

RECENZJA

rozprawy doktorskiej mgr inż. Mateusza Paciorka pt.:
"Scalable simulation of social phenomena based on signal propagation modeling"

Zawartość rozprawy

Rozprawa doktorska mgr inż. Mateusza Paciorka poświęcona jest zagadnieniom symulacji, a dokładniej dyskretnym, przestrzennym metodom modelowania i symulacji opartym na podejściu agentowym. Jest to niewątpliwie problem ważny i interesujący z teoretycznego punktu widzenia, a wymóg uwzględnienia struktury przestrzennej czyni go dodatkowo atrakcyjnym dla praktyków, jako modelu bliskiego rzeczywistym systemom.

Recenzowana rozprawa liczy 100 stron i składa się z ośmiu rozdziałów, wniosków, trzech dodatków zawierających przykłady zastosowań, spisów tabel i rysunków oraz bibliografii. W liczącej 76 pozycji bibliografii, zawierającej istotne pozycje z zakresu tematyki rozprawy, występują 6 współautorskich publikacji doktoranta.

Z merytorycznego punktu widzenia w rozprawie można wyróżnić pięć części oraz dodatki. Pierwsza, obejmująca rozdział 1 oraz 2, ma charakter wprowadzający. Przedstawiono w nim obecny stan wiedzy na temat problematyki systemów agentowych ze szczególnym uwzględnieniem metod symulacji z ich wykorzystaniem. Zamieszczono przegląd literaturowy współcześnie stosowanych metod i algorytmów. Dokonano także omówienia pokrewnych do omawianej tematyki zagadnień zrównoleglenia metod symulacji opartych na podejściu agentowym oraz zagadnień propagacji sygnału w dwuwymiarowych sieciach. Autor wskazuje na podobieństwa i różnice opisywanego podejścia w stosunku do teorii automatów komórkowych.

Cześć druga, obejmująca rozdziały 3 i 4, dotyczy wykazania poprawności definicji rozproszonego przestrzennego dyskretnego modelowania i symulacji opartych na podejściu agentowym, a także przedstawienia ulepszeń przedstawionych metod wykorzystywanych jako podstawa proponowanego rozwiązania. Ulepszenia te są związane głównie z uogólnieniem struktury modelu środowiska metody jak też z wymiarowością modeli.

Trzon pracy, rozdziały 5,6 i 7, stanowią część trzecią i prezentują główne wyniki autora. W rozdziale 5 przedstawione zostały rozwiązania problemów postawionych na początku rozprawy. Rozdział 6 obejmuje opis schematu testowania poprawności metody rozpraszania obliczeń i oceny rozwiązania przy użyciu metod statystycznych na bazie wybranego zagadnienia. Z kolei w rozdziale 7 zaprezentowane zostały wyniki testów skalowalności rozważanej metody.

Rozdział 8, czwarta część pracy, przedstawia zastosowania: modele oparte na rzeczywistych scenariuszach wykorzystujących zaproponowaną metodę. Przedstawiono także rozszerzenia związane z dalszymi rozszerzeniami rozważanych modeli.

Rozdział 9, krótka część piąta, prezentuje podsumowanie badań oraz wnioski. Całość uzupełniają trzy dodatki przedstawiające założenia przykładów wykorzystywanych do testów.

Główną tezę rozprawy jest stwierdzenie, że możliwe jest stworzenie silnie skalowalnej metody rozproszenia obliczeń dla dyskretnych modeli przestrzennych opartych na agentach i symulacjach, która nie wpływa na poprawność symulacji, zapewniając, że uzyskane wyniki nie będą się różnić od wersji nierozproszonej. Wykorzystanie koncepcji propagacji sygnału pozwala na reprezentowanie przy pomocy zaproponowanej metodologii różnorodnych modeli, a szczególnie modeli zjawisk społecznych takich jak środowiska miejskie czy symulacje sztucznego życia.

