

Descriptive Generalization of Subregular Languages

Dominik D. Freydenberger

Bex et al. 2010, as well as others, consider the learnability of certain subregular language classes which typically occur in XML specifications, the so called *chain regular expressions*, CHAREs, and *single occurrence regular expressions*, SOREs.

This previous literature contains a fast algorithm for generalizing finite sample sets to CHAREs, and a number of algorithms for generalizing to SOREs, providing a trade off between generalization speed and quality of the solution. For each of these algorithms, there exist sample sets which do not contain sufficient information to suggest a “best” expression. In these cases, the algorithms return an expression that generalizes the sample, but is not minimal for the given language class.

We address this problem in the framework of *Descriptive Generalization*, which was introduced by Freydenberger and Reidenbach 2010. For each of the two language classes, we give an *efficient* algorithm that returns a *minimal* generalization from the given finite sample to an element of the respective language class; such generalizations are called *descriptive*.

This is a joint work with Timo Kötzing (MPI Saarbrücken).