Parallels between Minds and Machines

Cognitive Science 16:185:605
Sample Syllabus 3 credits

Instructor: Qiong Zhang
Instructor Email:

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: 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. The one-paragraph commentary should be ended with a suggestion on what would be a good discussion question to have in class. 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 a number of in-class ***quizzes*** distributed randomly across the semester. They consist of short true and false questions which serve as attendance and attention check for that class.

Another component of the course is an individual/team ***project*** to assess the student's ability to put together the concepts and tools they have learned in the course (50% 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 project will be an excellent opportunity for students to be engaged in multi-disciplinary research.

Grade Evaluation

Commentaries (due midnight prior to each class)	20%
Paper presentations	20%
In-class quizzes/attendance	10%
Project mid-term report	20%
Project final report	30%

Schedule of Classes and Readings

Week 1
Course Overview
Review of key concepts in human cognition; history of cognitive modeling; human intelligence and machine intelligence

Week 2
Marr's three levels of analysis
- *Marr, D. (1982). Vision. San Francisco: W. H. Freeman. Chapter 1.*

Class project briefing

Overview of class projects and datasets

Week 3

Formal theories of cognition

- *van Rooij, I., & Blokpoel, M. (2020). Formalizing verbal theories: A tutorial by dialogue. Social Psychology, 51(5), 285–298. https://doi.org/10.1027/1864-9335/a000428*

Rational analysis

- *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.*)

Week 4

Rational analysis

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

Probabilistic models of cognition: Concept learning

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.*

Week 5

Probabilistic models of cognition: Memory

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*

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

Resource-rational analysis

Callaway, F., Norman, K., Griffiths, T., Zhang, Q.(2023). Optimal Metacognitive Control of Memory Recall. Psychological Review.

Active learning in humans

- *Markant, D. B., Settles, B., & Gureckis, T. M. (2016). Self-directed learning favors local, rather than global, uncertainty. Cognitive science, 40(1), 100-120.*

Week 7

Mechanistic models of cognition

Human decision making

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

Mechanistic models of cognition

Group cognition

- *Luhmann, C. C., & Rajaram, S. (2015). Memory transmission in small groups and large networks: An agent-based model. Psychological Science, 26(12), 1909-1917.*

Week 8

Neural network models of cognition

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

Complementary Learning Systems

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

Week 9

Human-machine comparison

- *Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Scientific reports, 6(1), 27755.*

Human-machine comparison

- *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. Advances in Neural Information Processing Systems, 33, 13073-13087.*

Week 10

Inductive bias

- *Hermann, K., Chen, T., & Kornblith, S. (2020). The origins and prevalence of texture bias in convolutional neural networks. Advances in Neural Information Processing Systems, 33, 19000-19015.*

Inductive bias

- *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

Contrastive learning

- *Konkle, T., & Alvarez, G. A. (2022). A self-supervised domain-general learning framework for human ventral stream representation. Nature communications, 13(1), 491.*

Nov 15 class canceled due to conference traveling.

Week 12

Brain-inspired Replay

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), 4069.

Curiosity-driven exploration

- *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

- *Ren, M., Iuzzolino, M. L., Mozer, M. C., & Zemel, R. S. (2021). Wandering within a world: Online contextualized few-shot learning. ICLR.*

Hierarchical memory

- *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

Event understanding in large language models

- *Michelmann, S., Kumar, M., Norman, K. A., & Toneva, M. (2023). Large language models can segment narrative events similarly to humans. arXiv preprint arXiv:2301.10297.*

Reasoning in large language models

- *Sap, M., LeBras, R., Fried, D., & Choi, Y. (2022). Neural theory-of-mind? on the limits of social intelligence in large lms. arXiv preprint arXiv:2210.13312.*

Week 15
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. 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/