Georgia Tech College of Computing
Computational Science and Engineering

CSE 6243

Advanced Topics in Machine Learning

Bo Dai
School of CSE, Georgia Tech

Course Introduction

Instructor

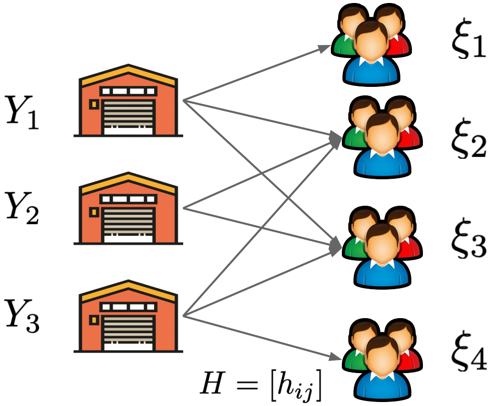
Bo Dai

Assistant Professor
School of Computational Science and Engineering

Address: Coda E1342A
Email: bodai@cc.gatech.edu
Homepage: https://bo-dai.github.io

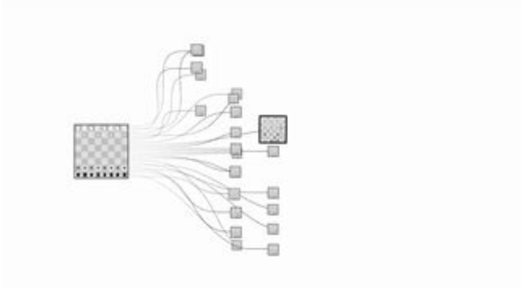
Teaching:
CSE6243 Advanced Machine Learning https://bo-dai.github.io/CSE6243-fall2023/

Decision AI

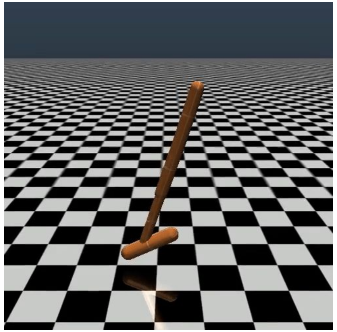


Order Fulfillment under Uncertainty

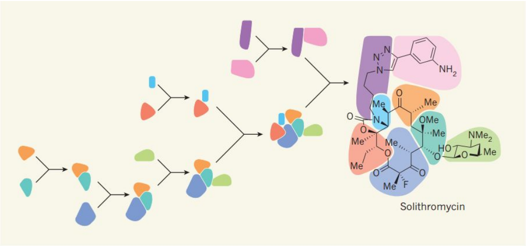
Stochastic Combinatorial Optimization



Searching/Decoding



Sequential Decision Making



Practical decision-making and planning algorithms with computational and statistical efficiency

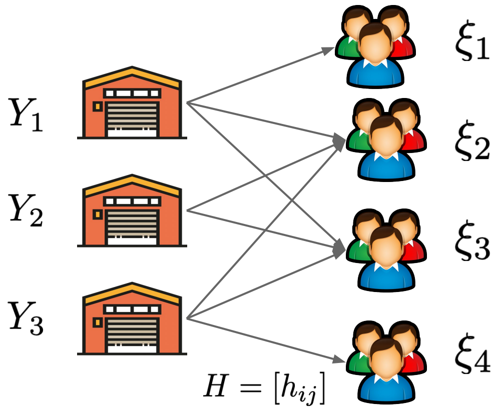
Decision AI

Learning-based methods to combat

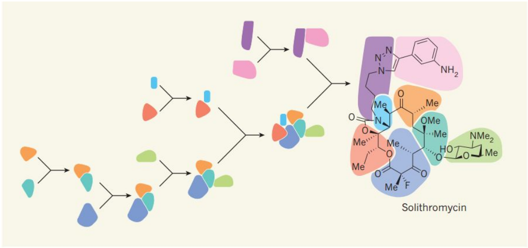
Combinatorial and Stochastic complexity

in decision-making process, focusing on

better decisions in Social and Physical science.



