

Recenzja rozprawy doktorskiej

Imię i nazwisko kandydata: Bartłomiej Kotyra

Tytuł rozprawy doktorskiej:

Wydajne algorytmy równoległe w modelowaniu hydrologicznym

Promotor: dr hab. inż. Przemysław Stipczyński

Recenzent: dr hab. inż. prof. IITIS PAN Krzysztof Grochla

Gr

1. Wybór tematu pracy

Rozprawa dogłębnie analizuje problem zastosowania obliczeń równoległych w przetwarzaniu modeli hydrologicznych. Doktorant w ramach prac nad rozprawą przygotował nowe algorytmy przetwarzania równoległego dla wybranych zagadnień z obszarów badań nad modelowaniem hydrologicznym zlewni rzek i systemów informacji geograficznej. Praca dotyczy tematu aktualnego i dobrze wpisuje się w trend zastosowania obliczeń równoległych w celu przyspieszenia przetwarzania modeli złożonych systemów. Doktorant skoncentrował się na algorytmach akumulacji spływu powierzchniowego, wyznaczania zlewni oraz identyfikacji najdłuższych ścieżek spływu. Doktorant wybrał te algorytmy ze względu na brak w literaturze naukowej przed podjęciem badań przez doktoranta propozycji algorytmów pozwalających efektywnie rozwiązać wskazane problemy za pomocą masywnie równoległego przetwarzania.

Temat podjęty w pracy doktorskiej jest aktualny i dotyczy obliczeń, dla których przetwarzanie złożonych zbiorów danych jest dużym wyzwaniem. Ze względu na zwiększenie dokładności modeli GIS i dostępność zbiorów danych pozwalających bardzo precyzyjnie opisać wysokość terenu, wyznaczenie zlewni i identyfikacja najdłuższych ścieżek spływu za pomocą klasycznych algorytmów sekwencyjnych wymaga bardzo czasochłonnych obliczeń. Dynamiczny rozwój mocy obliczeniowej kart graficznych oraz zwiększenie liczby rdzeni dostępnych we współcześnie stosowanych procesorach zwiększa dostępność serwerów i komputerów umożliwiających uruchamianie algorytmów rozwijanych w rozprawie doktorskiej. Zastosowanie przetwarzania równoległego za pomocą obliczeń realizowanych na kartach graficznych (np. w modelu CUDA) lub na procesorach wielordzeniowych daje szansę na istotne skrócenie czasu obliczeń dla wskazanych problemów modelowania hydrologicznego. Tym samym Doktorant dobrze dobrał metodę zwiększenia wydajności przetwarzania danych do wskazanego problemu badawczego.

2. Ocena układu rozprawy

Rozprawa składa się ze streszczenia, 5 rozdziałów, bibliografii oraz 3 artykułów naukowych stanowiących załączniki do rozprawy. Rozdziały głównej części pracy obejmują wstęp do problematyki badawczej, opis stanu wiedzy w zakresie obliczeń równoległych (rozdział 2) oraz modelowania hydrologicznego i systemów informacji przestrzennej GIS (rozdział 3), omówienie najważniejszych wyników przedstawionych w załączonych artykułach (rozdział 4) oraz podsumowanie.

Podział rozprawy na rozdziały jest typowy dla prac doktorskich stanowiących zbiór opublikowanych i powiązanych tematycznie artykułów naukowych. Doktorant czytelnie zdefiniował cel rozprawy i zakres realizowanych badań we wstępie. Następnie dobrze określił stan wiedzy oraz przedstawił pojęcia z zakresu hydrologii, ułatwiając czytelnikowi zrozumienie celu działania algorytmów rozwijanych dalej w pracy. Doktorant także bardzo klarownie podsumował najważniejsze rezultaty opisane w każdym z 3 artykułów stanowiących elementy rozprawy. Przyjęta metoda organizacji pracy w sposób jednoznaczny pozwala na wyróżnienie wkładu własnego doktoranta od opisu stanu wiedzy. Ostatni rozdział dobrze podsumowuje uzyskanie wyniki i prezentuje je w kontekście badań w obszarze systemów informacji

przestrzennej i geograficznej. Mniej starannie natomiast przygotowano streszczenie, które sprowadza się do kilku zdań podsumowania zakresu realizowanych prac oraz kopii opisu zawartości poszczególnych rozdziałów, a brak w nim np. informacji na temat uzyskanego przyrostu wydajności przetwarzania danych.

