Going for algorithm gold

The use of organized competition to evaluate algorithm performance would be very beneficial for small communities, not just large ones.

What athlete has not dreamed of an Olympic gold medal? This month, once again, we will be reminded of how this granddaddy of individual competitive prizes can push ever-higher levels of performance. Outside the Olympic stadium, competitions between engineers for large monetary prizes have sometimes been used to spur technological advances. And even when there is no medal or money at stake, scientists have jostled in organized competitions. So far these have been mostly organized by large scientific communities to evaluate the performance of prediction and data-analysis algorithms.

A prime example is CASP, or Critical Assessment of Techniques for Protein Structure Prediction, a community-wide competition to evaluate protein structure prediction algorithms. CASP's format allows anyone to submit their predictions for yet-to-be solved protein structures. After the predictions have been evaluated, the participants gather to discuss the results. CASP's success quickly spawned CAPRI, or Critical Assessment of Prediction of Interactions, and other communities have followed suit with competitions in the areas of genome annotation (EGASP), text mining (BioCreAtIve), microarray data analysis (CAMDA) and more.

These competitions are ideal for objectively defining the current state of the art in a methodology. Algorithmic methods are well suited for competitive evaluation as the use of digital data, rather than of biological samples, makes it trivial for different groups to run their methods on identical material. Allowing participants to run their own algorithm is important because many algorithms require adjustment for optimal performance.

CASP quickly highlighted the principal value of competition by separating fact from fiction in a field where many difficult-to-verify claims had been made. By bringing together competitors to discuss how various algorithms performed under a common frame of reference, such competitions also provide a context in which progress can be made quickly. New algorithms that perform better than expected by the community have a good chance at quickly influencing the direction of future developments.

Even if the outcome is not the crowning of a winner, but rather that all methods perform quite poorly, the result is a definitive answer that the field can use to move on. For instance, the finding by EGASP in 2005 that automatic gene-annotation methods could not reproduce the quality of human annotators strongly influenced the current scaling phase of the ENCODE project.

Beyond identifying winners and losers, a competition can contribute very valuable resources in the evaluation criteria and the training and benchmark datasets, which are far from trivial to define. For CASP, the organizers require the collaboration of biologists who provide organizers with unpublished structure data to evaluate the accuracy of predictions. In other areas, truth models are difficult to obtain, and testing data and procedures must be agreed upon ahead of time. Ensuring that the datasets and methodologies are unbiased and adequate for the question at hand is critical and rarely easy.

Because of these challenges, a large well-organized competition requires a substantial investment of time and effort from dedicated organizers. Setting up a website, follow up meetings and other necessary infrastructure is also costly. When the goal is to evaluate algorithms for an established community effort such as ENCODE, funding is not difficult to come by. But this may prove more difficult for smaller communities.

Yet we believe that small independent competitions are well worth the effort—and likely less costly to manage—when comparison of algorithms is otherwise challenging. On page 671, for example, Michael Saxton argues for such efforts in the single-particle tracking field.

Importantly, smaller competitions could afford more flexibility than their community-wide counterparts—for example, by adding a second round of competition when the participants meet, using new test data and a strict time limit. This would favor simple efficient algorithms over those requiring extensive tuning and analysis time. This may not be very important for structure prediction and the like but could be critical for algorithms intended to efficiently replace human data analysis.

Even one-on-one competitions could provide valuable results for the community. These could be most useful as trial competitions, generating interest in the community and providing preliminary data for a more inclusive effort.

The scale of the competition matters little. There may not be a medal or large monetary prize at stake, and the winner's bragging rights may not extend much further than the local bar. But by clearly establishing state-of-the-art and future goals, and sharing ideas on how to achieve them, competitions can expedite the development of algorithms and bring substantial benefit to the community. As Pierre de Coubertin, founder of the Olympic Committee, has said: "The important thing is not to win, but to take part."