Effects of Game Tournaments on Learning and Classroom Climate

Michael Wollowski, J. P. Verkamp

Rose Hulman Institute of Technology
5500 Wabash Ave.
Terre Haute, IN 47803, USA
wollowski@rose hulman.edu, verkamj@rose hulman.edu

Abstract

In this paper, we motivate and describe a term long game tournament. The tournament has different milestones, including breadth first search, heuristic search, knowledge representation and reasoning and the tournament proper. For this last milestone, students are encouraged to use machine learning techniques. We describe in detail each of the milestones as well as the game and its specifications. We discuss concerns related to competitiveness and classroom climate as well as ways in which we addressed them. We present and evaluate data gathered about the competitive nature of our students as well as their learning experience with the tournament.

Introduction

Our AI course has had a game tournament for at least ten years. Many students consider it the highlight of this course. We once eliminated the tournament from the course only to be told in the course evaluations to put it back. Our students are mostly male and mostly very competitive, as such it should come as no surprise that they enjoy the competitive aspects of a tournament. However, it is our desire to create a classroom climate that encourages cooperation and welcomes students whether they enjoy competition or not.

To study the effects of tournaments on a healthy learning environment, we employed the help of the Office of Institutional Research, Planning and Assessment (IRPA)—an office internal to our institution—to gather appropriate data and evaluate it.

In this paper, we describe the game tournament, its milestones, the format of the tournament, how it is graded and the data that was gathered. We evaluate the data, describe some of the challenges we faced and suggest ways of improving the tournament.

Game Tournament as a Motivating Factor

Many students enjoy computer games. The MacArthur Foundation announced a $2 Million competition to develop video games that teach science and math [MacArthur Foundation]. Our students spend a good amount of time playing them. According to a survey we took, 64% of the students in our class play between 1-4 hours of video games **per day**. Combined with our students' desire to do well, it seems natural to use games and game tournaments to maximally motivate students.

Competition does not have to be detrimental to classroom climate. There are those who distinguish between "healthy" and "unhealthy" competition. John Shindler characterizes *healthy competition* as follows [Shindler]:

1. The goal is primarily fun.
2. The competitive goal is not "valuable/real." And it is characterized that way.
3. The learning and/or growth goal is conspicuously characterized as valuable.
4. The competition has a short duration and is characterized by high energy.
5. There is no long-term effect from the episode.
6. All individuals or groups see a reasonable chance of winning.
7. The students all firmly understand points 1-5 above.
8. Examples include: Trivial contests, short-term competitions for a solely symbolic reward, lighthearted challenges between groups where there is no reward

Shindler goes on to characterize *unhealthy competition* as follows:

1. It feels real. The winners and losers will be affected.
2. The competitive goal/reward is "valuable/real," and is characterized that way.
3. The learning task is characterized as a means to an end (winning the competition).
4. Winners are able to use their victory as social or educational capital at a later time.
5. Competition implicitly or explicitly rewards the advantaged students.

6. Over time students develop an increasingly "competitive mindset."
7. Examples include: Long-term point systems, competition for grades, grading on a curve, playing favorites, Awards for skill related performance.

Our goal was to examine the tournament and the way we run it to study its effect on competitiveness and classroom climate.

Recent History of the Tournament

In the past, the tournament was known as the "game competition." We decided to change that name so that it helps us think about this assignment in the right frame of mind and were hoping that our students would do the same.

We typically assign a different game each year. This is primarily to keep the game fresh. A second reason is that our teaching assistants run the game tournament and they get a distinct sense of ownership when we pick new games.

Fall 2007

During the 2007 offering of this course, we used an independent game server for the first time. It was developed by one of the authors who was then a student in the course. We used the game of checkers. Several reasons influenced this choice, chiefly the fact that students can develop a solution that outperforms their own abilities.

The final competition proved to be interesting and showed a wide range of client abilities. The simplest clients showed little in the way of AI concepts while the most advanced could calculate every possible outcome as far as 10-12 moves ahead.

Fall 2008

The second iteration of the server was developed during the summer of 2008 to replace checkers with Hex. Hex was chosen both because of a higher branching factor and the relative simplicity of the rules. In addition, per proof by John Nash, it is impossible for a game of Hex to end in a tie. The only way to prevent one player from winning is for the other player to win instead.

