Smallest Explanations and Diagnoses of Rejection in Abstract Argumentation

Andreas Niskanen and **Matti Järvisalo**
HIIT, Department of Computer Science, University of Helsinki, Finland

Abstract

Deciding acceptance of arguments is a central problem in the realm of abstract argumentation. Beyond mere acceptance status, when an argument is rejected it would be informative to analyze reasons for the rejection. Recently, two complementary notions—explanations and diagnoses—were proposed for capturing underlying reasons for rejection in terms of (small) subsets of arguments or attacks. We provide tight complexity results for deciding and computing argument-based explanations and diagnoses. Computationally, we identify that smallest explanations and diagnoses for argumentation frameworks can be computed as so-called smallest unsatisfiable subsets (SMUSes) and smallest correction sets of propositional formulas. Empirically, we show that SMUS extractors and maximum satisfiability solvers (computing smallest correction sets) offer effective ways of computing smallest explanations and diagnoses.

1 Introduction

Formal argumentation is an active area of knowledge representation and reasoning research, where abstract argumentation frameworks (AFs) constitute a central formalism (Dung 1995). Syntactically AFs are directed graphs in which nodes represent arguments and directed edges conflicts (attacks) between arguments. Argumentation semantics define criteria for acceptable subsets of arguments called extensions.

A central fundamental reasoning task in AFs is to decide whether a given query argument is accepted credulously (i.e., contained in at least one extension of an AF) or skeptically (i.e., contained in all extensions) under a prescribed argumentation semantics. For credulous reasoning a "yes" answer is witnessed by an extension containing the query argument. However, a "no" answer requires proving that no such extensions exists. Our focus is on credulous reasoning, building on (Saribatur, Wallner, and Woltran 2020).[1]

Beyond mere acceptance status, in particular when an argument is rejected under credulous reasoning, it would be more informative to obtain information on the reasons for rejection. While automated reasoning can be harnessed for constructing proofs (Wetzler, Heule, and Hunt 2014) of rejection, such proofs can be exponentially large and do not directly provide informative ideas for the reasons of rejection. This intuition has led to recent proposals for explaining rejection (or non-acceptability) in terms of, e.g., minimal sets of arguments and attacks to be removed from an AF to enforce acceptability (Fan and Toni 2015)[2], and abduction in argumentation frameworks as modifications to an AF to enforce a status under a labeling-based view (Sakama 2018).

In this work, we focus on two complementary notions—explanations (Saribatur, Wallner, and Woltran 2020) and diagnoses (Ulbricht and Baumann 2019)—that, building in part on the notion of strong inconsistency in non-monotonic reasoning (Brewka, Thimm, and Ulbricht 2019), have been recently proposed for capturing underlying reasons for rejection in terms of (small) subsets of arguments or attacks which are enough to yield rejection. Intuitively, a diagnosis is a part of an AF the removal of which results in acceptance, and an explanation is a part which resists all additions allowed by the AF in terms of rejection of the query. While diagnoses were originally proposed without attention to acceptance of arguments (Ulbricht and Baumann 2019), we focus on rejection of a particular query argument.

We provide complexity results and practical algorithms for computing smallest and minimal explanations and diagnoses, with a stronger focus on the smallest counterparts which can be viewed as the most generic explanations and diagnoses. We identify that a standard reduction of propositional formulas in conjunctive normal form to AFs allows for obtaining tight complexity results for computing argument-based explanations and diagnoses. In particular, we identify a correspondence between smallest explanations and smallest unsatisfiable subsets (SMUSes) of propositional formulas, as well as smallest diagnoses and smallest correction sets of propositional formulas. This correspondence allows for harnessing Boolean satisfiability (SAT) based SMUS extractors (Ignatiev et al. 2015) and maximum satisfiability (MaxSAT) solvers for computing smallest explanations and smallest diagnoses, respectively, as well as MUS extractors and MCS extractors for computing minimal explanations

[1] Dually, for skeptical reasoning a "no" answer is witnessed by an extension, "yes" answer requires a proof that counterexamples do not exist.

