

mpi. RDITT
14.10.2022
M. J. 67

Kraków 30.09.2022

prof. dr hab. inż. Grzegorz J. Nalepa
Instytut Informatyki Stosowanej
Wydział Fizyki, Astronomii i Informatyki Stosowanej
Uniwersytet Jagielloński

Tytuł rozprawy: Metody analizy efektywności i jakości procesów produkcji dyskretniej wykorzystujące wzorce technologiczne i wybrane mechanizmy eksploracji danych

Autor rozprawy: Marek Drewniak

Promotorzy rozprawy: dr hab. inż. Rafał Cupek, prof. Pol. Śl.

Dziedzina: nauki techniczne

Dyscyplina: informatyka techniczna i telekomunikacja

1 Wstęp

Rozprawa doktorska mgra Marka Drewniaka wpisuje się w ważny dla informatyki technicznej obszar zastosowań w przemyśle. W ramach kolejnej, mającej obecnie miejsce, rewolucji przemysłowej nazywanej Przemysłem 4.0 (ang. *Industry 4.0*, *I4.0*), określanej również po angielski *Smart Industry*, obserwujemy przemianę procesów i instalacji przemysłowych dzięki szeroko rozumianej cyfryzacji i zaawansowanym technologiom informatycznym. Rozwiązania te to m.in. nowoczesna sensoryka i powiązane z nią rozproszone przetwarzanie danych w paradygmacie *Internte of Things*, robotyka i systemy cyberfizyczne, szerokopasmowa łączność bezprzewodowa, czy chmurowe przetwarzanie danych. Jednym z wyzwań powiązanych z użyciem tych technologii są bardzo duże, szybkozmienne i niezwykle zróżnicowane wolumeny danych, co z kolei łączy I4.0 z tzw. rewolucją *Big Data*. Analiza danych tego rodzaju wymaga z kolei zaawansowanych metod z obszaru sztucznej inteligencji (ang. *Artificial Intelligence*, *AI*), a szczególnie duże zapotrzebowanie jest na użycie i rozwijanie dedykowanych metod eksploracji danych (ang. *Data Mining*, *DM*), czerpiących z postępów w obszarze uczenia maszynowego (ang. *Machine Learning*, *ML*), gdzie z kolei opracowuje się algorytmy stanowiące trzon DM. Dzięki tym narzędziom w ramach I4.0 buduje się rozwiązania w obszarach takich jak monitorowanie zużycia zasobów (w tym maszyn i urządzeń), zarządzanie jakością, analiza łańcuchów dostaw, monitorowanie produktów, czy w końcu podnoszenia bezpieczeństwa pracy w złożonych środowiskach przemysłowych. Zakłady przemysłowe czerpiące z tych rozwiązań nazywa się niejednokrotnie inteligentnymi fabrykami (ang. *smart factories*).

Praca Doktoranta dotyczy w szczególności modelowania systemów przemysłowych i instalacji produkcyjnych, oraz poszerzania wiedzy o procesach przemysłowych w oparciu i techniki informatyczne i analityczne – wykorzystujące rozwiązania DM.

Dalsza część recenzji odpowiada strukturalnie wytycznym przekazanych przez Radę Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Śląskiej.

2 Cel, zakres, teza i charakter rozprawy

Badania Doktoranta obejmują swoim zakresem zastosowanie i poszerzenie wybranych metod DM w praktyce przemysłowej. Zidentyfikowane we wstępie obszary szczegółowe to przede wszystkim powiązanie danych dostarczanych przez monitorujące pracę instalacji kontrolery przemysłowe z wiedzą ekspercką związaną z wiedzą o procesach przemysłowych, za pomocą współczesnych technik analitycznych DM. Ponadto Autora interesuje automatyczne wykrywanie wariantów produktów w procesie produkcji wielowariantowej i krótkoseryjnej za pomocą analizy danych pod kątem wykrycia specyficznych wzorców technologicznych. Ponadto, trzecim obszarem zainteresowań i badań było wykrywanie anomalii w pracy instalacji pod kątem oceny jej efektywności pracy.

