Analysis of the PeerRank Method for Peer Grading

Joshua Kline

Advisors: Matthew Anderson and William Zwicker
Benefits of Peer Grading

- Reduces time instructors spend grading
- Provides faster feedback for students
- Increases student understanding through analysis of others
Potential Issues with Peer Grading

**Issues:**
- Students may be unreliable graders
  - Inexperience in grading
  - Lack of understanding of material
- Students may not care about grading accurately

**Ways to Address:**
- Make inaccurate graders count less toward final grade
- Provide graders with an incentive to grade accurately
PeerRank

- Algorithm developed by Toby Walsh
- Two factors in final grade:
  - Weighted combination of grades from peers
  - Individual’s own accuracy in grading others
- Same linear algebra foundations as Google PageRank
- Original application: Reviewing grant proposals
PeerRank

- Start with “initial seed” grade vector $X^0$
  - Average of grades received

- PeerRank equation is evaluated iteratively until fixed point is reached
  - $X^{n+1} \approx X^n$

\[
X_i^0 = \frac{1}{m} \sum_j A_{i,j}
\]

\[
X_i^{n+1} = (1 - \alpha - \beta) \cdot X_i^n + \frac{\alpha}{\sum_j x_j^n \cdot \sum_j X_j^n} \cdot A_{i,j} + \frac{\beta}{m} \cdot \sum_j 1 - |A_{j,i} - X_j^n|
\]
Problems with PeerRank

• **Walsh’s Assumption:**
  A grader’s accuracy is assumed to be equal to their grade
  • Unrealistic assumption?

• No way of specifying “correctness”
  • May produce incorrect results

Correct Result: [1,1,0,0,0]
Problems with PeerRank

• **Walsh’s Assumption:**
  A grader’s accuracy is assumed to be equal to their grade
  • Unrealistic assumption?

• No way of specifying “correctness”
  • May produce incorrect results

Correct Result: [1,1,0,0,0]
Actual Result: [0,0,1,1,1]
Project Goal

Modify and adapt the PeerRank algorithm so that it can better provide accurate peer grading in a classroom setting.
Incorporating “Ground Truth”

• Recall: There is no way of specifying “correctness” in PeerRank.

• In education, there is a notion of “ground truth” in assignments
  • Right answer vs. wrong answer
  • Correct proof
  • Essay with strong argument and no errors

• Ground truth is normally determined by instructor
Incorporating “Ground Truth”

• **Goal:** Give the instructor a role in the PeerRank process that influences the accuracy weights of the students

  - The instructor submits their own assignment with a known grade.
  - Each student grades the instructor’s assignment, and their grading error determines their accuracy.
  - Students do not know which assignment is the instructor’s.
  - Use these accuracies to produce a weighted combination of the peer grades.
Incorporating “Ground Truth”

**Goal:** Give the instructor a role in the PeerRank process that influences the accuracy weights of the students

**Solution:**
- The instructor submits their own assignment for which they know the correct grade
- Each student grades the instructor’s assignment, and their grading error determines their accuracy
  - Students do not know which assignment is instructor’s
- Use these accuracies to produce a weighted combination of the peer grades
Our Method vs. PeerRank

PeerRank:
• Accuracy equal to grade
  • Walsh’s assumption applies
• Iterative process
  • Final grades are fixed point

\[
X_i^0 = \frac{1}{m} \sum_j A_{i,j}
\]

\[
X_i^{n+1} = (1 - \alpha - \beta) \cdot X_i^n + \frac{\alpha}{\sum_j X_j^n} \cdot \sum_j X_j^n \cdot A_{i,j} + \frac{\beta}{m} \cdot \sum_j 1 - |A_{j,i} - X_j^n|
\]

Our Method:
• Accuracy determined by accuracy in grading the instructor
  • Walsh’s assumption no longer applies
• Non-iterative
  • Final grades are a weighted average of the peer grades, weighted by the accuracies

\[
ACC_i = 1 - |A_{I,i} - X_i|
\]

\[
\tilde{X} = \frac{1}{\|ACC\|_1} (A \cdot \overline{ACC})
\]
Majority vs. Minority Case

• Recall: If a group of incorrect students outnumber a group of correct students, the wrong grades are produced by PeerRank.

Correct Result: \[[1,1,0,0,0]\]
Actual Result: \[[0,0,1,1,1]\]
Majority vs. Minority Case

- Recall: If a group of incorrect students outnumber a group of correct students, the wrong grades are produced by PeerRank.
- What if the instructor submits a correct assignment in our system?

