔬菜的病虫害种类是可能不相同的, 同一种蔬菜有不同类型的病虫害。而每种病虫害都有它的相应发病条件, 如条件属性  $a_1$  (种子处理方法), 条件属性  $a_2$  (温度), 条件属性  $a_3$  (栽种季节), 条件属性  $a_4$  (周围蔬菜栽种蔬菜品种) 等。每种条件或多种条件相结合在一定的状态下会引起一定的病虫害, 用粗集理论为每种病害的发病条件确定一个范围, 只要这些条件特征与这个范围相稳合就给予报警且采取相应的措施预防。根据以往的统计资料和蔬菜可能出现的病虫害有  $d_1, d_2, d_3, d_4$  4 种, 要确定病虫害类型需要把后 3 个条件属性  $a_2, a_3, a_4$  所出现的情况分为 4 种状态, 设第  $i$  个条件属性的状态分为  $a_{i1}, a_{i2}, a_{i3}, a_{i4}$  ( $2 \leq i \leq 4$ ), 第一个条件属性  $a_1$  的状态分为 3 种  $a_{11}, a_{12}, a_{13}$ 。设  $R = \{a_1, a_2, a_3, a_4\}$  是条件属性,  $D = \{d_1, d_2, d_3, d_4\}$  为病虫害属性,  $Q$  是  $R$  的等价类。

第一步: 把条件属性病虫害属性编成信息表 1。

表 1 蔬菜病虫害预警信息表

论域 $U$	条件属性				病虫害属性 $D$
	$a_1$	$a_2$	$a_3$	$a_4$	
1	$a_{12}$	$a_{21}$	$a_{33}$	$a_{41}$	$d_2$
2	$a_{12}$	$a_{21}$	$a_{31}$	$a_{42}$	$d_2$
3	$a_{13}$	$a_{21}$	$a_{31}$	$a_{42}$	$d_2$
4	$a_{13}$	$a_{23}$	$a_{31}$	$a_{42}$	$d_1$
5	$a_{12}$	$a_{23}$	$a_{31}$	$a_{42}$	$d_1$
6	$a_{12}$	$a_{23}$	$a_{33}$	$a_{42}$	$d_1$
7	$a_{11}$	$a_{21}$	$a_{31}$	$a_{43}$	$d_4$
8	$a_{11}$	$a_{21}$	$a_{31}$	$a_{42}$	$d_4$
9	$a_{11}$	$a_{24}$	$a_{33}$	$a_{42}$	$d_4$
10	$a_{11}$	$a_{24}$	$a_{33}$	$a_{44}$	$d_3$
11	$a_{12}$	$a_{24}$	$a_{33}$	$a_{44}$	$d_3$
12	$a_{12}$	$a_{24}$	$a_{34}$	$a_{44}$	$d_3$

第二步: 条件属性约简

属性约简的方法是根据粗集理论考察是否存在  $r_R(Q) = 1$ , 判断行中的行是否由哪些条件唯一确定, 求  $R$  的  $Q$  的核, 删除多余的条件属性和重复信息, 得到条件属性简化表, 具体做法是去掉条件属性  $a_1$  得表 2, 去掉条件  $a_3$  得表 3

表 2 去掉条件属性  $a_1$  简化表

$U$	条件属性			病虫害属性
1	$a_{21}$	$a_{33}$	$a_{42}$	$d_2$
2	$a_{21}$	$a_{31}$	$a_{42}$	$d_2$
3	$a_{21}$	$a_{31}$	$a_{42}$	$d_2$
4	$a_{23}$	$a_{31}$	$a_{42}$	$d_1$
5	$a_{23}$	$a_{31}$	$a_{42}$	$d_1$
6	$a_{23}$	$a_{31}$	$a_{42}$	$d_1$
7	$a_{21}$	$a_{31}$	$a_{42}$	$d_4$
8	$a_{21}$	$a_{33}$	$a_{42}$	$d_4$
9	$a_{24}$	$a_{33}$	$a_{42}$	$d_4$
10	$a_{24}$	$a_{33}$	$a_{44}$	$d_3$
11	$a_{24}$	$a_{34}$	$a_{44}$	$d_3$
12	$a_{24}$	$a_{34}$	$a_{44}$	$d_3$