We wstępie do pracy Doktorant definiuje ogólny cel badań (s. 15) jako: „opracowanie zestawu metod do analizy danych produkcyjnych, których rezultaty mogą zostać wykorzystane do poprawy efektywności i jakości pracy stanowisk produkcyjnych w wielowariantowej produkcji dyskretniej.” Ponadto wskazuje użycie metod klasteryzacji i statystyki tak aby uwzględnić automatyczne rozpoznawanie wytwarzanych wariantów produktów. Podstawowe założenia dotyczące wyników badań są wskazane w tezach i ograniczeniach. W p. 1.3 rozprawy zdefiniowany jest główna teza rozprawy:

Zastosowanie modeli informacyjnych do powiązania danych procesowych z wiedzą technologiczną i wykorzystanie mechanizmów eksploracji danych pozwala na istotną poprawę efektywności pracy i utrzymania dostępności instalacji przemysłowych.

Mając na uwadze specyficzne osadzenie pracy w praktyce przemysłowej i płynące z tego jej uwarunkowania, sformułowane są poniższe tezy pomocnicze:

1. Agregacja danych wyliczanych w obrębie cykli produkcyjnych pozwala na zredukowanie złożoności obliczeniowej i skrócenie czasu realizacji algorytmów wyszukiwania wzorców w produkcji dyskretniej wielowariantowej.
2. Możliwe jest automatyczne wykrywanie i klasyfikacja profili produkcyjnych instalacji bez znajomości technologii a priori (tj. bez wiedzy na temat stosowanych urządzeń i specyfiki ich pracy, bez znajomości procesu produkcyjnego oraz bez wiedzy na temat aktualnie produkowanego typu produktu), jedynie z wykorzystaniem informacji o zużywanych energiach, czasach trwania i sygnałach binarnych definiujących aktywność urządzeń wykonawczych.
3. Zaproponowane podejście do badania efektywności i jakości procesów produkcyjnych, bazujące na agregacji danych i automatycznym tworzeniu profili umożliwia wykrywanie anomalii w pracy instalacji przemysłowych w produkcji dyskretniej.

Ponadto Autor wskazuje ograniczenia technologiczne związane z rozpatrywanymi instalacjami, dotyczące:

- instalacji dla produkcji dyskretniej z wykorzystaniem rozwiązań typu PLC i SCADA,
- możliwość pomiaru całkowitej zużywanej energii w jednym punkcie,
- dostępność danych procesowych poprzez właściwe interfejsy instalacji.

Następnie Doktorant wskazuje szczegółowe zagadnienia i kroki badawcze, które bliżej opisują w kolejnym punkcie recenzji.

Mając na uwadze cel, tezę pracy, oraz przyjętą metodologię pracy, w której ewaluacja ma charakter praktyczny, sama rozprawa ma charakter praktyczno-technologiczny. W wyżej wymienionym we wstępie do recenzji kontekście, rozprawa p. Drewniaka dobrze lokuje się w bieżącej tematyce związanej z zastosowaniami informatyki w przemyśle, ze szczególnym uwzględnieniem metod AI w tym DM.

3 Zawartość rozprawy

Rozprawa składa się z siedmiu ponumerowanych rozdziałów, z których pierwszy stanowi wstęp, gdzie Autor zawarł genezę, motywację, tezę i cele pracy, przedostatni zawiera podsumowanie, a ostatni jest wykazem bibliografii. Pracy towarzyszy krótki załącznik dotyczący eksperymentalnej instalacji o nazwie DC40. Rozprawa ma objętość 158 stron, z czego bibliografia obejmuje 127 pozycji bibliograficznych na 14 stronach.

Rozdział 2., o objętości 56 str., dotyczy opisu wybranych wyzwań dziedzinowych z obszaru I4.0 oraz rozwiązań z obszaru DM. Autor opisuje systemy realizacji produkcji, wspierane informatycznie obiektowe modelowanie usług i danych produkcyjnych, a także zagadnienia komunikacji w przemysłowych systemach informatycznych. Osobny podrozdział dotyczy metod analizy skupień w DM, ze szczególnym uwzględnieniem klasteryzacji metodą k-Means. Następnie omawiane są obszary w których Autor prezentuje własne wyniki, tj. detekcja wariantów i profili produkcyjnych, klasteryzacja danych pochodzących z przemysłowych systemach informatyki, a także wykrywanie anomalii produkcyjnych z wykorzystaniem technik DM.

