Thắc mắc về quy trình khai thác tiện ích cao bảo vệ quyền riêng tư và thuật toán FULD

Trần Khắc Bình

20/07/2024

1 Thắc mắc về quy trình khai thác tiện ích cao bảo vệ quyền riêng tư

Sau khi đọc paper, em hiểu mục đích của paper như sau: Trong quá trình khai thác tiện ích cao, các thông tin nhạy cảm của CSDL có thể bị rò rỉ, gây thiệt hại cho chủ của CSDL. Do đó, người ta quan tâm đến vấn đề khai thác tiện ích bảo vệ quyền riêng tư (privacy-preserving utility mining).

Để giải quyết vấn đề này, Paper đã đề xuất thuật toán FULD. FULD sẽ biến CSDL gốc (D) thành CSDL được làm sạch (D'), bằng cách ẩn tất cả các itemset tiện ích cao nhạy cảm của nó. Sau đó, quá trình khai thác tiện ích cao sẽ được tiến hành trên D'.

Thuật toán FULD được chia thành 3 thuật toán nhỏ:

- Thuật toán 1: Xây dựng $UTLDic$ từ CSDL D.
  INPUT: D, tập itemset tiện ích cao nhạy cảm S, tập itemset tiện ích cao không nhạy cảm NS.
  OUTPUT: $UTLDic$

- Thuật toán 2: Ẩn itemset tiện ích cao nhạy cảm.
  INPUT: $UTLDic$, ngưỡng tiện ích tối thiểu δ, tập itemset tiện ích cao nhạy cảm S, D.
  OUTPUT: $sanitized\_UTLDic$

- Thuật toán 3: Tạo CSDL D'.
  INPUT: $sanitized\_UTLDic$
  OUTPUT: D'

$\longrightarrow$ **Thắc mắc**: Mục đích ban đầu của chúng ta là khai thác tiện ích cao bảo vệ quyền riêng tư, hay nói cách khác là tìm tập itemset tiện ích cao không nhạy cảm NS. Nhưng **INPUT** của thuật toán 1 lại yêu cầu NS.

2 Thắc mắc về thuật toán FULD

$\longrightarrow$ **Thắc mắc**: Trong paper không trình bày thuật toán 3: Tạo CSDL D' từ $sanitized\_UTLDic$ (Chỉ nhắc tới tên và công dụng của nó mà không nói tới cách hoạt động)