

DELTA: Dynamically Evolving Long-Term Autonomy

CHIST-ERA Projects Seminar 2021 - LLIS 2016

13 April 2021

① Project description and objectives

② Dissemination Actions

③ Key Achievements

- Planning
- Exploration
- Task decomposition
- Integration

④ Conclusion

1 Project description and objectives

2 Dissemination Actions

3 Key Achievements

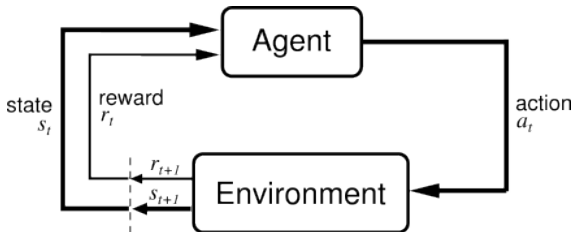
- Planning
- Exploration
- Task decomposition
- Integration

4 Conclusion

Partners and respective funding agency:

- Universitat Pompeu Fabra (UPF), Spain (AEI)
- Inria Lille - Nord Europe (Inria), France (ANR)
- Montanuniversität Leoben (MUL), Austria (FWF)
- Université de Liège (ULG), Belgium (FNRS)

Duration: 36 months (2018-2020)



- Formalized as a **Markov decision process** (MDP) $M = \langle S, A, P, R \rangle$
- At each time step t , learning agent observes state s_t , selects action a_t , transitions to state s_{t+1} and receives reward r_{t+1}
- Aim is to compute a **policy** $\pi : S \rightarrow A$ that maximizes some measure of **expected future reward**

Pros:

- Adapts to unknown environments
- Can optimise action selection from experience
- Achieves human-level performance in many tasks

Cons:

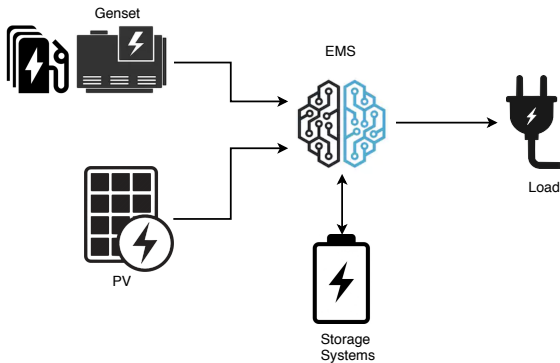
- Stationary environment model
- Fixed objective
- Trials end once the objective is met

Address three key issues related to lifelong RL:

- Planning
- Exploration
- Task decomposition

Key components of lifelong RL:

- Flexible environmental model based on factored states
- Allow state variables and actions to change over time
- Multiple objectives (some initially unknown)



- Microgrids have **non-stationary dynamics**
- Lifelong learning should **adapt** to changes in dynamics

- Develop an integrated autonomous system which is able to successfully interact with its environment during long periods of time without breaking down or failing to achieve its objectives.
- Improve on the state-of-the-art for active network management and microgrid management.
- Disseminate the novel RL algorithms developed as part of the project.

① Project description and objectives

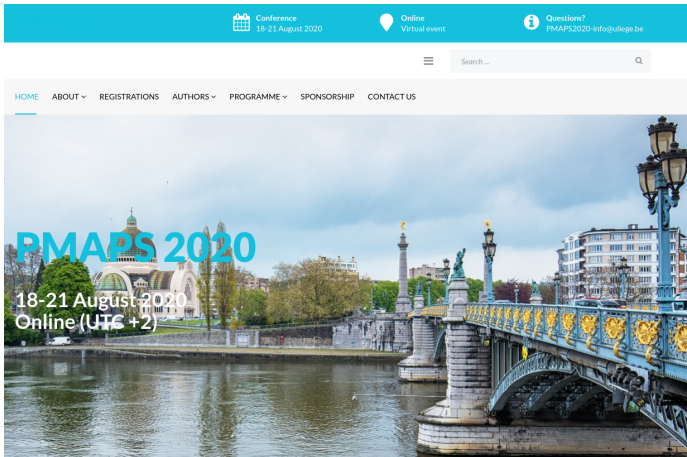
② Dissemination Actions

③ Key Achievements

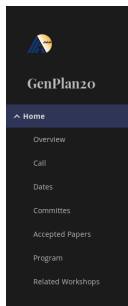
- Planning
- Exploration
- Task decomposition
- Integration