Reinforcement Learning, Generative Models, Optimization and Sampling



Teaching Assistant

Dmitry Shribak

Address: Coda East Wing
Email: shribak@gatech.edu

Logistics

Time: Monday/Wednesday 5:00-06:15 pm

Discussion & HW submission: Ed Discussion & Canvas

Office Hour:

- Instructor: Monday, 4:00 - 5:00PM, CODA E1342A
- TA: Wednesday, 4:00 - 5:00PM, TBD, CODA C1306

Prerequisite

- Graduate-level Machine Learning
  - Deep neural networks
  - Graphical models
  - Kernel methods…
- Probability and Statistics
  - Random variable, moment generating function
  - Bootstrap, delta method
  - MCMC sampling ...
- Numerical Linear Algebra & Optimization
  - Eigen decomposition, Singular value decomposition
  - Gram-Schmidt process
  - Convex function, duality…

Outline

- Module I: Background Knowledge
- Module II: Generative Model
- Module III: Differentiable Programming
- Module IV: Reinforcement Learning

Outline

- Module I: Background Knowledge
- Module II: Generative Model
- Module III: Differentiable Programming
- Module IV: Reinforcement Learning

Guest lectures from prestigious researchers

Advanced Sampling, Reinforcement Learning, and Foundation Models

Module I: Background Knowledge

- Convex Optimization
  - Convex function
  - Duality
  - Stochastic gradient descent
- Probabilistic graphical model
  - Directed graphical models (Bayes Nets)
  - Undirected graphical models (Markov Random Fields)
- Sampling
  - Metropolis–Hastings algorithm
  - Gibbs sampling
  - Hamiltonian Monte-Carlo
- Revisit Neural Network

Module II: Generative ~~Model~~

- Variational auto-encoder
- Autoregressive model
- Generative adversarial net
- Energy-based model
- Diffusion models

Module II: Generative ~~Model~~ AI

- Variational auto-encoder
- Autoregressive model : [ChatGPT,](ChatGPT) Bard, Claude…..
- Generative adversarial net
- Energy-based model
- Diffusion models : Midjourney, Stability AI, Imagen….

Module III: Differentiable Programming

- Differentiable bilevel optimization as NNs
- Differentiable sampler as NNs
- Differentiable algorithms as NNs

Module IV: Reinforcement Learning

- Markov decision process
- Approximate dynamic programming
- Policy gradient
- Linear Programing-based planning
- (Offline RL, Exploration)...

Textbooks

- Boyd & Vandenberghe. Convex Optimization. Cambridge University Press. 2003
- Bishop. Pattern Recognition and Machine Learning. Springer. 2006
- Mohri, Rostamizadeh, & Talwalkar. Foundations of Machine Learning. MIT Press. 2018
- Amos. Tutorial on amortized optimization.
- Putman. Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons, Inc. 1994

The material of the class may go beyond these books

Grading

- Participation (20%)
- Scribe Duties (40%)
- Final Project (40%)

Grading

- Participation (20%)
  - In-Class quiz 10%
  - Completing mid-semester evaluation 4%.
  - Machine Learning seminar 6%

This is an in person class, no zoom link, except the guest lectures.

Grading

- Scribe Duties (40%)
    - 2 students as a group
    - 24-26 lectures scribing with template
    - Submitted in 1 week on Canvas
    - Scribing slots
      https://docs.google.com/spreadsheets/d/1TTlvcGLdT3sJgRes-8mI1eD_k9EdnSPVISIP59FGVvU/edit?usp=sharing

Grading

- Final Project (40%)
  - 3-4 students as a group
  - Proposal : 2 pages excluding references (10%)
  - Midway Report : 3 pages excluding references (20%)
  - Presentation : oral presentation (20%)
  - Final Report : 5 pages excluding references (50%)
  - All write-ups should use the NeurIPS style

More details: https://bo-dai.github.io/CSE6243-fall2023/project/

What is Machine Learning?

Machine learning (**ML**) is an umbrella term for solving problems for which development of algorithms by human programmers would be cost-prohibitive, and instead the problems are solved by helping machines 'discover' their 'own' algorithms,[1] without needing to be explicitly told what to do by any human-developed algorithms.

-- Wikipedia

Machine learning is a subfield of artificial intelligence (AI) that focuses on developing algorithms and models that enable computers to learn from and make predictions or decisions based on data, without being explicitly programmed. The goal of machine learning is to enable computers to improve their performance on a task by learning from experience.

