Bold Hearts 2005

Building up Skills for the 3D scenario using Minimal Interaction Scenarios and an Information Parsimony Principle

Chaohua Zhu and Daniel Polani

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

Abstract. The design of Bold Hearts 2005 (3D) is motivated by the SIVE approach used in [12, 13], but expands it by developing a consistent build-up methodology for skills, based on bottom-up pattern identification restricted by information parsimony principles. The idea is to learn skills carrying a limited amount of complexity (in terms of required information processing resources, measurable with Shannon information) and a hierarchical buildup of skills from these principles.
This methodology naturally extends outside the RoboCup scenario and will serve as a basis for further systematic research in buildup of competence and skills for AI systems. A first implementation shows very encouraging results.

1 Introduction

The standard approach to construct strong RoboCup agents focuses on creating specific skills and capabilities. These are often constructed by the programmer with specific knowledge about properties of the world physics as simulated by the soccer server. This explicit knowledge by the developer limits the flexibility and robustness of the system in case of world changes, and, to some degree, it also defeats the original aim of Artificial Intelligence, where the main onus is on the system to learn how to operate in a given environment. It is needless to say that RoboCup is a most prominent example for scenarios where this latter perspective would be desirable — and, as we believe, achievable.

Having been part of the RoboCup endeavour for a long time, *learning* has been introduced as early as [31, 32], but also in combination with reinforcement learning [5, 20]. As we argued in [13], 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. To alleviate that, dedicated decompositions of the representation of the state space have to be performed that deconstruct the task hierarchically into manageable parts [10], 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 [14, 15]. 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) [33].

In [13], 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 [13], 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.

Note that here the 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. In SIVE, we desired to mimic some of the properties exhibited by human learning: extremely fast generalization and adaptation, "holistic" learning and the capability to combine skills. For this purpose, the SIVE method had been introduced combining ideas and approaches from different areas, in particular from pattern matching and information theory.

This year, we go a step further and systematically unroll the requirements that we believe are fundamental ingredients to information processing required by agents.

2 Information and its Role in Control

Since Barlow's and Linsker's important results [3, 22] about the self-organization of perception based on principles from information theory and the universal information quantities discovered by Shannon [29] and related to the fundamental physical quantity of entropy [16, 39], there have been repeated approaches to study the problem of information processing in the brain (or of biological systems in general) using information theory [1, 2, 4, 7, 23, 38]. The power of these approaches was, however, limited, because one essential element was missing: all these studies were concentrating on passive systems, systems that would take in information as it comes along, but that had no power to modify their universe.

Only recent research has begun to include the influence of agent actions on the information balance of a system under consideration [35, 36]. This has far-reaching consequences. In fact, 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 [17, 18, 24] 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* [30, 34].

While recovering aspects of the environment can be done using manifold representation [8], in general, there is no guarantee for continuity or symmetry that can be exploited. Informational structures, however, are omnipresent in a system that is not completely lacking structure; they can be used to infer sensomotoric maps from entirely uninterpreted sensoric input [25, 26, 27, 28] 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 [9, 37]. On the one hand, information processing is "convertible" into other basic biological currencies, e.g. ATP consumption [21], on the other hand, it provides tools by which information can be treated *directly* as a quantifiable, limited resource.

Together with the insights of the self-organized structuring of information from the information bottleneck principle [34] as well as from the information-flow studies [17, 18], 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. Moreover, if the hypothesis is correct, then the principle could act as an approach for developing AI systems that capture some of the spirit of living learning systems. This is the main motivation for our current methodology.

3 Building Skills via Information-Based Structuring and Information Parsimony Principles

We have discussed above how Shannon information can be used as fundamental quantity to establish the relationship between an agent and its environment. This will form the basis for the construction of our Bold Hearts 2005 team. At this point, we do not yet have a mechanism that allows to build up automatically a skill hierarchy, so we, as humans have to provide the system with a suitable skill achievement hierarchy. Even in nature, different skills are developed at different times which are genetically determined; in addition, there is a cultural component (teaching) which determines the order in which skills are best learnt. Here, we will externally determine the skill learning order and hierarchy. We strive, however, to automatize this in the future.

The essential philosophy of our approach is to build up a world model from interaction with the world. Given that in the 3D simulator, the world is realized by ODE which simulates real physics, the complexity of predicting the behaviour of the world, even given only the ego-agent's control, is significantly higher than in the 2D world.