④ Conclusion

- 39 publications in top-rated international AI & ML conferences (AAAI, AISTATS, ALT, COLT, ICAPS, ICML, NeurIPS, UAI, etc.)
- 4 journal publications (AIJ, JAIR) and several submissions under review (Energy, JMLR)
- 8 joint publications among project partners



Tutorial on “Reinforcement learning applications for active network management for electrical distribution networks and microgrid management”



[AAAI 2020](#) workshop on

Generalization in Planning

[Overview](#)

[Call](#)

[Dates](#)

[Program](#)
[Committee](#)

[Accepted Papers](#)

[Program](#)

[Related](#)
[Workshops](#)

Overview

Confirmed Invited Speakers

We are glad to announce the following speakers for GenPlan '20 workshop:

- [Hector Geffner](#) (invited talk)
- [Leslie Kaelbling](#) (invited talk)
- [Giuseppe De Giacomo](#) (highlights talk)
- [George Konidaris](#) (highlights talk)
- [Sheila McIlraith](#) (highlights talk)
- [Peter Stone](#) (highlights talk)



Bridging the Gap Between AI Planning and Reinforcement Learning (PRL)

While AI Planning and Reinforcement Learning communities focus on similar sequential decision-making problems, these communities remain somewhat unaware of each other on specific problems, techniques, methodologies, and evaluation.

Invited Speakers

[Playlist of all the invited talks.](#)

- ▶ [Will Dabney](#), DeepMind (recorded)
- ▶ [Alan Fern](#), Oregon State University (recorded)
- ▶ [Michael Littman](#), Brown University (recorded)
- ▶ [Julian Schrittwieser](#), DeepMind (recorded)
- ▶ [Peter Stone](#), The University of Texas at Austin (recorded)

① Project description and objectives

② Dissemination Actions

③ Key Achievements

- Planning
- Exploration
- Task decomposition
- Integration

④ Conclusion

➊ Project description and objectives

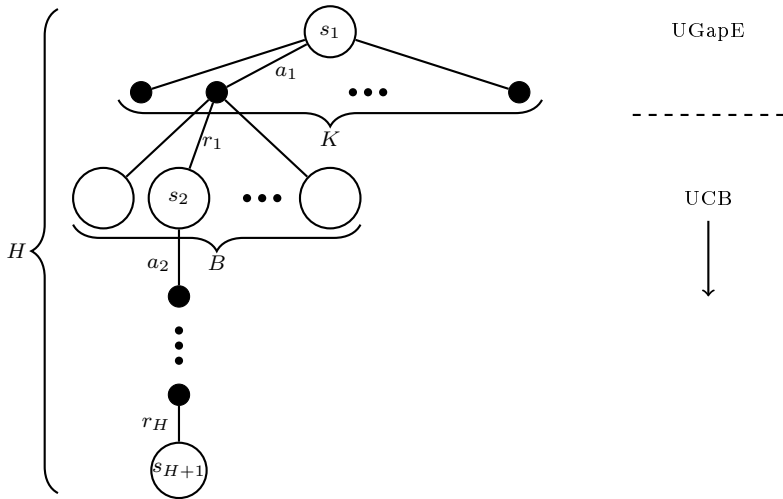
➋ Dissemination Actions

➌ Key Achievements

- Planning
- Exploration
- Task decomposition
- Integration

➍ Conclusion

- Novel algorithm for Monte-Carlo Tree Search with gap-dependent sample complexity [Jonsson et al., NeurIPS 2020]
- A kernel-based approach to non-stationary reinforcement learning in metric spaces [Domingues et al., AISTATS 2021]



