

Lublin, 5.08.2015 r.

Prof. dr hab. inż. Andrzej Kusz
Katedra Podstaw Techniki
Uniwersytet Przyrodniczy w Lublinie

**Ocena rozprawy doktorskiej mgr. inż. Marcina Rymarza
z tytułem
„Modelowanie jakości środowiska w języku sieci bayesowskich”**

1. Informacja o rozprawie

Rozprawa doktorska mgr. inż. Marcina Rymarza została wykonana na Wydziale Inżynierii Produkcji Uniwersytetu Przyrodniczego w Lublinie pod kierunkiem dr hab. Agnieszki Sujak. Rozprawa zawiera 99 stron formatu A4, łącznie ze spisem treści, wykazem cytowanej literatury (117 pozycji) oraz aneksem (str. 81-99). W pracy znajduje się 20 rysunków oraz 16 tabel, ilustrujących metodykę oraz wyniki przeprowadzonych badań.

2. Merytoryczna ocena rozprawy

Oceniana rozprawa dotyczy modelowania jakości środowiska przyrodniczego w języku sieci bayesowskich. Podjęty problem jest istotny i nawiązuje do aktualnych trendów związanych z konstruowaniem formalnych systemów reprezentacji wiedzy dla potrzeb zarządzania środowiskiem.

Podstawowy cel pracy, którym było „opracowanie funkcjonalnego modelu jakości środowiska ...” określa problem naukowy rozprawy a sformułowane hipotezy są jego uściśleniem. Zakres pracy obejmujący:

- opracowanie koncepcji modelu umożliwiającego ilościową ocenę stanu środowiska,
- wybór reprezentatywnego biomarkera,
- zaplanowanie i przeprowadzenie badań eksperymentalnych,
- analiza wyników i wybór zmiennych (elementów) modelu,
- implementacja modelu w środowisku BayesiaLab,
- walidacja modelu,

uwzględnia wszystkie etapy rozwiązywania problemów badawczych zgodnie z metodologią nauk empirycznych.

Na podkreślenie zasługuje oryginalność podejścia na etapie identyfikacji biomarkera oraz szeroki zakres badań eksperymentalnych. Wyniki tych badań oraz przeprowadzone analizy umożliwiły wyodrębnienie reprezentatywnego zbioru zmiennych modelu, wyznaczenie ich przedziałów

zmienności oraz określenie apriorycznych rozkładów prawdopodobieństwa. Od strony merytorycznej rozprawa, będąca przedmiotem recenzji, nie budzi zastrzeżeń.

Przegląd literatury, bardzo szeroki jest adekwatny do zakresu pracy i w sposób kompleksowy nawiązuje do zagadnień istotnie związanych z tematyką rozprawy doktorskiej. Cytowane pozycje odnoszą się zarówno do aspektów teoretycznych modelowania probabilistycznego oraz pokazują szerokie możliwości praktycznego wykorzystania tej technologii w różnych dziedzinach nauki ze szczególnym uwzględnieniem zastosowań w obszarze szeroko rozumianej problematyki przyrodniczej.

Końcowy aspekt przeglądu literatury dotyczy zagadnień związanych z degradacją środowiska naturalnego ze szczególnym uwzględnieniem źródeł emisji produktów szkodliwych dla środowiska a zwłaszcza metali ciężkich. Jak wynika z cytowanej literatury, poszukiwanie dla potrzeb monitoringu stanu środowiska naturalnego, charakterystycznego dla danego obszaru biomarkera środowiskowego coraz częściej jest ukierunkowane na dziko żyjące ptaki, które gniazdują kolonialnie, charakteryzują się długowiecznością i mobilnością. Naturalną konsekwencją tych rozważań jest wskazanie skorup jaj czapli siwej jako reprezentatywnego bioindykatora środowiska dla terenu województwa lubelskiego. Za tym wyborem przemawia fakt, że gatunek ten jest w stosunkowo niewielkim stopniu zagrożony wyginięciem, a ponadto objęty jest ochroną gatunkową czasową.

