

Rozdział 4

Metody sztucznej inteligencji w diagnostyce

Zintegrowana baza wiedzy diagnostycznego SE opiera się na kilku formach reprezentacji wiedzy wydobywanej i przetwarzanej z wykorzystaniem metod sztucznej inteligencji. W niniejszym rozdziale przedstawia się możliwości zastosowania wybranych elementów sztucznej inteligencji do detekcji i lokalizacji uszkodzeń w obiektach technicznych. Omawia się zastosowanie sieci neuronowych, zbiorów rozmytych i algorytmów genetycznych w budowie wybranych bloków systemu diagnostycznego.

Przeglądowo przedstawiono zastosowanie wielowarstwowych sieci jednokierunkowych typu perceptronowego MLP (ang. *MultiLayer Perceptron*), sieci samoorganizujących się oraz sieci dynamicznych. Sieci dynamiczne stanowią ważny element omawianego obszaru rozwoju sztucznej inteligencji w przypadku realizacji zadania diagnozowania złożonego procesu przemysłowego [44, 51].

W zadaniu monitorowania procesu przemysłowego często spotyka się zagadnienie gromadzenia i przetwarzania wiedzy niepewnej i nieprecyzyjnej. Temat ten przedstawiono w rozdziale 4 w kontekście poszukiwania odpowiedniej reprezentacji, która umożliwi wiedzę niepewną i nieprecyzyjną wykorzystać w diagnostycznym SE. Zadanie to można rozwiązać stosując logikę rozmytą. W przedstawianym rozdziale zawarto analizę zasad wykorzystania zbiorów rozmytych do budowy regulowej bazy wiedzy, modeli oraz klasyfikatorów stanów diagnozowanych obiektów. Omówiono techniki optymalizowania struktury i parametrów modeli i klasyfikatorów rozmytych. Ponadto omówiono właściwości systemów neuro-rozmytych oraz techniki uczenia sieci neuro-rozmytych.

Zadanie diagnozowania procesu przemysłowego ze zintegrowaną bazą wiedzy

proponuje się realizować na bazie SE. W końcowej części rozdziału przedstawiono zagadnienia struktury SE, metod wnioskowania oraz własności diagnostycznego SE z rozmytą reprezentacją sygnałów obiektowych.

4.1. Wstęp

Diagnostyka procesów przemysłowych może być realizowana z zastosowaniem różnych elementów bazy wiedzy. Zaliczyć do nich można wiedzę reprezentowaną z wykorzystaniem metod analitycznych, heurystycznych, sieci neuronowych lub logiki rozmytej. Każda z tych form reprezentacji wiedzy może w zadaniach diagnostycznych wiązać się z dedykowanymi zastosowaniami oraz wyznaczać specyficzne cechy systemu diagnostycznego [38].

Detekcja uszkodzeń z zastosowaniem metod analitycznych jest niekiedy niemożliwa, ponieważ techniki te wymagają dobrej znajomości modelu procesu. Przy nieprecyzyjnym modelu bardzo prawdopodobne jest powstawanie fałszywych alarmów wynikających z błędów estymacji zmiennych stanu, lub parametrów procesu [103]. Znane są prace przedstawiające zastosowanie w systemach diagnostycznych sieci neuronowych, pełniących funkcje modeli diagnozowanych obiektów oraz klasyfikatorów analizowanych stanów obiektu.

Dla złożonych procesów przemysłowych uzyskane w czasie monitorowania dane zazwyczaj zawierają nieprecyzyjności i niepewności zawartej informacji o stanie procesu. W takich sytuacjach zadawalającą jakość diagnozowania uzyskuje się tworząc bazę wiedzy z zastosowaniem reprezentacji opartej o logikę rozmytą [202, 243, 247].

4.2. Sieci neuronowe

Sztuczne sieci neuronowe cechują się kilkoma istotnymi własnościami [47]:

- ◇ zdolność realizacji złożonych obliczeń przez równoległe przetwarzanie informacji,
- ◇ efektywna aproksymacja dowolnych nieliniowości,
- ◇ uczenie i douczanie sieci na podstawie obserwacji (pomiarów) sygnałów obiektu,
- ◇ wysoki stopień odporności na zakłócenia,
- ◇ możliwość pracy z wiedzą niekompletną.

Wymienione cechy oraz to, że przy tworzeniu detektora neuronowego brak jest potrzeby znajomości modeli matematycznych diagnozowanych obiektów lub procesów, wymusiły znaczne zainteresowanie zastosowaniem sztucznych sieci neuronowych do budowy układów diagnostycznych [47, 106], szczególnie często spotykanych w przemyśle chemicznym, cechującym się nieliniowymi i niestacjonarnymi procesami.

Typowe zastosowanie detektora neuronowego wiąże się z funkcjami klasyfikatora stanów obiektu. Klasyfikator ten budowany jest na podstawie eksperymentalnych danych wejściowych i wyjściowych. Neuronowy detektor w zasadzie realizuje zadanie rozpoznawania obrazów traktowanych jako stany obiektu. Każdy obraz wykorzystywany do uczenia sieci zawiera zestaw danych wejściowych i odpowiadających im danych wyjściowych. W sieciach wiedza jest reprezentowana przez wartości wag i topologię połączeń neuronów. Takie sieci mogą uczyć się i adaptować do zachowania procesu, stąd mogą reprezentować złożone systemy techniczne, które są trudne do modelowania lub opisu w tradycyjnej bazie wiedzy systemu ekspertowego.

4.2.1. Sieci neuronowe typu perceptronowego

Perceptron jest siecią jednokierunkową, w której sygnały (obrazy wejściowe) są przesyłane od warstwy wejściowej poprzez warstwy ukryte do warstwy wyjściowej [53]. Z punktu widzenia systemów reprezentuje on statyczne układy nieliniowe, gdyż jego wyjścia w chwili t zależą od wejść tylko w tej samej chwili t . Perceptron wielowarstwowy tworzą neurony ułożone w warstwach o jednym kierunku przepływu sygnałów, o połączeniach międzywarstwowych jedynie między sąsiednimi warstwami.

Nieliniowe przetwarzanie neuronowe każdej warstwy można formalnie opisać funkcją $F\{\cdot\}$ aktywacji neuronów:

$$N_h \{u_h\} = F \{ \mathbf{W}^h \mathbf{u}_h \}, \quad h = 1, 2, \dots, H, \quad (4.1)$$

gdzie: \mathbf{u}_h oznacza wektor sygnałów wejściowych dla h -tej warstwy, N_h - operator przetwarzania neuronowego między $h - 1$ i h -tą warstwą.

Ogólne odwzorowanie wejścia-wyjścia tego typu sieci opisano zależnością:

$$y = N \{ \mathbf{u} \} = F^{wyj} \{ \mathbf{W}^{wyj} F \{ \mathbf{W}^h F \{ \mathbf{W}^{h-1} \dots F \{ \mathbf{W}^1 \mathbf{u} \} \dots \} \} \}, \quad (4.2)$$

gdzie: $y = [y_1, y_2, \dots, y_n]^T$ jest wektorem sygnałów wyjściowych, \mathbf{W}_{wyj} - macierzą wag połączeń warstwy wyjściowej z ostatnią warstwą ukrytą, F_{wyj} - operatorem przetwarzania neuronowego warstwy wyjściowej.

Sieć typu perceptronowego ma trzy grupy warstw: wejściową, ukrytą oraz wyjściową. Warstw ukrytych może być kilka. Neurony warstwy wejściowej realizują zadanie wstępnego przygotowania danych \mathbf{u} , może to obejmować normalizację lub skalowanie sygnałów. Przetwarzanie opisane zależnością (4.2) wykonywane jest w warstwach ukrytych i wyjściowej [76]. W warstwach tych każdy element jednej warstwy połączony jest z każdym elementem warstwy następnej. Połączenia te są realizowane z odpowiednimi współczynnikami wag w_{kl} , które w zależności od postawionego zadania są wyznaczane indywidualnie w procesie uczenia lub douczania.

Do uczenia sieci typu perceptronowego najczęściej stosuje się algorytm propagacji wstecznej, zaliczany do uczenia nadzorowanego. Jest to algorytm, oparty na minimalizacji sumy kwadratów błędów uczenia z wykorzystaniem optymalizacji metodą największego spadku [239]. W algorytmie tym zastosowano specyficzny sposób propagowania błędów uczenia sieci z wyjścia poprzez kolejne warstwy ukryte do wejścia [188]. Zaliczany jest on do najskuteczniejszych algorytmów uczenia sieci.

Konsekwentna realizacja wybranej metody minimalizującej błąd wymaga dokonywania zmian wag dopiero po zaprezentowaniu sieci wszystkich przygotowanych wzorców. W praktyce zmiany wag dokonuje się po każdej prezentacji pojedynczego wzorca. Dla nieliniowych funkcji aktywacji metoda ta nosi nazwę *uogólnionej reguły delty*.

Dla sieci wielowarstwowej należy określić zależności określające zmianę wartości wag między warstwami ukrytymi oraz warstwą wyjściową. Wyznaczanie pochodnej cząstkowej błędu powstałego dla elementów warstw ukrytych jest bardziej kłopotliwe, ponieważ zależność błędu sieci od odpowiedzi danej warstwy jest bardzo złożona [188].

4.2.2. Sieci samoorganizujące się

Dla poprawnego działania sieci samoorganizującej się muszą być spełnione dwa podstawowe założenia:

- ◇ obrazy wejściowe o określonych wspólnych cechach należą do tej samej klasy,
- ◇ sieć jest zdolna zidentyfikować te wspólne cechy w całym zestawie obrazów wejściowych.

Do tego typu sieci należy zaliczyć mapę cech Kohonena. Kohonen (1984) zaproponował dwuwymiarowy model układu uczącego się bez nadzoru [101]. Sieć jest zbudowana z warstwy elementów wejściowych, których jedynym zadaniem jest przekazywanie sygnałów wejściowych do warstwy elementów przetwarzających. Zatem

jest to sieć jednowarstwowa. Wyjścia wszystkich elementów warstwy wejściowej są połączone ze wszystkimi elementami sieci. Każdy element przetwarzający jest jednocześnie elementem wyjściowym. Wielkość oraz rodzaj połączenia (hamujące czy pobudzające) między elementami warstwy wejściowej i warstwy przetwarzającej zależą od odległości (fizycznej) między elementami.

Neurony położone bliżej elementu wysyłającego sygnały odbierają to oddziaływanie jako silniejsze, natomiast dla dalej położonych po przekroczeniu pewnej wartości progowej odległości - jako hamujące (kapelusz meksykański). Opierając się na tych regułach Kohonen zbudował obszary sąsiedztwa.

Sieci samoorganizujące uczone bez nadzoru są poddawane uczeniu konkurencyjnemu. Do każdego z elementów przetwarzających sieci przesyłana jest pełna informacja o prezentowanym obrazie wejściowym. Aktywacji podlega jedynie ten element przetwarzający, którego wektor wag jest najbardziej zbliżony do wektora wejściowego.

Do uczenia sieci stosuje się znormalizowane względem 1 wektory uczące. Uczenie odbywa się według zmodyfikowanej reguły Grossberga-Kohonena i należy do odmiany WTA (ang. *Winner Takes All*). Modyfikacja polega na tym, że uaktywnianie wag w każdym kroku uczenia dotyczy nie tylko neuronu zwycięzcy ale również wszystkich neuronów z nim sąsiadujących. Na starcie uczenia wszystkie wagi przyjmują wartości losowe i małe. Wektor wejściowy x uaktywnia współzawodnictwo neuronów. Zwycięzca uaktualnia swoje wagi oraz pozwala uaktualnić wagi neuronom z sąsiedztwa.

Sieć typu Kohonena można stosować do wykrywania cech charakterystycznych rozkładu danych wejściowych, wykrywania skupień danych wejściowych, kompresji danych itd. Biorąc pod uwagę te cechy można zastosować taką sieć do diagnozowania obiektów przemysłowych.

Mapa Kohonena jest niekiedy używana jako narzędzie diagnostyczne, ponieważ nie wymaga nadzorowania przy uczeniu. Przy uszkodzeniach maszyn (przerwy w pracy) często zebranie unikalnych danych uczących o różnych uszkodzeniach jest bardzo trudne albo niemożliwe i wtedy należy wybrać klasyfikator neuronowy oparty o sieci nienadzorowane.