We build up the world model from probing the world in selected scenarios and then using the information parsimony principle to minimize the description complexity of the system as to still being able to make sensible predictions. To do so, and to be able to treat the continuous action space, we use the algorithm from [6] to implement a projection algorithm that compacts data points into manifolds, thereby reducing the information necessary to represent the data: this is the principle that we call *information parsimony*. The price that is to be paid is an increasing distortion for less information [see the original article, 6], and it is traded off by judicious use of a (Lagrangian) parameter λ.

3.1 Learning Undisturbed Ball Movement

Obviously, the undisturbed ball movement is simple to code directly, but we use the information manifold approach consistently even here to show that the build-up of prediction and action skills can be carried through to all the required elements.

We consider random samples of sequences for the movement of the ball, where we measure raw positional data for the ball[1]: $x_t, x_{t+1}, x_{t+2}, \ldots$. One could compress this information into a manifold, being aware of the structure of affine space, of rotational symmetry and of the properties of friction; we will suggest here a possible mechanism to achieve that, based on the bowtie principle [11]. The implementation is currently still done manually.

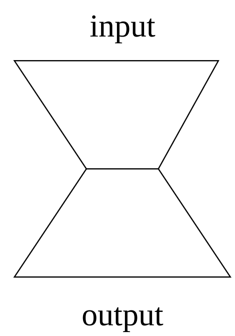

Fig. 1. The bowtie principle [11]. A complex system input is translated into a compact universal representation and then retranslated back into the original system language

In the *bowtie principle*, one considers complex systems and their large possible state set. The bowtie itself is then a kind of minimalistic interface which all these systems share and which all systems can use, only to translate the outputs back into the original representation. In our scenario, we will use the bowtie philosophy to manage symmetries. At the present point, the symmetries are still being identified by hand, but methods are in work to implement automatic discovery and utilization of these.

In the scenario of the unperturbed ball movement, the available symmetries (which we at this point use explicitly) are those of the affine space. The movement is invariant to any displacement of the starting positions. In particular, the movement can be preprocessed by calculating the sequence $v_t, v_{t+1}, v_{t+2} = x_{t+1} - x_t, x_{t+2} - x_{t+1}, x_{t+3} - x_{t+1}$, which can be considered the bowtie version, because it collapses the large sample space

[1] It should be noted that the data are calculated not objectively, but in the subjective coordinate system of the agent, and are thus prone to errors, agent displacement or slip and noise. Nevertheless, the method worked well.

for experiments with different starting positions into sequences starting at the same position. Knowing the starting position, we could reconstruct the original sequence (by reversing the operation).

A second bowtie operation is to rotate the vectors of the sequence individually as to make the first v_t to point in, say, x axis direction. The prediction of v_{t+1} from v_t then obviously becomes a much simpler task, namely $v_{t+1} = TfT^{-1}v_t$, where T denotes the rotation of the x axis vector ($\hat{e}_x$) to match v_t and f now only needs to operate on a vector along one axis, i.e. effectively on a one-dimensional quantity. Currently we still have to choose this operation by hand, but we are working on methods to automatically identify bowties via parsimony, possibly by combining the approach from [8] with the information manifold approach. In the present case, bowtying is nothing else than removing well-known spatial symmetries from the data, but in the general case, the presence of such symmetries cannot be assumed, and a more general bowtying approach will be called for.

At this point, one could try to reconstruct the generating manifold via Takens' theorem. Here, however, we consider the sequence of "bowtied" vectors, i.e. $T^{-1}v_t$ for all times t. At this point, there is no obvious bowtying any more. Now the joint v_t, v_{t+1} pairs are directly used as data samples, from which an interrelation in form of a compressed manifold is constructed. Through the repeated bowtying, the sparseness in the sample space has been reduced significantly, and the expressiveness of the data is strong, resulting in pronounced (in this case linear) manifolds which can then serve for prediction.

3.2 Jumping Ball

A very similar approach is used for the vertically moving ball, where the sequential velocities split into two groups, one that is relevant if the ball just undergoes a bounce, and the other one is for a free flying ball. (At this point, we still do not consider the influence of other agents on the ball). Again, an information manifold compression is undertaken and the resulting (partial) manifolds used for prediction.

3.3 Player Movement

For the player movement, we consider samples from two extreme cases, a player with full driving force and a player without any driving force. This results in two branches of the manifold structure, which result in two different sets of prediction parameters.

