ICAI 2023
12th International Conference on Applied Informatics
Eger, Hungary, 2023

On Stateless $5' \to 3'$ WK Automata

Benedek Nagy

1: Eastern Mediterranean University, Faculty of Arts and Sciences, Department of Mathematics, Famagusta, North Cyprus, Mersin-10, Turkey
2: Eszterházy Károly Catholic University
nbenedek.inf@gmail.com

Abstract

Finite automata are one of the oldest models of computing. The classes of both deterministic and nondeterministic variants are able to accept exactly the class of regular languages. Finite automata are very popular because of they are very simple comparing them to other more sophisticated models. During the last decades, many kinds of extensions of finite automata are studied and proven to be applicable to accept larger classes of languages than the class of regular languages, but still have a moderate complexity. One of the branches of DNA computing is working with automata models accepting DNA molecules (or their formal representations), these automata are named as Watson-Crick automata [6]. These automata have two reading heads, one for each strand of the double stranded DNA. On the other hand, the strings have two extremes, namely their beginning and ends, which gives the rise of the 2-head models proceeding the input from their beginning and their end in a kind of parallel manner. Some of these 2-head models are known as $5' \to 3'$ Watson-Crick automata by a biological motivation describing these models to accept DNA molecules instead of ordinary words [1, 2]. As usual at Watson-Crick automata, various extensions/restrictions could be applied on the model, e.g., string reading feature, or having only accepting states or having only accepting states. Generally, this 2-head model of computation, by finishing the computation at latest when the heads are in the same position, characterizes exactly the class of linear context-free languages. On the other hand, opposite to the ordinary finite state automata, the deterministic counterpart of the 2-head model is weaker, and the language class 2detLIN is accepted by them [5]. Recently two more variants of the model have been investigated: In the state deterministic $5' \to 3'$ Watson-Crick automata the state of the next configuration depends only on the actual state

and it does not depend on the read symbol(s) [4]. These automata can easily be characterized by their graphs. On the other hand, in quasi-deterministic $5' \to 3'$ Watson-Crick automata, in any computation, the state of the next configuration is deterministically computed, however the configuration itself is not [3]. These automata behave somewhat between the classical deterministic and nondeterministic models of the $5' \to 3'$ Watson-Crick automata.

From biological point of view, the stateless variants, i.e., $5' \to 3'$ Watson-Crick automata with a sole state make more sense than models with several states. Thus, in the abstract, we consider these variants. Let us define formally our model:

Definition 1. A Watson-Crick finite automaton (a WK automaton) is a 5-tuple $A = (T, Q, q_0, F, \delta)$, where: T is the (input) alphabet, (e.g., the letters standing for possible bases of the nucleotides), the finite set of states Q, the initial state $q_0 \in Q$ and the set of final (also called accepting) states $F \subseteq Q$, the transition mapping δ is of the form $\delta : Q \times T^* \times T^* \to 2^Q$, such that it is non-empty only for finitely many triplets $(q, u, v), q \in Q, u, v \in T^*$.

A configuration is a pair (q, w) containing q, the current state and w, the unprocessed part of the input. In sensing $5' \to 3'$ WK automata, for any $w', x, y \in T^*, q, q' \in Q$, we write a step of the computation between two configurations as follows: $(q, xw'y) \Rightarrow (q', w')$ if and only if $q' \in \delta(q, x, y)$. We denote the reflexive and transitive closure of the relation $\Rightarrow$ by $\Rightarrow^*$. Further, for an input $w \in T^*$, an accepting computation is a sequence of steps $(q_0, w) \Rightarrow^* (q_f, \lambda)$ for some $q_f \in F$. The language accepted by a sensing $5' \to 3'$ WK automaton consists of all words that are accepted by the automaton.

A Watson-Crick finite automaton is stateless if $Q = F = \{q_0\}$. The notation NWK is used for these automata. A Watson-Crick finite automaton is simple if $\delta : (Q \times ((\lambda, T^*) \cup (T^*, \lambda))) \to 2^Q$, i.e., at most one heads read in a step; while it is one-limited if $\delta : (Q \times ((\lambda, T) \cup (T, \lambda))) \to 2^Q$, i.e., exactly one letter is being read in each step. The notation NSWK and N1WK is used for stateless simple and stateless one-limited automata, respectively.

