Journées Automates Cellulaires 2010 (Turku), pp. 8-12

TILING-RECOGNIZABLE TWO-DIMENSIONAL LANGUAGES: FROM NON-DETERMINISM TO DETERMINISM THROUGH UNAMBIGUITY

DORA GIAMMARRESI

Dipartimento di Matematica, Università di Roma "Tor Vergata", via della Ricerca Scientifica, 00133 Roma, Italy
E-mail address: giammarr@mat.uniroma2.it

Abstract. Tiling recognizable two-dimensional languages, also known as REC, generalize recognizable string languages to two dimensions and share with them several theoretical properties. Nevertheless REC is not closed under complementation and the membership problem is NP-complete. This implies that this family REC is intrinsically non-deterministic. The natural and immediate definition of unambiguity corresponds to a family UREC of languages that is strictly contained in REC. On the other hand this definition of unambiguity leads to an undecidability result and therefore it cannot correspond to any deterministic notion. We introduce the notion of line-unambiguous tiling recognizable languages and prove that it corresponds or somehow naturally introduces different notions of determinism that define a hierarchy inside REC.

A *picture* (or two-dimensional string) is a two-dimensional arrays of symbols from a finite alphabet. A set of pictures is called *two-dimensional language*. Basic notations and operations can be extended from string to pictures. The size of a picture p is a pair (m, n) corresponding to the number of its rows and columns, respectively. Moreover there can be defined an operation of column-concatenation between pictures with the same number of rows and of row-concatenation between pictures with the same number of columns. By iteration, there can be also defined the corresponding row- and column- star operations.

The first generalization of finite-state automata to two dimensions can be attributed to M. Blum and C. Hewitt who in 1967 introduced the notion of a four-way automaton moving on a two-dimensional tape as the natural extension of a one-dimensional two-way finite automaton (see [7]). They also proved that the deterministic version corresponds to a language class smaller than the corresponding one defined by the non-deterministic model. Four-way automata was not a successful model since the corresponding language class does not satisfies important properties as closure under concatenation and star operations. Since then, many approaches have been presented in the literature in order to find the "right way" to generalize in 2D what regular languages are in one dimension: finite automata, grammars, logics and regular expressions (see for example [8, 14, 25, 18, 27]). Here we focus

Key words and phrases: automata and formal Languages, two-dimensional languages, tiling systems, unambiguity, determinism.

This work was partially supported by ESF Project *"AutoMathA"* (2005-2010).

on the family *REC* of *tiling recognizable picture languages* (see [14, 15] that have been widely investigated and that it is considered as a valid candidate to represent a counter part to 2D of regular string languages.

The definition of REC takes as starting point a characterization of recognizable string languages in terms of local languages and projections (cf. [11]). A picture language L is *local* if it is defined by a finite set of 2×2 pictures, called *tiles* that represent all allowed sub-pictures of size $(2, 2)$ for pictures in L. A pair composed by a local language over an alphabet Γ and an alphabetic projection $\pi : \Gamma \longrightarrow \Sigma$ is called *tiling system*. A picture language L over an alphabet Σ is recognized by a tiling system (given by a local language L' and π) if each picture $p \in L$ can be obtained as projection of a picture $p' \in L'$ (i.e. $p = \pi(p')$). A picture language is tiling recognizable if it is recognized by a tiling system. REC is the family of tiling recognizable picture languages. We point that languages of infinite picture (ω-pictures) were also studied in the setting of tiling systems in [1, 12, 13].

It can be verified that REC is closed under union and intersection, rotation and mirror and under column- and row- concatenation and star operations. Moreover, the definition of REC in terms of tiling systems turns out to be very robust: in [15, 17] it is shown that the family REC has a characterization in terms of logical formulas (a generalization of Büchi's theorem for strings to 2D). In [19], it is proved that REC has a counterpart as machine model in the *two-dimensional on-line tessellation acceptor (OTA)* introduced by K. Inoue and A. Nakamura in [18]. Other models of automata for REC are proposed in [4, 8, 24]. Tiling systems can be also simulated by domino systems [19] and Wang tiles [10] and grammars [9]. Further we remark that when pictures degenerate in strings (i.e. when considering only one-row pictures) recognizability by tiling systems corresponds exactly to recognizability by finite state string automata.

A crucial difference with the one-dimensional case lies in the fact that the definition of recognizability by tiling systems is intrinsically non-deterministic. Deterministic machine models to recognize two-dimensional languages have been considered in the literature: they always accept classes of languages smaller than the corresponding non-deterministic ones (see for example, [7, 18, 26]). This seems to be unavoidable when jumping from one to two dimensions. Further REC family is not closed under complementation and therefore the definition of any constraint to force determinism in tiling systems should necessary result in a class smaller than REC. Strictly connected with this problems are the complexity results on the recognition problem in REC. Let L be a language in REC defined by a tiling system composed by a local picture language L' and a projection π. To recognize that a given picture p of m rows and n columns belongs to L, one has to "rewrite" symbols in all positions in p to get a local picture p' that belongs to L' and such that $\pi(p') = p$. This can be done by scanning all positions of p in some order. The non-determinism implies that, once reached a given position one may eventually backtrack on all positions already visited, that is on $O(mn)$ steps. Moreover in [21] it is proved that the recognition problem for REC languages is NP-complete.

