Bold Hearts 2006

Using Universal Guiding Principles for Developing Agent Skills

Chin Foo Wong and Daniel Polani

Adaptive Systems and Algorithms Research Group
Dept. of Computer Science
University of Hertfordshire, UK

Abstract. In earlier RoboCup teams, we have made a point of using principled methods to learn behaviours. The main line of these methods is the use of information-theoretic and related algorithms to identify the structure of possible behaviours for an agent. In this endeavour, we strive to move further ahead, using some further new concepts.
In particular, we will use, in addition to existing information-theoretic methods described last year, the concept of *empowerment* [18, 19] to select desirable directions of activity.

1 Introduction

Our program of research concentrates on creating agent control from first principles. A particular interest is to avoid having to know the detailed physics of the world they live in. We thus wish to move away from the standard approach to construct strong RoboCup agents by creating specific skills and capabilities. Our research program in the last years has been devoted to develop techniques that allow self-organized emergence of control for artificial agents in very general settings [13, 14, 15, 16, 17, 18, 19, 25] and in real robots (AIBOs) [26, 27, 28, 29, 30, 31, 32, 33, 34].

Having been part of the RoboCup endeavour for a long time, *learning* has been introduced as early as [39, 40], but also in combination with reinforcement learning [6, 21]. As we argued in [10], reinforcement learning methods are attractive for learning approaches because they are highly general, mathematically accessible and well understood. This generality, however, comes at a price. In large search spaces, the learning algorithms are slow and their robustness and generalizability is not well controlled. Dedicated decompositions of the representation of the state space are sometimes performed that deconstruct the task hierarchically into manageable parts [9], and this still mostly requires manual decomposition. Only recently, approaches begin to emerge that show promising ways to reduce reinforcement learning complexity without human introspection [11, 12]. Still, a large number of learning steps is required to learn a more complex task. In addition, convergence problems can arise in continuous domains (as RoboCup) [41].

In the last teams, beginning with [10], a different approach had been used. It introduced SIVE, which was inspired by many different sources. Its original motivation stems from the observation that humans are able to attain a much steeper learning performance than computers when faced with a new task. As mentioned in [10], when facing an autonomous agent team, a human team playing with the OpenZeng interface at the GermanOpen 2001, while being technically and tactically inferior, showed a rapidly improving performance and thus a much steeper learning curve than any current available learning system.

This is particularly striking since human accuracy in estimating ball position and performing actions was nowhere as accurate as that of the autonomous team. This is a clear indication that the "exhaustive learning" character exhibited by typical automated learning algorithms is inadequate to obtain the directedness and generalization power that human learning exhibits. Human learning exhibits extremely fast generalization and adaptation, "holistic" learning and the capability to combine skills. To achieve this flexibility is our ultimate goal.

2 Information and Intelligent Agents

We emphasized in the past the importance of information theoretic approaches in understanding how biological systems achieve their goals. Barlow's and Linsker's results [4, 23] about the self-organization of perception based on principles from information theory (information maximization) have found a larger number of approaches to study the problem of information processing in the brain (or of biological systems in general) using information theory [1, 3, 5, 7, 24, 46]. The limited power of these approaches was due to the lack of one essential element: these studies concentrated on passive systems, systems that would take in information as it comes along, but that had no power of action.

Apart from an early insight by [2], only recent research has begun to include the influence of agent actions on the information balance of a system under consideration [43, 44]. In addition It could be recently shown that optimizing information-flows in the closed perception-action loop of an agent with given embodiment and limited informational resources acts as a self-organization process for information flows; the way information propagates through the system [15, 17, 25] organizes itself as to represent "essential" features of the environment. Using this principle, virtually no extra assumptions are introduced into the system beyond the natural embodiment of the agent and the requirement of information flow optimization, something that can be naturally defined (even if not necessarily to compute) for any type of agent. The observation that the limitation of resources can force (Shannon) information to "crystallize" into meaningful structures that capture essential properties of a system, has found its probably most striking incarnation in the *information bottleneck principle* [38, 42].

Manifold representation can recover aspects of the environment using [35], but this is no guarantee for continuity or symmetry that can be exploited. If a system is not completely lacking structure, than one can, however, always hope to identify informational structures; for instance one can infer sensomotoric maps from entirely uninterpreted sensoric input [30, 31, 32, 36] for a sensomotoric system as complex as real AIBO robot.

