Chapter 5. The Bootstrapping Approach to Developing Reinforcement Learning-based Strategies

in Reinforcement Learning for Adaptive Dialogue Systems, V. Riser

Course: Autonomous Machine Learning

Jung, Jaehoon

School of Computer Science and Engineering

Seoul National University

Motivation

- Reinforcement Learning (RL) based strategies outperform hand-coded strategies
  - “intelligently” adapt their strategies to the representation of the dialogue environment
- Correct representation of the learning environment is important
  - Data-driven methods are preferred

Motivation

- Major limitations of previous approaches
  - Relies on a large quantity of data
  - The optimal policy can only be discovered when it is present in the data set
- Simulated learning environments
  - Still hand-crafted
- Suggestion
  - Learn simulated environment from small amounts of Wizard-of-Oz (WOZ) data

Term Definition

- Bootstrap
  - Various different meanings in the various fields
  - Original meaning
    - Pulling oneself up by one's own bootstraps
  - Problem of how to learn an optimal strategy before a working system or prototype exits

Term Definition

- Bootstrap
  - Problem of how to learn an optimal strategy before a working system or prototype exits
  - Conceptually contrasts with dialogue strategies being manually “uplifted” by an human

Related Work

- Approaches to address the problem of lacking initial training data
  - Handcrafting the simulated components
    - Schatzmann et al, 2007a
  - Online learning
    - Chickering and Paek, 2007
  - Transferring policies from another domain
    - Lemon an Liu, 2007; Lemon et al, 2006a
  - Starting learning from a limited amount of WOZ data
    - Prommer et al, 2006; Williams and Young, 2004b

Related Work

- Use of WOZ data in the context of RL
  - Williams and Young
    - Discover the state and action space for the design of a Markov Decision Process
  - Prommer et al
    - Build a simulated user and noise model for simulation-based RL

Advantages for Learning from WOZ Data

- Data collection does not require a working prototype
  - Allows to apply RL to new application areas beyond the scope of existing dialogue systems
- Can be used to explore human strategies as a basis for automatic optimization
  - Allows less restricted exploration of dialogue policies

