

## POZYSKIWANIE WIEDZY Z BAZY DANYCH DLA POTRZEB DIAGNOZOWANIA OKRĘTOWEGO SILNIKA TŁOKOWEGO

Rafał PAWLETKO

Katedra Siłowni Okrętowych, Akademia Morska w Gdyni  
ul. Morska 81-87, 81-225 Gdynia

### Streszczenie

W artykule przedstawiono próbę wykorzystania indukcyjnych metod uczenia maszynowego, do pozyskania wiedzy dla potrzeb ekspertowego systemu diagnozowania okrętowego silnika tłokowego. Metody uczenia maszynowego zastosowano do uzyskania reguł diagnostycznych. Przykłady uczące do indukcji reguł stanowiły wyniki eksperymentu czynnego, przeprowadzonego na silniku laboratoryjnym. Oceny sprawności uzyskanych klasyfikatorów regułowych dokonano techniką *k-fold cross validation*. Wykorzystane techniki mogą zostać zastosowane między innymi do automatycznego pozyskiwania wiedzy dla potrzeb systemu ekspertowego.

Słowa kluczowe: diagnostyka techniczna, silniki spalinowe, klasyfikator regułowy.

### KNOWLEDGE ACQUISITION FROM DATABASE FOR MARINE DIESEL ENGINE DIAGNOSTIC

### Summary

In this paper automatic rule induction algorithms are used to knowledge acquisition from data base for marine diesel engine diagnostic expert system. Training and test data were acquired from experiment on marine engine Sulzer 3AL 25/30. 10-fold cross validation method was used to estimation classification efficiency for different rule induction algorithms.

Keywords: technical diagnostic, combustion engines, expert knowledge acquisition.

## 1. WPROWADZENIE

Rozwój systemów diagnozowania okrętowych silników tłokowych jest istotnym zagadnieniem, zarówno z punktu widzenia bezpieczeństwa statku, jak i uzyskania bezpośrednich korzyści ekonomicznych związanych z eksploatacją. Istnieje obecnie szereg metod oraz systemów diagnozowania silników okrętowych, opracowanych zarówno przez ośrodki badawcze, jak i producentów silników. Główną wadą większości tych rozwiązań jest to, iż 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, umożliwia opracowanie systemu otwartego, w którym wiedza diagnostyczna może być w łatwy sposób uaktualniana i rozszerzana.

Kluczowym problemem podczas budowy systemu ekspertowego jest pozyskanie wiedzy. W przypadku systemów ekspertowych z dziedziny diagnostyki technicznej, wiedza ta może być pozyskana przede wszystkim od specjalistów oraz z baz danych, w których są gromadzone informacje

o przebiegu eksploatacji obiektów technicznych. Wśród metod umożliwiających pozyskiwanie wiedzy z baz danych, dominują indukcyjne metody uczenia maszynowego. Metody te powinny stanowić uzupełnienie, a w niektórych sytuacjach nawet alternatywę dla mało efektywnego pozyskiwania wiedzy od specjalistów.

W artykule podjęto próbę, zastosowania indukcyjnych metod uczenia maszynowego, do pozyskania wiedzy dla potrzeb ekspertowego systemu diagnozowania silnika okrętowego.

## 2. POZYSKIWANIE WIEDZY Z DIAGNOSTYCZNYCH BAZ DANYCH

Rozwój systemów informatycznych i 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 ich działania 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 indukcje reguł. Większość istniejących rozwiązań bazuje na algorytmie generowania kolejnych pokryw opracowanym przez R. Michalskiego. Polega on na uczeniu się pojedynczej reguły, usuwaniu przykładów, które pokrywa, i powtarzaniu procesu dla pozostałych przykładów. Wynikiem działania algorytmu jest zbiór reguł pokrywających rozważany przykład. 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 [9] 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 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 indukcje 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.

#### 4. BADANIA DOŚWIADCZALNE

Celem badań było uzyskanie reguł, umożliwiających ocenę stanu okrętowego silnika spalinowego, na podstawie informacji o przebiegu eksploatacji zgromadzonych w bazie danych.

Zbiór reguł uzyskano za pomocą wybranych metod indukcji reguł. Porównano wyniki uzyskane klasycznym algorytmem LEM2 z algorytmem MODLEM, który umożliwia zastosowanie danych nie poddanych wcześniejszej dyskretyzacji.

Oceny przydatności algorytmów indukcji reguł, dokonano na danych pozyskanych w ramach eksperymentu czynnego na silniku laboratoryjnym.

Obiektem badań był czterosurowy silnik typu Sulzer 3A1 25/30 o mocy nominalnej  $N_n=408$  kW i prędkości obrotowej  $n=750$  obr/min doładowany turbosprężarką. 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 były mierzone 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 jeden poziom określonego uszkodzenia, następnie dokonywano pomiarów wszystkich parametrów, w zakresie pracy silnika od 50 do 250 kW. Doświadczenie nie uwzględniało występowania wielu uszkodzeń jednocześnie oraz różnego poziomu natężenia danego uszkodzenia.

Uwzględniono następujące uszkodzenia silnika:

- spadek wydajności sprężarki powietrza,
- zanieczyszczenie filtra powietrza,
- zanieczyszczenie chłodnicy powietrza,
- zużycie pompy wtryskowej na cyl. nr 2,
- zakokosowanie wtryskiwacza na cyl. nr 2,
- zanieczyszczenie traktu wydechowego.

