

## KNOWLEDGE ACQUISITION FROM DATABASE FOR MARINE DIESEL ENGINE DIAGNOSIS

Adam Charchalis, Rafał Pawletko

Gdynia Maritime University  
Faculty of Marine Engineering  
Morska Street 83, 81-225 Gdynia, Poland  
tel.: +48 58 6901305  
e-mail: pawletko@am.gdynia.pl

### Abstract

*This article presents an attempt to use inductive machine learning methods to knowledge acquisition from the databases for the purpose of marine diesel engine diagnostic's expert system. The existing methods of acquiring knowledge based on information about the exploitation of technical objects stored in databases was characterizes. The selected machine learning methods was used to obtain the relationship in the form of diagnostic rules. The results obtained with algorithms LEM2, MODLEM and EXPLORE was compared. MODLEM algorithm allows the use of numerical data directly without having to pre-discretization. Learning examples stored in the diagnostic database was obtained as a result of the active experiment, carried out on laboratory Sulzer engine 3AL 25/30. During the experiment the damages of the turbocharging system, fuel injection system and combustion chamber was simulated. Only the elementary states (single damage in the same time) in a variable load were included. 10-fold cross validation technique was used for evaluation of the obtained rules classifiers. The obtained diagnostic rules have also been assessed in substantive terms, including an analysis of the relationship between disability states and received symptoms. Used machine learning techniques can be used for automatic knowledge acquisition for the diagnostic expert system.*

**Keywords:** technical diagnostic, expert systems, combustion engines, expert knowledge acquisition

## POZYSKIWANIE WIEDZY Z BAZ DANYCH DLA POTRZEB DIAGNOZOWANIA OKRĘTOWEGO SILNIKA SPALINOWEGO

### Streszczenie

*W artykule przedstawiono próbę wykorzystania indukcyjnych metod uczenia maszynowego, do pozyskania wiedzy z baz danych dla potrzeb ekspertowego systemu diagnozowania okrętowego silnika tłokowego. Scharakteryzowano istniejące metody pozyskiwania wiedzy na podstawie informacji o przebiegu eksploatacji obiektów technicznych zapisanych w bazach danych. Zastosowano wybrane metody uczenia maszynowego do uzyskania relacji diagnostycznych w postaci reguł. Porównano wyniki uzyskane za pomocą algorytmów LEM2, MODLEM oraz EXPLORE. Algorytm MODLEM umożliwia wykorzystanie bezpośrednio danych numerycznych bez konieczności stosowania dyskretyzacji wstępnej. Przykłady uczące zapisane w diagnostycznej bazie danych uzyskano w wyniku realizacji eksperymentu czynnego, przeprowadzonego na silniku laboratoryjnym Sulzer 3Al 25/30. Podczas eksperymentu symulowano wybrane uszkodzenia układu wymiany czynnika roboczego, układu wtryskowego oraz komory spalania. Uwzględniono tylko stany elementarne (pojedyncze uszkodzenia w tym samym czasie) w warunkach zmiennego obciążenia. Oceny sprawności uzyskanych klasyfikatorów regułowych dokonano techniką 10-fold cross validation. Uzyskany zbiór reguł diagnostycznych został również poddany ocenie merytorycznej, obejmującą analizę związków pomiędzy diagnozowanymi stanami niezdatności a uzyskanymi symptomami. Wykorzystane techniki uczenia maszynowego mogą zostać zastosowane między innymi do automatycznego pozyskiwania wiedzy dla potrzeb systemu ekspertowego.*

**Słowa kluczowe:** diagnostyka techniczna, systemy ekspertowe, silniki spalinowe, pozyskiwanie wiedzy ekspertowej

### 1. Wprowadzenie

Rozwój systemów diagnozowania okrętowych silników tłokowych jest istotnym zagadnieniem, zarówno z punktu widzenia bezpieczeństwa statku jak i uzyskania bezpośrednich korzyści

ekonomicznych. Istnieje obecnie szereg systemów diagnozowania silników okrętowych, opracowanych zarówno przez ośrodki badawcze, jak i samych producentów silników. Główną wadą większości tych rozwiązań jest to, że są to systemy zamknięte. Oznacza to, że algorytmy oceny stanu technicznego zaimplementowane na etapie tworzenia systemu, nie mogą być rozwijane oraz modyfikowane w czasie późniejszej eksploatacji.

