



**ĐỀ CƯƠNG KHÓA LUẬN TỐT NGHIỆP**

**PHÂN TÍCH HIỆU NĂNG QUY TRÌNH TỪ  
DỮ LIỆU BẰNG PHƯƠNG PHÁP XEM XÉT  
NHIỀU ĐỐI TƯỢNG**

*(Analyzing process performance from data using object-centric method)*

## **1 THÔNG TIN CHUNG**

**Người hướng dẫn:**

- ThS. Trần Trung Kiên
- TS. Nguyễn Ngọc Thảo (Khoa Công nghệ Thông tin)

**Nhóm sinh viên thực hiện:**

1. Lê Thị Thùy Dương (MSSV: 20120063)
2. Nguyễn Dương Tuấn Phương (MSSV: 20120166)

**Loại đề tài:** Nghiên cứu.

**Thời gian thực hiện:** Từ 01/2024 đến 06/2024.

## 2 NỘI DUNG THỰC HIỆN

### 2.1 Giới thiệu về đề tài

Hiện nay, khai thác quy trình từ dữ liệu đang trở thành một xu hướng phổ biến trong các doanh nghiệp, được áp dụng nhằm mục đích tăng cường chất lượng của hoạt động kinh doanh. Và điều này đã dẫn đến sự xuất hiện của bài toán có ý nghĩa quan trọng là bài toán phân tích hiệu năng của quy trình nhằm tối ưu hóa quy trình.

Bài toán phân tích hiệu năng quy trình từ dữ liệu được phát biểu như sau:

- Cho đầu vào là dữ liệu log của một quy trình cần phân tích hiệu năng.
- Yêu cầu: Mô hình hóa được quy trình từ dữ liệu đầu vào, tính toán các chỉ số hiệu năng (ví dụ như thời gian chờ trung bình, thời gian thực thi trung bình, ...) cho từng bước trong quy trình và hiển thị các chỉ số này cùng với mô hình quy trình tương ứng.

Giải quyết được bài toán này sẽ đem đến một ứng dụng tiềm năng là hỗ trợ các doanh nghiệp, nhà đầu tư, và người quản lý trong việc có thể hiểu rõ hơn về quy trình hoạt động nghiệp vụ của họ. Đồng thời, họ có thể phát hiện các điểm thiếu hiệu quả trong quy trình vận hành, cũng như nhận biết các biến thể sai lệch so với thiết kế ban đầu. Những phát hiện này có thể giúp họ điều chỉnh lại quy trình làm việc, phân bổ tài nguyên hiệu quả hơn. Từ đó giảm chi phí, tối ưu hóa hoạt động, nâng cao chất lượng, tăng sự cạnh tranh và thích nghi nhanh chóng với sự biến đổi trong môi trường kinh doanh.

Tuy nhiên, bài toán cũng tồn tại một số khó khăn và thách thức nhất định. Đầu tiên là yêu cầu về chất lượng của dữ liệu, điều này đóng vai trò quan trọng trong khai thác quy trình, đặc biệt là với bài toán phân tích hiệu năng. Để đảm bảo hiệu quả và các kết quả phân tích hiệu năng chính xác thì đòi hỏi cần có nguồn dữ liệu đầy đủ và chất lượng cao. Sự không chính xác hay thiếu sót trong dữ liệu có

thể dẫn đến kết quả sai lệch về hiệu năng của quy trình, từ đó tạo ra những nhận thức sai lệch từ kết quả đó. Ngoài ra, các phương pháp tiếp cận truyền thống để mô hình hóa quy trình thường không thể hiện đầy đủ những gì thực tế xảy ra bên trong của quy trình. Với các quy trình phức tạp với nhiều hoạt động và đối tượng phức tạp, khó có thể thu được những hiểu biết hữu ích từ các mô hình quy trình sử dụng phương pháp truyền thống.

Trong thời gian gần đây, một phương pháp tiếp cận đạt được kết quả tốt trong bài toán phân tích hiệu năng quy trình đó là sử dụng phương pháp xem xét nhiều đối tượng (Object-Centric). Đây sẽ là phương pháp tiếp cận mà chúng em chọn để tìm hiểu cho đề tài.

