Distance-based Merging: A General Framework and some Complexity Results

Sébastien Konieczny
IRIT
Université Paul Sabatier
31602 Toulouse Cedex - France
`konieczny@irit.fr`

Jérôme Lang
IRIT
Université Paul Sabatier
31602 Toulouse Cedex - France
`lang@irit.fr`

Pierre Marquis
CRIL / Université d'Artois
rue de l'Université - S.P. 16
62307 Lens - France
`marquis@cril.univ-artois.fr`

Abstract

We present in this paper a new framework for propositional merging. *Distance-based merging operators*, parameterized by a distance between interpretations and two aggregation functions, are introduced. Many distances and aggregation functions can be used and many merging operators already defined in the literature (including both model-based ones and syntax-based ones) can be recovered as specific distance-based operators. Both logical and complexity properties of distance-based merging operators are studied. An important result is that (under very weak assumptions) query entailment from merged bases is "only" at the first level of the polynomial hierarchy when any of our distance-based operators is used. As a by-product, complexity results for several existing merging operators are derived as well.

1 INTRODUCTION

Belief merging is an important issue of many AI fields (see [Bloch and Hunter, 2001] for a panorama of applications of data and knowledge fusion).

Although particular requirements can be asked for each application, several pieces of information are usually brought into play when propositional base merging is concerned. In the following:

- A *knowledge set* $E = \{K_1, \ldots, K_n\}$ is a finite multi-set of knowledge bases, where each *knowledge base* K_i represents the set of beliefs from source i. Each K_i is a propositional formula, or more generally, a finite set of propositional formulas $\varphi_{i,j}$ encoding the explicit beliefs from source i.
- Some *integrity constraints* IC encoded as a propositional formula. IC represents some common knowledge on which all sources agree (e.g. some physical constraints, norms, etc.).

The purpose of merging E is to characterize a formula (or a set of formulas) $\triangle_{IC}(E)$, considered as the overall knowledge from the m sources given the integrity constraints IC. Recently, several families of such merging operators have been defined and characterized in a logical way [Revesz, 1997; Lin and Mendelzon, 1999; Liberatore and Schaerf, 1998; Konieczny and Pino Pérez, 1999; Benferhat *et al.*, 2000]. Among them are the so-called *model-based* merging operators [Revesz, 1997; Lin and Mendelzon, 1999; Liberatore and Schaerf, 1998; Konieczny and Pino Pérez, 1999] where the models of $\triangle_{IC}(E)$ are defined as the models of IC which are preferred according to some criterion depending on E. Often, such preference information take the form of a total pre-order on interpretations, induced by a notion of distance $d(\omega, E)$ between an interpretation ω and the knowledge set E. $d(\omega, E)$ is typically defined by aggregating the distances $d(\omega, K_i)$ for every K_i. Usually, model-based merging operators takes only into account consistent knowledge bases K_i. Other merging operators are *syntax-based* ones [Baral *et al.*, 1991; Baral *et al.*, 1992; Konieczny, 2000]. They are based on the selection of some consistent subsets of $\bigcup_{i=1}^{m} K_i$. This renders possible to take into account inconsistent knowledge bases K_i and to incorporate some additional preference information into the merging process[1] but the price to be paid is to give some importance to the syntax of knowledge bases. Moreover, since they are based on the set-theoretic union $\bigcup_{i=1}^{m} K_i$ of the bases, such operators usually do not take into account the frequency

[1] Indeed, as in belief revision, giving some importance to the syntax of K_i is a way to specify (implicitly but in a cheap way w.r.t. representation) that explicit beliefs are preferred to implicit beliefs [Nebel, 1989; Hansson, 1998].

of each explicit pieces of belief into the merging process (the fact that $\varphi_{i,j}$ is believed in one source only or in the m sources under consideration is not considered relevant, which is often counter-intuitive[2]).

In this paper, a new framework for defining propositional merging operators is provided. A family of merging operators parametrized by a distance d between interpretations and two aggregation functions f and g is presented. These parameters are used to define a notion of distance between an interpretation and a knowledge set E in a two-step fashion. Like in existing model-based approaches to merging, the models of the merging of E given some integrity constraints IC are exactly the models of IC that are as close as possible to E with respect to the distance. Moreover, the first aggregation step enables to take into account the syntax of knowledge bases within the merging process. This allows to handle inconsistent ones in a satisfying way.

The contribution of this work is many fold. First, our framework is general enough to encompass almost all model-based merging operators as specific cases. In addition, despite the model-theoretic ground of our approach, several syntax-based merging operators provided so far in the literature can be captured as well. We show that, by imposing few conditions on the parameters, several logical properties that are expected when merging operators are considered, are already satisfied.

Another very strong feature offered by our framework is that query entailment from $\triangle_{IC}(E)$ is guaranteed to lay at the first level of the polynomial hierarchy provided that d, f and g can be computed in polynomial time. Accordingly, improving the generality of the model-based merging operators framework through an additional aggregation step does not result in a complexity shift.

We specifically focus on some simple families of distances and aggregation functions. By letting the parameters d, f and g vary in these respective sets, several merging operators are obtained; some of them were already known and are thus recovered as specific cases in our framework, and others are new operators. In any case, we investigate the logical properties and identify the complexity of each operator under consideration. As a by-product, the complexity of several model-based merging operators already pointed out so far is also identified.

The full proofs of the results given in this article can be found in [Konieczny *et al.*, 2001].

[2]See [Konieczny, 2000] for one step in that direction.

2 FORMAL PRELIMINARIES

We consider a propositional language $PROP_{PS}$ built up from a finite set PS of propositional symbols in the usual way. An interpretation is a total function from PS to $BOOL = \{0, 1\}$. The set of all interpretations is denoted $\mathcal{W}$. An interpretation ω is a model of a formula iff it makes it true in the usual classical truth functional way. Provided that φ is a formula from $PROP_{PS}$, $Mod(\varphi)$ denotes the set of models of φ, i.e., $Mod(\varphi) = \{\omega \in \mathcal{W} \mid \omega \models \varphi\}$.

A knowledge base K_i is said to be consistent iff the conjunction $\bigwedge K_i$ of its formulas is consistent. Similarly, a knowledge set E is said to be consistent iff the conjunction $\bigwedge E$ of its knowledge bases is consistent. Two knowledge bases K_1 and K_2 are said to be logically equivalent ($K_1 \equiv K_2$) iff $\bigwedge K_1 \equiv \bigwedge K_2$, and two knowledge sets E_1 and E_2 are said to be equivalent ($E_1 \equiv E_2$) iff there is a bijection between E_1 and E_2 such that each knowledge base of E_1 is logically equivalent to its image in E_2. $\sqcup$ denotes the multi-set union. For every knowledge set E and for every integer n, E^n denotes the multi-set obtained by "unioning" E with itself n times.

