

Każdy może być odkrywcą

OLEK MICHALSKI

Dr Olek Michalski jest neurobiologiem, zajmuje się rozwijaniem, wdrażaniem i wykorzystaniem technik obliczeniowych w naukach ścisłych.

Niewystarczająca moc obliczeniowa do przetworzenia danych? Absurdalna liczba roboczogodzin konieczna do inspekcji surowych wyników? Naukowcom, którzy mają tego typu problemy, mogą przyjść z pomocą wolontariusze wyposażeni jedynie w komputer podłączony do sieci. Czyli prawie wszyscy.

Statystyka umożliwia wykrywanie regularności w analizowanych zbiorach danych, ale słabo radzi sobie ze zdarzeniami bardzo rzadkimi lub sytuacjami, w których podejrzewamy istnienie pewnych wzorców, jednak dopóki ich nie wykryjemy, nie mamy pojęcia, jak mogą wyglądać. Z tego rodzaju problemem spotykamy się, na przykład przeszukując sygnały radiowe pochodzące z kosmosu w poszukiwaniu przekazów wysłanych przez pozaziemskie inteligencje.

Koncepcja nawiązania radiowej łączności z pozaziemskimi cywilizacjami jest stara jak wynalazek radia i już Nikola Tesla, Guglielmo Marconi i lord Kelvin sugerowali, że technologia radiowa może posłużyć do skontaktowania się z mieszkańcami Marsa. Chociaż nasze wyobrażenie o pozaziemskiej inteligencji od tego czasu radykalnie się zmieniło, przekonanie, że – jeśli taka istnieje – możliwe jest odebranie od niej sygnału, pozostało. Nasłuch pozaziemskich sygnałów prowadzono od lat 60. XX wieku w USA i w ZSRR, a wraz z rozwojem radioastronomii nastąpił lawinowy wzrost ilości danych w projektach SETI (ang. Search for Extraterrestrial Intelligence, poszukiwanie inteligencji pozaziemskiej). Do ich analizy potrzeba było coraz więcej mocy obliczeniowej. W połowie lat 90. XX wieku badacze z Uniwersytetu Kalifornijskiego w Berkeley zwrócili uwagę, że alternatywę dla wyspecjalizowanych superkomputerów mogą stanowić komputery osobiste podłączone do Internetu. W maju 1999 roku otwarto projekt SETI@home (czytane „SETI at home”, czyli „SETI w domu”), który pozwalał każdemu użytkownikowi na instalację programu udostępniającego nieużywaną moc obliczeniową na potrzeby analizy sygnału pochodzącego z radioteleskopu w Arecibo.

SETI@home wyróżniło się nadzwyczajną popularnością – w ciągu roku przed udostępnieniem pierwszej wersji oprogramowania do projektu zgłosiło się 400 tysięcy chętnych,

a tydzień po jego udostępnieniu program ściągnęło i uruchomiło 200 tysięcy osób. Pomimo braku spektakularnego sukcesu w postaci znalezienia sygnału, który faktycznie pochodziłby od pozaziemskiej inteligencji, projekt utworzył drogę podobnym przedsięwzięciom. Naukowcy skorzystali z pomocy internautów przy badaniu różnego rodzaju problemów, od symulacji zderzeń cząstek o wysokiej energii (LHC@home), które służą jako punkt odniesienia dla faktycznych pomiarów prowadzonych w Wielkim Zderzaczu Hadronów, do modelowania oddziaływań pomiędzy białkami (Rosetta@home) wykorzystywanych do poszukiwania nowych leków.

Zapatrzeni w niebo

W projektach obliczeniowych pomoc wolontariuszy jest bardzo istotna, jednak ich udział ogranicza się do udostępnienia komputerów. Często jednak konieczny jest świadomy udział człowieka. Dotyczy to najczęściej analizy obrazów, na których trzeba odnaleźć obiekty o trudnych do opisanie lub zmiennych cechach. Dobrym przykładem są zdjęcia prawie miliona galaktyk zebranych w ramach Sloan Digital Sky Survey – przeglądu nieba prowadzonego przez Obserwatorium Apache Point w Nowym Meksyku. Zajmujący się klasyfikacją sfotografowanych obiektów Kevin Schawinski już po tygodniu pracy określił to zajęcie jako „odmóżdżające”. Postanowił wciągnąć do współpracy wolontariuszy i wspólnie z Chrisem Lintottem z Uniwersytetu w Oksfordzie stworzył serwis internetowy, dzięki któremu każdy mógł pomóc w oznaczaniu galaktyk na zdjęciach. Liczył na 20-30 tysięcy osób, a w projekt zaangażowało się ponad 100 tysięcy uczestników, którzy w pół roku sklasyfikowali 40 milionów obiektów.

Wolontariusze poradzili sobie z zadaniem znacznie szybciej, niż się spodziewano, i na dodatek szybko zaczęli zauważać różnego rodzaju nieregularności, które przy dokładniejszej analizie okazały się nowymi typami obiektów astronomicznych. Dla naukowców była to nieoceniona pomoc.

