

dr hab. Piotr Wiśniewski
Wydział Matematyki i Informatyki
Uniwersytet Mikołaja Kopernika w Toruniu
Chopina 12/18, 87-100 Toruń

Recenzja rozprawy doktorskiej Daniela Kostrzewy

Przeszukiwanie przestrzeni rozwiązań w optymalizacji planów zapytań do baz danych z wykorzystaniem heurystycznego algorytmu IWO

Recenzowana rozprawa doktorska wpisuje się w bardzo ważny nurt badań nad optymalizacją przetwarzania zapytań w bazach danych. Nurt ten jest niezwykle istotny, gdyż rozmiar i złożoność przetwarzanych danych rośnie znacznie szybciej niż wydajność komputerów.

W recenzowanej pracy doktorant zajął się dobraniem optymalnej kolejności złączeń tabel, dla uzyskania minimalnego czasu realizacji zapytania. Cel ten jest realizowany poprzez modyfikację algorytmu *Invasive Weed Optimization* nazywanym dalej w skrócie *IWO*.

Omówienie wyników rozprawy

Praca składa się z siedmiu rozdziałów oraz dwóch dodatków. Rozdział pierwszy jest krótkim wstępem naświetlającym ważność prowadzonych badań oraz będącym przewodnikiem po pracy.

Rozdział drugi jest pierwszym właściwym rozdziałem w pracy. Wprowadza on czytelnika w zagadnienia optymalizacji zapytań. Przypomina istotne dla badań pojęcia planu dobrego, planu akceptowalnego oraz planu nie do zaakceptowania. Dalej omawia schemat działania optymalizatora.

Rozdział trzeci omawia problemy optymalizacyjne. Autor wprowadza czytelnika najpierw w główne zagadnienie dla pracy, tj. optymalizację kolejności złączeń tabel. Następnie omawia krótko, ale wystarczająco problem minimum funkcji wielowymiarowej oraz problem komiwojażera. Problemy te zostaną dalej wykorzystane do testów jakości proponowanych rozszerzeń algorytmu *IWO*.

N.S.

Rozdział czwarty zatytułowany jest *Przegląd algorytmów optymalizacyjnych*. Tytuł ten okazuje się jednak trochę górnolotny. Rozdział ten przedstawia najpierw algorytm ewolucyjny – jako pierwszy algorytm inspirowany biologią, następnie przedstawia kluczowy dla pracy algorytm IWO. Zostaje on bardzo solidnie omówiony wraz z przykładowymi zastosowaniami. Ostatnie półtorej strony stanowi pobieżna prezentacja kilku algorytmów mająca zapewne uzasadnić tytuł rozdziału. Pewnym brakiem jest pominięcie w nim algorytmów, z którymi rozwiązanie w rozdziale szóstym porównywane jest autorskie.

Kluczowe rozważania pracy zawiera rozdział piąty. Zaprezentowana w nim jest autorska modyfikacja algorytmu IWO, która jak wykazują późniejsze testy, potrafi dać lepsze wyniki w optymalizacji złączeń niż oryginalny algorytm. Modyfikacja dotyczy czterech aspektów:

- Inicjalizacja populacji początkowej – autor decyduje się na wstępną selekcję zamiast losowania. W efekcie skraca czas pracy algorytmu. Podejście to czyni w opinii recenzenta implementację bardziej złożoną jednocześnie wprowadzając ciężkie obliczenia na starcie. Okazuje się jednak, że pozwala to na istotne skrócenie realizacji całego algorytmu.
- Rozproszenie ziaren wokół chwastu – na tym etapie autor upatruje najwięcej możliwości poszukiwań badawczych. Etap ten jest kluczowy dla poszukiwania najlepszych rozwiązań. Dla każdego ziarna losowana jest metoda, czy zostaje ono rozwiane, rozsiane, czy stoczone. Autor dokładnie opisuje każdą z tych dróg. Pewną słabością tego fragmentu jest słabe uwypuklenie różnic, pomiędzy oryginalnym podejściem algorytmu IWO, a modyfikacjami proponowanymi przez autora.
- Selekcja najlepszych chwastów – autor prezentuje metody selekcji, niestety również słabo uwypuklone są modyfikacje wobec algorytmu oryginalnego.
- Warunek zatrzymania algorytmu - autor proponuje dodać ograniczenie czasowe dla algorytmu. Oryginalnie algorytm zatrzymuje się po wykonaniu z góry założonej liczby kroków.

W dalszej części rozdziału autor prezentuje wykorzystanie zmodyfikowanego algorytmu IWO w znajdowaniu minimum funkcji wielomianowa nowych oraz w problemie komiwojażera.

