

Natural Language Understanding with Deep Learning

Course codes: COMP-764 and LING-782

Instructor: [Siva Reddy](#)
Assistant Professor
McGill University and MILA
siva@cs.mcgill.ca

Term: Winter 2020
When: Tuesdays and Thursdays
Time: 4:05 pm - 5:25 pm
Where: [MC 321](#)

Office hours: MC 324 Tuesdays 2:00 pm to 4:00 pm

MyCourses link (class announcements, discussions):
<https://mycourses2.mcgill.ca/d2l/home/434904>

Covid-19 changes:

Rest of the lectures will move to Zoom. For pending student presentations, I will record your one-on-one Zoom presentation with me and release it to the class. I have also updated the grading scheme.

Course Description:

The field of natural language processing (NLP) has seen multiple paradigm shifts over decades, from symbolic AI to statistical methods to deep learning. We review this shift through the lens of natural language understanding (NLU), a branch of NLP that deals with “meaning”. We start with what is meaning and what does it mean for a machine to understand language? We explore how to represent the meaning of words, phrases, sentences and discourse. We then dive into many useful NLU applications.

Throughout the course, we take several concepts in NLU such as meaning or applications such as question answering, and study how the paradigm has shifted, what we gained with each paradigm shift, and what we lost? We will critically evaluate existing ideas and try to come up with new ideas that challenge existing limitations. We will particularly work on making deep learning models for language more robust.

This is a seminar-style course, where the class as a whole will work together in running the course. In the first few lectures, I will provide an overview of NLU and highlight the challenges the field is facing.

Expected Outcomes:

By the end of the course, you should be able to meaningfully contribute to cutting-edge research in natural language understanding.

Grading:

Due to the covid-19 crisis, the university now allows students to choose between satisfactory/unsatisfactory or a grade. The deadline to exercise this option is May 22, 2020. You can also withdraw the course by April 15 2020.

The grading criteria has been updated to accommodate two weeks of time loss:

- **Summaries (15%)** For each class, you are expected to submit a critical summary of the papers that are to be presented in the class, at least an hour before the lecture on MyCourses. You will submit ~~5 such summaries~~ 4 such summaries. [If you haven't finished submitting 4 summaries yet, you can submit summaries for any of the papers from Lecture 19-24 by Apr 2nd.](#)
- **Reviews (10%)** You will review two summaries of others for ~~5 classes~~ 4 classes (same as the one you wrote summaries for). [You can submit all your pending reviews by Apr 7th.](#)
- **Presentation (20%)** You have to present papers along with your project partner on the topic of the lecture (at least once). Not the same topic as you wrote a critical summary for.
- **Project ~~(50%)~~ (55%)** You will do a project in groups of two. This involves
 - Project Proposal (what is the project about): 5%
 - Literature review (10%)
 - ~~○ Baselines/Replication (15%)~~
 - ~~○ Final paper with new experiments and code submission (20%)~~
 - Final report (baseline + novel model): 40%
- ~~● Poster (5%) End of term poster~~
- **Class participation (up to 3%)**: This is a bonus determined at the end of the term. If your percentage is >100, it will be rounded off to 100.

Topics of interest:

1. Word meaning
 - a. distributional semantics
 - b. word embeddings
 - c. evaluation
2. Phrase and sentence meaning
 - a. logical representation
 - b. sentence embeddings
 - c. evaluation
3. Meaning in context
 - a. word senses
 - b. contextual word embeddings

- c. fine-tuning
- 4. Interpretability
 - a. feature-based models vs deep learning models
 - b. linguistic tests
 - c. probing
- 5. Compositionality
 - a. syntax and semantic interfaces
 - b. inductive priors
 - c. tests for compositionality
 - d. limitations
- 6. Reasoning
 - a. inference
 - b. question answering
 - c. other applications
- 7. Discourse
 - a. conversational systems
- 8. Language and physical world
 - a. model-theoretic semantics
 - b. grounded environment
 - c. reinforcement learning
- 9. Bias
 - a. word association tests
 - b. probing

Prerequisites:

You are expected to have done one of the following courses at McGill: natural language processing (COMP/LING 550) or computational linguistics (COMP/LING 445) or applied machine learning (COMP 551). If you have done similar courses at other universities, feel free to take the course. If you are not sure, email me.

