

Uma extensão nebulosa para Teoria das Coberturas Parcimoniosas aplicada ao diagnóstico de doenças de plantas

Silvia Maria Fonseca Silveira Massruhá^{*, 1,2}, Sandra Aparecida Sandri^{**, 1}, Jacques Wainer^{**, 3}

(1) Área de Inteligência Artificial

Laboratório Associado de Computação e Matemática Aplicada
Instituto Nacional de Pesquisas Espaciais (INPE)

(2) Embrapa Informática Agropecuária

Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)

(3) Instituto de Computação

Universidade Estadual de Campinas (UNICAMP)

(*)Doutorado, Embrapa Informática Agropecuária, e-mail: silvia@nucleo.inpe.br ou silvia@cnptia.embrapa.br (**) Orientadores

Resumo

Este artigo visa descrever a validação de uma abordagem de diagnóstico que incorpora raciocínio temporal nebuloso no domínio de diagnóstico de doenças de plantas. As doenças são descritas através de um conjunto de manifestações possíveis e necessárias. Entre duas manifestações, a ocorrência dos intervalos de tempo esperados é modelada por conjuntos nebulosos. O artigo discute a medida de consistência temporal entre dados de manifestações de uma plantação de milho e o modelo de doenças, e define quando as manifestações da plantação podem ser explicadas por uma doença. Finalmente, são apresentados os resultados obtidos e os trabalhos de pesquisa a serem explorados no contexto de modelo de diagnósticos.

Palavras-Chave: diagnóstico, abdução, conjuntos nebulosos ou fuzzy, doenças de plantas

Introdução

O problema de diagnóstico consiste em encontrar as explicações para a presença de um conjunto de manifestações (sintomas), utilizando o conhecimento sobre o mundo. Os problemas de diagnóstico podem ser encontrados em diversas áreas, como o diagnóstico de sintomas de um paciente, de uma planta ou de um animal, a determinação de falha de um programa de um computador, a causa de ruídos em um sistema hidráulico, a localização de falhas num dispositivo eletrônico, entre outras [7].

Os sistemas de diagnóstico são usados em indústrias e têm sido uma área ativa de pesquisa em Inteligência Artificial – IA. Esta área tem gerado muitas metodologias, teorias e aplicações nas duas últimas décadas. Sistemas de diagnósticos variam desde sistemas baseados em regras, teorias baseadas em conjuntos, teorias baseadas em lógicas e raciocínio baseado em caso [1,4,5,7]. Adicionalmente, o raciocínio temporal tem sido considerado uma parte importante no raciocínio de diagnóstico [8].

Neste contexto, a Teoria das Coberturas Parcimoniosas – TCP [7] (*Parsimonious Covering Theory* – PCT) é apontada por seus autores como fundamento teórico para estes modelos na formalização da natureza abduativa do raciocínio de diagnóstico, assim como a lógica de primeira ordem e a teoria da probabilidade são os fundamentos teóricos de dedução baseada em regras e modelo estatístico de classificação, respectivamente.

Neste artigo está descrito a validação de um modelo que incorpora conjuntos nebulosos no raciocínio temporal no contexto de TCP, que é a base teórica para nossa abordagem de diagnóstico. No escopo deste trabalho, o domínio do problema refere-se à diagnose de doenças de plantas.

Informação temporal/categórica nebulosa em diagnóstico

Uma extensão para TCP é proposta por Wainer & Sandri [9] para incorporar conceitos nebulosos [2] em alguns conceitos de TCP temporal [8]. Em muitas aplicações de diagnóstico, uma doença inicia-se com um evento, tal como, a quebra de um componente de um sistema elétrico ou mecânico, infecção de um paciente ou uma lesão em uma planta. Este trabalho mostra que melhor do que representar a informação temporal como intervalos “clássicos” (*crisp*) é representá-la como intervalos nebulosos. Em [9] foram propostos índices para incorporar informação temporal, categórica e de intensidade nebulosa em diagnóstico.

