Genotypic versus Behavioural Diversity for Teams of Programs Under the 4-v-3 Keepaway Soccer Task

Stephen Kelly and Malcolm I. Heywood
Dalhousie University, Halifax, NS, Canada
{skelly,mheywood}@cs.dal.ca http://web.cs.dal.ca/~skelly

Abstract

Keepaway soccer is a challenging robot control task that has been widely used as a benchmark for evaluating multi-agent learning systems. The majority of research in this domain has been from the perspective of reinforcement learning (function approximation) and neuroevolution. One of the challenges under multi-agent tasks such as keepaway is to formulate effective mechanisms for diversity maintenance. Indeed the best results to date on this task utilize some form of neuroevolution with genotypic diversity. In this work, a symbiotic framework for evolving teams of programs is utilized with both genotypic and behavioural forms of diversity maintenance considered. Specific contributions of this work include a simple scheme for characterizing genotypic diversity under teams of programs and its comparison to behavioural formulations for diversity under the keepaway soccer task. Unlike previous research concerning diversity maintenance in genetic programming (GP), we are explicitly interested in solutions taking the form of teams of programs.

Introduction

Symbiotic Bid-Based GP (SBB) is a hierarchical framework for symbiotically coevoling teams of simple programs over two distinct cycles of evolution (Kelly, Lichodzijewski, and Heywood 2012). The first cycle produces a library of diverse, specialist teams with limited capability. The second cycle builds more general and robust policies by re-using the library, essentially building generalist teams from multiple specialists. Thus, diversity maintenance is critical during the first stage of evolution to ensure the identification of a wide range of specialist behaviours.

Keepaway soccer is a challenging benchmark task for multi-agent learning in which a team of K *keepers* must maintain possession of the ball while an opposing team of $K-1$ *takers* attempt to gain possession (Stone et al. 2006). The keepers must learn a policy that maximizes the length of play against the takers, which follow a pre-specified behaviour. The players' sensors and actuators are noisy, making the task partially observable and highly stochastic. The size of the playing region and number of keepers vs. takers may be adjusted to scale the difficulty of the task. In this work we are working with the 4-v-3 task configuration on a 25m x 25m field. The keepaway task has been dominated by value function-based approaches to reinforcement learning (Kalyanakrishnan and Stone 2009) and neuroevolution (Metzen et al. 2007). In the case of neuroevolution, genotypic diversity maintenance was shown to be a critical factor in the algorithm's success. However, the maintenance of *behavioural* diversity is increasingly found to be more effective than diversity mechanisms operating solely in genotype space, in particular when the domain is deceptive, for example, due to a noisy fitness function (Gomez 2009). In this work we introduce novel methods for measuring genotypic and behavioural diversity among teams of programs, and empirically compare their efficacy under the keepaway task.

Diversity Mechanisms

In the keepaway task, the fitness of a team is the mean episode length over all games played. In order to promote population diversity, each team's *novelty* must also factor into the selection process, where novelty refers to the mean genotypic or behavioural distance between a team and all other members of the same population. Several methods to balance fitness and novelty are possible, including fitness sharing, crowding, or multi-objective optimization. In this work we adopt a simple linear combination of fitness and novelty (Cuccu and Gomez 2011), thus each team's score is defined prior to selection: $score(tm_i) = (1-p)\cdot\overline{Fit(tm_i)} + p\cdot\overline{Nov(tm_i)}$, where $\overline{Fit}$ and $\overline{Nov}$ are the normalized fitness and novelty of team i, and p is a parameter to control the relative weight of novelty. The novelty component requires a method of measuring genotypic or behavioural distance between each pair of teams, discussed below.

Genotypic Diversity

Teams in SBB are comprised of multiple cooperative programs, or symbionts. At each decision point in a game, action selection is determined care of a bidding metaphor, where each symbiont produces a real-valued bid relative to the current state observation and the highest bidder determines the action taken. The genotype of a team can be characterized simply by noting which specific symbionts, each having a unique program and identifier, are active within the

team. A symbiont is considered active if it has scored a winning bid, and thus suggested an action for the task domain, at least once during the life of the team. Genotypic distance between teams is summarized as the ratio of active symbionts common to both teams. Formally, the genotypic distance between teams i and k is

$$dist(tm_i, tm_k) = 1 - \frac{Sym_{active}(tm_i) \cap Sym_{active}(tm_k)}{Sym_{active}(tm_i) \cup Sym_{active}(tm_k)} \quad (1)$$

where $Sym_{active}(tm_x)$ represents the set of active symbionts in team x.

