ISOM4030

Natural Language Processing in Business Using Deep Learning

Fall 2024

Instructor: **Dr. Jia JIA (justinjia@ust.hk)**

Office: LSK 5045

Office Hours: By appointment

Teaching Assistant: **Samuel LAI (imsamuel@ust.hk)**

Office: LSK 4063

Class Schedule

Section	Date	Time	Venue
Lecture	Tue./Thur.	4:30PM - 5:50PM	LSK1001
Lab	Tue.	6:00PM - 7:20PM	LSKG005

Course Website: https://canvas.ust.hk. All course materials and announcements will be posted on this site. You are advised to check it regularly throughout the course.

COURSE DESCRIPTION

The past two years have been hailed as the 'iPhone moment' for Artificial Intelligence (AI) due to the rapid advancements and groundbreaking achievements in the field of Large Language Models (LLMs). AI applications like ChatGPT and Sora, powered by LLMs, have demonstrated remarkable capabilities in natural language processing, understanding, and generation. Deep-learning-based NLP Applications have suddenly become ubiquitous, profoundly transforming the way we work, live and communicate. As a result, there is an imminent surge in demand for individuals who possess a thorough understanding of these powerful tools and can adeptly apply and adapt them to diverse business applications, including (but not restricted to) Web search, advertising, information extraction, customer service, language translation, virtual agents, etc.

This course aims to offer students a comprehensive introduction to cutting-edge neural approaches for NLP, using Python as the primary implementation tool. Students will gain the necessary skills to

not only comprehend but also design and implement their own deep learning models to tackle a wide range of NLP tasks. Additionally, they will get familiar with prominent business applications of these models in the real world, and learn how to fine-tune pre-trained LLMs to achieve state-of-the-art performance for their own business applications.

After taking this course, students should:

- Know the basics of NLP techniques and deep learning models well enough to interact with business analysts, data scientists, engineers, and machine learning developers.
- Be able to apply NLP techniques and deep learning approaches to analyze text for business intelligence.
- Gain hands-on experience in using popular NLP and deep learning tools and get ready for the job positions that require familiarities with the NLP and deep learning tools.

LEARNING OBJECTIVES

1. Approach real-world Natural Language Processing problems using Deep Learning methods (PILO: 2, 3, 5, 7)
2. Be familiar with various concepts and theoretical frameworks of Deep Learning and Natural Language Processing and understand how they interact to contribute to the recent AI boom. (PILO: 2, 3, 5, 7)
3. Be able to develop programming solutions to real-world NLP problems in the business domain (PILO: 3, 5, 7)
4. Be able to fine-tune off-the-shelf Large Language Models (e.g., BERT, GPTs, T5, etc.) and use them for their own business applications (PILO: 3, 5, 7).

PREREQUISITES

- Official prequisites: ISOM 3360 or ISOM 3400

- All labs and assignments will be in Python. If you need a refresher on Python, or are not very familiar with NumPy, I highly recommend attending the first several lab sessions. These will help you catch up and ensure you have a solid foundation in these essential tools.

- College Calculus and Linear Algebra (e.g. MATH 1003 or equivalent)
  You should be comfortable taking derivatives and understanding matrix/vector notation and operations. We will have labs specifically designed to reinforce these foundational concepts.

- Basic Probability and Statistics (e.g. ISOM 2500 or equivalent)
  You should know the basics of probabilities, gaussian/normal distributions, mean, standard deviation, etc.

MATERIALS

1. **CLASS WEBSITE**
   All course relevant materials, including but not limited to, reading materials, handouts, and programming files, will be uploaded to CANVAS. You are advised to check this site regularly throughout this course.

2. **MAIN READING**
   The course will not follow a specific book, but will draw from a number of sources. We list relevant books below. We will also put up links to relevant reading material for each class.
   - Speech and Language Processing by Daniel Jurafsky and James H. Martin
   - Neural Networks and Deep Learning by Michael A. Nielsen.
   - Dive Into Deep Learning by Aston Zhang, et al.
   - Natural Language Processing in Action by Hobson Lane et al. 2019
   - A Primer on Neural Network Models for Natural Language Processing by Yoav Goldberg

3. **SOFTWARE REQUIREMENT**
   - Google CoLab – Most deep learning models will be built and run using Google CoLab programming environment. You must have a Google account, and have access to Google and Google Drive, in order to use Google CoLab.
   - Jupyter Notebook – For Python and Machining Learning fundamentals, you may also install Anaconda and use Jupyter Notebook for faster and easier programming in Python.

