

Mateusz Perlik

Wydział Matematyki, Informatyki i Mechaniki

Cybermatematyk, czyli o automatycznym dowodzeniu twierdzeń

W listopadzie ubiegłego roku opublikowany został ChatGPT – chatbot stworzony przez OpenAI¹. Od razu zrobił furorę w internecie dzięki temu, jak znakomicie potrafi odpowiadać na pytania z wielu dziedzin. Jednym z zagadnień, z którymi sobie jednak zbyt dobrze nie radzi, jest matematyka. Z jednej strony często nie potrafi poprawnie wykonywać podstawowych działań arytmetycznych nawet na kilkucyfrowych liczbach. Z drugiej zaś nieraz przeprowadza niepoprawne dowody stwierdzeń (zarówno prawdziwych, jak i fałszywych) z matematyki szkolnej i akademickiej. Nie powinno to jednak być wielkim zaskoczeniem, albowiem ChatGPT nie jest do takich zadań przystosowany. Do przeprowadzania rozumowań matematycznych konieczna jest jakaś wewnętrzna reprezentacja symboli występujących w formułach tak, by można było nimi odpowiednio manipulować, prowadząc ostatecznie do poszukiwanego dowodu. Nasuwa się jednak pytanie: jak szybko i w jakim stopniu komputery będą w stanie zastąpić matematyków?

Na początek przydałoby się ustalić, czym tak właściwie jest dowód. We współczesnej matematyce działa się w systemach aksjomatycznych. Oznacza to, że ustalamy pewien zbiór założeń zwanych aksjomatami oraz reguły wnioskowania, które pozwalają na manipulację formułami. Przez dowód zaś rozumie się skończony ciąg zdań takich, że każde z nich jest albo aksjomatem, albo wynika z poprzedzających go zdań za pomocą reguł wnioskowania, przy czym ostatecznie zdanie jest dowodzonym twierdzeniem.

Warto nadmienić, że wykorzystanie komputerów do dowodzenia twierdzeń nie jest niczym nowym w matematyce. Pierwszym takim przypadkiem o istotnym znaczeniu był dowód twierdzenia o czterech barwach opublikowany w 1976 roku². Polegał on na sprowadzeniu tezy twierdzenia do ponad tysiąca przypadków, które należało rozważyć – właśnie tym zajął się komputer. Już wtedy powodowało to kontrowersje w środowisku matematycznym, gdyż wygenerowany dowód był tak obszerny, że nie sposób było go przeczytać w sensownym czasie i zweryfikować jego poprawność. Jednym z najnowszych przykładów dowodów wspomaganych komputerowo jest rozstrzygnięcie problemu boolowskich trójek pitagorejskich³. Nieskompresowany dowód ważył aż 200 terabajtów! Ręczne zweryfikowanie poprawności tego dowodu jest niemożliwe, nawet jeśli człowiek poświęciłby na to całe swoje życie.

Powyższe przykłady pokazują jedynie komputery w roli wykonywania bardzo wielu żmudnych obliczeń, które stanowią integralną część dowodu. Innym istotnym wykorzystaniem komputerów są systemy wspomagające dowodzenie twierdzeń. Najważniejszym przykładem takiego programu jest Coq⁴, którego pierwsza wersja powstała już w 1989 roku. Jest to tak naprawdę język programowania, w którym użytkownik wprowadza twierdzenie, które chce udowodnić, a następnie podaje *taktyki*, które odpowiadają kolejnym krokom dowodowym. Taktyki te to zazwyczaj odpowiedniki reguł dedukcji. Co istotne, Coq może wykonać jedynie

¹ <https://chat.openai.com>

² https://en.wikipedia.org/wiki/Four_color_theorem

³ https://en.wikipedia.org/wiki/Boolean_Pythagorean_triples_problem

⁴ <https://coq.inria.fr/>

przejścia poprawne w danym kontekście, zatem dowód wykonany w Coq jest na pewno poprawny (przy odpowiednich założeniach, w szczególności systemu logicznego, w którym dowód jest przeprowadzany), więc jest to znakomite narzędzie do weryfikacji poprawności już napisanych dowodów. Istnieją również taktyki, które służą automatycznemu dowodzeniu. Gdy się je stosuje, Coq stara się przeprowadzić dowód, próbując wykonać ciąg odpowiednich taktyk podstawowych. Możliwości automatycznych taktyk są jednak istotnie ograniczone.