Rozdział 3., o objętości 14 str., dotyczy badań Doktoranta nad modelami informacyjnymi do wsparcia utrzymania instalacji przemysłowych. Omówiony jest model technologii instalacji produkcyjnej w języku Automation ML na przykładzie instalacji DC40. Następnie zaproponowany jest opis wzorców technologicznych w postaci wektora cech charakterystycznego dla danego wariantu produkcyjnego. Autor opisuje też zagadnienia związane z integracją danych procesowych z wiedzą dziedzinową tj. technologiczną za pomocą przestrzeni adresowej interfejsu OPC UA.

Rozdział 4., o objętości 32 str., opisuje badania związane z dwoma problemami wskazanymi we wstępie: automatycznym wykrywania profili produkcyjnych, oraz problemu selekcji i agregacji danych do ich analizy. Dyskusja rozpoczyna się od zaprezentowania schematu 11-sto etapowej (etapy oznaczono literami a-k) metody wykrywania wzorców technologicznych która ilustrowana jest przykładami. Następnie Autor przechodzi do detekcji wzorców technologicznych za pomocą klasteryzacji k-Means, m. in. na s. 106 zaproponowany jest model reprezentacji danych na potrzeby detekcji wzorców. Automatyzacja wyznaczania liczby klastrów jest tematem ostatniego podrozdziału.

Rozdział 5., o objętości str. 12, jest poświęcony metodzie automatycznego wykrywania anomalii w pracy urządzeń i dotyczy badań powiązanych z dwoma problemami: wykorzystania mechanizmów DM do poprawy jakości i efektywności procesów produkcyjnych w produkcji dyskretniej oraz zastosowania metod analitycznych bazujących na danych zagregowanych do wykrywania anomalii. Doktorant najpierw proponuje podejście do powiązania obserwacji wartości parametrów instalacji do wzorców technologicznych, na którym oparte jest omówione dalej podejście do detekcji anomalii w pracy instalacji.

Rozdział 6., o objętości 5 str., zawiera syntetyczne zakończenie i podsumowanie pracy, w tym nawiązanie do sformułowanych we wstępie tez pracy. Rozdział 7. zawiera wykaz pozycji bibliograficznych.

Pracy towarzyszy załącznik o objętości 3 stron opisujący skonstruowane stanowisko DC40. Stanowisko jest pokazane na zdjęciu, opisany jest składający się z 4 liczb rekord danych i opis słowa wyjściowego do sterowania urządzeniem składający się z 16 liczb.

4 Ocena rozprawy

Ocenę pracy warto rozpocząć od komentarza dotyczącego związku prezentowanych w niej badań do obecnie prowadzonych na świecie prac z obszaru zastosowań informatyki w przemyśle. Nie mam wątpliwości, że prace Doktoranta są aktualne i dobrze ułożone tematycznie. Rozwiązania technologiczne dla inteligentnych fabryk i instalacji przemysłowych szczególnie często korzystają obecnie z narzędzi DM. W tym kontekście, teza rozprawy (wraz z tezami pomocniczymi) jest sformułowana poprawnie, a koncepcje i podejścia zaproponowane w pracy są w mojej opinii ciekawe i wg mojej wiedzy oryginalne.

Obok podjętej tematyki, kolejną zaletą pracy jest przejrzystość języka i przemyślany układ tematyczny. Ponadto, na podstawie lektury pracy można zbudować przekonanie, że Doktorant ma dużą wiedzę praktyczną i swobodę posługiwania się technologiami przemysłowymi, które wykorzystuje w badaniach. Wraz z Promotorem był zaangażowany w liczne prace badawcze i był współautorem szeregu publikacji o których pisze w dalszej recenzji.

