

Linard, Alexis; de la Higuera, Colin; Vaandrager, Frits

Learning unions of k -testable languages. (English) [Zbl 1425.68150](#)

Martín-Vide, Carlos (ed.) et al., Language and automata theory and applications. 13th international conference, LATA 2019, St. Petersburg, Russia, March 26–29, 2019, Proceedings. Cham: Springer. Lect. Notes Comput. Sci. 11417, 328-339 (2019).

Summary: A classical problem in grammatical inference is to identify a language from a set of examples. In this paper, we address the problem of identifying a union of languages from examples that belong to several *different* unknown languages. Indeed, decomposing a language into smaller pieces that are easier to represent should make learning easier than aiming for a too generalized language. In particular, we consider k -testable languages in the strict sense (k -TSS). These are defined by a set of allowed prefixes, infixes (sub-strings) and suffixes that words in the language may contain. We establish a Galois connection between the lattice of all languages over alphabet Σ , and the lattice of k -TSS languages over Σ . We also define a simple metric on k -TSS languages. The Galois connection and the metric allow us to derive an efficient algorithm to learn the union of k -TSS languages. We evaluate our algorithm on an industrial dataset and thus demonstrate the relevance of our approach.

For the entire collection see [\[Zbl 1423.68031\]](#).

MSC:

- [68Q32](#) Computational learning theory
- [06A15](#) Galois correspondences, closure operators (in relation to ordered sets)
- [68Q45](#) Formal languages and automata

Keywords:

grammatical inference; k -testable languages; union of languages; Galois connection

Full Text: [DOI](#) [arXiv](#)