Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

A decidable two-sorted quantified fragment of set theory with ordered pairs and some undecidable extensions

Domenico Cantone [a], Cristiano Longo [b,*]

[a] Department of Mathematics and Computer Science, University of Catania, Italy
[b] Network Consulting Engineering, Valverde (Catania), Italy

ARTICLE INFO

Article history:
Received 22 February 2013
Received in revised form 15 December 2013
Accepted 18 March 2014
Available online xxxx

Keywords:
Decision procedures
Set theory
Undecidability

ABSTRACT

In this paper we address the decision problem for a two-sorted fragment of set theory with restricted quantification which extends the language studied in [4] with pair-related quantifiers and constructs. We also show that the decision problem for our language has a nondeterministic exponential-time complexity. However, in the restricted case of formulae whose quantifier prefixes have length bounded by a constant, the decision problem becomes NP-complete. In spite of such restriction, several useful set-theoretic constructs, mostly related to maps, are still expressible. We also argue that our restricted language has applications to knowledge representation, with particular reference to *metamodeling* issues. Finally, we compare our proposed language with two similar languages in terms of their expressivity and present some undecidable extensions of it, involving any of the domain, range, image, and map composition operators.

1. Introduction

The intuitive formalism of set theory has helped providing solid and unifying foundations to such different areas of mathematics as geometry, arithmetic, analysis, and so on. Hence, positive solutions to the decision problem for fragments of set theory can have considerable applications to the automation of mathematical reasoning and therefore in any area which can take advantage of automated deduction capabilities.

The decision problem in set theory has been intensively studied in the context of *Computable Set Theory* (see [6,12,23]), and decision procedures or undecidability results have been provided for several sublanguages of set theory. *Multi-Level Syllogistic* (in short MLS, cf. [16]) and its extension MLSS with the singleton operator were the first unquantified sublanguages of set theory that have been shown to have a solvable satisfiability problem. We recall that MLS is the Boolean combinations of atomic formulae involving the set predicates $\in$, $\subseteq$, $=$, and the Boolean set operators $\cup$, $\cap$, $\setminus$. Numerous extensions of MLS with various combinations of operators (such as powerset, unionset, etc.) and predicates (on finiteness, transitivity, etc.) have been also proved to be decidable (see [7,8,5]). Sublanguages of set theory admitting explicit quantification (see for example [4,21,22,9]) are of particular interest, since, as reported in [4], they allow one to express various set-theoretical constructs using only the basic predicates of membership and equality among sets.

The most efficient decision procedures developed in the context of Computable Set Theory have been implemented in the inferential core of the system ÆtnaNova/Referee, described in [13,20,23].

* Corresponding author.
E-mail addresses: cantone@dmi.unict.it (D. Cantone), clongo@nce.eu (C. Longo).

http://dx.doi.org/10.1016/j.tcs.2014.03.021
0304-3975/© 2014 Published by Elsevier B.V.

Recently, applications to *knowledge representation* have been investigated in [11,9], where some interrelationships between (decidable) fragments of set theory and *description logics* (see Section 5) have been exploited. Description logics are logic-based knowledge representation languages which allow one to represent knowledge about a domain of interest in terms of *concepts*, *roles*, and *individuals*, intended to denote sets of domain elements, relationships between elements, and domain elements, respectively. In contrast with the set-theoretical languages based on the von Neumann standard cumulative hierarchy of sets, recalled in Section 2, description logics do not allow *metamodeling*, namely the ability to define *meta-concepts* (i.e., concepts containing other concepts and roles) or *meta-roles* (i.e., relationships among concepts or among roles). To overcome this limitation, a novel description logic has been proposed in [19], where the three sorts of variables—concepts, roles, and individuals—have been collapsed into a unique set of *names*. This description logic comes equipped with a *model-theoretic* semantics, called ν-semantics, which however is somewhat counterintuitive, as it contradicts the regularity and the extensionality axioms of Zermelo–Fraenkel set theory.

As knowledge representation mainly focuses on representing relationships among items of a particular domain, any set-theoretical language of interest to knowledge representation should include a suitable collection of operators on *multi-valued maps*.[1]

Nondeterministic exponential-time decision procedures for two unquantified fragments of set theory involving map related constructs have been provided in [17,14]. As in both cases the map-image operator is allowed together with all the constructs of MLS, it turns out that both fragments have an ExpTime-hard decision problem (cf. [10]). On the other hand, the somewhat less expressive fragment $\mathsf{MLSS}^{\times}_{2,m}$ has been shown to have an NP-complete decision problem in [10]. We recall that $\mathsf{MLSS}^{\times}_{2,m}$ is a two-sorted language with set and map variables, which involves various map constructs like Cartesian product, map restrictions, map inverse, and Boolean operators among maps, and predicates for single-valuedness, injectivity, and bijectivity of maps.