Rozwiązaniem tego problemu może być wykorzystanie systemu ekspertowego do diagnozowania silników okrętowych. Modułowa struktura takiego systemu, a przede wszystkim oddzielenie bazy wiedzy od reszty programu, umożliwi opracowanie systemu otwartego, w którym wiedza diagnostyczna może być w łatwy sposób uaktualniana i rozszerzana.

Wiedza dla potrzeb diagnostycznego systemu ekspertowego może być pozyskana od ekspertów (specjalistów w dziedzinie eksploatacji) oraz z diagnostycznych baz danych. Pozyskiwanie wiedzy od ekspertów jest mało efektywne. Wymaga żmudnych i czasochłonnych badań, problemem jest dostępność samych ekspertów oraz ocena ich kompetencji. Z tego powodu coraz większego znaczenia w ostatnim czasie zaczyna nabierać drugie źródło wiedzy – diagnostyczne bazy danych.

W artykule podjęto próbę pozyskania wiedzy z baz danych za pomocą indukcyjnych metod uczenia maszynowego.

## 2. Pozyskiwanie wiedzy z baz danych

Rozwój systemów informatycznych oraz ich powszechna dostępność spowodowały, że są one coraz częściej stosowane w siłowniach okrętowych. Współczesne systemy kontrolne, oprócz pomiaru szeregu parametrów pracy siłowni, umożliwiają także ich automatyczną rejestrację. W wyniku działania systemów pomiarowych powstają obszerne zbiory danych, których analiza oraz poprawna interpretacja coraz częściej przekracza możliwości człowieka. W związku z tym nastąpił rozwój metod i narzędzi informatycznych, wspomagających proces pozyskiwania wiedzy z baz danych.

Dla potrzeb diagnostyki technicznej szczególnie przydatne okazały się indukcyjne metody uczenia maszynowego. Umożliwiają one uzyskanie wiedzy reprezentowanej najczęściej w postaci reguł na bazie przykładów uczących.

Opracowano wiele metod umożliwiających indukcję reguł. Większość istniejących rozwiązań bazuje na algorytmie generowania kolejnych pokryć opracowanym przez R. Michalskiego [8]. Algorytm polega na uczeniu się pojedynczej reguły, usuwaniu przykładów, które ona pokrywa, i powtarzaniu procesu dla pozostałych przykładów. Wynikiem działania algorytmu jest zbiór reguł pokrywających rozważany zbiór przykładów. Na tej zasadzie bazują między innymi algorytmy serii AQ [3], INLEN oraz LEM [2].

Innym podejściem jest generowanie drzew decyzyjnych. Do tej grupy zaliczyć należy przede wszystkim algorytmy ID3, C4 zaproponowane przez Quinlana [7] oraz CN2. Otrzymane, w wyniku działania tych algorytmów, drzewo decyzyjne może być łatwo przekształcone do zbioru reguł.

Współczesne bazy danych bardzo często zawierają dane numeryczne, tj. liczby całkowite lub rzeczywiste. Większość algorytmów indukcji reguł wymaga natomiast przykładów uczących, reprezentowanych przez atrybuty dyskretne. W pracy [2] zaproponowano następujące dwa sposoby rozwiązania tego problemu:

- zastosowanie metod wstępnej dyskretyzacji danych,
- wybór specjalizowanego algorytmu, dostosowanego do indukcji reguł bezpośrednio z danych numerycznych.

