Dr inż. Dominik Żurek

Rozmowa z autorem pracy:
„Akceleracja obliczeń algorytmów uczenia maszynowego oraz wybranych populacyjnych algorytmów inteligencji obliczeniowej ze zredukowaną precyzją danych poprzez implementację w układach GPGPU"

Jak przybliżyłby Pan swoje badania osobie niezwiązanej z dziedziną?

Celem badań było opracowanie i zanalizowanie skutecznych i efektywnych metod przyśpieszania obliczeń nadających się do przetwarzania równoległego, związanych z algorytmami sztucznej inteligencji oraz algorytmami ewolucyjnymi. Łatwo zauważyć, że pojawiły się tu trzy aspekty, tj. akceleracja obliczeń, sztuczna inteligencja oraz algorytmy ewolucyjne, które postaram się pokrótce przybliżyć.

Pojęciu sztucznej inteligencji (ang. *Artificial Intelligence – AI*) przypisuje się kilka znaczeń, najczęściej jednak termin ten definiowany jest jako uczące się maszyny lub systemy informatyczne, które przetwarzają informację w oparciu o reguły ludzkiego rozumowania. Sztuczna inteligencja zajmuje się zagadnieniami, które nie są efektywnie algorytmizowane w oparciu o modelowanie wiedzy, więc do ich rozwiązania należy wprowadzić algorytmy posiadające znamiona inteligencji. Przez określenie „inteligencja" rozumie się w tym przypadku zdolność do samodzielnego przystosowania się do zmiennych warunków. Przez proces uczenia się systemu rozumie się dokonanie autonomicznej zmiany w systemie, zachodzącej na podstawie zdobytych doświadczeń i prowadzącej do poprawy jakości jego działania. Przy takiej definicji zakłada się, że istnieje możliwość oceny jakości podejmowanych decyzji, czyli umiejętność odróżnienia zmian korzystnych od niekorzystnych. Sztuczna inteligencja potrafi przyswajać wiedzę poprzez wyodrębnianie wzorców z surowych danych. Wprowadzając do algorytmu elementy ludzkiej inteligencji możliwe jest wytrenowanie go tak, by potrafił rozpoznawać obrazy, rozumiał język naturalny czy też był zdolny do logicznego rozumowania.

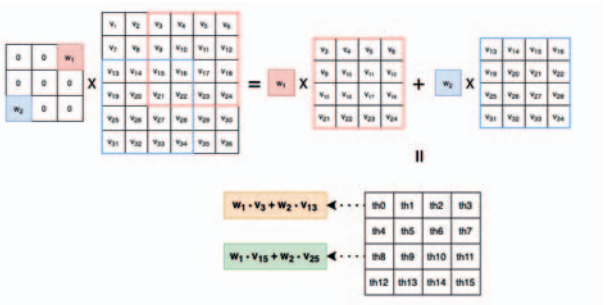
Jedną z poddziedzin sztucznej inteligencji jest tzw. inteligencja obliczeniowa (ang. *Computational Intelligence – CI*), która polega głównie na zdolności przystosowania się systemu do zmieniającego się środowiska. Kładzie ona duży nacisk na ulepszanie i rozwój aplikacji w świecie rzeczywistym. Cechą odróżniającą algorytmy tego typu od innych algorytmów sztucznej inteligencji jest brak korzystania ze zdefiniowanego modelu, lecz podejmowanie prób zbudowania go samodzielnie na podstawie dostarczonych zbiorów uczących. Algorytmy inteligencji obliczeniowej obejmują algorytmy wywodzące się ze sztucznej inteligencji związane z inteligentnym przetwarzaniem danych, w których istotną grupę stanowią algorytmy ewolucyjne. Ich inspiracją są procesy naturalnej ewolucji organizmów. Przenosząc inspiracje biologiczne na algorytmy ewolucyjne, populacja osobników jest odzwierciedleniem liczby rozwiązań (osobnik = rozwiązanie). Ewolucja rozwiązań algorytmu jest odpowiednikiem zmian występujących w populacji, za co odpowiedzialne są operatory mutacji (losowa modyfikacja) oraz rekombinacji (wymiana materiału genetycznego), co sprowadza się do wyszukiwania kolejnych rozwiązań.

Wspomniane grupy algorytmów do prawidłowego działania potrzebują ogromnych zasobów mocy obliczeniowej, stąd też rodzi się potrzeba opracowania metod, dzięki którym możliwe będzie szybsze ich wykonanie, czyli „akceleracja obliczeń".

Co zadecydowało o tym, że poświęcił Pan rozprawę doktorską zagadnieniu akceleracji w układach GPGPU?

