



**POLITECHNIKA
RZESZOWSKA**
im. IGNACEGO ŁUKASIEWICZA

al. Powstańców Warszawy 12, 35-959 Rzeszów
tel./fax: +48 17 854 12 60, tel.: +48 17 865 11 00
www.prz.edu.pl

dr hab. inż. Dominik Strzałka, prof. PRz

Rzeszów, dn. 05.12.2022 r.

Politechnika Rzeszowska

Wydział Elektrotechniki i Informatyki

Zakład Systemów Złożonych

Tytuł rozprawy: Zwiększenie niezawodności pracy urządzeń bloków energetycznych poprzez zastosowanie analityki predykcyjnej

Autor rozprawy: mgr inż. Marek Mołęda

Promotor rozprawy: dr hab. inż. Dariusz Mrozek, prof. PŚ

Dziedzina: nauki inżyniersko-techniczne

Dyscyplina: informatyka techniczna i telekomunikacja



ZAKŁAD SYSTEMÓW ZŁOŻONYCH
Wydział Elektrotechniki i Informatyki
ul. Wincentego Pola 2, 35-959 Rzeszów, tel. 17 865 1340
zsz.prz.edu.pl



Syntetyczne przedstawienie rozprawy

Cel pracy

Podstawowym celem pracy było zweryfikowanie i wskazanie w elektrowniach opalanych węglem tych obszarów, gdzie jest możliwe oraz będzie efektywne użycie technik predykcyjnego utrzymania ruchu (ang. predictive maintenance), dla których opracowano narzędzia oraz modele analityczne realizujące te zadania.

Jako szczegółowy cel badawczy wskazano opracowanie modelu matematycznego umożliwiającego przewidywanie nadchodzących usterek i awarii na podstawie danych zbieranych w systemach IT oraz systemach technologicznych. Na podstawie historycznych danych pomiarowych oraz informacji o awariach z Elektrowni Łaziska w Łaziskach Górnych i Elektrowni Jaworzno III zaproponowano heurystyczny model predykcyjny.

Zakres pracy

Rozprawa doktorska składa się z 5 zasadniczych rozdziałów, które kolejno obejmują: Rozdział 1 – wprowadzenie do pracy, Rozdział 2 – część teoretyczna, Rozdział 3 – opis zastosowanej metodologii, narzędzi oraz autorskich metod dla zaprojektowanych eksperymentów, Rozdział 4 – przedstawienie uzyskanych wyników badań w odniesieniu do zdefiniowanego problemu badawczego i postawionych tez, Rozdział 5 – podsumowanie.

Charakter rozprawy

Praca ma charakter teoretyczno-eksperymentalny ze szczególnym naciskiem na wskazanie jej wdrożeniowego charakteru. Przeprowadzone badania bazują na danych rzeczywistych pozyskanych z dwóch polskich przedsiębiorstw: Elektrowni Łaziska w Łaziskach Górnych oraz Elektrowni Jaworzno III.





Opinia:

- o poprawności o oryginalności postawionej tezy i stopniu w jakim została ona wykazana.

W pracy postawiono 3 tezy:

1. Zastosowanie opartych na regresji metod redundancji analitycznej operujących na dużych zbiorach danych z systemów przemysłowych pozwala na realizację zadania przewidywania awarii i usterek w sytuacji:

- ograniczonej liczby analizowanych cech i zdarzeń,
- zmiennych warunków pracy wynikających z częstych remontów i wpływu czynników zewnętrznych.

2. Algorytm adaptacyjnego okna przesuwonego oparty na analizie średnich wartości cząstkowych elementów tego okna pozwala zachować parametry jakościowe modelu predykcyjnego wykrywającego usterki, jednocześnie ograniczając potrzebę okresowej nadzorowanej aktualizacji modeli.

3. Proces inżynierii danych i tworzenia modeli predykcyjnych oparty na danych generowanych maszynowo z urządzeń przemysłowych może zostać zautomatyzowany dla zastosowania względem nowych urządzeń, zachowując zbliżone zdolności predykcyjne.