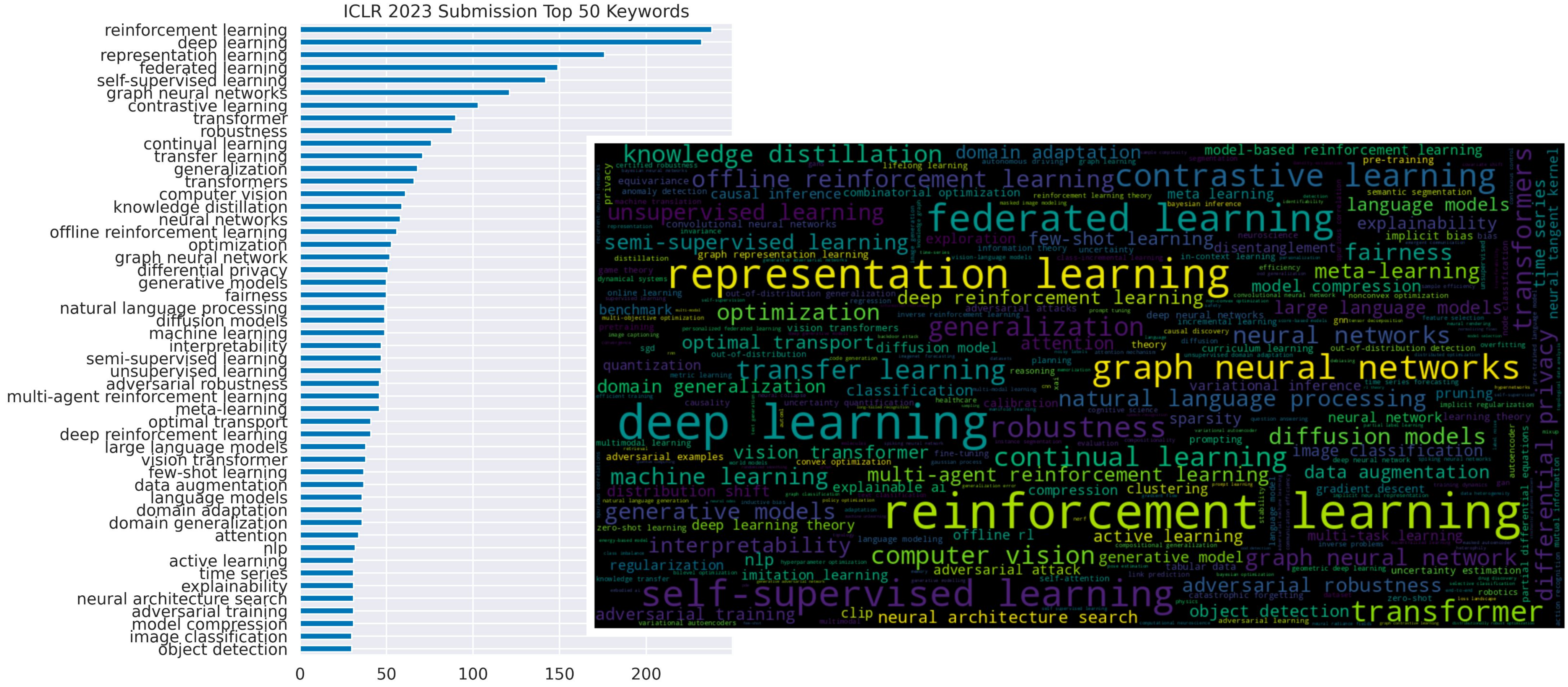
-- ChatGPT

Machine learning is a type of artificial intelligence (AI) that allows software applications to become more accurate in predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

-- Google Bard

Personal Opinion

- Machine Learning is a subfield of AI
- Machine Learning focuses on a special type of algorithm design
    - These algorithms consume data, generates a model for prediction and decision