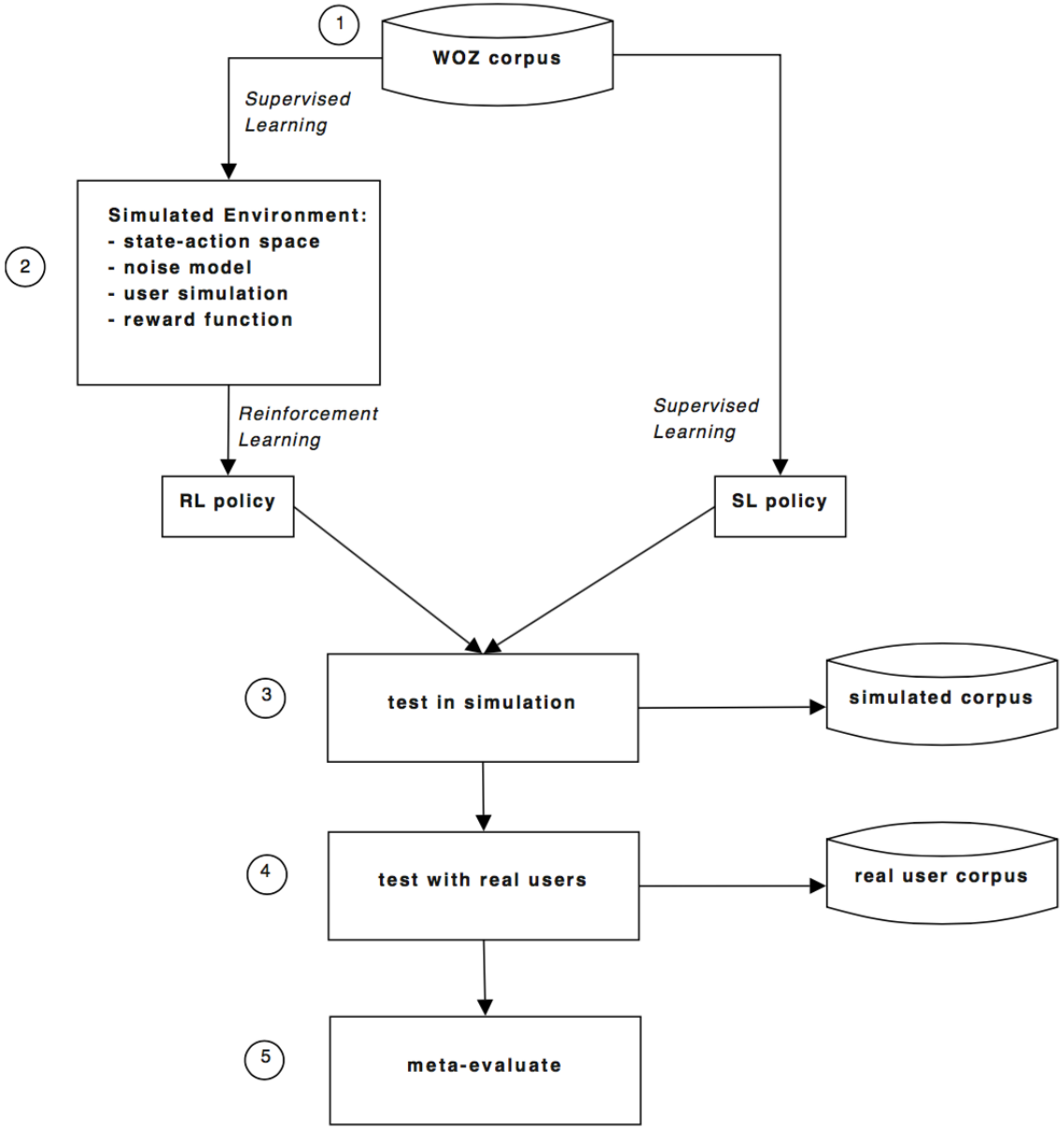
Challenges for Learning from WOZ Data

- How to learn simulated components from small data set
  - In previous, the simulated components are either hand-crafted or learned via Supervised Learning (SL).
  - Supervised Learning
    - Learning to infer a general mappings from unseen example
    - Tends to “overfit” if training examples are rare
  - Major challenges of simulated components
    - Realistic behavior
    - Wide-coverage behavior
  - Methods will be introduced in chapter 7

Challenges for Learning from WOZ Data

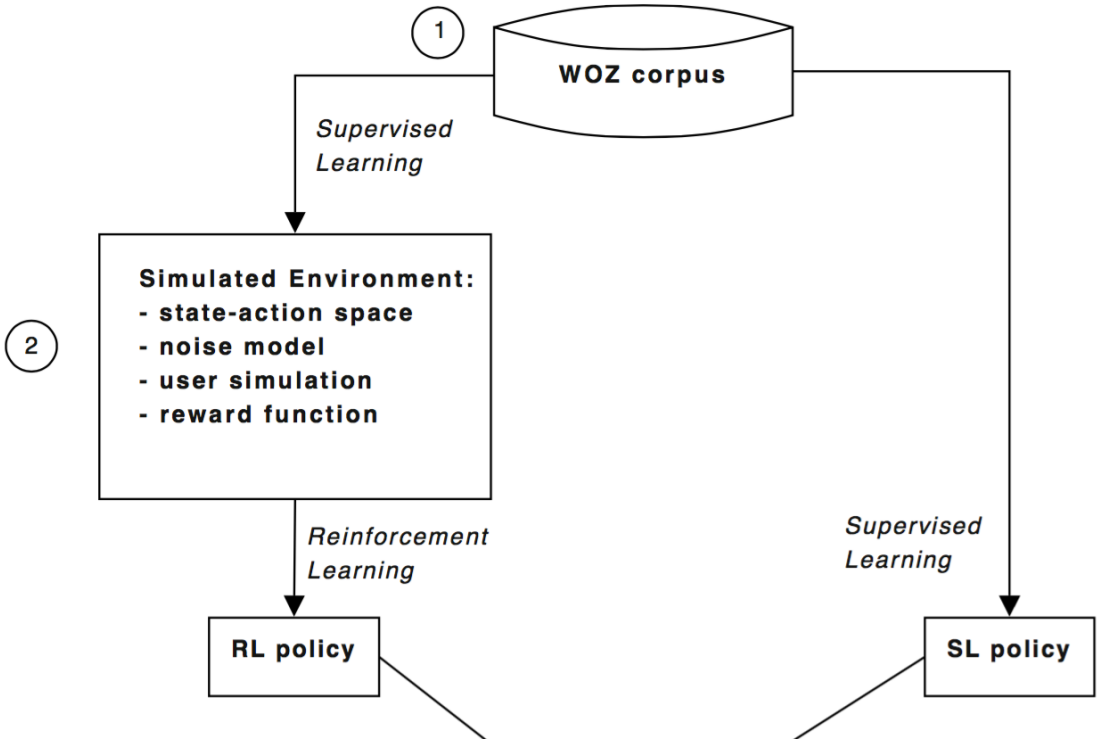
- How to account for the fact that a WOZ study only simulates real HCI
  - Simulated environment from WOZ data is “simulation of simulation”
  - Extra post-evaluating step in general simulation-based RL framework

