

Issues in Working on Biological Applications of Evolutionary Computation

Clare Bates Congdon

Department of Computer Science, Colby College,
5846 Mayflower Hill Drive, Waterville, ME 04901 USA
ccongdon@colby.edu

Abstract. Biological applications of evolutionary computation (BioGEC) is a relatively recent interdisciplinary research community. This paper seeks to outline some of the difficulties experienced by researchers in this area and to provide some initial thoughts about addressing these issues as a starting point for discussion during the workshop.

1 Introduction

Succeeding at interdisciplinary research is often as challenging as it is rewarding. In addition to requiring knowledge of two or more disparate fields, there are usually different assumptions about appropriate research methodology to follow, and different reporting conventions to be wrestled with. In addition, publication of interdisciplinary research typically requires a difficult balancing act, trying to explain relevant background and assumptions in multiple fields while communicating research contributions, often within a restricted page limit.

In this paper, I seek to outline some of the relevant issues in working with applications of evolutionary computation to biology while offering few concrete suggestions for how we should proceed as a community. Instead, it is my hope that during the workshop, these thoughts will provide the seeds for discussion and help us to begin to resolve these issues as a community and to establish methodological standards for our work.

2 Some Issues

2.1 Can two researchers working in Biological Applications of Genetic and Evolutionary Computation talk to each other?

This is perhaps the most telling question of the conundrum. On the one hand, working on BioGEC problems is a highly specialized area and a relatively small community. On the other, the breadth of problems that “BioGEC” might span is enormous. If I am working on phylogenetics and you are working on protein folding, we might have very little understanding of the background of each other’s work. This problem is exacerbated if my formal training is in computer science and your training is in biology, but exists even if our background is in the same field.

2.2 What is an appropriate research methodology?

Researchers with formal training in biology will naturally tend towards different research methodologies than those with formal training in computer science. For example, when using data to build a model, it is typical in the biology community to use all the data to build the model. The thought is that if we trust the modelling tools (such as statistics), we will get the strongest model by forming it using all of the available data. Computer scientists, on the other hand, are often more focused on the question of how much one can trust the tool, and the use of cross-validation techniques (building the model with some of the data and evaluating it on the remainder of the data) is the norm.

2.3 What is an appropriate reporting methodology?

Reporting BioGEC research presents an challenge for authors, primarily in the difficulty of determining how much relevant background to include, often while confronted with page limits. On the biology side, it is difficult to gauge the expected or reasonable background of readers, in order to determine how much information is required for others to understand the scope of the work. On the computer science side, it is difficult to know how many technical details to include, particularly if one's audience turns out to be more interested in the solution to the problem than the means of achieving it.

2.4 How can we keep abreast of relevant work, which happens in multiple research communities?

One of the difficulties of working in BioGEC is that researchers tend to have a home in a particular discipline (such as Biology or Computer Science) and are naturally more familiar with work published in journals or conferences primarily associated with that discipline. At GECCO, for example, the majority of researchers tend to come from computer science, and may be unfamiliar with related work published in Biology journals. While researchers may be disciplined enough to do literature searches in other fields, they often do not regularly read publications in other fields, so may fall behind in their knowledge of current results and practice.

2.5 What are responsibilities of editors and organizers in regards to these issues?

In part because BioGEC research is a relatively new niche and we are just beginning to understand these sorts of growing pains, editors and organizers tend to sidestep the problems. For example, calls for papers do not provide clarifying information about the intended audience, and papers are rejected for missing the mark although the research contribution may be sound.

3 Some Suggestions

I have only begun to wrestle with the issues to be faced in working on BioGEC research, and can't claim to have arrived at any concrete suggestions for improvements. Instead, I offer the following possible suggestions as starting points, with the full expectation that workshop attendees will have additional thoughts to contribute:

1. We should assume that other BioGEC researchers are familiar with the basics of evolutionary computation and the central dogma. Granted, not all researchers possess this background; my suggestion here is that they be expected to acquire it as they intend to present their work and absorb the work of others in the BioGEC community.
2. Authors should state up front their methodological assumptions. This deliberately sidesteps the issue of what is "right" or "wrong" in terms of assumptions inherited from parent disciplines, while helping to clarify for the audience how to regard the work.
3. Authors should assume basic background, as outlined in suggestion 1 (and mitigated by suggestion 5, below), and provide minimal relevant details as necessary beyond that, with sufficient references for readers to learn more about the domain and the techniques.
4. As a community, we should develop resources. For example, we should develop a website listing relevant conferences and journals, as well as specific publications; this listing should include publications in non-interdisciplinary venues.
5. Editors and organizers should acknowledge their responsibility in terms of addressing these problems. Calls for papers should be clear on the expected audience and the expected form of contributions.

4 Conclusions

It is not my intentions to solve the problems above in this short work. Nor do I suspect that I've identified all the issues we might need to talk about. Rather, my hope is to identify some of the relevant difficulties of working in the BioGEC area and to inspire a healthy conversation during the workshop.