

Załącznik nr 2: Autoreferat

1. Dane osobowe

Imię i nazwisko: Piotr Bilski

Data urodzenia:

Miejsce urodzenia:

Miejsce zatrudnienia: Instytut Radioelektroniki, Politechnika Warszawska, ul. Nowowiejska 15/19, 00-665 Warszawa

Wydział Zastosowań Informatyki i Matematyki, Szkoła Główna Gospodarstwa Wiejskiego, ul. Nowoursynowska 159, 02-776 Warszawa

Adres email: pbiłski@ire.pw.edu.pl, piotr_bilski@sggw.pl

Telefon kontaktowy:

2. Edukacja i uzyskane dyplomy

1995-1999 – studia 1. stopnia (inżynierskie): Politechnika Warszawska, Wydział Elektroniki i Technik Informatycznych, specjalność: inżynieria komputerowa

1999-2001 – studia 2. stopnia (magisterskie): Politechnika Warszawska, Wydział Elektroniki i Technik Informatycznych, Instytut Radioelektroniki (z wyróżnieniem)

2001-2006 – studia doktoranckie: Politechnika Warszawska, Wydział Elektroniki i Technik Informatycznych, Instytut Radioelektroniki (z wyróżnieniem)

3. Przebieg pracy zawodowej

2005-2006 – asystent, Politechnika Warszawska, Instytut Radioelektroniki

2006-obecnie – adiunkt, SGGW w Warszawie, Wydział Zastosowań Informatyki i Matematyki

2008-obecnie – adiunkt, Politechnika Warszawska, Instytut Radioelektroniki

4. Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki:

- a. Jako osiągnięcie naukowe, uzyskane po otrzymaniu stopnia doktora, stanowiące znaczny wkład w rozwój dyscypliny naukowej wskazuję monografię w języku angielskim:

P. Bilski, "Artificial intelligence methods in the diagnostics of analog systems," Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa, 2013, ISSN 0137-2343 (178 stron).

- b. Omówienie celu naukowego ww. pracy i osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania

Główną tematyką badawczą poruszaną przeze mnie od momentu rozpoczęcia studiów doktoranckich do chwili obecnej była diagnostyka systemów analogowych. W szczególności skupiłem się na zastosowaniach w tej dziedzinie metod sztucznej inteligencji, którą to tematykę zgłębiałem pod kierownictwem prof. Jacka Wojciechowskiego. Podstawowym problemem, jakim zajmowałem się w ciągu ostatnich dziesięciu lat jest możliwie dokładne określenie stanu badanego obiektu na podstawie analizy informacji (symptomów) pozyskiwanych z jego odpowiedzi. Cele diagnostyki realizowane przeze mnie obejmowały:

- Maksymalizację skuteczności diagnostycznej w warunkach szumu addytywnego i z uwzględnieniem tolerancji parametrów systemu, mierzonej przy pomocy błędu próbki, tj. procentowego udziału niepoprawnych decyzji klasyfikacyjnych na zbiorze testującym

- Automatyczny dobór optymalnych wartości parametrów heurystycznej metody klasyfikacji lub regresji w celu minimalizacji błędu próbki.
- Dobór optymalnego zbioru symptomów mierzonych w sygnałach odpowiedzi generowanych przez system. Powinien być to zbiór minimalny, zapewniający jednak dostatecznie wysoką efektywność klasyfikacji lub regresji.
- Analiza danych trenujących w celu znalezienia grup niepewności, utrudniających rozróżnienie pomiędzy poszczególnymi stanami systemu analogowego.

Dla każdego z opisanych celów zaproponowałem metodę przetwarzającą dane i pozwalającą osiągnąć wyniki lepsze od dotychczas uzyskiwanych lub uzyskać dodatkową wiedzę na temat badanych obiektów.

