

# Głębokie sieci neuronowe w automatycznym dowodzeniu twierdzeń

~STRESZCZENIE POPULARNONAUKOWE~

Bartosz Piotrowski

Logiczne podwaliny nowoczesnej matematyki zostały położone na przełomie XIX i XX w. Prace Fregego, Russela, czy Whiteheada dały nam pojęcie dowodu matematycznego jako formalnego rachunku logicznego. Późniejsze nastanie doby komputerów zrodziło dziedzinę automatycznego dowodzenia twierdzeń (*automated theorem proving* - ATP). Systemy ATP, realizujące zupełny rachunek dowodowy wzbogacony o heurystyki przeszukiwania, co do zasady mogą zaatakować dowolny formalnie postawiony problem matematyczny. W praktyce jednak, systemy te są dużo słabsze, niż matematycy.

Niedawno w systemach ATP zaczęto stosować paradygmat uczenia maszynowego. To zbliża pewne aspekty tychże systemów do ludzkiego sposobu uprawiania matematyki – ludzie dowodząc twierdzenia nabywają *intuicję*, która pomaga wybierać obiecujące kroki podczas konstruowania dowodu. Wydaje się rozsądnym korzystać z tego rodzaju *uczenia się* również w systemach ATP. Istotnie, rosnąca liczba eksperymentów wedle tej idei zdaje się potwierdzać, że jest to owocny kierunek badawczy. Duże biblioteki sformalizowanej matematyki zostały przetłumaczone na formalizm systemów ATP, zapewniając adekwatne zbiory danych treningowych, na których zaczęto trenować różne modele uczenia maszynowego. W szczególności, zaczęto stosować do tego celu metody głębokich sieci neuronowych.

Głębokie sieci neuronowe, jest to aktywnie rozwijana gałąź uczenia maszynowego. Dzięki nim dokonano przełomów w wielu różnych dziedzinach, m.in. w rozpoznawaniu obrazu czy też przetwarzaniu języka naturalnego. Zastosowanie głębokich sieci neuronowych do rozwiązywania problemów formalnych jest tematem, który nie przyciągnął jeszcze uwagi wielu badaczy, jednak kilka pionierskich prac z ostatnich lat prezentuje ciekawe, inspirujące wyniki.

Głębokie sieci neuronowe posiadają właściwości, które wydają się być szczególnie istotne w kontekście wnioskowania formalnego. Realizują one *uczenie hierarchiczne* – są w stanie wewnętrznie reprezentować złożone koncepty jako kombinację prostych wzorców. Na przykładzie rozpoznawanie obrazu jest to, na przykład, reprezentowanie konceptu słonia jako pewnych kombinacji prostych wzorców, jak linie. Znaczy to, że głębokie sieci neuronowe, ogólnie mówiąc, wykazują zdolność do przetwarzania surowych, prostych reprezentacji w abstrakcyjne reprezentacje o *semantycznym* charakterze. W logice istnieje kluczowe rozróżnienie między *składnią* (surową reprezentacją formuł), a *semantyką* (ich znaczeniem). Niewielka zmiana w składni formuły może kompletnie zmienić jej semantykę, jak również dwa bardzo różne wyrażenia logiczne mogą znaczyć to samo. Byłoby pożądane mieć modele uczenia maszynowego w jakimś stopniu poprawnie operujące na tym rozróżnieniu pomiędzy składnią i semantyką formuł.

Celem projektu, motywowanym przez powyższe uwagi, jest **systematyczne, dokładne zbadanie stosowalności głębokich sieci neuronowych w automatycznym dowodzeniu twierdzeń.**

Plan projektu zakłada dwie fazy. W pierwszej z nich zostanie przeprowadzone dokładne badanie poparte przez eksperymenty o elementarnym, izolowanym charakterze. Ich celem będzie zrozumienie, jakie są możliwości i ograniczenia głębokich sieci neuronowych co do uczenia się semantyki formuł i różnych relacji logicznych między nimi, takich jak wynikanie logiczne czy spełnianie w modelu. W drugiej fazie, w której wykorzystana będzie ekspertyza zdobyta w fazie pierwszej, zostanie przeprowadzone zintegrowanie istniejących systemów ATP z metodami głębokich sieci neuronowych. W szczególności zostanie zbadane, jak głębokie sieci neuronowe nadają się do zadania selekcji przesłanek (wybierania faktów z formalnej biblioteki, które są istotne do dowiedzenia danego twierdzenia) oraz do wewnętrznego kierowania krokami podejmowanymi przez system ATP podczas szukania dowodu.

Opisany projekt badawczy może przyczynić się zarówno dla uczenia maszynowego, jak i automatycznego dowodzenia twierdzeń. W przypadku tego pierwszego, osiągnięte rezultaty pomogą zrozumieć możliwości i ograniczenia głębokich sieci neuronowych. W przypadku tego drugiego, mamy nadzieję na ulepszenie istniejących metod ATP i spopularyzować stosowanie dla nich paradygmatu uczenia maszynowego. Jakikolwiek ulepszenia tutaj naturalnie przenoszą się na pokrewne dziedziny, takie jak interaktywne dowodzenie twierdzeń czy też weryfikację formalną – duże projekty formalizacyjne związane z nimi, zarówno matematycznymi, jak i przemysłowymi. W projektach tych często jest wiele małych uciążliwych kroków dowodowych, dla których systemy ATP są efektywnymi narzędziami. Ulepszenie systemów ATP umożliwi przeprowadzanie większych projektów formalizacyjnych, które z kolei są źródłem danych treningowych dla uczenia maszynowego – co formuje pozytywne sprzężenie zwrotne.