Leniwy nauczyciel

Wyobraź sobie, Czytelniku, że jesteś dość leniwym nauczycielem w podstawówce. Uczniowie mieli zadane, jako pracę domową, ogromne ilości przykładów z tabliczki mnożenia i dodawania w zakresie pięćdziesięciu. Biorąc pod uwagę liczbę dzieci w klasie, sprawdzanie tego to istna męczarnia. Ale przecież można trochę oszukać, np. spośród dziesiątków przykładów, spojrzeć na losowe pięć i przyjąć, że jeśli są dobrze, to reszta też jest pewnie dobrze. A czytania zostaje dużo mniej.

W informatyce powyższy trik działa zaskakująco dobrze, prowadząc do tzw. dowodów sprawdzanych losowo, zwanych w skrócie PCP od ang. *Probabilistically Checkable Proof*. Jako przykład weźmy problem znajdowania cyklu Hamiltona: pytamy, czy dany graf ma cykl, który przechodzi przez każdy wierzchołek dokładnie raz. Choć jest to problem trudny obliczeniowo, istotną własnością tego problemu jest to, iż łatwo przekonać kogoś, że dany graf rzeczywiście ma taki cykl: wystarczy go pokazać. Taki dowód – dla ustalenia uwagi, niech to będzie lista wierzchołków grafu w kolejności odwiedzenia przez cykl Hamiltona – jest jednak trudny do zweryfikowania przez leniwego nauczyciela, który czyta losowy, krótki jego kawałek. Na przykład, bardzo trudno będzie mu odróżnić listę, która opisuje cykl Hamiltona, od takiej, która opisuje dwa rozłączne cykle, które łącznie przechodzą przez każdy wierzchołek grafu.

Twierdzenie PCP, bardzo ważne twierdzenie informatyki teoretycznej z lat 90., mówi, że problem cyklu Hamiltona – i, w ogólności, wszystkie problemy z tzw. klasy NP – mają krótkie dowody PCP. To znaczy, że będąc uczniami leniwego nauczyciela, i mając za zadanie domowe sprawdzenie, czy dany graf G ma cykl Hamiltona, możemy zapisać pozytywną odpowiedź jako napis niewiele dłuższy niż wielkość G, taki że nauczyciel przeczyta (wybrane w sposób losowy) trzy bity naszej odpowiedzi oraz

- jeśli w grafie G istnieje cykl Hamiltona, to istnieje rozwiązanie, które mogliśmy dać nauczycielowi, takie że on zawsze zaakceptuje je jako poprawne;
- jeśli w grafie G nie istnieje cykl Hamiltona, to istnieje stała $\varepsilon > 0$ (niezależna od grafu G) taka, że niezależnie, jakie rozwiązanie spróbujemy oddać, nauczyciel zorientuje się, że go oszukujemy z prawdopodobieństwem co najmniej ε.

By docenić znaczenie tego twierdzenia, spójrzmy na nie z innej strony. Załóżmy, że jesteśmy uczniem, który zna algorytm sprawdzania nauczyciela (tj. wiemy, które trzy bity będzie czytał nauczyciel z jakim prawdopodobieństwem i które wartości tych bitów prowadzą do zaakceptowania pracy domowej), ale nie umie znajdować cyklu Hamiltona w grafach. Czy możemy jakoś oszukać ten system?

Załóżmy, że nauczyciel oczekuje rozwiązania złożonego z m bitów. Dla każdej pozycji i w rozwiązaniu stwórzmy zmienną binarną x_i. Załóżmy, że nauczyciel z prawdopodobieństwem p_{i_1,i_2,i_3} czyta bity na pozycjach i_1, i_2, i_3, i oczekuje wartości ze zbioru $S_{i_1,i_2,i_3} \subseteq \{0,1\}^3$. Dla nas to się tłumaczy jako następujący warunek na nasze zmienne: chcemy, by było $(x_{i_1}, x_{i_2}, x_{i_3}) \in S_{i_1,i_2,i_3}$, a jeśli tego nie spełnimy, to prawdopodobieństwo tego, że nauczyciel odrzuci naszą pracę domową, rośnie o p_{i_1,i_2,i_3}. Innymi słowy, problem znalezienia rozwiązania pracy domowej sprowadza się do tzw. *problemu z więzami*: mamy m zmiennych binarnych, i pewną liczbę więzów mówiących, że pewne trójki zmiennych mają przyjmować określone wartości; każdy z więzów ma określone prawdopodobieństwo, zwane dalej *wartością*.

Twierdzenie PCP mówi, że jeśli nauczyciel sprawdza pracę domową dla grafu G mającego cykl Hamiltona, to istnieje rozwiązanie, które go zawsze przekonuje – czyli istnieje rozwiązanie naszego problemu z więzami, które spełnia *wszystkie* więzy. Z drugiej strony, jeśli G nie ma cyklu Hamiltona, to nauczyciel odrzuca dowolne rozwiązanie z prawdopodobieństwem co najmniej ε – czyli każde rozwiązanie naszego problemu z więzami nie spełnia więzów o łącznym prawdopodobieństwie (wartości) co najmniej ε. Zapomnijmy o nauczycielu, i spójrzmy na to tak: zamieniliśmy problem sprawdzania, czy dany graf ma cykl Hamiltona, na problem rozróżniania, czy w danym problemie z więzami da się spełnić *wszystkie* więzy, czy też dowolne rozwiązanie nie spełnia jakiejś stałej, ustalonej wartości ε więzów. Czyli przetłumaczyliśmy problem cyklu Hamiltona na tzw. problem z dziurą: instancje dla grafów z cyklami Hamiltona są znacząco inne od tych dla grafów bez takich cykli – jest tam dziura o wartości ε.

Załóżmy teraz, że mamy $(1-\delta)$-aproksymacyjny algorytm dla naszego problemu z więzami: jeśli w danej instancji najlepsze rozwiązanie spełnia więzy o łącznej wartości μ, to nasz algorytm zawsze zwraca rozwiązanie o łącznej wartości co najmniej $(1-\delta)\mu$. Jeśli $\delta < \varepsilon$, to można użyć takiego algorytmu do rozstrzygania problemu istnienia cyklu Hamiltona, gdyż będzie on w stanie odróżnić instancje problemu z więzami, w których można spełnić wszystkie więzy, od tych, w których można spełnić więzy o łącznej wartości co najwyżej $(1-\varepsilon)$. Otrzymujemy następujący wniosek z twierdzenia PCP: stworzenie algorytmu $(1-\delta)$-aproksymacyjnego dla $\delta < \varepsilon$ dla naszego problemu z więzami jest co najmniej tak trudne jak znajdowanie cyklu Hamiltona w grafie.

Ten wniosek jest punktem wyjścia do całej dziedziny *trudności aproksymacji*, która jest współcześnie intensywnie rozwijana.

Marcin PILIPCZUK