W ramach pracy doktorskiej opracowałem metodę automatycznej diagnostyki wybranych systemów technicznych (takich jak filtry elektroniczne, serwomechanizmy, czy maszyny elektryczne) z wykorzystaniem logiki rozmytej. Główne zalety tej metody jako systemu ekspertowego to poprawna praca w warunkach niepewności pomiarowej, czytelna postać wiedzy (reguły) oraz zdolność do wykrywania wielokrotnych uszkodzeń parametrycznych. Mankamentem logiki rozmytej jest brak automatycznego algorytmu ekstrakcji wiedzy z danych trenujących. W pracy doktorskiej zaproponowano zastosowanie drzew decyzyjnych do tego celu. Następnie wykorzystałem strukturę drzewa do automatycznego zaprojektowania wszystkich elementów modułu diagnostycznego, tj. wejściowych i wyjściowych funkcji przynależności oraz reguł diagnostycznych. Zaprojektowany moduł został przetestowany na wybranych obiektach analogowych, pozwalając uzyskać satysfakcjonujące wyniki wykrywania i identyfikacji pojedynczych uszkodzeń parametrycznych. Eksperymenty wykazały również potencjalne możliwości wykorzystania architektury do analizy uszkodzeń wielokrotnych.

Po uzyskaniu stopnia doktora nauk technicznych kontynuowałem badania nad możliwościami wykorzystania innych metod sztucznej inteligencji w diagnostyce. Moje prace związane były z analizą dwóch grup algorytmów: opartych na regułach (ang. rule-based) oraz numerycznych (takich, jak sztuczne sieci neuronowe - ANN, czy maszyny wektorów podpierających - SVM). Wyniki badań były przedstawiane w szeregu publikacji konferencyjnych oraz artykułów w czasopismach, zebranych w załączniku nr 3 oraz w monografii przedstawionej jako osiągnięcie naukowe stanowiące znaczny wkład w rozwój dyscypliny naukowej. Poszczególne osiągnięcia związane z metodami opartymi na regułach obejmują:

Opracowanie metody automatycznej generacji logiki rozmytej alternatywnej do przedstawionej w pracy doktorskiej. W przeciwieństwie do tej ostatniej, opartej na konstrukcji drzewa decyzyjnego, zaproponowana metoda wykorzystuje informacje uzyskane dzięki zbiorom przybliżonym (ang. rough sets), które dokonują dyskretyzacji danych trenujących, a następnie pozyskują z nich reduktę, tj. minimalne zbiory atrybutów (symptomów) pozwalające rozróżnić wszystkie stany obiektu reprezentowane w zbiorze danych trenujących. Na tej podstawie opracowałem algorytm generacji reguł, wykorzystujący dyskretne numery przedziałów dla każdego symptomu oraz zredukowaną liczbę symptomów do generacji wszystkich elementów logiki rozmytej (analogicznie do podejścia wykorzystującego drzewa decyzyjne). Uzyskany algorytm ma podobne właściwości do przedstawionego w pracy doktorskiej, jednak lepiej nadaje się do wykrywania uszkodzeń wielokrotnych, ponieważ tworzone tutaj reguły są niezależne, podczas gdy zbudowane na podstawie drzewa decyzyjnego wykluczają się (mając wspólny przynajmniej jeden element przesłanek). Przeprowadzono szereg eksperymentów związanych z zastosowaniem różnych metod dyskretyzacji i redukcji. Metoda została zaprezentowana w punkcie 7.1.2 monografii, zaś charakterystyka logiki rozmytej oraz zbiorów przybliżonych w punkcie 6.3.

Opracowanie nowego algorytmu indukcji reguł, dopasowanego do ciągłych danych trenujących. Metoda powstała poprzez modyfikację algorytmu AQ, który w oryginalnej postaci przetwarza dane dyskretne. W celu poradzenia sobie z niepewnością pomiarową w danych (reprezentowaną przez szum addytywny oraz tolerancje parametrów składowych badanego systemu) progi rozdzielające poszczególne wartości symptomów zostały zastąpione przez wprowadzenie wymagania minimalnej różnicy pomiędzy przykładami. Dzięki temu udało się stworzyć szybką metodę diagnostyczną, w której postać wiedzy jest zrozumiała dla człowieka. Eksperymenty przeprowadzone na filtrze elektronicznym oraz silniku indukcyjnym pozwoliły stwierdzić, że jakość identyfikacji uszkodzeń jest wyższa, niż w przypadku drzewa decyzyjnego oraz wersji dyskretnej algorytmu indukcji reguł. Metoda została przedstawiona w punkcie 6.2.1 monografii.

