

Árvores de Decisão Aplicadas na Análise e no Alerta da Ferrugem do Cafeeiro



*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Informática Agropecuária
Ministério da Agricultura, Pecuária e Abastecimento*

Boletim de Pesquisa e Desenvolvimento 22

Árvores de decisão aplicadas na análise e no alerta da ferrugem do cafeeiro

Carlos Alberto Alves Meira
Luiz Henrique Antunes Rodrigues

Embrapa Informática Agropecuária
Campinas, SP
2009

Embrapa Informática Agropecuária

Av. André Tosello, 209 - Barão Geraldo
Caixa Postal 6041 - 13083-886 - Campinas, SP
Fone: (19) 3211-5700 - Fax: (19) 3211-5754
www.cnptia.embrapa.br
sac@cnptia.embrapa.br

Comitê de Publicações

Presidente: *Silvia Maria Fonseca Silveira Massruhá*

Membros: *Poliana Fernanda Giachetto, Roberto Hiroshi Higa, Stanley Robson de Medeiros Oliveira, Marcia Izabel Fugisawa Souza, Adriana Farah Gonzalez, Neide Makiko Furukawa, Suzilei Almeida Carneiro*

Membros suplentes: *Alexandre de Castro, Fernando Attique Máximo, Paula Regina Kuser Falcão, Maria Goretti Gurgel Praxedes*

Supervisor editorial: *Neide Makiko Furukawa, Suzilei Almeida Carneiro*

Revisor de texto: *Adriana Farah Gonzalez*

Normalização bibliográfica: *Maria Goretti Gurgel Praxedes*

Editoração eletrônica: *Neide Makiko Furukawa*

Secretária: *Suzilei Almeida Carneiro*

1ª edição on-line 2009

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei no 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Informática Agropecuária

Meira, Carlos Alberto Alves

Árvores de decisão aplicadas na análise e no alerta da ferrugem do cafeeiro / Carlos Alberto Alves Meira, Luiz Henrique Antunes Rodrigues. - Campinas : Embrapa Informática Agropecuária, 2009.

24 p. : il. - (Boletim de pesquisa e desenvolvimento / Embrapa Informática Agropecuária, ISSN 1677-9266; 22).

1. Mineração de dados. 2. Inteligência artificial. 3. Modelagem. 4. Café. 5. Doença de planta. I. Rodrigues, Luiz Henrique Antunes. II. Título. III. Série.

CDD: 006.3 (21.ed.)

© Embrapa 2009

Sumário

Resumo	5
Abstract	6
Introdução	7
Material e Métodos	8
Resultados e Discussão	12
Análise das epidemias da ferrugem do cafeeiro	12
Modelos de alerta da ferrugem do cafeeiro.....	16
Conclusões	21
Agradecimentos	22
Referências	22

Árvores de decisão aplicadas na análise e no alerta da ferrugem do cafeeiro

Carlos Alberto Alves Meira¹

Luiz Henrique Antunes Rodrigues²

Resumo

A ferrugem do cafeeiro é a principal doença da cultura do café. O conhecimento dos fatores que condicionam as epidemias de ferrugem e a sua previsão são importantes. Uma instância do processo de descoberta de conhecimento em bases de dados foi realizada para avaliar a aplicação de árvores de decisão na análise e no alerta da ferrugem. As classes do atributo meta foram definidas por intervalos da taxa de progresso da doença. Dados meteorológicos, a carga pendente de frutos e o espaçamento entre plantas serviram de atributos preditivos. As árvores de decisão obtidas auxiliaram na compreensão de quais variáveis, e como as interações dessas variáveis, conduziram a ferrugem no campo. O modelo de alerta para lavouras com alta carga pendente de frutos apresentou bom desempenho e pode ajudar na tomada de decisão referente ao controle da ferrugem do cafeeiro.

Termos para indexação: *Coffea arabica*, *Hemileia vastatrix*, modelos, predição de doenças de plantas, descoberta de conhecimento em bases de dados, classificação.

¹ Doutor em Engenharia Agrícola, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo, 13083-886, Campinas, SP, e-mail: carlos@cnptia.embrapa.br

² Phd em Engenharia Agrícola, Professor da Faculdade de Engenharia Agrícola da Universidade Estadual de Campinas - Unicamp, Caixa Postal 6011, Barão Geraldo, 13083-875, Campinas, SP, e-mail: lique@feagri.unicamp.br

Decision trees applied in the analysis and warning of coffee rust

Abstract

The coffee rust is the main disease of this crop. Knowledge about the factors that influence coffee rust epidemics and its prediction are important. A process instance of knowledge discovery in databases was performed to evaluate the use of decision trees in the analysis and warning of coffee rust. The target attribute classes were defined by ranges of the disease progress rate. Meteorological data, fruit load and space between plants were used as predictive attributes. The generated decision trees helped to understand which variables and how the interactions between these variables conducted the coffee rust in the field. The warning model for growing areas with large fruit load showed good performance and can aid the decision making related to coffee rust control.

Index terms: *Coffea arabica, Hemileia vastatrix, models, plant disease prediction, knowledge discovery in databases – KDD, classification.*

Introdução

A ferrugem, causada pelo fungo *Hemileia vastatrix*, é a principal doença do cafeeiro em todo o mundo, ocasionando decréscimos significativos na produção de café (ZAMBOLIM et al., 1997). O conhecimento dos fatores que determinam as epidemias da ferrugem é fundamental, e o desenvolvimento de modelos de alerta se justifica não apenas pelos prejuízos econômicos, mas pela variação na intensidade da doença entre os anos agrícolas e pela disponibilidade de medidas de controle economicamente viáveis.

Na epidemiologia de doenças de plantas, é comum o uso de regressão, a fim de se obter equações para prever uma ou mais características das epidemias, como é o caso para a ferrugem do cafeeiro (KUSHALAPPA et al., 1984; MORAES et al., 1976). Estudos mais recentes empregaram a análise de trilha (SILVA-ACUÑA et al., 1998) e redes neurais (PINTO et al., 2002) na descrição de epidemias da ferrugem.