Postawione tezy mają oryginalny i nowatorski charakter w odniesieniu do szerokiego kontekstu całej rozprawy, która ma: (i) charakter doktoratu wdrożeniowego, (ii) powstała w odniesieniu do konkretnych danych pozyskanych z polskich elektrowni opalanych węglem oraz (iii) odnosi się do realnej i bieżącej potrzeby wdrażania w polskim przemyśle rozwiązań bazujących na predykcyjnym utrzymaniu ruchu. Jest to możliwe nie tylko ze względu na wystarczającą moc obliczeniową systemów komputerowych przetwarzających dane, ale także jest wynikiem tego, iż w polskich przedsiębiorstwach w wielu miejscach gromadzi się bardzo duże ilości różnych danych, które niestety nie są dalej przetwarzane.

Analizując treść pracy, zakres wykonanych prac eksperymentalnych i badawczych oraz włożony indywidualny wkład autorski należy stwierdzić, że wszystkie 3 tezy zostały w pracy





udowodnione w odniesieniu do konkretnych, rozważanych w pracy przykładów obejmujących studia przypadków: 1) zespołu pomp zasilających blok energetyczny Elektrowni Łaziska w Łaziskach Górnych, 2) zespołu sprężarek natleniających Elektrowni Jaworzno III.

- o analizie źródeł:

W pracy użyto referencji do ponad 200 prac naukowych. Zdecydowana większość z nich to prace z ostatnich 8-10 lat, co wskazuje na bardzo dużą aktualność poruszanego tematu oraz jego duże znaczenie we współcześnie prowadzonych badaniach zarówno teoretycznych jak i eksperymentalnych także z naciskiem na wdrażanie proponowanych rozwiązań.

W przypadku własnych prac naukowych autora rozprawy doktorskiej mamy odwołania do 5-u publikacji, przy czym publikacja pt. *Evolution of maintenance approaches towards smart power industry - a review* jest nadal w recenzji. Nie są to samodzielne prace autora, ale powstały w zespole obejmującym Pana prof. Dariusza Mrozka (promotora) oraz Panią dr Alinę Momot – promotora pomocniczego. Trzy z tych artykułów zostały opublikowane w materiałach konferencyjnych. Prace naukowe mają związek z tematem i zakresem rozprawy doktorskiej.

- o pozycji rozprawy w stosunku do stanu wiedzy i/lub stanu techniki reprezentowanych przez literaturę światową

Jeżeli spojrzeć na bardzo duży dorobek publikacyjny z ostatnich 10-u lat w zakresie zastosowań metod sztucznej inteligencji, uczenia maszynowego, analizy dużych zbiorów danych oraz użycia dostępnej mocy obliczeniowej do zastosowań praktycznych także w zakresie utrzymania ciągłości produkcji oraz utrzymania ruchu, to praca idealnie wpisuje się w ten zakres. Generalnie, obszar predykcyjnego utrzymania ruchu wydaje się być w chwili obecnej jednym z najciekawszych w związku z tym, że z punktu widzenia działalności produkcyjnej wielu przedsiębiorstw jest najbardziej ekonomicznie opłacalny (wykrywanie z wyprzedzeniem usterek na podstawie danych procesowych) zaś w systemach infrastruktury krytycznej (np. produkcji energii elektrycznej) gwarantującym w największym stopniu nieprzerwaną pracę. Bardzo duża liczba odwołań do literatury użyta przez autora, wskazuje na bieżącą aktualność poruszanego tematu zaś fakt, iż był on w stanie skorzystać z rzeczywistych danych





pochodzących z polskich przedsiębiorstw wnosi tutaj dodatkowe walory poznawcze oraz aplikacyjne.