Sample complexity $\mathcal{O} \left(\sum_{a_1} \frac{(BK)^{H-1}}{(\Delta_1(s_1, a_1) \vee \epsilon)^2} \left[\ln \left(\frac{1}{\delta} \right) + BH \ln(BK) \right] \right)$

Non-Stationary Reinforcement Learning in Metric Spaces

- Episodic reinforcement learning in non-stationary MDPs
- Assumes the existence of a **metric** in state-action space
- Non-parametric kernel smoothing can handle non-stationarity by introducing time- and space-dependent kernels
- **First regret bound** for non-stationary RL in continuous environments

➊ Project description and objectives

➋ Dissemination Actions

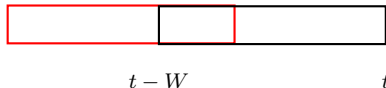
➌ Key Achievements

- Planning
- **Exploration**
- Task decomposition
- Integration

➍ Conclusion

- Algorithm that automatically adapts to the amount of change in the environment [Gajane et al., LLARLA 2018, best paper award]
- Novel algorithms for reward-free exploration [Kaufmann et al., ALT 2021; Ménard et al., submitted to ICML 2021]

Exploration that adapts to amount of change



- Use a **sliding window** to represent the most recent history
- Estimate a model only based on the most recent history
- Can handle **abrupt changes** in the environment
- Contribution: high-probability upper bound on the cumulative regret

Idea: optimally explore an episodic MDP with horizon H so as to learn an optimal policy **for any reward function r** [Jin et al. 20]

RF-UCRL: Maintain an **upper bound** $\bar{E}^t(s, a)$ on the approximation error

Sampling rule: policy π^t acts **greedily** with respect to $\bar{E}^t(s, a)$

Stopping rule: stop at time τ if $\bar{E}^\tau(s_1, \pi^\tau(s_1)) \leq \varepsilon/2$

Sample complexity $\tau = \tilde{\mathcal{O}}\left(\frac{S^2 A H^3}{\varepsilon^2}\right)$ (small ε); $\tau = \mathcal{O}\left(\frac{S A H^3}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right)\right)$ (small δ)

1 Project description and objectives

2 Dissemination Actions

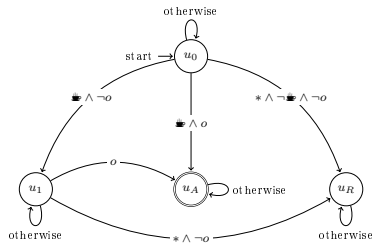
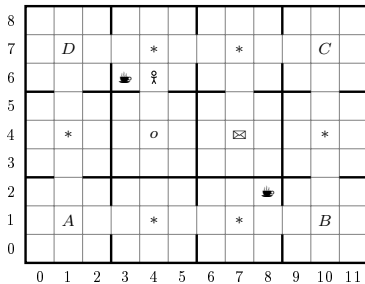
3 Key Achievements

- Planning
- Exploration
- Task decomposition
- Integration

4 Conclusion

- Autonomous discovery of subtask automata [Furelos et al., AAAI 2020 + JAIR 2021]
- Hierarchical width-based planning and learning [Junyent et al., ICAPS 2021]

Autonomous discovery of subtask automata



- Use reinforcement learning to gather **traces of observations**
- Inductive logic answer set programming (ILASP) generates minimal automata
- Subtasks are to satisfy conditions on automaton edges
- System maintains **library** of known subtasks

- Assume that each state s is factored into high-level s_h and low-level s_ℓ
- Plan and learn on two levels of abstractions!
- **Subtask**: change the current high-level state
- At the high level, plan on top of high-level states
- Maintain and update a library of known subtasks