A indução de árvores de decisão é uma técnica de modelagem alternativa. Árvores de decisão são de interesse para a descoberta de conhecimento em bases de dados ou KDD (*Knowledge Discovery in Databases*) devido à sua representação simbólica e interpretável (FAYYAD et al., 1996), permitindo a compreensão das fronteiras de decisão que existem nos dados e da lógica implícita neles (APTE; WEISS, 1997). O propósito da indução de árvores de decisão pode ser produzir modelos de predição precisos ou descobrir a estrutura preditiva do problema (BREIMAN et al., 1984). Esses dois propósitos não são excludentes, podendo aparecer juntos em um mesmo estudo.

Redes neurais, embora possam ter alta precisão, são difíceis de compreender quando comparadas com as árvores de decisão (FAYYAD et al., 1996). A indução de árvores de decisão não é prejudicada por multicolinearidade entre as variáveis independentes como na regressão (BUTT; ROYLE, 1990). Além disso, diversas variáveis, numéricas ou categóricas, podem ser analisadas ao mesmo tempo, sendo que o próprio algoritmo de indução se encarrega de selecionar as de maior importância.

A árvore de decisão é um modelo representado graficamente por nós e ramos, parecido com uma árvore, mas no sentido invertido (WITTEN;

FRANK, 2005). O nó raiz, no topo da estrutura, e os nós internos são nós de decisão. Cada um contém um teste sobre uma variável independente e os resultados desse teste formam os ramos da árvore. Os nós folhas, nas extremidades, representam valores de predição da variável dependente ou distribuições de probabilidade desses valores.

Paul e Munkvold (2004) usaram esse tipo de modelagem para prever categorias de severidade da cercosporiose do milho em estágio avançado do cultivo. Árvores de decisão também modelaram epidemias de giberela do trigo, procurando prever se a severidade da doença seria maior ou igual a 10% (MOLINEROS et al., 2005). Baker et al. (1993) desenvolveram uma árvore de decisão para prever o risco (alto ou baixo) de mortalidade de pínus em decorrência de podridão das raízes causada por *Heterobasidion annosum*.

Consideradas as vantagens mencionadas das árvores de decisão em relação às outras técnicas de modelagem e verificadas algumas de suas aplicações na fitopatologia, o objetivo deste trabalho foi aplicar e avaliar o potencial da indução de árvores de decisão para a análise de epidemias da ferrugem do cafeeiro e para o desenvolvimento de modelos de alerta desta doença.

Material e Métodos

Os dados brutos, coletados por Japiassú et al. (2007), utilizados na indução dos modelos em árvore de decisão, vieram do acompanhamento mensal da incidência da ferrugem do cafeeiro na fazenda experimental da Fundação Procafé, em Varginha, MG, de out/1998 a out/2006. Em setembro de cada ano, foram selecionadas duas lavouras de café adultas em produção, com idade entre 6 e 20 anos, para cada combinação de espaçamento (lavoura adensada e larga) e de carga pendente de frutos (alta e baixa). Dessas duas lavouras, uma foi da cultivar Catuaí e a outra da cultivar Mundo Novo. Não houve controle da doença nos oito talhões escolhidos. O método de amostragem da incidência, realizado no final de cada mês, foi aquele recomendado por Chalfoun (1997). Dados meteo-

rológicos foram registrados por uma estação meteorológica automática, instalada próximo dos locais de avaliação da incidência da ferrugem (Y).

A análise dos dados foi conduzida como um processo de descoberta de conhecimento em bases de dados (FAYYAD et al., 1996), de acordo com o modelo de processo de mineração de dados CRISP-DM (CHAPMAN et al., 2000). A instância do processo compreendeu as fases de compreensão do domínio, entendimento dos dados, preparação dos dados, modelagem e avaliação dos resultados obtidos (MEIRA, 2008).

Considerando um período de incubação (PI) estimado (MORAES et al., 1976), todo eventual dia de infecção (D_i) foi associado à taxa de progresso ou de infecção ($Y_i - Y_{i-1}$) para a qual possivelmente contribuiu (Figura 1), formando o período de infecção (PINF). Os atributos preditivos meteorológicos foram construídos para esses períodos de infecção, por meio de programas na linguagem de programação Perl (ActivePerl versão 5.8.7, ActiveState Corp.).

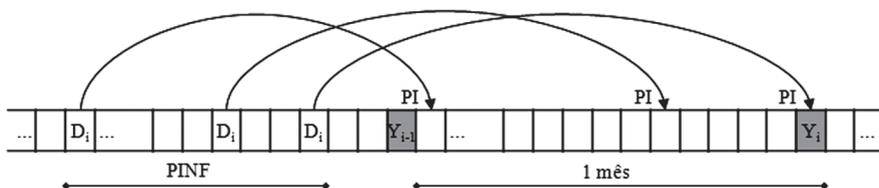


Figura 1. Esquema usado na preparação dos dados meteorológicos.

Fonte: Meira et al., 2008.

Os atributos meteorológicos foram construídos a partir do nível horário (registros da estação) até o nível do PINF, permitindo o seu relacionamento com a variável dependente ou atributo meta. No nível diário, além de médias e somatórias das variáveis meteorológicas, foram calculados valores estimados de molhamento foliar prolongado (mínimo de 6 horas), porque a germinação dos uredósporos de *H. vastatrix* só ocorre se a folha estiver molhada. O tempo mínimo necessário para a ocorrência de infecção foi avaliado em 6 horas de água livre na superfície da folha (KUSHALAPPA et al., 1983).

O número de horas com alta umidade relativa do ar (maior ou igual a 95%) foi utilizado como medida indireta de molhamento foliar contínuo.

Os períodos de molhamento foliar foram analisados tanto na extensão total como na fração noturna (das 20h às 8h), já que a infecção ocorre preferencialmente com pouca ou nenhuma luminosidade (MONTROYA; CHAVES, 1974). A temperatura média durante o período de molhamento foliar também foi considerada, uma vez que é o fator principal que determina o percentual de germinação dos esporos e de penetração enquanto a superfície da folha está molhada (KUSHALAPPA et al., 1983). O intervalo considerado como um dia se estendeu das 12h de um dia até 12h do dia seguinte, pois os períodos de molhamento ocorrem geralmente entre um dia e outro.

Os dias com precipitação maior ou igual a 1 mm foram considerados chuvosos segundo o mesmo critério usado por Kushalappa et al. (1983). A carga pendente de frutos e o espaçamento da lavoura completaram o conjunto dos atributos preditivos ou variáveis independentes (Tabela 1).

