

Aggregating Interval Orders by Propositional Optimization

Daniel Le Berre, Pierre Marquis, and Meltem Öztürk*

Université Lille Nord de France, F-59000 Lille, France

Université d'Artois, CRIL, F-62307 Lens, France

CNRS, UMR 8188, F-62307 Lens, France

{leberre, marquis, ozturk}@cril.univ-artois.fr

Abstract. Aggregating preferences for finding a consensus between several agents is an important issue in many fields, like economics, decision theory and artificial intelligence. In this paper we focus on the problem of aggregating interval orders which are special preference structures allowing the introduction of thresholds for the indifference relation. We propose to solve this problem by first translating it into a propositional optimization problem, namely the Binate Covering Problem, then to solve the latter using a MAX-SAT solver. We discuss some properties of the proposed encoding and provide some hints about its practicability using preliminary experimental results.

Keywords: Interval orders, preference modelling and aggregation, propositional reasoning, Boolean optimization

1 Introduction

Aggregating preferences for finding a consensus between several agents is an important issue in many fields, like economics, decision theory, and artificial intelligence. Given the preferences of a set of agents (or voters) over a set of alternatives (or candidates), where preferences are generally formulated as binary relations such as strict preference, indifference, etc., preference aggregation aims at determining a collective preference relation representing as much as possible the individual preferences.

However many works have shown through paradoxes and impossibility theorems that preference aggregation is not an easy task, the famous ones are Condorcet's paradox [3], Arrow's theorem [2].

A common approach is to consider a preference relation as a complete preorder (i.e., a reflexive and transitive relation). In the above results each voter is supposed to present a complete preorder over the set of alternatives. However, such a model for preferences does not prove adequate to all situations, and other models (generalizing the complete preorder one) have been pointed out. In particular, different structures have been introduced for defining thresholds as in the famous example given by Luce [10] about a cup

* This work has been supported in part by two Projects: ANR-05-BLAN-0384 and PEPS-09 37. This support is gratefully acknowledged by the authors. The authors also thank anonymous referees for their helpful comments and suggestions.

of coffee. Indeed, in contrast to the strict preference relation, the indifference relation induced by such structures is not necessarily transitive. Semiorders may form the simplest class of such structures and they appear as a special case of interval orders. The axiomatic analysis of what we call now interval orders has been given by Wiener [16], then the term “semiorders” has been introduced by Luce [10] and many results about their representations are available in the literature (for more details see [5,13]). Roughly speaking, within an interval order, alternative x_1 is strictly preferred to alternative x_2 if and only if the evaluation of x_1 is greater than the evaluation of x_2 plus a threshold. It is easy to see that preorders are special cases of interval orders where the value of threshold is fixed to zero.

In this paper we consider the interval order aggregation problem; to solve it, we propose a method based on the Kemeny distance which makes use of a translation into the Binate Covering Problem [4]. More precisely, we consider the case where the preferences of voters are interval orders and we try to find a final interval order which will be “as close as possible” to the set of voter’s preferences. Let us note that having an interval order as a result of an aggregation is not a drawback for pointing out an undominated alternative since it is known that when the asymmetric part of a binary relation is transitive, which is the case of interval orders, there is always at least one such undominated alternative [14]. Moreover it is natural to ask an interval order as a result when preferences of voters are interval orders. Finally, as we will show it, even when the input preferences are preorders, focusing on interval orders as outputs is a way to get an aggregation which is closer to the given preferences than when preorders are targeted (just because the set of all interval orders over the alternatives is a superset of the set of all preorders over the alternatives).

2 Aggregation as Optimisation

In this paper, we consider a finite set of alternatives A on which preference relations are applied ($|A| = n$), we represent with a, b, c, \dots specific elements of A and x_1, x_2, \dots or x, y, z, \dots variables ranging over the set A . We have a finite set of voters $V = \{v_1, \dots, v_m\}$ ($|V| = m$). Voters express their preferences by the help of two binary relations represented in an explicit way as n^2 -matrices: the notation $aP_i b$ (resp. $aI_i b$) means that the voter v_i prefers strictly alternative a to b (resp. is indifferent between a and b). $\#p(a, b)$ (resp. $\#i(a, b)$) is the number of voters v_i for whom $aP_i b$ (resp. $aI_i b$) holds. We call a *profile*, the set of voter’s preference relations and denote it by $X = \{\langle P_1, I_1 \rangle, \langle P_2, I_2 \rangle, \dots, \langle P_m, I_m \rangle\}$; its size is in $\mathcal{O}(m \cdot n^2)$.

