

TRƯỜNG ĐẠI HỌC KHN TẾ - KỸ THUẬT CÔNG NGHIỆP
KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN: HỆ THỐNG THÔNG TIN

ĐỀ CƯƠNG CHI TIẾT
HỌC PHẦN: KHAI PHÁ DỮ LIỆU

1. THÔNG TIN CHUNG

Tên học phần (tiếng Việt):	Khai phá dữ liệu
Tên học phần (tiếng Anh):	Data Mining
Mã môn học:	
Khoa/Bộ môn phụ trách:	Khoa CNTT/HTTT
Giảng viên phụ trách chính:	ThS. Trần Thanh Đại Email: ttdaiuneti@gmail.com
GV tham gia giảng dạy:	ThS. Hoàng Thị Minh Châu ThS. Vũ Mỹ Hạnh
Số tín chỉ:	3 (36,18,45,90)
Số tiết Lý thuyết:	36
Số tiết TH/TL:	18 $36+18/2 = 15$ tuần x 3 tiết/tuần
Số tiết Tự học:	90
Tính chất của học phần:	Tự chọn
Học phần học trước:	Cơ sở dữ liệu, Cấu trúc dữ liệu và giải thuật
Học phần tiên quyết :	Không
Các yêu cầu của học phần:	- Sinh viên phải có tài liệu học tập - Sinh viên phải có máy tính cá nhân

2. MÔ TẢ HỌC PHẦN

Cung cấp các kiến thức cơ bản về khai phá dữ liệu, giới thiệu các bài toán khai phá dữ liệu điển hình và ứng dụng của chúng trong thực tiễn, các thuật toán khai phá dữ liệu hiện đại, hướng dẫn sử dụng công cụ khai phá dữ liệu Weka.

3. MỤC TIÊU CỦA HỌC PHẦN ĐỐI VỚI NGƯỜI HỌC

Kiến thức

Vận dụng được các công cụ khai phá dữ liệu điển hình vào các ứng dụng cụ thể trong thực tiễn

Kỹ năng

Phân tích: mối quan hệ giữa các yêu cầu 0075 cụ thể trong thực tiễn với các bài toán khai phá dữ liệu đã được học.

Sử dụng: thành thạo các công cụ khai phá dữ liệu của Weka trong việc triển khai quá trình phát hiện tri thức phục vụ cho các yêu cầu trong thực tiễn.

Năng lực tự chủ và trách nhiệm

Nghiêm túc, trách nhiệm, chủ động, tích cực, chăm chỉ, cẩn thận.

4. CHUẨN ĐẦU RA HỌC PHẦN

Mã CDR	Mô tả CDR học phần <i>Sau khi học xong môn học này, người học có thể:</i>	CDR của CTĐT
G1	Về kiến thức	
<i>G1.1.1</i>	Hiểu được ý nghĩa của học phần KPDL và các lĩnh vực liên quan đến KPDL	1.2.1
<i>G1.1.2</i>	Hiểu được quy trình KPDL, mối liên hệ của các bài toán KPDL với thống kê và học máy	1.2.1
<i>G1.2.1</i>	Nắm được các lớp thuật toán KPDL và biết cách vận dụng chúng vào các yêu cầu KPDL thực tế	1.2.2
G2	Về kỹ năng	
<i>G2.1.1</i>	Biết các bước trong quy trình KPDL và vận dụng phù hợp với tình hình thực tế của yêu cầu KPDL	2.1.1
<i>G2.1.2</i>	Biết vận dụng các thuật toán đã học cho các yêu cầu KPDL trong thực tiễn	2.1.2
<i>G2.2.1</i>	Biết sử dụng công cụ Weka triển khai một số lớp thuật toán KPDL cơ bản	2.2.1
G3	Năng lực tự chủ và trách nhiệm	
<i>G3.1.1</i>	Tuân thủ bảo vệ tính riêng tư trong KPDL	3.1.1
<i>G3.1.2</i>	Tôn trọng quyền tác giả	3.1.2
<i>G3.2.1</i>	Chủ động trong công việc được giao	3.2.1
<i>G3.2.2</i>	Rèn luyện được áp lực công việc	3.2.1

5. NỘI DUNG MÔN HỌC, KẾ HOẠCH GIẢNG DẠY

Tuần thứ	Nội dung	Số tiết LT	Số tiết TH	Tài liệu học tập, tham khảo
1	Chương 1: Giới thiệu chung <i>1.1. Nhu cầu phát hiện tri thức</i> <i>1.2. Phát hiện tri thức trong CSDL</i> <i>1.3. Các kiểu dữ liệu</i> <i>1.4. Phân loại các bài toán KPDL</i> <i>1.5. Các lĩnh vực liên quan</i>	3		1, 2, 3, 4

