CS280r Final Project Report
Project Name

Anna Sophie Hilgard and Nicholas Hoernle

Abstract

1. Introduction

Communication is a costly resource in human-human and human-computer interaction [references]. Given a medical setting, Amir et al. (2015) report that study participants could not review necessary information in a timely manner, nullifying the effect of obtaining complete and correct information. Similar problems arise in the crowdsourcing setting. Hahn et al. (2016) show that crowdsourcing vendors consistently struggle to balance the amount of information needed convey to a worker to equip the worker to excel at his/her job without overburdening the process with too much information. An added complication is that in many settings, the ideal contextual information that is shared is subjective. Therefore, multiple parties have competing interests in having their contributions addressed. One possible solution is to allow for a single contributor or an outside controller to make these subjective decisions. However, experiences with content generators like Wikipedia with a strong hierarchical or dictatorial leadership (Benkler et al., 2015) have shown that the resulting content is often suboptimal from the viewpoint of the whole and heavily skewed to conform to the opinion(s) of the decision-maker(s). Schwartz (2015) stresses that those situations in which group members have different information and the actions of individuals are interdependent are the most critical to be collectively assessed.

Under these conditions, we see a strong case for adopting social budgeting techniques to crowdsource contextual points. As shown in Boutilier et al. (2015), if we assume adopt a utilitarian framework in which we hope to maximize the satisfaction of all group members, properly chosen voting rules can ensure that we minimize the maximum difference between the optimal possible satisfaction to all members and that selected by the voting rule in expectation (the regret), whereas it is clear that for a dictatorial selection this could be trivially equal to the worst case if the size of the alternative set is larger than two times the size of the set of options to be selected.

In particular, we will seek to test the effectiveness of the subset selection algorithm generated by Caragiannis et al. (2017), which approaches the problem as a variation on the maximin rule. In particular, the authors show that it is possible to derive an explicit utility function which maximizes regret while maintaining consistency with the votes, leading to the following expression for maximum regret for a subset selection T:

$$\max_{S \in A_k} \sum_{i=1}^{n} \frac{\mathbb{1}[S \succ_{\sigma_i} T]}{\sigma_i(S)} \tag{1}$$

Where $S \succ_{\sigma_i} T$ indicates that there is no alternative in T preferred to every alternative in S given the utility function σ_i, and $\sigma_i(S)$ is the ordinal ranking of the best alternative in set S in the ranking determined by the utility function σ_i.

Intuitively, any term in this maximization captures the lost satisfaction to the voters of not having the given set S_i chosen rather than T, weighted by how much he or she liked his or her best option in S_i. This will lead to a greater penalization for sets T that do not give many participants at least one of their top choices.

We seek the set T that minimizes this quantity.

$$\text{argmin}_{T \in A_k} \max_{S \in A_k} \sum_{i=1}^{n} \frac{\mathbb{1}[S \succ_{\sigma_i} T]}{\sigma_i(S)} \tag{2}$$

Shah et al. show that this can be solved through an ILP with nm variables and $nm^2 + \binom{n}{m}$ constraints, where n is the number of voters and m is the number of alternatives available.

For comparison, we also use plurality/knapsack voting, which has been used in real-world participatory budgeting programs (likely in part because of its computational simplicity and ease of understanding) (Goel et al., 2015) and is shown in Shah et al. to have empirical regret approaching that of the subset selection algorithm above for subset sizes greater than three, which will be the case in our experiment and should be generally true for problems of this nature.

We apply these different voting rules to the problem of subset selection and conclude that in practice the success of a voting rule may depend heavily on the difficulty (cognitively or subjectively) one has in comparing options.

2. Experiment Design

Three voting rules are compared to evaluate the success of the subset selection. To test the voting rules we pose a setting where participants are requested to select a number of points that may be relevant to a given topic. The authors compiled a set of 10 supporting points from popular *New York Times* opinion pieces, and used a web-based survey form to allow participants to make subset selections. The topics were presented in a 'debate prompt' style and the participants were asked to select the points that would contribute the most value to the presented argument. We also allowed participants to provide feedback on the subset selection styles they most and least enjoyed.

The interface [1] was designed to present participants with thee question (on three different articles) and each question would display a different subset selection protocol. The different sections consisted of

- 'ranking' selection by dragging and dropping alternatives into the correct order from most useful to least useful for the given argument.
- 'plurality' selection, where check-boxes are selected until 5 selections were made.
- 'cardinal' selection where each point was given a score out of 10 independently of the others.

The study consisted of 37 respondents and 111 subset choices over the three different voting rules. We solve 2 using integer linear programming as described in Caragiannis et al. (2017) to aggregate the ranking and ordinal results into an optimal subset. We induced a ranking over the cardinal results to obtain ranked values to use in the subset selection algorithm, breaking ties at random. For the plurality results, we selected the subset greedily using a majority rule approach. Selected subsets included four points each.

Finally, selected subsets were presented to a different group of 15 participants. These participants were simply asked to select the most relevant subset, also given the same topic. This data was also collected through a web-based survey [2].

3. Results

4. Extension: Cognitive Load

The results from Caragiannis et al. (2017) suggest that for a subset of size four, minimum regret should generally provide a lower upper bound on the regret of participants. That is, it is better in the case where we assume the utility function with highest possible distortion compatible with the selections. However, plurality voting has a very similar upper bound for a subset size of four and may have superior properties in other considerations, as we've seen above. We believe the success of plurality voting in our experiment has to do with two factors.

