



ĐỀ CƯƠNG CHI TIẾT MÔN HỌC: KHAI PHÁ DỮ LIỆU

1. Thông tin về các giảng viên môn học

STT	Họ và tên	Chức danh, học vị	Địa chỉ liên hệ	Điện thoại/Email	Ghi chú
1	Hà Quang Thụy	PGS.TS	BM HTTT	thuyhq@vnu.edu.vn	Trưởng môn học
2	Nguyễn Trí Thành	TS	BM HTTT	ntthanh@vnu.edu.vn	Giảng viên
3	Phan Xuân Hiếu	TS	BM HTTT	hieupx@vnu.edu.vn	Giảng viên

2. Thông tin chung về môn học

- Tên môn học: Khai phá dữ liệu (Data Mining)
- Mã số môn học: INT 3209
- Số tín chỉ: 3
- Giờ tín chỉ đối với các hoạt động (LT/ThH/TH): 45/0/0
- Môn học tiên quyết: INT2207-Cơ sở dữ liệu
- Các yêu cầu đối với môn học (nếu có):
- Bộ môn, Khoa phụ trách môn học: Bộ môn Hệ thống thông tin, Khoa CNTT

3. Mục tiêu môn học

Cung cấp cho sinh viên các kiến thức và kỹ năng sau:

Về kiến thức:

+ Hiểu và phân biệt được các khái niệm dữ liệu, thông tin, tri thức. Hiểu được vai trò của tri thức trong tổ chức và nền kinh tế. Nắm bắt được nội dung các bước cơ bản của quá trình phát hiện tri thức từ dữ liệu. Phân biệt được phương pháp khai phá dữ liệu (KPDL) với các phương pháp xử lý dữ liệu truyền thống. Hiểu được vai trò đặc biệt của chuyên gia miền ứng dụng, người sử dụng trong quá trình KPDL. Hiểu được lý do của các phương pháp KPDL và giải thích được vì sao các phương pháp như vậy không thích hợp trong các tiếp cận truyền thống. Biết được xu thế phát triển hiện đại của KPDL;

+ Hiểu và vận dụng được (i) các phương pháp hiệu dữ liệu; (ii) các bài toán và các phương pháp tiền xử lý dữ liệu; (iii) các thuật toán khai phá luật kết hợp điển hình; (iv) các thuật toán phân cụm điển hình; (v) các thuật toán phân lớp điển hình; (vi) các lỗi điển hình khi thực hiện dự án KPDL và phương hướng khắc phục; (vii) một vài nguồn tài nguyên điển hình về KPDL.

Về kỹ năng:

+ Phát biểu được một bài toán KPDL thực tiễn và thiết kế được mô hình giải quyết bài

toán,

+ Khai thác được tài nguyên KPDL để giải quyết một bài toán KPDL cụ thể có quy mô nhỏ.

4. Chuẩn đầu ra

Nội dung	Mục tiêu	Bậc 1 (biết)	Bậc 2 (hiểu và áp dụng)	Bậc 3 (phân tích và đánh giá)	Bậc 4 (sáng tạo)
1. Kiến thức					
1.1. Kiến thức chung của nhóm ngành					
Xu hướng mới trong CNTT: KPDL là bước phát triển mới của công nghệ CSDL; công nghệ tri thức trong nền kinh tế; xu hướng mới của KPDL			×		
1.2. Kiến thức ngành và bổ trợ					
Quản lý HTTT và tính lãnh đạo: Kinh tế tri thức, Công nghệ Tri thức hỗ trợ ra quyết định				×	
Quản lý dữ liệu và thông tin: Hiểu dữ liệu, chuẩn bị dữ liệu, trình diễn dữ liệu và mẫu trong KPDL			×		
Phân tích và thiết kế hệ thống; Thiết kế và thi hành giải pháp KPDL				×	
Quản lý dự án HTTT: Dự án KPDL			×		
Khai thác và sử dụng tri thức người dùng: Vai trò người dùng và chuyên gia miền ứng dụng trong quá trình KPDL					
1.3. Kiến thức thực tập					
Thực tập về quản lý dự án: Triển khai thi hành dự án KPDL			×		
Kiến thức miền ứng dụng: tự tìm hiểu như chuyên gia miền ứng dụng hoặc phối hợp với chuyên gia miền ứng dụng cho một ứng dụng KPDL cụ thể			×		
2. Kỹ năng					
2.1. Kỹ năng cứng					
2.1.1. Các kỹ năng nghề nghiệp chung					
Kỹ năng tìm kiếm, cập nhật, tổng hợp, khai thác thông tin				×	
Kỹ năng thiết kế cơ hội để cải tiến tổ chức dựa trên CNTT (Hình thành ý tưởng)			×		
Kỹ năng Thiết kế và thi hành giải pháp HTTT (Thiết kế và thi hành)			×		
Kỹ năng Quản lý hoạt động CNTT đang diễn ra (Vận hành)			×		
2.1.2. Kỹ năng lập luận tư duy và giải quyết vấn đề				×	
2.1.3. Kỹ năng nghiên cứu và khám phá tri thức				×	
2.1.4. Kỹ năng tư duy theo hệ thống			×		
2.1.5. Hiểu bối cảnh xã hội và ngoại cảnh			×		
2.1.6. Hiểu bối cảnh tổ chức			×		
2.1.7. Năng lực vận dụng kiến thức và kỹ năng vào thực tiễn				×	
2.1.8. Năng lực sáng tạo, phát triển dẫn dắt sự thay đổi				×	

