Notes on CS70 Discussion Solutions CrowdSourcing Model

Leah Dickstein

April 22, 2014

Current working model:

$$\begin{aligned} grade &= \frac{-time}{efficiency} + knowledge * time + reward * Pr[reward] \\ &= -at + bt + AP[A], 0 < a < 1 b > 1 \end{aligned}$$

Contents

1 Experiment 1

1.1 Action Plan

1. Diagnostic test tells us strength of students
2. Either: 1) Ask how much time it took each student (weaker students should take longer) OR 2) Set a fixed amount of time (weaker students with higher cost will produce lower quality bids) [The way we analyze/compare the following values will vary based on which method was chosen]
3. Calculate relationship between strength and cost
4. Calculate expected bid value
5. Collect actual bid values
6. Compare to theoretical bid value + other variables
7. Calculate expected value
8. Collect actual expected value
9. Compare values

1.2 Thought Process

We need to identify:

- # of bids is as expected
- $\mathbb{E}[U]$ as expected
- Bid values as expected

Variables we should know:

- A – reward
- c_i – cost
- x_i – bid

Need to accomplish:

1) Rank players by strength to determine their efficiency/knowledge coefficients
2) Assign costs: amount of time it takes to complete assignment
3) Evaluate bids: Quality of assignment is determined by 1] Correctness 2] Thoroughness (rough estimate = length)

NOTE: Need resolution: a regular way to assess student's grades (interpreting that as value/knowledge) for better precision on what activity/event/action corresponded to how large a change in grades

1) Strength

How to rank players by strength:

1. Ideal: Use Experiment 2's results and control the combination of activities of all the participants
   (a) Experiment 2 is supposed to tell you how much each activity matters
   (b) A student participating in 2 activities should be stronger than a student participating in just 1 (assuming they are equal in everything else, including starting knowledge) so we can rank students that way
2. Baseline diagnostic to assess familiarity with concepts (can be given on first day of class, or as part of Homework 1 along with sign in information)
3. Voluntary poll (can be given on first day of class, Piazza poll looking for voluntary participation) "Which of these concepts/buzzwords have you seen before?"

After that, place data points (students) in bins

Assume people stay in bins relative to each other (Not actually true!)
Else: We need regular calibration checks

2) Costs

Need to determine relationship between strength (efficiency/knowledge) and costs:

1. Ask students how much time they took
2. Only give students a fixed amount of time
   (a) Like the practice midterm!
   (b) For example "There is an extra credit opportunity with no penalty; the only cost is your time."
   (c) See how many students show up and how they fall in the strength-bins
3. Propose a relationship and see if it's true based on bid values/expected values (e.g. inverse As strength increases costs decrease with some constant factor)

3) Bids/Gain

- Collect bid values as data points, plot to see relationship with cost/strength/grades
- If relationship exists, calculate theoretical values of costs, etc. per bin and compare to model
- Calculate $\mathbb{E}[U_{students}]$, see distribution
- Check if it matches model
- How do students' grades change pre $\rightarrow$ post discussion event

2 Experiment 2

Key Point: Identify how to weight activities, or quantify $\mathbb{E}[value]$ of each activity
Later when we're using grades as our measure of value/knowledge, we know what "contribution" is from Discussion Solutions

2.1 Factors that affect knowledge/grades/expected value

In no particular order:

1. Midterm Performance (Required)
2. HW performance (Required)
3. Attendance of HW Parties
4. Attending Lecture
5. Attending Discussion
6. Doing Discussion problems (submitting solutions)
7. Having a Study Group
8. Going to Office Hours
9. Reading Lecture Notes
10. Posting Qs/Receiving Feedback from Piazza

    Not as important but worth mentioning:

11. Starting Knowledge
12. Stress from other classes → Assume this will be reflected by participation in material

    12.1 Time spent mulling

13. Enjoyment of material/intuition/being able to pick things up quickly → Assume not a problem / uniform

2.2 Ideal Experiment

Every single person is uniform: they learn the same way, same speed, same starting knowledge, are taking the same classes and therefore dedicate the same amount of time to CS70.

I change which of the above 8 activities (3. – 10.) the students participate in, then see based on grades which activities are helpful and by how much.

There are $\sum_{i=1}^{8} \binom{8}{i}$ possible subsets of the 8 activities, and I want at least 3 data points per subset (minimum needed to fit a line.) This means I would want a population of at least $3 \sum_{i=1}^{8} \binom{8}{i} = 3 * 254 = 762$. Although we're not quite there yet, with a CS70 student body of over 600 we might be close enough (over 79%).

2.3 Realistic Experiment

Require students to try everything out in the first 1 2 weeks.

- See what is more effective based on subsequent attendance (assuming students recognize what is best for them)
- Have students rate their experiences (assuming they recognize what is better for them)

Recap from 2014-03-30

3 Questions

What is it about the situation that caused it to match a crowdsourcing model? (What criterion defined it as a crowdsourcing problem?)
How can we improve the contest so it better fits the goal of the manager? Let's tweak parameters in controlled experiments and see how and by how much parameters affect the outcomes of the contest. Then, let's design the ideal contest for admin and discussion solutions. Let's also observe how students react to different contests to determine how they are learning and what learning styles best suit them.

