

Regras para classificação de ocorrências de ferrugem asiática da soja com base em variáveis meteorológicas

Guilherme A. S. Megeto¹, Stanley R. de M. Oliveira¹, Carlos A. A. Meira¹

¹ Embrapa Informática Agropecuária
Caixa Postal 6041, CEP 13083-970 – Campinas – SP – Brazil

guimegeto@gmail.com, stanley.oliveira@embrapa.br,
carlos.meira@embrapa.br

Abstract. *Asian soybean rust is one of the most important and aggressive diseases of soybean, especially in Brazil. The use of mathematical and computational models can assist decision making on the need and timing of an action for prevention of disease. Decision tree models based on temperature and precipitation were developed and rules were extracted to identify the main factors that could influence the occurrence of soybean rust. The final model showed to be interpretable and consistent with existing literature.*

Resumo. *A ferrugem asiática é uma das doenças mais importantes e agressivas da soja, especialmente no Brasil. O uso de modelos matemáticos e computacionais podem auxiliar a tomada de decisão do produtor sobre a necessidade e o momento de uma ação de prevenção à doença. Modelos em árvore de decisão baseados em temperatura e precipitação foram desenvolvidos e extraídas regras que permitiram identificar os principais fatores que influenciaram a ocorrência da ferrugem em lavouras comerciais. O modelo final se mostrou interpretável e coerente com a literatura existente.*

1. Introdução

A ferrugem asiática, causada pelo fungo *Phakopsora pachyrhizi* Syd. & P. Syd., é considerada uma das doenças mais importantes da soja no Brasil, sendo que a principal forma de controle é feita por meio de aplicações de fungicidas [Del Ponte e Esker, 2008]. Entretanto, as aplicações têm sido realizadas de forma calendarizada, sem levar em conta os fatores de risco [Godoy et al., 2009], prática que pode resultar em aplicações desnecessárias ou, também, diminuir a eficiência do controle, caso elas se iniciem com atraso. Assim, modelos de alerta ou apoio à tomada de decisão podem ser uma alternativa para auxiliar o produtor ao indicar a necessidade ou o momento de realizar práticas de prevenção e combate à doença em uma possibilidade de epidemia.

No caso da ferrugem asiática, já foram desenvolvidos modelos utilizando duas abordagens: i) simulação, com o objetivo de reproduzir os processos biológicos da doença; e ii) empírica, com base em relações estatísticas entre dados experimentais. A maior parte dos modelos foi desenvolvida utilizando dados provenientes de ambientes controlados ou um pequeno número de observações, além de utilizar técnicas tradicionais de modelagem [Del Ponte et al., 2006a]. Estudos sobre componentes específicos do ciclo da doença proporcionaram um conhecimento básico e importante sobre a epidemiologia da ferrugem asiática da soja, porém alguns dos modelos

desenvolvidos utilizaram variáveis com informações muito específicas sobre o hospedeiro ou o patógeno, o que pode dificultar a interpretação e a aplicação dos modelos [Kim et al., 2005]. Além disso, as interações entre os diversos fatores presentes em situação de campo podem ser mais complexas do que as simuladas em ambientes controlados [Del Ponte e Esker, 2008], fazendo com que sejam necessários mais estudos com dados provenientes de situações naturais.

Mesmo com esforços na modelagem a partir de dados de campo e uma literatura representativa sobre a epidemiologia da ferrugem asiática da soja, ainda existem lacunas no conhecimento, assim como dados e técnicas de modelagem ainda não explorados. Com o aumento da obtenção de dados e da capacidade computacional, técnicas de mineração de dados vêm se apresentando como complementares e alternativas às técnicas tradicionais.

Uma das técnicas mais utilizadas em mineração de dados é a indução de árvores de decisão, pois sua representação simbólica e interpretável pode ser utilizada tanto para predição quanto para conhecer a estrutura preditiva do problema [Witten et al. 2011]. Meira et al. (2009) utilizaram essa técnica para desenvolver modelos de alerta da ferrugem-do-cafeeiro em lavouras de café com alta carga pendente de frutos, sendo capazes de prever, com 81% de taxa de acerto, a taxa de infecção igual ou superior a 5 pontos percentuais para o prazo de um mês, utilizando informações meteorológicas e o espaçamento entre plantas.