表 3 去掉条件属性  $a_3$  简化表

$U$	条件属性			病虫害属性
1	$a_{12}$	$a_{21}$	$a_{42}$	$d_2$
2	$a_{12}$	$a_{21}$	$a_{42}$	$d_2$
3	$a_{13}$	$a_{21}$	$a_{42}$	$d_2$
4	$a_{13}$	$a_{23}$	$a_{42}$	$d_1$
5	$a_{12}$	$a_{23}$	$a_{42}$	$d_1$
6	$a_{12}$	$a_{23}$	$a_{42}$	$d_1$
7	$a_{11}$	$a_{21}$	$a_{42}$	$d_4$
8	$a_{11}$	$a_{21}$	$a_{42}$	$d_4$
9	$a_{11}$	$a_{24}$	$a_{42}$	$d_4$
10	$a_{11}$	$a_{24}$	$a_{44}$	$d_3$
11	$a_{11}$	$a_{24}$	$a_{44}$	$d_3$
12	$a_{12}$	$a_{24}$	$a_{44}$	$d_3$

从表 2 可以发现  $a_{21} \wedge a_{31} \wedge a_{42} \Rightarrow d_2, a_{21} \wedge a_{31} \wedge a_{42} \Rightarrow d_4$

$$a_{21} \wedge a_{33} \wedge a_{42} \Rightarrow d_2, a_{21} \wedge a_{31} \wedge a_{42} \Rightarrow d_4$$

这些决策规则是不相容的, 因为每种病虫害发生都有它自己固有的边界条件。显然条件属性  $a_1$  是必不可少的。

从表 3 可以看出, 相同的条件导致相同的病害发生, 决策规则是相容的, 说明条件属性  $a_3$  是冗余的。用同样的方法可检验  $a_2, a_4$  是  $Q$  不可省的, 所以  $a_1, a_2, a_4$  是  $R$  的  $Q$  的核和  $R$  的唯一简化。通过删除多余的条件属性及重复的样本信息得到属性约简表 4

表 4 属性约简表

$U$	条件属性			病虫害属性
1	$a_{12}$	$a_{21}$	$a_{42}$	$d_2$
2	$a_{12}$	$a_{21}$	$a_{42}$	$d_2$
3	$a_{13}$	$a_{23}$	$a_{42}$	$d_1$
4	$a_{12}$	$a_{23}$	$a_{42}$	$d_1$
5	$a_{11}$	$a_{21}$	$a_{42}$	$d_4$
6	$a_{11}$	$a_{24}$	$a_{42}$	$d_4$
7	$a_{11}$	$a_{24}$	$a_{44}$	$d_3$
8	$a_{12}$	$a_{24}$	$a_{44}$	$d_3$

第三步: 值约简求核值和简化值

对表 4 的每一个决策规则求出核值。具体做法是：以  $a_{12} \wedge a_{21} \wedge a_{42} \Rightarrow d_2$ ，为例说明：去掉  $a_{12}$ ，则  $a_{21} \wedge a_{42} \Rightarrow d_2$ ， $a_{21} \wedge a_{42} \Rightarrow d_4$ ，由于此两规则是不相容的，说明  $a_{21}$  是  $d_2$  不可省的，去掉  $a_{21}$ ， $a_{12} \wedge a_{42} \Rightarrow d_2$ ， $a_{12} \wedge a_{42} \Rightarrow d_1$ ，由于此两规则不相容，说明  $a_{42}$  是不可省的。去掉  $a_{42}$ ，则  $a_{12} \wedge a_{21} \Rightarrow d_2$  是相容的说明  $a_{42}$  是  $d_2$  可省略的。对其余决策规则按此法化简，得到决策规则的核值表 5。