ICLR 2023 Submission Top 50 Keywords
reinforcement learning
deep learning
representation learning
federated learning
self-supervised learning
graph neural networks
contrastive learning
transformer
robustness
continual learning
transfer learning
generalization
transformers
computer vision
knowledge distillation
neural networks
offline reinforcement learning
optimization
graph neural network
differential privacy
generative models
fairness
natural language processing
diffusion models
machine learning
interpretability
semi-supervised learning
unsupervised learning
adversarial robustness
multi-agent reinforcement learning
meta-learning
optimal transport
deep reinforcement learning
large language models
vision transformer
few-shot learning
data augmentation
language models
domain adaptation
domain generalization
attention
nlp
active learning
time series
explainability
neural architecture search
adversarial training
model compression
image classification
object detection
0
50
100
150
200

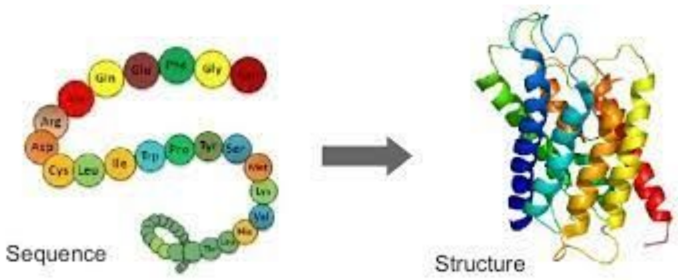
Machine Learning Paradigms

Supervised Learning $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n \in \mathcal{X} \times \mathcal{Y}$ $\mathrm{Alg}(\mathcal{D}) \Rightarrow f(\cdot) : \mathcal{X} \to \mathcal{Y}$

Unsupervised Learning $\mathcal{D} = \{x_i\}_{i=1}^n \in \mathcal{X}$ $\mathrm{Alg}(\mathcal{D}) \Rightarrow f(\cdot) : \mathcal{X} \to \mathcal{Z}$

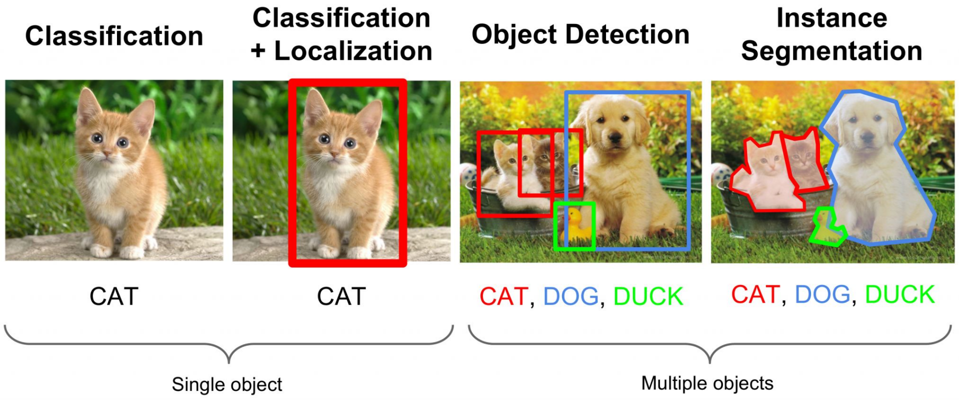
Reinforcement Learning

Machine Learning Paradigms



Supervised Learning $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n} \in \mathcal{X} \times \mathcal{Y}$ $\mathrm{Alg}(\mathcal{D}) \Rightarrow f(\cdot) : \mathcal{X} \to \mathcal{Y}$

Unsupervised Learning $\mathcal{D} = \{x_i\}_{i=1}^{n} \in \mathcal{X}$ $\mathrm{Alg}(\mathcal{D}) \Rightarrow f(\cdot) : \mathcal{X} \to \mathcal{Z}$

