Wrocław, dnia 19 maja 2017 r.

prof. dr hab. inż. Jerzy Świątek
Katedra Informatyki
Wydział Informatyki i Zarządzania
Politechnika Wrocławska

RECENZJA ROZPRAWY DOKTORSKIEJ
DLA RADY WYDZIAŁU CYBERNETYKI WOJSKOWEJ AKADEMII TECHNICZNEJ
im. Jarosława Dąbrowskiego w Warszawie

(Recenzja opracowana na podstawie zlecenia Przewodniczącego Rady Wydziału Cybernetyki WAT prof. dr hab. inż. Kazimierza Worwy pismem z dnia 12 kwietnia 2017 r., zgodnie z uchwałą Rady Wydziału Cybernetyki WAT z dnia 11 kwietnia 2017 r.)

**Temat**: „Models and Methods of Conflict Resolution in Data Fusion Process"

**Autor**: mgr inż. Walid Cherifi

**Promotor:** dr hab. inż. Bolesław Szafrański, prof. WAT

## 1. Cel i zakres pracy.

W ostatnich latach w literaturze światowej obserwujemy znaczący rozwój problemów dotyczących przetwarzania dużych woluminów danych w różnych dziedzinach. Jest to spowodowane z jednej strony rozwojem techniki pomiarowej, która udostępnia tanie i dokładne czujniki oraz urządzenia pomiarowe, które dostarczają precyzyjnych danych dotyczących rzeczywistych obiektów i procesów. Z drugiej strony współczesne platformy sprzętowe i programowe pozwalają na projektowanie złożonych systemów monitorowania i przetwarzania danych. W wyniku otrzymujemy ogromny przyrost danych dotyczących monitorowanego procesu, obiektu czy zjawiska w postaci strumieni danych. Problem ten można odnotować w różnych obszarach zastosowań, począwszy od monitorowania procesów technologicznych, chemicznych poprzez procesy ekonomiczne, społeczne, biomedyczne po monitorowanie obserwacji astronomicznych. Pojawia się problem wykorzystania zebranych danych w zadaniach wspomagania decyzji, projektowania, sterowania, zarządzania, w procesach diagnostycznych itp. W aktualnej, światowej literaturze ukształtował się obszar prac

naukowych nazwany: „fuzja danych (data fusion)". W tym obszarze można wyróżnić następujące problemy:

- Niskopoziomowa fuzja danych – dotyczy problemów związanych z układem pomiarowym, sygnałem, zakłóceniami pomiarowymi, danymi niekompletnymi itp.
- Średnio-poziomowa fuzja danych – dotyczy analizy cech sygnałów w określonym kontekście (redukcja cech, transformacja pomiarów),
- Wysokopoziomowa fuzja danych – dotyczy przetwarzania danych pochodzących z monitoringu w zadaniach wspomagania decyzji projektowych, sterowania, zarządzania itp.

Główny problem podjęty w recenzowanym opracowaniu leży w obszarze wysokopoziomowej fuzji danych dotyczącej analizy danych dostarczanych przez źródło celem wspomagania decyzji. Bardzo często dane dotyczące tego samego obiektu pochodzące z różnorodnych źródeł mogą być nieprecyzyjne, a w szczególności sprzeczne. Wykorzystanie takich danych może prowadzić do błędnych decyzji. Agregacja tego typu danych celem wypracowania decyzji może prowadzić do konfliktów. Powstaje ważny i aktualny problem opracowania algorytmów wspomagania decyzji w oparciu o różnorodne strumienie danych. Dodatkowo należy zwrócić uwagę na szybki napływ danych i konieczność ich przetworzenia w krótkim czasie przy ograniczonej pamięci. Nakłada to mocne wymagania na algorytmy przetwarzania danych strumieniowych. W tym kontekście zasadniczego znaczenia nabiera konieczność opracowania skutecznych metod rozwiązywania konfliktów w danych pochodzących z różnych źródeł podczas ich bieżącego napływu. W tym obszarze należy ulokować główny problem podjęty w recenzowanej pracy jasno i precyzyjnie określony przez cel pracy, który polega na opracowaniu metod i algorytmów, które pozwalają na unikanie konfliktów podczas przetwarzania danych strumieniowych pochodzących z różnych źródeł dostarczających sprzecznych informacji. Istotnym elementem tych algorytmów jest oszacowanie wartości miary opisującej jakość źródła danych. Na tym tle określono szczegółowy cel pracy, który polega na opracowaniu metod modelowania wiarygodności źródła danych z wykorzystaniem teorii funkcji przekonania (teorii Dempstera-Shafera).

