

Prof. dr hab. inż. Franciszek Seredyński
Uniwersytet Kardynała Stefana Wyszyńskiego w
Warszawie
Wydział Matematyczno-Przyrodniczy.SNS
Instytut Informatyki
f.seredynski@uksw.edu.pl

Warszawa, 22.11.2018

Recenzja rozprawy doktorskiej

mgr inż. Daniela Grzonki

zatytułowanej:

Inteligentne systemy monitoringu procesów harmonogramowania w rozproszonych środowiskach dużej skali w ujęciu wieloagentowym

1. Problem badawczy i jego znaczenie

Chmury obliczeniowe to aktualnie najbardziej reprezentatywna i intensywnie rozwijana koncepcja obliczeń równoległych i rozproszonych, która zakłada przeniesienie całego ciężaru stworzenia infrastruktury komputerowej oraz oprogramowania z użytkownika na właściciela udostępniającego użytkownikowi dostęp do zasobów obliczeniowych w postaci płatnych usług. Tworzenie i użytkowanie tej coraz bardziej powszechnej platformy obliczeniowej związane jest z koniecznością rozwiązania szeregu problemów związanych z zapewnieniem wysokiej wydajności systemu, bezpieczeństwa, sprawnym zarządzaniem rozproszonego systemu, itp.

Doktorant w swojej rozprawie proponuje podejście łączące kwestię zapewnienia wysokiej wydajności systemu z kwestią jego zarządzania. Zapewnienie wysokiej wydajności systemu związane jest z rozwiązaniem problemu harmonogramowania zadań użytkowników w rozproszonym systemie komputerowym. Problem ten należy do klasy problemów NP.-trudnych i dlatego doktorant proponuje użycie metaheurystyki, która w rozsądnym czasie określi przybliżone rozwiązanie tego problemu. Doktorant zakłada, że działanie systemu odbywa się w czasie rzeczywistym i w związku z tym istnieje konieczność wielokrotnego wykonywania algorytmu harmonogramowania, co musi być poprzedzone zbieraniem danych o rozproszonym systemie. W celu monitorowania stanu rozproszonego systemu doktorant proponuje system wieloagentowy, którego agenty charakteryzują się pewną dozą inteligencji przetwarzając lokalną informację o stanie poszczególnych maszyn i przesyłają tę informację w celu aktualizacji danych wejściowych dla problemu harmonogramowania.

Tak więc doktorant zaproponował interesujące rozwiązanie problemu monitorowania w czasie rzeczywistym systemu rozproszonego z użyciem podejścia wieloagentowego wspomagającego proces harmonogramowania zadań w takich systemach. Uzyskane wyniki potwierdzają tezę rozprawy o efektywności takiego podejścia dla środowisk obliczeniowych dużej skali jakimi są dzisiaj chmury obliczeniowe. Mają one istotne znaczenie teoretyczne i praktyczne dla współczesnej informatyki.

2. Wkład autora

Doktorant w swojej rozprawie zaproponował nowatorskie podejście do problemu dynamicznego zarządzania zasobami w rozproszonych środowiskach obliczeniowych dużej skali, o które ubiegają się użytkownicy. Istotą zaproponowanego podejścia jest system wieloagentowy, który monitoruje w czasie rzeczywistym procesy alokacji rozproszonych zasobów użytkownikom systemu. Główne wyniki własne doktoranta znajdują się w rozdziałach 6 i 7.

Za najważniejsze wyniki uzyskane przez doktoranta w rozprawie uważam:

- Opracowanie nowatorskiej koncepcji inteligentnego systemu zarządzania zasobami rozproszonego środowiska obliczeniowego z użyciem systemu wieloagentowego składającego się z 5 typów agentów (Rozdział 6)
- Opracowanie oryginalnego algorytmu ewolucyjnego wykorzystującego tzw. podejście Michigan do rozwiązania problemu optymalizacji kombinatorycznej związanego z tworzeniem harmonogramów alokacji zadań do wirtualnych maszyn systemu rozproszonego (agent Ag0) (Rozdział 6)
- Zastosowanie sieci neuronowej (agent Ag3) do przewidywania opóźnień wykonywania zadań w systemie (Rozdział 6)
- Wyczerpujące badania eksperymentalne zaproponowanego wieloagentowego systemu monitorowania i wspierania procesów harmonogramowania zadań w systemie rozproszonym (Rozdział 7).

Uzyskane przez doktoranta wyniki w sposób istotny rozszerzają aktualną wiedzę o możliwościach zastosowania inteligentnych systemów zarządzania rozproszonymi środowiskami obliczeniowymi, a w szczególności chmurami obliczeniowymi. Wyniki te były prezentowane na szeregu międzynarodowych konferencjach i publikowane w ich materiałach. O dużej wadze uzyskanych wyników świadczy również fakt publikacji ich w czasopiśmie, w tym w dwóch renomowanych czasopiśmie z tzw. Listy Filadelfijskiej.