4.2.3. Dynamiczne sieci neuronowe

Układy realizujące zadania diagnostyki procesów przemysłowych często mają strukturę zawierającą model badanego procesu. Dla obiektów o złożonej strukturze lub nieznanych własnościach są to modele spekulatywne zbudowane z wykorzystaniem odpowiedniej struktury sieci neuronowej. Modele budowane z wykorzystaniem sie-

ci typu perceptronowego zapewniają prawidłową aproksymację charakterystyki obiektu w stanie statycznym. Jednakże w ostatnich czasach obserwuje się znaczne zainteresowanie badaniami w zakresie zastosowań sztucznych sieci neuronowych w modelowaniu obiektów dynamicznych [105, 250].

W odróżnieniu od dobrych wyników zastosowania sieci neuronowych w zadaniach diagnostyki obiektów w stanach statycznych lub wolnozmiennych, stosowanie ich w procesie diagnozowania obiektu o dużej dynamice wymaga poszukania nowych struktur sieci. Pierwsze próby rozwiązania tego problemu podjęto przez zastosowanie sieci jednokierunkowych z wprowadzonymi sztucznie liniami opóźniającymi sygnały [138, 105]. Jednak takie sieci są nadal statyczne i zastosowanie ich do modelowania w zagadnieniach dynamicznych jest mocno ograniczone i niewystarczające.

Znacznie lepsze wyniki diagnostyczne w przedstawionym systemie można osiągnąć używając sieci rekurencyjne [43]. Dzięki wprowadzonym do sieci rekurencyjnym sprzężeniom zwrotnym możliwe jest gromadzenie informacji, a zatem praca z historią modelowanego obiektu, np. sieci Hopfielda [80]. Najbardziej ogólną architekturę sieci rekurencyjnej przedstawili Williams i Zipser (1989). Niestety, zastosowanie ich w problemie modelowania jest mocno ograniczone, głównie z powodu ich niestabilności i bardzo wolnej zbieżności procesu uczenia [252]. Stany stabilne odpowiadają minimum funkcji energii i nazywa się je *atraktorami*. Każdemu z nich można przyporządkować zbiór warunków początkowych, które inicjują ewolucję sieci kończącą się na *atraktorze*. Zbiór takich wejść nazywany jest *niecką przyciągania danego atraktora*. Rodzaj i liczba *atraktorów*, ich wzajemne oddalenie ściśle zależy od doboru wag połączeń między elementami. Doboru tego dokonuje się w zależności od obszaru zastosowań sieci Hopfielda. Sieć Hopfielda jest nieliniowa, to znaczy zbudowana z neuronów o nieliniowych funkcjach aktywacji. Parametry sieci dostrajane są w trakcie uczenia [47]. Sieci te mogą mieć szerokie zastosowania przy kojarzeniu, klasyfikacji i odtwarzaniu obrazów, rozwiązywaniu zagadnień optymalizacyjnych itp.

W ogólnym przypadku dobre wyniki w zagadnieniu modelowania dynamiki procesów osiągnąć można stosując sieci z modelami dynamicznych neuronów. Można je otrzymać poprzez wprowadzenie do modelu statycznego typu McCullocha-Pittsa [150] sprzężenia zwrotnego w synapsie [73]. W strukturze takiego modelu występuje blok sumacyjny, liniowy układ dynamiczny w postaci filtru IIR (ang. *Infinite Impulse Response*) oraz nieliniowego bloku aktywacji [6, 171].

Dynamika w przedstawionym modelu zostaje wprowadzona do neuronu w taki sposób, aby aktywność neuronu zależała od jego wewnętrznych stanów. Przy pomocy zastosowanego filtru liniowego każdy neuron w dynamicznej sieci neuronowej odtwarza przeszłe wartości sygnałów, mając do dyspozycji dwa sygnały: wejściowy $u_p(k)$, dla $p = 1, 2, \dots, P$ i sygnał wyjściowy $y(k)$. Do budowy dyna-

micznego modelu neuronu można stosować liniowe systemy dynamiczne różnych rzędów, pierwszego, drugiego i wyższych.

Modelowanie własności obiektu dynamicznego wymaga zastosowania określonej architektury sieci z dynamicznymi neuronowymi modelami. Sieć ta ma prostą strukturę podobną do struktury statycznej jednokierunkowej sieci wielowarstwowej. Miejsce standardowych statycznych modeli neuronu zajmują modele dynamiczne. Zaproponowana struktura sieci nie zawiera żadnego rodzaju globalnych sprzężeń zwrotnych, które komplikują architekturę sieci, a co najważniejsze, zależności adaptacyjnego algorytmu uczenia. Dynamiczna wielowarstwowa sieć neuronowa może być zastosowana do modelowania własności nieznanego nieliniowego obiektu dynamicznego [111]. W ogólnym przypadku zadanie polega na odwzorowaniu funkcji przejścia wejście-wyjście nieznanego obiektu, bazując na parach wartości wejście-wyjście.

W procesie uczenia sieci błąd obliczony na wyjściu sieci jest propagowany wstecz poprzez warstwy ukryte do warstwy wejściowej przez dynamiczne filtry warstw, podobnie jak w standardowym algorytmie wstecznej propagacji błędu [47]. Wynikiem jest rozszerzony algorytm dynamicznej propagacji wstecznej [6].

4.2.4. Zastosowania sieci neuronowych w wydobywaniu wiedzy i diagnostyce

Stosowanie sztucznych sieci neuronowych w diagnostyce wiąże się z następującymi działaniami [200]:

- definicja listy uszkodzeń,
- wyznaczanie typowych obrazów odpowiednich dla poszczególnych uszkodzeń i stanu normalnego,
- wybór struktury sieci i algorytmu uczenia,
- uczenia sieci na bazie obrazów uczących,
- testowanie sieci dla dowolnych stanów obiektu.

Wśród najnowszych praktycznych zastosowań sieci neuronowych w przemyśle można wymienić następujące przykłady [24]:

- system badania emisji akustycznej w stacji badawczej Freedom przygotowany przez NASA,
- system diagnostyczny dla silników spalinowych przygotowany przez Ford Motor Company oraz NASA,

- system CATCH sponsorowany przez Departament Sprawiedliwości USA, stosowany przy identyfikacji seryjnych zabójców,
- system eksploatacji i konserwacji elektrociepłowni w bazach US Navy,
- system klasyfikacji i identyfikacji sygnałów sonarowych opracowany przez General Dynamics dla US Navy,
- prognozowanie zapotrzebowania na moc elektryczną, wdrożony w elektrowni BC Aydro w Vancouver.

Sieci neuronowe stosuje się z sukcesem w wydobywaniu wiedzy z baz danych w zadaniach klasyfikacji oraz regresji. Niestety, mają one istotny mankament w przypadku potrzeby generowania powiązań danych w przetwarzanej bazie. Zagadnienie to ma istotne znaczenie w zadaniach wydobywania wiedzy, gdzie powiązania między bazą danych i powstałą w procesie przetwarzania informacją hurtownią danych ma istotne znaczenie. Sieci neuronowe postrzegane jako *czarne skrzynki* mają nieczytelną reprezentację wiedzy o wykrytych i zakodowanych poprzez wektory wag powiązaniach między danymi w bazie. Aby ograniczyć wadę *czarnej skrzynki*, proponuje się stosowanie różnych algorytmów, które są zdolne wydobywać symboliczne reguły z utworzonej sieci neuronowej.

Dobrą skuteczność wydobywania wiedzy przy wykorzystaniu sieci neuronowej można uzyskać stosując systemy hybrydowe. Ciekawym jest rozwiązanie wzorowane na procesie myślowym człowieka. W systemie tym proponuje się zintegrowanie działania sieci neuronowej typu perceptronowego z systemem rozumowania opartym na pamięci MBR (ang. *Memory-Based Reasoning*) [214]. Tak zintegrowana struktura wynika z procesu myślowego mózgu ludzkiego, który jest wspomagany przez pamięć i struktury sieci neuronów z wagami wyznaczonymi w procesie nauczenia.

Ważnym zagadnieniem do rozwiązania jest problem zdefiniowania cech wydobytej przez sieć neuronową wiedzy. Jednym z rozwiązań jest wyznaczanie cech na bazie wektorów wag sieci ANN. Do reprezentacji wiedzy zawartej w nauczonej sieci neuronowej zaproponowano [214] cztery metody wyznaczania cech: wrażliwość, aktywność, istotność, znaczenie (doniosłość).

Wrażliwość wejściowego węzła jest obliczana przy usuniętym wybranym węzle z nauczonej sieci ANN. Ocenę wrażliwości dokonuje się porównując odpowiedź zmodyfikowanej sieci z odpowiedzią sieci o oryginalnej strukturze. Wyznacza się ją w oparciu o następującą zależność:

$$S_i = \frac{\left(\sum_L \frac{|P^0 - P^i|}{P^0} \right)}{n}, \quad (4.3)$$

gdzie: P^0 oznacza wartość generowaną przez sieć w warunkach normalnych, P^i - wartość generowana przez zmodyfikowaną sieć przez usunięcie i -tego węzła, L i n są odpowiednio zbiorem i liczbą danych uczących.

Aktywność węzła jest wyznaczana jako wariancja poziomu aktywności dla danych uczących. Kiedy wartość aktywności węzła znacznie różni się dla jego danych wejściowych, aktywność tego węzła określa się jako wysoką. Aktywność węzła z_j w warstwie ukrytej można określić za pomocą następującej zależności:

$$A_j = \left(w_j^{(2)}\right)^2 \cdot \text{var} \left(g \left(\sum_{i=0}^d w_{ji}^{(1)} x_i \right) \right), \quad (4.4)$$

gdzie: $\text{var}(\cdot)$ oznacza funkcję wariancji, $w_j^{(2)}$ -waga między j -tym węzłem warstwy ukrytej a węzłem w warstwie wyjściowej, $w_{ji}^{(1)}$ -waga między i -tym wejściem a j -tym węzłem w warstwie ukrytej, $g(\cdot)$ -funkcja aktywacji oraz d -liczba wejść. Aktywność węzła w warstwie wejściowej wyznacza się za pomocą zależności:

$$A_i = \sum_{j=1}^M \left(\left(w_{ji}^{(1)}\right)^2 \cdot A_j \right), \quad (4.5)$$

gdzie: M oznacza liczbę węzłów w warstwie ukrytej.

Istotność wagi określa się jako estymatę drugiej pochodnej błędu względem wagi. Mierzona istotność wagi jest proporcjonalna do kwadratu wagi. Istotność węzła wejściowego określa się jako:

$$Istot_i = \sum_{j=1}^M \left(\left(w_{ji}^{(1)}\right)^2 \cdot \left(w_j^{(2)}\right)^2 \right). \quad (4.6)$$

Znaczenie jest określane jako wpływ wariancji wag połączeń wejściowych do węzła na jego odpowiedź. Wyzerowanie największej wagi węzła o dużym znaczeniu powoduje znaczny wzrost błędu odpowiedzi sieci. Znaczenie węzła ukrytego z_j można wyznaczyć z następującej zależności:

$$Zn_j = \left(w_j^{(2)}\right)^2 \cdot \text{var} \left(w_{ji}^{(1)} \right). \quad (4.7)$$

oraz znaczenie dla węzła wejściowego x_i opisuje zależność:

$$Zn_i = \left(\left(w_{ji}^{(1)}\right)^2 \cdot Zn_j \right). \quad (4.8)$$

Diagnozowanie procesów przemysłowych obejmuje również zagadnienie weryfikacji poprawności pracy systemu pomiarowego. Klasyczna metoda diagnozowania czujników bazuje na systemie z modelem. Do budowy modeli o zazwyczaj nieliniowych charakterystykach statycznych stosuje się sieci neuronowe. Przy budowie neuronowych modeli bardzo ważnym i trudnym zadaniem jest wybranie spośród pomiarowo dostępnych wielkości takiego podzbioru, na podstawie którego można będzie osiągnąć najwyższą dokładność wyznaczania odpowiedzi.