Unfortunately, Hex proved to be too simplistic for our students and many of the clients played at almost exactly the same level. As such, whichever player made the opening move almost always won despite the implementation of the pie rule to allow the second player to swap with the first player if he or she should choose to do so.

Fall 2009

During the fall quarter of 2009, we assigned a simplified version of "capture the flag," chosen because once basic game play is implemented, there exists a wide variety of possible extensions. This way, the same basic game framework can be used for several years, with a variety of extensions to keep the game interesting. The game server was re-written from scratch.

Capture the Flag

Basic Game

While a number of different variants of capture the flag are possible, most revolve around a core set of rules. Two teams divide up the playing field, each placing one or more flags on their own side of the map. Then the players attempt to find the opposing team's flags without their own being discovered.

Specifications

We used the following specifications. Not all specifications were introduced initially. Some of the constraints were added in later milestones of the project. More about this in the section entitled "Stages of Implementation."

- Fixed 30x30 board size
- One base per team.
- Three players per team with five flags per team. All of them are randomly placed.
- Arbitrarily long walls are randomly placed on the board.
- The board is symmetric. One side gets randomly generated and then flipped to complete the board.
- Enemy flags need to be captured and brought back to your own base.
- On your half of your board, running into an enemy player will return them to their base. If they carried a flag it will return to its spawn position.
- Fog of war aspect: A player can see their own position as well as spaces that are 4 or less spaces away, using Manhattan distance. Fog-of-war only applies to visibility of players, you will see all walls.
- Everyone can see the spawn positions and the eight spaces immediately adjacent to all bases and all flag spawn points at all times, whether the flag is there or not.
- You know the original position of the flag, but if a player captures it, you only see it if the player is in your field of vision.
- You can see over walls.
- If a flag gets returned to base, it re-spawns at its original spawn location.
- Each team gets a total of three minutes per game. You may decide how much time you spend on a move.
- Players can be on the spawn point of the flags they are to catch.
- Players cannot be on the spawn point of the flags that they are to defend.

To make matches fair, half of each map was generated and then mirrored across a vertical axis. Once players cross the axis, it is possible for an opponent player to run into them

and capture them, thereby returning them to their home base. Due to the presence of walls, it is not always possible to reach all of the flags. However, since flags re-spawn when turned in at a base, it is only necessary to have one accessible flag per side. In addition, if both parties agreed, a map could be skipped and another random map generated.

Re-spawning a flag and setting a time limit rather than a move limit on the game gives students the opportunity to employ various forms of machine learning.

The server handled move validation for the clients and would send back a distinct error message for each possible error, including but not limited to players colliding with walls or each other, players attempting to pick up more than one flag, and players attempting to capture on the wrong side of the board.

In addition to the game board, the current score for each player and the current player time (for timed games) were also stored in the server and could be requested using the predetermined protocol.

Stages of Implementation

While attending the 2009 MLeXAI workshop, based on discussion with attendees of the 2009 FLAIRS conference and based on student feedback, we decided to combine several of the programming assignments into one overarching one. The projects as found on the MLeXAI web-site [MLEXAI] are designed to last a term and to feature Machine Learning techniques. Due to the short duration of our terms, we decided to develop a project in which Machine Learning would be optional. We used to assign programming assignments covering basic NLP, heuristic search, theorem provers, and as final project the game tournament. We decided to integrate some of those assignments into the game tournament, making them milestones and avoiding start-up times across projects.

The "Uninformed search," "Heuristic search" and the "Alpha/Beta" assignments were individual assignments. The final two assignments were team assignments with a target team size of two students. Two teams had a size of three students.

Uninformed Search. For the first stage of the capture the flag game, students were asked to implement breadth-first search. For this version, the board had one player, one flag and zero or more walls. Students were asked to find an optimal path between the player's starting position and the flag. Due to the placement of walls, some problems were not solvable. The clients were required to identify such problems as well.

Heuristic Search. For the second stage, we added bases, multiple players and multiple flags. A number of more difficult maps with twisting branches and narrow corridors were added. Students were asked to develop several heuristics and implement and evaluate their performance. They had to conduct experiments and write up and evaluate their results.

Alpha Beta Pruning. To demonstrate the use of alpha/beta pruning, a rudimentary opponent was made available. This opponent was integrated into the server. As such, students just needed to connect to the server to interact with the opponent. This way, students get more and more experience interaction with the server. Players were required to return the flags to their base more quickly then their opponents. This assignment was an extra credit assignment.

