

TẠO DỮ LIỆU ẨN DANH PHỤC VỤ PHÂN LỚP DỮ LIỆU

Hoàng Anh Tú⁽¹⁾, Trần Minh Triết⁽¹⁾, Dương Anh Đức⁽²⁾

(1) Khoa Công nghệ Thông tin, Trường ĐH KHTN, ĐHQG-HCM

(2) Trường ĐH Công nghệ thông tin, ĐHQG-HCM

Tóm tắt

Khai khoáng dữ liệu có vai trò ngày càng quan trọng trong nhiều lĩnh vực khác nhau. Tuy nhiên, để thông tin rút trích được có ý nghĩa cần một bộ dữ liệu lớn được chia sẻ từ nhiều tổ chức khác nhau. Tuy nhiên, chia sẻ dữ liệu có thể tiết lộ thông tin nhạy cảm của người dùng. Xóa thông tin định danh trực tiếp là không đủ để bảo vệ thông tin nhạy cảm của người dùng. Sweeney đã chỉ ra rằng giải pháp này không thể bảo vệ người dùng khỏi cách tấn công liên kết nhiều dữ liệu khác nhau. Một số thuật toán đã được đưa ra để tạo dữ liệu ẩn danh có chất lượng dữ liệu tối ưu nhưng các thuật toán này có chi phí quá lớn và khó áp dụng vào thực tế nên một số thuật toán khác áp dụng heuristic để giải tìm dữ liệu ẩn danh nhưng vẫn còn một số hạn chế. Trong bài báo này, chúng tôi đề xuất một thuật toán tạo dữ liệu ẩn danh thỏa ràng buộc k -Anonymity với chi phí nhỏ hơn có thể áp dụng vào thực tế.