Linking Metric and Symbolic Levels in Autonomous Reinforcement Learning

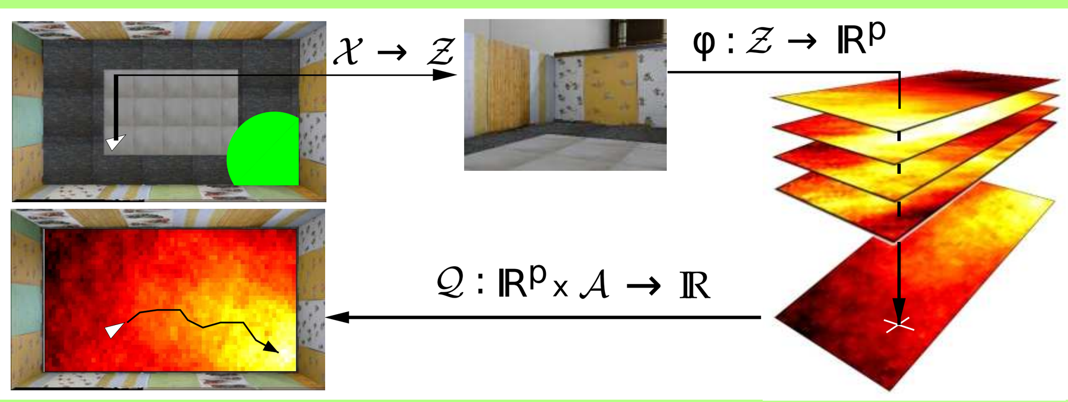
DFG
PPR 1527

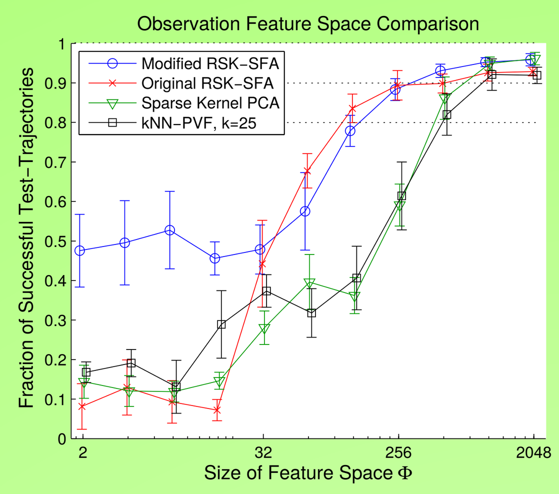
Wendelin Böhmer <wendelin@ni.tu-berlin.de>

