1.4.1: Mô tả lớp/khái niệm: Đặc tính hóa và sự phân biệt:

- Các mục dữ liệu có thể liên kết với các lớp hoặc khái niệm.

- Ví dụ như trong cửa hàng điện máy, các lớp sản phẩm bao gồm **‘TV’** và **‘tủ lạnh’**, và khái niệm của khách hàng bao gồm **‘chi trả’** và **‘lượng chi trả’**.

- Nó có thể hữu ích để mô tả các lớp riêng lẻ và các khái niệm bằng những thuật ngữ tóm tắt, ngắn gọn và chính xác. Những mô tả như vậy về một lớp hoặc một khái niệm được gọi là **‘class/concept descriptions’**.

- Những mô tả này có thể được rút ra bằng cách sử dụng (1) đặc tính dữ liệu, bằng cách tóm tắt dữ liệu của lớp đang nghiên cứu (thường được gọi là lớp mục tiêu) theo thuật ngữ chung hoặc (2) phân biệt dữ liệu, bằng cách so sánh lớp mục tiêu với một hoặc một tập hợp của các lớp so sánh (thường được gọi là các lớp tương phản), hoặc (3) cả đặc tính và phân biệt dữ liệu.

**Đặc tính dữ liệu** là sự tóm tắt các đặc điểm hoặc tính năng chung của lớp dữ liệu mục tiêu.

- Dữ liệu tương ứng với lớp do người dùng chỉ định thường được thu thập bằng một truy vấn. Ví dụ: để nghiên cứu đặc điểm của các sản phẩm phần mềm có doanh số tăng 10% trong năm trước, dữ liệu liên quan đến các sản phẩm đó có thể được thu thập bằng cách thực hiện truy vấn SQL trên cơ sở dữ liệu bán hàng.

- Có một số phương pháp để tóm tắt và mô tả dữ liệu hiệu quả. Tóm tắt dữ liệu đơn giản dựa trên các thước đo và sơ đồ thống kê được mô tả trong Chương 2. Hoạt động cuộn lên OLAP dựa trên khối dữ liệu (Phần 1.3.2) có thể được sử dụng để thực hiện tóm tắt dữ liệu do người dùng kiểm soát dọc theo một chiều được chỉ định.

- Đầu ra của đặc tính dữ liệu có thể được trình bày dưới nhiều hình thức khác nhau.

- Ví dụ bao gồm biểu đồ hình tròn, biểu đồ thanh, đường cong, khối dữ liệu đa chiều và bảng đa chiều, bao gồm cả bảng chéo.

**Phân biệt dữ liệu** là sự so sánh các đặc điểm chung của các đối tượng dữ liệu của lớp mục tiêu với các đặc điểm chung của các đối tượng từ một hoặc nhiều lớp tương phản.

- Người dùng có thể chỉ định các lớp mục tiêu và lớp tương phản và các đối tượng dữ liệu tương ứng có thể được truy xuất thông qua các truy vấn cơ sở dữ liệu.

- Ví dụ, một người dùng có thể muốn so sánh đặc điểm chung của các sản phẩm phần mềm có doanh số tăng 10% trong năm ngoái với các sản phẩm có doanh số giảm ít nhất 30% trong cùng kỳ. Các phương pháp được sử dụng để phân biệt dữ liệu cũng tương tự như các phương pháp được sử dụng để mô tả đặc tính dữ liệu.

1.4.1: Khai thác các mô hình, liên kết và mối tương quan thường xuyên:

- Các mẫu phổ biến, như tên cho thấy, là các mẫu xuất hiện thường xuyên trong dữ liệu.

- Có nhiều loại mẫu phổ biến, bao gồm tập mục phổ biến, chuỗi con thường xuyên (còn được gọi là mẫu tuần tự) và cấu trúc con phổ biến.

- Một tập mục thường xuyên thường đề cập đến một tập hợp các mục thường xuất hiện cùng nhau trong một giao dịch. tập dữ liệu—ví dụ: sữa và bánh mì, những thứ thường được nhiều khách hàng mua cùng nhau trong các cửa hàng tạp hóa.

- Một chuỗi xảy ra thường xuyên, chẳng hạn như mẫu mà khách hàng có xu hướng mua đầu tiên là máy tính xách tay, tiếp theo là máy ảnh kỹ thuật số và sau đó là thẻ nhớ, là một mẫu tuần tự (thường xuyên).

- Một cấu trúc con có thể đề cập đến khác nhau các dạng cấu trúc (ví dụ: đồ thị, cây hoặc mạng) có thể được kết hợp với tập mục hoặc chuỗi con.

=> Khai thác các mẫu phổ biến dẫn đến việc phát hiện ra các mối liên hệ và mối tương quan thú vị trong dữ liệu