

ILA APPLICATION IN SPECIALIZATION OPTIONS FOR STUDENTS OF INFORMATION TECHNOLOGY FACULTY

Nguyen Thu Huong, Le Thu Trang*

TNU - University of Information and Communication Technology

ARTICLE INFO	ABSTRACT
Received: 19/5/2022 Revised: 31/5/2022 Published: 31/5/2022 KEYWORDS Choose a major Classify Data ILA–Inductive Learning Algorithm Specialized	How to choose the right direction of study and research is one of the deciding factors for a student's career and future. For the purpose of partly supporting students of the Faculty of Information Technology in choosing a major study orientation after the third year of the Information Technology Engineer training program. The problem is to come up with classifications of orientation choices that are suitable for students based on the final transcript data of the whole course at the time the students make the selection decision. In this paper, we propose to use the inductive learning algorithm ILA - Inductive Learning Algorithm with data as learning results to classify the orientation of major selection in accordance with the orientation and desire of students. The obtained experimental results achieved a rate of >90% which coincided with the selection decisions of the students of Course 17 at the Faculty. This research partly helps students choose the appropriate majoring orientation in order to achieve good results in the next learning process.

ỨNG DỤNG ILA TRONG LỰA CHỌN ĐỊNH HƯỚNG CHUYÊN NGÀNH CHO SINH VIÊN KHOA CÔNG NGHỆ THÔNG TIN

Nguyễn Thu Hương, Lê Thu Trang*

Trường Đại học Công nghệ Thông tin và Truyền thông – ĐH Thái Nguyên

THÔNG TIN BÀI BÁO	TÓM TẮT
Ngày nhận bài: 19/5/2022 Ngày hoàn thiện: 31/5/2022 Ngày đăng: 31/5/2022 TỪ KHÓA Lựa chọn chuyên ngành học Phân loại Dữ liệu ILA–Inductive Learning Algorithm Chuyên ngành	Lựa chọn đúng định hướng học tập và nghiên cứu là một trong những yếu tố quyết định đến nghề nghiệp và tương lai sau này của sinh viên. Với mục đích phân nào hỗ trợ được sinh viên khoa Công nghệ thông tin trong việc lựa chọn định hướng chuyên ngành học tập sau năm thứ ba của chương trình đào tạo Kỹ sư Công nghệ thông tin. Bài toán đặt ra là cần đưa ra các phân loại lựa chọn định hướng cho phù hợp với sinh viên dựa trên dữ liệu bảng điểm tổng kết toàn khoá tại thời điểm sinh viên ra quyết định lựa chọn. Trong bài báo này, chúng tôi đề xuất sử dụng thuật giải học quy nạp ILA - Inductive Learning Algorithm với dữ liệu là kết quả học tập để phân loại định hướng lựa chọn chuyên ngành cho phù hợp với định hướng và mong muốn của sinh viên. Kết quả thực nghiệm thu được đạt tỷ lệ >90% trùng với các quyết định lựa chọn của sinh viên Khoá 17 tại Khoa. Việc nghiên cứu này phần nào giúp sinh viên lựa chọn định hướng chuyên ngành học tập cho phù hợp nhằm mục đích đạt kết quả tốt trong tiến trình học tập tiếp theo.

