Expert Systems with Applications 41 (2014) 2222–2230

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Betting system for formative code review in educational competitions

Manuel Palomo-Duarte *, Juan Manuel Dodero, Antonio García-Domínguez

Dept. of Computer Science, University of Cádiz, C/Chile 1, CP 11002, Spain

ARTICLE INFO

Keywords:
Assessment
Gamification
Competition
Software development
Code review

ABSTRACT

Using competitions to motivate students is a well-known practice that has proved to be successful. Nevertheless, grading students only through their results in the competition could unfairly limit the range of grades that each student can get: regardless of the quality of the different teams participating, one student must necessarily win the tournament, another must be second, and so on until the last place. In fact, player rankings are relative assessments that are conditioned by the performance of every student. In this paper, we propose solving this issue by making students do a code review before betting on the competition. By betting, the grade of students depends both on the performance of their own solution and the one they bet on. This way, grades represent not only coding skills, but also code analysis skills, widening the attainable range of grades and allowing for a fairer grade distribution. As a result, students that are not so proficient in coding are rewarded if they demonstrate they can do a good analysis of the source code written by others, which is a very valuable skill in the professional world. We provide a case study in an undergraduate course, showing positive results.

1. Introduction

Using competitions to motivate students is a common practice known as Competition-based Learning (CnBL). In these competitions, students are not only encouraged by what they can learn, but also by finding out how their skills are compared to their peers' skills (Burguillo, 2010). This prevents several problems in other grading systems, the most common being plagiarism (Arevalillo-Herráez, Benavent, & Ferris, 2009). Nevertheless, assessment remains as a challenge for gamification experiences (del Blanco et al., 2012). Specifically, if the grades only depend on the ranked position obtained by each student, it could unfairly limit the number of students that can obtain different grades, i.e. regardless of the quality of the different teams participating, one student must necessarily win the tournament, another must be second, and so on until the last place. In fact, the position of a player in the ranking is a relative assessment that is conditioned by the performance of her rivals. If most of the rivals are excellent, a good player would be in a rather low position. Conversely, the same player would be at the upper half of the ranking in a competition with average participants.

In this paper, we formally demonstrate how a typical grading system based on competition ranking limits the minimum and maximum number of teams that can get the different grades. We analyze different proposals to mitigate this limitation, including playing more than one tournament or rewarding other achievements, not only final ranking position, finding them all unsuitable for our course.

Then, we propose a methodology to solve this issue. It is based on including a betting assessment in the competition. This way, students have to perform a critical analysis of the code written by their peers to decide which system to bet for. As a result, grades then represent both coding and code analysis skills, widening the range of grades that students can get and allowing for a fairer grade distribution. Additionally, making a good analysis of code written by others is a very valuable skill in the professional world that is often overlooked in traditional computer science courses.

We provide a case study using a predefined board game in an undergraduate course, showing positive results. Students develop artificial intelligence modules to play the game (Palomo-Duarte, Dodero, Tocino, García-Domínguez, & Balderas, 2012). We show how a typical ranking-based grading system limited how many students could get each grade in previous years, and how our improvement solved this issue by factoring code analysis skills into the grade.

The rest of the paper is organized as follows. Section 2 analyzes the drawbacks of the classical competitive grading systems, which usually impose limitations on grade distribution. Section 3 introduces our proposal for including code review and bets to overcome this problem. The proposal is evaluated through the case study in Section 4 and the results are presented in Section 5. Section 6 comments on a selection of related works. Finally, Section 7 compiles our conclusions and outlines our future work.

* Corresponding author. Tel.: +34 956 01 54 83/57 84/57 80.
E-mail addresses: manuel.palomo@uca.es (M. Palomo-Duarte), juanma.dodero@uca.es (J. Manuel Dodero), antonio.garciadominguez@uca.es (A. García-Domínguez).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.09.020

2. Classical competitive grading systems

2.1. Limitations in classical competitive grading systems

The main hypothesis of our work is that simply grading according to the performance of a system in a competition, while being highly motivating, could also be unfair, as the final ranking depends not only on the skill of the student but also on a number of external factors.

Usually tournaments follow a playoff or round-robin format or a mix of the two. If we analyze a playoff, we observe that even a uniformly random initial ordering of the participants could result in unbalanced paths to the final. Some branches could include many skilled participants, eliminating some of them very early, while other branches could have average participants advancing until a late round.