The result of the aggregation is also expressed by two relations that we denote by P and I (P^{-1} represents the inverse of P : $\forall x, y \in A, xP^{-1}y$ iff yPx). aPb (resp. aIb) means that alternative a is preferred to alternative b (resp. a and b are indifferent) in the resulting order. We denote it as $f(\langle P_1, I_1 \rangle, \langle P_2, I_2 \rangle, \dots, \langle P_m, I_m \rangle) = \langle P, I \rangle$.

The pair $\langle P, I \rangle$ is called a *preference structure* if and only if P is asymmetric, I is reflexive and symmetric, $P \cup I$ is complete and $P \cap I$ is empty. Such a pair is an interval order if and only if it is a preference structure and satisfies a property called Ferrers relation.¹

¹ $\forall x, y, z, t \in A, xPy \wedge yIz \wedge zPt \Rightarrow xPt$.

Definition 1. Let P and I be binary relations on $A \times A$, $\langle P, I \rangle$ is an interval order if and only if

- i) $P \cup I \cup P^{-1} = A \times A$ (completeness),
- ii) $P \cap I = \emptyset$ (exclusivity),
- iii) P is asymmetric, I is symmetric and reflexive,
- iv) $P.I.P \subset P$ (Ferrers relation).

The numerical representation of interval orders is as in the following:

Proposition 1. [5] Let P and I be binary relations on $A \times A$, $\langle P, I \rangle$ is an interval order if and only if there exist a mapping g from A to \mathbb{R} and a mapping q from \mathbb{R} to \mathbb{R}^+ such that for any $x, y \in A$, we have:

$$xPy \Leftrightarrow g(x) > g(y) + q(g(y)).$$

$$xIy \Leftrightarrow g(x) \leq g(y) + q(g(y)).$$

Interval orders are quasi-orders (i.e., orders with a transitive asymmetric part). Gibbard ([6]) has showed that Arrow's theorem can be generalized to the case of quasi-orders, hence we have this impossibility result for interval orders. Pirlot and Vincke ([13]) have focused also on this theorem with a special attention to interval orders. Before presenting this theorem we first need the following definitions in order to state it formally:

weak unanimity an aggregation procedure satisfies the weak unanimity condition if and only if, for all voters $v_i \in V$ and for all $a, b \in A$, $aP_i b \implies aPb$;

non-dictatorship an aggregation procedure satisfies the non-dictatorship condition if and only if, for no voter $v_i \in V$ such that for all possible preferences of other voters and for all alternatives a and b $aP_i b \implies aPb$;

independence of irrelevant alternatives an aggregation procedure satisfies the independence of irrelevant alternatives condition if and only if $\forall (\langle P_1, I_1 \rangle, \dots, \langle P_m, I_m \rangle), (\langle P'_1, I'_1 \rangle, \dots, \langle P'_m, I'_m \rangle), \forall a, b \in A$,
 $(\langle P_1, I_1 \rangle, \dots, \langle P_m, I_m \rangle) / \{a, b\} = (\langle P'_1, I'_1 \rangle, \dots, \langle P'_m, I'_m \rangle) / \{a, b\} \implies$
 $(\langle P, I \rangle) / \{a, b\} = (\langle P', I' \rangle) / \{a, b\}$
 where $\langle P, I \rangle$ is the result on $(\langle P_1, I_1 \rangle, \dots, \langle P_m, I_m \rangle)$, and $\langle P, I \rangle / \{a, b\}$ is the restriction of $\langle P, I \rangle$ to $\{a, b\}$, etc.

Theorem 1 (Generalized Arrow's Theorem). [13] If $|A| \geq 4$, if X is the set of all n -tuples of interval orders on A and if Y is the set of all interval orders on A , then there is no $(X-Y)$ -aggregation procedure² satisfying simultaneously weak unanimity, non-dictatorship and independence conditions.

Note that if all the considered relations are complete preorders, Theorem 1 is exactly Arrow's theorem with $|A| \geq 3$. We need four alternatives for interval orders because of the definition of Ferrers relation.

