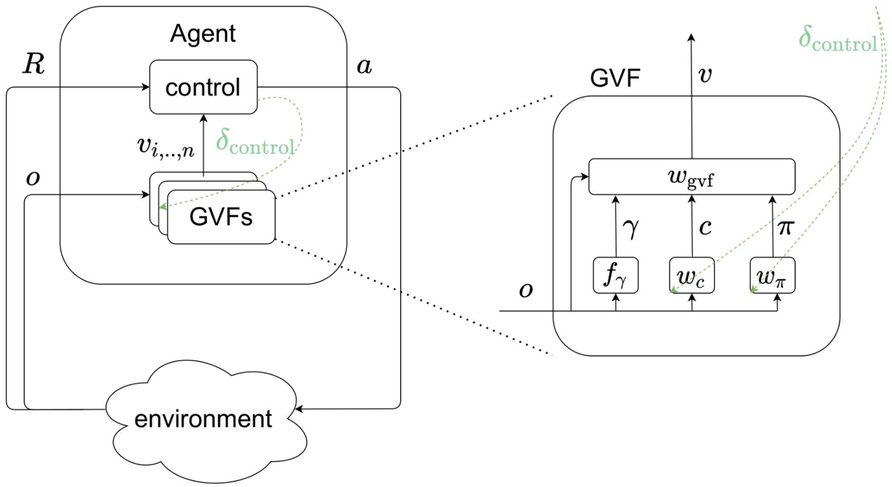
https://arxiv.org/pdf/2111.11212.pdf

Finding Useful Predictions by Meta-gradient Descent to Improve Decision-making

Alex Kearney 
Anna Koop 
Johannes Günther 
Patrick Pilarski 

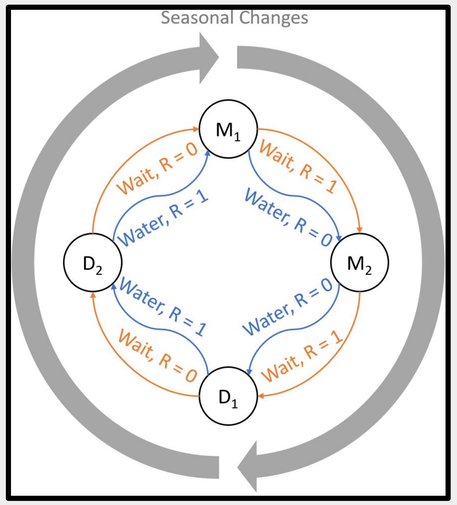
Introduction

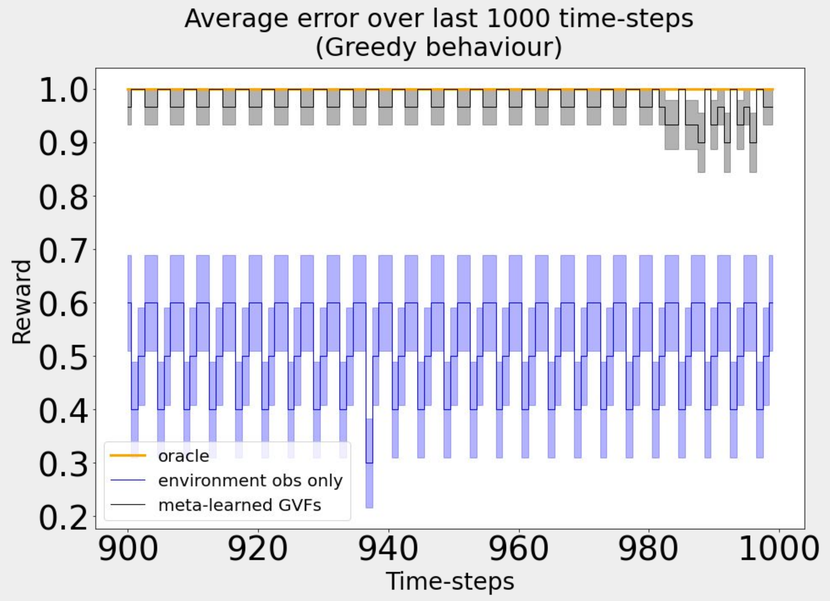
- A growing collection of work in Reinforcement Learning expresses an agent's knowledge of the world as a collection of predictions of future sensations—predictions expressed as *General Value Functions* (GVFs)
- Deciding *what to predict* is a challenge; predictions are often chosen by human designers.
- What if the agent could choose what to predict via Meta Gradient Descent (MGD)?



Learning what to predict

- Three scalars determine what a GVF is about: γ, π, and c.
- We define each question-parameter using a set of weights the agent can learn.
- Weights are learned by MGD: construct a gradient of the agent's TD error WRT the meta-parameters that specify what a GVF is about.





Experimental setup

- Can an agent learn what predictions are necessary in order to solve a POMDP?
- Two baselines: agent with oracle, agent using environment observations alone.
- Over 30 independent trials, an agent using MGD can achieve performance almost equivalent to oracle.

Funding & Acknowledgement This research was undertaken, in part, thanks to funding from the Canada Research Chairs program, the Canada CIFAR AI Chairs program, the Canada Foundation for Innovation, the Alberta Machine Intelligence Institute, Alberta Innovates, and the Natural Sciences and Engineering Research Council. AKK was supported by scholarships and awards from NSERC, Alberta Innovates, and Borealis AI.

Lessons learned

- We provide an example where ***learned GVFs are used directly as inputs to a control agent.***
- ***We enabled an agent to independently choose what to predict*** through meta-gradient descent.
- An **agent was able to choose predictions necessary** to solve a POMDP without human direction, **achieving performance comparable to an oracle agent.**