As classes do atributo meta foram definidas como intervalos da taxa de progresso da ferrugem. Para a análise das epidemias, foram definidas três classes: TX1(≤ 0), TX2($> 0 \leq 5$) e TX3(> 5). Para os modelos de alerta, o atributo meta foi binário: 1 para taxas maiores ou iguais a 5 pontos percentuais (p.p.) e 0, caso contrário – recomenda-se o limite de 5% de incidência para o controle da doença por via foliar (ZAMBOLIM et al., 1997). Os modelos de alerta foram específicos por carga pendente, aproveitando a característica bianual dos cafezais, que em anos de alta produção estão mais predispostos à ferrugem.

O conjunto de dados preparado totalizou 384 exemplos ou casos (8 anos x 12 meses x 4 combinações carga-espaçamento). Períodos de falha no registro da estação meteorológica motivaram a eliminação de 20 exemplos, encerrando o conjunto de dados para a modelagem, ou conjunto de treinamento, com 364 exemplos. Para os modelos de alerta, o conjunto de treinamento foi dividido em dois, resultando em 182 exemplos para cada opção de carga pendente.

As árvores de decisão foram geradas com a ferramenta “Decision Tree” do SAS® Enterprise Miner™ (versão 4.3, SAS Institute Inc.) e depois visualizadas e analisadas com a ferramenta SAS® Enterprise Miner™ Tree Desktop Application (versão 9.1.32, SAS Institute Inc.). O Enterprise Miner™ é a

Tabela 1. Variáveis independentes usadas na indução das árvores de decisão.

Variável	Tipo	Medida	Descrição
CARGA	binário	-	Carga pendente de frutos: ALTA ou BAIXA.
DCHUV_PINF	numérico	dias	Número de dias chuvosos (precipitação ≥ 1 mm) no PINF (período de infecção).
LAVOURA	binário	-	Espaçamento: lavoura ADENSADA ou LARGA.
MED_PRECIP_PINF	numérico	mm	Média das precipitações pluviométricas diárias no PINF.
NHNUR95_PINF	numérico	h	Média diária do número de horas noturnas (das 20h às 8h) com umidade relativa do ar $\geq 95\%$ no PINF.
NHUR95_PINF	numérico	h	Média diária do número de horas com umidade relativa do ar $\geq 95\%$ no PINF.
PRECIP_PINF	numérico	mm	Precipitação pluviométrica acumulada no PINF.
THUR95_PINF	numérico	°C	Temperatura média diária durante as horas com umidade relativa do ar $\geq 95\%$ no PINF.
TMAX_PINF	numérico	°C	Média das temperaturas máximas diárias no PINF.
TMAX_PI_PINF	numérico	°C	Média das temperaturas máximas diárias no período de incubação para os dias do PINF.
TMED_PINF	numérico	°C	Média das temperaturas médias diárias no PINF.
TMED_PI_PINF	numérico	°C	Média das temperaturas médias diárias no período de incubação para os dias do PINF.
TMIN_PINF	numérico	°C	Média das temperaturas mínimas diárias no PINF.
TMIN_PI_PINF	numérico	°C	Média das temperaturas mínimas diárias no período de incubação para os dias do PINF.
UR_PINF	numérico	%	Umidade relativa do ar média diária no PINF.

solução SAS para o processo de mineração de dados (SAS INSTITUTE, 2004).

A acurácia e a taxa de erro (WITTEN; FRANK, 2005) dos classificadores foram avaliadas. Essas medidas são estimativas dos percentuais de acertos e de erros do modelo na predição da classe de novos exemplos. Para os modelos de alerta, outras medidas de avaliação também foram calculadas (MONARD; BARANAUSKAS, 2002): sensibilidade, especificidade, confiabilidade positiva e confiabilidade negativa. Essas medidas de avaliação foram estimadas por validação cruzada, por meio de 10 partições aleatórias e estratificadas do conjunto de treinamento, conhecida como *10-fold crossvalidation* (WITTEN; FRANK, 2005).

Resultados e Discussão

Análise das epidemias da ferrugem do cafeeiro

O início da epidemia da ferrugem do cafeeiro, na média de todos os anos, foi no mês de dezembro e o pico no mês de junho, independente do espaçamento e da carga pendente de frutos. A partir de dezembro, as taxas de infecção atingiram níveis mais elevados, com o percentual de distribuição da classe TX3(>5) ultrapassando o das classes de menor nível (Figura 2). O percentual de distribuição das três classes de taxa de infecção, no conjunto de todos os meses do período analisado, foi 42% (154 exemplos) da classe TX1(≤ 0), 23% (82 exemplos) da classe TX2($> 0 \leq 5$) e 35% (128 exemplos) da classe TX3(>5).

A Figura 3 apresenta os três primeiros níveis da árvore de decisão que auxiliou na compreensão das epidemias da ferrugem do cafeeiro (MEIRA et al., 2008). As informações em cada nó da árvore representam, de cima para baixo, o número identificador do nó, a classe de taxa de infecção predominante e a distribuição percentual de cada classe no nó, na ordem TX1(≤ 0), TX2($> 0 \leq 5$) e TX3(>5).

O tom de cinza em cada nó indica a proporção de exemplos da classe TX3(>5) – mais escuro, maior a proporção, ou seja, melhores as condições

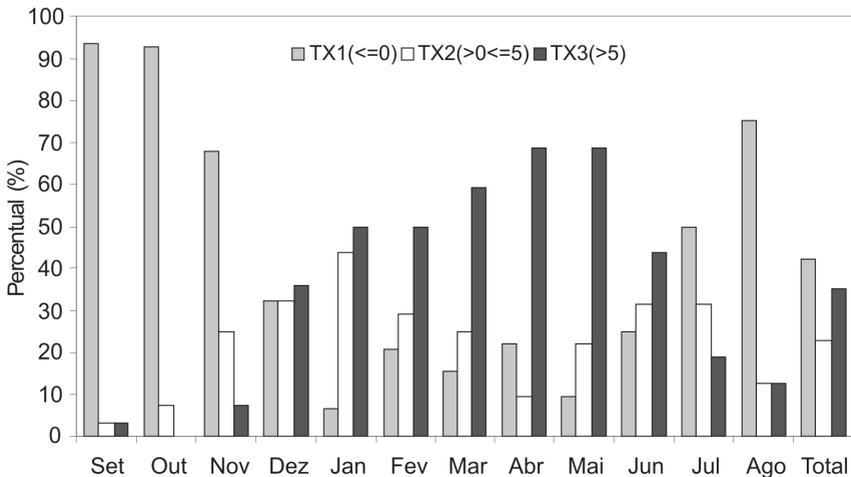


Figura 2. Distribuição percentual das três classes de taxa de infecção da ferrugem do cafeeiro. Fonte: Meira et al., 2008.

para um progresso mais acelerado da doença e vice-versa. Essa característica particular do Enterprise Miner, aliada à distribuição percentual entre as classes em cada nó, auxilia bastante na interpretação das regras da árvore de decisão por fitopatologistas especialistas do domínio.