Wyniki pomiarów zostały zarejestrowane w bazie danych a następnie przekształcone do postaci tablicy decyzyjnej. Taka forma reprezentacji danych jest bowiem wymagana przez zastosowane 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 215 przykładów uczących, każdy opisany 43 atrybutami typu

numerycznego. Przykłady obejmowały 6 symulowanych uszkodzeń silnika.

W związku z tym, iż algorytm LEM2 nie powinien być stosowany bezpośrednio do danych numerycznych, zastosowano dyskretyzację wstępną. Dyskretyzację zrealizowano zarówno metodą globalną, jak i metodą lokalną.

Badane algorytmy indukcji (LEM2 oraz MODLEM) zastosowano zarówno do danych nie poddanych dyskretyzacji, jak i poddanych dyskretyzacji metodą lokalną i globalną.

Do badań wykorzystano oprogramowanie opracowane przez Zakład Inteligentnych Systemów Wspomagania Decyzji Politechniki Poznańskiej o nazwie ROSE2 [10, 11].

## 5. OCENA WYNIKÓW EKSPERYMENTU

Ocena uzyskanych zbiorów reguł była realizowana w perspektywie klasyfikacji. Oznacza to, że każdorazowo na podstawie reguł budowano klasyfikator, który był poddawany ocenie.

Ocena sprawności działania klasyfikatora obejmowała zarówno przykłady uczące, które służyły do budowy klasyfikatora, jak i przykłady nowe – nie znane podczas nauki. Można w ten sposób zweryfikować zdolność klasyfikatora do uogólniania pozyskanej wiedzy. Najczęściej stosowanym rozwiązaniem jest podział zbioru dostępnych przykładów na część uczącą oraz testującą, służącą do estymacji wybranej miary oceny klasyfikatora. Sposób podziału zbioru jest losowy i zależy przede wszystkim od liczby dostępnych przykładów.

Na potrzeby pracy zastosowano technikę *k-fold cross validation*. W metodzie tej zbiór przykładów jest losowo podzielony na *k* podzbiorów

$U = E1 \cup \dots \cup Ek$ . W *i*-tej iteracji ( $1 \leq i \leq k$ ), zbiór uczący stanowi zbiór  $Eu = U \setminus Ei$ , a sam zbiór  $Ei$  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 tabeli 1 przedstawiono liczbę reguł oraz trafności klasyfikacji, uzyskane techniką *10-fold cross validation* dla badanych algorytmów indukcji reguł decyzyjnych.

Uzyskane wyniki potwierdzają wysoką skuteczność algorytmu MODLEM, dla danych nie poddanych wcześniejszej dyskretyzacji. Uzyskana trafność klasyfikacji estymowana techniką *10-fold cross validation* wyniosła 92%. Trafność klasyfikacji uzyskana algorytmem LEM2 wyniosła w tym przypadku 10%. W przypadku zastosowania dyskretyzacji wstępnej uzyskano zbliżone, bardzo wysokie, wyniki skuteczności klasyfikacji dla obydwu algorytmów. Należy również stwierdzić, że w badanym przypadku, zdecydowanie najlepsze rezultaty uzyskano przy zastosowaniu dyskretyzacji wstępnej metodą lokalną.

Niewątpliwą zaletą algorytmu MODLEM w stosunku do LEM2, jest możliwość wykorzystania bezpośredniego danych numerycznych, nie poddanych dyskretyzacji wstępnej. Z jednej strony, upraszcza to sam proces pozyskiwania wiedzy, z drugiej warunkuje bezpośrednio czytelność i łatwość interpretacji uzyskanych reguł. Użytkownik systemu ekspertowego ma w takiej sytuacji bezpośredni podgląd wartości cech zawartych w przesłankach reguły.

Tab. 1. Porównanie wyników trafności klasyfikacji uzyskanych algorytmami LEM2 oraz MODLEM

Rodzaj dyskretyzacji wstępnej	Algorytm indukcji reguł	Ilość uzyskanych reguł	Trafność klasyfikacji ( <i>10-fold cross validation</i> )
Bez dyskretyzacji	LEM2	132	10 %
	MODLEM	9	92 %
Dyskretyzacja wstępna metodą lokalną	LEM2	17	97 %
	MODLEM	14	97 %
Dyskretyzacja wstępna metodą globalną	LEM2	43	84 %
	MODLEM	45	79 %

## 5. PODSUMOWANIE

Na podstawie przeprowadzonych badań wstępnych, nad możliwością zastosowania indukcyjnych metod uczenia maszynowego, do pozyskania wiedzy diagnostycznej z baz danych, dla potrzeb diagnostowania 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 diagnostowania maszyn z zastosowaniem zbiorów rozmytych*, ZN Pol. 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 (1983), pp. 111-161
- [4] Moczulski W.: *Metody pozyskiwania wiedzy dla potrzeb diagnostyki maszyn*, ZN Pol. 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] B. Predki, R. Slowinski, J. Stefanowski, R. Susmaga, Sz. Wilk: *ROSE – Software Implementation of the Rough Set Theory*. In: L. Polkowski, A. Skowron, eds. *Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence*, vol. 1424. Springer-Verlag, Berlin (1998), 605-608.
- [11] B. Predki, Sz. Wilk: *Rough Set Based Data Exploration Using ROSE System*. In: Z. W. Ras, A. Skowron, eds. *Foundations of Intelligent Systems, Lecture Notes in Artificial Intelligence*, vol. 1609. Springer-Verlag, Berlin (1999), 172-180.



Mgr inż. **Rafał PAWLETKO** jest asystentem w Katedrze Siłowni Okrętowych, Wydziału Mechanicznego Akademii Morskiej w Gdyni. Zajmuje się diagnostyką okrętowych silników spalinowych.