Politechnika Śląska

UCZELNIA BADAWCZA
INICJATYWA DOSKONAŁOŚCI

Wydział Chemiczny
Katedra Inżynierii Chemicznej i Projektowania Procesowego

dr hab. inż.
Krzysztof Piotrowski
profesor Politechniki Śląskiej

Gliwice, 19.12.2024

Dr hab. inż. Krzysztof Piotrowski, prof. Pol. Śl.
Politechnika Śląska w Gliwicach
Wydział Chemiczny
Katedra Inżynierii Chemicznej i Projektowania Procesowego

Recenzja rozprawy doktorskiej mgr inż. Dawida Szpadzika

„Badanie możliwości wykorzystania narzędzi statystycznych w analizach jakościowych w obszarach produkcyjnych"

Przedstawiona rozprawa dotyczy ważnego zagadnienia związanego z opracowaniem programu (algorytmu) opartego na zależnościach statystycznych, który może być wykorzystany w analizach jakościowych i ilościowych związanych z procesami technologii chemicznej jako system ekspercki wspomagający procesy decyzyjne.

Charakterystyczną cechą tego nowatorskiego podejścia do problemu jest opracowanie w pełni funkcjonalnego algorytmu opartego o uczenie maszynowe i wykorzystującego wyłącznie standardowy pakiet oprogramowania MS Office. W związku z tym recenzowana rozprawa obejmuje następujące aspekty:

HR EXCELLENCE IN RESEARCH

Politechnika Śląska
Wydział Chemiczny
Katedra Inżynierii Chemicznej i Projektowania Procesowego

ul. ks. M. Strzody 7, pok. 194, 44-100 Gliwice
+48 32 237 19 00 / +48 605 647 949
krzysztof.piotrowski@polsl.pl

- poznawcze:
  a) identyfikacja możliwości redukcji liczby koniecznych analiz fizykochemicznych półproduktów w ramach procesu kontroli jakości i zastąpienia ich przez system ekspercki oparty o opracowany algorytm,
  b) sformułowanie koncepcji doboru modelu statystycznego jako algorytmu opartego na wykorzystaniu stosunkowo prostych procesów decyzyjnych,
  c) weryfikacja działania algorytmu na dostępnych danych przemysłowych,
  d) krytyczna analiza możliwości predykcyjnych systemu eksperckiego w oparciu o wybrane wskaźniki statystyczne,
- praktyczne:
  opracowanie i weryfikacja praktyczna przykładowego programu obliczeniowego – systemu eksperckiego, który umożliwia ograniczenie liczby koniecznych analiz fizykochemicznych półproduktu w ramach procesu kontroli jakości w konkretnym zakładzie przemysłu chemicznego.

  Zaprezentowane możliwości obliczeniowe programu mogą potencjalnie zostać uogólnione i zaadoptowane dla innych instalacji technologicznych, profili produkcji oraz różnych poziomów złożoności procesu technologicznego.

Systematyczne zwiększanie zdolności produkcyjnych instalacji technologicznych z jednoczesnym zapewnieniem coraz wyższej jakości produktów jest bezpośrednio związane z rozwojem zaawansowanych metod analityki przemysłowej. Konieczne jest tu wykonywanie wiarygodnych i szybkich rutynowych i standardowych analiz na różnych etapach procesu technologicznego.

Sprostanie rosnącym wymaganiom strategii Przemysłu 4.0 zakłada systematyczne wprowadzanie automatycznego, komputerowego sterowania procesem na różnych jego etapach. Złożoność cykli wytwórczych i związane z tym stopniowe zwiększanie stopnia skomplikowania różnych algorytmów predykcyjnych wymagają zastosowania coraz bardziej skomplikowanego, a tym samym kosztownego, profesjonalnego oprogramowania komercyjnego, sterującego procesami decyzyjnymi w oparciu o m.in.