3.4 Kick and Kick Distance

The kick is studied by having the agent run towards the (resting) ball and permanent kicking. Once the ball deviates from its prediction (namely, no movement), it can be determined how the kick strength and angle interact with the post-kick ball velocity. Again, this is achieved by looking at the resulting information manifold.

From this calculation, it is easy to determine the kick distance. Here, however, there is the problem that different runs of the simulation will yield different kick distances, a problem that caused Bold Hearts 2004 (3D) to waste kicks in some situations. This problem again can be solved by bowtying; however, since we do not wish to assume some exact nature of the variation of kick distance, the precise bowtying structure needs to be determined empirically. One of the tasks for the present team is to find out how this can be achieved.

4 Preliminary Results and Open Questions

We have shown that the information parsimony principle, implemented as information manifold compression of sample data obtained from an agent, can provide powerful mechanisms to create a world model for the agent. The resulting players had a very good idea of ball movement, agent/ball interaction (except for the cases where the kick distance was misjudged) and move smoothly and efficiently towards the ball.

At this point, still some aspects of the methodology require manual intervention which we intend to eliminate successively in the future. These are:

1. choosing the training scenario hierarchy (which, as we argued, is also in general not naturally given in humans and we believe to be the most difficult to get rid of)
2. defining some of the required bowtie transformations (this is ongoing research)
3. converting the obtained manifolds into simple and fast functions; for this, there are already some candidate approaches, but one open question is how to decompose naturally the several appearing manifold fragments. This is a current research question.

5 Acknowledgements

We wish to thank the UvA team for providing the basis code (UvA 2002) [19] 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 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., Cichoki, 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] Baddeley, R., Hancock, P., and Földiák, P., editors, [2000]. *Information Theory and the Brain*. Cambridge University Press.

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

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

[5] 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.

[6] Chigirev, D., and Bialek, W., [2004]. Optimal manifold representation of data: An information theoretic perspective. In *Advances in Neural Information Processing 16*. Cambridge: MIT Press.

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

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

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

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

[11] Doyle, J., [2004]. Presentation at ICCS 2004, Boston.

[12] Franco, S., and Polani, D., [2004]. Bold Hearts 2003: Skill Learning Via Information-Theoretical Decomposition of Behaviour Features. Technical Report 399, Faculty of Engineering and Information Sciences, University of Hertfordshire.

[13] 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).

[14] 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.

[15] 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*.

[16] Jaynes, E. T., [1957]. Information theory and statistical mechanics. *Phys. Rev.*, 106(4):620–630.

[17] 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*. IEEE Computer Society. Technical Report No. 400, Department of Computer Science, Faculty of Engineering and Information Sciences, University of Hertfordshire.

[18] Klyubin, A. S., Polani, D., and Nehaniv, C. L., [2004]. Tracking Information Flow through the Environment: Simple Cases of Stigmergy. In *Proc. Artificial Life IX*. Technical Report No. 402, Department of Computer Science, Faculty of Engineering and Information Sciences, University of Hertfordshire.

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

[20] 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.

[21] 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.

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

[23] 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.

[24] 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.

[25] Olsson, L., Nehaniv, C., and Polani, D., [2004]. The Effects on Visual Information in a Robot in Environments with Oriented Contours. In *Proc. Epigenetic Robotics*.

[26] Olsson, L., Nehaniv, C. L., and Polani, D., [2004]. Information Trade-Offs and the Evolution of Sensory Layouts. In *Proc. Artificial Life IX*.

[27] Olsson, L., Nehaniv, C. L., and Polani, D., [2004]. Sensory Channel Grouping and Structure from Uninterpreted Sensor Data. In *Proc. Evolvable Hardware*. Technical Report No. 401, Department of Computer Science, Faculty of Engineering and Information Sciences, University of Hertfordshire.

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

[29] Shannon, C. E., [1949]. The Mathematical Theory of Communication. In Shannon, C. E., and Weaver, W., editors, *The Mathematical Theory of Communication*. Urbana: The University of Illinois Press.

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

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

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

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

[34] 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*.

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

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

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

[38] 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.

[39] Zurek, W. H., editor, [1990]. *Complexity, Entropy and the Physics of Information*, Santa Fe Studies in the Sciences of Complexity, Reading, Mass. Addison-Wesley.