In [4], an extension of the quantified fragment $\forall_0$ (studied in the same paper—here the subscript '0' denotes that quantification is restricted) with *single-valued* maps, the map domain operator, and terms of the form $f(t)$, with t a function-free term, was considered. Pure $\forall_0$-formulae are propositional combinations of restricted quantified prenex formulae

$$(\forall y_1 \in z_1)\cdots(\forall y_n \in z_n)p,$$

where p is a Boolean combination of atoms of the types $x \in y$, $x = y$, and *quantified variables nesting* is not allowed, in the sense that any quantified variable y_i cannot occur at the right-hand side of a membership symbol $\in$ in the same quantifier prefix (roughly speaking, no z_j can be a y_i). More recently, a decision procedure for another fragment of set theory, called $\forall_0^\pi$, has been presented in [9]. The superscript "π" denotes the presence of operators related to ordered pairs. Formulae of the fragment $\forall_0^\pi$, to be reviewed in Section 4, involve the operator $\bar\pi(\cdot)$, which computes the collection of the nonpair members of its argument, and terms of the form $[x, y]$, for ordered pairs. The predicates $=$ and $\in$ allowed in it can occur only within atoms of the following three types $x = y$, $x \in \bar\pi(y)$, and $[x, y] \in z$, whereas quantifiers in $\forall_0^\pi$-formulae are restricted to the forms $(\forall x \in \bar\pi(y))$ and $(\forall [x, y] \in z)$, and, as in the case of the fragment $\forall_0$, quantified variables nesting is not allowed.

In this paper we solve the decision problem for the extension $\forall_{0,2}^\pi$ of the fragment $\forall_0$ with map variables and ordered pairs and prove that, under particular conditions, our decision procedure runs in non-deterministic polynomial time. In addition, we prove the undecidability of various extensions of $\forall_{0,2}^\pi$ with map-related constructs.

$\forall_{0,2}^\pi$ is a two-sorted (as indicated by the second subscript "2") quantified fragment of set theory which allows restricted quantifiers of the forms

$$(\forall x \in y), \qquad (\exists x \in y), \qquad \big(\forall [x, y] \in f\big), \qquad \big(\exists [x, y] \in f\big),$$

and literals of the forms

$$x \in y, \qquad [x, y] \in f, \qquad x = y, \qquad f = g,$$

where x, y are set variables (i.e., variables ranging over sets) and f, g are map variables (i.e., variables ranging over maps).

The language $\forall_{0,2}^\pi$ extends properly the language $\forall_0$ with pair-related operators and quantifiers (see also Section 6). Thus, it is possible to express multi-valued maps in $\forall_{0,2}^\pi$, whereas the extension of $\forall_0$ presented in [4] was limited to single-valued maps only. Though the operator $\bar\pi(\cdot)$ of the $\forall_0^\pi$ language is not allowed in $\forall_{0,2}^\pi$-formulae, still it is possible to express by $\forall_{0,2}^\pi$-formulae considerably many set-theoretic constructs in a very natural way, as shown in Table 1. In fact, the language $\forall_{0,2}^\pi$ is also an extension of $\mathsf{MLSS}^{\times}_{2,m}$ (cf. [10]). As shown in detail in Section 5, it also allows one to extend the very expressive description logic $\mathcal{DL}\langle\forall_0^\pi\rangle$, presented in [9], with some metamodeling-related features, without affecting the computational complexity of the resulting logic. However, the language $\forall_{0,2}^\pi$ is not strong enough to express inclusions like $x \subseteq \mathsf{dom}(f)$, $x \subseteq \mathsf{range}(f)$, $x \subseteq f[y]$, and $h \subseteq f \circ g$, but only those in which the operators domain, range, (multi-)image, and map composition are allowed to appear on the left-hand side of the inclusion operator $\subseteq$. These expressive limitations are justified by the undecidability results provided later in the paper.

[1] According to [24], we use here the term 'maps' to denote sets of ordered pairs.

Please cite this article in press as: D. Cantone, C. Longo, A decidable two-sorted quantified fragment of set theory with ordered pairs and some undecidable extensions, Theor. Comput. Sci. (2014), http://dx.doi.org/10.1016/j.tcs.2014.03.021