In formal language theory, an intermediate notion between determinism and non-determinism is the notion of unambiguity. In an unambiguous model, we require that each accepted object admits only one successful computation. Both determinism and unambiguity correspond to the existence of a *unique* process of computation, but while determinism is a "local" notion, unambiguity is a fully "global" one. *Unambiguous tiling recognizable* two-dimensional languages have been introduced in

[14], and their family is referred to as UREC. Informally, a picture language belongs to UREC if it admits an unambiguous tiling system, that is if every picture has a unique pre-image in its corresponding local language. In [5], the proper inclusion of UREC in REC is proved but it is also proved that it is undecidable whether a given tiling system is unambiguous. From a computational side, there are not known algorithms to recognize pictures that exploit the properties of UREC. This implies that, at each step of the recognition computation, it can be necessary to backtrack on all already visited positions.

A relevant goal is then to find subclasses for REC that inherit important properties but also allow feasible computations. Moreover an interesting result would be proving that, as for regular string languages, notions of some kind of unambiguity and determinism coincide.

Remark that another difference between unambiguity and determinism is that determinism is always related to a scanning strategy to read the input. In the string case the scanning is implicitly assumed to be left-to right and in fact deterministic automata are defined related to this direction. Moreover since deterministic, non-ambiguous and non-deterministic models are all equivalent there is no need to consider determinism from right-to-left (referred to as co-determinism). Nevertheless it is worthy to remark that not all regular string languages admits automata that are both deterministic and co-deterministic. In the two-dimensional case we have to consider all the scanning directions from left, right, top and bottom sides.

By exploiting the different possibilities of scanning for a two-dimensional array in [3, 2] there are introduced different notions of unambiguity we call here *line-unambiguity* where a line can be either a column or a row or a diagonal. We consider tiling systems for which the computations to recognize a given picture can have at each position a backtracking on at most $m + n$ steps. Such definitions lie between those of unambiguity and determinism (as long as we consider that a deterministic computation has zero backtracking steps at each position) while they all coincide with determinism when pictures degenerate in strings.

The informal definitions are very simple and natural. A tiling system is *column-unambiguous* if, when used to recognize a picture by reading it along a left-to-right or right-to left direction, once computed a local column, there is only one possible next local column. As consequence in a computation by a column-unambiguous tiling system to recognize a picture with m rows, the backtracking at each step is at most of m steps. Similarly there are defined *row-unambiguous* and *diagonal-unambiguous* tiling systems corresponding to computations that proceed by rows or by diagonals, respectively. The corresponding families of languages are denoted by *Col-UREC*, *Row-UREC* and *Diag-UREC*. In [3, 2] there are proved necessary conditions for a language to be in Col-UREC and in Row-UREC. Using such conditions one can show that families Col-UREC and Row-UREC are strictly contained in UREC. In a different set-up it is also shown that Diag-UREC is strictly included both in Col-UREC and Row-UREC. Moreover all those properties are decidable.

Very interestingly we can prove that diagonal-unambiguous tiling systems are equivalent to some deterministic tiling systems where the uniqueness of computation is guaranteed by certain conditions on the set of local tiles: the corresponding language family is denoted by DREC ([3]). Similar results hold for classes Col-UREC and Row-UREC whose union turns to be equivalent to another "deterministic" class named Snake-DREC [23]. All those classes are closed under complementation [2, 23]. As result, when we consider this line unambiguity we can prove equivalence with deterministic models and therefore we guarantee a recognition algorithm linear in the size (i.e. number of rows times number of columns) of the input.

References

[1] J.-H. Altenbernd, W. Thomas, and S. Wöhrle. Tiling systems over infinite pictures and their acceptance conditions. In *Developments in Language Theory 2002*, volume 2450 of *Lecture Notes in Computer Science*, pages 297–306. Springer, 2003.

[2] M. Anselmo, D. Giammarresi, M. Madonia. M. Anselmo, D. Giammarresi, and M. Madonia. Deterministic and unambiguous families within recognizable two-dimensional languages. *Fundamenta Informaticae*, 98(2-3):143–166, 2010.

[3] M. Anselmo, D. Giammarresi, M. Madonia. From determinism to non-determinism in recognizable two-dimensional languages. In *Procs. DLT 07*, T. Harju, J. Karhumaki and A. Lepisto (Eds.), LNCS 4588, Springer-Verlag, Berlin 2007.