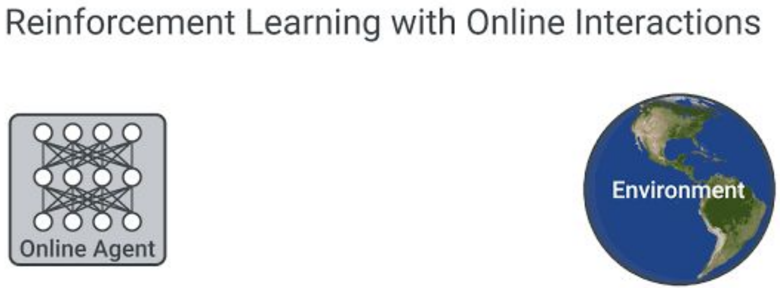


Machine Learning Paradigms

Passive Learning

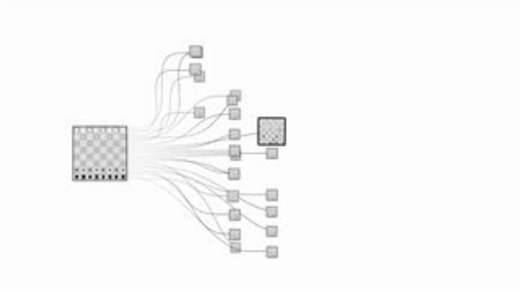
- Supervised Learning $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n} \in \mathcal{X} \times \mathcal{Y}$ $\quad \mathrm{Alg}(\mathcal{D}) \Rightarrow f(\cdot) : \mathcal{X} \to \mathcal{Y}$
- Unsupervised Learning $\mathcal{D} = \{x_i\}_{i=1}^{n} \in \mathcal{X}$ $\quad \mathrm{Alg}(\mathcal{D}) \Rightarrow f(\cdot) : \mathcal{X} \to \mathcal{Z}$

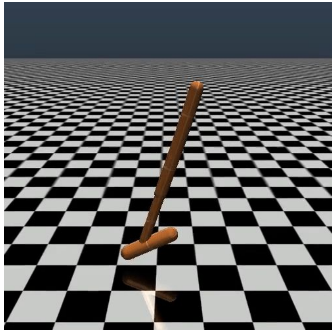
Reinforcement Learning



$$\mathrm{Alg}(\mathrm{Env}) \Rightarrow (\mathcal{D} = \{(s_i, a_i, r_i, s_i')\}_{i=1}^{T},\ \pi(\cdot|s) : \mathcal{S} \to \Delta(\mathcal{A}))$$

Machine Learning Paradigms





Reinforcement Learning

Reinforcement Learning with Online Interactions

$$\mathrm{Alg}(\mathrm{Env}) \Rightarrow (\mathcal{D} = \{(s_i, a_i, r_i, s_i')\}_{i=1}^{T},\ \pi(\cdot|s) : \mathcal{S} \to \Delta(\mathcal{A}))$$

Machine Learning Paradigms

Supervised Learning

Unsupervised Learning

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n} \in \mathcal{X} \times \mathcal{Y} \quad \mathcal{D} = \{x_i\}_{i=1}^{n} \in \mathcal{X}$$

Semi Supervised Learning

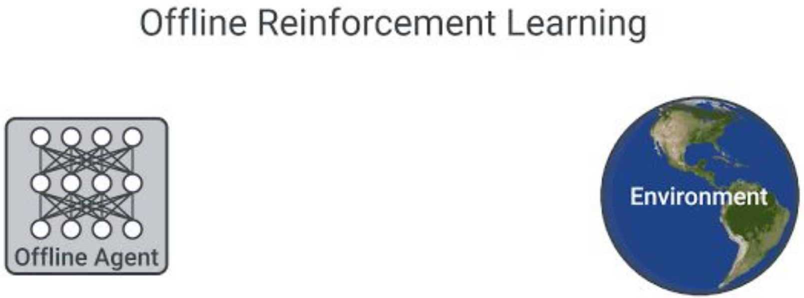
Reinforcement Learning

Machine Learning Paradigms

Supervised Learning

Unsupervised Learning

Reinforcement Learning



$$\mathcal{D} = \{(s_i, a_i, r_i, s_i')\}_{i=1}^{T} \qquad \mathrm{Alg}(\mathcal{D}) \Rightarrow \pi(\cdot|s) : \mathcal{S} \to \Delta(\mathcal{A})$$

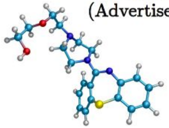
Machine Learning Paradigms

Graph Learning

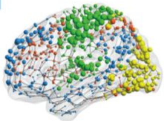
Passive Learning
- Supervised Learning
- Unsupervised Learning

Reinforcement Learning

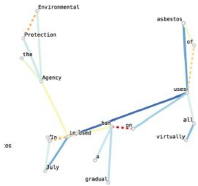
Social networks (Advertisement)



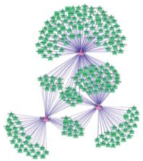
Drug/Material molecules (Chemistry)