As for the round-robin league, in which all participants play against each other, the final ranking depends not only on the individual performance of each system, but also on the relative performance of all systems. For instance, having a few brilliant participants would make getting a high position very difficult, while having many poor participants would make it easy to have an average position. Furthermore, in the case of non-zero-sum games (e.g. soccer, which awards victories with 3 points and ties with just 1 point for each team to encourage attacking tactics), the distribution of wins and ties throughout the matches can significantly impact the final ranking: a high number of ties reduces the amount of points distributed among opponents, what could increase other participants' ranking.

As a result, we can conclude that using the performance of a system in a competition as the only grading factor for its author can be unfair. Getting to a certain round in a playoff (quarter-finals, semi-finals, etc.) or finishing in a certain position of a league (upper half, in the top 4, etc.) is just a relative and not an absolute evidence of the skill of the student. As it is, this method is unsuitable for assessing the performance of a student, since it should be independent and fair and allow students to obtain any grade from the minimum to the maximum value.

2.2. Analysis by simulation

We can try to mitigate limitations in grade distribution assessing other aspects concerning player performance beyond final rank can lead to a more detailed grade. For example, we could consider how many matches were won, whether a minimum number of points was reached in the league or whether certain goals were achieved throughout the matches (e.g. scoring at least one goal in a soccer match or capturing more than two pieces in a chess match). Alternatively, we could provide a set of predefined teams to be beaten as a test bed and award students that defeat them with additional points. These actions can lead to a more detailed grading system that acknowledges other achievements besides final rank, or that at least makes it easier to get the high grades that students want. Nevertheless, usually teams in higher ranking positions are more likely to get *positive* achievements than those in lower ones, so the difference will probably persist somehow. In any case, those grading systems still depend on the relative performance of the rivals and therefore limit the grade distribution significantly.

To illustrate these issues, we decided to simulate a course which graded n students depending on the results of a league and a playoff. All students were assumed to have defeated a sparring team, earning 5 points. Each of the following achievements would provide an additional point, up to 10:

- Defeat any other opponent in the league and/or playoffs (DO). In practice, every student managed to obtain this extra point. Simulations with binomial distributions assuming an even chance of winning and losing and 15 + students produced the same results with a > 99% probability.
- End in the top half of the league (ML). The "top half" could be everyone in the degenerate case in which all matches were tied and everyone had the same points. Alternatively, if all students had different amounts of points, it would be exactly half of the number of participants (rounding down).
- Win the league (WL).
- Advance one round in the playoff (RP). If n were equal to 2^k for some integer k (i.e. if n were a power of two), exactly $n/2$ students would obtain this point. In a tournament with 16 participants, 8 would obtain the RP point.
  Otherwise, $2^k < n < 2^{k+1}$ and we would need to divide the first round into two parts. In the first part, we would reduce the number of participants from n to 2^k by playing $n - 2^k$ matches. In the second part, the winners from the first part and the $n - 2(n - 2^k) = 2^{k+1} - n$ students who did not play in the first part would compete against each other. From these formulas, we can conclude that at least $n - 2^k$ and at most $n - 2^k + 2^{k+1} - n = 2^k$ students could obtain the RP point.
  In a tournament with $n = 30$ ($16 < n < 32$) students, the first part would have 14 matches (reducing the participants to 16) and 2 students would rest. In the second part, these 2 students could win or lose, so the students with the RP point would range between 14 and 16.
- Win the playoff (WP).

These results are summarized in Table 1, indicating the number of students that may obtain each of the above achievements. Using these results, we constructed the formalism inspired on colored Petri nets shown in Fig. 1. The nodes in the network represent all the possible combinations of achievements: for instance, "MLWP" means "reached top half of the league and won the playoff" and is equal to 9 points. Initially, DS contains one differently colored token per student. Following a topological order from DO, each node propagates the tokens according to these steps:

- From every incoming edge labeled with $[m,M]$, pick a uniform random integer s between m and M (both inclusive) and extract a uniformly random subset of s tokens from the source. If s is

Table 1
Minimum and maximum students by achievement among n students.

Achievement	Node	Min	Max	Notes
Defeat sparring	DS	n	n	Everyone is required to defeat the sparring team
Defeat opponent	DO	n	n	In practice, everyone obtains it (confirmed with even binomial distributions with $n > 15$)
Top middle half of the league	ML	$n/2$	n	Maximum is degenerate case when all matches are tied
Win league	WL	1	1	
Advance one round of playoff	RP	$n/2$	$n/2$	$n = 2^k, k \in \mathbb{Z}$
		$n - 2^k$	2^k	$2^k < n < 2^{k+1}, k \in \mathbb{Z}$
Win playoff	WP	1	1	