The complexity results we give in this paper refer to some complexity classes which we now briefly recall (see [Papadimitriou, 1994] for more details), especially the classes Δ_2^p and Θ_2^p [Eiter and Gottlob, 1992; Wagner, 1987] from the polynomial hierarchy PH, as well as the class BH_2 from the Boolean hierarchy (see [Papadimitriou, 1994]). Given a problem A, we denote by $\overline{\text{A}}$ its complement. We assume the reader familiar with the classes P, NP et coNP and we now introduce the following three classes located at the first level of the polynomial hierarchy:

- BH_2 (also known as DP) is the class of all languages L such that $L = L_1 \cap L_2$, where L_1 is in NP and L_2 in coNP. The canonical BH_2-complete problem is SAT–UNSAT: given two propositional formulas φ and ψ, $\langle\varphi, \psi\rangle$ is in SAT–UNSAT if and only if φ is consistent and ψ is inconsistent.
- $\Delta_2^p = \mathsf{P}^{\mathsf{NP}}$ is the class of all languages that can be recognized in polynomial time by a Turing machine equipped with an NP oracle, where an NP oracle solves whatever instance of a problem NP in unit time.
- $\Theta_2^p = \Delta_2^p[\mathcal{O}(log\ n)])$ is the class of all languages that can be recognized in polynomial time by a Turing machine using a number of NP oracles bounded by a logarithmic function of the size of the input data.

Note that the following inclusions hold:

$$\mathsf{NP} \cup \mathsf{coNP} \subseteq \mathsf{BH}_2 \subseteq \Theta_2^p \subseteq \Delta_2^p \subseteq \mathsf{PH}.$$

3 DISTANCE-BASED MERGING

3.1 THE GENERAL FRAMEWORK

Defining a merging operator in our framework simply consists in setting three parameters: a distance d and two aggregation functions f and g. Let us first make precise what such notions mean in this paper:

Definition 1 (distances) *Let d be a total function from $\mathcal{W} \times \mathcal{W}$ to $I\!N$ s.t. (1) for every $\omega_1,\ \omega_2 \in \mathcal{W}$, $d(\omega_1, \omega_2) = d(\omega_2, \omega_1)$ and (2) $d(\omega_1, \omega_2) = 0$ iff $\omega_1 = \omega_2$. Such a d is called a* distance *between interpretations*[3]. *d induces a distance between any interpretation ω and any formula φ given by $d(\omega, \varphi) = \min_{\omega' \models \varphi} d(\omega, \omega')$.*

Definition 2 (aggregation functions) *Let f be a total function associating a nonnegative integer to every finite tuple of nonnegative integers and s.t.*

- *f is non-decreasing in each argument*[4], *and*
- *f satisfies (minimality) : for every n-uple $(x_1, \ldots, x_n)$ of nonnegative integers, $f(x_1, \ldots, x_n) = 0$ iff $x_1 = \ldots = x_n = 0$, and*
- *for every nonnegative integer x_1, $f(x_1) = x_1$.*

f is called an aggregation function[5].

We are now in position to define our distance-based merging operators:

Definition 3 (distance-based merging operators) *Let d be a distance between interpretations and f and g be two aggregation functions. For every knowledge set $E = \{K_1, \ldots, K_n\}$ and every integrity constraint IC, $\triangle_{IC}^{d,f,g}(E)$ is defined in a model-theoretical way by:*

$$Mod(\triangle_{IC}^{d,f,g}(E)) = \{\omega \in Mod(IC) \mid d(\omega, E) \textit{ is minimal}\}$$

where

$$d(\omega, E) = g(d(\omega, K_1), \ldots, d(\omega, K_n))$$

and for every $K_i = \{\varphi_{i,1}, \ldots, \varphi_{i,n_i}\}$

$$d(\omega, K_i) = f(d(\omega, \varphi_{i,1}), \ldots, d(\omega, \varphi_{i,n_i})).$$

Formulas appearing in a knowledge base K_i can have various possible meanings, for instance:

- *pieces of information provided by the source i*: when merging several beliefs stemming from different sources (sensors or experts, for example);
- *pieces of information pertaining to a criterion i*: when evaluating alternatives with respect to different criteria;
- *elementary goals expressed by the agent i*: when aggregating individual preferences in a group decision making context – see [Lafage and Lang, 2000]. In this case, the formulas $\varphi_{i,j}$ are no longer *beliefs* but *preferences* (which does not prevent us from using the same merging operators).

The reason why we use *two* (generally distinct) aggregation functions f and g is that both aggregation steps are of different nature. The first step is an *intra-source* aggregation: f aggregates scores w.r.t. the elementary (explicit) pieces of information contained in each K_i (it allows, in particular, to take inconsistent knowledge bases into account). The second step is an *inter-source* aggregation: g aggregates the "f-aggregated scores" pertaining to the different sources.

Interestingly, few conditions are imposed on d, f, and g. As we will see in the next section, many distances and aggregation functions can be used. Often, the aggregation functions f and g are required to be symmetric (i.e., no priority is given to some explicit beliefs in a knowledge base, and no priority is given to some knowledge bases in a knowledge set). However, this condition is not mandatory here and this is important when some preference information is available, especially when all sources i are not equally reliable. For instance, the *weighted sum* aggregation function can be used to give rise to (non-symmetric) merging operators.

Let us stress that, contrarily to usual model-based operators, our definition allows inconsistent knowledge bases to take (a non-trivial) part in the merging process.

Example 1 *Assume for example that we want to merge $E = \{K_1, K_2, K_3, K_4\}$ under the integrity constraints $IC = \top$, where*

[3] We slightly abuse words here, since d is only a pseudo-distance (triangular inequality is not required).

[4] I.e., if $x \leq y$, then $f(x_1, \ldots, x, \ldots, x_n) \leq f(x_1, \ldots, y, \ldots, x_n)$.

[5] The aggregation function f can take arbitrarily many arguments; more formally, a "function" f is a family $f = \{f_n \mid n \in I\!N\}$ of n-ary functions from $I\!N^n$ to $I\!N$. Slightly abusing notations, we write $f(x_1, \ldots, x_n)$ instead of $f_n(x_1, \ldots, x_n)$ since this can never be ambiguous.

- $K_1 = \{a, b, c, a \rightarrow \neg b\}$,
- $K_2 = \{a, b\}$,
- $K_3 = \{\neg a, \neg b\}$,
- $K_4 = \{a, a \rightarrow b\}$.