Klaus Obermayer <oby@ni.tu-berlin.de>

Neural Information Processing Group, Technische Universität Berlin

Inductive Learning

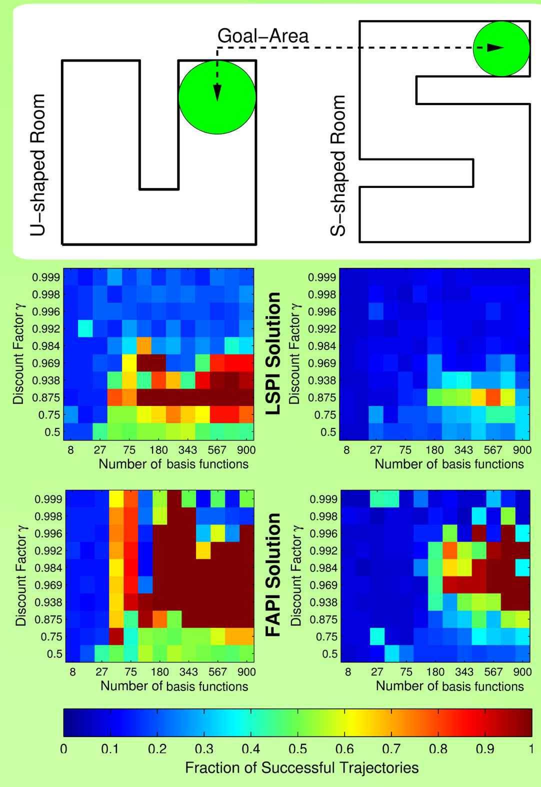


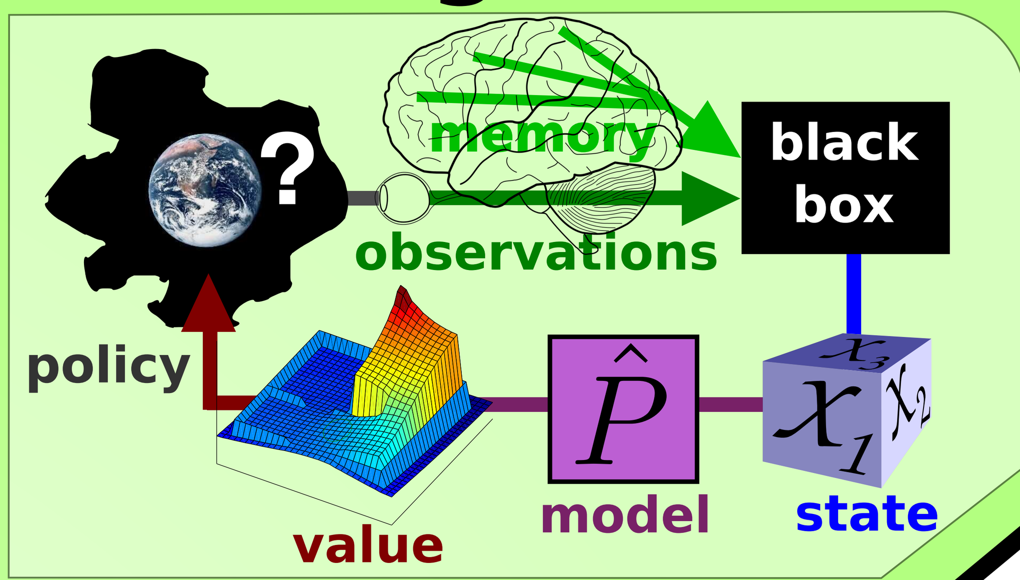


Theorem: SFA minimizes a bound on the approximation error of all LSTD value functions with the same transition model.

- Slow features encode transition model
- Values generalize near optimally
- Features are learned inductively
- **Not a solution to curse of insufficient samples**

Deductive Learning





- FAPI estimates values with models
- Generalizes everywhere equally
- Transition model estimation requires uniform samples
- **No solution either**

Project Publications

- Böhmer et. al. (2015). *Autonomous learning of state representation for control.* **KI** 29(4):353-362.
- Böhmer and Obermayer (2015). *Regression with linear factored functions.* Proceedings to **ECML/PKDD**, pages 119-134.
- Böhmer et al. (2013). *Construction of approximation spaces for reinforcement learning.* **JMLR** 14:2067-2118.
- Böhmer and Obermayer (2013). *Towards structural generalization: factored approximate planning.* **ICRA** workshop on autonomous learning.
- Böhmer et al. (2012). *Generating feature spaces for linear algorithms with regularized sparse kernel slow feature analysis.* **ML** 89(1):67-86.