Suponhamos uma base de conhecimento, contendo um conjunto de doenças $D = \{d_1, d_2, \dots, d_n\}$ associado às manifestações $M = \{m_1, m_2, m_3, \dots, m_m\}$ que elas causam, bem como a relação causal entre D e M . Adicionalmente, cada doença $d_i \in D$ possui um grafo temporal associado, no qual os nós representam os eventos de início ou término de uma manifestação rotulados por (m_k^b, m_k^e) e os arcos rotulados de $(\pi_i(e_i, e_j))$ representam os intervalos mínimos entre quaisquer dois eventos que possam ocorrer na progressão de uma doença d_i . O conjunto M^+ contém as manifestações apresentadas em um dado caso, o conjunto EV^+ contém

eventos conhecidos de início e fim de manifestações, e $TIME^+(m)$ indica quando ocorreu a manifestação m (este valor e' modelado por um intervalo nebuloso). O índice de consistência temporal é definido por

$$\alpha(dl) = \min_{e_i, e_j \in EV^+ \cap V(dl)} [h(DIST^+(e_i, e_j) \cap \pi_i(e_i, e_j))]$$

O conjunto nebuloso $DIST^+(e_i, e_j) = TIME^+(e_j) \ominus TIME^+(e_i)$ representa a diferença dos intervalos entre dois eventos, onde h representa a altura de um conjunto nebuloso, \cap denota a interseção de 2 conjuntos nebulosos, utilizando o operador \min e a diferença nebulosa $A \ominus B = \langle a_1 - b_4, a_2 - b_3, a_3 - b_2, a_4 - b_1 \rangle$ para $A = \langle a_1, a_2, a_3, a_4 \rangle$ e $B = \langle b_1, b_2, b_3, b_4 \rangle$. Para exemplificar a consistência temporal será apresentado dois casos sobre doenças de milho, baseados em [3,6].

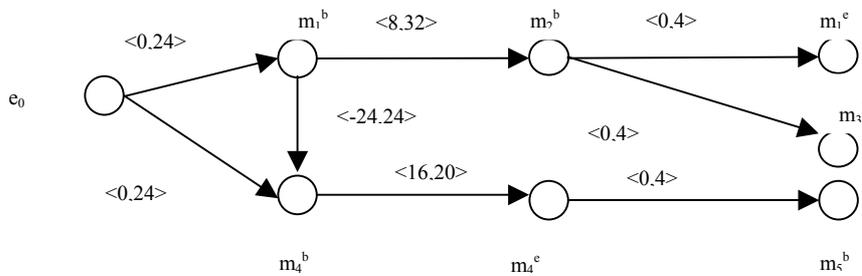


Figura 1: Modelo simplificado do grafo mínimo G1 de dl

Na Figura 1 está apresentado um modelo simplificado do grafo mínimo temporal da doença de milho denominada Ferrugem Comum (dl). Neste modelo simplificado são usados dados fictícios para modelar a informação temporal de modo a facilitar os cálculos. Seja a base de conhecimento seguinte: $D = \{dl\}$ onde dl = Ferrugem comum; $M = \{e0, m1, m2, m3, m4, m5\}$ onde $e0$: contaminação causada pelo fungo *Puccinia sorghi*; $m1$: formação de pústulas marrom-claras; $m2$: formação de fenda; $m3$: formação de pústulas alongadas marrom-escuras; $m4$: seca prematura das plantas; $m5$: baixa produção. Os intervalos “em menos de 24 semanas” e “a cerca de 8 a 32 semanas” são dados fictícios que correspondem aos intervalos $\langle 0,24 \rangle$ e $\langle 8,32 \rangle$, respectivamente.

a) Seja o caso hipotético 1 em plantação de milho no qual temos :

