Prof. dr hab. inż. Marian Gorzałczany
Wydział Elektrotechniki, Automatyki i Informatyki
Politechnika Świętokrzyska

Kielce, 30.05.2019r.

Recenzja rozprawy doktorskiej

Tytuł rozprawy:	Wspomaganie diagnozy medycznej na podstawie rozmytych reguł warunkowych
Autor rozprawy:	Mgr inż. Sebastian Porębski
Promotor rozprawy:	Dr hab. inż. Ewa Straszecka
Dziedzina:	Nauki techniczne
Dyscyplina:	Elektronika

RAU
Biuro Dziekana
Wpłynęło dnia 30.05.2019
Nr 583 / zał. .....

Niniejsza recenzja została przygotowana w odpowiedzi na pismo Dziekana Wydziału Automatyki, Elektroniki i Informatyki Politechniki Śląskiej, prof. dr hab. inż. Adama Czornika (pismo nr RAu/829/2018/2019 z dn. 10.04.2019r.) zawierające informację, że zostałem powołany na recenzenta w przewodzie doktorskim mgr. inż. Sebastiana Porębskiego.

Recenzowana rozprawa doktorska poświęcona jest budowie regułowych systemów wspomagania decyzji – bazujących na synergicznym połączeniu teorii Dempstera-Shafera, teorii zbiorów rozmytych i związanej z nią teorii możliwości – w problemach diagnozowania medycznego. Rozprawa składa się z sześciu rozdziałów oraz dodatku i spisu literatury (łącznie 155 stron druku). Pierwszy rozdział ma charakter wstępny i zawiera zarówno usystematyzowany przegląd zagadnień dotyczących wspomagania decyzji w diagnozowaniu medycznym jak również obszerny i szczegółowy przegląd narzędzi teoretycznych wykorzystywanych dotychczas w budowie tego typu systemów. Wspomaganie decyzji w diagnozowaniu medycznym charakteryzuje się m.in. koniecznością wykorzystywania i przetwarzania informacji zarówno o charakterze nieprecyzyjnym (np. przy opisie objawów) jak i niepewnym (np. podczas formułowania diagnozy). Doktorant dokonał przeglądu ponad 100 pozycji literatury światowej prezentującej różnorodne podejścia do wspomagania decyzji w medycynie, w tym szereg podejść wykorzystujących teorię zbiorów rozmytych, teorię Dempstera-Shafera wraz jej rozszerzeniem na zbiory rozmyte, drzewa decyzyjne, sztuczne sieci neuronowe i tzw. maszyny wektorów podpierających (ang. *support vector machines*). Podejścia te dotyczą zarówno reprezentowania informacji nieprecyzyjnej i niepewnej jak formułowania wiedzy diagnostycznej w formie rozmytych reguł warunkowych.

Dokonany przez Doktoranta bardzo obszerny i krytyczny przegląd literatury światowej dotyczącej jego obszaru badań zarówno świadczy o głębokiej wiedzy Doktoranta w reprezentowanej przez niego dyscyplinie naukowej jak i stanowi logiczny punkt wyjścia do

zaproponowanych przez niego rozwiązań będących odpowiedzią na określone mankamenty podejść dotychczas stosowanych. Doktorant stawiając sobie jako cel rozprawy opracowanie i przetestowanie metod syntezy zestawów reguł diagnostycznych na podstawie medycznych baz danych, z wykorzystaniem miar nieprecyzyjności i niepewności informacji, podjął się zadania trudnego, o dużym znaczeniu zarówno poznawczym jak i aplikacyjnym. Praktyczny wymiar tego zadania jest ważny również ze względu na znaczenie nowoczesnych systemów wspomagania diagnozowania medycznego w szeroko rozumianej ochronie zdrowia. Co więcej, w ostatnich latach intensywnie rozwija się rynek medycznych urządzeń elektronicznych służących do terapii lub monitorowania pacjenta w domu. Wyposażenie takich urządzeń w systemy wspomagające ocenę stanu pacjenta mogłyby poprawić rezultaty leczenia.