Natomiast co jako pierwsze zwraca uwagę w zawartości pracy to rozłożenie akcentów w tekście pracy. Zasadnicza treść pracy liczy 136 stron, z czego wstęp i zakończenie mają 12 stron, rozdział dotyczący przeglądu technologii, wybranych metod i literatury z zakresu DM liczy 56 stron, a sam opis wyników badań stron 68, którym towarzyszy załącznik o objętości 3 stron. Rzecz jasna same parametry ilościowe nie przesadzają o jakości wyników i nie ma w tym zakresie uznanych standardów, ale jak wskazuję poniżej, sam poziom szczegółowości opisu wyników jest niejednokrotnie zaskakująco skromny.

Przeeglądowy rozdział 2. jest niewątpliwie ciekawy i obszerny. Opisywane standardy technologiczne są później podstawą do opisu rozwiązań Autora. Jeśli chodzi o przegląd metod i wyników z obszaru DM, to jest on w mojej opinii przeprowadzony subiektywnie. Autor na str. 46 wspomina o kilku technikach DM, by potem płynnie przejść do stanowiska, że techniki klasteryzacji są tymi właśnie które mogą posłużyć do zidentyfikowanych przez Niego problemów. Trudno znaleźć ku temu przekonujące argumenty. W mojej opinii i wg mojej wiedzy, adekwatny dobór technik DM do konkretnych problemów bardzo silnie zależy od charakteru i ilości dostępnych danych (w tym danych etykietowanych), oraz uzupełniającej wiedzy dziedzinowej. Można odnieść wrażenie, że Autor *a priori* przyjął, że klasteryzacja jest tym zadaniem uczenia maszynowego do którego da się sprowadzić rozwiązywanie dostrzeżonych przez Niego problemów analitycznych. W pracy trudno mi było odnaleźć wskazanie klas systemów przemysłowych (i powiązanych z nimi typami i wolumenami danych) do których można by było stosować proponowane w pracy podejścia. Tym bardziej trudno być przekonanym, że tylko generyczny algorytm k-Means będzie adekwatny do proponowanych dalej w pracy analiz.

O wartości pracy przede wszystkim świadczą uzyskane przez Doktoranta wyniki opisane w rozdziałach 3, 4 i 5. Przedstawione podejścia uważam za ciekawe i potencjalnie użyteczne. Natomiast sposób ich opisu i poziom szczegółowości budzi wątpliwości.

W rozdziale 3. jest mowa o nowoczesnym modelu informacyjnym który może pełnić rolę cyfrowego bliźniaka. W podrozdziale 4.1 zaproponowane jest podejście do zamodelowania struktury instalacji w opartym o XML język Automation ML. W dalszej części rozdziału jest pokazane jak można ten opis strukturalny powiązać z opisem parametrów instalacji tym wartości mierzonych przez sensory i sygnałami sterującymi. Zaproponowane jest też podejście do zamodelowania wzorców technologicznych. Niestety model ten nie uwzględnia aspektu dynamicznego pracy instalacji, co za tym idzie jako użycie jako cyfrowego bliźniaka, np., do symulacji wydaje się być nie uprawnione. Cała dyskusja prowadzona jest wyłącznie w oparciu o jeden konkretny przykład instalacji opracowanej specjalnie na potrzeby badań – DC40. Ponadto modele nigdzie nie są pokazane w całości, Autor omawia jedynie ich fragmenty. Pojawia się zatem pytanie, czy to co jest prezentowane to powtarzalna metoda, która może znaleźć zastosowanie do jakiejś klasy przypadków, czy jedynie ogólna koncepcja zilustrowana wybiórczymi przykładami.