In this example, K_1 knows that c holds; since this piece of information is not involved in any contradiction, it can prove sensible to be confident in K_1 about the truth of c. Model-based merging operators can not handle this situation: inconsistent knowledge bases can not be taken into account. Thus, provided that the Hamming distance between interpretations is considered, the operator $\triangle^{\Sigma}$ [Revesz, 1997; Lin and Mendelzon, 1999; Konieczny and Pino Pérez, 1999] gives a merged base whose models are: $\{a, b, \neg c\}$ and $\{a, b, c\}$; the operator $\triangle^{Gmax}$ [Konieczny and Pino Pérez, 1999] gives a merged base whose models are: $\{\neg a, b, \neg c\}$, $\{\neg a, b, c\}$, $\{a, \neg b, \neg c\}$, and $\{a, \neg b, c\}$. In any of these two cases, nothing can be said about the truth of c in the merged base, which is counter-intuitive since no argument against it can be found in the input.

Syntax-based operators render possible the exploitation of inconsistent knowledge bases, but they do not care about the distribution of information. Consider the two standard syntax-based operators [Baral *et al.*, 1992], selecting the maximal subsets of $\bigcup_{i=1}^{m} K_i$ (one w.r.t. set inclusion and the other one w.r.t. cardinality). On the previous example, the first one returns a merged base equivalent to c and the second one to $c \wedge \neg a$. So, a is in the result for none of these two operators, whereas a holds in three over four input bases.

Our distance-based operators achieve a compromise between model-based operators and syntax-based operators, by taking into account the way information is distributed and by taking advantage of the information stemming from inconsistent knowledge bases. For instance, our operator $\triangle^{d_D,sum,sum}$ (cf. Section 3.2) gives a merged base whose single model is $\{a, b, c\}$, and $\triangle^{d_D,sum,lex}$ returns a merged base whose models are $\{\neg a, b, c\}$ and $\{a, \neg b, c\}$. So, with any of these two operators, we can deduce that c holds after the merging. Moreover, these operators exhibit typical merging behaviours. The first one is a majority operator: since three of four bases agree on a, a holds in the result. The second one is an arbitration operator; being more consensual, it gives that only one of a or b holds, to be as close as possible to each of the knowledge bases.

3.2 INSTANCIATING OUR FRAMEWORK

Let us now instantiate our framework and focus on some simple families of distances and aggregation functions.

Definition 4 (some distances) *Let ω_1, $\omega_2 \in \mathcal{W}$ be two interpretations.*

- *The* drastic distance d_D *is defined by*
$$d_D(\omega_1, \omega_2) = 0 \textit{ if } \omega_1 = \omega_2, \quad 1 \textit{ otherwise}$$
- *The* Hamming distance d_H *is defined by*
$$d_H(\omega_1, \omega_2) = |\{x \in PS \mid \omega_1(x) \neq \omega_2(x)\}|$$
- *Let q be a total function from PS to $\mathbb{N}^*$. The* weighted Hamming distance d_{H_q} *induced by q is defined by*
$$d_{H_q}(\omega_1, \omega_2) = \sum_{x \in PS \mid \omega_1(x) \neq \omega_2(x)} q(x)$$

These distances satisfy the requirements imposed in Definition 3.

The Hamming distance is the most usual distance considered in model-based merging[6]. It is very simple to express, but one has to keep in mind that it is very sensitive to the representation language of the problem (i.e., the choice of propositional symbols) and that numerous others distances can be used. Weighted Hamming distances are relevant when some propositional symbols are known as more important than others.

Definition 5 (some aggregation functions)

- *Let q be a total function from $\{1, \ldots, n\}$ to $\mathbb{N}^*$ s.t. $q(1) = 1$ whenever $n = 1$. The weighted sum WS_q induced by q is defined by $WS_q(e_1, \ldots, e_n) = \sum_{i=1}^{n} q(i)e_i$.*
- *Let q be a total function from $\{1, \ldots, n\}$ to $\mathbb{N}$ s.t. $q(1) = 1$ whenever $n = 1$, and $q(1) \neq 0$ in any case. The ordered weighted sum OWS_q induced by q is defined by $OWS_q(e_1, \ldots, e_n) = \sum_{i=1}^{n} q(i)e_{\sigma(i)}$ where σ is a permutation of $\{1 \ldots n\}$ s.t. $e_{\sigma(1)} \geq e_{\sigma(2)} \geq \cdots \geq e_{\sigma(n)}$.*

q is a *weight function*, that gives to each formula (resp. knowledge base) φ_i (resp. K_i) of index i its weight $q(i)$ denoting the formula (resp. knowledge base) reliability. With the slight difference that q is normalized (but without requiring that $q(1) = 1$ whenever $n = 1$), the latter family is well-known in multi-criteria decision

[6] In this context, it is also called *Dalal distance* [Dalal, 1988].

$$
\begin{array}{rcl}
\triangle^{d_D,max,max} & = & \top \\
\triangle^{d_D,max,sum}, \triangle^{d_D,max,lex}, \triangle^{d_H,max,sum} & = & a \wedge b \\
\triangle^{d_D,sum,max} & = & \neg b \\
\triangle^{d_D,sum,sum} & = & (\neg a \wedge \neg b) \vee (a \wedge b \wedge c) \\
\triangle^{d_D,sum,lex} & = & \neg a \wedge \neg b \\
\triangle^{d_H,sum,max}, \triangle^{d_H,sum,lex} & = & a \wedge \neg b \wedge c \\
\triangle^{d_H,max,max}, \triangle^{d_H,max,lex} & = & (\neg a \wedge b \wedge c) \vee (a \wedge \neg b \wedge c) \\
\triangle^{d_H,sum,sum} & = & a \wedge c
\end{array}
$$

Figure 1: Example 2

making under the terminology "Ordered Weighted Averages" (OWAs) [Yager, 1998]. When $q(i) = 1$ for every $i \in 1 \dots n$, WS_q is the usual sum (and OWS_q as well). When $q(1) = 1$ and $q(2) = \dots = q(n) = 0$ then $OWS_q(e_1, \dots, e_n) = max(e_1, \dots, e_n)$.

For the second aggregation step g, it is relevant to consider the well-known *leximax* ordering which compares two vectors of scores by focusing on the largest scores of each vector, and in case of equality, on the second largest scores, and so on. For the sake of homogeneity, we reformulate the leximax ordering so as to compare aggregated scores rather than vectors of scores. This can be done thanks to a specific aggregation function OWS_q:

Definition 5.1 (leximax)
Let M be an upper bound of the scores $d(\omega, K_i)$[7], i.e., for any ω we have $d(\omega, K_i) < M$. Now, let $q(i) = M^{n-i}$ for all i. The rank order on vectors of scores induced by OWS_q is the leximax *ordering, abbreviated by lex[8].*

Using the *leximax* aggregation for the first aggregation step (f) would also be possible, but leads to rather lengthy technical tricks to be defined properly in case where the second aggregation function g is not purely ordinal (i.e., g different from *max* and *leximax*) and we ignore this possibility here (see the long version of the paper [Konieczny *et al.*, 2001]).

All these functions satisfy the requirements imposed in Definition 3; all of them are symmetric but *weighted sum* (except when q is uniform).

