Universal RL: applications and approximations

Joel Veness
Adjunct, University of Alberta, Edmonton, Canada
veness@cs.ualberta.ca

July 17, 2013

Abstract

While the main ideas underlying Universal RL have existed for over a decade now (see [Hutter, 2012] for historical context), practical applications are only just starting to emerge. In particular, the direct approximation introduced by Veness et al. [2010, 2011] was shown empirically to compare favorably to a number of other model-based RL techniques on small, partially observable environments with initially unknown, stochastic dynamics. Since then, a variety of additional techniques have been introduced that allow for the construction of far more sophisticated approximations. This short paper collects together and reviews some of the main ideas that have the potential to lead to larger scale applications.

1 Introduction

Within universal reinforcement learning, some recent efforts [Veness et al., 2010, 2011] have been made towards approximating AIXI [Hutter, 2005], a Bayesian optimality notion for general reinforcement learning agents. Impressively, these agents have been shown to be able to learn, *from scratch*, to play TicTacToe, Pacman, Kuhn Poker, and other simple games by trial and error alone – even the rules of each game were not communicated to the agent. The mathematical framework used in these aforementioned works is a natural generalization of the statistical data compression setting to reinforcement learning. The distinguishing feature of this setting is an extra source of side information – namely, the history of actions chosen by some control algorithm – which is incorporated into a sequential, probabilistic framework that makes only minimal assumptions. This generality provides considerable flexibility when constructing models of the environment, which arguably makes Universal RL a leading candidate to complement the more restricted (PO)MDP formulations that have thus far driven the majority of model based reinforcement learning efforts. For this effort to be considered successful however, larger scale applications need to be demonstrated. We now provide a brief survey of some recent work relevant to this goal.

Handling large observation spaces via model averaging over factorizations. One promising approach for scaling model-based reinforcement learning to larger observation spaces is to consider various kinds of factored models of the environment. A good choice of factorization can make the model learning task much easier by decomposing it

into a number of more manageable sub-problems. Unfortunately, the success of this approach typically depends heavily on the particular choice of factorization. Bellemare et al. [2013] recently introduced a general technique that efficiently performs exact Bayesian model averaging over a large class of possible factorizations. This method comes with strong competitive guarantees with respect to the best factorization in its class, and was successfully demonstrated on an observation space consisting of 210×160 7-bit pixels.

Ensemble methods. Another promising approach for systematically improving the range and scope of existing model learning algorithms is to combine multiple such methods using various ensemble techniques. Veness et al. [2012b] discuss three such methods for the Universal RL setting, all of which come with relatively strong theoretical guarantees. These techniques provide a principled way to scale up existing approximations without adversely affecting the agent's existing abilities.

Efficient algorithms for non-stationary environments. The recent Context Tree Switching [Veness et al., 2012a] method, when combined with an appropriate base model (for example, see [Veness et al., 2013]), is one such example of a universal model for piecewise stationary k-Markov sources. This method can be generalized to the Universal RL setting the same way that the Context Tree Weighting algorithm [Willems et al., 1995] was generalized to the active case in [Veness et al., 2011]. These models lead to agents that are robust to abrupt changes in the environment, and form a key component of the Atari modeling work of Bellemare et al. [2013].

Efficient models for sparse memoryless sources. Veness and Hutter [2012] and more recently Hutter [2013], proposed principled memoryless models that can efficiently handle large alphabets in cases where only a small subset of symbols from the original alphabet are expected to occur. These models constitute powerful building blocks for more sophisticated model based reinforcement learning agents.

Assessing progress empirically. When constructing agents to work across a wide range of environments, it is important to have some empirical measure of success. The Universal RL theory leads to a natural definition of agent intelligence, which is approximated in [Legg and Veness, 2011] to rank a number of simple agents. Perhaps more interestingly, the publicly available Arcade Learning Environment [Bellemare et al., 2012] provides an interesting set of challenge problems for existing RL algorithms.