algorytmy sztucznej inteligencji, sztuczne sieci neuronowe, itp. Jednak z uwagi na stosunkowo wysokie koszty inwestycyjne, oprogramowanie tego typu może być zwykle stosowane tylko w dużych instalacjach technologicznych, gdzie wysokie zdolności produkcyjne uzasadniają ekonomicznie wysokie koszty zakupu licencji. Bariera kosztów powoduje, że dla mniejszych instalacji technologicznych wprowadzenie tego typu oprogramowania może być nieopłacalne, co skutecznie spowalnia powszechną informatyzację w technologii chemicznej. Sposobem skutecznego ominięcia bariery ekonomicznej we wdrażaniu zasad Przemysłu 4.0 w instalacjach technologii chemicznej może być indywidualne opracowanie oprogramowania sterującego lub pełniącego rolę systemów eksperckich. Mogą być tu zastosowane konfiguracje standardowych komputerów PC oraz pakiety podstawowego oprogramowania biurowego wykorzystujące bazy danych lub arkusze kalkulacyjne.

Dokonany przez Doktoranta krytyczny przegląd i analiza danych literaturowych dotyczących metod uczenia maszynowego w inżynierii chemicznej oraz metod klasyfikacji binarnej wskazał na potencjalną możliwość zastosowania standardowego komputera osobistego i stosunkowo prostego algorytmu klasyfikacji binarnej w procesie kontroli jakości w przemyśle chemicznym, z możliwością uogólnienia koncepcji podstawowego algorytmu decyzyjnego w odniesieniu do instalacji o dowolnym poziomie złożoności technologicznej.

Podjęcie przedstawionej, aktualnej tematyki przez Doktoranta jest więc w pełni uzasadnione, zarówno z naukowego, jak i praktycznego (wdrożeniowego) punktu widzenia.

Celem rozprawy było opracowanie algorytmu obliczeniowego opartego o metodę uczenia maszynowego (naiwny klasyfikator Bayesa) oraz określenie możliwości jego wykorzystania w kontroli analitycznej procesów technologii chemicznej.

Rozprawa doktorska mgr inż. Dawida Szpadzika składa się z następujących rozdziałów:

- Wstęp,
- Przegląd literatury,
- Część eksperymentalna,
- Wyniki badań,
- Dyskusja wyników,
- Wnioski,
- Literatura.

Układ rozprawy jako pracy naukowej jest poprawny formalnie.

Główne innowacyjne elementy rozprawy doktorskiej obejmują:

a) dobór metody klasyfikacji binarnej i jej oceny statystycznej (określenie wymagań stawianych aplikacji komputerowej mającej spełniać rolę systemu eksperckiego w procesie kontroli jakości półproduktu),

b) opracowanie algorytmu eksperckiego klasyfikacji binarnej oraz - na jego podstawie - opracowanie aplikacji komputerowej,

c) przeprowadzenie prac obliczeniowych związanych z weryfikacją działania programu (określenie możliwości klasyfikacyjnych algorytmu na podstawie dostępnych danych przemysłowych).

Ad. a)

Przeprowadzono analizę dostępnych w literaturze metod klasyfikacji binarnej, obejmującej algorytm typu regresji logistycznej, algorytm typu *k-najbliższych sąsiadów*, algorytm typu *drzewa klasyfikacyjne*, algorytm typu *losowy las decyzyjny* oraz naiwny klasyfikator Bayesa. Uwaga Doktoranta skupiła się na naiwnym klasyfikatorze Bayesa, stąd przedstawił On bardziej szczegółowe informacje

dotyczące prawdopodobieństwa warunkowego (twierdzenia Bayesa) oraz zasad jego zastosowania w opracowywanym algorytmie klasyfikacyjnym, ze szczegółowym uwzględnieniem zasad dyskretyzacji zmiennych. Informacje teoretyczne zostały uzupełnione przykładami obliczeniowymi.

Autor przedstawił także wybrane metody oceny klasyfikatorów binarnych. W przypadku tablicy pomyłek przedstawił definicje wskaźników porównań prognozy ze stanem faktycznym: TP, TN, FP, FN, AP, AN, PP, PN oraz relacje teoretyczne pomiędzy nimi.

Przedstawiono także definicje dokładności (ACC) i poziomu błędów (ERR), precyzji (PPV), czułości (TPR) i specyficzności (TNR), częstości fałszywych alarmów (FPR) oraz częstości fałszywych odkryć (FNR).

Autor podał definicję współczynnika korelacji Matthewsa (MCC) oraz współczynnika F_1-*score*, a także krzywej ROC.