[4] M. Anselmo, D. Giammarresi, M. Madonia. A computational model for tiling recognizable two-dimensional languages. *Theoretical Computer Science*, Vol. 410-37, 3520–3529 Elsevier 2009.

[5] M. Anselmo, D. Giammarresi, M. Madonia, A. Restivo. Unambiguous Recognizable Two-dimensional Languages. *RAIRO: Theoretical Informatics and Applications*, Vol. 40, 2, pp. 227-294, EDP Sciences 2006.

[6] M. Anselmo, M. Madonia. Deterministic and unambiguous two-dimensional languages over one-letter alphabet. *Theoretical Computer Science*, Vol. 410-16, 1477–1485 Elsevier 2009.

[7] M. Blum, C. Hewitt. Automata on a two-dimensional tape. *IEEE Symposium on Switching and Automata Theory*, pages 155–160, 1967.

[8] S. Bozapalidis, A. Grammatikopoulou, Recognizable picture series, *Journal of Automata, Languages and Combinatorics*, special vol. on *Weighted Automata, 2004*.

[9] S. Crespi Reghizzi and M. Pradella. Tile rewriting grammars and picture languages. *Theoretical Computer Science*, vol 340, n.2, pp. 257-272, Elsevier 2005.

[10] De Prophetis, L., Varricchio, S.: Recognizability of rectangular pictures by wang systems. Journal of Automata, Languages, Combinatorics. **2** (1997) 269-288

[11] S. Eilenberg. *Automata, Languages and Machines.* Vol. A, Academic Press, 1974.

[12] O. Finkel. On recognizable languages of infinite pictures. *Int. J. Found. Comput. Sci.*, 15(6):823–840, 2004.

[13] O. Finkel. Highly undecidable problems about recognizability by tiling systems. *Fundam. Inform.*, 91(2):305–323, 2009.

[14] D. Giammarresi, A. Restivo. Recognizable picture languages. *Int. Journal Pattern Recognition and Artificial Intelligence.* Vol. 6, No. 2& 3, pages 241 –256, 1992.

[15] D. Giammarresi, A. Restivo. Two-dimensional languages. *Handbook of Formal Languages*, G.Rozenberg, *et al.* Eds, Vol. III, pag. 215–268. Springer Verlag, 1997.

[16] D. Giammarresi, A. Restivo. Matrix-based complexity functions and recognizable picture languages. In *Logic and Automata: History and Perspectives.* E. Grader, J.Flum, T. Wilke Eds. ,pag 315-337. Texts in Logic and Games 2. Amsterdam University Press, 2007.

[17] D. Giammarresi, A. Restivo, S. Seibert, W. Thomas. Monadic second order logic over pictures and recognizability by tiling systems. *Information and Computation*, Vol 125, 1, pag 32–45, 1996.

[18] K. Inoue, A. Nakamura. Some properties of two-dimensional on-line tessellation acceptors. *Information Sciences*, Vol. 13, pages 95–121, 1977.

[19] K. Inoue, I. Takanami. A characterization of recognizable picture languages. In *Proc. Second International Colloquium on Parallel Image Processing*, A. Nakamura et al. (Eds.), LNCS 654, Springer-Verlag, Berlin 1993.

[20] M. Latteux and D. Simplot. Recognizable Picture Languages and Domino Tiling. *Theorethical Computer Science* 178(1-2): 275-283, 1997.

[21] K. Lindgren, C. Moore, M. Nordahl. Complexity of two-dimensional patterns. *Journal of Statistical Physics*, 91 (5-6), pag. 909–951, 1998.

[22] O. Matz.Regular expressions and Context-free Grammars for picture languages. *Proc. STACS'97* - LNCS 1200 pag. 283-294 - Springer Verlag 1997.

[23] V. Lonati, M. Pradella. Snake-Deterministic Tiling Systems. In *Proc. MFCS 2009*, LNCS, Vol. 5734, 549-560, Springer 2009.
[24] V. Lonati and M. Pradella. Picture-recognizability with automata based on Wang tiles. In *Proc. SOFSEM 2010*, LNCS, vol. 5901, 576-587. Springer, 2010.
[25] O. Matz. On piecewise testable, starfree, and recognizable picture languages. In *Foundations of Software Science and Computation Structures*, M. Nivat Ed., vol. 1378, Springer, 1998.
[26] A. Potthoff, S. Seibert, W. Thomas. Nondeterminism versus determinism of finite automata over directed acyclic graphs. *Bull. Belgian Math. Soc.* 1, 285–298, 1994.
[27] R. Siromoney. Advances in array languages. In *Graph-Grammars and Their Applications to Computer Science*, Ehrig et al. (Eds.), pages 549–563. Lecture Notes in Computer Science 291, Springer-Verlag, Berlin, 1987.