To compete or not compete? Ingredients for a successful competition

Maria Gini
Department of Computer Science and Engineering
University of Minnesota
gini@cs.umn.edu

Abstract

As the number of competitions in the AI/Robotics field grows larger the question of what is the value of participating becomes more relevant and harder to answer. We look at multiple competition we have participated to, make some observations on them, try to determine what ingredients are needed for a successful competition, and raise questions for the future.

1 Introduction

The number of international competitions in the AI and robotics field has grown over the years from a few to so many that making a choice about which competitions to participate as become imperative. We report briefly on our experience over the years in multiple competitions and provide comments on what helps a competition being successful and provide suggestions for the future.

2 Robotics competitions at AAAI/IJCAI: The early years

A team from the University of Minnesota has competed for multiple years in the AAAI/IJCAI robotic competition. Our first entry was in 1995 with a small robot, named Walleye after the state fish, built out of a remote car chassis controlled by a stack of home designed boards that included three 6811 microprocessors and a small black and white camera [Gini, 1997]. The task was to look for styrofoam cups and empty cans, pick them up, and deliver them to either a trash or a recycling bin.

Robot competitions are hard to plan and organize because of the huge differences among the hardware of the robots. It is very challenging to come it up with a task that is sufficiently difficult yet doable. Robotic competitions at AAAI/IJCAI soon ended up being uninteresting for the public, with the robots moving slowly to avoid penalties for running into obstacles and often failing miserably because of simple changes in the lighting conditions. Our entry in the 1995 competition, Walleye, lost in the final competition run because camera flashes from the public blinded it while it was searching for cups and places to deliver them. This type of competition has quickly become uninteresting and the major research groups stopped attending it, leaving it to teams of undergraduates.

From the comments above, we could think that organizing a competition with real robots is a bad idea. However, RoboCup is a counter example of an extremely successful competition. RoboCup has avoided the problem of being boring by starting with an extremely challenging task that is entertaining for the public and that puts two teams of multiple robots in direct opposition to each other. The multiplicity of robots makes the game dynamic, having two teams playing against each other creates close interactions among the robots that maintain the attention of the public. Each game is relatively short, so many games can be played during the competition maintaining the attention of the public.

3 The Trading Agent Competition for Supply Chain Management

The Trading Agent Competition for Supply Chain Management (TAC SCM) was a carefully designed competition added to the pre-existing Trading Agent Competition for travel (later renamed classic) to bring new problems and challenges to the community. TAC SCM is interesting to many researchers because it provides a competitive environment in which dynamic, agent-based supply chain methods can be evaluated without the costs and risks associated with a real-world supply chain.

Despite the careful design, the first competition in 2003 showed a serious consequence of the way orders were accepted on the first day, which made the outcome of the competition almost dictated by the random choice made at the start of the game of what agent orders to fulfill first. This prompted a redesign of the rules in 2006, which has remained stable until now [Collins *et al.*, 2005].

If we analyze what factors contributed to the growth of the TAC SCM community we can see a multiplicity of them:

1. Stability of the game specifications. This is very important since it enables to use previous years agents with more current agents.
2. Existence of an agent repository. The repository at http://www.sics.se/tac/showagents.php includes multiple agents that teams have made available to the community. Those are real complete agents that have competed in previous competitions. Most agents are avail-

able in binary form, but some are available in source code. Their availability allows new entrants to get up to speed fast and provides a more even competition ground.

3. Availability of software tools created by the community to help developers. In addition to the game server and the Java AgentWare, which is a complete agent made available to all the teams as an example of a fully running agent, various tools have been contributed by the TAC SCM community.

4. Availability of tools to support running experiments and collecting data. This is specially important to enable teams to collect empirical data for publications. In TAC SCM availability and prices of parts in the procurement market, and unmet demand and prices in the customer market, are influenced by both the mix of agents, known as the *profile space* [Wellman *et al.*, 2006], and by random variations in supply, demand, and other market parameters, known as the *market space* [Sodomka *et al.*, 2007]. The size of the profile space is large, but researchers have control in the choice of the set of competitors. The market space is not controlled by the researcher, but by the TAC SCM server. However, we have made available a controlled server [Sodomka *et al.*, 2007] to give control over the market space back into the researcher's hand. In TAC SCM the server runs with hard time constraints, sending out messages every 15 seconds, so the only variability between runs is caused by the random numbers generated by the server. Once the random number are captured and stored, then the game can be repeated multiple times, reusing the random numbers and enabling researchers to compare results and analyze strategies more easily.

5. The availability of a large number of scientific papers published in multiple venues. The papers describe the approaches taken by different teams, their solutions, analysis, comparisons, etc. The Trading Agent Competition has a particularly strong organizational structure, with a Web page that has links to all the relevant information and to published papers. TAC SCM dominates the other Trading Agent competitions with 54 papers published from 2003 to 2007. There 51 additional papers describing the individual agents. The papers are available at http://tradingagents.org/research-reports/ Even though the number of the papers does not say anything on the quality of the papers, a quick look at the list shows many publications in high-quality venues. Most of the papers are published by teams that have participated to the competition, but a few are published by researchers who have not been in the competion but have used the server and other agents to pursue their own research agenda.

6. The TADA workshop, which has been held every year at the time of the semi-final and final runs, has helped form the community by publishing research results and creating a forum for scientific discussion. The fact that TADA is a small single track workshop helps the community building.

