A Dual Process Theory of Optimistic Cognition

Peter Sunehag (peter.sunehag@anu.edu.au) and Marcus Hutter, Research School of Computer Science, Australian National University.

1 Abstract

Optimism is a prevalent bias in human cognition including variations like self-serving beliefs, illusions of control and overly positive views of one's own future. Further, optimism has been linked with both success and happiness. In fact, it has been described as a part of human mental well-being which has otherwise been assumed to be about being connected to reality. In reality, only people suffering from depression are realistic. Here we study a formalization of optimism within a dual process framework and investigate its usefulness beyond human needs in a way that also applies to artificial reinforcement learning agents. Optimism enables systematic exploration which is essential in an (partially) unknown world. The key property of an optimistic agent is that it is not contradicted when one acts greedily with respect to it, then one is well-rewarded even if it were wrong.

4 Decision Functions

• A decision function \( f : \mathcal{M} \to A \) (is the set of finite classes of environments) only depending on a class of environments \( \mathcal{M} \).

• The decision function is independent of the history

• However, the class \( \mathcal{M} \) led to the decision function introduces an indirect dependence

\[
\pi \in \arg \max_{\pi} \sum t \in \mathcal{M}_t V^\pi_t
\]

A special case is when \( |\mathcal{M}| = 1 \) and (1) becomes

\[
\pi \in \arg \max \sum t \in \mathcal{M}_t V^\pi_t
\]

where \( \nu \) is the environment in \( \mathcal{M} \).

5 Hypothesis-Generating Functions

• A decision function \( f : \mathcal{M} \to A \) is defined as a hypothesis generating function \( f \) and a decision function \( \tilde{f} \) by choosing action \( f(\nu) \) after seeing history \( h \).

Example 1 Suppose that \( \mathcal{M} \) is a finite class of deterministic environments and let \( f(\nu) = \arg \max \{ \nu \in \mathcal{M} \} \). If we combine \( f \) with the optimistic decision function we have defined the optimistic agents for finite classes of deterministic environments. We here extend the analysis to infinite classes by letting \( f(\nu) \) contain new environments that were not in \( \mathcal{M} \).

Example 2 The Model Based Interval Estimation (MBIE) method for Markov Decision Processes (MDPs) defines \( f(\nu) \) as a set of MDPs (for a given state space) with transition probabilities in confidence intervals calculated from history. This is combined with the optimistic decision function.

7 Combining Laws into Environments

• We define a hypothesis generating function from a countable enumerated class \( \mathcal{M} \) based on a budget function for \( \pi \)-inconsistency that is increasing and unbounded.

• When the number of \( \pi \)-inconsistency points is below budget we introduce the next environment in the class.

• This form of hypothesis generating function enables bounds on the number of errors made by optimistic agents and it implements the intuition that the agent should not introduce more environments when the existing ones are very contradictory.

6 Agents in a Dual Process Framework

An agent, i.e. a function from histories to actions, is defined from a hypothesis generating function \( f \) and a decision function \( \tilde{f} \) by choosing action \( f(\nu) \) after seeing history \( h \).

Example 3 (Determinate laws for fixed vector). Consider an environment with a constant binary feature vector of length \( m \). There are \( 2^m \) such environments. Every such environment can be defined by combining \( m \) out of a class of \( 2^m \) laws. Each law says what the value of one of the features is, one law for \( 0 \) and one for \( 1 \). In this example, a coherent set of laws is simply one feature for each coefficient. The generated environment is the constant vector defined by that vector and the set of all the generated environments is the full set of \( 2^m \) environments.

8 Conclusions

• Optimism enables sufficient exploration for short-sighted agents to achieve optimality. Strict rationality fails to guarantee this.

• Viewing environments as combinations of laws can improve bounds exponentially.

• Outlook: Milder form better related to human's with Reward Modulated Inference as in reward-modulated spike-timing plasticity


Paper2: Optimistic Agents are Asymptotically Optimal, Peter Sunehag and Marcus Hutter, AUSAI 2012

Paper3: Optimistic AIO, Peter Sunehag and Marcus Hutter, AGI 2012

Paper4: Axioms for Rational Reinforcement Learning, Peter Sunehag and Marcus Hutter, ALT 2011