Considerations for Multiagent Multi-Objective Systems

(Doctoral Consortium)

Logan Yliniemi
Oregon State University
Corvallis, OR, USA
logan.yliniemi@engr.orst.edu

ABSTRACT

Cooperative multiagent systems are an important area of research, with many applications including air traffic control, satellite communications, and extraplanetary exploration. These applications also lend themselves to multiple simultaneous objectives, but multi-objective multiagent systems have garnered little attention. We make contributions to three distinct aspects of multi-objective multiagent systems: integrating credit assignment into multi-objective systems; providing a computationally efficient technique for dealing with non-convex Pareto fronts; and providing a framework for outside knowledge to be integrated into the agent's reward structure in multi-objective problems. These three contributions together break down some of the largest barriers that have prevented multiagent multi-objective research from gaining traction.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems

Keywords

Multiagent Learning; Multi-Objective Problem

1. INTRODUCTION

Cooperative multiagent systems (MASs) focuses on producing a set of autonomous agents to achieve a system-level goal [6]. Multiagent frameworks have been used to study complex, real-world systems like air traffic [1, 6], teams of satellites, and extra-planetary rover exploration. In most cases, the goal is to optimize a single, well-defined objective function.

But, in many of these cases, the problems lend themselves more naturally to multiple objectives: for example, air travel should be as safe and as expedient as possible. Satellites may need to make observations for multiple separate institutions. Extra-planetary rovers should acquire multiple different types of scientific data. However, most research in multiagent systems does not address them as a multi-objective problem (MOP).

The contributions of this thesis are threefold: first, the integration of credit assignment from multiagent systems into multi-objective problems (Section 2); second, the development of a technique for dealing with Pareto fronts of unknown convexity (Section 3); and third, the a method for inclusion of outside knowledge to boost learning speed in multi-objective problems (Section 4).

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

2. ON CREDIT ASSIGNMENT IN MOPS

Developing successful agent policies in MASs can be challenging. One successful approach is to use adaptive agents with tools like reinforcement learning or evolutionary algorithms. Each agent seeks to maximize its own reward; with a properly designed reward signal, the whole system will attain desirable behaviors. This is the science of credit assignment: determining the contribution each agent had to the system as a whole. Clearly quantifying this contribution on a per-agent level is essential to multiagent learning. This is an issue that has not been studied within the context of MOPs.

Contribution.

We draw from the framework of difference rewards [1], and develop them in a multi-objective setting. We discuss some of the challenges that arise when using various *a priori* and *a posteori* multi-objective methods in tandem with difference rewards, and identify properties of algorithms that allow symbiotic gains.

Significance.

We show that credit assignment is incredibly important in MO-MASs; at least as important as the choice of multi-objective scalarization. Using proper credit assignment increases learning speed by up to 10x over a traditional global reward, and greatly increases robustness to unmodeled disturbances (up to 98% in some domains).

3. ON UNKNOWN PARETO FRONTS

In MOPs, there is no single optimal solution. Instead, there usually exists a set of optimal solutions which trade a loss on one objective for some amount of gains elsewhere. These solutions are called the Pareto front. The shape of the Pareto front in the objective space dictates which types of methods can be used to solve for solutions along the Pareto front. Specifically, some methods have a very difficult time finding regions of the Pareto front that are concave. This is unfortunate, as the shape of the Pareto front is not known until after optimization. Choosing an optimizer without the ability to find these concave solutions may mask the true nature of the Pareto front, and the system designer may never ever know that there exists a larger set of potential tradeoff solutions that are also Pareto optimal.

Techniques that are capable of solving for concave portions of Pareto fronts tend to be computationally expensive, however. Especially in tandem with multiagent techniques, this can present an insurmountable computational boundary.

Contribution.

One of the most computationally efficient methods is the linear combination of objectives, but it has been proven to be unable to find concave areas of the Pareto front [2]. We present a novel objective-space transformation that shapes the Pareto front such that it is non-concave in the transformed space. This transformation is known as the Pareto Concavity Elimination Transformation (PaCcET). Using a linear combination in this transformed space then leads to the discovery of solutions all along the Pareto front, even in the case of discontinuous or concave Pareto fronts.

The basic functionality of PaCcET is to keep the current approximation of the Pareto front P_I^*, based on all previous iterations. Using this approximation, we can create a transformation such that P_I^* lies on an equivaluable line to a linear combination. This means that any non-dominated point will be more desirable to an optimizer than a point on P_I^*, and a dominated point will be less valued. This succinctly addresses the issue of concavity for the linear combination, because the Pareto front in the transformed space is thus known to be globally non-concave. As the approximation of the Pareto front is pushed forward further and further, to better approximate the front, each local concavity of the Pareto front is transformed such that it is no longer. With a sufficiently powerful optimizer, this eventually results in the entire actual Pareto front existing on the equivaluable line.

