

Prof. dr hab. inż. Eulalia Szmidt
Instytut Badań Systemowych
Polskiej Akademii Nauk
ul. Newelska 6
01-447 Warszawa
E-mail: szmidt@ibspan.waw.pl

Warszawa, 17.06.2019r.

RAU	Biuro Dziekana	
	Wpłynęło dnia	24.06.2019
	Nr	638 / zał.

**Recenzja rozprawy doktorskiej mgr inż. Sebastiana Porębskiego
pt. „Wspomaganie diagnozy medycznej na podstawie
rozmytych reguł warunkowych”**

Niniejsza recenzja została przygotowana na prośbę Dziekana Wydziału Automatyki, Elektroniki i Informatyki Politechniki Śląskiej Pana prof. dr hab. inż. Adama Czornika wyrażoną w piśmie z dn. 10 kwietnia 2019r.

Recenzowana rozprawa doktorska dotyczy wydobywania wiedzy z danych medycznych, czyli ciągle aktualnego, ważnego i interesującego zagadnienia badawczego, zarówno w kontekście teoretycznym jak i aplikacyjnym. W niniejszej pracy chodzi o zaproponowanie i przetestowanie metod konstrukcji reguł decyzyjnych na podstawie medycznych baz danych, w których zwykle mamy do czynienia z niepewną i nieprecyzyjną wiedzą. Autor, na podstawie swoich wcześniejszych, opublikowanych badań, zastosował do reprezentowania niepewności i nieprecyzyjności danych metodę Dempstera-Shafera i teorię zbiorów rozmytych. Zaproponowane reguły decyzyjne mają nie tylko wspomagać podejmowanie diagnoz medycznych, ale co równie ważne, powinny być zrozumiałe i łatwe do zinterpretowania dla eksperta podejmującego decyzje. Oznacza to, że reguły nie powinny być skomplikowane i nie powinno być ich zbyt wiele. Powstaje więc niejako automatycznie konieczność szukania kompromisu między niezawodnością i złożonością reguł. Powyższy krótki opis pokazuje, że **wyбір tematu jest ambitny i stawia przed Doktorantem duże wyzwania**. Należy przy tym zauważyć, że chociaż w pracy rozpatrywane są dane medyczne, to rozwiązanie postawionego w rozprawie problemu może być przydatne przy rozwiązywaniu wielu

innych problemów opisanych za pomocą niepewnych i nieprecyzyjnych danych. Warto w tym miejscu podkreślić jeszcze jeden aspekt związany z rozprawą, a mianowicie, Autor powołuje się na swoje prace (3 artykuły z części A listy MNiSW w spisie literatury, np. z *Information Sciences*), które powstały na przestrzeni trzech ostatnich lat we współpracy z Promotorem. Pozwala to zauważyć, że proponowane rozwiązania zostały poddane weryfikacji międzynarodowej społeczności akademickiej co znacząco wzmacnia przedstawione w rozprawie idee i wyniki. Co więcej, prace te były cytowane 17 razy (Gogle Scholar), między innymi w tak prestiżowym czasopiśmie jak *IEEE Transactions on Fuzzy Systems*, 2018.

Praca liczy 155 stron, składa się 6 rozdziałów, dodatku, w którym opisane są testowane bazy danych oraz przykładowe zbiory reguł i funkcje przynależności, wykazu literatury obejmującego 138 pozycji (w tym kilku publikacji Doktoranta), oraz odrębnych spisów: rysunków, tabel i wykazu ważniejszych oznaczeń i skrótów.