The Bootstrapping Method



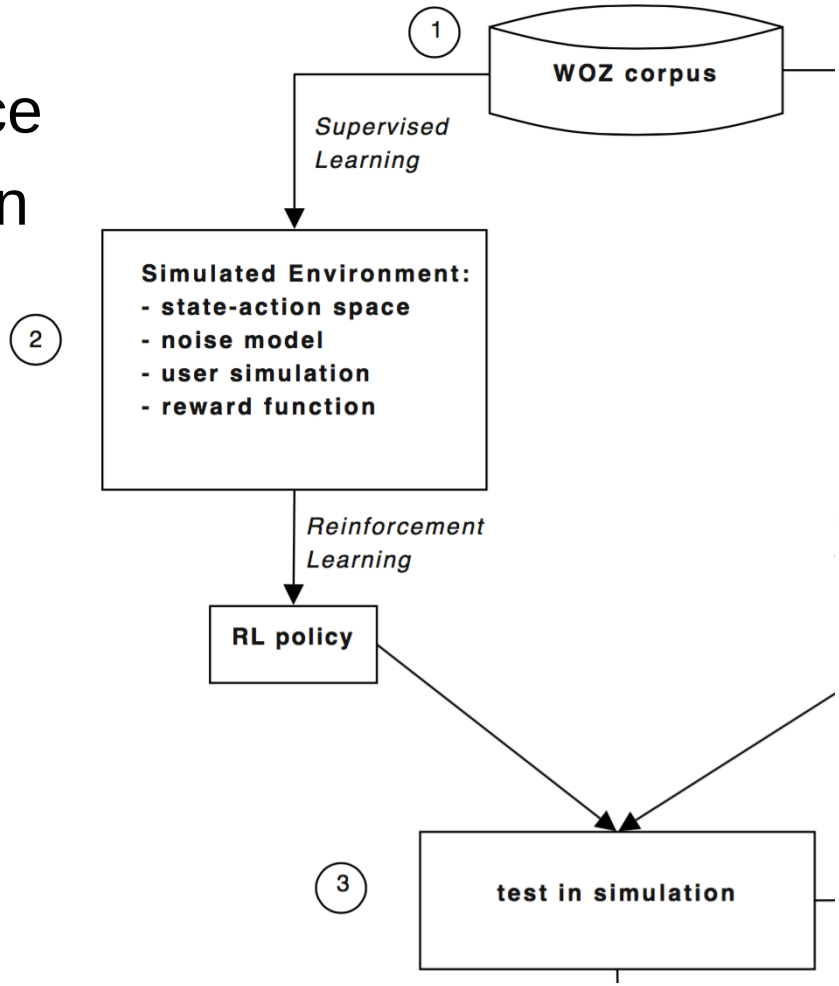
The Bootstrapping Method: Step1

- Data Collection in a WOZ Experiment
  - Some changes to the conventional WOZ setup
    - Wizard can interact freely with the user
    - Artificially introduced channel noise