Przedstawiono podstawowe wymagania stawiane funkcjonalności aplikacji komputerowej pełniącej funkcje eksperckie w zakresie kontroli jakości półproduktu na instalacji technologicznej.

Doktorant dokonał ponadto oceny dostępności oraz kompletności danych technologicznych, a także przeprowadził analizę szacowania ryzyka związanego z zastosowaniem proponowanego algorytmu (w formie programu komputerowego).

Z uwagi na wdrożeniowy charakter pracy, Autor uwzględnił w analizie definicje kryterium sukcesu w świetle celów pracy badawczej z jednoczesnym określeniem zakresu wymagań i potrzeb przedsiębiorcy.

Ad. b)

Na podstawie dostępnych informacji opracowana została koncepcja algorytmu eksperckiego klasyfikacji binarnej opartego o algorytm naiwnego klasyfikatora Bayesa. Wykorzystując dostępne archiwalne dane technologiczne, opracowana została aplikacja komputerowa w języku VBA *(Visual Basic for Applications)* pracująca w środowisku aplikacji *Microsoft Office Access*. Z uwagi na cel opracowania programu, jakim było zastąpienie elementu kontroli jakości, został opracowany interfejs użytkownika w formie graficznej.

Ad. c)

Opracowany program został poddany serii testów numerycznych weryfikujących wpływ poszczególnych parametrów na skuteczność klasyfikacji binarnej, będącej miarą przydatności praktycznej programu zastępującego element kontroli jakości półproduktu instalacji technologicznej.

Weryfikacji poddano parametry modelu klasyfikatora binarnego, zarówno nie związanego z charakterystyką próbek półproduktu, jak i parametrów charakteryzujących próbkę.

Przeanalizowano wpływ maksymalnej ogólnej liczby próbek uwzględnianej w obliczeniach, wymaganej minimalnej ogólnej liczby próbek, wymaganej minimalnej liczby próbek z klasy, a także wpływ dyskretnych parametrów algorytmu które nie charakteryzują bezpośrednio próbki.

Dokonano analizy wpływu modyfikacji parametrów, które charakteryzują próbki. Obejmowały one: pH, gęstość, lepkość, stężenie procentowe nadtlenku wodoru, stężenie wolnego chloru oraz wynik analizy suchej pozostałości.

Weryfikacji poddano także wpływ wartości parametrów dyskretnych, obejmujących cechy partii półproduktu, jak też porównanie półproduktu ze wzorcem.

Dokonano porównania wyników testów opartych na modyfikacji jednego parametru, a także opartych na jednoczesnej modyfikacji wszystkich parametrów jednocześnie.

Dokonano także analizy związanej z możliwością ograniczenia kosztochłonnych analiz fizykochemicznych.

Najważniejsze osiągnięcia rozprawy to:

1. Potwierdzenie możliwości zastosowania standardowego komputera osobistego i stosunkowo prostego algorytmu klasyfikacji binarnej (algorytm naiwnego klasyfikatora Bayesa) w jednym z elementów procesu kontroli jakości w przemyśle chemicznym, z możliwością potencjalnego uogólnienia koncepcji podstawowego algorytmu decyzyjnego (eksperckiego) w odniesieniu do instalacji o potencjalnie dowolnym poziomie złożoności.
2. Pozytywna weryfikacja możliwości wdrożenia opracowanego algorytmu w oparciu o założone wartości wskaźników klasyfikacji.
3. Podanie ogólnej koncepcji algorytmu klasyfikacji binarnej w oparciu o algorytm naiwnego klasyfikatora Bayesa w procedurach przemysłowej kontroli jakości, możliwego do adaptacji dla różnych instalacji technologicznych, dla którego implementacji wystarczy stosunkowo prosta konfiguracja sprzętowa komputera PC i standardowy pakiet oprogramowania *MS Office*.

Praca nasuwa następujące uwagi:

1. Autor podaje, że wykorzystywał wszystkie dostępne dane numeryczne dotyczące analizowanego procesu technologicznego jako dane treningowe. W tekście pracy nie ma informacji o ewentualnej weryfikacji poprawności zdolności klasyfikacyjnych programu za pomocą pewnego niezależnego podzbioru danych (o identycznej strukturze wektora informacji numerycznej jak wszystkie dane ze zbioru treningowego), które nie były zastosowane w procesie opracowania algorytmu - i tym samym byłyby prezentowane strukturze obliczeniowej po raz pierwszy. Jako rzeczywiste dane procesowe umożliwiłyby rzeczywistą, niezależną weryfikację testową poprawności klasyfikacji dla rzeczywistych sytuacji technologicznych.