DOI: <https://doi.org/10.34238/tnu-jst.6016>

* Corresponding author. Email: lttrang@ictu.edu.vn

1. Giới thiệu

Chọn đúng định hướng chuyên ngành đào tạo là yếu tố quan trọng đến hiệu quả học tập, rèn luyện, tiến độ học tập và định hướng nghề nghiệp, chuyên môn của sinh viên. Có rất nhiều lý do tác động đến việc lựa chọn định hướng chuyên ngành như: ý thích, niềm đam mê, dự kiến vị trí việc làm sau khi ra trường, các môn học yêu thích... Trong đó, kết quả học tập của ba năm học đầu trong chương trình đào tạo cũng góp phần ảnh hưởng đến quyết định lựa chọn này. Trong những kỳ học đầu của chương trình đào tạo sẽ cung cấp cho người học những kiến thức giáo dục đại cương, các môn học thuộc khối kiến thức cơ sở ngành để từ đó người học có thể tiếp nhận những kiến thức chuyên ngành trong thời gian đào tạo còn lại. Với mục đích hỗ trợ người học trong việc ra quyết định lựa chọn các định hướng chuyên ngành phù hợp với người học trong quá trình học tập tiếp theo của chương trình đào tạo Công nghệ thông tin (CNTT). Chúng tôi thực hiện ứng dụng nhằm đưa ra các luật khuyến nghị người học lựa chọn một trong các chuyên ngành đã được xây dựng trong chương trình đào tạo gồm: Định hướng Khoa học dữ liệu (KHDL); Định hướng Kỹ thuật phần mềm (KTPM); Định hướng Hệ thống thông tin (HTTT); Định hướng Khoa học máy tính (KHMT); Định hướng Mạng máy tính và truyền thông (MTT).

Trước đây, bài toán phân lớp dữ liệu đã được nhiều nhà nghiên cứu tiến hành nghiên cứu và thực nghiệm trong rất nhiều lĩnh vực: Giáo dục, Y tế... Tiếp đến, đã có nhiều kết quả nghiên cứu đề cập và thực hiện giải quyết các bài toán về phân loại kết quả học tập ứng dụng ở nhiều đơn vị khác nhau. Ngoài ra, cũng có nhiều kết quả của các nhà nghiên cứu về phân loại dựa trên kết quả học tập của người học bằng nhiều phương pháp khác nhau. Trong [1], các tác giả sử dụng công cụ BIDS của Microsoft SQL Server 2008 để khai phá dữ liệu điểm của sinh viên. Đồng thời, các tác giả sử dụng phương pháp hai thuật toán khai phá dữ liệu Logistic Regression, Naive Bayes đã được áp dụng để tìm ra mô hình tốt nhất cho việc dự báo tình trạng học tập [1]. Trong [2], các tác giả sử dụng giải thuật rừng ngẫu nhiên học từ dữ liệu để rút trích các môn học quan trọng trong chương trình đào tạo ngành CNTT. Trong [3], các tác giả sử dụng công cụ BIDS của Microsoft SQL Server 2008 để khai phá dữ liệu điểm của sinh viên. Trong [4], [5] trình bày các kết quả nghiên cứu về đánh giá dự đoán kết quả học tập của người học bằng kỹ thuật khai phá dữ liệu. Các kết quả nghiên cứu này đều nhằm mục đích hỗ trợ tư vấn cho sinh viên trong quá trình học tập. Các kết quả nghiên cứu cho thấy sự cần thiết của khai phá kết quả học tập nhằm mục đích tăng cường hỗ trợ, dự báo kết quả học tập của sinh viên, giúp sinh viên đạt quả cao trong quá trình học tập.

Tuy nhiên, với mục đích nghiên cứu và ứng dụng trong bài toán cụ thể nhằm hỗ trợ người học Khoa Công nghệ thông tin phần nào trong việc ra quyết định chọn định hướng chuyên ngành, bài báo sử dụng ILA nhằm hình thành các luật cho phép phân lớp lựa chọn định hướng chuyên ngành đào tạo. Với kết quả thực nghiệm dựa trên tập dữ liệu điểm tổng kết học tập của người học đã chỉ ra phần nào hiệu quả của phương pháp nhóm tác giả đã sử dụng.