[2] With our terminology explanations of (Fan and Toni 2015) correspond to minimal diagnoses (Ulbricht and Baumann 2019).

and diagnoses. Empirically, we show that SMUS extraction and MaxSAT are effective ways of computing smallest explanations and diagnoses for rejection in credulous reasoning on standard argumentation reasoning benchmarks, significantly outperforming a recently proposed answer set programming approach (Saribatur, Wallner, and Woltran 2020) to computing smallest explanations.

2 Preliminaries

We begin with background on abstract argumentation, explanations and diagnoses in AFs, and unsatisfiability and correction sets in propositional logic.

Abstract Argumentation An *argumentation framework* (AF) (Dung 1995) is a pair $F = (A, R)$ where A is a (finite) set of arguments and $R \subseteq A \times A$ is a set of attacks. The notions of admissibility and stable semantics (Baroni, Caminada, and Giacomin 2018) provide two ways for characterizing jointly acceptable arguments.[3] Given an AF $F = (A, R)$, a set $S \subseteq A$ is *conflict-free* if there are no $x, y \in S$ with $(x, y) \in R$. The collection of conflict-free sets of F is denoted by $cf(F)$. A set $S \in cf(F)$ is an *admissible set* if for each $(b, a) \in R$ with $a \in S$ there is $(c, b) \in R$ with $c \in S$. A set $S \in cf(F)$ is a *stable extension* if for each $a \in A \setminus S$ there is $b \in S$ with $(b, a) \in R$. Admissible sets and stable extensions of F are denoted by $adm(F)$ and $stb(F)$, respectively. Argument a is *credulously accepted* under semantics $\sigma \in \{adm, stb\}$ if there is $E \in \sigma(F)$ with $a \in E$; otherwise a is rejected in F (in terms of credulous reasoning).

Explanations and Diagnoses Following Saribatur, Wallner, and Woltran (2020), if a is not credulously accepted under semantics σ, a strongly rejecting subframework provides an *explanation* for not accepting a. A subset of arguments $A' \subseteq A$ gives rise to the subframework $F|_{A'} = (A', R \cap (A' \times A'))$, and a subset of attacks $R' \subseteq R$ gives rise to the subframework $F|_{R'} = (A, R')$.

Definition 1. For a set $A' \subseteq A$, the subframework $F|_{A'}$ *strongly rejects* a if a is not credulously accepted under σ in $F|_{A''}$ for any $A'' \supseteq A'$. Such A' is an *argument-based explanation* for rejecting a. For a set of attacks $R' \subseteq R$, the subframework $F|_{R'}$ *strongly attack-rejects* a if a is not credulously accepted under σ in $F|_{R''}$ for any $R'' \supseteq R'$. Such R' is an *attack-based explanation* for rejecting a.

Diagnoses (Ulbricht and Baumann 2019) are a dual notion of explanations for non-acceptability.[4] In particular, diagnoses are subsets of arguments and attacks whose removal from the framework results in a subframework—called a repair—where the argument is credulously accepted.

Definition 2. A set $A' \subseteq A$ is an *argument-based diagnosis* (of rejection) if a is credulously accepted under σ in the subframework $F|_{A \setminus A'}$. A set of attacks $R' \subseteq R$ is an *attack based-diagnosis* if a is credulously accepted under σ in the subframework $F|_{R \setminus R'}$.

Our focus is on *smallest* (and *minimal*) explanations and diagnoses in terms of set-cardinality.

MUSes and MaxSAT We recall standard notions for analyzing the unsatisfiability of propositional formulas central in our complexity analysis and algorithms for computing small explanations and diagnoses. We assume familiarity with propositional logic, in particular satisfiability (SAT) of formulas in conjunctive normal form (CNF). Let $\mathcal{F} = \{c_1, \ldots, c_m\}$ be a CNF formula.

Definition 3. A set $\mathcal{M} \subseteq \mathcal{F}$ is a *minimal unsatisfiable subformula (MUS)* (Bruni 2003; Liffiton and Sakallah 2008; Kleine Büning and Kullmann 2009) of an unsatisfiable CNF formula $\mathcal{F}$ if (i) $\mathcal{M}$ is unsatisfiable and (ii) every $\mathcal{M}' \subsetneq \mathcal{M}$ is satisfiable. A smallest MUS (SMUS) $\mathcal{M}$ is an MUS such that there is no MUS $\mathcal{M}'$ with $|\mathcal{M}'| < |\mathcal{M}|$.