- o znaczeniu uzyskanych wyników dla danej dyscypliny naukowej

Z punktu widzenia oceny uzyskanych wyników dla dyscypliny informatyka techniczna i telekomunikacja należy wskazać jako najważniejsze to, iż autor był w stanie bardzo szczegółowo pokazać, po drodze rozwiązując wiele praktycznych problemów, w jaki sposób dotychczas (tj. w tym konkretnym przypadku, od około 10-u lat) gromadzone dane w systemach przemysłowych można w efektywny sposób przetworzyć i uzyskać z nich dodatkową wiedzę, którą można zastosować w sposób praktyczny w utrzymaniu ruchu w przedsiębiorstwach będących elementem infrastruktury krytycznej produkcji energii elektrycznej w elektrowniach opalanych węglem w Polsce. W dyscyplinie informatyka techniczna i telekomunikacja tematyka predykcyjnego utrzymania ruchu nie jest nowa zaś zastosowane przez autora rozwiązania wpisują się w typowe podejścia w tym zakresie. Ale autor samodzielnie zaproponował kilka własnych modyfikacji istniejących algorytmów, które w tych konkretnych przypadkach w znaczący sposób sprawiły, że uzyskiwane wyniki końcowe były zdecydowanie lepsze względem innych możliwych do użycia metod.

Było to także możliwe dzięki temu, że autor rozprawy miał bezpośredni dostęp do dużych zbiorów danych rzeczywistych, czyli nie bazował na symulacjach lub ogólnodostępnych zbiorach testowych. Każdy, kto kiedykolwiek miał pomysł na potencjalną możliwość nawiązania współpracy z przemysłem, wie, że zdobycie zaufania przedstawicieli przemysłu (w tym przypadku przedsiębiorstw będących częścią krajowych systemów infrastruktury krytycznej) jest bardzo dużym wyzwaniem. Praca ma wybitnie praktyczny charakter odnoszący się dwojako: do istniejących w przedsiębiorstwach problemów wraz z potrzebami ich rozwiązywania oraz do budowania zaufania pomiędzy polskim przemysłem i polskimi jednostkami naukowymi. Autor w tym zakresie pokazuje, że jest możliwe uzyskanie dostępu do dużych zbiorów danych przemysłowych, wykonanie dla nich odpowiednich i wiarygodnych analiz oraz przedstawienie modelu biznesowego dla uzyskanych wyników i zaproponowanego rozwiązania. Ponadto, w pracy pokazano dla środowiska przemysłowego, w jaki sposób można używać technologii IT w przemyśle.





- o umiejętności autora do poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników.

Autor operuje w rozprawie dość zwięzłym jednak nie zawsze precyzyjnym, poprawnym i przekonującym językiem. Liczne usterki w odniesieniu do prezentacji danych w formie graficznej (szczegóły także w dalszej części recenzji) miejscami mogą nawet obudzić wątpliwości co do przekonującego przedstawienia uzyskanych wyników. Podczas lektury pracy można wskazać kilkanaście fragmentów, które u czytelnika mogą wzbudzić pewne wątpliwości a także pytania. Poniżej lista najważniejszych z nich:

Str. 16 Rysunek 2.1 przygotowany przez autora można zinterpretować w sposób, który zaprzeczyłby skuteczności podejścia predykcyjnego. Symbolem czerwonego X oznaczono „wystąpienie usterki”, ale jeżeli zastosować podejście prewencyjne lub predykcyjne, to znaczenie tego symbolu (podane w opisie) powinno być inne np. jako: potencjalne wystąpienie usterki, bowiem prewencja lub predykcja mają właśnie zapobiegać występowaniu awarii a rysunek sugeruje, że i tak one się pojawiły.

Str. 52, Autor napisał: *Obliczanie transformaty Fouriera w kolejnych przedziałach czasowych sprawia, że metoda ta jest skomplikowana obliczeniowo.* Istnieją bardzo szybkie algorytmy obliczania transformaty Fouriera, które nawet dla szeregów czasowych liczących milion rekordów zwracają wyniki w bardzo krótkim czasie.

Str. 71, Autor napisał: *Na wykresie (rys 3.3 – dod. DS.) widać okres, w którym uszkodzony był zawór minimalnego przepływu, stąd widoczny był duży błąd predykcji.* Raczej tego na tym wykresie nie widać lub też nie zostało to dobrze wyeksponowane.

Str. 73, Autor napisał: *W procesie czyszczenia poprawiane są wiersze, w których brakuje wartości (brakujące wartości ustawiane są na 0).* Dlaczego przyjęto wartość 0 i jaki to ma wpływ na uzyskiwane wyniki? Jak dużo było takich przypadków?