$M^+ = \{m1, m4\}$; $EV^+ = \{m1^b, m4^b\}$; $TIME^+(m1^b) = \langle 49,50 \rangle$ e $TIME^+(m4^b) = \langle 23,24 \rangle$; $DIST^+(m1^b, m4^b) = \langle 25,27 \rangle$. A partir do grafo mínimo, temos os intervalos mínimos entre dois eventos, por exemplo $\pi_i(m1^b, m4^b) = \langle -24,24 \rangle$ e consequentemente aplicando o índice de consistência temporal temos,

$\alpha(dl) = h(DIST^+(m1^b, m4^b) \cap \pi_i(m1^b, m4^b)) = h(\langle 25,27 \rangle \cap \langle -24,24 \rangle) = 0$. Portanto, o modelo e o caso são totalmente incompatíveis em termos temporais.

b) Seja o caso hipotético 2 em plantação de milho no qual temos :

$M^+ = \{m1, m4\}$; $EV^+ = \{m1^b, m4^b\}$; $TIME^+(m1^b) = \langle 48, 49,50,51 \rangle$ e $TIME^+(m4^b) = \langle 22,23,24,25 \rangle$; $DIST^+(m1^b, m4^b) = \langle 23,25,27,29 \rangle$. A partir do grafo mínimo $\pi_i(m1^b, m4^b) = \langle -24,24 \rangle$ e consequentemente

$\alpha(dl) = h(DIST^+(m1^b, m4^b) \cap \pi_i(m1^b, m4^b)) = h(\langle 23,25,27,29 \rangle \cap \langle -24,24 \rangle) = 0.5$ (veja Figura 2). Assim, dl poderia ser considerada uma explicação plausível para M^+ .

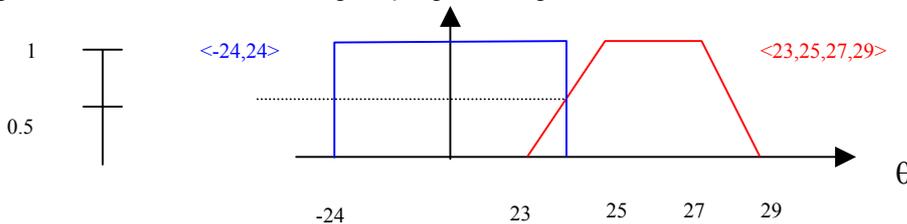


Figura 2: Gráfico exemplificando $\alpha(dl) = h(DIST^+(m1^b, m4^b) \cap \pi_i(m1^b, m4^b))$

Concluindo, a melhor explicação para o diagnóstico assumida em [9] é uma doença que tem consistência temporal, categórica e de intensidade com todos os sintomas e explica a presença deles no caso. Por exemplo, $\alpha(dl) > 0$ como no caso hipotético 2 (letra b).

A informação categórica é representada por um rótulo N ou P para cada nó no grafo, indicando se aquela manifestação é necessariamente ou somente possível ser causada pela doença d. Conforme apresentado em [9], a consistência categórica entre caso e o modelo de uma doença d_i refere-se ao fato que d_i não deve ser considerada uma boa explicação para o conjunto de sintomas se uma manifestação necessária de d_i não tenha ocorrido mesmo que haja transcorrido um intervalo de tempo suficiente para tal. Em algumas doenças é

importante quantificar a intensidade em que a manifestação ocorre. Por exemplo, para uma determinada doença a temperatura tem que ser alta para que ela desenvolva. Portanto, em um caso em que a temperatura é baixa, talvez diminua a possibilidade de ser aquela doença. A consistência categórica e de intensidade não estão apresentadas neste trabalho, entretanto, maiores detalhes podem ser encontrados em [9].

Conclusões e trabalho futuros

Através deste trabalho, foi possível aplicar o modelo apresentado em [9] no domínio de doenças de plantas de milho a fim de validar a abordagem de diagnóstico proposta em um domínio diferente da área médica. Em [9], além de ser apresentado o modelo para incorporar informação temporal nebulosa, informação categórica e informação de intensidade numa abordagem de diagnóstico, também foi mostrado como fazer previsões sobre manifestações futuras. Entretanto, a parte de previsões futuras não foi validada neste trabalho devido as heurísticas específicas do domínio agrícola que precisam ser melhor investigadas.