Schedule (subject to change):

Lecture	Date	Topic	Presented Papers	Additional Readings
1	Jan 7	Introduction to NLU and applications	Lecture	Computational Linguistics and Deep Learning A case for deep learning in semantics

				Extending Machine Language Models toward Human-Level Language Understanding
2	Jan 9	Course projects/ How to write a research paper/ Mechanical Turk	Lecture	Write the Paper First How to Write Your First Research Paper How to use MTurk for tagging training data
3	Jan 14	Deep Learning Basics (hands on practice moved to next lecture)	Guest Lecture: Sarath Chandar Recurrent Nets and Vanishing Gradient Problem	Deep Learning for NLP with Pytorch Debugging Python Code
4	Jan 16	Word meaning Presenter: Siva Reddy	Mikolov et al. (2013): Distributed Representations of Words and Phrases and their Compositionality Baroni et al. (2014): Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors	Turney and Pantel 2010: From Frequency to Meaning: Vector Space Models of Semantics Pennington et al. (2014): GloVe: Global Vectors for Word Representation
5	Jan 21	Phrase meaning Presenter: Emily Goodwin	Reddy et al. (2011): An Empirical Study on Compositionality in Compound Nouns Baroni and Zamparelli	Mitchell and Lapata (2010): Composition in Distributional Models of Semantics

			(2010): Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space	Erk (2012): Vector Space Models of Word Meaning and Phrase Meaning: A Survey
6	Jan 23	Sentence meaning Presenters: Andrea Jang	Supervised Learning of Universal Sentence Representations from Natural Language Inference Data (This paper is more close to what we have been discussing in the class) Banarescu et al. (2013): Abstract Meaning Representation for Sembanking	Skip-thought vectors Senteval: An evaluation toolkit for universal sentence representations
7	Jan 28	Project brainstorming (optional)	An opportunity to present your idea to the class and refine it before the project proposal deadline	
		Meaning in context Presenter: Siva Reddy	Exemplar-Based Models for Word Meaning In Context Deep contextualized word representations	Dynamic and Static Prototype Vectors for Semantic Composition
8	Jan 30	Fine-tuning/ Transfer learning Presenter: Meng Cao, Faizan Khan	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding Learning and Evaluating General Linguistic Intelligence	To Tune or Not to Tune? Adapting Pretrained Representations to Diverse Tasks ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators
9	Feb 4	Interpretability	Assessing the Ability of	BLiMP: A

		Presenter: Fanny Salvail-Bérard, Suhail Kandanur	LSTMs to Learn Syntax-Sensitive Dependencies What do you learn from context? Probing for sentence structure in contextualized word representations	Benchmark of Linguistic Minimal Pairs for English A Structural Probe for Finding Syntax in Word Representations BERT Rediscovered the Classical NLP Pipeline
10	Feb 6	Interpretability Presenter: Aanika Rahman	Visualizing and Understanding Neural Models in NLP Generalization without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks	Universal Adversarial Triggers for Attacking and Analyzing NLP Linguistic generalization and compositionality in modern artificial neural networks
	Feb 10	Project Proposals + Literature Review due (11:59 pm)		
11	Feb 11	Proposal presentations	5-minute madness	3 minutes for presentation, two minutes for feedback
		Syntax and semantics	Learning to Map Sentences to Logical Form: Structured Classification with Probabilistic Categorical Grammars Parsing Natural Scenes and Natural Language with Recursive Neural Networks	Transforming Dependency Structures to Logical Forms Learning to Compose Neural Networks for Question Answering

12	Feb 13	General-purpose semantic parsing Presenter: Devendra Singh Sachan	Semantics as a Foreign Language Linguistically-Informed Self-Attention for Semantic Role Labeling	Jointly Predicting Predicates and Arguments in Neural Semantic Role Labeling A Discriminative Graph-Based Parser for the Abstract Meaning Representation
13	Feb 18	Other forms of semantic representations Presenters: Zhi Wen, Xing Han Lu	Large-Scale QA-SRL Parsing Ordered Neurons: Integrating Tree Structures into Recurrent Neural Networks	Do latent tree learning models identify meaningful structure in sentences?
14	Feb 20	Relation extraction and knowledge graphs Presenter: Dora Jambor, Paul Wu	Distant supervision for relation extraction without labeled data Zero-Shot Relation Extraction via Reading Comprehension	Graph Convolution over Pruned Dependency Trees Improves Relation Extraction
15	Feb 25	Executable Semantic Parsing Presenters: Zhaocheng Zhu, Jingyi He	Language to Logical Form with Neural Attention Global Reasoning over Database Structures for Text-to-SQL Parsing	Editing-Based SQL Query Generation for Cross-Domain Context-Dependent Questions
16	Feb 27	Multi-turn semantic parsing (closely related to goal-oriented dialogue systems) Presenter: Ryan Kun Ni, Louis Clouatre-Latraverse	Learning to Map Context-Dependent Sentences to Executable Formal Queries CoSQL: A Conversational Text-to-SQL Challenge Towards Cross-Domain Natural Language Interfaces to Databases	Complex Sequential Question Answering: Towards Learning to Converse Over Linked Question Answer Pairs with a Knowledge Graph

