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

Tom Anthony, Daniel Polani, and Chrystopher L. Nehaniv

Adaptive Systems Research Group, University of Hertfordshire, UK

Abstract. *Empowerment* is a promising concept to begin explaining how some biological organisms may assign *a priori* value 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; here we explore impoverished versions achieved by limiting 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 indicates a path towards and intrinsic preselection for preferred behaviour sequences and helps to suggest more biologically plausible approaches.

1 Introduction

Methods to provide an agent embodied in an environment with strategies to behave intelligently when given no specific goals or tasks are of great interest in Artificial Life. However, 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 information an agent receives from its environment (Atick, 1992; Attneave, 1954; Barlow, 1959, 2001), following the hypothesis that embodied agents benefit from an adaptive and evolutionary advantage by informationally optimising their sensory and neural configurations for their environment.

Information-based predictions could provide organisms/agents with intrinsic motivation based on *predictive information*(Ay et al., 2008; Bialek et al., 2001; Prokopenko et al., 2006). In this paper we will concentrate on *empowerment* (Klyubin et al., 2005a,b), an information theoretic measure for the external 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 the algorithms used to calculated it tend not to operate using an equally plausible process. It implicitly requires a notion of the richness and full size of the space it searches whatever algorithm is used to determine it. In this paper we thus introduce the assumption of a limit on the richness of the action repertoire.

G. Kampis, I. Karsai, and E. Szathmáry (Eds.): ECAL 2009, Part II, LNCS 5778, pp. 294–301, 2011.

1.1 Information Theory

First we give a very brief introduction to information theory, introduced by Shannon (1948). The first measure is *entropy*, a measure of uncertainty given by $H(X) = -\sum_x p(x) \log p(x)$ where X is a discrete random variable with values x from a finite set $\mathcal{X}$ and $p(x)$ is the probability that X has the value x. We use base 2 logarithm and measure entropy 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 two random variables: $I(X;Y) = H(Y) - H(Y|X)$.

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 measures the channel capacity for the external component of a perception-action loop to identify states that are advantageous for an agent embodied within an environment. It assumes that 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 for the agent to visit.

Empowerment is based on the information theoretic perception-action loop formalism introduced by Klyubin et al. (2004, 2005a,b), 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.

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.

Empowerment measures the maximum *potential* information flow, this can be modelled by the channel capacity (Shannon, 1948) for a discrete memoryless channel: $C(p(s|a)) = \max_{p(a)} I(A;S)$.

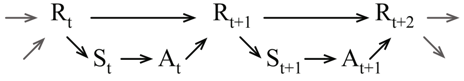

Fig. 1. Bayesian network representation of the perception-action loop

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

An agent that maximises its empowerment will position itself in the environment in a way as to maximise its options for influencing 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 at possible outcomes for sequences 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 can perform 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 agent's control and thus its empowerment.
- Redundancy - Often there are multiple action (or sequences) 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.

Due to redundancy there are many cases where bandwidth for action sequences can be reduced with little or no impact on achievable information flow. Beyond this there may be scenarios with a favourable trade off between a large reduction in action bandwidth only resulting in a small reduction in empowerment (or utility).

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 each moving the agent one space into the corresponding cell, provided it is not occupied by a wall. In the scenario the state of the world is solely the position of the agent, which is all that is detected by the agent's sensors.

2.3 Algorithm

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

To do this we use the information bottleneck method (Tishby et al., 1999). Having calculated the empowement we have two distributions: $p(a_t^n)$ is the capacity achieving distribution of action sequences and $p(s_{t+n}|a_t^n)$ is the channel that represents the results of an agent's iteractions with the environment.

We now look for a new "compact" distribution $p(g|a_t^n)$, where g are groups of 'alike' action sequences with $g \in G$ where $|G| \leq |A_t^n|$ and the cardinality of G corresponds to our desired bandwidth limit. A colloquial, though not entirely accurate, way to think of this is as grouping together action sequences that have similar outcomes (or represent similar 'strategies'). The information bottleneck works by first choosing a cardinality for G and then maximising $I(G; S_{t+n})$ (the empowerment of the reduced action set) using S_{t+n} as a relevance variable.

This results in a conditional distribution $p(g|a_t^n)$, from which we must derive a new distribution of our action sequences (with an entropy within the specified bandwidth limit). In order to end up with a subset of our original action sequences to form this new action policy for the agent, we must use an algorithm to 'decompose' the conditional distribution into a new distribution $p(\hat{a}_t^n)$ which has an entropy within the specified bandwidth limit (and usually contains only a subset of the original action sequences).

In the spirit of empowerment, for each g we want to select the action sequences which are most likely to map to that g (i.e the highest value of $p(g|a_t^n)$ for the given g) and provide the most towards our empowerment (i.e the highest value of $I(a_t^n; S_{t+n})$). This results in collapsing strategies to their dominant action sequence and maximises an agent's ability to select between strategies.

