Trường Đại học Bách Khoa - ĐHQG Tp.HCM
Khoa: Khoa Khoa học và Kỹ thuật Máy tính
Khoa/Bộ môn quản lý MH: Công nghệ Phần mềm

Tp.HCM, ngày ..... tháng ..... năm .....

Đề cương môn học Sau đại học

CÁCH TIẾP CẬN HIỆN ĐẠI TRONG XỬ LÝ NGÔN NGỮ TỰ NHIÊN
(MODERN APPROACHES FOR NATURAL LANGUAGE PROCESSING)

Mã số MH: CO5256

Số tín chỉ:	**Tc (LT.BT&TH.Tự Học): 3**						**ECTS: 6**				
Số tiết -Tổng:	**75**	LT:	**30**	BT:	**0**	TH:	**0**	ĐA:		BTL/TL:	**45**
Đánh giá:	Bài tập lớn:			25%							
	Bài tập trên lớp:			15%							
	Thi:			60%							
- Môn tiên quyết:											
- Môn học trước:											
- Môn song hành:											
- CTĐT ngành (Mã ngành):	Khoa Học Máy Tính (8480101)										
- Ghi chú khác:											

1. <u>Mục tiêu môn học:</u>

Đây là một khoá học nâng cao về xử lý ngôn ngữ tự nhiên (NLP). Mục tiêu của khoá học là nhằm trang bị cho người học các kiến thức về:

- biểu diễn và xử lý văn bản

- các hướng tiếp cận học sâu cho NLP

- các công việc chính của NLP và cách xử lý chúng bằng học sâu

- hiện thực và kiểm tra các công việc NLP bằng học sâu

Aims:

This is an advanced course on natural language processing. The course objective is to equip the learners with knowledge of

- text representation and processing

- deep learning approaches for NLP

- core tasks of NLP and how to handle them by deep learning

- implementation and testing of NLP tasks by deep learning

2. Nội dung tóm tắt môn học:

Đây là một khoá học ứng dụng tập trung vào các thành tựu mới của xử lý ngôn ngữ tự nhiên (NLP) sử dụng các hướng tiếp cận hiện đại như học sâu. Khoá học trước tiên cung cấp các kiến thức và khái niệm cơ bản về việc biểu diễn và xử lý văn bản. Tiếp đó, cách biểu diễn sự phân bố từ bằng các vector được trình bày. Sau đó, người hoc sẽ được học các mô hình học sâu phổ biến, chẳng hạn như CNN và RNN cùng với các mô hình mở rộng của chúng. Một số mô hình tiên tiến như BERT cũng được trình bày. Cuối cùng, các công việc và ứng dụng nền tảng của NLP và cách giải quyết chúng bằng học sâu sẽ được trình bày.

Course outline:

It is an applied course focussing on recent advances in natural language processing (NLP) using modern approachs such as deep learning. The course firstly gives fundamental knowledge and concepts of text representation and processing. Then, distributed representation of word vectors is covered. Next, popular deep learning models, such as CNN and RNN and their enhanced variants, are discussed. Some advanced models like BERT are also introduced. The course is concluded by the introduction of some core tasks and applications of NLP and how to handle them by deep learning.

3. Tài liệu học tập:

[1] Quản Thành Thơ, *Mạng Nơ-ron nhân tạo: từ hồi quy đến học sâu*, Nhà Xuất Bản ĐHQG TPHCM, 2021.

[2] Mitchell, T.M., Học máy, Nhà xuất bản McGraw-Hill, 1997.

[3] Russell, S.J. and Norvig, P., Trí tuệ nhân tạo: Một hướng tiếp cận hiện đại, Nhà xuất bản Prentice Hall, 2002.

[4] Ian Goodfellow, Yoshua Bengio, và Aaron Courville, Deep Learning, MIT Press, http://www.deeplearningbook.org, 2016

4. Các hiểu biết, các kỹ năng cần đạt được sau khi học môn học:

STT	Chuẩn đầu ra môn học (CĐRMH)	Công cụ đánh giá CĐRMH	Đóng góp CĐR Chương trình (CĐRCT)		
			Ứng dụng	Nghiên cứu	
CĐRMH.1	Hiểu được các khái niệm căn bản của biểu diễn văn bản	Bài tập trên lớp, Thi	a		1.3
CĐRMH.2	Hiểu các định nghĩa về các mô hình mạng neural dùng cho NLP	Bài tập lớn, Bài tập trên lớp	a		1.3
CĐRMH.3	Biết cách phối hợp các mô hình học sâu để phát triển các hệ thống NLP hiện đại	Bài tập lớn, Bài tập trên lớp, Thi	a		1.3
CĐRMH.4	Biết cách hiện thực và đánh giá các hệ thống cho các công việc NLP	Bài tập lớn, Thi	a		4.5