	<p>1.6. Các vấn đề thách thức thời sự</p> <p>1.7. Công cụ Weka</p>			
2	<p>Chương 2: Chuẩn bị dữ liệu</p> <p>2.1. Hiểu dữ liệu</p> <p>2.2. Làm sạch dữ liệu</p> <p>2.3. Biến đổi dữ liệu</p> <p>2.4. Tích hợp dữ liệu</p> <p>2.5. Thu gọn dữ liệu</p> <p>2.6. Chuẩn bị dữ liệu bằng Weka</p>	3		1, 2, 3, 4
3	<p>Chương 3: Phát hiện luật kết hợp</p> <p>3.1. Các khái niệm cơ bản</p> <p>3.1.1. Tập phổ biến và các đơn vị đo cơ bản</p> <p>3.1.2. Các tính chất của tập phổ biến</p> <p>3.1.3. Tập phổ biến cực đại</p> <p>3.1.4. Tập phổ biến đóng</p> <p>3.2. Thuật toán Apriori</p> <p>3.2.1. Phương pháp xây dựng ứng viên</p> <p>3.2.2. Phương pháp loại bỏ các ứng viên</p> <p>3.2.3. Thuật toán</p> <p>3.2.4. Phân tích ưu và nhược điểm</p>	3		1, 2, 3, 4
4	<p>3.3. Thuật toán FP-Growth</p> <p>3.3.1. Phương pháp tổ chức và lưu trữ của cây FP</p> <p>3.3.2. Phương pháp duyệt cây FP</p> <p>3.3.3. Thuật toán</p> <p>3.3.4. Nhận xét và đánh giá</p>	3		1, 2, 3, 4
5	<p>3.4. Sử dụng công cụ Weka để phát hiện luật kết hợp</p> <p>3.4.1. Chuẩn bị dữ liệu</p> <p>3.4.2. Khai phá luật kết hợp bằng thuật toán Apriori</p> <p>3.4.3. Khai phá luật kết hợp bằng thuật toán FP-Growth</p>	3		1, 2, 3, 4
6	Chữa bài tập + Kiểm tra	0	6	1, 2, 3, 4
7	<p>Chương 4: Phân lớp</p> <p>4.1. Tổng quan</p> <p>4.1.1. Các tình huống phân lớp dữ liệu</p> <p>4.1.2. Các khái niệm cơ bản</p> <p>4.1.3. Các giải thuật phân lớp</p> <p>4.2. Phương pháp phân chia dữ liệu</p> <p>4.2.1. Ma trận gây lẫn</p> <p>4.2.2. Phương pháp Holdout</p>	3		1, 2, 3, 4

	4.2.3. Phương pháp Cross validation			
8	4.2. Giải thuật phân lớp bằng cây quyết định 4.2.1. Các khái niệm cơ bản 4.2.2. Cây ID3	3		1, 2, 3, 4
9	4.3. Giải thuật phân lớp bằng mạng Bayesian 4.3.1. Các khái niệm cơ bản 4.3.2. Giải thuật phân lớp Bayesian	3		1, 2, 3, 4
10	Chữa bài tập + Kiểm tra	0	6	1, 2, 3, 4
11	4.4. Sử dụng công cụ Weka để phân lớp dữ liệu 4.4.1. Chuẩn bị dữ liệu 4.4.2. Phân lớp theo cây quyết định 4.4.3. Phân lớp theo mạng Bayes	3		1, 2, 3, 4
12	Chương 5: Phân cụm 5.1. Phương pháp K-means 5.1.1. Cơ sở lý thuyết 5.1.2. Giải thuật phân cụm K-mean	3		1, 2, 3, 4
13	5.2. Một số phương pháp phân cụm khác 5.2.1. Phương pháp phân cụm dựa trên HAC 5.2.2. Phương pháp phân cụm dựa trên mật độ	3		1, 2, 3, 4
14	5.3. Sử dụng công cụ Weka để phân cụm dữ liệu 5.3.1. Chuẩn bị dữ liệu 5.3.2. Phân cụm theo phương pháp K-Mean 5.3.3. Phân cụm theo phương pháp Expectation - Maximization	3		1, 2, 3, 4
15	Chữa bài tập + Kiểm tra	0	6	1, 2, 3, 4

6. MA TRẬN MỨC ĐỘ ĐÓNG GÓP CỦA NỘI DUNG GIẢNG DẠY ĐỂ ĐẠT ĐƯỢC CHUẨN ĐẦU RA CỦA HỌC PHẦN

Mức 1: Thấp

Mức 2: Trung bình

Mức 3: Cao

(Lưu ý: Khi đánh giá mức độ đóng góp từng “nội dung giảng dạy” tới các tiêu chuẩn (Gx.x.x) sẽ ảnh hưởng tới việc phân bổ thời lượng giảng dạy của từng phần nội dung giảng dạy và mức độ ưu tiên kiểm tra đánh giá nội dung đó).