Str. 80, Autor napisał: *... zakładając, iż dystrybucja błędów ma rozkład zbliżony do normalnego $N(0, 1)$ – ale czy tak jest na pewno w tym przypadku?*





Str. 98, Autor napisał: ... *odczytywaniu i interpretowaniu pomiarów z aparatury pomiarowej widocznej na rys. 4.2* – jednak taka aparatura na tym rysunku nie jest widoczna, bo nie jest zaznaczona.

Str. 100, Autor napisał: *Istotną cechą zbioru danych jest występowanie znaczących korelacji pomiędzy poszczególnymi grupami sygnałów*, ale znów należało podkreślić, że są to korelacje liniowe.

Str. 110, Autor napisał: *Znacząca korelacja pomiędzy ...*, jednak biorąc pod uwagę założone podejście była to *znacząca korelacja liniowa* pomiędzy ...

Str. 115, Autor napisał: ... *pomiędzy różnymi zmiennymi istnieją znaczące zależności* – należałoby dodać słowa *zależności liniowe*, bowiem w tym samym akapicie mamy odwołanie do obliczeń współczynnika korelacji Pearsona.

Str. 118, Autor napisał: *W przypadku wyższych stopni wielomianu model predykcyjny ujawnia więcej nietypowych sytuacji, jednak działa bardzo niestabilnie i jest podatny na zakłócenia*. Jak należy to rozumieć, czym m.in. są „nietypowe sytuacje” i czy nie można z takiego stwierdzenia wyciągnąć wniosku, że są to anomalie i/lub awarie?

Str. 120, Autor napisał: ... *histogram czerwony jest zbliżony do rozkładu normalnego* – na jakiej podstawie pada takie stwierdzenie?

Str. 122, Autor napisał: *pokrywają się z większymi odchyleniami wskazanymi przez model predykcyjny ...*, ale analiza rysunku 4.14 nie do końca przekonuje do takiego stwierdzenia, bowiem wynika z niego, że nie do końca się pokrywają, np. dwie ostatnie awarie (czerwone oznaczenia) na rysunku górnym oraz awarie 3, 4 i 5 na rysunku dolnym.

Str. 123, Autor napisał: *uzyskując dobre jakościowo wyniki predykcji*, ale autor nie objaśnia, co to wg niego znaczy i jak można to uzasadnić np. liczbowo.

Str. 127, Autor napisał: ... *Wybór tych sygnałów podyktowany był niewielkimi wzajemnymi korelacjami pomiędzy wartościami wybranych pomiarów* jednak dla uzasadnienia tego sformułowania brakuje macierzy takich jak na rys. 4.3 i 4.11.





Od strony redakcyjnej praca napisana jest bardzo starannie, choć należy zwrócić uwagę na kilka drobnych niedociągnięć technicznych (nazwijmy je literówkami):

- str. 66 tworzono model do oczekiwanych -> tworzono **modelu** do oczekiwanych
- str. 75 iż nie wyznaczana jest -> iż nie **jest** wyznaczana
- str. 100 dla których wartość zasilania -> dla których wartość **prądu** zasilania
- str. 103 przypadkowo podzielony akapit
- str. 126 działania algorytmy -> działania **algorytmu**
- str. 129 analiza ilości fałszywych alarmów -> analiza **liczby** fałszywych alarmów
- str. 135 które mają być wykryte nie był ustalany statycznie -> które mają być wykryte, nie był ustalany statycznie
- str. 135 Zarównow -> Zarówno **w**
- str. 140 ... wykrywania usterek w sytuacji ograniczonych informacjach o historycznych usterekach ... -> wykrywania usterek w sytuacji ograniczonych **informacji** o historycznych **awariach**
- str. 112 użycie zwrotu 'liczby danych'; powinno być 'ilości danych', bo dane to rzeczownik niepoliczalny.

Główne wady rozprawy i jej słabe strony.

W pracy nie ma bardzo poważnych usterek, ale można wskazać na kilka słabych stron natury edytorskiej oraz dbałości o odpowiedni przekaz informacji (zwłaszcza poprzez rysunki) do czytelnika, są to m.in.

