

模糊推理算法及信息不确定条件下的诊断应用

房建东^{1,2}, 赵于东³

(1. 天津大学电子信息工程学院, 天津 300072; 2. 内蒙古工业大学信息工程学院, 呼和浩特 010051;
3. 内蒙古自治区农牧业厅信息中心, 呼和浩特 010068)

摘要:针对系统动态特征信息的随机性、模糊性和不完备性,借助模糊数学理论的逻辑推理方法,构造一种可实现农作物病虫害模糊推理诊断模型,给出在已知症状信息条件下的求解算法。仿真算例表明,模糊逻辑推理诊断模型及算法能够有效完成农作物病虫害在多因素、多症状及症状信息不完备条件下的诊断推理,具有一定智能化程度、简单实用等主要技术特点,表现出一定的诊断可靠性。

关键词:模糊逻辑;信息不确定;智能决策

Fuzzy Reasoning Method and Its Diagnosis Application with Incomplete Certain Information

FANG Jian-dong^{1,2}, ZHAO Yu-dong³

(1. College of Electronic Information Engineering, Tianjin University, Tianjin 300072;
2. College of Information Engineering, Inner Mongolia University of Technology, Huhhot 010051;
3. Information Center of Agriculture and Animal Husbandry Department of Inner Mongolia, Huhhot 010068)

【Abstract】 For the randomness, fuzziness and incompleteness in the of system dynamic information, a fuzzy logic diagnosis model to the crops diseases and insect pests is built using the logic reasoning method in mathematical theory, and an algorithm is given to the known symptoms. Simulation example SHOES, the fuzzy logic algorithm can effectively complete the diagnosis on insect pests with the lack of many factors, symptoms and complete information, which is intelligent, simple and practical.

【Key words】 fuzzy logic; incomplete certain information; intelligent decision

1 概述

人类智能和生物智能的最大差别,就是人能够通过语言文字等方式传承千百年来积累的知识^[1]。智能的不确定性必然反映在知识中。知识的不确定性,还反映在常识知识上,也称为元知识,是其他专业知识的基础。知识的不确定性有多个方面,主要有随机性、模糊性和不完备性。

农作物病虫害诊断中也存在着不确定性。随机性主要反映客观世界的不确定性,常用概率统计方法进行研究。模糊性主要反映人为主观理解上的不确定性,它表现在边界的亦此亦彼性,主要是由农作物病虫害症状表现的模糊性造成的。不完备性是由于对农作物病虫害发生机理认识及检测手段等客观因素的限制,导致病虫害症状信息或诊断信息不完整^[2]。

随着科学技术的发展,人们逐渐认识到,若将存在不确定性事物人为地精确化,不仅会以方法的复杂性为代价,而且会降低结果的有意义性,使其在实际中毫无用处。不相容原理指出:当系统的复杂程度提高时,对其进行精确而有意义的认识和描述的能力会随之下降^[3]。合理借助模糊性可以对复杂事物做出正确、高效率的判断和处理。

2 模糊诊断模型的建立

农作物病虫害的模糊诊断问题中,“病症-病虫害”之间的对应关系可以定义为

$$P=(D, M, C, R, M_S) \quad (1)$$

其中, $D=(d_1, d_2, \dots, d_n)$ 表示农作物病虫害的有限非空集合;
 $M=(m_1, m_2, \dots, m_k)$ 表示农作物病虫害症状的有限非空集合;

$C=\{c_{11}, c_{12}, \dots, c_{mk}\}$, c_{ij} 表示农作物病虫害 d_i 引起病虫害症状 m_j 的可信度; $R \subseteq D \times M \times C$, $\langle d_i, m_j, c_{ij} \rangle \in R$ 表示农作物病虫害 d_i 引起病虫害症状 m_j 的可能性是 c_{ij} , 其定义域为 $T(R)=D$, 而值域为 $F(R)=M$ 。

给定 D, M, C, R , 可定义

$$effect(d_i) = \{m_j | \langle d_i, m_j, c_{ij} \rangle \in R, \forall d_i \in D\}$$