Deciding whether a given CNF formula $\mathcal{F}$ has an unsatisfiable subformula of size $\leq k$ for a given integer k is known to be Σ_2^p-complete (Liberatore 2005).

Minimal correction sets (Marques-Silva et al. 2013) are a dual notion of MUSes (Reiter 1987).

Definition 4. A set $\mathcal{M} \subseteq \mathcal{F}$ is a *minimal correction set (MCS)* of an unsatisfiable CNF formula F if (i) $\mathcal{F} \setminus \mathcal{M}$ is satisfiable and (ii) for all $\mathcal{M}' \subsetneq \mathcal{M}$, $\mathcal{F} \setminus \mathcal{M}'$ is unsatisfiable.

The task of computing a smallest MCS of a given CNF formula $\mathcal{F}$ is equivalent to finding an optimal MaxSAT solution (Li and Manyà 2009) to $\mathcal{F}$, i.e., a truth assignment that satisfies as many clauses of $\mathcal{F}$ as possible; in particular, the clauses left unsatisfied by an optimal MaxSAT solution are a smallest MCS of $\mathcal{F}$.

A standardly used and implemented generalization of MUSes and MCSes is to allow for declaring both a set of hard clauses $\mathcal{H}$ and soft clauses $\mathcal{S}$, and defining MUSes (resp. MCSes) as minimal subsets $\mathcal{M} \subseteq \mathcal{S}$ of the soft clauses which together with the hard clauses constitute an unsatisfiable CNF formula $\mathcal{M} \cup \mathcal{H}$ (resp. such that $\mathcal{H} \cup (\mathcal{S} \setminus \mathcal{M})$ is satisfiable). In particular, MUS and MCS extraction algorithms readily support this generalization which has no impact on computational complexity.

3 Complexity Results

Saribatur, Wallner, and Woltran (2020) showed that deciding whether there is a subframework strongly rejecting a queried argument is in Σ_2^p and NP- and coNP-hard under admissibility. We close this gap by showing Σ_2^p-hardness under admissibility and establish Σ_2^p-completeness under stable semantics. Furthermore, we show that verification of minimal explanations is DP-complete.

Theorem 1. *Given an AF $F = (A, R)$, $a \in A$, $\sigma \in \{adm, stb\}$, and an integer $k \geq 0$, the following hold. (i) Deciding whether there is an argument-based explanation $A' \subseteq A$ with $|A'| \leq k$ for rejecting a in F under σ is Σ_2^p-complete. (ii) Verifying that a given $A' \subseteq A$ is a minimal argument-based explanation for rejecting a in F under σ is DP-complete.*

[3] Note that credulous acceptance under admissibility coincides with credulous acceptance under complete and preferred semantics. Thus admissible and stable semantics cover credulous reasoning under all standard AF semantics (Dung 1995) except grounded.

[4] Ulbricht and Baumann (2019) consider diagnoses and repairs in the context where no argument is accepted, i.e. $\bigcup \sigma(F) = \emptyset$.

Proof. (i) Membership is by guessing a subset $A' \subseteq A$ and checking in coNP if $F|_{A'}$ strongly rejects a under σ (Saribatur, Wallner, and Woltran 2020).

For hardness, we reduce from the Σ_2^p-complete (Liberatore 2005) problem of deciding whether there is an unsatisfiable subset of size at most k. Let $\mathcal{F} = \{c_1, \ldots, c_m\}$ be a CNF formula over variables $\mathcal{X} = \{x_1, \ldots, x_n\}$ and $k \geq 0$. Consider the *standard reduction from CNF formulas to AFs* (Dimopoulos and Torres 1996; Dvorák and Dunne 2018), i.e., $F = (A, R)$ with $A = A_x \cup A_c \cup \{\varphi\}$, where $A_x = \{x_i, \overline{x}_i \mid i = 1..n\}$ and $A_c = \{\overline{c}_j \mid j = 1..m\}$, and $R = \{(x_i, \overline{x}_i), (\overline{x}_i, x_i) \mid i = 1..n\} \cup \{(x_i, \overline{c}_j) \mid x_i \in c_j\} \cup \{(\overline{x}_i, \overline{c}_j) \mid \neg x_i \in c_j\} \cup \{(\overline{c}_j, \varphi) \mid j = 1..m\}$. It is well-known that φ is credulously accepted under σ iff $\mathcal{F}$ is satisfiable. Suppose φ is not credulously accepted in F, and let $A' \subseteq A$. We make the following observations: (a) The set A' is an explanation for rejecting φ under *adm* iff A' is an explanation for rejecting φ under *stb*. (b) If A' is an explanation for rejecting φ under σ, then so is $A' \cap A_c$. By observation (a) it suffices to consider one of *adm* or *stb* to prove the claim.

