aha-five

Empowerment & ASP

- Empowerment — An Introduction https://arxiv.org/pdf/1310.1863.pdf 👍
- Keep your options open: An information-based driving principle for sensorimotor systems
  - It measures the capacity of the agent to influence the world in a way that this influence is perceivable via the agent's sensors.
  - Concretely, we define empowerment as the maximum amount of information that an agent could send from its actuators to its sensors via the environment, reducing in the simplest case to the external information channel capacity of the channel from the actuators to the sensors of the agent.
  - An individual agent or an agent population can attempt and explore only a small fraction of possible behaviors during its lifetime.
  - universal & local
- What is intrinsic motivation? A typology of computational approaches
- Variational Information Maximisation for Intrinsically Motivated Reinforcement Learning [2015] https://arxiv.org/abs/1509.08731
- 强化学习如何使用内在动机
  - Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning
- A survey on intrinsic motivation in reinforcement learning https://arxiv.org/abs/1908.06976
- Efficient Exploration via State Marginal Matching https://arxiv.org/pdf/1906.05274.pdf
- Empowerment: A Universal Agent-Centric Measure of Control https://uhra.herts.ac.uk/bitstream/handle/2299/1114/901241.pdf?sequence=1&isAllowed=y
- SMiRL: Surprise Minimizing Reinforcement Learning in Dynamic Environments https://arxiv.org/pdf/1912.05510.pdf [new]

  In the real world, natural forces and other agents already **offer unending novelty**. The second law of thermodynamics stipulates **ever-increasing entropy**, and therefore perpetual novelty, without even requiring any active intervention.

  Paired Open-Ended Trailblazer (POET): Endlessly Generating Increasingly Complex and Diverse Learning Environments and Their Solutions Enhanced POET: Open-Ended Reinforcement Learning through Unbounded Invention of Learning Challenges and their Solutions 🙂

- INTRINSIC MOTIVATION AND AUTOMATIC CURRICULA VIA ASYMMETRIC SELF-PLAY https://arxiv.org/pdf/1703.05407.pdf [起飞 ASP] 🔥 🔥 👍
- Keeping Your Distance: Solving Sparse Reward Tasks Using Self-Balancing Shaped Rewards https://papers.nips.cc/paper/9225-keeping-your-distance-solving-sparse-reward-tasks-using-self-balancing-shaped-rewards.pdf [ASP] 🔥 🧗

  Our method introduces an auxiliary distance-based reward based on pairs of rollouts to encourage diverse exploration. This approach effectively prevents learning dynamics from stabilizing around local optima induced by the naive distance-to-goal reward shaping and enables policies to efficiently solve sparse reward tasks.

- Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm https://arxiv.org/pdf/1712.01815.pdf 💧
- Learning Goal Embeddings via Self-Play for Hierarchical Reinforcement Learning https://arxiv.org/pdf/1811.09083.pdf [ASP] 🔥 👍
- Generating Automatic Curricula via Self-Supervised Active Domain Randomization https://arxiv.org/pdf/2002.07911.pdf [ASP]

Meta

🐸 🐯 🐌 🐍 🐫 🐯 🐢 🐦 🐜 🐷 🐶 🐞 🐔 🐀 🐃 🐄 🐋 🐟 🐬 🐳 🐈 🐡 🐉 🐲 🐐 🐙 🐜 🐢 🐊 🐤

- A Meta-Transfer Objective for Learning to Disentangle Causal Mechanisms 2020 https://arxiv.org/pdf/1901.10912.pdf Yoshua Bengio 🌀 🔥 🔥 👍 🍁 🍃 [contrative loss on causal mechanisms?]

  We show that under this assumption, the correct causal structural choices lead to faster adaptation to modified distributions because the changes are concentrated in one or just a few mechanisms when the learned knowledge is modularized appropriately.