3. Poprawność

Rozprawa ma charakter teoretyczno-eksperymentalny. Doktorant opracował szereg modeli formalnych dotyczących różnych aspektów zaproponowanego systemu wieloagentowego monitorowania procesu alokacji zasobów oraz kilka algorytmów zorientowanych na rozwiązanie poszczególnych kwestii związanych z opracowanymi modelami. W celu weryfikacji zaproponowanego systemu wieloagentowego monitorowania procesu alokacji zasobów doktorant zrealizował symulator, który następnie użył do przeprowadzenia wyczerpujących badań symulacyjnych.

Badany symulator posiada znaczną liczbę parametrów, w tym parametrów mających charakter zmiennych losowych. W celu uzyskania wiedzy o symulowanym systemie doktorant przygotował szczegółowy plan przeprowadzenia eksperymentów, opracował szereg wariantów realizacji eksperymentów, przygotował odpowiednie dane testowe, a następnie wykonał badania eksperymentalne. W celu uzyskania statystycznej wiarygodności wyniki eksperymentów były uśredniane na podstawie kilkudziesięciu uruchomień symulatora dla tych samych danych

wejściowych. Z całą pewnością można sądzić, że dzięki temu uzyskane wyniki są poprawne i wiarygodne.

Uzyskane wyniki zostały przedstawione w formie ciekawych wykresów dających łatwy wgląd w różne aspekty działania modelowanego systemu.

4. Wiedza kandydata

Doktorant przygotowując swoją rozprawę wykazał się dużą wiedzą dotyczącą obszarów związanych z tematyką rozprawy.

W Rozdziale 3 znajdujemy obszerny przegląd literaturowy aktualnego stanu problematyki związanej z monitorowaniem rozproszonych środowisk obliczeniowych ze szczególnym uwzględnieniem chmur obliczeniowych. Doktorant szczególną uwagę poświęca rozwiązaniom problemu monitorowania wykorzystującym koncepcje systemów wieloagentowych. Przytacza aktualnie istniejące taksonomie systemów monitorowania chmur obliczeniowych oraz formułuje własny model systemu monitorującego.

Rozdział 4 poświęcony jest systemom wieloagentowym. Doktorant na podstawie istniejącej literatury przedstawia koncepcję autonomicznego agenta, środowisko działania agenta oraz dokonuje przeglądu zaproponowanych w literaturze typów agentów. Rozdział kończy się przedstawieniem istniejących koncepcji systemów wieloagentowych.

W Rozdziale 5 doktorant dokonuje przeglądu literaturowego problematyki harmonogramowania zadań w rozproszonych środowiskach obliczeniowych. Przedstawia klasyfikacje problemów harmonogramowania zadań, podaje definicję problemu harmonogramowania oraz kryteria rozwiązania tego problemu optymalizacyjnego. Formułuje model zadania oraz model jednostki obliczeniowej oraz przedstawia koncepcję zastosowania algorytmu ewolucyjnego do rozwiązania problemu harmonogramowania niezależnych zadań.

W podrozdziale 6.5.3 odnajdujemy przegląd literaturowy problematyki sieci neuronowych, a w podrozdziale 7.1 przegląd literaturowy rozwiązań problemu harmonogramowania, w tym rozwiązań z użyciem algorytmów ewolucyjnych.

Cytowana w rozprawie literatura obejmuje 129 pozycji bibliograficznych.

5. Uwagi krytyczne

Patrząc całościowo na strukturę rozprawy można stwierdzić, że jest ona logiczna, a przedstawienie kolejnych zagadnień jest na ogół precyzyjne. Jednak niektóre kwestie należałoby bardziej szczegółowo opracować w celu lepszego ich zrozumienia. Z przyjemnością zapoznałbym się z bardziej szczegółowymi wyjaśnieniami dotyczącymi następujących kwestii:

- Doktorant wielokrotnie w pracy podkreśla kwestię funkcjonowania systemu w czasie, patrz, np. str. 77: „, zakłada się ciągłość pracy systemu w czasie, tj. wysyłanie przez użytkowników zadań, ich gromadzenie, formowanie pakietów, następnie harmonogramowanie...”. Nigdzie jednak nie precyzuje tego w jakich momentach czasowych napływają zadania, czy te momenty mają charakter deterministyczny czy podlegają być może jakiemuś rozkładowi