Reasoning. For this stage, students were asked to use a knowledge representation formalism to represent board states and to develop and implement rules of inference that determine moves to be made. This stage was tested on a number of scenarios that were already available for uninformed and heuristic search. Bases were not included in these tests and no further changes were made to the server.

Tournament. The largest number of changes were made to the server for the final programming assignment: the full tournament. In this version, players were allowed to capture enemies on their side of the map, a rudimentary fog of war was implemented, flags re-spawned after being captured, and a game timer was implemented. Perhaps due to these changes, this version of the server was the most error prone; however, by the end of the tournament, all of the major bugs had been fixed.

We used a Swiss style tournament. This format was decided based on student input. A Swiss style tournament takes place over several rounds. During each round, each player or team is matched against each other player or team. The final rankings are based on the cumulative number of wins for each player or team.

Overview of Server

The current implementation of the AI Game Server serves as the center of a client/server network designed to facilitate game play between student written clients. The server will accept incoming messages from clients and will notify the clients of any updates made to the game state.

The server was implemented in C# using Microsoft's .NET framework and is fully compatible with the Mono project allowing execution under Linux or OSX. Since the packets sent between clients and the server are all in a plain text format that is easily decodable, it is possible to write clients in a variety of languages. A sample network and protocol decoding framework was provided for Java and C# to allow students to focus on their game AI rather than underlying implementation details.

Grading

In the past, we used a grading scheme that favored good performance in the tournament. In an attempt to create healthy competition (see the earlier section on "Game

Tournament as a Motivating Factor"), we changed the grading criteria considerably, adding additional rubrics on teamwork and sportsmanship resulting in the following grading scheme:

- *Performance* (35%): 15 percentage points, if your software plays without crashing. 10 percentage points, if your software has some strategy, rather than randomly making moves. 10 percentage points are based on the tournament performance.
- *Teamwork* (15%): How well does your team work together. Please use the team evaluation form.
- *Sportsmanship* (15%): Things that make the tournament fun, non-competitive, and a good learning experience. Consider sharing ideas and lessons learned. Sharing code is not a good idea as all the code you turn in should be your own.
- *Features* (35%): Interesting and non-standard features of your system that you implemented or experimented with. Please provide a detailed write-up of your features and indicate where in your code those features are implemented.

An initial set of percentage points was determined by us, but adjusted based on class discussion.

The score of the game tournament assignment was worth as much as the three mandatory assignments combined. The 10 percentage points assigned to standings in the tournament account for 2.5% of a student's course grade. As such, the performance in the tournament is mostly a symbolic reward when it comes to course grades. Without a doubt, there are bragging rights associated with doing well in the tournament.

However, a much larger portion of the course grade was determined by teamwork and sportsmanship. Our students typically do well when it comes to teamwork, so that portion of their grade is a free-bee. Nevertheless, we do adjust this portion in case students do not contribute equitably to their team.

The sportsmanship category was introduced based on prior experience with sportsmanship like behavior. During past offerings of the course, some students helped other students to improve their software, partially to ensure that they had a good opponent to test their own software. More importantly, we wanted our students to contribute to a positive learning environment. It turned out that the addition of this category did exactly that as discussed in the "Conclusions."

Due to technical issues with the server, we were not able to play all rounds of the tournament. During the last day of classes, we proposed to the class that the 10 percentage points originally assigned to the standings in the tournament be distributed over other portions of the assignment. This proposal was accepted. Students received extra credit for any game they played and some additional extra credit if they won the game. However, no more games were played after this announcement. This is likely due to the fact that students prepared for final exams.

Data Gathered

In order to study the effect of a game tournament on competitiveness among students, the Office of Institutional Research, Planning and Assessment (IRPA), an office internal to our institution, gathered appropriate data and evaluated it. Shannon Sexton, the director of Assessment located a survey designed and widely used to assess competitiveness. The survey [Housten and Smither] consists of 20 true/false questions and is called the "Competitive Index" (CI). The referenced survey also provides a scoring key.

Scoring our students' responses revealed that 46% of our students have a competitive index of high, 36% have one in the middle range and 5% are in the low range. This confirms that our students are very competitive.

Near the end of the tournament, a second survey was offered. In it, we asked a few questions in which we attempted to ascertain how students spend their free time.

