Parker, Austin J.; Yancey, Kelly B.; Yancey, Matthew P.
Definitions and properties of entropy and distance for regular languages. (English)

Zbl 07154763

Summary: This paper addresses a need that has arisen for constructing a practical and intuitive distance function over regular languages. Some of the previously constructed distance functions are not well-defined; and we construct examples showing when they fail to meet intuitive expectations. We present thorough mathematical analysis and general reasoning for how and why our definitions fix these issues.

As regular languages are sets, most proposed distances between regular languages \( L_1 \) and \( L_2 \) are based on the “size” of \( L_1 \triangle L_2 \). As a related issue, this paper also addresses the need to construct a practical and intuitive function that describes the size of a regular language, which we name language entropy.

There exists a well-known map from the space of regular languages to sofic shifts, but theorems about sofic shifts have only been pulled back to the space of regular languages under strong assumptions. One of the contributions of this paper is a new map from regular languages to sofic shifts; in this new map the entropy of a regular language equals the topological entropy of the associated sofic shift.

For the entire collection see [Zbl 1426.37005].

MSC:

37B10 Symbolic dynamics
68Q45 Formal languages and automata
37B40 Topological entropy
68Q70 Algebraic theory of languages and automata

Keywords:
deterministic finite automata; sofic shift; regular language

Full Text: DOI

References:
[7] Chomsky, Noam; Miller, George A., Finite state languages, Information and Control, 1, 91-112 (1958) · Zbl 0081.14503


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.