Reducing the Complexity of Reinforcement Learning Using Localization and Factorization

Peter Sunehag and Marcus Hutter
Presented by Jan Leike

Australian National University

AGI 2015

A Generic Reinforcement Learning Agent

- $\mathcal{M} :=$ class of finitely many deterministic environments
- Sunehag and Hutter (2012): Optimistic agent picks policy π and environment $\nu \in \mathcal{M}$ that promise highest reward and follows π until contradictions

A Generic Reinforcement Learning Agent

- $\mathcal{M} :=$ class of finitely many deterministic environments
- Sunehag and Hutter (2012): Optimistic agent picks policy π and environment $\nu \in \mathcal{M}$ that promise highest reward and follows π until contradictions
- Number of ε-errors $\leq \frac{|\mathcal{M}|}{1-\gamma} \log \frac{1}{\varepsilon(1-\gamma)}$.

A Generic Reinforcement Learning Agent

- $\mathcal{M} :=$ class of finitely many deterministic environments
- Sunehag and Hutter (2012): Optimistic agent picks policy π and environment $\nu \in \mathcal{M}$ that promise highest reward and follows π until contradictions
- Number of ε-errors $\leq \frac{|\mathcal{M}|}{1-\gamma} \log \frac{1}{\varepsilon(1-\gamma)}$.
- Sunehag and Hutter (2013, 2014): Extension to growing classes $\mathcal{M}$: agent switches policy if newly introduced environment promises more reward
- true environment $\mu \in \mathcal{M} \Longrightarrow$ number of ε-errors $\leq \mathit{const} + \frac{N_t}{1-\gamma} \log \frac{1}{\varepsilon(1-\gamma)}$ where N_t is number environments introduced at time t

A Generic Reinforcement Learning Agent

- $\mathcal{M}$:= class of finitely many deterministic environments
- Sunehag and Hutter (2012): Optimistic agent picks policy π and environment $\nu \in \mathcal{M}$ that promise highest reward and follows π until contradictions
- Number of ε-errors $\leq \frac{|\mathcal{M}|}{1-\gamma} \log \frac{1}{\varepsilon(1-\gamma)}$.
- Sunehag and Hutter (2013, 2014): Extension to growing classes $\mathcal{M}$: agent switches policy if newly introduced environment promises more reward
- true environment $\mu \in \mathcal{M} \Longrightarrow$ number of ε-errors $\leq const + \frac{N_t}{1-\gamma} \log \frac{1}{\varepsilon(1-\gamma)}$ where N_t is number environments introduced at time t
- $\Longrightarrow$ weakly asymptotically optimal agent

Combining Deterministic Laws

- hypothesis class was $\mathcal{M}$ = set of environments
- Instead, a class of laws $\mathcal{T}$ is more efficient (Sunehag and Hutter, 2013, 2014): partial (factorization) predictions under some circumstances (localization).

Combining Deterministic Laws

- hypothesis class was $\mathcal{M}$ = set of environments
- Instead, a class of laws $\mathcal{T}$ is more efficient (Sunehag and Hutter, 2013, 2014): partial (factorization) predictions under some circumstances (localization).

- Example: Newton's three laws of motion and the law of universal gravitation forms Newton's mechanical universe

Combining Deterministic Laws

- hypothesis class was $\mathcal{M}$ = set of environments
- Instead, a class of laws $\mathcal{T}$ is more efficient (Sunehag and Hutter, 2013, 2014): partial (factorization) predictions under some circumstances (localization).

  - Example: Newton's three laws of motion and the law of universal gravitation forms Newton's mechanical universe
  - Contradiction of a law is a contradiction of a lot of environments

Combining Deterministic Laws

- hypothesis class was $\mathcal{M}$ = set of environments
- Instead, a class of laws $\mathcal{T}$ is more efficient (Sunehag and Hutter, 2013, 2014): partial (factorization) predictions under some circumstances (localization).