Why is information such a useful quantity? One reason is that it is universal — any exchange of data, no matter whether in artificial or in biological systems, is subjected to Shannon's laws. The second is that, in absence of other costs (or if variation keeps other costs unchanged), biological systems tend to exploit the available “information space” to its limits [8, 45]. On the one hand, information processing is “convertible” into other basic biological currencies, e.g. ATP consumption [22], on the other hand, it provides tools by which information can be treated *directly* as a quantifiable, limited resource.

Including with the insights of the self-organized structuring of information from the information bottleneck principle [42] as well as from the information-flow studies [15, 17], this leads us to the hypothesis that we might be able to gain access to some of the principles that underlie biological information processing by understanding what happens in terms of information flows instead of trying to reverse engineer the concrete biological implementation in detail. This hypothesis that the principle could act as an approach for developing AI systems that capture some of the spirit of living learning systems is the main motivation for our overall methodology.

3 Empowerment as Guiding Principle

We wish to further increase our tools in the bottom-up creation of skills. The approach which we attempt to incorporate into our team learning strategy is the concept of *empowerment* (introduced in [18, 19]).

One of the central problems of teaching skills to autonomous agents is the construction of a suitable reward function that will spread over the whole state/action space. Often, unless heavy human intervention is involved, true autonomous agent scenarios have only a very sparse reinforcement feedback which takes long time to spread back over a system. A special problem is also the identification of salient and interesting features.

The concept of empowerment had been introduced to alleviate that problem. Empowerment is basically a measure of how strongly an agent can influence its environment. It is calculated using information theory (see above). What it achieves is that it constructs a dense utility function throughout a perception-action space; it thereby solves the sparsity problem.

In addition, in a number of scenarios we have found that empowerment has the ability to identify “interesting”, salient points in the state/action landscape. In our 2006 team, we wish to employ empowerment as guiding principle to find interesting states to train by the agents. This will not just be applied to individual agent skills as ball control (something for which empowerment promises to be very well suited).

A key issue will be to control empowerment with respect to teammates *and* to opponent agents. This way, we can create a collective utility function for the team (it has to work using the information of the individual agent, though, and thus will only have a subjective view of the world). Using the empowerment measure will (in conjunction with applying a minimax principle for the opponent agents) serve us to create as far as possible a bottom-up strategy for the interaction of the agents. The technique for applying a minimax principle has already been experimented with in the Lucky Lbeck 2000 team; this time, however, we also have a suitable dense utility function that may help to improve the quality of the found solutions.

One problem that still has to be solved at this time is the incorporation of explicit higher level goals. For this, a balance between local empowerment optimization and global goals will have to be constructed. This is, at the current point, not solved in a conceptually consistent way, but can, in any case, be achieved by incorporating the reward for achieving a goal (the ultimate task of the agent, that, however, cannot be easily seen by the pure empowerment measure which can only identify points of interest via the agent embodiment) explicitly into the calculation — at this point, it will still require direct human intervention.

4 Acknowledgements

We wish to thank the UvA team for providing the basis code (UvA 2002) [20] for the 2D team in which the SIVE predecessor of the current method for training the 3D team was based. In addition, we would like to thank Joschka Boedecker, Oliver Obst, Pat Riley, Tom Howard and all the other developers of the 3D simulator; their merit in advancing the Simulation League and RoboCup itself is appreciated to the highest degree.

Bibliography

[1] Amari, S., Cichocki, A., and Yang, H., [1996]. A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems*, vol. 8, 757–763. Cambridge, MA: MIT Press.

[2] Ashby, W. R., [1952]. *Design for a Brain*. New York: Wiley & Sons.

[3] Baddeley, R., Hancock, P., and Földiák, P., editors, [2000]. *Information Theory and the Brain*. Cambridge University Press.

[4] Barlow, H. B., [1989]. Unsupervised Learning. *Neural Computation*, 1:295–311.

[5] Becker, S., [1996]. Mutual Information Maximization: Models of Cortical Self-Organization. *Network: Computation in Neural Systems*, 7:7–31.

[6] Buck, S., and Riedmiller, M., [2000]. Learning situation dependent sucess rates of actions in a robocup scenario. In *Proceedings of PRICAI '00, Melbourne, Australia, 28.8.-3.9.2000*, 809.

[7] Comon, P., [1991]. Independent Component Analysis. In *Proc. Intl. Signal Processing Workshop on Higher-order Statistics, Chamrousse, France*, 111–120.

[8] de Ruyter van Steveninck, R. R., and Laughlin, S. B., [1996]. The Rate of Information Transfer at Graded-Potential Synapses. *Nature*, 379:642–645.

[9] Dietterich, T. G., [2000]. Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research*, 13:227–303.

