

Metody uczenia się maszyn w diagnostyce układów biegowych pojazdów szynowych za pomocą sygnałów wibroakustycznych

W artykule przedstawiono rezultaty pracy badawczej związanej z diagnozowaniem układów biegowych pojazdów szynowych za pomocą sygnałów wibroakustycznych. Po krótkim omówieniu metod uczenia się maszyn podane zostały podstawowe wiadomości o najnowszych systemach obliczeniowych, tzw. sieciach neuronowych i drzewach decyzyjnych. W części podstawowej artykułu zaprezentowana została metoda przetwarzania sygnałów wibroakustycznych zarejestrowanych podczas wstępnych badań poligonowych oraz wyniki zastosowania sztucznych sieci neuronowych i drzew decyzyjnych do klasyfikacji uszkodzeń. W zakończeniu wskazano na pozytywne rokowania metody.

1. Wprowadzenie

Artykuł omawia wstępne wyniki zastosowania algorytmów uczących się (metoda propagacji wstecznej w sieciach neuronowych i metoda uczenia drzew decyzyjnych) do diagnostyki układów biegowych pojazdów szynowych na bazie sygnałów wibroakustycznych.

Ogólne koncepcje omawianej tu problematyki diagnostyki pojazdów szynowych opisane zostały w artykule [1]. Celem prowadzonych prac jest zbadanie możliwości diagnozowania stanu układów biegowych pojazdów szynowych na podstawie sygnałów wibroakustycznych zbieranych z szyn w czasie przejazdu pojazdów. Wykazanie takich możliwości mogłoby w późniejszym czasie doprowadzić do realizacji systemu diagnostycznego wykrywającego uszkodzenia w przejeżdżającym pojeździe.

Prace, których wyniki przedstawiono w niniejszym artykule, realizowane są w ramach projektu badawczego nr 9 T12C 017 15 pt. „Wybór sygnałów wibroakustycznych oraz metod ich przetwarzania i analizowania dla oceny stanu technicznego układu biegowego pojazdów szynowych”.

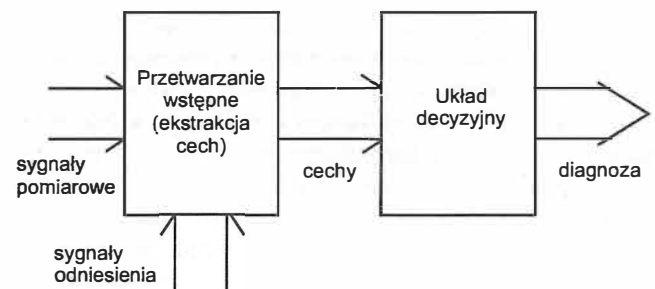
Sygnały wibroakustyczne mierzone były przez czujniki przyspieszenia umieszczone w szynie. Na szynie umieszczano przeszkodę w postaci metalowej płytki lub klina. Wjazd i spadek koła z przeszkody wymuszały drgania w szynie, które mierzono i przetwarzano.

Przetwarzanie otrzymanych sygnałów zrealizowano dwuetapowo (patrz rys. 1). Pierwszym etapem była ekstrakcja cech istotnych, a etapem drugim było podejmowanie decyzji o występowaniu uszkodzenia na podstawie otrzymanych cech. Algorytmami zastosowanymi w pierwszym etapie były filtracja cyfrowa i FFT mierzonego sygnału czasowego oraz parametryzacja sygnału otrzymanego po tych transformacjach. Algorytmami w drugim etapie były warstwowe sieci neuronowe o wagach wyznaczonych podczas procesów uczenia, a także drzewa decyzyjne skonstruowane na podstawie danych pomiarowych.

Występowanie dwóch podstawowych grup uszkodzeń układów biegowych (uszkodzeń układu zawieszenia i pęknięć kół) o bardzo różnej specyfice wymagać będzie przy-

puszczalnie dwóch systemów diagnostycznych o różnych algorytmach działania. W pierwszym etapie badań ograniczono się do jednego uszkodzenia, którym było zablokowanie sprężyny wózka.

Udział algorytmów przetwarzania uzyskanych na drodze uczenia w stosunku do udziału algorytmów uzyskanych na drodze analizy zależy od znajomości modeli matematycznych zachowania się pojazdów uszkodzonych i generacji przez nie sygnałów wibroakustycznych.



Rys.1

Posiadanie dokładnych modeli matematycznych wyeliminowałoby w znacznym stopniu konieczność prowadzenia uczenia, natomiast odwrotnie, niezajomość modeli matematycznych zmusza do korzystania z metod uczenia. Modele matematyczne pozwoliłyby także, w przypadku zastosowania metod uczenia, na częściowe przynajmniej zastąpienie rzeczywistych pomiarów przez eksperymenty symulacyjne na komputerze.

W przedstawionej pracy nie wykorzystano informacji o modelu matematycznym pojazdu – oparto się wyłącznie na otrzymanych pomiarach. W dalszej części pracy omówiono skrótowo zastosowane metody uczenia oraz przedstawiono uzyskane przy ich pomocy wyniki.

2. Metody uczenia się maszyn

Uczenie się komputerów jest jednym z ważnych sposobów pozyskiwania wiedzy lub uzyskiwania algorytmów

(optymalnych lub suboptymalnych). W czasie uczenia systemowi uczącemu się przedstawiane są zadania do rozwiązania i ewentualnie odpowiedzi, a system stopniowo ulepsza swoje działanie generując coraz lepsze rozwiązania. Uczenie prowadzone na podstawie obserwacji nazywa się uczeniem *indukcyjnym*.

Rozróżnia się trzy rodzaje uczenia. W czasie uczenia z *nauczycielem* dysponuje się zbiorem par uczących składających się z zadania i rozwiązania. Na wejście systemu podawane są zadania, a generowane przez system odpowiedzi porównywane są z rozwiązaniami. Różnice pomiędzy odpowiedziami i rozwiązaniami wpływają na zmianę algorytmu działania systemu. W trakcie uczenia z *krytykiem* (ang. *reinforcement learning*) nie są dostępne prawidłowe odpowiedzi, a jedynie łączna ocena po całej serii działań (np. rezultat partii szachów, mandat za nieprawidłową jazdę czy ocena zysku przedsiębiorstwa po określonym okresie czasu). Przy nauce *bez nauczyciela* wymagane są tylko obrazy wejściowe – system na ich podstawie samodzielnie optymalizuje swoje działanie. Systemy bez nauczyciela nazywane też są systemami *samoorganizującymi się*.

Innym podziałem metod uczenia się maszyn jest postać uzyskiwanej wiedzy. Może to być wiedza niejawna, uzyskiwana w postaci parametrów jakiegoś systemu przetwarzającego informację, np. wagi sieci neuronowej, parametry funkcji przy stosowaniu algorytmu genetycznego czy parametry funkcji przynależności uczących się systemów rozmytych. System z takimi parametrami przejawia oczekiwane, optymalne działanie, można więc mówić o uczeniu się działania. Drugą postacią uzyskiwanej wiedzy jest wiedza w postaci jawnej – wyrażona np. w postaci zbioru faktów i reguł lub drzewa decyzyjnego.

Jeszcze innym podziałem metod uczenia jest podział na metody iteracyjne wymagające wielokrotnego podawania danych uczących, w czasie którego system stopniowo poprawia swoje działanie i metody jednorazowe, w których na podstawie danych uczących wyznaczane są parametry systemu w jednorazowym procesie obliczeniowym [patrz np. 9,12].

3. Warstwowe sieci neuronowe i ich uczenie

Przy podejmowaniu decyzji na podstawie danych pomiarowych bardzo przydatne są układy obliczeniowe nazywane *sztucznymi sieciami neuronowymi*. Układy te wnoszą zasadnicze zmiany zarówno w sposobie przygotowania obliczeń jak i w metodach ich wykonania przez powszechnie stosowany szeregowy komputer [2,10,16].

Sieci neuronowe składają się z dużej liczby wzajemnie połączonych układów zwanych *sztucznymi neuronami*, które realizują proste przekształcenia sygnałów wejściowych w wyjściowe.