Dodatковым utrudnieniem uzyskania dobrej jakości modelu jest wybór odpowiednich ciągów uczących. Źle dobrane ciągi uczące (np. mało zmienne mierzone wielkości procesowe w stanie poprawnej pracy, przy złożonych charakterystykach obserwowanego procesu) powodują, że uzyskany model wykazuje wrażliwość bardzo różniącą się od rzeczywistego obiektu i równocześnie dobre wyniki testów zgodności. Zazwyczaj jest to następstwem złego skomponowania ciągów uczących i szczegółowe badania wykazują, że dobra jakość modelu jest zachowana tylko w ograniczonym obszarze stanów procesu. Ponadto zwrócono uwagę, że przy budowie modelu neuronowego poza oceną dokładności aproksymacji istotna jest również wrażliwość modelu na poszczególne sygnały wejściowe. Bardziej szczegółową analizę tego zagadnienia można znaleźć w artykule Kornackiego i Jankowskiej [119].

Obserwacja niezbędnych w procesie diagnozowania sygnałów obiektowych jest często utrudniona lub wręcz niemożliwa, aby utworzyć model wejściowo-wyjściowy. W takich sytuacjach często stosuje się do modelowania właściwości podzespołów sieci komórkowe uczone bez nauczyciela. Przykładem jest sieć samoorganizująca się Kohonena. Tego typu sieci stosuje się często do wydobywania wiedzy z baz danych. Zarchiwizowane dane zawierające wiedzę o normalnym lub uszkodzonym trybie pracy badanego obiektu stanowią bazę ukrytej wiedzy o cechach obiektu, które mogą być wykorzystane do przygotowania wzorców identyfikujących stan normalny lub uszkodzony obiektu. Sieci samoorganizujące Kohonena pozwalają przygotować niezbędne w procesie diagnostycznym wzorce wykorzystywane w diagnostycznych SE [14].

Po nauczaniu sieć tworzy kilka grup skojarzonych z normalną pracą i z uszkodzeniami pojawiającymi się z różnych przyczyn. Istotnym zagadnieniem projektanta takiej sieci jest definicja *a priori* liczby elementów przyjętego modelu. Wstępnie przyjmuje się, że liczba elementów sieci powinna być większa od liczby wejść. Wtedy po nauczaniu sieć, w większości przypadków, klasyfikuje uszkodzenia prawidłowo, gdy wzorce testujące są bliskie uczącym. Potrafi ona wyróżnić grupy nawet wtedy, gdy pewne cechy wzorców dla różnych uszkodzeń są takie same, ale węzły w mapie Kohonena są bliskie sobie. Niestety, czasami sieć ta tworzy grupy nie rozwijane dla danego wzorca. Dla przykładu: wzorce opisujące różne uszkodzenia są klasyfikowane do tych samych elementów warstwy Kohonena. Zwiększenie liczby elementów warstwy Kohonena zwiększa niezawodność, jednakże nawet kilkakrotne zwiększenie liczby elementów nie gwarantuje niezawodnej klasyfikacji uszkodzeń.

Kiedy z sukcesem uczona mapa jest używana do klasyfikacji, to na kilkaset wzorców testujących zazwyczaj kilka jest klasyfikowanych nieprawidłowo i kilkanaście jest klasyfikowanych do pustych grup. Przedstawione zagadnienia przedstawił na przykładzie maszyny papierniczej Koivo [103].

Obszar diagnozowania procesów przemysłowych często obejmuje zagadnienia analizy własności dynamicznych. Zagadnienie modelowania własności dynamicznych diagnozowanych obiektów może być realizowane w oparciu o sieć Hopfielda (pamięć autoskojarzeniowa) [80] jak również z zastosowaniem, ostatnio dynamicznie rozwijanych, sieci typu perceptron z dynamicznymi neuronami. Rozwiązanie drugie pozwala zastosować dobrze znane techniki uczenia sieci, przy dodatkowych parametrach opisujących moduł opóźnienia każdego neuronu [166, 169, 170].

Sieć neuronowa już z jednym dynamicznym neuronem i filtrem drugiego rzędu może być modelem obiektu dynamicznego drugiego rzędu. Patan i in. [171] zastosowali tego typu sieć neuronową jako model obiektu dynamicznego drugiego rzędu z nieliniowością typu nasycenie. Uzyskane wyniki identyfikacji nieparametrycznej obiektu uznano za zadawalające pomimo zastosowania tylko jednego dynamicznego neuronu.

Przy omawianiu cech sieci neuronowych z dynamicznymi neuronami można zadać sobie pytanie: czy można je stosować w układach diagnostycznych dla złożonych procesów przemysłowych? Odpowiedź na to pytanie można znaleźć w przykładzie ukazującym zasadę zastosowania dynamicznych sieci neuronowych w procesie diagnozowania systemu dwóch zbiorników z elementem opóźnienia transportowego [171].

Bazując na strukturze systemu FDI z dynamicznymi neuronami, uzyskano poprawne wyniki detekcji wprowadzonych w obiekcie uszkodzeń. Każda z sieci modelowała prawidłowo zachowanie obiektu w wybranym stanie, dając w efekcie zerową wartość residuum dla nauczonego stanu obiektu. Jako klasyfikator residuów zastosowano klasyczną statyczną sieć wielowarstwową jednokierunkową typu $N_{4,5,4,3}^3$. Zaproponowany system FDI dobrze realizuje postawione zadania diagnostyczne. Jakkolwiek dobre rezultaty uzyskano po kilku procesach uczących sieć kaskadową.

4.3. Zbiory rozmyte i ich obszar zastosowań

W diagnostyce procesów najczęściej używa się układy z modelem [87]. Można wyróżnić modele zbudowane na podstawie znajomości fizyki zjawisk zachodzących w procesie lub modele abstrakcyjne. Pierwsze nazywane są modelami matematycznymi, drugie logicznymi. Po roku 1970 można zauważyć szybki rozwój modeli logicznych budowanych z wykorzystaniem osiągnięć w dziedzinie sztucznej inteligencji.

Logika rozmyta stanowi narzędzie, które może być stosowane z powodzeniem do budowy obu typów modeli [220]. Wykorzystanie podzbiorów rozmytych umożliwia rozpoczęcie realizacji zamiaru Zadeha, złagodzenia żądania precyzji przy tworzeniu modelu. Drugim obszarem wykorzystania pojęcia rozmycia w zagadnieniu modelowania jest rozmyte funkcjonowanie operatorów i spójników.

Obecnie zbiory rozmyte stosuje się głównie w klasie modeli logicznych, zwanych modelami, na podstawie reguł lub modelami na podstawie wiedzy. Na rozwój tego rodzaju modeli silny wpływ miały prace prowadzone przez specjalistów od sztucznej inteligencji w dziedzinie systemów ekspertowych. Mamdani i jego współpracownicy [147] połączyli ideę regułowej bazy wiedzy z wykorzystaniem parametrów rozmytych do budowy regulatora modelującego pracę człowieka-operatora. W szczególności nowym aspektem modeli o regułach rozmytych jest szerokie stosowanie operacji logicznych na spójnikach, których znaczenie i parametry mają wartości rozmyte.

Systemy oparte na logice rozmytej wymagają definicji kilku istotnych parametrów systemu rozmytego. Zaliczyć do nich należy: zasady rozmywania danych pomiarowych [40], techniki przetwarzania zmiennych lingwistycznych w procesie wnioskowania jak również zasady wyostrzania wyniku opisanego w formie rozmytej [202, 210, 243]. Proces strojenia systemu rozmytego wymaga niekiedy zdefiniowania kilku podstawowych parametrów *a priori* (na bazie wiedzy eksperta), jak również sprowadzenia zadania do grupy problemów optymalizacyjnych, takich jak poszukiwanie optymalnej struktury rozmytego systemu lub określenia parametrów rozmytego modelu lub klasyfikatora [191]. Zatem w tym przypadku pojawia się problem optymalizacji wieloparametrycznej.

4.3.1. Podstawowe zagadnienia teorii zbiorów rozmytych

Pojęcie zbioru rozmytego wprowadził po raz pierwszy L. A. Zadeh (1965) [246] jako uogólnienie pojęcia zbioru klasycznego. Podzbiór rozmyty uważa się jako predykat o wartości logicznej z przedziału jednostkowego $[0, 1]$. Logiczną podstawą pojęcia zbioru rozmytego jest logika wielowartościowa. Zbiór rozmyty umożliwia opisywanie pojęć, których granica między posiadaniem pewnej własności i jej brakiem jest rozmyta [179, 188].

Definicja: Niech X będzie przestrzenią obserwacji. Zbiór rozmyty A ze zbiorem X jest związany funkcją charakterystyczną:

$$\mu_A : X \longrightarrow [0, 1], \quad (4.9)$$

W teorii zbiorów rozmytych funkcja charakterystyczna jest na ogół nazywana *funkcją przynależności*. Funkcja ta określa stopień przynależności każdego elementu x do zbioru A .

Zbiory rozmyte są szczególnie użyteczne do reprezentacji pojęć o niesprecyzowanych granicach. Zdolność bardziej naturalnej reprezentacji nieprecyzyjnego pojęcia jest szczególnie ważna w procesie projektowania systemów wyposażonych w elementy sztucznej inteligencji. Przy budowie systemów ekspertowych można stosować pojęcia jakościowe, jak np. *niskie, wysokie, małe, średnie, duże* itd.

4.3.2. Wnioskowanie i rozmyta baza wiedzy

Proces wnioskowania obejmuje zagadnienia oceny spełnienia (prawdziwości) przesłanek poszczególnych reguł w bazie. Obliczanie stopnia prawdziwości przesłanek złożonych z przesłanek prostych nazywa się **agregacją**. Funkcję przynależności konkluzji poszczególnych reguł wyznacza się w oparciu o znajomość stopni spełnienia ich przesłanek. Operację wyznaczania stopni aktywacji konkluzji nazywa się **wnioskowaniem** i realizuje się ją z użyciem operatorów implikacji. W wyniku realizacji wnioskowania dla poszczególnych reguł wyznacza się stopień aktywacji konkluzji, na podstawie których należy utworzyć jedną wynikową funkcję przynależności konkluzji całej bazy reguł. Operację wyznaczania wypadkowej konkluzji nazywa się **akumulacją**.

Dla rozmytej reprezentacji wiedzy stopień spełnienia przesłanek, w odróżnieniu od logiki klasycznej może przyjmować wartości ułamkowe z przedziału $[0, 1]$ [247]. Wartość zero oznacza, że dana reguła nie jest aktywna. Sposób obliczania stopnia spełnienia przesłanki prostej jest równoznaczny z wyznaczeniem stopnia przynależności wartości wejściowej do zbioru rozmytego zastosowanego w przesłance.

Implikacja rozmyta jest regułą R , która w swej najprostszej formie przyjmuje postać:

$$\text{Jeśli } (x = A) \text{ to } (y = D),$$

gdzie: $(x = A)$ jest przesłanką a część $(y = D)$ jest konkluzją.

W przypadku implikacji rozmytej, podobnie jak przy określaniu spełnienia przesłanki, dopuszcza się częściową prawdziwość przesłanki i konkluzji, mieszczącą się w ciągłym przedziale $[0, 1]$.

Funkcja przynależności implikacji $\mu_{A \rightarrow B}(x, y)$ jest podstawą tzw. wnioskowania rozmytego, umożliwiającego obliczanie wyjścia modelu rozmytego przy danym stanie wejść. Wartość tej funkcji wyznacza się stosując odpowiedni operator implikacji. Najczęściej stosowany jest operator implikacji Mamdaniego oparty na operatorze *t-normy* typu **MIN**. Drugim z często używanych operatorów rozmytej implikacji jest operator Larsena typu **PROD**.

Proces wyznaczania końcowej rozmytej odpowiedzi modelu realizuje się za pomocą operacji akumulacji. Istnieje wiele metod przeprowadzania akumulacji rozmytych konkluzji. Związane to jest z mnogością operatorów stosowanych w logice

rozmytej. Najczęściej relację wynikową całego procesu wnioskowania uzyskuje się stosując operator *s-normy* **MAX**.

4.3.3. Optymalizacja bazy wiedzy

Rozmyta reprezentacja wiedzy wymaga definicji wielu parametrów. Ważnym zagadnieniem jest definicja zasad rozmywania wejść i wyjść opisywanego obiektu. Od liczby zbiorów rozmytych definiujących wejścia obiektu zależy liczba reguł w bazie wiedzy systemu. Przyrost liczby reguł i parametrów funkcji przynależności wraz z liczbą wejść obiektu jest tak szybki, że w literaturze bywa nazywany *przekleństwem* wymiarowości (ang. *curse of dimensionality*). Dlatego należy dążyć do maksymalnego upraszczania opisu (wymaga to jednak zachowania rozsądku z punktu widzenia dokładności tworzonego modelu). Dokonując modyfikacji zasad opisu rozmytej reprezentacji wiedzy należy dokonać analizy podstawowych cech otrzymanej bazy wiedzy. Zaliczamy do nich: kompletność, niesprzeczność i nadmiarowość.

