Available online at www.sciencedirect.com

SciVerse ScienceDirect

Electronic Notes in Discrete Mathematics 42 (2013) 57–64

Electronic Notes in DISCRETE MATHEMATICS

www.elsevier.com/locate/endm

Complexity of computing median linear orders and variants

Olivier Hudry

Telecom ParisTech
46, rue Barrault, 75634 Paris Cedex 13, France
hudry@enst.fr

Abstract

Given a finite set X and a collection Π of linear orders defined on X, computing a median linear order (Condorcet-Kemeny's problem) consists in determining a linear order minimizing the remoteness from Π. This remoteness is based on the symmetric distance, and measures the number of disagreements between O and Π. In the context of voting theory, X can be considered as a set of candidates and the linear orders of Π as the preferences of voters, while a linear order minimizing the remoteness from Π can be adopted as the collective ranking of the candidates with respect to the voters' opinions. This paper studies the complexity of this problem and of several variants of it: computing a median order, computing a winner according to this method, checking that a given candidate is a winner and so on. We try to locate these problems inside the polynomial hierarchy.

Keywords: Complexity, Turing transformation, NP-completeness, NP-hardness, polynomial hierarchy, linear order, Condorcet-Kemeny problem, Slater problem, voting theory, pairwise comparison method, median order, linear ordering problem, feedback arc set, majority tournament.

[1] Research supported by the ANR project "Computational Social Choice" ANR-09-BLAN-0305

http://dx.doi.org/10.1016/j.endm.2013.05.146

1 Introduction

In an election, assume that we are given a finite set X of n candidates and a collection (or multi-set) $\Pi = (O_1, O_2, ..., O_m)$, called a profile, of the preferences O_i of m voters ($1 \leq i \leq m$) who want to rank the n candidates. Assume moreover that the individual preferences O_i ($1 \leq i \leq m$) of the m voters are linear orders over X. Note that the linear orders involved in the profile may be the same: two different voters may share the same preference. In order to aggregate these m linear orders into a linear order which can be considered as the collective ranking, Condorcet [8] suggested to compute, for each pair of candidates x, y (with $x \neq y$), the number m_{xy} of voters who prefer x to y and the number m_{yx} of voters who prefer y to x. Then x is collectively preferred to y if we have $m_{xy} > m_{yx}$. Unfortunately, as pointed out by Condorcet himself, the relation thus defined does not necessarily provide a linear order. More precisely (see the example below), a majority may prefer a candidate x to another candidate y, another majority may prefer y to a third candidate z, and still another majority may prefer z to x. This is the well-known "voting paradox" or also "Condorcet effect" [12].

When such a situation occurs, one possibility to define the collective preference consists in computing a linear order which summarizes the individual preferences as well as possible, more precisely which minimizes the number of disagreements with respect to Π (see below). A linear order minimizing this number of disagreements is called a *median linear order* [3], or sometimes a Kemeny order (though the problem considered by Kemeny deals in fact with complete preorders, see [18]). The candidate who beats the other candidates in such a median order will be called a *winner* in the following.

The problem that we consider here consists in studying the complexity of computing such a median order or such a winner. The studied problems are more precisely defined in Section 3, after some definitions and notation specified in Section 2. The complexity results are summarized, without their proofs, in Section 4.

2 Definitions and notation

2.1 Symmetric difference distance, remoteness, median order

Let X be a finite set. If R is a binary relation defined on X and if x and y are two elements of X, we write xRy if x is in relation with y with respect to R. Let R and S be two binary relations defined on X. The symmetric difference distance $\rho(R, S)$ between R and S is defined by, where Δ denotes the usual symmetric difference between sets:

$$\delta(R, S) = |R \Delta S|,$$

i.e.

$$\delta(R, S) = |\{(x, y) \in X^2 \text{ s.t. } [xRy \text{ and not } xSy] \text{ or } [\text{not } xRy \text{ and } xSy]\}|.$$

This distance, which owns good axiomatic properties (see [2]), measures the number of disagreements between R and S. From this distance, we may define a *remoteness* ρ between the profile $\Pi = (O_1, O_2, ..., O_m)$ and any linear order O defined on X by:

$$\rho(\Pi, O) = \sum_{i=1}^{m} \delta(O_i, O).$$

Thus $\rho(\Pi, O)$ measures the total number of disagreements between Π and O. A *median linear order of* Π is a linear order O^* which minimizes the remoteness from Π:

$$\rho(\Pi, O^*) = \min_{O \in \Omega(X)} \rho(\Pi, O),$$

where $\Omega(X)$ denotes the set of all the linear orders defined on X; $\mu(\Pi)$ will denote this minimum value:

$$\mu(\Pi) = \min_{O \in \Omega(X)} \rho(\Pi, O).$$

2.2 Complexity classes

As it is usual, we will distinguish between decision problems (i.e. problems for which a question is set of which the answer is "yes" or "no") and the other types of problems (as optimization problems or search problems). The usual classes P and NP are assumed to be known, as well as the concept of NP-complete or NP-hard problems (see for instance [11] for their definitions). The class P^{NP} or $P(NP)$, or Δ_2^P (or simply Δ_2) contains the decision problems which can be solved by applying, with a polynomial (with respect to the size of the instance) number of calls, a subprogram able to solve an appropriate problem belonging to NP (usually, an NP-complete problem). In other words, P^{NP} contains the decision problems $\mathcal{P}$ such that there exists a problem $\mathcal{Q}$ belonging to NP with $\mathcal{P} <_T \mathcal{Q}$, where $<_T$ denotes the Turing transformation. Such a problem $\mathcal{P}$ is sometimes called NP*-easy* (though it can be NP-hard as well; a problem which is simultaneously NP-easy and

NP-hard is said to be NP-*equivalent*: this means that the complexity of an NP-equivalent problem is the same, up to some polynomials, as the complexity of NP-complete problems). This class is usually considered as the first step of the polynomial hierarchy above NP and co-NP (with this respect, the notation Δ_2 is more usual when dealing with this polynomial hierarchy; anyway, we shall keep the notation P^{NP}, more informative and of which the meaning is easier to memorize). Indeed, P^{NP} contains NP obviously as well as the class co-NP: $NP \cup$ co-$NP \in P^{NP}$. It also contains the class L^{NP}, also denoted by Θ_2^P, which contains the decision problems that can be solved by applying, a logarithmic (still with respect to the size of the instance) number of times, a subprogram able to solve an appropriate problem belonging to NP (usually, an NP-complete problem). This class contains the classes NP and co-NP and is contained in the class P^{NP}. It also contains the class $P^{NP[1]}$, that we shall note 1^{NP} in the sequel for the homogeneity of the notation, of the problems that can be solved by applying once a subprogram able to solve an appropriate problem belonging to NP (usually, an NP-complete problem); note that 1^{NP} contains NP and co-NP. All in all, we have the following inclusions: $NP \cup$ co-$NP \subseteq 1^{NP} \subseteq L^{NP} \subseteq P^{NP}$.

For the problems which are not decision problems (sometimes called "function problems"), we generalize these classes by adding "F" in front of their names (see [17]). For example, the class FP^{NP} or $F\Delta_2^P$ (respectively the class FL^{NP}) contains the optimization problems and the search problems which can be solved by the application of a subprogram able to solve an appropriate problem belonging to NP a polynomial (respectively logarithmic) number of times.

3 Complexity results

We may now specify the problems that we consider and the complexity results related to them.

The NP-hardness of the computation of a median linear order of a profile of linear orders has been known for a long time if m is assumed to be large enough with respect to n (see for instance [4], [14], [15]; more generally, see also [7]). More precisely, the decision problem associated with the computation of $\mu(\Pi)$ is NP-complete. More recently, C. Dwork et alii [10] have shown that the computation of a median linear order remains NP-hard if m is equal to 4 (hence we deduce easily that it is NP-hard for all given even number m with $m \geq 4$; on the other hand, the problem is polynomial for $m = 2$, see [7]; the complexity for m odd and small is unknown). Moreover, E. Hemaspaandra et alii [13] have also been interested in the complexity of the problem consisting in verifying whether a given candidate is a winner (see below).

We now pay attention to the complexity of the following seven problems, related to the aggregation of the profile of linear orders into a median linear order:

PROBLEM (P_1). Given a profile $\Pi = (O_1, O_2, ..., O_m)$ of linear orders, compute the value of $\mu(\Pi)$.

PROBLEM (P_2). Given a profile $\Pi = (O_1, O_2, ..., O_m)$ of linear orders, compute a median order $O^*(\Pi)$ of Π.

PROBLEM (P_3). Given a profile $\Pi = (O_1, O_2, ..., O_m)$ of linear orders, compute all the median order $O^*(\Pi)$ of Π.

PROBLEM (P_4). Given a profile $\Pi = (O_1, O_2, ..., O_m)$ of linear orders, compute a of Π.

PROBLEM (P_5). Given a profile $\Pi = (O_1, O_2, ..., O_m)$ of linear orders, compute all the winners of Π.

PROBLEM (P_6). Given a profile $\Pi = (O_1, O_2, ..., O_m)$ of linear orders and an element x of X, determine whether x is a winner of Π.

PROBLEM (P_7). Given a profile $\Pi = (O_1, O_2, ..., O_m)$ of linear orders and a linear order O, determine whether O is a median linear order of Π.

To study the complexity of these problems, we use the NP-hardness of Slater's problem, which can be stated as follows [19]:

SLATER'S PROBLEM. Given a profile Π containing only one tournament defined on X, compute a median linear order of Π.

Slater's problem is known to be NP-hard (see [1], [5], [9], [16]). From this NP-hardness, we may draw the following theorems:

THEOREM 1. Problems (P_1) to (P_6) are NP-hard.