A temperatura média nos períodos de molhamento foliar (THUR95_PINF) foi a primeira variável de teste na árvore de decisão (Figura 3, nó 1). Temperaturas inferiores a 17°C produziram taxas de infecção negativas ou nulas na maioria dos casos (73%), enquanto temperaturas maiores ou iguais a 17°C resultaram em taxas de infecção positivas na maior parte dos casos (28% de TX2(>0<=5) e 57% de TX3(>5)). Montoya e Chaves (1974) indicaram que o ponto mínimo de germinação de *H. vastatrix* seria encontrado em temperaturas inferiores a 18°C.

Temperaturas abaixo de 16°C foram mais desfavoráveis à infecção (Figura 3, nó 8). Kushalappa et al. (1983) consideraram 14°C como o limite mínimo de atividade do patógeno. O valor máximo de THUR95_PINF foi 20,3°C, abaixo da temperatura ótima de germinação estimada em 23,7°C (MONTAYA; CHAVES, 1974). Por essa razão, a árvore de decisão não deve ter identificado efeito negativo de temperaturas acima da ótima no poder germinativo de *H. vastatrix*.

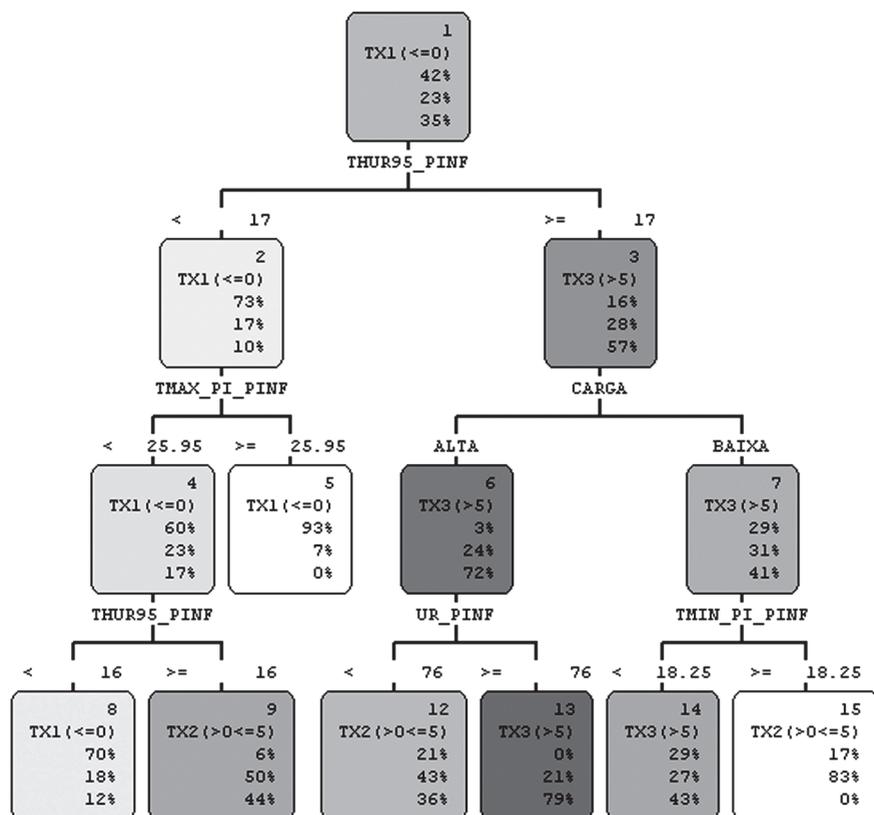


Figura 3. Três primeiros níveis da árvore de decisão que auxiliou na compreensão das epidemias da ferrugem do café.

Fonte: Meira et al., 2008.

A importância da temperatura durante o período de molhamento foliar no progresso da ferrugem do café é reconhecida na literatura (KUSHALAPPA; ESKES, 1989; ZAMBOLIM et al., 1997; ZAMBOLIM et al., 2002). Enquanto a superfície da folha está molhada, a temperatura é o fator principal que determina o percentual de germinação dos esporos e de penetração do agente etiológico da ferrugem (KUSHALAPPA et al., 1983). Na árvore de decisão, a temperatura durante o molhamento foliar, medida indiretamente pela temperatura média nos períodos de alta umidade relativa do ar (THUR95_PINF), foi a variável mais importante na determinação da classe de taxa de infecção. Além do teste no nó raiz, foi escolhida para outros dois testes nos níveis intermediários da árvore de decisão (MEIRA et al., 2008).

O próximo teste da árvore de decisão (Figura 3, nó 2) foi sobre a média das temperaturas máximas diárias no período de incubação (TMAX_PI_PINF). Temperaturas maiores ou iguais a 25,95°C resultaram em taxas de infecção negativas ou nulas (93% dos casos). Moraes et al. (1976) observaram que médias altas de temperatura máxima no período de incubação ocasionaram efeito depressivo sobre o desenvolvimento de *H. vastatrix*. Médias altas de temperatura mínima diária no período de incubação (TMIN_PI_PINF) também tiveram efeito negativo nas taxas de infecção (nó 15).

A decisão com base na carga pendente de frutos (Figura 3, nó 3) confirmou a predisposição das plantas à infecção de *H. vastatrix* devido à alta produção (KUSHALAPPA, 1989). Taxas de infecção em nível mais elevado ocorreram em cafeeiros com alta carga pendente, em comparação com os de baixa carga. Segundo Zambolim et al. (2002), quanto maior a produção, maiores a incidência e a severidade da ferrugem do cafeeiro.

