

Patryk Orzechowski – autor pracy „Zagregowane algorytmy biklasteryzacji oraz ich efektywna implementacja”. Opracował kryteria i zakresy stosowalności dostępnych algorytmów dwugrupowania wraz z oceną możliwości ich przyspieszenia w środowisku komputerów dużej mocy.

Od jak dawna zajmuje się Pan naukami ścisłymi?

Właściwie od czasów liceum. Dzięki temu, że uczęszczałem do klasy o profilu matematyczno-fizycznym UJ w V LO w Krakowie, mogłem odkryć swoje pasje oraz zalety pracy w zespole. Już wtedy wyobrażałem sobie siebie jako naukowca.

Czym kierował się Pan wybierając temat pracy doktorskiej?

Zdecydowałem, że swoje poszukiwania rozszerzę na dziedziny biologii i genetyki, wykorzystując do badań aparat matematyczny i statystyczny. Ostatecznie skupiłem się na algorytmach dwugrupowania, czy jak to określiłem w rozprawie, biklasteryzacji. Dzięki mojemu opiekunowi, jego doświadczeniu, radom i wsparciu wielu innych osób doprowadziłem badania do końca.

Jakie narzędzia były przez Pana wykorzystywane?

Były to narzędzia typowo informatyczne – środowisko R do analizy statystycznej, różne bazy danych wyszukane w internecie. Analizując algorytmy starałem się określić ich najsilniejsze strony – właściwości, które wyróżniały je spośród innych. Obliczenia początkowo wykonywałem używając własnego komputera, lecz oczekiwanie na wyniki jednego zadania trwało nawet kilkanaście godzin. Z tego powodu (jeśli to tylko było możliwe) do dalszych obliczeń był wykorzystywany superkomputer Zeus znajdujący się w Cyfronecie.

Więc praca z wykorzystaniem komputerów udostępnionych przez Cyfronet przyczyniła się do przyspieszenia obliczeń...

Zdecydowanie tak. Dzięki szkoleniu przygotowującemu do pracy z Zeusem dowiedziałem się, jak w praktyce można przyspieszyć obliczenia wykorzystując karty GPGPU. Nauczyłem się także lepiej szacować czas, który był potrzebny na wykonanie poszczególnych zadań. To przełożyło się na szybsze ich rozpoczynanie i znaczne skrócenie oczekiwania na rezultaty. Wykorzystałem również możliwość jednoczesnego uruchamiania nawet kilkuset obliczeń na raz. Zaimplementowane przeze mnie skrypty pozwalały automatycznie dodawać kolejne zadania do kolejki – wystarczyło właściwie tylko obserwować postęp ich wykonania.

Czy otrzymane wyniki były zgodne z Pańskimi oczekiwaniami?

Czasami – tak, a czasem – nie. Przykładowo: oczekiwałem, że połączenie najważniejszych cech dwóch algorytmów pozwoli otrzymać lepsze wyniki, a nie zawsze miało to miejsce. Zauważyłem jednak pozytywną stronę – choć najlepszy wynik był nieco gorszy niż zakładałem, znacznie zwiększyła się liczba wyników dobrych i bardzo dobrych.

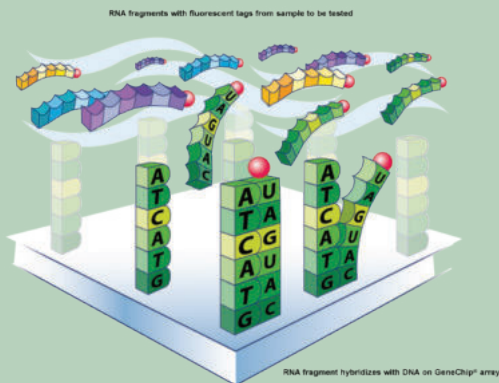
Co uważa Pan za swój największy sukces?

Było wiele pomniejszych sukcesów, które tak naprawdę złożyły się na ten największy, czyli obronę przedstawionych w rozprawie tez. Dużym impulsem, który zmotywował mnie do zdwojenia wysiłków, było powstanie pierwszego sensownie działającego algorytmu, a później spostrzeżenie, że celem badań nie musi być uzyskanie najlepszego wyniku w każdym teście. Nie istnieją przecież algorytmy, które dla każdego zbioru danych dawałyby najlepsze wyniki. Sporym osiągnięciem jest także generalna poprawa jakości uzyskanych wyników czy większa powtarzalność dobrych wyników. Wiele radości dała mi możliwość wykorzystania wyników pracy w praktyce, np. w medycynie personalizowanej, biologii, serwisach społecznościowych czy serwisach internetowych.

Co poradziłby Pan przyszłym naukowcom?

To w sumie niełatwe pytanie, ponieważ każda dziedzina nauki jest inna, każdego badacza fascynuje coś innego. Wydaje mi się, że najważniejsze to znaleźć taki temat badań, który dostarcza satysfakcji. Radziłbym przyszłym badaczom, aby mieli jasno określony cel, do którego będą dążyć i aby nie poddawali się, gdy przyjdzie zmierzyć się z niepowodzeniami. Jeśli jakieś zadanie jest zbyt skomplikowane, warto podzielić je na mniejsze podzadania. Trzeba też być wytrwałym i mieć otwarty umysł – wyniki nie zawsze pojawiają się tam, gdzie ich oczekujemy.

Bardzo dziękuję za rozmowę.



Proces hybrydyzacji (rysunek dzięki uprzejomości Affymetrix)



Najczęściej wykorzystywane słowa w tym samym tweecie z tagiem #data