Note that P_7 is not known to be NP-hard. More precisely, we may show that P_7 belongs to co-NP, but is not known to be co-NP-complete:

THEOREM 2. Problems (P_7) belongs to co-NP.

Under the usual hypothesis, i.e. $P \neq NP$, Theorem 1 shows that the exact resolution of Problems (P_1) to (P_6) requires an exponential time. In other words, it provides a lower bound of the complexity of Problems (P_1) to (P_6). Theorem 3 provides an upper bound of this complexity:

THEOREM 3. Problems (P_1), (P_2), (P_4), (P_5) belong to FP^{NP}. Problem (P_6) belongs to L^{NP}.

Note that E. Hemaspaandra *et alii* studied the complexity of Problem (P_6) in [13]: they prove that (P_6) is L^{NP}-complete. In other words, (P_6) belongs to L^{NP} and, inside this class, it belongs to the most difficult problems (in the usual meaning of complexity theory). This result incites to state the following conjectures:

CONJECTURES. Problems (P_1), (P_2), (P_4), (P_5) are FP^{NP}-complete; (P_7) is co-NP-complete.

For Problem (P_3), note that there are some cases with m even for which the number of median linear orders is equal to $n!$: in other words, all the linear orders defined on X are median. When m is odd, the maximum number of median linear orders is not known precisely, but we know (see [6], [7], [20]) that, when n is a power of 3, it lies between $\exp[\frac{\ln 3}{4}(3n - 2\log_3 n - 3)]$ and $\frac{\alpha n\sqrt{(n)}n!}{2^n}$, where α is a constant.

References

[1] N. Alon: Ranking tournaments. SIAM Journal on Discrete Mathematics 20 (1), 137–142, 2006.

[2] J.-P. Barthélemy: Caractérisations axiomatiques de la distance de la différence symétrique entre des relations binaires, Mathématiques et Sciences humaines 67,

85–113, 1979.

[3] J.-P. Barthélemy, B. Monjardet: The median procedure in cluster analysis and social choice theory, Mathematical Social Sciences 1, 235–267, 1981.

[4] J.J. Bartholdi III, C.A. Tovey, M.A. Trick: Voting schemes for which it can be difficult to tell who won the election. Social Choice and Welfare 6, 157–165, 1989.

[5] P. Charbit, S. Thomasse, A. Yeo: The minimum feedback arc set problem is NP-hard for tournaments. Combinatorics, Probability and Computing 16 (1), 1–4, 2007.

[6] I. Charon, O. Hudry: Slater orders and Hamiltonian paths of tournaments. Electronic Notes in Discrete Mathematics 5, 60–63, 2000.

[7] I. Charon, O. Hudry: An updated survey on the linear ordering problem for weighted or unweighted tournaments. Annals of Operations Research 175, 2010, 107–158, 2010.

[8] M.J.A.N. Caritat, marquis de Condorcet: Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Imprimerie royale, Paris, 1785.

[9] V. Conitzer: Computing Slater Rankings Using Similarities Among Candidates. Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06), Boston, MA, USA, 613–619, 2006. Early version: IBM Research Report RC23748, 2005.

[10] C. Dwork, R. Kumar, M. Naor, D. Sivakumar: Rank aggregation methods for the Web. Proceedings of the 10th International Conference on World Wide Web (WWW10), 613–622, 2001.

[11] M.R. Garey, D.S. Johnson: Computers and intractability, a guide to the theory of NP-completeness. Freeman, New York, 1979.

[12] G.Th. Guilbaud: Les théories de l'intérêt général et le problème logique de l'agrégation. Économie appliquée 5 (4), 501–584, 1952. Reprint in Éléments de la théorie des jeux, Dunod, Paris, 1968. English translation: Theories of the general interest and the logical problem of aggregation, Electronic Journal for History of Probability and Statistics 4, 2008.

[13] E. Hemaspaandra, H. Sparowski, J. Vogel: The complexity of Kemeny elections, Theoretical Computer Science 349, 382–391, 2005.

[14] O. Hudry: Recherche d'ordres médians : complexité, algorithmique et problèmes combinatoires. PhD thesis, Telecom ParisTech, Paris, 1989.

[15] O. Hudry: NP-hardness results on the aggregation of linear orders into median orders. Annals of Operations Research 163 (1), 63–88, 2008.

[16] O. Hudry: On the complexity of Slater's problems. European Journal of Operational Research 203, 216–221, 2010.

[17] D.S. Johnson: A catalog of complexity classes. In J. van Leeuwen (ed.), Handbook of Theoretical Computer Science Vol. A: Algorithms and Complexity, Elsevier, Amsterdam, 67–161, 1990.

[18] J.G. Kemeny: Mathematics without numbers, Daedalus 88, 577–591, 1959.

[19] P. Slater: Inconsistencies in a schedule of paired comparisons, Biometrika 48, 303–312, 1961.

[20] F. Woirgard: Recherche et dénombrement des ordres médians des tournois. PhD thesis, Telecom ParisTech, Paris, 1997.