Impoverished Empowerment: 'Meaningful' Action Sequence Generation through Bandwidth Limitation

Tom Anthony, Daniel Polani, Chrystopher L. Nehaniv

University of Hertfordshire

Abstract. *Empowerment* is a promising concept to begin explaining how some biological organisms may assign *apriori* values expectations to states in taskless scenarios. Standard empowerment samples the full richness of an environment and assumes it can be fully explored. This may be too aggressive an assumption and here we explore impoverished versions achieved through a limit on the bandwidth of the empowerment generating action sequences. It turns out that limited richness of actions concentrate on the "most important" ones with the additional benefit that the empowerment horizon can be extended drastically into the future. This may indicate a path towards and intrinsical preselection for preferred behaviour sequences and may help to suggest more biologically plausible approaches.

1 Introduction

Methods to provide an embodied agent with strategies to behave intelligently in an previously unknown environment or without specific goals or tasks are of great interest in Artificial Life. However, in order to do this embodied agents require some method by which they can differentiate available actions and states in order to decide on how to proceed. In the absence of no specific tasks or goals it can be difficult to decide what is and is not important to an agent.

One set of approaches examines processing and optimising the Shannon-type information an agent receives from it's environment (Attneave, 1954; Barlow, 1959, 2001; Atick, 1992), following the hypothesis that embodied agents benefit from an adaptive and evolutionary advantage by informationally optimising their sensoric and neural configurations for their environment.

Information based predictions could provide organisms/agents with intrinsic motivation (Prokopenko et al., 2006; Bialek et al., 2001; Ay et al., 2008); each use similar approaches based on *predictive information*. In this paper we will concentrate on *empowerment* (Klyubin et al., 2005b,a), an information theoretic measure for the efficiency of a *perception-action loop*.

One shortcoming of empowerment is that whilst it provides behaviours and results which seem to align it with processes that may have resulted from evolution they tend not to operate using an equally plausible process. In an artificial setting empowerment is calculated with the Blahut-Arimoto algorithm, but the

problem remains that it implicitly retains a notion of the richness and full size of the space it searches whatever process creates it. In this paper we assume a limit on the richness of the action repetoire.

1.1 Information Theory

Here we give a very brief introduction to information theory, introduced by Shannon (1948). The first measure is *entropy*, a measure of uncertainty given by $H(X) = -\sum p(x) \log p(x)$ where X is a discrete random variable with values $x \in \mathcal{X}$ and $p(x)$ is the probability mass function such that $p(x) = Pr(X = x)$. We use base 2 logarithm and measure in *bits*.

If Y is another random variable jointly distributed with X the *conditional entropy* is $H(Y|X) = -\sum_x p(x) \sum_y p(y|x) \log p(y|x)$. This measures the remaining uncertainty about the value of Y if we know the value of X. Finally, this also allows us to measure the *mutual information* between to random variables:

$$I(X;Y) = H(Y) - H(Y|X). \tag{1}$$

Mutual information can be thought of as the reduction in uncertainty about the variable X or Y, given that we know the value of the other.

1.2 Empowerment

Essentially empowerment uses the channel capacity for the external component of a perception-action loop to identify areas that are advantageous for an agent embodied within an environment. It assumes situations with a high efficiency of the perception-action loop should be favoured by an agent. Based entirely on the sensors and actuators of an agent, empowerment intrinsically encapsulates an evolutionary perspective; namely that evolution has selected which sensors and actuators a successful agent should have, which in turn implies which states are most advantageous to be visited.

Empowerment is based on the information theoretic perception-action loop formalism introduced by Klyubin et al. (2005b,a, 2004), as a way to model embodied agents and their environments. The model views the world as a communication channel; when the agent performs an action, it is injecting Shannon information into the environment, which may or may not be modified, and subsequently the agent re-acquires part of this information from the environment via its sensors.

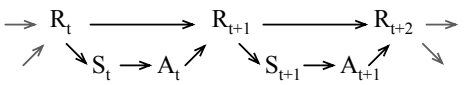

Fig. 1. Bayesian network representation of the perception-action loop.

In Fig.1 we can see the perception-action loop represented by a Bayesian network, where the random variable R_t represents the state of the environment, S_t the state of the sensors, and A_t the actuation selected by the agent at time t. It can be seen that R_{t+1} depends only on the state of the environment at time t, and the action just carried out by the agent.

