Lifelong Credit Assignment with the Success-Story Algorithm

Jürgen Schmidhuber
IDSIA, Galleria 2, 6928 Manno-Lugano, Switzerland
University of Lugano & SUPSI, Switzerland

Abstract

Consider an embedded agent with a self-modifying, Turing-equivalent policy that can change only through active self-modifications. How can we make sure that it learns to continually accelerate reward intake? Throughout its life the agent remains ready to undo any self-modification generated during any earlier point of its life, provided the reward per time since then has not increased, thus enforcing a lifelong success-story of self-modifications, each followed by long-term reward acceleration up to the present time. The stack-based method for enforcing this is called the success-story algorithm. It fully takes into account that early self-modifications set the stage for later ones (learning a learning algorithm), and automatically learns to extend self-evaluations until the collected reward statistics are reliable... a very simple but general method waiting to be re-discovered! Time permitting, I will also briefly discuss more recent mathematically optimal universal maximizers of lifelong reward, in particular, the fully self-referential Gödel machine.

Note: This is a summary of earlier work (Schmidhuber 1994; Schmidhuber, Zhao, and Schraudolph 1997; Schmidhuber 1987; 1993; Wiering and Schmidhuber 1996; Schmidhuber, Zhao, and Wiering 1997; Schmidhuber and Zhao 1997; Hutter 2005; Schmidhuber 2006; 2005; 2009).

References

Hutter, M. 2005. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Berlin: Springer. (On J. Schmidhuber's SNF grant 20-61847).

Schmidhuber, J., and Zhao, J. 1997. Multi-agent learning with the success-story algorithm. In Weiss, G., ed., *Distributed Artificial Intelligence meets Machine Learning*. Springer, Berlin. 82–93.

Schmidhuber, J.; Zhao, J.; and Schraudolph, N. 1997. Reinforcement learning with self-modifying policies. In Thrun, S., and Pratt, L., eds., *Learning to learn*. Kluwer. 293–309.

Schmidhuber, J.; Zhao, J.; and Wiering, M. 1997. Shifting inductive bias with success-story algorithm, adaptive Levin search, and incremental self-improvement. *Machine Learning* 28:105–130.

Schmidhuber, J. 1987. Evolutionary principles in self-referential learning, or on learning how to learn: the meta-meta-... hook. Institut für Informatik, Technische Universität München. http://www.idsia.ch/~juergen/diploma.html.

Schmidhuber, J. 1993. A self-referential weight matrix. In *Proceedings of the International Conference on Artificial Neural Networks, Amsterdam*, 446–451. Springer.

Schmidhuber, J. 1994. On learning how to learn learning strategies. Technical Report FKI-198-94, Fakultät für Informatik, Technische Universität München. See (Schmidhuber, Zhao, and Wiering 1997; Schmidhuber, Zhao, and Schraudolph 1997).

Schmidhuber, J. 2005. Completely self-referential optimal reinforcement learners. In Duch, W.; Kacprzyk, J.; Oja, E.; and Zadrozny, S., eds., *Artificial Neural Networks: Biological Inspirations - ICANN 2005, LNCS 3697*, 223–233. Springer-Verlag Berlin Heidelberg. Plenary talk.

Schmidhuber, J. 2006. Gödel machines: Fully self-referential optimal universal self-improvers. In Goertzel, B., and Pennachin, C., eds., *Artificial General Intelligence*. Springer Verlag. 199–226. Variant available as arXiv:cs.LO/0309048.

Schmidhuber, J. 2009. Ultimate cognition *à la* Gödel. *Cognitive Computation* 1(2):177–193.

Wiering, M., and Schmidhuber, J. 1996. Solving POMDPs with Levin search and EIRA. In Saitta, L., ed., *Machine Learning: Proceedings of the Thirteenth International Conference*, 534–542. Morgan Kaufmann Publishers, San Francisco, CA.