W pierwszym rozwiązaniu przed użyciem algorytmu indukcji reguł realizuje się dyskretyzację wstępną. Polega ona na zamianie atrybutów numerycznych atrybutami symbolicznymi. Proces ten realizuje się poprzez podział oryginalnej dziedziny atrybutu numerycznego na pewną liczbę podprzedziałów i przypisaniu tym przedziałom kodów symbolicznych. Dotychczas zaproponowano wiele metod dyskretyzacji wstępnej, których klasyfikacja bazuje przede wszystkim na rodzaju informacji wykorzystywanych podczas przetwarzania danych. Metody,

które wykorzystują informacje o przydziale obiektów do poszczególnych klas zalicza się do nadzorowanych w przeciwieństwie do metod nienadzorowanych, w których ta informacja nie jest wykorzystywana. Podział na metody globalne i lokalne jest różnicowany z zależności od tego czy dyskretyzacja jest realizowana jednakowo na całej dziedzinie atrybutu (globalna) czy też, w różny sposób dla różnych jej obszarów (lokalna) [9].

Stosowanie dyskretyzacji wstępnej, nie jest konieczne w przypadku zastosowania algorytmu umożliwiającego indukcję reguł bezpośrednio z danych numerycznych. Niewątpliwą korzyścią takiego rozwiązania jest uproszczenie procesu pozyskiwania wiedzy, poprzez wyeliminowanie etapu wstępnej dyskretyzacji danych. Przykładem algorytmu umożliwiającego indukcję bez stosowania dyskretyzacji wstępnej jest algorytm MODLEM [2].

Należy podkreślić, że dla różnych zbiorów danych, skuteczne są różne strategie pozyskiwania wiedzy. Istnieje konieczność doboru właściwej metody dla konkretnych danych na drodze eksperymentalnej. Nie istnieje bowiem uniwersalna skuteczna metoda dla dowolnych zastosowań.

### 3. Badania doświadczalne

Celem badań doświadczalnych było uzyskanie zbioru przykładów reprezentujących wybrane stany niezdatności silnika okrętowego.

Obiektem badań był czterosuwowy silnik typu Sulzer 3A1 25/30 o mocy nominalnej  $N_n = 408$  kW i prędkości obrotowej  $n = 750$  obr/min. Silnik został wyposażony w układ pomiarowy, umożliwiający rejestrację podstawowych parametrów roboczych, takich jak ciśnienia i temperatury spalin, powietrza doładowującego, wody chłodzącej oraz oleju smarowego. Dodatkowo mierzono przebiegi ciśnień szybkozmiennych w cylindrach silnika oraz w przewodach paliwowych. Wszystkie parametry były automatycznie zapisywane w bazie danych zintegrowanej z systemem pomiarowym.

Program badań zrealizowano zgodnie z planem eksperymentu czynnego. Podczas eksperymentu symulowano wybrane uszkodzenia układu wtryskowego, wymiany czynnika roboczego oraz komory spalania. Wykaz symulowanych stanów silnika przedstawiono w Tab. 1.

Tab. 1. Engine faults simulated during experiment  
Tab. 1. Uszkodzenia silnika symulowane podczas eksperymentu

L.p.	Układ funkcjonalny silnika	Stan niezdatności
1.	Układ wymiany czynnika roboczego	Spadek wydajności sprężarki powietrza
2.		Zanieczyszczenie filtra turbosprężarki
3.		Zanieczyszczenie chłodnicy powietrza doładowującego
4.		Zanieczyszczenie traktu wylotowego
5.	Układ wtryskowy	Nieszczelność pompy wtryskowej
6.		Zmniejszenie ciśnienia otwarcia wtryskiwacza
7.		Niedrożne otwory wtryskiwacza
8.		Rozkalibrowane otwory wtryskiwacza
9.	Komora spalania	Nieszczelność głowicy

Wyniki pomiarów zostały zarejestrowane w bazie danych a następnie przekształcone do postaci tablicy decyzyjnej. Taka forma reprezentacji danych jest wymagana przez algorytmy indukcji reguł. Poszczególne przykłady uczące są w takiej sytuacji opisane w wierszach tablicy, za

pomocą zbioru atrybutów. Jeden z tych atrybutów jest atrybutem decyzyjnym określającym przynależność przykładu do określonej klasy decyzyjnej [2].