Learning outcomes:

No.	Course learning outcomes (CLO)	CLO assessment	Matching with PLO		
			Coursework	Research	
L.O.1	Understand basic concepts of text representation	Exercises, Final exam	a		1.3

No.	Course learning outcomes (CLO)	CLO assessment	Matching with PLO		
			Coursework	Research	
L.O.2	Understand the definition of a range of neural network models applied for NLP	Assignment, Exercises	a		1.3
L.O.3	Know how to combine those models to build state-of-the-art NLP systems	Assignment, Exercises, Final exam	a		1.3
L.O.4	Be able to implement and evaluate systems for core NLP tasks	Assignment, Final exam	a		4.5

Bảng ánh xạ chuẩn đầu ra môn học và chuẩn đầu ra chương trình ứng dụng:

	Chuẩn đầu ra của chương trình (CĐRCT)										
Chuẩn đầu ra môn học (CĐRMH)	a	b	c	d	e	f	g	h	i	j	k
CĐRMH.1	✓										
CĐRMH.2	✓										
CĐRMH.3	✓										
CĐRMH.4	✓										

Bảng ánh xạ chuẩn đầu ra môn học và chuẩn đầu ra chương trình nghiên cứu:

	Chuẩn đầu ra của chương trình (CĐRCT)										
Chuẩn đầu ra môn học (CĐRMH)	a	b	c	d	e	f	g	h	i	j	k
CĐRMH.1											
CĐRMH.2											
CĐRMH.3											
CĐRMH.4											

5. <u>Hướng dẫn cách học - chi tiết cách đánh giá môn học:</u>

Đọc kỹ tài liệu

Giải các bài tập được yêu cầu trong các phần hướng dẫn

Làm các bài tập lớn

Tập trung trong các giờ giảng trên lớp

Learning strategies & Assessment Scheme:

Read the textbook intensively.

Solve the required exercises in advance of tutorial sessions

Implement the assignment

Attend the lectures as concentrative as possible

6. <u>Nội dung chi tiết:</u>

Tuần/ Buổi	Chủ đề (chương)	Nội dung	Chuẩn đầu ra môn học	Tài liệu
1	Chương 1. Giới thiệu	Học máy và NLPSự phát triển của mạng neural với NLPKhông gian vectors và trọng số tf.idf	CĐRMH.1, CĐRMH.2	[1],[2]
2	Chương 2. Biểu diễn từ bằng vector	Biểu diễn từ bằng one-hot vectorsPhương pháp word2vecPhương pháp Glove	CĐRMH.1	[3]
3	Chương 3. Mạng neural	- Perceptron- Mạng neural- Sử dụng mạng neural để xử lý văn bản	CĐRMH.2	[1],[2]
4,5	Chương 4. Mạng RNN	- Mạng RNN- GRU và LSTM- Ứng dụng cho các bài toán NLP	CĐRMH.2	[1],[3]
6,7	Chương 5. Cơ chế attention và seq2seq	- Cơ chế attention- Kiến trúc seq2seq- Ứng dụng cho các bài toán NLP	CĐRMH.2	[1],[3]
8	Chương 6. Mạng CNN cho các bài toán NLP	- Mạng CNN- Áp dụng CNN để xử lý văn bản- Mô hình kết hợp CNN và RNN	CĐRMH.2, CĐRMH.3	[1],[3]
9	Chương 7. Sử dụng các mô hình pre-trained	- Cơ chế transfer learning- Các mô hình pre-trained: GMT, ULMFit- Mạng Transformer và BERT	CĐRMH.3, CĐRMH.4	
10	Ôn tập		CĐRMH.1, CĐRMH.2, CĐRMH.3, CĐRMH.4	

7. <u>Giảng viên tham gia giảng dạy:</u>

CBGD chính:	PGS.TS Quản Thành Thơ
CBGD tham gia:	TS. Nguyễn Đức Dũng

XÁC NHẬN
CỦA HỘI
ĐỒNG XÂY
DỰNG
CHƯƠNG
TRÌNH ĐÀO
TẠO VÀ KHOA

Tp. Hồ Chí Minh, ngày ..... tháng ..... năm .....

GIẢNG
VIÊN
LẬP ĐỀ
CƯƠNG

PGS.TS
Quản
Thành
Thơ