The 5th International Conference on Cognitive Systems CogSys2012 February 22-23 2012, TU Vienna, Austria

Exploiting Past Success Failure for Effective and Robust Task Learning

John Nassour[1,2], Fathi Ben Ouezdou[2], and Gordon Cheng[1]
nassour@tum.de

(1) Institute for Cognitive Systems (ICS)
Technical University of Munich - Germany

(2) Laboratoire d'Ingénierie des Systèmes de Versailles (LISV)
University of Versailles - France

Objectives

-To propose a learning mechanism that is able to learn from negative and positive feedback with reward coding adaptively.

- To produce an early warning mechanism that can help to avoid repeating past errors in the generation of walking patterns of a humanoid robot. The notion of reward adaptation is introduced in order to qualify the walking task in term of energy.

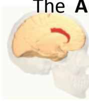
Bio Inspiration

The **OrbitoFrontal Cortex (OFC)**

-The key reward structure of the brain.

- Adaptivity in coding the reward according to the available rewards that changed in every block of trials.

The **Anterior Cingulate Cortex (ACC)**



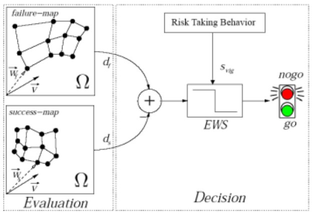
- Avoiding repeated mistakes.
- Early warning system
- External error feedback.
- Internal error detection.

OFC & ACC

are involved in cognitive decision-making process

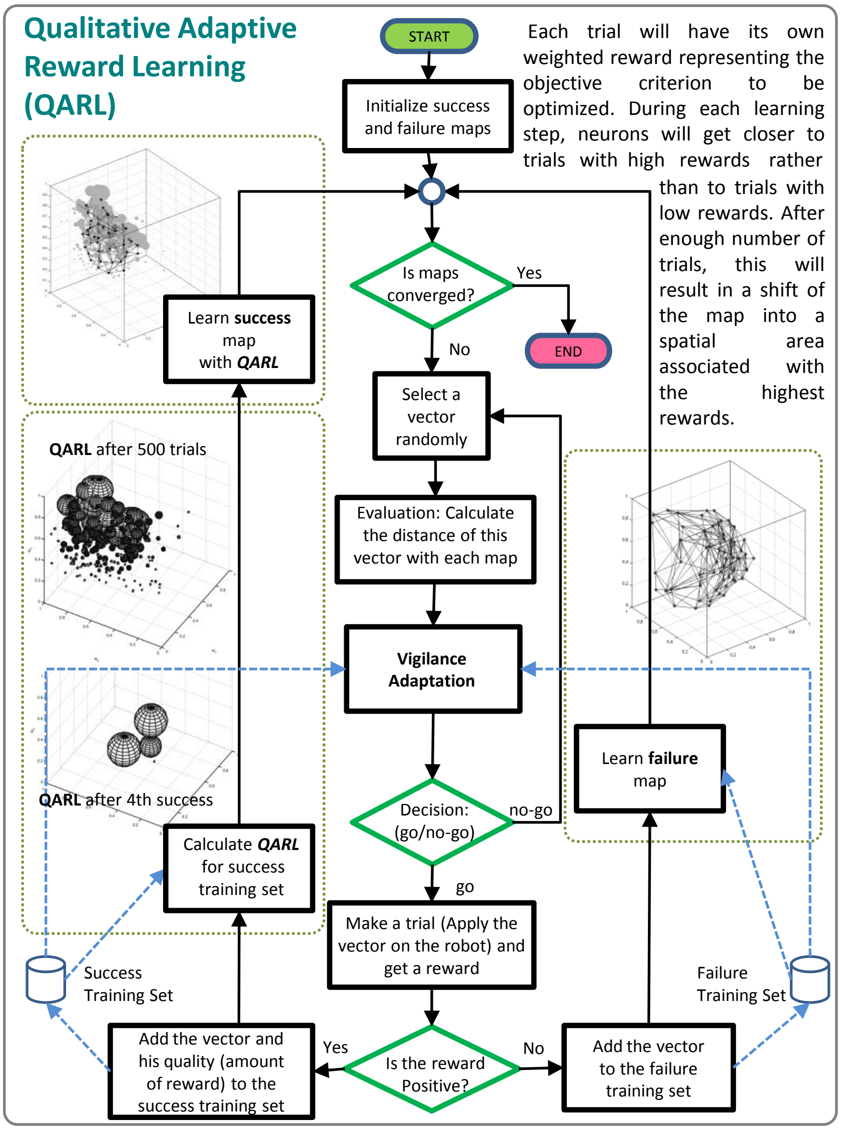
Success-Failure Learning

- Self Organizing Map
- Early Warning System



Qualitative Adaptive Reward Learning (QARL)

Each trial will have its own weighted reward representing the objective criterion to be optimized. During each learning step, neurons will get closer to trials with high rewards rather than to trials with low rewards. After enough number of trials, this will result in a shift of the map into a spatial area associated with the highest rewards.

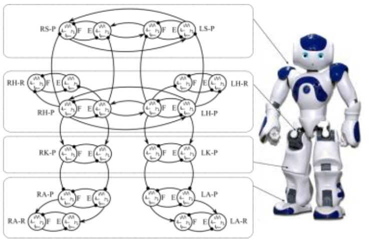


START
Initialize success and failure maps
Is maps converged?
Yes
END
No
Select a vector randomly
Evaluation: Calculate the distance of this vector with each map
Vigilance Adaptation
Decision: (go/no-go)
no-go
go
Make a trial (Apply the vector on the robot) and get a reward
Is the reward Positive?
Yes
No
Add the vector and his quality (amount of reward) to the success training set
Add the vector to the failure training set
Success Training Set
Failure Training Set
Calculate ***QARL*** for success training set
Learn **success** map with ***QARL***
Learn **failure** map
QARL after 500 trials
QARL after 4th success

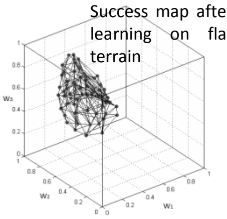
Walking Task

simultaneously learn and optimize walking

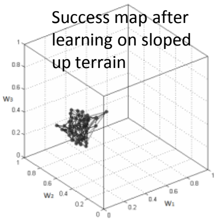
$$efficiency = \frac{mechanical\ work\ done}{metabolic\ energy\ consumed}$$



Learn Different Conditions



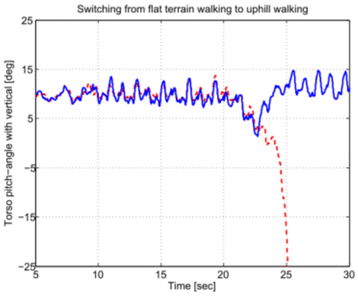
Success map after learning on flat terrain

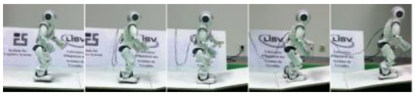


Success map after learning on sloped up terrain

Slope adaptation

When the torso pitch angle reach a pre-defined threshold, switching occurs gradually between neurons of success maps





Summary

We proposed a neurobiological inspired learning algorithm. The notion of qualitative adaptive reward was introduced in order to simultaneously learn and optimize. The objectives of the mechanism were to learn from mistakes and to avoid making them again. This was done by building on experiences of past mistakes and successes. We showed how these two experiences could build themselves through the stages of evaluation, decision and then trials. Learning succeeded trials with reward related walking efficiency make success map match trials where efficiency is high. It can be said that negative experiences is as importance as positive experiences.