A Framework for a Peer-to-Peer Scoring System in Collaborative Competitions

ASU Luminosity Lab

April 1, 2023

Abstract

This paper introduces a mathematical framework for a peer-to-peer scoring system tailored for digital collaborative competitions. The system is predicated on the concept of the Wisdom of the Crowds and aims to encourage unbiased and accurate evaluations by the participants. A feature of this system is that it mandates participants to evaluate a minimum number of submissions from their peers to qualify for receiving a score. Additionally, it integrates a penalty mechanism for those whose evaluations deviate significantly from the aggregate score. This penalty is reflected in the final score of the participants own submission. By incentivizing honest evaluations and fostering active engagement, this system addresses traditional shortcomings such as biases and lack of transparency. The scoring system is intended to facilitate better peer-to-peer scoring, and thus help make digital competitions more scalable through self-relience. This paper presents the mathematical underpinnings of the scoring algorithm, its implementation, and the implications for collaborative competitions. Through this approach, the system enhances the evaluation process, ensuring fairness, encouraging participation, and ultimately enhancing the quality of collaborative competitions.

1 Introduction

1.1 Background

Collaborative competitions are an amalgamation of collaboration and competition, and have proven to be efficacious in spurring innovation, creativity, and problem-solving. In these competitions, teams work together to solve challenges, and their submissions are evaluated to ascertain winners. The evaluation process, however, has been subject to scrutiny, primarily due to biases, lack of transparency, and varying standards of judgment.

Historically, evaluations in competitions have been conducted by a panel of experts or judges who assess submissions based on predetermined criteria. While this approach has been widely adopted, it often falls short in ensuring impartiality. Expert judges, despite their expertise, can exhibit biases, whether conscious or subconscious. Additionally, traditional judging can be resource-intensive, requiring significant time and effort from experts who may be limited in number.

One of the core advantages of adopting a peer-to-peer scoring system, as described in this paper, is that it significantly enhances the scalability and autonomy of competitions. Traditional competitions, especially those at a large scale, require significant human resources for organizing and judging, which often restricts the ability to scale. In contrast, the peer-to-peer scoring system automates the evaluation process, thereby drastically reducing the need for human intervention. This not only allows for competitions to be scaled to a much larger extent but also makes them more autonomous, with lesser dependence on organizers and expert judges for evaluation.

Moreover, the system's inherent scalability and autonomy are particularly advantageous in online competitions, where participants from diverse geographic locations and domains can contribute. The system can efficiently handle large volumes of submissions and evaluate them in a timely manner, making it highly suitable for global and diverse competitions.

Given the rise in the scale and diversity of collaborative competitions, especially in the online space, it is essential to reexamine and improve evaluation methodologies. Addressing the inherent limitations of traditional evaluation processes is crucial for ensuring the credibility of competitions and for fostering an environment where innovation can thrive.

1.2 The Wisdom of the Crowds

The concept of the Wisdom of the Crowds was popularized by James Surowiecki in his book, The Wisdom of Crowds, which posits that the collective intelligence of a large group of non-experts can often provide outcomes that are superior to the outputs of a small group of experts. This phenomenon can be attributed to the aggregation of information in groups, which results in decisions that are often remarkably accurate.

In the context of collaborative competitions, the Wisdom of the Crowds suggests that a large and diverse set of participants could be leveraged to evaluate submissions. Instead of relying solely on a small panel of experts, competitions could employ a methodology where participants evaluate and score submissions from their peers. When these scores are aggregated, the average is likely to provide a more objective and unbiased assessment than an expert panel.

However, the efficacy of Wisdom of the Crowds as a decision-making tool is contingent on certain conditions being met, such as diversity of opinion, independence, decentralization, and a method for aggregating responses. In collaborative competitions, it is important to ensure that these conditions are satisfied for the Wisdom of the Crowds to be effectively harnessed. Independence is particularly crucial, as participants should give their assessments independently without being influenced by the opinions of others.

The adoption of a peer-to-peer scoring system based on the Wisdom of the Crowds is a significant paradigm shift from traditional evaluation methodologies.

It holds the promise of not only enhancing objectivity and fairness but also democratizing the evaluation process by giving all participants a voice in the assessment of submissions.

2 Framework for Peer-to-Peer Scoring System

This section presents a mathematical framework for implementing a peer-to-peer scoring system that aims to foster fairness and reduce biases in collaborative competitions. The essence of this framework is to ensure that participants are encouraged to provide honest and accurate scores by implementing a mechanism to detect and penalize significant deviations from the collective judgment of peers. The framework is structured into three key components: the scoring mechanism, the identification of outliers through the Interquartile Range (IQR) method, and the penalty mechanism for scores deemed as outliers. This approach contributes to making the competitions scalable, autonomous, and impartial.