表 5 决策规则的核值表

U	$a_1$	$a_2$	$a_4$	d
1	$a_{12}$	$a_{21}$	—	$d_2$
2	$a_{13}$	$a_{21}$	—	$d_2$
3	—	$a_{23}$	—	$d_1$
4	—	$a_{23}$	—	$d_1$
5	$a_{11}$	—	—	$d_4$
6	—	—	$a_{42}$	$d_4$
7	—	—	$a_{43}$	$d_3$
8	—	—	—	$d_3$

从表 5 可以看出  $d_2$  和  $d_1$  决策的核值的集合已简化，因为  $a_{12} \wedge a_{21} \Rightarrow d_2$ ， $a_{13} \wedge a_{21} \Rightarrow d_2$  为真。但  $d_4$  和  $d_3$  决策，核值不能形成值的简化，因为  $a_{11} \Rightarrow d_4$ ，与  $a_{11} \Rightarrow d_3$ ， $a_{42} \Rightarrow d_4$  与  $a_{43} \Rightarrow d_1$ ，是不相容的。由核值表得出决策规则的简化表 6

表 6 决策规则简化表

U	$a_1$	$a_2$	$a_4$	D
1	$a_{12}$	$a_{21}$	—	$d_2$
2	$a_{13}$	$a_{21}$	—	$d_2$
3	—	$a_{23}$	—	$d_1$
4	—	$a_{23}$	—	$d_1$
5	$a_{11}$	—	$a_{41}$	$d_4$
5'	$a_{11}$	$a_{21}$	—	$a_4$
6	—	$a_{42}$	$a_{41}$	$d_4$
6'	$a_{11}$	—	$a_{41}$	$a_4$
7	—	—	$a_{43}$	$d_3$
8	—	$a_{21}$	$a_{43}$	$d_3$
8'	$a_{12}$	—	$a_{43}$	$d_3$
8''	$a_{12}$	$a_{21}$	—	$d_3$

因此，病虫害预测的最快融合算法为：

$$\begin{aligned}
 &a_{12} \wedge a_{21} \vee a_{13} \wedge a_{21} \Rightarrow d_2 \\
 &a_{23} \Rightarrow d_1 \\
 &a_{11} \wedge a_{41} \vee a_{11} \wedge a_{24} \wedge a_{41} \Rightarrow d_4 \\
 &a_{43} \Rightarrow d_3
 \end{aligned}$$

### 3 蔬菜病虫害防治方法

前面已利用粗集理论对所采集到的样本进行处理，得到了最快的融合算法，以最少的信息能迅速为可能出现的病虫害类型报警，然后根据该种病虫害类型应采取的预防措施进行预防。无需去建立模型库，且能够融合不完整和不精确的信息。这对提高融合速度，增强预测能力及栽种知识的传播，丰产丰收，发展绿色蔬菜生产，稳定市场供应具有重大的意义。

#### 参考文献：

- [1] 李宝栋,冯东昕. 黄瓜病虫害防治新技术[M].北京:金盾出版社 1993.
- [2] PAWLAK Z. Rough sets[J]. Int. J. Communication of the ACM, 1995, 38(11): 89 - 95.
- [3] PAWLAK. Z. Rough sets[J]. Int. J. Of Computer and information Science, 1982, 11: 341 - 356.
- [4] 曾黄麟. 粗集理论及其应用[M]. 重庆:重庆大学出版社, 1996.

## Alarm of Blight in the Process of Vegetables Cultivation on the Bases of Rough Sets Theory

YI Zheng-jun<sup>1</sup>, HUANG Han-ming<sup>2</sup>, HUANG Xi-yue<sup>2</sup>

(1. College of science, Chongqing university, Chongqing 400044, China;

2. College of automation, Chongqing university, Chongqing 400044, China)

**Abstract:** This essay applies rough sets theory into the alarm of blight in the process of cultivation of vegetables. One kind of the simplest knowledge extraction rules can be constructed, so that the quickest information fusion algorithms. can be got and thus a optimized alarm for the prevention of blight in vegetable cultivation is provided. This bears great significance to the spread of planting knowledge, the production of green vegetables, high yield and bumper harvest, and stabilizing the market.

**Key words:** rough sets; information fusion; vegetable cultivation; alarm

(责任编辑 吕赛英)