Let $\mathcal{M} \subseteq \mathcal{F}$, $|\mathcal{M}| \leq k$ be an unsatisfiable subformula. We show that $F|_{A'}$ with $A' = \{\overline{c}_j \mid c_j \in \mathcal{M}\}$ strongly rejects φ under *adm*, which in turn implies $|A'| \leq k$. Suppose on the contrary that for some $A'' \supseteq A'$, φ is credulously accepted in $F|_{A''}$. Now φ is credulously accepted also in $F' = (A_x \cup A'', \{(x_i, \overline{x}_i), (\overline{x}_i, x_i) \mid i = 1..n\} \cup (R \cap (A'' \times A'')))$. Further, F' corresponds to the formula $\mathcal{M}' = \{c_j \setminus (\{x_i \mid x_i \notin A''\} \cup \{\neg x_i \mid \overline{x}_i \notin A''\}) \mid \overline{c}_j \in A''\}$, implying that $\mathcal{M}'$ is satisfiable. However, since $\mathcal{M}$ is obtained from $\mathcal{M}'$ by removing clauses and adding literals to clauses, $\mathcal{M}$ must also be satisfiable, which is a contradiction.

Let $A' \subseteq A$, $|A'| \leq k$, and $F|_{A'}$ be a strongly rejecting subframework for φ under *adm*. Via observation (b), $A' \cap A_c$ also yields such a strongly rejecting subframework. We show that $\mathcal{M} = \{c_j \mid \overline{c}_j \in A' \cap A_c\}$ is an unsatisfiable subformula, implying $|\mathcal{M}| \leq k$. Suppose on the contrary that $\mathcal{M}$ is satisfiable. This implies that φ is credulously accepted under *adm* in the framework $F'' = (A'', R \cap (A'' \times A''))$, where $A'' = A_x \cup (A' \cap A_c) \cup \{\varphi\}$. But $A'' \supseteq A' \cap A_c$, which contradicts strong rejection.

(ii) Membership in DP is by checking first in NP if for all $a' \in A'$ the argument a is credulously accepted under σ in $F|_{A' \setminus \{a'\}}$, and then in coNP whether $F|_{A'}$ strongly rejects a under σ (Saribatur, Wallner, and Woltran 2020).

For hardness, we reduce from the DP-complete problem of verifying whether for a CNF formula $\mathcal{F}$ a given subformula $\mathcal{M}$ is an MUS (Papadimitriou and Wolfe 1988). Consider again the reduction from CNF formulas to AFs (recall part (i)).

Let $\mathcal{M} \subseteq \mathcal{F}$ be an MUS. The set $A' = \{\overline{c}_j \mid c_j \in \mathcal{M}\}$ is now an argument-based explanation for rejecting φ under $\sigma \in \{adm, stb\}$. Minimality of A' follows from the fact that for any $c_j \in \mathcal{M}$, the subframework $F|_{A' \setminus \{\overline{c}_j\}}$ does not strongly reject φ, since $\mathcal{M} \setminus \{c_j\}$ is satisfiable.

Let $A' \subseteq A$ be a minimal argument-based explanation for rejecting φ under $\sigma \in \{adm, stb\}$. Since A' is minimal, $A' \subseteq \{\overline{c}_j \mid j = 1..m\}$, and hence $\mathcal{M} = \{c_j \mid \overline{c}_j \in A'\}$ is an unsatisfiable subformula. Minimality of $\mathcal{M}$ follows from the fact that for any $\overline{c}_j \in A'$ the subformula $\mathcal{M} \setminus \{c_j\}$ is satisfiable, since $F|_{A' \setminus \{\overline{c}_j\}}$ does not strongly reject φ. □

We continue with complexity results for argument-based diagnoses. In particular, we establish NP-completeness for deciding whether there exists a diagnosis of at most a given size. Further, we extend the DP-completeness result of verifying minimal diagnoses for semantical collapse ($\bigcup \sigma(F) = \emptyset$) (Ulbricht and Baumann 2019) to argument rejection.

