

ĐỀ TÀI KC.01.15/06-10

Thành phần khai mô tri thức: mô hình trực quan
cây quyết định cho dự báo dịch hại

Đỗ Thanh Nghị, Lê Quyết Thắng

Khoa Công nghệ thông tin, ĐHCT
Số 1 Lý Tự Trọng, Ninh Kiều, Cần Thơ
dtngghi@cit.ctu.edu.vn

Ngày 13 tháng 3 năm 2011



Giới thiệu

Xây dựng thành phần khai mở tri thức

Mô hình trực quan cây quyết định cho dự báo dịch hại

Kết luận và hướng phát triển



Nội dung công việc thực hiện

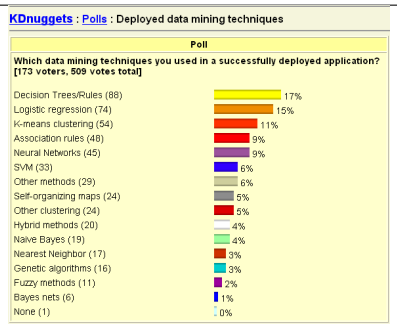
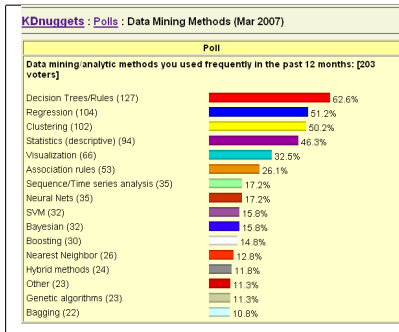
Thành phần khai mở tri thức

- ▶ xây dựng thư viện các hàm khai mở tri thức
- ▶ xây dựng dịch vụ web khai mở tri thức phòng chống dịch hại
- ▶ hướng dẫn cài đặt và sử dụng
- ▶ mô hình trực quan cây quyết định cho dự báo dịch hại



Giải thuật quan trọng của khai mở dữ liệu (KDNuggets)

Top 10 giải thuật khai mở dữ liệu (Wu & Kumar, 09)



Xây dựng thư viện các hàm khai mở tri thức

Các giải thuật hiệu quả

- ▶ k láng giềng (Fix & Hodges, 52)
- ▶ Bayes thơ ngây (Good, 65)
- ▶ cây quyết định (Breiman et al., 84), (Quinlan, 93)
- ▶ phương pháp tập hợp mô hình như bagging (Breiman, 96),
- ▶ boosting (Freund & Schapire, 95)
- ▶ rừng ngẫu nhiên (Breiman, 01)
- ▶ máy học véctơ hỗ trợ SVM (Vapnik, 95)
- ▶ giải thuật gom cụm kMeans (MacQueen, 67)



Xây dựng thư viện các hàm khai mở tri thức

Tính hiệu quả

- ▶ tốc độ xử lý: Bayes thơ ngây, cây quyết định, kMeans
- ▶ dễ dịch kết quả: Bayes thơ ngây, cây quyết định, kMeans
- ▶ độ chính xác cao: bagging, boosting, rừng ngẫu nhiên, SVM
- ▶ kết hợp nhiều phương pháp hỗ trợ lẫn nhau
- ▶ sử dụng phương pháp hiển thị, tương tác, trực quan để nâng cao hiệu quả sử dụng của mô hình khai mở dữ liệu



Xây dựng dịch vụ web khai mở tri thức

Kiến trúc hướng dịch vụ SOAP

- ▶ chuẩn cho nghi thức truy xuất đối tượng đơn giản
- ▶ nghi thức giao tiếp giữa các ứng dụng thông qua internet
- ▶ định dạng cho việc gửi thông điệp
- ▶ độc lập với nền
- ▶ độc lập với ngôn ngữ lập trình
- ▶ dựa trên ngôn ngữ XML
- ▶ đơn giản, dễ mở rộng



Giải thuật cây quyết định

Tại sao cây quyết định?

- ▶ xây dựng cây nhanh, đơn giản
- ▶ cho độ chính xác cao so với giải thuật khác
- ▶ kết quả dễ hiểu cho người sử dụng: chuyên gia nông nghiệp
- ▶ rút trích luật dự báo: **IF ... THEN ...**
- ▶ điều kiện **và**: đường dẫn từ nút gốc
- ▶ dự báo **lớp**: nhãn của nút lá