   Czynności przedstawione przez Doktoranta w podrozdziale *5.4 Model testowania klasyfikacji*, opisane jako *„(...) przeprowadzono szereg testów, które w sposób iteracyjny modyfikowały wartość każdej nastawy względem punktu odniesienia (...)"* nie reprezentują procesu testowania, rozumianego klasycznie jako niezależnej weryfikacji poprawności działania programu przy użyciu nowych, rzeczywistych danych opisujących zachowanie obiektu analizy. Są one raczej rutynową procedurą badania wrażliwości (czułości) modelu wielu zmiennych kolejno względem poszczególnych zmiennych (tzw. *model sensitivity*).

   W takiej sytuacji należy także wziąć pod uwagę, że badając wrażliwość modelu wielu zmiennych względem próbkowania przyrostowego wybranej jednej zmiennej, konieczne staje się założenie pewnej kombinacji stałych wartości pozostałych zmiennych. Dla innej kombinacji wartości tych pozostałych zmiennych wrażliwość modelu względem tej samej analizowanej zmiennej może być już zupełnie inna – jest więc bezpośrednio uzależniona od zadanego wektora wartości stałych (tzn. ich pewnej kombinacji). Czy Doktorant wziął pod uwagę możliwość istnienia zróżnicowanej wrażliwości modelu względem tej samej zmiennej dla różnych kombinacji pozostałych parametrów (jako wartości stałych)?

2. Czy *trening* znaczył jakieś iteracyjne modyfikacje wartości parametrów programu? Jaka była charakterystyka samego procesu uczenia – czy wykorzystywał on reguły optymalizacji funkcji wielu zmiennych, np. poszukiwania minimum (maksimum) lokalnego lub minimum (maksimum) globalnego, jak ma to miejsce w przypadku uczenia sztucznych sieci neuronowych?
3. Czy został krytycznie przeanalizowany problem reprezentatywności danych przypadków (wektorów danych) dla rozpatrywanego procesu technologicznego? Jakie kryteria odnośnie reprezentatywności danych dla danego problemu obliczeniowego zostały w takim przypadku wzięte pod uwagę?
4. Dla poprawności działania programu kluczowy może być równomierny rozkład przypadków treningowych na klasy (bez „*lokalnych kumulacji*" przypadków reprezentujących wyłącznie jedną klasę). Czy Doktorant dokonał wstępnej analizy będących do dyspozycji danych pomiarowych, np. szacunkowej weryfikacji równomierności rozkładu wartości liczbowych poszczególnych parametrów?
5. **Str. 8**: Autor pisze, że „(...) *Potwierdzono możliwości wykorzystania narzędzi statystycznych w kontroli jakości w przemyśle chemicznym (cel badawczy pracy doktorskiej)* (...)". Metody statystyczne są już od wielu lat powszechnie stosowane w różnych zagadnieniach związanych z procesami produkcyjnymi, istnieje też dość bogata literatura naukowa dotycząca tych zagadnień. Przykładowe ogólnodostępne pozycje literaturowe: Grzegorz Kończak „*Metody statystyczne w sterowaniu jakością produkcji*", Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice 2009 - opisująca m.in. przyjęte rozwiązania z zakresu sterowania jakością w czasie procesu produkcyjnego, zagadnienia kontroli materiału będącego efektem tej produkcji, z zastosowaniem analizy porównawczej różnych rozwiązań. Inna przykładowa pozycja to: Andrzej Iwasiewicz „*Statystyczna kontrola jakości w toku produkcji. Systemy i procedury*", Wydawnictwo Naukowe PWN, Warszawa 1985, która omawia metody statystyczne możliwe do zastosowania w sterowaniu jakością. Ponadto w ogólnodostępnej ofercie szkoleń firmy *STATSOFT* pojawiają się szkolenia dotyczące podjętej przez Doktoranta problematyki, jak m.in.: „*Statystyka w walidacji metod analitycznych*", „*Analiza danych i sztuczna inteligencja w Przemyśle 4.0*", „*Statystyka w przemyśle – kurs podstawowy*" i inne.