Quanto à validação da abordagem de diagnóstico, obtiveram-se resultados positivos em relação ao mapeamento das categorias de consistência no domínio de doenças de milho. Todavia, pode-se observar que algumas extensões se tornam necessárias devido às peculiaridades do domínio agrícola. Um ponto crucial também em diagnóstico é identificar quais as manifestações prioritárias a serem descobertas para obter um diagnóstico mais rápido. No domínio agrícola, tal descoberta pode ser vital, inclusive comprometendo toda uma produção ou um replanejamento agrícola. Isto pode ser feito através de um função de utilidade em cima das manifestações de modo a reduzir hipóteses. No mesmo domínio, ainda se tem um outro problema que não foi considerado neste trabalho: as condições ambientais que interferem na severidade da doença. É o caso, por exemplo, da temperatura, da umidade, do vento, dentre outras.

Um outro trabalho de pesquisa futuro é investigar melhor a informação categórica nebulosa, como por exemplo “na doença d_i a manifestação m_j raramente ocorre”. Isto provocaria que N e P fossem substituídos por distribuições de possibilidades. Um outro ponto é modelar a incerteza de uma determinada manifestação ter ocorrido ou não que é diferente de modelar a intensidade da manifestação como foi apresentado em [9].

É válido lembrar a importância de explorar a possibilidade de ter um conjunto de doenças para explicar as manifestações de uma caso particular e não só de uma doença como proposto em [9]. Isto se torna mais interessante em domínios em que um conjunto de doenças tem manifestações em comum como é o caso de doenças de plantas. Finalmente, cabe ressaltar que, nos trabalhos futuros desta linha de pesquisa, pretende-se continuar a avaliar e, possivelmente, desenvolver modelos de diagnósticos que são mais apropriados para as áreas de veterinária e fitopatologia e que resultem num paradigma de desenvolvimento de sistemas de diagnósticos mais eficientes.

Referências

- [1] Buchanan, J. R.; Shortliffe, E.H. *Rule-Based Expert Systems: The MYCIN Experiments of the Standard Heuristic Programming Project*. Addison-Wesley Publishing Company, 1984.
- [2] Dubois D; Prade, H. *Possibility Theory: na approach to computadorized processing of uncertainty*. Plenum Press. 1988.
- [3] Fernandes, F.T.; Oliveira, E. *Principais doenças na cultura do milho*. Brasília: EMBRAPA-CNPMS / EMBRAPA-SPI, 1997. 80p.
- [4] Koton, P. A. *Using Experience in Learning and Problem Solving*. Phd Thesis, MIT Laboratories of Computer Science, MIT/LCS/TR441, 1988.
- [5] Lucas, P. Modelling Interactions for diagnosis. In *Proceedings of CESA'96 IMACS Multiconference: Symposium and Modelling, Analysis and Simulation*, vol.1, pg.541-546, 1996.
- [6] Massruhá, S.M.F.S; Cruz, S.A.B.; Romani, L.A.S.; Souza, E. Remote Diagnosis: An expert system for diagnosis of the corn diseases by web.– *AWAPAP'99 Proceedings*. Sidney, Australia, December, 1999.
- [7] Peng, Y.; Reggia J.A. *Abductive Inference Models for Diagnostic Problem-Solving*, Springer-verlag, 1990.
- [8] Wainer, J.; Rezende, A.M. A temporal extension to the Parsimonius Covering Theory. *Artificial Intelligence in Medicine*. Elsevier Science B. V, 1997. v.10, p. 235-255.
- [9] Wainer, J.; Sandri, S. A Fuzzy Temporal/Categorical Information in diagnosis. *Special Issue on Intelligent Temporal Information Systems in Medicine, JIIS*, 1998.