4 Multi-agent search and rescue competition at RoboCup

This competition started in 2001 and was modeled after the 1995 Kobe earthquake [Kitano and others, 1999]. The idea is to provide emergency decision support after a urban disaster. The competition has multiple parts, namely a simulation part and a Real Robots part that is run in the NIST rescue arenas. The Real Robots competition does not require the robots to be fully autonomous.

The simulation part is divided into two parts, one done with Virtual Robots, who are simulated robots operating in a 3D simulated environment, the other is an Agents Simulation, where agents such as police cars, ambulances, and fire brigades operate in a city to rescue civilians trapped into collapsed buildings.

This latter simulation would seem to be of great interest to the search and rescue community, yet participation is limited to a relatively small number of teams. Having multiple competitions with similar names adds confusion. The fact that the RoboCup Rescue Agents Simulation is divided into two competitions, one called Infrastructure Competition and the other Rescue Agents Competition does not help. This years 16 teams have qualified for the Agent simulation and 5 for the Infrastructure competition.

The competition is not as well organized as other parts of RoboCup. The publications resulting from the competition are few, with no organized effort to share them and make them accessible on the web. We started a web page where we list the publications we have found. So far we have found a dozen of papers that are accessible. We are likely missing some, but overall the numbers are not as large as for TAC SCM. RoboCup runs a large symposium each year where relevant papers are published, so fewer papers are published in other venues. This limits the visibility of the competition to the community that is already involved into it, and makes it harder for new people to know about it and consider entering.

Changes to the rules and to the implementation of the communication makes it impossible to run agents from previous years. While it is important to improve the simulation, this limits significant the reusability of previous years agents and makes comparing results across years impossible. Things are even worse, since there are multiple versions of the system that are incompatible. The documentation is spotty, specifically for the new additions.

So far we have not been able so able to create a controlled version of the server to enable repeatability of experiments, as we did for TAC SCM.

5 Metrics for success

No attempt has been made so far in establishing metrics to measure success of competitions. Given the significant effort that is required to enter a competition, it is time to try to get the community to agree on metrics.

1. An obvious metric is number of papers published on work related to the competition. Since competitions are mostly designed to spur research activities, lack of publications should be a warning sign. Counting papers alone

is obviously not good, since we need to look at the quality of the publications. Metrics such as selectivity rate could also be used, but have their own limitations. The number of citations could also be used but it harder to track them and they are not always significant. Finding a way of measuring the influence of the publications generated by the competition on research outside the competition would be valuable but it would be hard to automate it (how to decide when research has been influenced by the competition) and hard to decide how to quantify it (are major influences to be counted more than minor and how to scale them).

2. Number of participants. The more teams participate the more successful the competition is. This would seem a reasonable measure, The trend is to split the competition into smaller competitions by creating new competitions that focus on different aspects of the problem. While the creation of new competitions is a sign of the vitality of the community, it is also damaging since it reduces the number of agents in each competition and hence limits the value of competing.

6 Conclusions and open questions

As seen in the examples described, the choice of task and competition conditions matter significantly. Having an interesting task is not sufficient to engage a broad community, as shown in part by the the RoboCup Rescue Agents Simulation.

To start a competition a significant effort is needed up front to design the rules of the competition and to ensure they give rise to interesting scientific problems. For competitions done in simulation, the design of the server is critical and so is the consistency of its implementation with the rules.

The availability of the server, either by downloading or by connecting via Web, is also important, since teams need access to experiment. Some competitions, such as the AAAI Poker competition (see http://www.cs.ualberta.ca/ games/poker/), have chosen not to allow the agents to connect to a server, but run the agents code at the site that runs the competition. While this choice reduces the chances of malevolent software and enables large runs, it makes the competition more opaque.

Having a group that maintains control of the game, its specifications, and coordinates the community, as done in TAC SCM, is an important factor that helps the creation and growth of the community.

A question that remains open is how to decide when it is time to stop a competition. Simply waiting for the community to disappear is not a good choice. Should a decrease in the number of publications be considered a sign that the competition has run its course?

References

[Collins *et al.*, 2005] John Collins, Raghu Arunachalam, Norman Sadeh, Joakim Ericsson, Niclas Finne, and Sverker Janson. The supply chain management game for the 2006 trading agent competition. Technical Report CMU-ISRI-05-132, Carnegie Mellon University, Pittsburgh, PA, November 2005.

[Gini, 1997] M. Gini. Designing and building autonomous minirobots. *Computer Science Education*, 7(4):223–237, 1997.

[Kitano and others, 1999] H. Kitano et al. Robocup-rescue: Search and rescue for large scale disasters as a domain for multi-agent research. In *Proc. of IEEE Conference SMC*, 1999.

[Sodomka *et al.*, 2007] Eric Sodomka, John Collins, and Maria Gini. Efficient statistical methods for evaluating trading agent performance. In *Proc. of the 22th Nat'l Conf. on Artificial Intelligence*, pages 770–775, 2007.

[Wellman *et al.*, 2006] Michael P Wellman, Patrick R. Jordan, Christopher Kiekintveld, Jason Miller, and Daniel M. Reeves. Empirical game-theoretic analysis of the tac market games. In *AAMAS06: Workshop on Trading Agent Design and Analysis (TADA/AMEC)*, 2006.