Learning in Humans and Machines

Cognitive Science/Computer Science
3 credits
Spring 2022

Instructor: Qiong Zhang
Course Modality: in-person (if not possible, synchronous remote)
Prerequisites: basic programming in Python
Meeting Times: TBD by the course scheduling system
Office Hours: TBD, after meeting times are determined
Email: qiongz@princeton.edu

Course Description

This interdisciplinary seminar course explores the parallels between human learning and machine learning. The central link between the two is the set of shared computational problems faced by humans and machines which includes making complex decisions; predicting future events; storing and retrieving information efficiently; and generalizing knowledge to new situations. By examining such problems, we will see that

1. solutions drawn on methods developed from machine learning can help us gain insights about human cognition, and conversely,
2. knowledge about how humans solve these problems can inform the development of more intelligent machines.

The first half of the course covers the application of machine learning to explain how human cognition works. We will explore the landscape of computational models of human cognition and discuss the insights these models reveal into how people learn, remember, and make complex decisions in everyday situations. The methods discussed include neural networks, symbolic approaches, Bayesian statistics, information-theoretic approaches, and more. The applications discussed include perception, skill learning, memory, categorization, and decision making.

In the second half of the course we will draw parallels between human learning and machine learning. Specifically, we will explore how our understanding of human cognition can explain and inform advances in machine learning. We will accomplish this by examining recent advances in neural networks and reinforcement learning from a psychologist's perspective.

Each class will start with a short lecture covering the necessary machine learning techniques and cognitive science concepts to understand the readings. Following this is a student presentation of the reading. We will end with an instructor-led discussion around the short lecture and the readings.

Learning Objectives

By the end of the course, students will

1. understand the basics of Bayesian inference, neural networks and other computational approaches,
2. understand the basics of the key aspects of human cognition such as memory and decision making,

3. be able to characterize the relationship between computational approaches to cognition and machine learning research,
4. be able to identify ways in which computational models can be experimentally tested as models of cognition, and
5. acquire skills to test simple models of cognition.

Textbook/Resources

Lecture slides are self-contained. There is no required textbook. There will be a number of cognitive science and computer science papers for discussion, available as PDF files through the class website.

Who should take this course

The course is designed for graduate students in cognitive science, psychology, or computer science who are interested in developing computational models of human cognition and exploring the parallels between human learning and machine learning. Prerequisites are a basic familiarity with programming languages.

Coursework Requirements

Students are expected to actively participate in class discussions, and sign up for at least one paper presentation. There will be a reading assignment for every class, and you are expected to arrive in class with ideas and questions to discuss. To help you develop these ideas, you are required to write short commentaries before classes– one paragraph is typical.
A commentary might take one of the following forms: mention a claim that doesn't seem right to you; describe how the work could be usefully extended; draw a connection between the reading and something else that has been discussed previously. Commentaries are graded pass/fail. If you submit and pass all commentaries, you will receive full credit for this component of the course.

A large component of the course is a team project to assess the student's ability to put together the concepts and tools they have learned in the course. The class project will be an independent research project presenting a simple experiment, testing a new cognitive/machine learning model, or analyzing an existing model. Each team has a total of three students, and ideally have at least one student from cognitive science or psychology major, and one student from computer science major. The team project will be a great opportunity for students to be engaged in multi-disciplinary research and learn new practical skills from other team members.

Grade Evaluation

Commentaries	20%
Paper presentations	20%
Project proposal	10%
Project mid-term report	20%
Project final report	20%
Final presentation	10%

Schedule of Classes and Readings
Week 1
Course introduction
Review of key concepts in human cognition; and the history of cognitive modeling

- *Marr, D. (1982). Vision. San Francisco: W. H. Freeman. Chapter 1.*

Week 2
Approaches to modeling cognition I: neural network models
Overview of neural networks basics

- *McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T.T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to understanding cognition. Trends in Cognitive Sciences, 14, 348-356.*
- *Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, & J. L. McClelland (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations (pp. 318- 362). Cambridge, MA: MIT Press.*

Week 3
Approaches to modeling cognition II: cognitive architectures

- *Anderson, J. R. (1996). ACT: A simple theory of complex cognition. American Psychologist, 51, 355-365.*
- *Newell, A., Rosenbloom, P. S., & Laird, J. E. (1989). Symbolic architectures for cognition. In M. I. Posner (ed.), Foundations of cognitive science, 93-131. Cambridge, MA: MIT Press.*

