



Khoa Công Nghệ Thông Tin Trường Đại Học Cần Thơ



Phương pháp hiển thị dữ liệu Information visualisation

Đỗ Thanh Nghị
dtnghi@cit.ctu.edu.vn

Cần Thơ
02-12-2008

Nội dung

- Giới thiệu về info viz
- Info viz trong khai mỏ dữ liệu
- Kết luận và hướng phát triển

Nội dung

- Giới thiệu về info viz
- Info viz trong khai mỏ dữ liệu
- Kết luận và hướng phát triển

- Giới thiệu về info viz
- Info viz trong khai mỏ dữ liệu
- kết luận và hướng phát triển

Hiển thị dữ liệu

■ hiển thị dữ liệu

- quan trọng trong quá trình khai mỏ dữ liệu
- trong tiền xử lý dữ liệu : giúp xem sơ lược về dữ liệu, phát hiện một vài tính chất tổng quát, chọn giải thuật để khai mỏ dữ liệu và có 1 vài ý tưởng cho việc chọn các tham số
- trong khai mỏ dữ liệu, có thể thay thế hoặc phối hợp các phương pháp học tự động (visual data mining), người sử dụng làm trung tâm, sử dụng được khả năng nhận dạng mẫu của con người, có thể sử dụng được ý kiến chuyên gia khi xây dựng mô hình, giúp người dùng dễ hiểu mô hình xây dựng bởi vì chính họ trực tiếp tham gia xây dựng mô hình
- trong hậu xử lý : giúp giải thích các kết quả sinh ra trong quá trình học tự động

Nội dung

- Giới thiệu về info viz
- **Info viz trong khai mỏ dữ liệu**
- Kết luận và hướng phát triển

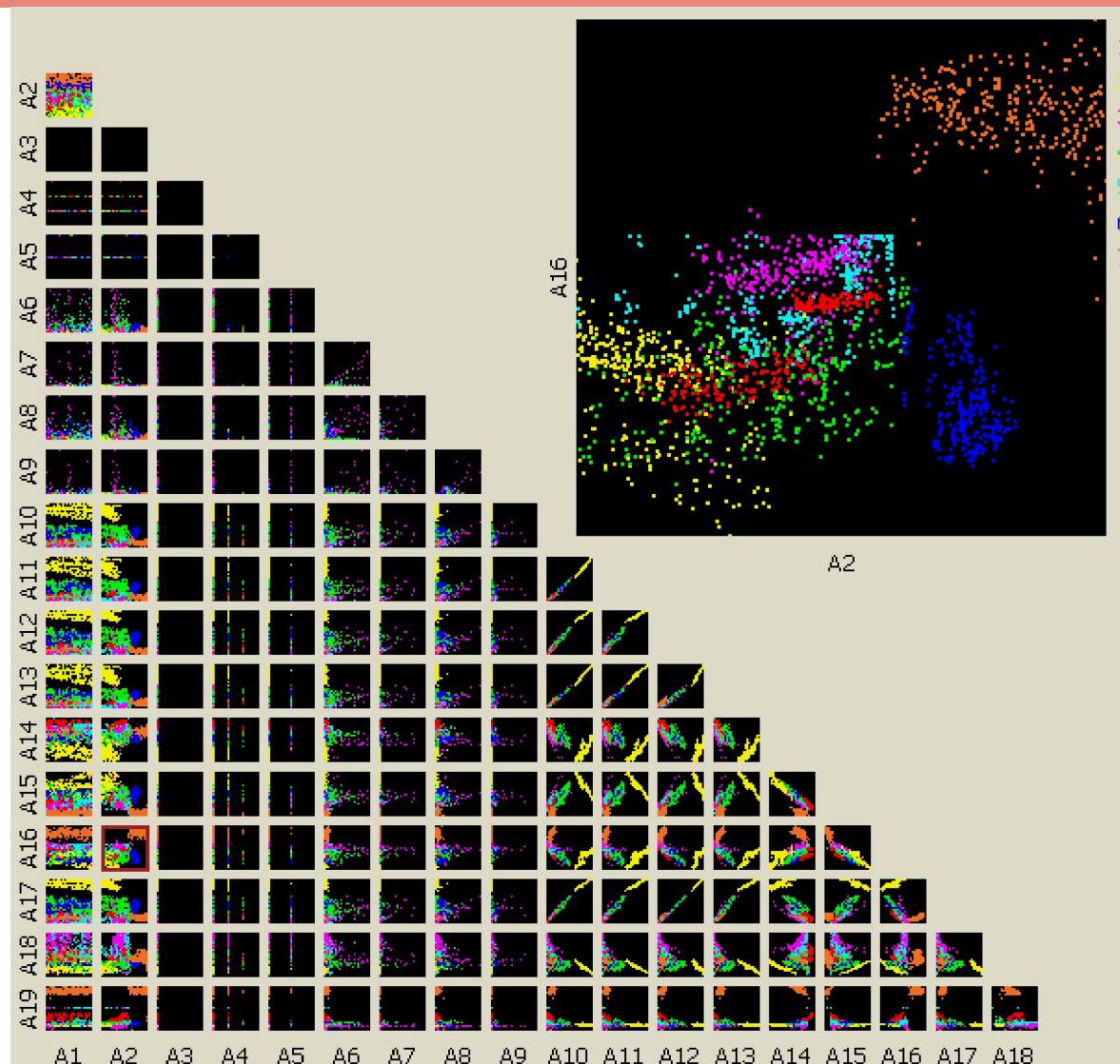
- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
kết luận và hướng phát triển

Phương pháp hiển thị dữ liệu (Keim, 1996) (Fayyad et al., 2001)

- phương pháp hình học : scatterplots (Cleveland, 1993), parallel coordinates (Inselberg, 1985)
- phương pháp pixel : bar viz (Ankerst, 1999), circle segment (Keim, 1996)
- phương pháp icon : Chernoff faces (Chernoff, 1973), Figure baton (Pickett, 1970)
- phương pháp phân cấp : dimensional stacking (LeBlanc, 1990)

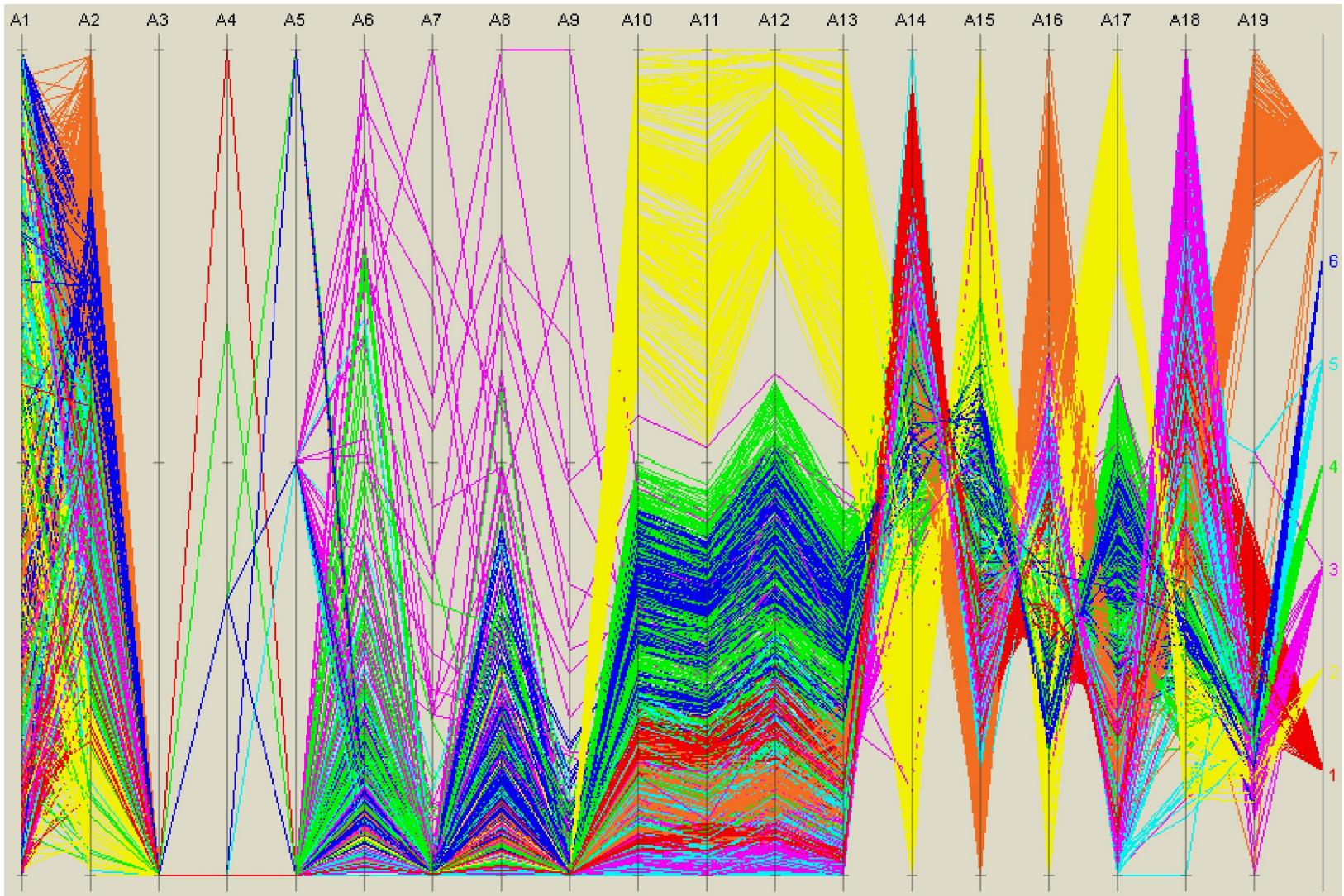
- Giới thiệu về info viz
- Info viz trong khai mỏ dữ liệu
- kết luận và hướng phát triển