4.3.3.1 Kompletność modelu rozmytego

W literaturze spotyka się różne definicje kompletności bazy reguł. Większość z nich dotyczy kompletności numerycznej [97]. Podstawową definicję kompletności modelu rozmytego sformułowano następująco:

Definicja 4.2 [202]

Model rozmyty jest kompletny jeżeli dla każdego stanu wejść $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ przyporządkowuje pewien stan wyjścia y^* . Model rozmyty nie spełnia zasady kompletności, gdy istnieje choć jedna wartość wejścia \mathbf{x}^* , dla której nie zdefiniowano żadnej wartości wyjścia y^* .

Dla rozmytego modelu występują również pojęcia kompletności podziału obszaru wejściowego oraz kompletności bazy reguł.

Definicja 4.3

Rozmyty podział zakresu zmiennej wejściowej x_i jest kompletny wtedy gdy spełniona jest zależność:

$$\sum_{j=1}^m \mu_{A_{ji}}(x_i^*) > 0, \text{ dla } x_i^* \in X_i, \quad (4.10)$$

gdzie: m oznacza liczbę zbiorów rozmytych $\{A_{ji}\}$ przyjętych dla zmiennej X_i .

Definicja 4.4

Baza reguł rozmytego modelu jest lingwistycznie kompletna, jeżeli lingwistycznemu stanowi $(A_{1k}, A_{2l}, \dots, A_{np})$ przyporządkowano co najmniej jeden lingwistyczny

stan wyjścia B_j .

Godny rozważenia jest związek między kompletnością lingwistyczną a numeryczną. Według definicji lingwistycznie kompletna baza reguł może być numerycznie niekompletna, jeśli rozmyty podział przestrzeni wejść jest niekompletny. Natomiast lingwistycznie niekompletna baza reguł może być numerycznie kompletna, jeśli zastosowano funkcje przynależności o odpowiednio szerokich nośnikach [202].

4.3.3.2 Niesprzeczność bazy reguł

Definicja 4.5

Baza reguł jest niesprzeczna jeśli nie zawiera reguł sprzecznych, tzn. reguł o jednakowych przesłankach i różnych konkluzjach.

W przypadku rozmytej regułowej reprezentacji wiedzy wyróżnia się różne poziomy sprzeczności reguł. Jeśli konkluzje sprzecznych reguł opisane są za pomocą znacznie różniących się wyjściowych zbiorów rozmytych, mamy do czynienia z *silną* sprzecznością reguł. *Słaba* sprzeczność reguł obejmuje przypadki, gdy konkluzje opisane są za pomocą zbiorów rozmytych o zbliżonej wartości modalnej (punkt lub zbiór punktów przestrzeni rozważań, które uzyskują maksymalną wartość funkcji przynależności. Dla zbioru punktów wartość modalną wyznacza się jako wartość średnią).

Definicja 4.6

Baza m reguł R_j , $j = 1, \dots, m$, jest **kompletna i niesprzeczna**, gdy spełniona jest zależność: [202]

$$\sum_{j=1}^m \mu^j(\mathbf{x}^*) = 1, \quad \forall \mathbf{x}^* \in \mathbf{X}, \quad (4.11)$$

Jeżeli suma stopni spełnienia przesłanek reguł (4.11) jest mniejsza od 1, to baza reguł dla stanu wejść \mathbf{x}^* jest niekompletna. Jednakże jeśli ta suma jest większa od 1, to mamy do czynienia ze sprzecznością w regułach.

4.3.3.3 Nadmiarowość bazy reguł

Nadmiarowość bazy reguł występuje wtedy, gdy w bazie występuje kilka a nawet kilkanaście reguł o takich samych przesłankach i konkluzjach. Taka sytuacja może wynikać z pomyłki eksperta lub inżyniera wiedzy albo powstaje jako wzmocnienie konkluzji nadmiarowych reguł wygenerowanych przez samouczący się (samoorganizujący się) model rozmyty.

Jeśli nadmiarowość wynika z błędu człowieka, to nadmiarowe reguły należy usunąć. Jednakże w przypadku celowego wprowadzenia nadmiarowych reguł należy przeanalizować ich funkcje i dokonać syntezy powstałej bazy w celu akumulacji efektów wzmacniania konkluzji.

Dla przykładu rozpatrzmy dwie identyczne reguły R_2 i R_3 w postaci:

$$\begin{aligned} R_2 &: (\text{Jeśli } (x = n) \text{ to } (y = D)) \\ R_3 &: (\text{Jeśli } (x = n) \text{ to } (y = D)), \end{aligned}$$

które można zastąpić jedną regułą R_2^* o zsumowanej logicznie konkluzji w postaci:

$$R_2^* : (\text{Jeśli } (x = n) \text{ to } (y = D \cup D)).$$

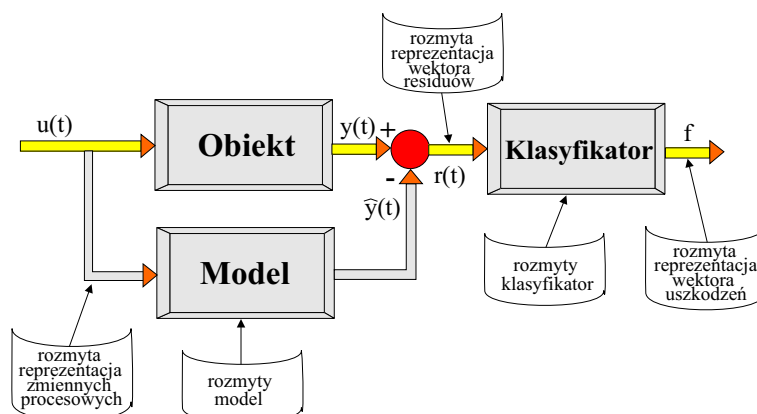
Efektywność wzmocnienia konkluzji reguły wypadkowej R_2^* uzależniona jest od zastosowanego operatora s -normy. Jeżeli sumowanie będzie realizowane za pomocą operatora MAX uzyskamy zbiór wynikowy $D^* = D \cup D = D$. W przypadku użycia innego operatora wartość konkluzji zostanie wzmocniona w stosunku zależnym od zastosowanego operatora.

Przedstawione w tej części rozdziału zagadnienia uzasadniają potrzebę budowy zintegrowanych SE, ze szczególnym zwróceniem uwagi na efektywność systemu zawierającego elementy wiedzy reprezentowanej w logice rozmytej. Dla wiedzy w takiej reprezentacji występuje większe prawdopodobieństwo uzyskania jej kompletności, szczególnie dla złożonych procesów przemysłowych, dla których należy analizować szeroki zakres zmienności obserwowanych sygnałów. Problem sprzeczności reguł, uzyskiwanych na drodze automatycznego wydobywania wiedzy z baz danych, przy rozmytej reprezentacji przyjmuje łagodniejszy wymiar. W takim przypadku istnieje możliwość wyznaczenia reguły wypadkowej z odpowiednią wartością poziomu przynależności. Nadmiarowość bazy wiedzy w reprezentacji rozmytej może być eliminowana bazując na odpowiednim operatorze t -normy. Niekiedy nadmiarowość pomaga we wzmocnieniu istotności danej reguły (wzrost wartości funkcji przynależności).

4.3.4. Zbiory rozmyte w diagnostyce

Zastosowanie logiki rozmytej wyznacza nowe podejście do jakościowych technik diagnostycznych [125, 156, 217]. Procesy detekcji uszkodzeń opierają się często na wiedzy heurystycznej skorelowanej z rozmytą reprezentacją sygnałów opisujących stan diagnozowanego obiektu [7, 22, 132, 187].

Rozmytą reprezentację wiedzy można stosować w realizacji różnych zadań postawionych przed systemem diagnostycznym (rys.4.1) [42, 60]. W etapie wstępnym spotyka się z rozmytą reprezentacją wartości zmiennych procesowych, co pozwala



Rys. 4.1. Zastosowanie logiki rozmytej w systemie diagnostycznym

prować proces diagnozowania z wykorzystaniem doradczych systemów ekspertowych [27, 186]. Bardzo ważnym i często realizowanym w oparciu o rozmytą reprezentację wiedzy, zadaniem jest generowanie reszduów. Unormowanej wartości reszduum przyporządkowuje się etykiety lingwistyczne w postaci zbiorów rozmytych o zadanych kształtach (najczęściej trapezoidalne). Trapezoidalne funkcje przynależności pozwalają wydzielić strefy odpowiadające pracy prawidłowej i awarii, a także obszar, w którym decyzja o uszkodzeniu jest niepewna [123]. Otrzymane w ten sposób wektory reszduów pozwalają budować macierze diagnostyczne z rozmytą reprezentacją powiązań symptom-uszkodzenie. Daje to bardziej elastyczny system diagnostyczny w porównaniu z systemem opartym o binarne macierze diagnostyczne [126].

4.3.4.1 Rozmyta reprezentacja symptomów uszkodzeń

W wyniku realizacji testów diagnostycznych generowany jest zbiór wartości symptomów (sf), które w postaci rozmytej przyjmują postaci [11]:

$$sf = \{\mu_N(r_{n1}), \mu_N(r_{n2}), \dots, \mu_N(r_{nJ})\} \quad (4.12)$$

$$\forall j \in \{1 \dots J\},$$

gdzie: r_{nj} oznacza unormowane reszduum dla j -tego uszkodzenia, $\mu_N(r_{nj})$ jest wartością funkcji przynależności j -tego negatywnego wyniku testu diagnostycznego oraz J oznacza liczbę testów diagnostycznych.

Tak zdefiniowaną rozmytą reprezentację symptomów uszkodzeń można zastosować do określenia miary skuteczności detekcyjnej zbioru testów diagnostycznych.

Może nią być unormowana moc zbioru:

$$|S_f|_n = \frac{1}{J} \sum_{j=1}^J \mu_N(r_{nj}). \quad (4.13)$$

Im większa jest wartość mocy unormowanej (bliższa wartości 1) tym większa jest pewność braku uszkodzeń.

Innym ważnym aspektem diagnostycznym jest wyznaczenie stopnia pewności wystąpienia uszkodzenia lub uszkodzeń, który wyznacza się z zależności [11]:

$$\bigcup_{j=1}^J \mu_N(r_{nj}) > 0. \quad (4.14)$$

Ostatecznie, w pracy [11], proponuje się do oceny stopnia pewności braku uszkodzenia stosować unormowaną moc zbioru (4.13), zaś za pomocą wartości $\mu_N(r_{nj})$ określać stopień pewności wystąpienia j -tego uszkodzenia.

Kolejnym istotnym zadaniem systemu diagnostycznego jest lokalizacja występującego uszkodzenia. Zastosowanie rozmytej reprezentacji symptomów wymusza zdefiniowanie rozmytej reprezentacji sygnatur wzorcowych dla każdego identyfikowanego uszkodzenia. Zgodność wyników testów diagnostycznych i sygnatur wzorcowych w przypadku rozmytej reprezentacji otrzymano w wyniku przecięcia mnogościowego odpowiednich zbiorów [11].

4.3.5. Modelowanie rozmyte

W zaproponowanym przez autora diagnostycznym SE wykorzystuje się różne obszary i formy zastosowania regułowej rozmytej wiedzy. Niestety w diagnostyce złożonych procesów taka forma reprezentacji wiedzy zazwyczaj jest niewystarczająca. Wzrost efektywności procesu diagnostycznego uzyskuje się w układach z modelem [1]. Modele obiektów w stanie pełnej zdatności uzyskiwane są na podstawie danych eksperymentalnych z zastosowaniem różnych technik uczenia. Zaletą modeli rozmytych i rozmytych sieci neuronowych jest możliwość połączenia wiedzy eksperta oraz danych pomiarowych [118].

