

Combined Reasoning with Sets and Aggregation Functions

Markus Bender

Universität Koblenz-Landau

Institut für Informatik

Universitätsstrasse 1

56070 Koblenz

mbender@uni-koblenz.de

Abstract: We developed a method that allows to check the satisfiability of a formula in the combined theories of sets and the bridging functions *card*, *sum*, *avg*, *min*, *max* by using a prover for linear arithmetic. Since abstractions of certain verification tasks lie in this fragment, this method can be used in program verification.

1 Introduction

In [2, 1], Kuncak et al. give a sound, complete and terminating method to check the satisfiability of formulae in the theory of Boolean algebras and Presburger arithmetic (BAPA), which is equivalent to the combined theory of sets and cardinalities. They reduce the given problem to a problem in pure Presburger arithmetic and then use a prover for Presburger arithmetic. We extended this method, such that in addition to considering BAPA, we can deal with additional bridging functions between the theory of sets and the theory of Presburger arithmetic, namely the functions *sum* that calculates the sum of all elements of a set, *avg* that calculates the average of all elements of a set and *min* and *max* that calculate the minimal and maximal element of a set, respectively.

This variety of additional bridging functions allows us to deal with a broader range of verification tasks.

2 Preliminaries

As we need the concepts of atomic sets and atomic decomposition, these are introduced as follows:

For n given sets $A_1, \dots, A_n$, there are 2^n atomic sets $S_0, \dots, S_{2^n-1}$ that are defined as follows:

$$S_i = \bigcap_{j=0}^n A_j^{d(j,i)}, \text{ for } 0 \leq i < 2^n.$$

where $d(i, j)$ is the j -th binary digit of i and A^0 is defined as A and A^1 is defined as the complement of A , $\bar{A}$.

Due to the construction of the atomic sets, all S_i are disjoint.

We define that an atomic set $S_i = \bigcap_{j=0}^n A_j^{d(j,i)}$, for $0 \leq i < 2^n$ is contained in a set A if and only if there is a j such that $A_j^{d(j,i)} = A$.

Each given set A_i and its complement $\bar{A}_i$ can now be represented uniquely as *atomic decomposition*, which is the union of all atomic sets that are contained in the set.

As the developed approach relies on the method developed by Kuncak et al. [3], we give a short introduction to the latter.

3 Boolean Algebra and Presburger Arithmetic

The method of Kuncak et al. [2, 1] follows these 5 steps: a) replace equations of sets $A_1 \approx A_2$ with subset relations in both directions $A_1 \subseteq A_2 \wedge A_2 \subseteq A_1$, b) replace every subset relation between two sets $A_1 \subseteq A_2$ with a relation on the cardinality $\text{card}(A_1 \cap \bar{A}_2) \approx 0$, c) create all atomic decompositions for the sets appearing in the given formula and represent all set expressions by equivalent unions of atomic sets, d) replace cardinality of unions with sum of cardinalities of atomic sets, e) purify by introducing new constants of sort natural numbers for cardinalities of atomic sets. After these steps, a prover for Presburger arithmetic can be used to check the satisfiability of the derived formula. With this method a formula in the combined theory BAPA is reduced to a formula in Presburger arithmetic in a sound and complete way. This works nicely as the only attributes of the sets that needs to be considered is their size, i.e specific elements are not of interest. We extended this method to deal with the additional bridging functions.

4 Combined Reasoning with Sets and Aggregation Functions

In contrast to the method of Kuncak et al. [2, 1], we are considering not only the size of the sets but their elements as well. This makes the method applicable for checking the satisfiability of a formula in the combined theories of sets of numbers and the aggregation functions *sum*, *avg*, *min*, *max* more complex.

In our approach, the bridging functions *min* and *max* have to be considered together, and the handling of *avg* relies on the function *sum*, i.e. we have three different theories, called BAPA_M (for *min* and *max*), BAPA_A (for *avg*) and BAPA_S (for *sum*), or combinations thereof.

The method involves the following parts, which are explained afterwards: a) enrichment of the given formula with axioms, b) transformation of the formula to pure linear arithmetic, c) computation of a model of this formula and construction of a new formula in pure linear arithmetic from the model, and finally d) computing a model of this formula. If we have both models, we can construct a model of the original formula.