W Rozdziale 1 dobrze i ciekawie przedstawiona jest motywacja, zakres i cel rozprawy. Po jasnym sformułowaniu tezy, Autor wymienia szczegółowe zadania, których wykonanie ma potwierdzić słuszność tezy. Rozdział kończy syntetyczne przedstawienie treści rozprawy. W związku z faktem, że w rozprawie szczególna uwaga poświęcona jest zarówno niepewności jaki i nieprecyzyjności towarzyszącym zazwyczaj dostępnej wiedzy przy podejmowaniu decyzji (w szczególności, określaniu diagnoz medycznych), **praca wpisuje się w nurt prac nowatorskich.**

W Rozdziale 2 pracy doktorskiej przedstawiono podstawowe pojęcia i metody związane z przetwarzaniem niepewności. Rozdział rozpoczyna przykład z zakresu medycyny ilustrujący potrzebę uwzględniania niepewności i nieprecyzyjności w zagadnieniach medycznych. Dalej Autor wprowadza czytelnika w podstawy teorii zbiorów rozmytych ze szczególnym uwzględnieniem rodzajów funkcji przynależności, następnie omawia teorię możliwości i teorię Dempstera-Shafera. Zagadnienia te są ściśle związane z badaniami prowadzonymi przez Autora rozprawy, są opisane w sposób przejrzysty, pozwalający stwierdzić, że Autor zna odpowiednią literaturę.

W Rozdziale 3 przedstawiono metodę konstruowania rozmytych reguł warunkowych reprezentujących wiedzę medyczną i model wspomaganie diagnozy bazujący na miarach przekonania wprowadzonych w teorii Dempstera-Shafera. Autor proponuje cztery funkcje przynależności, które mogą być wykorzystane w przesłankach reguł. Proponuje także nowy wskaźnik oceny reguł diagnostycznych. Rozdział kończy przykład ilustrujący omawiane podejścia. Rozdział napisany jest w sposób przejrzysty.

Rozdział 4 opisuje **trzy iteracyjne metody wydobywania reguł diagnostycznych** z danych medycznych. **Są to oryginalne propozycje Autora** (algorytm eliminacji reguł, algorytm selekcji reguł, algorytm warunkowej selekcji reguł). Algorytmy są przedstawione w sposób zrozumiały i uzupełnione o przejrzyste schematy blokowe. **Warto podkreślić, że proponowane podejście jest o wiele bardziej zrozumiałe dla użytkownika (lekarza)** niż metody wykorzystujące sieci neuronowe czy algorytmy genetyczne. Możliwość prześledzenia i uzasadnienia każdego kroku algorytmu może być czynnikiem decydującym o wykorzystaniu proponowanych metod przez lekarzy w placówkach medycznych. W dalszej części rozdziału przedstawione są znane z literatury metody badania efektywności reguł decyzyjnych oraz metody badania złożoności zbioru reguł. Działanie zaproponowanych trzech algorytmów zilustrowano i porównano na podstawie dobrze znanej bazy referencyjnej (*Iris flowers*). Oceniano liczbę reguł i błędów dla każdego algorytmu. Na zakończenie uzasadniono celowość wprowadzenia i **zaproponowano nowy wskaźnik oceny reguł**, którego zadaniem jest ocena efektywności i złożoności reguł.

Rozdział 5 zawiera eksperymenty numeryczne, pozwalające na ocenę skuteczności działania zaproponowanych wcześniej metod. Do testów wykorzystano 5 dobrze znanych medycznych baz referencyjnych oraz jedną **rzeczywistą bazę szpitalną**. Ocenie poddano zbiory wydobytych reguł. Zbadano wpływ kształtów funkcji przynależności na uzyskane wyniki, przetestowano warunek zatrzymania algorytmu eliminacji reguł i oceniono złożoność reguł diagnostycznych. Uzyskane wyniki pozwoliły na wskazanie różnic między zaproponowanymi metodami. Zbadano też różnice między metodami oceny rozmytych elementów ogniskowych. W drugiej części rozdziału zbadano jakość reguł diagnostycznych na zbiorach testowych, czyli praktyczną przydatność reguł, co pozwoliło wskazać najlepsze z proponowanych

w rozprawie rozwiązań. Ponadto, proponowane metody zostały porównane z kilkoma innymi, znanymi z literatury metodami. Warto tutaj podkreślić, że porównanie z innymi metodami nie ogranicza się do zacytowania wyników zaczerpniętych z artykułów, ale Autor przeprowadził własne obliczenia korzystając z pakietu KEEL, dbając o zapewnienie dokładnie tych samych warunków we wszystkich eksperymentach (do porównań wykorzystał 7 znanych metod). **Rozdział dowodzi dobrego teoretycznego warsztatu doktoranta, a badania i eksperymenty, stanowiące istotny element rozprawy, świadczą o Jego docieklivości i sumienności. Uzyskane wyniki potwierdzają skuteczność proponowanych autorskich rozwiązań.**