W rozdziale 4. pokazany jest schemat metody wykrywania wzorców, w postaci ilustracji która ma wprowadzić wartość poglądową wraz z towarzyszącym jej tekstem, ale trudno zrozumieć ją jako sformalizowaną procedurę. Brak też informacji o jej ewentualnej implementacji. Sama metoda wydaje się poprawna. Znowu brak jest informacji o jej użyciu na innym przykładzie niż DC40. Dalsza część dyskusji w tym rozdziale jest nadal prowadzona wyłącznie w oparciu o przykłady i fragmenty. Autor mówi o przeprowadzanych eksperymentach i występujących w nich zbiorach danych. Trudno ocenić ich wartość w ewaluacji metody, bo nie są one w pełni opisane. Poziom precyzji jest czasem zaskakujący: „spośród około 130 zarejestrowanych cykli produkcyjnych” (str. 100). Inny przykład: na str. 111 można znaleźć tabelę prezentującą wyniki klasteryzacji. Można się domyślić, że nazwy cech są zbieżne z opisem na str. 158, ale już interpretacja konkretnych wartości liczbowych nie jest możliwa.

W podrozdziale 4.3 omówiona jest koncepcja metody automatyzującej wyznaczanie liczby klastrów przy użyciu algorytmu k-Means. Jest to sam w sobie ciekawy wynik, jednak brak: a) formalizacji metody w postaci algorytmu, b) jej ewaluacji na czymś więcej niż przykładowych danych dot. DC40, c) porównania do innych metod wyznaczania parametru k. Omówienie jej użycia ogranicza się do wzorców technologicznych w kontekście DC40.

W rozdziale 5 jest mowa o detekcji anomalii, co jest niezwykle szerokim tematem badawczym w obszarze analityki danych przemysłowych. Podobnie jak poprzednio, to co opisuje Doktorant to bardziej pomysł niż metoda, przy braku elementarnej formalizacji opisu i jej użycia na czymś więcej niż jeden eksperyment z DC40. Wspomniane schematy pracy nie są w pełni scharakteryzowane. Proponowane podejście statystyczne bazujące na informacji o klasteryzacji pozwala na wykrywanie anomalii w konkretnym opisywanym przypadku, ale jego skuteczność dla innego typu instalacji nie jest pokazana. Nie ma również porównania do innych metod wykrywania anomalii. Ponadto brak jest systematycznej typologii anomalii które mogą wystąpić w opisywanej instalacji.

W pracy brak jest systematycznej, przejrzystej ewaluacji otrzymanych wyników. Koncepcje metod są jedynie ilustrowane wykresami i tabelami pokazującymi wybrane fragmenty pojedynczych eksperymentów dla jednego dedykowanego stanowiska DC40. Trudno przesądzić jaka jest praktyczna wartość otrzymanych wyników. We wstępie Autor wskazuje na pewne ogólne wytyczne do jakiego typu instalacji mogą być stosowane jego metody, ale trudno ku temu znaleźć w pełni przekonującą argumentację, czy potwierdzającą ją ewaluację eksperymentalną. W obszarze I4.0 dla różnych problemów (np. detekcja anomalii, cyfrowe bliźniaki) można znaleźć publiczne zbiory danych stanowiące punkt odniesienia i wyjścia do porównania, jednak Autor z nich nie korzysta.

Przyjęta w pracy metodologia wydaje się nieco arbitralna. Zastanawia na przykład założenie, dotyczące pracy tylko z surowymi (jak można rozumieć z opisu) danymi z instalacji. W praktyce przemysłowej bardzo często do analizy pracy urządzeń korzysta się danych przetworzonych do postaci dzienników zdarzeń (ang. *event logs*). Autor nie wspomina o takim podejściu, pomimo że również pozwala na uproszenie metod analitycznych i ograniczenie ilości danych.

Ostatnia wątpliwość dotyczy nie tylko samego oparcia *gros* argumentacji na dedykowanym stanowisku DC40, ale jego niezwykle skąpy opis w załączniku. Brak jest precyzyjnego schematu, pomimo, że struktura stanowiska jest potem modelowana. Brak jest szczegółowego opisu komponentów i danych których są źródłem. W jakim sensie i zakresie DC40 może uogólniać instalacje przemysłowe i jakiego typu?

Praca jest zredagowana ogólnie poprawnie. Szczegółowe usterki wskazuję w dalszej części recenzji.

Pomimo wskazanych powyżej niedostatków pracy, uważam przedstawione koncepcje za oryginalne i ciekawe, a moja końcowa ocena rozprawy jest pozytywna.