There exist a number of papers addressing the aggregation issue for binary relations as an optimization problem. Typically, a 0/1 linear program is targeted. Contrastingly,

² X represents here the set of voter's preferences and Y the resulting order.

in our approach, we associate to each profile of binary relations an instance of BCP, the so-called *Binare Covering Problem* [4], where the set of constraints is not any set of 0/1 linear inequations but a SAT instance. This problem has been studied for decades by the circuit community where it is important for logic synthesis (minimizing the number of components needed to perform a given operation).

From a theoretical standpoint, like 0/1 linear programming, BCP is an NP-hard optimization problem (and the associated decision problem is in NP) (see e.g. [12]). In practice, each clause can be translated into an equivalent 0/1 linear inequation, but the converse does not hold. The specific format of the constraints considered in BCP (compared to 0/1 linear programs) enables us to take advantage of the power of existing MAX-SAT solvers in order to solve its instances in a more efficient way from the practical side.

To our knowledge there is a limited number of studies related to the aggregation of interval orders. Pirlot and Vincke [13] have shown that the schemes that work well for complete preorders such as lexicographic procedure or Borda's sum of ranks do not lean themselves easily to the generalization with interval orders. They proposed two types of aggregation procedures: one consisting in aggregating numerical representations into a "global evaluation" function, and the other inspired from pairwise comparison methods.

In this paper we propose a hybrid approach consisting in finding an interval order being optimal in the sense of minimal Kemeny distance [8] to the input profile. Intuitively, ranking the alternatives according to Kemeny's rule can be seen as the best compromise since on average it gives the "closest" social preference to the individual preferences. Our idea can be summarized as in the following:

1. Determine all pairwise comparisons for which all the voters have the same opinion and build a partial order that preserves those comparisons.
2. Search within the set of feasible interval orders in order to find a closest one to the input profile.

The first step can be easily achieved the following way:

$$\forall v_i \in V, \forall x, y \in A, xP_i y, \implies xPy,$$

$$\forall v_i \in V, \forall x, y \in A, xI_i y, \implies xIy.$$

The resulting $\langle P, I \rangle$ is a partial order.

From partial order to interval orders Naturally this step provides in the majority of cases many interval orders. The worst case that we may expect is when the partial order provided in the first step is empty. In this case we have to find all the interval orders containing n objects (n being the cardinality of A). This case gives an idea on the number of interval orders that we may have. Stanley [15] has precised the number of interval orders with n elements; for this he has made use of relations between interval orders and hyperplanes arrangement. The coefficient of the following polynomial provides the number of interval orders:

$$z = \sum_{k \geq 0} c_k \frac{x^k}{k!}$$

$$z = 1 + x + 3 \frac{x^2}{2!} + 19 \frac{x^3}{3!} + 195 \frac{x^4}{4!} + 2831 \frac{x^5}{5!} + 53703 \frac{x^6}{6!} + 1264467 \frac{x^7}{7!} + \dots$$

z is the unique power series satisfying $\frac{z'}{z} = y^2$, $z(0) = 1$ where $1 = y(2 - e^{xy})$. The value c_k of the serie z is the number of interval orders on k alternatives. This number grows exponentially on the number of alternatives: for instance, with just 7 alternatives we have more than one million interval orders. However, we will see in the following that we do not need to represent those interval orders explicitly. We denote by $\langle P^{(1)}, I^{(1)} \rangle, \langle P^{(2)}, I^{(2)} \rangle, \dots$ these interval orders.

Discriminating interval orders In our approach, the distance of an interval order $\langle P^{(i)}, I^{(i)} \rangle$ to the input profile X , $D(\langle P^{(i)}, I^{(i)} \rangle, X)$, will be calculated as the sum of its distance to each voter's order $\langle P_j, I_j \rangle$. Let us denote this distance by $d(\langle P^{(i)}, I^{(i)} \rangle, \langle P_j, I_j \rangle)$:

$$D(\langle P^{(i)}, I^{(i)} \rangle, X) = \sum_{\langle P_j, I_j \rangle \in X} d(\langle P^{(i)}, I^{(i)} \rangle, \langle P_j, I_j \rangle)$$

