Learning in Humans and Machines

Cognitive Science 185:601 session 02
Computer Science 198:598 session 01
3 credits
Spring 2022

Instructor: Qiong Zhang
Course Modality: in-person (subject to university policy regarding the temporary use of remote instruction)
Prerequisites: basic familiarity with programming languages
Meeting Times: M 1210 PM – 0130; H 1210 PM - 0130
Venue: HLL 009
Email: qiong.z@rutgers.edu
(Please add 598 in the email title, e.g. [598 project] [598 attendance] [598 presentation])

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, 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 neuroscience and 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 a discussion around the reading.

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, and
4. be able to identify ways in which computational models can be experimentally tested as 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, computer science, or engineering 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 or several of the following forms: questions you have that you would like to discuss further in class; describe the part of the reading that you find most interesting or surprising; 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

Attendance	10%
Commentaries (due midnight prior to each class)	20%
Paper presentations	20%
Project proposal (due Feb 6)	10%
Project mid-term report (due March 21)	10%
Project final report (due May 8)	20%
Final presentation	10%

Schedule of Classes and Readings

Week 1

Course Overview (Jan 20)

Review of key concepts in human cognition; history of cognitive modeling; human intelligence and machine intelligence

Week 2

Marr's three levels of analysis (Jan 24)

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

Class project briefing (Jan 27)

Overview of class projects and datasets

Week 3

Rational analysis (Jan 31)

- *Schooler, L. J., & Anderson, J. R. (2017). The Adaptive Nature of Memory. In J. H. Byrne (Ed.) Learning and Memory: A Comprehensive Reference, 2nd edition. Amsterdam, Elsevier. (Originally: Anderson, J. R. (1990). The adaptive character of thought. Hillsdale, NJ: Erlbaum. Chapter 1.*)

Rational analysis (Feb 3)

- *Griffiths, T. L., Steyvers, M., & Firl, A. (2007). Google and the mind: Predicting fluency with PageRank. Psychological science, 18(12), 1069-1076.*

Week 4

Probabilistic models of cognition: Concept learning (Feb 7)

Bayesian inference with a discrete space of hypotheses

- *Tenenbaum, J. B. (2000). Rules and similarity in concept learning. Advances in neural information processing systems, 12, 59-65.*

Probabilistic models of cognition: Memory (Feb 10)

Bayesian inference with a continuous space of hypotheses

- *Huttenlocher, J., Hedges, L.V., & Vevea, J.L. (2000). Why do categories affect stimulus judgment? Journal of Experimental Psychology, General, 129, 220-241*

Week 5

Probabilistic models of cognition: Hindsight bias (Feb 14)

Mixture models

- *Wilson, S. A., Arora, S., Zhang, Q., & Griffiths, T. (2021). A rational account of anchor effects in hindsight bias. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 43, No. 43).*

Probabilistic models of cognition: Anchoring bias (Feb 17)

Resource-rational analysis

- *Lieder, F., Griffiths, T. L., & Goodman, N. D. (2012, December). "Burn-in, bias, and the rationality of anchoring". In NIPS (pp. 2699-2707).*

Week 6

Mechanistic models of cognition (Feb 21)

Human decision making

- *Modelling response times for two-choice decisions. Pyschological Science, 9, 347–356.*

Mechanistic models of cognition (Feb 24)

Human memory search

- *Foundations of human memory. Kahana, M. J. (2012) New York, NY, US: Oxford University Press Chapter 7.*

Week 7

Cognitive architectures (Feb 28)

- *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.*

Cognitive architectures (Mar 3)

- *Gunzelmann, G., & Anderson, J. R. (2003). Problem solving: Increased planning with practice. Cognitive systems research, 4(1), 57-76.*

Week 8

Neural network models of cognition (Mar 7)

- *Hinton, G. E., Plaut, D. C., & Shallice, T. (1993). Simulating brain damage. Scientific American, 269(4), 76-82.*

Neural network models of cognition (Mar 10)

- *Lu, Q., Hasson, U., & Norman, K. A. (2021). When to retrieve and encode episodic memories: a neural network model*

Mid-semester break

Week 9

Human-machine comparison (Mar 21)

- *Elsayed, G. F., Shankar, S., Cheung, B., Papernot, N., Kurakin, A., Goodfellow, I., & Sohl-Dickstein, J. (2018). Adversarial examples that fool both computer vision and time-limited humans. arXiv preprint arXiv:1802.08195.*

Human-machine comparison (March 24)

- *Dapello, J., Marques, T., Schrimpf, M., Geiger, F., Cox, D. D., & DiCarlo, J. J. (2020). Simulating a primary visual cortex at the front of CNNs improves robustness to image perturbations. BioRxiv.*

Week 10

Inductive bias (Mar 28)

- *K. L. Hermann, T. Chen, S. Kornblith, The origins and prevalence of texture bias in convolutional neural networks. arXiv:1911.09071 (29 June 2020).*

Inductive bias (Mar 31)

- *Lake, B. M., Linzen, T., and Baroni, M. (2019). Human few-shot learning of compositional instructions. In Proceedings of the 41st Annual Conference of the Cognitive Science Society*

Week 11
Brain-like learning: Contrastive learning (Arp 4)

- *Konkle, T., & Alvarez, G. A. (2020). Instance-level contrastive learning yields human brain-like representation without category-supervision. bioRxiv.*

Brain-like learning: Replay (Arp 7)

- *Roscow, E. L., Chua, R., Costa, R. P., Jones, M. W., & Lepora, N. (2021). Learning offline: memory replay in biological and artificial reinforcement learning. Trends in neurosciences, 44(10), 808-821.*

Week 12
Curiosity-driven exploration (Apr 11)

- *Dubey, R., & Griffiths, T. L. (2020). Reconciling novelty and complexity through a rational analysis of curiosity. Psychological Review, 127(3), 455.*

Curiosity-driven exploration (Apr 14)

- *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 13
Contextual memory (Apr 18)

- *Jacques, B., Tiganj, Z., Howard, M. W., & Sederberg, P. B. (2021). 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.*

Hierarchical memory (Apr 21)

- *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*

Week 14
Final project presentations (Apr 25, Apr 28)

Class policies

If you need to attend an in-person class remotely or cannot make it to a class, please email the instructor *before* the class to avoid penalty on the attendance points.

Grades

Final grades will be calculated according to these guidelines:
A = 89.5-100
B+ = 84.5-89.49
B = 79.5-84.49
C+ = 74.5-79.49
C = 69.5-74.49
D = 59.5-69.49
F = 0-59.49

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