

dr hab. inż. Julian Szymański, prof. PG
Politechnika Gdańska
Wydział Elektroniki Telekomunikacji i Informatyki

RECENZJA

Rozprawy doktorskiej Pana mgr. inż. Rafała Palaka
pt.: The impact of centralisation on the accuracy of collective prediction
promotor: prof. dr hab. inż. Ngoc Thanh Nguyen

1. Formalna podstawa wykonania recenzji

Recenzja przygotowana została na zlecenie przewodniczącego Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Politechniki Wrocławskiej, prof. dr hab. inż. Michała Woźniaka (pismo RDN-ITT/138/2022).

2. Uwagi ogólne

Rozprawa poświęcona jest problemom badania wpływu struktury kolektywu na jego decyzje. W literaturze przedmiotu brak jest dogłębnych studiów nad ilościową oceną wpływu rozproszenia kolektywu na jakość dokonywanych z jego użyciem prognoz, co wskazano w zawartym w rozprawie przeglądzie literatury. Zaproponowany w rozprawie model kolektywu jest interesującym wkładem w dziedzinę kolektywnej inteligencji a autorskie miary centralizacji stanowią rozwinięcie istniejących podejść i wnoszą do dziedziny poprawę szybkości przetwarzania struktur kolektywnych. Wykonane eksperymenty przeprowadzone zarówno na syntetycznych, jak i rzeczywistych danych wskazują na potencjał aplikacyjny zaproponowanego podejścia.

Uważam, że wybór tematu rozprawy jest trafny i uzasadniony aktualnym stanem wiedzy oraz potrzebami praktycznymi.

WPLYNĘŁO

09-09-2022

RDN-ITT/231/2022

3. Struktura rozprawy, ocena merytoryczna i metodyki badawczej

Rozprawa obejmuje 151 stron, ma właściwy układ, podzielona jest logicznie na 7 rozdziałów. Rozpoczyna się od wprowadzenia w tematykę kolektywnej inteligencji i przeglądu literatury zawierającego opis modeli kolektywnych i miar wykorzystywanych do ich charakteryzacji. Rozdział 4 opisuje zaproponowany przez doktoranta model kolektywu, w którym obok szczegółów teoretycznych zawarty został również projekt implementacji. W kolejnym rozdziale zdefiniowano wymagania stawiane miarom centralizacji wykorzystywanymi do opisu kolektywu. W rozdziale tym przedstawiono również opis autorskich miar centralizacji używanych do charakteryzacji kolektywów reprezentowanych poprzez grafy skierowane i nieskierowane.

Zaznaczyć należy, że praca zawiera błędy edytorskie (np.: Avail- able), stylistyczne jak i literowe (np. migration -> mitigation, underacted -> undirected, devise -> diverse), które miejscami utrudniają odbiór merytoryczny rozprawy i wymagają od czytającego domyślenia się intencji Autora.

Praca napisana jest w języku angielskim, co zdecydowanie zwiększa jej zasięg odbioru, jednak zawiera niezgrabności językowe, w szczególności używane są kalki językowe np.: Prediction performance -> prediction accuracy.

Rozprawa zawiera 127 przypisów bibliograficznych, w większości są one aktualne i związane z dziedziną. Obrazują to dobrą znajomość tematyki Autora. Zaznaczyć należy, że część przypisów jest niekompletna np. [94] (praca Autora) nie ma miejsca wydania, roku – nie wiadomo czy jest ona w druku czy jest to niedopatrzenie Doktoranta.

Tytuł pracy „*The impact of centralisation on the accuracy of collective prediction*” dobrze odzwierciedla tematykę rozprawy.

Teza postawiona w rozprawie brzmi: „Wzrost centralizacji kolektywu prowadzi do wzrostu zmienności kolektywnej predykcji”. Przy stawianiu hipotezy badawczej zabrakło mi doszczegółowienia co ma Autor na myśli pisząc o centralizacji oraz zmienności predykcji. Zaznaczyć należy, że w dalszej części pracy jest to wyjaśnione, niemniej jednak odniesienia do definicji kluczowych pojęć dla dysertacji powinny znaleźć się we wprowadzeniu.

W dysertacji zabrakło mi również dogłębnej dyskusji konkluzji wynikających z udowodnionej tezy, np.: czy wzrost zmienności predykcji jest pożądaną własnością przy wykorzystaniu systemów kolektywnych do prognozowania, czy można to mitygować, czy zależność ta pojawia się we wszystkich kolektywach, czy też zależna jest od typów powiązań i jest własnością strukturalną.