Scatterplots : hiển thị cặp các thuộc tính dữ liệu,
màu = lớp



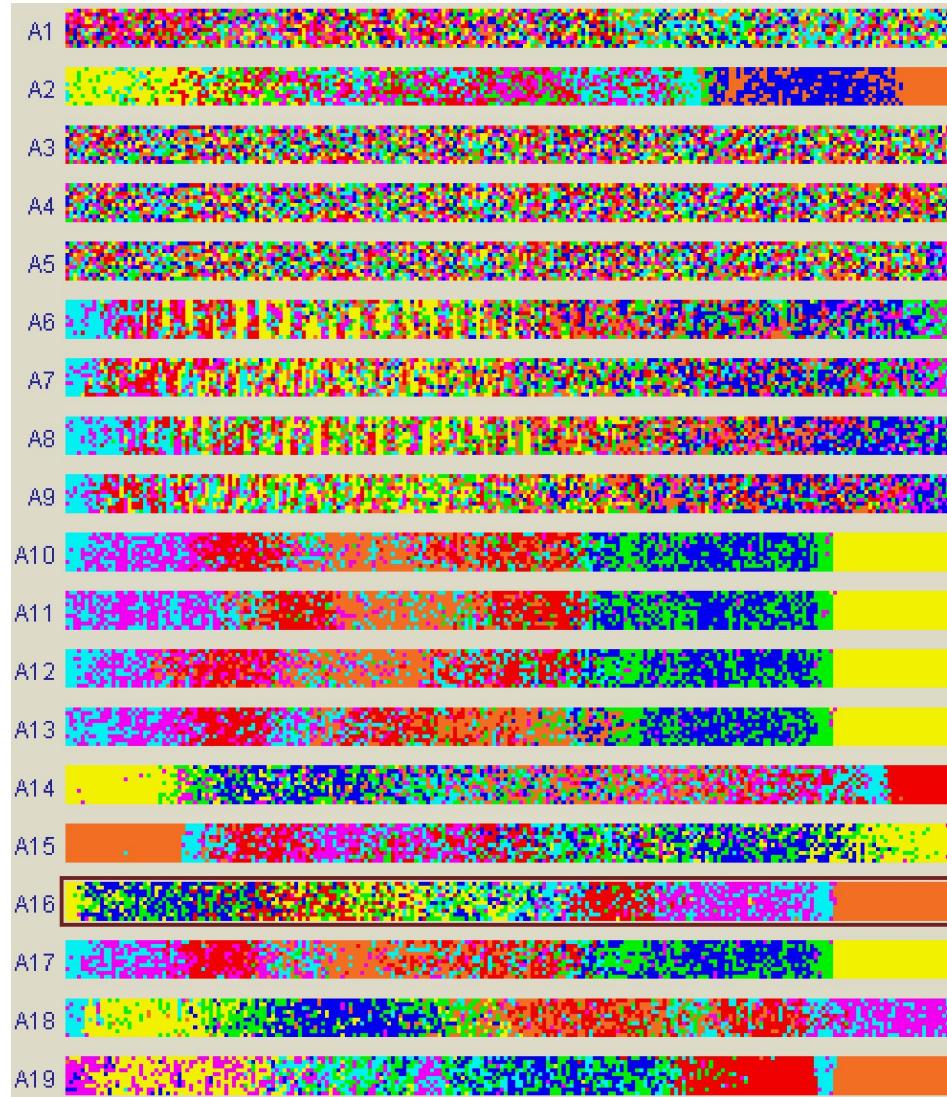
- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

Parallel coordinates : điểm = đường, màu = lớp



- Giới thiệu về info viz
- Info viz trong khai mỏ dữ liệu
- kết luận và hướng phát triển

Phương pháp Bar viz. : bar = thuộc tính, sắp xếp theo thứ tự, màu = lớp



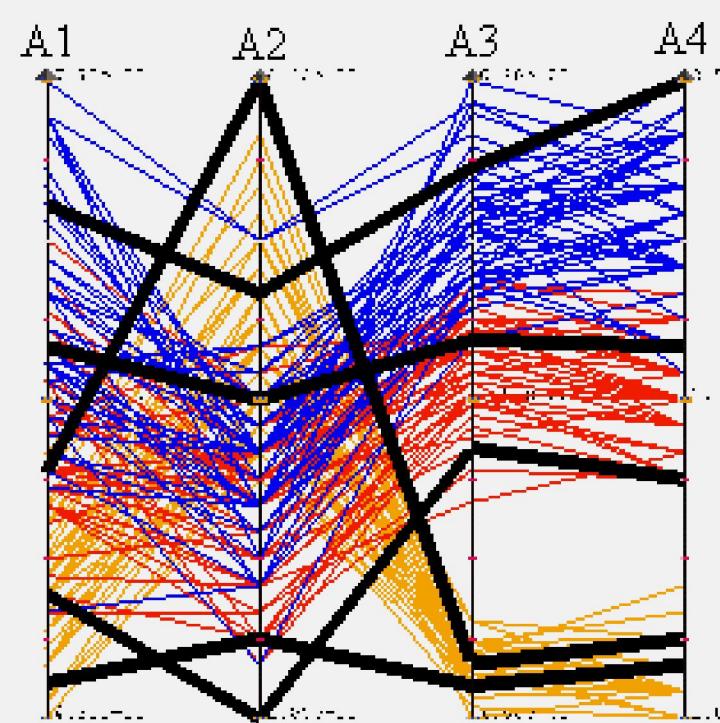
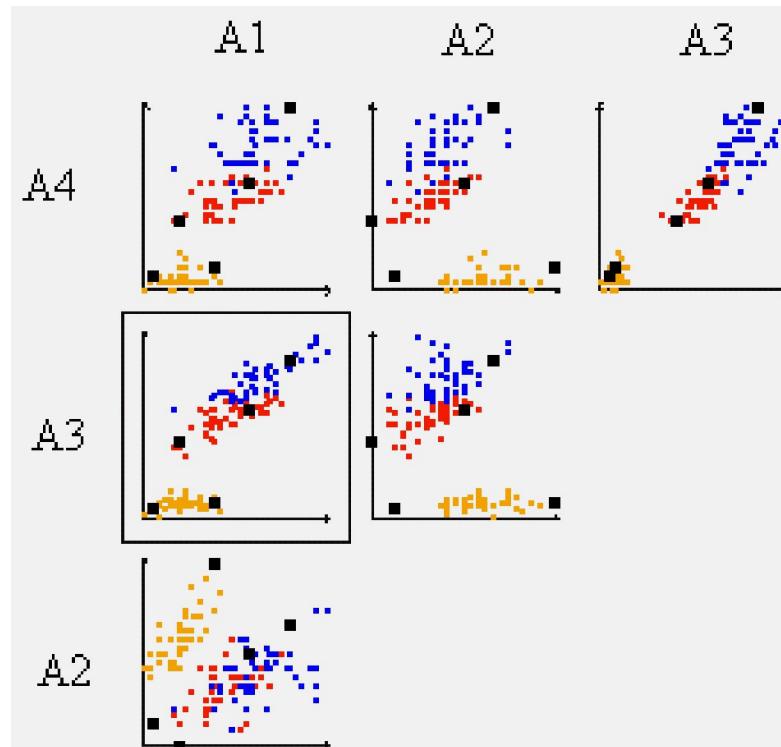
- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