Theorem 2. *Given an AF $F = (A, R)$, $a \in A$, $\sigma \in \{adm, stb\}$, and $k \geq 0$, the following hold. (i) Deciding whether there is an argument-based diagnosis $A' \subseteq A$ with $|A'| \leq k$ of rejecting a in F under σ is NP-complete. (ii) Verifying that a given $A' \subseteq A$ is a minimal argument-based diagnosis for rejecting a in F under σ is DP-complete.*

Proof. (Sketch.) (i) For membership, guess a subset of arguments A' and a set $E \subseteq A'$ containing a, and verify in polynomial time whether $|A'| \leq k$ and $E \in \sigma(F|_{A \setminus A'})$. For hardness, reduce from credulous acceptance under σ (an NP-complete problem) by setting $k = 0$.

(ii) Membership follows from checking in NP whether a is credulously accepted in $F|_{A \setminus A'}$ and in coNP whether for all $a' \in A$, a is not credulously accepted in $F|_{A \setminus (A' \setminus \{a'\})}$. For hardness, reduce from the DP-complete problem of verifying an MCS (Chen and Toda 1995), utilizing again the standard reduction from CNF to AFs. In particular, arguments $\overline{c}_j$ form a minimal diagnosis for rejecting φ under σ if and only if clauses c_j form an MCS. □

4 Smallest Explanations as Smallest MUSes

Moving from complexity to computation, we show how to encode the problem of computing a smallest explanation of credulous rejection as the problem of computing a smallest MUS. This is in particular motivated by the fact that deciding the existence of small argument-based explanations (Theorem 1) and small MUSes are both Σ_2^p-complete. The same approach also allows for computing minimal argument-based explanations.

Let $F = (A, R)$ be an AF, and $\sigma \in \{adm, stb\}$. We declare variables x_a for each $a \in A$ corresponding to the inclusion of a in a σ-extension of F, and y_a for each $a \in A$ corresponding to the existence of argument a in a subframework. We define the propositional formulas $\varphi_{cf}^A(F) = \bigwedge_{(a,b) \in R} ((y_a \wedge y_b) \rightarrow (\neg x_a \vee \neg x_b))$, $\varphi_{adm}^A(F)$ as

$$\varphi_{cf}^A(F) \wedge \bigwedge_{(b,a) \in R} ((y_a \wedge y_b \wedge x_a) \rightarrow \bigvee_{(c,b) \in R} (y_c \wedge x_c)), \text{ and}$$

$$\varphi_{stb}^A(F) = \varphi_{cf}^A(F) \wedge \bigwedge_{a \in A} ((y_a \wedge \neg x_a) \rightarrow \bigvee_{(b,a) \in R} (y_b \wedge x_b)).$$

Proposition 3. *Given $F = (A, R)$, $q \in A$, and $\sigma \in \{adm, stb\}$, for the formula $\mathcal{H}^A \cup \mathcal{S}^A$ with $\mathcal{H}^A = \varphi_\sigma^A(F) \wedge \neg x_q$ and $\mathcal{S}^A = \bigwedge_{a \in A} y_a$, the subformula $\bigwedge_{a \in A' \subseteq A} y_a$ is a (smallest) MUS of $\mathcal{H}^A \cup \mathcal{S}^A$ if and only if A' is a minimal (resp. smallest) argument-based explanation for rejecting a in F under σ.*

