Easychair as a Pedagogical Tool

Engaging Graduate Students in the Reviewing Process

Kartik Talamadupula and **Subbarao Kambhampati**

Arizona State University
Department of Computer Science & Engineering
Tempe, AZ 85287 USA
{krt,rao} @ asu.edu

Abstract

One of the more important aims of graduate artificial intelligence courses is to prepare graduate students to critically evaluate the current literature. The established approaches for this include either asking a student to present a paper in class, or to have the entire class read and discuss a paper. However, neither of these approaches presents incentives for student participation beyond the posting of a single summary or review. In this paper, we describe a class project that uses the popular Easychair conference management system as a pedagogical tool to enable engagement in the peer review process. We report on the deployment of this project in a medium- sized graduate AI class, and present the results of this deployment. We hope that the success of this project in engaging students in the peer review process can be used better train and bolster the future corps of AI reviewers.

1 Introduction

One of the aims of graduate level courses – especially in artificial intelligence (AI) areas – is preparing students to critically evaluate current literature that describes the state-of-the-art. The normal approaches for facilitating this involve either asking the students to present a paper in class, or having the entire class discuss the paper. While these approaches do sometimes work, they are only effective in small class settings. For larger classes, an alternative that several instructors use involves having the students enter their discussion comments relating to the paper in an online forum. While this ensures some level of participation, the forums often tend to become "write-only" due to the large amount of input – there is little incentive for the students to do critical analyses of reviews other than their own, or to defend their own evaluations and positions.

It occurred to us that such incentives are inherently (and amply) provided in all the best-of-the-breed online conference management systems (CMS). Thus the question we posed was this — what if we turned the paper reading process into a paper reviewing exercise, with the students not only writing a review of the paper, but also defending their review in peer-to-peer discussions? We evaluated the effectiveness of this idea in Fall 2013 as part of a graduate level AI course, by designing a paper reviewing course project for students.

In order to structure the project in a manner that was as close to current day conference reviewing processes as possible, we used the 'free' license provided by the Easychair CMS. Easychair offers a wide variety of customizations and options for conference organizers' use, while at the same time providing a standardized interface. Easychair also provides a host of administrative options that allow organizers to collect statistics and other information on various processes related to the peer review process, like paper bidding, review submission, review discussion etc. (see Section 2).

In the following, we describe the setup of our pedagogical experiment, as well as its results. Although our main intent in pursuing this idea was to find an effective means of getting graduate students to follow the current AI literature, we are not unaware of the consequences of this project in training future reviewers. Graduate students rarely get any explicit training in paper reviewing, and instead learn the art "on the job". This project will go a small way towards improving this situation by acquainting students with the peer review process in a setting that is close to the way that major AI conferences are run currently.

2 Project Setup

2.1 Paper Selection

As the first phase of the project, the students selected one paper each from the list of papers accepted to two recently concluded top-tier AI conferences – AAAI 2013 (desJardins and Littman 2013) and IJCAI 2013 (Rossi 2013). The selected papers together formed the global pool of papers that the students would eventually review. There are of course a few shortcomings to selecting papers that have already undergone thorough peer review; the most glaring one is the obviously reduced lack of variability in the quality of the papers. However, we felt that these papers – representing the very best work in AI over the past year – would keep the students engaged, while also providing a deeper understanding of selected topics.

2.2 Paper Bidding & Assignment

Once the global pool of papers was selected, the papers in that pool were then uploaded to the Easychair system. Stu-

Metric	Lowest	Highest
Overall	-3 (Strong Reject)	3 (Strong Accept)
Confidence	1 (None)	5 (Expert)
Clarity	1 (Very Poor)	5 (Excellent)
Soundness	1 (Very Poor)	5 (Excellent)
Novelty	1 (Very Poor)	5 (Excellent)
Format	1 (Poster)	3 (Full)
Best Paper	1 (No)	2 (Yes)

Table 1: Review Metrics

dents registered as reviewers for the mock conference, and bid on the papers in a bidding phase. The instructions to the students were the following:

- Bid 'yes' on the one paper that you selected.
- Bid 'maybe' on *at least* one other paper in the pool; bidding on more than one is strongly recommended to avoid being assigned a paper that you are not interested in.