## 2.2 Mục tiêu đề tài

- Hiểu rõ bối cảnh nghiên cứu về bài toán phân tích hiệu năng quy trình từ dữ liệu (hiện nay, đã xuất hiện những phương pháp tiếp cận nào để giải quyết vấn đề này? Mỗi phương pháp có ý tưởng, kết quả, cũng như ưu và nhược điểm như thế nào?). Từ đó chọn ra một phương pháp tốt (ứng với một bài báo uy tín), có tiềm năng phát triển trong tương lai và cũng khả thi để hoàn thành trong thời gian của khóa luận để tìm hiểu sâu.
- Hiểu rõ về lý thuyết của phương pháp tiếp cận đã lựa chọn (trên cơ sở hiểu rõ lý thuyết nền tảng về khai thác quy trình).
- Cài đặt lại thuật toán theo phương pháp tiếp cận đã chọn để ra được các kết quả trong bài báo tương ứng; có thể tiến hành thêm một số thí nghiệm ngoài bài báo để thấy rõ hơn về ưu/nhược điểm của thuật toán theo phương pháp tiếp cận được sử dụng.
- Trên cơ sở đã hiểu rõ thuật toán của phương pháp tiếp cận, nếu còn thời gian thì có thể xem xét các cải tiến có thể có.
- Rèn luyện được các kỹ năng: Tư duy logic, lên kế hoạch, làm việc nhóm, trình

bày, ...

## 2.3 Phạm vi của đề tài

- Để giải quyết bài toán phân tích hiệu năng quy trình, việc cần thiết nhất là sử dụng dữ liệu thực tế như một nguồn thông tin cơ sở. Trong nghiên cứu này, chúng em sẽ sử dụng đầu vào là dữ liệu log từ các quy trình thực tế, thường được sử dụng phổ biến trong các bài toán về quy trình. Cụ thể là bộ dữ liệu về quy trình đăng ký khoản vay của một viện Tài chính Hà Lan, được cung cấp bởi Hội thảo Quốc tế về Business Process Intelligence (BPI) lần thứ 13 diễn ra vào năm 2017.
- Đề tài sẽ tập trung vào việc hiểu và tái cài đặt thuật toán từ một bài báo uy tín. Bên cạnh đó, chúng em cũng có kế hoạch thực hiện các thí nghiệm bổ sung ngoài phạm vi của bài báo để có cái nhìn rõ ràng hơn về ưu/nhược điểm của thuật toán. Chúng em quyết định giới hạn phạm vi của đề tài vì hai lý do chính: (i) việc hiểu thuật toán và các kiến thức nền tảng cũng như tái cài đặt thuật toán mất rất nhiều thời gian, và (ii) chúng em nhận thấy rằng, chỉ thông qua việc hiểu rõ thuật toán và các kiến thức nền tảng, chúng em mới có thể đạt được các cải tiến thực sự trong tương lai và áp dụng thuật toán cho các bài toán khác. Tất nhiên, trong quá trình làm khóa luận, nếu có đủ thời gian, chúng em sẽ xem xét và thử nghiệm các cải tiến; tuy nhiên, điều này không phải là mục tiêu chính.

## 2.4 Cách tiếp cận dự kiến

Dưới đây chúng em sẽ trình bày một số phương pháp để giải quyết bài toán phân tích hiệu năng quy trình mà chúng em đã tìm hiểu được cho đến thời điểm hiện tại, cũng như là phương pháp mà chúng em dự kiến sẽ chọn để tập trung tìm hiểu sâu.

- Một nghiên cứu đáng chú ý trong nhóm các phương pháp truyền thống là “Performance analysis of business processes through process mining” [?]. Nghiên cứu này bắt đầu bằng việc tạo ra một mô hình quy trình dạng Petri Net từ dữ liệu log. Tiếp theo, các sự kiện trong dữ liệu log được phát lại (replay) trên mô hình quy trình này và sau đó các chỉ số hiệu năng được tính toán và ghi lại. Tuy nhiên, điểm yếu của nghiên cứu này và của nhóm các phương pháp truyền thống là giả định rằng mỗi sự kiện trong dữ liệu log chỉ thuộc về một trường hợp duy nhất của việc thực hiện quy trình cho một đối tượng [?]. Và trong thực tế, một sự kiện có thể thuộc về nhiều trường hợp thực hiện quy trình của nhiều đối tượng khác nhau. Điều này đòi hỏi chúng ta sẽ cần phải tiền xử lý dữ liệu log trước tiên, để làm cho mỗi sự kiện chỉ thuộc về đúng một trường hợp thực hiện quy trình của một đối tượng khi muốn áp dụng các phương pháp truyền thống. Việc này có thể sẽ làm cho mô hình quy trình không phản ánh được thực tế một cách chính xác, và do đó có thể sẽ làm cho kết quả tính toán các chỉ số hiệu năng không được chính xác. Ngoài ra, các chỉ số hiệu năng liên quan đến việc tương tác giữa các đối tượng cũng không được xem xét tới.
- Nhận thấy nhược điểm của các phương pháp truyền thống, bài báo “OPerA: Object-Centric Performance Analysis” [?] đã đề xuất một phương pháp phân tích hiệu năng quy trình từ dữ liệu mà có xem xét nhiều đối tượng. Cách làm của phương pháp này tương tự như bài báo [?], nhưng thay vì sử dụng mô hình quy trình dạng Petri Net, phương pháp sẽ sử dụng Object-Centric Petri Net [?]. Bằng cách này, mô hình quy trình có khả năng phản ánh thực tế chính xác hơn, từ đó cho ra kết quả tính toán các chỉ số hiệu năng của quy trình sẽ chính xác hơn. Ngoài ra, phương pháp [?] cũng đề xuất thêm các chỉ số hiệu năng liên quan đến sự tương tác của các đối tượng.