5 Uwagi dyskusyjne i słabe strony rozprawy

W nawiązaniu do przedstawionej powyżej oceny rozprawy, za najważniejsze słabe strony rozprawy należy w mojej opinii uznać:

- a) brak systematycznej ewaluacji zarówno w postaci większej liczby eksperymentów, jak i odniesienia się do innych zbiorów danych przemysłowych, czy innych instalacji, czy w końcu porównania do innych metod.
- b) brak elementarnej usystematyzowania opisywanych koncepcji metod w postaci procedur czy algorytmów.
- c) brak informacji o implementacji omawianych koncepcji (jest nie rozwinięta zmianka o obliczeniach w Matlabie).
- d) niewystarczająco dokładny opis stanowiska DC40 stanowiącego jedyną podstawę ewaluacji.

Poniżej formułuję szereg uwag technicznych i szczegółowych.

1. s. 5: „dane zbierane” – kolokwializm,
2. s. 14: podwójna „,” w tezie 2.,
3. s. 70: rys 2.7.1 – skąd pochodzi sam podział i ilustracja?,
4. s. 82: jest mowa o wektorze V_{MC} – o jaki wektor chodzi? Dalej, na str. 107 będzie mowa o wektorze V_{MCn} ,
5. s. 82: literówka „odpowiadających”,
6. s. 95: literówka na rysunku: System rejestracji dancyh,
7. s. 106,7: numery nie są unikalne,
8. s. 107,8: formuła nie ma numeru,
9. s. 109: przykładowy zbiór danych jest tylko wizualizowany po przekształceniu ale nie jest opisany,
10. s. 109: rysunek nie ma legendy,
11. s. 148: błąd składu a pozycjach 51,52 ,
12. s. 149: jw. w pozycji 77,
13. s. 150: jw. w pozycjach 81, 84,
14. s. 153: jw. w pozycji 107.

Powyższe słabe strony rozprawy i uwagi krytyczne nie podważają mojej finalnej pozytywnej oceny rozprawy.

6 Działalność publikacyjna

We wstępie (rozdział 1.) i podsumowaniach rozdziałów 3, 4 i 5 Autor rozprawy wskazuje artykuły, których jest współautorem, a w których publikowane były wyniki zawarte w rozprawie, oraz ew. poszerzające niektóre wątki badawcze. Wskazywany dorobek publikacyjny obejmuje 3 artykuły w czasopiśmie z listy JCR, w tym International Journal of Computer Integrated Manufacturing (IF 4.9) i Enterprise Information Systems (IF 5.12), oraz 10 artykułów na relatywnie rozpoznawanych konferencjach międzynarodowych dotyczących głównie przemysłowych zastosowań informatyki oraz inteligentnych metod analizy danych. We wszystkich, poza jedną publikacją pierwszym autorem jest Promotor, a Doktorant należy do zespołu 3-4 współautorów.

W bazie DBLP można znaleźć w sumie 4 artykuły w czasopiśmie i 17 artykułów konferencyjnych, których współautorem jest Autor rozprawy. Doktorant nie ma założonego profilu w serwisie Google Scholar.

W mojej opinii aktywność publikacyjna Pana Drewniaka jest duża i zdecydowanie wyróżniająca się.

7 Podsumowanie i wniosek końcowy

Recenzowana rozprawa doktorska zawiera oryginalne wyniki badań naukowych w ważnym obszarze zastosowań informatyki i AI w przemyśle. Autor udało się zrealizować postawione cele badawcze oraz wykazać prawdziwość sformułowanych w rozprawie tez. Doktorant wykazał się umiejętnościami analizy złożonych problemów praktycznych i użycia do nich rozwiązań koncepcyjnych, w tym algorytmów z obszaru eksploracji danych. Ponadto wraz z promotorem opublikował szereg swoich wyników w artykułach naukowych.

Podsumowując, stwierdzam, że recenzowana przez mnie rozprawa spełnia wymagania stawiane rozprawom doktorskim przez obowiązujące przepisy i wnoszę o dopuszczenie Pana mgra inż. Marka Drewniaka do dalszych etapów przewodu doktorskiego, w tym publicznej obrony.