Many other possible choices for f and g can be found in the literature of multi-criteria decision making (and to a smaller extent in the literature of group decision theory). Noticeably, the usual aggregation functions used in these fields are all polynomially computable, which makes the following complexity results applicable when instantiating f and g with these functions.

Note that functions such as the purely utilitarian *sum* or *weighted sum* allow for compensations between scores (and lead to majority-like operators), while the egalitarian functions *max* and *lex* do not.

By letting the parameters d, f and g vary in these respective sets, several merging operators are obtained; some of them were already known and are thus recovered as specific cases in our framework, and others are new operators. Thus, $\triangle^{d_D,max,max}$ is the *basic* merging operator, giving $\bigwedge E \wedge IC$ if consistent and IC otherwise. $\triangle^{d_D,max,sum}$ is the *drastic* merging operator which amounts to select the models of IC satisfying the greatest number of knowledge bases from E. It is equivalent to the drastic majority operator as defined in [Konieczny, 2000] when working with deductively closed knowledge bases. $\triangle^{d_D,sum,sum}$ corresponds to the intersection operator of [Konieczny, 2000]. $\triangle^{d_D,WS_q,max}$ corresponds to an operator used in [Lafage and Lang, 2000] in a group decision context. When singleton knowledge bases are considered [9] – recall that in this case f is irrelevant – every $\triangle^{d_H,f,max}$ operator is a $\triangle^{Max}$ operator [Revesz, 1997], every $\triangle^{d_H,f,sum}$ operator is a $\triangle^{\Sigma}$ operator [Revesz, 1997; Lin and Mendelzon, 1999; Konieczny and Pino Pérez, 1999], and every $\triangle^{d_H,f,lex}$ operator is a $\triangle^{GMax}$ operator [Konieczny and Pino Pérez, 1999]. Still with singleton knowledge bases, taking $d = d_D$ and $f = WS_q$, $\triangle^{d_H,f,WS_q}$ is a penalty-based merging (where one minimizes the sum of the penalties $q(i)$ attached to the K_i's) [Pinkas, 1995], and taking $d = d_D$ and $f = WMAX_q$ (defined by $WMAX_q(x_1, \dots, x_n) = \max_{i=1 \dots n} \min(q(i), x_i)$) we get[10] a possibilistic merging operator [Benferhat *et al.*, 2000].

[7] For instance, when $d = d_H$ and $f = \max$ we can choose $M = |PS| + 1$; when $d = d_H$ and $f = \sum$ we can choose $M = |PS|^2 + 1$.

[8] Namely, we have $OWS_q(e_1, \dots, e_n) \geq OWS_q(e'_1, \dots, e'_n)$ iff $(e_{\sigma(1)} > e'_{\sigma'(1)})$ or $(e_{\sigma(1)} = e'_{\sigma'(1)}$ and $e_{\sigma(2)} > e'_{\sigma'(2)})$ or etc.

[9] Or, equivalently, when each K_i is replaced by $\{\bigwedge K_i\}$ before merging.

[10] The scales used for scores are different but it is obvious to show that this difference has no impact.

Table 1: $\triangle^{d_H,sum,lex}$ Operator

	$a \wedge b \wedge c$	$a \rightarrow \neg b$	$a \wedge b$	$\neg a \wedge \neg b$	$\neg b$	a	$a \rightarrow b$	$\mathbf{K_1}$	$\mathbf{K_2}$	$\mathbf{K_3}$	$\mathbf{K_4}$	**E**
$(0,0,0)$	3	0	2	0	0	1	0	3	2	0	1	3210
$(0,0,1)$	2	0	2	0	0	1	0	2	2	0	1	2210
$(0,1,0)$	2	0	1	1	1	1	0	2	1	2	1	2211
$(0,1,1)$	1	0	1	1	1	1	0	1	1	2	1	2111
$(1,0,0)$	2	0	1	1	0	0	1	2	1	1	1	2111
$(1,0,1)$	1	0	1	1	0	0	1	1	1	1	1	1111
$(1,1,0)$	1	1	0	2	1	0	0	2	0	3	0	3200
$(1,1,1)$	0	1	0	2	1	0	0	1	0	3	0	3100

We will now illustrate the behaviour of these different operators on an example.

Example 2 *Consider the following knowledge set* $E = \{K_1, K_2, K_3, K_4\}$ *that we want to merge under the integrity constraints* $IC = \top$.

- $K_1 = \{a \wedge b \wedge c, a \rightarrow \neg b\}$,
- $K_2 = \{a \wedge b\}$,
- $K_3 = \{\neg a \wedge \neg b, \neg b\}$,
- $K_4 = \{a, a \rightarrow b\}$.

The result of the merging of E *according to the different operators with* $d \in \{d_D, d_H\}$, $f \in \{max, sum\}$ *and* $g \in \{max, sum, lex\}$ *under no constraints (i.e.* $IC = \top$*) is indicated figure 1. See table 1 for an example of calculation with the* $\triangle^{d_H,sum,lex}$ *operator. In this table the interpretation* $(1,0,0)$ *for example is the one mapping* a *to true and* b *and* c *to* $false$. *The result of the merging* $\triangle_{\top}^{d_H,sum,lex}(E)$ *is the interpretation that is the closest to* E, *that is the one at a distance* 1111, *i.e. the one mapping* a *and* c *to true and* b *to false.*

The wide variety of obtained results show the degree of freedom given by this framework. This example illustrates several aspects of merging operators : the knowledge base K_1 is not consistent, but it is the only base that gives an information about c, so it can be sensible to take c as true in the result of the merging. K_3 is logically equivalent to $\neg a \wedge \neg b$, but replacing K_3 by this formula would lead to different results for merging. Syntax is relevant for distance-based merging operators since one has to consider that different formulae of a same base are distinct reasons to believe in a same information. Taking syntax into account is important from the point of view of representation of beliefs (or goals), but the operators can then "choose" to take or not this information into account. So the point in this framework is that, unlike classical model-based merging operators, the connector "," is not the same that the connector "$\wedge$".

4 COMPUTATIONAL COMPLEXITY

Let us now turn to the complexity issue. We obtained the following result:

Proposition 1 *Let* $\triangle^{d,f,g}$ *be a distance-based merging operator. Given a knowledge set* E *and two formulas* IC *and* α:

- *If* d, f *and* g *are computable in polynomial time, then determining whether* $\triangle_{IC}^{d,f,g}(E) \models \alpha$ *holds is in* Δ_2^p.
- *If* d, f *and* g *are computable in polynomial time and are polynomially bounded, then determining whether* $\triangle_{IC}^{d,f,g}(E) \models \alpha$ *holds is in* Θ_2^p.

A sketch of proof is given in the Appendix. See [Konieczny *et al.*, 2001] for a detailed proof.

As shown by the previous proposition, improving the generality of the model-based merging operators framework through an additional aggregation step does not result in a complexity shift (the decision problem for query entailment is still at the first level of PH).

