Linguistics 751: Neural Networks and Natural Language Fall 2017

Tue 12:05–13:25, EDUC 539
Thu 13:05–14:25, WILSON 110

Instructor

Tim O'Donnell
timothy.odonnell@mcgill.ca
1085 Dr. Penfield, 304
Office hours: By Appointment

1 Overview

In recent years, Artificial Neural Networks (ANNs), especially "deep" ANNs, have been applied to a wide variety of problems in natural language processing with sometimes astonishing success. Standard natural language processing problems such as automatic speech recognition and machine translation have seen the state of the art advance by magnitudes that would have been thought impossible just a few years ago. As language scientists, how are we to interpret these results? Do ANNs designed to solve such engineering problems capture human-like linguistic generalization? How can we probe this question rigorously? How and where should ANNs be integrated into the linguist's toolkit? How can linguists contribute to the engineering of ANN-based natural language processing systems and help advance the state of the art?

2 Goals

The goal of this course is to begin laying the foundations on which answers to such questions can be developed and on which linguists and other language scientists can participate in the discussion surrounding such modeling approches.

Students will:

- Gain familiarity with the basic toolset underlying modern ANNs
- Learn about a number of network architectures that have been used for linguistic applications
- Gain practical experience with using ANNs
- Develop skills to critically evaluate work using ANNs from a linguistic perspective.

3 Prerequisites

Programming: programming experience equivalent to at least one college-level course in computer science. You should have some experience programming in Python in particular, or the ability to learn quickly.

Math: Basic experience with probability theory and calculus. Recommended: experience with linear algebra.

4 McGill Policy Statements

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).

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.

Instructor generated course materials (e.g., handouts, notes, summaries, exam questions, etc.) are protected by law and may not be copied or distributed in any form or in any medium without explicit permission of the instructor. Note that infringements of copyright can be subject to follow up by the University under the Code of Student Conduct and Disciplinary Procedures.

5 Logistics

Piazza website: https://piazza.com/mcgill.ca/fall2017/ling751/home

Course GIT Repository: https://bitbucket.org/timo_bitbucket/ling751-fall-2017

Materials: Readings will be made available electronically.

Lunch: Feel free to bring a lunch to class.

Handouts/slides: Slides, ipython notebooks, and other materials will be posted online after each class in the GIT repository and when, appropriate, on piazza.

6 Readings

The course is based primarily on the excellent mongraph of Goldberg (2017).[1] Also useful for the course is the recent textbook of Goodfellow et al. (2016) as well as the course notes by Cho (2015). Other readings will be provided as the term progresses.

7 Course structure

The course consists of two main parts.

Part 1 of the course will give:

- A high-level overview of the essentials of artificial neural networks.
- Some practical experience with using neural network toolkits, tentatively `keras` and `dynet`.

The goals of this part of the course are to provide you with enough conceptual familiarity with ANN concepts, and familiarity with practical implementation, to pick a topic of interest (by reading relevant literature) which your group will focus on in the remainder of the course.

During this part of the course, you will also begin reading the ANN literature and preparing for the mini-course in Part 2. In particular, each student will do a triage review of 6 papers from the literature. These reviews will focus on three kinds of information about the paper:

- What is the problem addressed by the paper (including the specific task and dataset), what are the techniques used, and what is the novel contribution?
- What earlier literature does this paper rely on most crucially?
- What concepts are needed to understand this paper?

Towards the end of Part 1 of the course, each student will pick a single paper (either from their 6 above, or a new one) and develop a mini-course outline for that paper outlining what background reading other students should do to understand the paper, the sections of the proposed mini-course, examples, toy datasets, etc.

In Part 2 of the course, each group (2–5 students) will develop a mini-course on its topic, in two classes, culminating in a polished iPython notebook and any addition materials such as datasets needed to run it, along with background reading needed to understand it. Each group will both lead a mini-course (their own) and 'take' the other groups' mini-courses. This will happen in two sessions after each of which you will receive feedback from the other members of the class as well as myself. After each presentation, you will improve your mini-course improving in ways that include: level of polish, depth, breadth, and/or sophistication of examples. Note that it will be up to you to focus the improvement on the right area based on the feedback you receive from the rest of the class. Your final mini-course should work as an isolated unit.

Between Part 1 and Part 2 we will have several 'perspective' classes, where we step back and discuss broader issues about neural network models and their relationship to human language, such as:

- Are deep learning methods really solving the linguistic problems they purport to solve?
- Do these models give new insights into human language? (This is an orthogonal dimension to better performance on engineering tasks.)

[1] Special thanks to Yoav Goldberg for making various additional course materials available.

8 Evaluation

Note that details below are subject to change.

Lit Review and mini-course prep (30%)

- 6 triage reviews: 18% (3% each)
- One mini-course outline focused on a specific paper 6%
- Two Ipython notebook stub problems 6% (3% each)

Mini-course (50%)

- Two-page proposal/plan/lit review for your mini-course by your group: 10%
- Two class sessions: 15% (7.5% each)
- Final product: 25%

Participation (20%)

- Preparation for class (do any reading, come prepared): 5%
- In-class participation: 5%
- Feedback, commentary, and suggestions on my and other groups presentations: 10%

9 Course schedule (tentative)

Part 1: Introduction to neural networks

- T Sep 5: Class Overview
- R Sep 7: Background and Practicalities
- T Sep 13: NO CLASS
- R Sep 14: ANN Basics
- T Sep 19: Deep Networks
- R Sep 21: Gradient Descent
  - ✎: Triage papers 1 and 2.
- T Sep 26: NO CLASS
- R Sep 28: Word Embeddings
  - ✎: Paper triage 3 and 4
- T Oct 3: RNNs
  - ✎: First code example.
- R Oct 5: Conditioned Generation
  - ✎: Paper triage 5 and 6
- T Oct 10: Neural Parsing
- R Oct 12: RxNNs
  - ✎: Detailed paper mini-course outline
- T Oct 17: Cascaded Architectures/Semi-Supervised Learning
- R Oct 19: Multi-task Learning

Interlude: Perspective

- T Oct 24
  - ✎: Second code example
- R Oct 26
  - ✎: Group proposals
- T Oct 31

Part 2: Mini-courses

- R Nov 2: Group 1
- T Nov 7: Group 2
- R Nov 9: Group 3
- T Nov 14: Group 4
- R Nov 16: Group 5
- T Nov 21: Group 1
- R Nov 23: Group 2
- T Nov 28: Group 3
- R Nov 30: Group 4
- T Dec 4: Group 5

References

Cho, K. (2015). Natural language understanding with distributed representation. arXiv, (arXiv:1511.07916v1 [cs.CL]).

Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing. Morgan & Claypool.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep Learning. MIT Press.