表示所有病虫害 d_i 可能引起的病虫害症状集合。

$$cause(m_j) = \{d_i | \langle d_i, m_j, c_{ij} \rangle \in R, \forall m_j \in M\}$$

表示所有可能引起病虫害症状 m_j 的病虫害集合。

$M_S = (m_1, m_2, \dots, m_k)$ 是 M 的一个模糊子集, 为已知农作物病虫害症状集合, 对于每一个 $m_j \in M_S$, $\mu(m_j) \in [0, 1]$ 是 m_j 在 M_S 中的隶属度。

根据模糊诊断问题的假设, 定义农作物病虫害模糊诊断解的相对似然函数为

$$L(D_I, M_S) = L_1(D_I, M_S) \cdot L_2(D_I, M_S) \quad (2)$$

其中, $L(D_I, M_S)$ 表示已知诊断信息包括的病虫害集合 D_I 的相

基金项目: 国家电子信息产业发展基金资助项目(XBY2007-329); 内蒙古自治区自然科学基金资助项目(200408020812)

作者简介: 房建东(1966 -), 女, 博士, 主研方向: 智能决策及智能控制; 赵于东, 研究员

收稿日期: 2007-08-20 **E-mail:** fangjd@imut.edu.cn

对似然函数，为农作物病虫害集合 D 能够解释模糊病症集 M_S 的程度； $L_1(D_I, M_S)$ 表示 D_I 有多大可能引起 M_S 中病症存在形成的加权； $L_2(D_I, M_S)$ 表示由 D_I 的先验概率组成的加权。

$$L_1(D_I, M_S) = \prod_{m_j \in M_S} \mu(m_j) p(m_j | D_I) = \prod_{m_j \in M_S} \mu(m_j) (1 - \prod_{d_i \in D_I} (1 - c_{ij}))$$

$$L_2(D_I, M_S) = \prod_{d_i \in D_I} \frac{p_i}{1 - p_i}$$

相对似然函数用于度量在已知病虫害症状模糊度的情况下诊断解出现的相对似然度即相对可能程度。实际诊断中，相对似然度随着已存在的病虫害症状的隶属度及病虫害症状与病虫害之间的因果强度的增加而增大。如果存在某种病虫害症状的可能性很大而且病虫害 d_i 与病虫害症状 m_j 之间有很强的相关性因果强度，则病虫害 d_i 出现的可能性相对较大；如果病虫害 d_i 与某病虫害症状 m_j 之间没有太大的相关性，那么即使该症状存在的可能性较大，病虫害 d_i 出现的可能性也不大；反之，如果病虫害症状存在的可能性较小，那么即使病虫害与症状之间的相关性很大，病虫害 d_i 出现的可能性也不会太大。似然函数 L 体现出了病虫害症状模糊度对病虫害诊断结果的正影响，体现了病虫害和病虫害症状之间的因果强度对病虫害诊断结果的正影响。

关于病虫害症状模糊隶属度的选取，可由用户根据实际情况来确定或修正。本文对每个病症构造了5个等级：病虫害症状完全相同：{1}；病虫害症状相近：{0.8}；病虫害症状基本相近：{0.6}；病虫害症状不太相近：{0.4}；病虫害症状相似度很小：{0.1}。

3 模糊诊断推理算法

作物的生长过程是一个随时间和环境条件而变的变化过程。不同地理位置的环境条件、水肥条件和病虫害因素等，都对作物的生长产生影响。因而在推理过程中应以事实作为条件对问题进行求解。

推理算法作为作物病虫害模糊诊断系统的核心，其主要任务就是在问题求解过程中适时地决定知识的选择和运用。推理的控制策略确定知识的选择。模糊诊断推理算法的基本思想是，在建立元知识库、规则库、数据库的基础上，从用户提供的已有事实，对作物所产生症状的初始原因做试探性假设，然后在当前的假设引导下寻求进一步的信息，从而推出新的结果。

首先引入符号定义：

M_1 ： M_S 的子集，目前已知的病虫害症状集合；

D_I ： $cause(M_1)$ ，能够引起已知症状的所有病虫害的集合；

S ：对已知症状的试探解集合，为描述多病虫害的症状特征， s 用笛卡尔积表示。 \times 表示笛卡尔积，初始状态为 $M_1 = \Phi$ ， $D_I = \Phi$ ， $S=D$ ，终止条件为 $M_1=M$ ，即找不出可处理的症状。