Nie tylko galaktyki

Sukces projektu ochrzczonego Galaktycznym Zoo doczekał się kolejnych, rozszerzonych edycji i skłonił Lintotta do stworzenia platformy Zooniverse, w której wolontariusze mogą wybierać z całego spektrum projektów – od przeglądania rejestrowanych w LHC trajektorii cząstek, poprzez opisywanie nagrań wideo z foteopułapek ustawionych w dżungli, aż do pomocy w digitalizacji i analizie dzienników wojskowych z pierwszej wojny światowej. Zadania te sprowadzają się do wyszukiwania i klasyfikacji wzorców na zdjęciach, w czym człowiek jest nadal znacznie lepszy niż komputer.



Projekt GalaxyZoo zatrudnił wolontariuszy do klasyfikacji obiektów takich jak galaktyki na zdjęciach nieba. Na zdjęciu galaktyka NGC-1309 widziana przez teleskop Hubble'a; sierpień-wrzesień 2005.



Zrzut ekranu działającego programu Foldit. Widoczne jest zadanie 1153 – świńska saponyna, białko służące jako aktywator enzymów degradujących lipidy

Czy jednak jest to jedyny sposób, w jaki można wykorzystać pracę wolontariuszy w nauce?

Jednym z najważniejszych problemów nauk biomedycznych jest ograniczona wiedza o strukturze białek. Chociaż łatwo poznać sekwencję aminokwasów, czyli cegiełek budujących białko, mamy trudności z ustaleniem, jaki kształt przybierze taki łańcuch w procesie nazywanym zwiżaniem białka, a od tego zależy jego funkcja. Liczba możliwych kształtów jest ogromna, ponieważ łańcuchy są długie i bardzo elastyczne. Sprawdzenie wszystkich układów po kolei w celu wybrania najbardziej prawdopodobnego trwa bardzo długo mimo stosowania wyrafinowanych metod obliczeniowych. Problem ten spróbowano rozwiązać, tworząc grę Foldit, która wykorzystuje umiejętność rozwiązywania pewnego rodzaju problemów przestrzennych i charakterystyczny dla człowieka sposób poszukiwania rozwiązań.

Foldit to gra komputerowa umożliwiająca użytkownikowi własnoręcznie zwiżanie białka. Program pokazuje model cząsteczki, pozwala modyfikować jej kształt przy użyciu różnych narzędzi i na bieżąco punktuje jakość uzyskanej konformacji przestrzennej. Od maja 2008 roku z takimi zadaniami zmierzyło się ponad 57 tysięcy graczy, którzy osiągnęli znaczne sukcesy. Dzięki ich pomocy udało się na przykład opracować strukturę proteazy retrowirusowej wirusa Masona-Pfizera wywołującego małpi odpowiednik AIDS, przy czym wcześniej kształt tego białka, kluczowego dla funkcjonowania wirusa, próbowano bezskutecznie ustalić od ponad dekady. Tak zachęcające wyniki nie zmieniają celu twórców Foldita, którym jest ustalenie, czy taki sposób poszukiwania optymalnej struktury białka jest skuteczniejszy niż zalgorytmizowane metody obliczeniowe.

Rewolucja w świecie nauki?

Mimo znacznej pomocy ze strony wolontariuszy ich udział w tworzeniu nauki budzi wątpliwości, szczególnie w kwestii dokładności dostarczanych wyników. Nie na wszystkich uczestnikach można polegać w tym samym stopniu, jak

zauważył Andrew Westphal korzystający z ich pomocy w projekcie Stardust@home poświęconym poszukiwaniu śladów pyłu kosmicznego. Żeby wspiąć się w rankingu liczby przejranych prób, niektórzy wolontariusze po prostu otwierali pliki z kolejnymi danymi, zaniedbując ich analizę. Zaczęto więc oceniać jakość pracy użytkowników i na podstawie zmian tych wartości w czasie ustalono minimalną liczbę użytkowników, którzy muszą wydać zgodną opinię na temat danej próbki, żeby została ona zaakceptowana. Z drugiej strony twórcy projektu Galaxy Zoo zapewniają, że uzyskane wyniki są wiarygodne, choć zaobserwowali pewne systematyczne odchylenia w oznaczaniu galaktyk spiralnych jako prawo- bądź lewoskrętnych. Wynika ono ze specyficznych właściwości ludzkiego postrzegania i żeby oszacować wielkość tego błędu, zaprezentowali wolontariuszom niektóre zdjęcia w lustrzanym odbiciu.

Pozostaje jeszcze kwestia etyczna wykorzystania pracy wolontariuszy. W artykule opublikowanym niedawno w „Journal of Medical Ethics” autorzy zwracają uwagę na ryzyko uzależnienia od uczestnictwa w projektach obliczeniowych i postulują, że projekty tego rodzaju powinny podlegać nadzorowi komisji bioetycznych. Większość głosów jest jednak wobec tego rodzaju przedsięwzięć mniej sceptyczna. Upatruje się w nich nie tylko rozwiązania problemu braku mocy obliczeniowej i niedoskonałości algorytmów, ale także pomostu między światem nauki i szeroką publicznością, a także sposobu na wciągnięcie do tego świata młodzieży. ■

Chcesz wiedzieć więcej?

- www.setiathome.ssl.berkeley.edu – Seti@home
- www.galaxyzoo.org – Galaxy Zoo
- <http://fold.it> – Foldit
- E. Hand (2010). Citizen science: People power. *Nature* 466 (7307), 685–7. doi:10.1038/466685a, PMID 20686547.
- Khatib F., DiMaio F., Foldit Contenders Group i in. (2011). Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural and Molecular Biology* 18, 1175–1177.