

小麦赤霉病 CBR 预测模型参数的优化

关东¹, 陈莉^{1*}, 张沙沙¹, 丁克坚¹, 朱诚¹, 刘家成², 郑朝阳²
(1.安徽农业大学植物保护学院, 合肥 230036; 2.安徽省农业委员会, 合肥 230061)

摘要: 基于类比推理人工智能方法 CBR 原理建立了以时间序列为轴线的小麦赤霉病滚动预测模型。为提高模型预测的准确性, 利用安徽省农作物病虫害数据库, 结合专家经验会商的结果, 优化筛选该预测模型中的预测单元、预测阶段、各预测阶段的权重、气象因子及其权重等各关键预测参数, 并检验其预测的准确性。结果表明, 运用优化后的参数, 小麦赤霉病 CBR 预测模型常年预测准确率可达 84.21%。故优化后的预测参数可用于小麦赤霉病 CBR 预测模型。

关键词: 案例推理; 预测因子; 权重; 优化

中图分类号: S435.121.45

文献标识码: A

文章编号: 1672-352X (2014)01-0082-05

Parameter optimization for CBR model of wheat scab

GUAN Dong¹, CHEN Li¹, ZHANG Shasha¹, DING Kejian¹, ZHU Cheng¹, LIU Jiacheng², ZHENG Zhaoyang²
(1. School of Plant Protection, Anhui Agricultural University, Hefei 230036; 2. Agriculture Committee in Anhui Province, Hefei 230061)

Abstract: A rolling forecasting model of wheat scab was set up using CBR theory with time series. To improve the accuracy and reliability of this model, impact factors and parameters, such as model unit, prediction stages, meteorological factors and its weight were optimized by using the database system of crops pests in Anhui province and the results of expert consultation. The accuracy of CBR model of wheat scab with optimized parameters was tested, and the results showed that the precision rate was 84.21% in a normal year. The optimized factors and parameters could be applied to the forecasting model of wheat scab.

Key words: case-based reasoning; impact factors; weight; optimization

赤霉病是我国小麦生产上重要的病害, 其产生的毒素严重威胁人类和动物健康^[1-2]。在长江中下游地区, 小麦赤霉病几乎每两三年就有一次中度流行或大流行, 给粮食生产造成严重威胁。小麦赤霉病的发生与气象因子关系密切^[3-4], Moschini 等用 SAS 软件分析了气象因子对小麦赤霉病发生的影响, 并以此建立了预测模型^[5]。Madgwick 等分析了小麦花期与小麦赤霉病发生有关的气象因子, 并以此为基础构建了美国小麦赤霉病预测模型^[6]。我国对小麦赤霉病的预测也有很长时间的的发展, 梁震中等通过分析了气候、病菌活动状况等因子研究小麦赤霉病分段预测模型^[7]。关海鸥等运用 BP 神经网络算法的一种改进对平均相对湿度、降雨日数和日照时数 3 个气象主要因素分析, 完成小麦赤霉病

的预测^[8]。李军等用地面气象因子、500 hPa 高空环流因子和北太平洋海温场因子, 选取相关系数较高且稳定性好的预测因子, 然后在逐步回归和逐步判别分析方法基础上, 集成建立了小麦赤霉病发生程度的动态预测模式^[9]。霍治国等小麦赤霉病发生的气象因子, 建立了我国小麦赤霉病最大熵谱预报模式^[10]。张旭晖等对小麦赤霉病发生具有影响的气象因子进行统计分析, 建立了小麦赤霉病气象等级预报模式^[11]。以上预测模型多数是基于数理统计的方法和原理建立的静态预测模型, 受数据来源的时空动态限制强, 缺乏植保站专业技术人员的参与和互动环节, 植保部门在病虫害预测过程中很难长期实际应用。基于案例推理 (Case-based reasoning, CBR) 是由美国耶鲁大学 Roger Schank 教授首次提

