

### THÔNG TIN VỀ LUẬN ÁN TIẾN SĨ

1. Họ và tên nghiên cứu sinh: PHẠM THANH HUYỀN... 2. Giới tính: Nữ.....
3. Ngày sinh: 01/06/1981 ..... 4. Nơi sinh: Hải Dương .....
5. Quyết định công nhận nghiên cứu sinh số: 489/QĐ-ĐT, ngày 25 tháng 6 năm 2015 của Hiệu trưởng Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội.
6. Các thay đổi trong quá trình đào tạo: Không có
7. Tên đề tài luận án: ***Phát triển mô hình tập thô phủ, tập thô mờ và áp dụng tập thô vào khai phá dữ liệu***
8. Chuyên ngành: Hệ thống thông tin..... 9. Mã số: 9480104.01 .....
10. Cán bộ hướng dẫn khoa học:

Hướng dẫn 1: PGS. TS Hồ Thuần, Viện CNTT, Viện HLKH&CN Việt Nam

Hướng dẫn 2: PGS. TS Nguyễn Trí Thành, Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

11. Tóm tắt các **kết quả mới** của luận án:

Luận án đã đạt được một số kết quả mới như sau:

- Thứ nhất, đề xuất bảng quyết định phủ, thuật toán và mô hình phân lớp đa nhãn CDTML - KNN khai thác phụ thuộc nhãn dựa trên bảng quyết định phủ, đề xuất thuật toán phân lớp đa nhãn bán giám sát MULTICS2 khai thác phụ thuộc nhãn từ ma trận quan hệ nhãn – nhãn và thuật toán rút gọn đặc trưng FRR - RED dựa trên tập thô – mờ cho phân lớp đa nhãn. Triển khai và đánh giá mô hình thực nghiệm thuật toán MULTICS2 trên hai tập dữ liệu Enron1 và Hotel1, mô hình thực nghiệm thuật toán CDTML - KNN trên ba tập dữ liệu Enron, Medical và Hotel.

- Thứ hai, đề xuất bảng quyết định dàn phủ, thuật toán rút gọn dàn điều kiện Reduct\_Finding, thuật toán thích nghi dàn quyết định Fitting\_Finding, thuật toán phân lớp đa nhãn CLML dựa trên bảng quyết định dàn phủ CLML. Triển khai và đánh giá mô hình thực nghiệm áp dụng hai thuật toán Reduct\_Finding và Fitting\_Finding vào lọc cộng tác hướng người dùng trên tập dữ liệu MovieLens100K.

- Thứ ba, đề xuất quy tắc suy diễn hợp thành tập mờ bức tranh PFL - CRI, định nghĩa Hệ thống thông tin (HTTT) bội và HTTT mờ bức tranh, quy trình hình thành HTTT mờ bức tranh, không gian xấp xỉ mờ bức tranh từ một HTTT bội, mô hình giải pháp đánh giá và xếp hạng ứng viên tuyển dụng sử dụng quy trình hình thành HTTT mờ bức tranh từ một HTTT bội.

12. Khả năng ứng dụng trong thực tiễn

13. Những hướng nghiên cứu tiếp theo

Trong thời gian tiếp theo, nghiên cứu sinh sẽ tiếp tục nghiên cứu các hướng giải quyết cho các hạn chế còn tồn tại của luận án và tiếp tục triển khai các đề xuất để hoàn thiện hơn các giải pháp Phát triển mô hình và áp dụng mô hình tập thô phủ và tập thô – mờ trong khai phá dữ liệu:

- Tiếp tục triển khai thực nghiệm mô hình CDTML - kNN trên một số tập dữ liệu đa nhãn thuộc Thư viện Java phân lớp đa nhãn Mulan (như đã được đề cập với tập dữ liệu Medical), tiến hành thêm thực nghiệm và phân tích lỗi trên tập dữ liệu Hotel để làm sáng tỏ vấn đề hiệu năng của CDTML - kNN. Từ các kết quả nghiên cứu bổ sung nâng cấp bài báo đã gửi đăng Hội nghị quốc tế ACIIDS 2022 thành một bài báo gửi đăng một tạp chí khoa học uy tín.

