Ordinal Approximation in Matching and Social Choice

ELLIOT ANSHELEVICH
Rensselaer Polytechnic Institute

In this note, we discuss several settings in which algorithms can provide good outcomes when given only ordinal information. We focus especially on voting mechanisms in settings with spacial preferences, and on the notion of distortion.

Categories and Subject Descriptors: J.4 [**Computer Applications**]: Social and Behavioral Sciences; F.2 [**Theory of Computation**]: Analysis Of Algorithms

General Terms: Algorithms, Theory

Additional Key Words and Phrases: Matching, Social Choice, Metric Utility

1. INTRODUCTION

Traditional approximation algorithms attempt to find good solutions under the constraint that they are also computable efficiently, as compared with an optimal solution which could be obtained using unbounded resources. There are many other notions of approximation, such as finding good solutions without knowing the future in online algorithms, being able to access the input a limited number of times in streaming algorithms, etc. In this letter we will focus on algorithms which are only given *ordinal* information, and yet must compete with algorithms which know the "ground truth" numerical information.

To illustrate when such constraints can arise, consider the following simple voting scenario. As in classic voting and social choice literature, a set N of agents (voters) have preferences over a set of alternatives (candidates) A. As is done in the utilitarian view of social choice (see [Boutilier et al. 2012; Harsanyi 1976]), and in much of spatial preference literature (e.g., [Merrill and Grofman 1999; Enelow and Hinich 1984; Conitzer 2009]), we assume that these preferences actually result from underlying utilities or cost functions, with $c_i(a)$ being the cost that agent i assigns to alternative $a \in A$ occurring. For example, this can occur when both agents and alternatives are points in some metric space, such as a (overly simplistic) one- or two-dimensional space of "liberal-conservative" and "libertarian-authoritarian" spectrum of opinions. In such a case the happiness of an agent with a particular alternative can simply correspond to the distance between them in this space, which represents how close their views are to each other. More generally, both agents and alternatives can be points in a very high-dimensional space (for example a space in which each dimension is a different political issue), with $c_i(a)$ still being the distance between i and a in this space, or maybe some simple increasing function of this distance. Given such explicit numerical costs, we can define the truly optimal

This material is based partially on work supported by NSF awards CCF-1101495, CNS-1218374, and CCF-1527497. Author's address: eanshel@cs.rpi.edu

alternative as the one which minimizes $\sum_i c_i(a)$ (or some other reasonable objective function); this is in contrast to the case when there are no numerical costs and no clear optimal alternative may exist.

Now, if our goal is to find this optimal candidate, and we have access to all the numerical $c_i(a)$ values, then this problem is trivial. In many social choice settings, however, while it may be reasonable to assume some underlying utility structure on the preferences, it is often *un*reasonable to assume that we know these numerical values exactly; it is much more likely that we only know the preference ordering over the candidates for each voter. In other words, we only know the *ordinal* preferences for each agent i over alternatives in A; we do not know the actual values $c_i(a)$ and thus the *strength* of agent i's preferences.[1] This can be true because the ordering is much easier to specify than numerical values; in fact agents i themselves may not know their exact utility or cost values, but can still make pairwise comparisons between alternatives. Given only this limited ordinal information, we still want to choose a good candidate, i.e., one that has small $\sum_i c_i(a)$ compared to the optimal one. This requires an approximation algorithm: one that is limited not by its computational power (or at least not *only* by its computational power), but is instead limited by only knowing ordinal information, while trying to compete with the optimum solution according to the true numerical information. Perhaps surprisingly, one can often create such algorithms with approximation guarantees of a small constant, without knowing anything about the underlying values $c_i(a)$ other than the ordinal preferences they result in, and the fact that these values form an arbitrary metric.

In this letter we describe some approaches to forming and analyzing such *ordinal approximation algorithms*, for both the social choice setting and other settings such as matching and clustering. We will focus specifically on the case when the underlying numerical costs (or utilities) form a metric space. It is worth noting that this is not the only type of correlation between the numerical costs which can lead to interesting results. Specifically, several works consider the case when these costs are non-metric, but are *normalized*, e.g., assuming for example that the largest cost is always equal to 1 for each participant. In this setting, [Caragiannis et al. 2016; Boutilier et al. 2012; Procaccia and Rosenschein 2006] analyze the quality of various social choice mechanisms, and [Filos-Ratsikas et al. 2014] analyze mechanisms for one-sided matching.