Sztuczny neuron jest układem statycznym o schemacie blokowym i symbolu przedstawionym na rys. 2a,b. Sygnał wyjściowy neuronu dany jest zależnością:

$$y = f(\text{net}), \quad \text{net} = \mathbf{w}^t \mathbf{x} = \left(\sum_{i=1}^n w_i x_i \right) \quad (1)$$

gdzie sygnał *net* będący ważoną sumą sygnałów wejściowych nazywany jest *pobudzeniem łącznym*, funkcja $f(\text{net})$ oddziałująca na pobudzenie łączne nazywa się *funkcją aktywacji*, oraz

$$\mathbf{w} \stackrel{\text{def}}{=} [w_1 \ w_2 \ \dots \ w_n]^t, \quad \mathbf{x} \stackrel{\text{def}}{=} [x_1 \ x_2 \ \dots \ x_n]^t$$

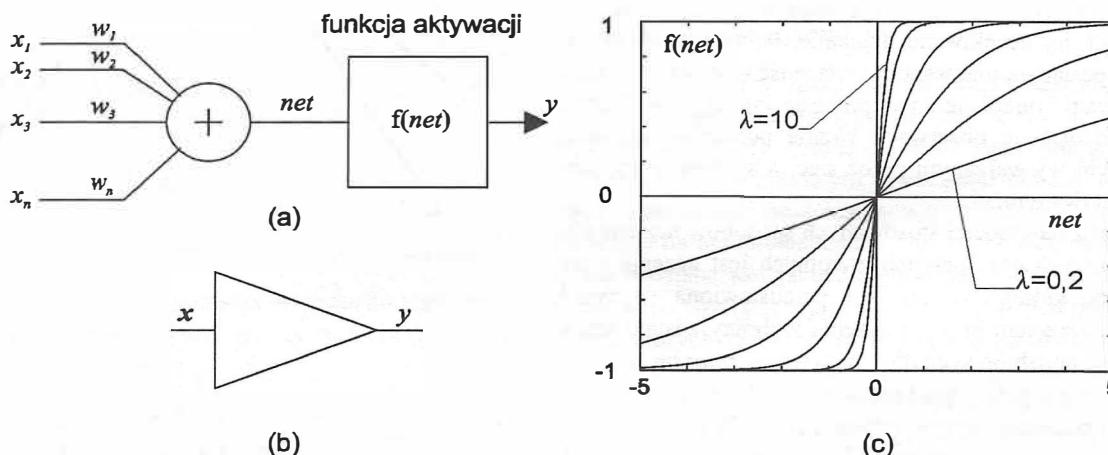
są odpowiednio wektorem wag i wektorem wejściowym.

Typową funkcją aktywacji jest

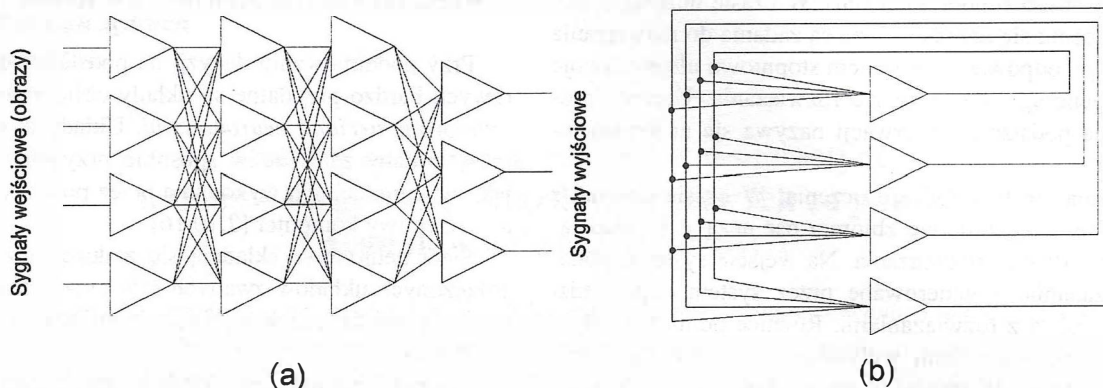
$$f(\text{net}) = \frac{2}{1 + \exp(-\lambda \text{net})} - 1, \quad \lambda > 0 \quad (2)$$

zwana ciągłą funkcją *bipolarną*. Jej wykresy dla różnych wartości parametru λ pokazane są na rys. 2c.

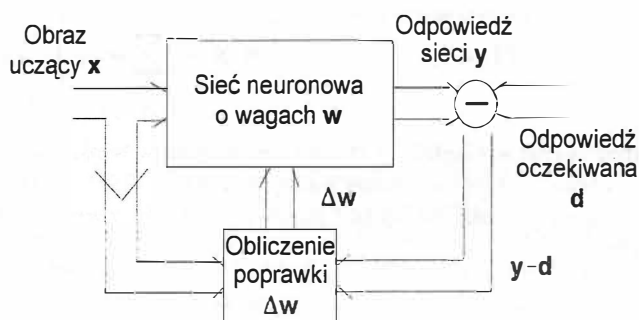
Neurony łączone są w większe struktury, spośród których można wyróżnić dwie podstawowe – sieci warstwowe bez sprzężeń zwrotnych oraz sieci ze sprzężeniami zwrotnymi (rys. 3).



Rys. 2 Sztuczny neuron: (a) schemat blokowy, (b) symbol i (c) bipolarna funkcja aktywacji dla różnych wartości



Rys. 3 Podstawowe struktury sieci neuronowych: (a) sieć warstwowa bez sprzężeń zwrotnych i (b) sieć ze sprzężeniami zwrotnymi



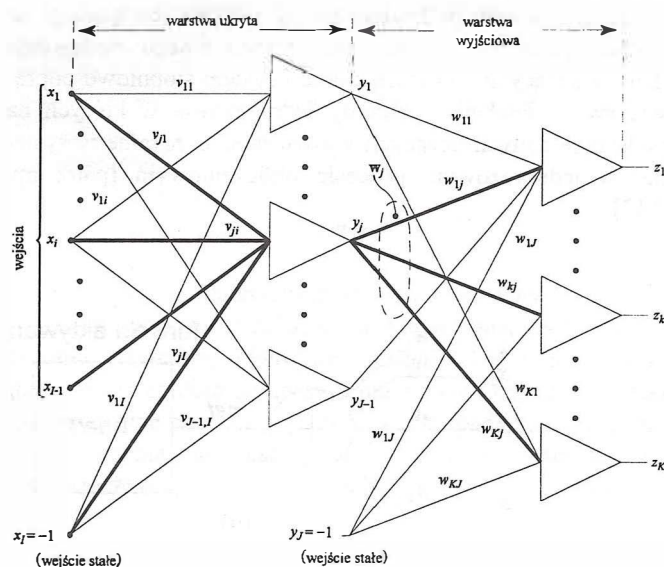
Rys. 4 Zasada nauki sieci warstwowej z nauczycielem

Sieci warstwowe bez sprzężeń zwrotnych można uważać za układy statyczne realizujące złożone nieliniowe odwzorowanie, którego kształt zależy od przyjętej funkcji aktywacji, stałej λ i wektorów wag w w wszystkich neuronów sieci. Potencjalny zakres zastosowań sieci warstwowych jest nadzwyczaj szeroki. Mogą one być stosowane we wszelkich zadaniach klasyfikacji, rozpoznawania, diagnostyki, monitorowania i sterowania.

Sieci neuronowe mogą być realizowane w postaci specjalizowanych układów elektronicznych lub symulowane przy pomocy uniwersalnych komputerów. Przygotowanie sieci do rozwiązywania określonych zadań odbywa się na drodze jej uczenia, w którym wyznaczone zostają wagi neuronów sieci powodujące jej oczekiwane działanie. Istnieje wiele metod nauki ale posiadają one wspólną własność – w czasie uczenia sieci jej wagi zmieniają się w procesie iteracyjnym, najczęściej na podstawie obserwacji różnic pomiędzy sygnałem wyjściowym wytwarzanym przez sieć, a sygnałem wyjściowym oczekiwanym.

Jednym z najczęściej stosowanych sposobów uczenia sieci warstwowych bez sprzężeń zwrotnych jest uczenie z nauczycielem, którego zasada jest przedstawiona na rys. 4. Przed rozpoczęciem procesu uczenia wybrany zostaje reprezentatywny podzbiór sygnałów nazywany *zbiorem uczącym* składający się z par: sygnał wejściowy (obraz) x i odpowiadający mu poprawny sygnał odpowiedzi d . W czasie uczenia na wejście sieci podawane są po kolei sygnały x należące do zbioru uczącego. Po podaniu sygnału x obserwowana jest odpowiedź sieci y i porównywana z odpowiedzią poprawną d systemu (stąd nazwa: uczenie z nauczycielem). Otrzymana

wartość różnicy $y - d$ wykorzystywana jest do korekty wartości wektora wag w , która dokonywana jest w taki sposób, aby zmniejszyć różnicę $y - d$. Po dokonaniu korekty wag na wejście sieci podawany jest następny sygnał wejściowy i ponownie dokonywana jest korekta wag. Proces ten jest kontynuowany, aż do wyczerpania wszystkich sygnałów uczących, po czym proces uczenia powtarza się dla tych samych sygnałów ze zbioru uczącego, najlepiej podanych w innej kolejności. Etap uczenia, w którym podane są wszystkie sygnały zbioru uczącego nazywa się *cyklem*. Proces uczenia trwa, aż zsumowane w cyklu różnice pomiędzy odpowiedziami sieci, a sygnałami oczekiwanymi nie opadną poniżej wartości przyjętej arbitralnie przez użytkownika. Najczęściej stosowaną metodą korekty wag jest odmiana metody gradientowej zwana *metodą propagacji wstecznej błędu*, którą poniżej krótko omówiono dla przypadku sieci dwuwarstwowej.



