



Wrocław, 22 czerwca 2022 roku

dr hab. inż. Marcin Hernes, prof. UEW
Katedra Zarządzania Procesami
Uniwersytet Ekonomiczny we Wrocławiu

RECENZJA

rozprawy doktorskiej mgr inż. Thanh Ngo Nguyena
pod tytułem „Developing efficient methods for mining inter-transaction
patterns” (“Opracowanie efektywnych metod eksploracji częstych wzorców
międzytransakcyjnych”)

Promotor: prof. dr hab. inż. Ngoc Thanh Nguyen

Kopromotor: prof. dr Loan T.T. Nguyen

1. Podstawa sporządzenia recenzji

Podstawą sporządzenia recenzji jest pismo Przewodniczącego Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja, prof. dra hab. inż. Michała Woźniaka, Politechnika Wrocławska, informujące o powołaniu mnie na recenzenta rozprawy doktorskiej mgr inż. Thanh Ngo Nguyena. W piśmie poinformowano mnie również, że przewód doktorski przeprowadzany jest w dyscyplinie informatyka (przewód realizowany zgodnie z przepisami Ustawy o stopniach naukowych i tytule naukowym oraz stopniach i tytule w zakresie sztuki, Dz. U.2003, nr 65 poz.565 ze zm.).

2. Określenie trafności wyboru i oryginalności problemu badawczego podjętego w rozprawie

Problem badawczy podjęty w pracy doktorskiej dotyczy eksploracji częstych wzorców międzytransakcyjnych. Problem ten dotyczy szerszego zagadnienia eksploracji reguł asocjacyjnych, którego celem jest znalezienie związku między elementami w bazie danych transakcji. Należy zauważyć, że reguły asocjacyjne wykorzystywane są we wspomaganie realizacji procesów w wielu obszarach działalności, m.in. w promocji i marketingu produktów, e-learningu, bankowości. Zatem problem analizowany w niniejszej pracy doktorskiej jest istotny zarówno z punktu widzenia badań naukowych, jak i praktyki gospodarczej. Jednak większość metod eksploracji częstych wzorców koncentruje się tylko na wzorcach występujących w ramach tych samych transakcji w bazie danych transakcji, tj. wzorcach wewnątrz transakcji. Problem wykorzystania eksploracji częstych wzorców

międzytransakcyjnych jest analizowany w dotychczasowej literaturze przedmiotu w sposób niewystarczający. Zatem uzasadnione są badania w celu opracowania metod eksploracji częstych wzorców międzytransakcyjnych, charakteryzujących się niską złożonością obliczeniową oraz zmniejszeniem przestrzeni wyszukiwania w celu ograniczenia użycia pamięci.

Tym samym stwierdzam, że problem badawczy podjęty w rozprawie badawczej jest aktualny i ważny, a jego oryginalność polega na opracowaniu nowych metod eksploracji częstych wzorców międzytransakcyjnych bardziej wydajnych pod względem zużycia zasobów, niż istniejące metody.

W mojej ocenie obszar teoretyczny obejmujący eksplorację danych, jak i obszar aplikacyjny dotyczący usprawniania procesu analizy danych w różnych obszarach działalności (sprzedaż, wypadki drogowe) dobrze pozycjonują recenzowaną rozprawę w dyscyplinie informatyka. Autor wykazał się dobrym rozeznaniem i szeroką znajomością omawianych zagadnień.

3. Ocena poprawności i kompletności celów oraz hipotez badawczych

Cel główny pracy został określony w następujący sposób:

„Celem niniejszej pracy jest opracowanie skutecznych metod eksploracji częstych wzorców międzytransakcyjnych, eksploracji częstych zamkniętych wzorców międzytransakcyjnych oraz eksploracji częstych zamkniętych maksymalnych wzorców międzytransakcyjnych”.

Sformułowano również 4 cele pomocnicze.

Ponieważ recenzowana rozprawa doktorska ma charakter metodologiczny, w mojej ocenie cel główny oraz cele pomocnicze zostały sformułowane prawidłowo oraz zostały osiągnięte w toku rozważań prowadzonych w ramach recenzowanej dysertacji. Zawierają zarówno elementy teoretyczne, metodyczne, jak i praktyczne, zatem są spójne i kompletne.

Moja uwaga dotyczy braku podziału celów pomocniczych na cele teorio-poznawcze, metodyczne, utylitarne.

Hipoteza została sformułowana następująco: „Możliwe jest zwiększenie wydajności pod względem czasu obliczeń i wykorzystania pamięci podczas eksploracji wzorców międzytransakcyjnych poprzez zastosowanie zwartych struktur danych i skutecznych strategii przycinania podczas procesu eksploracji.”.