By modelling this as a communication channel, we can employ information-theoretic methods, which are the basis for empowerment. Empowerment measures the maximum potential information flow, this can be modelled by the channel capacity (Shannon, 1948) for a discrete memoryless channel:

$$C(p(s|a)) = \max_{p(a)} I(A;S). \tag{2}$$

The random variable A represents the distribution of messages being sent over the channel, and S the distribution of received signals. The channel capacity is measured as the maximum mutual information taken over all possible input distributions, $p(a)$, and depends only on $p(s|a)$, which is fixed. One algorithm that can be used to find this maximum is the iterative Blahut-Arimoto algorithm (Blahut, 1972).

Empowerment can be intuitively thought of as a measure of how many observable modifications an embodied agent can make to his environment, either immediately, or in the case of n-step empowerment, over a given period of time.

In the case of n-step empowerment, we first construct a compound random variable of the last n actuations, labelled A_t^n. We now need to maximise the mutual information between this variable and the sensor readings at time $t+n$, represented by S_{t+n}. Here we consider empowerment as the channel capacity between these:

$$\mathfrak{E} = C(p(s_{t+n}|a_t^n)) = \max_{p(a_t^n)} I(A_t^n; S_{t+n}). \tag{3}$$

An agent that maximises its empowerment will position itself in the environment in a way as to maximise its options for influencing its relationship with the environment (Klyubin et al., 2005a).

2 Empowerment with limited action bandwidth

2.1 Goal

We wanted to introduce a bandwidth constraint into empowerment, specifically $n-step$ empowerment where an agent must look ahead and possible outcomes for sequencess of actions, and even with a small set of actions these sequences can become very numerous.

An agent's empowerment is bounded by that agent's memory; empowerment measures the agent's ability to exert influence over it's environment and an agent that knows only 4 distinct actions can have no more than 2 bits of empowerment per step. However, there are two factors which normally prevent empowerment from reaching this bound:

– Noise - A noisy / non-deterministic / stochastic environment means that from a given state an action has a stochastic mapping to the next state. This reduces an agents control and thus it's empowerment.
– Redundancy - Often there are multiple actions available which map from a given state to the same resultant state. This is especially true when considering multi-step empowerment. e.g Moving North then West or moving West then North.

Therefore in many cases bandwidth for actions can be reduced with little or no impact on achievable information flow. Beyond this there may be scenarios where a reduction in empowerment/utility is acceptable and is desirable to achieve further reductions in action bandwidth.

2.2 Scenario

To run tests we constructed a simple scenario; an embodied agent is situated within a 2 dimensional infinite gridworld and has 4 possible actions in any single time step. The actions the agent can execute are North, South, East and West and provided the cell in the corresponding direction is free; it may be that the target cell is occupied by a wall, in which case the action is executed but the agent does not move from its current cell. In the scenario the state of the world is solely the position of the agent, and this is all that is detected by the agent's sensors.

2.3 Approach

We hypothesise, given that for short sequences of actions it is manageable to cheaply examine all sequences, that we could approach an agent's bandwidth divided into two parts; a 'working' memory and a 'long term' memory. The constraints we were to apply should be on an agent's 'long term' memory.

The agent to examines all possible options for $n - step$ empowerment for small values of n (typically $n < 6$) and then selects a subgroup of the available sequences to be retained (the number of which corresponding to the bandwidth limit).

To do this we use the information bottleneck method (Tishby et al., 1999) to select which actions to retain. Having calculated the empowement we have two distributions; $p(a)$ is the capacity achieving distribution of actions and $p(s|a)$ is the channel that represents the results of an agent's iteractions with the environment.

We now look for a new "compact" distribution $p(g|a)$, where g are groups of alike action sequences and $|G| < |A|$ where the cardinality of G corresponds to our desired bandwidth limit. A colloquial, though not entirely accurate, way to think of this is as grouping action sequences that have similar outcomes (or represent similar 'strategies').

The information bottleneck is used to select $p(g|a)$ and from this mapping of action sequences to groups we select a new distribution $p(\hat{a})$ where $\hat{a}$ is the reduced set of action sequences (for implementation details see section 4).

