Towards a Physio-Cognitive Model of the Exploration Exploitation Trade-off

Department of Computer Science, Bucknell University, Lewisburg, Pa.

Introduction

What type of food do you order at a restaurant? Do you get your favorite meal that guarantees satisfaction or try something new on the menu which could potentially be better? This example demonstrates the essence of the exploration exploitation trade-off. It concerns how people use and acquire knowledge of their environment. It is the balance between utilizing current knowledge to perform a task and looking for more information to derive a better solution. Many researchers have studied this phenomenon from both symbolic and physiological levels. This project aims to link models from each category to create an in depth simulation of how the trade-off affects both the body and mind.

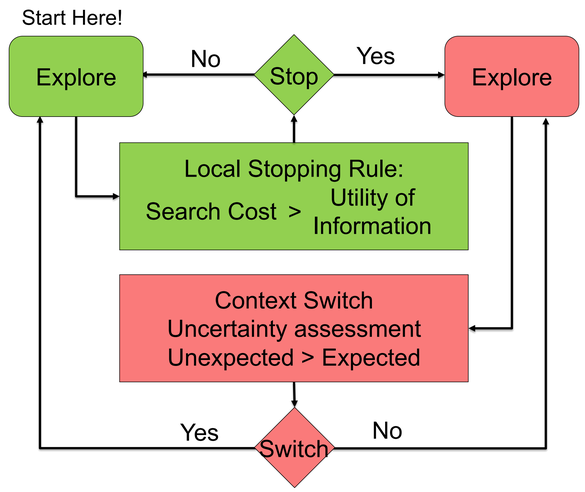
Theory

We have constructed a two part model:

1. **Satisficing base** (labelled in green) (Fu 2007)
   - Handles the transition from exploring to exploiting
   - Transition occurs when either:
     - Value of information exceeds cost of search
     - Continued search is too costly
2. **Context switching mechanism** (labelled in red)
   - Handles transition from exploiting to exploring
   - Based on assessments of expected and unexpected uncertainty
     - Represented by acetylcholine and norepinephrine (Yu and Dayan 2005).

Types of uncertainty:

- Expected – agent accounts for readily
- Unexpected – due to gross changes in the task



Body and Mind

We are using ACT-R/Φ (Dancy, 2013; Dancy et al., 2015) to develop our model. It is a hybrid architecture that bridges the ACT-R theory of cognition and HumMod physiological simulations.
This system enables our model to:

- Represent context switches with physiological parameters
- Show how modification to other hormones and parameters affects the management of the trade-off

The latter point refers to the linked properties of HumMod's hormone and parameter calculations. A change in one hormone may cause a residual effect that alters how the agent manages the trade-off. With ACT-R Φ we can model those effects.

Changing Behavior

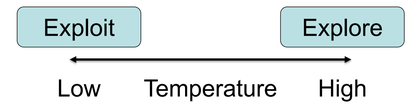
Alterations in decision making behavior occur by dynamically modifying the temperature in the model.

Productions (rules) in ACT-R are selected by their utility value.

The calculation factors in:

- Previous reinforcement
- How well the production suits the current situation
- Noise

The amount of noise is altered by the temperature.

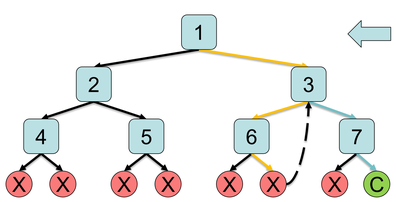


Experiment

The agent is being tested with a symbolic maze task (Fu and Anderson 2006). To show that ACT-R/Φ can be used in a complex environment, we built the maze in Project Malmo (Johnson et al., 2016), a modification to the game minecraft which allows artificial agents to be tested.

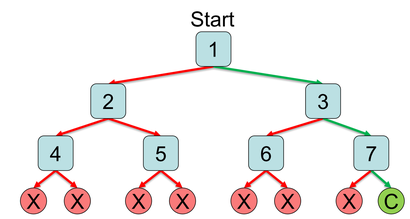
The task: complete the maze by making three correct decision.

The player receives feedback about their decisions when they find the exit or hit a dead end. It is up to them to track which decisions and associations are correct.



A path through the maze. Orange lines show path before reset, blue lines are after. The dotted line depicts the reset.

A room configuration



Structure of the maze. Boxes are rooms. Green lines show the correct path. Red X's are dead ends and the green C is the exit.

Future Work

Going forward we have to:

- Optimize the simulation
- Determine if using Malmo adjusts how people play
  - Reveals if the base model should be modified
- Add the context switching mechanism
- Gather data from subjects and the simulation
- Analyze the effectiveness of the model

Conclusion

We are developing an in depth model of the exploration exploitation trade-off using ACT-R/Φ. We have added a physiological controlled context switch to a satisficing model to better represent how the trade-off is managed by both the body and mind. We have chosen to use ACT-R/ Φ to make the physiological component of the model more thorough. Lastly, we are testing the model in the Project Malmo AI platform.

Acknowledgements

Special thanks to:

- Ryan Hope for developing the ACT-R JNI module.
- Son Pham for writing the code that enables ACT-R/Φ agents to be tested in Malmo

References

1. Dancy, C.L (2013) ACT-RΦ; A cognitive architecture with physiology and affect. Biologically Inspired Cognitive Architectures, 6(1), 40-45.
2. Dancy, C. L., Ritter, F. E., Berry, K. A., & Klein, L. C. (2015). Using a cognitive architecture with a physiological substrate to represent effects of a psychological stressor on cognition. Computational and Mathematical Organization Theory, 21(1), 90-114.
3. Fu, W. (2007). A Rational-Ecological Approach to the Exploration/Exploitation Trade-offs. In W.D. Gray (ed.), Integrated Models of Cognitive Systems (Vol. 1, pp 165-179). New York, NY: OUP.
4. Fu, W., & Anderson J. R. (2006). From recurrent choice to skill learning: A reinforcement-learning model. Journal of Experimental Psychology: General, 135(2), 184-206.
5. Johnson, M., Hofmann, K., Hutton, T., & Bignell, D. (2016). The Malmo platform for artificial intelligence experimentation. In proceedings of Twenty-Fifth International joint conference on artificial intelligence (IJCAI), New York, NY, 4246-4247.
6. Yu, A. J, & Dayan, P. (2005). Uncertainty, Neuromodulation, and Attention. Neuron, 46(4), 681-692.