Chương	Nội dung giảng dạy	Chuẩn đầu ra học phần														
		G1.1.1	G1.2.1	G1.2.2	G1.2.3	G1.2.4	G1.2.5	G.2.1.1	G.2.1.2	G.2.2.1	G.2.2.2	G2.2.3	G3.1.1	G3.1.2	G3.2.1	
1	Chương 1: Tổng quan															
	1.1. Nhu cầu phát hiện tri		1	1		2	2	2	2	2			2	2	3	

	thức													
	1.2. Phát hiện tri thức trong CSDL	1	1				2	1	2			2		3
	1.3. Các kiểu dữ liệu	2	2				1	2	1			2	2	3
	1.4. Phân loại các bài toán KPDL	3	1				1	1	2			2		3
	1.5. Các lĩnh vực liên quan	2	2				2	1	1			2	2	3
	1.6. Các vấn đề thách thức thời sự	1	1				1	2	2			2		3
	1.7. Công cụ Weka	2	3				2	1	2			2		3
Chương 2: Tiền xử lý dữ liệu														
2	2.1. Hiệu dữ liệu	1	1				2	1	2			2		3
	2.2. Làm sạch dữ liệu	2	2				1	2	1			2	2	3
	2.3. Biến đổi dữ liệu	3	1				1	1	2			2		3
	2.4. Tích hợp dữ liệu	2	2				2	1	1			2	2	3
	2.5. Thu gọn dữ liệu	1	1				1	2	2			2		3
	2.6. Chuẩn bị dữ liệu bằng Weka	2	3				2	1	2			2		3
Chương 3: Phát hiện luật kết hợp														
	3.1. Các khái niệm cơ bản	3												
	3.2. Thuật toán Apriori	3												
	3.3. Thuật toán FP-Grow		3											
	3.4. Sử dụng công cụ Weka để phát hiện luật kết hợp					2								
Chương 4: Phân lớp dữ liệu														
	4.1. Phương pháp phân	3												

	chia dữ liệu													
	4.2. Phân lớp dữ liệu bằng cây quyết định		3											
	4.3. Phân lớp dữ liệu bằng mạng Bayesian			3										
	4.4. Sử dụng công cụ Weka để phân lớp dữ liệu		3				2							
Chương 5: Phân cụm dữ liệu														
	5.1. Phương pháp K-means		3											
	5.2. Phương pháp HAC			3										
	5.3. Sử dụng công cụ Weka để phân cụm dữ liệu						2							

7. PHƯƠNG THỨC ĐÁNH GIÁ HỌC PHẦN

TT	Điểm thành phần	Quy định (Theo QĐ Số: 686/QĐ- ĐHKTKTCN)	Chuẩn đầu ra học phần											
			G1.1.1	G1.1.2	G1.2.1	G1.2.2	G2.1.1	G2.1.2	G2.2.1	G2.2.2	G3.1.1	G3.2.1	G3.2.1	
1	Điểm quá trình (40%)	1. Kiểm tra thường xuyên + Hình thức: <i>Tham gia thảo luận, kiểm tra 15 phút, hỏi đáp</i> + Số lần: <i>Tối thiểu 1 lần/sinh viên</i> + Hệ số: <i>1</i>	x	x	x	x	x	x	x	x	x	x	x	x
		2. Kiểm tra định kỳ lần 1 + Hình thức: <i>Tự luận</i>	x	x	x	x	x					x	x	x

		+ Thời điểm: <i>Tuần</i> ... (sau khi học xong chương 3) + Hệ số: 2											
		3. Kiểm tra định kỳ lần 2 + Hình thức: Tự luận + Thời điểm: <i>Tuần</i>(sau khi học xong chương 4) + Hệ số: 2	x	x	x			x	x	x	x	x	x
		4. Kiểm tra định kỳ lần..... + Hình thức: Tự luận + Thời điểm: <i>Tuần</i>(sau khi học xong chương 5) + Hệ số: 2	x	x	x	x	x	x	x	x	x	x	x
												