Rys.5

Oznaczenia przyjęto jak na schemacie sieci przedstawionym na rys. 5. Oznaczając pożądaną sygnał wyjściowy jako

$$\mathbf{d} = [d_1 \ d_2 \ \dots \ d_k]^t$$

błąd sieci dla jednego wektora wejściowego x_l , $l = 1, 2, \dots, p$, wynosi

$$E_l = \frac{1}{2} \sum_{k=1}^K (d_{lk} - z_{lk})^2 = \frac{1}{2} \|\mathbf{d}_l - \mathbf{z}_l\|^2, \quad (3)$$

gdzie \mathbf{d}_l i \mathbf{z}_l , oznaczają odpowiednio pożądane i aktualne sygnały wyjściowe, gdy na wejście podano obraz \mathbf{x}_l .

Idea metody propagacji wstecznej polega na przesuwni wektora wag, po każdorazowym podaniu wektora wejściowego, w kierunku powodującym zmniejszanie błędu (3). Do wyznaczenia przyrostu wag stosuje się metodę gradientową szukania ekstremum funkcji. Stosując ją do błędu (3) otrzymujemy ogólne wzory na korekcję wag w warstwie pierwszej i drugiej

$$\begin{aligned} \Delta w_{ki} &= -\eta \frac{\partial E}{\partial w_{ki}}, & k &= 1, 2, \dots, K, & j &= 1, 2, \dots, J \\ \Delta v_{ij} &= -\eta \frac{\partial E}{\partial v_{ij}}, & j &= 1, 2, \dots, J-1, & i &= 1, 2, \dots, I \end{aligned} \quad (4)$$

gdzie η jest stałym współczynnikiem, błąd E jest zdefiniowany wzorem (3), wskaźnik zaś pominięto dla uproszczenia zapisu.

Zdefiniujmy sygnały błędu delta jako

$$\delta_{zk} \stackrel{\text{def}}{=} -\frac{\partial E}{\partial net_k} \quad \delta_{yj} \stackrel{\text{def}}{=} -\frac{\partial E}{\partial net_j}$$

gdzie net_i oznacza pobudzenie łączne neuronu i . Po prostych przekształceniach otrzymuje się wyrażenia opisujące zasadę korekcji wag w obu warstwach

$$\begin{aligned} w'_{kj} &= w_{kj} + \eta \delta_{zk} y_j, & \delta_{zk} &= (d_k - z_k) f'(net_k), \\ k &= 1, 2, \dots, K, & j &= 1, 2, \dots, J, \\ v'_{ji} &= v_{ji} + \eta \delta_{yj} x_i, & \delta_{yj} &= f'_j(net_j) \sum_{k=1}^K \delta_{zk} w_{kj}, \\ j &= 1, 2, \dots, J-1, & i &= 1, 2, \dots, I, \end{aligned} \quad (5)$$

gdzie $f'(net)$ oznacza pochodną funkcji aktywacji; dla bipolarnej funkcji aktywacji danej wzorem (2) mamy

$$f'(net) = \frac{2 \exp(-net)}{[1 + \exp(-net)]^2} = \frac{1}{2} (1 - y^2)$$

Jak wynika z wzoru (5), korekcja wag dochodzących do j -tego neuronu w warstwie pierwszej jest proporcjonalna do sumy ważonej wszystkich wartości δ w warstwie wyjściowej. Wagi w_{kj} , $k = 1, 2, \dots, K$, w warstwie wyjściowej, które wpływają na błąd δ_{yj} , oraz wagi v_{ji} , $i = 1, 2, \dots, I$, na które ma wpływ błąd δ_{yj} , zostały wyróżnione na rys. 5 pogrubionymi liniami. Gdy charakter klasyfikowanych danych jest znany tylko w niewielkim stopniu, strukturę sieci (liczbę warstw i

liczby neuronów w warstwach) dobiera się empirycznie. W takim przypadku można do uczenia sieci zastosować (co robi się coraz częściej) algorytm genetyczny, który oprócz doboru wag, może także optymalizować strukturę sieci.

Po nauczeniu, sieć może rozpocząć właściwe działanie. Nauczona sieć oprócz poprawnych reakcji na sygnały zbioru uczącego powinna także reagować poprawnie na sygnały nie podawane jej w czasie nauki. Taka własność sieci nosi nazwę *generalizacji*.

Podsumowując, sztuczne sieci neuronowe:

- zbudowane są z dużej liczby wzajemnie połączonych prostych jednostek – neuronów
- wykonują obliczenia równoległe, jednocześnie w całej swej strukturze,
- przygotowanie ich do działania odbywa się poprzez uczenie, a nie poprzez programowanie,
- mają zastosowania w zadaniach, które potrafią rozwiązywać ludzie, a bardzo trudno do ich rozwiązania zaprogramować komputery.

Korzyści z zastosowania sieci neuronowych:

- zastąpienie programowania przez uczenie,
- duża szybkość działania (w przypadku realizacji sprzętowych).

Wady zastosowania sieci neuronowych:

- czasami bardzo długa nauka bez osiągnięcia dostatecznie małego błędu lub zła generalizacja,
- trudność w zrozumieniu przez człowieka algorytmu działania nauczonej sieci (postacią algorytmu sieci jest złożona funkcja nieliniowa, a nie zrozumiałe dla człowieka reguły przyczynowo–skutkowe).

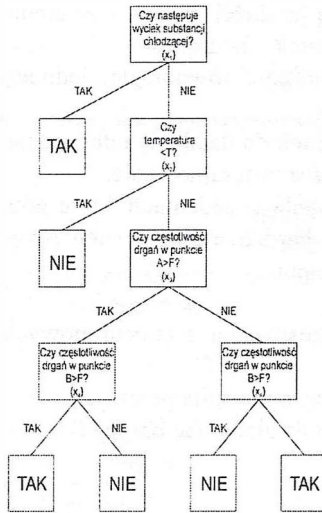
Oprócz zastosowania sieci neuronowych warstwowych uczonych metodą propagacji wstecznej lub algorytmami genetycznymi zastosowanie w omawianym problemie diagnostyki mogą mieć sieci samoorganizujące się, które same, bez udziału nauczyciela dokonują klasyfikacji na podstawie podobieństwa cech.

2.2 Algorytmy genetyczne

Algorytmy genetyczne są metodami umożliwiającymi znalezienie optymalnych algorytmów rozwiązujących postawione zadania [3,4]. Zasada działania algorytmu genetycznego polega na początkowym utworzeniu pewnej liczby różnych algorytmów rozwiązujących postawione zadanie, a następnie na przeprowadzeniu wielu cykli uczenia. W czasie cyklu oceniana jest jakość algorytmów, wybierane są algorytmy najlepsze (selekcja) i tworzone są z nich pary, na podstawie których tworzone są algorytmy nowe (krzyżowanie) oraz dokonywane są losowe zmiany algorytmów (mutacja). Dzięki tym operacjom w kolejnych cyklach otrzymuje się algorytmy coraz lepsze i dotyczy to zarówno całej populacji jak i algorytmu w danej chwili najlepszego. Po przeprowadzeniu pewnej liczby cykli proces uczenia się kończy i z populacji algorytmów wybiera się algorytm najlepszy.

2.3. Uczenie drzew decyzyjnych

Drzewo decyzyjne jest algorytmem który na podstawie podanych dyskretnych wartości atrybutów generuje konkluzję będącą decyzją typu *tak/nie* [12,9]. Liście drzewa decyzyjnego reprezentują konkluzje, pozostałe węzły reprezentują atrybuty, a krawędzie reprezentują wartości atrybutów. Drzewa decyzyjne są specyficznym zapisem funkcji logicznych. Przykład drzewa przedstawiono na rys. 6.



Rys. 6 Przykład drzewa decyzyjnego. Drzewo generuje sygnał alarmu w pewnym urządzeniu zależności od wartości binarnej czterech parametrów: występowania wycieku substancji chłodzącej (x_1), temperatury $< T(x_2)$, częstotliwości drgań w punkcie pomiarowym $A > F(x_3)$ oraz częstotliwości drgań w punkcie pomiarowym $B > F(x_4)$

Celem uczenia jest uzyskanie drzewa decyzyjnego na podstawie zadanej pewnej liczby przykładów (obrazów) zawierających wartości atrybutów danych i odpowiadające im decyzje. Uczenie drzew decyzyjnych należy do metod uczenia, w których uzyskiwana wiedza wyrażona jest w postaci jawnej – a ponadto typowe metody uczenia drzew decyzyjnych są jednorazowe (nie są iteracyjne).

Podstawową zasadą uczenia indukcyjnego jest tzw. zasada *brzytwy Ockhama* mówiąca, że: „bytów nie należy mnożyć więcej niż potrzeba” co tutaj można wyrazić jako: najbardziej prawdopodobną hipotezą jest hipoteza najprostsza, która jest zgodna z największą liczbą obserwacji. W myśl tej zasady na każdym etapie uczenia (tworzenia) drzewa decyzyjnego poszukuje się atrybutu, który ma największy wpływ na podejmowaną decyzję (np. jego określona wartość pozwala na jednoznaczną decyzję). Drzewo decyzyjne tworzy się począwszy od korzenia. Na początku uczenia atrybut o największym wpływie staje się korzeniem budowanego drzewa, a następnie wyznaczane są atrybuty tworzące węzły potomne (będące korzeniami poddrzew). Przy wyznaczaniu kolejnego węzła potomnego bierze się pod uwagę tylko te obrazy uczące, których atrybuty zapewniają dojście do tego węzła.

