Improving peer review with ACORN: ACO algorithm for Reviewer's Network

Mark Flynn and Melanie Moses

Abstract

Peer review, our current system used by journals and conferences to determine which papers to accept and which to reject, has limitations that impair the quality of scientific communication. Under the current system, reviewers have only a limited amount of time to devote to evaluating papers and each paper receives an equal amount of attention regardless of how good the paper is. We propose to implement a new system for computer science peer review based on ant colony optimization (ACO) algorithms. In our model, each reviewer has a set of ants that goes out and finds articles. The reviewer reviews the articles the ants bring and the ant deposits pheromone that is proportional to the quality of the review. The reviewer assesses the paper according to the criteria specified by the conference organizers. Each subsequent ant then samples the pheromones and probabilistically selects the next article based on the strength of the pheromones. We used an agent-based model to determine if an ACO-based paper selection system will direct reviewers' attention to the best articles and if the average quality of papers increases with each round of reviews. We will also conduct an experiment in conjunction with the Computer Science Graduate Student Association conference and compare the results of our simulation.

Introduction

The peer review system is the cornerstone of modern scientific communication. It is the method for determining which research is suitable for dissemination and where it should appear. Despite the success of the current system, there are some disadvantages of peer review (Neff BD 2006). There are questions of fairness and bias towards established authors, and how big a role chance plays in determining whether a paper is accepted. Computer science publishing is based on conference proceedings. A small group of reviewers is tasked with determining which papers are suitable for presentation at the conference and later inclusion in the proceedings and also assigning them to groups based on the paper's topic. The restricted pool of reviewers means that each reviewer must assess many papers and each paper can only be seen by a few reviewers. Furthermore, each paper receives the same amount of attention from the reviewers regardless of how good the paper is.

We propose to implement a new system for computer science peer review based on ant colony optimization (ACO) algorithms. ACO algorithms have been used to efficiently allocate dynamically changing resources, such as the traveling salesperson problem and network routing (Marco Dorigo 2006). Ants communicate through the exchange of chemical signals called pheromones. Ants lay these volatile compounds on the ground if they find a resource that is useful to their nest, e.g. food, a new nesting site, building materials. The distribution of pheromones mirrors the distribution of the resource (Marco Dorigo 2006).

We used this property of ant colonies to direct the attention of the ants as inspiration for our modification to the current peer review system. In our model, each reviewer has a set of ants that goes out and finds articles. The reviewer reviews the articles the ants bring and the ant deposits pheromone that is proportional to the quality of the review. The reviewer assesses the paper according to the criteria specified by the conference organizers. Each subsequent ant then samples the pheromones and probabilistically selects the next article based on the strength of the pheromones. We used an agent-based model to determine if an ACO-based paper selection system will direct reviewers' attention to the best articles and if the average quality of papers increases with each round of reviews. Additionally, If the goal is to determine which papers will be accepted, the model can also be used to determine which papers exceed a given cutoff. For example, if the conference can only accept the top 40% of the papers they receive; those papers that are closest to the cutoff would receive the most scrutiny. We also looked at the sensitivity of the model to amount of agreement on paper quality and the degree of trust among the reviewers on convergence of the model on the target paper quality.

Methods

To test whether our ACO network would be useful for evaluating and sorting papers for a CS conference, we simulated the peer review process using an agent-based model. The agents in this model are ants that search through the papers and bring them to the reviewers. Papers were modeled by giving each one a single quality score which could be considered the "ground truth". This would be the paper's score if it was reviewed by a large number of reviewers such that further reviews would be unlikely to change the score. As shown in figure 1, each of these means was drawn from an overall normal distribution in order to reflect the diversity of papers a conference is likely to receive, and the amount of diversity was determined by the distribution's standard deviation. To account for the fact that only a finite number of reviewers will see the paper, this quality score was used as the mean of a normal distribution. The standard deviation for each paper reflects the probability that different reviewers will give the same score for a paper. This was a parameter of the model that we varied to examine the effect of reviewer agreement on the ability of the network to converge on the target paper score. To ensure that each paper is reviewed at least once, each ant randomly selects a paper until all papers have received one review. For subsequent rounds of reviews, the ants sample the pheromone trails and select the next paper for review probabilistically based on the quality of the reviews up to that point. The probability of a paper being select was determined by normalizing to the mean paper score and using this adjusted score as the exponent in the equation for calculating the probability (eqn1). The bias factor reflects how much the reviewers trust the opinions of the other reviewers.