Dokładne zbadanie przydatności zbiorów przybliżonych do diagnostyki systemów analogowych. Idea wykorzystania metody została zasugerowana w pracy doktorskiej. Na tej podstawie zaproponowałem schemat wykorzystania zbiorów przybliżonych w diagnostyce. Realizacja pomysłu wymagała przeprowadzenia testów w celu określenia skuteczności algorytmów dyskretyzacyjnych i redukcyjnych. Ponieważ zbiory przybliżone przetwarzają informacje dyskretne, ich skuteczność w największym stopniu jest zależna od zastosowanej metody generacji przedziałów dla każdego symptomu. W wyniku eksperymentów metoda równego podziału częstości (ang. equal frequency binning) została określona jako najbardziej przydatna. Wymaga ona jednak podania liczby przedziałów jako parametru wywołania. Najskuteczniejszą metodą redukcyjną jest algorytm genetyczny. Choć najbardziej czasochłonny, generuje maksymalną liczbę reduktów, zapewniającą najwyższą skuteczność diagnostyki podczas głosowania reguł. Eksperymenty zostały przedstawione w punkcie 8.3 monografii, zaś podstawy operacji wykorzystywanych w badaniach, tj. dyskretyzacji i zbiorów przybliżonych znajdują się, odpowiednio, w punktach 5.1 oraz 6.3.2.

Prace badawcze dotyczące zastosowania inteligentnych metod numerycznych w diagnostyce obejmowały:

Zastosowanie algorytmu optymalizacyjnego symulowanego wyżarzania do automatycznego doboru parametrów klasyfikatora zbudowanego na bazie maszyn wektorów podpierających. Metoda ta jest uważana za lepszą wersję sztucznych sieci neuronowych, optymalnie rozdzielając przykłady należące do różnych kategorii przy uwzględnieniu niepewności pomiarowej w danych. Do jej mankamentów należy konieczność doboru parametrów jądra przekształcającego oryginalną dziedzinę przykładów trenujących w nową, w której separacja między poszczególnymi kategoriami uszkodzeń jest łatwiejsza. Problem stanowi zarówno wybór najlepszej funkcji jądra, jak i jej parametrów. Ponieważ optymalna konfiguracja zależy od danych trenujących, jej poszukiwanie musi być przeprowadzane za każdym razem od nowa. Ze względu na ciągły charakter parametrów jądra, przeszukiwanie wyczerpujące może być przeprowadzane tylko na zdyskretyzowanym zakresie wartości. Jest ono również długotrwałe, co sugerowało zastosowanie heurystycznych metod optymalizacyjnych. Symulowane wyżarzanie zostało przetestowane w wersji jednorodnej i niejednorodnej. Określono parametry algorytmu (tj. początkową i końcową temperaturę oraz schemat schładzania) zapewniające najszybsze znalezienie optymalnych wartości. Skuteczność metody została sprawdzona dla układu elektronicznego oraz silnika prądu stałego. Implementację algorytmu optymalizacyjnego przedstawiono w punkcie 6.6 oraz 8.2 monografii.

Zastosowanie metody maszyn wektorów podpierających do identyfikacji parametrów silnika indukcyjnego. Zadanie to zostało sprowadzone do problemu regresji wielu zmiennych opisujących stan obiektu. Dzięki niej możliwa jest dokładniejsza analiza uszkodzeń, niż w przypadku dyskretnych

