Grouping Students in Educational Settings

Rakesh Agrawal  
rakesha@microsoft.com

Behzad Golshan  
Behzad@cs.bu.edu

Evimaria Terzi  
evimaria@cs.bu.edu
Introduction

- Consider a class of students
  - Different ability levels (single scores)
    - Example: GRE, TOEFL, SAT, ...

How to form study groups?
Introduction

- Classical methods
  - Ability-Based Grouping
    - Grouping students with similar abilities together
  - Pseudo-Random Grouping
    - Grouping students based on some arbitrary ordering
    - Alphabetically, FCFS, ...
Introduction

- Classical methods
  - Ability-Based Grouping
    - Grouping students with similar abilities together
  - Pseudo-Random Grouping
    - Grouping students based on some arbitrary ordering
    - Alphabetically, FCFS, ...

Which method to use?
Inconclusive verdict from empirical studies
(Kulik 92, Loveless 13, McPartland 87)
Let’s take a computational approach
Our Results

- Grouping strong students with not much weaker students
Our Results

- Grouping strong students with not much weaker students
Our Results

- Similar structure with different distributions of abilities

![Graphs showing Normal Distribution, Uniform Distribution, and Pareto Distribution.](image)
Our Results

- Classical methods are not optimal
  - With respect to our objective
Framework

- Set of $n$ students with abilities $\theta_1, \theta_2, \ldots, \theta_n$
  - Ability scores are real number ($\theta_i \in R$)

- Collective Ability of a team $T$
  - Represents the group ability
  - Expected Ability
    - Choose a random student and ask him

$$\hat{\Theta}_T = \frac{1}{|T|} \sum_{i \in T} \theta_i$$
Framework

- Two groups of students in a study group
  - Students below the collective ability
  - Students above the collective ability
Framework

- Two groups of students in a study group
  - Students below the collective ability
  - Students above the collective ability

- Followers ($F_T$)
  - Mostly learn from other members of the group

- Leaders ($L_T$)
  - Mostly improve by teaching others
Framework

- Two groups of students in a study group
  - Students below the collective ability
  - Students above the collective ability

Followers \((F_T)\)
- Mostly learn from other members of the group

Leaders \((L_T)\)
- Mostly improve by teaching others

Our Focus
- Maximize the number of such students
Problem

- Partitioning students into study groups
  - Partition students into $l$ groups of size $k$ to maximize the gain
    - Gain = sum of the number of followers in each group

Theorem:
- NP-hard to solve
- PARTITION problem reduces this problem
Algorithm

- Partitioning students into study groups
  - Partition students into $l$ groups of size $k$ to maximize the gain

- Algorithm:
  - Find the best team of size $k$ from the pool of students
  - Remove the team from the pool
  - Repeat until all groups are formed
### Algorithm

- **Partitioning students into study groups**
  - Partition students into $l$ groups of size $k$ to maximize the gain

- **Algorithm:**
  - **Find the best team** of size $k$ from the pool of students
  - Remove the team from the pool
  - Repeat until all groups are formed

- **Best Team**
  - Team with the maximum gain (i.e., number of followers)
  - How to find the best team?
Finding the Best Team

Observation 1

Pick the best students
Observation 2

The followers are consecutive
Finding the Best Team

**Algorithm**

- How many leaders?
  - Try all values of $x$ (i.e., number of leaders)
- Who are the followers?
  - Try moving the sliding window
Finding the Best Team

Algorithm

- How many leaders?
  - Try all values of $x$ (i.e., number of leaders)
- Who are the followers?
  - Try moving the sliding window
- Satisfying condition?

- Test $O(n \log(k))$ groupings
Experiments (Count-\(lG\))

- Different distribution of student abilities
Alternative Formulations

- General Framework

\[ A(T) = \sum_{i \in F_T} A_f(i, T) + \sum_{i \in L_T} A_\ell(i, T) \]

- Other Gain functions
  - How much do followers learn?
  - See the paper for more details

Gain Function
Gain (leader)
Gain (follower)
Summary

- Traditional methods are not optimal
- Different objectives leads to different team structures
- Computation approaches can reveal such optimal structures

Future Work

- Richer gain functions
  - Gain for the leaders
  - Non-linear gain functions
- Incorporating constraints due to socio-emotional factors
- Evaluating our method in real-life settings
Thanks

KDD 2014
Thanks
Single Group

Observation 2

Strongest follower is weaker than collective team ability

\[ \hat{\Theta}_T = \left( \sum_{i \in L} \theta_i + \sum_{i \in F} \theta_i \right) / k > \max_{i \in F} \theta_i \]

Or

\[ \sum_{i \in F} \theta_i > k \times \max_{i \in F} \theta_i - \sum_{i \in L} \theta_i \]

1) Maximize this

2) Pick students weaker than \( \max_{i \in F} \theta_i \)

Feasibility Condition
Framework

- Two groups of students
  - Different benefits from participating

\[
A(T) = \sum_{i \in F_T} A_f(i, T) + \sum_{i \in L_T} A_\ell(i, T)
\]

Gain Function ➔ Gain (follower) ➔ Gain (leader)
Framework

- Two groups of students
  - Different benefits from participating

\[ A(T) = \sum_{i \in F_T} A_f(i, T) + \sum_{i \in L_T} A_\ell(i, T) \]

- Gain Function
- Gain (leader)
- Gain (follower)

- Our focus: Gain for the followers