Valores médios diários mais elevados de umidade relativa do ar (UR_PINF) corresponderam a níveis mais elevados da taxa de infecção (Figura 3, nó 13). A umidade relativa parece ter expressado melhor a importância das chuvas do que as variáveis independentes relacionadas com a precipitação (MEIRA et al., 2008). As estações chuvosas estão frequentemente associadas com alta umidade relativa do ar (KUSHALAPPA, 1989).

O espaçamento entre as plantas é considerado um fator de interferência no desenvolvimento da ferrugem do cafeeiro, provavelmente influenciando as condições microclimáticas dentro da lavoura (KUSHALAPPA; ESKES, 1989). Entretanto, o espaçamento das lavouras não se mostrou significativo na determinação da classe de taxa de infecção da ferrugem – a variável LAVOURA não apareceu em nenhum teste da árvore de decisão.

Os períodos de molhamento foliar prolongado (NHUR95_PINF e NHNUR95_PINF), presentes em praticamente todos os períodos de infecção, não serviram à árvore de decisão para identificar aqueles com maiores ou menores taxas de infecção. Isso pode estar relacionado à característica de que, no Brasil, é frequente a presença de água livre na superfície das folhas do cafeeiro, mesmo no inverno, estação seca, devido principalmente ao orvalho (KUSHALAPPA; ESKES, 1989). Sendo assim,

as baixas temperaturas, enquanto a folha está molhada, tornam-se o fator limitante para a germinação e a penetração do fungo.

Os nós 5 e 15 da Figura 3 são nós folhas da árvore de decisão completa que auxiliou na compreensão das epidemias da ferrugem do cafeeiro. O caminho de decisão entre o nó raiz e um nó folha pode ser traduzido para uma regra na forma 'SE <condição> ENTÃO <decisão>'. Por exemplo, o caminho até o nó 5 se traduz na regra 'SE (THUR95_PINF < 17°C) e (TMAX_PI_PINF >= 25,95°C) ENTÃO TAXA = TX1(<=0)'. Os outros nós da Figura 3 tiveram ramificações adicionais, até os demais nós folhas de predição da classe de taxa de infecção (MEIRA et al., 2008).

A árvore de decisão apresentou acurácia de 73% pela validação cruzada. O modelo classificou corretamente 283 exemplos dos 364 do conjunto de treinamento. Foram classificados corretamente 88% dos exemplos da classe TX1(<=0), 57% dos exemplos da classe TX2(>0<=5) e 79% dos exemplos da classe TX3(>5). O menor desempenho para a classe TX2(>0<=5) deve estar relacionado com o menor número de exemplos dessa classe no conjunto de treinamento (Figura 2), permitindo que as outras duas classes prevalecessem na distribuição dos exemplos nos nós folhas.

Modelos de alerta da ferrugem do cafeeiro

Os alertas da ferrugem do cafeeiro são considerados quando a taxa de progresso da doença, prevista para o prazo de um mês, atingir ou ultrapassar o limite de 5 p.p. Esse tipo de alerta pode ser útil na tomada de decisão de quais medidas e qual o melhor momento de aplicá-las no controle da doença. Alertas emitidos corretamente e situações acertadas em que os alertas não são emitidos permitem identificar os momentos oportunos para a aplicação de fungicidas.

O modelo de alerta para lavouras com alta carga pendente de frutos é apresentado na Figura 4. Os nós da árvore de decisão indicam o seu número identificador e a distribuição de probabilidade das duas classes. Eles são coloridos em tons de cinza com base na proporção de exemplos da classe 1 – quanto mais escuro, maior a proporção. O nó raiz (nó 1) indica a distribuição percentual dos exemplos entre as classes no conjunto

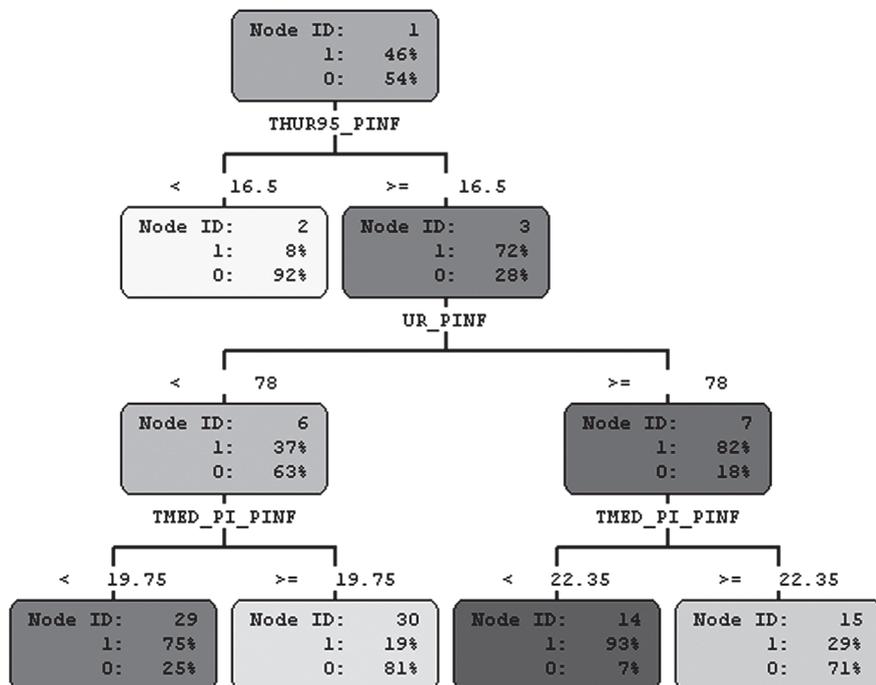


Figura 4. Árvore de decisão para alerta da ferrugem do cafeeiro em lavouras com alta carga pendente de frutos.

Fonte: Meira et al., 2009.

de treinamento: 46% (84 exemplos) da classe 1 e 54% (98 exemplos) da classe 0. A classe de predição em um nó folha é a que apresenta a maior probabilidade.