Przetwarzania sygnałów oparte o logikę rozmytą można zaliczyć do metod nowoczesnych. Wykorzystywane są nie tylko do realizacji zadań sterowania ale również pozwalają uzyskiwać zadawalające wyniki w zadaniach modelowania nieliniowych procesów [63, 99, 243]. Własności obiektów przemysłowych można modelować za pomocą różnego typu modeli. Do znanych zaliczyć należy modele wejściowo-wyjściowe, które można zrealizować bazując na elementach sztucznej inteligencji

takich jak: sztuczne sieci neuronowe (ANN), zbiory rozmyte i systemy ekspertowe [37, 184]. Korzyścią wynikającą z zastosowania zbiorów rozmytych, zamiast ANN, jest mniejsza liczba zbiorów danych pomiarowych, niezbędna do utworzenia dobrego modelu nieliniowego procesu [176, 183, 211].

Model rozmyty wykazuje jednakże kilka stopni swobody, które wymagają zdefiniowania w procesie strojenia modelu. Do najważniejszych należy zaliczyć: kształty krzywych rozmywających [152], liczbę i rozmieszczenie zbiorów rozmytych stosowanych w procesie rozmywania wejść i wyjść, zastosowane operatory agregacji przesłanek i wnioskowania [202, 182].

Definiując strukturę rozmytego modelu, dokonano analizy typu zastosowanego rozmytego modelu. W literaturze [40, 202, 211] wyróżnić można dwa typy modeli: Mamdaniego i TSK (Takagi-Sugeno-Kanga).

W modelu Mamdaniego stosowana jest klasyczna rozmyta reprezentacja odpowiedzi, przy zastosowaniu od kilku do kilkunastu, a niekiedy kilkudziesięciu zbiorów rozmytych opisanych funkcjami przynależności różnego typu (najczęściej jest to singleton, krzywe typu trójkątnego lub krzywe Gaussa) [40].

Model TSK, nazywany modelem liniowym, bazuje na opisie wyjścia za pomocą funkcji liniowych stosowanych w konkluzjach poszczególnych reguł [216].

4.3.5.1 Struktury i parametry rozmytych modeli

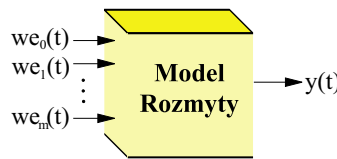
Dla systemów wielowymiarowych typu MISO (ang. *Multi Input Single Output*) liczba parametrów podlegająca strojeniu niejednokrotnie przekracza możliwości dostępnych danych pomiarowych. W takiej sytuacji w pracy tej zaproponowano stosowanie struktur uproszczonych lub rozproszonych [200, 201, 210].

W dalszej części monografii przedstawiono trzy struktury modeli rozmytych [190, 191]:

- struktura kompleksowa,
- struktura równoległa,
- struktura kaskadowa.

Struktura kompleksowa reprezentuje klasyczny model wejściowo-wyjściowy, w którym wiedzę reprezentuje się w formie rozmytej (rys. 4.2). Wektor wejść opisuje się za pomocą określonej w etapie wstępnym liczby zbiorów rozmytych stosowanych w procesie rozmywania wyostrzonych wartości wejść. Liczba stosowanych w procesie rozmywania zbiorów, ich rozmieszczenie oraz kształty funkcji przynależności

definiuje się niezależnie dla każdego wejścia. Baza wiedzy może być przedstawiona w postaci wielowymiarowej tablicy. Wymiar tablicy zależy od liczby sygnałów wejściowych. Rozmiary indeksów elementów tablicy definiowane są podczas określania ilości zbiorów rozmytych stosowanych w trakcie rozmywania wybranego wejścia. Dla układów o wielu wejściach powstaje problem wyznaczenia zawartości dużej liczby komórek tablic decyzyjnych. Ponadto dla tablic o dużych rozmiarach stosować należy rozbudowany algorytm wyostrzania rozmytej reprezentacji odpowiedzi modelu. Dla obiektów o wielu wejściach o podobnym stopniu intensywności wpływu na odpowiedź modelu w większości przypadków jest to jedyna możliwa do zastosowania struktura.



Rys. 4.2. Kompleksowa struktura modelu obiektu wielowymiarowego

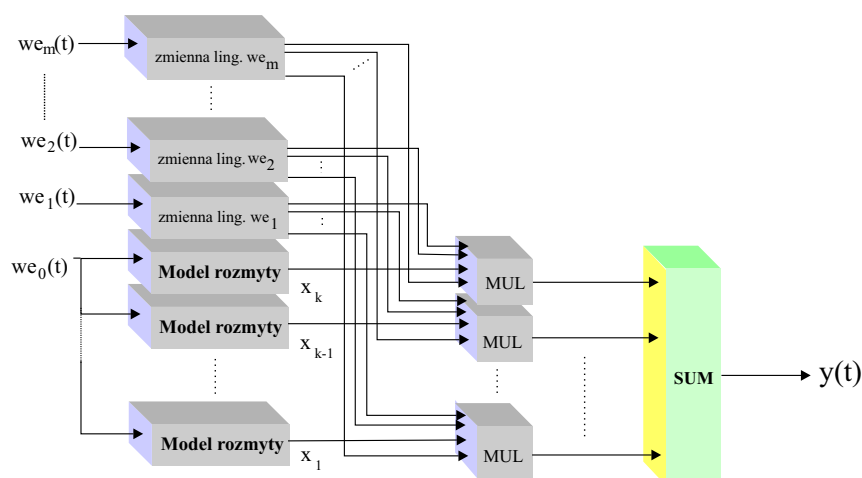
Struktura równoległa stanowi uproszczoną reprezentację zdekomponowanej struktury kompleksowej (rys. 4.3). Istotą tej struktury jest wyróżnienie różnej wrażliwości odpowiedzi obiektu na analizowane sygnały wejściowe. Pozwala to na zredukowanie liczby zbiorów rozmytych stosowanych do opisu wybranego wejścia. Omawiana struktura bazuje na grupach modeli rozmytych typu SISO (ang. *Single Input Single Output*) realizujących funkcje przetwarzania typu wejście-wyjście dla sygnału wejściowego o maksymalnym wpływie. Pozostałe sygnały wejściowe opisują w formie rozmytej modyfikację podstawowej funkcji przetwarzania.

Wyostrzoną wartość wyjścia modelu o strukturze równoległej opisano za pomocą następującej zależności:

$$y = \sum_{i=1}^k \prod_{j=1, l \in L_j}^m \mu_{A_l}(we_j) x_{i,j,l}, \quad (4.15)$$

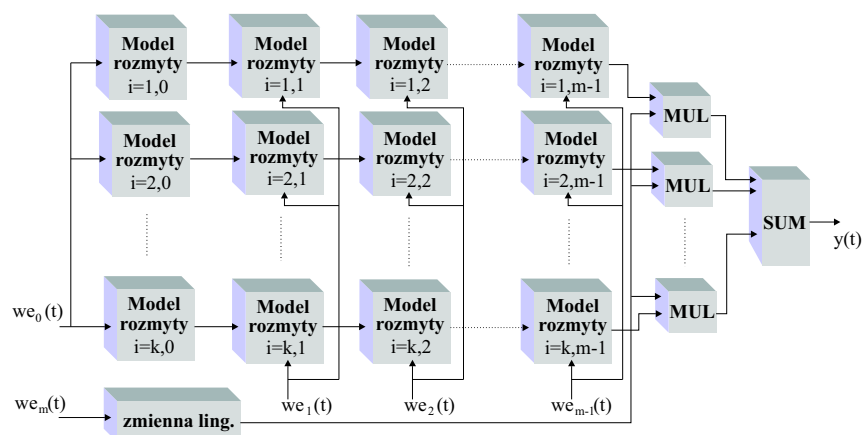
gdzie: $\mu_{A_l}(we_j)$ oznacza wartości funkcji przynależności do zbiorów rozmytych opisujących j -ty sygnał wejściowy we_j , $x_{i,j,l}$ oznacza odpowiedź prostego modelu rozmytego (aproksymującego charakterystykę obiektu dla wybranych wartości lingwistycznych opisujących wejścia pomocnicze), j indeks wejścia pomocniczego, m liczba wejść pomocniczych, l indeks funkcji przynależności do zbioru rozmytego A , L_j liczba zbiorów rozmytych opisujących j -te wejście, i indeks prostego rozmytego modelu oraz $k = L_1 * L_2 * \dots * L_m$ liczba prostych rozmytych modeli.

Struktura kaskadowa stanowi kolejną uproszczoną reprezentację zdekomponowanej struktury kompleksowej (rys. 4.4). Istotą tej struktury jest wyróżnienie grup



Rys. 4.3. Równoległa struktura modelu obiektu wielowymiarowego

rodzin charakterystyk, które aproksymuje rozmyty model. Kaskadowa struktura znajduje zastosowanie dla układów wielowejsciowych, w których można wyróżnić różne stopnie wrażliwości odpowiedzi obiektu na poszczególne wejścia.

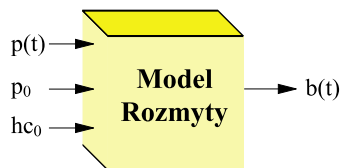


Rys. 4.4. Kaskadowa struktura modelu obiektu wielowymiarowego

Struktura kaskadowa modelu rozmytego składa się z kilku prostych modeli jednowejsciowych, z dominującym wejściem we_0 , aproksymujących charakterystykę reprezentującą określoną rodzinę charakterystyk definiowanych dla określonych wartości lingwistycznych pozostałych wejść pomocniczych. W kolejnych etapach przetwarzania sygnałów wejściowych stosuje się bloki wiążące odpowiedź modelu prostego z wartościami pomocniczych wejść. W tych blokach następuje przesunię-

cie wstępnie wybranej charakterystyki dla określonych wartości lingwistycznych wejść pomocniczych w kierunku zdefiniowanym aktualną wartością tych sygnałów wyjściowych. W systemie zastosowano kilka bloków, każdy dla wstępnie wybranej wartości lingwistycznej wejścia o najmniejszym wpływie na odpowiedzi obiektu w_{em} . Ostatnie proste bloki realizują proste operacje *t-normy* i *s-normy*. Pełnią one funkcje ważenia odpowiedzi wcześniej omówionych bloków w zależności od aktualnej wartości wejścia w_{em} .

Do oceny porównywanych struktur rozmytych modeli pod względem jakościowym i ilościowym przyjęto obiekt przemysłowy o trzech wejściach (komora reaktora termicznego stosowanego w instalacji chłodniczej bloku elektrowni jądrowej) (rys. 4.5). Sygnał główny $p(t)$ opisuje przebieg ciśnienia stosowanego medium. Dwa pozostałe sygnały zawierają informacje o stanie procesu w chwili wystąpienia zakłócenia pracy procesu (rozruch, wyłączenie, uszkodzenie). Dla reaktora stan początkowy analizy poprawności pracy obiektu można opisać za pomocą ciśnienia początkowego p_0 i początkowego poziomu chłodziwa hc_0 . Omawiany model generuje wartość współczynnika wzmocnienia liniowego obserwatora procesu zachodzącego w reaktorze.



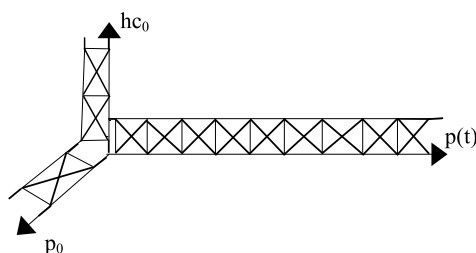
Rys. 4.5. Kompleksowy model reaktora termicznego

MODEL O STRUKTURZE KOMPLEKSOWEJ

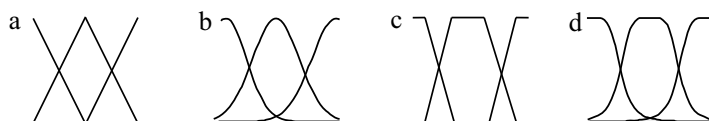
W strukturze kompleksowej przyjęto, że wektor $(p(t), p_0, hc_0)$ stanowi wejście do bloku zawierającego wiedzę o transformacji wejścia w wyjście [201]. W celu przygotowania rozmytej reprezentacji wektora wejściowego należy przeprowadzić operację rozmywania sygnałów obiektowych.