Behavioural Diversity

The behaviour of a host is summarized with respect to behavioural attributes recorded over the course of each game. At each timestep, the observation, 19 real-valued state variables in the case of 4-v-3 keepaway, and subsequent action taken by the keeper are recorded. Each state variable is discretized to $[0, 1, 2]$, or low, medium, high. Thus, for each training episode a *profile* vector is recorded, $\vec{p} = [\{a(t), s(t)\}, t \in T]$, where $a(t)$ is the discrete, domain-specific action taken at each timestep, $s(t)$ is the discretized state observation, and T represents every timestep in the associated episode. Note that this method of characterizing behaviour is task-independent. In every generation, each team is evaluated in 10 new games. Teams maintain a historical record of the profile $\vec{p}$ for 100 games. Once the team's historical record is full, it is no longer evaluated. However, it stays in the population as long as it is not marked for deletion by the selection process. We define $\vec{P}$ to be the concatenation of all profile vectors $\vec{p}$ in a team's historical record. Behavioural distance between a pair of teams can now be summarized as the Normalized Compression Distance (Gomez 2009) between their corresponding concatenated profile vectors: $dist(tm_i, tm_k) = ncd(\vec{P}_i, \vec{P}_k)$

Results

Three experimental cases are considered. In each case, hierarchical policies are developed over two cycles of evolution with 250 generations in the first cycle and 125 generations in the second. During the first cycle, case 1 employs no diversity maintenance, case 2 employs genotypic diversity, and case 3 employs behavioural diversity. No diversity maintenance is used during the second cycle of evolution, where the goal is strictly exploitative. The parameter p was fixed at 0.1 and 0.4 for genotypic and behavioural diversity respectively. These values were determined experimentally and are not necessarily optimal. Figure 1 reports the test performance of the champion team from each experimental case at the end of the second evolutionary cycle. It is apparent that policies with behavioural diversity in the first cycle are able to achieve significantly better performance in the second cycle, or exploitation phase. This suggests that behavioural diversity is most effective when constructing a library of reusable specialist policies. Wilcoxon rank-sum test confirms statistical significance between groups of all $10,000$ test outcomes from behavioural vs. no diversity and behavioural vs. genotypic ($p < 0.05$).

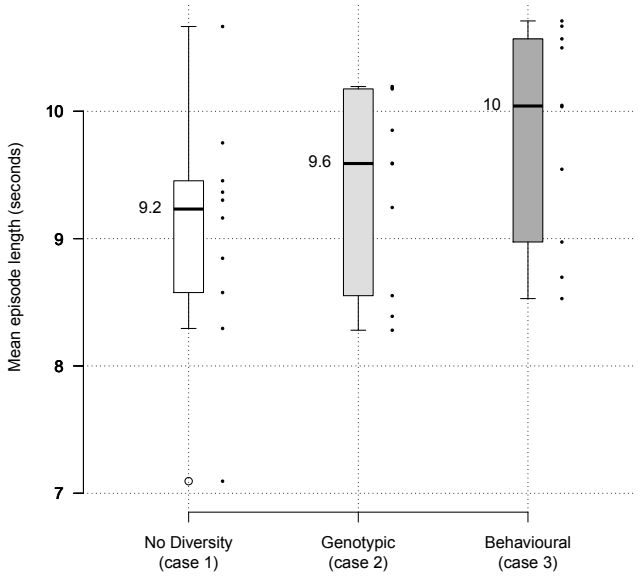

Figure 1: Average performance of champion policy against 1,000 test games. Box plot reflects the quartile distribution and scatter plot the actual performance points from 10 runs. Numerical value reports the median of the 10 runs.

Conclusion

We have introduced a simple genotypic diversity measure applicable to teams of programs and compared this with a behavioural, task-independent diversity characterization. While both diversity mechanisms are beneficial relative to the case of no diversity, the development of hierarchical polices has been shown to benefit more from behavioural diversity maintenance than from the genotypic formulation.

References

Cuccu, G., and Gomez, F. 2011. When novelty is not enough. In *EvoApplications – Part I*, volume 6624 of *LNCS*.

Gomez, F. 2009. Sustaining diversity using behavioral information distance. In *ACM Conference on Genetic and Evolutionary Computation*.

Kalyanakrishnan, S., and Stone, P. 2009. An empirical analysis of value function-based and policy search reinforcement learning. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*.

Kelly, S.; Lichodzijewski, P.; and Heywood, M. I. 2012. On run time libraries and hierarchical symbiosis. In *IEEE Congress on Evolutionary Computation*.

Metzen, J. H.; Edgington, M.; Kassahun, Y.; and Kirchner, F. 2007. Performance evaluation of EANT in the robocup keepaway benchmark. In *IEEE International Conference on Machine Learning and Applications*.

Stone, P.; Kuhlmann, G.; Taylor, M.; and Liu, Y. 2006. Keepaway soccer: From machine learning testbed to benchmark. In *RoboCup 2005: Robot Soccer World Cup IX*.