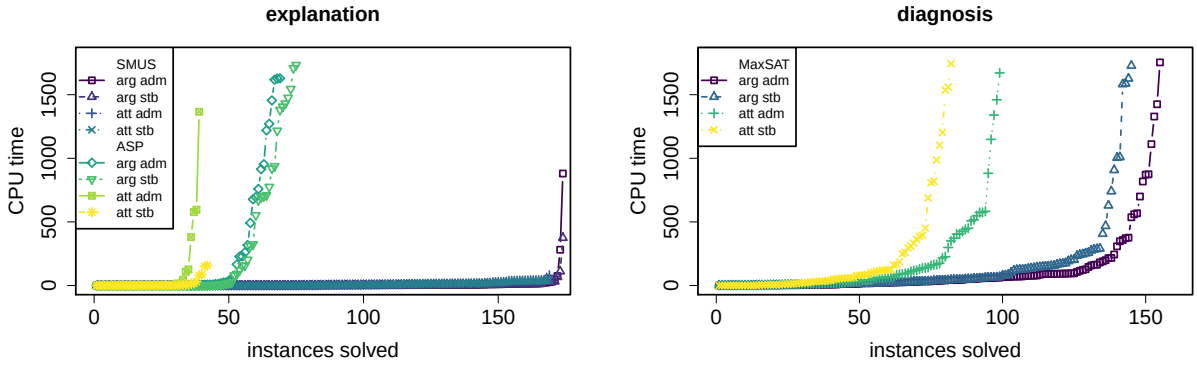

Figure 1: Left: SMUS vs ASP on computing a smallest explanation (ICCMA'19 instances). Right: MaxSAT on computing a smallest diagnosis (ICCMA'17 instances).

For attack-based explanations, we declare variables x_a for each $a \in A$ corresponding to a σ-extension, and furthermore variables $r_{a,b}$ for each $(a,b) \in R$ corresponding to the existence of attack (a,b) in a subframework. We consider the formulas $\varphi^R_{cf}(F) = \bigwedge_{(a,b)\in R} (r_{a,b} \rightarrow (\neg x_a \vee \neg x_b))$,

$$\varphi^R_{adm}(F) = \varphi^R_{cf}(F) \wedge \bigwedge_{(b,a)\in R} ((r_{b,a} \wedge x_a) \rightarrow \bigvee_{(c,b)\in R} (r_{c,b} \wedge x_c)),$$

and $$\varphi^R_{stb}(F) = \varphi^R_{cf}(F) \wedge \bigwedge_{a\in A} (\neg x_a \rightarrow \bigvee_{(b,a)\in R} (r_{b,a} \wedge x_b)).$$

.

Proposition 4. *Given $F = (A,R)$, $q \in A$, and $\sigma \in \{adm, stb\}$, for the formula $\mathcal{H}^R \cup \mathcal{S}^R$ with $\mathcal{H}^R = \varphi^R_\sigma(F) \wedge \neg x_q$ and $\mathcal{S}^R = \bigwedge_{(a,b)\in R} r_{a,b}$, the subformula $\bigwedge_{(a,b)\in R' \subseteq R} r_{a,b}$ is a (smallest) MUS of $\mathcal{H}^R \cup \mathcal{S}^R$ if and only if R' is a minimal (resp. smallest) attack-based explanation for rejecting a in F under σ.*

5 Smallest Diagnoses via Smallest MCSes

We turn to the computation of smallest (and minimal) diagnoses via MCS extraction. In particular, standard MCS extraction algorithms can then be used to extract minimal diagnoses, and MaxSAT solvers are applicable for extracting smallest diagnoses. Again, the related decision problems of small argument-based (Theorem 2) and small-cost MaxSAT solutions coincide. The following propositions also establish that computing smallest both argument and attack-based diagnoses is polynomial-time computable with access to a logarithmic number of NP oracle calls.

Here we again make use of the formulas $\varphi^A_\sigma(F)$ and $\varphi^R_\sigma(F)$ for capturing both argument-based and attack-based diagnosis, as detailed by the following propositions.

Proposition 5. *Given $F = (A,R)$, $a \in A$, and $\sigma \in \{adm, stb\}$, consider the CNF formula $\mathcal{H}^A \cup \mathcal{S}^A$ with hard clauses $\mathcal{H}^A = \varphi^A_\sigma(F) \wedge \neg x_a$ and soft clauses $\mathcal{S}^A = \bigwedge_{a\in A} y_a$. Now $\bigwedge_{a\in A' \subseteq A} y_a$ is an MCS of $\mathcal{H}^A \cup \mathcal{S}^A$ if and only if A' is a minimal argument-based diagnosis for rejecting a in F under σ. In particular, τ is an optimal MaxSAT solution to $\mathcal{H}^A \cup \mathcal{S}^A$ if and only if the set $\{a \in A \mid \tau(y_a) = 0\}$ is a smallest argument-based diagnosis for rejecting a in F under σ.*

Similarly, MaxSAT can be used for computing smallest attack-based diagnoses.