- Phát triển thuật toán và triển khai thực nghiệm mô hình áp dụng tập thô mờ bức tranh vào bài toán đánh giá và xếp hạng ứng viên tuyển dụng, chú trọng giải pháp áp dụng quy tắc hợp thành suy diễn trong thuật toán. Viết một bài báo khoa học dựa trên kết quả nghiên cứu này.

- Khảo sát mô hình và giải pháp lựa chọn đặc trưng trong phân lớp đa nhãn để tìm ra các yếu tố phù hợp nhằm nâng cấp thuật toán FRR - RED và tích hợp nó vào thuật toán MULTICS2.

- Tiếp tục triển khai thực nghiệm mô hình áp dụng bảng quyết định dàn phủ vào hệ tư vấn.

14. Các công trình đã công bố có liên quan đến luận án:

[1] Quang-Thuy Ha, Thi-Ngan Pham, Van-Quang Nguyen, Minh-Chau Nguyen, Thanh-Huyen Pham, Tri-Thanh Nguyen. *A New Text Semi-supervised Multi-label Learning Model Based on Using the Label-Feature Relations*. ICCCI (1) 2018: 403-413. (**DBLP, Scopus**).

[2] Bui Cong Cuong, Thanh-Huyen Pham, Chien Pham Van, Pham Van Hai. *Some Fuzzy Inference Processes in Picture Fuzzy Systems*. KSE 2019 (IEEE Xplore), pp.1-5. (**DBLP, Scopus**).

[3] Pham Thanh Huyen, Ho Thuan. *A New Feature Reduction Algorithm Based on Fuzzy Rough Relation for the Multi-Label Classification*. VNU Journal of Science: Computer Science and Communication Engineering, Vol. 36, No. 1, pp. 17-24, 2020.

[4] Thanh-Huyen Pham, Thi-Cam-Van Nguyen, Thi-Hong Vuong, Thuan Ho, Quang-Thuy Ha, Tri-Thanh Nguyen. *A Definition of Covering Based Decision Table and Its Sample Applications*. ICISA2020 (Hyuncheol Kim, Kuinam J. Kim, Suhyun Park. *Information Science and Applications*. LNEE, volume 739), pp. 175-187. (**Scopus**).

[5] Thanh-Huyen Pham, Thi-Ngan Pham, Thuan Ho, Thi-Hong Vuong, Tri-Thanh Nguyen, Quang-Thuy Ha. *Solving Reduction Problems in Cover Lattice Based Decision Tables*. ACIIDS (Companion) 2021: 55-64. (**DBLP, Scopus**).

Ngày ... tháng ..... năm 20.....  
**Xác nhận của cán bộ hướng dẫn**  
(Kí và ghi rõ họ tên)

Ngày ..... tháng .... năm 20.....  
**Nghiên cứu sinh**  
(Kí và ghi rõ họ tên)

**PGS. TS. Nguyễn Trí Thành**

**Phạm Thanh Huyền**

### INFORMATION ON DOCTORAL THESIS

1. Full name : PHAM THANH HUYEN..... 2. Sex: Female .....
3. Date of birth: 01/06/1981 ..... 4. Place of birth: Haiduong.....
5. Admission decision number: 489/QĐ-ĐT Dated June 25, 2015 .....
6. Changes in academic process: .....
7. Official thesis title: *Developing the cover rough set model, the fuzzy rough set model and applying the rough set to data mining*
8. Major: Information System ..... 9. Code: 9480104.01 .....
10. Supervisors: .....

Assoc. Prof. Dr. Ho Thuan, Institute of Information Technology, Vietnam Academy of Science and Technology.

Assoc. Prof. Dr. Nguyen Tri Thanh, VNU University of Engineering and Technology, Vietnam National University, Hanoi

11. Summary of the **new findings** of the thesis: .....

This thesis has obtained the major contributions as follows:

- Firstly, the thesis proposes the definition of covering based decision table, multi – label classification model and **an** algorithm CDTML – KNN which exploits the label relationship from label - label relationship matrix and the feature reduction algorithm FRR - RED based on the fuzzy rough set for multi - label classification. Thesis also implements and evaluates **the performance in comparison with the baselines** (the MULTICS2 and CDTML-KNN algorithms) on severnal datasets of Enron1, Hotell, Enron, Medical and Hotel.