kategorii. Zaprojektowanie modułu wymaga znacznego wysiłku od projektanta. W przypadku opisywanego obiektu zidentyfikowano siedem parametrów, co wymusza proces uczenia tyluż modułów SVM. Zaproponowałem metodę regresji samodzielnej, w której wszystkie moduły działają jednocześnie, usiłując zidentyfikować wartości parametrów, za które są odpowiedzialne. Z kolei regresja połączona z uprzednią klasyfikacją wymaga użycia tylko jednego modułu w danej chwili, wyznaczanego przez klasyfikator, który wstępnie określa kategorię uszkodzenia. W opisywanym eksperymencie zastosowano w tym celu drzewo decyzyjne, które wybrano ze względu na szybkość działania i prostą implementację algorytmu uczącego. W obu przypadkach konieczne było uczenie modułów SVM w celu znalezienia najlepszej konfiguracji jądra, zapewniającej optymalne odwzorowanie wartości parametrów diagnozowanego systemu. Ponownie okazało się, że jest to proces długotrwały, w którym wykorzystano zarówno optymalizację heurystyczną, jak i przeszukiwanie wyczerpujące. Miarą dokładności identyfikacji parametrów był pomiar błędu średniokwadratowego (MSE) dla danych testujących. Metoda regresji samodzielnej wymagała modyfikacji standardowego MSE do postaci uwzględniającej dwa ważone składniki – błąd dla odwzorowania wartości nominalnych oraz dla wartości parametru uszkodzonego. Dobór wag był konieczny, ponieważ jądro zapewniające mały błąd dla wartości w stanie nominalnym prowadzi jednocześnie do dużego błędu odwzorowania wartości w stanie uszkodzonym i vice versa. Regresja samodzielna okazała się mniej skuteczna od regresji połączonej z klasyfikatorem, uzyskane dla niej wartości MSE były o dwa lub trzy rzędy wielkości większe. Dokładność tej drugiej zależy jednak od jakości klasyfikatora. Moduł SVM został scharakteryzowany w punkcie 6.4.2, hierarchiczne podejście do zadania diagnostycznego – w punkcie 7.4, zaś eksperymentalne wyniki - w punkcie 8.4 monografii.

Na potrzeby modułów klasyfikujących generujących wartości binarne jako kody uszkodzeń zaproponowałem nowy sposób kodowania kategorii. Istniejące kody należy analizować pod kątem liczby klasyfikatorów binarnych niezbędnych do rozróżnienia wszystkich kategorii, zdolności do wykrywania uszkodzeń wielokrotnych oraz minimalizacji prawdopodobieństwa popełnienia błędu (wskazania niewłaściwego uszkodzenia). Istniejące rozwiązania eliminują zwykle tylko część problemów, w większości przypadków odległość Hamminga pomiędzy dowolnymi dwiema kategoriami jest identyczna. Nie istniał dotąd kod, w którym odległości między kategoriami byłyby różne, w zależności od stopnia błędu popełnianego przez klasyfikator. Zaproponowałem rozwiązanie umożliwiające wykrywanie uszkodzeń wielokrotnych (w przeciwieństwie np. do kodowania „one-vs-all”) oraz rozróżniające wskazanie niewłaściwego parametru od wskazania niewłaściwego odchylenia od wartości nominalnej i zapewniające minimalne ryzyko wyboru niewłaściwej kategorii (jak w przypadku np. kodowania ECOC). Zaproponowany schemat kodowania nadaje się zarówno dla ANN, jak i SVM. Schemat kodowania został przedstawiony w punkcie 6.4.1 monografii.

Oprócz wymienionych eksperymentów zaproponowałem następujące badania w celu rozwiązania konkretnych problemów diagnostycznych:

Zaimplementowałem fuzję klasyfikatorów do analizy czujnika przyspieszenia (akcelerometru). Zauważyłem, że wiedza pozyskiwana przez poszczególne metody uczenia maszynowego ma inną postać dla każdego algorytmu. Wyniki identyfikacji nie są identyczne, bo każda metoda popełnia błędy na innych przykładach. Z tego powodu zastosowanie kilku podejść jednocześnie mogłoby zwiększyć skuteczność diagnostyczną. W ramach eksperymentu zaimplementowałem trzy metody różniące się strukturą przechowywanej wiedzy: sztuczne sieci neuronowe, drzewa decyzyjne oraz naiwny klasyfikator Bayesa. Zaproponowane metody charakteryzują się stosunkowo niewielką złożonością obliczeniową podczas podejmowania decyzji, co