Assumptions on state representation $\mathcal{X}$

- factorizing transition models

$$P(d\vec{y}|\vec{x}) = \prod_{k=1}^{d} P_k(dy_k|\vec{z}_k)$$

- sparse state interdependencies

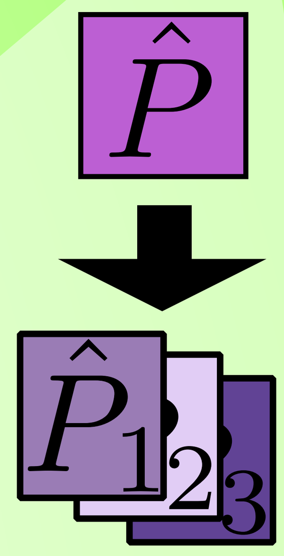
$$\vec{z}_k \subset \vec{x}$$

- linear factored functions

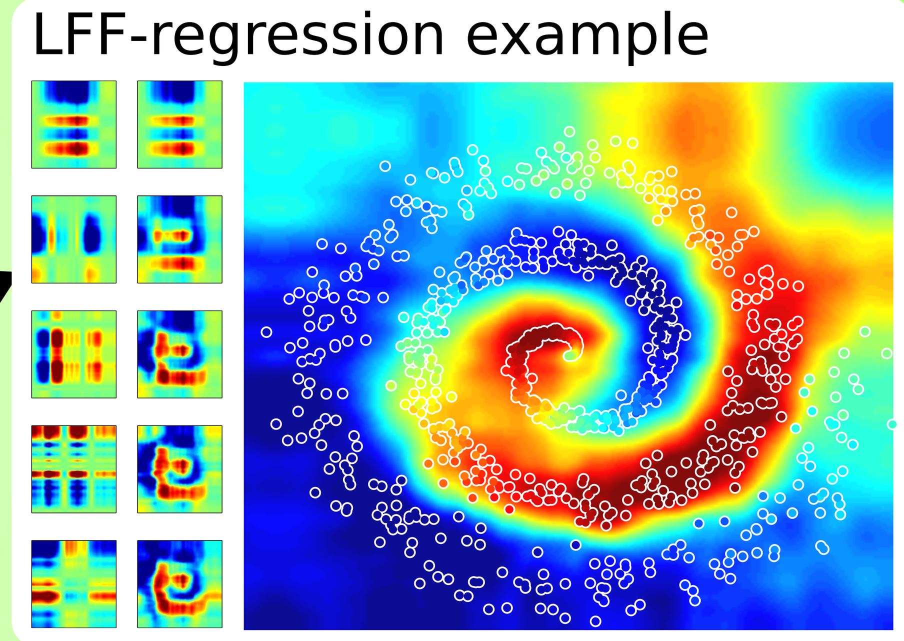
$$f(\vec{x}) = \sum_{i=1}^{m} a_i \prod_{k=1}^{d} \sum_{j=1}^{m_k} B^k_{ji} \phi^k_j(x_k)$$

(LFF)