Powyżej określony cel przekłada się na następujące zadania badawcze i projektowe:

1. Opracowanie modelu wiarygodności dowodowej informacji pochodzących z różnych źródeł z uwzględnieniem ich jakości (rzetelności) przy wykorzystaniu teorii funkcji przekonania.
2. Opracowanie nadzorowanych algorytmów estymacji wartości miary jakości dowodowej źródeł składowych z wykorzystaniem ciągu uczącego (na podstawie danych historycznych).

3. Opracowanie nienadzorowanych algorytmów estymacji wartości miary jakości dowodowej źródeł składowych.
4. Opracowanie przyrostowej procedury estymacji wartości miary jakości dowodowej źródeł składowych dla bieżąco napływających informacji.

W świetle wcześniej przedstawionych uwag mogę stwierdzić, że podjęty temat recenzowanej rozprawy jest ważny i aktualny, a opracowane zadania mają ważny aspekt praktyczny i wnoszą istotny wkład w rozwój informatyki, a w szczególności obszarze projektowania systemów wspomagania decyzji z wykorzystaniem metod fuzji danych tj.: analizy i przetwarzania strumieni danych

## 2. Przegląd treści pracy

Wyniki pracy Autor przedstawił w rozprawie doktorskiej napisanej w języku angielskim, która składa się z 8. rozdziałów oraz wykazu literatury, a treść pracy poprzedzona jest wykazem rysunków, tabel, algorytmów oraz skrótów. Całość liczy 172. strony.

Rozdział pierwszy stanowi wstęp, w którym Autor wprowadza czytelnika w zakres rozprawy, lokuje jej treść na tle aktualnie stosowanych metod opisu nieprecyzyjnych danych. Następnie omawia cel, najważniejsze wyniki pracy oraz przedstawia przegląd jej zawartości. W rozdziale drugim przedstawiono na tle literatury podstawowe pojęcia i przybliżenie aktualnych problemów zadania fuzji danych, zwracając uwagę na dane nieprecyzyjne prowadzące do decyzji konfliktowych. Kolejno rozdział trzeci zawiera wprowadzenie do teorii funkcji przekonania (teorii Dempstera-Shafera), w tym podstawowe pojęcia teorii, zasady łączenia źródeł dowodów, metody podejmowania decyzji w oparciu o przyjętą teorię, a przede wszystkim pojęcia istotne dla treści rozprawy. Zwrócono również uwagę na zalety i wady prezentowanej teorii. Istotne wyniki pracy zaczynają się w rozdziale czwartym. Przedstawiono tutaj oryginalny model wiarygodności dowodowej źródła z wyraźnym wskazaniem na jego odrębność na tle istniejących modeli. Istotnym elementem opisu jest uwzględnienie miary jakości (poprawności, rzetelności) źródła informacji. W rozdziale tym zaprezentowano również algorytm estymacji wartości miary jakości źródła na podstawie danych historycznych. Wskazano na wykorzystanie informacji o jakości źródła do podejmowania decyzji. Proponowany mechanizm jest podstawą do rozwiązania konfliktów, które mogą wystąpić przy nieprecyzyjnych danych generowanych przez źródła. Przedstawione problemy są ilustrowane przykładami. Kolejny rozdział piąty zawiera opis nienadzorowanego algorytmu estymacji wartości miary jakości źródeł, którego istota polega na iteracyjnym modyfikowaniu wartości