- Causal Reasoning from Meta-reinforcement Learning 2019 😉 😶
- Discovering Reinforcement Learning Algorithms https://arxiv.org/pdf/2007.08794.pdf 👍

  This paper introduces a new meta-learning approach that discovers an entire update rule which includes both '**what to predict**' (e.g. value functions) and '**how to learn from it**' (e.g. bootstrapping) by interacting with a set of environments.

- Zhongwen Xu (DeepMind)

  👂 Discovering Reinforcement Learning Algorithms Attempte to discover the full update rule 👍

  👂 What Can Learned Intrinsic Rewards Capture? How/What value function/policy network 👍

  lifetime return:A finite sequence of agent-environment interactions until the end of training defined by an agentdesigner, which can consist of multiple episodes.

  👂 Discovery of Useful Questions as Auxiliary Tasks 😕

  Related work is good! (Prior work on auxiliary tasks in RL + GVF) 🔥 👍

  👂 Meta-Gradient Reinforcement Learning discount factor + bootstrapped factor 🍃

- Unsupervised Meta-Learning for Reinforcement Learning https://arxiv.org/pdf/1806.04640.pdf [Abhishek Gupta, Benjamin Eysenbach, Chelsea Finn, Sergey Levine] 😕 😉

Meta-RL shifts the human burden from algorithm to task design. In contrast, our work deals with the RL setting, where the environment dynamics provides a rich inductive bias that our meta-learner can exploit.

👂 UNSUPERVISED LEARNING VIA META-LEARNING 😉 We construct tasks from unlabeled data in an automatic way and run meta-learning over the constructed tasks.

👂 Unsupervised Curricula for Visual Meta-Reinforcement Learning [Allan Jabriα; Kyle Hsu] 👍 💧

However, relying solely on discriminability becomes problematic in environments with high-dimensional (image-based) observation spaces as it **results in an issue akin to mode-collapse in the task space**. This problem is further complicated in the setting we propose to study, wherein the policy data distribution is that of a meta-learner rather than a contextual policy. We will see that this can be ameliorated by specifying **a hybrid discriminative-generative model** for parameterizing the task distribution.

We, rather, will **tolerate lossy representations** as long as they capture discriminative features useful for stimulus-reward association.

- Asymmetric Distribution Measure for Few-shot Learning https://arxiv.org/pdf/2002.00153.pdf 👍

  feature representations and relation measure.

-

HRL

- SUB-POLICY ADAPTATION FOR HIERARCHICAL REINFORCEMENT LEARNING https://arxiv.org/pdf/1906.05862.pdf 👎

  👂 STOCHASTIC NEURAL NETWORKS FOR HIERARCHICAL REINFORCEMENT LEARNING
- HIERARCHICAL RL USING AN ENSEMBLE OF PROPRIOCEPTIVE PERIODIC POLICIES https://openreview.net/pdf?id=SJz1x20cFQ 👎
- LEARNING TEMPORAL ABSTRACTION WITH INFORMATION-THEORETIC CONSTRAINTS FOR HIERARCHICAL REINFORCEMENT LEARNING https://openreview.net/pdf?id=HkeUDCNFPS 🔥 👍

  we maximize the mutual information between the latent variables and the state changes.

CITING D...

- Latent Space Policies for Hierarchical Reinforcement Learning 2018 https://arxiv.org/pdf/1804.02808.pdf
- LEARNING SELF-IMITATING DIVERSE POLICIES ICLR2019 https://arxiv.org/pdf/1805.10309.pdf 👍

  Although the policy π(a|s) is given as a conditional distribution, its behavior is better characterized by the corresponding state-action visitation distribution ρπ(s, a), which wraps the MDP dynamics and fully decides the expected return via η(π) = Eρπ [r(s, a)]. Therefore, distance metrics on a policy π should be defined with respect to the visitation distribution ρπ.

  - replay memory may be not useful
  - sub-optimal

- stochasticity: 2-armed bandit problem

improving exploration with stein variational gradient 💢 💢 💢

One approach to achieve better exploration in challenging cases like above is to simultaneously learn **multiple diverse policies** and enforce them to explore different parts of the high dimensional space.