O objetivo deste trabalho foi identificar situações favoráveis ou não à ocorrência da ferrugem asiática da soja por meio de regras de classificação extraídas de modelos em árvore de decisão obtidos a partir de dados de ocorrências da doença em situação de campo no Brasil e de variáveis derivadas de temperatura e precipitação.

2. Material e métodos

2.1. Bases de dados

Os dados de ocorrência da ferrugem da soja foram obtidos junto ao Consórcio Antiferrugem. Foram utilizadas as ocorrências da doença relatadas em lavouras comerciais entre as safras 2007/08 e 2010/11 (4 safras), totalizando 7.810 registros, incluindo-se ocorrências múltiplas para um mesmo município em uma mesma safra.

Os dados meteorológicos, de temperaturas mínima e máxima e de precipitação, em escala diária, foram obtidos junto ao sistema Agritempo (Sistema de Monitoramento Agrometeorológico) a partir de estações virtuais [Romani et al., 2007]. Essas estações têm por principal objetivo suprir a falta de dados reais (medidos) ao interpolar dados de temperatura de estações agrometeorológicas físicas mais próximas ao município definido e utilizar dados de precipitação provenientes do radar a bordo do satélite TRMM - *Tropical Rainfall Measuring Mission*.

2.2. Variáveis dependente e preditoras

Para a definição da variável dependente (atributo meta ou classe), cada registro da doença foi considerado como uma ocorrência verdadeira (confirmada pelo laboratório), rotulado como “Oc” e relacionado diretamente à data da ocorrência. Em seguida, para cada uma das ocorrências, foi criado um registro correspondente de não-ocorrência (“NaoOc”). Esta classe foi definida como sendo o trigésimo dia anterior ao dia do relato da ocorrência, sendo assumida a presença do inóculo, porém com condições ambientais desfavoráveis à doença.

A partir das temperaturas mínima e máxima e da precipitação, foram derivadas nove variáveis preditoras (Tabela 1). Cada uma das variáveis foi avaliada em cinco janelas de tempo, sendo o período latente e 5, 10, 15 e 20 dias anteriores à possível data de infecção. Neste trabalho, o período latente foi o número de dias entre a detecção da ocorrência e a possível data de infecção, calculado pela equação descrita em Alves et al. (2006). O conjunto resultante totalizou 12.591 registros de ocorrência e não-ocorrência com 45 variáveis meteorológicas preditoras.

Tabela 1. Variáveis preditoras derivadas de temperatura e precipitação

| Variável | Descrição |
|----------|---|
| TMIN | Média da temperatura mínima. |
| DTMIN15 | Número de dias com temperatura mínima menor que 15°C. |
| TMAX | Média da temperatura máxima. |
| DTMAX30 | Número de dias com temperatura máxima maior que 30°C. |
| TMED | Média da temperatura média diária. |
| PMED | Média da precipitação (em mm). |
| PTOTAL | Precipitação (em mm) total ou acumulada. |
| DP1 | Número de dias com precipitação maior do que 1mm. |
| PMAX | O máximo evento de precipitação pluvial (em mm). |

Os dados foram preparados com rotinas em linguagem de programação *Python* e no software R. As árvores de decisão foram induzidas no software WEKA 3.6, por meio do classificador J48. Foram realizados diversos experimentos, variando-se o conjunto de treinamento (com e sem a presença de *outliers*) e um dos parâmetros de poda do classificador J48, com o objetivo de obter modelos mais genéricos e interpretáveis. A remoção de *outliers* foi feita com o filtro *RemoveMissclassified*, dado que, em geral, a presença destes aumenta a complexidade do modelo [Witten et al., 2011]. Foi realizada a pré-poda das árvores com a variação do parâmetro “minNumObj”, responsável por parar o desenvolvimento do modelo quando se atinge um número mínimo de registros que estão sendo avaliados.

