Integrating Learning into a BDI Agent for Environments with Changing Dynamics

Dhirendra Singh$^1$  Sebastian Sardina$^1$  Lin Padgham$^1$  Geoff James$^2$

$^1$RMIT University, Melbourne, Australia  $^2$CSIRO Energy Technology, Sydney, Australia

Summary
This paper extends our earlier work integrating learning to improve plan selection in the popular Belief, Desire, Intentions (BDI) agent paradigm.

Here we address the problem that learning in deployed agents must be continuous rather than a one-off process.

Our main contribution is a novel confidence measure which allows the agent to adjust its reliance on the learning dynamically, facilitating in principle infinitely many (re)learning phases.

We demonstrate the benefits of the approach in an example battery controller for energy management.

BDI Learning Framework
Our learning framework augments plan’s context conditions with decision trees, allowing plan applicability to be learnt from experience.

Using a probabilistic plan selection function, the agent balances exploration and exploitation of plans.

Acting and learning are interleaved in an online manner, i.e., current learning influences ongoing choices that impact subsequent learning.

Confidence in Learning
We build confidence from observed performance of a plan by evaluating how well-informed were the recent decisions, or stability-based measure, and how well we know the worlds we are witnessing, or world-based measure.

Plan selection weight, that dictates exploration, is then calculated using the predicted likelihood of success and the dynamic confidence measure.

Modular Battery Controller
A building with local generation and loads is to restrict power consumption to a set range, using a modular battery system that can be charged or discharged as needed.

A programmed solution is not ideal since battery performance is susceptible to change over time.

We design a learning BDI controller that works to initial specification but also adapts to ongoing changes in the battery system.

Scenario 1: Recovery from deterioration in module capacities at 5k episodes.

Scenario 2: Recovery from individual module failures during [0, 20k], [20k, 40k] episodes.

Scenario 3: Recovery from complete system failure during [0, 5k] episodes.

The above experiments plot average success in configuring the battery correctly (y-axis) over the number of episodes (x-axis) for various changes in the environment dynamics.

D. Singh, S. Sardina, L. Padgham, G. James, Integrating Learning into a BDI Agent for Environments with Changing Dynamics. In Proceedings of International Joint Conference on Artificial Intelligence (IJCAI), Barcelona, Spain, 2011.