收稿日期: 2013-06-24

基金项目: 公益性行业 (农业) 科研专项 (201203016) 资助。

作者简介: 关东, 硕士研究生。E-mail: guandong880316@163.com

* 通信作者: 陈莉, 副教授。E-mail: chenlii@ahau.edu.cn

出的一种适用于机械产品再设计的类比推理人工智能方法^[12]。其基本原理是基于人的认知心理过程, 用先前求解问题的经验和方法, 通过类比、联想和推理来解决当前相似问题的一种推理方法。目前, 基于案例推理的技术已经广泛应用于法律咨询、医疗诊断、工程规划和设计、园林病虫害诊断、农作物病虫害诊断、天气预报以及故障诊断等一些已积累丰富经验的领域中^[13-14]。本研究基于 CBR 原理建立了小麦赤霉病预测模型, 并对其参数进行了优化, 以提高该模型预测的准确性, 使其在农业生产中发挥更大的作用。

1 材料与方 法

1.1 小麦赤霉病 CBR 预测模型的组建

根据案例推理 (CBR) 原理建立了基于时间序列数据的相似年分析的小麦赤霉病预测模型, 该模型针对病害发生的复杂性, 进行案例推理时, 首先利用历史测报数据 (包括气象和作物、病虫害等相关数据等) 生成时间序列历史案例库, 再按照指定或预先设定匹配规格和参数因子权重来构建相似度函数, 然后逐一计算目标案例与历史案例的匹配程度, 最终选取与目标案例相似程度最高的历史案例的结论作为预测结论; 在预测过程中, 未来相关的预报数据也可加入目标案例的时间序列中。此外, 针对病害数据为时间序列的特性, 该预测方法采用了滑动窗口分段匹配和预设模式匹配方法, 以获得相对较高的可信度和执行效率。

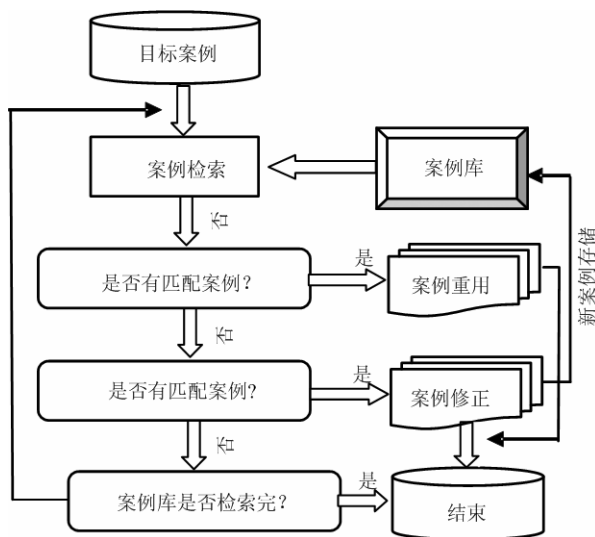


图 1 小麦赤霉病 CBR 预测模型组建的步骤

Figure 1 Process for the establishment of forecasting model of wheat scab

参数因子向量: $X = (\text{气温, 气温权重; 气温距平, 气温距平权重; 降水量, 降水量权重; 降水距平, 降水距平权重; 日照时数, 日照时数权重})$

前期向量: $QQ = (X_{qq1}, X_{qq2}, X_{qq3} \cdots X_{qqn})$; 前期权重 W_q 。

关键期向量: $GJQ = (X_{gjq1}, X_{gjq2}, X_{gjq3} \cdots X_{gjqm})$; 关键期权重 W_{gjq} 。

滚动预测期向量: $YB = (X_{yb1}, X_{yb2}, X_{yb3} \cdots X_{ybk})$; 预报期权重 W_{yb} 。

其中 $X_{qq1}, X_{qq2}, \cdots X_{qqn}$ 为前期相应预测单元的气象因子; $X_{gjq1}, X_{gjq2}, \cdots X_{gjqm}$ 为关键期相应预测单元的气象因子; $X_{yb1}, X_{yb2}, \cdots X_{ybk}$ 为各滚动预测期的气象因子。