Os modelos induzidos a partir do conjunto de treinamento original, com *outliers*, foram avaliados utilizando o método de validação cruzada em 10 partes, considerando a taxa de acerto (acurácia) e o número de folhas (número de regras). Os modelos induzidos utilizando o conjunto de treinamento reduzido, sem possíveis *outliers*, tiveram suas regras extraídas na forma de expressões condicionais, aplicadas sobre o conjunto original e avaliadas segundo a taxa de acerto e cobertura de cada regra.

3. Resultados e Discussão

O modelo em árvore de decisão escolhido para interpretação foi obtido a partir do conjunto de treinamento reduzido, sem a presença de possíveis *outliers*, com 7.129 registros. O classificador utilizado no filtro *RemoveMissclassified* e na fase de modelagem foi o J48, com os parâmetros *default*, exceto pelo número mínimo de objetos por folha (“minNumObj”), sendo utilizado 40. Esse modelo apresentou 28 regras e foi avaliado utilizando o conjunto original (12.591 registros), obtendo uma taxa de acerto de 73,4%. Com essa avaliação, o modelo como um todo pode não ser

adequado para utilizar como modelo de alerta ou para predição, mas regras específicas podem ser avaliadas e interpretadas. Nesse caso, todas as regras foram extraídas e aplicadas sobre o conjunto original, avaliando-se suas respectivas cobertura e taxa de acerto. As seis melhores regras estão na Tabela 2.

Tabela 2 – Principais regras extraídas do modelo em árvore de decisão.

| nº | Regras em destaque | Classe | Cobertura (%) | Acerto (%) |
|----|--|--------|---------------|------------|
| 1 | $DTMIN15_{20} \leq 0 \ \& \ DTMAX30_{20} \leq 3$ | Oc | 14,1 | 82,8 |
| 2 | $DTMIN15_{20} \leq 0 \ \& \ DTMAX30_{20} > 3 \ \& \ TMIN_{15} \leq 18,3$ | NaoOc | 6,1 | 77,4 |
| 3 | $DTMIN15_{20} \leq 0 \ \& \ DTMAX30_{20} > 3 \ \& \ TMIN_{15} > 18,3 \ \& \ TMAX_{15} \leq 31,89 \ \& \ DTMAX30_L \leq 0$ | Oc | 6,1 | 74,6 |
| 4 | $DTMIN15_{20} \leq 0 \ \& \ DTMAX30_{20} > 3 \ \& \ TMAX_{15} \leq 31,89 \ \& \ 0 < DTMAX30_L \leq 6 \ \& \ TMIN_L > 18,7 \ \& \ DTMAX30_{15} > 4 \ \& \ TMIN_{15} > 19,93 \ \& \ TMIN_5 \leq 21,91 \ \& \ TMED_{15} \leq 26,12 \ \& \ 28,96 < TMAX_{20} \leq 30,83$ | Oc | 7,7 | 74,7 |
| 5 | $DTMIN15_{20} > 0 \ \& \ PMED_L \leq 17,67 \ \& \ TMIN_{10} \leq 18,34 \ \& \ TMAX_{15} > 26,12 \ \& \ DP1_{20} \leq 11 \ \& \ TMAX_{20} > 27,15$ | NaoOc | 7,0 | 80,0 |
| 6 | $DTMIN15_{20} > 0 \ \& \ PMED_L \leq 17,67 \ \& \ TMIN_{10} > 18,34$ | NaoOc | 10,2 | 80,6 |

Especialistas em epidemiologia de doenças de plantas auxiliaram o processo de modelagem e interpretação do modelo. Os principais resultados englobam a avaliação e interpretação das melhores regras. O nó raiz da árvore de decisão é considerado a variável mais importante para o modelo, pois é a que melhor explica o conjunto, considerando a classe, utilizando o conceito de ganho de informação [Witten et al., 2011]. No caso, foi a variável $DTMIN15_{20}$, que representa o número de dias com temperatura mínima abaixo de 15 °C em 20 dias antes da possível data de infecção, separando os 12.591 registros em 72,6% para $DTMIN15_{20} \leq 0$ e 27,4% para $DTMIN15_{20} > 0$.