Miscellaneous. Effective exploration within a setting as general as Universal RL is extremely challenging, to say the least. That said, recent theoretical work by Lattimore et al. [2013] suggests that the situation may not be entirely hopeless. Furthermore, optimistic variants of the AIXI optimality notion are discussed in [Sunehag and Hutter, 2012a,b]; these may also prove interesting to approximate in restricted settings, and may be less sensitive to issues of exploration. In terms of ensemble methods, another promising model combination technique that needs further investigation in the RL context is geometric mixing, which was recently shown by Mattern [2012, 2013] to be a key component of the well-known PAQ [Mahoney, 2005] family of data compressors. Alternative measures of agent intelligence have also been proposed [Dowe and Hajek, 1998, Hernández-Orallo and Dowe, 2010], which may be interesting to explore in the future.

References

M. G. Bellemare, Y. Naddaf, J. Veness, and M. Bowling. The arcade learning environment: An evaluation platform for general agents. *CoRR*, abs/1207.4708, 2012.

M. G. Bellemare, J. Veness, and M. Bowling. Bayesian learning of recursively factored environments. *Proceedings of the International Conference on Machine Learning (ICML)*, 2013.

D. L. Dowe and A. R. Hajek. A non-behavioural, computational extension to the Turing Test. In *Intl. Conf. on Computational Intelligence & multimedia applications (ICCIMA'98), Gippsland, Australia*, pages 101–106, February 1998.

J. Hernández-Orallo and D. L. Dowe. Measuring universal intelligence: Towards an anytime intelligence test. *Artificial Intelligence*, 174(18):1508 – 1539, 2010. ISSN 0004-3702. doi: DOI: 10.1016/j.artint.2010.09.006. URL `http://www.sciencedirect.com/science/article/pii/S0004370210001554`.

M. Hutter. *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*. Springer, 2005.

M. Hutter. One decade of universal artificial intelligence. In *Theoretical Foundations of Artificial General Intelligence*, pages 67–88. Atlantis Press, 2012.

M. Hutter. Sparse adaptive Dirichlet-multinomial-like processes. *COLT*, 2013.

T. Lattimore, M. Hutter, and P. Sunehag. The sample complexity of general reinforcement learning. In *International Conference on Machine Learning*. 2013.

S. Legg and J. Veness. An Approximation of the Universal Intelligence Measure. *CoRR*, abs/1109.5951, 2011.

M. Mahoney. Adaptive weighing of context models for lossless data compression. Technical report, Florida Institute of Technology, 2005.

C. Mattern. Mixing strategies in data compression. In *DCC*, pages 337–346, 2012.

C. Mattern. Linear and geometric mixtures - analysis. In *DCC*, pages 301–310, 2013.

P. Sunehag and M. Hutter. Optimistic AIXI. In *Proc. 5th Conf. on Artificial General Intelligence (AGI'12)*, volume 7716 of *LNAI*, pages 312–321. Springer, Heidelberg, 2012a.

P. Sunehag and M. Hutter. Optimistic agents are asymptotically optimal. In *Proc. 25th Australasian Joint Conference on Artificial Intelligence (AusAI'12)*, volume 7691 of *LNAI*, pages 15–26, Sydney, Australia, 2012b. Springer.

J. Veness and M. Hutter. Sparse Sequential Dirichlet Coding. *CoRR*, abs/1206.3618, 2012.

J. Veness, K. S. Ng, M. Hutter, and D. Silver. Reinforcement Learning via AIXI Approximation. In *AAAI*, 2010.

J. Veness, K. S. Ng, M. Hutter, W. T. B. Uther, and D. Silver. A Monte-Carlo AIXI Approximation. *Journal of Artificial Intelligence Research (JAIR)*, 40:95–142, 2011.

J. Veness, K. S. Ng, M. Hutter, and M. H. Bowling. Context Tree Switching. In *DCC*, pages 327–336, 2012a.

J. Veness, P. Sunehag, and M. Hutter. On Ensemble Techniques for AIXI Approximation. In *AGI*, pages 341–351, 2012b.

J. Veness, M. White, M. Bowling, and A. György. Partition Tree Weighting. In *DCC*, pages 321–330, 2013.

F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens. The Context Tree Weighting Method: Basic Properties. *IEEE Transactions on Information Theory*, 41:653–664, 1995.