An Exploring Study on Building Affective Artificial Intelligence by Neural-Symbolic Computing (Extended Abstract)

Jonathan C.H. Tong[1], **Yung-Fong Hsu**[2], **Churn-Jung Liau**[3]

[1]Department of International Business, National Taiwan University, Taipei 106, Taiwan
[2]Department of Psychology, National Taiwan University, Taipei 106, Taiwan
[3]Institute of Information Science, Academia Sinica, Taipei 115, Taiwan
d03323003@ntu.edu.tw, yfhsu@ntu.edu.tw, liaucj@iis.sinica.edu.tw

The goal-directed decision-making behavior of agents with natural intelligence is a self-organizing cognitive process. Based on preferences and memories, agents choose one of the most advantageous actions for achieving their goals among alternatives. The outcomes of these choices are then incorporated into the agent's memory through a kind of reinforcement learning process, forming the basis for judgments in the subsequent decision-making.

Under risky situations where the results of choices can lead to rewards or punishments, the entire cognitive process of natural intelligence becomes more complex, setting it apart significantly from artificial intelligence agents. First, natural intelligence agents might experience emotions before making a choice due to the uncertainty of rewards or punishments associated with the outcomes. Emotions may also arise post-choice based on whether the results align with or differ from expectations, and then embed into their learning and memory. Second, natural intelligence agents may have more than one thinking process, as suggested by the dual-process model in which intuitive and analytical processes may lean towards different, sometimes even conflict choices. Emotions might function as arbitrators in such cases.

The aim of our research is to construct an artificial intelligence model with inherent emotional functionality and the ability with dual-process thinking by Neural-Symbolic computing method. Such an artificial intelligence system would be more human-like in terms of thinking and behavior, thereby aiding in the development of Artificial General Intelligence (AGI) while also contributing to the study of human mental processes.

Our model is based on behavioral and brain imaging data collecting in a lottery gambling experiment. Both external behaviors and internal neural activities of humans demonstrate that a subjects current decision strongly depends on experiences of reward or punishment resulting from his/her prior decisions. Moreover, past experiences and the current risk context can result in distinct judgments and conflicts within dual decision-making processes, namely Process 1 and Process 2, and thus trigger the subjects emotional responses. Emotions play a crucial role in the formation and retrieval of human experiences, thereby exerting considerable influence on the dual process thinking and subsequent choice behavior. Hence, the Affective Artificial Intelligence agent developed in this study comprises a much human-like learning module with the interaction between emotions and experiences being embedded.

There are two major modules in our model. One is for decision and the other is for learning. We use Probabilistic Logic Programming (ProbLog) (Raedt and Kimmig 2015) to model the subjects' dual process thinking in the decision module and designate cognitive biases as the difference between their subjective probability judgments and the real probability. Meanwhile we couple the decision module with a learning module consisting of 2 neural networks where parameters of the former depend on outputs of the latter. Additionally, the study utilizes symbolic logics to incorporate emotions into the model based on the OCC theory (Adam, Herzig, and Longin 2009).

The two neural networks in the learning module are a Bayesian Neural Network (BNN) (Magris and Iosifidis 2023) and a Causal Reinforcement Learning (CRL) network (Gasse et al. 2021). The former forms and adjusts the beliefs of the AAI agent through the learning process, serving as the subjective probability input for constructing the decision module in ProbLog. The latter is based on a Convolutional Neural Network (CNN) to construct the Q-learning mechanism, maximizing reward or regret to learn the Structural Causal Model (SCM) by do-calculus and counterfactual inferences. It is then incorporated into the decision module as a meta-rule for the agent to choose applicable policies or rules during the decision-making phase. In addition, emotions triggered by past experiences and the current risk context will affect the thresholds in both BNN and CNN, thereby influencing the outputs of the learning module. Subsequently, these outputs will further impact the agent's decision-making behavior by modifying parameters of the agent's decision policy.

To summarize, this study is grounded in (i) empirical data on human decision-making behavior under risk, (ii) features such as dual process and emotions underpinning decision and learning, and (iii) the employment of appropriate theories and tools to build a computational model for human-like AAI agents decision and learning. It paves the way for the construction of a more human-like artificial intelligence model, thereby contributing to the understanding of human cognition and behavior, as well as the development of human-like machine learning.

Acknowledgments

This work is partially supported by National Science and Technology Council of Taiwan under Grants MOST 110-2221-E-001-022-MY3.

References

Adam, C.; Herzig, A.; and Longin, D. 2009. A logical formalization of the OCC theory of emotions. *Synthese*, 168(2): 201–248.

Gasse, M.; Grasset, D.; Gaudron, G.; and Oudeyer, P. 2021. Causal Reinforcement Learning using Observational and Interventional Data. *CoRR*, abs/2106.14421.

Magris, M.; and Iosifidis, A. 2023. Bayesian learning for neural networks: an algorithmic survey. *Artificial Intelligence Review*, 56(10): 11773–11823.

Raedt, L. D.; and Kimmig, A. 2015. Probabilistic (logic) programming concepts. *Machine Learning*, 100(1): 5–47.