Nội dung	Mục tiêu	Bậc 1 (biết)	Bậc 2 (hiểu và áp dụng)	Bậc 3 (phân tích và đánh giá)	Bậc 4 (sáng tạo)
2.2. Kỹ năng mềm					
2.2.1. <i>Các kỹ năng cá nhân:</i> Tư duy sáng tạo; Kỹ năng phê phán, phản biện; Cập nhật thể giới công nghệ; Quản lý tài nguyên và thời gian của cá nhân; Kỹ năng học suốt đời. Tích hợp tri thức				×	
2.2.2. <i>Làm việc theo nhóm:</i> Tạo lập nhóm; Hoạt động nhóm; Phát triển và tiến hóa nhóm; Lập nhóm đa ngành và nhóm kỹ thuật	×				
2.2.3. <i>Kỹ năng giao tiếp:</i> Đặt câu hỏi, lắng nghe và hội thoại về các chủ đề HTTT			×		
2.2.4. <i>Kỹ năng giao tiếp sử dụng ngoại ngữ:</i> Nắm vững thuật ngữ tiếng Anh về KPDL và biết phiên bản tiếng Việt phổ biến; dịch được tài liệu tiếng Anh về KPDL			×		
2.2.5. <i>Các kỹ năng mềm khác:</i> Đương đầu với thách thức, rủi ro; Thích nghi đa văn hóa.			×		
3. Phẩm chất					
Phẩm chất, đạo đức nghề nghiệp:					
- Có trách nhiệm với công việc			×		
- Tinh thần phát huy sáng kiến và tính sẵn sàng quyết định khi còn có sự không chắc chắn			×		
Phẩm chất, đạo đức xã hội: Tính riêng tư và khai phá dữ liệu đảm bảo tính riêng tư					
				×	

5. Tóm tắt nội dung môn học

Giới thiệu môn học; Giới thiệu về khai phá dữ liệu; Công nghệ tri thức và phát hiện tri thức từ dữ liệu; Hiểu dữ liệu và tiền xử lý dữ liệu; Khai phá luật kết hợp; Phân cụm dữ liệu; Phân lớp dữ liệu; Tài nguyên cho khai phá dữ liệu; Xu thế của khai phá dữ liệu.

6. Nội dung chi tiết môn học

Chương 1. Giới thiệu về khai phá dữ liệu

- 1.1. Giới thiệu môn học
- 1.2. Nhu cầu phát hiện tri thức từ dữ liệu
- 1.3. Quá trình phát hiện tri thức trong Cơ sở dữ liệu (KDD)
- 1.4. Khai phá dữ liệu và xử lý CSDL truyền thống
- 1.5. Kiểu dữ liệu trong KPDL
- 1.6. Một số lĩnh vực ứng dụng KPDL điển hình
- 1.7. Các bài toán KPDL điển hình
- 1.8. Tính liên ngành của KPDL

Chương 2. Công nghệ tri thức và phát hiện tri thức từ dữ liệu

- 2.1. Vai trò của Công nghệ Thông tin trong kinh tế tri thức
- 2.2. Công nghệ Tri thức
- 2.3. Mô hình phát hiện tri thức từ dữ liệu
- 2.4. Độ đo hấp dẫn trong KPDL

Chương 3. Chuẩn bị dữ liệu

- 3.1. Giới thiệu
- 3.2. Hiểu dữ liệu
- 3.3. Tiền xử lý dữ liệu
- 3.4. Làm sạch dữ liệu
- 3.5. Tích hợp dữ liệu
- 3.6. Chuyển đổi dữ liệu
- 3.7. Thu gọn dữ liệu

Chương 4. Phát hiện luật kết hợp

- 4.1. Giới thiệu về luật kết hợp
- 4.2. Phương pháp khai phá tập mục phổ biến
- 4.3. Thuật toán FP-Growth
- 4.4. Một số ứng dụng của luật kết hợp
- 4.5. Một số thuật toán song song
- 4.6. Khai phá mẫu kết hợp nâng cao