Uwzględniając różną wrażliwość procesu na poszczególne wejściowe sygnały w przeprowadzonych analizach zaproponowano zastosować różne ilości zbiorów rozmytych w rozmywaniu poszczególnych sygnałów wektora wejściowego (rys. 4.6).

Sygnał główny $p(t)$ rozmyto przy użyciu kilkunastu zbiorów rozmytych, sygnał pomocniczy ciśnienie początkowe p_0 rozmyto za pomocą dwóch zbiorów rozmytych (*niskie*, *wysokie*) oraz sygnał pomocniczy, poziom w stanie początkowym hc_0 rozmyto za pomocą trzech zbiorów rozmytych (*niski*, *średni*, *duży*).



Rys. 4.6. Rozmywanie sygnałów wejściowych



Rys. 4.7. Funkcje przynależności stosowane w badaniach autora [202]

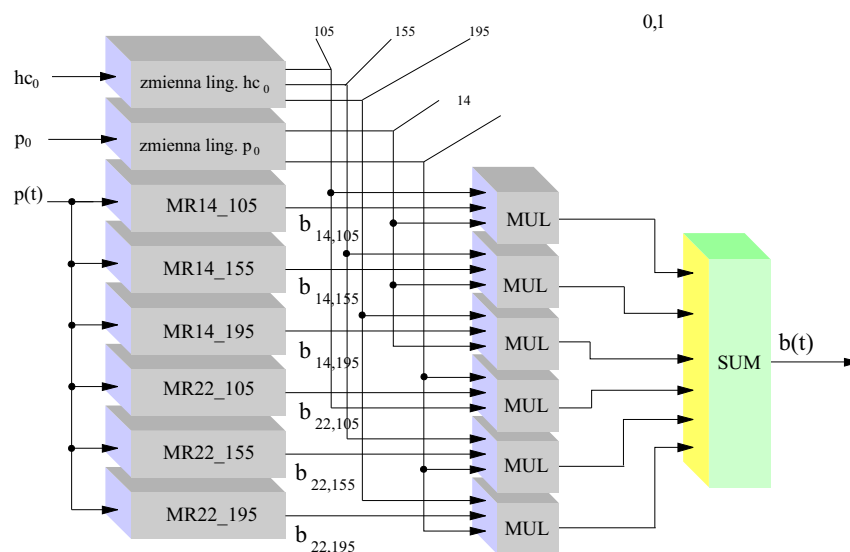
Ważnym zagadnieniem procesu modelowania rozmytego jest dobór kształtu krzywych stosowanych w procesie rozmywania sygnałów wejściowych. Przeprowadzona przez autora analiza wpływu kształtu funkcji przynależności obejmowała najczęściej spotykane w literaturze [202, 243] funkcje przynależności (rys. 4.7): trójkątną (a), Gaussa (b), trapezową (c), dzwonową (d).

MODEL ROZMYTY O STRUKTURZE RÓWNOLEGŁEJ

Zaproponowana przez autora struktura równoległa modelu rozmytego (rys. 4.8) składa się z:

- Prostych jednowymiarowych modeli o wejściu głównym $p(t)$ aproksymujących charakterystykę dla określonego, za pomocą pomocniczych danych wejściowych p_0 i hc_0 , punktu pracy modelowanego procesu;
- Specjalnego rozmytego modelu dla wyznaczenia wartości funkcji przynależności zmiennych pomocniczych p_0 i hc_0 ;
- Dodatkowego algorytmu do określenia wartości sygnału wyjściowego. Algorytm ten określa wagowy udział odpowiedzi poszczególnych prostych jednowymiarowych modeli w zależności od punktu początkowego opisanego danymi pomocniczymi p_0 i hc_0 .

Na rys. 4.8 $MR_{p_0_{hc_0}}$ oznacza model rozmyty dla danych warunków początkowych, p_0 ciśnienie początkowe (w eksperymentach przyjęto 14 i 22 barów) oraz hc_0 początkową wartość opadającego poziomu (w eksperymentach przyjęto 105, 155 i 195 cm).



Rys. 4.8. Model rozmyty o strukturze równoległej

Wyostrzoną wartość wyjścia modelu o strukturze równoległej opisano za pomocą następującej zależności:

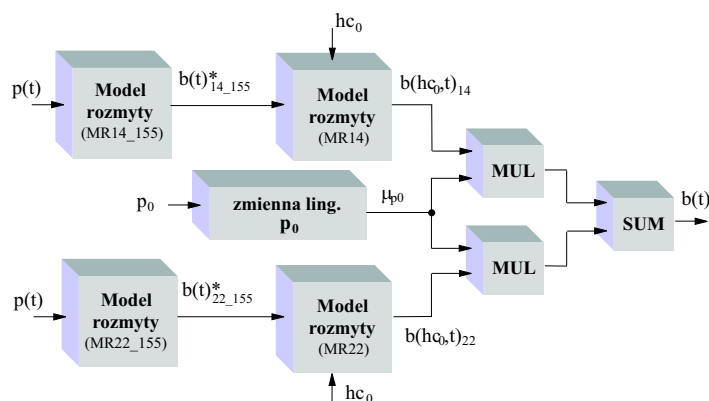
$$b = \sum \mu_{p_0} \mu_{hc_0} b_{p_0_{hc_0}}, \quad (4.16)$$

gdzie: μ_{p_0} , μ_{hc_0} oznaczają wartości funkcji przynależności do zbiorów rozmytych opisujących sygnały wejściowe p_0 i hc_0 , a $b_{p_0_{hc_0}}$ oznacza odpowiedź prostego modelu rozmytego (aproxymującego charakterystykę obiektu dla wybranych warunków początkowych).

MODEL ROZMYTY O STRUKTURZE KASKADOWEJ

Struktura kaskadowa modelu rozmytego [46] (rys.4.9) składa się z następujących bloków:

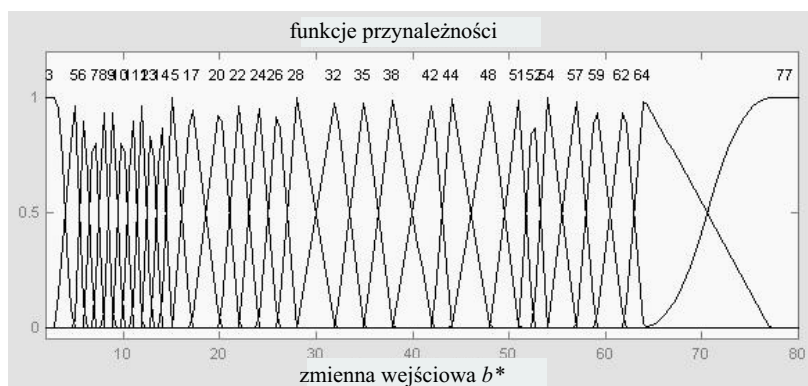
- Prostych modeli jednowejściowych z wejściem $p(t)$ aproxymujących charakterystykę reprezentującą określoną rodzinę charakterystyk definiowanych dla określonych warunków początkowych p_0 i hc_0 . W omawianym układzie zastosowano dwa modele, każdy dla charakterystycznej wartości p_0 ;
- Bloku wiążącego odpowiedź modelu prostego z wartością początkowego poziomu hc_0 . W tym bloku następuje przesunięcie wstępnie wybranej charakterystyki dla określonej wartości hc_0 w kierunku zdefiniowanym aktualną wartością początkowego poziomu hc_0 . W systemie zastosowano dwa bloki, każdy dla wstępnie wybranej wartości początkowego poziomu p_0 ;



Rys. 4.9. Rozmyty model o strukturze kaskadowej z dwoma równoległymi torami sygnałowymi, gdzie: $b(t)_{p_0, hc_0}$ oznacza odpowiedź pierwszego bloku dla określonych warunków początkowych p_0 i hc_0 , MR_{p_0} drugi blok kaskady dla warunku początkowego p_0 oraz $b(hc_0, t)_{p_0}$ odpowiedź toru sygnałowego dla warunku początkowego p_0

- Ostatnie trzy proste bloki realizują proste operacje *t-normy* i *s-normy*. Pełnią one funkcje ważenia odpowiedzi dwóch wcześniej omówionych bloków w zależności od aktualnej wartości p_0 .

Bloki jednowejściowe mają strukturę i parametry identyczne z prostymi lokalnymi modelami struktury równoległej wyznaczonymi dla określonych warunków początkowych p_0 i hc_0 .

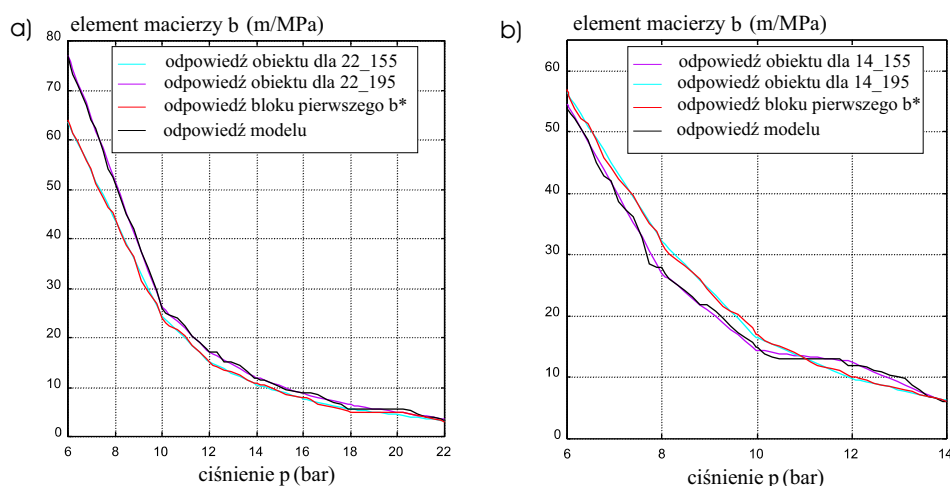


Rys. 4.10. Rozmywanie wejścia b_{p_0, hc_0} dla bloku drugiego obu kaskad

Drugi blok każdej kaskady realizuje zadanie aktualizacji aproksymowanej cha-

rakterystyki. Kierunek przesunięcia wzorcowej charakterystyki zależy od wartości sygnału warunków początkowych poziomu hc_0 . Przy realizacji drugiego bloku zastosowano skomplikowany proces rozmywania wejścia $b_{p_0_{hc_0}}$ (rys. 4.10). Podyktowane to zostało potrzebą wygładzenia odpowiedzi drugiego bloku. Pozostała grupa trzech prostych bloków może być porównywana z algorytmem ważenia wprowadzonym w strukturze równoległej, z tą różnicą, że proces ważenia struktury kaskadowej oparto na rozmytej wartości sygnału początkowego p_0 .

Wyjścia obu kaskad są używane do określenia odpowiedzi rozmytego modelu. Algorytm ważenia może być interpretowany jako rozmyta interpolacja między rozmytymi wartościami początkowego ciśnienia *niskie* i *wysokie*. Aby pokazać zasadę pracy modelu o strukturze kaskadowej na rys. 4.11 przedstawiono odpowiedzi bloków wewnętrznych modelu.



Rys. 4.11. Odpowiedzi pierwszego i drugiego bloku dla $p_0 = 22 \text{ bar}$ i $hc_0 = 195 \text{ mm}$ (a) oraz $p_0 = 14 \text{ bar}$ i $hc_0 = 105 \text{ mm}$ (b)

4.3.5.2 Strojenie rozmytych modeli

Uzyskanie dobrej jakości modelowania wymaga zestrojenia parametrów wstępnie zdefiniowanego rozmytego modelu. Strojenie modelu jest zadaniem zaliczanym do grupy zadań optymalizacji wieloparametrycznej. Optymalizację tak postawionego zadania można wykonać stosując klasyczne metody gradientowe lub metody przeszukiwania bezpośredniego.

W początkowej fazie zaproponowano przeprowadzenie wstępnej analizy wpły-

wu poszczególnych parametrów zaproponowanych struktur rozmytych modeli metodą kolejnych prób według kryterium optymalizacji zdroworozsądkowej (historycznie pierwsze techniki).