Sformułowana hipoteza badawcza jest spójna z celami rozprawy. Hipoteza została poprawnie zweryfikowana w drodze przedstawionych badań oraz analiz. Pewien niedosyt budzi jednak zastosowanie w hipotezie zwrotu „Możliwe jest...”, ponieważ sugeruje to weryfikację hipotezy w sposób 0/1 (możliwe/nieosiągalne). Hipotezę można było sformułować w sposób sugerujący w większym zakresie poszerzenie aktualnej wiedzy.

4. Ocena poprawności struktury rozprawy

Recenzowana rozprawa doktorska składa się ze wstępu, trzech rozdziałów właściwych, zakończenia oraz spisów: bibliografii tabel i rysunków. Ogółem rozprawa liczy 164 strony, w tym tekstu głównego: 152 strony. Strona tytułowa, spisy treści, tabel i rysunków oraz skrótów i akronimów liczą 12 stron. Praca została skonstruowana zgodnie ze standardami prac naukowych z dyscypliny

informatyka i prawidłowo odzwierciedla etapy procesu prowadzenia badań naukowych, począwszy od rozważań teoretycznych oraz nakreślenia kontekstu badań, poprzez opracowanie autorskich metod eksploracji częstych wzorców międzytransakcyjnych po ich weryfikację z wykorzystaniem baz danych z rzeczywistych procesów.

W odniesieniu do drugiego rozdziału (Literature review) przyjęto perspektywę teoretyczno-poznawczą, natomiast rozdziały 3 i 4 dotyczą badań przeprowadzonych przez Autora, dotyczących konstrukcji, zastosowania i weryfikacji opracowanych metod.

W mojej ocenie kolejność rozdziałów jest prawidłowa i umożliwia prowadzenie wywodu w zakresie realizacji celów rozprawy oraz weryfikacji hipotezy. Z punktu widzenia metodologii badań naukowych tytuły rozdziałów i podrozdziałów są sformułowane prawidłowo.

Moja uwaga dotyczy jednak tytułów następujących podrozdziałów/paragrafów w rozdziale 2 i 3:

2.2.7. N-List based algorithms for mining frequent patterns (Deng et al., 2012)

2.3.4 The ITP-Miner algorithm (Lee, A. J. T. & Wang, C. S., 2007)

2.3.5 The PITP-Miner algorithm (C. S. Wang & Chu, 2011)

2.3.6 The ICMiner algorithm (Lee et al., 2008)

3.2.1 DITP-Miner algorithm (Nguyen, T. N., Nguyen, L. T. T., & Nguyen, N. T., 2017)

3.2.2 NL-ITP-MINER Algorithm (Nguyen, T. N., Nguyen, L. T. T., Vo, B., Nguyen, N. T. et al., 2020).

W tytułach podrozdziałów/paragrafów nie powinno być odwołań do literatury przedmiotu (odwołania te należało zrobić w tekście podrozdziałów/paragrafów).

5. Ocena znajomości metodologii badań oraz przyjętych i zastosowanych metod badawczych

Autor zastosował metody badawcze takie, jak analiza literatury przedmiotu, metody modelowania, metody symulacyjne, metody projektowania systemów informatycznych, eksperymenty badawcze.

Jednakże niedosyt budzi brak opisu metodyki badawczej (kroki badawcze oraz metody badawcze) w odniesieniu do realizacji celów pracy. Metodyka badawcza powinna zostać scharakteryzowana w rozdziale 1 (Introduction).

Podsumowując, zastosowane w pracy metody badawcze uważam za prawidłowe gdyż umożliwiły w konsekwencji realizację celów oraz hipotez sformułowanych w rozprawie. Jednakże ze względu na brak w recenzowanej rozprawie opisu metodyki badań formułuję pytanie do Doktoranta:

Jaka metodyka badawcza została zastosowana w pracy i jakie jest uzasadnienie zastosowanych metod badawczych?

6. Ocena znajomości, doboru, analizy i interpretacji źródeł literaturowych

Zestawienie bibliograficzne pracy liczy 93 pozycje literaturowe. Są to głównie pozycje publikowane w czasopismach i wydawnictwach o zasięgu międzynarodowym. Tematyka cytowanych prac związana jest ściśle z obszarami badawczymi podjętymi w rozprawie doktorskiej. Pozycje literatury są aktualne. Analiza i interpretacja źródeł literaturowych jest prawidłowa. Autor w trakcie

rozważań w rozprawie dokonuje ich analizy i syntezy oraz przedstawia własne wnioski i krytyczną ocenę analizowanych pozycji literatury.

Podsumowując, moja ocena znajomości, doboru, analizy i interpretacji źródeł literaturowych przez Autora jest pozytywna.