👂 Self-Imitation Learning ICML2018 https://arxiv.org/pdf/1806.05635.pdf

interpreted as **cross entropy loss** (i.e., classification loss for discrete action) with sample weights proportional to the gap between the return and the agent's value estimate

👂 Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm 2016 https://arxiv.org/pdf/1608.04471.pdf

- EPISODIC CURIOSITY THROUGH REACHABILITY [reward design]

  In particular, inspired by curious behaviour in animals, observing something novel could be rewarded with a bonus. Such bonus is summed up with the real task reward — making it possible for RL algorithms to learn from the combined reward. We propose a new curiosity method which uses episodic memory to form the novelty bonus. 💧 **To determine the bonus, the current observation is compared with the observations in memory.** Crucially, the comparison is done based on how many environment steps it takes to reach the current observation from those in memory — which incorporates rich information about environment dynamics. This allows us to overcome the known "couch-potato" issues of prior work — when the agent finds a way to instantly gratify itself by exploiting actions which lead to hardly predictable consequences.

- Combing Skills & **KL regularized expected reward objective**

  👂 [the option keyboard Combing Skills in Reinforcement Learning]()

  We argue that a more robust way of combining skills is to do so directly in **the goal space**, using pseudo-rewards or cumulants. If we associate each skill with a cumulant, we can combine the former by manipulating the latter. This allows us to go beyond the direct prescription of behaviors, working instead in the space of intentions. 😕

  Others: 1. in the space of policies -- over actions; 2. manipulating the corresponding parameters.

  👂 [Scaling simulation-to-real transfer by learning composable robot skills]() 🔥 👍 💢

  we first use simulation to jointly learn a policy for a set of low-level skills, and a **"skill embedding"** parameterization which can be used to compose them.

  👂 [LEARNING AN EMBEDDING SPACE FOR TRANSFERABLE ROBOT SKILLS]() 🔥 👍 💢

  our method is able to learn the skill embedding distributions, which enables interpolation between different skills as well as discovering the number of distinct skills necessary to accomplish a set of tasks.

  👂 [CoMic: Complementary Task Learning & Mimicry for Reusable Skills]() 🔥 💢

  We study the problem of learning reusable humanoid skills by imitating motion capture data and joint training with complementary tasks. **Related work is good!**

  👂 [Learning to combine primitive skills: A step towards versatile robotic manipulation]() 👍

  RL(high-level) + IM (low-level)

  👂 [COMPOSABLE SEMI-PARAMETRIC MODELLING FOR LONG-RANGE MOTION GENERATION]()
  👍

Our proposed method learns to model the motion of human by combining the complementary strengths of both non-parametric techniques and parametric ones. Good EXPERIMENTS!

[LEARNING TO COORDINATE MANIPULATION SKILLS VIA SKILL BEHAVIOR DIVERSIFICATION]

Our method consists of two parts: (1) acquiring primitive skills with diverse behaviors by mutual information maximization, and (2) learning a meta policy that selects a skill for each end-effector and coordinates the chosen skills by controlling the behavior of each skill. **Related work is good!**

[Information asymmetry in KL-regularized RL]

In this work we study the possibility of leveraging such repeated structure to speed up and regularize learning. We start from the **KL regularized expected reward objective** which introduces an additional component, a default policy. Instead of relying on a fixed default policy, we learn it from data. But crucially, we **restrict the amount of information the default policy receives**, forcing it to learn reusable behaviours that help the policy learn faster.

[Exploiting Hierarchy for Learning and Transfer in KL-regularized RL]

The KL-regularized expected reward objective constitutes a convenient tool to this end. It introduces an additional component, a default or prior behavior, which can be learned alongside the policy and as such partially transforms the reinforcement learning problem into one of behavior modelling. **In this work we consider the implications of this framework in case where both the policy and default behavior are augmented with latent variables.** We discuss how the resulting hierarchical structures can be exploited to implement different inductive biases and how the resulting modular structures can be exploited for transfer. Good Writing / Related-work!