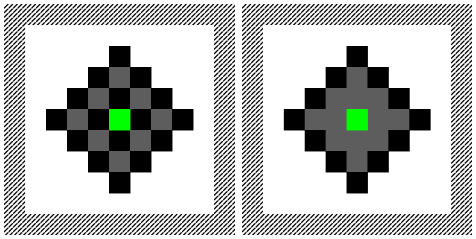

Fig. 2. Selected trajectories in a 3-step scenario showing redundancy elimination. On the left 16 selected action sequences, on the right 12 selection action sequences.

In section 2.1 we discussed redundancy as one factor which should be eliminated first in order to maintain empowerment whilst reducing bandwidth. In fig. 2 we can see a scenario in which the agent uses sequences of 3 actions; there are $4^3 = 64$ possible actions that can reach 16 different end states. The walls are represented by patterned grey, the starting position of the agent is the light center square, and the selected trajectories by the dark lines with black marking their end.

On the left we see results for an agent with a bandwidth of 4 bits resulting in the selection of 16 sequences; the best result is successfully achieved with a single trajectory to each of the 16 possible end states. We will present another example before returning to the right side of this figure.

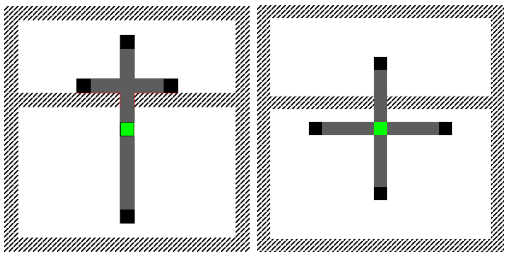

Fig. 3. Typical behaviours where 4 action sequences were selected from 4^6 possibilities.

Fig. 3 shows two further typical outcomes of this algorithm; in this example we have a bandwidth constraint of 2 bits, operating on sequences of 6 actions. What can be seen as to have emerged is of interest; the sequences chosen can immediately be seen to be non-trivial and a brief examination reveals that the end points each have only a single sequence (of the available 4,096) that reaches them.

Returning to fig. 2, the righthand result is the same setup exactly as the left, but now the bandwidth has been reduced to 3.58 bits corresponding to 12

allowed action sequences. We can see that, of the 16 sequences from before, the agent now 'forgets' the 4 which led to the states having higher redundancy.

If we extrapolate this process of eliminating trajectories to 'easier to reach' states then it follows that, exactly as in fig. 3, the last states the agent will retain are the entirely unique states that have only a single sequence that reaches them.

It appears that choosing to retain a limited number of explored sequences and this tendency for the agent to value 'unique' sequences indicates a first step towards a solution for extending the sequences beyond what was computationally possible before. This in turn indicates that it may point to a plausible process for a biological organism to undertake. We discuss this in section 3.

2.4 Noise induced behvaiour modifications

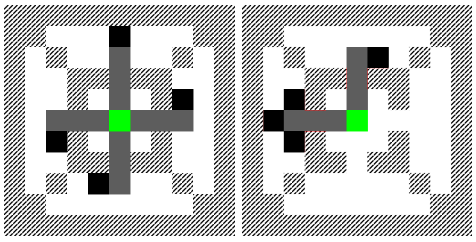

Fig. 4. Randomly selected behaviours; 4 steps with a 2 bit bandwidth constraint.

Fig. 4, a 4-step scenario with a bandwidth constraint of 2 bits corresponding to 4 actions, shows there is not always a neat division of the world into what we would probably recognise as the 4 main 'strategies' (one trajectory into each of the 4 rooms). However, there is no push for the agent to do this or to consider the geographical distinctions between states.

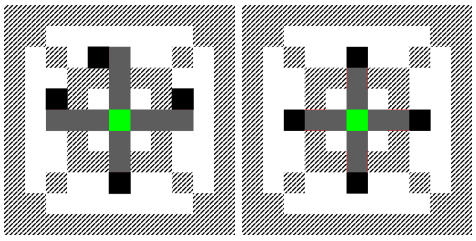

Fig. 5. Two randomly selected behaviours in a 4-step scenario with a 2 bit bandwidth constraint and 5% noise per step.