Za najważniejsze wyniki zamieszczone w rozprawie uważam:

1. Opracowanie definicji poprawności rozproszonego ABMS (Agent-Based Modeling and Simulation, modelowania i symulacji opartych na podejściu agentowym).
2. Uogólnienie struktury modelu środowiska w metodzie Xinuk core (zaproponowanej w pracy J. Bujas, D. Dworak, W. Turek, and A. Byrski. High-performance computing framework with desynchronized information propagation for large-scale simulations. *Journal of Computational Science*, 32:70–86, 2019.)
3. Skonstruowanie skalowalnej zuniwersalizowanej metody symulacji. Głównym wkładem Autora jest tu metoda symulacyjna dla modeli agentowych zapewniająca dobrą skalowalność, zachowującą także ograniczenia modelu (tzn. jest poprawna) oraz obsługę szerokiego zakresu przypadków poddawanych symulacjom.
4. Przedstawienie wyników badań eksperymentalnych rozważanych metod. W rozprawie doświadczalnie wykazano skuteczność zaproponowanego podejścia agentowego (choć niestety nie sformułowano teoretycznych oszacowań jakości możliwych do uzyskania wyników, tj. teoretycznych wskaźników realizmu samych symulacji).
5. Przedstawienie przykładów praktycznych użycia omawianych w pracy problemów i algorytmów w zastosowaniu do zagadnień dla problemów ewolucji i symulacji: (1) królików i sałaty, (2) organizmów jednokomórkowych, (3) ewakuacji pożarowej.

Zarówno strona informatyczna, jak i matematyczna rozprawy nie budzi zastrzeżeń. Zaproponowane metody i miary ich efektywności są bezpośrednio wykorzystywane w konstrukcjach algorytmów. Bibliografia zawiera najważniejsze obecnie publikacje z dziedziny systemów agentowych w odniesieniu do zagadnień symulacji. Należy także podkreślić umiejętność zaplanowania i przeprowadzenia przez Autora rozprawy złożonych eksperymentów obliczeniowych. Dysertacja zawiera ponadto wiele rysunków, tablic i wykresów, które dodatkowo ilustrują i ułatwiają analizę uzyskanych wyników.

Część zamieszczonych w rozprawie wyników była prezentowana na konferencjach oraz została opublikowana w pracach współautorskich zamieszczonych w 3 specjalistycznych czasopismach (w tym 2 posiadających Impact Factor) oraz materiałach konferencji międzynarodowych.

Rozprawa została napisana poprawnie, jednak Autor nie ustrzegł się pewnych niedociągnięć, np.

1. W Rozdziale 3, odnoszącym się do wykazania poprawności definicji „rozproszonego przestrzennego dyskretnego modelowania i symulacji opartych na podejściu agentowym” (*Correctness in distributed discrete spatial ABMS*) Autor swoje rozważania opiera na

wynikach eksperymentów opracowanych w formie wykresów pudełkowych. Można się zastanawiać czy tak przyjęta metodologia wykazania poprawności jest wystarczająca pod względem metodologicznym, gdyż – jak zawsze przy posługiwaniu się przykładami testowymi – wszelkie wnioski odnoszą się tylko do rozpatrywanych przykładów i raczej nie mają charakteru ogólnego. Chyba, że zastosowuje się aparat statystyczny (np. testy istotności różnic), tak jak go użyto w testach w rozdziale 6.