[CompILE: Compositional Imitation Learning and Execution]

[Strategic Attentive Writer for Learning Macro-Actions]

[Synthesizing Programs for Images using Reinforced Adversarial Learning]

[Neural Task Graphs: Generalizing to Unseen Tasks from a Single Video Demonstration]

[Hierarchical Cooperative Multi-Agent Reinforcement Learning with Skill Discovery]

[Motion Planner Augmented Action Spaces for Reinforcement Learning]

[The Emergence of Individuality in Multi-Agent Reinforcement Learning]

- Acquiring Diverse Robot Skills via Maximum Entropy Deep Reinforcement Learning [Tuomas Haarnoja, UCB] https://www2.eecs.berkeley.edu/Pubs/TechRpts/2018/EECS-2018-176.pdf

Galaxy

- Deep Reinforcement Learning amidst Lifelong Non-Stationarity https://arxiv.org/pdf/2006.10701.pdf
- Learning Robot Skills with Temporal Variational Inference https://arxiv.org/pdf/2006.16232.pdf

- Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design https://arxiv.org/pdf/0912.3995.pdf [icml2020 test of time award] 😕 ❓
- On Learning Sets of Symmetric Elements https://arxiv.org/pdf/2002.08599.pdf [icml2020 outstanding paper awards] 😕 ❓
- Non-delusional Q-learning and Value Iteration https://papers.nips.cc/paper/8200-non-delusional-q-learning-and-value-iteration.pdf [NeurIPS2018 Best Paper Award]
- SurVAE **Flows**: Surjections to Bridge the Gap between VAEs and Flows [Max Welling] https://arxiv.org/pdf/2007.02731.pdf

  👂 Normalizing Flows: An Introduction and Review of Current Methods 👍 ; Citing: Normalizing Flows for Probabilistic Modeling and Inference 👍 💥 💥 💥 ; lil-log: Flow-based Deep Generative Models ; Jianlin Su: f-VAES 🍬 ; Deep generative models 🍬 ;

  👂 Deep Kernel Density Estimation (Maximum Likelihood, Neural Density Estimation (Auto Regressive Models + Normalizing Flows), Score Matching (MRF), Kernel Exponential Family (RKHS), Deep Kernel);
- Self-Supervised Learning lil-log 🍬 ;

  👂 Self-Supervised Exploration via Disagreement 😕 👎
- **Bisimulation**: Representation learning for control based on bisimulation does not depend on reconstruction, but aims to group states based on their behavioral similarity in MDP. lil-log 🍬

  👂 Equivalence Notions and Model Minimization in Markov Decision Processes http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.61.2493&rep=rep1&type=pdf : refers to an equivalence relation between two states with similar long-term behavior. 😕

  BISIMULATION METRICS FOR CONTINUOUS MARKOV DECISION PROCESSES. https://www.cs.mcgill.ca/~prakash/Pubs/siamFP11.pdf

  👂 DeepMDP: Learning Continuous Latent Space Models for Representation Learning https://arxiv.org/pdf/1906.02736.pdf simplifies high-dimensional observations in RL tasks and learns a latent space model via minimizing two losses: **prediction of rewards** and **prediction of the distribution over next latent states.** 📣 🎯 😕 💥 💣 💥

  👂 DBC: Learning Invariant Representations for Reinforcement Learning without Reconstruction https://arxiv.org/pdf/2006.10742.pdf 💥 💥 💥

  👂 LEARNING INVARIANT FEATURE SPACES TO TRANSFER SKILLS WITH REINFORCEMENT LEARNING https://arxiv.org/pdf/1703.02949.pdf 🔥 👍

  differ in state-space, action-space, and dynamics.