Zadaniem systemu automatycznego dowodzenia twierdzeń (ATP, ang. automated theorem prover) jest znalezienie dowodu dla podanego zdania w ustalonym systemie aksjomatycznym. Coq ma więc podstawową funkcjonalność ATP. Istotne jest to, że system aksjomatyczny, w którym dowód ma zostać przeprowadzony, w znaczącym stopniu determinuje to, jak trudnym zadaniem staje się jego znalezienie. Przykładowo ustalenie, czy dana formuła w klasycznym rachunku zdań jest tautologią (tj. jest prawdziwa przy każdym wartościowaniu zmiennych zdaniowych) jest problemem rozstrzygalnym – wystarczy sprawdzić, czy wszystkie wartościowania powodują, że formuła ta staje się prawdziwa (algorytm ten jest niestety wykładniczy względem liczby zmiennych zdaniowych). Z drugiej strony, jeśli mamy do czynienia z systemem, w którym zawarta jest aksjomatyka liczb naturalnych, to z twierdzenia Gödla o niezupełności istnieje w nim zdanie takie, że nie da się udowodnić ani jego, ani jego zaprzeczenia (o ile się wierzy w niesprzeczność tegoż systemu). Jednakże kwestia ta jest problemem raczej teoretycznym – większość stwierdzeń, które się spotyka w praktyce, będą dowodliwe (lub ich zaprzeczenie).

Podstawowym problemem automatycznego dowodzenia twierdzeń jest: jak właściwie zabrać się za znalezienie dowodu? Zazwyczaj nie jest tak, że wystarczy siłowo sprawdzić pewną skończoną liczbę przypadków – przestrzeń stanów, w których można się znaleźć w trakcie przeprowadzania dowodu, jest najczęściej nieskończona. Jednakże prawie wszystkie z nich są bez sensu, tj. nie przybliżają do ukończenia dowodu. Z tego powodu stosuje się algorytmy heurystyczne (tzn. bez gwarancji sukcesu), które mają pomóc w ustaleniu tych bardziej sensownych ścieżek, którymi można w dowodzie podążać.

W ostatnich latach do gry wkroczyło uczenie maszynowe. Szybki rozwój tej dyscypliny, w szczególności w przetwarzaniu języka naturalnego oraz w grach (np. w szachach) pomógł w wytwarzaniu lepszych ATP⁵. Głównym usprawnieniem jest dobór heurystyk – kiedyś zazwyczaj były wymyślane przez ludzi, teraz sztuczna inteligencja jest w stanie się ich dużo lepiej nauczyć. W ten sposób (z grubsza) ATP stworzonemu ostatnio przez zespół Meta AI udało się rozwiązać aż 10 zadań pochodzących z Międzynarodowej Olimpiady Matematycznej⁶. Te zadania stanowią wyzwanie dla uczniów szkół średnich z całego świata. Jest to więc ogromny sukces automatycznego dowodzenia, nawet jeśli są to jedne z łatwiejszych olimpijskich zadań.

Do utworzenia ATP mogącego konkurować ze współczesnymi matematykami jeszcze daleka droga. Wyzwaniem w tej dziedzinie pozostaje radzenie sobie modelów z rozbijaniem zadania na kilka prostszych zadań, co jest zazwyczaj kluczowe w „ludzkich” rozumowaniach matematycznych. Jednakże dowody pisane przez ludzi są pewnym podzbiorem wszystkich możliwych dowodów. Charakteryzują się one zazwyczaj tym, że są dosyć zwarte i ustrukturyzowane, co wynika po części z tego, że ludzie mają słabą pamięć „podręczną” – komputery nie mają takiego problemu. Inną charakterystyczną rzeczą, którą robią ludzie w procesie wymyślania dowodu, jest myślenie o problemie w sposób uproszczony – z

⁵ <https://www.quantamagazine.org/how-close-are-computers-to-automating-mathematical-reasoning-20200827/>

⁶ <https://ai.facebook.com/blog/ai-math-theorem-proving/>

doświadczenia wiedzą, co w nim jest istotne, a co nie. Ponadto przykładowo w geometrii ludzie mają wyrobione intuicje dzięki obserwacji otaczającego nas świata. Intuicje te potrafią być jednak zawodne, gdy zagadnienie dotyczy przestrzeni o większym wymiarze niż 3, choć i w takiej sytuacji mogą okazać się cenne. Wydaje się, że te kwestie (i nie tylko one) mogą się okazać bardzo problematyczne dla rozwoju tej dziedziny. Z drugiej strony może dzięki tym różnicom komputery będą w stanie stosować techniki dowodzenia, które byłyby niewykonalne dla nas ze względu na ludzkie ograniczenia? Aby to osiągnąć, trzeba by postawić większy nacisk na „samodzielne” uczenie się tych modeli, niż na próbach kopiowania ludzkiego myślenia o matematyce. Tak czy inaczej, prowadzenie dowodu wymaga pewnej dozy kreatywności matematycznej, której współczesnym modelom sztucznej inteligencji niewątpliwie nadal brakuje.

Trudno przewidzieć, co dokładnie nas czeka, ale szybki rozwój automatycznego dowodzenia w ostatnich latach pozwala przypuszczać, że w nieodległej przyszłości zauważymy zwiększenie znaczenia tej dziedziny w matematyce. Nie zapowiada się też, by w bardziej odległej przyszłości (jeśli kiedykolwiek) maszyny mogły w pełni zastąpić matematyków. To się jednak okaże.