The distance d is computed using the difference between pairwise comparisons in the following way:

$$d(\langle P^{(i)}, I^{(i)} \rangle, \langle P_j, I_j \rangle) = \sum_{(x,y) \in A^2} \delta_{\langle P^{(i)}, I^{(i)} \rangle, \langle P_j, I_j \rangle}(x, y)$$

$$\delta_{\langle P^{(i)}, I^{(i)} \rangle, \langle P_j, I_j \rangle}(x, y) = \begin{cases} p2p & \text{if } (xP^{(i)}y \text{ and } yP_jx) \text{ or } (yP^{(i)}x \text{ and } xP_jy) \\ 0 & \text{if } (xP^{(i)}y \text{ and } xP_jy) \text{ or } (xI^{(i)}y \text{ and } xI_jy) \\ p2i & \text{otherwise} \end{cases}$$

Here $p2p$ and $p2i$ are nonnegative constant numbers. The rationale for this definition of d is to put a penalty when there is a discrepancy of preference relation between the comparison given by a voter and the one of the interval order. Naturally, a discrepancy of a strict preference (for instance xPy) to the inverse of this preference (yPx) is at least as problematic as a discrepancy of a strict preference (for instance xPy) to an indifference (xIy) for this reason we suggest that $p2p \geq p2i$. Even more one can impose the strict inequality ($p2p > p2i$) which will guarantee to have as a result aIb when the profile with three voters is aP_1b , aI_2b and bP_3a . Note that the distance used by Hudry ([7]) imposes $p2p = p2i = 1$ and provides as a result three interval orders (aPb , bPa and aIb) for this example.

We propose to represent the set of interval orders to be *implicitly* considered in the second step using propositional constraints (clauses). Then, computing the interval orders closest to the profile is encoded as minimizing an objective function. Accordingly, we reduce our interval order optimization problem to the BCP one.

3 Translation into the Binat Covering Problem

We first need propositional variables v_{xPy} and v_{xIy} to represent all pairs of the form xPy ($\forall x \neq y \in A$) and xIy ($\forall x, y \in A, x \leq y$). As a consequence, $n^2 - n + \frac{n \times (n-1)}{2} + n = \frac{3 \times n^2 - n}{2}$ variables must be considered. For instance, for 4 alternatives, we need 22 propositional variables. For 16 alternatives, we need 376 propositional variables.

3.1 Implicit representation of interval orders

Structural constraints The following constraints express that the result of the aggregation must be an interval order. They do not depend on the voters.

- $P \cup I$ is complete: $\forall x < y \in A, v_{xPy} \vee v_{xIy} \vee v_{yPx}$,
- P is asymmetric: $\forall x < y \in A, \neg(v_{xPy} \wedge v_{yPx}) \equiv \neg v_{xPy} \vee \neg v_{yPx}$,
- P and I are exclusive: $\forall x \neq y \in A, \neg(v_{xPy} \wedge v_{xIy}) \equiv \neg v_{xPy} \vee \neg v_{xIy}$,
- I is symmetric by construction because a single propositional variable v_{xIy} represents both xIy and yIx .
- I is reflexive: $\forall x \in A, v_{xIx}$ is forced to be true,
- $P \cup I$ is Ferrers: $\forall x, y, z, t \in A, x \neq y, z \neq t, x \neq t, y \neq t, x \neq z (v_{xPy} \wedge v_{yIz} \wedge v_{zPt}) \Rightarrow v_{xPt}$,

Note that we need to generate $2n(n-1) + n + n(n-1)(n-2)^2 = n(n^3 - 5n^2 + 10n - 5)$ structural constraints plus the unit clauses needed to preserve unanimity (see below). For 4 alternatives, it means at least 76 constraints. For 16 alternatives, it means at least 47536 constraints. The $O(n^4)$ space required by the above encoding is clearly dominated by the cost of ensuring Ferrers condition.

Unanimity constraints Those additional constraints encode unanimity for both P and I . They are generated according to the votes. Since they force the truth value of some variables, they simplify in practice the computation of the best interval order.

- Unanimity for P : $\forall x \neq y \in A$, if $\#p(x, y) = |V|$ then xPy is forced to be true,
- Unanimity for I : $\forall x \neq y \in A$, if $\#i(x, y) = |V|$ then xIy is forced to be true.

3.2 Distance between interval orders and the profile

The coefficient associated to each variable is computed according the individual penalty δ defined earlier and the number of voters that disagree with the interval order.