Kết hợp các phương pháp

- mỗi phương pháp đều có ưu khuyết điểm
 - dễ hiểu
 - dữ liệu lớn (số chiều, số phần tử)
 - mất thông tin
 - nhìn thấy clusters
 - kết hợp các phương pháp để tận dụng các ưu điểm
 - multiview, linking, brushing

- Giới thiệu về info viz
- Info viz trong khai mỏ dữ liệu
- kết luận và hướng phát triển

Scatterplots, parallel coordinates

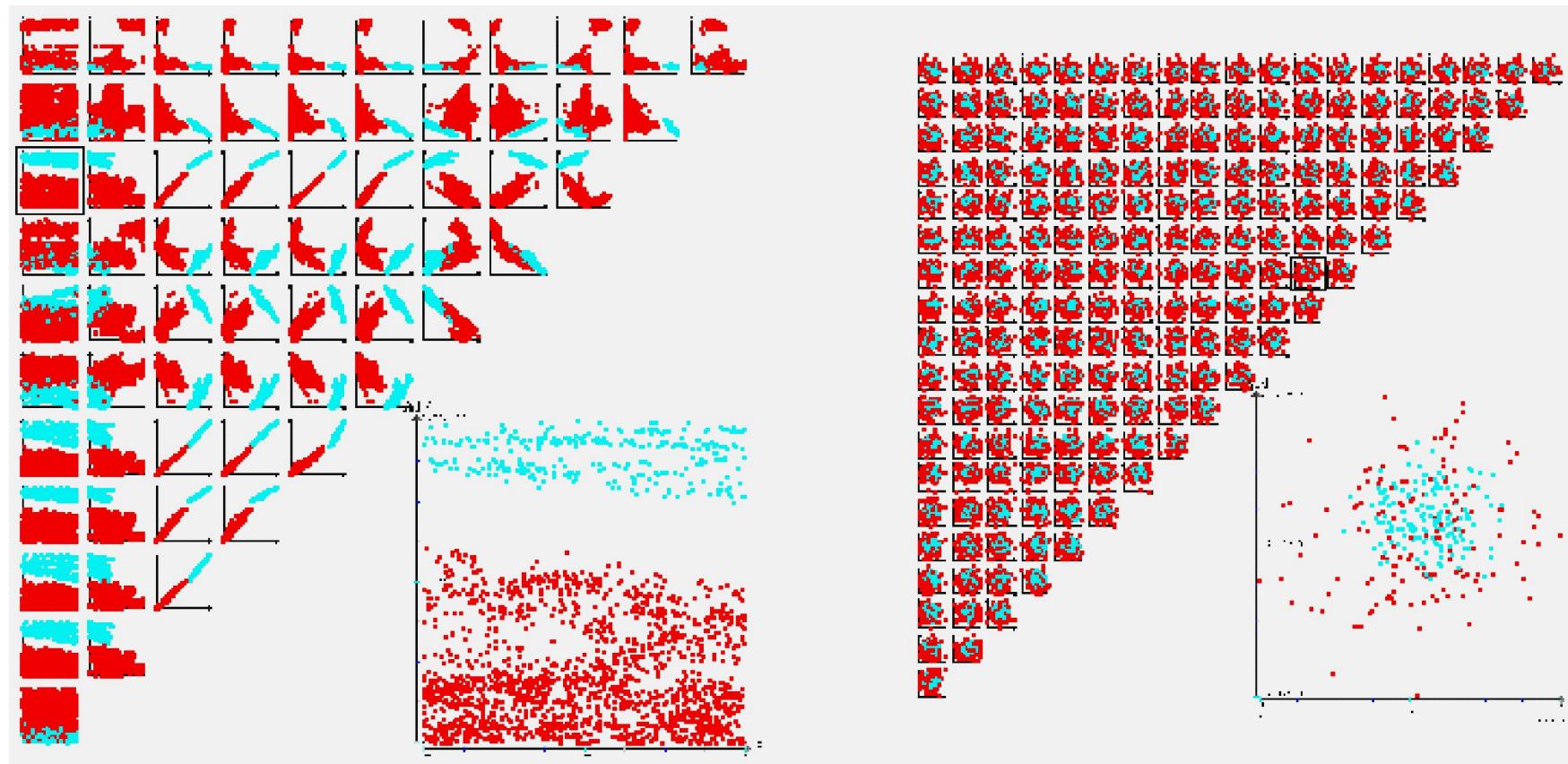


- Giới thiệu về info viz
- Info viz trong khai mỏ dữ liệu
- kết luận và hướng phát triển

Hiển thị dữ liệu hỗ trợ cho tiền xử lý

- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

Scatterplots : linear ?



Hiển thị dữ liệu hỗ trợ cho khai mỏ dữ liệu

- Giới thiệu về info viz
- Info viz trong khai mỏ dữ liệu
- Kết luận và hướng phát triển

- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- Kết luận và hướng phát triển

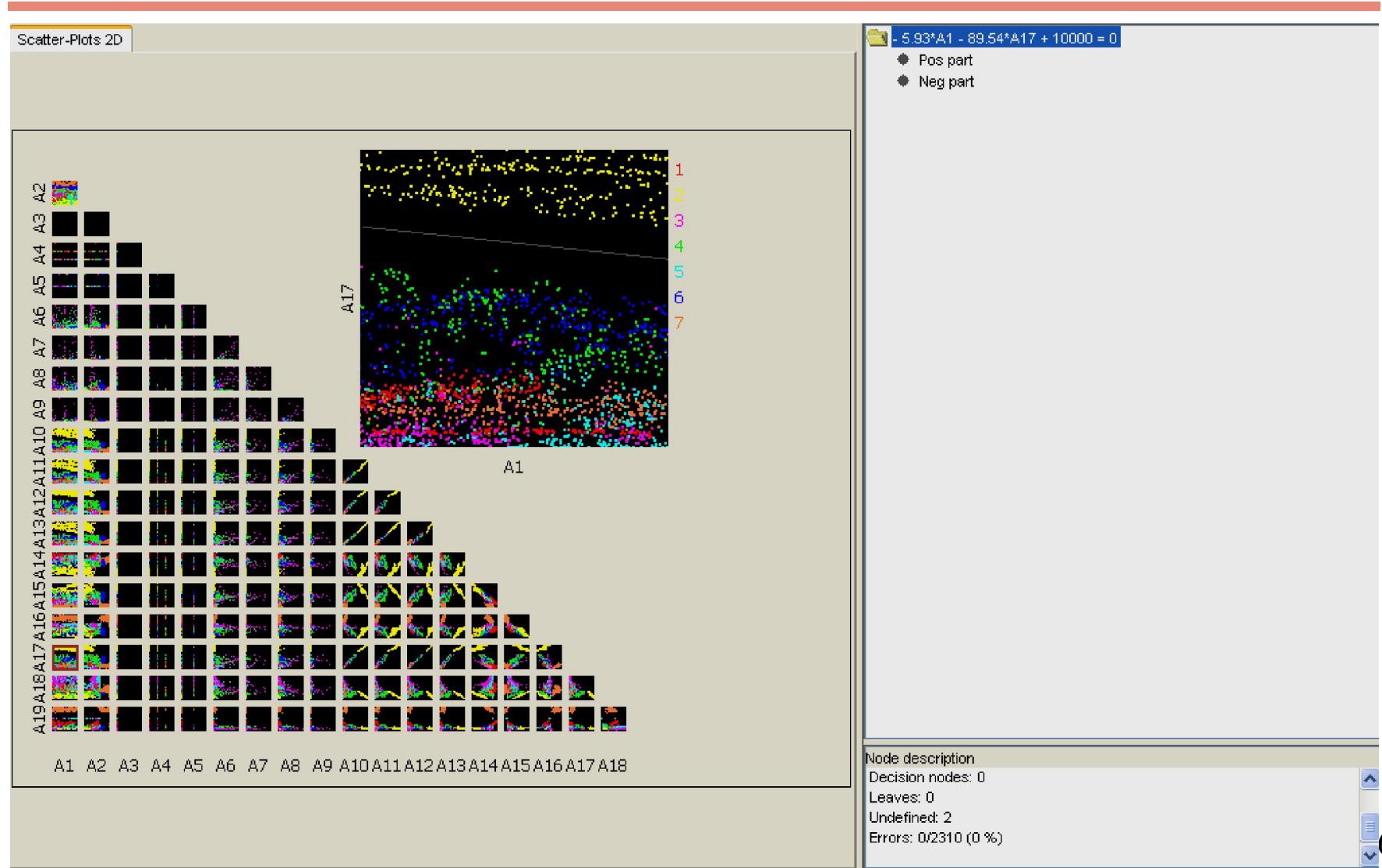
PBC (Ankerst, 1999), CIAD (Poulet, 2001)