	Mar 3	Study Days		
	Mar 5	Study Days		
17	Mar 10	Reading comprehension/ Question answering Presenters: Jules Gagnon-Marchand, Ian Porada	Teaching Machines to Read and Comprehend Neural Module Networks for Reasoning over Text	Reading Wikipedia to Answer Open-Domain Questions Latent Retrieval for Weakly Supervised Open Domain Question Answering
18	Mar 12	Conversational question answering Presenters: Akshatha Arodi Nagaraja, Ashita Diwan	CoQA: A Conversational Question Answering Challenge Can You Unpack That? Learning to Rewrite Questions-in-Context	QuAC: Question Answering in Context FlowQA: Grasping Flow in History for Conversational Machine Comprehension
	Mar 16 Mar 18	Project Baselines + Baseline Results (11:59pm)	Initial results of your own model are highly encouraged.	
19	Mar 17	Chit-chat dialogue Presenters: Silan He, Magesh Sreedhar	A Neural Conversational Model TransferTransfo: A Transfer Learning Approach for Neural Network Based Conversational Agents	Neural Approaches to Conversational AI What makes a good conversation? How controllable attributes affect human judgments
20	Mar 19	Language and vision	Guest Lecture: Harm de Vries, Element AI GuessWhat?! Visual object discovery through	Compositional Attention Networks for Machine Reasoning

			multi-modal dialogue FiLM: Visual Reasoning with a General Conditioning Layer	
21	Mar-24	<p>Language in grounded environments</p> <p>Presenters: Yutong Yan, Irene (Qizhen) Zhang</p>	TOUCHDOWN: Natural Language Navigation and Spatial Reasoning in Visual Street Environments Interactive Language Learning by Question Answering	Embodied Question Answering Habitat: A Platform for Embodied AI Research
22	Mar-26	<p>Systematic Generalization</p>	<p>Guest Lecture: Dzmitry Bahdanau, Element AI</p> <p>Systematic Generalization: What Is Required and Can It Be Learned?</p> <p>CLOSURE: Assessing Systematic Generalization of CLEVR Models</p>	<p>BabyAI: A Platform to Study the Sample Efficiency of Grounded Language Learning</p> <p>CLUTRR: A Diagnostic Benchmark for Inductive Reasoning from Text</p>
<p>You can still submit summaries to the cancelled classes if you haven't submitted four summaries yet.</p> <p>Student presentations of Lec 19 and Lec 21 will be posted online.</p> <p>Lec 23 and Lec 24 will be presented in a live Zoom session.</p>				
23	Mar 31	<p>Bias in language</p> <p>Presenters: Siva Reddy</p>	<p>Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings</p> <p>Measuring Bias in Contextualized Word Representations</p>	<p>Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But do not Remove Them</p>

				Queens are Powerful too: Mitigating Gender Bias in Dialogue Generation
24	Apr 2	Knowledge distillation Presenter: Carolyne Pelletier, Zichao Yan	DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter BAM! Born-Again Multi-Task Networks for Natural Language Understanding	Distilling the Knowledge in a Neural Network
25	Apr 7	Wrap up	Lecture	
26	Apr 11	Poster session	Poster session	
	Apr 14 21	Final papers due (11:59 pm)		

FAQs:

Question: Which book do you recommend for learning basics of linguistics?

Answer: [Bender 2013: Linguistic Fundamentals for Natural Language Processing](#) (You should be able to download this book for free on McGill network)

Question: Any suggestions for books on deep learning for NLP?

Answer:

1. [Goldberg 2017: Neural Network Methods for Natural Language Processing](#) (You should be able to download this book for free on McGill network)
2. [Jurafsky and Martin 2019: Speech and Language Processing](#)
3. [Eisenstein 2019: Introduction to Natural Language Processing](#) (a draft copy can be downloaded [here](#))

Question: Do I have to write a summary for more than 5 classes?

Answer: No, you don't have to. You are required to write a summary for only up to 5 classes.

Question: Are we supposed to write a summary for one paper or two papers for each class?

Answer: Your summary should be up to 500 words. You can summarize either of the papers or both the papers but try to stick to 500 words. I leave it to your discretion to go above 500 words. When you are writing a summary for two papers, it would be great if you can describe the connection between the two papers.

Language of Submission:

In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.

Academic Integrity:

McGill University values academic integrity. Therefore, all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures (see www.mcgill.ca/students/srr/honest/ for more information)

Inclusivity

As the instructor of this course I endeavor to provide an inclusive learning environment. However, if you experience barriers to learning in this course, do not hesitate to discuss them with me and the Office for Students with Disabilities, 514-398-6009.