W tym miejscu nasuwa się również pytanie, czy prawdziwe będą wnioski z zanegowania przesłanek tezy: czy predykcja bardziej zdecentralizowanego kolektywu będzie mniej zmienna (bardziej stabilna)?

Teza dysertacji udowodniona została w eksperymentach wykonanych na autorskim modelu kolektywu, którego implementacja świadczy o dobrej znajomości technologii informacyjnych Doktoranta.

W pracy zawarty został szczegółowy przegląd modeli kolektywnych, oraz skupiono się na metodach ich opisu, w szczególności dogłębnie opisano miary służące do opisu centralizacji, gdzie Doktorant wymienia szereg metryk stosowanych w teorii grafów.

Wykonany przegląd istniejących podejść do pomiaru kolektywu wskazał ich braki, na podstawie czego zaproponowane zostały nowe funkcje centralizacji charakteryzujące się niską złożonością obliczeniową i spełnianiem zdefiniowanych wymagań dla tego typu miar.

Do wkładu pracy w dziedzinę kolektywnej inteligencji zaliczyć również należy zdefiniowanie nowej funkcji agregacji, nazwaną łączoną funkcją agregacji umożliwiającą wykorzystanie wybranych cech i ich wpływu na końcową prognozę. Wykorzystanie opracowanej funkcji centralizacji umożliwia określenie wpływu określonych charakterystyk członków na działanie kolektywu. Zaproponowana miara spełnia przedstawione w rozprawie wymagania dla miar centralizacji dla modeli kolektywu opartego na grafach a dodatkowo, jak zostało to dowiedzione, charakteryzuje się niższą niż istniejące miary złożonością obliczeniową, co należy uznać za jej zaletę.

W rozprawie zabrakło mi bardziej szczegółowej dyskusji ograniczeń zaproponowanej miary centralizacji dla różnych rozmiarów i różnych topologii kolektywów. Przykładowo, można by było wykonać analizę jak będzie się zachowywała zaproponowana miara dla dużych kolektywów zawierających różne gęstości połączeń wierzchołków. Przykład ten jest analogią do procesów grupowania (wspominanych w rozprawie w rozdziale 4.3.1) w których typowo wykorzystuje się jedynie miary oparte na odległościach, które nie radzą sobie z tego typu strukturami, a metryki oparte na miarach gęstościowych (np. DBSCAN) potrafią kompensować ten problem.

W pracy przedstawiono porównanie zaproponowanej miary centralizacji do dwóch istniejących technik wskazując jej przewagę wydajnościową i zbliżone wyniki jakościowe, jak również zachowanie wymagań stawianych miarom centralizacji. Dysertacja zawiera szczegółowe wyniki badań wykonane dla grafów posiadających 5 węzłów. Przedstawiono szczegółową analizę z szeregiem przykładów obrazujących działanie poszczególnych miar dla różnych grafów tej klasy. Z uzyskanych rezultatów eksperymentów wynika, że zaproponowana miara centralizacji daje zbliżone wyniki jakości do miar istniejących w literaturze jednak uzyskiwane są one w niższym czasie.

W rozprawie widzę brak analizy zaproponowanej miary dla grafów o wyższych liczbach wierzchołków. Oczywiście trudno by było taką analizę wykonać tak szczegółowo dla dużych struktur, jak przeprowadzone to zostało dla grafów 5 wierzchołkowych, jednakże ocena nawet losowej próbki większych grafów dałaby dokładniejszy pogląd jak miara centralizacji zachowuje się dla innych grafów i pozwoliłoby to wyciągnąć statystycznie wiarygodne, bardziej ogólne wnioski.

Ocena wpływu centralizacji na kolektywną prognozę zrealizowana została dwójako: teoretycznie i eksperymentalnie, co uważam za dobrą metodologię.

Teoretyczna analiza wskazała, że kolektywy z wysokim stopniem centralizacji są silnie zależne od ich członków, analogicznie pokazano że kolektywy o niskim poziomie centralizacji są słabo zależne od pojedynczych członków.

Przeprowadzone eksperymenty potwierdziły poprawność teoretycznych rozważań i są dowodem na poparcie tezy doktoratu. Pewien niedosyt budzi wykonanie eksperymentów na małe różnorodnych zbiorach danych, w szczególności rozmiary kolektywów poniżej 100 węzłów są istotnym zawężeniem wykonanego eksperymentu.