[10] Franco, S., and Polani, D., [2004]. Skill Learning Via Information-Theoretical Decomposition of Behaviour Features. In Polani, D., Browning, B., Bonarini, A., and Yoshida, K., editors, *RoboCup 2003: Robot Soccer World Cup VII*, vol. 3020 of *LNCS*. Springer. Team Description (CD supplement).

[11] Goel, S., and Huber, M., [2003]. Subgoal Discovery for Hierarchical Reinforcement Learning Using Learned Policies. In *Proceedings of the 16th International FLAIRS Conference, St. Augustine, FL*, 346–350.

[12] Jacob, D., Polani, D., and Nehaniv, C. L., [2004]. Improving Learning for Embodied Agents in Dynamic Environments by State Factorisation. In *TAROS 2004, Towards Autonomous Robotic Systems, September 6th - 8th, 2004 University of Essex, Colchester, UK*.

[13] Klyubin, A., Polani, D., and Nehaniv, C., [2005]. Representations of Space and Time in the Maximization of Information Flow in the Perception-Action Loop. *Neural Computation*. Submitted.

[14] Klyubin, A., Polani, D., and Nehaniv, C. L., [2005]. Decomposition of Information Flows. In preparation.

[15] Klyubin, A. S., Polani, D., and Nehaniv, C. L., [2004]. Organization of the Information Flow in the Perception-Action Loop of Evolved Agents. In *Proceedings of 2004 NASA/DoD Conference on Evolvable Hardware*, 177–180. IEEE Computer Society.

[16] Klyubin, A. S., Polani, D., and Nehaniv, C. L., [2004]. Organization of the Information Flow in the Perception-Action Loop of Evolved Agents. Computer

Science Technical Report 400, Faculty of Engineering and Information Sciences, University of Hertfordshire. January.

[17] Klyubin, A. S., Polani, D., and Nehaniv, C. L., [2004]. Tracking Information Flow through the Environment: Simple Cases of Stigmergy. In [37], 563–568. Available as Technical Report No. 402, Department of Computer Science, Faculty of Engineering and Information Sciences, University of Hertfordshire.

[18] Klyubin, A. S., Polani, D., and Nehaniv, C. L., [2005]. All Else Being Equal Be Empowered. In *Advances in Artficial Life, European Conference on Artificial Life (ECAL 2005)*, vol. 3630 of *LNAI*, 744–753. Springer.

[19] Klyubin, A. S., Polani, D., and Nehaniv, C. L., [2005]. Empowerment: A Universal Agent-Centric Measure of Control. In *Proc. IEEE Congress on Evolutionary Computation, 2-5 September 2005, Edinburgh, Scotland (CEC 2005)*, 128–135. IEEE.

[20] Kok, J., and de Boer, R., [2002]. UvA Trilearn. Software.
`http://carol.wins.uva.nl/ jellekok/robocup/`, October 2003

[21] Lauer, M., and Riedmiller, M., [2000]. An Algorithm for Distributed Reinforcement Learning in Cooperative Multi-Agent Systems. In *Proc. 17th International Conf. on Machine Learning*, 535–542. Morgan Kaufmann, San Francisco, CA.

[22] Laughlin, S. B., de Ruyter van Steveninck, R. R., and Anderson, J. C., [1998]. The metabolic cost of neural information. *Nature Neuroscience*, 1(1):36–41.

[23] Linsker, R., [1988]. Self-Organization in a Perceptual Network. *Computer*, 21(3):105–117.

[24] Luttrell, S., [1989]. Self-Organization: a derivation from first principles of a class of learning algorithms. In *Proceedings 3rd IEEE Int. Joint Conf. on Neural Networks*, vol. 2, 495–498. IEEE Neural Networks Council, Washington.

[25] Nehaniv, C. L., Polani, D., Olsson, L. A., and Klyubin, A., [2005]. Evolutionary Information-Theoretic Foundations of Sensory Ecology: Channels of Organism-Specific Meaningful Information. In da Fontoura Costa, L., and Müller, G. B., editors, *Modeling Biology: Structures, Behaviour, Evolution*, Vienna Series in Theoretical Biology. MIT press. Invited lecture at the 10th Altenberg Workshop in Theoretical Biology, July 9-11, 2004, Konrad Lorenz Institute for Evolution and Cognition Research, Altenberg, Austria. Submitted.

[26] Olsson, L., Nehaniv, C., and Polani, D., [2005]. Discovering Motion Flow by Temporal-Informational Correlations in Sensors. In Berthouze, L., Kaplan, F., Kozima, H., Yano, H., Konczak, J., Metta, G., Nadel, J., Sandini, G., Stojanov, G., and Balkenius, C., editors, *Fifth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems (EpiRob2005)*, vol. 123, 117–120.