By far the most listed item is "playing video games" with 73% of the respondents listing it. The next two items in terms of frequency were "sleeping" with 50%, and "watching TV/movies" with 50%. As mentioned before, students play between 1 to 4 hours of video games a day. The most poplar kinds of games are strategy games with 68%, action games with 64% and role playing games with 59%.

We then asked some specific questions about the tournament and gave our students the ability to provide us with feedback. Below are the remaining questions. All of them required a response of one of the following: Strongly agree, agree, disagree, and strongly disagree.

7. The game tournament for CSSE 413 encourages teamwork.
8. The game tournament for CSSE 413 encourages competition among classmates.
9. The game tournament challenges me to do my best academically.
10. Because of the tournament, I felt challenged to spend more effort on the final project than I otherwise would have.
11. The grading requirements for this tournament were clearly explained to me.
12. The grading requirements for this project are appropriate.

Responses

Table 1 provides the mean responses for questions 7 to 12, using a scale of 1-4; strongly disagree to strongly agree.

It is good to learn that the game tournament encourages teamwork. We were hoping that it would not encourage as much competition among students as the numeric responses indicate. Perhaps, students do not distinguish among healthy and unhealthy competition. Furthermore, as the software was indeed playing against each other, the tournament did create competition. In the future, we will

rephrase this question so as to distinguish between competition and competitive behavior.

Survey Item	Mean
The game tournament for CSSE 413 encourages teamwork.	3.33
The game tournament for CSSE 413 encourages competition among classmates.	2.94
The game tournament challenges me to do my best academically.	3.11
Because of the tournament, I felt challenged to spend more effort on the final project than I otherwise would have.	2.72
The grading requirements for this tournament were clearly explained to me.	2.94
The grading requirements for this project are appropriate.	2.94

Table 1: Mean responses to survey items.

The tournament seems to challenge our students to do their best academically; this is certainly one of the objectives of holding a tournament. Apparently though, the tournament in itself does not challenge students to exert more effort on the final project. A follow-up question should be asked to determine whether this is perhaps due to the fact that they would have put in a 100% anyway.

The grading requirements should be more explicitly spelled out, perhaps listing concrete examples of sportsmanship and desirable features. When it comes to the appropriateness of the grading criteria, we should have invited out students to suggest what they consider more appropriate requirements.

We then analyzed the responses based on the overall project score, the scores of the four grading categories and our students' competitive index scores. Table 2 summaries significant correlations.

It is good to see that those students who found the tournament to be academically challenging put more effort into it. In the future, we need to advertise its academic challenges better.

Another important result is that students' competitiveness does not seem to affect their attitude towards the tournament. In this context, we remind the reader that our results are tentative due to the low sampling size of 18 students. However, we like to think that they are encouraging and deserving of further studies.

A positive correlation from the standpoint of the instructor is that the standing in the tournament is directly related to the effort put into the game project.

The positive correlation between the sportsmanship scores and the overall game project scores comes as a pleasant surprise. For future incarnations of the tournament, it would be useful to provide even more opportunities for students to contribute in this category.

Affect of Tournament Standing
Students who had a higher standing in the game tournament believed that it encouraged competition among classmates more than students who had lower standings in the tournament.
Students who found the game tournament assignment to be academically challenging also found it encouraged teamwork among classmates. They also reported putting more effort into their final project than students who did not see the game tournament assignment as academically challenging.
Students who thought the game tournament assignment fostered competition reported putting more effort into their final project than students who did not see the game tournament assignment as competitive.
There is no difference in tournament standing or attitude toward the tournament for students scoring mid or high on the competitiveness index. Their competitive predisposition does not seem to make a difference.
There is a significant positive correlation between the game tournament project score and the sportsmanship score.

Table 2: Correlation between survey items and scores

We closed the survey with three open ended questions:

13. How could the game tournament be improved?
14. What do you like best about the game tournament?
15. What do you like least about the game tournament?

We summarize the responses to those questions in the next three sections.

Positive Aspects

Five out of 18 students explicitly mentioned that the tournament was fun.

Several students indicated that it was interesting to see different kinds of AI perform when pitted against each other. Based on their response, they clearly saw the educational value of studying the performance of different kinds of AI in a real time environment.

Some students indicated that they performed a good amount of research into an AI technique of their interest.

Some students felt a sense of accomplishment because their software was used rather than just graded and filed away. In that context, some students felt that they had to think outside the box and be concerned about efficiency.