umożliwia ich zastosowanie w warunkach czasu rzeczywistego na komputerze przemysłowym o ograniczonych zasobach obliczeniowych. Podjęcie decyzji przez moduł odbywało się w wyniku głosowania większościowego. W wyniku analizy udało się uzyskać lepsze wyniki, niż dla każdego klasyfikatora osobno. Badania planowane w przyszłości będą uwzględniać większą liczbę metod uruchamianych jednocześnie. Eksperyment został przedstawiony w punkcie 8.5 monografii, zaś poszczególne metody wchodzące w skład schematu diagnostycznego zostały opisane w punktach 6.2.2, 6.4.1 oraz 6.5.1.

Zastosowałem szereg metod uczenia bez nadzoru do wstępnej analizy danych ze zbioru uczącego w celu znalezienia grup niepewności (ang. ambiguity groups). Pomimo, iż przykłady (wyniki symulacji modeli obiektów) są etykietowane (tzn. opisane przez rzeczywistą kategorię uszkodzenia), wartości ich atrybutów mogą wskazywać na podobieństwo do przykładów należących do innych kategorii. Takie grupy są trudne do rozróżnienia i mogą powodować zmniejszenie skuteczności modułu diagnostycznego. Powinny one zostać wykryte, a następnie albo wyeliminowane ze zbioru, albo osobno przetworzone przez algorytm uczący. Do wykrywania grup niepewności zastosowałem metodę grupowania grafowego oraz sieć samoorganizującą (ang. Self-Organizing Map), co pozwoliło porównać uzyskane przez nie wyniki. Określiłem metody doboru parametrów algorytmów. W przypadku grupowania grafowego było to obliczenie progów rozróżniającego poszczególne podzbiory od siebie. W przeciwieństwie do wcześniej stosowanych metod, wykorzystujących odległość euklidesową, zastosowałem grafy cząstkowe, tworzone osobno dla każdego symptomu. Wierzchołki w grafie są tworzone przez przykłady, zaś krawędzie między nimi oznaczają podobieństwo. Dzięki temu tylko przykłady podobne do siebie „pod każdym względem” (tzn. takie, między którymi istnieje krawędź dla każdego podgrafu), były uznawane za należące do tej samej kategorii. Określenie wartości progów oddzielających poszczególne grupy określiłem na podstawie analizy histogramów. W przypadku samoorganizujących się map wykorzystano różne architektury sieci, odkrywając, że najprostsza ich postać, wykorzystująca strategię WTA (ang. Winner Takes All) jest w zupełności wystarczająca do przedstawianego zadania. Dodatkowo wykorzystałem szereg miar odległości (Euklidesa, Manhattan, Czebyszewa itp.), uzyskując informacje na temat różnic w grupowaniu między nimi. Metodologia została przedstawiona w punkcie 6.7 i 7.2, zaś wyniki eksperymentów w punkcie 8.1 monografii.

W efekcie przeprowadzonych badań przetwarzania udało mi się znaleźć grupy przykładów trudne do rozróżnienia. Wiedza pozyskana tą metodą została wykorzystana do następujących celów:

- Ocena stopnia trudności poszczególnych systemów analogowych, reprezentowanych przez zbiory trenujące i testujące. Określana jest ona przez procentowy udział w nich przykładów o trudno rozróżnialnych kategoriach. Miara taka pozwala przewidywać jakość diagnostyki realizowanej przez wybraną metodę inteligentną.
- Określenie metody porównania skuteczności heurystycznych metod diagnostycznych. Dzięki wiedzy na temat stopnia trudności zbiorów uczącego i testującego można ustalić, która metoda lepiej nadaje się do analizy wybranego obiektu analogowego.
- Oznaczenie przykładów wchodzących w skład grup niepewności jako „trudnych”, co pozwala na modyfikację metody uczącej. Wiedzę taką wykorzystano do modyfikacji algorytmu generującego logikę rozmytą (wykorzystanego w pracy doktorskiej) i poprawienie jego skuteczności.