- $\forall x, y \in A$, satisfying $I(x, y)$ entails that voters that strictly prefer x to y or y to x disagree with that fact, with a simple individual penalty of $p2i$. As a consequence, the coefficient of the variables is exactly $p2i(\#p(x, y) + \#p(y, x))$,
- $\forall x, y \in A$, satisfying $P(x, y)$ entails that voters that are indifferent between x and y disagree with that fact with a simple penalty of $p2i$, while the voters that strictly prefer y to x disagree with that fact with an individual penalty of $p2p$. So the coefficient of those variables is exactly $p2i\#i(x, y) + p2p * \#p(y, x)$.

The objective function of the binate covering problem associated with X is denoted by $score_X(\langle P, I \rangle)$ and is

$$\sum_{x \leq y \in A} p2i(\#p(x, y) + \#p(y, x))v_{xIy} + \sum_{x \neq y \in A} (p2i\#i(x, y) + p2p * \#p(y, x))v_{xPy}.$$

Thus the space needed to represent the objective function is in $\mathcal{O}(n^2 \cdot \log_2(m))$. Interestingly, the space needed by the encoding (constraints and objective function)

is only logarithmic in the number of voters. This renders the approach feasible for a large number of voters. On the other hand, the space needed by the encoding is in $\mathcal{O}(n^4)$; considering that MAX-SAT solvers are currently able to solve *some* instances with millions of variables, it might be possible to solve aggregation problems up to roughly 40 alternatives (which leads to 2 millions of clauses using the above encoding).

The result of the aggregation step is any interval order which minimizes the value of the objective function. An important issue is to determine whether it makes sense to use of sophisticated SAT engine (or 0/1 linear program solver) to solve those specific BCP instances stemming from a translation from instances of the aggregation problem. [7] gave a positive answer to this query, by identifying the complexity of the following decision problem: SCORE:

Input: A finite profile X of binary relations $\langle P, I \rangle$ on A and a nonnegative integer k .

Question: Does there exist an interval order $\langle P, I \rangle$ on A such that $score_X(\langle P, I \rangle) \leq k$?

In a nutshell Hudry showed that SCORE is NP-complete as soon as the number of voters m is "sufficiently" large compared to the number n of alternatives, even in the restricted case when X consists of linear orders only, provided that $p2p = p2i = 1$. This justifies to take advantage of algorithms running in exponential time (as MAX-SAT solvers) in the worst case, since polynomial time ones are hardly expected.

Hudry's NP-hardness result extends easily to our framework when the parameters $p2p$ and $p2i$ are such that $p2p = p2i$ since linear orders are interval orders; on the other hand, the membership to NP of the SCORE problem is obvious in our setting: in order to determine that an instance of this decision problem is positive, it is enough to guess a binary relation $\langle P, I \rangle$ on A (its size is $\mathcal{O}(n^2)$), then to check that it is an interval order (this can be easily achieved in polynomial time in the size of the relation), and finally to compute in polynomial time $score_X(\langle P, I \rangle)$ in order to compare it with k .

Our MAX-SAT algorithm for the BCP problem is a branch-and-bound algorithm. During the search, each time a (partial) assignment is found that satisfies all the constraints, the corresponding score is computed (each unassigned variable is set to 0) and a constraint which eliminates all the assignments leading to a greater bound is added, so that whenever a partial assignment leads to a score which is worse than this bound, a backtrack occurs. Its worst-case time complexity is simply exponential in the number of variables under consideration (hence linear in the size of X) and its space complexity is linear in the size of the constraints (hence quadratic in the size of X).

3.3 Examples

As a matter of illustration, let us consider the following examples. For these examples we suppose that $p2i = 1$ and $p2p = 2$.

Example 1. Consider first a case with 5 voters and 4 alternatives with the preferences of voters shown in Table 1.