Segundo a árvore de decisão, os alertas devem ser emitidos quando as condições das variáveis meteorológicas entre o nó raiz e os nós folhas 14 ou 29 forem satisfeitas. Por exemplo, o caminho até o nó 14 se traduz na regra 'SE (THUR95_PINF \geq 16,5°C) e (UR_PINF \geq 78%) e (TMED_PI_PINF < 22,35°C) ENTÃO TAXA = 1'. Caso contrário, é provável que a taxa de infecção da ferrugem do cafeeiro fique abaixo de 5 p.p.

Conforme mostra o modelo da Figura 4, as influências da temperatura e da umidade relativa do ar se revelaram mais importantes no progresso da ferrugem do cafeeiro. Temperaturas médias mais baixas durante o molha-

mento foliar (THUR95_PINF) foram desfavoráveis às taxas de infecção maiores ou iguais a 5 p.p. (nó 2). Temperaturas médias diárias no período de incubação (TMED_PI_PINF) mais elevadas também exerceram efeito negativo nas taxas de infecção da ferrugem do cafeeiro (nós 15 e 30). Umidade relativa média diária (UR_PINF) mais alta foi favorável às taxas de infecção maiores ou iguais a 5 p.p. (nó 7).

Dependendo das condições para a infecção, menos favoráveis (Figura 4, nó 6) ou mais favoráveis (nó 7), a influência da temperatura no período de incubação foi diferenciada. Meira et al. (2008) observaram que infecções ocorridas em condições menos favoráveis foram mais sensíveis ao efeito da temperatura durante o período de incubação.

O modelo da Figura 4 é simples e compacto, com cinco regras e com base apenas em três variáveis de teste. O modelo foi bem avaliado, com acurácia estimada em 81,3% e valores próximos a este para as outras medidas (Tabela 2). Comparando com outras aplicações de árvores de decisão como modelos de predição de doenças de culturas agrícolas, os resultados obtidos foram promissores, uma vez que os maiores valores de acurácia relatados não ultrapassaram 80% (PAUL; MUNKVOLD 2004; MOLINEROS et al. 2005).

A acurácia dá a noção da proporção de acertos que o modelo pode ter caso venha a ser aplicado no problema real. As outras medidas de avaliação também são importantes: a sensibilidade estima a capacidade do modelo de acertar nas situações em que se deve emitir um alerta; a con-

Tabela 2. Avaliação dos modelos de alerta da ferrugem do cafeeiro.

Medida de avaliação	Alta carga pendente	Baixa carga pendente
Acurácia	81,3±4,8%	71,8±2,5%
Taxa de erro	18,7±4,8%	28,2±2,5%
Sensibilidade	79,9±6,9%	46,3±6,9%
Especificidade	82,6±4,3%	82,1±2,1%
Confiabilidade positiva	79,4±6,1%	49,9±5,9%
Confiabilidade negativa	83,9±4,7%	79,4±2,1%

fiabilidade positiva, por sua vez, estima a capacidade do modelo de emitir corretamente os alertas; a especificidade e a confiabilidade negativa são correspondentes às duas primeiras, mas para as situações em que o alerta não é necessário ou não é emitido.

O modelo de alerta da ferrugem do cafeeiro para lavouras com baixa carga pendente de frutos é apresentado na Figura 5. A distribuição percentual dos exemplos entre as duas classes no conjunto de treinamento não foi balanceada, conforme indica o nó raiz (nó 1): 29% (53 exemplos) da classe 1 e 79% (129 exemplos) da classe 0. Os nós da árvore de decisão também são coloridos em tons de cinza com base na proporção de exemplos da classe 1. Os alertas devem ser emitidos quando as condições de uma entre cinco regras forem satisfeitas: regras correspondentes aos nós folhas 13, 16, 19, 23 ou 27. Caso contrário, é provável que a taxa de infecção fique abaixo de 5 p.p.

O primeiro atributo de teste da árvore de decisão foi o mesmo do modelo de alerta para lavouras com alta carga pendente, mas com o limite de decisão 1°C acima. Temperaturas médias diárias durante o molhamento foliar (THUR95_PINF) abaixo de 17,55°C foram menos favoráveis à infecção (Figura 5, nó 2), e maiores ou iguais a esse limite foram mais favoráveis às taxas de infecção maiores ou iguais a 5 p.p. (nó 3). A indução do modelo ramificou o nó à esquerda do nó raiz, utilizando a média diária da temperatura máxima no período de incubação (TMAX_PI_PINF) e a temperatura média diária no período de infecção (TMED_PINF). Essa ramificação identificou melhor os exemplos das duas classes, contribuindo para reduzir a taxa de erro do modelo.

A precipitação acumulada no período de infecção (PRECIP_PINF) apareceu com influência no modelo da Figura 5, mas esta variável sozinha não foi suficiente para indicar uma condição mais ou menos favorável à taxa de infecção da ferrugem do cafeeiro. Foi preciso incluir outras variáveis para compor as regras de decisão, como a temperatura média durante o molhamento foliar (THUR95_PINF), o espaçamento da lavoura (LAVOURA) e o molhamento foliar noturno (NHNUR95_PINF). A média do número de horas de molhamento foliar (NHUR95_PINF), com valores a partir de 14h, causou efeito depressivo nas taxas de infecção (nó 15). O modelo parece ter capturado efeito negativo do molhamento foliar sobre a doença, se-