Results

Our goal was to determine whether an ACO algorithm could direct reviewers' attention to the papers that were deemed most important, depending upon whatever criteria were considered relevant. First, we tested the ability of our algorithm to determine which papers were the best and how sensitive this determination was to variation in the parameters. We found that there was a positive correlation between how many papers a reviewer has evaluated and the quality of the papers the reviewer receives. As seen in figure 2, after each round of reviews the better papers were more likely to be selected for review. We also found a positive correlation between the quality of papers and the number of reviews the paper received. Figure 3 shows that the number of ants that visited a paper depended on the quality of the paper.

Two important factors that affect the ability of the model to select the best papers are 1) the amount of agreement the reviewers have on the quality of the papers and 2) how much they need to rely on each other's judgment for the algorithm to work. Too much disagreement would mean that the reviewers would not see an improvement in the paper quality, or if the goal is to decide which papers to accept, convergence on the target paper quality. We modeled this by varying the standard deviation for the normal distribution that paper scores are drawn from. This interacts with the degree of trust between the reviewers. While too little trust would mean that reviewers would receive randomly selected papers, too much trust could amplify the effects of disagreement among reviewers. To explore these interactions, we ran all combinations of bias factors between 1 and 2, in increments of 0.1 (completely random and completely deterministic paper selection) and paper standard deviations of the integer values between 0 and 4 (complete agreement and complete disagreement). 20 simulations were run with each combination to average out the stochastic nature of the paper selection processl. Figures 4 shows the effects of these interactions on the correlation of the number of papers a reviewer has evaluated up to that point and the quality of the papers reviewed. Figure 5 shows the corresponding effects on the total number of reviews a paper receives. While the effect on the improvement in quality of the papers over time was fairly insensitive to the level of trust and agreement of the reviewers, the number of reviews a paper received based on its quality was very dependent on agreement. As the amount of variability in reviewer scores increased, the correlation between paper quality and the number of reviews a paper received decreased.

Conclusion

We found that our system for changing the peer review process can successfully direct reviewers' attention to the papers that are either the best, or the closest to the cutoff for acceptance. The number of reviews a paper received was more sensitive to the amount of agreement among the reviewers than the change in quality over time. Changing the goal to selecting papers nearest to a cutoff for acceptance is equivalent to selecting the best paper, since it is only a matter of rescaling the values so that the papers that are favored are now the ones closest to the cutoff instead of the papers that received the best reviews.

Some things we would like to improve for the future would be explicitly modeling the reviewers. In our present model, all reviewers are identical, but it might be more realistic to model the potential idiosyncrasies of reviewers by incorporating a factor that skews different reviewers' opinions differently. Some reviewers might for instance consistently grade on a different scale than others. Another possibility is that reviewers change their behavior over time. We would like to incorporate these factors to more accurately reflect the reviewing process. We would also like to include the variance of the scores in the selection process. The overall evaluation of papers that receive very similar scores are unlikely to change with further reviews and will be removed so that the reviewer's attention can focus on the papers whose fate is more uncertain. We would also like to include some of the best practices from the peer review process among the journals. For example, the Artificial Intelligence conference includes the reviewer's assessment of their experience with the topic of the paper, so that the views of experts and those not as familiar with a particular topic can have their evaluations properly weighted. Another practice we will investigate is the use of multiple levels of reviews, where the results of the review are passed to a higher level for further review.