Uzyskana tablica zawierała 450 przykładów uczących, każdy opisany 43 atrybutami typu numerycznego. Przykłady obejmowały 10 symulowanych stanów silnika (wraz ze stanem wzorcowym).

#### 4. Wyniki eksperymentu

Do pozyskania wiedzy z bazy danych zastosowano indukcyjne metody uczenia maszynowego. Celem pozyskiwania wiedzy było uzyskanie reguł, umożliwiających ocenę stanu okrętowego silnika spalinowego, na podstawie informacji o przebiegu eksploatacji zgromadzonych w bazie danych. Bazę danych stanowiły wyniki uzyskane w ramach eksperymentu diagnostycznego.

W ramach badań porównano wyniki uzyskane klasycznym algorytmem LEM2 z dwoma nowymi algorytmami MODLEM oraz EXPLORE. W związku z tym, iż algorytmy LEM2 i EXPLORE nie powinny być stosowane bezpośrednio do danych numerycznych, zastosowano dyskretyzację wstępną. Do badań wykorzystano oprogramowanie opracowane przez Zakład Inteligentnych Systemów Wspomagania Decyzji Politechniki Poznańskiej o nazwie ROSE2 [10, 11].

Ocenę uzyskanych zbiorów reguł przeprowadzono w perspektywie klasyfikacji symulowanych stanów niezdatności. Do bezpośredniego wyznaczenia sprawności klasyfikacji zastosowano technikę *k-fold cross validation*. W metodzie tej zbiór przykładów jest losowo podzielony na  $k$  podzbiorów  $U = E_1 \cup \dots \cup E_k$ . W  $i$ -tej iteracji ( $1 \leq i \leq k$ ), zbiór uczący stanowi zbiór  $E_u = U \setminus E_i$ , a sam zbiór  $E_i$  jest zbiorem przykładów testowych. Trafność klasyfikowania jest wyliczana jako wartość średnia z trafności estymowanych w każdej iteracji. Warunkiem stosowania tej techniki jest liczba przykładów powyżej 100. Dobór parametru  $k$  powinien być uzależniony od liczebności przykładów. Przyjęto wartość  $k=10$ . W Tab. 2 przedstawiono liczbę reguł oraz trafności klasyfikacji, uzyskane techniką *10-fold cross validation* dla badanych algorytmów indukcji reguł decyzyjnych.

Tab. 2. Classification results receives with different algorithms  
Tab. 2. Wyniki klasyfikacji uzyskane różnymi algorytmami

Lp.	Dyskretyzacja wstępna	Algorytm indukcji	Liczba uzyskanych reguł	Udział poprawnie sklasyfikowanych przykładów [%]	Udział błędnie sklasyfikowanych przykładów [%]	Udział przykładów niesklasyfikowanych [%]
1.	Brak	LEM2	178	24	32	44
2.		MODLEM	35	87	2	11
3.		EXPLORE	5	21	76	3
4.	Metoda lokalna	LEM2	56	91	9	0
5.		MODLEM	46	91	9	0
6.		EXPLORE	300	74	26	0

Stwierdzono zdecydowaną przewagę algorytmu MODLEM nad LEM2 oraz EXPLORE w odniesieniu do danych rzeczywistych nie poddanych dyskretyzacji wstępnej. Uzyskana średnia trafność klasyfikacji estymowana metodą *10-fold cross validation* dla algorytmu MODLEM wyniosła aż 87% w porównaniu z 24% LEM2 oraz 21% EXPLORE. Po zastosowaniu dyskretyzacji wstępnej brak jest istotnych różnic pomiędzy algorytmami LEM2 oraz MODLEM (91%). Najniższą sprawność klasyfikacji uzyskano algorytmem EXPLORE (74%).