Politechnika Śląska

UCZELNIA BADAWCZA

Dlatego też w celu uwypuklenia i wyraźnego podkreślenia elementów nowości pracy, tytuł rozprawy doktorskiej powinien raczej bardziej oddawać rzeczywisty zakres i cel pracy, jak np. „Zastosowanie algorytmu naiwnego klasyfikatora Bayesa w problematyce binarnej klasyfikacji decyzyjnej uwzględniającej specyfikę złożonych procesów inżynierii chemicznej". Wydaje się również, że Autor powinien raczej ukierunkować opis swych prac badawczych, obliczeniowych i symulacyjnych na wielokrotnie podkreślaną przez Niego w tekście problematykę minimalizacji struktury algorytmu (np. optymalizacja konfiguracji algorytmu jako celowe upraszczanie wstępnie rozbudowanej nadmiarowo tzw. superstruktury lub stopniowe rozszerzanie struktury minimalnej, modele równoległego przetwarzania danych, modele kaskadowe, itp.) przy zachowaniu w pełni wymaganej funkcjonalności i efektywności procesu predykcji i klasyfikacji binarnej koniecznej dla spełnienia celów praktycznych w konkretnej instalacji technologicznej.

Ponadto Autor powinien (w dyskusji wyników) bardziej wyeksponować wyraźne możliwości uogólnienia struktury algorytmu dla jego implementacji w kontroli jakości produktu (lub półproduktu) z dowolnej instalacji technologicznej. Zamiast tego Doktorant koncentruje swą uwagę raczej wyłącznie na specyficznych szczegółach konkretnej instalacji służącej za punkt odniesienia w weryfikacji algorytmu i powstałego na jego podstawie programu obliczeniowego. Jednak z punktu widzenia nowości naukowej jest to tylko „*studium konkretnego przypadku*" z własną, charakterystyczną i unikatową strukturą danych i konkretnych wymagań procedur kontroli jakości, których dotyczą wszystkie omawiane przez Autora szczegóły obliczeń i zdolności predykcji – bez wyraźnie zaznaczonego charakteru omówienia ukierunkowanego na pewne uogólnienie proponowanego algorytmu klasyfikacji binarnej i krytycznej weryfikacji możliwości jego stosowania dla zróżnicowanych zestawów kryteriów kontroli jakości (pół)produktów.

Nawet w przypadku założenia pełnej reprezentatywności użytych archiwalnych danych numerycznych, dodatkowa weryfikacja tego samego programu np. dla innej, podobnej instalacji produkcyjnej mogłaby już dostarczyć pewnych informacji związanych z możliwością

uogólnienia, wskazania ewentualnych ograniczeń, jak i wskazówek dla potencjalnie wszechstronnego zastosowania proponowanego podejścia numerycznego, niemniej bardzo interesującego z poznawczego i aplikacyjnego punktu widzenia.

Uwagi szczegółowe:

6. **Str. 8:** Autor podaje sformułowanie „*Populacji testów*" które jest niejasne.
7. **Str. 8**: Autor pisze: „*Prace optymalizacyjne wykazały (...)*" , jednak nie jest znane jakie było przyjęte kryterium optymalizacji i jaka metoda optymalizacji została wybrana dla celów obliczeń.
8. **Str. 15**: Podane sformułowanie „*Model testowania klasyfikacji (...), tzw. trenowania algorytmu*" jest niejasne, gdyż trenowanie struktury obliczeniowej nie jest równoznaczne z jej testowaniem. Czy chodziło Autorowi o zasady stosowane np. podczas uczenia sztucznych sieci neuronowych, gdzie trening jest zwykle stosowany wraz z równolegle prowadzoną walidacją, a dopiero po zakończeniu treningu (jak i walidacji) następuje niezależne testowanie na podstawie zupełnie nowych wektorów danych?
9. **Str. 15**: Podane w tekście rozprawy sformułowanie: „*Opisano realizację modelu testowania...*" jest niejasne.
10. **Str. 15**: Sformułowanie „*(...) optymalne wartości nastaw*" powinno raczej mieć formę „*(...) optymalne wartości parametrów programu*".
11. **Str. 22**: Doktorant podaje, że „*(...) w danych, które służą do trenowania algorytmów nie muszą być zdefiniowane oraz oznaczone wzorce (...)*" – zgodnie z tym zapisem oznacza to, że w danych służących do opracowania algorytmu nie ma możliwości zidentyfikowania żadnych wielowymiarowych schematów powiązań pomiędzy zmiennymi, a więc nie wnoszą one żadnego wkładu merytorycznego do tego algorytmu. Natomiast dla poprawnego skonstruowania modelu dane treningowe rozumiane są jako „*informacja o relacjach w zbiorze danych podana w formie reprezentatywnych przykładów*". Tym samym wzorce te są podane „*pośrednio*" – jako „*ukryte*" w wielowymiarowej strukturze wartości numerycznych reprezentujących dane archiwalne.