We have also identified the complexity of query entailment from a merged base for the following distance-based merging operators:

Proposition 2 *Given a knowledge set* E *and two formulas from* $PROP_{PS}$ IC *and* α, *the complexity of* $\triangle_{IC}^{d,f,g}(E) \models^? \alpha$ *is reported in Tables 2, 3 and 4 (when* X *is a complexity class,* X*-c means* X*-complete).*

Table 2: Complexity results ($d = d_D$)

f/g	max	sum	lex	WS_q	OWS_q
max	BH_2-c	Θ_2^p-c	Θ_2^p-c	Δ_2^p-c	Θ_2^p-c
sum	Θ_2^p-c	Θ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
WS_q	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
OWS_q	Θ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c

Table 3: Complexity results ($d = d_H$)

f/g	max	sum	lex	WS_q	OWS_q
max	Θ_2^p-c	Θ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
sum	Θ_2^p-c	Θ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
WS_q	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
OWS_q	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c

Sketches of the proofs are given in the Appendix (again, see [Konieczny *et al.*, 2001] for fully detailed proof). It is worth adding that in the case $d = d_{H_q}$, Δ_2^p-hardness still holds whenever E is a singleton $\{K\}$, K is a singleton $\{\varphi\}$ and φ is a conjunction of variables (in this case, neither f nor g plays a significant role in the elaboration of the distance to E). As to the case $d = d_H$, Δ_2^p-hardness still holds when each explicit belief is a conjunction of variables, and Θ_2^p-hardness results hold whenever E is a singleton $\{K\}$, K is a singleton $\{\varphi\}$ and φ is a conjunction of variables.

Looking at the tables above, we can observe that the choice of the distance d has a great influence on the complexity results. Thus, whenever $d = d_H$ or $d = d_{H_q}$, the complexity results for inference from a merged base coincide whenever f (or g) is a WS_q function or a OWS_q function. This is no longer the case when $d = d_D$ is considered.

Together with Proposition 1, the complexity of many model-based merging operators already pointed out in the literature are derived as a by-product of the previous complexity results. To the best of our knowledge, the complexity of such operators has not been identified up to now[11], hence this is an additional contribution of this work. We can also note that, while the complexity of our distance-based operators is not very high (first level of PH, at most), finding out significant tractable restrictions seems a hard task since intractability is still the case in many restricted situations. Finally, our results show that some syntax-based merging operators (based on set inclusion instead of cardinality and "located" at the 2^{nd} level of PH) cannot be encoded in polynomial time as distance-based operators (unless PH collapses).

[11] However, $(\triangle_{IC}^{d_H,sum,sum}(E) \models^? \alpha) \in \Delta_2^p$ can be recovered from a complexity result given in [Liberatore and Schaerf, 2000], page 151.

Table 4: Complexity results ($d = d_{H_q}$)

f/g	max	sum	lex	WS_q	OWS_q
max	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
sum	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
WS_q	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c
OWS_q	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c	Δ_2^p-c

5 LOGICAL PROPERTIES

Since we aim at investigating the logical properties of our family of merging operators, a set of properties must first be considered as a base line. In [Konieczny and Pino Pérez, 1999], a study of logical properties that "good" merging operators should satisfy (in the case where all the knowledge bases are equally reliable) is carried on. The following set of postulates was proposed:

Definition 6 (IC merging operators) *Let E, E_1, E_2 be knowledge sets, K_1, K_2 be consistent knowledge bases, and IC, IC_1, IC_2 be formulas from $PROP_{PS}$. $\triangle$ is an* IC merging operator *iff it satisfies the following postulates:*

(IC0) $\triangle_{IC}(E) \models IC$

(IC1) *If IC is consistent, then $\triangle_{IC}(E)$ is consistent*

(IC2) *If $\bigwedge E$ is consistent with IC, then* $\triangle_{IC}(E) \equiv \bigwedge E \wedge IC$

(IC3) *If $E_1 \equiv E_2$ and $IC_1 \equiv IC_2$, then* $\triangle_{IC_1}(E_1) \equiv \triangle_{IC_2}(E_2)$

(IC4) *If $K_1 \models IC$ and $K_2 \models IC$, then $\triangle_{IC}(K_1 \sqcup K_2) \wedge K_1$ is consistent iff $\triangle_{IC}(K_1 \sqcup K_2) \wedge K_2$ is consistent*

(IC5) $\triangle_{IC}(E_1) \wedge \triangle_{IC}(E_2) \models \triangle_{IC}(E_1 \sqcup E_2)$

(IC6) *If $\triangle_{IC}(E_1) \wedge \triangle_{IC}(E_2)$ is consistent , then* $\triangle_{IC}(E_1 \sqcup E_2) \models \triangle_{IC}(E_1) \wedge \triangle_{IC}(E_2)$

(IC7) $\triangle_{IC_1}(E) \wedge IC_2 \models \triangle_{IC_1 \wedge IC_2}(E)$

(IC8) *If $\triangle_{IC_1}(E) \wedge IC_2$ is consistent, then* $\triangle_{IC_1 \wedge IC_2}(E) \models \triangle_{IC_1}(E)$

Two sub-classes of IC merging operators have also been defined. Majority operators that aim at resolving conflicts by listening the majority wishes, and arbitration operators that have a more consensual behaviour:

Definition 7 (majority and arbitration) *A* majority operator *is an IC merging operator that satisfies the following majority postulate:*

$$\textbf{(Maj)} \quad \exists n \ \triangle_{IC}(E_1 \sqcup E_2{}^n) \models \triangle_{IC}(E_2)$$

An arbitration operator *is an IC merging operator that satisfies the following postulate:*

$$\textbf{(Arb)} \quad \left.\begin{array}{l} \triangle_{IC_1}(K_1) \equiv \triangle_{IC_2}(K_2) \\ \triangle_{IC_1 \Leftrightarrow \neg IC_2}(K_1 \sqcup K_2) \equiv (IC_1 \Leftrightarrow \neg IC_2) \\ IC_1 \not\models IC_2 \\ IC_2 \not\models IC_1 \end{array}\right\} \Rightarrow \quad \triangle_{IC_1 \vee IC_2}(K_1 \sqcup K_2) \equiv \triangle_{IC_1}(K_1)$$

See [Konieczny and Pino Pérez, 2002; Konieczny and Pino Pérez, 1999] for more explanations about those two postulates and the behaviour of the two subclasses.