2. DISTORTION OF SOCIAL CHOICE MECHANISMS

As discussed above, in our setting one can think of a social choice mechanism as an approximation algorithm which attempts to choose the best possible alternative (maximize social welfare or minimize social cost), but only has access to limited information (ordinal preferences induced by the underlying costs c_i instead of the actual numerical costs c_i). To denote the approximation factor of a social choice mechanism, [Procaccia and Rosenschein 2006] introduced the term *distortion* which

[1] Another important concern is the truthfulness of the voters in reporting their preferences; while e.g., [Filos-Ratsikas et al. 2014; Feldman et al. 2015] consider truthful mechanisms with only ordinal knowledge, in this letter we focus on the issues arising from ordinal knowledge as opposed to the ones arising from strategic voting.

we will continue to use in this letter (the definition below is slightly different from the one in [Procaccia and Rosenschein 2006]; for more about distortion see [Boutilier and Rosenschein 2016]). Formally, the distortion of an alternative $a \in A$ for given ordinal preference profiles σ_i for each $i \in N$ is

$$\max_{c_i \in C(\sigma_i), \forall i} \left[\frac{\sum_i c_i(a)}{\min_{b \in A} \sum_i c_i(b)} \right],$$

where $C(\sigma_i)$ is the set of all numerical cost functions consistent with preferences σ_i. In other words, the distortion of an alternative is an upper bound on how bad this alternative can be compared to the optimum one, no matter what the true underlying costs c_i actually are. As is common with approximation algorithms of other types, the distortion of a social choice mechanism is then defined to be the worst-case ratio of the social cost of the alternative selected by the mechanism, compared to the cost of the optimal alternative.

To understand how well mechanisms which only know ordinal information can perform compared to mechanisms which know the underlying numerical information, let us first consider the extremely simple setting of $|A| = 2$ alternatives. Call these alternatives a and b, and suppose that k voters prefer a to b, and $n-k$ voters prefer b to a (let's assume that no ties are possible). Even if only a single voter prefers a to b, it is still possible that a is the alternative which minimizes social cost: for example the cost of this voter for a could be 0, while the other $n-1$ voters only slightly prefer b to a, and are essentially indifferent. Therefore, it is *impossible* to determine the true optimum alternative based on only ordinal information. Consider, however, the obvious mechanism in which we choose a if and only if $k \geq n/2$. It is not difficult to show that the distortion of this mechanism is at most 3: the worst case has k voters being in the middle between a and b in our metric space, and $n-k$ voters being exactly on top of b (i.e., $c_i(b) = 0$), which leads to a ratio of at most 3 between the total cost of a and the total cost of b. Therefore, without knowing anything about the true costs c_i except that they form an arbitrary metric space, and that the ordinal preferences are induced by these costs, we are able to always choose an alternative which is within a factor of 3 away from the optimum alternative. It is also easy to see that the same example guarantees that no deterministic mechanism can have worst-case distortion better than 3.

In [Anshelevich et al. 2015], we consider many common social choice mechanisms for this setting, and quantify their worst-case distortion. For common positional scoring rules such as plurality, Borda, k-approval, and veto, we prove that the worst case distortion can be high: either $2m-1$ or $2n-1$ where m is the number of alternatives and n is the number of agents/voters. For the Copeland social choice rule, however, we prove that the distortion is always at most 5. This means that, although the Copeland social choice mechanism knows nothing about the metric costs other than the ordinal preferences induced by them, and cannot possibly find the true optimal alternative, it nevertheless *always* selects an alternative whose quality is only a factor of 5 away from optimal! Moreover, due to our lower bound, no deterministic mechanism can do much better than Copeland for the sum objective, because the distortion lower bound for any deterministic mechanism is 3. In [Anshelevich and Postl 2016], we expand our search to randomized mechanisms, and establish tight bounds on their expected distortion. It is not difficult to show that randomized dictatorship has expected distortion strictly better than 3, and thus better than any deterministic mechanism. We also consider specific cases such as the 1-Euclidean metric setting (see for example [Elkind and Faliszewski 2014] and [Procaccia and Tennenholtz 2009]), and give randomized mechanisms which are either optimal or close-to-optimal with respect to their expected distortion. Recently, [Feldman et al. 2015] also considered the distortion of randomized social choice functions: they specifically focus on truthful mechanisms (i.e., the "strategic" setting), and give an extremely interesting truthful mechanism for the 1-Euclidean case with distortion of 2.

In addition to the sum objective function which defines the social cost of an alternative as the sum of all agent costs for that alternative, we also consider the *median objective*: the cost of an alternative is the median of agent costs for that alternative. This captures the objective that the best alternative is the one in which the cost of the median voter is minimized, instead of the average voter. Here the results are even nicer: no deterministic mechanism can always have distortion better than 5, and Copeland achieves this bound exactly [Anshelevich et al. 2015]. If we allow randomized mechanisms, then a more complex mechanism can be shown to have expected distortion of at most 4; the lower bound for randomized mechanisms with the median objective becomes 3, however, so better mechanisms may still be possible [Anshelevich and Postl 2016].