Rangę rozprawy podnosi oryginalna teza badawcza, którą można streścić następująco: synergiczne połączenie teorii Dempstera-Shafera i teorii zbiorów rozmytych w celu wykorzystania do budowy iteracyjnych algorytmów syntezy rozmytych reguł warunkowych podnosi jakość wspomagania diagnozowania medycznego rozumianą zarówno w sensie dokładności jak i przejrzystości otrzymanego zbioru reguł. Należy wyraźnie podkreślić, że powyższa teza rozprawy jest oryginalna i ważna zarówno z naukowego jak i praktycznego punktu widzenia. Przedstawienie metod i algorytmów oraz wyników praktycznych eksperymentów pozwalających na wykazanie jej prawdziwości wniesie wkład w rozwój zarówno teorii jak i praktyki budowy systemów wspomagania decyzji z wykorzystaniem informacji nieprecyzyjnej i niepewnej.

W następnym, drugim rozdziale rozprawy Doktorant przedstawił krótki przegląd podstawowych pojęć z zakresu teorii zbiorów rozmytych, związanej z nią teorii możliwości oraz teorii Dempstera-Shafera. Pojęcia te są wykorzystywane przez niego w kolejnych trzech rozdziałach, składających się na zasadniczą część rozprawy, w której przedstawia on proponowane przez siebie rozwiązania. Rozdz. 3 prezentuje, bazującą na teorii Dempstera-Shafera, metodę konstruowania reguł warunkowych oraz analizę sposobów ich oceny, w tym zaproponowany przez Doktoranta nowy wskaźnik oceny reguł. W rozdz. 4 Doktorant przedstawia trzy oryginalne metody i algorytmy syntezy reguł diagnostycznych z danych medycznych. Rozdz. 5 poświęcony jest wszechstronnej eksperymentalnej weryfikacji proponowanych rozwiązań teoretycznych z wykorzystaniem zarówno pięciu powszechnie dostępnych benchmarkowych medycznych baz danych jak i rzeczywistej, szpitalnej bazy danych dotyczącej diagnozy stopnia zwłóknienia wątroby pacjentów, u których stwierdzono wirusowe zapalenie wątroby typu C. Rozdz. 6, zawierający jasno sformułowane wnioski końcowe, stanowi podsumowanie pracy. W dodatku Doktorant zamieścił opis wykorzystywanych baz danych oraz przykładowe zbiory reguł wraz z postaciami funkcji

przynależności zbiorów rozmytych. Zarówno sekwencja rozdziałów rozprawy jest właściwa jak i prawidłowe są proporcje pomiędzy przeglądem faktów znanych z literatury (rozdziały 1 i 2), a prezentacją i weryfikacją metod zaproponowanych lub rozwiniętych przez Doktoranta (rozdziały 3, 4 i 5 oraz rozdział podsumowujący – 6).

Do szczególnie wartościowych i o oryginalnym charakterze elementów rozprawy należy zaliczyć:

1. Opracowanie zmodyfikowanego algorytmu eliminacji reguł diagnostycznych z początkowego, nadmiarowego zbioru tych reguł w drodze wielokrokowej procedury usuwania rozmytych elementów ogniskowych (definiujących przesłanki reguł) charakteryzujących się najniższą wartością bazowego prawdopodobieństwa.
2. Opracowanie algorytmu selekcji reguł diagnostycznych z początkowego, nadmiarowego zbioru tych reguł w drodze wielokrokowej procedury wyboru pojedynczych najlepszych reguł ze zbioru początkowego i przenoszenie ich do nowo tworzonego zbioru reguł.
3. Opracowanie modyfikacji algorytmu z punktu 2 w postaci tzw. algorytmu warunkowej selekcji reguł diagnostycznych poprzez dodanie etapu warunkowego usuwania reguł, które nie poprawiają dokładności diagnozowania.
4. Demonstracja praktycznego wykorzystania proponowanych algorytmów w realnych zagadnieniach diagnozowania medycznego wraz z wszechstronną eksperymentalną ich weryfikacją
5. Wprowadzenie nowego wskaźnika oceny jakości poszczególnych generowanych rozmytych reguł diagnostycznych. Wskaźnik ten uwzględnia dwa zasadnicze aspekty jakości reguły, a mianowicie jej dokładność oraz interpretowalność. Zastosowanie tego wskaźnika w ramach zaproponowanych i wymienionych wyżej algorytmów wydobywania reguł diagnostycznych z danych medycznych pozwoliło otrzymywać mniejszą liczbę reguł z zachowaniem zbliżonej dokładności systemu, w porównaniu z dotychczasowym kryterium oceny w postaci wartości bazowego prawdopodobieństwa.
6. Opracowanie metody wyznaczania kształtów funkcji przynależności zbiorów rozmytych reprezentujących rozmyte elementy ogniskowe oraz zbadanie ich wpływu na efektywność zaproponowanych algorytmów. Rozważono niemal wszystkie typowe klasy funkcji przynależności zbiorów rozmytych, a mianowicie funkcje trójkątne, trapezowe, uogólnione dzwonowe i Gaussowskie.
7. Wprowadzenie nowego wskaźnika oceny zestawów reguł diagnostycznych generowanych przez wymienione wyżej algorytmy. Wskaźnik ten uwzględnia kompromis pomiędzy dokładnością a złożonością otrzymanych zestawów reguł.
8. Wykorzystanie zmodyfikowanej miary przekonania i porównanie jej z tradycyjną miarą przekonania w generowaniu decyzji diagnostycznych przez wymienione wyżej algorytmy.