[27] Olsson, L., Nehaniv, C., and Polani, D., [2005]. From Unknown Sensors and Actuators to Visually Guided Movement. In *4th IEEE International Conference on Development and Learning (ICDL-05)*, 1–6. IEEE Computer Society Press.

[28] Olsson, L., Nehaniv, C., and Polani, D., [2005]. Sensor Adaptation and Development in Robots by Entropy Maximization of Sensory Data. In *IEEE Computational Intelligence in Robotics & Automata (IEEE CIRA'05), special session on Ontogenetic Robotics, Espoo, Finland*, 587–592.

[29] Olsson, L., Nehaniv, C., and Polani, D., [2006]. Measuring Informational Distances Between Sensors and Sensor Integration. In M.Rocha, L., Bedau, M., Floreano, D., Goldstone, R., Vespignani, A., and Yaeger, L., editors, *Proc. Artificial Life X*. (In Press).

[30] Olsson, L., Nehaniv, C. L., and Polani, D., [2004]. The Effects on Visual Information in a Robot in Environments with Oriented Contours. In Berthouze, L., Kozima, H., Prince, C. G., Sandini, G., Stojanov, G., Metta, G., , and Balkenius, C., editors, *Proceedings of the Fourth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, August 25-27th Genoa, Italy*, 83–88. Lund University Cognitive Studies.

[31] Olsson, L., Nehaniv, C. L., and Polani, D., [2004]. Information Trade-Offs and the Evolution of Sensory Layouts. In [37]. Available as University of Hertfordshire Technical Report 403, Feb. 2004.

[32] Olsson, L., Nehaniv, C. L., and Polani, D., [2004]. Sensory Channel Grouping and Structure from Uninterpreted Sensor Data. In *IEEE NASA/DoD Conference on Evolvable Hardware 2004*, 153–160. IEEE Computer Society.

[33] Olsson, L., Nehaniv, C. L., and Polani, D., [2005]. Measuring Informational Distances Between Sensors and Sensor Integration. Technical Report 431, University of Hertfordshire.

[34] Olsson, L., Nehaniv, C. L., and Polani, D., [2006]. From Unknown Sensors and Actuators to Actions Grounded in Sensorimotor Perceptions. *Connection Science*, 18(2). Special Issue on Developmental Robotics, Douglas Blank and Lisa Meeden, editors. (In Press).

[35] Philipona, D., O'Regan, K., and Nadal, J.-P., [2003]. Is there something out there? Infering space from sensorimotor dependencies. *Neural Computation*, 15(9):2029–2049.

[36] Pierce, D., and Kuipers, B., [1997]. Map learning with uninterpreted sensors and effectors. *Artificial Intelligence Journal*, 92:169–229.

[37] Pollack, J., Bedau, M., Husbands, P., Ikegami, T., and Watson, R. A., editors, [2004]. *Artificial Life IX: Proceedings of the Ninth International Conference on Artificial Life*. MIT Press.

[38] Slonim, N., Friedman, N., , and Tishby, T., [2001]. Agglomerative Multivariate Information Bottleneck. In *Neural Information Processing Systems (NIPS 01)*.

[39] Stone, P., [2000]. *Layered Learning in Multiagent Systems: A Winning Approach to Robotic Soccer.* MIT Press.

[40] Stone, P., and Veloso, M., [1998]. A layered approach to learning client behaviors in the RoboCup soccer server. *Applied Artificial Intelligence*, 12.

[41] Sutton, R. S., and Barto, A. G., [1998]. *Reinforcement Learning*. Cambridge, Mass.: MIT Press.

[42] Tishby, N., Pereira, F. C., and Bialek, W., [1999]. The Information Bottleneck Method. In *Proc. 37th Annual Allerton Conference on Communication, Control and Computing, Illinois*.

[43] Touchette, H., and Lloyd, S., [2000]. Information-Theoretic Limits of Control. *Phys. Rev. Lett.*, 84:1156.

[44] Touchette, H., and Lloyd, S., [2004]. Information-theoretic approach to the study of control systems. *Physica A*, 331:140–172.

[45] van Hateren, J. H., [1992]. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *J.Comp.Physiol. A*, 171:157–170.

[46] Van Hulle, M. M., [1996]. Topographic map formation by maximizing unconditional entropy: a plausible strategy for 'online' unsupervised competitive learning and nonparametric density estimation. *IEEE Transactions on Neural Networks*, 7(5):1299–305.