Information and Computation ••• (••••) •••–•••

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Decision problems of tree transducers with origin

Emmanuel Filiot [a], Sebastian Maneth [b,*], Pierre-Alain Reynier [c], Jean-Marc Talbot [c]

[a] Université Libre de Bruxelles & FNRS, Belgium
[b] University of Bremen, Germany
[c] Aix-Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France

ARTICLE INFO

Article history:
Received 3 August 2015
Available online xxxx

Keywords:
Top-down and MSO tree transducers
Origin
Decidability
Equivalence
Injectivity
Query determinacy

ABSTRACT

A tree transducer with origin translates an input tree into a pair of output tree and origin information. The origin information maps each node in the output tree to the unique node in the input tree that created it. In this way, the implementation of the transducer becomes part of its semantics. We show that the landscape of decidable properties changes drastically when origin information is added. For instance, equivalence of nondeterministic top-down and MSO transducers with origin becomes decidable. Both problems are undecidable without origin. The equivalence of deterministic top-down tree-to-string transducers is decidable with origin, while without origin it has (until very recently) been a long standing open problem. With origin, we can decide if a deterministic macro tree transducer can be realized by a deterministic top-down tree transducer; without origin this is an open problem.

1. Introduction

Tree transducers were invented in the early 1970's as a formal model for compilers and linguistics by Thatcher [37] and Rounds [35]. They are being applied in many fields of computer science, such as syntax-directed translation [23], databases [33,25], linguistics [30,6], programming languages [40,32], and security analysis [27]. The most essential feature of tree transducers is their good balance between expressive power and decidability.

Bojańczyk [5] introduces (string) transducers with origin. For "regular" string-to-string transducers with origin he presents a machine independent characterization which admits Angluin-style learning and the decidability of natural subclasses. These results indicate that classes of translations with origin are mathematically even better behaved than their origin-less counter parts.

We initiate a rigorous study of tree transducers with origin by investigating the decidability of *equivalence*, *injectivity*, and *query determinacy* on the following models:

- top-down tree-to-tree transducers [37,35],
- top-down tree-to-string transducers [17], and
- MSO definable tree-to-string transducers, see, *e.g.*, [15].

* Corresponding author.
E-mail address: maneth@uni-bremen.de (S. Maneth).

https://doi.org/10.1016/j.ic.2018.02.011
0890-5401/

Please cite this article in press as: E. Filiot et al., Decision problems of tree transducers with origin, Inf. Comput. (2018), https://doi.org/10.1016/j.ic.2018.02.011

Table 1
Decidability of equivalence.

	Top-down tree-to-tree		Top-down tree-to-string		MSOT tree-to-string	
	det	nd	det	nd	det	nd
no origin	+ [19]	− [24]	[36]	− [24]	+ [15]	−
with origin	+	+	+	−	+	+

Unlike the string transducers of Bojańczyk [5], we will see that equivalent models of tree-to-string transducers do *not* remain equivalent in the presence of origin. This motivates the study of *subclass definability* problems (definability of a translation from a class in a subclass) when considering the origin semantics.

Table 1 summarizes our results on equivalence; the words non-/deterministic are abbreviated by nd/det and decidable/undecidable by +/−. The first change from − to + is the equivalence of nondeterministic top-down tree transducers. In the non-origin case this problem is already undecidable for restricted string-to-string transducers, as shown by Griffith [24]. In the presence of origin it becomes decidable for tree transducers, because origin implies that any connected region of output nodes with the same origin is generated by one single rule. Hence, the problem reduces to letter-to-letter transducers as studied by Andre and Bossut [1]. What about nondeterministic top-down *tree-to-string* transducers (column four in Table 1)? Here output patterns cannot be treated as letters. By deferring output generation to a leaf they can simulate non-origin translations with undecidable equivalence (*i.e.*, finite-state string transducers) [24]. Finally, let us discuss column three. Here the origin information induces a structure on the output strings: recursive calls of origin-equivalent transducers must occur in similar "blocks", so that the same children of the current input node are visited in the same order (but possibly with differing numbers of recursive calls). This block structure allows to reason over single input paths, and to reduce the problem to deterministic tree-to-string transducers with monadic input. The latter can be reduced [31] to the famous HDT0L sequence equivalence problem.

Injectivity for deterministic transducers is undecidable for all origin-free models of Table 1. With origin, we prove undecidability in the tree-to-string case and decidability in the MSO and top-down tree cases. The latter is again due to the rigid structure implied by origins. We can track if two different inputs, over the *same* input nodes, produce the same output tree. We use the convenient framework of recognizable relations to show that the set of trees for which a transducer with origin produces the same output can be recognized by a tree automaton.

We present two results on subclass definability. Recall from [14] that a deterministic top-down tree-to-string translation is MSO definable if and only if it is of linear size increase. As mentioned above, a top-down tree-to-string transducer can defer its output generation to the leaves of the input tree. For instance, it can realize the identity on monadic input trees, by producing a state call for each input node; each such state call ignores the rest of the tree until it reaches the unique input leaf. Thus, even for a translation of linear size increase, the number of output nodes with same origin can be unbounded. In contrast, for an MSO transducer the number of nodes with same origin is always bounded ("bounded origin property"). Thus, origin translations of deterministic top-down tree-to-string transducers form a strict superclass of those of MSO transducers (the same holds for the string-to-string "regular" translations studied by Bojańczyk [5]). We prove that for a given deterministic top-down tree-to-string transducer it is decidable whether an origin-equivalent MSO transducer exists; this is done by deciding the bounded origin property.

The second subclass definability result concerns a more powerful model of tree transducer: the *macro tree transducer* [18]. This model can be seen as a generalization of top-down tree transducers by adding context-parameters (of type output tree) to the states. We show that, under origin semantics, it is decidable for a given total deterministic macro tree transducer, whether it can be realized by a top-down tree transducer. This is an open problem in the non-origin setting. The proof relies on two properties: (1) the origin translation must be order-preserving, *i.e.*, the origins of output descendant nodes must be descendants, and (2) on each path of the output tree, the number of nodes with same origin must be bounded.

Motivation

Clearly, the more information we include in a translation, the more properties become decidable. Consider invertability: on the one extreme, if all reads and writes are recorded (under ACID), then any computation becomes invertible. The question then arises, how much information needs to be included in order to be invertible. This problem has recently deserved much attention in the programming language community, see, *e.g.*, [39]. Our work here was inspired by the very similar view/query determinacy problem. This problem asks for a given view and query, whether the query can be answered on the output of the view. It was shown decidable by Benedikt, Engelfriet, and Maneth [2] for views that are linear extended tree transducers, and queries that are deterministic MSO or top-down transducers. For views that include copying, the problem quickly becomes undecidable [2]. Our results show that such views *can* be supported, if origin is included. Consider for instance a view that regroups a list of publications into sublists of books, articles, etc. A tree transducer realizing this view *needs copying* (*i.e.*, needs to process the original list multiple times). Without origin, we do not know a procedure that decides determinacy for such a view. With origin, we prove that determinacy is decidable for views with origin and (origin-less) queries, where the view and query are either given by deterministic top-down transducers or by deterministic

Please cite this article in press as: E. Filiot et al., Decision problems of tree transducers with origin, Inf. Comput. (2018), https://doi.org/10.1016/j.ic.2018.02.011