These preferences of voters can be compactly represented in a matrix where $\forall x_i, x_j$, $P(x_i, x_j) = \alpha$ means that there are α voters who prefer alternative x_i to alternative

V_1	a	b	c	d	V_2	a	b	c	d	V_3, V_4, V_5	a	b	c	d
a	I	P	P	P	a	I	P	I	P	a	I	P	P	P
b	P^{-1}	I	I	P	b	P^{-1}	I	P^{-1}	P	b	P^{-1}	I	I	P
c	P^{-1}	I	I	P	c	I	P	I	P	c	P^{-1}	I	I	P
d	P^{-1}	P^{-1}	P^{-1}	I	d	P^{-1}	P^{-1}	P^{-1}	I	d	P^{-1}	P^{-1}	P^{-1}	I

Table 1. Pairwise comparisons on 4 alternatives given by 5 voters

x_j . Table 2 represents the matrix related to the previous example. Accordingly, this matrix contains all the information needed for running our aggregation procedure. Our method find as a result the following interval order: $aPb, cPa, dPa, cPb, dPb, cPd$ (its distance to the profile is 12).

P	a	b	c	d
a	0	5	1	2
b	0	0	0	2
c	3	4	0	5
d	3	3	0	0

Table 2. The number of voters agreeing for a strict preference

Example 2. Here is a second example; Table 3 shows the pairwise comparisons given by three voters on three alternatives (a, b, c).

V_1	a	b	c	V_2	a	b	c	V_3	a	b	c
a	I	P	I	a	I	P	P	a	I	I	I
b	P^{-1}	I	P^{-1}	b	P^{-1}	I	I	b	I	I	I
c	I	P	I	c	P^{-1}	I	I	c	I	I	I

Table 3. The profile of Example 2

The result of our aggregation procedure provides a unique interval order as close as possible to the input profile. It is not a preorder (a is preferred to b and all the other comparisons are indifference), despite the fact that each preference relation in the input profile is a preorder.

3.4 More than one solution is often the case

Clearly enough, there is no guarantee in general that a unique interval order $\langle P, I \rangle$ exists, leading to a minimal value s^* for the objective function $score_X(\langle P, I \rangle)$. This problem is inherent to the fact that voters may have different preferences, and it may happen in very simple scenarios, for instance when A consists of two alternatives a and b and X consists of two interval orders $\langle P_1, I_1 \rangle$ and $\langle P_2, I_2 \rangle$ on A so that aP_1b and aI_2b : in

such a case, both $\langle P_1, I_1 \rangle$ and $\langle P_2, I_2 \rangle$ lead to the minimal value $s^* = p2i$, but not to the same sets of undominated alternatives. Nevertheless, this plurality is problematic since decisions made using only one of such optimal interval orders are not necessarily robust, in the sense that the choice of another optimal interval order could question them. Typically decisions are made by comparing alternatives or determining undominated ones. While robustness is a complex notion, a sufficient condition for a comparison to be robust is when it holds for every optimal interval order, and similarly an alternative is robustly undominated when it is undominated for all optimal interval orders. Formally, the following decision problems have to be considered: NEC-COMP(R):

Input: A finite profile X of binary relations $\langle P, I \rangle$ on A and two alternatives a, b from A .

Question: Is it the case that every interval order $\langle P, I \rangle$ on A satisfying $score_X(\langle P, I \rangle) = s^*$ is such that aRb ? (where $R = P$ or $R = I$)?

NEC-UNDOM:

Input: A finite profile X of binary relations $\langle P, I \rangle$ on A and an alternative a from A .

Question: Is it the case that for every interval order $\langle P, I \rangle$ on A such that $score_X(\langle P, I \rangle) = s^*$, we have $a(P \cup I)b$ for every $b \in A$?

Those decision problems are "mildly" hard, since they belong to the complexity class Θ_2^P , consisting of all decision problems which can be solved in deterministic polynomial time using logarithmically many calls to an NP oracle. In order to prove the membership of NEC-COMP(R) and NEC-UNDOM to Θ_2^P , we consider the complementary problems and show them in Θ_2^P as well (this class is closed under complementation). We have already seen that SCORE is in NP. Now, the value of s^* can be computed by binary searching it within the bounds 0 and $m.n^2.max(p2p, p2i)$ which is a (rough) upper bound of s^* , and has a value linear in the size of X since $max(p2p, p2i)$ is a constant. Hence, s^* can be computed in deterministic polynomial time using logarithmically many calls to an NP oracle (used to solve the SCORE instances encountered during the search, associated to the successive values of k). Once this is done, it remains to guess a binary relation $\langle P, I \rangle$ on A using a last call to the NP oracle, check that it is an interval order such that $score_X(\langle P, I \rangle) = s^*$, and finally check that $a\bar{R}b$ (resp. that there exists a $b \in A$ such that bPa). We conjecture that NEC-COMP(R) and NEC-UNDOM are Θ_2^P -complete. Noticeably, when s^* is part of the input, the complexity of NEC-COMP(R) and NEC-UNDOM falls down to coNP. From the practical side, when several instances of NEC-COMP(R) or NEC-UNDOM sharing the same profile X are to be solved, it can prove useful to compute s^* once for all during a pre-processing phase, then to exploit it in order to solve those instances in a more efficient way.