Rozdział 6 stanowi umiejętne podsumowanie treści rozprawy i uzyskanych w niej wyników. Nie ulega wątpliwości, że **teza rozprawy została udowodniona, a wszystkie zaplanowane zadania cząstkowe sformułowane na początku rozprawy, zostały zrealizowane.** Są to następujące autorskie rozwiązania:

- Opracowanie zmodyfikowanego algorytmu eliminacji reguł.
- Zaproponowanie metody wyznaczania kształtów funkcji przynależności dla elementów ogniskowych i zbadanie ich wpływu na efektywność zaproponowanego algorytmu eliminacji reguł.
- Zaproponowanie wskaźnika oceny reguły diagnostycznej, który w wyniku działania proponowanych algorytmów, pozwolił otrzymać mniejszą liczbę reguł z zachowaniem zbliżonej efektywności wyznaczania diagnozy.
- Opracowanie algorytmu selekcji reguł.
- Zmodyfikowanie algorytmu selekcji reguł i uzyskanie w efekcie algorytmu warunkowej selekcji reguł.
- Zaproponowanie wskaźnika oceny wyniku działania proponowanych algorytmów wydobywania reguł diagnostycznych, który uwzględnia kompromis pomiędzy efektywnością wyznaczania poprawnej diagnozy a złożonością zbioru reguł diagnostycznych.
- Wykorzystanie do wspomaganie diagnozy zmodyfikowanej miary przekonania.

Pracę doktorską mgr inż. Sebastiana Porębskiego czyta się dobrze. Autor przedstawia zagadnienia poruszane w pracy w sposób jasny, spójny i logiczny, posługując się w wybranych miejscach trafnie dobranymi przykładami.

Uwaga szczegółowa

Rys. 2.3. został zamieszczony przez pomyłkę zamiast innego rysunku (na pewno Rys. 2.3. nie odnosi się do Rys. 2.2. co było zamierzeniem Autora). Innych błędów nie udało się znaleźć, bo praca jest napisana bardzo starannie.

Chciałabym podkreślić, że **moja ogólna ocena pracy jest bardzo pozytywna**, a Doktorant wywiązał się bardzo dobrze z postawionych na początku zadań.

Podsumowując, uważam, że praca jest poprawna merytorycznie, zawiera oryginalny dorobek Doktoranta, zaproponowane nowe algorytmy zostały oprogramowane i przetestowane na znanych danych referencyjnych oraz rzeczywistych, uzyskując lepsze rezultaty niż wcześniej stosowane podejścia (szczególnie w zakresie liczby, złożoności i transparentności reguł decyzyjnych). **Rozprawa ma** zarówno ważne walory poznawcze jak również **potencjał użytkowy i wdrożeniowy**. Praca świadczy o dużej erudycji Autora i jego znajomości tematyki związanej z szeroko rozumianą dziedziną obliczeń miękkich, logiką rozmytą, niepewnością i nieprecyzyjnością danych, klasyfikacją, realnymi problemami programowania, a także o umiejętności jasnego i precyzyjnego przedstawiania wyników.

W związku z powyższym, moim zdaniem, praca spełnia wszelkie wymagania zawarte w obowiązujących przepisach dotyczących rozpraw doktorskich. Wnoszę zatem o jej przyjęcie i dopuszczenie pana mgr inż. Sebastiana Porębskiego do publicznej obrony.

Jednocześnie, biorąc pod uwagę nowatorstwo, zakres i rangę uzyskanych wyników, wnoszę o jej wyróżnienie.