

Zastosowanie algorytmów genetycznych do generowania gramatyk bezkontekstowych z próbek słów

Adrianna Gietka
Instytut Informatyki
Uniwersytet Gdański

Matematyczna teoria wyuczalności jest działem informatyki teoretycznej zajmującym się poszukiwaniem metod wyuczania języków. Wyuczanie języka polega na określeniu gramatyki, która wyprowadza dany język. Poszukiwane mechanizmy wyuczania opierają się na modelu uczenia języków naturalnych: ludzie uczą się języka, którym posługuje się ich otoczenie poprzez przyswajanie i analizowanie przykładów pozytywnych (słów z języka) i negatywnych (słów, które do danego języka nie należą). W procesie naturalnego wyuczania języka bierze również udział postać nauczyciela, który znając gramatykę naszego języka potrafi wydać osąd czy dane słowo należy czy nie do wyuczanego języka. Cały cykl nauki kończy się sprawdzeniem czy gramatyka przez nas przedstawiona wyprowadza analizowany przez nas język.

W 2004 roku Sakakibara zaproponował metodę identyfikacji języka bezkontekstowego. Algorytm przez niego zaproponowany wykorzystuje strukturę tablicową analogiczną do zastosowanej w *algorytmie CYK weryfikacji przynależności słowa do języka* oraz *algorytm genetyczny* w celu znalezienia gramatyki bezkontekstowej w określonej postaci normalnej Chomsky'ego o minimalnej liczbie stanów. Metoda wyuczania klasy języków bezkontekstowych przedstawiona przez Y.Sakakibara wygląda bardzo obiecująco ze względu na ograniczoną przestrzeń rozwiązań problemu i uniwersalność dla każdego języka z klasy. Przestrzeń poszukiwań rozwiązania problemu jest niestety jeszcze zbyt duża, żeby można mówić o dużej skuteczności algorytmu, mimo pozytywnych wyników przedstawionych przez autora algorytmu.

W pracy zostanie przedstawiona zmodyfikowana wersja algorytmu bazującego na metodzie Sakakibara. Modyfikacje obejmują następujące zmiany w algorytmie genetycznym:

- zastosowane zostały nowe struktury osobników ograniczone do wywodów pojedynczych ścieżek wyvodu słów ze zbioru przykładów pozytywnych,
- w procesie generowania populacji początkowej wykorzystane zostały pewne własności charakterystyczne dla języków bezkontekstowych,
- zastosowane zostały nowe operacje w algorytmie genetycznym (starzenie się osobników, funkcja epidemii, mutacja „nowa ścieżka”),
- zmodyfikowana została funkcja przystosowania osobnika.

Powyższe modyfikacje pozwalają na przyspieszenie procesu generowania gramatyki, co więcej, pozwalają na zastosowanie algorytmu dla bogatszych języków.