Techniki oceny klasyfikatorów, takie jak *10-fold cross validation* czy *leave one out* pozwalają

na efektywną ocenę skuteczności klasyfikacji zbioru reguł uzyskanego w trakcie indukcji. Należy jednak pamiętać o zagrożeniach, jakie są związane z automatycznymi metodami uczenia maszynowego. W trakcie procesu uczenia korzysta się zazwyczaj ze wszystkich dostępnych atrybutów opisujących działanie obiektów. Algorytmy uczenia pozwalają na efektywną selekcję tzw. atrybutów relewantnych, które niosą informację diagnostyczną. Taka selekcja uwzględnia niezbędną ilość informacji do rozróżnienia klasyfikowanych obiektów. Brak jest natomiast możliwości automatycznego uwzględnienia rzeczywistych zależności przyczynowo-skutkowych typu symptomy – stany niezdatności. W trakcie uczenia istnieje niebezpieczeństwo uwzględnienia atrybutów, które nie zawierają informacji diagnostycznej, ale wynikają z warunków działania obiektu czy zakłóceń losowych.

W związku z powyższym podjęto próbę analizy reguł uzyskanych podczas automatycznej indukcji, pod kątem merytorycznym. Nie jest oczywiście celowa analiza przyjętych progów warunków w przesłankach reguł, ponieważ wynikają one z możliwości rozróżnienia poszczególnych stanów i trudno byłoby znaleźć ich logiczne uzasadnienie, możliwa jest jednak próba uzasadnienia wpływu danego stanu (symulowanego uszkodzenia) na atrybuty zawarte w przesłankach reguł.

W celu identyfikacji spadku wydajności sprężarki powietrza wyznaczone zostały cztery reguły. Reguła nr 8 obejmuje 20 przykładów (44%), reguła nr 9 - 9 przykładów (20%), reguła nr 10 - 9 przykładów (20%), reguła nr 11 - 18 przykładów (40%). Reguły w formacie ROSE2 zostały przytoczone poniżej:

rule 8.  $(a22 < 145.625) \ \& \ (a35 \geq 506.25) \Rightarrow (\text{Dec} = 1); [20, 20, 44.44\%, 100.00\%]$

rule 9.  $(a22 < 103.125) \ \& \ (a37 \geq 387.45) \Rightarrow (\text{Dec} = 1); [9, 9, 20.00\%, 100.00\%]$

rule 10.  $(a4 \geq 8.75) \ \& \ (a22 < 75) \Rightarrow (\text{Dec} = 1); [9, 9, 20.00\%, 100.00\%]$

rule 11.  $(a34 \geq 517.2) \Rightarrow (\text{Dec} = 1); [18, 18, 40.00\%, 100.00\%]$

Przesłanki reguł odwołują się do następujących atrybutów:

- a4 - Maksymalne ciśnienie spalania - cylinder 3,
- a22 - Ciśnienie powietrza doładowującego,
- a34 - Temperatura spalin za cylindrem 1,
- a35 - Temperatura spalin za cylindrem 2,
- a37 - Temperatura spalin przed turbiną.

Przesłanka reguły nr 8 składa się z dwóch warunków elementarnych. Odwołują się one do atrybutów a22 - ciśnienia powietrza doładowującego oraz a35 - temperatury spalin za cylindrem 2. Symptodem symulowanego spadku wydajności sprężarki powietrza powinna być zmiana wartości ciśnienia doładowania. Jest to parametr diagnostyczny, który powinien reagować na uszkodzenia tego elementu. Spadek wartości ciśnienia doładowania skutkuje natomiast mniejszą ilością powietrza doprowadzoną do cylindra oraz zaburzeniami przebiegu procesu roboczego. Efektem tych zaburzeń może być spadek lub wzrost wartości temperatury spalin za cylindrami. Przesłankę reguły numer 8 należy zatem uznać za poprawną. Podobne uzasadnienie można wskazać w stosunku do reguły nr 9 oraz nr 11, których przesłanki odwołują się również do wartości temperatur spalin za cylindrami oraz ciśnienia powietrza doładowującego. W przesłance reguły nr 10 pojawia się dodatkowo maksymalne ciśnienie spalania – cylinder nr 3. Można uznać, że zmiana tego parametru jest również symptodem zaburzeń w przebiegu procesu spalania, spowodowanego mniejszą ilością powietrza doładowującego.