  - Example: Newton's three laws of motion and the law of universal gravitation forms Newton's mechanical universe
  - Contradiction of a law is a contradiction of a lot of environments
  - $|\mathcal{M}|$ is replaced by $|\mathcal{T}|$ in the error bound

Semi-determinism: Deterministic Laws and Probabilistic Background Knowledge

- Combining laws making partial deterministic predictions and separately learnt correlations between the entries within a feature vector (background knowledge)

Semi-determinism: Deterministic Laws and Probabilistic Background Knowledge

- Combining laws making partial deterministic predictions and separately learnt correlations between the entries within a feature vector (background knowledge)
- Predict as much as possible with deterministic laws and conditioning on background knowledge

Semi-determinism: Deterministic Laws and Probabilistic Background Knowledge

- Combining laws making partial deterministic predictions and separately learnt correlations between the entries within a feature vector (background knowledge)
- Predict as much as possible with deterministic laws and conditioning on background knowledge
- truth is in the class $\Longrightarrow$ optimistic agent has the same error bounds as before

Stochastic Laws: Learning Correlations

- Here we introduce the formal notion of stochastic laws

Stochastic Laws: Learning Correlations

- Here we introduce the formal notion of stochastic laws
- Under a domination assumption stochastic laws merge with the truth

Stochastic Laws: Learning Correlations

- Here we introduce the formal notion of stochastic laws
- Under a domination assumption stochastic laws merge with the truth
- Using dominant stochastic laws replaces the need to provide correlations as background

Stochastic Laws: Learning Correlations

- Here we introduce the formal notion of stochastic laws
- Under a domination assumption stochastic laws merge with the truth
- Using dominant stochastic laws replaces the need to provide correlations as background
- Example: Context Tree Weighting can be broken up into laws for each context

Mixing Stochastic and Deterministic Laws

- New hypothesis class = mix of deterministic and stochastic laws
- Fall back on dominant stochastic laws when all deterministic laws fail

Mixing Stochastic and Deterministic Laws

- New hypothesis class = mix of deterministic and stochastic laws
- Fall back on dominant stochastic laws when all deterministic laws fail
- deterministic learning: exclusion
  stochastic learning: merging

Mixing Stochastic and Deterministic Laws

- New hypothesis class = mix of deterministic and stochastic laws
- Fall back on dominant stochastic laws when all deterministic laws fail
- deterministic learning: exclusion
  stochastic learning: merging
- Relying only on determinism when there noise breaks the agent

Mixing Stochastic and Deterministic Laws

- New hypothesis class = mix of deterministic and stochastic laws
- Fall back on dominant stochastic laws when all deterministic laws fail
- deterministic learning: exclusion
  stochastic learning: merging
- Relying only on determinism when there noise breaks the agent

Combining stochastic laws with deterministic laws (predictions) for each context and that are used until contradictions is highly beneficial if some aspects of the environment are deterministic but others are not

Conclusions

- Starting with axioms of rational and optimistic general RL agents; error bounds, localization, and factoring (as in e.g. Baum's economy of agents) through relying on laws.
- Here we added dominant stochastic laws

Conclusions

- Starting with axioms of rational and optimistic general RL agents; error bounds, localization, and factoring (as in e.g. Baum's economy of agents) through relying on laws.
- Here we added dominant stochastic laws
- The rest is just coming out in a journal paper (Sunehag and Hutter, 2015)

Conclusions

- Starting with axioms of rational and optimistic general RL agents; error bounds, localization, and factoring (as in e.g. Baum's economy of agents) through relying on laws.
- Here we added dominant stochastic laws
- The rest is just coming out in a journal paper (Sunehag and Hutter, 2015)
- Peter is now in perfect position at Google DeepMind to implement, but instead tries to serve YouTube recommendations with Deep-RL

Conclusions

- Starting with axioms of rational and optimistic general RL agents; error bounds, localization, and factoring (as in e.g. Baum's economy of agents) through relying on laws.
- Here we added dominant stochastic laws
- The rest is just coming out in a journal paper (Sunehag and Hutter, 2015)
- Peter is now in perfect position at Google DeepMind to implement, but instead tries to serve YouTube recommendations with Deep-RL
- If that sounds like more fun and you got strong CS/math/stat/ML (DL and/or RL), email Peter at `sunehag@google.com`. We are hiring!
- If that sounds like less fun than mathematical theory, Marcus Hutter is also hiring! Ask the speaker.

References

Peter Sunehag and Marcus Hutter. Optimistic agents are asymptotically optimal. In Australasian Joint Conference on Artificial Intelligence, pages 15–26. Springer, 2012.

Peter Sunehag and Marcus Hutter. Learning agents with evolving hypothesis classes. In Artificial General Intelligence, pages 150–159. Springer, 2013.

Peter Sunehag and Marcus Hutter. A dual process theory of optimistic cognition. In Annual Meeting of the Cognitive Science Society, pages 2949–2954, 2014.

Peter Sunehag and Marcus Hutter. Rationality, optimism and guarantees in general reinforcement learning. Journal of Machine Learning Research, 2015.