

Reinforcement Learning and Beyond

Part II: **Transfer Learning in RL**

Section 3: Conclusions and Future Work

Outline

- Selection of future challenges
 - Towards a theoretical foundation of transfer in RL
 - Enabling fully autonomous transfer
 - Constructing effective task learning sequences
- Concluding summary

Theoretical Foundation

- **RL is more complex than supervised learning**
- Transfer in RL introduces specific scenarios and issues
- Recent theoretical results in RL shows that it has strong connections with statistical learning theory results (see (Munos&Szepesvari, 2008))
- Recent theoretical results in TL in supervised learning shows the effectiveness of TL w.r.t. single task learning
 - Inductive bias learning
 - Multi-task learning

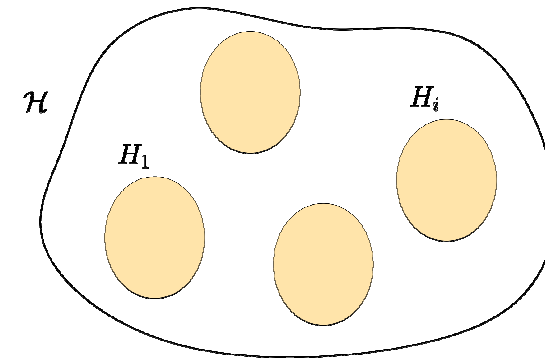
State of the Art

	Supervised Learning	Reinforcement Learning
Single Task	(Vapnik&Chervonenkis, 1971)	(Munos&Szepesvari, 2008) (Farahmand et al., 2008)
Multi Task	(Baxter, 2000) (Ben-David&Shuller, 2008)	Nothing!

Inductive Bias Learning

- “A model of inductive bias learning” (Baxter, 2000)
- Scenario
 - Distribution \mathcal{Q} over task space \mathcal{P}
 - Training set m samples from each of n tasks
- *Objective*: find a hypothesis space H which contains **good hypotheses for all the tasks** in \mathcal{P} **on average** (according to distribution \mathcal{Q})

Inductive Bias Learning



Extension of complexity measures (e.g., VC, covering numbers) to the set of hypothesis spaces \mathcal{H}

Inductive Bias Learning

- If
 - **enough tasks** are provided to the learner and
 - **enough samples** per task are collected
 - the set of hypothesis spaces is **not too big**
- Then
 - The **generalization error** of the hypothesis space H **can be bounded** on new tasks drawn from \mathcal{Q}
 - The **generalization** performance is **better** than learning independently
 - The number of **samples decreases with the number of tasks**

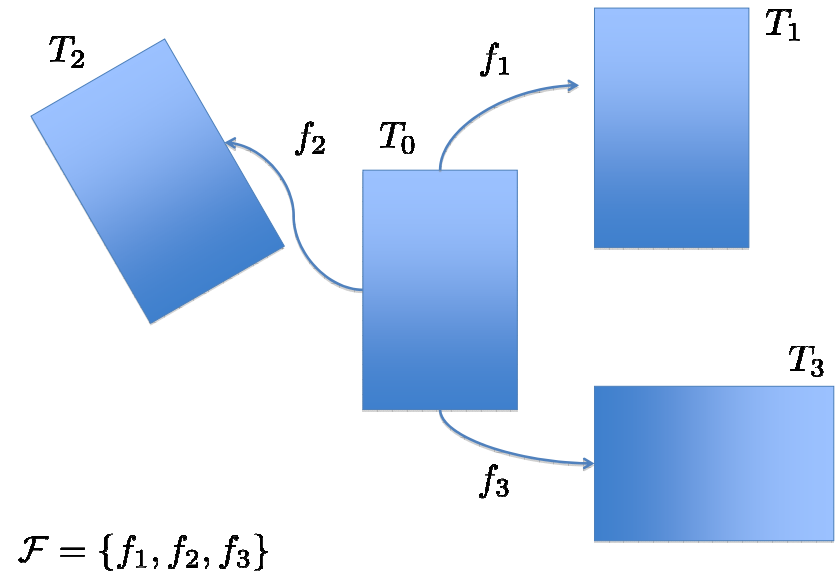
Inductive Bias Learning in RL

- Pros
 - The scenario could be easily adapted to RL domains (distribution of MDPs/value functions)
 - RL could be decomposed in a sequence of supervised learning problems (e.g., FQI)
- Cons
 - Not straightforward generalization of Baxter’s result across different iterations (e.g., FQI, policy improvement)
 - Similar MDPs does not imply similar solutions