However, with the introduction of noise this changes. Fig. 5 shows two more randomly selected behaviours from the same scenario but with the introduction of noise, where each action has a 5% probability of being replaced with a random action. In order to maintain as much empowerment as possible, the agent must ensure that in attempting one strategy it doesn't accidentally employ another, and in this environment that translates to being 'blown off course' and adds a push for a geographical distinction between end states.

Note in this figure that 5 of the 8 sequences shown appear to be only 3 steps long. However, this appears to be a strategy employed by the agent, and what is actually happening is the agent uses an action to push against the wall while passing through the doorway.

3 Building long action sequences

The current formulation for $n-step$ empowerment utilises an exhaustive search of the action space for $n-steps$. It can be seen that this is a highly unlikely approach for biological organisms to employ, especially for large values of n and in rich environments.

Following the result above from bandwidth limited empowerment it became apparent that retaining only a small subset of investigated action sequences lends itself to the idea of then searching further from the final states of such sequences.

To begin, this is obvious when applied to the cases where the bandwidth has been constrained just enough to retain empowerment but eliminate all redundancy; it is essentially realising the Markovian nature of such sequence based exploration. Having arrived at a state to explore then how you arrived is not of consequence to further exploration.

However, the results seem to suggest that even beyond this point of retained empowerment, where the bandwidth is severely limiting to the achievable empowerment and selection of sequences the iterative build up still produce noteworthy behaviours.

The approach was to set a target length for a sequence, for example 15-step empowerment, then the problem is broken down in to i iterations of $n-step$ empowerment where $ni = 15$. Standard $n-step$ empowerment is performed, and then the above presented bandwidth-reduction algorithm is run to reduced the action set to a small subset. Each of these action sequences is then extended with n additional steps. These are then again passed through the bandwidth-reduction algorithm and this repeated a total of i times.

If we selected $n = 5$, $i = 3$ and a bandwidth limit of 4 bits (16 actions) then the total states searched in our gridworld scenario would be reduced from 4^{15} to $33,792$, which is a search space over $3 \cdot 10^4$ times smaller.

Fig. 6 shows the results of such a scenario with the selected action sequences and there are several important aspects to note. Firstly, the agent continues to reach certain states that are of obvious consequence, most notably the 4 cardinal directions, but also over half of the 8 further corner points. Further more the pattern of trajectories has a somehwat 'fractal' nature and appear to divide

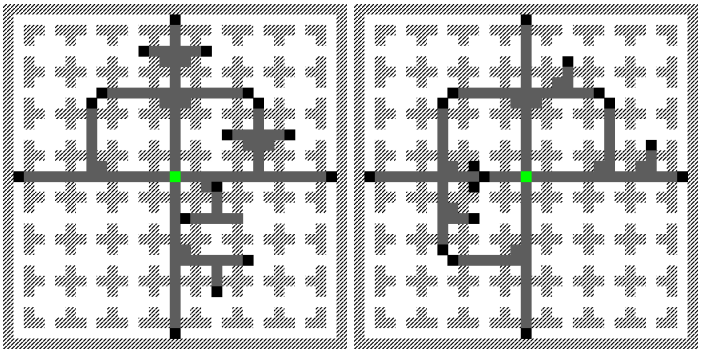

Fig. 6. Iteratively built sequences of 15 steps, with a bandwidth constraint of 4 bits.

the search space up systematically. These results are of interest because these states and behaviours are far beyond the horizon of a single iteration of standard $n - step$ empowerment. Space does not permit but initial results also indicate that interesting locations of the environment, such as door and bridges, are also handled by such iterative sequence building.

4 Algorithm details

We chose to implement the Information Bottleneck combined with a secondary algorithm to 'decompose' the produced conditional distribution $p(g|a)$ into a new distribution for $p(a)$ which has an entropy within the specified bandwidth limit (and usually contains only a subset of the original actions).

Using the information bottleneck we were able to reduce entropy of our actions by choosing a cardinality for G and then maximising $I(G;S)$ (the empowerment of the reduced action set) using A as a relevancy variable. This results in retaining empowerment whilst compressing the action sets entropy.

However, this does not result in a one to one mapping of a subset of A to G, but rather results in a conditional probability distribution. Therefore, in order to end up with a subset of our original actions to form a new action policy for the agent, we must apply some method of selecting an a for each value of g (which are essentially meta actions).