3. BEYOND SOCIAL CHOICE: ORDINAL APPROXIMATION FOR MATCHING AND CLUSTERING

The questions and techniques described above go far beyond social choice. In matching settings, it is natural to suppose that the participating agents have underlying numerical valuations for who they want to be matched with, but the algorithm forming the assignment is only aware of ordinal information. Works such as [Filos-Ratsikas et al. 2014] and [Anshelevich and Sekar 2016] analyze such settings; for example [Anshelevich and Sekar 2016] defines a 1.6-approximation algorithm for the maximum-weight metric matching problem in the presence of only ordinal information. This algorithm uses a mix of greedy and random matchings in order to form matchings with provably high quality. More generally, other graph and clustering problems allow good approximations based only on ordinal information; [Anshelevich and Sekar 2015] provides such approximations for Max k-sum, Densest k-subgraph, and Maximum Traveling Salesman problems.

4. CONCLUSION

In this note we discussed algorithms which only have access to ordinal information, and yet perform almost as well as algorithms with access to the full numerical information. This work, and hopefully future work on this topic, adds to the basic understanding of the *fundamental power of ordinal information,* by determining under which settings and conditions ordinal information is enough to approximate the numerical truth, and when such approximations are impossible. More generally, the same questions can be asked for other types of limited information: how much information about the true input is enough to produce good results?

REFERENCES

ANSHELEVICH, E., BHARDWAJ, O., AND POSTL, J. 2015. Approximating optimal social choice under metric preferences. In *Proc. 29th AAAI Conf. Artificial Intell.* AAAI Press, 777–783.

ANSHELEVICH, E. AND POSTL, J. 2016. Randomized social choice functions under metric preferences. In *IJCAI-16: Proc. 25th Intl. Joint Conference on Artificial Intelligence*.

ANSHELEVICH, E. AND SEKAR, S. 2015. Blind, greedy, and random: Ordinal approximation algorithms for graph problems. In *arXiv:1512.05504 preprint*.

ANSHELEVICH, E. AND SEKAR, S. 2016. Blind, greedy, and random: Algorithms for matching and clustering using only ordinal information. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*.

BOUTILIER, C., CARAGIANNIS, I., HABER, S., LU, T., PROCACCIA, A. D., AND SHEFFET, O. 2012. Optimal social choice functions: A utilitarian view. In *Proceedings of the 13th ACM Conference on Electronic Commerce*. ACM, 197–214.

BOUTILIER, C. AND ROSENSCHEIN, J. S. 2016. Chapter 10: Incomplete information and communication in voting. In *Handbook of Computational Social Choice*, F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, Eds. Cambridge University Press, 223–257.

CARAGIANNIS, I., NATH, S., PROCACCIA, A. D., AND SHAH, N. 2016. Subset selection via implicit utilitarian voting. In *IJCAI-16: Proc. 25th Intl. Joint Conference on Artificial Intelligence*.

CONITZER, V. 2009. Eliciting single-peaked preferences using comparison queries. *Journal of Artificial Intelligence Research 35*, 161–191.

ELKIND, E. AND FALISZEWSKI, P. 2014. Recognizing 1-euclidean preferences: An alternative approach. In *Proc. 7th Int. Symp. Algorithmic Game Theory*. Springer, 146–157.

ENELOW, J. M. AND HINICH, M. J. 1984. *The Spatial Theory of Voting: An Introduction.* Cambridge University Press, New York, NY.

FELDMAN, M., FIAT, A., AND GOLOMB, I. 2015. On voting and facility location. In *arXiv:1512.05868 preprint*.

FILOS-RATSIKAS, A., FREDERIKSEN, S. K. S., AND ZHANG, J. 2014. Social welfare in one-sided matchings: Random priority and beyond. In *Proceedings of the 7th International Symposium on Algorithmic Game Theory (SAGT 2014)*. 1–12.

HARSANYI, J. C. 1976. *Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility.* Springer.

MERRILL, S. AND GROFMAN, B. 1999. *A unified theory of voting: Directional and proximity spatial models.* Cambridge University Press.

PROCACCIA, A. D. AND ROSENSCHEIN, J. S. 2006. The distortion of cardinal preferences in voting. In *Proc. 10th Int. Workshop Cooperative Inform. Agents X*. Springer, 317–331.

PROCACCIA, A. D. AND TENNENHOLTZ, M. 2009. Approximate mechanism design without money. In *Proc. 10th ACM Conf. Electron. Commerce*. ACM, 177–186.