Najczęściej we wczesnych etapach tworzenia drzewa decyzyjnego nie ma zmiennych, których określona wartość pozwala na jednoznaczną decyzję. Wtedy jako miarę wpływu

zmiennej na podejmowaną decyzję przyjmuje się entropię rozkładu prawdopodobieństwa decyzji wynikających z przyjęcia tej zmiennej jako węzła drzewa decyzyjnego (algorytm ID3).

Entropia rozkładu prawdopodobieństwa p_i , $i = 1, 2, \dots, n$ wynosi

$$E = - \sum_{i=1}^n p_i \log_2 p_i$$

W przypadku nieznanymi prawdopodobieństw p_i można je zastępować liczebnościami zdarzeń.

Przyjmijmy oznaczenia: $i = 1, 2, \dots, I$ – numer atrybutu, $j = 1, 2, \dots, J_i$ – indeks wartości przyjmowanych przez i -ty atrybut, $k = 0, 1$ – możliwe decyzje, L – liczba elementów zbioru uczącego, L_{ijk} – liczba decyzji o wartości k , gdy wybrano wartość o indeksie j atrybutu o numerze i (patrz tablica na rysunku 7a).

Zastępując prawdopodobieństwa liczebnościami, entropia rozkładu prawdopodobieństwa decyzji, gdy wybrano wartość j atrybutu i wynosi

$$E(D_{ij}) = - \frac{L_{ij0}}{L_{ij0} + L_{ij1}} \log_2 \frac{L_{ij0}}{L_{ij0} + L_{ij1}} - \frac{L_{ij1}}{L_{ij0} + L_{ij1}} \log_2 \frac{L_{ij1}}{L_{ij0} + L_{ij1}} \quad (6)$$

gdzie D_{ij} oznacza poddrzewo decyzyjne wychodzące z atrybutu i wzdłuż krawędzi j (po wyborze wartości j atrybutu i).

Podobnie, entropia rozkładu prawdopodobieństwa decyzji gdy wybrano atrybut i jest ważoną sumą entropii dla poszczególnych wartości atrybutów:

$$E(D_i) = \sum_{j=1}^{J_i} \frac{L_{ij0} + L_{ij1}}{L} E(D_{ij}) \quad (7)$$

gdzie D_i jest poddrzewem decyzyjnym o korzeniu będącym atrybutem i .

Zasada wyboru atrybutu: wybierany jest atrybut, dla którego entropia dana wzorem (7) przyjmuje wartość minimalną. Gdy pewna wartość wybranego atrybutu jednoznacznie określa wartość funkcji, to wychodząca z tego atrybutu krawędź drzewa opisana tą wartością kończy się liściem drzewa decyzyjnego, a w przeciwnym wypadku kończy się innym atrybutem o minimalnej entropii.

W tablicy na rys. 7a podany jest przykładowy zestaw danych uczących.

Na podstawie wzoru (6) będzie

$$E(D_{11}) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1.0000$$

i następnie

$$E(D_{12}) = 0.6500, \quad E(D_{13}) = 1.0000$$

oraz

$$E(D_{21}) = 0.9710, \quad E(D_{22}) = 0.8631,$$

$$E(D_{31}) = 0.5917, \quad E(D_{32}) = 0.9183, \quad E(D_{33}) = 1.0000$$

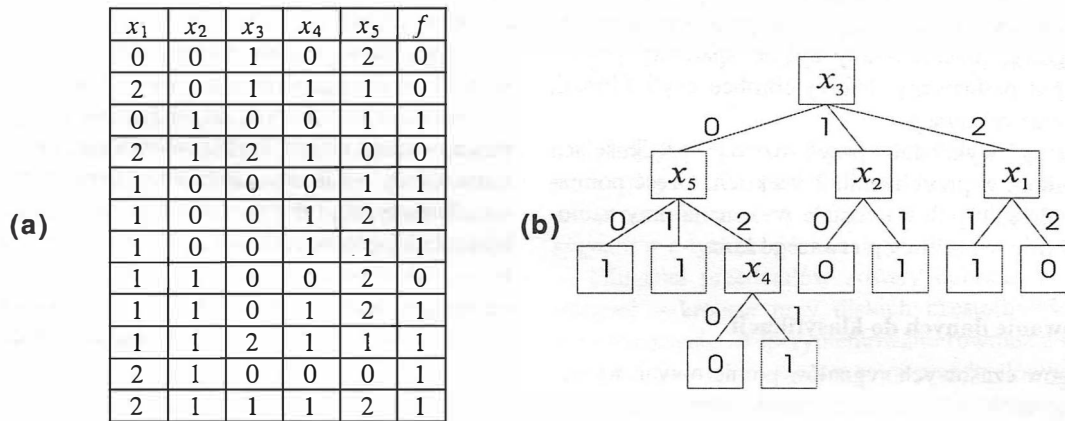
$$E(D_{21}) = 0.9710, \quad E(D_{22}) = 0.8631,$$

$$E(D_{21}) = 1.0000, \quad E(D_{22}) = 0.7219, \quad E(D_{33}) = 1.9710$$

Podstawiając otrzymane wyniki do wzoru (7) otrzymujemy:

$$E(D_1) = \frac{2}{12} \cdot 1 + \frac{6}{12} \cdot 0.65 + \frac{4}{12} \cdot 1 = 0.8250$$

i następnie $E(D_2) = 0.9080$, $E(D_3) = 0.7414$, $E(D_4) = 0.8720$. Stąd widać, że korzeniem tworzonego drzewa powinien być atrybut x_3 , ponieważ poddrzewo decyzyjne z korzeniem x_3 posiada najmniejszą entropię. Postępując analogicznie dla kolejnych poddrzew uzyskuje się drzewo decyzyjne przedstawione na rys. 7b.



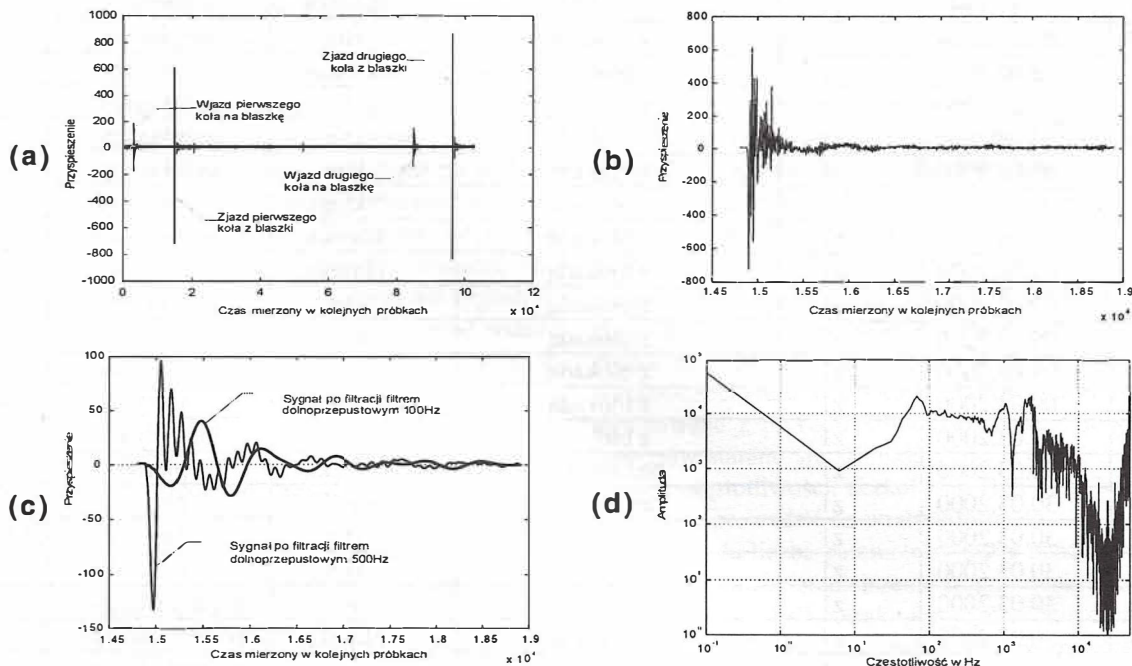
Rys.7 Przykład uczenia drzewa decyzyjnego: (a) Zbiór uczący. Stosując przyjęte oznaczenia mamy: $I = 5$, $J_1 = J_3 = J_5 = 3$, $J_2 = J_4 = 2$, $L = 12$, oraz np. $L_{110} = 1$, $L_{111} = 1$, $L_{120} = 1$, $L_{121} = 5$, $L_{130} = 2$, oraz $L_{131} = 2$, (b) Uzyskane drzewo decyzyjne, gdy w czasie uczenia ustalano kolejności wyboru atrybutów według minimalnych entropii

3. Przetwarzanie danych wibroakustycznych

3.1. Przebieg pomiarów

Pomiary przeprowadzono na specjalnie do tego celu przystosowanym izolowanym odcinku pomiarowym. Fragment

toru o długości około 160 cm posadowiony został na betonowym bloku o grubości około 1m. Długość odcinka została dobrana w ten sposób, aby uniemożliwić jednoczesne pojawienie się na nim obu osi wózka. W połowie odcinka na jednym toku szynowym umieszczono element wymuszający



Rys. 8 Wybrane sygnały pomiarowe: (a) przebieg w czasie sygnału pomiarowego nr2 (patrz tablica 1), (b) wybrany fragment przebiegu czasowego z rys. (a) - zjazd pierwszego koła z blaszki, (c) przebieg czasowy z rysunku (b) po filtracji dolnoprzepustowej 500 i 100 Hz, (d) widmo przebiegu czasowego z rys. (b) (8192 próbek)

– blaszkę o wymiarach 100x30x8 mm lub klin o końcowej wysokości ok. 15 mm. Czujniki pomiarowe były przytwierdzone do szyny w okolicach spadku koła z blaszki lub klina. Użyto trzech typów czujników: indukcyjnego, piezoelektrycznego oraz pojemnościowego. Wszystkie trzy czujniki mierzyły przyspieszenia pionowe w szynie. Dodatkowo zastosowano fotooptyczny rejestrator prędkości wagonu.