Jednym z moich ostatnich osiągnięć jest **zastosowanie metod selekcji atrybutów ze zbioru trenującego do przetworzenia wstępnego zbiorów, które następnie są analizowane przez metodę**

uczającą. Celem eksperymentu było określenie minimalnej liczby symptomów koniecznych do rozróżnienia wszystkich stanów badanego systemu. Pozwoliłoby to uprościć przechowywaną wiedzę oraz przyspieszyć proces uczenia, a być może również poprawić efektywność diagnostyczną. W przeciwieństwie do innych metod stosowanych dotychczas, takich jak PCA, zastosowana metoda minimalizuje liczbę węzłów pomiarowych, z których zbierane są symptomy. Dzięki temu możliwe jest przyspieszenie pracy systemu diagnostycznego oraz zmniejszenie kosztów całego procesu. Do eksperymentów wybrano dwie metody statystyczne, tj. Hellwiga oraz Współczynników Korelacji Wielorakiej (WKW), stosowane dotychczas w analizach ekonometrycznych. Obie metody połączono z testem quasi-stacjonarności w celu odrzucenia symptomów, których wartości praktycznie nie ulegają zmianie (nie mają znaczenia dla procesu diagnostycznego). W wyniku badań przeprowadzonych na filtrze elektronicznym piątego rzędu udało się zminimalizować liczbę atrybutów z 54 do 12 dla metody Hellwiga i 18 dla WKW. Weryfikację metod przeprowadzono wykorzystując ANN w zadaniu klasyfikacyjnym. Ich użycie wymagało doboru optymalnej struktury (tj. liczby warstw i neuronów w każdej warstwie), a także metody uczenia. Uzyskane wyniki pozwoliły stwierdzić, iż zmniejszenie liczby symptomów pozwala na utrzymanie skuteczności diagnostycznej na takim samym poziomie, jak w przypadku danych oryginalnych, często wymagając do tego celu prostszej struktury sieci. Nie udało się natomiast poprawić skuteczności diagnostyki. Metody wstępnego przetwarzania zbiorów danych zostały przedstawione w punkcie 5.2 monografii, zaś wyniki eksperymentalne w punkcie 8.1 monografii.

Ponieważ opisywane metody są przeznaczone do pracy w warunkach rzeczywistych, ich implementacja na komputerze przemysłowym wymaga analizy zapotrzebowania odnośnie mocy obliczeniowej procesora i zasobów pamięci. Dokonano w tym celu analizy czasowej typowych elementów składowych aplikacji pomiarowo-obliczeniowej w środowisku LabVIEW, które jest obecnie standardem w informatyce przemysłowej. Opracowano metodologię wykonywania pomiarów czasu trwania poszczególnych funkcji pod kontrolą systemu operacyjnego czasu rzeczywistego. Uzyskano wiedzę na temat czasochłonności najważniejszych operacji przetwarzania danych. Analiza została przedstawiona w punkcie 7.5 monografii.

Metody stosowane w diagnostyce systemów analogowych zostały również przeze mnie wykorzystane w dziedzinie pokrewnej, tj. identyfikacji i analizie geotechnicznej gruntów wykonywanej na potrzeby budownictwa. Zadanie polegało na określeniu kategorii gruntów na podstawie pomiarów z sond geotechnicznych, takich jak CPT lub DMT. Ponieważ dokładna informacja na temat struktury gruntu jest pozyskiwana w wyniku wierceń i długotrwałych pomiarów w laboratorium, zmniejszenie ich liczby lub całkowita eliminacja mogłyby przyczynić się do zmniejszenia kosztów analizy i jej znaczącego przyspieszenia. Tradycyjnie identyfikacja kategorii gruntów na podstawie pomiarów z sond wymaga użycia nomogramów, czyli wykresów wskazujących kategorię występującą na przecięciu wartości zmierzonych parametrów. Wykresy te nie są jednak dokładne i nie mogą być stosowane w tym samym stopniu na całym świecie (szczególnie w Polsce). Z tego powodu zaproponowałem następujący schemat:

- Pogrupowanie przykładów ze zbioru uczącego według ich podobieństwa poprzez zastosowanie wybranych metod analizy skupień.
- Odkrycie w pogrupowanych danych zależności, których nie da się zauważyć podczas analizy próbek ziemi pozyskanych w wyniku odwiertów oraz obliczenie na tej podstawie tzw. indeksów, czyli parametrów geotechnicznych gruntu.
- Wykorzystanie wiedzy pozyskanej z grupowania do stworzenia reguł klasyfikacyjnych, które można wykorzystać do analizy danych pochodzących z innych lokalizacji.

Do realizacji pierwszego i drugiego zadania wykorzystałem metody uczenia bez nadzoru: grupowanie grafowe oraz mapy samoorganizujące się, wcześniej zaimplementowane do wykrywania grup niepewności. Dzięki nim udało się stworzyć profile geotechniczne, które następnie porównałem w celu ustalenia, w jakim stopniu wyniki uzyskiwane przez obie metody są podobne. Określiłem również wartości parametrów wymaganych przez metody analizy skupień, analogicznie jak w przypadku zadania poszukiwania grup niepewności. Uzyskane wyniki pozwoliły zaobserwować zależności w danych, których nie było widać wcześniej. Dzięki temu poprawiłem dokładność analizy i zautomatyzowałem część procesu decyzyjnego, tworząc moduł wspomagający inżyniera geotechnika.

Realizacja zadania trzeciego wymagała zastosowania metod uczenia z nadzorem. Zaproponowałem w tym celu algorytm generacji reguł, który wydobyl wiedzę z tym razem już etykietowanych danych. Została ona wykorzystana do analizy pomiarów pochodzących z innych lokalizacji. Udało się dzięki temu określić uniwersalność pozyskanej tą metodą wiedzy. Okazało się, że ma ona raczej charakter lokalny, ponieważ te same kategorie gruntów w różnych miejscach reagują inaczej na sondy, w związku z czym są widziane przez metody grupowania i podejmowania decyzji jako inne kategorie. Ponadto metoda jest wrażliwa na ilość danych dostępnych w procesie uczenia. Zwiększenie liczby zbiorów pozwala wydatnie poprawić jej dokładność i zmniejszyć liczbę nierozpoznanych przykładów.

Prace prowadzone nad analizą danych geotechnicznych są nadal przeze mnie rozwijane. Na ich kontynuację uzyskałem grant NCN nr 2011/03/D/ST8/04309 (wymieniony w punkcie 6 załącznika nr 4), którego jestem kierownikiem.

Część metod stosowanych w diagnostyce wykorzystałem również do analizy kursów akcji pochodzących z giełdy papierów wartościowych. Zaproponowałem zastosowanie transformaty falkowej do odsumowania szeregów czasowych, co ułatwia zadanie wyszukiwania trendów. Wykorzystałem także algorytm grupowania grafowego do identyfikacji podobnych wydarzeń na giełdzie. Wyniki przedstawiłem w dwóch artykułach w czasopiśmie Polish Journal of Environmental Studies (których kserokopie załączone są do niniejszej dokumentacji).

c. Omówienie pozostałych osiągnięć naukowo-badawczych

Od momentu rozpoczęcia studiów doktoranckich równoległe z pracami nad rozwojem metod diagnostyki technicznej prowadziłem, we współpracy z prof. Wiesławem Winieckim, badania nad analizą i projektowaniem komputerowych systemów pomiarowo-sterujących. Na pierwszym etapie moich badań zajmowałem się metodami oceny efektywności przyrządów wirtualnych (takich jak analizator widma, zaprojektowany przeze mnie w ramach pracy magisterskiej). Tematyka ta obejmowała w szczególności:

- Opracowanie metodyki pomiaru efektywności czasowej aplikacji pomiarowych działających w systemie operacyjnym ogólnego przeznaczenia (takim, jak Microsoft Windows). Określiłem wymagania konfiguracji sprzętowo-programowej w celu uzyskania trybu miękkiego czasu rzeczywistego (ang. Soft Real Time). W ramach tego zadania zbadałem właściwości bufora cyklicznego wykorzystywanego jako główna struktura danych w aplikacji pomiarowej komunikującej się ze światem zewnętrznym. Badania wykazały jednoznacznie niemożność uzyskania trybu twardego czasu rzeczywistego w takim systemie. Wyniki badań zostały przeze mnie opublikowane w czasopiśmie IEEE Transactions on Instrumentation and Measurement.
- Opracowanie metodyki projektowania przyrządów wirtualnych pracujących pod kontrolą systemu czasu rzeczywistego. Badania obejmowały zarówno analizę kodu uruchamianego

pod kontrolą systemu operacyjnego, jak i komunikację przy użyciu deterministycznej sieci komputerowej. Rezultatem badań było określenie możliwości przetwarzania oraz przesyłania danych o określonym rozmiarze.

W ciągu ostatnich czterech lat uczestniczyłem w pracach zespołu projektującego algorytmy kryptograficzne na potrzeby bezpiecznych rozproszonych systemów pomiarowych. Moje główne osiągnięcia na tym polu obejmują:

- Opracowanie równoległej realizacji systemu kryptograficznego AES w środowisku LabVIEW, pracującej pod kontrolą systemu czasu rzeczywistego. Projekt był istotny ze względu na konieczność zabezpieczenia danych transmitowanych i odbieranych z serwera pomiarowego. Stworzony system wykorzystuje wielowątkowość środowiska LabVIEW, która umożliwia wykonywanie wielu fragmentów kodu jednocześnie na procesorze wielordzeniowym. Dokonano pomiarów efektywności zaimplementowanego algorytmu zarówno na poziomie zrównoleglenia podstawowych operacji AES, jak i szyfrowania oraz deszyfrowania całych bloków z wykorzystaniem trybów CDC, EBC i innych. Uzyskałem dokładne informacje na temat efektywności mojego rozwiązania, którą porównałem z innymi implementacjami, wykorzystujących np. matryce FPGA. Wyniki badań zostały opublikowane w czasopiśmie Measurement.
- Opracowanie realizacji systemu kryptograficznego AES w środowisku LabVIEW przeznaczonego na modułowy komputer przemysłowy wyposażony w matrycę FPGA. Jest to druga część projektu związanego z zabezpieczeniem rozproszonego systemu pomiarowego przed atakiem w celu wykradzenia danych lub przejęcia kontroli nad systemem. Tym razem chodziło o zabezpieczenie drugiego węzła komunikacji między serwerem i klientem, tj. tego drugiego. Jest on zwykle realizowany w postaci niewielkiego komputera przemysłowego lub systemu wbudowanego o niewielkich możliwościach obliczeniowych. Z tego względu konieczna była optymalizacja kodu. System AES został zaprojektowany w dwóch wersjach: na procesor pracujący pod kontrolą systemu operacyjnego czasu rzeczywistego oraz na matrycę FPGA. Dzięki temu możliwe było porównanie efektywności obu rozwiązań oraz optymalne połączenie operacji szyfrowania z akwizycją danych przez moduł pomiarowy. Wyniki badań zostały opublikowane w materiałach konferencyjnych konferencji I2MTC.
- Opracowanie koncepcji zastosowania do rozproszonego systemu pomiarowego kryptografii kwantowej opartej na położeniu. Tematem analizy było zastosowanie kwantowego kanału informacyjnego do przesłania klucza potrzebnego do zabezpieczenia systemu opartego na położeniu. Obecnie rozważany jest ten schemat na ogólnym poziomie teoretycznym, ja zaś opracowałem koncepcję jego zastosowania do systemu pomiarowego z uwzględnieniem aktualnych możliwości technicznych. Przeprowadzono studium wykonalności systemu z uwzględnieniem zastosowania laserów jako źródeł kubitów oraz urządzeń do ich rejestracji (fotodiod). Wyniki analizy zostały opublikowane w czasopiśmie Measurement.

Robi Gilbi