7. Ocena zawartości merytorycznej rozdziałów

Rozprawę rozpoczyna Wstęp, w którym zawarto uzasadnienie wyboru tematyki badań, cele, hipotezę, metodykę badań oraz strukturę rozprawy doktorskiej.

Rozdział drugi zatytułowano „Przegląd literatury”. W ramach niniejszego rozdziału Autor charakteryzuje kluczowe pojęcie określone w temacie pracy, czyli eksplorację częstych wzorców oraz eksplorację częstych wzorców międzytransakcyjnych. Na uwagę zasługuje dobrze przeprowadzone zdefiniowanie formalne pojęć oraz wskazanie zalet i ograniczeń istniejących metod eksploracji częstych wzorców międzytransakcyjnych.

Pewne wątpliwości budzi niejednoznaczność użytych pojęć. We wstępie pracy Autor używa sformułowania „metody eksploracji częstych wzorców”, podczas gdy w tytułach podrozdziałów 2.2. i 2.3 używa sformułowania „algorytmy częstych wzorców”. Uważam, że w tytułach tych podrozdziałów należało użyć sformułowania: „metody eksploracji częstych wzorców” i w tekście pracy wyjaśnić, że metody te opracowane są z wykorzystaniem algorytmów.

Dodatkowo, w podrozdziale 2.4 zatytułowanym „Podsumowanie” Autor przedstawił zawartość rozdziału. W rozprawie doktorskiej bardziej właściwe byłoby przedstawienie wniosków z rozdziału.

Rozdział 3 zatytułowano „Metody eksploracji częstych wzorców międzytransakcyjnych”. Wskazano różnice pomiędzy eksploracją częstych wzorców a eksploracją częstych wzorców międzytransakcyjnych. Następnie przedstawiono autorskie metody eksploracji częstych wzorców międzytransakcyjnych. Na uwagę zasługują opracowane podstawy teoretyczne metod, czyli twierdzenia, dowody twierdzeń, definicje. Algorytmy zostały opracowane na odpowiednim poziomie szczegółowości. Działanie metod zostało zilustrowane na przykładach. Przedstawione zostały wyniki eksperymentów przeprowadzonych z wykorzystaniem danych z repozytorium Frequent Itemset Mining Dataset Repository. Wyniki eksperymentów pozwalają wyciągnąć wnioski, że opracowana metoda NL-ITP-Miner umożliwia mniejsze zużycie zasobów pamięci przy krótszym czasie działania w porównaniu z pozostałymi metodami.

W podrozdziale 3.1. tekst na stronie 55 zaczynający się od słów “However, most of the methods for mining frequent patterns only focus on mining...” Jest powtórzeniem tekstu z podrozdziału 1.3. Autor mógł choćby zmienić treść przykładowych reguł R1 i R2.

W rozdziale czwartym, zatytułowanym „Metody eksploracji częstych wzorców międzytransakcyjnych w zbiorach skompresowanych” zostały przedstawione autorskie metody eksploracji częstych zamkniętych wzorców międzytransakcyjnych oraz eksploracji częstych maksymalnych wzorców międzytransakcyjnych. Podobnie jak w rozdziale 3 opracowano podstawy teoretyczne metod oraz algorytmy. Działanie metod zostało zilustrowane na przykładach. Przedstawione zostały wyniki eksperymentów przeprowadzonych, podobnie jak w rozdziale 3, z

wykorzystaniem danych z repozytorium Frequent Itemset Mining Dataset Repository. Wyniki eksperymentów pozwalają wyciągnąć wnioski, że w odniesieniu do problematyki eksploracji częstych zamkniętych wzorców międzytransakcyjnych opracowana metoda NCITM-Miner umożliwia mniejsze zużycie zasobów pamięci przy krótszym czasie działania w porównaniu z pozostałymi metodami. Z kolei w odniesieniu do problematyki eksploracji częstych maksymalnych wzorców międzytransakcyjnych opracowana metoda nMITP-Miner umożliwia mniejsze zużycie zasobów pamięci przy krótszym czasie działania w porównaniu z pozostałymi metodami.

W mojej ocenie wystąpiły pewne nieścisłości w dowodzie twierdzenia 4.2. (str. 95) Autor napisał " ...zgodnie z hipotezą...", z tym, że nie jest jasne, z którą hipotezą (hipoteza główna rozprawy nie dotyczy bowiem warunku określonego w twierdzeniu 4.2). Ten sam problem dotyczy twierdzenia 4.3 (str. 121). Dodatkowo w rozdziałach 3 i 4 Autor nie odnosił się do zagadnień związanych z weryfikacją hipotezy głównej. Powinien wskazać (na przykład w podsumowaniach tych rozdziałów), które wyniki potwierdzają hipotezę główną rozprawy.

