Poster: Improving Formation of Student Teams: A Clustering Approach

Shoaib Akbar
Department of Computer Science
University of North Carolina Charlotte
Charlotte, NC, USA
sakbar1@uncc.edu

Edward F. Gehringer
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
efg@ncsu.edu

Zhewei Hu
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
zhu6@ncsu.edu

ABSTRACT

Today's courses in engineering and other fields frequently involve projects done by teams of students. An important aspect of these team assignments is the formation of the teams. In some courses, teams select different topics to work on. Ideally, team formation would be included with topic selection, so teams could be formed from students interested in the same topics. Intuitive criteria for a team formation algorithm are that students should be assigned to (1) a topic which they have interest and (2) a team of students with similar interests in their topic. We propose an approach to meeting these criteria by mining student preferences for topics with a clustering approach and then matching them in groups to topics that suit their shared interests. Our implementation is based on hierarchical k-means clustering and a weighting formula that favors increasing overall student satisfaction and adding members until the maximum allowable team size is reached.

KEYWORDS

student team formation, bidding, clustering, peer assessment system, MOOCs

1 INTRODUCTION

Learning teaming skills is an important part of education. Good teaming skills are in great demand in many fields, particularly engineering [7]. How these teams should be formed is an important research question [2]. Some approaches, like Team-Based Learning [6], argue that teams should be formed by the instructor and remain static throughout the course. Other philosophies, like Agile Methodologies [3], posit that teams should form eclectically and change over time.

But how should eclectic teams be formed? If the teams are working on different topics, it seems reasonable for students to indicate which topics they are interested in. Then, teams could be formed algorithmically based on student preferences. When students do multiple team projects during a course, the team-formation algorithm can be run for each, likely assigning students to different teammates each time. This mitigates the problem of students working just with their friends. A diversity of topics and a diversity of interests should tend to lead to the formation of diverse teams. Moreover, in large classes, or distance-ed classes, there are bound to be some students who know few, if any, other class members. These students would be disadvantaged in forming self-selected teams, but when teams are formed based on topic preferences, all have an equal chance to find partners. This argues against a first-come first-served (FCFS) approach, which would give greater control to students who are able to select early. In distance-education classes, for example, work and family responsibilities may prevent students from selecting topics right away.

ICSE '18 Companion, May 27-June 3, 2018, Gothenburg, Sweden

ACM ISBN 978-1-4503-5663-3/18/05.
https://doi.org/10.1145/3183440.3195057

2 TEAM FORMATION APPROACHES

2.1 First-Come First-Serve (FCFS)

Expertiza [4] is an online peer-assessment system developed and based at North Carolina State University, which is used at 17 other institutions. It provides several services including team formation, topic selection, teammate review, calibrated peer review [11], peer review collusion detection [10], and a reputation system [9]. Initially, Expertiza allowed students to select their own teammates and employed an FCFS method for teams to reserve a topic. In practice, this led to individual students "grabbing" all topics quickly and then searching for teammates. Students who were not able to choose early encountered no available topics, and long waiting lists for everything. In order to induce students to form teams of the requisite size, the instructor often had to threaten students that any team that was not up to minimum strength by a particular date and time would be manually dropped from the topic it held.

2.2 Instructor-Selected Teams

Instructor-selected teams is a logical approach to team formation. Some instructors have specific criteria they would like to optimize for in forming teams, such as matching students with new team members or students with complementary skills; as a result, other methods may be unable to satisfy their conditions. However, instructor-selected teaming comes with some disadvantages: students will likely have difficulty providing input regarding their interests and often the search space is simply too extensive to find an optimal solution. And in larger classes, this approach would become infeasible, specifically MOOC courses which could potentially involve hundreds of teams and proposals [5].

3 INTELLIGENT TEAM ASSIGNMENT

We aim to use an automated solution which can scale efficiently with course size. Our approach asks each student to rank topics in order of preference and records the topic bids. It is not necessary for each student to bid on each topic; if a student could not be assigned one of the topics (s)he had requested, they can then select one of the unclaimed topics.

Using the recorded topic preferences, priorities are weighted by determining cost and benefit of building clusters (teams) as though they had been assigned solely with data from that priority. In other words, if all students had rated the same topic as their priority-2, the weight of second priority bids would be significantly decreased as it does not provide practical data for partitioning the dataset into teams. The benefit and cost of using any particular priority to coalesce groups can be calculated as student satisfaction and use of available space respectively.

3.1 Weighting formula

A student satisfaction metric should reflect that satisfaction may increase in a nonlinear fashion as students are assigned topics of higher priority, i.e. the utility of assigning primary choices over secondary choices is likely to be greater than the utility of assigning secondary choices over tertiary choices. This pattern closely resembles Zipf's law [8], which is utilized in the benefit formula, where the benefit of each priority (P_i) is calculated in Equation 1.

