



THALES

# AUTOMATED DISCOVERY OF OPTIONS IN FACTORED REINFORCEMENT LEARNING

Olga KOZLOVA, Olivier SIGAUD and Christophe MEYER

[olga.kozlova@isir.upmc.fr](mailto:olga.kozlova@isir.upmc.fr)

[olivier.sigaud@upmc.fr](mailto:olivier.sigaud@upmc.fr)

[christophe.meyer@thalesgroup.com](mailto:christophe.meyer@thalesgroup.com)



**ISIR**

### AUTOMATED DISCOVERY OF OPTIONS IN FACTORED REINFORCEMENT LEARNING

Éric Granger, Olivier Pietroni & Sébastien Bubeck

<https://arxiv.org/abs/1806.02832v1>

**Factorial Bandit Problem**

(MAB)  $n$  arms,  $n$  actions  
 • Model of expected cumulative reward of the actions is defined by the unknown matrix  $R$  ( $n \times n$  real numbers)

**Factorial Dynamical**

(FDMAB)  $n$  arms,  $n$  actions  
 • Model of the expected cumulative reward of the actions is defined by the unknown matrix  $R$  ( $n \times n$  real numbers)

**Factorial Multi-Armed Bandit Problem**

(FMMAB)  $n$  arms,  $n$  actions  
 • Model of the expected cumulative reward of the actions is defined by the unknown matrix  $R$  ( $n \times n$  real numbers)

**Factorial Submodular Learning (FSL)**

• Model of the expected cumulative reward of the actions is defined by the unknown matrix  $R$  ( $n \times n$  real numbers)

• Learning upper bound of the expected cumulative reward of the actions is defined by the unknown matrix  $R$  ( $n \times n$  real numbers)

• Finding a policy that maximizes the expected cumulative reward of the actions is defined by the unknown matrix  $R$  ( $n \times n$  real numbers)

**ISIR (Incrementally Scalable ISIR)**

• Model of the expected cumulative reward of the actions is defined by the unknown matrix  $R$  ( $n \times n$  real numbers)

• Learning upper bound of the expected cumulative reward of the actions is defined by the unknown matrix  $R$  ( $n \times n$  real numbers)

• Finding a policy that maximizes the expected cumulative reward of the actions is defined by the unknown matrix  $R$  ( $n \times n$  real numbers)

**Experiments: The test problem**

• Grid world with 10 states, 4 actions  
 • Grid world with 10 states, 4 actions  
 • Grid world with 10 states, 4 actions

Algorithm	Upper Bound	Lower Bound	Upper Bound	Lower Bound
ISIR	0.95	0.95	0.95	0.95
ISIR	0.95	0.95	0.95	0.95
ISIR	0.95	0.95	0.95	0.95

**Future work:**

- Local model learning
- Experimental evaluation of the ISIR algorithm
- Incremental optimization of options in ISIR

**ISIR**  
 Institut des Systèmes Intelligents et de Robotique  
 June 18, 2018, Bordeaux, France

**UPMC** **Thales** **THALES**