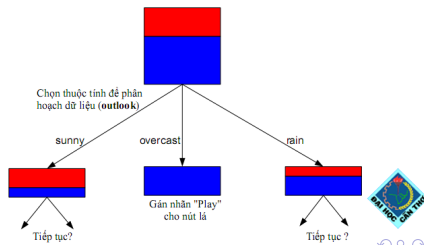


Giải thuật cây quyết định

Học có giám sát

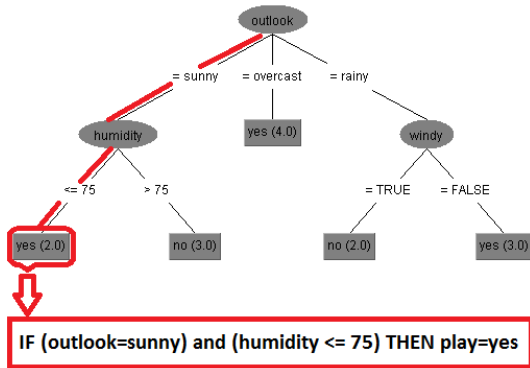
- ▶ xây dựng cây: hàm phân hoạch (entropy, gini)
- ▶ gán nhãn cho nút lá: luật số đông
- ▶ cắt nhánh: tránh học vẹt

outlook	temp.	hum.	windy	Play?
sunny	85	85	false	No
sunny	80	90	true	No
overcast	83	78	false	Yes
rain	70	96	false	Yes
rain	68	80	false	Yes
rain	65	70	true	No
...	...	...	...	...
...	...	...	...	...



Mô hình cây quyết định cho dự báo chơi Golf

Một luật dự báo chơi Golf



Cây quyết định dự báo dịch hại lúa

Xây dựng mô hình cây dựa trên các thuộc tính

- ▶ giống lúa
- ▶ mật độ sạ (kg/ha)
- ▶ gốc thuốc sâu
- ▶ thời điểm bón phân
- ▶ lượng phân đạm
- ▶ thiên địch
- ▶ độ tuổi rầy
- ▶ giai đoạn sinh trưởng lúa
- ▶ mật độ rầy, hướng gió
- ▶ dự báo **nhễm, cháy, lan truyền rầy**



Cây quyết định dự báo dịch hại lúa

Chuẩn bị dữ liệu

- ▶ 40000 mẫu tin
- ▶ 34 thuộc tính
- ▶ dự báo **niêm rầy, cháy rầy**
- ▶ dữ liệu có nhãn lan truyền theo hướng đông bắc và hướng tây nam => khó dự báo lan truyền: lấy mẫu tăng thêm
- ▶ dự báo **lan truyền rầy**



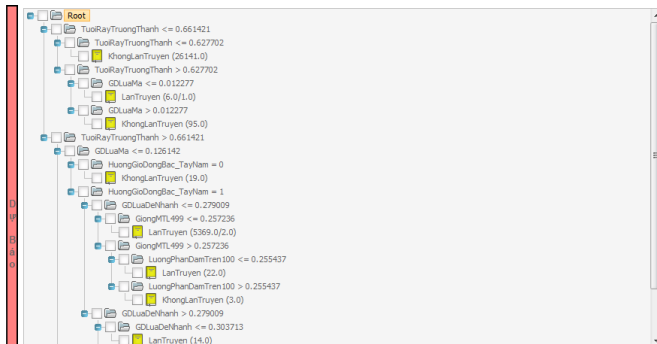
Cây quyết định dự báo dịch hại lúa

Hiển thị trực quan, tương tác

- ▶ hiển thị cây theo cấu trúc phân cấp
- ▶ nhấp chuột vào nút lá: rút trích luật quyết định tương ứng
- ▶ nhấp chuột vào nút trong: rút trích tập luật quyết định tương ứng với các nút lá dẫn xuất từ nút trong
- ▶ dự báo trực tuyến
- ▶ chuyên gia dễ kiểm chứng kết quả



Cây quyết định dự báo lan rầy theo hướng đông bắc



[Mô cây]

[Đóng cây]

[Dự báo]

Sử dụng:

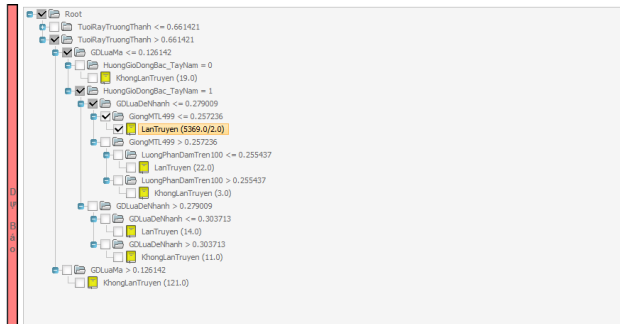
Để nút trích luật quyết định từ cây, chỉ cần nhấp chuột vào từng nút của cây. Nếu là nút lá thì ta có một luật quyết định là đường dẫn từ nút gốc (Root) đến nút lá. Nếu là nút trong ta có tập luật tương ứng với các nút lá của nút trong. Doubleclick menu Dự báo cho phép nhập dữ liệu cho dự báo.

Mở đến nút 27

<= Tập luật



Rút trích 1 luật dự báo (nút lá)



[Mở cây]

[Đóng cây]

[Dự báo]

Sử dụng:

Để rút trích luật quyết định từ cây, chỉ cần nhấp chuột vào từng nút của cây. Nếu là nút lá thì ta có một luật quyết định là đường dẫn từ nút gốc (Root) đến nút lá. Nếu là nút trong ta có tập luật tương ứng với các nút lá của nút trong. Doubleclick menu Dự báo cho phép nhập dữ liệu cho dự báo.

