1 Introduction

- explanations in constraint satisfaction problems TODO:cite: xai special issue easy, understandable, human-interpretable
- Explanations can be small, but also difficult (combinations of constraints/clues)
- TODO:which complexity is extracting an OMUS ?
- use a/different proxy(s) to qualify interpretability of an explanation
- extension SMUS to OMUS based on enumeration of optimal hitting set
- TODO:approach tested on Boolean Satisfiability instances and on a high level problem

2 Background

The algorithm presented in this section is based on the key ideas and observations of Ignatiev et al. presented in [Ignatiev et al., 2015]. The algorithm is adapted to incorporate an optimality criterion in order to guide the search not in the direction of the SMUS, but towards the OMUS. To do so, we first define the objective function f:

Definition 1. *Given a CNF Formula $\mathcal{F}$, let $f : 2^{\mathcal{F}} \rightarrow \mathbb{R}$ be a mapping of a set of clauses to a real number. f is said to be a valid objective function if for any subsets $\mathcal{A}$, $\mathcal{B}$ of $\mathcal{F}$ if $\mathcal{A} \subseteq \mathcal{B}$ then $f(\mathcal{A}) \leq f(\mathcal{B})$. function f is said to be set increasing if ...*

For an unsatisfiable CNF formula $\mathcal{F}$, a subset of clauses that are still unsatisfiable, but if any of the clauses are removed then the reduced formula becomes satisfiable. Formally, we define this set as a Minimum unsatisfiable Subset (MUS):

Definition 2. *A subset $\mathcal{U} \subseteq \mathcal{F}$ is a* ***minimal unsatisfiable subset*** *(MUS) if $\mathcal{U}$* unsatisfiable *and $\forall\, \mathcal{U}' \subset \mathcal{U}$, $\mathcal{U}'$ is* satisfiable*. An MUS of $\mathcal{F}$ with an optimal value w.r.t an* objective function *f is called an* ***optimal MUS*** *(OMUS).*

Definition 3. *A subset $\mathcal{C}$ of $\mathcal{F}$ is an* ***minimal correction subset*** *(MCS) if $\mathcal{F} \setminus \mathcal{C}$ is* satisfiable *and $\forall \mathcal{C}' \subseteq \mathcal{C} \wedge \mathcal{C}' \neq \emptyset$, $(\mathcal{F} \setminus \mathcal{C}) \cup \mathcal{C}'$ is* unsatisfiable.

Definition 4. *A* satisfiable subset *$\mathcal{S} \subseteq \mathcal{F}$ is a* ***Maximal Satisfiable Subset*** *(MSS) if $\forall\, \mathcal{S}' \subseteq \mathcal{F}'$ s.t $\mathcal{S} \subseteq \mathcal{S}'$, $\mathcal{S}'$ is* unsatisfiable.

An MSS can also be defined as the complement of an MCS (and vice versa). If $\mathcal{C}$ is a MCS then $\mathcal{S} = \mathcal{F} \setminus \mathcal{C}$ is a MSS. On the other hand, MUSes and MCSes are related by the concept of minimal hitting set.

Definition 5. *Given a collection of sets Γ from a universe $\mathbb{U}$, a hitting set on Γ is a set such that $\forall\ \mathcal{S} \in \Gamma, h \cap S \neq \emptyset$.*

Proposition 1. *Given a CNF formula $\mathcal{F}$, let MUSes($\mathcal{F}$) and MCSes($\mathcal{F}$) be the set of all MUSes and MCSes of $\mathcal{F}$ respectively. Then the following holds:*

1. *A subset $\mathcal{U}$ of $\mathcal{F}$ is an MUS iff $\mathcal{U}$ is a minimal hitting set of MCSes($\mathcal{F}$)*
2. *A subset $\mathcal{C}$ of $\mathcal{F}$ is an MCS iff $\mathcal{U}$ is a minimal hitting set of MUSes($\mathcal{F}$)*

3 OMUS Algorithm

Note that if we assign a unit weight to every element in the subset, we reduce the problem of finding an OMUS back to finding a SMUS.

Definition 6. *Let Γ be a collection of sets and $HS(\Gamma)$ the set of all hitting sets on Γ and let f be an valid objective function. Then a hitting set $h \in \Gamma$ is said to be an* ***optimal*** *hitting set if $\forall\ h' \in HS(\Gamma)$ we have that $f(h) \leq f(h')$.*

Property 1. *The* ***optimal*** *hitting set of a collection of sets Γ is denoted by $OHS(\Gamma)$.*

The algorithm is based on the following observation:

Proposition 2. *A set $\mathcal{U} \subseteq \mathcal{F}$ is an OMUS of $\mathcal{F}$ if and only if $\mathcal{U}$ is an optimal hitting set of MCSes($\mathcal{F}$)*

Algorithm 1 can be seen as the dual of the algorithm presented in [Davies and Bacchus, 2011]. $\mathcal{H}$ represents a collection of sets, where each set is a Minimal Correction Set on $\mathcal{F}$. The algorithm is given $\mathcal{F}$ is a CNF formula of sets, f a valid objective function and optionally a set of MCSes $\mathcal{H}_0$. Each element in $\mathcal{H}$ is a clause from $\mathcal{F}$. At the beginning, $\mathcal{H}$ is empty of initialised with a collection of MCSes. At every iteration, the optimal hitting set is computed on $\mathcal{H}$ guided by objective function f (line TODO:ref line) and all clauses are added into $\mathcal{F}'$ (line TODO:ref line). The resulting formula is tested for satisfiability at line TODO:ref line. If $\mathcal{F}'$ is satisfiable, it is grown to a maximum satisfiable subset with and its complement, MCS C is added to $\mathcal{H}$. If instead $\mathcal{F}'$ is unsatisfiable, then $\mathcal{F}'$ is guaranteed to be an OMUS of $\mathcal{F}$ as the following lemma states:

Lemma 1. *Let $\mathcal{K} \subseteq$ MCSes($\mathcal{F}$). Then a subset $\mathcal{U}$ of $\mathcal{F}$ is an OMUS if $\mathcal{U}$ is a optimal hitting set on $\mathcal{K}$ and $\mathcal{U}$ is unsatisfiable*

Proof. Since $\mathcal{U}$ is unsatisfiable it means it means it already hits every MCS in MCSes($\mathcal{F}$) 1. $\mathcal{U}$ is also an optimal hitting set on MCSes($\mathcal{F}$), since it is an optimal hitting set for $\mathcal{K}$ and no other added MCS can reduce its optimality criterion. Moreover, all other hittings sets can have their objective value stay equal or increase. Thus, by proposition 2 $\mathcal{U}$ must be an OMUS. □

Algorithm 1: OMUS($\mathcal{F}, f, \mathcal{H}_0$) [Ignatiev et al., 2015]

input : *$\mathcal{F}$ a CNF formula*
input : *f a cost function*
optional input: *$\mathcal{H}_0$ initial collection of disjoint Minimum Correction Sets*
output : *$\mathcal{OMUS}(\mathcal{F})$*

begin
    $\mathcal{H} \leftarrow \mathcal{H}_0$;
    while *true* **do**
        `// cost minimal hitting set w.r.t cost function f`
        $h \leftarrow$ `OHS`$(\mathcal{H}, f)$;
        `// set with all unique clauses from hitting set`
        $\mathcal{F}' \leftarrow \{c_i | e_i \in h\}$;
        if ***not*** *SAT($\mathcal{F}'$)* **then**
            return $\mathcal{OMUS} \leftarrow \mathcal{F}'$;
        end
        `// written as grow(`$\mathcal{F}'$`) which is Minimum Correction Set of` $\mathcal{F}'$
        `// find the biggest one with the biggest cost`
        $\mathcal{C} \leftarrow \mathcal{F} \setminus \mathcal{F}'$;
        $\mathcal{H} \leftarrow \mathcal{H} \cup \{\mathcal{C}\}$;
    end
end

For the set of clauses $\mathcal{C} = \{c_1, ...c_{|C|}\}$ with weights $\mathcal{W} = \{w_1, ...w_{|C|}\}$ in the collection of sets $\mathcal{H}$. For Example:

$$\begin{aligned} \mathcal{C} &= \{c_1, ...c_6\} \\ \mathcal{W} &= \{w_1 = 20, w_2 = 20, w_3 = 10, w_4 = 10, w_5 = 10, w_6 = 20\} \\ \mathcal{H} &= \{c_3\},\ \{c_2, c_4\},\ \{c_1, c_4\},\ \{c_1, c_5, c_6\} \end{aligned} \tag{1}$$

The optimal hitting set can be formulated as an integer linear program.

$$min \sum_{i \in \{1..|C|\}} w_i \cdot x_i \tag{2}$$

$$\sum_{i \in \{1..|C|\}} x_i \cdot h_{ij} \geq 1,\ \forall\ j \in \{1..|\mathcal{H}|\} \tag{3}$$

$$x_i = \{0, 1\} \tag{4}$$

- w_i is the input cost/weight associated with clause i in
- x_i is a boolean decision variable if constraint/clause c_i is chosen or not.
- Equation 3, h_{ij} is a boolean input variable corresponding to if constraint/clause i is in set to hit j.

4 Future Work

- TODO:@holygrail : for every literal
- TODO:@holygrail : efficient computation
- TODO:Can we derive any properties from OCS and OSS for the algorithm ?
- TODO:Can we engineer it to be faster, parallelize parts ?

Definition 7. *A* minimal correction subset $\mathcal{C}$ *of* $\mathcal{F}$ *is an* ***Optimal Correction Subset*** *of* $\mathcal{F}$*, OCS(*$\mathcal{F}$*) if* $\forall\ \mathcal{C}' \in MCSes(\mathcal{F}') : f(\mathcal{C}) \leq f(\mathcal{C}')$.

Definition 8. *A* maximal satisfiable subset $\mathcal{S} \subseteq \mathcal{F}$ *is an* ***Optimal Satisfiable Subset*** *(OSS) if* $\forall\ \mathcal{S}' \in MSSes(\mathcal{F}') : f(\mathcal{S}) \leq f(\mathcal{S}')$.

References

[Davies and Bacchus, 2011] Davies, J. and Bacchus, F. (2011). Solving maxsat by solving a sequence of simpler sat instances. In *International conference on principles and practice of constraint programming*, pages 225–239. Springer.

[Ignatiev et al., 2015] Ignatiev, A., Previti, A., Liffiton, M., and Marques-Silva, J. (2015). Smallest mus extraction with minimal hitting set dualization. In *International Conference on Principles and Practice of Constraint Programming*, pages 173–182. Springer.