2	Điểm thi kết thúc học phần (60%)		x	x	x	x	x	x	x	x	x	x	x

Ghi chú: Số lần kiểm tra định kỳ bằng số tín chỉ học phần

8. PHƯƠNG PHÁP DẠY VÀ HỌC

✓ Giảng viên giới thiệu học phần, tài liệu học tập, tài liệu tham khảo, các địa chỉ website để tìm tư liệu liên quan đến môn học. Nêu nội dung cốt lõi của chương và tổng kết chương, sử dụng bài giảng điện tử và các mô hình giáo cụ trực quan trong giảng dạy. Tập trung hướng dẫn học, tư vấn học, phản hồi kết quả thảo luận, bài tập lớn, kết quả kiểm tra và các nội dung lý thuyết chính mỗi chương.

✓ Các phương pháp giảng dạy có thể áp dụng: Phương pháp thuyết trình; Phương pháp thảo luận nhóm; Phương pháp mô phỏng; Phương pháp minh họa; Phương pháp miêu tả, làm mẫu.

✓ Sinh viên chuẩn bị bài từng chương, làm bài tập đầy đủ, trau dồi kỹ năng làm việc nhóm để chuẩn bị bài thảo luận.

✓ Trong quá trình học tập, sinh viên được khuyến khích đặt câu hỏi phản biện, trình bày quan điểm, các ý tưởng sáng tạo mới dưới nhiều hình thức khác nhau.

9. QUY ĐỊNH CỦA HỌC PHẦN

9.1. Quy định về tham dự lớp học

- ✓ Sinh viên/học viên có trách nhiệm tham dự đầy đủ các buổi học. Trong trường hợp nghỉ học do lý do bất khả kháng thì phải có giấy tờ chứng minh đầy đủ và hợp lý.
- ✓ Sinh viên vắng quá 50% buổi học dù có lý do hay không có lý do đều bị coi như không hoàn thành khóa học và phải đăng ký học lại vào học kỳ sau.
- ✓ Tham dự các tiết học lý thuyết
- ✓ Thực hiện đầy đủ các bài tập được giao trong cuốn tài liệu học tập.
- ✓ Tham dự kiểm tra giữa học kỳ
- ✓ Tham dự thi kết thúc học phần
- ✓ Chủ động tổ chức thực hiện giờ tự học

9.2. Quy định về hành vi lớp học

- ✓ Học phần được thực hiện trên nguyên tắc tôn trọng người học và người dạy. Mọi hành vi làm ảnh hưởng đến quá trình dạy và học đều bị nghiêm cấm.
- ✓ Sinh viên phải đi học đúng giờ quy định. Sinh viên đi trễ quá 15 phút sau khi giờ học bắt đầu sẽ không được tham dự buổi học.
- ✓ Tuyệt đối không làm ồn, gây ảnh hưởng đến người khác trong quá trình học.
- ✓ Tuyệt đối không được ăn uống, nhai kẹo cao su, sử dụng các thiết bị như điện thoại, máy nghe nhạc trong giờ học.

10. TÀI LIỆU HỌC TẬP, THAM KHẢO

10.1. Tài liệu học tập:

[1] Nguyễn Hà Nam, Nguyễn Chí Thành, Hà Quang Thụy, Giáo trình Khai Phá Dữ Liệu, NXB Đại học quốc gia Hà Nội 2013

10.2. Tài liệu tham khảo:

[2] Lê Văn Phùng-Quách Xuân Trường, Khai phá dữ liệu, NXB Thông tin và truyền thông 2017.

[3] Hà Quang Thụy, Phan Xuân Hiếu, Đoàn Sơn, Giáo trình khai phá dữ liệu Web, NXB Giáo dục 2011.

[4] Đỗ Trung Tuấn, Phân tích thống kê và khai phá dữ liệu, NXB Đại học Quốc gia Hà Nội 2018.

11. HƯỚNG DẪN THỰC HIỆN

- ✓ Các Khoa, Bộ môn phổ biến đề cương chi tiết cho toàn thể giáo viên thực hiện.
- ✓ Giảng viên phổ biến đề cương chi tiết cho sinh viên vào buổi học đầu tiên của học phần.
- ✓ Giảng viên thực hiện theo đúng đề cương chi tiết đã được duyệt.

Hà Nội, Ngày tháng năm 2018

Trưởng khoa
(Ký và ghi rõ họ tên)

Trưởng bộ môn
(Ký và ghi rõ họ tên)

Người biên soạn
(Ký và ghi rõ họ tên)