Với những gì đã trình bày ở trên, chúng em quyết định sẽ tập trung tìm hiểu sâu

phương pháp xem xét nhiều đối tượng được đề xuất bởi [?] để giải quyết bài toán phân tích hiệu năng quy trình từ dữ liệu.

## 2.5 Kết quả dự kiến của đề tài

- Tổng hợp đầy đủ, phân tích rõ ràng, cụ thể về các phương pháp phân tích hiệu năng quy trình.
- Cài đặt lại được từ đầu phương pháp được đề xuất trong bài báo [?].
- Có được các kết quả thí nghiệm cho thấy việc tự cài đặt cho ra được các kết quả tương tự như bài báo [?].
- Có được các kết quả thí nghiệm ngoài bài báo [?] để thấy rõ hơn về ưu/nhược điểm của phương pháp.
- Nếu có thời gian, có thể nghiên cứu cải tiến phương pháp và có được các kết quả thí nghiệm tương ứng.

## 2.6 Kế hoạch thực hiện

Từ ngày	Đến ngày	Nội dung thực hiện	Phân công
01/01/2024	07/01/2024	Thống nhất ý kiến, xác định các mục tiêu, yêu cầu và lên kế hoạch thực hiện khóa luận.	Dương, Phương
08/01/2024	22/01/2024	Tìm hiểu về tình hình nghiên cứu của bài toán phân tích hiệu năng quy trình.	Dương, Phương
23/01/2024	30/01/2024	Tìm hiểu về các phương pháp tiếp cận xem xét nhiều đối tượng cho bài toán phân tích hiệu năng quy trình.	Dương, Phương

31/01/2024	06/02/2024	Phân tích ưu/khuyết điểm của phương pháp tiếp cận và kết luận.	Dương, Phương
07/02/2023	21/02/2024	Tìm hiểu về các bài báo có liên quan đến bài toán và phương pháp tiếp cận đã chọn, chọn được bài báo chính để tiến hành tìm hiểu và nghiên cứu.	Dương, Phương
22/02/2024	29/02/2024	Đọc hiểu bài báo chính và tìm hiểu thêm các bài báo mới có liên quan đến bài toán.	Dương, Phương
01/03/2024	14/03/2024	Đọc và hiểu sâu hơn về bài báo chính và đọc hiểu thêm các bài báo đã chọn lọc.	Dương, Phương
15/03/2024	31/03/2024	Tìm hiểu về thuật toán được sử dụng cho bài báo và các công trình liên quan.	Dương, Phương
01/04/2024	25/04/2024	Dựa trên ý tưởng các công trình hiện có và kiến thức đã tìm hiểu được, xây dựng và cài đặt lại thuật toán giải quyết bài toán của khóa luận.	Dương, Phương
25/04/2024	07/05/2024	So sánh, nhận xét, đánh giá kết quả và tinh chỉnh thuật toán cho phù hợp.	Dương, Phương
08/05/2024	22/05/2024	Bổ sung và hoàn thành viết cuốn luận.	Dương, Phương

23/05/2024	31/05/2024	Hoàn tất viết cuốn luận, tiến hành làm slide thuyết trình cho khóa luận.	Dương, Phương
01/06/2024	07/06/2024	Kiểm tra lại cuốn luận, mã nguồn và các tài nguyên liên quan.	Dương, Phương