

Khám phá phần tử ngoại lai trong cơ sở dữ liệu :

Luận văn ThS / Đặng Thị Thu Hiền ; Nghd. : TS. Hoàng Xuân
Huân . - H. : ĐHCN, 2005 . - 121 tr. + Đĩa CD - Rom

Tóm tắt: Khái quát về khám phá tri thức trong cơ sở dữ liệu và phần tử ngoại lai.
Trình bày các thuật toán tìm kiếm phần tử ngoại lai dựa trên khoảng cách. Đi sâu
nghiên cứu về phần tử ngoại lai, xác định tham số p và D , đồng thời giới thiệu các
phép biến đổi. Xác định các phần tử ngoại lai cục bộ dựa vào mật độ

	Trang
Trang phụ bìa	0
Lời cảm ơn	1
Mục lục	2
Danh mục từ viết tắt	5
Mở đầu	6
CHƯƠNG 1: KHÁM PHÁ TRI THỨC TRONG CƠ SỞ DỮ LIỆU VÀ PHẦN TỬ NGOẠI LAI	8
1.1 Khám phá tri thức	8
1.2 Các ứng dụng sử dụng kỹ thuật khai thác dữ liệu	10
1.3 Phần tử ngoại lai.	11
1.4. Mối quan hệ giữa phần tử ngoại lai và khai thác dữ liệu.	12
1.5 Ứng dụng của các phần tử ngoại lai	12
CHƯƠNG 2: CÁC THUẬT TOÁN TÌM KIẾM PHẦN TỬ NGOẠI LAI DỰA TRÊN KHOẢNG CÁCH	14
2.1 Định nghĩa các phần tử ngoại lai dựa trên khoảng cách	14
2.2 Thuật toán Nested-Loop	14
2.2.1 Tư tưởng thuật toán	14
2.2.2 Mô tả thuật toán NL	15
2.2.3 Đánh giá độ phức tạp của thuật toán NL	16
2.3 Thuật toán đánh giá theo ô	17

2.3.1 Các khái niệm và tính chất liên quan	17
2.3.2 Thuật toán FindAllOutsM cho các tập dữ liệu trong bộ nhớ chính	18
2.3.2.1 Tư tưởng thuật toán	18
2.3.2.2 Mô tả thuật toán FindAllOutsM (Find All Outliers in Memory)	19
2.3.2.3 Đánh giá độ phức tạp thuật toán trong không gian hai chiều	21
2.3.2.4 Tổng quát cho trường hợp nhiều chiều	23
2.3.2.5 Đánh giá độ phức tạp trong không gian nhiều chiều	24
2.3.3. Tìm kiếm các phần tử ngoại lai DB(p,D) trong các tập dữ liệu lớn, ở bộ nhớ ngoài	25
2.3.3.1 Phân tích tổng quát	25
2.3.3.2 Thuật toán FindAllOutsD cho các phần tử ngoại lai nằm trong bộ nhớ ngoài	27
2.4 Xử lý thực nghiệm.	31
2.4.1 Thiết lập thực nghiệm	31
2.4.2 Thay đổi cỡ của tập dữ liệu	32
2.4.3 Thay đổi giá trị của p	33
2.4.4 Thay đổi chiều dữ liệu và số lượng ô	34
2.5 Tóm tắt	34
CHƯƠNG 3: TRI THỨC SÂU VỀ PHẦN TỬ NGOẠI LAI	36
3.1 Các khái niệm về độ mạnh của các phần tử ngoại lai DB	36
3.2 Các định nghĩa và thuật ngữ	37
3.2.1 Các phần tử ngoại lai <i>mạnh nhất</i> và <i>mạnh nhất Top-u</i>	39
3.2.2 Các phần tử ngoại lai yếu và các phần tử ngoại lai tầm thường	40
3.3 Ví dụ minh họa -các cầu thủ NHL	43
3.4 Thuật toán tìm kiếm các phần tử ngoại lai không tầm thường (FindNonTrivialOuts)	46
3.5 Tóm tắt	49
CHƯƠNG 4: XÁC ĐỊNH THAM SỐ p, D VÀ CÁC PHÉP BIẾN ĐỔI	50
4.1 Giới thiệu	50
4.2 Lấy mẫu và biến đổi thành dạng xấp xỉ chuẩn	51
4.2.1 Lấy mẫu và biến đổi	51
4.2.1.1 Lấy mẫu	51
4.2.1.2 Biến đổi thành dạng xấp xỉ chuẩn	53

4.2.2 Phép tích phân Monte-Carlo	56
4.2.3 Các bộ ước lượng mạnh đơn biến	58
4.2.3.1 Shorth	58
4.2.3.2 MAD	59
4.2.4 Thuật toán KDD_Outliers	61
4.3 Các phép biến đổi không gian mạnh	63
4.3.1 Tổng quát hoá các hàm khoảng cách, các phép biến đổi không gian mạnh	63
4.3.1.1 Các hàm khoảng cách	63
4.3.1.2 Các phép biến đổi không gian mạnh	65
4.3.2 Các bộ ước lượng nhị biến mạnh: hiệp phương sai và hiệp phương sai mạnh	67
4.3.3 Bộ ước lượng Donoho-Stahel (DSE) trong 2-D	68
4.3.3.1 Bộ ước lượng Donoho-Stahel (DSE) trong 2-D	68
4.3.3.2 Các kết quả thí nghiệm 2-D	70
4.3.4 So sánh sự khác nhau giữa các bộ ước lượng, các thuộc tính chính của DSE	78
4.3.4.1 So sánh sự khác nhau giữa các bộ ước lượng	78
4.3.4.2 Các thuộc tính chính của DSE	80
4.3.5 Bộ ước lượng Donoho-Stahel k-D	83
4.3.5.1 Thuật toán Fixed-Angle k-D	83
4.3.5.2 Thuật toán lấy mẫu phụ (SubSampling) k-D	84
4.3.5.3 Các thuật toán ngẫu nhiên đối với việc tính toán DSE k-D	86
4.3.6 Sự đánh giá thí nghiệm	89
4.3.6.1 Các tham số bên trong của các thuật toán	89
4.3.6.2 Xác định tỷ lệ đã cho về mức gọi lại (recall)	93
4.3.6.3 So sánh kích cỡ và số mặt cắt(patches) với tỷ lệ xung đột	94
4.4 Tổng kết	95
CHƯƠNG 5: XÁC ĐỊNH CÁC PHẦN TỬ NGOẠI LAI CỤC BỘ DỰ VÀO MẬT ĐỘ	97
5.1. Một số trở ngại của các cách tiếp cận trước đây.	97
5.2. Định nghĩa các phần tử ngoại lai theo cách nhìn địa phương.	98
5.3 Tính chất của các phần tử ngoại lai cục bộ	101

5.3.1 LOF của các đối tượng nằm sâu trong một vùng	101
5.3.2 Cận trên và cận dưới của LOF	102
5.3.3 Tính chặt chẽ của các cận LOF	104
5.3.4 Cận trên và cận dưới của đối tượng có các lân cận trực tiếp nằm trong nhiều vùng	106
5.4. Sự ảnh hưởng của tham số Minpts.	109
5.4.1 Sự phụ thuộc của LOF theo Minpts	109
5.4.2 Xác định miền của Minpts	110
5.5. Thực nghiệm	112
5.6 Đánh giá độ phức tạp của thuật toán xác định giá trị LOF	114
KẾT LUẬN	116