模糊诊断推理的基本算法描述如下：

(1)从已知模糊病虫害症状集合 M_S 中得到一个症状 m_j ；/*输入一个目前已知症状*/

(2) $M_1 = M_1 \cup m_j$ ；/*在目前已知的农作物病虫害症状子集 M_1 中加入新增症状*/

(3) $D_I = D_I \cup cause(m_j)$ ；/*在已知诊断信息集合 D_I 中加入新增可能引起症状的诊断信息*/

(4) $S'' = S' - S$ ， $S' = S$ ， $S = S \cap cause(m_j)$ ；/*构造农作物所产生病虫害症状试探性集合*/

(5)if $S = \Phi$ then{ $S_1 = S' \times cause(m_j)$ }/*若已知症状的试探解集合为空，则构造试探集合 S_1 */

if $effect(s'') \cup effect(\{d_i | d_i \in cause(m_j)\})$ 是目前已知病虫害症状集 M_1 的最小覆盖

then $S_2 = \{d_i\} \times S''$ /*若已知症状试探解集合是 M_1 的最小覆盖，则构造试探集合 S_2 */

else $S_2 = \Phi$ $s = S_1 \cup S_2$ }；/*若不是最小覆盖，则构造试探集合 S */

(6)if $M_1=M_S$ then end/*如果目前症状集合与已知集合一致则算法结束*/

if $M_1=M$ then end/*如找不出可处理的症状，诊断终止*/ else go to (1)

4 模糊推理算法在实际诊断中的应用

4.1 模糊诊断推理算例

考虑一个田间玉米虫害的动态诊断问题，它们都是玉米生长过程中常见而有可能发生的虫害。玉米虫害原始数据由植保专家提供，症状集合的模糊边界及相应隶属函数应根据实测数据整理分析，如果所掌握的实测数据较少，在症状与虫害间因果强度的确定过程中可暂时加入一些主观经验，以后再根据实际情况进一步修正^[4]。取玉米虫害集合和虫害症状集合如表1、表2所示。

表1 虫害症状集合

代号	症状说明
m_{01}	玉米抽穗后钻蛀茎秆，蛀孔处易倒折
m_{02}	幼虫食心叶或蛀食未展开心叶，造成花叶
m_{03}	穗期蛀食雌穗、嫩粒，造成籽粒缺损霉烂
m_{04}	叶片变黄或发红
m_{05}	心叶产生黑色霉状物
m_{06}	叶片变黄枯死，叶面生霉变黑
m_{07}	结果期缩短，降低粒重
m_{08}	叶片呈灰白色或枯黄色细斑
m_{09}	穗期受害遇风易折倒形成瘪穗和秕粒
m_{10}	幼虫食叶
m_{11}	初孵幼虫取食叶肉，残留表皮
m_{12}	将叶片吃成缺刻或仅留叶脉，叶片呈网状
m_{13}	地面上咬幼苗
m_{14}	主茎硬化可爬到上部为害生长点
m_{15}	造成缺苗断垄，重则毁种

表2 作物虫害集合和症状类型

代号	虫害类型	症状类型 $effect(d_i)$
d_{01}	玉米螟	$m_{01} m_{02} m_{03} m_{07}$
d_{02}	玉米蚜	$m_{04} m_{07}$
d_{03}	玉米截形叶螨	$m_{04} m_{07} m_{08} m_{13}$
d_{04}	粟灰螟	$m_{01} m_{07} m_{09}$
d_{05}	粘虫	$m_{04} m_{07} m_{10} m_{13}$
d_{06}	草地螟	$m_{07} m_{11} m_{12}$
d_{07}	白边地老虎	$m_{07} m_{13} m_{14} m_{15}$