Proposition 6. *Given $F = (A,R)$, $a \in A$, and $\sigma \in \{adm, stb\}$, consider the CNF formula $\mathcal{H}^R \cup \mathcal{S}^R$ with hard clauses $\mathcal{H}^R = \varphi^A_\sigma(F) \wedge \neg x_a$ and soft clauses $\mathcal{S}^R = \bigwedge_{(a,b)\in R} r_{a,b}$. Now $\bigwedge_{(a,b)\in R' \subseteq R} r_{a,b}$ is an MCS of $\mathcal{H}^R \cup \mathcal{S}^R$ if and only if R' is a minimal attack-based diagnosis for rejecting a in F under σ. In particular, τ is an optimal MaxSAT solution to $\mathcal{H}^R \cup \mathcal{S}^R$ if only if the set $\{(a,b) \in R \mid \tau(r_{a,b}) = 0\}$ is a smallest attack-based diagnosis for rejecting a in F under σ.*

6 Experiments

We evaluate the empirical efficiency of computing smallest attack and argument-based explanations and diagnoses for credulously rejecting an argument using state-of-the-art implementations for computing smallest MUSes and optimal MaxSAT solutions. We use FORQES (Ignatiev et al. 2015) as the SMUS extractor, and RC2 (Ignatiev, Morgado, and Marques-Silva 2019) as the MaxSAT solver. For smallest explanations we compare to the recent answer set programming (ASP) based approach presented in (Saribatur, Wallner, and Woltran 2020), which uses the ASP solver CLINGO (Gebser et al. 2016). To the best of our knowledge, our system is the first one for computing smallest diagnoses.

For explanation benchmarks, we followed (Saribatur, Wallner, and Woltran 2020), and used ICCMA'19 competition AFs. We used the original query arguments for "no" instances. For original "yes" instances, we chose at random an alternative query and used it in case it gave a "no" instance. This gave 211 instances. The ICCMA'19 instances turned out to be too easy for diagnosis (inline with lower computational complexity of diagnosis), so for diagnosis we used the ICCMA'17 competition benchmark set B instances that yield "no" instances under admissibility, giving 193 instances. The experiments were run on nodes with 8-core Intel Xeon E5-2670 2.6-GHz CPUs and 64-GB memory under a per-instance 1800-second time and 16-GB memory limit.

Figure 1 (left) shows that SMUS extraction with FORQES vastly outperforms the ASP approach to explanation. FORQES solved $> 80\%$ of the instances for both argu-

ment and attack-based explanations and both semantics. CLINGO solved $\approx$ 40% for argument-based and $\approx$ 20% for attack-based explanations. FORQES solves instances for which smallest explanations have over 4000 elements, while CLINGO only solves instances with up to 4 arguments and 3 attacks in the smallest explanations. Instances with more than 50 elements in smallest explanations arose only from the AdmBuster instance family consisting of crafted instances (Mailly and Maratea 2019). We hypothesize that for this reason also the corresponding explanations may have compact representations. Figure 1 (right) shows that MaxSAT is effective in finding smallest diagnoses. Argument-based diagnosis appears easier than attack-based, possibly due to the larger search space over attacks, the number of which can be quadratic in the number of arguments.

7 Conclusion

Computing minimal explanations and diagnoses of rejecting a credulous query in abstract argumentation tightly correspond to the established concepts of minimal unsatisfiable subsets and minimal correction sets of propositional formulas. Identifying this connection yields new complexity results for computing and verifying explanations and diagnoses. Furthermore, harnessing SAT-based algorithmic approaches to computing smallest unsatisfiable subsets and correction sets yields efficient ways of computing smallest explanations and diagnoses for credulous rejection of arguments in argumentation frameworks. Establishing the computational complexity of attack-based explanations and diagnoses is an important aspect of future work.

Acknowledgments

This work was financially supported by Academy of Finland (grants 322869, 328718) and University of Helsinki Doctoral Programme in Computer Science DoCS. The authors thank Alexey Ignatiev for advice on Forqes and Johannes P. Wallner for assistance on computing explanations with the ASP approach. Computational resources were provided by Finnish Grid and Cloud Infrastructure (urn:nbn:fi:research-infras-2016072533).

