On Boolean closed full trios and rational Kripke frames

Markus Lohrey[1] and Georg Zetzsche[2]

[1] Universität Siegen
lohrey@informatik.uni-leipzig.de
[2] Technische Universität Kaiserslautern
zetzsche@cs.uni-kl.de

Abstract. A Boolean closed full trio is a class of languages that is closed under Boolean operations (union, intersection, and complement) and rational transductions. It is well-known that the regular languages constitute such a Boolean closed full trio. We present a result stating that every such language class that contains any non-regular language already contains the whole arithmetical hierarchy.
Our construction also shows that there is a fixed rational Kripke frame such that assigning an arbitrary non-regular language to some variable allows the interpretation of any language from the arithmetical hierarchy in the corresponding Kripke structure.

Given alphabets X and Y, a *rational transduction* is a rational subset of the monoid $X^* \times Y^*$. For a language $L \subseteq Y^*$ and a rational transduction R, we write $RL = \{x \in X^* \mid \exists y \in L : (x, y) \in R\}$.

A class $\mathcal{C}$ of languages is called a *full trio* if it is closed under (arbitrary) homomorphisms, inverse homomorphisms, and regular intersections. It is well-known [2] that a class $\mathcal{C}$ is a full trio if and only if it is closed under rational transductions, i.e., for every $L \in \mathcal{C}$ and every rational transduction R, we have $RL \in \mathcal{C}$. We call a language class *Boolean closed* if it is closed under all Boolean operations (union, intersection, and complementation).

For any language class $\mathcal{C}$, we write $\mathsf{RE}(\mathcal{C})$ for the class of languages accepted by some Turing machine with an oracle $L \in \mathcal{C}$. Furthermore, let REC denote the class of recursive languages. Then the *arithmetical hierarchy* (see, for example, [3]) is defined as

$$\Sigma_0 = \mathsf{REC}, \quad \Sigma_{n+1} = \mathsf{RE}(\Sigma_n) \text{ for } n \geq 0.$$

Languages in $\bigcup_{n\geq 0} \Sigma_n$ are called *arithmetical*. It is well-known that the class of regular languages constitutes a Boolean closed full trio. These closure properties of the regular languages allow for a rich array of applications. Aside from the insights gained from the individual closure properties, this particular collection is exploited, for example, in the theory of automatic structures, since it implies that in such structures, every first-order definable relation can be represented by a regular language. Since emptiness is decidable for regular languages, one can therefore decide the first-order theory of these structures.

Hence, the question arises whether there are languages classes beyond the regular languages that enjoy these closure properties. Our first main result states that every such language class already contains the whole arithmetical hierarchy and thus loses virtually all decidability properties.

Theorem 1. *Let $\mathcal{C}$ be a Boolean closed full trio. If $\mathcal{C}$ contains any non-regular language, then $\mathcal{C}$ contains the arithmetical hierarchy.*

Actually, it turns out that a fixed set of rational transductions suffice to construct all arithmetical languages from any non-regular language:

Theorem 2. *There exist a fixed alphabet X and a list $R_1, \ldots, R_n \subseteq X^* \times X^*$ of rational transductions such that for every non-regular language $L \subseteq X^*$, every arithmetical language can be constructed from L using Boolean operations and applications of the rational transductions $R_1, \ldots, R_n$.*

It should be noted that Theorem 1 and 2 do not mean that there is no way of developing a theory of automatic structures beyond regular languages. It might well be that some smaller collection of closure properties suffices to obtain all first-order definable relations and still admits a decision procedure for the emptiness problem.

A large number of grammar and automata models is easily seen to produce only recursively enumerable languages. Hence, Theorem 1 also implies that the corresponding language classes are never Boolean closed full trios.

Theorem 2 can be also restated in terms of multimodal logic. A *Kripke structure* (or edge- and node-labeled graph) is a tuple

$$\mathcal{K} = (V, (E_a)_{a \in A}, (U_p)_{p \in P}),$$

where V is a set of nodes (also called worlds), A and P are finite sets of actions and propositions, respectively, for every $a \in A$, $E_a \subseteq V \times V$, and for every $p \in P$, $U_p \subseteq V$. The tuple $\mathcal{F} = (V, (E_a)_{a \in A})$ is then also called a *Kripke frame*. We say that $\mathcal{K}$ (and $\mathcal{F}$) is *word-based* if $V = X^*$ for some finite alphabet X. Formulas of *multimodal logic* are defined by the following grammar, where $p \in P$ and $a \in A$:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \Box_a \varphi \mid \Diamond_a \varphi.$$

The semantics $[\![\varphi]\!]_{\mathcal{K}} \subseteq V$ of formula φ in $\mathcal{K}$ is inductively defined as follows:

$$\begin{aligned}
[\![p]\!]_{\mathcal{K}} &= U_p, \\
[\![\neg\varphi]\!]_{\mathcal{K}} &= V \setminus [\![\varphi]\!]_{\mathcal{K}}, \\
[\![\varphi \wedge \psi]\!]_{\mathcal{K}} &= [\![\varphi]\!]_{\mathcal{K}} \cap [\![\psi]\!]_{\mathcal{K}}, \\
[\![\varphi \vee \psi]\!]_{\mathcal{K}} &= [\![\varphi]\!]_{\mathcal{K}} \cup [\![\psi]\!]_{\mathcal{K}}, \\
[\![\Box_a \varphi]\!]_{\mathcal{K}} &= \{v \in V \mid \forall u \in V : (v, u) \in E_a \to u \in [\![\varphi]\!]_{\mathcal{K}}\}, \\
[\![\Diamond_a \varphi]\!]_{\mathcal{K}} &= \{v \in V \mid \exists u \in V : (v, u) \in E_a \wedge u \in [\![\varphi]\!]_{\mathcal{K}}\}.
\end{aligned}$$

A word-based Kripke frame $\mathcal{F} = (X^*, (E_a)_{a \in A})$ is called *rational* if every E_a is a rational transduction. A word-based Kripke structure $\mathcal{K} = (X^*, (E_a)_{a \in A}, (U_p)_{p \in P})$

is called *rational* if every relation E_a is a rational transduction and every U_p is a regular language. The closure properties of regular languages imply that for every rational Kripke structure $\mathcal{K}$ and every multimodal formula φ, the set $[\![\varphi]\!]_{\mathcal{K}}$ is a regular language that can be effectively constructed from φ and (automata describing the structure) $\mathcal{K}$. Using this fact, Bekker and Goranko [1] proved that the model-checking problem for rational Kripke structures and multimodal logic is decidable. Our reformulation of Theorem 2 in terms of multimodal logic is:

Theorem 3. *There exist a fixed alphabet X and a fixed rational Kripke frame $\mathcal{F} = (X^*, R_1, \ldots, R_n)$ such that for every non-regular language $L \subseteq X^*$ and every arithmetical language $A \subseteq X^*$ there exists a multimodal formula φ such that $A = [\![\varphi]\!]_{\mathcal{K}}$, where $\mathcal{K} = (X^*, R_1, \ldots, R_n, L)$.*

References

1. W. Bekker and V. Goranko. Symbolic model checking of tense logics on rational kripke models. In M. Archibald, V. Brattka, V. Goranko, and B. Löwe, editors, *ILC*, volume 5489 of *Lecture Notes in Computer Science*, pages 2–20. Springer, 2007.
2. J. Berstel. *Transductions and context-free languages.* Teubner, Stuttgart, 1979.
3. D. C. Kozen. *Automata and computability.* Springer-Verlag, New York, 1997.