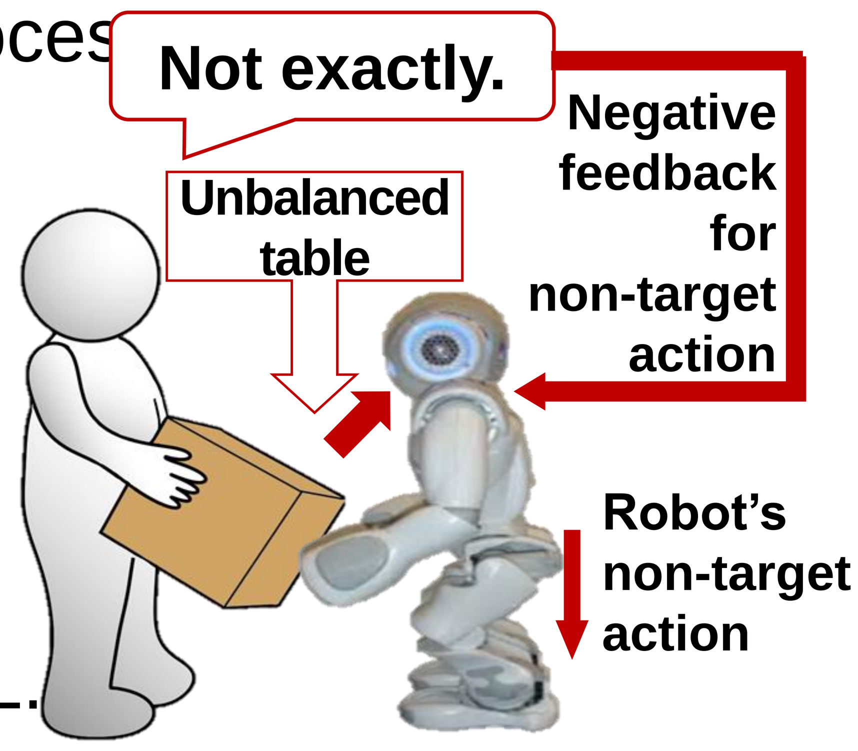
Interactive Reinforcement Learning for Table Balancing Robot

Haein Jeon, Yewon Kim and Bo-Yeong Kang

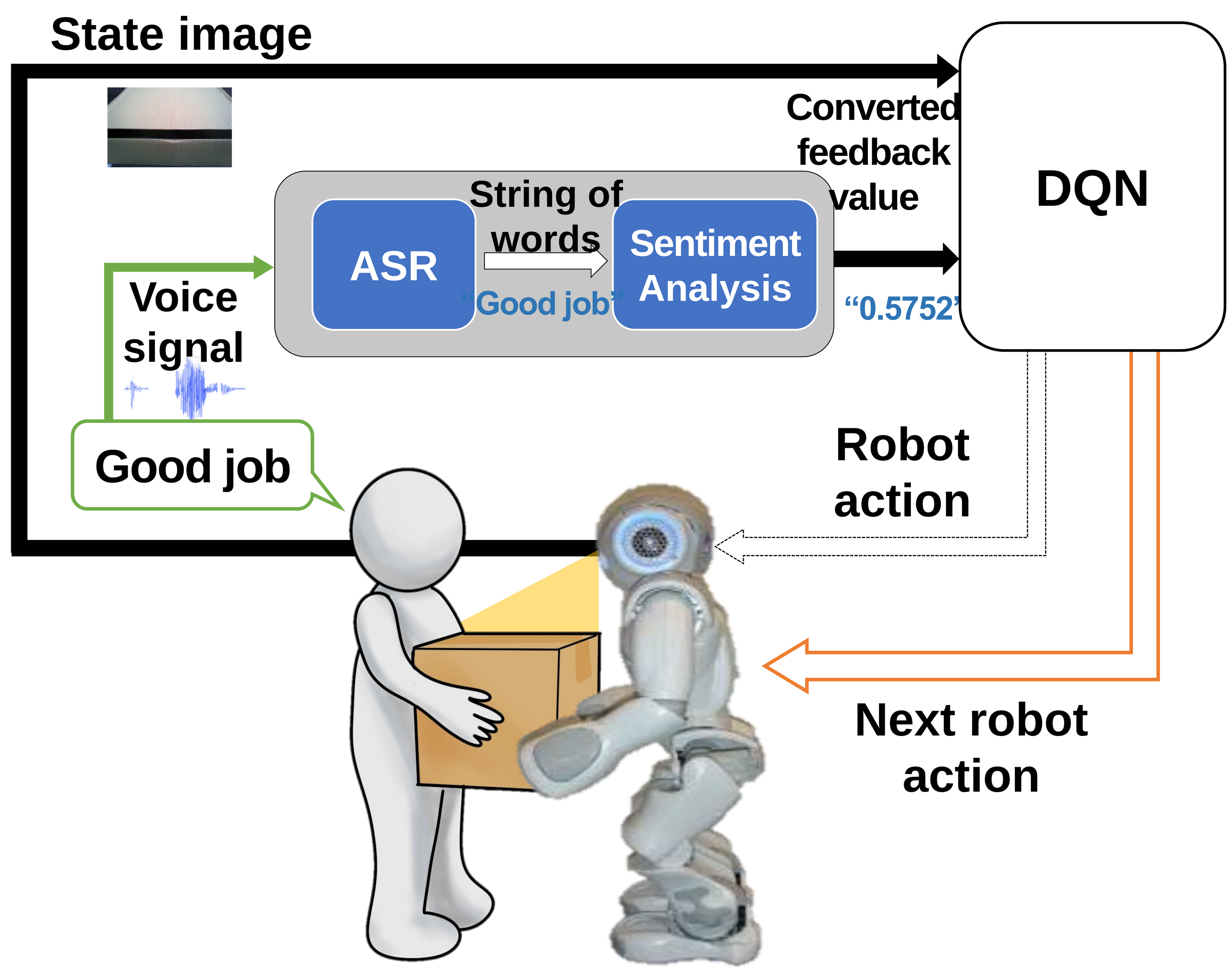
Artificial Intelligence Robot Laboratory