Każdy pomiar polegał na przetaczaniu wagonu typu 415W pchanego przez lokomotywę przez odcinek pomiarowy i rejestracji sygnału mierzonego przez czujnik w trakcie przejazdu pierwszego wózka wagonu. Sygnał pomiarowy w postaci analogowej przekazywany był do aparatury pomiarowej, gdzie był poddawany dalszej obróbce czyli filtracji, próbkowaniu oraz rejestracji.

Eksperymenty wykonano przy różnych prędkościach przejazdu wagonu, w przybliżeniu 2 i 4 km/h. Część pomiarów – w przybliżeniu połowa została wykonana przy zablokowanej sprężynie w maźnicy pierwszego koła.

3.2. Przygotowanie danych do klasyfikacji

Z przebiegów czasowych sygnałów pomiarowych wybrane zostały fragmenty, w których początek odpowiada chwili

wymuszenia (spadku pierwszego koła wózka z blaszki lub klina), czasie trwania 0.16384 s. i różnej liczbie próbek w zależności od częstotliwości próbkowania sygnału, odpowiednio 16384 próbki dla 100 kHz, 8192 dla 50 kHz oraz 3280 dla 20 kHz. Długości fragmentów zostały dobrane w taki sposób, by z jednej strony obejmowały tylko okres tłumienia drgań aż do ich całkowitego zaniku, z drugiej zawierały dostateczną liczbę próbek sygnału pozwalającą na wyznaczenie dokładnych wartości widma, zwłaszcza dla niskich częstotliwości. Liczba okresów drgań niskoczęstotliwościowych w przedziale tłumienia często jednak okazywała się zbyt mała dla wiarygodnej oceny odpowiadających im wartości widma, dlatego dalej analizowane są tylko drgania o częstotliwościach większych od 20 Hz. Na wybór liczby próbek miała również wpływ szybkość obliczania transformaty FFT, która zależy od liczby dzielników pierwszych liczby próbek transformaty, stąd dobrym rozwiązaniem było przyjęcie liczby próbek będącej całkowitą potęgą liczby 2. Z ogólnej liczby sygnałów pomiarowych odrzucono te, które były obciążone dużym, widocznym błędem związanym np. z przesunięciem blaszki/klina lub przeskalowaniem aparatury pomiarowej.

Zestawienie pomiarów wykorzystywanych do uczenia i testowania sieci neuronowych i drzew decyzyjnych

Tablica 1

Lp.	data pomiaru	typ zdarzenia	typ uszkodzenia	rodzaj wymuszenia	typ czujnika	prędkość [km/h]
1	19.11.1999	z1	bez blokady	blaszka	piezo	2
2-6	19.11.1999	z1	bez blokady	blaszka	piezo	4
7-9	19.11.1999	z1	z blokadą	blaszka	piezo	4
10-13	19.11.1999	z1	z blokadą	blaszka	piezo	2
14-15	08.03.2000	z1	z blokadą	klin	indukcyjny	2
16-18	08.03.2000	z1	z blokadą	blaszka	indukcyjny	2
19-22	08.03.2000	z1	z blokadą	blaszka	indukcyjny	4
23-24	08.03.2000	z1	z blokadą	klin	indukcyjny	4
25-27	08.03.2000	z1	z blokadą	klin	indukcyjny	2
28	09.03.2000	z1	z blokadą	klin	indukcyjny	2
29-31	09.03.2000	z1	z blokadą	blaszka	indukcyjny	2
32-36	09.03.2000	z1	z blokadą	blaszka	indukcyjny	4
37	09.03.2000	z1	z blokadą	klin	indukcyjny	2
38-40	09.03.2000	z1	z blokadą	klin	indukcyjny	4
41	09.03.2000	z1	z blokadą	klin	indukcyjny	0
42	09.03.2000	z1	z blokadą	klin	indukcyjny	2
43	09.03.2000	z1	z blokadą	blaszka	indukcyjny	2
44-46	30.03.2000	z1	bez blokady	blaszka	piezo	2
47-49	30.03.2000	z1	bez blokady	blaszka	piezo	4
50-52	30.03.2000	z1	bez blokady	klin	piezo	2
53-57	30.03.2000	z1	bez blokady	klin	piezo	4
58-60	30.03.2000	z1	bez blokady	blaszka	pojemnościowy	2
61-64	30.03.2000	z1	bez blokady	blaszka	pojemnościowy	4
65-67	30.03.2000	z1	bez blokady	klin	pojemnościowy	2
68-72	30.03.2000	z1	bez blokady	klin	pojemnościowy	4

W sumie uzyskano 72 przykładowe sygnały ułożone w kolejności chronologicznej (pomiar z 19.11.1999, 8-9.03.2000). Sporządzono bazę fragmentów sygnałów pomiarowych, w której oprócz samego sygnału znalazły się dodatkowe atrybuty takie, jak blokada lub brak blokady sprężyny w zawieszeniu, prędkość przejazdu wagonu (2 lub 4 km/h), rodzaj wymuszenia (blaszka, klin), rodzaj czujnika (piezoelektryczny, indukcyjny lub pojemnościowy) oraz typ zdarzenia (w1, z1, w2, z2) oznaczający odpowiednio wjazd i zjazd pierwszego koła na/z wymuszenie/a oraz wjazd i zjazd drugiego koła). W przypadku ostatniego atrybutu ograniczono się na tym etapie do zjazdu pierwszego koła. Ograniczono się również do jednej klasy uszkodzeń – pęknięcia sprężyny w zawieszeniu, symulowanego jej zablokowaniem. W ten sposób każdy z 72 przykładów należy do jednej z dwóch klas: z blokadą lub bez blokady zawieszenia, odpowiednio klasy „0” i „1”. Zestawienie wszystkich stosowanych w uczeniu i testowaniu przykładów przedstawiono w tabelicy 1.

Ze względu na małą liczbę przykładowych pomiarów ograniczono się do eksperymentów na całym ich zbiorze. Gdy dokonana zostanie większa liczba pomiarów możliwe stanie się zbadanie jakości klasyfikacji w zależności od zastosowanego typu czujnika lub wymuszenia.

3.3. Parametryzacja sygnałów pomiarowych

Wyznaczone fragmenty przebiegów czasowych sygnałów można już poddać klasyfikacji, jednak ze względu na dużą złożoność problemu i małą liczbę przykładów klasyfikacja w oparciu o nieprzetworzone przebiegi czasowe jest praktycznie niemożliwa. Spowodowane jest to nadmiarem informacji i tym samym dużym prawdopodobieństwem oparcia się klasyfikacji na cechach wspólnych dla zbioru przykładowych sygnałów ale zupełnie nieistotnych dla problemu tzn. nie zapewniających prawidłowej klasyfikacji obrazów nieznanymi. Konieczne jest więc wprowadzenie dodatkowej wiedzy do rozwiązywanego problemu klasyfikacyjnego czyli wyboru parametrów lub cech istotnych dla tego zadania. Wybór takich cech może być dokonany drogą ewolucyjną na podstawie dużej liczby podobnych, wcześniej rozwiązanych problemów, jednak na tym etapie może opierać się jedynie na intuicji projektanta lub literaturze związanej z tematem. Parametry powinny być możliwe do wyznaczenia dla sygnałów jeszcze nie znanych oraz maksymalnie niezależne od wszelkich przesunięć i przeskalowań sygnałów, związanymi np. z błędami aparatury pomiarowej.