Przedstawiona w następnym rozdziale koncepcja modelu oceny środowiska sprowadza się do implementacji procedury obliczeniowej umożliwiającej wyznaczenie Globalnego Wskaźnika Jakości Środowiska (GWJŚP) jako sumy ważonej wskaźników lokalnych. Znormalizowane wagi wskaźników lokalnych wyznaczono uwzględniając obszar żerowania poszczególnych kolonii i ich liczebność. Wskaźniki lokalne reprezentują wyróżnione lokalizacje kolonii czapli siwej, w których pobierano próbki biomarkera (skorupki jaj) w celu wyznaczenia stężeń pierwiastków toksycznych. Na podstawie znajomości wartości tych stężeń wyznaczono wskaźniki lokalne jako sumę ważoną tych stężeń.

Realizacji tej koncepcji wymagała wyznaczenia miejsc występowania kolonii czapli siwej na terenie województwa lubelskiego oraz zaplanowania i przeprowadzenia badań laboratoryjnych zawartości pierwiastków, w tym metali ciężkich i pierwiastków śladowych w skorupkach jaj w poszczególnych lokalizacjach. Pierwotny obszar badań obejmujący pięć stanowisk kolonii czapli siwej, zredukowano do trzech lokalizacji uwzględniając podobieństwo ze względu na rodzaj żerowania i dostępność pokarmu. Wybrane lokalizacje kolonii czapli siwej stanowią przekrój ekologicznego oddziaływania środowiska od obszarów nieskażonych działalnością przemysłu (miejsco-

wości Kosmów i Stulno), poprzez tereny na obszarze zlokalizowanym wzdłuż Wisły (Wólka Michowska i Chodlik) a kończąc na terenach typowo rolniczych (Lipisko).

Skorupki jaj czapli siwej pobierano z miejsca występowania kolonii. W ramach badań laboratoryjnych określono stężenia 16 pierwiastków, przeprowadzając trzydzieści niezależnych pomiarów każdego z nich z każdej z wydzielonych lokalizacji. Pomiary zawartości metali ciężkich przeprowadzono zgodnie z obowiązującymi standardami. Na podstawie analizy korelacji wytypowano 10 pierwiastków (As, Ba, Cd, Cr, Hg, Ni, Pb, Sc, Sr i V), których stężenia są reprezentowane jako wartości zmiennych w modelu bayesowskim. Wartości ciągłych zmiennych losowych poddano normalizacji i dyskretyzacji na pięć jednakowych wartości przedziałowych (0-0,2, ..., 0,8-1). Każdej znormalizowanej wartości przedziałowej odpowiadają przedziały wartości stężeń rzeczywistych specyficzne dla każdej badanej lokalizacji kolonii czapli siwej.

Na podstawie badań ankietowych, określono znormalizowane wagi przypisane poszczególnym zmiennych (pierwiastkom). Zakres przeprowadzonych badań laboratoryjnych, oryginalność podejścia polegająca na wyborze skorupki jaj czapli siwej jako lokalnego bioindykatora degradacji środowiska uzasadniają pozytywną ocenę tej części pracy.

Dane pozyskane w wyniku badań laboratoryjnych zostały wykorzystane do edycji zmiennych reprezentujących stężenia poszczególnych metali oraz do wyznaczenia apriorycznych rozkładów prawdopodobieństwa. W przypadku zmiennej reprezentującej stężenie ołowiu Hg założono, że ma ona rozkład normalny o bardzo małym odchyleniu standardowym (0,001) i wartości oczekiwanej wynoszącej odpowiednio: 0,0; 0,5; 0,7; 0,9; 0,9. Wartość średnia tej zmiennej jest przypisywana jako wynik działania formuły logicznej w zależności od zaobserwowanej liczby wystąpień tego pierwiastka reprezentowanej w modelu przez dyskretną zmienną losową Liczba_wystąpień_Hg. Zmienna ta może przyjmować pięć wartości (0,1,2,3,4) odpowiadających zaobserwowanej liczbie wystąpień w konkretnej lokalizacji.