Cây quyết định

- CIAD xây dựng cây quyết định
 - dữ liệu segment (2310 phần tử, 19 chiều, 7 lớp)
 - dùng scatterplots 2D
 - split (bi-dim)
 - người sử dụng tự xây dựng cây quyết định qua tương tác với máy tính
 - dễ hiểu
 - có kết hợp các phương pháp để tận dụng các ưu điểm
 - multiview, linking, brushing

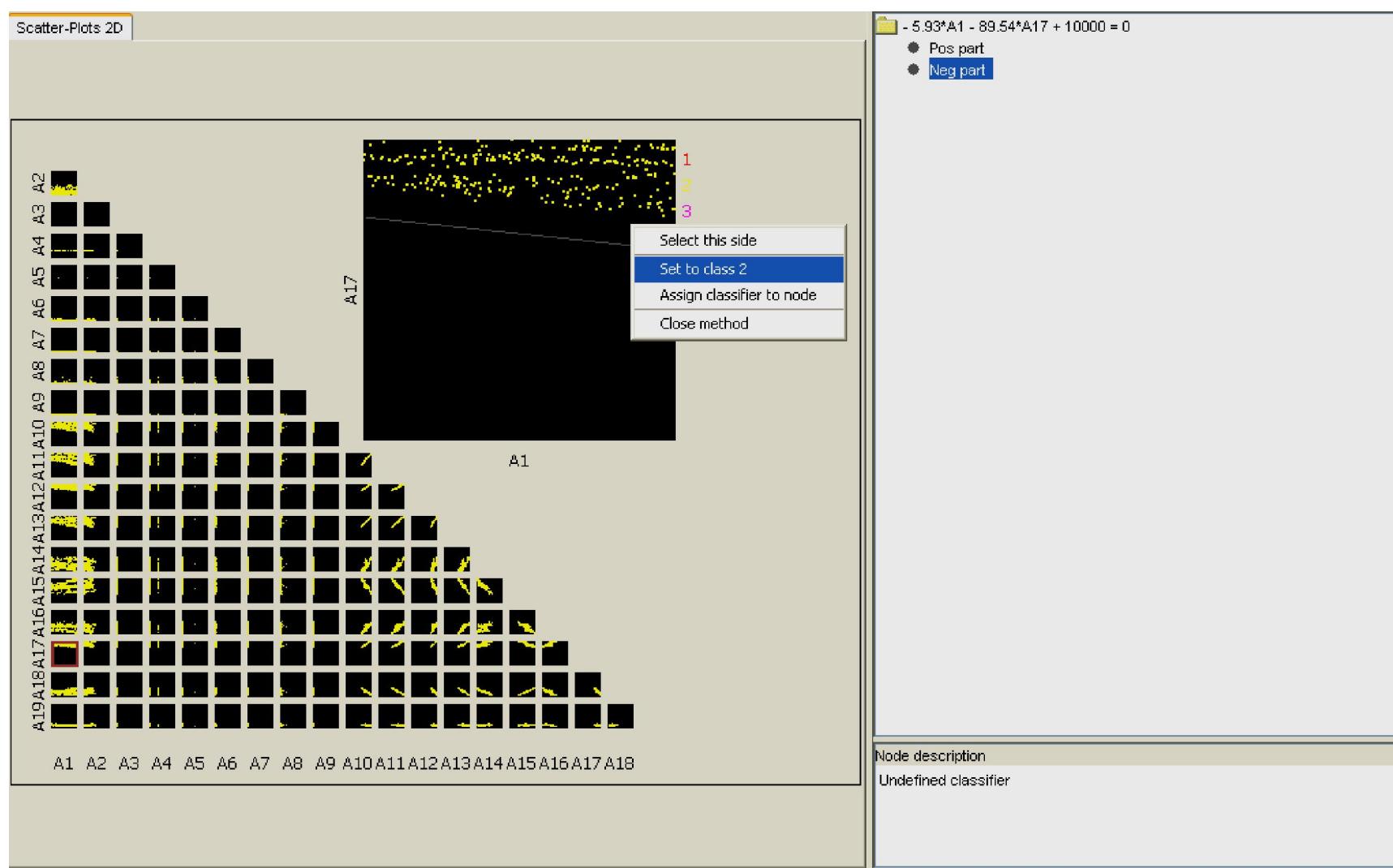
- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

CIAD, Cây quyết định trên dữ liệu segment



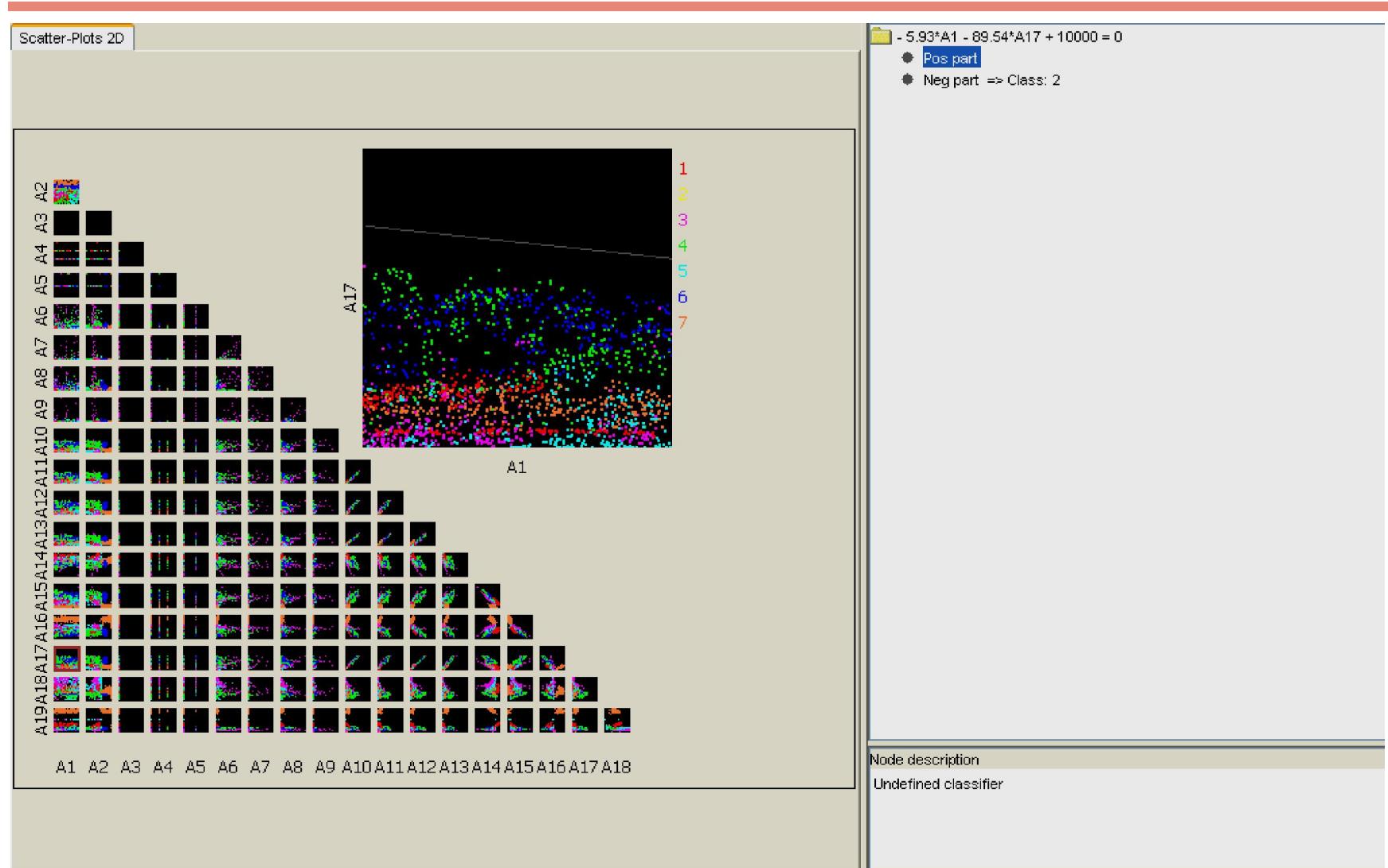
- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