We have the following result:

Proposition 3 $\triangle^{d,f,g}$ *satisfies (IC0), (IC1), (IC2), (IC7), (IC8). The other postulates are not satisfied in the general case.*

Clearly enough, it is not the case that every distance-based merging operator is an IC merging operator (not satisfying some postulates is deliberate since we want to give some importance to the syntax in order to take into account inconsistent knowledge bases). Let us introduce some properties to be satisfied by aggregation functions f:

1) $f(x_1, \ldots, x_n) = 0$ iff $x_1 = \ldots = x_n = 0$ **(minimality)**

2) If $\varphi_1 \wedge \ldots \wedge \varphi_n$ is consistent, then $f(d(w, \varphi_1), \ldots, d(w, \varphi_n)) = f(d(w, \varphi_1 \wedge \ldots \wedge \varphi_n))$ **(and)**

3) For any permutation σ, $f(x_1, \ldots, x_n) = f(\sigma(x_1, \ldots, x_n))$ **(symmetry)**

4) If $f(x_1, \ldots, x_n) \leq f(y_1, \ldots, y_n)$, then $f(x_1, \ldots, x_n, z) \leq f(y_1, \ldots, y_n, z)$ **(composition)**

5) If $f(x_1, \ldots, x_n, z) \leq f(y_1, \ldots, y_n, z)$, then $f(x_1, \ldots, x_n) \leq f(y_1, \ldots, y_n)$ **(decomposition)**

Now, if one wants to recover the full set of postulates (IC0)-(IC8):

Proposition 4 *A distance-based merging operator* $\triangle^{d,f,g}$ *satisfies (IC0)-(IC8) if and only if the function* f *satisfies (minimality) and (and), and the function* g *satisfies (minimality), (symmetry), (composition) and (decomposition).*

Concerning the operators examined in the previous section, we have identified the following properties:

Proposition 5 $\triangle^{d,f,g}$ *satisfies the logical properties stated in Tables 5 and 6. Since all these operators are already known to satisfy (IC0), (IC1), (IC2), (IC7) and (IC8) (cf. Proposition 3), we avoid repeating such postulates here. For more readability, postulate* (IC_i) *is noted* i *and* M *(resp.* A*) stands for (Maj) (resp. (Arb)).*

Table 5: Logical properties ($d = d_D$)

f/g	max	sum	lex	WS_q
max	3,4,5,A	3,4,5,6,M,A		5,6,M
sum	5,A	5,6,M	5,6,A	5,6,M
$WS_q - OWS_q$	5,A	5,6,M	5,6,A	5,6,M

Table 6: Logical properties ($d = d_H$ or $d = d_{H_q}$)

f/g	max	sum	lex	WS_q
max	5,A	5,6,M	5,6,A	5,6,M
sum	5,A	5,6,M	5,6,A	5,6,M
$WS_q - OWS_q$	5,A	5,6,M	5,6,A	5,6,M

The tables above show our operators to exhibit different properties. We remark that among our operators, only $\triangle^{d_D,max,sum}$ satisfies all listed properties. Failing to satisfy (IC3) (*irrelevance to the syntax*) in many cases is not surprising, since we want to allow our operators to take syntax into account. (IC4) imposes that, when merging two knowledge bases, if the result is consistent with one knowledge base, it has to be consistent with the other one – this fairness postulate is irrelevant when working with non-symmetric operators (so, unsurprisingly, it is not satisfied for $g = WS_q$). This postulate is not satisfied by any operator for which d is Hamming distance since cardinalities of the knowledge bases have an influence on f, and more generally, it is hardly satisfiable when working with syntax-dependent operators. (IC5) and (IC6) are related to Pareto dominance in social choice theory and are really important for multi-source aggregation; so it is worth noting that almost all operators satisfy them (only operators for which $g = max$ do not satisfy (IC6)).

We do not put the operators with $g = OWS_q$ in the tables because they gather many aggregation functions and so they do not satisfy a lot of logical properties. Moreover, some properties (as (IC5) and (IC6)) require to be able to cope with knowledge sets of different sizes, whereas $g = OWS_q$ operators have to specify exactly the size of the knowledge sets. It is possible to generalize the definition of those operators to cope with these cases but it is out of the scope of this paper.

6 CONCLUSION

The major contribution of this paper is a new framework for propositional merging. It is general enough to encompass many existing operators (both model-based ones and syntax-based ones) and to enable the definition of many new operators (symmetric or not). Both the logical properties and the computational properties of the merging operators pertaining to our framework have been investigated. Some of our results are large-scope ones in the sense that they make sense under very weak conditions on the three parameters that must be set to define an operator in our framework. By instantiating our framework and considering several distances and aggregation functions, more refined results have also been obtained.

This work calls for several perspectives. One of them consists in analyzing the properties of the distance-based operators that are achieved when some other aggregation functions or some other distances are considered. For instance, suppose that a collection of formulas of interest (topics) is available. In this situation, the distance between ω_1 and ω_2 can be defined as the number of relevant formulas on which ω_1 and ω_2 differs (i.e., such that one of them satisfies the formula and the other one violates it). Several additional distances could also be defined and investigated (see e.g. [Lafage and Lang, 2001] for distances based on Choquet integral).

Acknowledgements

The third author has been partly supported by the IUT de Lens, the Université d'Artois, the Région Nord/Pas-de-Calais under the TACT-TIC project, and by the European Community FEDER Program.

Appendix: Proof sketches of the complexity results

Sketch of Proof of Proposition 1 : These results are consequences of the two following lemmata:

Lemma 1 *Let k be an integer; if d, f and g are computable in polynomial time, then the problem of determining whether* $\min_{\omega \models IC} d(\omega, E) \leq k$ *given IC, E and k is in* NP.

Proof : It is sufficient to consider the following non-deterministic algorithm:

1. guess an interpretation ω and N interpretations $\omega_{i,j}$ ($i = 1..m$, $j = 1..n_i$) over $Var(E \cup \{IC\})$, where $N = \Sigma_{i=1..n} n_i$ is the total number of formulas $\varphi_{i,j}$ in E;
2. check that $\omega \models IC$ and that $\omega_{i,j} \models \varphi_{i,j}$ for all $i = 1..m$ and all $j = 1..n_i$;
3. compute $d(\omega, \omega_{i,j})$ for all i and all j;
4. compute $d(\omega, K_i)$ for all i;
5. compute $d(\omega, E)$ and check that $d(\omega, E) \leq k$.

This algorithm runs in polynomial time in the size of the input (E, IC and k represented in binary) since d, f, g are computable in polynomial time.

□

Lemma 2 *If for all $\omega \in \mathcal{W}$ the value of $d(\omega, E)$ is bounded by the value $h(|E| + |IC|)$ (where h is a function with values in $\mathbb{N}$) then the value* $\min_{\omega \models IC} d(\omega, E)$ *can be computed using* $\lceil \log h(|E| + |IC|) \rceil$ *calls to an* NP *oracle.*

Proof : $min = \min_{\omega \models IC} d(\omega, E)$ can be computed using binary search on $\{0, \ldots, h(|E| + |IC|)\}$ with at each step a call to an NP oracle to check whether $\min_{\omega \models IC} d(\omega, E) \leq k$ (that is in NP from lemma 1). Since a binary search on $\{0, \ldots, h(|E| + |IC|)\}$ needs at most $\lceil \log h(|E| + |IC|) \rceil$ steps, the result follows.