4 Some Theoretical Results

We analyze here some expected properties for aggregation procedures such as respect of unanimity, independance, majority, etc., and our objective is to determine whether or

not our approach satisfies some of them. We begin by the properties at work in Arrow's theorem:

Universality An aggregation procedure is universal if it accepts all configurations for the input profile. Since the input of our procedure can be any finite set of interval orders, we can conclude that our procedure is universal.

Transitivity Arrow's theorem imposes the transitivity of the preference and the indifference relation. Our procedure provides an interval order which has a transitive preference relation P but the indifference relation I is not necessarily transitive. However as we mentioned in the introduction, in order to find an undominated alternative, transitivity of P is enough.

Weak-unanimity Our procedure satisfies the weak unanimity condition since unanimity is imposed by our formulation as a hard constraint to be respected.

Non-dictatorship Our procedure obviously satisfies the non-dictatorship condition.

Independence Our procedure does not satisfy the condition of independence of irrelevant alternatives: let us show it on a new example. The set of alternatives is $A = \{a, b, c, d\}$ and we have two different profiles X and X' which have the same votes on the subset $A' = \{c, d\}$ of A :

Example 3. Table 3 shows the compact matrix of each profile. Our procedure con-

X	a	b	c	d
a	0	0	0	1
b	0	0	6	2
c	4	0	0	6
d	8	6	4	0

X'	a	b	c	d
a	0	3	0	0
b	3	0	0	3
c	8	1	0	6
d	2	2	4	0

Table 4. The compact matrix of profile X and X' 4

cludes that for the profile X there are two optimum solutions, in the first one c is indifferent to d and in the second one d is preferred to c . However, even if the profile X' has the same votes for the comparison between c and d , our procedure concludes for X' that c is preferred to d .

We consider now some other properties that an aggregation procedure should preferably satisfy.

Anonymity The result of the aggregation depends only on the preferences of voters (and for instance, not on the age, sex or seniority of candidates). Let \mathcal{P} be the set of permutations of A , π one element of \mathcal{P} . We denote by $\pi(R)$, the binary relation such as $\pi(a)\pi(R)\pi(b) \iff aRb$. An aggregation procedure is anonymous if and only if $\forall \pi \in \mathcal{P}, f(R_1, R_2, \dots, R_m) = \pi(f(\pi(R_1), \pi(R_2), \dots, \pi(R_m)))$.

It is easy to see that our procedure is anonymous.

Loyalty If there is just one voter the procedure must provide as a result the same preference as her: $m = 1 \implies f(R_1) = \{a \in A : aR_1b, \forall b \in A\}$. Again, it is easy to check that our procedure satisfies the loyalty condition

Majority condition If there is a majority of voters who prefers a to b then the result of the aggregation procedure must agree with this comparison: f satisfies the *majority condition* if and only if $\forall(R_1, R_2, \dots, R_m) \in X, \forall a, b \in A$

$$\#p(a, b) > \#p(b, a) \implies aPb,$$

$$\#p(a, b) = \#p(b, a) \implies aIb.$$

Our aggregation procedure does not satisfy the majority condition as the following example shows it.

Example 4. Table 5 gives the number of votes for pairwise comparisons between four alternatives given by 11 voters

P	a	b	c	d
a	0	6	5	1
b	2	0	8	3
c	6	2	0	1
d	4	3	3	0

Table 5. The number of voters agreeing for a strict preference

Even if the majority of voters prefer c to a , the result of our procedure concludes that a is preferred to c (the output is the interval order such that $aPb, aPc, aId, aIa, bPc, bId, bIb, cId, cIc, dId$ and its distance to the profile is 39).