1. W odniesieniu do wielu grafik (np. Rys. 2.3, 2.5, 3.1, 3.3, 3.7, 3.11, 4.1, 4.9, 4.13, 5.2) zaprezentowanych w pracy należy zauważyć, że nie są one przedstawione w zbyt dobrej





rozdzielczości; być może przydałaby się grafika wektorowa, nieco większe rysunki; pionowy (inny) układ rysunku.

2. W wielu przypadkach brakuje legend wyjaśniających np. znaczenie użytych kolorów w rysunkach:
 - 2.1 Rys. 4.3 i 4.11 – jak należy liczbowo interpretować odcienie kolorów niebieskiego i czerwonego – to są bardzo ważne rysunki w kontekście zaproponowanego podejścia.
 - 2.2 Rys. 2.4 – jak interpretować kolory zielone i czerwono pomarańczowe,
 - 2.3 Rys. 2.5 – dla arkusza (tabeli) można było podać np. charakter użytych skali (punktów) oceny dotkliwości, prawdopodobieństwa, wykrywalności, bo nie jest wiadome, w jaki sposób interpretować podane liczby, np. zazwyczaj prawdopodobieństwo jest wyrażane w skali 0 – 1.
 - 2.4 Wydaje się, że podpisy pod rys. 4.5 i 4.6 powinny brzmieć: Różnica między rzeczywistą a szacunkową wartością **temperatury** łożyska nr 1 oraz Różnica między rzeczywistą a szacunkową wartością **temperatury** łożyska nr 2 – brak słowa ‘temperatury’.
 - 2.5 Rys. 4.9 – brak legendy odnoszącej się do kolorów sygnałów pokazanych na rysunku oraz względem notacji KKS użytej w pracy, w treści pracy mamy tylko tekst: *Kolor linii na wykresie wskazuje sygnał, dla którego występuje największe odchylenie względne (NRE_{max}).*

Jeżeli chodzi stronę poprawności językowej, to autor używa języka polskiego w sposób poprawny, choć niestety nieco niepokojące jest nadużywanie słowa „wykorzystanie” (w pracy jest to zrobione kilkadziesiąt razy) w odniesieniu niemal do wszystkiego, co autor potrafił w pracy osiągnąć, co będzie stosowane, używane, co będzie miało lub ma jakiegokolwiek aplikacje. Zgodnie ze słownikiem języka polskiego słowo ‘wykorzystanie’ ma raczej negatywne znaczenie, tymczasem autor np. na str. 113 tylko w jednym akapicie pisze: ***Wykorzystanie danych z systemów technologicznych do celów innych niż były pierwotnie przeznaczone. Dane wykorzystywane do nadzoru procesu produkcji wykorzystano do diagnostyki i predykcji stanów awaryjnych.***





Uwagi szczegółowe

Uwagi szczegółowe w odniesieniu do pracy zostaną wyrażone w postaci pytań do autora rozprawy doktorskiej. Nie mają one charakteru krytycznego, będą jedynie próbą rozwiania ewentualnych wątpliwości powstałych m.in. w wyniku wyżej wskazanych usterek lub niewłaściwego sposobu przekazania istoty zagadnień, co do przyjętej metodyki, nieprecyzyjnego języka, który użyto miejscami w pracy.

