



BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC



ĐỀ CƯƠNG LUẬN VĂN THẠC SĨ

Chuyên ngành: Công nghệ Thông tin

**ĐÁNH GIÁ CHẤT LƯỢNG HỌC SINH TRƯỜNG PHỔ THÔNG TRUNG HỌC
..... SỬ DỤNG THUẬT TOÁN ID3 VÀ BAYES**

HỌC VIÊN THỰC HIỆN:

HƯỚNG DẪN KHOA HỌC:



..... 3/2017



1. Đặt vấn đề

1.1. Tính cấp thiết của đề tài

Trong những năm gần đây, công tác đánh giá chất lượng cho học sinh tại các trường phổ thông trung học còn nhiều hạn chế và bất cập, chưa đáp ứng được nhu cầu thực tế trong đánh giá chất lượng lớp học. Những khó khăn cụ thể như: Số lượng học sinh từ 1000-1400 học sinh với đầy đủ các thông tin về bản thân, lớp học, môn học, giáo viên giảng dạy, được xếp loại về học lực Giỏi, Khá, Trung Bình, Yếu, Kém; hạnh kiểm Tốt, Khá, Trung bình, Yếu; Thẻ chất đạt, không đạt; đánh giá đạo đức... Trong quá trình học tập, học sinh đạt được sự tiến bộ khác nhau. Khi đó việc đánh giá chất lượng học sinh bằng phương pháp thủ công dẫn đến tình trạng khó đánh giá, khó kiểm soát được thông tin học sinh, dễ bị rối. Thay vào đó chúng ta ứng dụng công nghệ thông tin số hóa toàn bộ dữ liệu trên hệ thống máy tính với khả năng truy suất, tìm kiếm, thống kê nhanh, chính xác hơn. Tuy nhiên khối lượng dữ liệu là không đổi cho dù quản lý bằng cách nào đi nữa. Việc nắm bắt và hiểu được khối lượng dữ liệu khổng lồ đó vẫn là một thách thức không nhỏ.

Có thể nói, bài toán phân lớp dữ liệu có hai cách tiếp cận: Một là cách tiếp cận truyền thống, từ số liệu nhiều năm trước chúng ta sử dụng các phương pháp như hồi quy (phương pháp kinh điển) cho các thời điểm tương lai. Những phương pháp đó khá phổ biến cho các bài toán dự báo thời tiết, dự báo quy luật tăng trưởng dân số, dự báo mức nước biển, nước sông v.v... Phương pháp tiệm cận này thường sử dụng trong bối cảnh dài, có tính chất tương đối ổn định và có quy luật tương đối rõ.

Các tiếp cận khác, ngược lại thường dùng khi tất cả các hiện tượng, sự việc có nhiều biến động lớn như thay đổi chính sách, tác động của môi trường quá mạnh rất khó thực hiện theo quy luật. Đó là phương pháp phân lớp theo các đặc trưng của đối tượng, sự việc. Có nhiều phương pháp phân lớp, trong đó đã được đề xuất như: Phân lớp cây quyết định (Decision Tree Classification: ID3,

Quinlan, C4.5, C5.0), phân lớp Bayes (Bayesian Classifier), phân lớp K_NN (K_Nearest Neighbor Classifier)...

Trong nội dung đề tài chọn bài toán cần giải quyết với hy vọng tìm ra quy luật, các đặc trưng tác động đến công tác phân loại chất lượng học sinh. Đây là một hướng có ứng dụng thực tiễn tốt góp phần giúp các nhà quản lý và bản thân học sinh thấy được kết quả học tập của một đơn vị dữ liệu hay một nhóm dữ liệu học sinh trên toàn bộ tập dữ liệu. Giải quyết bài toán này sẽ giúp ích lớn cho quá trình cải thiện chất lượng dạy và học.

1.2. Mục tiêu nghiên cứu

Đề tài sẽ tập trung nghiên cứu, tìm hiểu thuật toán ID3 và Bayes để rút ra quy luật phân loại chất lượng học sinh theo số liệu thực tế của Trường phổ thông trung học

Từ việc phân loại chất lượng học sinh đưa ra các đề xuất, kiến nghị, tham mưu cho Ban Giám hiệu và các giáo viên chủ nhiệm có phương pháp để nâng cao chất lượng học sinh.

1.3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu của đề tài dựa trên học sinh của Trường phổ thông trung học; trong đó mỗi khối chọn 1-2 lớp điển hình.

Giới hạn Thử nghiệm cho từ 5-8 đặc trưng đầu vào được cho là cơ bản nhất; sau đó có thể mở rộng thêm các đặc trưng khác.

2. Phương pháp nghiên cứu

2.1. Về lý thuyết:

- Nghiên cứu thuật toán học ID3 và Bayes.
- Nghiên cứu các tiêu chí đánh giá chất lượng phân loại.