2. Thuật giải học quy nạp ILA - Inductive Learning Algorithm

Thuật giải ILA [6], [7] được dùng để xác định các luật phân loại cho tập hợp các mẫu học. Thuật giải này thực hiện theo cơ chế lặp, để tìm luật riêng đại diện cho tập mẫu của từng lớp. Sau khi xác định được luật, ILA loại bỏ các mẫu liên quan khỏi tập mẫu, đồng thời thêm luật mới này vào tập luật. Kết quả có được là một danh sách có thứ tự các luật phụ thuộc vào tần suất của giá trị thuộc tính trong tập mẫu. Các ưu điểm của thuật giải này có thể được trình bày như sau:

- Dạng các luật sẽ phù hợp cho việc khảo sát dữ liệu, mô tả mỗi lớp một cách đơn giản để dễ phân biệt với các lớp khác.
- Tập luật được sắp thứ tự, riêng biệt – cho phép quan tâm đến một luật tại thời điểm bất kỳ.

2.1. Xác định dữ liệu mẫu

1. Tập gồm 100 mẫu là kết quả học tập của sinh viên được lấy ngẫu nhiên trong 205 kết quả học tập của sinh viên K17 ngành CNTT được liệt kê trong bảng 1, với mỗi dòng tương ứng một mẫu, và mỗi cột thể hiện thông tin về tổng số tín chỉ tích lũy và điểm trung bình toàn khóa tính đến năm thứ ba, trong đó có một thuộc tính “Chuyên ngành” là thuộc tính quyết định. Với thuộc tính quyết định có 5 giá trị tương ứng với 5 định hướng chuyên môn của chương trình đào tạo CNTT và đây là số lớp của tập mẫu.

Bảng 1. Dữ liệu kết quả học tập toàn khóa

S T T	Mã SV	Họ và tên	Ngày sinh	Tổng số tín chỉ (TC)	Tổng số TC tích lũy	Điểm hệ số 10	Điểm hệ số 4	Điểm trung bình trung (TBT) tích lũy hệ 10	Điểm TBT tích lũy hệ 4	Chuyên ngành
1	DTC1854802010136	Vũ Hoàng Long	17/07/2000	90	90	8,65	3,68	8,65	3,68	KHDL
2	DTC1854802010232	Trần Minh Tuấn	15/08/2000	93	93	8,43	3,57	8,43	3,57	KTPM
3	DTC1854802010263	Chu Thị Thanh Thủy	11/02/2000	93	93	8,20	3,46	8,20	3,46	HTTT
4	DTC1854802010032	Nguyễn Văn Dũng	04/01/2000	93	91	8,21	3,45	8,33	3,53	KTPM
5	DTC1854802010116	Phạm Hoàng Kim	13/05/2000	93	93	8,23	3,39	8,23	3,39	KHDL
6	DTC18H4802010084	Vũ Ngọc Hoàng	12/06/1999	93	93	8,05	3,31	8,05	3,31	HTTT
7	DTC1854802010002	Trần Đình Chung	08/01/2000	93	93	7,98	3,29	7,98	3,29	KTPM
8	DTC18H4802010258	Bùi Thái Dương	10/06/2000	93	93	7,94	3,24	7,94	3,24	KTPM
9	DTC1854802010049	Triệu Minh Đức	22/12/2000	93	93	8,05	3,23	8,05	3,23	KTPM
10	DTC1854802010195	Kiều Hoàng Phúc	23/06/2000	93	93	8,04	3,20	8,04	3,20	KHDL
11	DTC1854802010060	Lương Xuân Hải	17/07/2000	93	93	7,89	3,18	7,89	3,18	KTPM
12	DTC1854802010120	Vũ Ngọc Khanh	14/02/2000	93	93	7,76	3,12	7,76	3,12	KHDL
13	DTC18H4802010272	Dương Thị Hải Yến	10/09/2000	93	93	7,68	3,05	7,68	3,05	HTTT
14	DTC1854802010258	Nguyễn Thị Huyền Thơ	11/02/2000	93	93	7,82	3,05	7,82	3,05	HTTT
15	DTC1854802010229	Nguyễn Đức Tuấn	26/12/2000	93	93	7,62	3,05	7,62	3,05	KTPM