  Our method uses the skills that were learned by both agents to train **invariant feature spaces** that can then be used to transfer other skills from one agent to another.

  👂 UIUC: CS 598 Statistical Reinforcement Learning (S19) NanJiang 👍 💸 💸
- **mutual information**:

  👂 MINE: Mutual Information Neural Estimation 👍 💧 🔥

  👂 Deep InfoMax: LEARNING DEEP REPRESENTATIONS BY MUTUAL INFORMATION ESTIMATION AND MAXIMIZATION 👍 💧

  👂 ON MUTUAL INFORMATION MAXIMIZATION FOR REPRESENTATION LEARNING 🍬 👍

  👂 Deep Reinforcement and InfoMax Learning 🍬 👍 😕 💧

Our work is based on the hypothesis that a model-free agent whose **representations are predictive of properties of future states** (beyond expected rewards) will be more capable of solving and adapting to new RL problems, and in a way, incorporate aspects of model-based learning.

👂 OUYANG:

小王爱迁移,

Self-Supervised Representation Learning From Multi-Domain Data, 🔥 👍 👍

The proposed mutual information constraints encourage neural network to extract common invariant information across domains and to preserve peculiar information of each domain simultaneously. We adopt tractable **upper and lower bounds of mutual information** to make the proposed constraints solvable.

Unsupervised Domain Adaptation via Regularized Conditional Alignment, 🔥 👍

Joint alignment ensures that not only the marginal distributions of the domains are aligned, but the labels as well.

Domain Adaptation with Conditional Distribution Matching and Generalized Label Shift, 🔥 💥 💥 🧠

In this paper, we extend a recent upper-bound on the performance of adversarial domain adaptation to multi-class classification and more general discriminators. We then propose **generalized label shift (GLS)** as a way to improve robustness against mismatched label distributions. GLS states that, conditioned on the label, **there exists a representation of the input that is invariant between the source and target domains**.

Learning to Learn with Variational Information Bottleneck for Domain Generalization,

Through episodic training, MetaVIB learns to gradually narrow domain gaps to establish domain-invariant representations, while simultaneously maximizing prediction accuracy.

Deep Domain Generalization via Conditional Invariant Adversarial Networks, 👍

On Learning Invariant Representation for Domain Adaptation 🔥 💥 🧠

- Distritutional RL Hao Liang, CUHK slide 🧠 🧠

  👂 C51: A Distributional Perspective on Reinforcement Learning 🧠

- Continual Learning

  👂 Continual Learning with Deep Generative Replay 💧 😶

  👂 online learning; regret 🧠

- Active Domain Randomization http://proceedings.mlr.press/v100/mehta20a/mehta20a.pdf 🔥 💥 🔥

  Our method looks for the most **informative environment variations** within the given randomization ranges by **leveraging the discrepancies of policy rollouts in randomized and reference environment instances**. We find that training more frequently on these instances leads to better overall agent generalization.

  Domain Randomization; Stein Variational Policy Gradient;

  Bhairav Mehta On Learning and Generalization in Unstructured Task Spaces 🧠 🧠

  👂 VADRA: Visual Adversarial Domain Randomization and Augmentation 🔥 👍 generative + learner

  👂 Which Training Methods for GANs do actually Converge? 👍 💧 ODE: GAN

  👂 Robust Adversarial Reinforcement Learning 😶

Our proposed method, Robust Adversarial Reinforcement Learning (RARL), jointly trains a pair of agents, a protagonist and an adversary, where the protagonist learns to fulfil the original task goals while being robust to the disruptions generated by its adversary.

[Closing the Sim-to-Real Loop: Adapting Simulation Randomization with Real World Experience]

[POLICY TRANSFER WITH STRATEGY OPTIMIZATION]

[https://lilianweng.github.io/lil-log/2019/05/05/domain-randomization.html]

- Generalization

[Automatic Data Augmentation for Generalization in Deep Reinforcement Learning]

Across different visual inputs (with the same semantics), dynamics, or other environment structures

[Image Augmentation Is All You Need: Regularizing Deep Reinforcement Learning from Pixels]

[Fast Adaptation to New Environments via Policy-Dynamics Value Functions]

PD-VF explicitly estimates the cumulative reward in a space of policies and environments.