A Watson-Crick finite automaton is deterministic, if any of its possible configurations there is at most one possible step to continue the computation. It is state-deterministic, if for each of its state $q \in Q$, if there is a transition from q and it goes to state p, i.e., $p \in \delta(q, u, v)$, then every transition from q goes to p. Further, it is quasi-deterministic, if for each possible configuration (q, w), if $(q, w) \Rightarrow (pu)$ and also $(q, w) \Rightarrow (r, v)$, then $p = r$ must hold.

As an example, the deterministic NWK accepts the language $\{0^n 1^{3n}\}$ having only transition by $(0, 111)$.

Theorem 2. *Let A be an NWK over alphabet T. Then there is a finite alphabet V and there exist two morphisms $\varphi, \mu : V \to T^*$ such that the language accepted by A can be written as $\{\varphi(x)\mu(x^R) \mid x \in V^*\}$, where x^R is the reversal of the word x.*

We may refer to φ and μ as the forward and the backward morphisms of the automaton A and its accepted language L. Clearly, for each word $w \in T^*$ accepted by the automaton, there exists $x \in V^*$ with the above property. Then, based on the words $x^i \in V^*$ we can do the following insertion/repetition.

Theorem 3. *Let A be an NWK over T. For any word w accepted by A, there is a factorisation $w = u \cdot v$, such that $u^i v^i$ is also accepted by A for any $i \in \mathbb{N}$.*

As we have seen, there are some non-regular languages that are accepted with deterministic NWK, on the other hand, the regular language a^*bba^* is not accepted by any NWK, as it does not satisfy the conditions of the previous theorem.

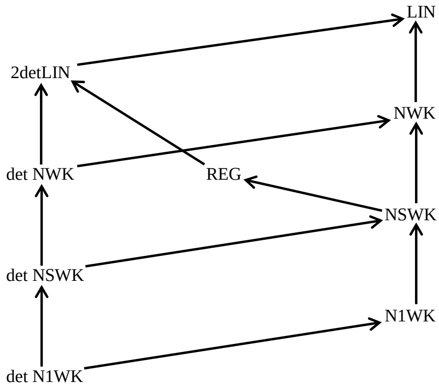

Figure 1. A hierarchy of language classes accepted by stateless sensing $5' \to 3'$ WK automata.

Figure 1 shows the hierarchy of the language classes of our model in a Hasse diagram (the automaton class here denoting the accepted language class). Classes not having directed path between them are incomparable under set theoretic inclusion relation. REG shows the class of regular, LIN, the class of linear languages (this is the class that is accepted by sensing $5' \to 3'$ WK automata) and 2detLIN the class of languages accepted by deterministic sensing $5' \to 3'$ WK automata.

References

[1] B. Nagy: *On $5' \to 3'$ Sensing Watson-Crick Finite Automata*, in: DNA13, vol. 4848, LNCS, Springer, 2008, pp. 256–262, doi: 10.1007/978-3-540-77962-9_27.

[2] B. Nagy: *On a hierarchy of $5' \to 3'$ sensing Watson-Crick finite automata languages*, J. Log. Comput. 23.4 (2013), pp. 855–872, doi: 10.1093/logcom/exr049.

[3] B. Nagy: *Quasi-deterministic $5' \to 3'$ Watson-Crick Automata*, in: Proceedings NCMA 2022, vol. 367, EPTCS, 2022, pp. 160–176, doi: 10.4204/EPTCS.367.11.

[4] B. Nagy: *State-deterministic $5' \to 3'$ Watson-Crick automata*, Nat. Comput. 20.4 (2021), pp. 725–737, doi: 10.1007/s11047-021-09865-z.

[5] B. Nagy, S. Parchami: *On deterministic sensing $5' \to 3'$ Watson-Crick finite automata: a full hierarchy in 2detLIN*, Acta Informatica 58 (2021), pp. 153–175, doi: 10.1007/s00236-019-00362-6.

[6] G. Paun, G. Rozenberg, A. Salomaa: *DNA Computing - New Computing Paradigms*, Springer, 1998, doi: 10.1007/978-3-662-03563-4.