W Zakończeniu dokonano podsumowania przeprowadzonych badań. Wskazano również wkład przeprowadzonych badań do nauki, ograniczenia tych badań oraz dalsze prace badawcze. W zakończeniu zabrakło mi również odniesienia do wyników potwierdzających weryfikację hipotezy głównej oraz wskazania implikacji dla nauki i praktyki. Stąd moje pytanie do Doktoranta:

Jakie są uwarunkowania implementacyjne opracowanych metod, w jakich klasach systemów informatycznych mogą być wykorzystane i jakie przykładowe procesy/decyzje mogą być wspomagane z wykorzystaniem tych metod.

Podsumowując, zawartość merytoryczna rozprawy jest spójna i pozwala na realizację celów oraz weryfikację hipotezy badawczej. Autor sposób sprawny przedstawia cały przeprowadzony cykl badawczy, jak również przejrzystość prezentuje wnioski z badań.

Bardzo pozytywnie oceniam publikację fragmentów wyników badań prezentowanych w rozprawie doktorskiej w międzynarodowych czasopiśmie i ich prezentację na międzynarodowych konferencjach. Umożliwiło to weryfikację uzyskanych przez Doktoranta wyników badań w międzynarodowym środowisku naukowym.

8. Ocena formalnej i redakcyjnej strony rozprawy

Praca napisana jest poprawnym językiem, a styl prowadzonego wywodu odpowiada kanonowi opracowań naukowych. W ujęciu formalnym całość prezentuje się zatem dobrze, spełniając wymogi przewidziane dla prac awansowych na stopień naukowy doktora. Autor nie ustrzegł się jednak pewnych nieścisłości, błędów językowych, terminologicznych i redakcyjnych, których przykłady wskazałem w dalszej części niniejszego punktu recenzji.

Autor często używa pojęcia "koszty obliczeniowe" ("computational costs"). Jest to sformułowanie niepoprawne znaczeniowo, gdyż koszt definiowany jest zawsze w ujęciu wartościowym (suma pieniędzy). W odniesieniu do problematyki rozpatrywanej w niniejszej rozprawie powinno używać się pojęcia „zużycie zasobów obliczeniowych”.

W rozdziale 2 brakuje odwołania w tekście pracy do rysunków od 2.10 do 2.19.

Występuje niejednolita interpunkcja po numerach rozdziałów/podrozdziałów/paragrafów.
Pozycje literatury powinny być ponumerowane.

9. Konkluzja

Podsumowując niniejszą recenzję stwierdzam, że problem badawczy podjęty w rozprawie jest istotny, aktualny i rozwiązany w oryginalny sposób. Zdefiniowane przez Doktoranta cele rozprawy doktorskiej zostały osiągnięte w sposób zgodny z kanonem procesu poznania naukowego. Autor rozprawy opanował umiejętność posługiwania się metodami wnioskowania naukowego, jak też zdobył niezbędny warsztat badawczy umożliwiający projektowanie i prowadzenie badań naukowych.

Do mocnych stron rozprawy zaliczam szczegółowe, wyczerpujące oraz wnikliwe opracowanie, i ewaluację autorskich metod eksploracji częstych wzorców międzytransakcyjnych. Do słabych stron rozprawy zaliczam niewystarczające przedstawienie metodyki badawczej.

Sformułowane przeze mnie uwagi krytyczne i pytania problemowe nie pomniejszają mojej bardzo pozytywnej oceny recenzowanej rozprawy doktorskiej i mają głównie charakter polemiczny, redakcyjny lub prowokujący do dalszych prac badawczych. Recenzowana rozprawa w mojej ocenie jednoznacznie posiada walory poznawcze, metodyczne i aplikacyjne. Autor wykazał się dobrą znajomością metodyki badań ilościowych.

W mojej opinii, recenzowana praca wnosi istotne wartości do dyscypliny informatyka. Stanowi ona oryginalne rozwiązanie problemu naukowego oraz wykazuje ogólną wiedzę teoretyczną Doktoranta w dyscyplinie informatyka oraz umiejętność samodzielnego prowadzenia pracy naukowej przez Autora.

Stwierdzam, że przedstawiona do oceny rozprawa doktorska mgr inż. Thanh Ngo Nguyena pod tytułem „Developing efficient methods for mining inter-transaction patterns” („Opracowanie efektywnych metod eksploracji częstych wzorców międzytransakcyjnych”) napisana pod kierunkiem prof. dr hab. inż. Ngoc Thanh Nguyena oraz prof. dr Loan T.T. Nguyen spełnia warunki określone w art. 13 Ustawy z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz stopniach i tytule w zakresie sztuki (Dz.U. z 2017 r., poz. 1789 z późniejszymi zmianami) i wnioskuję o dopuszczenie jej do dalszych etapów przewodu doktorskiego.

Manu Henies