Układy GPGPU specjalizują się w obliczeniach wysoce równoległych, poprzez przeprowadzanie obliczeń zgodnie z ar-

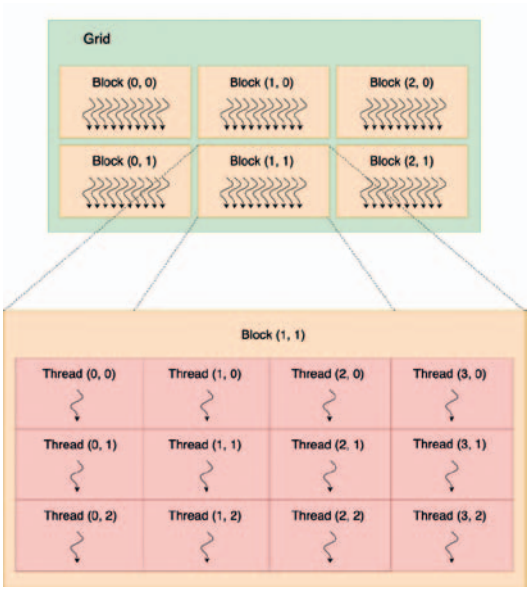
chitekturą SIMD (ang. *single instruction stream multiple data*). Jak zostało wyżej wspomniane, w moich badaniach skupiłem się na algorytmach uczenia maszynowego oraz algorytmach ewolucyjnych, ze względu na ich szerokie spektrum zastosowań. Obie te grupy algorytmów wykazują tendencję do równoległości, stąd wybór kart graficznych jako akceleratora sprzętowego wydawał się najbardziej naturalny.



Obliczanie konwolucji na układach GPGPU przy użyciu operacji na macierzach rzadkich

Spróbujmy spojrzeć w przyszłość: jak badania mogą się przełożyć na wdrożenia praktyczne? Czy jest dziedzina szczególnie oczekująca na przełom w tej materii?

W moim przekonaniu zdecydowanie taką dziedziną jest neuroewolucja, którą obecnie się zajmuję, dzięki wiedzy i doświadczeniu zdobytym podczas pracy nad doktoratem. Neurowolucja zajmuje się rozwojem sieci neuronowej poprzez użycie algorytmów ewolucyjnych (czyli dotyka obu grup algorytmów, które były przedmiotem moich badań). Rozwój ten może odbywać się na dwóch płaszczyznach, tj. na poziomie poszukiwania optymalnej architektury sieci neuronowej, bądź za jego przyczyną może odbywać się sam proces uczenia sieci (mutacja wag). W obu przypadkach wymagana jest ogromna moc obliczeniowa. W związku z tym metody potrafiące przetwarzać w sposób bardziej efektywny algorytmy sztucznej inteligencji oraz algorytmy ewolucyjne na pewno znajdą przełożenie na wdrożenia praktyczne w tej materii.



Schemat grupowania wątków na karcie graficznej

Jak w kontekście badań w dziedzinie akceleracji i obróbki danych ocenia Pan rolę Cyfronetu?

W swoich badaniach użyłem bardzo wydajnej i nowoczesnej karty graficznej, tj. Nvidia Tesla V100-SXM2-32GB. Dzięki niej dla niektórych z podjętych problemów możliwe stało się osiągnięcie ogromnych przyśpieszeń względem starszych kart oraz względem wielordzeniowych procesorów ogólnego przeznaczenia. Do wszystkich tych akceleratorów miałem dostęp dzięki Cyfronetowi, więc z całą pewnością mogę powiedzieć, że bez Cyfronetu nie byłoby możliwe uzyskanie tak imponujących wyników.

Czy miałby Pan jakieś rady dla kolegów i koleżanek rozważających podjęcie studiów doktoranckich?

Na początku swojej przygody naukowej trafiłem do zespołu „Akceleracji obliczeń" w Cyfronecie, gdzie poznałem wybitnych specjalistów, którzy okazali się również wspaniałymi ludźmi. Dzięki temu obudziła się we mnie pasja oraz głęboka chęć bliższego zaznajomienia się z tematyką, którą podjąłem w rozprawie doktorskiej i którą zajmuję się do dziś. Dlatego też młodym naukowcom rekomenduję nawiązanie współpracy z zespołem naukowym zajmującym się interesującą ich tematyką. Dzięki temu staną się częścią grupy biorącej udział w skomplikowanych projektach badawczych, a co za tym idzie, będą mogli szybciej zdobywać cenne doświadczenie i ubogacać swoją wiedzę.