Probabilities of Regular Languages of Infinite Trees

Paweł Parys

In a recent publication The Probabilistic Rabin Tree Theorem we have shown how to compute, given a regular language L of infinite trees, the probability that a random infinite tree belongs to L. The problem is reduced to solving a system of polynomial equations over real numbers (which may have multiple solutions), together with a formula in first-order logic over real numbers saying which solution of the system of equations should be taken.

The goal of the project is to find interesting examples: languages L for which the solution is nontrivial. We miss interesting examples.

We have easy examples like "the label of the root is a"; the probability of this event is 1/(alphabet size). A warm-up question is: can we obtain any rational number (in the interval $[0,1]$) as the probability, when the alphabet size is 2? Probably we can...

A similar question is: can we obtain any algebraic number in $[0,1]$? An algebraic number is (by definition) a root of a polynomial, and we expect that the language L can be created so that the probability is indeed a solution of this polynomial. This might be easier when all coefficients of this polynomial are in $[0,1]$, and when it has precisely one root in the interval $[0,1]$. What if not? Maybe some algebraic numbers cannot be obtained as probabilities?

Another (maybe more interesting) goal is to obtain a language L where the system of polynomial equations (that we obtain from the reduction) has indeed multiple solutions, and the first-order formula that we obtain selects the appropriate solution in a nontrivial way. This choice corresponds to behaviour of the language "in the limit", as opposed to properties holding locally near the root. However typical properties talking about limitary behaviour (like "there are infinitely many labels a in the tree) have usually probability 0 or 1, so they are not really interesting in our context. It would be nice to find a "limitary" property that really matters for the probability. Or maybe such properties do not exist, and the whole algorithm can be simplified? Maybe this can lead to establishing a better upper bound for the complexity of the algorithm...

Summing up, this project is purely theoretical, it deals with regular languages of infinite trees, and has several possible goals (some might be easier, other more difficult). There is no concrete theorem to be proven; the goal is rather to "better understand the situation".

The offer is addressed to students strong in theoretical aspects of computer science, who feel comfortable with topics such as automata theory or logic. On the other hand, no prior knowledge about automata on infinite trees is needed; this can be quickly learned during the internship.