References

Baroni, P.; Caminada, M.; and Giacomin, M. 2018. Abstract argumentation frameworks and their semantics. In *Handbook of Formal Argumentation*. College Publications. chapter 4, 159–236.

Brewka, G.; Thimm, M.; and Ulbricht, M. 2019. Strong inconsistency. *Artif. Intell.* 267:78–117.

Bruni, R. 2003. Approximating minimal unsatisfiable subformulae by means of adaptive core search. *Discret. Appl. Math.* 130(2):85–100.

Chen, Z., and Toda, S. 1995. The complexity of selecting maximal solutions. *Inf. Comput.* 119(2):231–239.

Dimopoulos, Y., and Torres, A. 1996. Graph theoretical structures in logic programs and default theories. *Theor. Comput. Sci.* 170(1-2):209–244.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.* 77(2):321–358.

Dvorák, W., and Dunne, P. 2018. Computational problems in formal argumentation and their complexity. In *Handbook of Formal Argumentation*. College Publications. chapter 13, 631–687.

Fan, X., and Toni, F. 2015. On explanations for non-acceptable arguments. In *TAFA Revised Selected Papers*, volume 9524 of *LNCS*, 112–127. Springer.

Gebser, M.; Kaminski, R.; Kaufmann, B.; Ostrowski, M.; Schaub, T.; and Wanko, P. 2016. Theory solving made easy with Clingo 5. In *Tech. Commun. ICLP*, volume 52 of *OASICS*, 2:1–2:15. Leibniz-Zentrum für Informatik.

Ignatiev, A.; Previti, A.; Liffiton, M. H.; and Marques-Silva, J. 2015. Smallest MUS extraction with minimal hitting set dualization. In *CP*, volume 9255 of *LNCS*, 173–182. Springer.

Ignatiev, A.; Morgado, A.; and Marques-Silva, J. 2019. RC2: An efficient MaxSAT solver. *J. Satisf. Boolean Model. Comput.* 11(1):53–64.

Kleine Büning, H., and Kullmann, O. 2009. Minimal unsatisfiability and autarkies. In *Handbook of Satisfiability*, volume 185 of *FAIA*. IOS Press. chapter 11, 339–401.

Li, C. M., and Manyà, F. 2009. MaxSAT, hard and soft constraints. In *Handbook of Satisfiability*, volume 185 of *FAIA*. IOS Press. chapter 19, 613–631.

Liberatore, P. 2005. Redundancy in logic I: CNF propositional formulae. *Artif. Intell.* 163(2):203–232.

Liffiton, M. H., and Sakallah, K. A. 2008. Algorithms for computing minimal unsatisfiable subsets of constraints. *J. Autom. Reason.* 40(1):1–33.

Mailly, J., and Maratea, M. 2019. Assessment of benchmarks for abstract argumentation. *Argument Comput.* 10(2):107–112.

Marques-Silva, J.; Heras, F.; Janota, M.; Previti, A.; and Belov, A. 2013. On computing minimal correction subsets. In *IJCAI*, 615–622. IJCAI/AAAI.

Papadimitriou, C. H., and Wolfe, D. 1988. The complexity of facets resolved. *J. Comput. Syst. Sci.* 37(1):2–13.

Reiter, R. 1987. A theory of diagnosis from first principles. *Artif. Intell.* 32(1):57–95.

Sakama, C. 2018. Abduction in argumentation frameworks. *J. Appl. Non-Class. Log.* 28(2-3):218–239.

Saribatur, Z. G.; Wallner, J. P.; and Woltran, S. 2020. Explaining non-acceptability in abstract argumentation. In *ECAI*, FAIA. IOS Press.

Ulbricht, M., and Baumann, R. 2019. If nothing is accepted - repairing argumentation frameworks. *J. Artif. Intell. Res.* 66:1099–1145.

Wetzler, N.; Heule, M.; and Hunt, W. A. 2014. DRAT-trim: Efficient checking and trimming using expressive clausal proofs. In *SAT*, volume 8561 of *LNCS*, 422–429. Springer.