As stated, in the spirit of empowerment, for each g we want to select the actions which are most likely to map to that g (i.e the highest value of $p(g|a)$ for the given g). This results in collapsing strategies to their dominant action sequence and maximises an agent's ability to select between strategies.

5 Discussion

We have identified several challenges to the recently introduced concept of empowerment which endows an agent's environmental niche with a concept distin-

guishing desirable from less desirable states. Empowerment essentially measures the range in environmental change imprinted by possible action sequences whose number grows exponentially with the length of the sequence. It is virtually impossible to compute it algorithmically for longer sequences, and, likewise, it is implausible that any adaptive or evolutionary natural process would be able to indirectly map this whole range.

Therefore, here we have, consistently with the information-theoretic spirit of our study, applied informational limits on the richness of the action sequences that generate the empowerment. In doing so, we found that: 1. the information bottleneck reduces redundant sequences; 2. in conjunction with the complexity reduction through the collapse of action sequences, particularly "meaningful" action sequences that explore important features of the environment, e.g. principal directions, doors and bridges, are retained, and finally, that *significantly* longer action sequences than before can be realistically handled. This promises important insights for understanding the possible emergence of useful long-term behavioural patterns. Note that in this study we have relinquished the computation of empowerment as measure for the desirability of states in favour of filtering out desirable action patterns.

Bibliography

Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing. *Network: Computation in Neural Systems*, 3(2):213–251.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61(3):183–193.

Ay, N., Bertschinger, N., R. Der, F. G., and Olbrich, E. (2008). Predictive information and explorative behavior of autonomous robots. *European Physical Journal B*. (Accepted).

Barlow, H. B. (1959). Possible principles underlying the transformations of sensory messages. In Rosenblith, W. A., editor, *Sensory Communication: Contributions to the Symposium on Principles of Sensory Communication*, pages 217–234. The M.I.T. Press.

Barlow, H. B. (2001). Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12(3):241–253.

Bialek, W., Nemenman, I., and Tishby, N. (2001). Predictability, complexity, and learning. *Neural Comp.*, 13(11):2409–2463.

Blahut, R. (1972). Computation of channel capacity and rate distortion functions. *IEEE Transactions on Information Theory*, 18(4):460–473.

Der, R., Steinmetz, U., and Pasemann, F. (1999). Homeokinesis - a new principle to back up evolution with learning. In Mohammadian, M., editor, *Computational Intelligence for Modelling, Control, and Automation*, volume 55 of *Concurrent Systems Engineering Series*, pages 43–47. IOS Press.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2004). Organization of the information flow in the perception-action loop of evolved agents. In Zebulum, R. S., Gwaltney, D., Hornby, G., Keymeulen, D., Lohn, J., and Stoica, A., editors, *Proceedings of 2004 NASA/DoD Conference on Evolvable Hardware*, pages 177–180. IEEE Computer Society.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2005a). All else being equal be empowered. In Capcarrère, M. S., Freitas, A. A., Bentley, P. J., Johnson, C. G., and Timmis, J., editors, *Advances in Artificial Life: Proceedings of the 8th European Conference on Artificial Life*, volume 3630 of *Lecture Notes in Artificial Intelligence*, pages 744–753. Springer.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2005b). Empowerment: A universal agent-centric measure of control. In *Proceedings of the 2005 IEEE Congress on Evolutionary Computation*, volume 1, pages 128–135. IEEE Press.

Prokopenko, M., Gerasimov, V., and Tanev, I. (2006). Evolving spatiotemporal coordination in a modular robotic system. In Nolfi, S., Baldassarre, G., Calabretta, R., Hallam, J., Marocco, D., Meyer, J.-A., and Parisi, D., editors, *From Animals to Animats 9: 9th International Conference on the Simulation of Adaptive Behavior (SAB 2006), Rome, Italy, September 25-29 2006*, volume 4095 of *Lecture Notes in Computer Science*, pages 558–569. Springer.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.

Tishby, N., Pereira, F., and Bialek, W. (1999). The information bottleneck method. In *Proceedings of the 37-th Annual Allerton Conference on Communication, Control and Computing*, pages 368–377.