Introduction

- The need for anyone to easily train robots .
- Applying reinforcement learning(RL) to robot training.
- Giving verbal instruction is one of the important methods that humans have for teaching a task.
- Reward Shaping(RS): in RL, external trainer gives intermediate rewards to a learning agent to guide its learning process
- **Idea**: allow the robot to learn a cooperative table balancing task from its interaction with human trainer through voice feedback. by applying RS to RL.



Approach



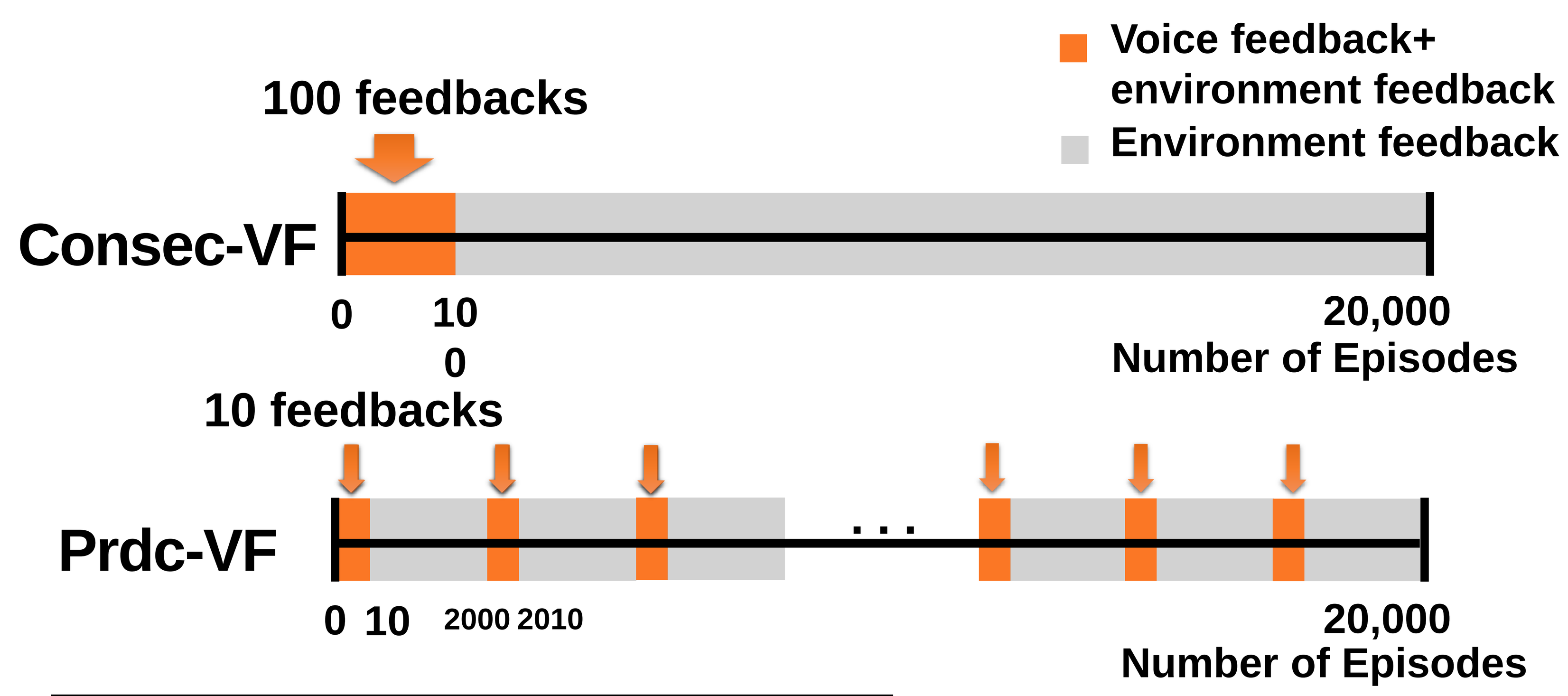
- 5 states and 5 actions are defined by the direction and degree of table movement and robot joint angle drive.
- State image taken by robot camera.
- After executing an action, the robot receives evaluative feedback from human on the action.
- The voice feedback is input via the robot's microphone, transcribed into character string by automatic speech recognition module.
- Then the transcribed feedback is converted to numeric value via sentiment analysis module.
- Converted feedback value is incorporated into the environmental rewards of the DQN algorithm.