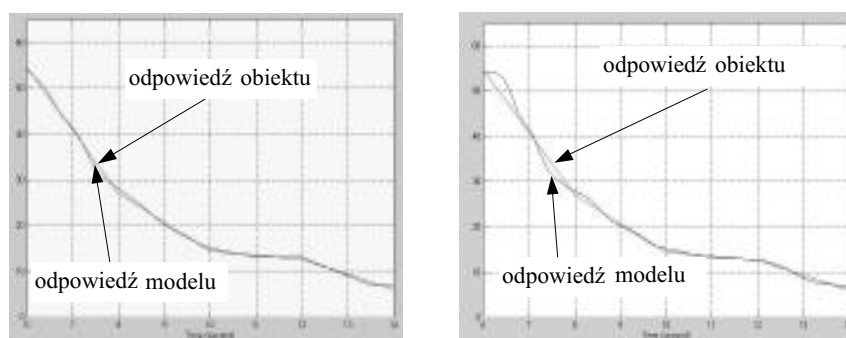
Do oceny jakości aproksymacji charakterystyk obiektu zastosowano kryterium sumy kwadratu błędu, SSE (ang. *Sum Square Error*) określonego zależnością:

$$e = \sum_k [b_R(k) - b_{FM}(k)]^2, \quad (4.17)$$

gdzie: b_R oznacza wartość oczekiwaną, b_{FM} - odpowiedź rozmytego modelu oraz k numer próbki.

MODEL O STRUKTURZE KOMPLEKSOWEJ

Proces strojenia rozpoczęto od badań wpływu kształtu krzywych rozmywających. Przeprowadzono je dla zadania rozmywania sygnału głównego $p(t)$ [201].



Rys. 4.12. Przebiegi odpowiedzi modelu rozmytego z trójkątnymi (a) i dzwonowymi (b) funkcjami przynależności

Na rysunkach 4.12a i 4.12b przedstawiono odpowiedzi modelu rozmytego o strukturze kompleksowej dla rozmywania wejścia $p(t)$ przy użyciu funkcji przynależności typu trójkątnego i dzwonowego [189]. W tabeli 4.1 zamieszczono porównanie wartości błędu SSE dla różnych kształtów funkcji przynależności stosowanych w procesie rozmywania wejścia $p(t)$. Badania przeprowadzono dla różnych warunków początkowych p_0 , hc_0 procesu.

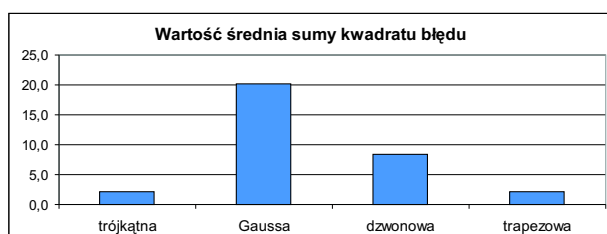
Wartość średnia w tabeli 4.1 określa średnią arytmetyczną błędu uzyskanego dla danego kształtu funkcji przynależności przy różnych warunkach początkowych badanego procesu. Analiza wyników pozwala wnioskować, że przy rozmywaniu sy-

Tabela. 4.1. Suma kwadratu błędu dla różnych kształtów funkcji przynależności

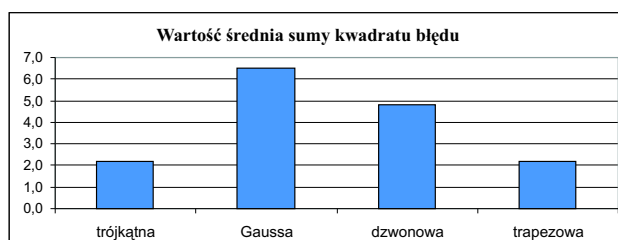
krzywa	Model						średnio
	MR14_105	MR14_155	MR14_195	MR22_105	MR22_155	MR22_195	
trójkątna	3.05	1.25	1.97	2.34	2.73	3.44	2.46
Gausa	8.36	8.47	14.65	5.40	7.38	6.91	8.53
dzwonowa	15.18	14.67	24.76	11.89	13.60	17.16	16.21
trapezowa	26.99	25.75	42.02	24.10	25.41	36.78	30.18

gdzie: $MRp_0_hc_0$ oznacza model rozmyty dla warunków początkowych p_0 , hc_0 .

gnałów wejściowych zaleca się stosować krzywe o skończonym nośniku typu trójkątnego.



Rys. 4.13. Diagram wpływu kształtu krzywych rozmywających wejście p_0 na dokładność aproksymacji nieliniowych charakterystyk obiektu



Rys. 4.14. Diagram wpływu kształtu krzywych rozmywających wejście hc_0 na dokładność aproksymacji nieliniowych charakterystyk obiektu

Po przeprowadzeniu badań technik rozmywania wejść pomocniczych p_0 i hc_0 uzyskano podobne wyniki jak dla wejścia $p(t)$ (rys. 4.13 i 4.14). W tym przypadku również zaleca się stosowanie funkcji przynależności typu trójkątnego, chociaż podobną dokładność aproksymacji uzyskano dla funkcji przynależności typu trapezowego.

Podobne badania przeprowadzono w celu zdefiniowania wpływu kształtu funk-

cji przynależności opisujących wyjście modelu Mamdaniego. Wyniki badań przedstawiono w tabeli 4.2. Na podstawie analizy wyników stwierdzono, że zmiana typu funkcji przynależności rozmywających wyjście modelu ma niewielki wpływ na dokładność aproksymacji.

Tabela. 4.2. Wartości błędu SSE dla różnych form rozmywania wyjścia modelu o strukturze kompleksowej

krzywa	Model						średnio
	MR14_105	MR14_155	MR14_195	MR22_105	MR22_155	MR22_195	
trójkątna	3.05	1.25	1.97	2.34	2.73	3.44	2.46
Gaussa	3.11	1.32	2.02	2.47	2.74	3.54	2.53
dzwonowa	2.77	0.96	1.82	2.11	2.22	3.09	2.16
trapezowa	2.89	0.99	1.79	2.06	2.62	3.02	2.23

Kolejny stopień swobody rozmytego modelu obejmuje zagadnienie wyboru typu modelu. W tym celu przeprowadzono badania dokładności aproksymacji dla modelu Mamdaniego i TSK. Wyniki przedstawiono w tabeli 4.3.

Tabela. 4.3. Porównanie dokładności aproksymacji charakterystyk obiektu nieliniowego dwóch typów rozmytych modeli

Typ modelu	Model Mamdaniego o strukturze kompleksowej					
Punkt pracy	MR14_105	MR14_155	MR14_195	MR22_105	MR22_155	MR22_195
Błąd SSE	2.29	1.16	3.05	1.86	2.70	2.20
Typ modelu	Model TSK o strukturze kompleksowej					
Punkt pracy	MR14_105	MR14_155	MR14_195	MR22_105	MR22_155	MR22_195
Błąd SSE	25.07	33.01	25.74	61.03	91.58	75.81

Analiza wyników pozwala wnioskować o dużej dokładności aproksymacji uzyskanej w układzie z modelem Mamdaniego. Model TSK aproksymuje nieliniowe charakterystyki obiektu z błędem SSE o rząd większym. W konkluzjach bazy wiedzy modelu TSK zastosowano 14 funkcji liniowych.

Mała dokładność aproksymacji przy użyciu modeli TSK może wynikać z dużej wrażliwości modelu na zmiany parametrów funkcji liniowych opisujących wyjścia. Strojenie tego typu modelu metodą kolejnych przybliżeń jest mało efektywne przy tak dużej liczbie parametrów modelu.

Dla aproksymacji nieliniowych charakterystyk za pomocą modeli rozmytych ważnym czynnikiem jest rozmieszczenie zastosowanych w procesie rozmywania

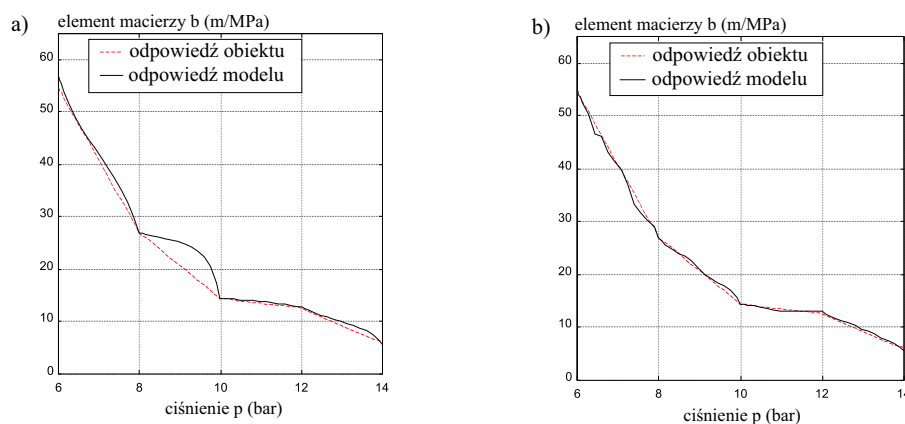
funkcji przynależności. Przeprowadzone badania wskazują na potrzebę asymetrycznego rozmieszczenia tych funkcji. Zagęszczenie ich powinno być stosowane w obszarach gdzie aproksymowana charakterystyka wykazuje duże zmiany nachylenia.

MODEL O STRUKTURZE RÓWNOLEGŁEJ

W przypadku struktury równoległej zadanie strojenia parametrów modelu autor proponuje realizować dwuetapowo. W pierwszym etapie strojeniu podlega każdy prosty model niezależnie. W drugim etapie strojone są parametry bloków rozmywających wejścia pomocnicze p_0 i hc_0 .

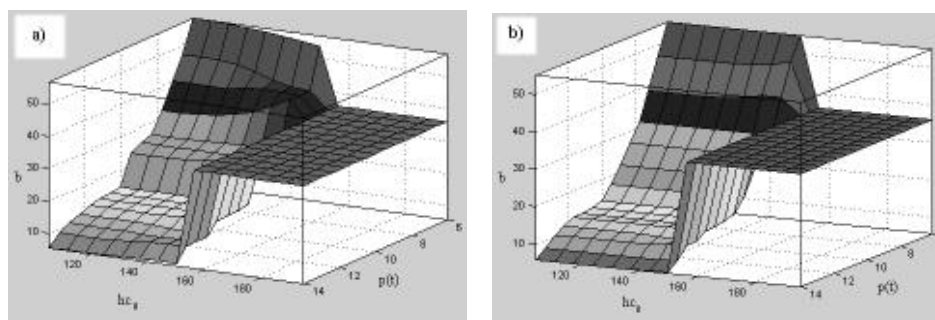
Rozmywanie sygnału głównego dla prostych modeli wymaga kilku lub kilkunastu zbiorów rozmytych. Podobnie jak w przypadku modelu kompleksowego, dla modelu prostego należy określić sposób rozmieszczenia funkcji przynależności użytych w procesie rozmywania wejścia głównego $p(t)$.

Na rys. 4.15 pokazano wpływ rozmieszczenia funkcji przynależności w procesie rozmywania $p(t)$ na dokładność aproksymacji nieliniowej charakterystyki. Zagęszczenie tych funkcji należy stosować w miejscach załamania aproksymowanej charakterystyki. Większe dopasowanie dwuwymiarowej charakterystyki modelu (po zastosowaniu dwóch sygnałów wejściowych: $p(t)$ i hc_0) uzyskano dla asymetrycznego rozmywania wejścia głównego $p(t)$ (rys. 4.16).



Rys. 4.15. Odpowiedzi prostego rozmytego modelu z symetrycznym (a) i niesymetrycznym rozmieszczeniem funkcji przynależności w procesie rozmywania wejścia głównego $p(t)$

Kształt krzywych rozmywających wejście modelu prostego stanowi ważne zagadnienie, które należy zdefiniować na etapie projektowania rozmytego modelu. W przeprowadzonych przez autora badaniach analizy kształtu funkcji rozmywających



Rys. 4.16. Dwuwymiarowa charakterystyka dla symetrycznego (a) i niesymetrycznego (b) rozmywania wejścia głównego $p(t)$

uzyskano wyniki zbliżone do przedstawionych dla modelu kompleksowego.

MODEL O STRUKTURZE KASKADOWEJ

Dla struktury kaskadowej zaproponowano zadanie strojenia parametrów modelu wykonać również dwuetapowo. W etapie pierwszym należy stroić parametry pierwszych bloków (bloki strojone w oparciu o algorytm stosowany w strukturze równoległej). Etap drugi obejmuje strojenie parametrów bloku drugiego.