References

Marco Dorigo, M. B., and Thomas St¨utzle (2006). "Ant Colony Optimization: Artificial Ants as a Computational Intelligence Technique." IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE 1(4): 39.

Neff BD, O. J. (2006). "Is Peer Review a Game of Chance?" Bioscience 56(4): 333-340.

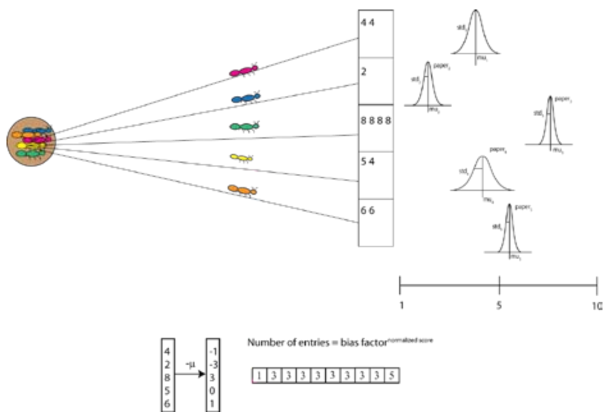

Figure 1. ACO peer review algorithm

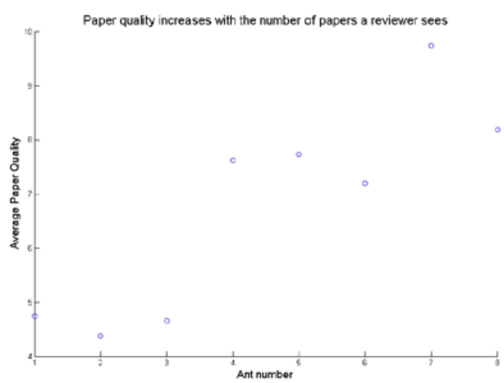

Figure 2. Paper quality improves with time

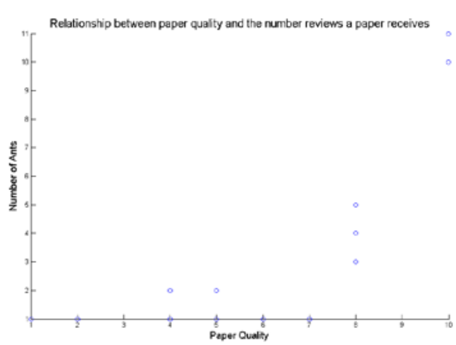

Figure 3. Higher quality papers receive more reviews

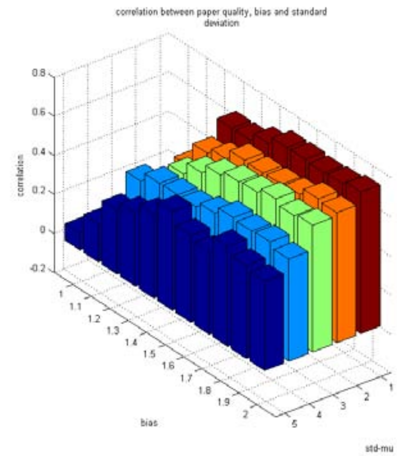

Figure 2. Decreasing trust and agreement among reviewers decreases correlation between paper quality and the increase in paper quality over time

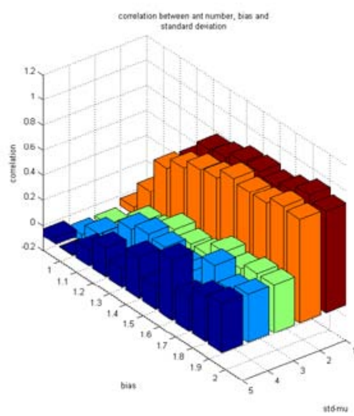

Figure 3. Decreasing trust and agreement decreases correlation between paper quality and the number of reviews a paper receives