□

• *Point 1. of Proposition 1*
If d, f and g are computable in polynomial time, then for every knowledge set E and every $\omega \in \mathcal{W}$, the binary representation of $d(\omega, E)$ is bounded by $p(|E| + |IC|)$, where p is a polynomial. Hence, the value of $d(\omega, E)$ is bounded by $2^{p(|E|+|IC|)}$. From lemma 2, we can conclude that min can be computed using a polynomial number of calls to an NP oracle. Now, let E be a knowledge set, IC be a formula, k be an integer and α be a formula, it can be shown that the problem of determining whether there exists a model ω of IC such that $d(\omega, E) = k$ and such that $\omega \not\models \alpha$ is in NP. So we can show that $\triangle_{IC}^{d,f,g}(E) \not\models \alpha$ using first a polynomial number of calls to an NP oracle in order to compute min, and then using an additional call to an NP oracle in order to determine whether there exists a model ω of IC s.t. $d(\omega, E) = min$ and $\omega \not\models \alpha$. Hence the membership to

Δ_2^p for this problem, and hence for its complement.

• *Point 2. of Proposition 1*
When d, f and g are polynomially bounded, the proof is similar to the one of point 1., but the computation of $\min_{\omega \models IC} d(\omega, E)$ needs only a logarithmic number of steps since h is polynomially bounded, hence the membership to Θ_2^p.

□

Sketch of Proof of Proposition 2 :

1. Membership
Membership-to-Δ_2^p results are direct consequences of Proposition 1 since both distances and aggregation functions can be computed in polynomial time.
Membership-to-Θ_2^p results are also consequences of Proposition 1, except those for which f or g is an OWS_q (including lex) when the drastic distance d_D is considered; these cases are briefly discussed now:

- *case $d = d_D$, $f = max$ and $g = OWS_q$.* We first establish that $d(\omega, E)$ can only take only a polynomial number of different values, and that this set of possible values can be computed in polynomial time. Indeed, if $k_E(\omega)$ is the number of belief bases K_i from E s.t. $\omega \models K_i$, we have $d(\omega, E) = \sum_{i=k(\omega)+1}^{m} q_i$; which makes $|E| + 1$ different values, computable in polynimial time. The rest of the proof is similar to the proof of membership to Θ_2^p in the cases where g and g are polynomially bounded, the difference being here that the minimal value $min = \min_{\omega \in \Omega} d(\omega, E)$ is computed through binary search using the precomputed $|E|+1$ different possible values for $d(\omega, E)$.
- the *case $d = d_D$, $f = OWS_q$ and $g = max$* is similar, the main difference is that $d(\omega, E)$ can only take at most $max_{i \in 1 \dots n} card(K_i)$ different values.

Finally, as to the basic merging operator (d_D, max, max), determining whether a formula α is a logical consequence of the merged base E given IC can be achieved using the following algorithm:
`if` $sat(E \cup \{IC\})$ `then` return($unsat(E \cup \{IC, \neg\alpha\})$) `else` return($unsat(\{IC, \neg\alpha\})$), which shows membership of the decision problem to BH_2.

2. Hardness:

- *table 2, Θ_2^p-hardness results*: they are direct consequences of hardness results for cardinality-maximizing base revision $\circ_C$ (Theorem 5.14 from [Nebel, 1998]) since we have $\Delta_{IC}^{d_D,f,g}(\{\{\varphi_1\}, \dots, \{\varphi_n\}\}) \equiv \{\varphi_1, \dots, \varphi_n\} \circ_C IC$ for any $\langle f, g\rangle \in \{\langle max, max\rangle, \langle max, lex\rangle, \langle sum, max\rangle, \langle sum, sum\rangle\}$. Since sum is a specific OWS_q function, the corresponding results still hold in the cases ($f = OWS_q, g = max$) and ($f = max, g = OWS_q$).
- *table 2, case $d = d_D$, $f = OWS_q$, $g = sum$*: Δ_2^p-hardness is established by considering the following polynomial reduction from the Δ_2^p-complete problem MAX-SAT-ASG$_{odd}$ [Wagner, 1987]. MAX-SAT-ASG$_{odd}$ is the following decision problem: given a propositional formula Σ s.t. $Var(\Sigma) = \{x_1, \dots, x_n\}$ and a strict ordering $x_1 < x_2 < \dots < x_n$ on $Var(\Sigma)$ inducing the lexicographic ordering $\preccurlyeq$ on Ω, is the greatest model ω of Σ w.r.t. $\preccurlyeq$ such that $\omega(x_n) = 1$? We just give here the reduction: to Σ s.t. $Var(\Sigma) = \{x_1, \dots, x_n\}$, we associate the tuple $M(\Sigma) = \langle E, IC, \alpha\rangle$, where $E = \{K_i \mid i \in 1 \dots n\}$, $IC = \Sigma$, $\alpha = x_n$ and for each $i \in 1 \dots n$, $K_i = \{\bigwedge_{k=1}^{n+2-j} x_i \mid j \in 1 \dots n+2-i\}$ (each K_i contains $n+2-i$ formulas that are syntactically distinct but all equivalent to x_i), and we consider the OWS_q function f induced by q s.t. $q(1) = 1$ and for every $j > 1$, $q(j) = 2^{i-2}$.
- *table 2, Δ_2^p-hardness results in the case $f = sum$*: hardness in the case ($d = d_D, f = sum, g = lex$) is easily derived by taking advantage of the Δ_2^p-hardness result in the case where each K_i is a singleton reduced to a conjunction of atoms (hence f is irrelevant), $g = lex$ and $d = d_H$. Since sum is a specific WS_q function and lex is a specific OWS_q function, this hardness result can be extended to the rest of the table, except for the cases where f is a WS_q function and $g \in \{max, sum\}$) and where g is a WS_q function. In the latter case, the Δ_2^p-hardness of linear base revision $\circ_L$ (Theorem 5.9 from [Nebel, 1998]) can be used to obtain the desired result: indeed, it is sufficient to consider belief bases K_i reduced to singletons; we have $\triangle_{IC}^{d_D,f,g}(\{K_1, \dots, K_n\}) \equiv \{K_1, \dots, K_n\} \circ_L IC$, where g is the weighted sum induced by q s.t. $q(i) = 2^{n-i}$, and each K_i is viewed as the unique formula it contains. Here, the preference ordering over $\{K_1, \dots, K_n\}$ is s.t. $K_1 < K_2 < \dots < K_n$.
- *table 2, case* ($d = d_D, f = g = max$): it is sufficient to consider the following polynomial reduction M from SAT-UNSAT: to a pair of formulas $\langle\varphi, \psi\rangle$ which do not share variables (this can be assumed without loss of generality), we let $M(\langle\varphi, \psi\rangle) = \langle E = \varphi, IC = new, \alpha = \varphi \wedge new \wedge$

$\neg\psi\rangle$ where new is a new variable and we check that $\langle\varphi, \psi\rangle \in$ SAT-UNSAT iff α is a logical consequence of the merged base E given IC.