Chương 5. Phân cụm dữ liệu

- 5.1. Giới thiệu
- 5.2. Một số độ đo cơ bản dùng trong phân cụm
- 5.3. Thuật toán phân cụm phẳng
- 5.4. Thuật toán phân cụm phân cấp
- 5.5. Thuật toán phân cụm dựa trên mật độ
- 5.6. Thuật toán phân cụm dựa trên mô hình
- 5.7. Đánh giá các thuật toán phân cụm
- 5.8. Một số thuật toán phân cụm nâng cao

Chương 6. Phân lớp dữ liệu

- 6.1. Giới thiệu
- 6.2. Thuật toán phân lớp cây quyết định
- 6.3. Đánh giá thuật toán phân lớp
- 6.4. Thuật toán phân lớp Naive Bayes
- 6.5. Thuật toán phân lớp máy vector hỗ trợ
- 6.6. Thuật toán phân lớp k-NN
- 6.7. Một số ứng dụng của thuật toán phân lớp
- 6.8. Phân lớp bán giám sát

Chương 7. Tài nguyên và khuynh hướng hiện đại của KPDL

- 7.1. Một số bài học về KPDL
- 7.2. Một số lỗi điển hình trong KPDL
- 7.3. Tài nguyên KPDL
- 7.4. Khuynh hướng hiện đại của KPDL

7. Học liệu

7.1. Học liệu bắt buộc

- [1]. Nguyễn Hà Nam, Nguyễn Trí Thành, Hà Quang Thụy (2013). Giáo trình khai phá dữ liệu (Các chương: 1-6, 10). NXB ĐHQGHN.
- [2]. J. Han, M. Kamber, and Jian Pei (2011). Data Mining: Concepts and Techniques (3rd edition). Morgan Kaufmann.

7.2 Học liệu tham khảo

- [3]. Robert Nisbet, John Elder, and Gary Miner (2009). Handbook of Statistical Analysis and Data Mining, Elsevier.
- [4]. Phan Xuân Hiếu, Đoàn Sơn, Nguyễn Trí Thành, Hà Quang Thụy, Nguyễn Thu Trang, Nguyễn Cẩm Tú (2009). Giáo trình khai phá dữ liệu Web, NXBGD.

8. Hình thức tổ chức dạy học

8.1. Phân bổ lịch trình giảng dạy trong 1 học kỳ (15 tuần)

Hình thức dạy	Số tiết/tuần	Từ tuần ... đến tuần...	Địa điểm
Lý thuyết	3	1-15	Giảng đường
Thực hành	×	5-15	Tự thực hành ở nhà theo chỉ dẫn của giảng viên (không tính giờ môn học)
Tự học bắt buộc	×	1-15	Tự làm tiểu luận ở nhà, phòng thí nghiệm (không tính giờ môn học)

8.2 Lịch trình dạy cụ thể

Tuần	Nội dung giảng dạy lý thuyết/thực hành	Nội dung sinh viên tự học
1	Chương 1. Giới thiệu về KPDL: Giới thiệu môn học; Nhu cầu phát hiện tri thức từ dữ liệu; Quá trình phát hiện tri thức trong CSDL	Đọc chương đầu tiên của học liệu bắt buộc [1,2]
2	Chương 1. Giới thiệu về KPDL (tiếp): Khai phá dữ liệu và xử lý CSDL truyền thống; Kiểu dữ liệu trong KPDL; Một số lĩnh vực ứng dụng KPDL điển hình; Các bài toán KPDL điển hình; Tính liên ngành của KPDL	Tìm hiểu tài liệu tiểu luận (nội dung học liệu tham khảo [3,4] là một phương án)
3	Chương 2. Công nghệ tri thức và phát hiện tri thức từ dữ liệu	Đọc chương 2 của HLBB [1]
4	Chương 3. Chuẩn bị dữ liệu: Hiểu dữ liệu; Tiền xử lý dữ liệu; Làm sạch dữ liệu	Đọc chương 3 của HLBB [1] và chương 2 của HLBB [2]
5	Chương 3. Chuẩn bị dữ liệu (tiếp): Tích hợp dữ liệu; Chuyển đổi dữ liệu; Thu gọn dữ liệu	Đọc chương 3 của HLBB [1] và chương 3 của HLBB [2]
6	Chương 4. Phát hiện luật kết hợp: Giới thiệu về luật kết hợp, Phương pháp khai phá tập mục phổ biến	Đọc chương 4 của HLBB [1] và các chương 6,7 của HLBB [2]