3. Cel pracy

Celem badań zrealizowanych w ramach rozprawy doktorskiej było opracowanie nowych wydajnych algorytmów przetwarzania równoległego, przeznaczonych do realizacji 3 wybranych problemów z obszaru modelowania hydrologicznego i systemów informacji geograficznej. Wybrane problemy obejmują zagadnienia obliczeniowe o dużej złożoności, dla których możliwe jest zastosowanie przetwarzania równoległego i znaczące przyśpieszenie obliczeń dzięki zastosowaniu systemów wielordzeniowych i kart graficznych. Wybrane zagadnienia modelowania są szeroko stosowane w hydrologii, a algorytmy znajdują zastosowanie m.in. w analizie obszarów zlewni rzek lub w wyznaczaniu obszaru objętego zagrożeniem powodziowym. Na konieczność opracowania nowych i bardziej wydajnych algorytmów wpływa także dostępność coraz bardziej dokładnych pomiarów wysokościowych terenu. Dzięki zastosowaniu siatki wysokościowej o większej dokładności możliwe jest wykonanie bardziej precyzyjnej analizy hydrologicznej, jednak złożoność obliczeniowa klasycznych algorytmów rośnie wraz kwadratem liczby elementów siatki wysokościowej. Dlatego niezbędne jest opracowanie nowych, bardziej wydajnych algorytmów modelowania hydrologicznego. Potwierdza to, że cel pracy jest aktualny, a badania mają duże znaczenie społeczne.

4. Analiza źródeł i stanu wiedzy

Bibliografia głównej części rozprawy obejmuje odwołania do 55 artykułów naukowych, prac magisterskich oraz książek z dziedziny przetwarzania równoległego i systemów GIS. Są to prace opisujące tło prowadzonych badań, w tym podstawowe koncepcje związane z równoleglizacją obliczeń, modelowaniem hydrologicznym lub przetwarzaniem danych przestrzennych. Doktorant zawarł także odwołania do prac opisujących wcześniejsze algorytmy tworzenia rastrów kierunku spływu, akumulacji spływu powierzchniowego i wyznaczania zlewni. Nieznacznie bardziej niż w głównej części rozprawy została rozbudowana bibliografia artykułów: zawiera ona odpowiednio 66, 65 i 52 pozycje literaturowe. Istotna część cytowanych pozycji literaturowych jest wspólna dla wszystkich 3 artykułów i głównej części rozprawy, jednak w załączonych artykułach szerzej niż w głównej części rozprawy odniesiono się do prac definiujących alternatywne algorytmy pozwalające rozwiązać ten sam problem modelowania hydrologicznego.

Cytowane prace są związane z tematem rozprawy, a odwołania zostały umieszczone adekwatnie do treści pracy. Doktorant przedstawił stan wiedzy w zakresie systemów obliczeń równoległych w sposób poprawny, jednak pewien niedosyt budzi brak odniesień do nowszych technologii i bibliotek pozwalających na realizację obliczeń równoległych, niż standard OpenMP i platforma CUDA, jak np. OpenCL lub oneAPI. Natomiast odniesienia do stanu wiedzy zawarte w poszczególnych 3 artykułach opisujące alternatywne algorytmy służące do

akumulacji spływu powierzchniowego, wyznaczania zlewni i identyfikacji najdłuższych ścieżek spływu są przedstawione w sposób wyczerpujący i świadczą, że autor dogłębnie przeanalizował stan wiedzy i przed przystąpieniem do tworzenia własnych algorytmów dobrze zapoznał się z informacjami na temat opisanych wcześniej algorytmów modelowania hydrologicznego.

5. Wartość naukowa rozprawy

Praca stanowi istotny przyczynek do rozwoju modelowania hydrologicznego za pomocą systemów masywnych obliczeń równoległych. Doktorant bardzo rzetelnie podszedł do postawionego problemu badawczego, dobrze zaprojektował i zaimplementował nowe algorytmy przetwarzania dużych zbiorów danych. Doktorant poprawnie postawił tezę rozprawy i udowodnił przeprowadzonymi badaniami, że opracowane przez niego metody równolegliczacji obliczeń mogą pozwolić na istotną poprawę wydajności obliczeń w stosunku do obecnie stosowanych rozwiązań.