Przedstawioną analizę merytoryczną przeprowadzono dla wszystkich uzyskanych reguł diagnostycznych.

## 5. Podsumowanie

Na podstawie przeprowadzonych badań wstępnych, nad możliwością zastosowania indukcyjnych metod uczenia maszynowego, do pozyskania wiedzy diagnostycznej z baz danych,

dla potrzeb diagnozowania silnika okrętowego można stwierdzić:

- Zastosowane algorytmy indukcji reguł pozwoliły na pozyskanie wiedzy z diagnostycznej bazy danych zawierającej wyniki eksperymentu.
- Uzyskane klasyfikatory regułowe, charakteryzują się wysoką sprawnością klasyfikacji.
- Stwierdzono zdecydowaną przewagę algorytmu MODLEM nad LEM2 w odniesieniu do danych rzeczywistych nie poddanych dyskretyzacji wstępnej. Uzyskana średnia trafność klasyfikacji estymowana metodą *10-fold cross validation* dla algorytmu MODLEM wyniosła aż 92% w porównaniu z 10% LEM2.
- Po zastosowaniu dyskretyzacji wstępnej brak jest istotnych różnic pomiędzy badanymi algorytmami.
- Zastosowanie algorytmu MODLEM upraszcza proces pozyskiwania wiedzy (nie ma potrzeby stosowania dyskretyzacji danych) oraz ułatwia interpretację uzyskanych reguł.
- Algorytmy automatycznej indukcji, mogą być wykorzystywane do pozyskiwania wiedzy z baz danych, dla potrzeb diagnostycznego systemu ekspertowego.

## Literatura

- [1] Cholewa, W., *Metoda diagnozowania maszyn z zastosowaniem zbiorów rozmytych*, ZN Politechniki Śląskiej Nr 764, Seria: Mechanika Z. 79, Gliwice 1983.
- [2] Stefanowski, J., *Algorytmy indukcji reguł decyzyjnych w odkrywaniu wiedzy*, Wydawnictwo Politechniki Poznańskiej, Rozprawy Nr 361, Poznań 2001.
- [3] Michalski, R. S., *A theory and methodology of inductive learning*, Artificial Intelligence 20, pp. 111-161, 1983.
- [4] Moczulski, W., *Metody pozyskiwania wiedzy dla potrzeb diagnostyki maszyn*, ZN Politechniki Śląskiej Nr 1382, Seria: Mechanika Z. 130, Gliwice 1997.
- [5] Mulawka, J., *Systemy ekspertowe*, WNT, Warszawa 1996.
- [6] Żółtowski, B., Cempel, Cz., *Inżyniera Diagnostyki Maszyn*, Polskie Towarzystwo Diagnostyki Technicznej, Instytut Technologii Eksploatacji PIB Radom, Warszawa, Bydgoszcz, Radom 2004.
- [7] Quilian, J. R., *Induction of decision trees*, Machine Learning, 1, 1986.
- [8] Michalski, R. S., *A theory and methodology of inductive learning*, Artificial Intelligence 20, 1983.
- [9] Cichosz, P., *Systemy uczące się*, WNT, Warszawa 2007.
- [10] Predki, B., Słowinski, R., Stefanowski, J. Susmaga, R., Wilk, Sz., *ROSE - Software Implementation of the Rough Set Theory*. In: Polkowski, L., Skowron, A., eds. Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, Vol. 1424, 605-608, Berlin 1998.
- [11] Predki, B., Wilk, Sz., *Rough Set Based Data Exploration Using ROSE System*. In: Ras, Z. W., Skowron, A., eds. Foundations of Intelligent Systems, Lecture Notes in Artificial Intelligence, Vol. 1609, 172-180, Berlin 1999.