In the enrichment step, a certain set of axioms for each of the theories is added to the given formula, to model some properties of the involved theories, resulting in the so called *enriched problem*. In the transformation step, we use an extended version of the transformation of Kuncak et al. with additional steps for treating each of the aggregation functions. The result of this step is called *transformed enriched problem*. We then need a prover for linear arithmetic for checking the satisfiability of the transformed enriched problem. In Kuncak et al.'s approach, this is a prover for Presburger arithmetic, as the value for the cardinality of a set is a natural number. As we are considering not only the size of the sets, but their elements as well, and as those can be elements of $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$, we need a prover that can deal with these domains and combinations thereof. The choice of the universe affects the completeness of the method. This is discussed in a later paragraph.

With a model of the enriched formula, we have values for the size of each of the involved sets and an assignment for the different instances of the involved aggregation functions, but no assignment of the elements of the sets.

To be able to generate assignments for the elements of the sets with the help of a prover for linear arithmetic, we use the values from the constructed model to build a new formula, which we call *set constraint*. A template on how to use the values from the models is used to incorporate the properties of sets and the involved aggregation function. There is one distinct template for each of the theories $BAPA_S$, $BAPA_A$ and $BAPA_M$.

A model of the set constraint defines assignments for the sets, i.e. defines which elements are in which of the sets. Together with the assignments gathered from the model of the transformed enriched problem, we can construct a model for the original formula.

$BAPA_S$. The method for dealing with the function sum is proven to be sound, i.e. a model of the enriched transformed model and a model of the set constraint can always be used to construct a model of the original problem. Completeness is proven for the case that the model of the transformed enriched problem has the property that for all individuals of the domain c there exist infinitely many individuals a and b in the domain such that $a \neq b \wedge a + b = c$.

If this property is not fulfilled, then the method is not able to generate a model of the set constraint and therefore no model of the original formula can be constructed.

$BAPA_A$. As avg relies on sum the same considerations concerning completeness and soundness apply.

$BAPA_M$. Informal considerations lead to the result that the method for dealing with min and max is sound. If a dense ordering on the domain of the model of the transformed enriched problem exists, the method is complete. If the considered model does not have this property, the effect is equivalent to the one for the appropriate case for sum.

Termination of the method is obvious.

Combination of the three theories $BAPA_S$, $BAPA_A$ and $BAPA_M$ are possible as well. For generating the transformed enriched problem, a combination of the transformation methods is used, and the set constraint is a conjunction of the involved set constraints, i.e. the template for constructing the set constraint in the combined case is a union of the templates used for the involved theories.

5 Conclusion and Ongoing Work

We have developed a way for checking satisfiability of formulae in $BAPA_S$, $BAPA_A$, $BAPA_M$, or combinations thereof. This method simplifies certain verification task and only relies on provers for arithmetic, which are well established and reliable tools.

As a next step we will formalize the proofs of soundness, completeness and termination for the methods for reasoning in $BAPA_M$.

Based on the snippet presented in the appendix of [2], we are implementing a system that allows to use the presented methods and a prover for linear arithmetic for checking satisfiability of formulae in the presented theories.

To increase the flexibility of the developed approach, we will consider an extension that works in the following way: Instead of applying the aggregation functions on the elements of the sets, we change the method in such a way that a function f can be supplied so that the aggregation functions do not consider an element e but $f(e)$. This allows to reason about properties of elements of a set.

With this extension, we will have a verification tool with a variety of use cases in verification with data structures.

Acknowledgements

We would like to thank Viorica Sofronie-Stokkermans and Matthias Horbach for fruitful discussions.

References

- [1] Viktor Kuncak, Huu Hai Nguyen, and Martin C. Rinard. An algorithm for deciding BAPA: Boolean algebra with presburger arithmetic. In Robert Nieuwenhuis, editor, *CADE-20, Proceedings*, volume 3632 of *LNCS*, pages 260–277. Springer, 2005.
- [2] Viktor Kuncak and Martin Rinard. The first-order theory of sets with cardinality constraints is decidable. Technical Report 958, MIT CSAIL, July 2004.
- [3] Viktor Kuncak and Martin C. Rinard. Towards efficient satisfiability checking for boolean algebra with presburger arithmetic. In Frank Pfenning, editor, *CADE-21, Proceedings*, volume 4603 of *LNCS*, pages 215–230. Springer, 2007.