12. **Str. 23**: Sformułowanie *„Zespół badawczy może zwrócić szczególną uwagę na opracowywanie algorytmu oraz optymalizowanie jego predykcji, zamiast poświęcać czas na przygotowanie zbiorów danych"* jest niejasne, ponieważ wszystkie parametry programu obliczeniowego wyznaczane są na podstawie pewnego zbioru danych – wiarygodność i reprezentatywność elementów tego zbioru (wynikająca z właściwego ich doboru, innymi słowy – przygotowania, selekcji, wstępnej weryfikacji, itp.) warunkuje bezpośrednio przydatność obliczeniową opracowanego na ich podstawie algorytmu obliczeniowego.
13. **Str. 23**: Sformułowanie *„Modele regresji (liniowej oraz wielozmiennej)"* jest nie do końca jasne, sugerując, że modele regresji wielu zmiennych są z natury modelami nieliniowymi. Funkcja wielu zmiennych również może być liniowa – jest to tak zwana *regresja wieloraka*, stosowana w przypadku, gdy analiza opiera się o więcej niż jedną zmienną niezależną. W takich przypadkach równanie liniowe ma formę: $Y = b_0 + b_1X_1 + b_2X_2 + ... + b_kX_k$
    w którym to równaniu k oznacza liczbę tzw. *predyktorów* (*zmiennych objaśniających*). W przypadku takiej formy równania określane współczynniki regresji ($b_1 ... b_k$) wyrażają niezależny udział każdej zmiennej niezależnej w określaniu wartości liczbowej poszukiwanej zmiennej zależnej. Alternatywnie można stwierdzić, że zmienna niezależna X_1 jest skorelowana ze zmienną zależną Y uwzględniając jednocześnie wpływ wszystkich innych zmiennych niezależnych $X_2 ... X_k$ (jest to tzw. *korelacja cząstkowa*).

    (*materiały szkoleniowe StatSoft Electronic Statistics Textbook, https://www.statsoft.pl/textbook/stathome_stat.html).*
14. **Str. 40**: Doktorant podaje: *„Integralną częścią tworzenia narzędzi na bazie uczenia maszynowego jest weryfikacja ich działania. (...) Odpowiednia ocena dostarcza informacji zwrotnej o efektach trenowania algorytmu oraz sygnalizuje użytkownikowi poziom jakości (skuteczności) prezentowanej klasyfikacji"* - czy taka weryfikacja była przeprowadzona w trakcie trenowania (tzn. jako *walidacja*), czy już po zakończeniu procesu trenowania (tzn. jako *niezależne testowanie*)?
15. **Str. 48**: Użyte sformułowanie *„(...) dające się recyklingować (...)"* powinno brzmieć *„(...) podatne na procesy recyklingu (...)"*.