2.1 Scoring Mechanism

Let S_i represent the score assigned by participant i to a given submission, and let N be the number of participants who scored that submission. The aggregate score, A, for that submission is given by the mean of all scores:

$$A = \frac{1}{N} \sum_{i=1}^{N} S_i$$

2.2 Identification of Outliers using Interquartile Range (IQR)

To determine if a participants score significantly deviates from the aggregate score, we employ the Interquartile Range (IQR) method. Firstly, the scores are sorted, and the first quartile (Q_1) and third quartile (Q_3) are determined. The IQR is calculated as the difference between Q_3 and Q_1:

$$\text{IQR} = Q_3 - Q_1$$

A score S_i is considered an outlier if it falls outside of the range defined by the following lower and upper bounds:

$$\text{Lower Bound} = Q_1 - 1.5 \times \text{IQR}$$

$$\text{Upper Bound} = Q_3 + 1.5 \times \text{IQR}$$

2.3 Penalty Mechanism

If a participants score S_i is found to be an outlier based on the IQR method, a penalty P is applied to the participants own submission score. The penalty can be a fixed value or a function of the deviation from the aggregate score.

For example, the penalty P can be defined as a proportion of the deviation:

$$P = k \times |S_i - A|$$

Where k is a constant factor.

The final score of the participants submission, F, after applying the penalty is:

$$F = \text{Original Score} - P$$

This mechanism encourages participants to provide scores that align with the collective judgment of their peers. The use of the IQR method ensures robust identification of scores that significantly deviate from the rest, thereby minimizing bias and promoting fair competition.

3 Algorithm Implementation

This section presents the pseudocode for the implementation of the peer-to-peer scoring system discussed in this paper. The algorithm comprises several steps including the calculation of aggregate scores, identification of outliers using the Interquartile Range (IQR) method, and the application of penalties.

It's worth noting that the algorithm below assumes the availability of scores provided by other users. The specific implementation details of how the scores are collected and provided to the algorithm may vary depending on the application or system in which this algorithm is utilized.

Algorithm 1 Peer-to-Peer Scoring System

```
procedure PeerToPeerScoring(submissions, scores)
    aggregates ← empty list
    for each submission in submissions do
        s ← scores for submission
        aggregate ← mean(s)
        add aggregate to aggregates
    end for

    Q_1 ← first quartile of aggregates
    Q_3 ← third quartile of aggregates
    IQR ← Q_3 − Q_1
    lowerBound ← Q_1 − 1.5 × IQR
    upperBound ← Q_3 + 1.5 × IQR

    for each submission in submissions do
        aggregate ← aggregate score for submission
        if aggregate < lowerBound or aggregate > upperBound then
            penalty ← calculatePenalty(aggregate, lowerBound, upperBound)
            apply penalty to submission
        end if
    end for
end procedure

function calculatePenalty(aggregate, lowerBound, upperBound)
    if aggregate < lowerBound then
        return (lowerBound − aggregate) × penalty factor
    else
        return (aggregate − upperBound) × penalty factor
    end if
end function
```

4 Discussion

In this paper, we introduced a peer-to-peer scoring system in collaborative competitions based on the Interquartile Range (IQR) method for outlier detection. This section discusses the implications, strengths, and potential limitations of this approach, as well as considerations for future work.

4.1 Implications and Advantages

The implementation of the IQR-based method for detecting outliers brings several implications and advantages:

- **Scalability and Autonomy**: By engaging participants in the scoring process, competitions can be scaled more effectively. The autonomous nature of the scoring system reduces the administrative burden and allows for a larger number of submissions to be evaluated.

- **Robustness Against Extreme Scores**: The IQR method is less sensitive to extreme scores compared to using standard deviations. This leads to more robust identification of genuinely deviating scores.
- **Encouragement of Honest Scoring**: The penalty mechanism deters participants from providing dishonest or biased scores, as they would risk penalization on their own submissions.
- **Community Engagement**: The scoring system can foster a sense of community, as participants become stakeholders in the evaluation process, potentially leading to richer feedback and discussions.

4.2 Potential Limitations

However, the proposed approach is not without potential limitations:

- **Over-Penalization**: There is a risk of over-penalization for honest but unconventional evaluations. This could discourage critical independent thinking and creativity among participants.
- **Data Dependency**: The effectiveness of the IQR method depends on the distribution and characteristics of the scores. In some cases, an alternative method might be more appropriate.
- **Collusion Among Participants**: Participants might collude to give each other high scores or deliberately downgrade others. Such behavior could affect the integrity of the competition.

4.3 Recommendations for Future Work

Given the aforementioned implications and limitations, we propose the following recommendations for future work:

- **Dynamic Thresholds**: Develop adaptive mechanisms for setting the IQR multiplier based on the characteristics of the data and competition, rather than using a fixed value.
- **Score Calibration**: Implement methods for calibrating participants biases. For example, if a participant consistently scores lower than the average, their scores could be adjusted upwards uniformly to reduce their bias. This requires some initial rounds of scoring for calibration.
- **Anti-Collusion Measures**: Implement algorithms to detect patterns indicative of collusion among participants and apply appropriate penalties.

- **Reputation-Based Weighting**: Introduce a reputation system where participants who consistently provide fair and accurate scores gain reputation points. These reputation points could affect the weight of their scores in future competitions, providing an incentive for sustained honest participation.

4.4 Conclusion

The IQR-based peer-to-peer scoring system introduced in this paper offers a potentially scalable and robust method for evaluating submissions in collaborative competitions. By involving the community of participants in the scoring process, it not only alleviates administrative burdens but also encourages honest and fair evaluations. While mindful of its limitations, with continuous refinement and adaptation, this approach holds significant promise for fostering innovation, collaboration, and fairness in competitive settings.