1. W ramach pracy, już w streszczeniu autor napisał, że (cyt., str. 1) *szczególna uwaga została skupiona na metodach modelowania (...) pozwalających osiągnąć skalowalne rozwiązania*, ale patrząc na treść pracy należy wskazać, że ten problem nie jest do końca dobrze przeanalizowany. W szczególności widzimy w pracy skupienie się na konkretnych studiach przypadków, dla których uzyskane wyniki są bardzo dobre, ale z tych studiów przypadków nie można uzyskać odpowiedzi na następujące pytanie: jak zaproponowane rozwiązania mogą być skalowalne w sytuacji, gdy liczba sygnałów pomiarowych z urządzeń będzie większa niż np. 20 lub też sama liczba śledzonych urządzeń będzie bardzo duża?
2. W pracy mieliśmy do czynienia tylko z sytuacją, gdy sygnałów wejściowych było 22 (pompy wody) i 10 (sprężarki). Stąd wyznaczenie macierzy korelacji i ich interpretacja była dość łatwa (także obliczeniowo), ale gdyby sygnałów było 500 lub więcej, to tego typu zadanie nie byłoby już takie proste zwłaszcza gdy szeregi czasowe liczyłyby setki tysięcy i więcej rekordów. Co prawda, w analizowanych przypadkach mieliśmy do czynienia z elementami systemu, które stanowiły krytyczne fragmenty infrastruktury sprzętowej, ale można sobie wyobrazić, że takich krytycznych (z różnych względów) urządzeń w elektrowni i spływających sygnałów informujących o ich stanie jest zapewne o wiele więcej. Zatem rodzi się pytanie, na ile zaproponowane podejście byłoby obliczeniowo wykonalne, gdyby sygnałów (inaczej analizowanych metryk) było kilkaset lub kilka tysięcy? Które z nich należałoby wybrać do zaproponowanego podejścia, czy dla każdego z nich (lub wybranych grup sygnałów) należałoby zastosować podobną metodykę? W pracy, jak sam autor napisał (str. 143): *Występujące*





istotne korelacje pomiędzy sygnałami procesowymi sprawiły, iż zaproponowana metoda daje rezultaty o bardzo dobrej jakości, ale trzeba pamiętać, że były to korelacje liniowe.

3. W sytuacji, gdy liczba analizowanych sygnałów (metryk) byłaby bardzo duża, jak należałoby szukać korelacji pomiędzy tak znaczącą liczbą sygnałów (np. użyć analizy skupień)? Czy podejście zakładające redundancję analityczną i śledzenie różnicy pomiędzy rzeczywistą wartością wyjściową (Y) a wartością oszacowaną (Y') byłoby wystarczające? Czy system mógłby pracować w czasie rzeczywistym lub generując wyniki z niewielkim opóźnieniem?
4. W pracy wybrano podejście, w którym rozważono tylko macierz korelacji Pearsona, ale ona wskazuje tylko na liniowe zależności. Czy zbadano możliwość występowania innych (nieliniowych) korelacji? Co należałoby zdaniem autora rozprawy uczynić, gdyby takie korelacje wystąpiły? Jest to ważne biorąc pod uwagę, że we wstępie rozprawy czytamy, iż docelowym (cyt. str. 3) *w obszarze wspierania decyzji jest zastosowanie metod opartych na sztucznej inteligencji pozwalających na całkowite zautomatyzowanie procesu decyzyjnego.*

Wniosek końcowy

Po analizie przedłożonej rozprawy doktorskiej Pana mgra inż. Marcina Molędy stwierdzam, że wnosi ona cenny wkład w rozwój dyscypliny informatyka techniczna i telekomunikacja. Stanowi ona oryginalne rozwiązanie problemu naukowego, potwierdzając wiedzę, bardzo dobre przygotowanie merytoryczne, umiejętności praktyczne oraz umiejętność samodzielnego prowadzenia badań naukowych przez Doktoranta. Podjęta tematyka jest aktualna i ma bardzo duży potencjał aplikacyjny. Recenzowana rozprawa doktorska spełnia ustawowe wymagania stawiane kandydatom do stopnia doktora, określone ustawą o stopniach i tytułach naukowych – uwzględnione w art. 13 ust. 1 Ustawy z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki w związku z art. 179 Ustawy z dnia 3 lipca 2018 r. Przepisy wprowadzające ustawę - Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2018 r., poz. 1669). Na tej podstawie wnioskuję do Rady Dyscypliny Informatyka





**POLITECHNIKA
RZESZOWSKA**
im. IGNACEGO ŁUKASIEWICZA

al. Powstańców Warszawy 12, 35-959 Rzeszów
tel./fax: +48 17 854 12 60, tel.: +48 17 865 11 00
www.prz.edu.pl

Techniczna i Telekomunikacja Politechniki Śląskiej w Gliwicach o dopuszczenie jej do publicznej obrony.

Dominik Stronka



ZAKŁAD SYSTEMÓW ZŁOŻONYCH

Wydział Elektrotechniki i Informatyki
ul. Wincentego Pola 2, 35-959 Rzeszów, tel. 17 865 1340
zsz.prz.edu.pl