2. Tập luật R các giá trị khởi tạo là $\emptyset$.

3. Tất cả các cột trong bảng ban đầu chưa được đánh dấu (kiểm tra).

2.2. Thuật giải ILA

Thuật giải học quy nạp ILA gồm các bước trình bày trong Thuật toán 1 dưới đây:

Thuật toán 1. Các bước chính của thuật giải ILA

Đầu vào	Tập mẫu đã lựa chọn liệt kê dưới dạng bảng gồm m mẫu, k thuộc tính trong đó có một thuộc tính “Chuyên ngành” là thuộc tính quyết định.
Đầu ra	Tập R các luật có dạng If <điều kiện> Then <Kết luận>
1:	Chia bảng m mẫu ban đầu thành n bảng con. Mỗi bảng con ứng với một giá trị của thuộc tính phân lớp của tập mẫu.
2:	(* thực hiện các bước 2 đến 8 cho mỗi bảng con*) Khởi tạo bộ đếm kết hợp thuộc tính j, j=1
3:	Với mỗi bảng con đang khảo sát, phân chia danh sách các thuộc tính theo các tổ hợp phân biệt, mỗi tổ hợp ứng với j thuộc tính phân biệt.
4:	Với mỗi tổ hợp các thuộc tính, tính số lượng các giá trị thuộc tính xuất hiện theo cùng tổ hợp thuộc tính trong các dòng chưa được đánh dấu của bảng con đang xét (mà đồng thời không xuất hiện với tổ hợp thuộc tính này trên các bảng còn lại). Gọi tổ hợp đầu tiên (trong

	bảng con) có số lần xuất hiện nhiều nhất là tổ hợp lớn nhất.
5:	Nếu tổ hợp lớn nhất bằng ϕ , tăng j lên 1 và quay lại bước 3
6:	Đánh dấu các dòng thoả tổ hợp lớn nhất của bảng con đang xử lý theo lớp
7:	Thêm luật mới vào tập luật R, với vế trái là tập các thuộc tính của tổ hợp lớn nhất (kết hợp các thuộc tính bằng toán tử AND) và vế phải là giá trị thuộc tính quyết định tương ứng.
8:	Nếu tất cả các dòng trong bảng con đang xét đã được đánh dấu phân lớp, tiếp tục thực hiện từ bước 2 cho các bảng con còn lại. Ngược lại (nếu chưa đánh dấu hết các dòng) thì quay lại bước 4. Nếu tất cả các bảng con đã được xét thì kết thúc, kết quả thu được là tập luật cần tìm.

3. Kết quả thực nghiệm và đánh giá

Dữ liệu được sử dụng trong thực nghiệm là kết quả học tập của 205 sinh viên ngành Công nghệ thông tin K17 sau khi kết thúc kỳ học 5 (Bảng 1). Với kết quả học tập tổng kết bao gồm các thông tin: Tổng số tín chỉ (TONGTC); Tổng số tín chỉ tích lũy (TONGTCTL); Điểm hệ số 10 (TBM10); Điểm hệ số 4 (TBM4); Điểm TBT tích lũy hệ 10 (TBM10TL); Điểm TBT tích lũy hệ 4 (TBM4TL) và Chuyên ngành (CHUYENNGANH). Tiếp đến nhóm tác giả tiến hành tiền xử lý dữ liệu và thu được bảng dữ liệu sau:

Bảng 2. Tập dữ liệu sau tiền xử lý

TONGTC	TONGTCTL	TBM10	TBM4	TBM10TL	TBM4TL	CHUYENNGANH
TB	Thap	Kem	Kem	TB	TB	MTT
TB	Thap	Kem	Kem	TB	TB	MTT
Thap	Thap	Kem	Yeu	TB	TB	HTTT
TB	TB	Yeu	Yeu	TB	TB	MTT
TB	TB	Yeu	Yeu	TB	TB	KHDL
TB	TB	Yeu	Yeu	TB	TB	HTTT
TB	TB	Yeu	Yeu	TB	TB	MTT
TB	TB	Yeu	Yeu	TB	TB	KHMT
TB	TB	Yeu	Yeu	TB	TB	KHDL
Cao	Cao	Kha	Gioi	Kha	Gioi	KTPM
Cao	Cao	Kha	Gioi	Kha	Gioi	KTPM
Cao	Cao	Kha	Gioi	Kha	Gioi	KTPM
Cao	Cao	Kha	Gioi	Kha	Gioi	KTPM
Cao	Cao	Kha	Gioi	Kha	Gioi	HTTT
Cao	Cao	Kha	Gioi	Kha	Gioi	KHDL
Cao	Cao	Kha	Gioi	Kha	Gioi	KHDL
Cao	Cao	Kha	Gioi	Kha	Gioi	HTTT
Cao	Cao	Gioi	Gioi	Kha	Gioi	KTPM
Cao	Cao	Gioi	XuatSac	Gioi	XuatSac	KHDL

Sử dụng giải thuật ILA (Thuật toán 1) với bảng dữ liệu sau tiền xử lý (Bảng 2) và thuộc tính quyết định “CHUYENNGANH”. Với giải thuật này, chúng tôi thực hiện trên dữ liệu điểm học tập của 205 sinh viên sau tiền xử lý và thu được tập luật sau:

Luật 1: IF TBM4 là Yếu AND TBM10TL là Trung bình AND TongTCTL là Thấp THEN chọn Chuyên ngành Mạng truyền thông.

Luật 2: IF TBM4 là Yếu AND TBM10TL là Trung bình AND TongTCTL là Trung bình AND TONGTC là Trung bình AND TBM10 là không Trung bình THEN chọn Chuyên ngành Mạng truyền thông.

Luật 3: IF TBM4 là Yếu AND TBM10TL là Trung bình AND TongTCTL là Trung bình AND TONGTC là Thấp THEN chọn Chuyên ngành Khoa học máy tính.

Luật 4: IF TBM4 là Yếu AND TBM10TL là Trung bình AND TongTCTL là Trung bình AND TONGTC là Cao THEN chọn Chuyên ngành Hệ thống thông tin.

Luật 5: IF TBM4 là Yếu AND TBM10TL là Trung bình AND TongTCTL là Trung bình AND TONGTC là Trung bình AND TBM10 là Trung bình THEN chọn Chuyên ngành Khoa học máy tính.

Luật 6: IF TBM4 là Yếu AND TBM10TL là Trung bình AND TongTCTL là Vừa phải THEN chọn Chuyên ngành Mạng truyền thông.

Luật 7: IF TBM4 là Yếu AND TBMTL10 là Trung bình AND TongTCTL là Cao THEN chọn Chuyên ngành Khoa học máy tính.

Luật 8: IF TBM4 là Yếu AND TBMTL10 là Khá THEN chọn Chuyên ngành Hệ thống thông tin.

Luật 9: IF TBM4 là Yếu AND TBMTL10 là Giỏi hoặc Yếu THEN chọn Chuyên ngành Mạng truyền thông.

Luật 10: IF TBM4 là Trung bình THEN chọn Chuyên ngành Hệ thống thông tin.

Luật 11: IF TBM4 là Kém THEN chọn Chuyên ngành Mạng truyền thông.

Luật 12: IF TBM4 là Khá AND TBM10 là Trung bình THEN chọn Chuyên ngành Hệ thống thông tin.

Luật 13: IF TBM4 là Khá AND TBM10 là Giỏi THEN chọn Chuyên ngành Kỹ thuật phần mềm.

Luật 14: IF TBM4 là Giỏi THEN chọn Chuyên ngành Kỹ thuật phần mềm.

Luật 15: IF TBM4 là Xuất sắc THEN chọn Chuyên ngành Khoa học dữ liệu.