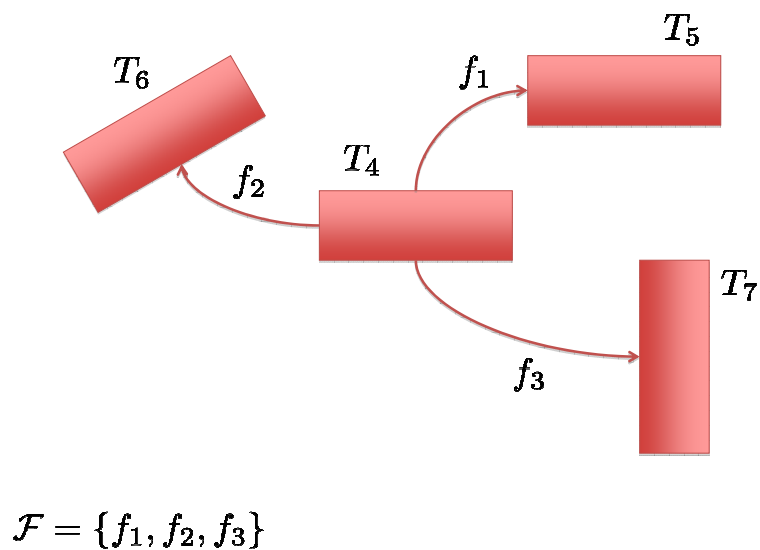
Transformation-Based MTL

- “A Notion of Task relatedness Yielding Provable Multiple-task Learning Guarantees” (Ben-David & Shuller, 2008)
- Scenario
 - Multi-task learning on n tasks
 - Training set: m samples from each of n tasks
 - **Assumption**: all the tasks pair-wise f -related, with f a transformation in a set of possible transformations \mathcal{F}
- *Objective*: given a target task, use all the samples to find the *high-level* characteristics of the solution and use the target samples to learn the *task-specific* solution

Transformation-Based MTL



Transformation-Based MTL



Transformation-Based MTL

- *Phase1*: use all the samples to identify which class the target task belongs to



- *Phase2*: use only samples of the target task to identify the best solution



Transformation-Based MTL

- If
 - enough samples per task are collected
 - enough target samples are collected
 - if the set of transformations is not too big
- Then
 - The performance for (any!) target task is better than learning independently
 - The number of samples decreases with the number of tasks

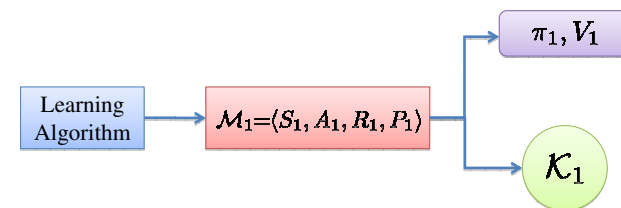
Transformation-Based MTL in RL

- Pros
 - RL could be decomposed in a sequence of supervised learning problems (e.g., FQI)
- Cons
 - Not straightforward definition of transformation in RL domains
 - Not straightforward generalization of Ben-David's bounds across different iterations

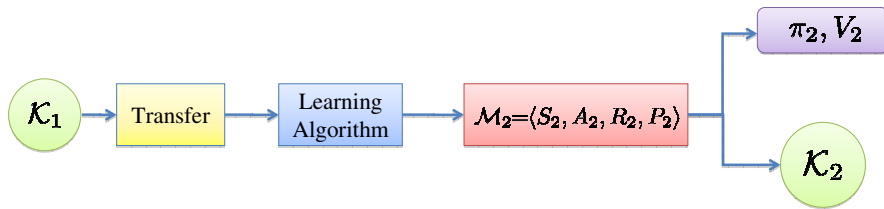
Theoretical Foundation

- Similarities of RL and supervised learning
- Promising line of research
- Several issues still unsolved (even in TL in supervised learning!)

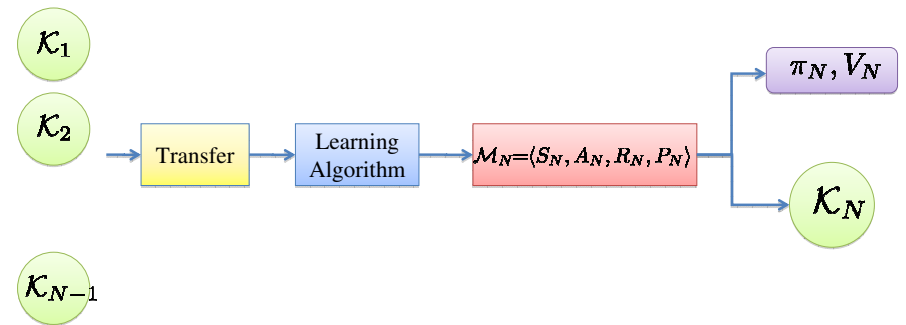
Fully Autonomous Transfer



Fully Autonomous Transfer



Fully Autonomous Transfer

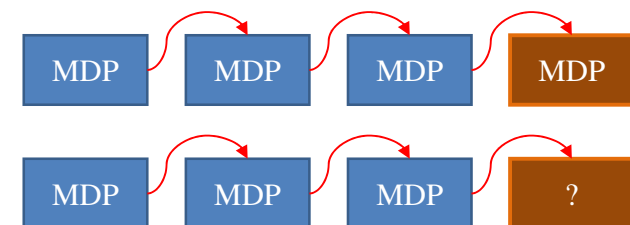


Fully Autonomous Transfer

- The full transfer problem
 - Different SxA
 - N-to-1 transfer
- Challenges
 - Learn the mapping
 - Select source tasks
 - Transfer effectively

Constructing Task Sequences [Taylor, 2009]

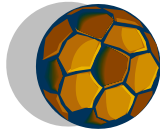
- Humans can selecting a training sequence
- Results in faster training / better performance
- Meta-planning problem for agent learning



Useful Information for Sequence Construction

- **Common Sense**

- Soccer balls roll after being kicked
- Friction reduces an object's speed



- **Domain Knowledge**

- It is easier to complete short passes than long passes

- **Algorithmic Knowledge**

- State space size can impact learning speed