Parametry sygnałów wibroakustycznych można podzielić na dwie grupy: parametry w dziedzinie czasu i w dziedzinie widma. Parametry w dziedzinie czasu okazały się w tym wypadku trudne do wyznaczenia ze względu na problemy z określeniem początku zdarzenia – zjazdu koła z blaszki lub klina oraz występowania w widmie sygnału składowych o bardzo niskiej częstotliwości bliskich składowej stałej. Opracowano wprowadzić algorytm wyznaczenia punktu początkowego zdarzenia czyli uderzenia koła o szynę, jednak z powodu znacznej zawartości drgań wysokoczęstotliwościowych w mocy sygnału jak również występowania przeduderzeń spowodowanych podnoszeniem się klina nie zawsze był on sku-

teczny, a fragmenty wycinane w sposób przybliżony z różnych przebiegów mogły być przesunięte względem siebie. Składowe niskoczęstotliwościowe w okolicach 0.5 Hz utrudniały analizę przejść przez zero, a po odfiltrowaniu prowadziły do utraty informacji.

W tej sytuacji skupiono się na parametrach widma sygnału. Wyznaczono średnie amplitudy w 8 przedziałach widma sygnału w zakresie 20–1000 Hz znormalizowanych względem średniej amplitudy całego widma według wzoru

$$P_i = \frac{n_{i2} - n_{i1}}{N_2 - N_1} \frac{\sum_{j=n_{i1}}^{n_{i2}} |X_j|}{\sum_{j=N_1}^{N_2} |X_j|} \quad (8)$$

gdzie P_i , $i = 1, 2, \dots, 8$ jest wartością i -tego parametru, X_j , $j = N_1, N_1+1, \dots, N_2$ jest widmem pomierzonego sygnału, n_{i1} , n_{i2} oznaczają numery pierwszej i ostatniej próbki widma sygnału w i -tym przedziale częstotliwości.

Długości przedziałów zostały dobrane w skali logarytmicznej – krótsze przy niskich częstotliwościach, dłuższe przy wyższych. Eksperymentowano również z większą liczbą przedziałów oraz z ciągiem grup przedziałów pokrywających zakres częstotliwościowy widma o stopniowo rosnącej gęstości. Zrezygnowano na tym etapie z innych parametrów zalecanych w literaturze np. momentów widmowych.

Ocena stopnia złożoności problemu

Redukcja sygnału do małej liczby parametrów (cech) bardziej istotnych ale wybranych a priori jako charakterystycznych w całej klasie podobnych problemów nie pozwala jeszcze na optymalne rozwiązanie konkretnego problemu klasyfikacji uszkodzeń układów biegowych. Przydatne jest obliczenie poziomu istotności parametrów w oparciu o matematyczny model badanego układu lub dokonanie automatycznej klasyfikacji bez konieczności wcześniejszej weryfikacji parametrów. Określenie poziomu istotności parametrów jest o tyle trudne, że dotyczy korelacji wartości parametrów w obrębie wszystkich przykładów znanych i nieznanymi należących do jednej klasy, co przekłada się zdolność systemu do uogólniania (generalizacji). W badanym problemie takiej weryfikacji można poddać niektóre parametry częstotliwościowe, zwłaszcza dotyczące wyższych częstotliwości w granicach 100–1000 Hz, kojarzonych zazwyczaj z drganiami strukturalnymi koła lub szyny. Nie mając jednak dokładnego, niemal idealnego modelu układu niezwykle trudno jest stwierdzić, czy zróżnicowanie wartości takich drgań nie ma źródła w badanych uszkodzeniach poprzez np. składowe harmoniczne, które w niskich częstotliwościach nie mogą być analizowane z powodu krótkiego okresu tłumienia. Z tych powodów starano się arbitralnie nie zawężać przedziału badanych częstotliwości, aczkolwiek zbyt duża liczba parametrów może prowadzić do utraty informacji.

Zbyt mała liczba parametrów może zwiększyć złożoność problemu poprzez zmniejszenie stopnia separowalności zbioru przykładów. Jeśli potraktujemy przykłady do nauki złożone z N parametrów jako punkty w przestrzeni N -wymiarowej, to klasy przykładów są liniowo separowalne, gdy istnieją takie hiperpłaszczyzny $N-1$ wymiarowe (po jednej dla każdego z 2 klas), które rozdziela punkty należące do różnych klas. Mniejsza liczba parametrów zawęża przestrzeń przykła-

dów do mniejszej liczby wymiarów i tym samym zmniejsza liczbę możliwości podziału tej przestrzeni czyli prawdopodobieństwo osiągnięcia separowalności liniowej.

Nie ograniczając się tylko do separowalności liniowej jedną z bardziej znanych miar separowalności jest kryterium odległościowe [14]:

$$Q = \frac{\min_{ij} D_{ij}}{\max_i d_i} > 1 \quad (9)$$

gdzie D_{ij} – odległość pomiędzy środkami ciężkości klas, d_i średnia odległość pomiędzy elementami klasy i .

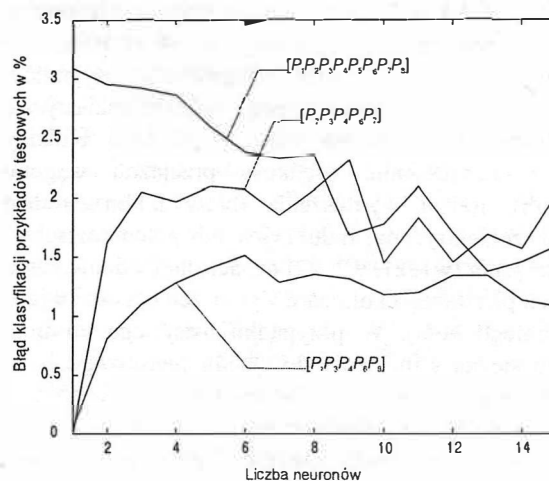
Spełnienie tego kryterium oznacza dobrą separowalność. Obliczono wartość kryterium (9) dla wszystkich możliwych 255 grup parametrów. W przypadku grupy $[P_1 P_2 \dots P_8]$ złożonej ze wszystkich 8 parametrów $Q=0.8034$. Lepsze wyniki uzyskano dla innych grup parametrów, paradoksalnie najlepszy, $Q=1.2627$ dla grupy $[P_7 P_8]$ odpowiadającej przedziałom najwyższych częstotliwości. Separowalność jest jednak tylko pewną miarą złożoności problemu, a tym bardziej jej kryteria. Bardziej empiryczną metodą sprawdzenia istotności parametrów jest przeprowadzenie nauki i testowania sieci neuronowej dla wszystkich 255 możliwych do utworzenia grup parametrów. Po przeprowadzeniu 10 powtórzeń eksperymentu najlepsza okazała się grupa 5 parametrów $[P_2 P_3 P_4 P_6 P_7]$.

3.5. Klasyfikacja uszkodzeń przy użyciu sztucznych sieci neuronowych

Na całość klasyfikacji złożyły się trzy etapy: podział zbioru przykładów na dwa podzbiory: do nauki i testowania, nauka oraz testowanie sieci neuronowej. Ze względu na małą liczbę przykładów pozyskanych z sygnałów pomiarowych w różnych warunkach, tzn. przy różnych prędkościach, różnych rodzajach czujników i różnych typach wymuszenia podział na podzbiory do nauki i testowania musiał zostać przeprowadzony w sposób zapewniający w obu podzbiorach odpowiednią liczbę przykładów z obu klas i posiadających wszystkie wartości wyżej wymienionych atrybutów. Dobrym rozwiązaniem okazał się arbitralny podział przykładów pod względem numeracji w tablicy 1 na parzyste i nieparzyste: odpowiednio 36 przykładów do testowania i 36 do nauki. Chronologiczne uporządkowanie pomiarów w tabeli sprawia, że atrybuty pomiarów o kolejnych numerach niewiele się pomiędzy sobą różnią.

Do klasyfikacji użyto sieci dwuwarstwowej, jednokierunkowej, zbudowanej z neuronów o ciągłej bipolarnej funkcji aktywacji danej wzorem (2). Sieć dwuwarstwowa wystarcza do poprawnej klasyfikacji dowolnego problemu o skończonej liczbie przykładów. Na wejście sieci podawano parametry $P_1 P_2 \dots P_8$. W drugiej warstwie – wyjściowej znajdował się jeden neuron, który po dyskretyzacji zwracał jedną z dwóch wartości: 0 dla klasy „z blokadą” i 1 dla klasy „bez blokady”.