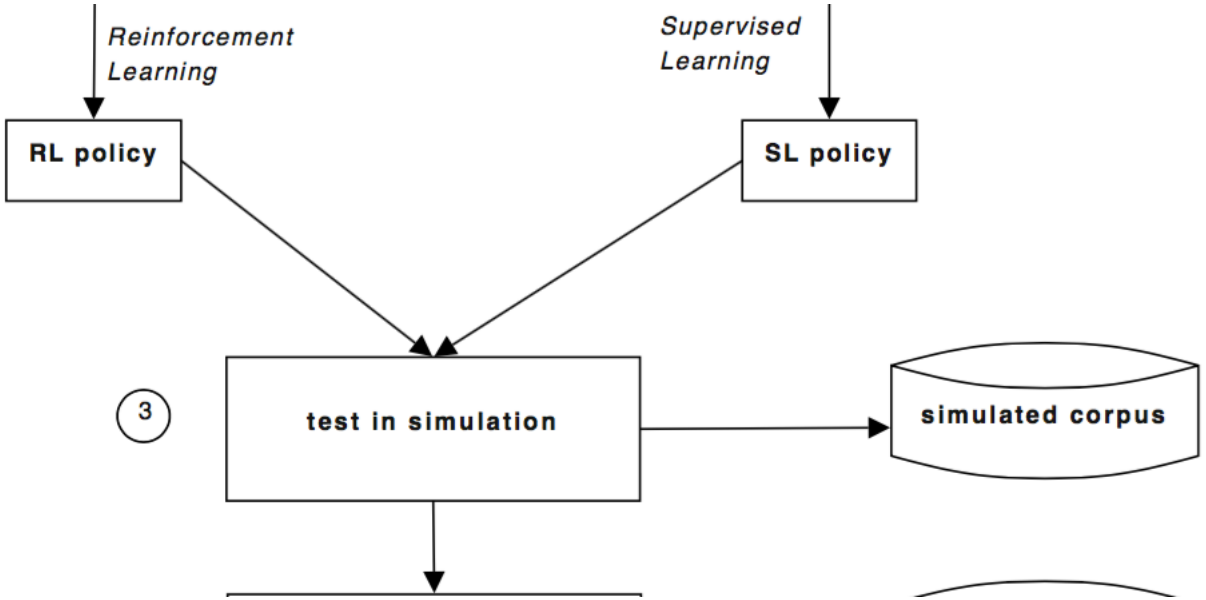
The Bootstrapping Method: Step2

- Build a Simulated Learning Environment
  - Model the components
    - Action set and state space
    - User and noise simulation
    - Objective function



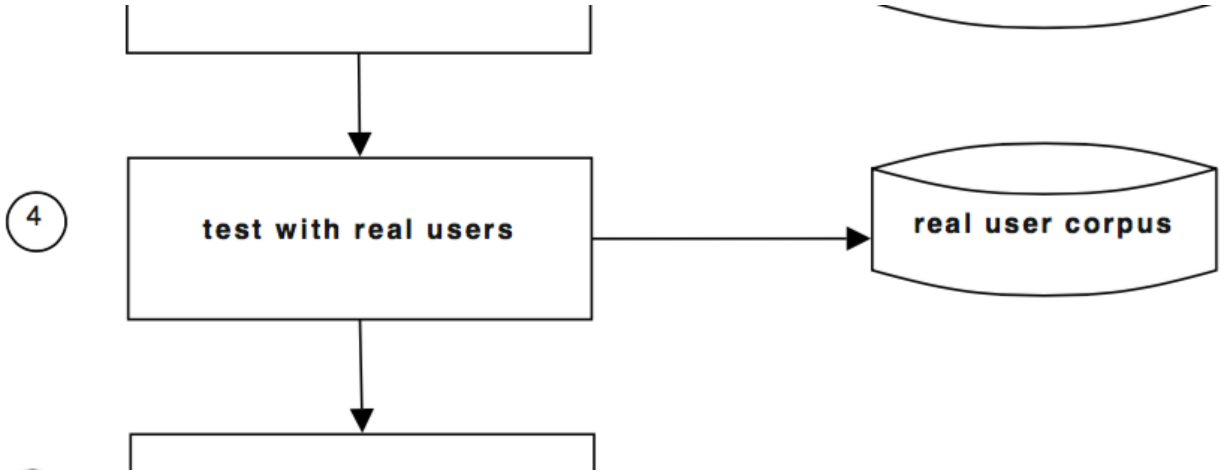
The Bootstrapping Method: Step3

- Train and test a strategy in simulation
  - Test RL-based strategy against SL-based strategy
  - SL is used to capture the average human wizard strategy



The Bootstrapping Method: Step4

- Test with Real Users



The Bootstrapping Method: Step5

- Post-Evaluation
  - Address the problem that a WOZ experiment is only a simulation of real HCI
    - Compare the WOZ study, the dialogues generated in simulation, and the final user tests
    - Explore whether the objective function used for learning is a realistic estimate of real user

Summary

- Bootstrapping an optimized strategy from WOZ data using RL

- Step1: Data collection in a WOZ environment
- Step2: Build a simulated learning environment
- Step3: Train and test a strategy in simulation
- Step4: Test with real users
- Step5: Post-evaluation