计算每一条历史案例的向量与本年度向量的相似度, 取相似度最大 (XSDall 最小) 的历史案例的发病测报值作为预测值。相似度计算使用欧拉相似度计算公式:

$$XSD = \sqrt{\sum_i (X_i \times W_{xi} - Y_i \times W_{yi})^2}$$

其中 X_i 为案例数据, W_{xi} 为案例数据的权重, Y_i 为当前数据, W_{yi} 为当前数据的权重。

1.2 数据来源及其处理

1.2.1 数据来源 病害观测值来源于安徽省病虫害监测预警管理系统中 1986-2011 年的年报数据。气象数据包括旬平均气温、旬气温距平、旬降雨量、旬降雨距平、旬日照时数, 来源于安徽省气象局。

1.2.2 数据选择和处理 本文所用气象数据和小麦赤霉病病情数据选取自贵池、桐城、阜南、颍东区等 4 个站点, 病情观测值为病穗率。对基层植保工作人员的直接或间接田间调查获得的年报数据, 整理并消除量纲后作为有效数据使用。

1.3 小麦赤霉病 CBR 预测模型的参数优化

CBR 预测模型中预测参数的优化分 3 步进行: 一是 CBR 模型中预测因子的分段划分优化筛选; 二是 CBR 模型中阶段权重的优化筛选; 三是 CBR 预测模型中各气象因子权重的优化筛选。

1.3.1 预测参数的初优化 (预测单元的确定) 为筛选出 CBR 预测模型中小麦赤霉病预测前期和关键期的阶段划分, 先保持预测分段的阶段权重和气象因子权重不变, 对预测分段进行优化筛选。试验设置的起始点为 1 月上旬, 变动预测前期和预测终点, 共设置 27 个处理 (表 1)。每个处理设置 2 个组, 分别为组 I 和组 II, 2 组同时进行。组 I 的设置 为预测前期每 3 旬作为 1 个预测单元, 关键期以 一旬为 1 个预测单元, 而组 II 中预测前期和关键期

均是一旬为1个预测单元。

1.3.2 CBR模型中小麦赤霉病预测阶段权重的优化 在分段筛选的基础上,对2个预测阶段(即前期和关键期)的权重进行优化,共设置4个处理(表2)。

1.3.3 CBR模型中气象因子及其权重的筛选 在阶段划分和阶段权重筛选的基础上,对气象因子及其权重进行了优化筛选,试验处理设置如表3。

1.3.4 参数优化的指标 将试验设置的处理运用于CBR预测模型对小麦赤霉病的病穗率进行预测,并计算预测值与历史真实值之间的误差标准差,选择误差标准差小的作为优化后的预测模型的参数。误差标准差的计算公式如下:

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

其中 X_i 为预测值与真实值的差值, \bar{X} 为误差的均值, n 为年份数量。

1.4 预测符合率检验

利用贵池、桐城、阜南、颖东区等4个站点的数据等,对优化后的小麦赤霉病CBR预测模型参数

进行检验,将预测值与历史真实值完全符合和相差一个病级的视为准确的,计算预测准确率。

2 结果与分析

2.1 CBR模型中小麦赤霉病预测阶段的确定

从表1可以看出,组I中误差标准差最小的是将1月上旬至3月中旬作为病害预测的前期,3月中旬至4月上旬作为病害预测的关键期的处理,其误差标准差为0.11228;组II中误差标准差最小的是1月上旬至3月中旬为前期,3月中旬至4月下旬为关键期的处理,其误差标准差为0.0928,小于组I的误差标准差。因此,在小麦赤霉病的CBR预测模型中,预测阶段的划分采用1月上旬至3月中旬为前期,3月中旬至4月下旬为关键期。