5 Conclusion

In this paper, we have presented an optimization-based approach to interval orders aggregation. In this approach, to every profile of interval orders, one associates an instance of a propositional optimization problem (namely the Binate Covering Problem); solving the latter gives in a straightforward way an interval order (the "closest" to the input profile in some sense), which is considered as the aggregation looked for. Among other things, we have computed an upper bound of the size of the BCP instance associated to every profile (showing that it is only logarithmic in the number of voters), identified some properties satisfied (or not) by the aggregation approach. An interesting feature of such an optimization-based approach to aggregation is that it can be easily tuned to fit with other preference structures (e.g. preorders, semiorders, etc.). Indeed, it is enough to point out the corresponding hard constraints. Investigating in more depth such extensions is a perspective for further research.

The Binate Covering Problem can be seen as a very specific case of an Integer Linear Program in which case efficient ILP frameworks exist (e.g. CPLEX). However, it looks that tools dedicated to Boolean reasoning are better suited to solve such problems: Weighted Partial MAX SAT [1] and Pseudo Boolean Optimization [11] engines are currently receiving a lot of attention since international evaluations are organized regularly and many systems are freely available for the research community.

We designed a proof of concept tool based on the SAT4J library[9], a library of Boolean search engines dedicated to solving SAT, MAX SAT and Pseudo Boolean problems. That tool can be downloaded from <http://sat4j.ow2.org/>.

In order to have an idea of the applicability of our approach on a real scenario, we used the publicly available results of the SAT RACE 2006³. It is a competitive event between 16 SAT solvers on a set of 100 benchmarks. Here each benchmark is a voter and each solver is an alternative. A given benchmark b prefers the SAT solver x to the SAT solver y iff x solved b faster than y . A given benchmark b is indifferent between the SAT solvers x and y iff none of x and y solved b or both of them solved b but with a roughly the same CPU time (the difference is less than 1 second). By definition, each vote is an interval order. Computing the aggregation of such votes means solving a binate covering problem with 376 variables and 47536 clauses. Our aggregator takes less than one second to generate the BCP from the compact matrix. SAT4J takes one second to find a solution, but fails to prove it is optimal even after running for several hours.

We plan to test several MAXSAT and Pseudo Boolean engines on aggregation of real interval orders instances (including LP based ones).

References

1. Josep Argelich, Chu-Min Li, Felip Many, and Jordi Planes. The first and second max-sat evaluations. *Journal on Satisfiability, Boolean Modeling and Computation (JSAT)*, 4:251–278, 2008.
2. K.J. Arrow. *Social choice and individual values*. J. Wiley, New York, 1951. 2nd edition, 1963.
3. 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.
4. Olivier Coudert. On solving covering problems. In *Design Automation Conference*, pages 197–202, 1996.
5. P.C. Fishburn. *Interval Orders and Interval Graphs*. J. Wiley, New York, 1985.
6. A. Gibbard. Social choice and the arrow conditions. *unpublished*, 1969.
7. O. Hudry. Np-hardness results for the aggregation of linear orders into median orders. *Annals of Operations Research*, 163:63–88, 2008.
8. J.G. Kemeny. Mathematics without numbers. *Daedalus*, 88:575–591, 1959.
9. Daniel Le Berre and Anne Parrain. SAT4J, a SATisfiability library for java. <http://www.sat4j.org>.
10. R.D. Luce. Semi-orders and a theory of utility discrimination. *Econometrica*, 24, 1956.
11. V. Manquinho and O. Roussel. The first evaluation of pseudo-boolean solvers (pb'05). *Journal on Satisfiability, Boolean Modeling and Computation (JSAT)*, 2:103–143, 2006.
12. CH. Papadimitriou and I. Steiglitz. *Combinatorial Optimization: algorithms and complexity*. Prentice-Hall, Englewood Cliffs, 1982.
13. M. Pirlot and Ph. Vincke. *Semi Orders*. Kluwer Academic, Dordrecht, 1997.
14. A.K. Sen. *Collective Choice and Social Welfare*. North Holland, Amsterdam, 1970.
15. R. Stanley. Hyperplanes arrangements, interval orders and trees. *Proc. Nat. Acad. Sci.*, 93:2620–2625, 1996.
16. N. Wiener. A contribution to the theory of relative position. *Proc. of Cambridge Philosophical Society*, 17:441–449, 1914.

³ <http://fmv.jku.at/sat-race-2006/>