"Well done" → 0.8

"That is not how you do it" → -0.699

"Try again" → -0.5

Experimental Settings



Type of actions	Reward
Reaching the target state	+0.5
Returning undefined action	-0.5
Reaching non-target states	-0.3

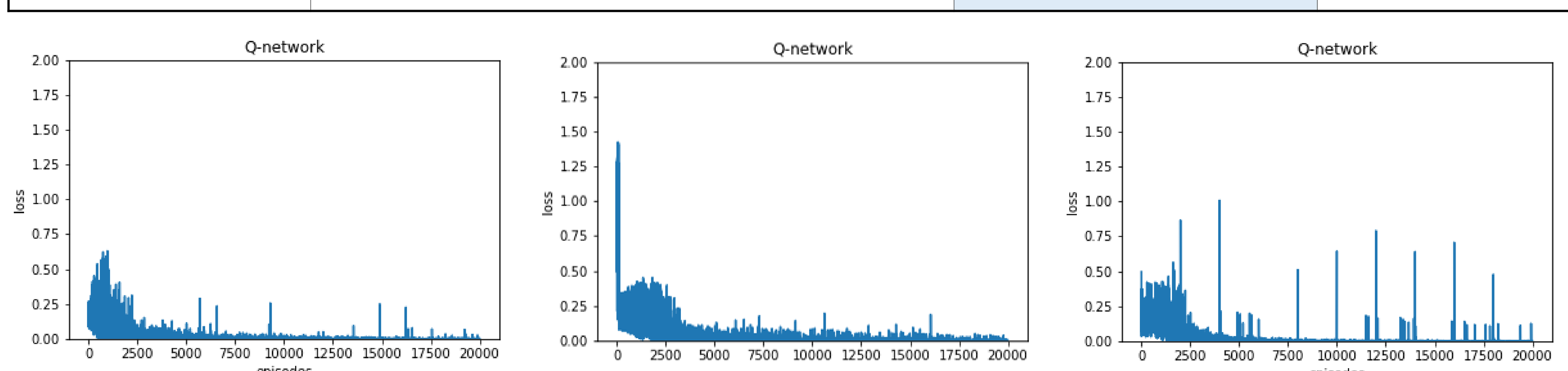
- Learning rate α=0.001
- Discount factor γ=0.9
- Epsilon ϵ=20
- Number of episodes: 20,000
- Number of voice feedbacks: 100 (50 positive and 50 negative)

- Training was mainly done in simulation, but also done on a physical robot as a proof of concept.

Results

Table of 3 tested model's optimal policy convergence rate

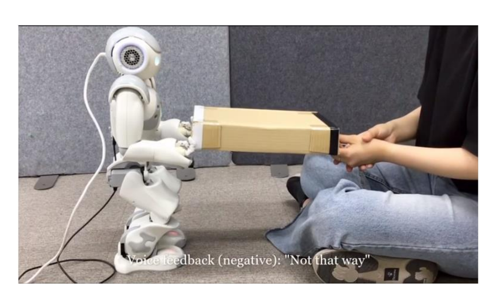
Optimizer	Only environment reward (baseline)	Consec-VF	Prdc-VF
SGD	80%	86%	80%
Adam	73%	96%	60%



(a) Baseline (b) Consec-VF (c) Prdc-VF

Loss graph of 3 tested models

Optimizer	Baseline	Consec-VF
SGD	80%	86%
Adam	73%	96%
Adagrad	43%	56%
Adadelta	63%	76%



Robot training video can be found at http://air.knu.ac.kr/index.php/evolutionary-cooperative-robot-development-using-distributed-deep-reinforcement-learning

- Consec-VF model learned optimal policies better than baseline and Prdc-VF model.
- In all experiments Consec-VF showed improved optimal policy learning compared to the baseline DQN.