Brain connectivity (Neuroscience)

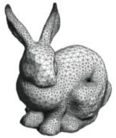


Words relationships (NLP)



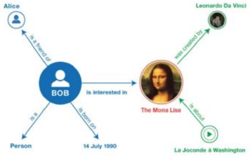
Gene Regulatory Network

Recommender systems (Amazon, Netflix)

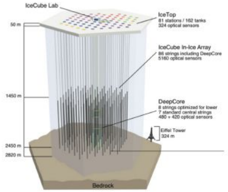


3D Meshes (Computer Graphics)

Transportation networks

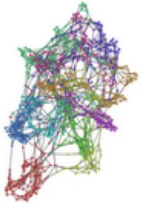


Knowledge graph (Causality)

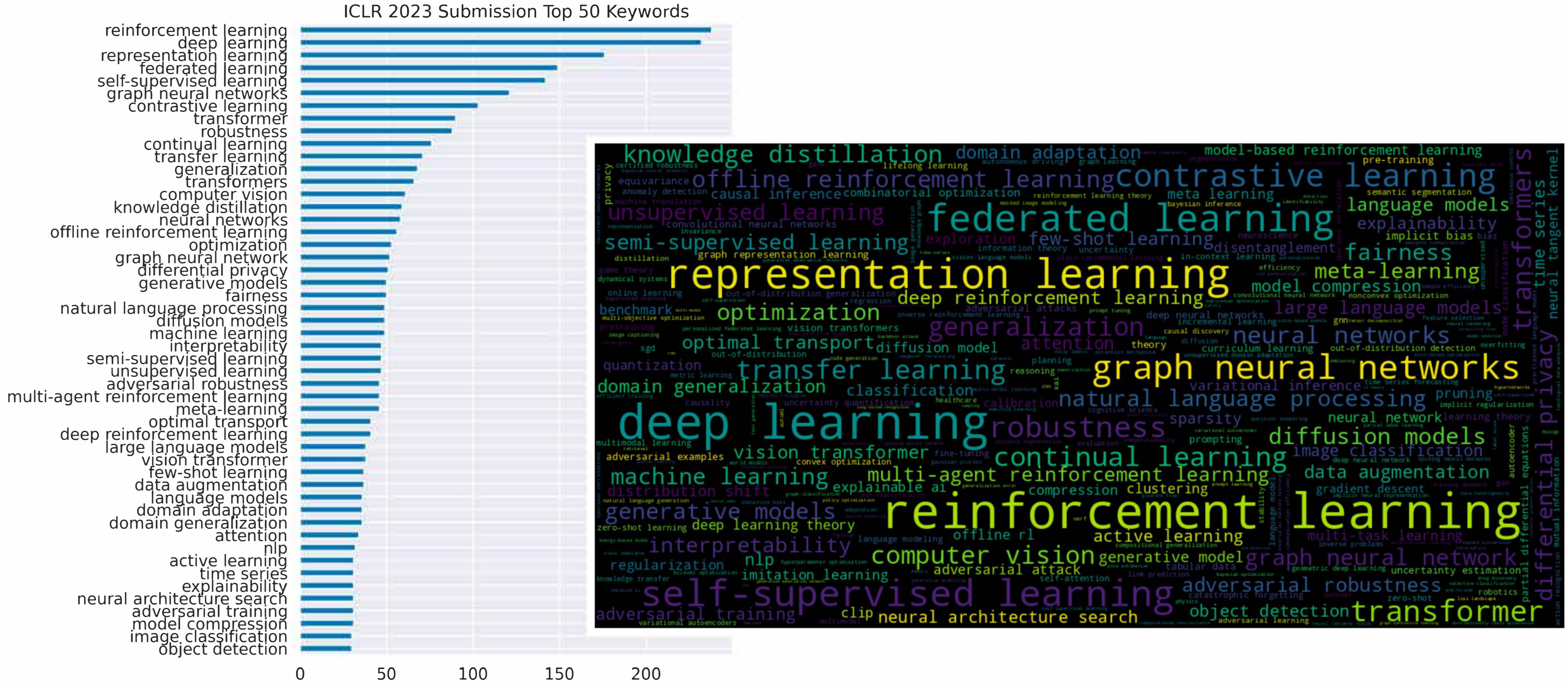


Neutrino detection (High-energy Physics)