miary jakości źródła z uwzględnieniem kolejnych wartości wiarygodności dowodowej źródeł. Działanie prezentowanego algorytmu zilustrowano na przykładzie praktycznym. Zaprezentowano również wyniki eksperymentalnych badań, na rzeczywistych bazach danych, które miały na celu porównanie proponowanej metody z innymi podejściami znanymi z literatury. Następnie w rozdziale szóstym, Autor proponuje przyrostowy algorytm wyznaczania wartości miary jakości źródła dedykowany do analizy napływających na bieżąco danych. Rozdział siódmy przedstawia przykładowe możliwości zastosowań opracowanych metod w różnych obszarach. W zakończeniu - rozdział siódmy zawiera podsumowanie wyników pracy oraz propozycję kierunków badań nad wykorzystaniem przedstawionej teorii do przetwarzania i analizy nieprecyzyjnych informacji.

Bibliografia zawiera wykaz 112 trafnie dobranych pozycji literaturowych. Cytowane są głównie artykuły w czasopismach oraz opracowania monograficzne. Większość pozycji literaturowych stanowią pozycje aktualne z ostatnich 10. lat. Informacje literaturowe ściśle dotyczą tematyki rozprawy, a proponowany zbiór jest ściśle ukierunkowany na wyznaczony cel pracy.

**3. Uwagi redakcyjne**

Na podstawie lektury pracy mogę stwierdzić, że podział treści jest logiczny i uporządkowany, styl oraz poziom językowy jest wysoki, a szata graficzna jest staranna i dopracowana. Sposób przekazywania treści jest bardzo dobry. Treść pracy jest wzbogacona dobrze dobranymi rysunkami, wykresami oaz tabelami zawierającymi wyniki badań eksperymentalnych, ocen i porównań. Pracę czyta się z przyjemnością. Informacja jest dobrze wyważona. Czytając lekturę pracy odnosi się wrażenie o wysokiej kompetencji merytorycznej Autora, który potrafi przedstawić swoje wyniki w sposób przyjazny dla czytelnika.

**Uwagi redakcyjne krytyczne:**

W pracy dostrzegłem kilka szczegółowych uwag redakcyjnych:

1. W definicja funkcji $m_{ij}^{\Psi}$ pojawia się nieuzasadniona funkcja $m_{ij}^{\Omega}$ (str. 67.)
2. W opisie przykładu (4.4) jest mowa o dziewięciu obiektach, a wymienia się ich dziesięć tj.: $o_1, o_2, \cdots o_{10}$ (str. 88.).
3. Wielkości $m_{ij}^{\Theta,D}(R)$ i $m_{ij}^{\Theta,D}(\bar{R})$ nie są określone.
4. W rozdziale siódmym Autor ogranicza się do przytoczenia możliwości zastosowania prezentowanego podejścia. W tym miejscu czytelnik spodziewa się bardziej

szczegółowego Autorskiego opracowania. Autor zapowiada takie opracowania w dalszej perspektywie.

Przedstawione powyższe uwagi redakcyjne nie umniejszają merytorycznej wartości pracy.

**4. Oryginalne wyniki**

Wykonanie założonego celu wiąże się z wykonaniem następujących zadań szczegółowych, które uważam za oryginalne wyniki pracy:

- Opracowanie oryginalnego modelu matematycznego wiarygodności źródeł z wykorzystaniem teorii przekonania, a w szczególności uwzględnienia w opracowanym modelu miary rzetelności źródła danych. Proponowany model pozwala na opracowanie oryginalnego podejścia do rozwiązywania problemu konfliktu danych.
- Opracowanie algorytmów estymacji wartości miary jakości źródła danych w tym:
    - algorytmu estymacji wartości miary jakości źródła danych w oparciu o ciąg uczący,
    - nienadzorowanego, rekurencyjnego algorytmu estymacji wartości miary jakości źródła danych,
    - adaptacyjnej procedury estymacji, która pozwala na analizę i rozstrzyganie konfliktów w bieżąco napływających strumieniach danych.
- Wskazanie na możliwości zastosowania przedstawionego modelu i opracowanych algorytmów