4.2 仿真实验结果

针对玉米作物的虫害进行诊断，假设 $M_S=(m_{04}, m_{13}, m_{14})$ ，相应的模糊隶属度分别为

$$\mu(m_{04})=0.8, \mu(m_{13})=0.4, \mu(m_{14})=0.2, D_I=7, p_i=1/7$$

根据模糊诊断推理算法，可以求出各个集合的关系变化，如表3所示。

表3 作物虫害集诊断推理结果

症状集合	初始状态	m_{04}	m_{13}	m_{14}
M_1	Φ	$\{m_{04}\}$	$\{m_{04}, m_{13}\}$	$\{m_{04}, m_{13}, m_{14}\}$
D_I	Φ	$\{d_{02}, d_{03}, d_{05}\}$	$\{d_{02}, d_{03}, d_{05}, d_{07}\}$	$\{d_{02}, d_{03}, d_{05}, d_{07}\}$
S'	Φ	D	$\{d_{02}, d_{03}, d_{05}\}$	$\{d_{03}, d_{05}\}$
S''	Φ	Φ	$\{d_{01}, d_{04}, d_{06}\}$	$\{d_{02}, d_{03}\}$
S	D	$\{d_{02}, d_{03}, d_{05}\}$	$\{d_{03}, d_{05}\}$	$d_{07} \times \{d_{03}, d_{05}\}$

根据表3的诊断结果，得出已知症状的试探解集合

$$S = d_{07} \times \{d_{03}, d_{05}\} = (\{d_{03}, d_{07}\}, \{d_{05}, d_{07}\})$$

采用最大似然值原则求出诊断解的可能程度：

$$L_1(\{d_{03}, d_{07}\}) = \mu(m_{04})^{(1 - (1 - c_{03,04})(1 - c_{07,04}))} \times \mu(m_{13})^{(1 - (1 - c_{03,13})(1 - c_{07,13}))} \times \mu(m_{14})^{(1 - (1 - c_{03,14})(1 - c_{07,14}))} = 0.064$$

$$L_2(\{d_{03}, d_{07}\}) = \frac{p_{03} \times p_{07}}{(1 - p_{03})(1 - p_{07})} = 0.028$$

同理，有

$$L_1(\{d_{05}, d_{07}\}) = \mu(m_{04})^{(1 - (1 - c_{05,04})(1 - c_{07,04}))} \times \mu(m_{13})^{(1 - (1 - c_{05,13})(1 - c_{07,13}))} \times \mu(m_{14})^{(1 - (1 - c_{05,14})(1 - c_{07,14}))} = 0.031$$

$$L_2(\{d_{05}, d_{07}\}) = \frac{p_{05} \times p_{07}}{(1 - p_{05})(1 - p_{07})} = 0.028$$

$$\frac{L(\{d_{03}, d_{07}\})}{L(\{d_{03}, d_{07}\}) + L(\{d_{05}, d_{07}\})} = \frac{L_1(\{d_{03}, d_{07}\})L_2(\{d_{03}, d_{07}\})}{L_1(\{d_{03}, d_{07}\})L_2(\{d_{03}, d_{07}\}) + L_1(\{d_{05}, d_{07}\})L_2(\{d_{05}, d_{07}\})} = 0.674$$

$$\frac{L(\{d_{05}, d_{07}\})}{L(\{d_{03}, d_{07}\}) + L(\{d_{05}, d_{07}\})} = \frac{L_1(\{d_{05}, d_{07}\})L_2(\{d_{05}, d_{07}\})}{L_1(\{d_{03}, d_{07}\})L_2(\{d_{03}, d_{07}\}) + L_1(\{d_{05}, d_{07}\})L_2(\{d_{05}, d_{07}\})} = 0.326$$

(上接第173页)

4 光滑支持向量机未来可能的发展方向

光滑支持向量机在分类问题上已显示出了它的优势，但作为一种新的模式分类方法，还存在很多尚未解决或尚未充分解决的问题。本文对一般支持向量机作了回顾，对光滑支持向量机的理论进行了综述，未来可能的发展方向是：(1)设法寻找更多的光滑函数，使新光滑函数的性能优于以往的光滑函数。(2)用新的光滑函数对非光滑的支持向量机模型做光滑处理，提出新的光滑支持向量机模型，并证明其全局收敛的速度更快。(3)研究光滑支持向量机的一般形式，使其理论更完备、应用更有效。(4)探索求解光滑支持向量机自身的新算法，进一步解决分类速度较慢的问题。

