

Nina Gierasimczuk

Algorytmiczne podejście do problemu uczenia się języka

Jednym z wielkich problemów epistemologii jest wyjaśnienie zjawiska przyswajania wiedzy i umiejętności, a w szczególności nabywania zdolności językowych. W jaki sposób człowiek uczy się swego pierwszego języka? Jak to możliwe, że korzystając ze szczątkowych i często wadliwych danych udaje mu się wykryć strukturę języka na tyle poprawnie, by móc się skutecznie porozumiewać? W jaki sposób zdobywa umiejętność konstruowania zdań, z którymi nigdy wcześniej się nie zetknął? Jak wreszcie opisać fragment umysłu odpowiadający za zdolności językowe, aby wyjaśnić fakt, że człowiek równie dobrze może jako pierwszego języka nauczyć się chińskiego jak i polskiego? Z tymi pytaniami mierzył się w swoich badaniach Noam Chomsky — twórca współczesnego paradygmatu badań nad językiem. Postulował on istnienie wrodzonego mechanizmu umysłowego, umożliwiającego nabywanie zdolności językowych. W pracach *Syntactic Structures* oraz *The Logical Structure of Linguistic Theory* zaproponował generatywistyczny model języka, który umożliwia analizowanie składni. Problem wyuczalności języka podjął E.M. Gold w pracy *Identification in the Limit* (Gold, 1967). Jego propozycja daje wyobrażenie o możliwym kształcie wrodzonego mechanizmu nabywania zdolności językowych.

W moim wystąpieniu skoncentruję się głównie na problemie nabywania wiedzy składniowej - przedstawię model identyfikowalności w granicy (Gold, 1967), który wyznacza zakres ogólnie pojętej wyuczalności.

Człowiek nie uczy się pierwszego języka poprzez zapamiętywanie reguł gramatycznych podanych *explicite*. Odbywa się to przez pewien rodzaj generalizacji zasłyszanych elementów języka. Od algorytmu uczącego się oczekujemy tego samego— przyswajania zbioru reguł gramatycznych na podstawie szczątkowych danych, przypadkowych elementów języka. Model identyfikacji w granicy ukazuje możliwy przebieg tego procesu. Model działa dla ustalonej klasy języków. Zostaje z niej wybrany jeden język. Uczeń otrzymuje informacje na jego temat. Ukazane są one na jeden z możliwych sposobów — według przyjętej metody prezentacji danych. Zadanie ucznia polega na odgadnięciu nazwy wybranego języka. Nazwami języków są ich opisy, na przykład gramatyki. Model identyfikacji w granicy polega więc na wyszukiwaniu odpowiedniej gramatyki opisującej zaprezentowany ciąg informacji dotyczącej wybranego języka. W każdym kolejnym kroku procedury uczniowi prezentowana jest jednostka danych dotycząca nieznanego języka. Każdorazowo uczeń ma do dyspozycji tylko skończony zbiór danych. W każdym kroku uczeń wybiera nazwę języka. Ta procedura przebiega w nieskończoność. Język jest identyfikowany w granicy, gdy po pewnym czasie uczeń wybiera poprawnie ciągle tę samą nazwę

(gramatykę). Właściwa identyfikowalność dotyczy klas języków. Cała klasa języków jest identyfikowalna w granicy, gdy istnieje algorytm zgadujący (uczeń) taki, że identyfikuje on w granicy dowolny język z tej klasy. Warto zaznaczyć, że uczeń nie ma dostępu do informacji, kiedy jego zgadnięcia są prawidłowe. Przetwarza informacje w nieskończoność, ponieważ nie może stwierdzić, czy w kolejnym kroku nie dostanie informacji, która zmusi go do zmiany wyboru.

Model uczenia się zaproponowany przez Golda to propozycja, w której:

1. Uczenie się jest procesem nieskończonym.
2. Efektem uczenia się języka jest ustabilizowanie ciągu wyników na jednej, poprawnej gramatyce.
3. Pojęcie identyfikowalności w granicy dotyczy określonych klas języków.
4. Uczeń posiada umiejętność testowania ciągów za pomocą automatów i generowania ciągów za pomocą gramatyk.
5. Uczenie się języka nie wymaga istnienia jakichkolwiek wrodzonych uczniowi struktur gramatycznych.
6. Żadna klasa, która mogłaby zawierać język naturalny, nie jest wyuczalna z danych pozytywnych (tekstu).

W dalszej części wystąpienia zaprezentuję algorytmiczny model uczenia się przez zapytania (Angluin, 1987), (Sakakibara, 1990). Opiszę algorytm L^* , który uczy się języków regularnych oraz algorytm LA uczący się gramatyk bezkontekstowych. Oba algorytmy wykorzystują w uczeniu się jednocześnie negatywne i pozytywne dane na temat poszukiwanego języka – kładą nacisk na udział nauczyciela w procesie przyswajania języka. Po dokonaniu prezentacji omówionych wyżej modeli skupię się na analizie metodologicznej ich założeń. Zbadam, w jakim stopniu koncepcje te wyjaśniają naturalny proces uczenia się języka. Podkreślę znaczenie modelowania procesu przyswajania języka dla epistemologii. Końcowa część wystąpienia poświęcona będzie problemowi użyteczności omówionych w pracy narzędzi do modelowania procesu przyswajania semantyki języka.

Bibliografia

- Angluin, D. (1987): "Learning regular sets from queries and counterexamples", *Information and Computation* 75/2(1987), 87–106.
- Chomsky, N. (1957): "Syntactic structures", Mouton, 1957.
- Gold, E. M. (1967): "Language identification in the limit", *Information and Control* 10(1967), 447–474.
- Sakakibara, Y. (1990): "Learning context-free grammars from structural data in polynomial time", *Theoretical Computer Science* 75(1990), 223 – 242.