CIAD, Cây quyết định trên dữ liệu segment



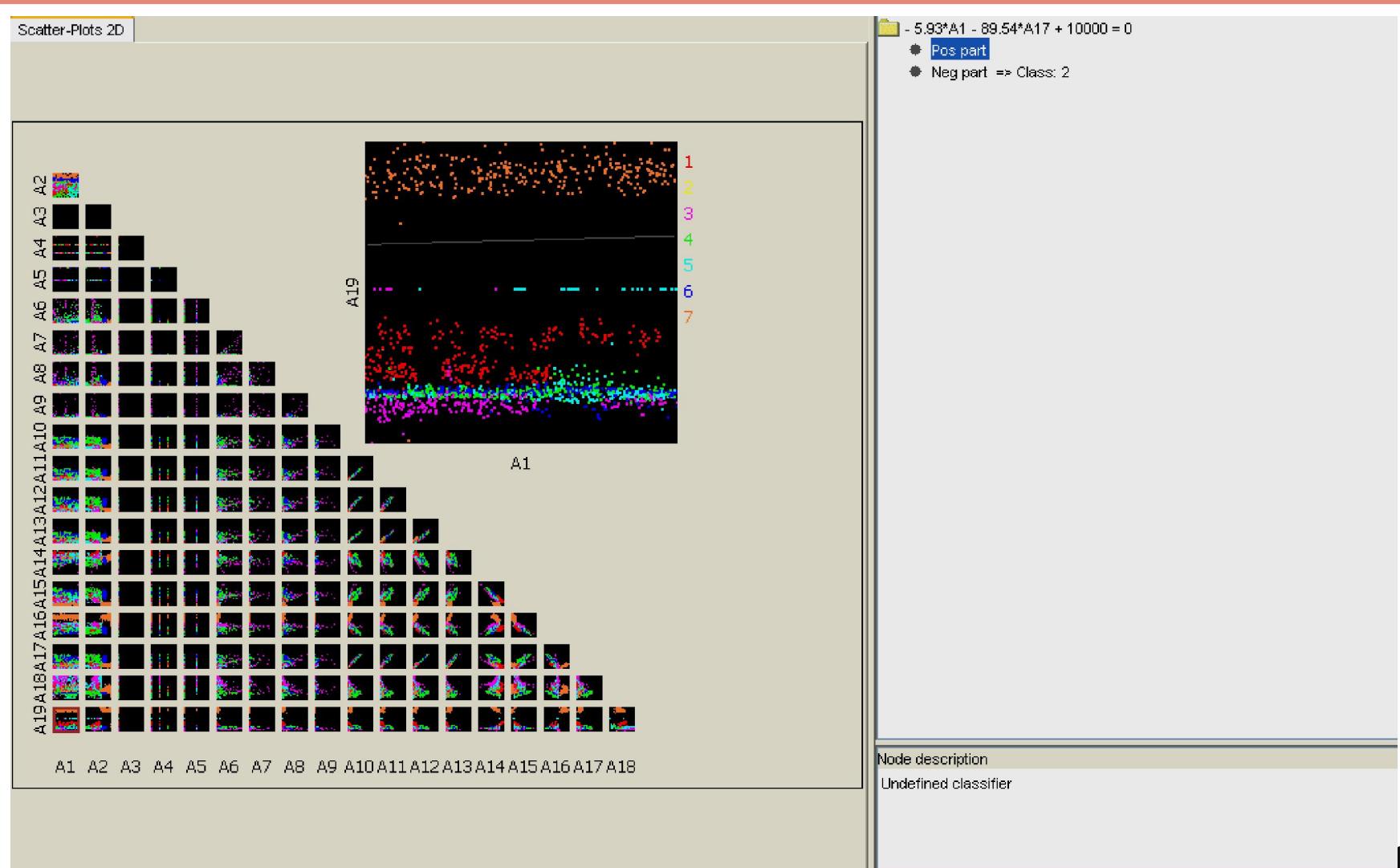
- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

CIAD, Cây quyết định trên dữ liệu segment



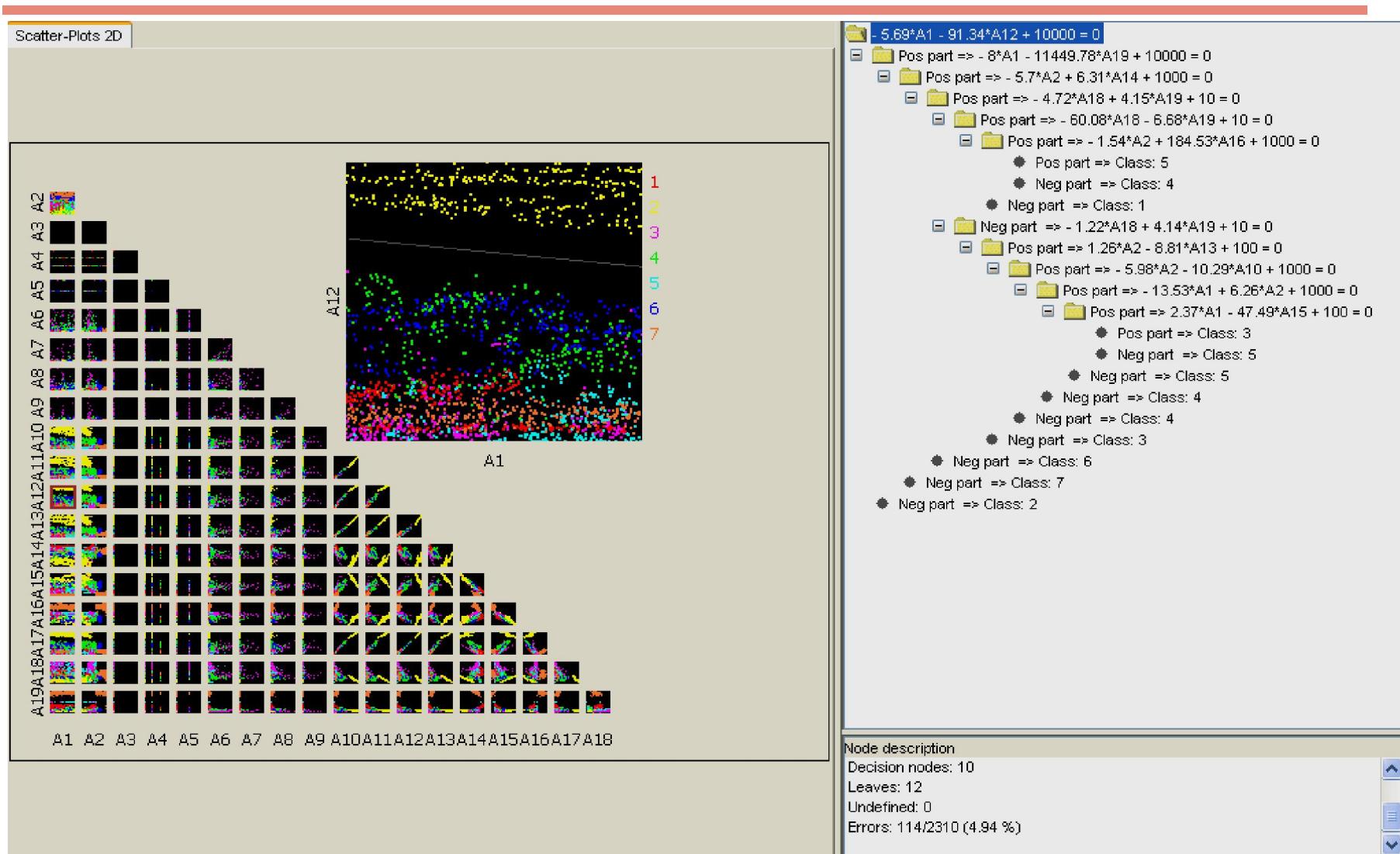
- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

CIAD, Cây quyết định trên dữ liệu segment



- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

CIAD, Cây quyết định trên dữ liệu segment



Hiển thị dữ liệu hỗ trợ cho dịch và giải thích kết quả

- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- Kết luận và hướng phát triển

- Giới thiệu về info viz
- Info viz trong khai mỏ dữ liệu
- kết luận và hướng phát triển

Viz-SVM (Do, 2004), Tree-view (Do & Pham, 2007)

■ Viz-SVM

- dịch và giải thích kết quả sinh ra bởi SVM
- phát hiện các thuộc tính quan trọng của mô hình SVM