1. NẾU

(TuoRayTruongThanh > 0.661421) VÀ
(GDLuaMa <= 0.126142) VÀ
(HuongGioDongBac_TayNam = 1) VÀ
(GDLuaDeNhanh <= 0.279009) VÀ
(GiongMTL499 <= 0.257236)

THÌ

LanTruyen (5369.0/2.0)

<= Tập luật



Rút trích nhiều luật dẫn từ nút trong



[Đóng cây]

Sử dụng:

Mẫu là tốt

Nếu là nút lá thì ta có một loại quyết định là đường dẫn từ nút gốc (Root) đến nút lá. Nếu là nút trong ta có tập luật tương ứng với các nút lá của nút trong. Doubleclick menu Dự báo cho phép nhập dữ liệu cho dự báo.

<= Tập luật



Dự báo trực tuyến

Tỉ lệ	0.05	0.2	0.04	0.1	0.45	0.1	0.06
Mật độ sạ	Dưới 120		Trên 120				
Tỉ lệ	0.45		0.55				
Góc thuốc	Clorpyrifos		Cypermethrin		Khác		
Tỉ lệ	0.45		0.23		0.32		
Ngày xịt	Trên 40 ngày		Dưới 40 ngày				
Tỉ lệ	0.34		0.66				
Thời điểm bón phân	Đúng		Sai				
Tỉ lệ	0.55		0.45				
Lượng phân đạm	< 90		90 - 100		> 100		
Tỉ lệ	0.57		0.18		0.25		
Thiên địch	Nhiều		Ít				
Tỉ lệ	0.6		0.4				
Tuổi rầy	Trùng	1 - 3 ngày	4 - 5 ngày	Trưởng thành			
Tỉ lệ	0	0	0.36	0.63			
Giai đoạn	Mạ	Đẻ nhánh	Đông trổ	Chín			
Tỉ lệ	0	1	0	0			
Mật độ rầy	< 5000		5000 - 10.000		> 10.000		
Tỉ lệ	1		0		0		
Hướng gió	Tây Nam		Đông Bắc				
Tỉ lệ	0		1				
[Dự báo]	Kết quả: LanTruyen (6.0/1.0)						

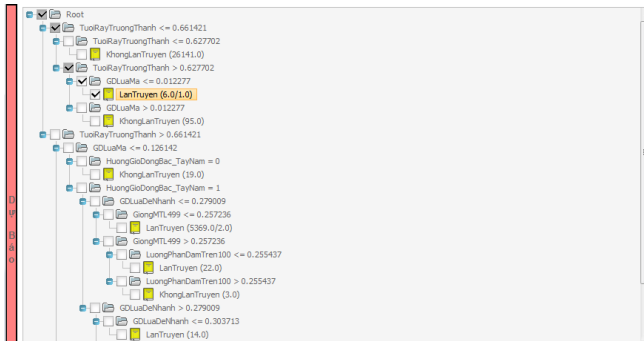
Sử dụng:

Để rút trích luật quyết định từ cây, chỉ cần nhấp chuột vào từng nút của cây. Nếu là nút lá thì ta có một luật quyết định là đường dẫn từ nút gốc (Root) đến nút lá. Nếu là nút trong ta có tập luật tương ứng với các nút lá của nút trong. Doubleclick menu Dự báo cho phép nhập dữ liệu cho dự báo.

<= Tập luật



Luật sử dụng để dự báo



[Mở cây]

[Đóng cây]

Dự báo

Sử dụng:

Để rút trích luật quyết định từ cây, chỉ cần nhấp chuột vào từng nút của cây. Nếu là nút lá thì ta có một luật quyết định là đường dẫn từ nút gốc (Root) đến nút lá. Nếu là nút trong ta có tập luật tương ứng với các nút lá của nút trong. Doubleclick menu Dự báo cho phép nhập dữ liệu cho dự báo.

1. NẾU

(TuoiRayTruongThanh ≤ 0.661421) VÀ
(TuoiRayTruongThanh > 0.627702) VÀ
(GDLuaMa ≤ 0.012277)

THÌ

LanTruyen (6.0/1.0)

$\leq$ Tập luật



Kết luận

Thành phần khai mở tri thức

- ▶ xây dựng thư viện các hàm khai mở tri thức
- ▶ xây dựng dịch vụ web khai mở tri thức phòng chống dịch hại
- ▶ mô hình trực quan cây quyết định cho dự báo dịch hại
- ▶ hiển thị cây theo cấu trúc phân cấp
- ▶ rút trích luật quyết định tương ứng
- ▶ dự báo trực tuyến
- ▶ chuyên gia dễ kiểm chứng kết quả
- ▶ truy cập: <http://www.phongchongdichhai.org.vn/kdd>



Hướng phát triển²

Thành phần khai mở tri thức

- ▶ tích hợp thêm các phương pháp hiển thị
- ▶ giải thích mô hình
- ▶ chuyên gia dễ kiểm chứng kết quả