Praktycznie jedyną istotną zmienną, od której mogła zależeć jakość klasyfikacji była liczba neuronów w pierwszej warstwie – tzw. warstwie ukrytej. Nie mogła być zbyt mała, gdyż nie zapewniałaby wtedy wstępnej separacji klas, a tym

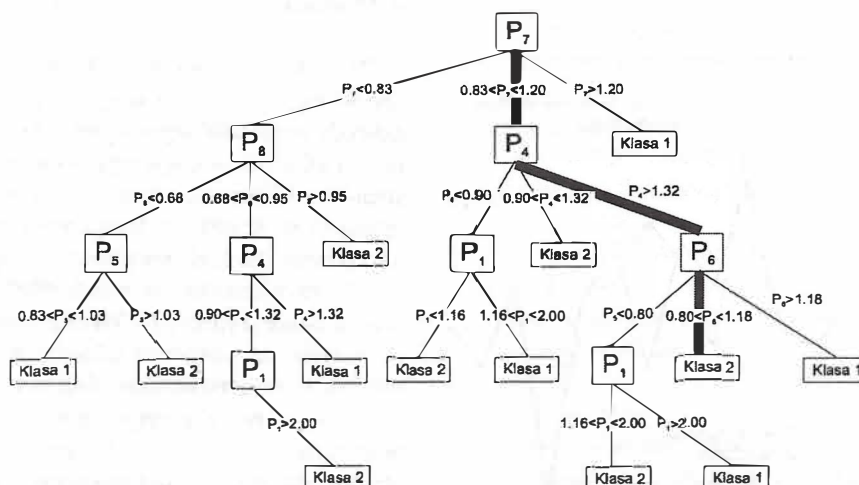


Rys. 9 Błędy testu dla trzech wybranych grup parametrów wejściowych w zależności od liczby neuronów w warstwie ukrytej

samym nie zapewniłaby nauczania sieci i prowadziłyby do systematycznych błędów klasyfikacji. Zbyt duża prowadziła by do uwypuklenia mniej istotnych szczegółów w rozkładzie wartości parametrów i tym samym spadku zdolności sieci do uogólniania tzn. poprawnej klasyfikacji przykładów testowych, nieuwzględnionych w czasie nauki. Każdy neuron w pierwszej warstwie można uważać za jedną hiperpłaszczyznę separującą, im jest ich więcej, tym łatwiej wydzielić pewne skupiska punktów należących do jednej klasy tak, by neuron w drugiej warstwie mógł je ostatecznie zgrupować w obszary spójne, odpowiadające klasom.

Przeprowadzono uczenie sieci przy użyciu algorytmu Levenberga–Marquardta, będącym rozszerzeniem metody propagacji wstecznej [6]. Uczenie przeprowadzono dla kilku grup parametrów P_i , dla różnych liczb neuronów w pierwszej warstwie zmieniających się od 1–15, powtarzając każdy proces uczenia 600-krotnie. Otrzymane średnie wyniki testowania nauczonych sieci przedstawiono na rys. 9. Jak widać z rysunku średnie błędy testu wahają się od 0 do 3%, co sprowadza się do przeciętnie mniej niż jednego błędu na 36 przykładów testujących.

Okazało się też, że bardzo dobre wyniki klasyfikacji uzyskano już przy jednym neuronie w pierwszej warstwie. Na 600 prób większość zakończyła się zerowym błędem nauki i klasyfikacji, co świadczy o liniowej separowalności badanego zbioru przykładów. Należy jednak pamiętać o tym, że przeprowadzono małą liczbę eksperymentów z tym samym wagonem, tym samym zestawem kołowym, w tym samym miejscu i tylko dla jednej klasy uszkodzeń. Ponadto wagon był nieobciążony, co w przeciwnym wypadku może prowadzić do znacznego zwiększenia amplitud drgań zwłaszcza w dziedzinie górnych częstotliwości i tym samym zwiększyć złożoność zadania. Dlatego dobrym rozwiązaniem jest przyjęcie większej liczby neuronów w warstwie ukrytej.



Rys. 10 Drzewo decyzyjne o trzech przedziałach wartości parametrów. Grubą linią zaznaczono przebieg klasyfikacji pomiaru numer 1 (w tablicy 1), którego wartości parametrów wynoszą: $P_1 = 2.50$, $P_2 = 3.74$, $P_3 = 2.41$, $P_4 = 1.74$, $P_5 = 1.25$, $P_6 = 1.17$, $P_7 = 0.93$, $P_8 = 0.67$. Klasa 1 oznacza uszkodzenie (zablokowana sprężyna), klasa 2 oznacza brak uszkodzenia

3.6. Klasyfikacja uszkodzeń przy zastosowaniu drzew decyzyjnych

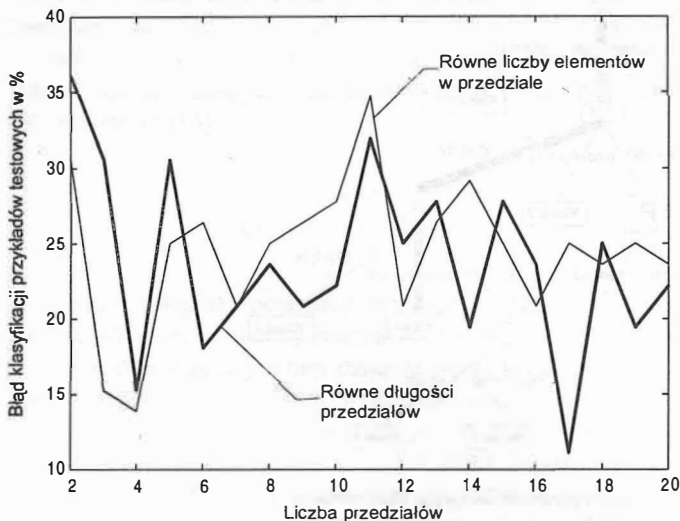
Celem wykorzystania drzew decyzyjnych jako metody klasyfikacji uszkodzeń było sprawdzenie przydatności metody oraz próba oceny jakości parametrów opisujących każdy pomiar. Dodatkową zaletą drzew decyzyjnych jest możliwość wydobywania reguł decyzyjnych przydatnych w celu weryfikacji systemu. Eksperyment składał się z 4 etapów: wyboru przykładów do budowy i testowania drzewa, dyskretyzacji parametrów, budowy i testowania drzewa decyzyjnego.

W pierwszym etapie dokonano podziału zbioru przykładów według numeracji w tablicy 1 na parzyste do konstrukcji drzewa i nieparzyste do testowania.

Drugi etap polegał na wyznaczeniu skończonej liczby przedziałów wartości każdego parametru i zastąpieniu ciągłych wartości danego parametru numerami przedziałów, w których dane wartości były zawarte. W przypadku 8 parametrów wyznaczonych według wzoru (8) każdy przedział ograniczony był dolną i górną amplitudą widma. Cały zbiór przedziałów pokrywał wszystkie wartości danego parametru dla wszystkich przykładów wykorzystanych do konstrukcji drzewa. Problemem okazał się sposób ustalenia granic każdego z przedziałów. Zastosowano dwie metody. Pierwsza polegała na wyznaczeniu przedziałów o równych długościach pokrywających wartości danego parametru dla całego zbioru 36 przykładów do budowy drzewa. Wadą tej metody jest pokrywanie dużej liczby przykładów przez jeden przedział, podczas gdy pozostałe są nieomal puste. Prowadzi to do małego zróżnicowania (względem przykładów) wartości każdego z parametrów po dyskretyzacji i w konsekwencji pojawia się wiele przykładów ze sobą sprzecznych tzn. o tych samych dyskretnych wartościach parametrów i różnych klasach. Sposobem na rozwiązanie tego problemu jest druga metoda polegająca na wyborze przedziałów o porównywalnych liczbach pokrywanych przez nie przykładów.

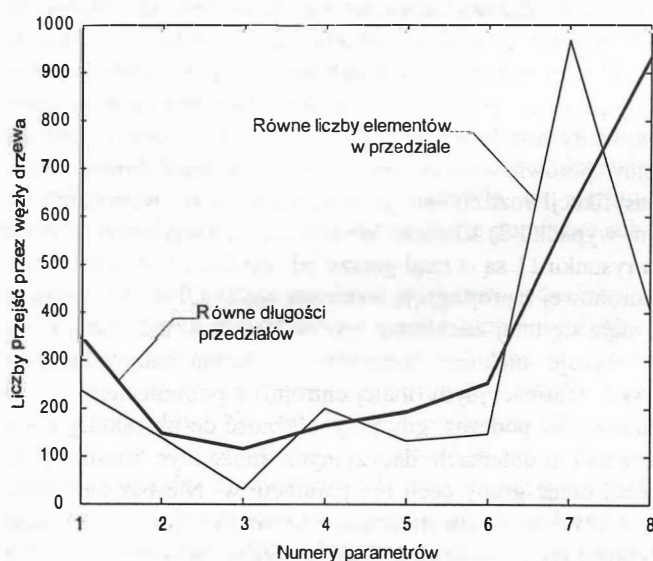
Do budowy drzewa użyto algorytmu ID3, zastępując prawdopodobieństwa wystąpienia poszczególnych wartości parametrów częstościami ich występowania. Liczbę przedziałów wartości parametrów próbowano wyznaczyć eksperymentalnie budując drzewa dla 2–20 przedziałów wartości wszystkich parametrów, przy czym liczba przedziałów wartości odpowiadała maksymalnemu stopniowi każdego węzła drzewa. Na rys. 10 przedstawiono drzewo uzyskane przy podziale wartości parametrów na trzy przedziały. Pewnym problemem okazało się występowanie przykładów sprzecznych w tym większej liczbie im mniej przyjęto przedziałów wartości dla każdego parametru. Jedynym rozsądnym rozwiązaniem w tym wypadku było przydzielenie ich do trzeciej klasy – nieokreślonej.