A primeira regra da Tabela 2 cobriu cerca de 1.775 registros (14,1% de 12.591 registros), classificando-os como “Oc” e obtendo um acerto de 82,8%, que pode caracterizar condições importantes para a ocorrência da doença. A interpretação é que a inexistência de dias com temperatura mínima abaixo de 15 °C em 20 dias antes da possível infecção e com no máximo 3 dias com temperaturas acima de 30 °C nesse mesmo período, pode favorecer a ocorrência da doença. Segundo Del Ponte e Esker (2008), o processo de germinação do fungo só ocorre com a presença de água livre na superfície da folha, porém, quando a umidade não é limitante, a germinação é afetada principalmente pela temperatura.

Embora não haja um consenso absoluto sobre uma faixa fixa de temperatura

considerada ótima para a germinação, a infecção e o desenvolvimento da doença, temperaturas abaixo de 15 e acima de 30 °C podem ser consideradas como desfavoráveis. Alves et al. (2006) encontraram que o fungo foi capaz de germinar em temperaturas entre 8 e 30 °C, sendo a temperatura ótima entre 15 e 25 °C. Bonde et al. (2007) coletaram isolados de diferentes partes do mundo durante 21 anos e concluíram que não há diferença significativa no processo de germinação e infecção entre eles, sendo que a temperatura ótima para os processos infecciosos ficou entre 17 e 28 °C.

Em concordância com as afirmações anteriores, a interpretação da segunda regra é que mais de três dias com temperaturas acima de 30 °C e com temperatura mínima média em 15 dias abaixo de 18,3 °C formam um ambiente desfavorável à ocorrência da ferrugem asiática. Outros trabalhos variaram a temperatura e o tempo de molhamento foliar, sendo que Marchetti et al. (1976) concluíram que a situação ótima para infecção é com temperaturas entre 20 e 25 °C com molhamento foliar de 10 a 12 horas. Melching et al. (1989) verificaram que é necessário no mínimo seis horas de molhamento foliar para temperaturas entre 18 e 26,5 °C para o desenvolvimento da doença.

Além de considerar períodos anteriores à possível data de infecção, as regras 3 e 4 consideraram atributos no chamado período latente. Segundo Alves et al. (2006), a temperatura é a principal variável que afeta esse período, e temperaturas não ótimas nesse período podem ocasionar um atraso no desenvolvimento e no aparecimento dos sintomas da ferrugem, fazendo com que a ocorrência da doença não seja detectada na data esperada. A regra 4 apresentou o maior número de antecedentes (restrições), o que não diminuiu significativamente sua cobertura ou taxa de acerto. As variáveis apresentaram alguma variação nos períodos considerados e nas temperaturas, contudo, ainda de acordo com as características já discutidas.

Uma das diferenças das duas últimas regras da Tabela 2 foi em relação ao teste sobre o nó raiz (DTMIN15_20), pois considera que houve ao menos um dia com temperatura abaixo de 15 °C no período, que não favorece a ocorrência da doença. Dos 27,4% dos registros do conjunto original que seguem esse ramo, cerca de 82% foram classificados como “NaoOc”, reiterando a importância da variável para não-ocorrências.

A regra 6 apresentou um fato diferente ao considerar a classe “NaoOc” relacionada com eventos $TMIN_{10} > 18,34$, contrastando com a literatura consultada e, inclusive, com outras regras discutidas. A maior parte das regras apresentou que temperaturas mínimas mais elevadas estão relacionadas à classe “Oc”.

Outra diferença das duas últimas regras é a utilização de atributos relacionados à precipitação que não haviam sido considerados nas regras discutidas anteriormente. O atributo PMED_L é relacionado à precipitação média no período latente, porém, a temperatura é a variável que mais afeta esse período [Alves et al., 2006].