Some students liked the competitive aspect of the tournament, mentioning that it did indeed increase their desire to do well in it.

Some students liked the teamwork that went into the final project.

Negative Aspects

Eleven out of 18 students explicitly stated that the issues we have had with the server was their least liked aspect of the tournament.

Some students indicated that they did not have sufficient software development experience to implement their ideas or deal with the complexity of the software to be developed.

Improvements

Almost all of the students stated that they would have liked a stable server from the time the tournament was assigned.

Some students stated that they did not put as much effort into the tournament, because the server was unstable.

Several students would have liked to know the rules of game play at the beginning of the term. We introduced them gradually, making game play more and more complex. In the future, we can introduce the complete rule set at the beginning of the term, but ask them to implement limited functionality for the earlier milestones.

Conclusions

From the general state of the responses, the concept of a tournament appeals to the students. Positive aspects included the ability to play against other students in the class, interactions between teams beyond simple competitiveness, and the relatively open setup of the tournament.

Students seemed to enjoy the idea of competing against other students in their class, supporting the evidence that almost half of the class ranked "high" in the competitiveness index.

While not reflected in the comments, based on conversations with our students, they had an appreciation of interactions beyond the tournament itself. Students were encouraged to help other groups with debugging problems and with interesting features and strategies. Several of them did so. One student made a GUI available to aid in testing. A number of students, including some students who otherwise rarely participated in class, posted scores of responses to the newsgroup associated with the tournament. In this context, they aided in debugging both the server and other student's code. All in all, we had about 158 postings to the newsgroup related to the tournament.

Most of the specifications for the game play were discussed in class. A number of changes to the initial specifications, including the fog of war aspect, purely symmetric maps, and structured flag re-spawning came about through discussions in class. In this way, students have a feeling of participation in the project and as such a certain ownership in the tournament.

Another example of student ownership was the decision to move from a the tree based tournament of years passed to a Swiss style tournament where each team plays other teams at their level. We feel that the ability of even less strong AIs to continue playing in later levels of the competition helped to improve student attitudes.

As mentioned, the server was unstable, primarily because of adding some of the features rather late in the term. As it is, the interactive nature of the development—involving the students—proved to be a double edged sword, causing more problems than would have cropped up otherwise. A number of these problems persisted into the mid/late stages of the tournament, reducing the number of rounds that could be successfully completed.

On a positive note, the server is now in a state such that it can be successfully used in future years with only minor tweaks. Documentation and source code can be found in [Wollowski and Verkamp].

As documented, the server problems affected student's attitudes towards the game tournament. Students put a lot of effort into their software and want to see results. However, they did not get the satisfaction of this experience. Certainly, for some students this was just bad planning. Every team should have submitted at the very least a slightly modified version of their "reasoning" milestone. Teams that did well were generally those that did modify their "reasoning" milestone, making it more robust and adding some small amount of learning to it, primarily concerned with optimizing the length of paths. However, two teams chose to develop a very radical architecture, choosing NN and GAs as their knowledge representation. The performance of their code was hurt by the lack of testing due to the server being unstable.

Without a doubt, we were very excited about the tournament. We put a lot of thought into it including the whole-sale revision of the programming assignments so that they are milestones of the tournament. As part of this excitement, we overreached, asking students to solve problems that they were not ready to solve. For the next offering of this course, we will draw out some of the milestones, simplify the game specification, and discuss implementation related issues in class. But most of all, we intend to continue to collect data so as to be able to get more reliable data and to improve the tournament.

References

Houston, John M. and Smither, Robert D. C*ompetitiveness Index* http://tars.rollins.edu/psychology/faculty/houston/CI99.HTML

MacArthur Foundation. *$2 Million Competition Seeks Ideas to Transform Learning.* http://www.macfound.org/site/c.lkLXJ8MQKrH/b.4462309/apps/s/content.asp?ct=7682383

Russell, Ingrid and Markov, Zdravko. *Project MLeXAI*. http://uhaweb.hartford.edu/compsci/ccli/

Shindler, John. *Transformative Classroom Management.* http://www.calstatela.edu/faculty/jshindl/cm/competition.htm

Wollowski, Michael and Verkamp, J.P. *Capture the Flag Assignment* http:/www.rose-hulman.edu/class/csse/csse413/archive/CTF