Doktorant udowodnił tezę na przykładzie 3 wybranych algorytmów modelowania hydrologicznego. Zaproponowane algorytmy zostały poddane wnikliwej ocenie i porównane z innymi dostępnymi w literaturze naukowej metodami. Przedstawione w rozprawie wyniki wskazują na istotną przewagę algorytmów opracowanych przez doktoranta, zapewniając wielokrotne przyspieszenie względem najlepszych wcześniej znanych metod, dzięki skutecznemu wykorzystaniu architektur wielordzeniowych. Analiza wydajności dla wszystkich 3 algorytmów została przeprowadzona bardzo rzetelnie, dla różnych rozmiarów analizowanego zbioru danych i z porównaniem do dobrze wybranego algorytmu referencyjnego. Wyniki działania algorytmów realizowanych równoległe zostały zestawione z wynikami klasycznych algorytmów sekwencyjnych wykonywanych na różnych zbiorach danych, co potwierdziło poprawność działania pod względem funkcjonalnym opracowanych metod modelowania hydrologicznego. Doktorant także rzetelnie wykonał analizę skalowalności opracowanych metod, np. dla algorytmu akumulacji przepływów prezentując wpływ ilości rdzeni na gniazdo na wydajność przetwarzania.

Bardzo dobrze, że doktorant udostępnił kody źródłowe wszystkich 3 algorytmów opracowanych w ramach prac nad rozprawą doktorską. Pozwala to na weryfikację i odtworzenie uzyskanych wyników, a także umożliwia naukowcom zajmującym się modelowaniem hydrologicznym, którzy nie mają tak dużej wiedzy z dziedziny przetwarzania równoległego, na wykorzystanie opracowanych algorytmów we własnych badaniach. O dojrzałości naukowej doktoranta świadczy także świadomość ograniczeń realizowanych prac zawarta w podsumowaniu, w którym autor celnie zauważa, że przedstawione wyniki dotyczą 3 wybranych zagadnień w obszarze przetwarzania danych GIS, w którym algorytmów przetwarzających duże zbiory danych jest znacznie więcej.

Doktorat został oparty na 3 artykułach: 1 opublikowanym w czasopiśmie *Computers & Geosciences* (100 pkt. na liście ministerialnej) oraz 2 artykułach opublikowanych w



czasopiśmie *Environmental Modelling & Software* (140 pkt. na liście ministerialnej). Liczba artykułów naukowych włączonych w skład rozprawy nie jest duża, lecz są to prace w czasopismach o istotnym znaczeniu i dużym zasięgu wśród naukowców zajmujących się modelowaniem hydrologicznym. Doktorant jest wyłącznym autorem artykułu pt. „High-performance watershed delineation algorithm for GPU using CUDA and OpenMP”. Opis wkładu autorów w każdym z pozostałych artykułów potwierdza, że doktorant był odpowiedzialny za opracowanie algorytmów i jest autorem kluczowych koncepcji prowadzących do rozwiązania problemu badawczego.

6. Poprawność redakcyjna rozprawy

Rozprawa została przygotowana w sposób bardzo staranny i nie zawiera znaczących błędów językowych lub redakcyjnych. Rozprawa jest napisana w sposób czytelny i zrozumiały. Jedyną wadą rozprawy jest niewielka szczegółowość przedstawionego opisu wyników prac badawczych. W omówieniu uzyskanych wyników w rozdziale 4 doktorant skrupulatnie opisał cel poszczególnych prac i przeprowadzone eksperymenty potwierdzające poprawność działania algorytmów, jednak opisy uzyskanych wyników eksperymentów (np. czasów realizacji obliczeń) zostały przedstawione ze zbyt małą ilością danych i bez odniesień do poszczególnych wykresów lub tabel.

Trzy artykuły naukowe, na których opiera się rozprawa, nie zawierają błędów redakcyjnych. Jednak w dwóch chronologicznie pierwszych artykułach doktoranta algorytmy zostały zaprezentowane w postaci kodu źródłowego. Z punktu widzenia czytelnika posiadającego chęć zrozumienia koncepcji danego algorytmu, czytelniejszym rozwiązaniem byłoby przedstawienie algorytmów w postaci pseudokodu, diagramów lub schematów blokowych, tak jak to zrobiono w artykule dotyczącym najdłuższej ścieżki spływu. W/w drobne wady jednak nie umniejszają wysokiego poziomu redakcyjnego całości rozprawy.