4 Description

Players: Students
Incentive: Extra Credit (Is prize value relatively unknown to students? They don't know HOW much extra credit they get until they bid and win for the first time.)
Manager: Professor, GSIs, Readers = admin
Incentive/Goal for Manager: Get the right discussion solutions and get maximum number of students to learn by participating *with effort.* A third possible goal is to differentiate who the strongest players are.
Cost of prize to manager: Virtually nothing if we consider grade inflation to be minimal, thus Utility = value of function f($\overrightarrow{x}$).

5 Comparison

5.1 Differences with Gireeja's model

1. Players do NOT know each other's strengths
   - However, based on their grades, they know relatively where they are in the class
   - Since it's Berkeley/maybe people in general, many players assume they are "bad" compared to everyone else (assumption??)
2. In model, there was a Nash equilibrium where only 2 players bid. Players 3 through n of less strength bid 0 with probability 1. In the current setting 3 people get extra credit, so there should be 3 to 4 players who bid.

5.2 Connection to Gireeja's paper

The task is a combination of a selective task and market creation. It's selective because there's a "right" answer to solutions. Even if students post different ways of approaching the problem, there's still a "correct end goal." It's market creation because the managers want as many people to participate as possible with effort, so the more students that participate the better.

For selective task, $U_task = max(x_1, x_2, \dots x_n) - A$

$$\mathbb{E}[U_task] = \frac{A}{6}\left(\frac{3c_2 + c_1}{c_2^2}\right) - A$$

This expected utility is positive iff $\frac{3c_2 - c_1}{6c_2^2} - 1 > 0$
If $c_2 \gg c_1$ then $c_2 < \frac{1}{2}$
If $3c_2 - c_1 = \epsilon$ is small (low difference) then $c_2^2 < \frac{\epsilon}{6}$ ensures positive utility
This is based on $c_1 < c_2$

For market creation, $f(\overrightarrow{x}) = \alpha n + \beta$
Manager utility doesn't depend on utility of players at all

5.3 What this means

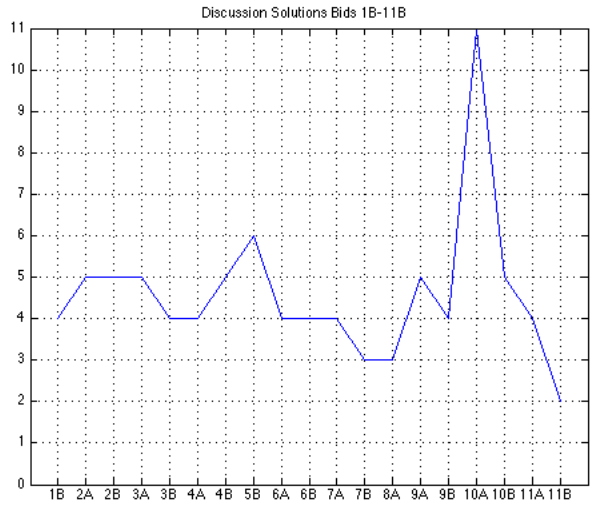
If the managers gave everyone extra credit simply for participating, everyone would only put in the minimal bid. This is why the manager wants participation *with effort.* The manager wants to challenge everyone.

Since there are elements of both selective task and market creation in the Discussion Solutions situation, we need to decide which one is weighted more and/or how to combine the two to create a model for this scenario.

6 Data: Updated to 11B

6.1 Discussion Solutions

Discussion	# of bids	Players
1B	4 bids	Hansong Zhang, Kevin Chen, Daniel Suryakusuma, Andrew Luo
2A	5 bids	Viraj Mahesh, Lingtian Cheng, Hansong Zhang, Hongling Lu, Arnav Dugar
2B	5 bids	Lingtian Cheng, Hongling Lu, Melanie Cebula, Myra Haqqi, Arnav Dugar
3A	5 bids	Hongling Lu, Viraj Mahesh, Brian Chu, Myra Haqqi, Arnav Dugar
3B	4 bids	Chris Dock, Hansong Zhang, Hongling Lu, Myra Haqqi
4A	4 bids	Kevin Mawhorter, Hongling Lu, Melanie Cebula, Cong Chen
4B	5 bids	Arnav Dugar, Anurag Ajay, Myra Haqqi, Hongling Lu, Lingtian Cheng
5A	0 bids	Midterm so no solution thread posted
5B	6 bids	Hongling Lu, Arnav Dugar, Aditya Challa, Anurag Ajay, Myra Haqqi, Alex Yang
6A	4 bids	Arnav Dugar, Hongling Lu, Myra Haqqi, Aditya Challa
6B	4 bids	Aditya Challa, Myra Haqqi, Anurag Ajay, Arnav Dugar
7A	4 bids	Aditya Challa, Jong Yun Lee, Myra Haqqi, Max Kanwal
7B	3 bids	Aditya Challa, Max Kanwal, Myra Haqqi
8A	3 bids	Myra Haqqi, Aditya Challa, Liuxiao Zhang
8B	0 bids	Midterm so no solution thread posted
9A	5 bids	Albert Lin, Derek Ahmed, Aditya Challa, Myra Haqqi, Max Kanwal
9B	4 bids	Derek Ahmed, Albert Lin, Aditya Challa, Myra Haqqi
10A	11 bids	Paul Bramsen, Adithyavairavan Murali, Max Kanwal, Josh Zarrabi, Tony Chen, Aditya Challa, Claire Wa
10B	5 bids	Janet Chu, Aditya Challa, Yang Rui, Myra Haqqi, Max Kanwal
11A	4 bids	Janet Chu, Aditya Challa, Myra Haqqi, Max Kanwal
11B	2 bids	Aditya Challa, Myra Haqqi