➊ Project description and objectives

➋ Dissemination Actions

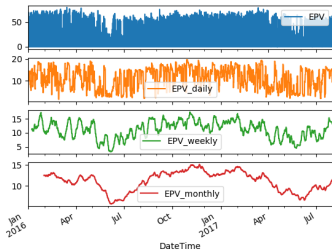
➌ Key Achievements

- Planning
- Exploration
- Task decomposition
- **Integration**

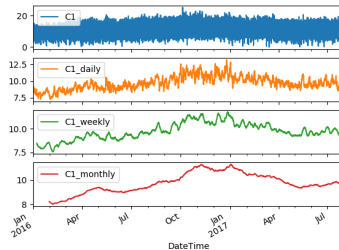
➍ Conclusion

- Deep reinforcement learning for microgrid management [Totaro et al., submitted to Energy]
- Stochastic Kalman Gradient Descent [Totaro & Jonsson, L4DC 2021]

Deep reinforcement learning for microgrid management



PV production



Electrical load

- 18 months' data from El Espino microgrid, Bolivia
- Train policy on first 12 months, evaluate on remaining 6 months

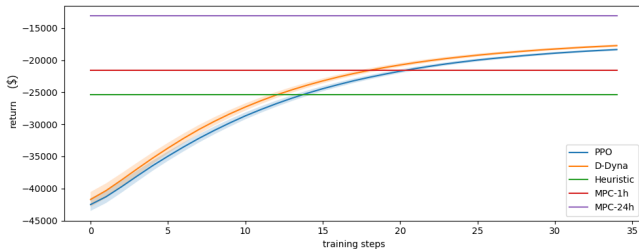
Two different types of non-stationarity:

- **Gradual**, e.g. photovoltaic production and electrical load
- **Sudden**, e.g. battery failure

Model-based reinforcement learning algorithm that can handle both types

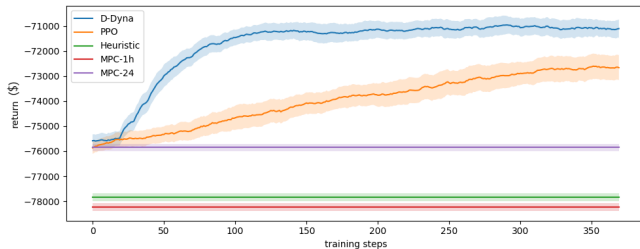
Limitation: Once trained, policy is **fixed**, not adaptive!

Performance under gradual change



- Two benchmarks: heuristic and model predictive control w/ perfect info
- Reinforcement learning yields a 25% cost reduction compared to heuristic

Robustness to sudden change



- Heuristic and MPC have no way of adapting to battery failure
- On the contrary, reinforcement learning can adapt to sudden changes!

Nonlinear extension of Kalman Temporal Difference for reinforcement learning

Allows **deep reinforcement learning** to handle **non-stationarity**

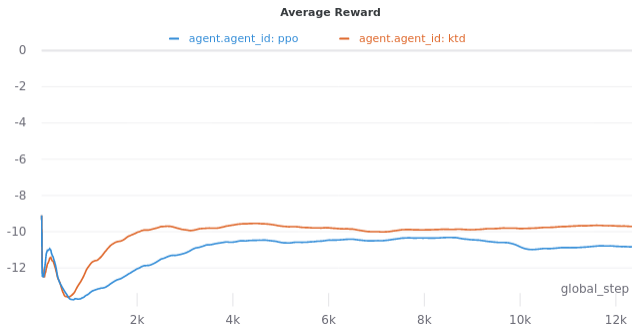
Addresses the drawbacks of Extended Kalman filtering:

- Only propagates uncertainty around the mean estimate
- The updated estimation error covariance may not lie in the feasible set
- The update rules have time complexity $O(|\theta|^2)$

Two novel components:

- Row dropout regularization
- Low-rank approximation

Results for microgrid simulator



- Trained a separate network that predicts the load and PV generation

① Project description and objectives

② Dissemination Actions

③ Key Achievements

- Planning
- Exploration
- Task decomposition
- Integration

④ Conclusion

Project objectives:

- For each objective, several novel algorithms with theoretical guarantees
- Academically, largely achieved the targeted outcomes

Spin-off effects:

- Microgrid simulator itself, open-source of academic or commercial use
- Adaptive control for microgrids as a service, integrating novel algorithms

Exploitation:

- ULG has a wide network of collaborators in electrical distribution networks
- Contact companies that develop products related to off-grid microgrid control