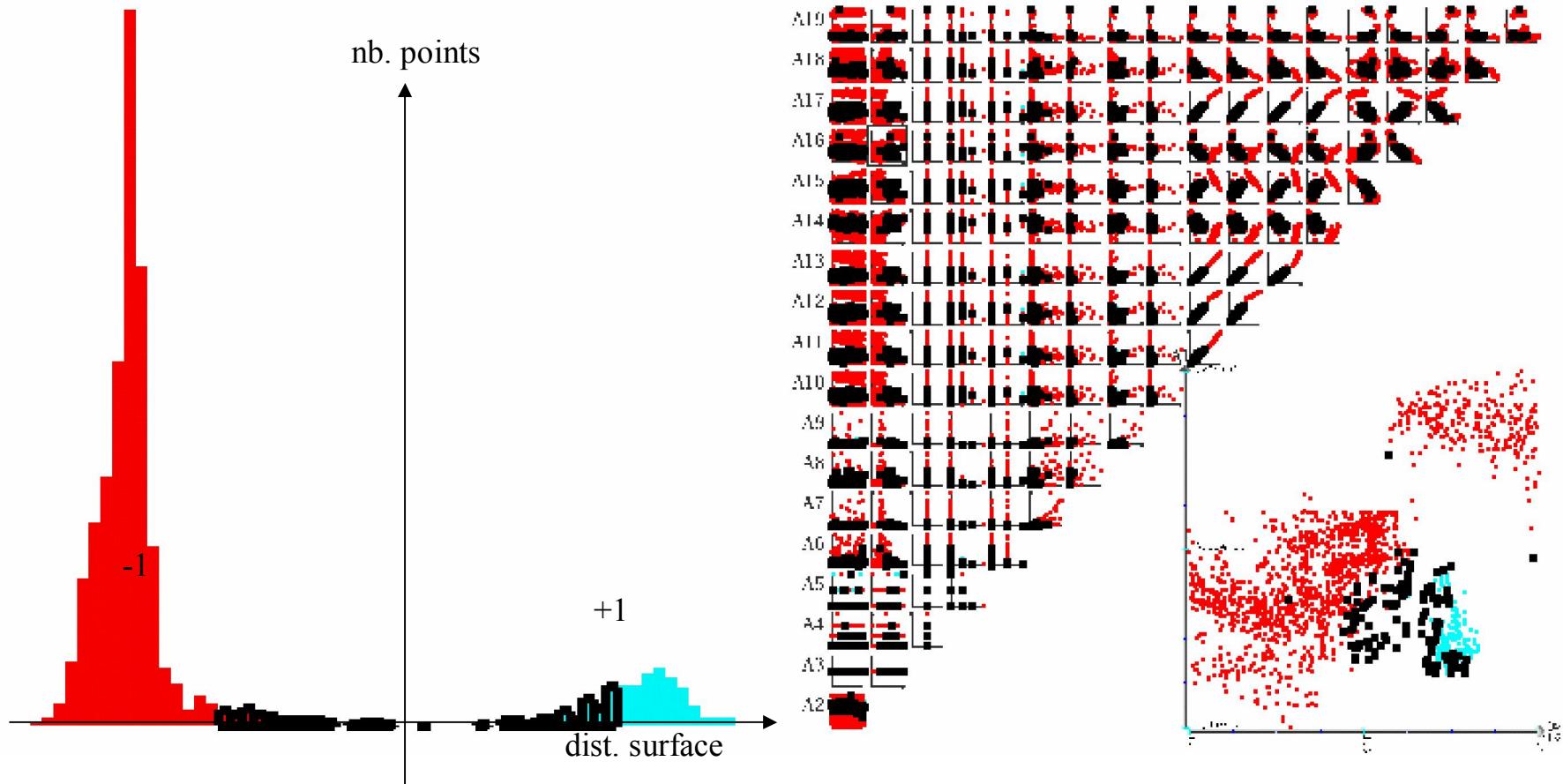
■ Tree-view

- duyệt và khám phá các luật quan trọng sinh ra từ cây quyết định lớn vài trăm đến vài ngàn nút
- giúp cắt nhánh

- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

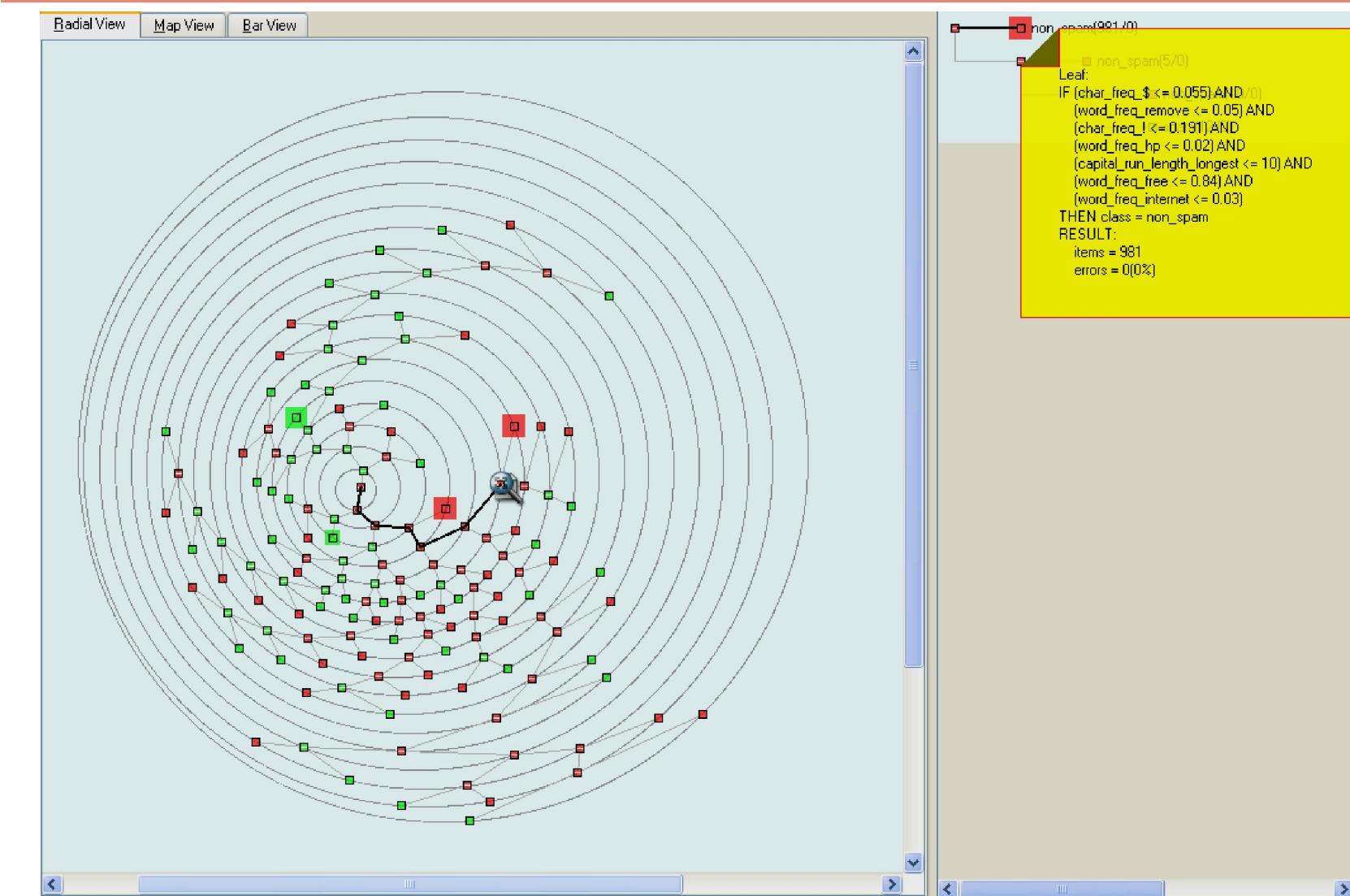
Viz-SVM (Do, 2004)

Histogram + Scatterplots, etc.



- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

Tree-View (Do & Pham, 2007)



Nội dung

- Giới thiệu về info viz
- Info viz trong khai mỏ dữ liệu
- **Kết luận và hướng phát triển**

- Giới thiệu về info viz
- Info viz trong khai mỏ dữ liệu
- kết luận và hướng phát triển

Kết luận

■ hiển thị dữ liệu

- sử dụng khả năng nhận dạng mẫu của người
- hỗ trợ rất nhiều cho khám phá tri thức : trong tiền xử lý dữ liệu, khai mỏ dữ liệu, hậu xử lý
- tuy nhiên khả năng xử lý dữ liệu lớn là rất khó, trong khi sinh ra mô hình dễ hiểu hoặc giúp diễn dịch, giải thích kết quả
- riêng phương pháp học tự động thì có khả năng xử lý dữ liệu lớn, nhưng lại cho kết quả khó hiểu
- ví dụ như xử lý dữ liệu ovarian cancer (253 phần tử, 15154 chiều)
- không có phương pháp hiển thị nào có thể hiển thị được tập dữ liệu, nhưng mô hình xây dựng của SVM sinh ra hyperplane có số chiều 15154 thì không thể diễn dịch

- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

Hướng phát triển[?]

- hợp tác phương pháp hiển thị + học tự động
 - xử lý dữ liệu lớn
 - kết quả dễ hiểu

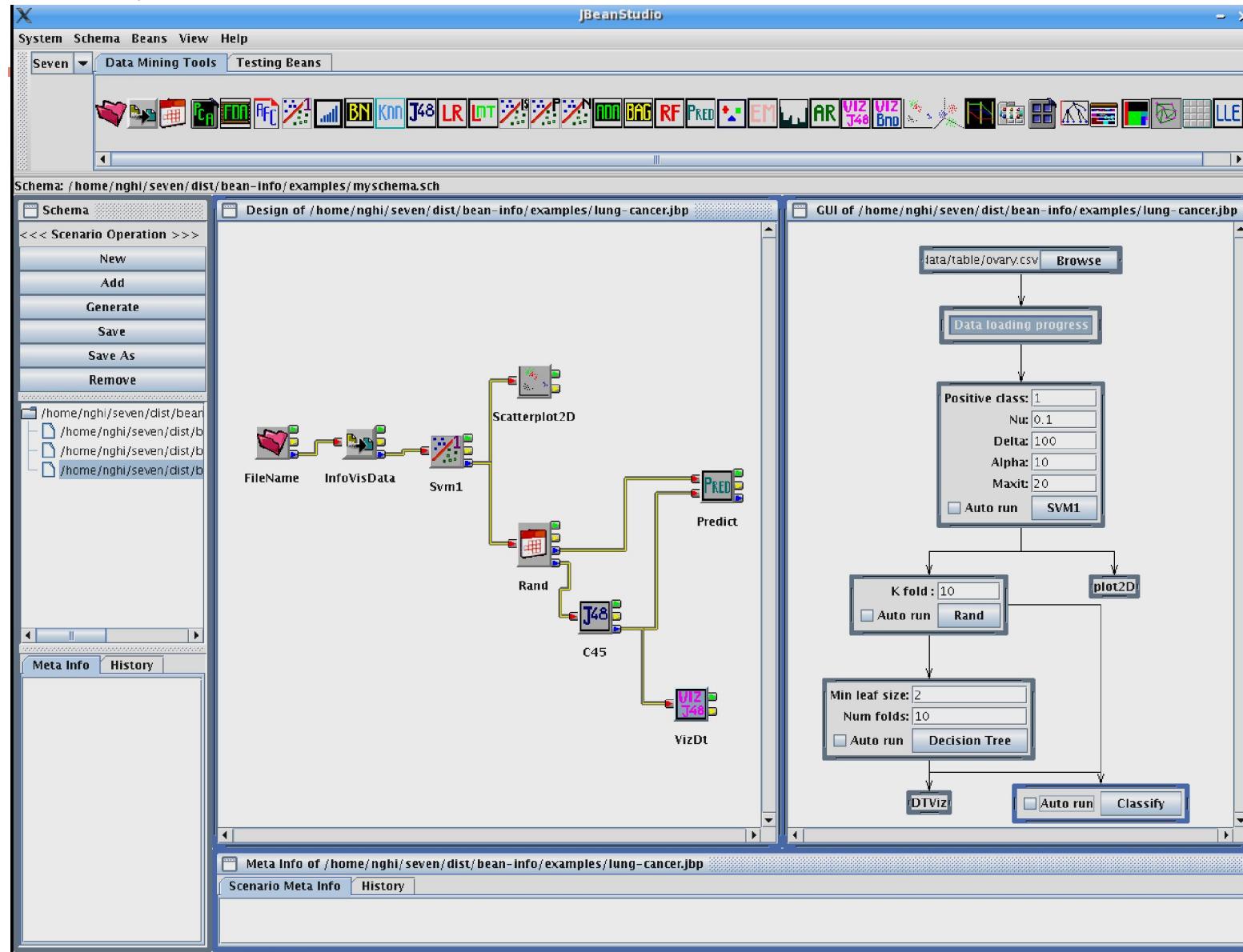
- Giới thiệu về info viz
- Info viz trong khai mỏ dữ liệu
- kết luận và hướng phát triển

V4Miner (Do & Fekete, 2007)

- hợp tác phương pháp hiển thị + học tự động
 - visual programming
 - không cần lập trình
 - khai mỏ dữ liệu = drag/drop/connect components
 - feedback => update params => results => viz => feedback ...

- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

Xử lý dữ liệu Ovarian Cancer



- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

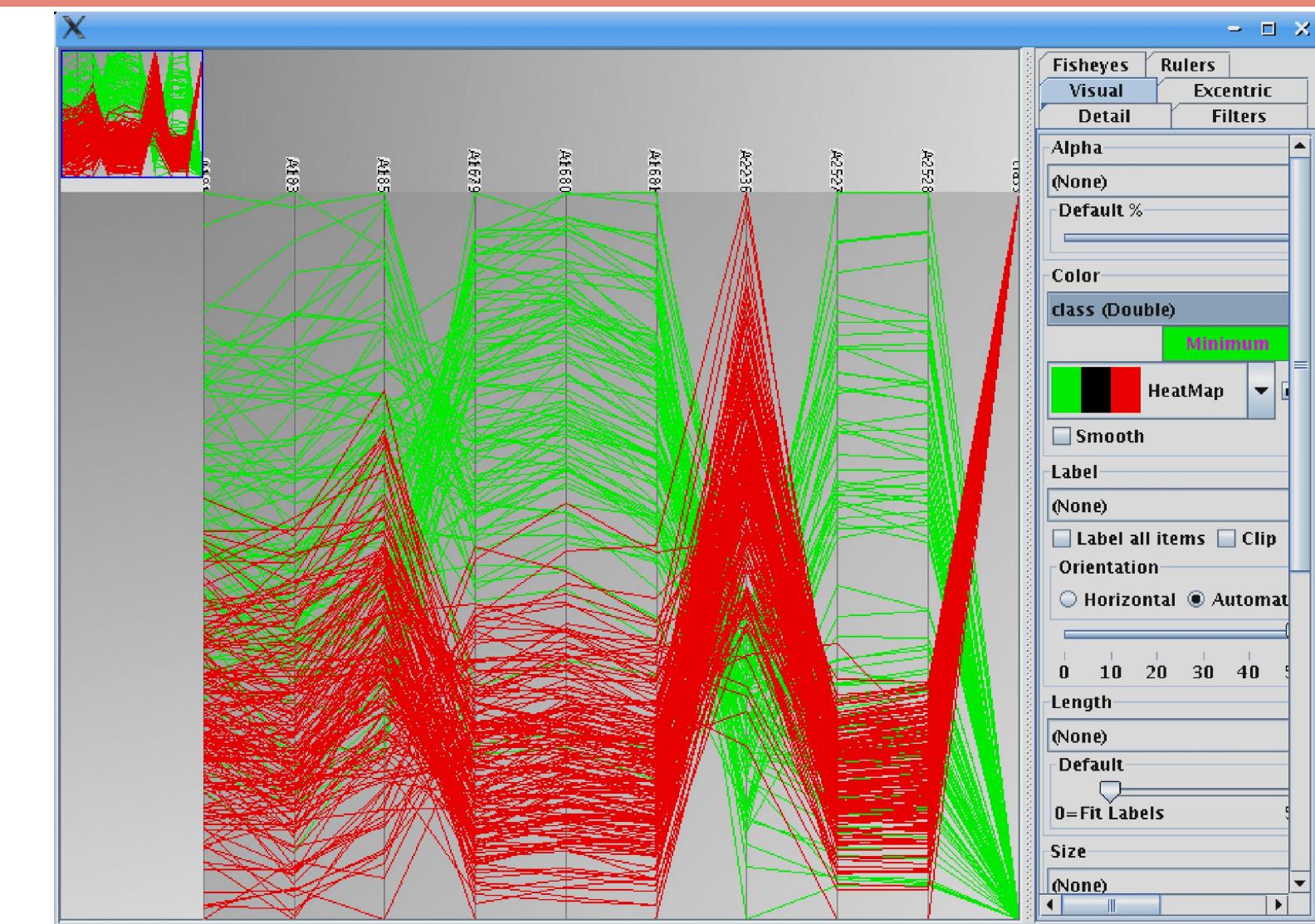
Xử lý dữ liệu Ovarian Cancer

■ Results

- without feature selection (15154 dims): **without error**
- feature selection with SVM-1: **9 dims**
- visualization with scatterplots, parallel coordinates: **without error**
- classification with decision tree C4.5