– *table 3, Θ_2^p-hardness results.* They still hold in the situation where E contains only one belief base K and K itself contains only one formula that is a conjunction of atoms. This merely shows that our hardness result is independent from f and g (since they are irrelevant whenever E and K are singletons) but is a consequence of the distance that is used (Hamming). Indeed, in this restricted case, $\triangle_{IC}^{d_H,f,g}(\{K\})$ is equivalent to $K \circ_D IC$ where $\circ_D$ is Dalal's revision operator. The fact that the inference problem from $K \circ_D IC$ is Θ_2^p-hard (even in the restricted case where K is a conjunction of atoms) concludes the proof (see Theorem 6.9 from [Eiter and Gottlob, 1992]).

– *table 3, Δ_2^p-hardness results.* We show that these Δ_2^p-hardness results hold in the restricted case where each K_i is a singleton, reduced to a conjunction of literals (hence f is irrelevant) when $g = lex$ by the following polynomial reduction M from MAX-SAT-ASG$_{odd}$: to any propositional formula Σ s.t. $Var(\Sigma) = \{x_1, \ldots, x_n\}$ we associate $M(\Sigma) = \langle E = \{K_i = \{x_i \wedge \bigwedge_{j=i+1}^{2n-i+1} new_j\} \mid i \in 1 \ .. \ n\}, IC = \Sigma \wedge \bigwedge_{j=2}^{2n} \neg new_j, \alpha = x_n\rangle$ where each new_j ($j \in 2 \ .. \ 2n$) is a new variable.

– *table 4.* We show that Δ_2^p-hardness holds in the very restricted case where E contains only one belief base K and K itself contains only one formula that is a conjunction of atoms. This merely shows that our hardness result is independent from f and g (since they are irrelevant whenever E and K are singletons) but is a consequence of the family of distances that is used (weighted Hamming). This is done by the following polynomial reduction M from MAX-SAT-ASG$_{odd}$: to any Σ s.t. $Var(\Sigma) = \{x_1, \ldots, x_n\}$ we associate $M(\Sigma) = \langle E = \{\{\bigwedge_{i=1}^{n} x_i\}\}, IC = \Sigma, \alpha = x_n\rangle$ and the weighted Hamming distance d_{H_q} induced by q s.t. $\forall i \in 1 \ .. \ n, q(x_i) = 2^{n-i}$.

□

References

[Baral *et al.*, 1991] C. Baral, S. Kraus, and J. Minker. Combining multiple knowledge bases. *IEEE Transactions on Knowledge and Data Engineering*, 3(2):208–220, 1991.

[Baral *et al.*, 1992] C. Baral, S. Kraus, J. Minker, and V. S. Subrahmanian. Combining knowledge bases consisting of first-order theories. *Computational Intelligence*, 8(1):45–71, 1992.

[Benferhat *et al.*, 2000] S. Benferhat, D. Dubois, S. Kaci, and H. Prade. Encoding information fusion in possibilistic logic: a general framework for rational syntactic merging. In *Proc. of ECAI'00*, pages 3–7, 2000.

[Bloch and Hunter, 2001] I. Bloch and A. Hunter, editors. *Fusion: General Concepts and Characteristics*, volume 16 of *International Journal of Intelligent Systems*. Wiley, 2001. Special Issue on Data and Knowledge Fusion.

[Dalal, 1988] M. Dalal. Investigations into a theory of knowledge base revision: preliminary report. In *Proc. of AAAI'88*, pages 475–479, 1988.

[Eiter and Gottlob, 1992] T. Eiter and G. Gottlob. On the complexity of propositional knowledge base revision , updates, and counterfactuals. *Artificial Intelligence*, 57(2-3):227–270, 1992.

[Hansson, 1998] S. O. Hansson. Revision of belief sets and belief bases. *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Vol. 3*, pages 17–75, 1998.

[Konieczny and Pino Pérez, 1999] S. Konieczny and R. Pino Pérez. Merging with integrity constraints. In *Proc. of ECSQARU'99*, LNAI 1638, pages 233–244, 1999.

[Konieczny and Pino Pérez, 2002] S. Konieczny and R. Pino Pérez. On the frontier between arbitration and majority. In *Proc. of KR'02*, 2002.

[Konieczny *et al.*, 2001] S. Konieczny, J. Lang, and P. Marquis. Distance-based merging: a general framework and some complexity results. Technical report, IRIT, `ftp://ftp.irit.fr/pub/IRIT/RPDMP/DBMC.ps.gz`, 2001.

[Konieczny, 2000] S. Konieczny. On the difference between merging knowledge bases and combining them. In *Proc. of KR'00*, pages 135–144, 2000.

[Lafage and Lang, 2000] C. Lafage and J. Lang. Logical representation of preferences for group decision theory. In *Proc. of KR'00*, pages 457–468, 2000.

[Lafage and Lang, 2001] C. Lafage and J. Lang. Propositional distances and preference representation. In *Proc. of ECSQARU'01*, pages 48–59, 2001.

[Liberatore and Schaerf, 1998] P. Liberatore and M. Schaerf. Arbitration (or how to merge knowledge bases). *IEEE Transactions on Knowledge and Data Engineering*, 10(1):76–90, 1998.

[Liberatore and Schaerf, 2000] P. Liberatore and M. Schaerf. Brels: a system for the integration of knowledge bases. In *Proc. of KR'00*, pages 145–152, 2000.

[Lin and Mendelzon, 1999] J. Lin and A. O. Mendelzon. Knowledge base merging by majority. In *Dynamic Worlds: From the Frame Problem to Knowledge Management*. Kluwer, 1999.

[Nebel, 1989] B. Nebel. A knowledge level analysis of belief revision. In *Proc. of KR'89*, pages 301–311, 1989.

[Nebel, 1998] B. Nebel. How hard is it to revise a belief base? *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Vol. 3: Belief Change*, pages 77–145, 1998.

[Papadimitriou, 1994] C. H. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.

[Pinkas, 1995] G. Pinkas. Reasoning, nonmonotonicity and learning in connectionist networks that capture propositional knowledge. *Artificial Intelligence*, 77:203–247, 1995.

[Revesz, 1997] P. Z. Revesz. On the semantics of arbitration. *International Journal of Algebra and Computation*, 7(2):133–160, 1997.

[Wagner, 1987] K. W. Wagner. More complicated questions about maxima and minima, and some closures of NP. *Theoretical Computer Science*, 51:53–80, 1987.

[Yager, 1998] R. R. Yager. On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18:183–190, 1998.