Equation 1: Benefit formula. $B(P_i) = \frac{1}{i}$.

The use of "available space", a measurement of the distance between the number of bids and the max team size (*mts*), can be calculated with Equation 2, as the *mts* and the number of bids for all N topics (t) at priority P_i.

Equation 2: Cost Formula. $C(P_i) = \frac{1}{N}\sum_{t=1}^{N}\left|mts - P_{i_t}\right|$.

So, if mts = 4 and 3 students have made topic t their priority-2 bid, then the available space on the topic-t team is 1. The available space overall is the sum of the available space for all the topics.

Equation 3: Weighting formula. $W(P_i) = \frac{B(P_i)}{C(P_i)}$.

The overall weight can then be determined with Equation 3, where weight is directly proportional to benefit and inversely proportional to cost.

3.2 Team and Topic Assignment

Student bids are mapped to an n-dimensional vector and Hierarchical k-means [1] clustering recursively splits the students into teams. The algorithm is run with two centroids which bisect the data set and k-means is then called for each cluster until the resultant cluster size is less than or equal to *mts*. k-means clustering is run iteratively to allow more fine-tuned control over the team sizes. Clustering users based on topic preference lets us assign a single topic that satisfies the group.

3.3 Matching Teams to Topics

The clustering algorithm is hosted on a separate web service. After the service is run, it returns the team assignments to the client system to be assigned to topics using a stable matching algorithm. First, team bids are determined by having team members "vote" on the priority of the bid for each topic. So, if three team members were to place topic 1 as their primary bid and one member was to place topic 1 as their secondary bid, then topic 1 would be recorded as the team's primary bid. Teams are then sorted by size, allowing the larger teams a higher priority to increase the overall satisfaction. Insofar as possible, each team is assigned the first available topic of the highest priority. Teams that lose the bid for all topics they bid on are disbanded and members either need to (manually) join a team that already holds a topic, or form a new team and select a topic that has not yet been assigned to a team.

4 RESULTS

Intelligent Assignment was utilized to assign teams and topics to students for course projects in a graduate Software Engineering course. Upon completing the assignment, 26 out of 64 students completed a voluntary survey to evaluate the use of Intelligent Assignment. Results showed that 73% of responders received one of their top three priority choices.

REFERENCES

[1] Kohei Arai and Ali Ridho Barakbah. 2007. Hierarchical K-means: an algorithm for centroids initialization for K-means. *Reports of the Faculty of Science and Engineering* 36, 1 (2007), 25–31.

[2] Ryan Cavanaugh, Matt Ellis, Richard Layton, and Mark Ardis. 2004. Automating the process of assigning students to cooperative-learning teams. *age* 9 (2004), 1.

[3] Torgeir Dingsøyr, Sridhar Nerur, VenuGopal Balijepally, and Nils Brede Moe. 2012. A decade of agile methodologies: Towards explaining agile software development. (2012).

[4] Edward Gehringer, Luke Ehresman, Susan G Conger, and Prasad Wagle. 2007. Reusable learning objects through peer review: The Expertiza approach. *Innovate: Journal of Online Education* 3, 5 (2007), 4.

[5] Katy Jordan. 2014. Initial trends in enrolment and completion of massive open online courses. *The International Review of Research in Open and Distributed Learning* 15, 1 (2014).

[6] Larry K Michaelsen, Michael Sweet, and Dean X Parmelee. 2011. *Team-Based Learning: Small Group Learning's Next Big Step: New Directions for Teaching and Learning, Number 116*. Vol. 103. John Wiley & Sons.

[7] ME Natishan, LC Schmidt, and P Mead. 2000. Student focus group results on student team performance issues. *Journal of Engineering Education* 89, 3 (2000), 269–272.

[8] Mark EJ Newman. 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary physics* 46, 5 (2005), 323–351.

[9] Yang Song, Zhewei Hu, and Edward F Gehringer. 2015. Pluggable reputation systems for peer review: A web-service approach. In *Frontiers in Education Conference (FIE), 2015 IEEE*. IEEE, 1–5.

[10] Yang Song, Zhewei Hu, and Edward F Gehringer. 2017. Collusion in educational peer assessment: How much do we need to worry about it?. In *2017 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–8.

[11] Yang Song, Zhewei Hu, Edward F Gehringer, Julia Morris, Jennifer Kidd, and Stacie Ringleb. 2016. Toward Better Training in Peer Assessment: Does Calibration Help?. In *EDM (Workshops)*.