W kolejnej części rozdziału omówione zostanie wykorzystanie algorytmu w głównym zastosowaniu pracy, czyli ustaleniu kolejności złączeń w realizacji zapytań w bazach danych. Ziarnem jest kolejność złączeń, oceną jakości jest ilość przetwarzanych danych w realizacji złączenia w danej kolejności.

Ostatni fragment rozdziału dokładnie analizuje złożoność obliczeniową realizacji algorytmu. Analiza ta jest gruntownie i dokładnie dokonana. W proponowanym zastosowaniu jest ona niezwykle ważna, pozwala ocenić, czy przypadkiem koszt dobrania właściwej kolejności nie jest wyższy niż zysk uzyskany w wyniku optymalizacji.

Rozdział szósty omawia wyniki testów.

W pierwszej części autor prezentuje funkcje wielowymiarowe, na których testowany jest omawiany algorytm. Funkcje te dobrane są tak, aby poziom skomplikowania optymalizacji rósł. Pierwsze z nich służą do wstępnej weryfikacji, czy zaproponowane zmiany algorytmu mają w ogóle sens. Jakość optymalizacji jest porównywana z wynikami otrzymanymi dla oryginalnej wersji IWO oraz dla algorytmu APSO. Zaprezentowane eksperymenty pokazują, że zmodyfikowany algorytm IWO dostarcza lepszej jakości wyniki niż porównywane algorytmy. Nie do końca jasne jest czy doktorant porównuje proponowane rozwiązania z własną implementacją algorytmów referencyjnych, czy też z wynikami z cytowanych prac. W etykietach wykresów porównania z algorytmem APSO pojawia się pewna nieścisłość. W pracy proponowana wersja algorytmu jest oznaczana jako exIWO, gdy tymczasem na etykietach wykresów w porównaniu z APSO pojawia się IWO, co w pracy używane jest do oznaczenia oryginalnego algorytmu. Z kontekstu natomiast widać, że autor ma na myśli exIWO.

Druga część porównuje efektywność zmodyfikowanej wersji algorytmu w zastosowaniu dla problemu komiwojażera. Autor prezentuje podejścia do problemu komiwojażera i porównuje wyniki dla exIWO z wynikami innych algorytmów. Podejście to bardzo dobrze prezentuje jakość otrzymanej optymalizacji. Interesująca byłaby własna implementacja przynajmniej części algorytmów prezentowanych do porównania. Możliwe byłoby wówczas zbadanie kosztu optymalizacji. Implementacja taka prawdopodobnie byłaby jednak zbyt czasochłonna i przekraczała rozmiar doktoratu. W tej części nieścisłość dotycząca nazw IWO i exIWO pojawia się dość często.

Trzecia część rozdziału dotyczy wyników eksperymentów dla najważniejszym, w kontekście pracy, zagadnieniu tj. dobieraniu kolejności złączeń. Doktorant wprowadza w dobór danych testowych, następnie prezentuje metodykę testów. Pewną słabością przeprowadzonych testów jest sprzęt na którym testy były realizowane - laptop. W opinii recenzenta przeprowadzenie testów na architekturze bardziej serwerowej (np. na komputerze, na którym realizowane były testy problemu komiwojażera) mogłyby dać trochę inne wyniki. Jednakże nie powinny one raczej wpłynąć istotnie na końcowy obraz. Algorytm exIWO okazuje się czasem być efektywniejszym, czasem mniej efektywnym. Najciekawszym wynikiem wyłaniającym się z testów jest rozdzielenie jakości

Ur

optymalizacji dokonanej przez algorytmy zależnie od kształtu grafu budowanego przez złączenia. Fakt ten powinien zostać bardziej uwypuklony w treści pracy.

Rozdział siódmy rozprawy podsumowuje uzyskane wyniki i kończy pracę.

Dodatki prezentują szczegółowe wyniki testów i omawiają aplikacje testowe.

Podsumowanie rozprawy

Przedstawiona do recenzji rozprawa porusza istotny problem optymalizacyjny. Opracowane w niej rozszerzenie algorytmu IWO okazuje się być efektywnym narzędziem nie tylko w optymalizacji kolejności złączeń tabel w bazach danych, ale również w innych klasycznych problemach optymalizacyjnych. Otrzymane wyniki zostały zaprezentowane w kilkunastu artykułach naukowych o zasięgu międzynarodowym. Jeden z artykułów ukazał się w czasopiśmie o współczynniku IF 1,73. Wspominane w omówieniu wyników rozprawy uwagi są drobiazgami wobec całości pracy wykonanej przez autora i nie mają istotnego wpływu na pozytywną ocenę wyników.

Konkluzja

Recenzowana rozprawa spełnia wszystkie formalne i zwyczajowe normy stawiane w naszym środowisku rozprawom doktorskim. Wnoszę o dopuszczenie magistra Daniela Kostrzewy do dalszych etapów przewodu doktorskiego.



Piotr Wiśniewski

Piotr Wiśniewski