7	Chương 4. Khai phá luật kết hợp (tiếp): Thuật toán FP-Growth; Một số ứng dụng của luật kết hợp; Một số thuật toán song song; Khai phá mẫu kết hợp nâng cao	Thực hiện bài tiểu luận
8	Chương 5. Phân cụm: Giới thiệu; Một số độ đo cơ bản dùng trong phân cụm; Thuật toán phân cụm phẳng; Thuật toán phân cụm phân cấp	Đọc chương 5 của HLBB [1] và chương 10 của HLBB [2]
9	Chương 5. Phân cụm (tiếp): Thuật toán phân cụm dựa trên mật độ; Thuật toán phân cụm dựa trên mô hình; Đánh giá phân cụm	Đọc chương 5 của HLBB [1] và chương 10 của HLBB [2]
10	Chương 5. Phân cụm (tiếp): Một số thuật toán phân cụm nâng cao	Đọc chương 11 của HLBB [2] Hoàn thiện và nộp bài tiểu luận
11	Chương 6. Phân lớp: Giới thiệu; Thuật toán phân lớp cây quyết định; Đánh giá thuật toán phân lớp; Thuật toán phân lớp Naive Bayes	Đọc chương 6 của HLBB [1] và chương 8 của HLBB [2]
12	Chương 6. Phân lớp (tiếp): Thuật toán phân lớp máy vector hỗ trợ; Thuật toán phân lớp k-NN	Đọc chương 6 của HLBB [1] và chương 9 của HLBB [2]
13	Chương 6. Phân lớp (tiếp): Một số ứng dụng của thuật toán phân lớp; Phân lớp bán giám sát	Đọc chương 9 của HLBB [2]
14	Tài nguyên và khuynh hướng hiện đại của KPD	Đọc chương 10 của HLBB [1] và chương 13 của HLBB [2]
15	Tổng hợp nội dung môn học và thảo luận về vấn đề liên quan	

9. Chính sách đối với môn học và các yêu cầu khác của giảng viên

- Khuyến khích sinh viên tham gia xây dựng bài học theo nhiều hình thức (i) trả lời câu hỏi trên lớp, (ii) đặt câu hỏi, nêu vấn đề liên quan nội dung bài học, (iii) các hình thức khác tăng cường chất lượng học tập chung của lớp (chuyên cần học, tạo môi trường học tập...).

10. Phương pháp, hình thức kiểm tra, đánh giá kết quả học tập môn học

10.1. Mục đích và trọng số kiểm tra, đánh giá

Hình thức	Phương pháp	Mục đích	Trọng số
Tham gia xây dựng bài học	Số lượng (i) trả lời đúng câu hỏi của giáo viên, (ii) câu hỏi tốt mà sinh viên đặt ra về nội dung bài giảng và chất lượng giải bài tập/tiểu luận.	Đánh giá năng lực phát hiện vấn đề, phân tích và lựa chọn giải pháp sinh viên	25%

Hình thức	Phương pháp	Mục đích	Trọng số
Chuyên cần	Số lượng buổi tham dự trên lớp, thái độ học trên lớp và làm tiểu luận ở nhà	Đánh giá tinh thần, thái độ làm việc (bao gồm làm việc nhóm)	15%
Thi kết thúc môn học	Thi vấn đáp nếu số sinh viên dưới 40 người, ngược lại thi viết	Đánh giá toàn diện kiến thức, kỹ năng sinh viên đạt được khi kết thúc môn học	60%
Tổng			100%

10.2. Tiêu chí đánh giá

- Tiêu chí đánh giá cụ thể với từng đầu điểm của môn học:

- Tham gia xây dựng bài học: Số lượng và chất lượng câu hỏi, câu trả lời của sinh viên tại các buổi học trên lớp;
- Chuyên cần: Số lượng buổi tham dự trên lớp, thái độ học trên lớp và chất lượng bài tiểu luận
- Kết thúc môn: nắm được kiến thức, kỹ năng đã học trong cả 15 tuần của học kỳ.

- Cụ thể việc đánh giá kiến thức, kỹ năng của sinh viên theo các mức đáp ứng được chuẩn đầu ra, mức khá, mức giỏi, mức trung bình:

- Giỏi: Mọi yêu cầu về kiến thức, kỹ năng và phẩm chất đối với môn học (mục 4. Chuẩn đầu ra môn học) đều được đảm bảo;
- Khá: Mọi yêu cầu về kiến thức và kỹ năng ở bậc 1, 2 đối với môn học được đảm bảo còn các yêu cầu ở bậc 3 được đảm bảo ở bậc 2;
- Trung bình: Mọi yêu cầu về kiến thức, kỹ năng và phẩm chất đối với môn học đều được đảm bảo ít nhất ở bậc 1 trong đó hơn nửa các yêu cầu bậc 2,3 được đảm bảo ở bậc 2.

10.3. Lịch thi và kiểm tra

Hình thức thi và kiểm tra	Thời gian
Nộp Tiểu luận	Tuần 10-11
Thi cuối kỳ	Theo lịch của Trường

Duyệt

Chủ nhiệm Khoa

Chủ nhiệm bộ môn