Obok wymienionych wyżej zalet rozprawy, jej lektura może nasuwać następujące uwagi ogólne:

1. Doktorant rozważa dwa aspekty dotyczące jakości generowanych przez proponowane algorytmy zestawów rozmytych reguł diagnostycznych, a mianowicie – z jednej strony – ich złożoność, czytelność i interpretowalność oraz – z drugiej strony i według terminologii Doktoranta – ich efektywność. Efektywność jest pojęciem bardzo pojemnym. Używanie tego pojęcia bez doprecyzowania jego znaczenia w rozważanym kontekście może prowadzić do nieporozumień odnośnie tego co faktycznie Doktorant ma na myśli. Dopiero na str. 66 rozprawy (4 i 5 wiersz od dołu) Doktorant jasno stwierdza (cytat): "...Efektywność jako stosunek liczby poprawnie zdiagnozowanych pacjentów do ich całkowitej liczby w zbiorze ...". A zatem, efektywność w ujęciu Doktoranta to procent poprawnych decyzji lub też, krótko, dokładność systemu. W literaturze dotyczącej systemów regułowych, w tym w renomowanych czasopismach anglojęzycznych, dokładność (ang. *accuracy*) i interpretowalność (ang. *interpretability*) są standardowymi miarami jakości tego typu systemów. Stąd też podejście Doktoranta jest całkowicie poprawne i łatwo można było się domyślić co kryje się pod wspomnianym pojęciem. Natomiast wydaje się, że wskazane byłoby albo używanie pojęć precyzyjnych albo doprecyzowanie pojęć bardziej ogólnych.
2. Wymienione wyżej miary i kryteria oceny systemów regułowych mają wzajemnie sprzeczny charakter, co Doktorant wielokrotnie stwierdza. W związku z tym, rozsądnym wydaje się pomysł wykorzystania metod wielokryterialnej optymalizacji w problemie syntezy zbioru reguł z danych. W ostatnich latach, przykładowo, w przypadku syntezy reguł rozmytych z danych, często wykorzystywane są metody wielokryterialnej optymalizacji ewolucyjnej (ang. *multi-objective evolutionary optimization algorithms*). Metody te generują zbiór rozwiązań charakteryzujących się różnym poziomem zoptymalizowanego kompromisu pomiędzy dokładnością a interpretowalnością systemu. Dla każdego z rozwiązań można następnie obliczyć jego dokładność dla danych testowych, a następnie – spośród rozwiązań o najwyższej dokładności – wybrać rozwiązanie o najmniejszej złożoności mając, przynajmniej po części, gwarancję, że jest to najlepsze możliwe rozwiązanie.
3. Interpretowalność rozmytych systemów regułowych powinna być rozpatrywana zarówno w kontekście złożoności systemu (liczba i złożoność reguł) jak i w jego wymiarze semantycznym. O ile Doktorant wielokrotnie odnosi się do pierwszego aspektu interpretowalności systemu, o tyle nie ma bezpośrednich odniesień w rozprawie do drugiego aspektu. Tymczasem, zaproponowany przez Doktoranta algorytm generowania funkcji przynależności zbiorów rozmytych (patrz punkt 6 przedstawionej wcześniej listy

szczególnie wartościowych elementów rozprawy) generuje zestawy zbiorów rozmytych (zamieszczone w dodatku rozprawy) bardzo dobrze spełniające sformułowane w literaturze wymogi w tym zakresie, tzn. – w terminologii anglojęzycznej – *coverage, normalization, distinguishability, complementarity.* [1]