- Secondly, **the** thesis proposes the definition of cover lattice decision table, the algorithm of condition lattice reduction Reduct\_Finding, the algorithm of condition lattice reduction Fitting\_Finding, multi-label classification algorithm CLML with cover lattice based the decision table. The thesis also implements and evaluates the Reduct\_Finding and Fitting\_Finding to user-oriented collaborative filtering on the dataset of MovieLens100K.

- Thirdly, the thesis proposes the Compositional Rule of Inference in Picture Fuzzy Logic PFL - CRI, the definition of Multiple information systems and the picture fuzzy information systems, the process of forming picture fuzzy information systems, picture fuzzy approximation space from the multiple information systems, model of evaluating and ranking recruitment candidates using the process of forming of the picture fuzzy information systems from the multiple information systems.

12. Practical applicability, if any: .....

13. Further research directions, if any: .....

In the future, the PhD student will continue to research solutions to the remaining limitations of the thesis and implement the proposals for improving and applying the covering rough set and fuzzy rough set models on data mining:

- Continue to carry out the experiments of CDTML - KNN model on the multi - label datasets based on multi - label classification - Java library of Mulan (as mentioned above of Medical dataset), deploy more experiments and errors analyses on the Hotel dataset to evaluate the performance problem of CDTML – KNN, then upgrade research results of the article in ACIIDS 2022 international conference to an article of a prestigious scientific journal.

- Develop algorithms and experiments for applying picture fuzzy rough set model on evaluating and ranking recruitment candidates focusing on the Compositional Rule of Inference. Write a scientific paper based on the research results.

- Investigate model and the solutions of feature selection in the multi - label classification to find suitable factors to upgrade the FRR – RED algorithm and integrate it into MULTICS2 algorithm.

- Continue to implement experiments of applying cover lattice based decision table to the recommender systems.

14. Thesis-related publications: .....

[1] Quang-Thuy Ha, Thi-Ngan Pham, Van-Quang Nguyen, Minh-Chau Nguyen, Thanh-Huyen Pham, Tri-Thanh Nguyen. *A New Text Semi-supervised Multi-label Learning Model Based on Using the Label-Feature Relations*. ICCCI (1) 2018: 403-413. **(DBLP, Scopus)**.

[2] Bui Cong Cuong, Thanh-Huyen Pham, Chien Pham Van, Pham Van Hai. *Some Fuzzy Inference Processes in Picture Fuzzy Systems*. KSE 2019 (IEEE Xplore), pp.1-5. **(DBLP, Scopus)**.

- [3] Pham Thanh Huyen, Ho Thuan. *A New Feature Reduction Algorithm Based on Fuzzy Rough Relation for the Multi-Label Classification*. VNU Journal of Science: Computer Science and Communication Engineering, Vol. 36, No. 1, pp. 17-24, 2020.
- [4] Thanh-Huyen Pham, Thi-Cam-Van Nguyen, Thi-Hong Vuong, Thuan Ho, Quang-Thuy Ha, Tri-Thanh Nguyen. *A Definition of Covering Based Decision Table and Its Sample Applications*. ICISA2020 (Hyuncheol Kim, Kuinam J. Kim, Suhyun Park. *Information Science and Applications*. LNEE, volume 739), pp. 175-187. **(Scopus)**.
- [5] Thanh-Huyen Pham, Thi-Ngan Pham, Thuan Ho, Thi-Hong Vuong, Tri-Thanh Nguyen, Quang-Thuy Ha. *Solving Reduction Problems in Cover Lattice Based Decision Tables*. ACIIDS (Companion) 2021: 55-64. **(DBLP, Scopus)**.

Date: .....  
 Signature: .....  
 Full name: Assoc. Prof. Dr. Nguyen Tri Thanh

Date: .....  
 Signature: .....  
 Full name: Phạm Thanh Huyền