=



Graphs/ Networks


ICLR 2023 Submission Top 50 Keywords
reinforcement learning
deep learning
representation learning
federated learning
self-supervised learning
graph neural networks
contrastive learning
transformer
robustness
continual learning
transfer learning
generalization
transformers
computer vision
knowledge distillation
neural networks
offline reinforcement learning
optimization
graph neural network
differential privacy
generative models
fairness
natural language processing
diffusion models
machine learning
interpretability
semi-supervised learning
unsupervised learning
adversarial robustness
multi-agent reinforcement learning
meta-learning
optimal transport
deep reinforcement learning
large language models
vision transformer
few-shot learning
data augmentation
language models
domain adaptation
domain generalization
attention
nlp
active learning
time series
explainability
neural architecture search
adversarial training
model compression
image classification
object detection
0
50
100
150
200

Machine Learning Paradigms

Module I
Basic
Knowledge

Module II
Generative
Models

Module III
Differentiable
Programming

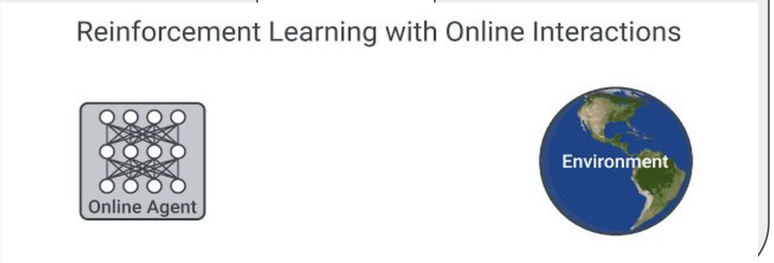
Supervised Learning

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n} \in \mathcal{X} \times \mathcal{Y} \quad \text{Alg}(\mathcal{D}) \Rightarrow f(\cdot) : \mathcal{X} \rightarrow \mathcal{Y}$$

Unsupervised Learning

$$\mathcal{D} = \{x_i\}_{i=1}^{n} \in \mathcal{X} \quad \text{Alg}(\mathcal{D}) \Rightarrow f(\cdot) : \mathcal{X} \rightarrow \mathcal{Z}$$

Reinforcement Learning



$$\text{Alg}(\text{Env}) \Rightarrow (\mathcal{D} = \{(s_i, a_i, r_i, s_i')\}_{i=1}^{T},\ \pi(\cdot|s) : \mathcal{S} \rightarrow \Delta(\mathcal{A}))$$

Machine Learning Paradigms

Module I
Basic
Knowledge

Module II
Generative
Models

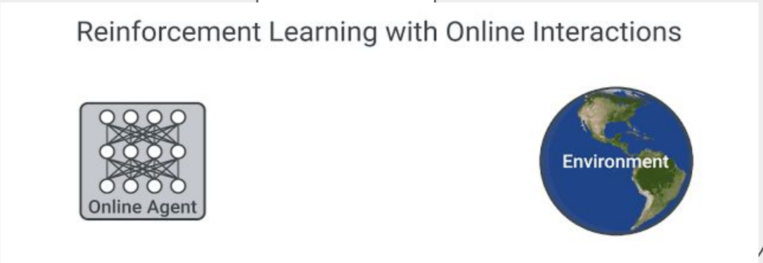
Module III
Differentiable
Programming

Supervised Learning

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n \in \mathcal{X} \times \mathcal{Y} \qquad \mathrm{Alg}(\mathcal{D}) \Rightarrow f(\cdot) : \mathcal{X} \to \mathcal{Y}$$

Unsupervised Learning

$$\mathcal{D} = \{x_i\}_{i=1}^n \in \mathcal{X} \qquad \mathrm{Alg}(\mathcal{D}) \Rightarrow f(\cdot) : \mathcal{X} \to \mathcal{Z}$$



Modele III
Reinforcement Learning

$$\mathrm{Alg(Env)} \Rightarrow (\mathcal{D} = \{(s_i, a_i, r_i, s_i')\}_{i=1}^T,\ \pi(\cdot|s) : \mathcal{S} \to \Delta(\mathcal{A}))$$

Q&A