2.2 小麦赤霉病CBR预测模型中预测单元的确定

组I中前期每3旬作为1个预测单元,关键期1旬为1个预测单元,组II中预测前期和关键期均以1旬为1个预测单元。分别计算各组的误差标准差,结果见表1。

表1 小麦赤霉病CBR模型中预测阶段和预测单元的优化

Table 1 Optimization of prediction unit and stages in CBR model of wheat scab

前期和关键期划分 Division of prophase and critical period	误差标准差 <i>SD</i>		前期和关键期划分 Division of prophase and critical period	误差标准差	
	组 I Group I	组 II Group II		组 I Group I	组 II Group II
1月上—2月下—4月上	0.11572	0.11920	1月下—3月中—4月中	0.11984	0.12202
1月上—3月上—4月上	0.11525	0.11844	1月下—2月下—4月下	0.11255	0.12597
1月上—3月中—4月上	0.11228	0.12007	1月下—3月上—4月下	0.11932	0.09979
1月上—2月下—4月中	0.11782	0.12033	1月下—3月中—4月下	0.11495	0.09849
1月上—3月上—4月中	0.12248	0.12052	1月中—2月下—4月上	0.11560	0.12732
1月上—3月中—4月中	0.11877	0.10654	1月中—3月上—4月上	0.12288	0.11575
1月上—2月下—4月下	0.11343	0.10866	1月中—3月中—4月上	0.11873	0.11564
1月上—3月上—4月下	0.11781	0.11233	1月中—2月下—4月中	0.11612	0.13208
1月上—3月中—4月下	0.11242	0.09280	1月中—3月上—4月中	0.11971	0.12132
1月下—2月下—4月上	0.11660	0.12812	1月中—3月中—4月中	0.11790	0.12091
1月下—3月上—4月上	0.12372	0.11618	1月中—2月下—4月下	0.11285	0.12205
1月下—3月中—4月上	0.12512	0.11612	1月中—3月上—4月下	0.12635	0.10412
1月下—2月下—4月中	0.11892	0.13257	1月中—3月中—4月下	0.11396	0.10399
1月下—3月上—4月中	0.11896	0.12251	平均 Mean	0.11778	0.11644

表2 小麦赤霉病CBR预测模型中预测阶段权重的筛选

Table 2 Selecting weight of prediction stages in CBR model of wheat scab

处理编号 Treatment code	阶段权重 Weight of prediction stage	误差标准差 <i>SD</i>	
		组 I Group I	组 II Group II
A	前期 30%, 关键期 70%	0.11228	0.09280
B	前期 40%, 关键期 60%	0.11254	0.09563
C	前期 35%, 关键期 65%	0.11255	0.09485
D	前期 45%, 关键期 55%	0.11294	0.09531

在选择的 1 月上旬至 3 月中旬为前期, 3 月中旬至 4 月下旬为关键期的预测阶段处理中, 组 I 的误差标准差为 0.11242, 组 II 的为 0.09280, 说组 II 中的预测参数较组 I 中的预测参数预测结果误差更稳定。因此, 在对赤霉病的预测中采取前期和关键期均为一句作为一个预测单元的划分方法。

2.3 CBR 预测模型中预测阶段权重的确定

分别计算各处理组的误差标准差, 结果见表 2。组 I 和组 II 中误差标准差最小的均为处理 A, 因此, 在小麦赤霉病的 CBR 预测模型中予以前期 30%, 关键期占 70% 的预测权重。

2.4 CBR 模型中气象因子及其权重的确定

在预测阶段和预测单元划分的基础上, 将小麦赤霉病的预测阶段前期定为 1 月上旬至 3 月中旬, 并赋予 30% 的权重, 关键期为 3 月中旬至 4 月下旬, 所占权重为 70%, 前期和关键期均设置一句作为一个预测单元, 运用 CBR 预测模型的对小麦赤霉病的病穗率进行预测, 并计算预测值与真实值之间的误

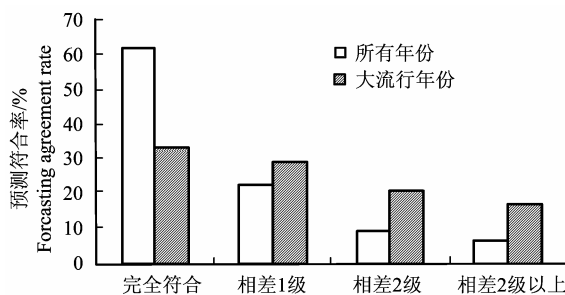
差标准差。结果 (表 3) 表明, 误差标准差最小的是处理 h, 因此在运用该模型对小麦赤霉病进行预测时, 将各气象因子的权重设置为旬平均气温权重占 15%、旬气温距平占 15%、旬降水量占 35%、旬降水距平占 35%。

