Learning in Humans and Machines

Cognitive Science 185:601 ~~session 02~~ Section 01
Computer Science 198:598 session 01
3 credits
Fall 2023

Instructor: Qiong Zhang
Meeting Times:
Office hour:
Venue:
Email: qiong.z@rutgers.edu

Course Description

This interdisciplinary graduate-level 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, or computer science who are interested in developing computational models of human cognition and exploring the parallels between human learning and machine learning.

Coursework Requirements

Students are expected to actively participate in class discussions and sign up for at least one paper ***presentation*** (20% of total grade).

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 to two paragraphs is typical (20% of total grade). 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.

Students are expected to attend all classes and take notes on the most basic and important concepts discussed in each class. There will be ten in-class ***quizzes*** distributed randomly across the semester (20% of total grade). They are conducted at the end of a class and consist of short true and false questions which serve as attendance and attention check for that class.

Another 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 (40% of total grade), delivered by a mid-term report, a final report, and a final presentation. The class project will be an independent research project analyzing an experiment, testing a new cognitive/machine learning model, or analyzing an existing model. The team project will be an excellent opportunity for students to be engaged in multi-disciplinary research and learn new practical skills from other team members.

Grade Evaluation

Commentaries (due midnight prior to each class)	20%
Paper presentations	20%
In-class quizzes	20%
Project mid-term report	10%
Project final report	15%
Final presentation	15%

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

- *Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. Psychological review, 115(4), 893.*

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)

Parallel Distributed Processing

- *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)

Complementary Learning Systems

- *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)

- *Barto, A. G., Singh, S., & Chentanez, N. (2004, October). Intrinsically motivated learning of hierarchical collections of skills. In Proceedings of the 3rd International Conference on Development and Learning (pp. 112-19).*

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)

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. Violations of academic integrity will be treated in accordance with university policy, and sanctions for violations may range from no credit for the assignment, to a failing course grade to (for the most severe violations) dismissal from the university. Details policies can be found here: http://academicintegrity.rutgers.edu

These principles forbid plagiarism and require that every Rutgers University student:
• properly acknowledge and cite all use of the ideas, results, or words of others
• properly acknowledge all contributors to a given piece of work
• make sure that all work submitted as his or her own in a course or other academic activity is produced without the aid of unsanctioned materials or unsanctioned collaboration
• treat all other students in an ethical manner, respecting their integrity and right to pursue their educational goals without interference. This requires that a student neither facilitate academic dishonesty by others nor obstruct their academic progress (reproduced from: http://academicintegrity.rutgers.edu/academic-integrity-at-rutgers/).

Students with Disabilities

Our community values diversity and seeks to promote meaningful access to educational opportunities for all students. If you believe that you need accommodations for a disability, please follow these procedures outlined at http://disabilityservices.rutgers.edu/request.html
Since accommodations may require early planning and are not provided retroactively, please initiate this process as soon as possible.

Rutgers CS Diversity and Inclusion Statement

Rutgers Computer Science Department is committed to creating a consciously anti-racist, inclusive community that welcomes diversity in various dimensions (e.g., race, national origin, gender, sexuality, disability status, class, or religious beliefs). We will not tolerate micro-aggressions and discrimination that creates a hostile atmosphere in the class and/or threatens the well-being of our students. We will continuously strive to create a safe learning environment that allows for the open exchange of ideas while also ensuring equitable opportunities and respect for all of us. Our goal is to maintain an environment where students, staff, and faculty can contribute without the fear of ridicule or intolerant or offensive language. If you witness or experience racism, discrimination micro-aggressions, or other offensive behavior, you are encouraged to bring it to the attention to the undergraduate program director, the graduate program director, or the department chair. You can also report it to the Bias Incident Reporting System http://inclusion.rutgers.edu/report-bias-incident/