

# THÔNG TIN LUẬN ÁN

Đề tài nghiên cứu: **Phát hiện những điểm thay đổi và chuỗi con bất thường trên dữ liệu chuỗi thời gian**

Chuyên ngành:	Khoa học máy tính
Mã số chuyên ngành:	62.48.01.01
Họ và tên nghiên cứu sinh:	Huỳnh Thị Thu Thủy
Tập thể hướng dẫn:	PGS. TS. Dương Tuấn Anh TS. Võ Thị Ngọc Châu
Cơ sở đào tạo:	Trường Đại học Bách Khoa – ĐHQG TP. HCM

## 1. TÓM TẮT

Dữ liệu chuỗi thời gian hiện diện trong rất nhiều lĩnh vực ứng dụng như y học, tài chính, thương mại điện tử, khí tượng thủy văn, v.v... Khai phá dữ liệu và khám phá tri thức từ dữ liệu chuỗi thời gian ngày càng được quan tâm; trong đó phát hiện chuỗi con bất thường nhất trên dữ liệu chuỗi thời gian là một bài toán có nhiều thách thức cần được xem xét. Những thách thức này là: (1). cần biết trước chiều dài chuỗi con bất thường, (2). kết quả chuỗi con bất thường tìm được phải chính xác với chi phí hữu hiệu, (3). đặc biệt là ứng phó với dữ liệu siêu lớn và yêu cầu đáp ứng tức thời. Hiện tại, các công trình nghiên cứu về bài toán phát hiện chuỗi con bất thường nhất trong dữ liệu chuỗi thời gian vẫn chưa giải quyết các thách thức trên hiệu quả, nhất là đối với dữ liệu chuỗi thời gian dạng luồng.

Do đó, mục tiêu của luận án là phát triển giải pháp hiệu quả cho bài toán tìm chuỗi con bất thường nhất trên dữ liệu chuỗi thời gian trong ngữ cảnh của những thách thức đã nêu trên. Định hướng của các giải pháp là dựa vào các điểm thay đổi để phân đoạn chuỗi thời gian và phát hiện chuỗi con bất thường dựa vào các phân đoạn được rút trích.

Thực nghiệm cho thấy hướng tiếp cận dựa vào phân đoạn giúp xác định chính xác chuỗi con bất thường với chi phí hữu hiệu và không cần biết trước chiều dài chuỗi con bất thường. Ngoài ra, hướng tiếp cận dựa vào phân đoạn còn giúp giải hiệu quả bài toán phát hiện bất thường cả trên dữ liệu chuỗi thời gian dạng luồng. Luận án có 09 bài báo đã công bố, gồm

02 bài đăng trong danh mục tạp chí quốc tế thuộc (01 bài thuộc danh mục SCIE-Q2 và 01 bài thuộc danh mục SCIE-Q4) và 07 bài đăng trong các kỷ yếu hội nghị khoa học quốc tế.

## 2. CÁC ĐÓNG GÓP CHÍNH CỦA LUẬN ÁN

- **Các giải thuật đề xuất cải tiến: I-HOTSAX, Hash\_DD, và KBF\_GPU** giúp cải thiện đáng kể hiệu suất của các giải thuật phát hiện chuỗi con bất thường nhất trên dữ liệu chuỗi thời gian theo hướng tiếp cận dựa vào cửa sổ trượt. Riêng giải thuật KBF\_GPU dựa vào công nghệ GPU có tốc độ thực thi cao, giúp thích ứng với quy mô dữ liệu lớn.
- **Đề xuất mới các giải thuật EP-ILeader, EP-ALeader, EP-Leader-DTW, SEP-Leader-DTW** giúp giải hiệu quả cho bài toán phát hiện chuỗi con bất thường nhất trên dữ liệu chuỗi thời gian dạng tĩnh và dữ liệu chuỗi thời gian dạng luồng theo hướng tiếp cận dựa vào phân đoạn trên độ đo Euclid và khoảng cách DTW.
- **Đề xuất mới các giải thuật TopK-EP-ILeader, TopK-EP-ALeader và TopK-EP-Aleader-S** nhằm phát hiện  $k$  chuỗi con bất thường nhất trên dữ liệu chuỗi thời gian dạng tĩnh và dạng luồng theo hướng tiếp cận dựa vào phân đoạn.
- **Đề xuất hướng tiếp cận mới EPL\_S\_X** ứng dụng kết quả phát hiện chuỗi con bất thường nhất vào bài toán dự báo dữ liệu chuỗi thời gian.
- **Đề xuất cải tiến giải thuật gom cụm I-Leader và A-Leader** hiệu quả cho bài toán gom cụm gia tăng dữ liệu chuỗi thời gian.
- **Đề xuất mới độ đo PALS** đánh giá chất lượng của các phương pháp phân đoạn chuỗi thời gian.

## 3. NHỮNG VẤN ĐỀ CÒN BỎ NGỎ CẦN TIẾP TỤC NGHIÊN CỨU

Các kết quả nghiên cứu trong luận án đã giải quyết được hết các mục tiêu nghiên cứu được đề ra ban đầu. Tuy nhiên, luận án vẫn còn một số hướng có thể phát triển thêm:

- Mở rộng KBF\_GPU để có thể làm việc được với khoảng cách DTW và có thể khai phá motif (chuỗi được lặp lại nhiều nhất) trên dữ liệu chuỗi thời gian.
- Nghiên cứu áp dụng hướng lập trình phân tán dựa vào Apache Spark hay Map Reduce cho giải thuật KBF.
- Mở rộng hướng tiếp cận EPL\_S\_X để *dự báo chuỗi thời gian dạng luồng* thuộc nhiều miền ứng dụng cần dự báo theo thời gian thực.
- Áp dụng hướng tiếp cận EPL\_S\_X cho phương pháp dự báo dựa trên *mạng nơ ron học sâu* (deep neural networks).
- Mở rộng phương pháp EP-Leader-DTW để phát hiện các ảnh bất thường trong cơ sở dữ liệu hình ảnh trong đó hình ảnh đã được chuyển thành chuỗi thời gian. Ứng dụng này dự kiến sẽ sử dụng khoảng cách DTW bất biến xoay vòng. Bên cạnh đó, giải thuật EP-Leader-DTW cũng sẽ được cải thiện bằng cách sử dụng bộ kỹ thuật cho khoảng cách DTW có tên là UCR-DTW để tăng tốc độ tính toán khoảng cách DTW hơn nữa.

CÁN BỘ HƯỚNG DẪN 1

CÁN BỘ HƯỚNG DẪN 2

NGHIÊN CỨU SINH

PGS.TS. Dương Tuấn Anh

TS. Võ Thị Ngọc Châu

Huỳnh Thị Thu Thủy