7. Uwagi krytyczne

Doktorant przygotował algorytmy rozwiązujące wybrane trzy z wielu znanych problemów hydrologicznych. Pewien niedosyt budzi brak w doktoracie szerszej analizy, które z problemów modelowania hydrologicznego można jeszcze zrównoleglić i dlaczego wybrano wskazane w rozprawie algorytmy akumulacji spływu powierzchniowego, wyznaczania zlewni oraz identyfikacji najdłuższych ścieżek spływu. Doktorant wskazał, że zagadnienia te zostały wybrane, ponieważ ich wspólnym problemem była niska wydajność obliczeniowa istniejących algorytmów i narzędzi, jednak czytelnik oczekiwałby szerszej dyskusji, dla których metod modelowania hydrologicznego już są dostępne algorytmy równoległe o wysokiej wydajności, dla których je opracowano w ramach doktoratu, a dla których są niezbędne dalsze badania.

Wydajność opracowanych algorytmów w sposób istotny zależy od możliwości ujęcia całości danych w pamięci jednego węzła obliczeniowego (w przypadku algorytmów realizowanych na CPU) lub pamięci karty (w przypadku algorytmów realizowanych za pomocą GPU). Dobrze, że autor wykonał pomiar czasu realizacji zadania obliczeniowego dla różnych rozmiarów zbioru

danych. Jednak przyjęty sposób prezentacji wyników oparty na współczynniku określającym część użytego zbioru (np. na rycinie 1 w pracy „High-performance watershed delineation algorithm for GPU using CUDA and OpenMP”) jest mniej czytelny niż bezpośrednio podanie rozmiaru zbioru danych. Brakuje w pracy pogłębienia analizy złożoności pamięciowej opracowanych metod, zwłaszcza w kontekście możliwego uruchomienia algorytmów na klastrach obliczeniowych lub maszynach wyposażonych w wiele kart działających w modelu CUDA. Pozwoliłoby to odpowiedzieć na pytanie, jaki jest maksymalny rozmiar przetwarzanych zbiorów danych bez znacznego spadku wydajności związanego z koniecznością wymiany danych pomiędzy węzłami obliczeniowymi lub kartami graficznymi.

W 2 z 3 artykułów skoncentrowano się na technologii CUDA, a więc na bibliotekach umożliwiających uruchomienie kodu na kartach graficznych firmy NVidia. Czy możliwe jest przeniesienie opracowanych algorytmów na systemy obliczeniowe innych producentów niż NVidia (AMD lub Intel), np. poprzez wykorzystanie HIP SDK do konwersji kodu dla kart AMD lub wykorzystanie oneAPI/UXL Foundation w celu przygotowania kodu kompilowalnego do dowolnej platformy obliczeniowej?

W/w uwagi krytyczne stanowią jednak bardziej pytania i wskazówki, jak należałoby rozszerzyć pracę w przyszłości, a nie podważają wysokiej jakości i poziomu naukowego ujętych w pracy doktorskiej badań.

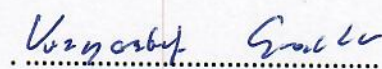
8. Podsumowanie i ocena końcowa

Praca doktorska pt. „Wydajne algorytmy równoległe w modelowaniu hydrologicznym” została przygotowana przez Pana magistra Bartłomieja Kotyra rzetelnie i wykazuje zdolność kandydata do prowadzenia pracy naukowej w sposób samodzielny. Praca jest poprawna pod względem metodologicznym, a opracowane algorytmy przetwarzania równoległego danych GIS wykraczają poza stan wiedzy i stanowią istotny przyczynek do rozwoju dziedziny modelowania hydrologicznego. Badania zostały poprawnie zaplanowane i zrealizowane, a ich wyniki szczegółowo i precyzyjnie opisane. Praca tym samym potwierdza, iż kandydat posiada wymaganą wiedzę teoretyczną w dyscyplinie Informatyka Techniczna i Telekomunikacja.

Ja, niżej podpisany stwierdzam, że recenzowana rozprawa doktorska Pana Bartłomieja Kotyra spełnia warunki określone w art. 187 ust. 1 i 2 Ustawy z dnia 20 lipca 2018r. Prawo o szkolnictwie wyższym i nauce (Dz.U. 2018 r., poz. 1668 z późn. zmianami) i wnioskuję do Rady Naukowej Instytutu Informatyki UMCS o dopuszczenie Pana Bartłomieja Kotyra do dalszych etapów przewodu doktorskiego.

1.08.2024

.....
data sporządzenia recenzji


.....
podpis recenzenta