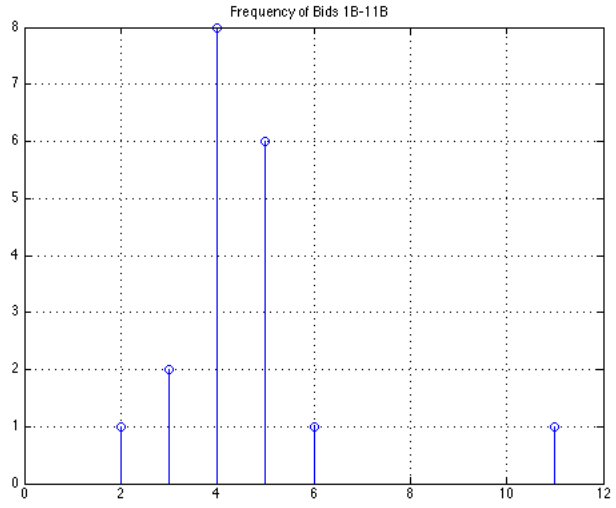



```
data = [4, 5, 5, 5, 4, 4, 5, 6, 4, 4, 4, 3, 3, 5, 4, 11, 5, 4, 2];
```

```
plot(data)
grid on
set(gca,'xtick',1:19);
ylim([0 11])
set(gca,'XTickLabel',{'1B','2A','2B','3A','3B','4A','4B','5B','6A','6B','7A','7B','8A','9A','9B','10A','10B','
title('Discussion Solutions Bids 1B-11B')
```

Average bids per contest: $4\frac{4}{7}$

# of bids	Frequency
2	1
3	2
4	8
5	6
6	1
11	1



The below is updated to 11A.

Player	# of bids	Bid dates
Aditya Challa	10	5B, 6A, 6B, 7A, 7B, 8A, 9A, 9B, 10A, 10B
Adithyavairavan Murali	1	10A
Albert Lin	2	9A, 9B
Alex Yang	1	5B
Andrew Luo	1	1B
Anurag Ajay	3	4B, 5B, 6B
Arnav Dugar	7	2A, 2B, 3A, 4B, 5B, 6A, 6B
Brian Chu	1	3A
Chris Dock	1	3B
Claire Watanabe	1	10A
Cong Chen	1	4A
Daniel Suryakusuma	1	1B
Derek Ahmed	3	9A, 9B, 10A
Hansong Zhang	3	1B, 2A, 3B
Hongling Lu	8	2A, 2B, 3A, 3B, 4A, 4B, 5B, 6A
Janet Chu	1	10B
Jong Yun Lee	1	7A
Josh Zarrabi	1	10A
Kevin Chen	1	1B
Kevin Mawhorter	1	4A
Lingtian Cheng	3	2A, 2B, 4B
Liuxiao Zhang	1	8A
Max Kanwal	5	7A, 7B, 9A, 10A, 10B
Melanie Cebula	2	2B, 4A
Myra Haqqi	14	2B, 3A, 3B, 4B, 5B, 6A, 6B, 7A, 7B, 8A, 9A, 9B, 10A, 10B
Paul Bramsen	1	10A
Tony Chen	1	10A
Tricia Fu	1	10A
Viraj Mahesh	2	2A, 3A
Yang Rui	1	10A

Average bids per player: 2.7
After removing the 4 outliers: 1.577
Total bid opportunities (contests): 18
Total players: 30
Returning players > 1 instance: 12 = a little under half!

6.2 HW Write Your Own Problem

In this case, students have already been forced to bid. Thus, posting online comes only at the cost of going online. There is potential cost in making a good problem while writing HW.

HW #	# of Bids
2	23
3	11
4	4
5	9
6	7
7	1
8	7
9	7
10	4
11	6

Average number of bids per HW: 7.9

These results indicate the majority of the class aren't submitting their own problems. In addition, the difficulty of the problems being submitted appears to have decreased, although I haven't examined them carefully. That could just be hindsight bias.

7 Moving Forward

1. Design ideal experiment for Leah
2. Make ideal experiment realistic
3. Compare already submitted solutions to see how they're chosen for reward and how that relates to bid value/the other variables