Luật 16: IF TBMTL4 là Trung bình AND TBM10 là Trung bình AND TONGTC là cao AND TBMTL10 là Khá AND TBM4 là Trung bình THEN chọn Chuyên ngành Khoa học dữ liệu.

Trong tập các luật thu được ở trên ta thấy, các luật vừa ở dạng tổng quát, vừa ở dạng chi tiết. Với luật 11 có thể được sử dụng để tư vấn chọn chuyên ngành Mạng truyền thông, nếu cần cụ thể hơn nữa về các yếu tố chọn chuyên ngành này chúng ta có thể sử dụng luật 1, luật 2 hoặc luật 9.

Căn cứ tập luật thu được ở trên, chúng ta có thể đưa ra lời khuyên hay dự đoán cho sinh viên trong quá trình lựa chọn định hướng chuyên môn sâu như: Nếu sinh viên có điểm Trung bình môn hệ số 4 là Giỏi thì chọn chuyên ngành Kỹ thuật phần mềm hoặc nếu điểm Trung bình môn hệ số 4 là Khá mà mong muốn chọn chuyên ngành Kỹ thuật phần mềm thì phải có điểm Trung bình môn hệ số 10 là Giỏi.

Với các luật như trên cho thấy, có >90% sinh viên K17 trùng kết quả thực tế đã lựa chọn chuyên ngành, do đó chúng ta có thể sử dụng tập các luật này để định hướng sinh viên trong việc lựa chọn chuyên ngành cho phù hợp với năng lực của người học.

4. Kết luận

Bài báo đã ứng dụng giải thuật ILA nhằm hình thành các luật cho phép phân lớp lựa chọn định hướng chuyên ngành đào tạo. Tập các luật được xây dựng dựa trên cơ sở dữ liệu là kết quả học tập của sinh viên ngành Công nghệ thông tin K17 sau khi kết thúc kỳ học 5. Kết quả thử nghiệm mà nhóm tác giả thực hiện đã đưa ra tập các luật nhằm hỗ trợ người học Khoa Công nghệ thông tin phần nào trong việc ra quyết định chọn định hướng chuyên ngành. Qua kết quả nghiên cứu này, nhóm tác giả nhận thấy nghiên cứu tiếp theo cần phải thử nghiệm với bộ cơ sở dữ liệu với nhiều yếu tố quan tâm hơn nữa nhằm nâng cao chất lượng của các luật thu được từ giải thuật.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] C. D. Dinh, "Researching and applying data mining techniques on university student data for academic advisors," Master's thesis, University of Engineering and Technology - VNU, 2018.
- [2] T. U. Nguyen, and M. T. Nguyen, "Predicting student's academic performance by applying data mining technique," *Journal of Science, Vinh University*, vol. 48, no. 3A, pp. 68-73, 2019.
- [3] T. N. Do, N. K. Pham, and M. T. Nguyen, "Detection of the key courses affecting the learning outcomes of information technology students," *Science Journal of Can Tho University*, vol. 33, pp. 49-57, 2014.
- [4] M. S. Amirah, H. Wahidah, and A. R. Nuraini, "A Review on Predicting Student's Performance using Data Mining Techniques," *Procedia Computer Science*, vol. 72, pp. 414-422, 2015.
- [5] A. K. Pal and S. Pal, "Classification Model of Prediction for Placement of Students," *International Journal of Modern Education and Computer Science*, vol. 5, no. 11, pp. 49-56, 2013.
- [6] M. R. Tolun and S. M. Abu-Soud, "ILA: an inductive learning algorithm for rule extraction," *Expert Systems With Applications*, vol. 14, pp. 361-370, 1998.
- [7] M. R. Tolun, H. Sever, M. Uludag, and S. M. Abu-Soud, "ILA-2: an inductive learning algorithm for knowledge discovery," *An International Journal*, vol. 10, pp. 609-628, 1999.