- [Exploration Strategies in Deep Reinforcement Learning [chinese]]

[VIME: Variational Information Maximizing Exploration] BNN

the agent should take actions that maximize the reduction in uncertainty about the dynamics.

[Self-Supervised Exploration via Disagreement]

an ensemble of dynamics models and incentivize the agent to explore such that the disagreement of those ensembles is maximized.

[DORA THE EXPLORER: DIRECTED OUTREACHING REINFORCEMENT ACTION-SELECTION]

We propose **E-values**, a generalization of counters that can be used to evaluate the propagating exploratory value over state-action trajectories. [The Hebrew University of Jerusalem]

[EXPLORATION BY RANDOM NETWORK DISTILLATION] [medium]

based on random network distillation (**RND**) bonus

[Randomized Prior Functions for Deep Reinforcement Learning]

[Large-Scale Study of Curiosity-Driven Learning]

[NEVER GIVE UP: LEARNING DIRECTED EXPLORATION STRATEGIES]

episodic memorybased intrinsic reward using k-nearest neighbors; self-supervised inverse dynamics model; Universal Value Function Approximators; different degrees of exploration/exploitation; distributed RL;

[Self-Imitation Learning via TrajectoryConditioned Policy for Hard-Exploration Tasks]

[Planning to Explore via Self-Supervised World Models]

a selfsupervised reinforcement learning agent that tackles both these challenges through a new approach to self-supervised exploration and fast adaptation to new tasks, which need not be known during exploration. **unlike prior methods which retrospectively compute the novelty of observations after the agent has already reached them**, our agent acts efficiently by leveraging planning to seek out expected future novelty.

- Offline RL

  [Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems](#)

  https://danieltakeshi.github.io/2020/06/28/offline-rl/

- Pareto Multi-Task Learning https://arxiv.org/pdf/1912.12854.pdf

  we proposed a novel Pareto Multi-Task Learning (Pareto MTL) algorithm to generate a set of well-distributed Pareto solutions with different trade-offs among tasks for a given multi-task learning (MTL) problem.

  [Efficient Continuous Pareto Exploration in Multi-Task Learning](#) [zhihu](#)

- BNN

  [Auto-Encoding Variational Bayes](#)

MARL

- MARL https://cloud.tencent.com/developer/article/1618396
- A Survey on Transfer Learning for Multiagent Reinforcement Learning Systems

Others

- Gaussian Precess, Kernel Method, [EM](#), [Conditional Neural Process](#), [Neural Process](#), (Deep Mind, ICML2018)
- [Ising model](#), Gibbs distribution,
- [f-GAN](#), [GAN-OP](#), [ODE: GAN](#),
- [Wasserstein Distance](#), [Statistical Aspects of Wasserstein Distances](#), [Optimal Transport and Wasserstein Distance](#),
- [MARKOV-LIPSCHITZ DEEP LEARNING](#),
- [Hindsight](#), [Rainbow](#) ,

Blogs & Corp. & Legends

[Lil'Log](#),

[covariant](#),

UCB: [Tuomas Haarnoja](#), [Pieter Abbeel](#), [Sergey Levine](#), [Abhishek Gupta](#) , [Coline Devin](#), [YuXuan (Andrew) Liu](#), [Rein Houthooft](#),

Standord: [Chelsea Finn](#),

NYU: [Rob Fergus](#),

MIT: [Bhairav Mehta](#),

DeepMind: [Yee Whye Teh](#) [[Homepage]](#), [Alexandre Galashov](#), [Leonard Hasenclever](#) [[GS]](#), [Siddhant M. Jayakumar](#),

Zhongwen Xu,