Figure 1 shows this process visually. The grid of points in the first figure is the same grid of points in the second figure, but in the second they have been transformed such that the red points of P_I^* all lie on the portion of the space that evaluates to a linear combination of 1. The space is distorted, but not broken, and any single-objective optimizer can be used to optimize in this space.

Significance.

MOP and MAS research have each independently developed a set of tools for answering the challenges that either type of problem will encounter independently. Likewise, MOP and MAS research has developed an understanding for when an inferior method will provide results that are good enough for the application at hand. Combining the two upsets both of these: using a complex tool from each branch of research in tandem can create an insurmountable computation barrier, while using an inferior method that would normally produce desirable results can be upset by the complications introduced by the other branch.

PaCcET allows a low-computational method for finding arbitrary Pareto fronts without sacrificing the quality of solutions that are generated, when compared to more sophisticated multi-objective algorithms. This allows various complex techniques from the MAS body of research to be used without the computational barrier.

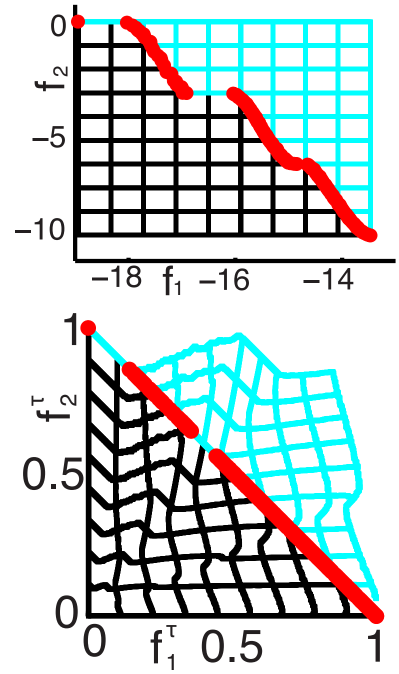

Figure 1: A grid of points before (top) and after (bottom) being transformed by PaCcET to guarantee the non-concavity of the Pareto front (red dots).

4. ON THE INCLUSION OF PRIOR KNOWLEDGE

Tabula rasa learning is arguably the worst possible approach for applying adaptive agents into the real world [5]. Though they offer high flexibility, there is often a lot of information that the system designer can provide to make the learning process easier. A framework for bestowing this knowledge to the agent is Potential Based Reward Shaping (PBRS) [3]. Often, a system designer will know something about how the agent can achieve its goals, and PBRS allows the agent to receive this knowledge. PBRS is well-developed for a single agent and single objective. Some work has been done in extending this into MASs [3], but none has addressed the complexities that arise with multi-objective PBRS.

Contribution.

We developed Difference Rewards including PBRS (DRiP), combining the agent-specificity of the difference reward with the ability of PBRS to include designer knowledge into the reward signal [4]. Using the multi-objective extension to this, MO-DRiP, we can do the same in a multi-objective setting.

Significance.

We show that this increases learning speed over using Difference Rewards or PBRS alone, and can potentially learn otherwise insurmountable tasks by leveraging the knowledge that a system designer has about how the system should work. Using DRiP does not change the Nash equilibria in the system beyond any disturbances created by Difference rewards alone.

5. REFERENCES

[1] A. Agogino and K. Tumer. Evolving distributed agents for managing air traffic. In *Proceedings of the Genetic and Evolutionary Computation Conference*, 2007.

[2] I. Das and J. E. Dennis. A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multicriteria optimization problems. *Structural Optimization*, pages 63–69, 1997.

[3] S. Devlin and D. Kudenko. Theoretical considerations of potential-based reward shaping for multi-agent system. *Autonomous Agents and Multiagent Systems*, 2011.

[4] S. Devlin, L. Yliniemi, D. Kudenko, and K. Tumer. Potential-based difference rewards for multiagent reinforcement learning. *Autonomous Agents and Multiagent Systems*, 2014.

[5] S. Whiteson, N. Kohl, R. Miikkulainen, and P. Stone. Evolving keepaway soccer players through task decomposition. *Genetic and Evolutionary Computation Conference*, 2003.

[6] M. Wooldridge. *An Introduction to MultiAgent Systems*. John Wiley and Sons, 2002.