COS 445 - Strategy Design 1

Sasha Khoretonenko, Ernest McCarter, and Aidan Ward

February 19, 2024

Part A

Description

What should a good strategy do when $T = 0$? Why?

Response

We found that the synergist approach was the best approach for when $T = 0$, which is equivalent to the holist approach when $T = 0$ due to how the holist lists universities according to the highest values of the sum of their quality and synergy, but the quality of every university is zero when $T = 0$. This is because, in this case, university quality does not matter. In other words, $Q_i = Q_j \forall i, j : 0 \leq i, j < N$.

This case is incentive compatible for students. This means that no student can gain any benefit or advantage from lying. This is because the student's preference is only based on synergy, and the universities' preferences are based on the sum of synergy and aptitude of the student.

Proof: The algorithm happens in two stages: (1) students report their top 10 preferences and (2) universities propose to the students in order of their preferences based on which students applied to them and students accept or reject universities in the order the universities propose.

If a student misreported their preferences and listed a school with lower synergy as their favorite, they would only ever be approached by universities they prefer less and also have a worse chance of getting into, since the universities also prefer students with a higher synergy and the aptitude of the student cannot be changed by changing the preferences. Thus, changing the preferences of the student does not change the order of the universities' preferences, and therefore the student will not benefit nor be proposed to by universities it prefers more by misreporting its preferences.

Furthermore, once universities begin proposing the student, the student will never benefit from misreporting their preferences because they will choose to reject universities that they have higher synergies with in favor of the universities that they misreported as their favorites. Therefore, the student cannot ever match with a university that it prefers more by misreporting their preferences than by reporting their true preferences.

Part B

Description

What should a good strategy do when $W = 0$? Why? If you like, you may assume for this part that all students (including you) know A_t for all students t (rather than just knowing S).

Response

When $W = 0$, this means that there is no difference in any synergies, since they are all zero. This is to say, all universities have the same preferences of students and all students have the same preferences of universities. This is very similar to problem 1 from PSET 1, but with the added condition that universities also have identical preferences. We can denote each students s_i where i represents the number of students that the universities prefer to s_i. Similarly, we can denote each university u_j where j represents the number of universities that the students prefer to u_j. It is clear that s_0 will get matched to u_0 if they apply, s_1 will get matched to u_1 if they apply, and in general that s_i will get matched to u_i if they apply, assuming that every $s_j : j < i$ gets matched to u_j.

Since, for a student s_i, the best possible university that they can get matched to assuming that every other student acts in their best interest is u_i and all students know all student and university preferences, it is in student s_i's best interest to apply for university u_i. Assuming that every student uses the optimal strategy, the rest of the 9 schools that you apply to shouldn't matter, but since we don't know every other student's aptitude in our code, we use the formula $Math.round(N(1-\frac{A_s}{S}))$ to estimate the number of students with a better aptitude than us (which is equivalent to our aptitude ranking). Let's call this integer r. We then select our 10 universities from $[r-5, r+4]$ (assuming the universities are listed in decreasing order).

Part C

Description

Provide a brief justification for your strategy. Focus on convincing the grader that it is a good strategy, by explaining the main ideas and why you chose this strategy. You should aim to keep this under one page. This will not be strictly enforced, but the grader may choose not to read beyond one page. You should not think of this merely as a documentation explaining only what your code does. Instead, try to imagine that it's purpose is to convince your guidance counselor why they should adopt your strategy.

Response

Our strategy handled three cases: the cases from part a, part b, as well as the general case. Because the part a and part b cases are explained above, we will only explain the general case here.

We arranged all universities in order of their quality and set bounds from upon which to select the most preferable universities.

The general strategy was to pick a target number to estimate approximately where we stood against universities when they were ranked by their quality. This allows us to pick an area approximately where we think we are at without overshooting. This method follows the logic found in part b, with the main difference being that a tuned constant o is included. We used this equation to calculate our target:
$M = max(0, \lfloor N(1 - \frac{A_s}{S}) - o \rfloor)$
We round down because we are selecting an integer index, and we take the max between that index and 0 because we cannot select a negative index. The o value is the tuned constant we call the offset that we found best to set as $o = 10$ from testing. This bumped our average performance from around 0.506 to 0.576. Our guess for why this works is because a lot of students are not being as aggressive as they could be.

Beyond this we also define another factor, which we refer to as our range, or r. Range represents how much we want to vary our preferences based on the proportions of W and T. Building off from part a and part b, we knew that when $W >> T$ this means that university quality is much less important compared to synergy, and therefore we should choose from the wider ranges of universities as synergy becomes more important to correspond with what was found in part a. When $T >> W$, like in part b, we realized that we want to pick based on where approximately we think we should be ranked by our own aptitude, since synergy is much less important than university quality. Putting this all together, we get this formula for r:
$r = max(5, \frac{kW}{T})$ k here is another constant we tuned through testing at $k = 5$. We also decided that we wanted to set a lower bound of 5 because we wanted to select from a range of at least 10 schools to choose from.

From this range, we selected the best schools according to $U_i + S_{s,u}$ and ordered them accordingly. Through more testing, we found that this bound was not enough to perform well as we wanted. We decided to also include every school that had a lower quality to capture schools that might have a higher preference (due to a particularly higher synergy) but were not captured by the bounds since we know that these schools would also prefer us based on this higher synergy. We thus chose the 10 highest schools according to preference on the bounds $[M - r, N - 1]$.

The final part of our design includes choosing our most preferential university and ranking it as our number one choice and pushing all other choices back by one index. In testing, we saw that this created a very marginal improvement, and it created a slight improvement on the leader boards as well. Our guess as to why this is is that it pays off to be more aggressive in these scenarios as other teams might not be using these techniques.