4. Złożoność systemu regułowego zależy nie tylko od liczby reguł ale również od złożoności poszczególnych reguł związanej z liczbą przesłanek, liczbą użytych określeń słownych, itp. W związku z tym, warto rozważyć rozszerzenie wskaźnika *QI* (4.16) (str. 77 rozprawy) tak aby uwzględniał również złożoność reguł, a nie tylko ich liczbę.

Nieznaczne zastrzeżenia można mieć do edycyjnego przygotowania rozprawy, np.:

a) na str. IV (spis treści; 16-ty wiersz od dołu) – powinno być "... zwłóknienia ...",
b) na str. 23 (14,15-ty wiersz od dołu) – "...można stwierdzić, że stwierdzenia..." – niezręczne sformułowanie,
c) na str. 37 (21-szy wiersz od dołu) – zamiast (3.21) powinno być (3.15),
d) na str. 41 (19-ty wiersz od dołu) – powinno być "strojenie",
e) na str. 43, rys. 3.4 i na str. 47, rys. 3.7ab – brak opisu osi poziomej,
f) na str. 76, rys. 4.9 – w opisie osi pionowej zamiast "suma liczby reguł" prawdopodobnie powinno być "sumaryczna liczba reguł",
g) na str. 86 (14-ty wiersz od dołu) – powinno być "... uzyskania ostatecznego zbioru reguł..."
h) na str. 93-95, rysunki 5.9-5.13 – symbole $c^{(j)}$ na wykresach słupkowych w lewej części rysunków prawdopodobnie powinny zostać zastąpione symbolami $\theta^{(j)}$,
i) na str. 97 (13-ty wiersz od góry) – powinno być "... Uzyskane dotychczas wyniki pokazują...",
j) na str. 112 (10-ty wiersz od góry) – powinno być "... zwłóknienia ...".

Powyższe uwagi, poza redakcyjnymi, mają charakter dyskusyjny (uwaga nr 2 sugeruje ponadto kierunek ewentualnych dalszych badań). Uwagi te nie kwestionują wyników zaprezentowanych w rozprawie. Należy podkreślić, że wszystkie określone w rozdz. 1 rozprawy cele zostały w pełni osiągnięte, a otrzymane i zaprezentowane w rozprawie wyniki stanowią potwierdzenie słuszności sformułowanej przez Doktoranta tezy badawczej.

Rozprawa napisana jest w sposób przejrzysty i stanowi logiczną całość. Zawiera przy tym wszystkie wymagane elementy rozprawy naukowej, tzn. obszerny i krytyczny przegląd literatury, wskazanie – na tym tle – zagadnień wymagających rozwiązania, sformułowanie celu i tezy rozprawy, propozycję nowych rozwiązań oraz ich eksperymentalną weryfikację. Rozprawa stanowi oryginalne rozwiązanie problemu naukowego potwierdzające zarówno ogólną wiedzę teoretyczną Doktoranta, jak i umiejętność samodzielnego prowadzenia pracy naukowej.

---

[1] M.J. Gacto, R. Alcala, F. Herrera, Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. Information Sciences 181 (2011), pp. 4340-4360.

Podsumowując przedstawioną wyżej ocenę rozprawy doktorskiej mgr. inż. Sebastiana Porębskiego, stwierdzam, że spełnia ona z nadmiarem warunki stawiane kandydatom do stopnia doktora przez obowiązującą ustawę, wobec czego wnioskuję o dopuszczenie jej do publicznej obrony.

Biorąc dodatkowo pod uwagę fakt, że mgr inż. Sebastian Porębski jest współautorem 6 publikacji związanych z tematyką rozprawy, w tym 3 w czasopismach znajdujących się w bazie JCR o sumarycznej wartości wskaźnika *impact factor* równej 8.431, co w przypadku osoby ubiegającej się o stopień doktora jest wynikiem imponującym, wnioskuję również o wyróżnienie tej rozprawy.