Following the end of the bidding process, students were assigned two papers each – one of these was the paper that they had selected initially, while the other was assigned by Easychair using its global assignment process.

2.3 Review Form & Submission

Since the students were reviewing papers that had been accepted and published in archival conferences, the author identities were known to the students. However, the reviewers' (students') identities were anonymized from each other. The review form included two kinds of evaluation items: multiple choice, where the students assigned relative scores to the paper; and text, where the students provided a detailed review that justified the scores assigned by them, along with other comments about the paper. The multiple choice questions are listed in Table 1. The students were told beforehand that they would be expected to defend their reviews in the discussion phase of the project.

2.4 Review Discussion

After submitting their reviews, the students engaged in a review discussion phase. In this phase, the students read the other reviewers' comments on the paper with an eye towards the following points: (1) to note points that the other reviewers missed in their reviews; (2) to point out why their scores disagreed with the other reviewers' scores; and (3) to defend their own scores and review comments from the above criticisms.

3 Evaluation

In all, the students selected 36 papers to review (c.f. Section 2.1). In the bidding phase, there were 39 'yes' bids and 230 'maybe' bids; the average number of bids among the 39 students who input bids was thus 6.9 per student. Of the 36 selected papers, only one paper did not receive any reviews (all students assigned to it did not submit their reviews). 2 papers received one review each; 25 papers received two reviews; and 8 papers received three reviews.

On the numeric portions of the reviews, students were quite generous with their ratings of the papers. The *lowest* overall average score assigned to a paper was 1 (which corresponds to 'weak accept'), while there were three papers that received a score of 3 on all their reviews. The average reviewer score for the 35 papers evaluated was 2.08 when weighted by reviewer confidence, and fractionally lower at 2.07 when not. These scores seem to indicate a positive bias on the part of the students. However, it is hard to judge from the numeric scores alone if these high scores were due to a confirmation bias (the students knew that the papers they were reviewing had successfully undergone peer review and been published), or the high quality of the papers in the selected pool.

The text reviews contributed by the students were themselves quite long – the average review length (without taking discussion comments into account, but including the confidential remarks to the program committee) was around 428 words per review, or 930 words per paper. Additionally, students generated 193 comments during the discussion phase, for an average of 5.51 comments per paper. The longest single-paper discussion thread contained 16 comments, generated by three reviewers.

While this is not a controlled experimental study, student feedback and our experience in administering this project indicate that at the very least, the project was successful in turning the reviewing process from "write-only" mode towards more discussion. Space precludes the publication of all the reviews that were generated. However, in order to give readers an idea of the kind of reviews that were received, as well as the discussion comments, we highlight one particular review in the following text file: `http://bit.ly/1fRYRQY`.

Our hope is that broad-based implementation of this project (or variants) in other graduate level AI classes will have a positive impact on student engagement in the paper reading and reviewing process. We also plan to improve the representativeness of the set of papers chosen for review by including papers from more conferences and workshops in future iterations of this project.

Acknowledgements

We thank the students from the Fall 2013 iteration of the `CSE 571` Artificial Intelligence class who took part in this project and provided the data for our deployment. We also extend heartfelt thanks to the Easychair administrators. This research is supported in part by the ARO grant W911NF-13-1-0023, the ONR grants N00014-13-1-0176 and N0014-13-1-0519, and a Google Research Grant.

References

desJardins, M., and Littman, M. L., eds. 2013. *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, July 14-18, 2013, Bellevue, Washington, USA*. AAAI Press.

Rossi, F., ed. 2013. *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*. IJCAI/AAAI.