Correct Result: [1,1,0,0,0,1]
Majority vs. Minority Case

• Recall: If a group of incorrect students outnumber a group of correct students, the wrong grades are produced by PeerRank.

• What if the instructor submits a correct assignment in our system?

Correct Result: [1,1,0,0,0,1]

Accuracies: [1,1,0,0,0,1]
Majority vs. Minority Case

• Recall: If a group of incorrect students outnumber a group of correct students, the wrong grades are produced by PeerRank.

• What if the instructor submits a correct assignment in our system?

Correct Result: [1,1,0,0,0,1]
Accuracies: [1,1,0,0,0,1]
Actual Result: [1,1,0,0,0,1]
Implementation

- Algorithms for PeerRank and our method implemented in Sage
  - Based on Python
  - Additional math operations, including matrices and vectors
  - Graphing packages

\[
X_i^0 = \frac{1}{m} \sum_j A_{i,j}
\]

\[
X_i^{n+1} = (1 - \alpha - \beta) \cdot X_i^n + \frac{\alpha}{\sum_j X_j^n} \cdot \sum_j X_j^n \cdot A_{i,j} + \frac{\beta}{m} \cdot \sum_j 1 - |A_{j,i} - X_j^n|
\]

def GeneralPeerRank(A, alpha, beta):
    m = A.nrows()
    Xlist = [0] * m
    for i in range(0, m):
        sum = 0.0
        for j in range(0, m):
            sum += A[i,j]
        X_i = sum / m
        Xlist[i] = X_i
    X = vector(Xlist)
    fixedpoint = False
    while not fixedpoint:
        oldX = X
        X = (1-alpha-beta)*X + \
            (alpha/X.norm(1))*A*X
        for i in range(0, m):
            X[i] += beta - \
            (beta/m)*((A.column(i) - \
            oldX).norm(1))
        difference = X - oldX
        if abs(difference) < 10**-10:
            fixedpoint = True
    return X
Simulating Data

- Real grade data is not easily accessible
- Data was simulated using statistical models
  - Ground truth grades drawn from bimodal distribution
  - Accuracies drawn from normal distributions centered at grader’s grade
  - Peer grades drawn from uniform distributions using ground truth grade and accuracies
Experiments

• Experiments consisted of generating class/grade data and comparing the performance of PeerRank and our modified version against the ground truth grades.

• Variables:
  • Class size
  • Grade distribution means, standard deviations
  • Percentage of students in each group
  • Accuracy distribution standard deviation

![Graph showing comparison of grades]
Reducing Connection Between Grade and Accuracy

- Recall: The original version of PeerRank assumes that the grader’s grade is equal to their grading accuracy.
  - Unrealistic assumption?

- Our method does assume any connection between grade and accuracy.

- How do the two versions compare as we reduce the connection between grade and accuracy?
  - We can model this reduction by increasing the standard deviation around the graders’ grades when drawing their accuracies.
Reducing Connection Between Grade and Accuracy

Standard Deviation = 0.02

Avg. Error Reduction < 0.1%
Reducing Connection Between Grade and Accuracy

- **Standard Deviation** = 0.02
- **Avg. Error Reduction** < 0.1%

- **Standard Deviation** = 0.10
- **Avg. Error Reduction** ≈ 0.2%
Reducing Connection Between Grade and Accuracy

- Standard Deviation = 0.02
- Avg. Error Reduction < 0.1%

- Standard Deviation = 0.50
- Avg. Error Reduction ≈ 2.3%

- Standard Deviation = 0.10
- Avg. Error Reduction ≈ 0.2%
Reducing Connection Between Grade and Accuracy

- Standard Deviation = 0.02
  - Avg. Error Reduction < 0.1%

- Standard Deviation = 0.50
  - Avg. Error Reduction ≈ 2.3%

- Standard Deviation = 1.0
  - Avg. Error Reduction ≈ 4.0%

- Standard Deviation = 0.10
  - Avg. Error Reduction ≈ 0.2%
Conclusions

• When grading accuracy is strongly correlated with the grader’s grade (Walsh’s assumption), our method produces grades extremely close to PeerRank.

• When grading accuracy is unrelated to the grader’s grade, our method produces more accurate grades than PeerRank.

Correct Grades
Grades from Our Method
PeerRank Grades

Standard Deviation = 1.0
Avg. Error Reduction ≈ 4.0%

Standard Deviation = 0.02
Avg. Error Reduction < 0.1%
Future Work

• Implementation of a “partial grading” scheme
  • Ignore missing grades?
  • Fill in missing grades based on known grades?
  • Best way of dividing the class?

• Additional methods of integrating ground truth
  • Instructor grades a certain number of students with a high accuracy score

Questions?