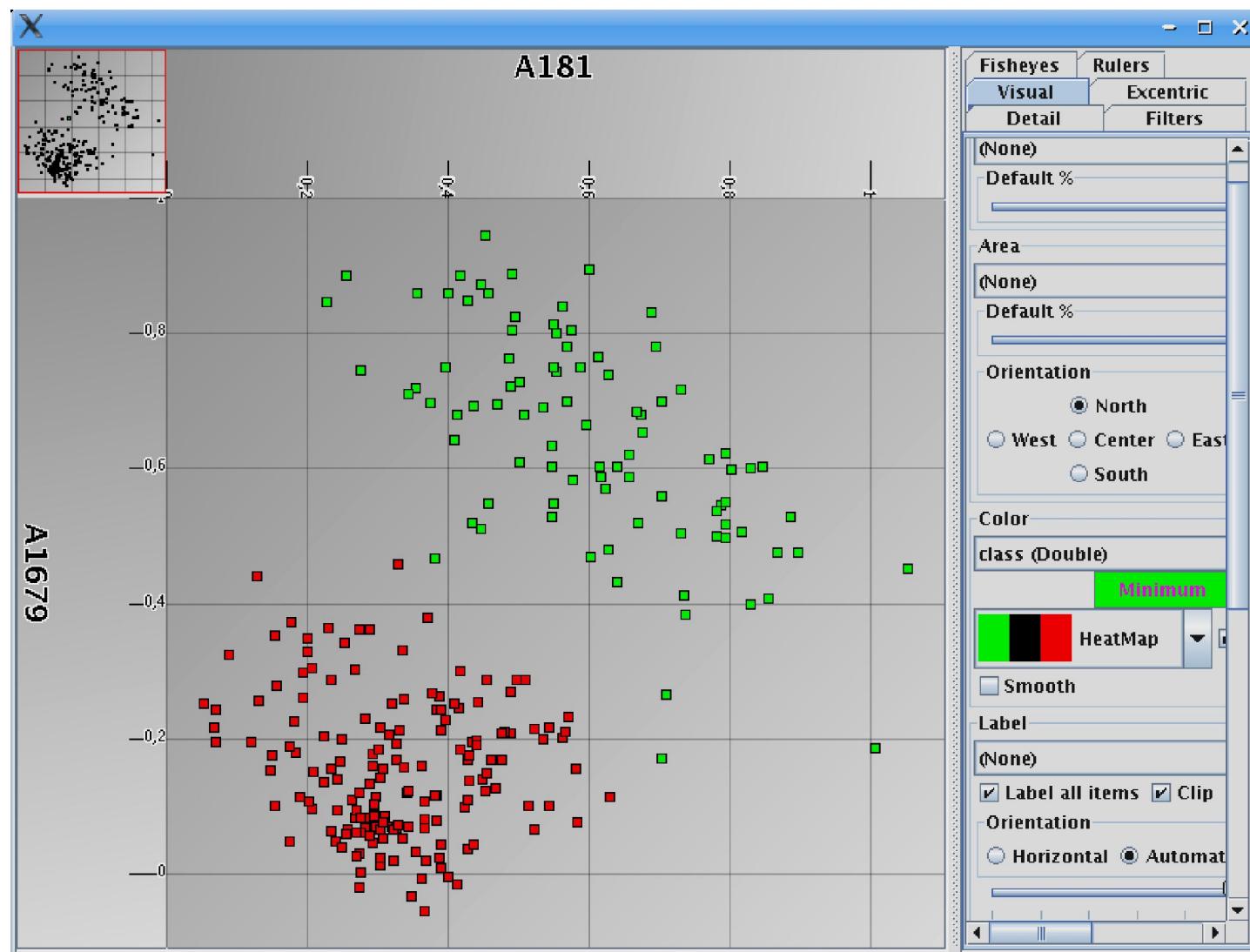
- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

Xử lý dữ liệu Ovarian Cancer



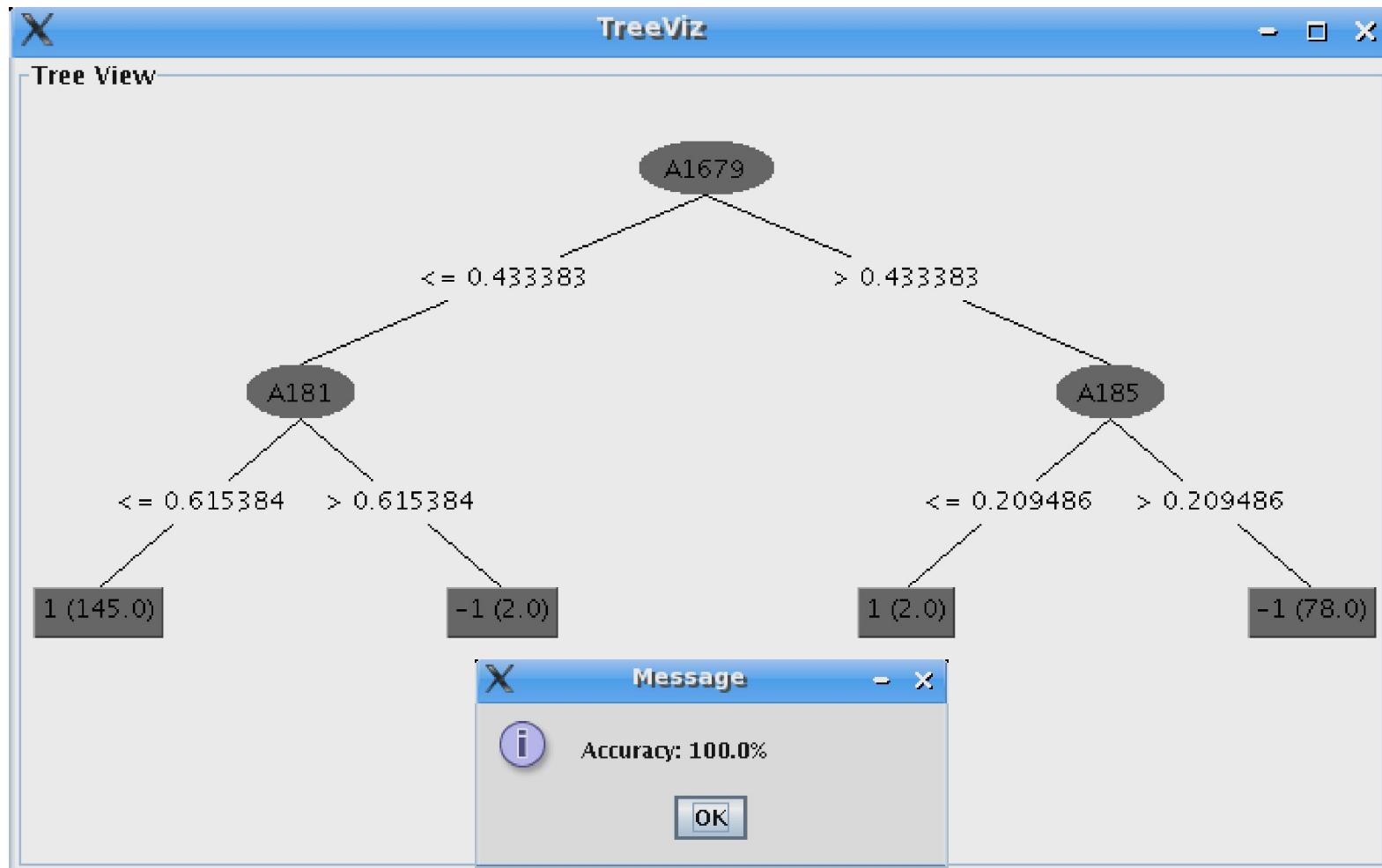
- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

Xử lý dữ liệu Ovarian Cancer



- Giới thiệu về info viz
- Info viz trong khai mở dữ liệu
- kết luận và hướng phát triển

Xử lý dữ liệu Ovarian Cancer





Cám ơn !