2.4 Results

Fig. 2 shows three typical outcomes of this algorithm; in this example we have a bandwidth constraint of 2 bits, operating on sequences of 6 actions. 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 a black cell marking the end location of the sequence. The result that emerges is of interest;

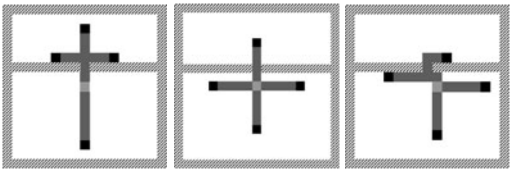

Fig. 2. Typical behaviours where 4 action sequences were selected from 4^6 possibilities

the sequences chosen can immediately be seen to be non-trivial and a brief examination reveals that the end points of each sequence each have only a single sequence (of the available 4,096) that reaches them.

In section 2.1 we discussed redundancy as one factor which should be eliminated first in order to maintain empowerment whilst reducing bandwidth. If we extrapolate this process of eliminating trajectories to 'easier to reach' states then it follows that, exactly as in Fig. 2, 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 and may point to a plausible process for a biological organism to undertake. We discuss this in section 3.

2.5 Noise Induced Behaviour Modifications

Figures 3 A & B, a 4-step scenario with a bandwidth constraint of 2 bits corresponding to 4 action sequences, show 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 pressure for the agent to do this or to consider the geographical distinctions between states, only for it to select unique end points.

However, with the introduction of noise this changes. Figures 3 C & D show two more randomly selected behaviours from the same scenario but with the introduction of noise, where each action in the sequence has a 5% probability of

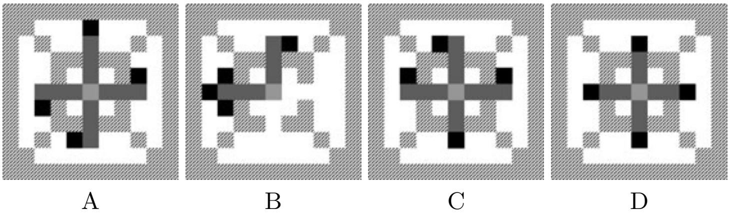

Fig. 3. Randomly selected behaviours; 4 steps with a 2 bit bandwidth constraint. A & B have no noise, C & D have 5% noise per step.

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 does not accidentally employ another, and in this environment that translates to being 'blown off course' and adds a drive for a geographical distinction between end states.

Note that some of the sequences appear to be only 3 steps long. This is 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, possibly as a way to minimise the effect of noise.

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.

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. In section 2.3 we evaluated all possible short sequences in a 'working' memory, then retained only a subset for the agent's 'long term' memory according to our bandwidth constraint.

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.

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: when arriving at a state to explore, how you arrived is not of consequence to further exploration. The results, however, seem to suggest that even *beyond* this point of retained empowerment, where the bandwidth severely limits the achievable empowerment and selection of sequences, the iterative approach still produces 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 $n \cdot i = 15$. 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 select $n = 5$, $i = 3$ and a bandwidth limit of 4 bits (16 action sequences) then the total sequences evaluated in our gridworld scenario is reduced from 4^{15} to $33,792$, which is a search space more than $3 \cdot 10^4$ times smaller.

Figure 4 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

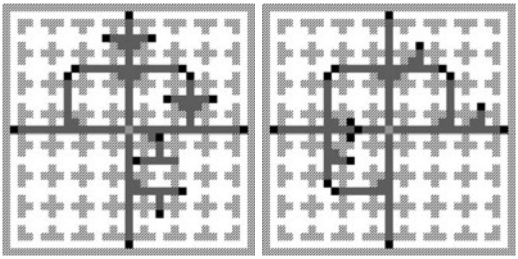

Fig. 4. Iteratively built sequences of 15 steps, with a bandwidth constraint of 4 bits

reach certain states that are of obvious consequence, most notably the 4 cardinal directions, but also over half of the furthest reachable corner points. Furthermore the pattern of trajectories has a somehwat 'fractal' nature and appears to divide 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 us to give details 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 Discussion

We have identified several challenges to the recently introduced concept of empowerment which endows an agent's environmental niche with a concept distinguishing 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 feasibly handled. This in itself already suggests 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.

References

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

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

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

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

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

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

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

Klyubin, A.S., Polani, D., Nehaniv, C.L.: 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., Stoica, A. (eds.) Proceedings of 2004 NASA/DoD Conference on Evolvable Hardware, pp. 177–180. IEEE Computer Society Press, Los Alamitos (2004)

Klyubin, A.S., Polani, D., Nehaniv, C.L.: All else being equal be empowered. In: Capcarrère, M.S., Freitas, A.A., Bentley, P.J., Johnson, C.G., Timmis, J. (eds.) ECAL 2005. LNCS (LNAI), vol. 3630, pp. 744–753. Springer, Heidelberg (2005a)

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

Prokopenko, M., Gerasimov, V., Tanev, I.: Evolving spatiotemporal coordination in a modular robotic system. In: Nolfi, S., Baldassarre, G., Calabretta, R., Hallam, J., Marocco, D., Meyer, J.-A., Parisi, D. (eds.) SAB 2006. LNCS (LNAI), vol. 4095, pp. 558–569. Springer, Heidelberg (2006)

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

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