2. Na początku Rozdziału 4 Autor pisze: *“It is impossible to rearrange the cells into a single 2D plane so that all edges are represented by a neighborhood between two cells.”* Być może jest to stwierdzenie zbyt stanowcze, bowiem niektóre grafy wydające się być na pierwszy rzut oka nie-planarne, okazują się jednak możliwe do przedstawienia na płaszczyźnie – szczególnie, gdy połączenia pomiędzy „poziomymi 2D” występują rzadko (schody, drabiny, etc.). np. dla konfiguracji z rysunku.1 ale z połączeniem pomiędzy poziomami w narożnej komórce. Oczywiście jest to sytuacja trywialna, ale lecz użycie sformułowania „jest niemożliwe” wymaga dowodu lub choćby silnej argumentacji, której tu zabrakło.
3. Przyjęcie w rozdziale 7.2 jako czasu referencyjnego, sekwencyjnego T_1 , wartości $240T_{240}$ skutkuje uzyskiwaniem przyspieszenia ponadliniowego (zob. Rys. 7.1). Trudno odnieść się do tego rezultatu gdyż chcąc nie chcąc uwzględnia on (240 razy) narzuty związane z rozproszeniem obliczeń wersji 240-rdzeniowej (komunikację etc.), których nie posiada normalnie kod sekwencyjny. Skutkiem jest zawyżona wartość wyznaczonych przyspieszeń. Pytanie – o ile zawyżona. Oczywiście inną kwestią jest czy Autor miał technicznie możliwość wykonania porównania do sekwencyjnej wersji kodu. Podsumowując – ta uwaga ma charakter raczej dyskusji otrzymanych wyników (konkretnie przyspieszenia większego niż p przy tak liczonym przyspieszeniu) a nie zarzutu merytorycznego.
4. Metodologia weryfikacji poprawności działania metody rozproszonej zaprezentowana w rozdziale 6. oparta jest na testach statystycznych i studium przypadku (królików i sataty). Dyskusyjna może być kwestia, na ile te wyniki są możliwe do uogólnienia na inne problemy (o innej strukturze) oraz na ile przyjęta statystyczna metodologia *gwarantuje* poprawność (tj. identyczność wyników wersji sekwencyjnej i zrównoleglonej) wyników.

Wymienione powyżej uwagi merytoryczne i redakcyjne nie mają zasadniczego wpływu na wagę i jakość przedstawionych w rozprawie wyników i nie wpływają też w znaczący sposób na ogólną bardzo pozytywną jej ocenę. Autor zastosował metody badawcze właściwie dobrane do poruszanej problematyki, a dobór cytowanej literatury nie budzi zastrzeżeń.

Jedną z głównych tez pracy było stwierdzenie, że możliwe jest opracowanie efektywnych metod rozproszenia obliczeń w symulacji systemów agentowych dla dyskretnych modeli przestrzennych. Cel ten został osiągnięty poprzez zaproponowanie metod zrównoleglania obliczeń i analizie wyników badań empirycznych.

Stwierdzam ponadto, że recenzowana praca ma charakter naukowy i może być przedmiotem rozprawy doktorskiej. Wyniki uzyskane przez Autora w istotny sposób uzupełniają istniejący stan wiedzy. Należy więc uznać, że podjęcie zawartego w tezach pracy problemu badawczego jest w pełni uzasadnione.

Konkluzja

Uważam, że rozprawa stanowi solidne, oparte na wielu podejściach, rozwiązanie rozważanego problemu naukowego. Praca zawiera szereg pomysłowych rozwiązań, a niektóre z przedstawionych

podejść wymagały zastosowania zaawansowanego formalnie modelowania matematycznego. Stwierdzam, że rozprawa spełnia wymogi Ustawy z dnia 20 lipca 2018 r. - Prawo o szkolnictwie wyższym i nauce (Dz.U.2022.0.574 Art. 186. - Warunki nadania stopnia doktora) i wnoszę o dopuszczenie mgr inż. Mateusza Paciorka do dalszych etapów przewodu doktorskiego. Ponadto, biorąc pod uwagę fakt opublikowania części wyników Autora w prestiżowych czasopismach i materiałach konferencyjnych, a także silną aplikowalność uzyskanych wyników, wnoszę o wyróżnienie rozprawy.

Uojacek Boryło

Lista uwag szczegółowych:

1. str. 21 i dalej, brak interpunkcji we wzorach
2. str. 21, równanie (2.3) jest niezrozumiałe (brak równości lub nierówności), podobnie akapit pod równaniem (nie wiadomo co to znaczy spełniać równanie (2.3))
3. str. 25, skróty HMS, FSF, etc. w tabelach 3.1-3.3 nie są opisane, ich znaczenie pojawia się dopiero na stronie 91
4. str. 32, „To this end, the set of cells C is defined...” – niezrozumiałe (“do tego końca”?)
5. str. 41-42, opisywany jest algorytm, ale brak jest jego pseudokodu lub schematu blokowego, które byłyby bardzo przydatne w jego analizie