16. **Str. 54**: W sformułowaniu podanym przez Doktoranta *„(...) problem badawczy niniejszej rozprawy ma na celu wykazać, że jest możliwa cyfryzacja jednego z elementów procesu kontroli jakości (...)"* – czy słowo *„cyfryzacja"* oznacza w tym ujęciu wspomaganie procesów kontroli jakości procesami numerycznymi, czy też ich całkowite zastąpienie?
17. **Str. 63**: Użyte niewłaściwe sformułowanie: *„W ramach tych czynności wykonywano wystandaryzowane analizy parametrów fizykochemicznych"* które powinno być podane jako *„analizy parametrów fizykochemicznych zgodnie z określonymi standardami"*.
18. **Str. 64**: W sformułowaniu *„Baza danych produkcyjnych (...) zawierała (...) dane o 538 półproduktach, całkowita liczba próbek wyniosła 44236"* brak jest kryterium ścisłego rozróżnienia, jaka jest definicja *„półproduktu"*, a jaka *„próbki"*? Brak jest informacji, jakie były przyjęte relacje pomiędzy tymi pojęciami.
19. **Str. 67**: Autor podaje, że: *„Wyświetlane są (...) dane, na podstawie których zostały one obliczone. Po każdej zapisanej klasyfikacji są one aktualizowane"* – czy każdy kolejny przypadek będzie tym samym wpływał na systematyczną korektę aktualnych wartości parametrów programu? Czy program cały czas się uaktualnia (*„uczy"*) w trakcie użytkowania (tzn. z wciąż poszerzanej i uaktualnianej na bieżąco bazy danych historycznych)? Natomiast na **str. 69** Autor pisze, że: *„(...) istniejącą aplikację produkcyjną zabezpieczono przed modyfikacją."* – czy zabezpieczenie odnosi się więc tylko do struktury programu klasyfikacji binarnej, a zbiór danych z którego program korzysta może być niezależnie wciąż niezmiennie uaktualniany wraz z kolejnymi przeprowadzanymi analizami?
20. **Str. 74**: Na Rysunku 5.7 przedstawiającym okno robocze programu podano wartości parametrów *Kp(Zawrócić)* oraz *Kp(Ściek)* z dokładnością do 15 miejsc znaczących – czy tak duża dokładność jest uzasadniona?
21. **Str. 85**: Sformułowanie: *„Wraz z inkrementacją wartości parametru (...)"*, **str. 86:** *„Parametr był modyfikowany inkrementacyjnie (...)"*, *„(...) wraz z kolejnymi inkrementacjami parametru"* powinno mieć formę: *„wraz z (systematycznym) przyrostem wartości parametru"*.

22. **Str. 93**: Autor podaje, że: „*Zgodnie z matrycą testów (...) przeprowadzono 90 testów*" nie podając żadnych bardziej szczegółowych danych dotyczących planu badań ani jakimi przesłankami kierował się planując tę liczbę testów.
23. **Str. 106**: Doktorant podaje, że: „*(...) po fazie optymalizacji każdego parametru osobno wysterowano wszystkie parametry jednocześnie (...)*" – w tym miejscu pojawia się pytanie, czy optymalizując wartości każdego parametru indywidualnie, a potem stosując kombinację otrzymanych w ten sposób wartości uzyska się także wartość optymalną? Optymalne wartości indywidualnych parametrów (ważne tylko i wyłącznie dla zadanych, stałych wartości wszystkich pozostałych parametrów) nie uwzględniają złożonych powiązań wzajemnych pomiędzy wszystkimi parametrami układu (synergia), których jednoczesne uwzględnienie dopiero umożliwia określenie rzeczywistego optimum. Problem ten dotyczy także zdania ze **str. 58**: „*(...) Następnie wysterowane zostały wszystkie atrybuty jednocześnie uwzględniając optymalne wartości z poprzednich testów (...)*".
24. **Str. 110**: Użyte sformułowanie: „(...) *W każdym przypadku optymalnego wysterowania obserwowane były lepsze wartości współczynników względem punktu odniesienia*" jest niejasne.

Uwagi te nie umniejszają końcowej pozytywnej oceny rozprawy.

Stwierdzam, że przedłożona rozprawa doktorska mgr inż. Dawida Szpadzika spełnia wymagania Ustawy z dnia 20 lipca 2018 r. „Prawo o szkolnictwie wyższym i nauce" Dz. U. z 2024 r. poz. 1571.

Wnoszę do Rady Naukowej Dyscypliny Inżynieria Chemiczna Politechniki Warszawskiej o przyjęcie rozprawy i dopuszczenie jej do publicznej obrony.

Krzysztof Piotrowski

Dr hab. inż. Krzysztof Piotrowski, prof. Pol. Śl.