**5. Uwagi merytoryczne – pytania dyskusyjne**

W trakcie analizy treści pracy nasuwają się pytania (uwagi), o których wyjaśnienie poproszę na publicznej obronie pracy:

1. Proponowany nienadzorowany algorytm oszacowania wartości miary jakości źródła jest inicjowany z zadanego punktu początkowego. Autor podaje różne możliwości wyboru punktu początkowego. Jaki ma wpływ dobór punktu początkowego na zbieżność algorytmu?
2. Przyrostowy algorytm dokonuje analizy napływających bieżących danych. Czy można określić zależność pomiędzy szybkością zmienności danych, a taktem algorytmu. Jakie warunki musi spełniać strumień danych aby analiza mogła być prowadzona w czasie rzeczywistym?

**6. Ocena wiedzy doktoranta w zakresie dyscyplin naukowych związanych z tematyką rozprawy**

1. Problematyka rozprawy obejmuje wiedzę z dziedziny nauk technicznych w zakresie dyscypliny informatyka, a w szczególności fuzji danych tj.: przetwarzania strumieni danych z zastosowaniem teorii Dempstera-Shafera do analizy danych nieprecyzyjnych i wykrywania konfliktów.
2. Doktorant swobodnie posługuje się aparatem matematycznym, właściwie formułuje hipotezy badawcze i wykazuje umiejętność dokonywania trafnych wyborów i właściwego wnioskowania. Formułowane przez doktoranta wnioski są przemyślane, logiczne i spójne.
3. W stopniu biegłym opanował tematykę rozprawy w warstwie nie tylko teoretycznej, ale także praktycznej, w oparciu o dobre rozeznanie problemów związanych z przetwarzaniem strumieni danych.
4. Lektura rozprawy upoważnia mnie do stwierdzenia, że doktorant swobodnie porusza się po literaturze przedmiotu, dysponuje wymaganym do prowadzenia badań naukowych zasobem wiedzy z zakresu informatyki, czyli tego obszaru nauki i techniki, którego dotyczy temat rozprawy.

**7. Podsumowanie.**

Na podstawie lektury recenzowanej rozprawy mogę stwierdzić, że doktorant:

- podjął ważny i aktualny problem w obszarze informatyki,
- uzyskał oryginalne wyniki w zakresie zastosowania metod wysokopoziomowej fuzji danych do projektowania systemów wspomagania decyzji,
- przedstawił interesujące możliwości zastosowań prezentowanych wyników,

a tym samym wniósł istotny wkład w rozwój dyscypliny informatyka.

W podsumowaniu mojej oceny stwierdzam, że opiniowana praca mgra inż. Walida Cherifi pt.: „Models and Methods of Conflict Resolution in Data Fusion Process " jest kompletna i nie wymaga żadnych zmian ani uzupełnień. Spełnia ona zwyczajowe i ustawowe wymagania stawiane rozprawom doktorskim określone w „Ustawie o stopniach naukowych i tytule naukowym oraz stopniach i tytule w zakresie sztuki" z dnia 14 marca 2003 roku.

Wnioskuję o przyjęcie niniejszej pracy, jako rozprawy doktorskiej i wnoszę o dopuszczenie mgra inż. Walida Cherifi do publicznej dyskusji nad przedłożoną pracą.

Biorąc pod uwagę bardzo aktualną problematykę i praktyczną wagę uzyskanych wyników oraz fakt, że przedstawione wyniki były publikowane w materiałach renomowanych konferencji o zasięgu międzynarodowym wnioskuję o wyróżnienie pracy.

prof. dr hab. inż. Jerzy Świątek