An automata characterisation for weighted multiple context-free grammars

Tobias Denkinger

Faculty of Computer Science, Technische Universität Dresden, 01062 Dresden, Germany
tobias.denkinger@tu-dresden.de

Multiple context-free grammars (MCFGs) and equivalent formalisms are currently prominent in natural language processing as they are able to express the non-projective constituents and dependencies present in natural language yet remain parsable in polynomial time. Although there are many parsing algorithms [3, Sec. 7] for MCFGs, no automaton model has been proven equivalent to them yet. Thread automata, introduced by Villemonte de la Clergerie [4], already come close to such an automaton model, and a construction of thread automata from ordered simple range concatenation grammars (which are equivalent to multiple context-free languages) was given [4, Sec. 4]. Based on the idea of thread automata, we introduce *tree stack automata* that possess, in addition to the usual finite state control, the ability to manipulate a tree-shaped stack which has the tree's root at its bottom. Our recent investigations of tree stack automata produced the following results:

(R1) We can show that a restricted form of tree stack automata is equivalent to MCFGs.

(R2) For a complete commutative strong bimonoid $\mathcal{A}$, we can show that a restricted form of $\mathcal{A}$-weighted tree stack automata is equivalent to $\mathcal{A}$-weighted MCFGs.

R1 is achieved by construction. For R2 we use weight separation for both weighted automata with storage [2, Thm. 6] and weighted MCFGs [1, Lem. 15].

References

[1] T. Denkinger. "A Chomsky-Schützenberger representation for weighted multiple context-free languages". In: *Proc. FSMNLP'15*. 2015. URL: http://aclweb.org/anthology/W/W15/W15-4803.pdf.

[2] L. Herrmann and H. Vogler. "A Chomsky-Schützenberger Theorem for Weighted Automata with Storage". In: *Proc. CAI'15*. Springer, 2015, pp. 90–102. DOI: 10.1007/978-3-319-23021-4_11.

[3] L. Kallmeyer. *Parsing beyond context-free grammars*. 2010. DOI: 10.1007/978-3-642-14846-0.

[4] É. Villemonte de la Clergerie. "Parsing Mildly Context-Sensitive Languages with Thread Automata". In: *Proc. COLING'02*. ACL, 2002, pp. 1–7. DOI: 10.3115/1072228.1072256.