Topologia modelu, w warstwie zawierającej zmienne reprezentujące surowe dane, uwzględnia dwa węzły dodatkowe (Sub_1 i Sub_2), które wprowadzono w celu redukcji złożoności obliczeniowej. Stosując zasadę łączności i przemienności dodawania, założono istnienie trzech grup pierwiastków toksycznych. Grupa pierwsza I (Ba, Cd, Sc, V) i druga II (As, Cr, Sr, Ni) zawierają po cztery pierwiastki dobrane w taki sposób, aby suma wag tych pierwiastków była podobna i sumarycznie nie większa niż 0.3. Węzły Sub_1 i Sub_2, zbierają dane z grupy I i II. W grupie III znajdują się pierwiastki najbardziej toksyczne czyli Pb i Hg.

Przedstawiony model w postaci sieci bayesowskiej umożliwiający wyznaczanie GWJŚ na podstawie wskaźników lokalnych (Bioindykatorów) jest poprawny pod względem logicznym a

sposób implementacji procedury obliczeniowej nie budzi zastrzeżeń natury formalnej i merytorycznej. Jego istotną zaletą jest fakt, że wszystkie zmienne oraz rozkłady prawdopodobieństwa nad zbiorem wartości tych zmiennych wyznaczono na podstawie wyników badań laboratoryjnych. Analiza rozkładów wartości prawdopodobieństwa odpowiadającym różnym stężeniom pierwiastków wskazuje na zasadność wydzielenia trzech specyficznych dla danego terenu obszarów lokalizacji miejsc zbierania danych.

Kolejny rozdział pracy to walidacja modelu. Wykorzystując mechanizmy wnioskowania typowe dla sieci bayesowskich przeprowadzono analizę wrażliwości badając odpowiedzi modelu na zadane dane wejściowe, a w szczególności badając czułość modelu na skrajne zmiany rozkładu wartości zmiennych reprezentujących stężenia metali o najwyższych i najniższych wagach. Na tym etapie stosowano zarówno wnioskowanie prognostyczne oraz wnioskowanie wsteczne. We wszystkich analizowanych przypadkach (analizowano 11 różnych wariantów) odpowiedzi modelu na zmianę wartości poszczególnych zmiennych należy uznać za prawidłowe i zgodne z logiką a ilościowa ocena tych zmian pozwala pozytywnie ocenić czułość modelu. Należy zauważyć, że przedstawione w pracy scenariusze działania nie wyczerpują wszystkich możliwości obliczeniowych modelu.

W kolejnym punkcie (IV) przytoczono szczegółowe informacje o przeprowadzonych badaniach eksperymentalnych zawartości metali ciężkich w skorupkach jaj czapli siwej. Przedstawiona dyskusja wskazuje na zasadność wydzielenia trzech różnych lokalizacji kolonii ze względu na zaobserwowane różnice wartości średnich i median stężeń poszczególnych analizowanych pierwiastków w zależności od lokalizacji kolonii.

Punkty VI i VII to Dyskusja wyników oraz Wnioski końcowe. Doktorant ma świadomość, że zaimplementowany w sieci bayesowskiej proces obliczeniowy realizowany jest z dokładnością do rozkładu prawdopodobieństwa, a fakt że zmienne ciągłe są reprezentowane przez zadaną z góry liczbę przedziałów przekłada się na dokładność obliczeń. Ale znajomość łącznego rozkładu prawdopodobieństwa będąca istotą modelowania w technologii sieci bayesowskich jest kompletną informacją w przeciwieństwie do metod klasycznej statystyki matematycznej bazującej na średnich i wariancjach.

Występujące w modelu węzły obliczeniowe reprezentują zmienne będące deterministycznymi funkcjami losowych argumentów. Takie węzły również reprezentują zmienne losowe o warunkowych rozkładach prawdopodobieństwa zmiennych będących ich argumentami. Stąd na sieć tego typu można patrzeć jak na program obliczeniowy na rozkładach prawdopodobieństwa.

Pragnę zauważyć, że sporządzanie wykresów powierzchni w sytuacji kiedy dysponujemy tylko skończonym zbiorem punktów pomiarowych, a nie znamy przebiegu zależności pomiędzy tymi punktami nie zawsze prawidłowo oddaje charakter występujących zmian.

Wszystkie wyniki, a także ich analiza zostały przedstawione przejrzysto i adekwatnie do sformułowanego celu. Poprawny dobór rysunków oraz tabel ułatwia czytelnikowi śledzenie prowadzonych rozważań.