2.5 预测符合率检验

提取安徽贵池、桐城、阜南、颍东区等地 1986 至 2011 年共 25 年的小麦赤霉病情和相关气象数据, 运用参数优化后的小麦赤霉病 CBR 预测模型对小麦赤霉病发生程度进行回检, 检验其预测符合率 (图 1), 预测值与观测值在同一病级和误差一个病级的均视为预测准确。在四地所有 25 年的 76 组可用数据中, 预测值与实际发生值完全符合的达到 61.84%, 相差一个病级的占 22.37%, 其预测准确率可达 84.21%。在可用的 24 组小麦赤霉病大发生年份数据中, 预测值与实际发生值完全符合的占 33.33%, 相差一个病级的占 29.17%, 其预测准确率为 62.50%。

表 3 小麦赤霉病 CBR 预测模型中气象因子及其权重的优化
Table 3 Optimization of meteorological factor and its weight in CBR model of wheat scab

处理编号 Code	气象因子的权重/% Weight of meteorological factor				误差标准差 <i>SD</i> (组 II, group II)
	旬平均气温 Ten-day average temperature	旬气温距平 Ten-day average temperature departure	旬降水量 Ten-day average precipitation	旬降水距平 Ten-day average precipitation departure	
a	20	20	30	30	0.09280
b	25	25	25	25	0.13295
c	15	20	35	30	0.09274
d	15	20	30	35	0.09255
e	20	15	30	35	0.09196
f	20	15	35	30	0.09225
g	15	15	40	30	0.09153
h	15	15	35	35	0.09134
i	10	15	45	30	0.09214
j	10	15	50	25	0.09202
k	15	15	20	50	0.09186



完全符合 Fully; 相差 1 级 Difference between grade 1; 相差 2 级 Difference between grade 2; 相差 2 级以上 Difference of 2 above

图 2 小麦赤霉病的预测符合率

Figure 2 Forecasting agreement rate of wheat scab

3 小结与讨论

病害的预测可为科学有效防治提供依据。目前, 用于病虫害预测预报的模型多种多样, 常见的预测模型有逐步判别法、投影寻踪、灰色系统、神经网络等^[15-16], 它们各有优点, 不足之处在于预测分析过程中过多的侧重数学方法, 缺乏植保技术人员的参与和互动, 导致这些预测方法的应用受到不同程度的限制。本研究利用安徽省农作物病虫害数据库, 采用 CBR 方法, 让植保领域专家参与, 确定病害发生的影响因子、参数权重, 然后通过预测计算、检

验, 分别进行影响因子筛选和参数优化。其特点是利用历史病害发生的数据和气象数据进行现阶段与历史同期气象数据相似度比较, 实现对现阶段病害发生程度的预测; 通过数理统计分析, 获得不同置信度下的异常病害发生程度信息; 利用可视化技术实现对数据的直观多维度多视角分析; 充分利用数据库的数据和知识, 和植保部门的专家经验, 实现病虫害发生程度的异地在线会商, 同时结合预测平台的优势, 开展以时间序列为主轴的滚动预报, 结合病原菌的发育和侵染特点, 对气象数据的时间量纲进行优化匹配, 不易量化的影响因子如: 菌量、肥力、抗感性等, 发挥当地专业人员的作用, 以页面输入的形式参与预测, 因此, 该预测模型具有较高的应用潜力。

小麦赤霉病的发生程度与气象条件关系密切。以菌丝体越冬的小麦赤霉病菌于 3-4 月开始产生子囊壳, 小麦抽穗扬花前后成熟并释放子囊孢子, 侵染小穗, 麦收后病菌又以腐生方式在植物残体上越冬或继续侵染其他寄主, 成为翌年主要的侵染源。不同时间段的气象因子对病害发生程度的影响有所不同, 因此, 本文采取分段预测的方法, 赋予不同时间段的预测因子以不同的权重, 对相关预测参数进行了筛选、优化, 使该预测模型更具科学性。优化后的参数经回检, 常规年份的符合率可达 84.21%, 用于病害的实际预测具有一定的可行性。