First, the worst case bounds do not imply anything about the average case, and the two worst case bounds are so similar that it is probable that the average case regret for plurality is in fact better. Second, our qualitative feedback suggests that knapsack/plurality selection is significantly easier for respondents. This leads us to consider the possible effects of cognitive load on the various voting algorithms. In fact, Caragiannis et al. (2017) explicitly mentions that the analysis has yet to investigate effects of cognitive load although the authors stress in other works, such as Benade et al.

[1] Accessed at: http://nick-and-sophie-harvard-cs280r.s3-website-us-east-1.amazonaws.com/index.html
[2] Accessed at: http://nick-and-sophie-harvard-cs280r.s3-website-us-east-1.amazonaws.com

(2017) that the entire point of voting mechanisms is to reduce the unacceptable cognitive load of eliciting a full utility function.

One good reason to do this is that people are likely to make mistakes when presented with a large cognitive load. In particular, in the Sushi dataset from Kamishima (2003), 70% of rankings and ratings of the same subsets contain contradictions. That is, the cardinal values in the rating set do not map to the ordinal values in the ranking set. Then it can be assumed that respondents have reported erroneous preferences in one of the two cases, possibly due to excessive cognitive load of the reporting mechanism. If it can be expected that voters will occasionally make errors in reporting their preferences, we should also be interested in the robustness of these selection algorithms to errors. Previous work has shown that the worst case robustness of both minimax and plurality voting is generally better than many other voting methods when considering the worst case for a single winner and that the worst case for a larger subset is bounded by the worst case for a single winner Procaccia et al. (2007). Here, we consider the empirical average case for a variety of subset sizes.

To set up the experiment, we take the Sushi dataset mentioned above and calculate for the rating and ranking problems on the same subsets (these are 'sushi3b.5000.10.order' and 'sushi3b.5000.10.score') the minimum number of flips of adjacent rankings required to bring the ranking into agreement with a ranking induced over the ratings (flips corresponding to equally rated items are not included in the count, as either ranking of such items is consistent with the rating assuming randomized tie breaking). We tally the distribution of the number of errors throughout all 5000 respondents to the sushi survey. Then, we use the third sushi dataset, 'sushi3a.5000.10.order', which contains only 10 types of sushi rather than the 100 in other subsets to bootstrap voting profiles. Because each of the users in the 100 sushi dataset were each only presented with 10 sushi out of the 100, we feel it is fair to assume the same cognitive load would be true for only 10 total sushi types. However, using this dataset allows us to simulate voting over a set of only 10 objects. To create the profiles, we repeatedly draw a set number of voting profiles at random from the rows of the file. We create 100 of these voting profile sets for each trial. Then, we loop through each item in each of the profile sets and induce a number of errors (flips of adjacent rankings) corresponding to a draw from the error distribution. We then perform plurality and minimum regret subset selection on each of the 100 correct voting profile sets and their corresponding profile sets with induced errors and report the number of times the answers matched, in spite of the errors. The results are reported below for subsets of size 1 to 9 and profile sets of 10 voters each.

We find that in general, plurality voting is much more robust to errors in voting than minimum regret. Then, based on our findings, we would expect that the additional cognitive load of ranking induces more errors in voting and that the algorithm is less robust to these errors.

4.1. Citations

Here are two examples of how to cite a paper properly:

- **?** shows that ...
- Prior work has shown that ... (**?**).

5. Related Work

Discussion of previous important, similar work in the area with comparison to the particular approach taken and results of the paper. Avoid simply providing a laundry list of other work that is somehow related to the subject of the paper. This section should contain brief, in depth discussions of the work most similar to your project, i.e., to research that takes an approach to the problem or produces results with which your project should be compared. As is always the case with written work, throughout the paper you should have citations to work that you draw on. For example, if you have adapted a system, include a citation to the system when you first mention it; if you are extending a formalization, include a citation to the original on first mention. If you are unclear about whether a simple citation suffices or an extended discussion is needed in the Related Work section, look at the papers read for class this semester for models. If you are still unsure, check with the teaching staff.

6. Conclusion

Describes the insights that can be taken away from the work reported in the paper.

7. Future work

Suggests extensions or challenges raised by the project.

O. Amir, B. J. Grosz, K. Z. Gajos, S. M. Swenson, L. M. Sanders, From care plans to care coordination: Opportunities for computer support of teamwork in complex healthcare, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM, 1419–1428, 2015.

N. Hahn, J. Chang, J. E. Kim, A. Kittur, The Knowledge Accelerator: Big picture thinking in small pieces, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, ACM, 2258–2270, 2016.

Y. Benkler, A. Shaw, B. M. Hill, Peer production: A Form of collective Intelligence, Handbook of Collective Intelligence 175.

R. Schwartz, How to Design an Agenda for an Effective Meeting, Harvard Business Review .

C. Boutilier, I. Caragiannis, S. Haber, T. Lu, A. D. Procaccia, O. Sheffet, Optimal social choice functions: A utilitarian view, Artificial Intelligence 227 (2015) 190–213.

I. Caragiannis, S. Nath, A. D. Procaccia, N. Shah, Subset selection via implicit utilitarian voting, Journal of Artificial Intelligence Research 58 (2017) 123–152.

A. Goel, A. K. Krishnaswamy, S. Sakshuwong, T. Aitamurto, Knapsack voting, Collective Intelligence .

G. Benade, S. Nath, A. D. Procaccia, N. Shah, Preference Elicitation For Participatory Budgeting, in: Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI). Forthcoming, 2017.

T. Kamishima, Nantonac collaborative filtering: recommendation based on order responses, in: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 583–588, 2003.

A. D. Procaccia, J. S. Rosenschein, G. A. Kaminka, On the robustness of preference aggregation in noisy environments, in: Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems, ACM, 66, 2007.