Jestem przekonany, że sformułowany cel ocenianej rozprawy, dobór metod badawczych, poprawna analiza, przeprowadzenie eksperymentów laboratoryjnych i symulacyjnych oraz umiejętne wnioskowanie stanowią o rozwoju naukowym Doktoranta. Przedstawione wyniki przyczyniają się do poszerzenia wiedzy naukowej oraz stanowią podstawę do praktycznych zastosowań.

3. Ocena formy rozprawy

Forma pracy jest poprawna. Można mieć pewne uwagi do strukturalizacji materiału, ale generalnie uznaję, że został on poprawnie zorganizowany. Pomimo widocznej staranności Doktoranta o przejrzystość pracy, stronę edytorską i poprawność językową to jednak Doktorant nie uniknął potknięć i braków. Oto niektóre z nich:

- Niekonsekwencja w wyróżnianiu akapitów – brak wyróżnienia lub stosowanie różnych form.
- Na str. 25 pominięto termin „liczby”,
- Na str. 25 (2 akapit od dołu) jest „Węzeł Pb_ Kos_Stu...” powinno być „węzeł Hg_Kos_Stu...”.
- Na str.63 (ostatni akapit) powinno być „Czyni to czapkę siwą ...”.
- Podobne usterki występują na stronach 3, 4, 5, 11 i 39.
- Nie wszystkie pozycje w Bibliografii są umieszczone w porządku alfabetycznym – pozycje 49 i 51 powinny być w innym miejscu.
- Rys 5, 6 i 7 zyskały by na „czytelności” jeśli monitorki reprezentujące rozkłady prawdopodobieństwa w postaci wykresów słupkowych byłyby usytuowane w tej samej kolejności.
- Zastrzeżenia budzi strona interpunkcyjna w odniesieniu do podpisów tabel i rysunków oraz zasad odwoływania się do poszczególnych rysunków.
- Materiał graficzny nie zawsze jest starannie opracowany.

4. Szczegółowe uwagi i komentarze

W trakcie czytania pracy nasunęły mi się następujące uwagi:

- Analizując podział treści uważam, że pkt. IV i V powinny być przedstawione jako jeden rozdział i umieszczone wcześniej, gdyż przedstawiane tam zagadnienia miały istotne znaczenie na etapie konkretyzacji modelu.
- Czy ocena stanu jakości na podstawie pomiaru stężeń wybranych pierwiastków w skorupkach jaj czapli siwej może być uznana za ocenę o charakterze globalnym czy też raczej jako element bardziej złożonych systemów oceny stanu jakości środowiska przyrodniczego?
- Analizując rozkłady prawdopodobieństwa wyznaczone na podstawie danych eksperymentalnych można zauważyć, że w przypadku niektórych pierwiastków (V_Lip, Sc_Lip, Cd_Kos_Stu i Cr_Kos_Stu) praktycznie cała masa prawdopodobieństwa jest skupiona w skrajnych przedziałach. Czy można to w jakiś sposób wyjaśnić?

Uwagi te nie obniżają końcowej pozytywnej oceny pracy. Doktorant zastosował prawidłowe, nowoczesne metody badawcze, wykazał się umiejętnością planowania i prowadzenia badań eksperymentalnych oraz opanował umiejętność implementacji procedur obliczeniowych w środowisku sieci bayesowskich.

5. Ocena końcowa

Uważam, że rozprawa doktorska mgr. inż. Marcina Rymarza „*Modelowanie jakości środowiska w języku sieci bayesowskich*” zawiera oryginalne rozwiązanie problemu naukowego mieszczącego się w dyscyplinie inżynieria rolnicza oraz potwierdza, że Doktorant dysponuje znaczną wiedzą i umiejętnościami niezbędnymi do prowadzenia badań naukowych w ramach wspomnianej dyscypliny. Oceniana przeze mnie praca spełnia wymagania stawiane rozprawom doktorskim określone w Ustawie o stopniach naukowych i tytule naukowym w związku z tym uważam, że Doktorant może być dopuszczony do dalszych etapów przewodu doktorskiego.