EVALUATION

Components	Percentage of the grade
A. Lab submissions	10%
B. Assignment 1	15%
C. Assignment 2	15%
D. Assignment 3	15%
E. Final exam – Concepts	25%
F. Final exam – Applications	20%
TOTAL:	**100%**

A. Lab Submissions

Labs are an essential part of the learning process. You are NOT going to learn without practicing data analysis and modeling yourselves. During lab sessions, we expect you to be able to actively link the empirical results to the concepts covered in the lectures, and learn necessary skills for completing the homework assignments. Lab submissions will be graded and count towards your final grade.

B. Assignments

3 individual assignments will be given throughout the semester. Each assignment has a specific due date and time. Late submission **within 24 hours** after the specified due date and time will be accepted with a **25% penalty**. Late submission **beyond 24 hours** will NOT be accepted.

C. Final Exam

There will be ONE open-book final exam scheduled in the final examination period. It comprises two parts labeled as *Final exam – Concepts* (25%) and *Final exam – Applications* (20%) in the evaluation section. Details of the exam will be provided later in the semester.

ACADEMIC INTEGRITY

Academic integrity is a critical value of the university community. Integrity violations destroy the fabric of a learning community and the spirit of inquiry that is vital to the effectiveness of the University. I have absolutely no tolerance for cheating and there are no acceptable excuses. Anyone caught cheating, plagiarizing, and any other form of academic dishonesty will have their course grade lowered by at least one letter grade. Any unethical behavior or evidence of dishonesty in this course will be reported to the University. Please remember the current university rule: "If a student is discovered cheating however minor the offence, the course grade will appear on the student's record with an X, to show that the grade resulted from cheating. This X grade stays on the record until graduation. If the student cheats again and "earns" another X grade, the student will be dismissed from the University."

Plagiarism is copying anything (text or ideas) from another source without citing that source. If you use another person's idea you must cite it, even if you rewrite the idea in your own words. Extreme care must be taken to avoid passing of other's work as one's own. You are required to provide appropriate citations when you use ideas and arguments or otherwise draw on others' work. If you use research from another source or from the Web you MUST cite the source. This is true even if you use only the general idea and not the exact words.

OTHERS

Email Policy

Please put **[ISOM4030] at the beginning of the subject line of your email along with your email subject**. Failure to do so may result in a longer response time.

As expected, there will be numerous emails when it is closer to the due dates. If you need any assistance, raise them **as early as possible**, and/or take advantage of the office hours of the instructor and the TA. Note that **neither the instructor nor the TAs will provide direct answers to the assignments**.

Learning Environment

We welcome feedbacks on our teaching throughout the semester. You are encouraged to contact any of them at any time when you have any questions, suggestions, concerns, or would like to ask for advice. Please remember, we are here to help you learn. Therefore, please do NOT hesitate to contact us at any time, so we could do our jobs better!

TENTATIVE LECTURE SCHEDULE

(Will be adjusted as needed. Please visit Canvas site for the latest schedule, readings, and assignments)

Week	Date	Topics	Remarks
1	Sep. 3	Course overview	
	Sep. 5	Machine learning basics 1: Loss function and gradient descent	
2	Sep. 10	Machine learning basics 2: Multi-class classification	
	Sep. 12	Machine learning basics 3: Model generalization	
3	Sep. 17	Machine learning basics 4: Regularization	HW1 release
	Sep. 19	Feedforward networks	
4	Sep. 24	Model training and optimization algorithms	
	Sep. 26	Train a network faster and better: Normalization and regularization	
5	Oct. 1	*Public Holiday*	HW1 due
	Oct. 3	Language modeling and recurrent neural networks (RNNs)	
6	Oct. 8	Advanced RNN models: LSTMs and GRUs	
	Oct. 10	Bidirectional and multi-layer RNN models	
7	Oct. 15	Word embeddings	HW2 release
	Oct. 17	Machine translation and sequence-to-sequence models	
8	Oct. 22	Attention-based sequence-to-sequence models	
	Oct. 24	Transformers 1	
9	Oct. 29	Transformers 2	HW2 due
	Oct. 31	Transformer-based speech recognition	
10	Nov. 5	Transformer-based computer vision	
	Nov. 7	Large language models (LLMs): BERT, T5, GPTs, and more	HW3 release
11	Nov. 12	Turn LLMs to intelligent assistants: Prompting engineering	
	Nov. 14	Turn LLMs to intelligent assistants: Instruction fine-tuning	
12	Nov. 19	Parameter-efficient fine-tuning	
	Nov. 21	Retrieval augmented generation for information retrieval	HW3 due
13	Nov. 26	Multi-modal models: Language-vision alignment	
	Nov. 28	Revision	