

# THÔNG TIN VỀ LUẬN ÁN

Tên đề tài luận án	<b>Phát triển các thuật toán khai thác mẫu và mẫu đóng trên cơ sở dữ liệu định lượng</b>
Chuyên ngành	Khoa học máy tính
Mã số ngành	62-48-01-01
Họ tên nghiên cứu sinh	Bùi Danh Hùng
Người hướng dẫn khoa học	PGS.TS. Võ Đình Bảy PGS.TS Nguyễn Hoàng Tú Anh
Cơ sở đào tạo	Đại học Công nghệ Thông tin – ĐHQG TP.HCM

## 1. TÓM TẮT

Luận án trình bày tổng quan, tìm hiểu cơ sở lý thuyết về khai thác mẫu phổ biến có trọng số trên cơ sở dữ liệu định lượng. Trên cơ sở đó, luận án đề xuất các phương pháp mới hiệu quả để giải quyết các bài toán và vượt qua các thách thức hiện có trong khai thác mẫu phổ biến có trọng số như là tập mẫu kết quả quá lớn, vấn đề luật dư thừa hay vấn đề khai thác mẫu theo định hướng người dùng hay theo thời gian thực. Bốn bài toán cụ thể được tập trung nghiên cứu giải quyết trong luận án là các bài toán khai thác mẫu phổ biến có trọng số, khai thác mẫu phổ biến đóng có trọng số, khai thác top-rank-k mẫu phổ biến có trọng số và khai thác mẫu phổ biến có trọng số theo luồng dữ liệu. Các thuật toán được đề xuất trong luận án cho hiệu quả vượt trội so với các thuật toán hiện có trên các mặt thời gian chạy, bộ nhớ sử dụng và khả năng mở rộng khi khai thác các dạng mẫu phổ biến hướng trọng số trong các bài toán nói trên.

Kết quả nghiên cứu của luận án đã được công bố trên 4 bài tạp chí SCIE bao gồm Expert Systems with Applications (Q1, 2018), Applied Intelligence (Q2, 2020), Knowledge-Based Systems (Q1, 2020), IEEE Access (Q1, 2021) và 3 bài hội thảo trong nước và quốc tế (SMC-2016, FAIR-2016 và @-2017).

## 2. CÁC ĐÓNG GÓP KHOA HỌC CỦA LUẬN ÁN

Các đóng góp khoa học của luận án bao gồm:

- Đề xuất cấu trúc cây WN-tree và cấu trúc dữ liệu WN-list, từ đó đưa ra thuật toán NFWI để khai thác hiệu quả mẫu phổ biến có trọng số. Cấu trúc WN-list có một số ưu điểm như phép giao WN-list có độ phức tạp tuyến tính, WN-list có khả năng tự rút gọn khi thực hiện phép giao, độ hỗ trợ trọng số của mẫu được tính nhanh dựa trên WN-list của mẫu đó.
- Đề xuất định lý tia nhánh dựa trên quan hệ tổ tiên WN-list, từ đó đưa ra thuật toán NFWCI khai thác hiệu quả mẫu phổ biến đóng có trọng số dựa trên cấu trúc WN-list.
- Đề xuất thuật toán TFWIN<sup>+</sup> khai thác hiệu quả top-rank-k mẫu phổ biến có trọng số dựa trên cấu trúc WN-list và các chiến lược tăng ngưỡng và tia nhánh sớm.
- Đề xuất cấu trúc SWN-tree, cải tiến từ cấu trúc cây WN-tree, để lưu trữ và bảo trì hiệu quả thông tin của các cửa sổ dữ liệu khi trượt trên luồng dữ liệu. Từ đó đưa ra thuật toán FWPODS khai thác hiệu quả mẫu phổ biến có trọng số theo luồng dữ liệu dựa trên mô hình cửa sổ trượt.

### **3. CÁC HƯỚNG PHÁT TRIỂN TRONG TƯƠNG LAI**

Hướng phát triển nghiên cứu trong tương lai sẽ tập trung giải quyết một số bài toán khai thác mẫu trên cơ sở dữ liệu có trọng số như là: khai thác tập đối đại có trọng số, khai thác mẫu phổ biến đóng có trọng số trên cơ sở dữ liệu tăng trưởng, khai thác mẫu phổ biến có trọng số trên cơ sở dữ liệu không chắc chắn và nghiên cứu triển khai các giải pháp khai thác mẫu phổ biến có trọng số trên hệ thống đa nhân và hệ thống phân tán. Chúng tôi cũng sẽ nghiên cứu triển khai các ứng dụng sử dụng nền tảng khai thác mẫu có trọng số như khai thác đồ thị, khai thác mạng xã hội, khai thác dữ liệu văn bản và khai thác dữ liệu IoT.

**CÁN BỘ HƯỚNG DẪN**

**NGHIÊN CỨU SINH**

**PGS.TS. Võ Đình Bảy**

**PGS.TS. Nguyễn Hoàng Tú Anh**

**Bùi Danh Hùng**