Week 4
Principle of rationality

- *Anderson, J. R. (1990). The adaptive character of thought. Hillsdale, NJ: Erlbaum. Chapter 1.*
- *Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment in memory. Psychological Science, 2, 396-408.*

Week 5
Approaches to modeling cognition III: Bayesian inference
Brief overview of Bayesian statistics

- *Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. Trends in Cognitive Sciences, 14, 357-364.*
- *Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. Psychological Science, 17, 767-773.*

Week 6
Approaches to modeling cognition III: Bayesian inference
Brief overview of concepts in human learning, memory and decision making

- *Huttenlocher, J., Hedges, L.V., & Vevea, J.L. (2000). Why do categories affect stimulus judgment? Journal of Experimental Psychology, General, 129, 220-241*
- *Feldman, N. H., & Griffiths, T. L. (2007). A rational account of the perceptual magnet effect. Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society.*

Week 7
Approaches to modeling cognition IV: an information-theoretic view
Brief overview of concepts in information theory

- *Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. Psychological review, 119(4), 807.*
- *Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. Nature, 407(6804), 630-633.*

Mid-semester break

Week 8
Learning in humans and machines

- *Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. Behavioral and brain sciences, 40.*
- *Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. nature, 529(7587), 484-489.*

Week 9
Meta-learning

- *Tomov, M. S., Schulz, E., & Gershman, S. J. (2021). Multi-task reinforcement learning in humans. Nature Human Behaviour, 1-10.*
- *Wang, J. X., Kurth-Nelson, Z., Tirumala, D., Soyer, H., Leibo, J. Z., Munos, R., ... & Botvinick, M. (2016). Learning to reinforcement learn. arXiv preprint arXiv:1611.05763.*

Week 10
Brain-inspired learning

- *O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary learning systems. Cognitive science, 38(6), 1229-1248.*
- *van de Ven, G. M., Siegelmann, H. T., & Tolias, A. S. (2020). Brain-inspired replay for continual learning with artificial neural networks. Nature communications, 11(1), 1-14.*

Week 11
Context

- *Manning, J. R., Norman, K. A., and Kahana, M. J. (2015). The role of context in episodic memory. In M. Gazzaniga (Ed.), The cognitive neurosciences, fifth edition. MIT Press.*
- *Ren, M., Iuzzolino, M. L., Mozer, M. C., & Zemel, R. S. (2020). Wandering within a world: Online contextualized few-shot learning. arXiv preprint arXiv:2007.04546.*

Week 12
Hierarchical Memory

Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive psychology, 4(1), 55-81.
Ignasi Sols, Sarah DuBrow, Lila Davachi, and LluiÃÅs Fuentemilla. Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in long-term memory. Current Biology, 27(22):3499-3504, 2017.
Lampinen, A. K., Chan, S. C., Banino, A., & Hill, F. (2021). Towards mental time travel: a hierarchical memory for reinforcement learning agents. arXiv preprint arXiv:2105.14039.

- *Andrew G Barto and Sridhar Mahadevan. Recent advances in hierarchical reinforcement learning. Discrete Event Dynamic Systems, 13(4):341–379, 2003.*

Week 13
Curiosity

- *Dubey, R., & Griffiths, T. L. (2020). Understanding exploration in humans and machines by formalizing the function of curiosity. Current Opinion in Behavioral Sciences, 35, 118-124.*

- *D. Pathak, P. Agrawal, A. A. Efros, T. Darrell, Curiosity-driven exploration by self-supervised prediction, in: International Conference on Machine Learning (ICML), volume 2017, 2017.*

Week 14
Final project presentations

Academic Integrity Policies

Rutgers University regards acts of dishonesty (e.g. plagiarism, cheating on examinations, obtaining unfair advantage, and falsification of records and official documents) as serious offenses against the values of intellectual honesty. These policies are detailed here:

https://nbprovost.rutgers.edu/academic-integrity-students

In addition, the Computer Science departments has established policies for academic integrity that pertain specifically to programming assignments:

https://www.cs.rutgers.edu/academics/undergraduate/academic-integrity-policy/programming-assignments