病虫害预测的时效期越短, 预测的准确性也越高。小麦赤霉病发生关键期气象因子对病害发生程度具有十分重要的影响, 在小麦赤霉病 CBR 预测模型中, 为提高对关键期气象因子的利用质量, 采用以时间序列为主轴的滚动预报, 有效提高了病害预测的准确性。但小麦赤霉病的预测一直是个难题, 尤其是大发生年份的预测。本研究优化后的参数应用与小麦赤霉病常规发生年份的预测, 符合率较高, 但应用于大发生年份, 其符合率也仅为 62.5%, 因此, 如将此模型应用于大流行年份小麦赤霉病的预测, 其参数尚需进一步优化。

参考文献:

- [1] Audenaert K, Van Broeck R, Bekaer B, et al. Fusarium head blight (FHB) in Flanders: population diversity, inter-species associations and DON contamination in commercial winter wheat varieties [J]. *Eur J Plant Pathol*, 2009, 125: 445-458.
- [2] Korn U, Müller T, Ulrich A, et al. Impact of aggressiveness of *Fusarium graminearum* and *F.culmorum* isolates on yield parameters and mycotoxin production in wheat[J]. *Mycotox Res*, 2011, 27: 195-206.
- [3] Xu X M, Monger W, Riten A. Effect of temperature and duration of wetness during initial infection periods on disease development. Fungal biomass and mycotoxin concentrations on wheat inoculated with single or combinations of *Fusarium* species[J]. *Plant Pathology*, 2007, 56(6): 943-956.
- [4] 刘寿东, 王恩眷, 郭安红, 等. 2003 年江苏省小麦赤霉病大流行的天气条件分析[J]. *中国农学通报*, 2009, 25(23): 370-374.
- [5] Moschini R C and Fortugno C. Predicting wheat head blight incidence using models based on meteorological factors in Pergamino, Argentina[J]. *European Journal of Plant Pathology*, 1996, 102: 211-218.
- [6] Madgwick J W, West J S, White R P, et al. Impacts of climate change on wheat anthesis and fusarium ear blight in the UK[J]. *Eur J Plant Pathol*, 2011, 130: 117-131.
- [7] 梁振中, 黄建森. 小麦赤霉病分段预测模型的研究[J]. *植物保护学报*, 1991, 18(2): 117-122.
- [8] 关海鸥, 许少华, 左豫虎, 等. BP 神经网络算法的一种改进及在小麦赤霉病预测中的应用[J]. *黑龙江八一农垦大学学报*, 2009, 21(3): 87-90.
- [9] 李军, 蒋耀培, 蒋建, 等. 小麦赤霉病发生程度的气象动态预测模型[J]. *西南农业大学学报*, 2004, 26(4): 776-780.
- [10] 霍治国, 姚彩文, 姜瑞中, 等. 我国小麦赤霉病最大熵谱预报模式研究[J]. *植物病理学报*, 1996, 26(2): 117-122.
- [11] 旭晖, 高苹, 居为民, 等. 小麦赤霉病气象等级预报模式研究[J]. *安徽农业科学*, 2008, 36(23): 10030-10032.
- [12] Lee J K, Lee N. Least modification principle for case-based reasoning: a software project planning experience[J]. *Expert Systems with Applications*, 2006, 30(2): 190-202.
- [13] 杨振刚, 邓飞其. CBR 中案例相似性测度研究[J]. *计算机应用与软件*, 2008, 25(6): 222-223.
- [14] 安继芳, 孙振元. 案例推理式园林植物病虫害诊断系统的实现[J]. *计算机时代*, 2010(2): 20-21.
- [15] 李祚泳, 彭荔红. 基于人工神经网络的农业病虫害预测模型及其效果检验[J]. *生态学报*, 1999, 19(5): 759-762.
- [16] 陈怀亮, 张宏, 李有. 农作物病虫害发生发展气象条件及预报方法研究综述[J]. *中国农业气象*, 2007, 28(2): 212-216.