只要在某种程度上解决了这些问题，一方面能够建立新的支持向量机模型及其新的求解算法，进一步提高分类的性能和效率，丰富机器学习的理论，另一方面可为已有的应用领域提供更好的理论支持和技术方法。

致谢：在本文的撰写过程中曾得到彭宏教授和胡天明博士的热情帮助，在此表示感谢。

参考文献

[1] Joachims T. Make Large-scale Support Vector Machine Learning Practical Advances in Kernel Methods-support Vector Learning[M].

当输入玉米虫害症状为 $\{m_{04}, m_{13}, m_{14}\}$ 时，诊断出的虫害集合为 $\{d_{03}, d_{07}\}, \{d_{05}, d_{07}\}$ ，由最大似然值原则得出2种诊断解集的可能性分别为67.4%和32.6%。因此，可以得出结论：虫害 d_{03} (即玉米截形叶螨)为该症状集合的最佳解。由此可以看出，通过采用模糊隶属度的方法对农作物病虫害诊断问题的求解，不仅可以提示用户搜集信息，而且可以根据已知症状给出符合农作物病虫害发生特点的诊断结果。

5 结束语

由于在农作物病虫害诊断过程中同时存在具有随机性、模糊性和症状信息不完备性等不确定因素，单纯用概率统计或模糊数学的方法并不利于准确影射病虫害的发生特性。通过对玉米作物病虫害的仿真诊断测试，表明本文所采用的模糊诊断推理算法对农作物病虫害诊断问题的求解是一种有效尝试，该方法实现了符号推理和模糊推理的结合，具有一定的诊断可靠性和实用性。模糊诊断推理算法不仅可用于农作物病虫害诊断问题，还可在其他诊断场合借鉴应用。

模糊诊断推理模型还具有一定的开放性，随着对农作物病虫害发生机理认识的深入和检测手段的进步，对于病虫害症状和病虫害的划分，以及隶属度和可信度的确定都可进行相应的修正，从而使诊断模型更符合实际情况，也可使诊断正判率得到进一步的提高。

参考文献

[1] 李道亮, 傅泽田, 田东. 智能系统基础方法及其在农业中的应用[M]. 北京: 清华大学出版社, 2004.

[2] Oerke E C, Dehne H W. Safeguarding Production—Losses in Major Crops and the Role of Crop Protection[J]. Crop Protection, 2004, 23(2): 275-285.

[3] Fox J. Strand Some Agrometeorological Aspects of Pest and Disease Management for the 21st Century[J]. Agricultural and Forest Meteorology, 2000, 103(1/2): 73-82.

[4] 李鸿吉. 模糊数学基础及实用算法[M]. 北京: 科学出版社, 2005. Cambridge, MA: MIT Press, 1999: 185-208.

[2] Platt J. Fast Training of Support Vector Machines Using Sequential Minimal Optimization[M]. Cambridge, MA: MIT Press, 1999: 185-208.

[3] Lee Y J, Mangasarian O L. SSVM: A Smooth Support Vector Machine[J]. Computational Optimization and Applications, 2001, 20(1): 5-22.

[4] Mangasarian O L, Musicant D R. Lagrangian Support Vector Machines[J]. Journal of Machine Learning Research, 2001, 22(1): 161-177.

[5] Lee Y, Mangasarian O L. RSVM: Reduced Support Vector Machines [C]//Proceedings of the 1st SIAM International Conference on Data Mining. Chicago, USA: [s. n.], 2001.

[6] Fung G, Mangasarian O L. Proximal Support Vector Machine Classifiers[C]//Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining. San Francisco, California, USA: [s. n.], 2001: 77-86.

[7] 袁玉波, 严杰, 徐成贤. 多项式光滑的支撑向量机[J]. 计算机学报, 2005, 28(1): 9-17.

[8] 熊金志, 胡金莲, 袁华强, 等. 一类光滑支持向量机新函数的研究[J]. 电子学报, 2007, 35(2): 366-370.