Factored Models



LFF-regression example

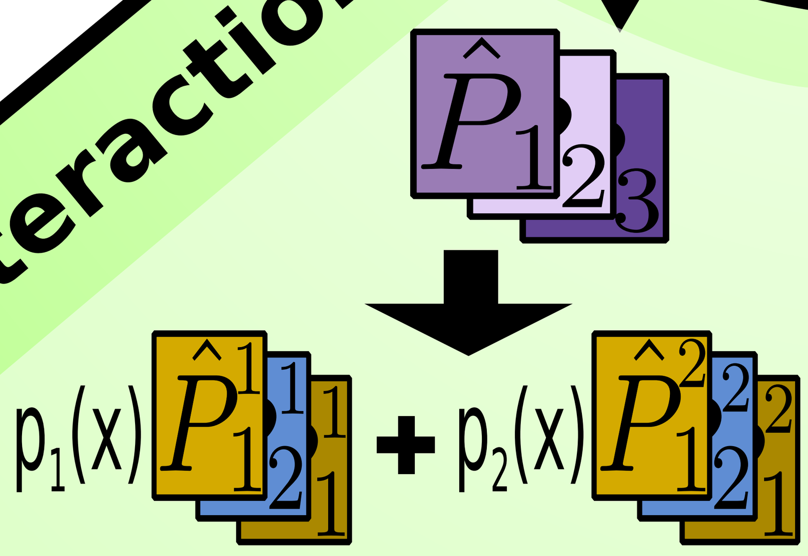


- Estimating factored transition models for LFF is regression

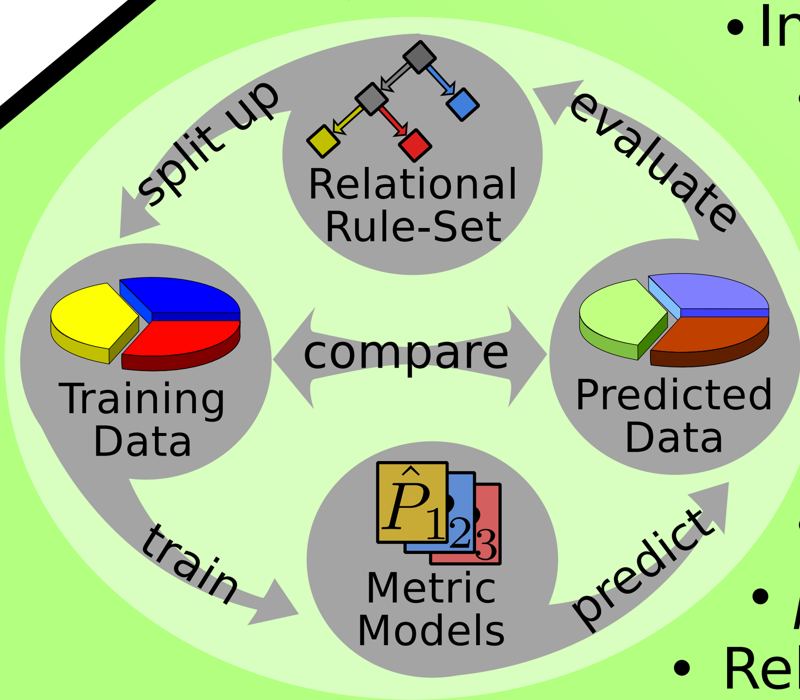
$$\hat{P}[f](\vec{x}) = \sum_{i=1}^{m} a_i \prod_{k=1}^{d} \sum_{j=1}^{m_k} B^k_{ji} \underbrace{\hat{P}_k[\phi^k_j](\vec{z}_k)}_{\text{target}}$$

- Greedy basis const.
- Non-convex CCD
- appears robust

Symbolic Interactions



- Models seldom factorize sparsely
- Interactions are often relational
- Mixture-of-experts approach

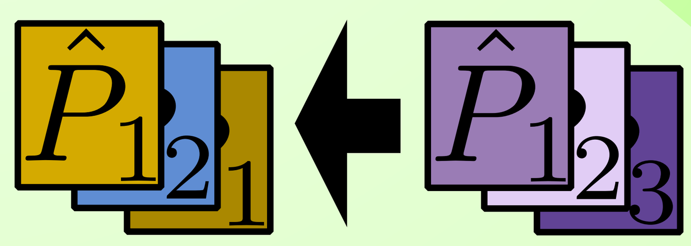


$$\hat{P}_k(dy_k\,|\,\vec{x}) = \sum_i p_i(\vec{x})\, \hat{P}^i_k(dy_k\,|\,\vec{z}^i_k)$$

$$\text{s.t.} \qquad \sum_i p_i(\vec{x}) \leq 1\,, \quad \forall x \in \mathcal{X}$$

- Conditions $p_i(x)$ are relational
- $p_i(x)$ is approximated as LFF
- Relational rules from samples

Object Orientation



- Classes can pool training data
- Classification by clustering
- Generalization over environments
- Deductive value estimation

$$\inf_f \Big\| f - r - \gamma \overbrace{\prod_k \hat{P}_k \big[\underbrace{\hat{\Gamma}_\pi[f]}_{\text{LFF}} \big]}^{\text{LFF}} \Big\|^2_\nu$$

- **Breaks curse for suitable $\mathcal{X}$**
- Deductive values will have errors
  - Shaping: inductive error correction
  - Regularization: prior for inductive RL