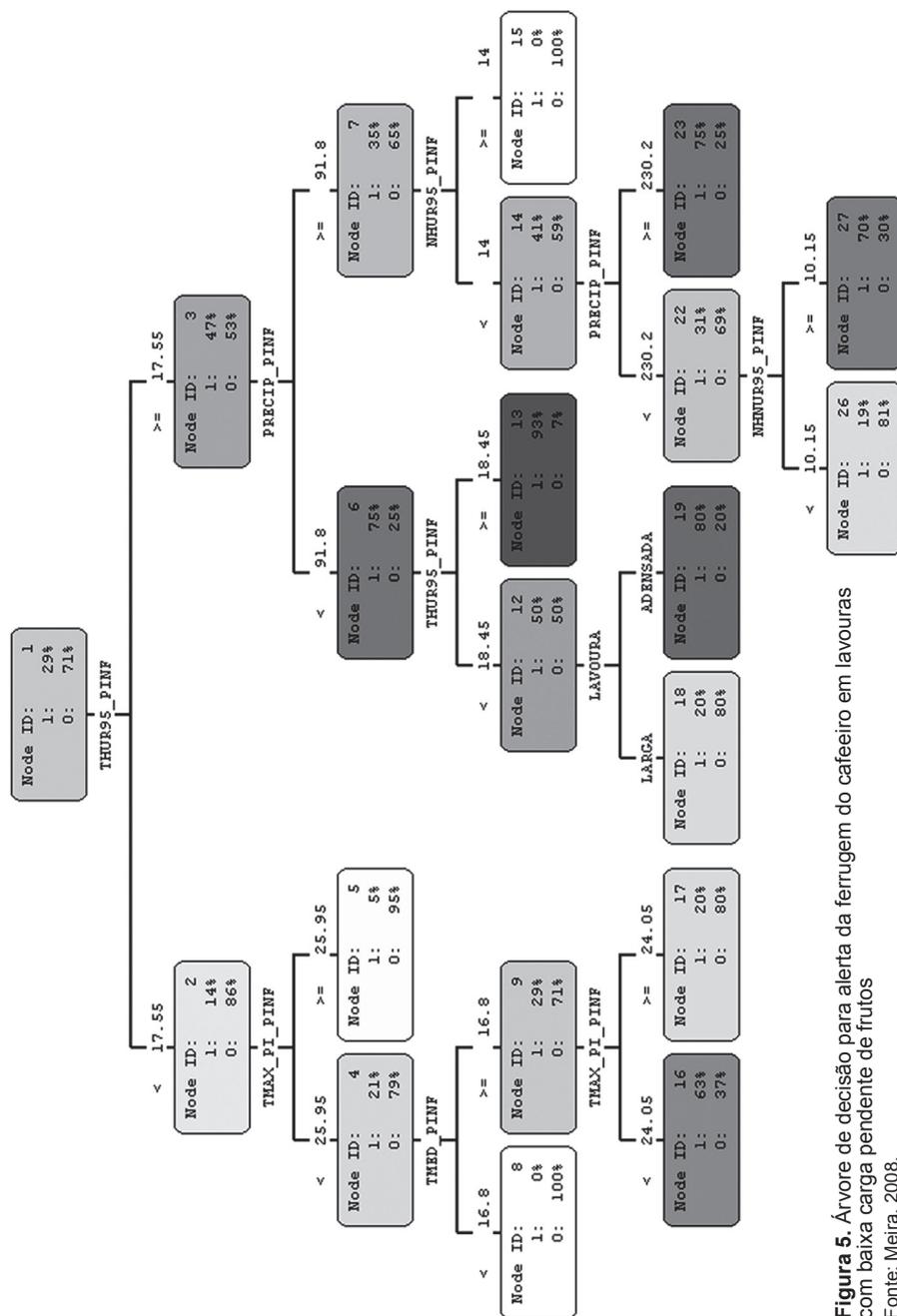


Figura 5. Árvore de decisão para alerta da ferramenta do café em lavouras com baixa carga pendente de frutos
Fonte: Meira, 2008.

melhante ao observado por Silva-Acuña et al. (1998), também em ano de baixa carga pendente de frutos.

O modelo da Figura 5, com 11 regras e 7 atributos de teste, não foi bem avaliado como o modelo de alerta para lavouras com alta carga pendente (Tabela 2). A acurácia foi de 71,8%, equivalente à acurácia de 71% de um classificador que atribuísse a classe majoritária 0 a todos os exemplos. Não houve equilíbrio entre as medidas de avaliação. A especificidade (82,1%) e a confiabilidade negativa (79,4%) foram melhores do que a sensibilidade (46,3%) e a confiabilidade positiva (49,9%), provavelmente em decorrência da distribuição desbalanceada dos exemplos entre as classes no conjunto de treinamento. Ou seja, apesar de predizer corretamente alguns alertas (classe 1), o modelo teve a tendência de predizer a classe 0 da maioria dos exemplos do conjunto de treinamento.

A principal limitação dos modelos de alerta desenvolvidos está relacionada com a sua abrangência. O uso desses modelos deve ficar restrito à região onde os dados foram coletados ou a regiões com clima parecido. Outras regiões podem apresentar condições meteorológicas que não foram representadas nos dados analisados e que podem condicionar o progresso da ferrugem do cafeeiro de maneira diferente da capturada pelos modelos. Cabe ressaltar ainda que, antes da adoção dos modelos apresentados, é importante que seja realizada uma etapa de validação desses modelos.

Conclusões

1. As árvores de decisão auxiliaram na compreensão de quais variáveis e como as interações dessas variáveis conduziram as epidemias da ferrugem do cafeeiro. As variáveis mais importantes foram a temperatura média nos períodos de molhamento foliar, a carga pendente de frutos, a média das temperaturas diárias no período de incubação e a umidade relativa do ar.
2. O modelo de alerta para lavouras com alta carga pendente de frutos apresentou bom desempenho e tem potencial para auxiliar na decisão

de quando e quais medidas devem ser tomadas no controle da doença no campo. Nos anos de alta carga pendente, não são recomendadas atomizações tardias, após a constatação de nível de incidência da ferrugem do cafeeiro maior do que 5%. O modelo de alerta para lavouras com baixa carga pendente não apresentou tão bom desempenho.

3. Os modelos em árvore de decisão facilitam a interpretação de suas regras e podem contribuir para a confiança do tomador de decisões em adotá-los como ferramenta de suporte na análise do risco de epidemias da ferrugem do cafeeiro.
4. A técnica de indução de árvores de decisão se mostrou adequada na modelagem proposta e pode ser uma boa alternativa para a epidemiologia de doenças de plantas.

Agradecimentos

À Fundação Procafé pela concessão dos dados. Ao SAS Brasil pela concessão da licença de uso do SAS Enterprise Miner por meio de seu Programa Acadêmico.

Referências

- APTE, C.; WEISS, S. Data mining with decision trees and decision rules. **Future Generation Computer Systems**. v. 13, n. 2-3, p. 197-210, Nov. 1997.
- BAKER, F. A.; VERBYLA, D. L.; HODGES, C. S.; ROSS, E. W. Classification and regression tree analysis for assessing hazard of pine mortality caused by *Heterobasidion annosum*. **Plant Disease**. St. Paul, v. 77, n. 2, p. 136-139, 1993.
- BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. **Classification and regression trees**. Boca Raton: CRC Press, 1984. 358 p.
- BUTT, D. J.; ROYLE, D. J. Multiple regression analysis in the epidemiology of plant diseases. In: KRANZ, J. (Ed.) **Epidemics of plant diseases: mathematical analysis and modeling**. 2nd ed. Berlin: Springer-Verlag, 1990. p. 143-180.