- Nie jest jasne jak należy interpretować, z punktu widzenia tworzenia harmonogramu rozdziału zadań między wirtualnymi maszynami, założenie o rozkładzie normalnym dostępnych zasobów obliczeniowych systemu (Tab. 7.1). Rozumiem, że w momencie utworzenia pakietu zadań przypisuje się im zapotrzebowania na zasoby obliczeniowe zgodnie z Tab. 7.2, generuje się zgodnie Tab. 7.1 informację o dostępnych zasobach obliczeniowych, a następnie rozwiązuje się deterministyczny problem przydziału zadań do wirtualnych maszyn. W momencie sformowania kolejnego pakietu, jak rozumiem, ta procedura jest powtarzana. Prawdopodobnie ponownie uruchomiany jest proces udostępnienia zasobów obliczeniowych zgodnie z Tab. 7.1 ? Co się dzieje jeżeli dostępne zasoby są niewystarczające do sprostania zapotrzebowania ? Czy istnieją założenia o maksymalnych rozmiarach zasobów obliczeniowych rozproszonego systemu ?
- W zaproponowanym systemie monitorującym wykorzystywane są agenty Ag3 oraz Ag4 w celu przewidywania rzeczywistych opóźnień jednostek obliczeniowych. Nie jest jasne skąd się biorą opóźnienia jeżeli, jak rozumiem, rozwiązywany jest problem deterministycznego przydziału zadań do maszyn; nie jest jasne w jaki sposób generuje się te opóźnienia i z użyciem jakiego formalnego mechanizmu
- Doktorant zaproponował użycie do rozwiązania problemu harmonogramowania ewolucyjny algorytm wykorzystującego tzw. podejście Michigan, który jeżeli był stosowany to raczej w uczeniu maszynowym. Nie są mi znane prace dotyczące zastosowania tego podejścia do optymalizacji, ani też na czym polega ewentualny „engine search” tego podejścia. Nie jest jasne jakie były motywacje zastosowania podejścia Michigan zamiast podejścia Pitt. W podejściu Michigan, które zastosował doktorant osobniki populacji tworzą jedno rozwiązanie, tak jak to jest np. w Strategiach ewolucyjnych (SE) czy w symulowanym wyżarzaniu (SA). Aby mówić o mechanizmie ewolucji darwinowskiej w tym algorytmie musiałyby, moim zdaniem istnieć dodatkowe rozwiązanie konkurujące z rozwiązaniem z populacji, tak jak to jest w ES czy SA. Operatory genetyczne selekcji, krzyżowania i mutacji używane przez doktoranta w stosowanym podejściu prawdopodobnie mają lokalny charakter i w swoim działaniu mają charakter podobny do swoistej mutacji, która modyfikuje bieżące rozwiązanie
- Jak wygląda pojedynczy przebieg algorytmu ewolucyjnego dla typowych instancji problemów harmonogramowania ? Celowe byłoby porównanie efektywności tego algorytmu z podejściem typu Pitt, czy innymi metaheurystykami tej samej klasy.

6. Uwagi redakcyjne

Praca jest napisana poprawnie językowo. Zauważyłem kilka niedociągnięć językowych:

- Str. 3: „...generowania bardziej optymalnych harmonogramów”
- Str. 10: „system monitoringu może być opisany za pomocą procesu”
- Str. 71: „zaproponowano reprezentację populacji dla procesu ewolucyjnego” – proponuje się raczej reprezentację rozwiązania
- Str. 72: „Przyjętą funkcją oceny populacji jest” – ocenia się zwykle osobniki z populacji, ale w tym przypadku po prostu rozwiązanie
- Str. 73: „Celem algorytmu jest znalezienie populacji minimalizującej” – raczej znalezienie rozwiązania, a nie populacji
- Str. 73: „Operator mutacji jest pomijany”, a na str. 82 mamy: „zaproponowano operację mutacji”
- Str. 74: „jednym z ... założeń jest jego umiejscowienie w czasie”
- Str. 82: „nie ma bezpośredniego pływu”

7. Podsumowanie

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez artykuł 13 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym (z późniejszymi zmianami)¹ moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

A. Czy rozprawa zawiera oryginalne rozwiązanie problemu naukowego? (wybierz jedną opcję stawiając znak X)

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka ?

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

C. Czy kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej?

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

Ponadto, biorąc pod uwagę takie czynniki jak ważność tematyki badawczej rozprawy, innowacyjne podejście do rozwiązania postawionego problemu, oryginalność uzyskanych wyników potwierdzonych publikacjami w renomowanych czasopismach z listy filadelfijskiej rekomenduję wyróżnienie rozprawy doktorskiej.



Podpis

¹ http://www.nauka.gov.pl/g2/oryginal/2013_05/b26ba540a5785d48bee41aec63403b2c.pdf