KHAI PHÁ DỮ LIỆU

BUỔI THỰC HÀNH 5

Trần Mạnh Tuấn

Bộ môn Hệ thống thông tin, Khoa CNTT

Trường đại học Thủy Lợi

NỘI DUNG

- Giới thiệu về luật kết hợp
- Luật kết hợp trên weka

Giới thiệu về luật kết hợp

Khai phá luật kết hợp:

- Tìm tần số mẫu, mối kết hợp, sự tương quan, hay các cấu trúc nhân quả giữa các tập đối tượng trong các cơ sở dữ liệu giao tác, cơ sở dữ liệu quan hệ, và những kho thông tin khác.

Tính hiểu được: dễ hiểu

Tính sử dụng được: Cung cấp thông tin thiết thực

Tính hiệu quả: Đã có những thuật toán khai thác hiệu quả

Các ứng dụng:

- Phân tích bán hàng trong siêu thị, cross-marketing, thiết kế catalog, loss-leader analysis, gom cụm, phân lớp, ...

Giới thiệu về luật kết hợp

Các khái niệm

Cho $I = \{I_1, I_2, \ldots, I_m\}$ là tập các đơn vị dự liệu. Cho D là tập các giao tác, mỗi giao tác T là tập các đơn vị d dữ liệu sao cho $T \subseteq I$

ịnh nghĩa 1: Ta gọi giao tác T chứa X, với X là tập các đơn vị dữ liệu của I, nếu $X \subseteq T$

ịnh nghĩa 2: Một luật kết hợp là một phép suy diễn có dạng $X \rightarrow Y$, trong đó $X \subset I$, $Y \subset I$ và $X \cap Y = \varnothing$

ịnh nghĩa 3: Ta gọi luật $X \rightarrow Y$ có mức xác nhận(support) là s trong tập giao tác D, nếu có s% giao tác trong D chứa $X \cup Y$.

Ký hiệu: $Supp(X \rightarrow Y) = s$

Giới thiệu về luật kết hợp

Định nghĩa 4:Ta gọi luật $X \rightarrow Y$ là có độ tin cậy c (Confidence) trên tập giao tác D,

Ký hiệu: $c = Conf(X \rightarrow Y) = Supp(X \rightarrow Y)/Supp(X)$

Nhận xét: Các xác nhận và độ tin cậy chính là các xác suất sau:

$Supp(X \rightarrow Y) = P(X \cup Y)$: Xác suất của $X \cup Y$ trong D

$Conf(X \rightarrow Y) = P(Y/X)$: Xác suất có điều kiện

Định nghĩa 5: Cho trớc Min_Supp=s_0 và Min_Conf=c_0

Ta gọi luật $X \rightarrow Y$ là xaỷ ra nếu thỏa:

$Supp(X \rightarrow Y) > s_0$ và $Conf(X \rightarrow Y) > c_0$

Giới thiệu về luật kết hợp

- Thuật toán Apriori
- Thuật toán FP-growth

Luật kết hợp trên weka

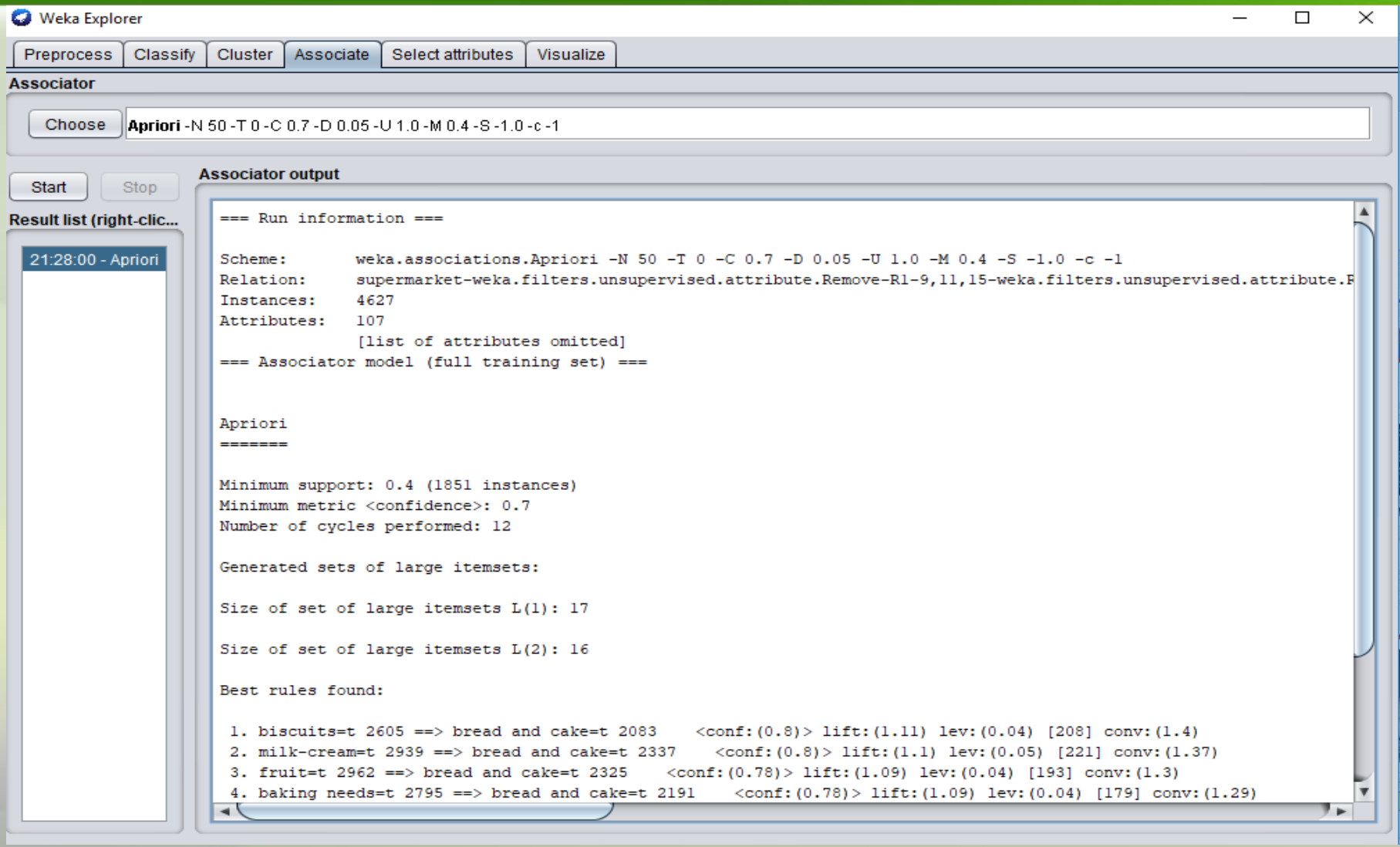
Luật kết hợp trên weka

- Là một chức năng của Explorer
- Hỗ trợ người dùng huấn luyện và kiểm chứng các thuật toán luật kết hợp cơ bản

Các bước thực hiện luật kết hợp

- Bước 1: tại tab Preprocess, chọn tập dữ liệu và tiền xử lý dữ liệu: các trường dữ liệu dạng Nominal. Nếu ở dạng khác thì dùng bộ lọc để chuyển về: NumericToNominal
- Bước 2: Chọn thuật toán luật kết hợp và tham số
- Bước 3: Tiến hành thực hiện thuật toán
- Bước 4: Ghi nhận và phân tích kết quả

Giao diện chính của luật kết hợp



Tổng hợp so sánh luật kết hợp

- **Chạy 1 bộ dữ liệu với các phương pháp thuật toán khác nhau**
- **Chạy thuật toán Apriori với các bộ dữ liệu khác nhau**

THỰC HÀNH