2.2. Về thực nghiệm

- Dự kiến thử nghiệm mỗi khối 1-2 lớp mẫu cho bài toán phân loại chất lượng học sinh Trường phổ thông trung học sử dụng thuật toán ID3 và Bayes.

- Xây dựng bộ dữ liệu mẫu dùng để kiểm tra, thử nghiệm chương trình và đưa ra nhận xét, so sánh, đánh giá kết quả đạt được và các đề xuất ứng dụng.

3. Nội dung nghiên cứu

Dự kiến các nội dung nghiên cứu chính với bố cục như sau:

MỞ ĐẦU

Nêu lý do chọn đề tài, phương pháp nghiên cứu, giới hạn đề tài và tóm tắt bố cục của luận văn.

CHƯƠNG 1. GIỚI THIỆU VỀ PHƯƠNG PHÁP ĐÁNH GIÁ CHẤT LƯỢNG HỌC SINH

- 1.1. Giới thiệu về Trường phổ thông trung học
- 1.2. Các tiêu chí phân loại học sinh
- 1.3. Phát biểu bài toán. Giới hạn của vấn đề nghiên cứu
- 1.4. Bố cục của luận văn

CHƯƠNG 2. THUẬT TOÁN ID3, BAYES

- 2.1. Các phương pháp phân lớp dữ liệu
- 2.2. Các đặc trưng dữ liệu dùng cho phân lớp
- 2.3. Thuật toán ID3
- 2.4. Thuật toán Bayes
- 2.5. Các chỉ tiêu đánh giá phân lớp

CHƯƠNG 3. PHÂN LỚP DỮ LIỆU HỌC SINH SỬ DỤNG THUẬT TOÁN ID3, BAYES

- 3.1. Thu thập và phân tích, tiền xử lý dữ liệu cho bài toán phân lớp
- 3.2. Phân tích, lựa chọn các đặc trưng cơ bản cho phân loại học sinh
- 3.3. Chọn công cụ phần mềm cho các thử nghiệm (phần mềm Weka);
- 3.4. Phân loại học sinh sử dụng thuật toán ID3;
- 3.5. Phân loại học sinh sử dụng thuật toán Bayes;
- 3.6. So sánh và đánh giá kết quả đạt được. Đề xuất.

KẾT LUẬN

Kết quả đạt được; hướng phát triển tiếp theo của luận văn.

Kết quả: - báo cáo luận văn

- Bộ dữ liệu để đánh giá và các đề xuất
- Tham gia 1 bài báo khoa học Tạp chí KHCN đại học

4. Kế hoạch thực hiện

STT	Nội dung nghiên cứu	Thời gian bắt đầu	Thời gian kết thúc	Kết quả nghiên cứu dự định đạt được
1	Viết đề cương chi tiết Bảo vệ đề cương chi tiết	01/03/2017	31/03/2017	Xong phần tổng quan, đề cương luận văn
2	Xác định rõ bài toán cần giải quyết	01/04/2017	30/04/2017	Xong phần mở đầu luận văn và danh mục Tài liệu tham khảo
3	Thu thập dữ liệu giải quyết bài toán	01/05/2017	30/05/2017	Xong phần khảo sát
4	Chương 1	01/06/2017	31/06/2017	Viết xong chương 1
5	Chương 2	01/07/2017	30/07/2017	Viết xong chương 2
6	Thử nghiệm	01/08/2017	31/08/2017	Viết xong chương 3. Xong thử nghiệm
7	Đánh giá kết quả	01/09/2017	31/09/2017	Viết kết luận và định hướng nghiên cứu tiếp theo
8	Tổng hợp, tích hợp các chương luận văn	01/10/2017	30/10/2017	Luận văn hoàn chỉnh
9	Dự kiến bảo vệ luận văn	01/11/2017	31/11/2017	Bảo vệ luận văn

5. Tài liệu tham khảo

1. Nguyễn Quang Hoan (2015), *Các hệ dựa trên trí thức*, giáo trình, Học viện Bưu chính Viễn thông.
2. Hoàng Kiếm, Đỗ Phúc, Đỗ Văn Nhơn (2002), *Các hệ cơ sở tri thức*, giáo trình, Nhà xuất bản Đại học Quốc gia Thành Phố Hồ Chí Minh.
3. Đỗ Thanh Nghị (2011), *Khai mở dữ liệu*, giáo trình, Nhà xuất bản Đại học Cần Thơ.
4. Đỗ Thanh Nghị, Phạm Nguyên Khang (2012), *Nguyên lý máy học*, giáo trình, Nhà xuất bản Đại học Cần Thơ.

....., ngày 30 tháng 03 năm 2017

HỌC VIÊN

HƯỚNG DẪN KHOA HỌC

TRƯỞNG KHOA

CHỦ TỊCH HĐKH