Do testowania użyto przykładów nie biorących udziału w konstrukcji drzewa, co przy ogólnie małej liczbie przykładów do budowy drzewa objawiało się często brakiem odpowiedniej struktury drzewa umożliwiającej klasyfikację niektórych przykładów testowych, których wartości parametrów wychodziły poza zakres uwzględniony przy budowie drzewa. Takie przykłady zostały uznane jako niesklasyfikowane, jednak dla celów porównawczych, przy obliczaniu uogólnionego błędu klasyfikacji rozdzielono je równomiernie po wszystkich (w tym wypadku 2) klasach. Wyniki klasyfikacji przedstawione na rysunku 11 są o rząd gorsze od wyników testowania sieci neuronowej z propagacją wsteczną błędów. Główną przyczyną wydaje się tutaj zachłanny sposób konstrukcji drzewa, który faworyzuje niektóre parametry o dużym indywidualnym zysku informacyjnym (małej entropii) z pominięciem innych parametrów podczas, gdy przynależność do określonej klasy w wielu problemach decyzyjnych może być rozstrzygana raczej przez grupy cech lub parametrów. Nie bez znaczenia była również utrata informacji ze względu na konieczność dyskretyzacji parametrów na potrzeby wykorzystania ich przez drzewa decyzyjne podczas, gdy wartości średnich amplitud widma są z natury ciągłe.



Rys. 11 Błędy klasyfikacji przykładowych testowych przy użyciu drzew decyzyjnych dla dwóch metod wyboru przedziałów wartości parametrów, w zależności od liczby przedziałów wartości wszystkich parametrów

Przy okazji testowania drzew decyzyjnych możliwe stało się przybliżone określenie istotności poszczególnych parametrów dla rozwiązania problemu klasyfikacji. Obliczono mianowicie liczbę przejść wszystkich przykładowych przez poszczególne węzły drzewa odpowiadające parametrom i zsumowano je dla drzew o 2 do 20 przedziałach wartości parametrów. Wyniki przedstawione na rysunku 12 wykazują dużą przydatność parametrów P_1 , P_7 i P_8 , a więc związanych z najniższymi i najwyższymi częstotliwościami z pominięciem środka widma. Wyniki te nie uwzględniają jednak w pełni istotności grup parametrów.



Rys. 12 Liczby przejść przez węzły odpowiadające kolejnym parametrom wszystkich testowanych drzew decyzyjnych przez wszystkie przykłady

4. Wnioski

W pracy omówiono zastosowanie sieci neuronowych i drzew decyzyjnych do diagnostyki uszkodzeń układów biegowych pojazdów szynowych. Do uczenia i testowania zarówno sieci neuronowych jak i drzew decyzyjnych wykorzystano po 36 przekształconych przebiegów czasowych przyspieszeń w szynie, z których 37 reprezentowało pojazdy uszkodzone (blokada sprężyny), a 35 pojazdów nieuszkodzone.

W przypadku zastosowania sieci neuronowych uzyskano bardzo dobre wyniki (0 – 3% błędów klasyfikacji przykładowych testowych nie biorących udziału w uczeniu). Takie wyniki wskazują, że uszkodzenia sprężyny w istotny sposób wpływają na mierzone przyspieszenia, co stanowi pierwsze potwierdzenie możliwości diagnozowania pojazdów na podstawie sygnałów wibroakustycznych zbieranych z szyny.

Wyniki uzyskane przez drzewa decyzyjne są znacznie gorsze (15 – 35% błędnych klasyfikacji w zależności od liczby przedziałów wartości poszczególnych parametrów (od 2 do 20)). Prawdopodobnymi przyczynami takiej dysproporcji wyników pomiędzy sieciami, a drzewami jest większy wpływ grup parametrów niż samych parametrów na jakość klasyfikacji, jak również ciągłe wartości parametrów, których dyskretyzacja w przypadku drzew decyzyjnych prowadzi do znacznej utraty informacji. Trudno jednak zdecydowanie odrzucić metodę drzew decyzyjnych z uwagi na małą liczbę dotychczas pozyskanych przykładowych jak i małą liczbę zbadanych parametrów, w tym również parametrów dyskretnych związanych np. z warunkami pomiarów, typem i konstrukcją wózka wagonu. Takie parametry mogą zostać uwzględnione w przyszłości o ile możliwy będzie ich pomiar w warunkach roboczych. Wiarygodność otrzymanych wniosków osłabia fakt, że dobre wyniki klasyfikacji dotyczą tylko jednej klasy uszkodzeń jednego wagonu.

Przygotowywane dalsze badania dla większej liczby pomiarów dla różnych uszkodzeń i różnych warunków pomiarów dadzą możliwość weryfikacji otrzymanych wniosków. Przy tych nowych pomiarach problem klasyfikacji stanie się znacznie bardziej złożony, jednak wtedy można będzie poprawić jakość zastosowanych algorytmów, stosując następujące rozszerzenia, które nie były do tej pory wykorzystywane:

- Zwiększenie liczby parametrów reprezentujących cechy podawanych na wejście sieci neuronowej. Przypuszczalnie wykorzystywane będą także parametry w dziedzinie czasu.
- Wykorzystanie modeli matematycznych zachowania się pojazdów uszkodzonych i nie uszkodzonych do optymalnej ekstrakcji cech.
- Wykorzystanie modeli matematycznych do uzupełnienia danych pomiarowych o dane uzyskane z symulacji komputerowych.
- Dekompozycja problemu klasyfikacji różnych typów uszkodzeń na oddzielne algorytmy wykrywające poszczególne typy uszkodzeń i algorytm decydujący o typie uszkodzeń lub ich braku.
- Zastosowanie do uczenia sieci metod algorytmów genetycznych optymalizujących oprócz wag, także strukturę sieci.

- Zastosowanie, oprócz drzew decyzyjnych, innych metod wydobywania wiedzy w postaci jawnej jak metodę generowania pokryć Michalskiego [8], czy metodę zbiorów przybliżonych Pawlaka [15].
- Budowa systemu ekspertowego opartego o sieci bayesowskie [11,5,7], do którego – wymagane byłyby dane statystyczne dotyczące występowania uszkodzeń oraz wiedza specjalistów dotycząca powiązań przyczynowo-skutkowych w pojazdach szynowych.
- Budowa jednego systemu hybrydowego wykorzystującego wymienione wyżej rozszerzenia.
- Wykonanie pomiaru przyspieszeń samego wagonu, możliwego obecnie w warunkach laboratoryjnych, co powinno dostarczyć sygnał niosący znacznie więcej informacji o uszkodzeniu.

Omawiane zagadnienie diagnostyki wydaje się być bardzo złożone. Dalsze pomiary wraz z rozszerzeniem zastosowanych algorytmów dadzą ostateczną odpowiedź dotyczącą możliwości zastosowania metody diagnostycznej, której istotną część stanowią algorytmy rozpoznawania cech sygnałów i decyzyjne, wprowadzone do sieci neuronowych.

Literatura

- [1] Durzyński Z., (1998), *Diagnostyka układów biegowych pojazdów szynowych na bazie sygnałów wibroakustycznych*, III Krajowa Konferencja Naukowo-Techniczna: Diagnostyka Procesów Przemysłowych, Jurata, 7–10 września, 1998, str. 355–360.
- [2] Korbicz J., Obuchowicz A., Uciński D., (1994) *Sztuczne sieci neuronowe*, Akademicka Oficyna Wydawnicza PLJ, Warszawa.
- [3] Michalewicz Z.: *Algorytmy genetyczne + struktury danych = programy ewolucyjne*, WNT, Warszawa 1996.
- [4] Goldberg D.E., (1995), *Algorytmy genetyczne i ich zastosowania*, WNT, Warszawa.
- [5] Heckerman D., Wellman M.P., (1995), *Bayesian networks*, *Communications of the ACM*, vol. 38, No. 3, p. 27–30, March 1995
- [6] Jang J-S.R., Sun C-T., Mizutani E., (1997), *Neuro-Fuzzy and Soft Computing*, Prentice Hall, Upper Saddle River.
- [7] Li H., (1998) *An introduction to belief networks*, *Technical research report*, *The Center for Satellite and Hybrid Communication Network*, internet: <http://www.isr.umd.edu/CSHCN/>
- [8] Michalski R.S., Kubat M., Bratko I., *Machine Learning & Data Mining: Methods & Applications*, J. Wiley & Sons, Chichester 1998.
- [9] Mulawka J.J., *Systemy ekspertowe*, WNT, Warszawa 1996.
- [10] Osowski S., *Sieci neuronowe w ujęciu algorytmicznym*, WNT, Warszawa 1996.
- [11] Pearl J., (1988) *Probabilistic reasoning in intelligent systems*, Morgan Kaufmann, San Mateo, Cal.
- [12] Russel S., Norvig P., *Artificial Intelligence*, Prentice-Hall, London 1995.
- [13] Rutkowska D., Piliński M., Rutkowski L., (1997) *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*, PWN, Warszawa–Łódź.
- [14] Tadeusiewicz R., (1998), *Sygnał mowy*, WKiŁ, Warszawa.
- [15] Ziarko W., (1999) *Discovery through rough set theory*, *Communications of the ACM*, vol. 42, No. 11, p. 55–57, November 1999.
- [16] Żurada J., Barski M., Jędruch W., *Sztuczne sieci neuronowe*, PWN, Warszawa 1996.