Przeprowadzone badania nie zostały również odniesione do innych modeli predykcyjnych – dla wykonanych eksperymentów nie mamy perspektywy na ile kolektywne decyzje dają lepsze rezultaty od znanych metod prognozowania np.: opartych na podejściu wykorzystującym uczenie maszynowe.

Zawarta w rozprawie dyskusja ograniczeń opracowanego modelu jest bardzo skromna i ogranicza się jedynie do dwóch aspektów, co sprawia wrażenie, że model jest bardzo ogólny, jednakże uważam, że jest szereg istotnych zagadnień związanych z kolektywnymi systemami, które tu powinny być przedyskutowane.

Przykładowo pominięte zostały rozważania odnośnie wzajemnego wpływu na siebie członków w czasie (model temporarny), czy zaproponowany model ma możliwość ujęcia zmieniających się cech członków kolektywu pod wpływem innych?

W cytowanym w pracy artykule Surowiecki wskazuje na założenia związane z inteligencją kolektywną, między innymi zakłada niezależność opinii jednostek – nie są one determinowane otaczającymi ją członkami. Uważam to za bardzo silne, redukujące założenie, gdyż w rzeczywistości członkowie kolektywów, w szczególności ludzie bardzo często podejmują decyzje na podstawie opinii innych. W związku z tym rodzą się pytania:

czy zaproponowany model umożliwiłby rozszerzenie o analizę wpływu otoczenia jednostek kolektywu na jakość końcowej predykcji.

czy w zaproponowanym modelu widziałby Doktorant możliwość wprowadzenia istotności węzłów kolektywu np.: typu *hub* i *authority*?

W dobie systemów rozproszonych, istotne jest również rozważenie aspektów związanych z bezpieczeństwem – niemyim założeniem opracowanego modelu jest, że podejmowana przez członków kolektywu decyzja nie jest fałszowana, co w rzeczywistych wypadkach nie koniecznie musi być prawdziwe: informacje o wartościach cech poszczególnych członków mogą być np. celowo zmieniane np.: przez sabotaż albo wynikać z wadliwego działania aparatury np.: niewłaściwych pomiarów wykonywanych przez stacje metrologiczne. W takim przypadku rodzi się pytanie do jakiego stopnia naruszenie integralności kolektywu pozwala zachować poprawność prognoz.

Praca zawiera przykład zastosowania opracowanego modelu do rozwiązania praktycznego problemu prognozowania. Jako że praca jest realizowana na uczelni technicznej proszę

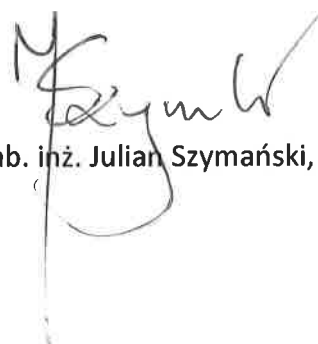
Doktoranta o szersze spojrzenie na praktyczne zastosowanie opracowanego modelu i wskazanie dziedzin, gdzie mogą znaleźć zastosowanie uzyskane w ramach pracy wyniki, w których wykorzystuje się kolektywną decyzję. Czy zaproponowany model nadawałby się np.: do zastosowań w metodach uczenia typu *ensemble learning*, czy mógłby on wnieść nowy wkład do istniejących metod uzyskiwania konsensusu.

4. Uwagi końcowe

Dorobek Doktoranta na podstawie baz Scopus, DBLP i Web of Science wskazuje na dwa artykuły w czasopismach z listy JCR i 11 publikacji konferencyjnych co uważam za dobre osiągnięcie na tym etapie rozwoju naukowego.

Pozytywnie oceniam zawartość merytoryczną rozprawy, dobre wprowadzenie w tematykę badawczą i sposób sformułowanych hipotez badawczych. Wykorzystana metodologia badawcza jest poprawna, a sposób przedstawienia wyników jest przejrzysty i popierający postawioną hipotezę badawczą.

Konkludując stwierdzam, że przedstawiona rozprawa doktorska mgr. inż. Rafała Palaka zatytułowana „The impact of centralisation on the accuracy of collective prediction” stanowi oryginalne rozwiązanie problemu naukowego, potwierdzając wiedzę oraz umiejętność samodzielnego prowadzenia badań naukowych przez Doktoranta i w mojej opinii spełnia wymagania stawiane stosownej ustawie. Na tej podstawie wnioskuję do Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Politechniki Wrocławskiej o dopuszczenie jej do dalszych etapów przewodu doktorskiego.



dr hab. inż. Julian Szymański, prof. PG