

# Preparação de dados para a modelagem de ocorrências da ferrugem asiática da soja

Guilherme Augusto Silva Megeto, estagiário, Laboratório de Inteligência Computacional  
Carlos Alberto Alves Meira, supervisor

V Mostra de Trabalhos de Estagiários e Bolsistas  
Campinas, SP – 26 a 30 de outubro de 2009

## Introdução

A ferrugem asiática da soja é uma das principais doenças da cultura. As perdas estimadas em grãos com a doença no Brasil são significativas. Os dados brutos se constituíram de registros de ocorrência da ferrugem em diferentes locais do Brasil, de 2004 a 2009, obtidos do Consórcio Antiferrugem, e de registros meteorológicos da base de dados do AgriTempo. A tarefa de mineração de dados escolhida inicialmente foi a classificação, e os casos de ocorrência da ferrugem definiram a classe positiva. Os casos da classe negativa foram criados para uma data anterior à de ocorrência da ferrugem, supondo-se não ter havido ocorrência antes da identificação confirmada.

## Objetivo

O objetivo do trabalho foi preparar dados para a modelagem de ocorrências de ferrugem asiática da soja. O intuito dessa modelagem é a descoberta de regras que permitam prever o risco de aparecimento de focos da doença com base em condições meteorológicas e dados sobre a cultura.

## Material e Métodos

Os dados do Consórcio Antiferrugem abrangem cinco safras (2004/05 a 2008/09) e contêm diversas informações, como a localização, o estágio fenológico da planta, a data em que foi verificada a ocorrência, o cultivar da soja etc. Os dados meteorológicos diários, obtidos do AgriTempo, incluem os atributos temperatura mínima, máxima e precipitação, medidos em diversas estações meteorológicas distribuídas pelo país. Para estudar os possíveis padrões que influenciaram a infecção, fez-se necessário o cálculo do período latente, estimado pela fórmula de Alves *et al* (2006):

$$PL = 0,11T^2 - 5,20T + 69,53$$

em que *PL* é o período latente em dias e *T* é a temperatura.

Assim, subtraiu-se o *PL* a partir da data de ocorrência ou não-ocorrência, e foram analisadas as variáveis meteorológicas de 3 dias (*m*) anteriores à possível infecção, como mostram as Figuras 01 e 02.

## Resultados e Discussão

O resultado da etapa de preparação dos dados foi o arquivo final contendo atributos meteorológicos, informações sobre a cultura e a classe de predição. Com a criação da classe negativa, duplicou-se o número de registros, gerando assim o conjunto de treinamento para realizar a tarefa de classificação. A partir desse arquivo, *softwares* específicos como o *Weka* e o *Rattle* serão utilizados com algoritmos de árvore de decisão para tentar extrair os padrões desejados.

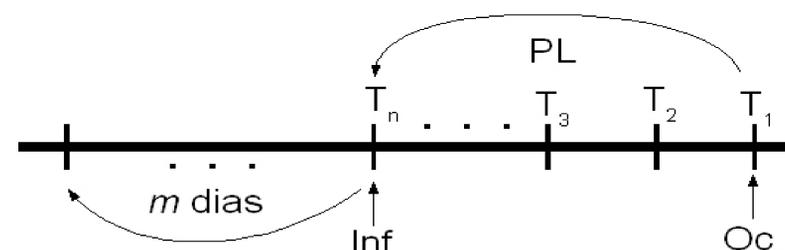


Figura 01 - Esquema da preparação dos dados meteorológicos. Oc. ocorrência da ferrugem; PL. período latente; T1 a Tn. temperaturas médias nos dias do PL; Inf. provável dia de infecção; m dias. período considerado para derivar os atributos meteorológicos.

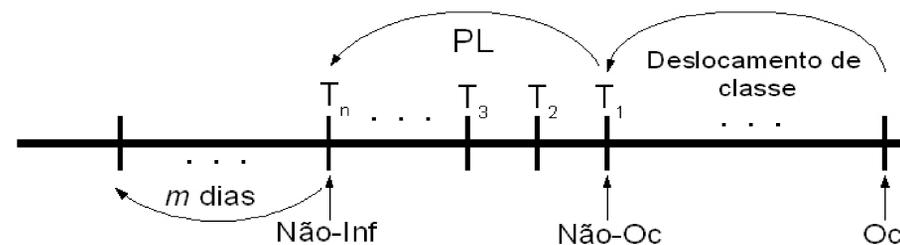


Figura 02 - Esquema de criação da classe negativa. Oc. data da ocorrência; Não-Oc. data da suposta não ocorrência, considerando um deslocamento de classe; PL. período latente; Não-Inf. data da suposta não infecção; m. dias anteriores à não infecção que foram analisados.

## Conclusões

A falta de dados meteorológicos em muitas das cidades reduziu bastante a quantidade de exemplos para a fase seguinte de modelagem, no entanto, não inviabiliza o prosseguimento do trabalho proposto. Para contornar esse problema, há a possibilidade de uso de dados estimados pelo radar do satélite TRMM para precipitação e pelo AgriTempo para as temperaturas máxima e mínima, com uma área de cobertura mais abrangente.