CHALFOUN, S. M. **Doenças do cafeeiro**: importância, identificação e métodos de controle. Lavras, MG: UFLA/FAEPE. 1997.

CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. **CRISP-DM 1.0**: step-by-step data mining guide. [Illinois]: SPSS, 2000. 78 p.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, Menlo Parck. v. 17, p. 37-54, 1996.

JAPIASSÚ, L. B.; GARCIA, A. W. R.; MIGUEL, A. E.; CARVALHO, C. H. S.; FERREIRA, R. A.; PADILHA, L.; MATIELLO, J. B. Influência da carga pendente, do espaçamento e de fatores climáticos no desenvolvimento da ferrugem do cafeeiro. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 5., 2007, Águas de Lindóia, SP. **Anais...** Brasília, DF: Embrapa Café, 2007. 1 CD-ROM.

KUSHALAPPA, A. C. Biology and epidemiology. In: KUSHALAPPA, A. C.; ESKES, A. B. (Ed.) **Coffee rust**: epidemiology, resistance, and management. Boca Raton: CRC Press, 1989. p. 13-80.

KUSHALAPPA, A. C.; ESKES, A. B. Advances in coffee rust research. **Annual Review of Phytopathology**, Palo Alto. v. 27, p. 503-531, 1989.

KUSHALAPPA, A. C.; AKUTSU, M.; OSEGUERA, S. H.; CHAVES, G. M.; MELLES, C. Equations for predicting the rate of coffee rust development based on net survival ratio for monocyclic process of *Hemileia vastatrix*. **Fitopatologia Brasileira**, Brasília, DF. v. 9, p. 255-271, jun. 1984.

KUSHALAPPA, A. C.; AKUTSU, M.; LUDWIG, A. Application of survival ratio for monocyclic process of *Hemileia vastatrix* in predicting coffee rust infection rates. **Phytopathology**, St. Paul. v. 73, p. 96-103, 1983.

MEIRA, C. A. A.; RODRIGUES, L. H. A.; MORAES, S. A. Modelos de alerta para o controle da ferrugem-do-cafeeiro em lavouras com alta carga pendente. **Pesquisa Agropecuária Brasileira**, Brasília, DF. v. 44, p. 233-242, 2009.

MEIRA, C. A. A.; RODRIGUES, L. H. A.; MORAES, S. A. Análise da epidemia da ferrugem do cafeeiro com árvore de decisão. **Tropical Plant Pathology**, Brasília, DF. v. 33, p. 114-124, 2008.

MEIRA, C. A. A. **Processo de descoberta de conhecimento em bases de dados para a análise e o alerta de doenças de culturas agrícolas e sua aplicação na ferrugem do cafeeiro**. 2008. 198 p. Tese (Doutorado), Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas.

MOLINEROS, J. E.; DE WOLF, E. D.; FRANCL, L.; MADDEN, L.; LIPPS, P. Modeling epidemics of fusarium head blight: trials and tribulations. **Phytopathology**, St. Paul, v. 95:S71. 2005.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. In: REZENDE, S. O. (Org.). **Sistemas inteligentes**: fundamentos e aplicações. Barueri: Ed. Manole, 2002. p. 89-114.

MONTOYA, R. H.; CHAVES, G. M. Influência da temperatura e da luz na germinação, infectividade e período de geração de *Hemileia vastatrix* Berk. & Br. **Experientiae**, Viçosa, MG, v. 18, p. 239-266, 1974.

MORAES, S. A.; SUGIMORI, M. H.; RIBEIRO, I. J. A.; ORTOLANI, A. A.; PEDRO JR., M. J. Período de incubação de *Hemileia vastatrix* Berk. et Br. em três regiões do Estado de São Paulo. **Summa Phytopathologica**, Botucatu, v. 2, p. 32-38, 1976.

PAUL, P. A.; MUNKVOLD, G. P. A model-based approach to preplanting risk assessment for gray leaf spot of maize. **Phytopathology**, St. Paul, v. 94, p. 1350-1357, 2004.

PINTO, A. C. S.; POZZA, E. A.; SOUZA, P. E.; POZZA, A. A. A.; TALAMINI, V.; BOLDINI, J. M.; SANTOS, F. S. Descrição da epidemia da ferrugem do cafeeiro com redes neuronais. **Fitopatologia Brasileira**, Brasília, DF, v. 27, p. 517-524, 2002.

SAS INSTITUTE. **Getting started with SAS® Enterprise Miner™ 4.3**. Cary, NC: SAS Institute Inc., 2004. 126 p.

SILVA-ACUÑA, R.; ZAMBOLIM, L.; CRUZ, C. D.; VALE, F. X. R. Estudo epidemiológico da ferrugem do cafeeiro (*Hemileia vastatrix*) utilizando a análise de trilha. **Fitopatologia Brasileira**, Brasília, DF, v. 23, p. 425-430, 1998.

WITTEN, I. H.; FRANK, E. **Data mining: practical machine learning tools and techniques**. 2nd ed. San Francisco: Morgan Kaufmann, 2005. 525 p.

ZAMBOLIM, L.; VALE, F. X. R.; COSTA, H.; PEREIRA, A. A.; CHAVES, G. M. Epidemiologia e controle integrado da ferrugem-do-cafeeiro. In: ZAMBOLIM, L. (Ed.). **O estado da arte de tecnologias na produção de café**. Viçosa, MG: UFV, 2002. p. 369-449.

ZAMBOLIM, L.; VALE, F. X. R.; PEREIRA, A. A.; CHAVES, G. M. Café (*Coffea arabica* L.): controle de doenças – doenças causadas por fungos, bactérias e vírus. In: VALE, F.X.R.; ZAMBOLIM, L. (Ed.). **Controle de doenças de plantas: grandes culturas**. Viçosa, MG: UFV, 1997. v. 1, p. 83-139.



Informática Agropecuária

Ministério da
Agricultura, Pecuária
e Abastecimento