Em outros trabalhos a precipitação é considerada um fator importante para o desenvolvimento e a severidade da doença em plantações já infectadas [Del Ponte et al., 2006b]. Neste trabalho, os períodos utilizados são anteriores à detecção da doença, além de considerar apenas o evento de ocorrência ou não-ocorrência da doença, não sua severidade. Essas diferenças podem influenciar a importância da precipitação para os eventos considerados.

4. Conclusão

Utilizando dados derivados de temperatura e precipitação foi possível identificar situações que favoreceram ou não a ocorrência da ferrugem asiática da soja. A técnica

de modelagem em árvore de decisão se mostrou capaz de captar o principal padrão encontrado nos dados de forma interpretável e coerente com a literatura existente.

5. Referências

- ALVES, S.A.M.; FURTADO, G.Q.; BERGAMIN FILHO, A. 2006. Influência das condições climáticas sobre a ferrugem da soja. p.37-59. In: Zambolim, L. (ed.). Ferrugem Asiática da Soja. Viçosa: Universidade Federal de Viçosa, Departamento de Fitopatologia. 2006.
- BONDE, M. R.; BERNER, D. K.; NESTER, S. E.; FREDERICK, R. D. Effects of temperature on urediniospore germination, germ tube growth, and initiation of infection in soybean by phakopsora isolates. *Phytopathology*, v. 97, n. 8, p. 997-1003, ago 2007.
- DEL PONTE, E.M.; ESKER, P.D. Meteorological factors and Asian soybean rust epidemics: a systems approach and implications for risk assessment. *Scientia Agricola*, v. 65, n. spe, p. 88-97, dez 2008.
- DEL PONTE, E.M.; GODOY, C.V.; CANTERI, M.G.; REIS, E.M.; YANG, X.B. Models and applications for risk assessment and prediction of Asian soybean rust epidemics. *Fitopatologia Brasileira*, v. 31, n. 6, p. 533-544, 2006a.
- DEL PONTE, E.M.; GODOY, C.V.; LI, X.; YANG, X.B. Predicting severity of asian soybean rust epidemics with empirical rainfall models. *Phytopathology*, v. 96, n. 7, p. 797-803, 2006b.
- GODOY, C.V.; FLAUSINO, A.M.; SANTOS, L.C.M.; DEL PONTE, E.M. Eficiência do controle da ferrugem asiática da soja em função do momento de aplicação sob condições de epidemia em Londrina, PR. *Tropical Plant Pathology*, v. 34, n. 1, p. 056-061, 2009.
- KIM, K. S.; WANG, T. C.; YANG, X. B. Simulation of apparent infection rate to predict severity of soybean rust using a fuzzy logic system. *Phytopathology*, v. 95, n. 10, p. 1122-31, out 2005.
- MARCHETTI, M.A.; MELCHING, J.S.; BROMFIELD, K.R. The effects of temperature and dew period on germination and infection by uredospores of *Phakopsora pachyrhizi*. *Phytopathology*, n. 66, p. 461-463, 1976.
- MEIRA, C.A.A.; RODRIGUES, L.H.A.; MORAES, S.A.D. Modelos de alerta para o controle da ferrugem-do-cafeeiro em lavouras com alta carga pendente. *Pesquisa Agropecuária Brasileira*, v. 44, n. 3, p. 233-242, mar 2009.
- MELCHING, J. S.; DOWLER, W. M.; KOOGLE, D. L.; ROYER, M. H. Effects of Duration, Frequency, and Temperature of Leaf Wetness Periods on Soybean Rust. *Plant Disease*, v. 73, p. 117-122, 1989.
- ROMANI, L.A.S.; OTAVIAN, A.F.; EVANGELISTA, S.R.M.; ASSAD, E.D. Modelo de estações virtuais com estimativa de precipitação e temperatura para aprimoramento dos mapas no Agritempo. XV Congresso Brasileiro de Agrometeorologia. Anais... Aracaju - SE , 2007.
- WITTEN, I.H.; FRANK, E.; HALL, M.A. Data mining: Practical machine learning tools and techniques. 3ed. San Francisco: Morgan Kaufmann, 2011.