Po wykonaniu zadania strojenia parametrów przeprowadzono badania określające dokładność aproksymacji charakterystyk obiektu. Ponownie osiągnięto wyniki zbliżone do zamieszczonych w tabeli 4.1 oraz 4.2.

4.3.6. Optymalizacja struktury i parametrów systemów rozmytych

Systemy rozmyte w klasycznej strukturze niestety nie mają zdolności uczenia się. We wstępnej fazie badań nad możliwością modelowania rozmytego rozkład parametrów dobierany był w oparciu o wiedzę eksperta, wiedzy zdroworozsądkowej inżyniera wiedzy lub suboptymalnego rozkładu parametrów modeli przygotowanych wcześniej dla podobnych problemów [204].

W ostatnich latach do strojenia modeli rozmytych stosuje się, w większości przypadków z dobrym skutkiem, techniki sprawdzone przy strojeniu sieci neuronowych (sieci neuro-rozmyte) oraz algorytmy ewolucyjne. Sieci neuro-rozmyte stanowią połączenie techniki modelowania rozmytego z metodami uczenia sieci neuronowych [210].

W jednej z technik, zaproponowanej przez Horikawę [81], część lewa odpowia-

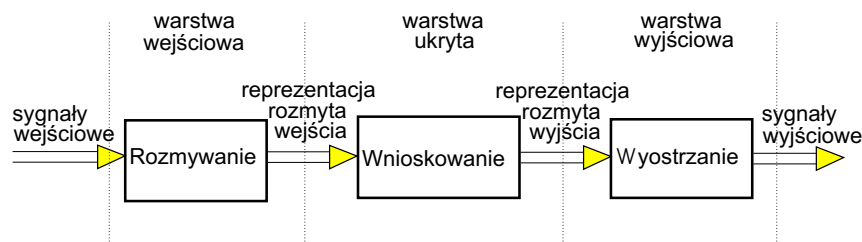
dająca przesłankom ma identyczną strukturę dla wszystkich typów sieci, różnice występują w części prawej realizującej konkluzje. W prawej części tego typu sieci można wyróżnić trzy wersje rozwiązań. Ze względu na elementy zastosowane w tej części wyróżniamy:

- wersja I - wartości stałe (singletony),
- wersja II - równania liniowe (TSK),
- zbiory rozmyte o różnych funkcjach przynależności.

Należy zauważyć, że FNN jest zapisem procesu wnioskowania rozmytego w postaci struktury sieci neuronowej, której parametry są reprezentowane za pomocą wag połączeń. W odróżnieniu od sieci neuronowych FNN nie jest *czarną skrzynką*, można ją z powodzeniem rozpisać na zbiór reguł rozmytych i rozpatrywać jak zwykły rozmyty model.

Obserwując burzliwy rozwój sieci neuronowych, można zauważyć, że w dużej mierze spowodowany on jest podstawową cechą sieci - zdolnością uczenia się z przykładów. Z drugiej strony badanie zasad wnioskowania systemów z rozmytą reprezentacją wiedzy pozwala zauważyć dużą zbieżność obu systemów - neuronowego i rozmytego. Układ z rozmytą reprezentacją wiedzy można przedstawić za pomocą sieci o trzech warstwach (rys.4.17):

- » rozmywanie wejść - warstwa wejściowa,
- » wnioskowanie - warstwa ukryta,
- » wyostżanie - warstwa wyjściowa.



Rys. 4.17. Sieciowa struktura rozmytego modelu

Strojenie parametrów sieci neuronowej wiąże się z doбором wag połączeń między neuronami w poszczególnych warstwach sieci. Najczęściej, w sieciach typu MLP, parametry pozostałych stopni swobody (liczba warstw ukrytych, liczba neuronów w poszczególnych warstwach, funkcje aktywacji neuronu) określa się *a priori*.

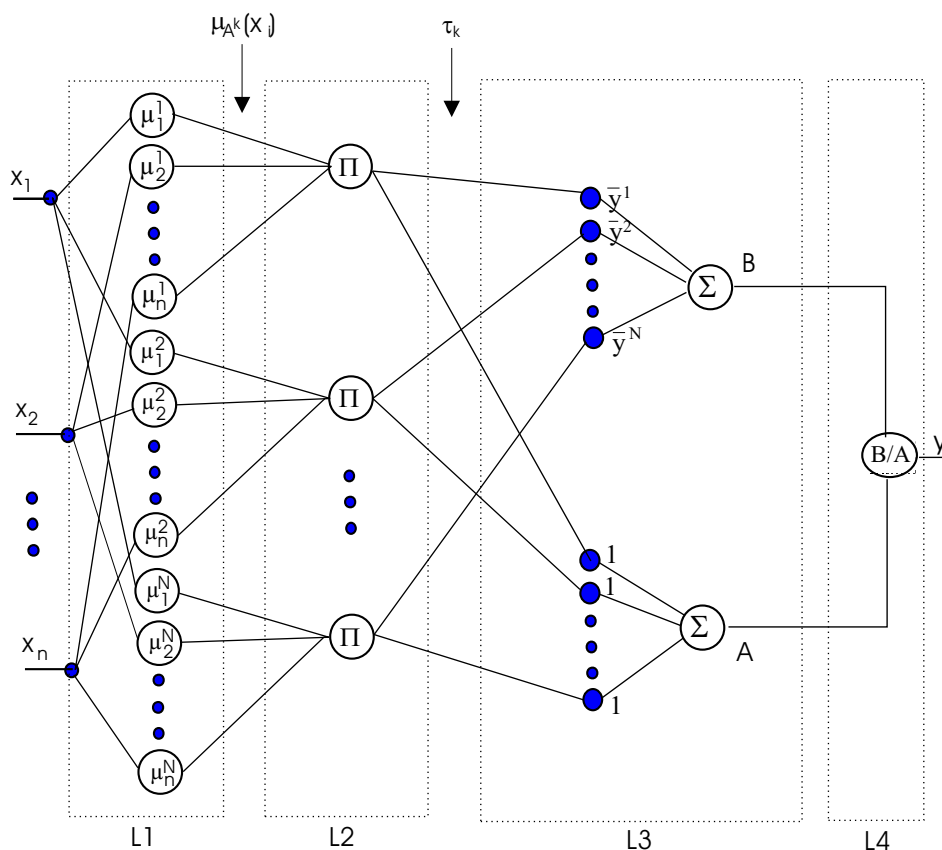
Analogicznie, proces optymalizacji parametrów systemu rozmytego należy rozpocząć od wstępnej definicji wybranych stopni swobody (dla modelu należy określić punkt pracy). Najczęściej dotyczy to typu modelu rozmytego (Mamdani, Larsen), krzywych opisujących proces rozmywania wejść i wyjść oraz zasady wyostrzania rozmytej reprezentacji wyjścia. W dalszej analizie zaproponowano zastosować *a priori* budowę rozmytego modelu typu Larsena (rozmyta implikacja typu iloczyn), rozmywanie wyjścia typu singleton oraz COA, (ang. *Center Of Area*) [243], jako metodę wyostrzania. Otrzymano podstawową strukturę systemu **neuro-rozmytego** (rys.4.18) opisanego zależnością [210]:

$$\bar{y} = \frac{\sum_{k=1}^N \bar{y}^k \prod_{i=1}^n \mu_{A_i^k}(\bar{x}_i)}{\sum_{k=1}^N \prod_{i=1}^n \mu_{A_i^k}(\bar{x}_i)} \quad (4.18)$$

W strukturze przedstawionej na rys. 4.18 wydzielono kilka warstw, ukazując podobieństwo do struktury sieci neuronowych. W warstwie oznaczonej L1 realizuje się operację wyznaczania wartości funkcji przynależności zbiorów rozmytych A_i^k , $i = 1, 2, \dots, n$, $k = 1, 2, \dots, N$, gdzie: n oznacza liczbę wejść, a N - liczbę reguł. Na wyjściach warstwy L1 występuje rozmyta reprezentacja sygnałów wejściowych w postaci wartości funkcji przynależności poszczególnych wejść do zdefiniowanych dla każdego wejścia zbiorów rozmytych. W przedstawionej na omawianym rysunku strukturze przyjęto jednakową liczbę zbiorów rozmytych dla każdego wejścia. W przypadku ogólnym może wystąpić przypadek rozmywania każdego wejścia inną liczbą zbiorów rozmytych. W warstwie L2 wykonywana jest operacja agregacji przesłanek poszczególnych reguł zawartych w bazie wiedzy rozmytego modelu. W przypadku modelu Larsena wykonywana jest operacja typu *prod*. Wyjście warstwy L2 opisuje zatem poziomy zapłonu poszczególnych reguł. W warstwie L3 i L4 realizowana jest operacja wyostrzania rozmytej reprezentacji wyjścia. Struktura tych warstw uzależniona jest od przyjętej metody wyostrzania.

Dobre rezultaty optymalizacji opartej na gradiencie funkcji kosztu uzyskano stosując techniki używane przy optymalizacji parametrów sieci neuronowych. Powstałe w ten sposób układy, opisane w reprezentacji rozmytej i strojone w oparciu o metody gradientowe, nazywane są układami neuro-rozmytymi [133, 134, 210, 211].

Sieci neuronowe skutecznie można uczyć, czyli optymalnie dobierać wagi połączeń, stosując metodę wstecznej propagacji błędów. Ideę tę można również przenieść na sieć neuro-rozmytą przedstawioną na rys. 4.18. Podobieństwo tej sieci do sieci neuronowych polega przede wszystkim na warstwowej strukturze. Zatem w sieci neuro-rozmytej można również dokonać propagacji błędów wstecz poprzez kolejne warstwy. Otrzymane w ten sposób zależności rekurencyjne mogą stanowić podstawę do budowy algorytmów strojenia rozmytego modelu. Utworzony algorytm realizuje zadanie optymalnego wyznaczania wartości parametrów funkcji



Rys. 4.18. Neuronowa reprezentacja rozmytego modelu

przynależności na podstawie *ciągu uczącego*, czyli kolejnych par wartości wejściowych x_i oraz odpowiednich wartości wyjściowych y_{ref} .

W algorytmach strojenia zaproponowanych struktur modeli rozmytych zastosowano wartość błędu aproksymacji opisanego zależnością [194]:

$$e = \frac{1}{2} (\bar{y} - y_{ref})^2. \quad (4.19)$$

W procesie strojenia wstępnie arbitralnie dobierano wartości parametrów funkcji przynależności, następnie algorytm uczenia modyfikował tak parametry funkcji przynależności aby błąd e określony wzorem (4.19) przyjmował wartość minimalną. Wartości początkowe często dobiera się także na podstawie wcześniej omówionej techniki zdroworozsądkowej. Modyfikacje wartości parametrów, zgodnie z algorytmem wstecznej propagacji błędu, realizowano w oparciu o wartości poprzednie i wartość pochodnej błędu e względem aktualnie modyfikowanego parametru. Dla

przykładu modyfikacji położenia funkcji przynależności, typu singleton, opisującej rozmytą reprezentację wyjścia \bar{y}^k wykonuje się zgodnie z następującą zależnością:

$$\bar{y}^k(t+1) = \bar{y}^k(t) - \eta \frac{\partial e(t)}{\partial \bar{y}^k(t)}, \quad (4.20)$$

gdzie: t oznacza dany krok algorytmu, natomiast $\eta \in (0, 1)$ jest współczynnikiem definiującym szybkość uczenia.

Dla wcześniej zaproponowanych struktur rozmytych modeli opracowano, w oparciu o własności modeli neuro-rozmytych, algorytmy strojenia automatycznego.

4.3.6.1 Model neuro-rozmyty o strukturze kompleksowej

Rozmyty model o bazie wiedzy przedstawionej na rys. 4.19 może być reprezentowany jako wielowarstwowa sieć neuronowa (rys. 4.20) [195].

Jeśli $p(t)$ jest s_1 i p_0 jest niskie i hc_0 jest niski wtedy b jest D_1
 Jeśli $p(t)$ jest s_2 i p_0 jest niskie i hc_0 jest średni wtedy b jest D_2

 Jeśli $p(t)$ jest s_k i p_0 jest wysokie i hc_0 jest duży wtedy b jest D_n

Rys. 4.19. Baza wiedzy rozmytego modelu o strukturze kompleksowej

Proces wyostrzania odpowiedzi modelu opisano jako:

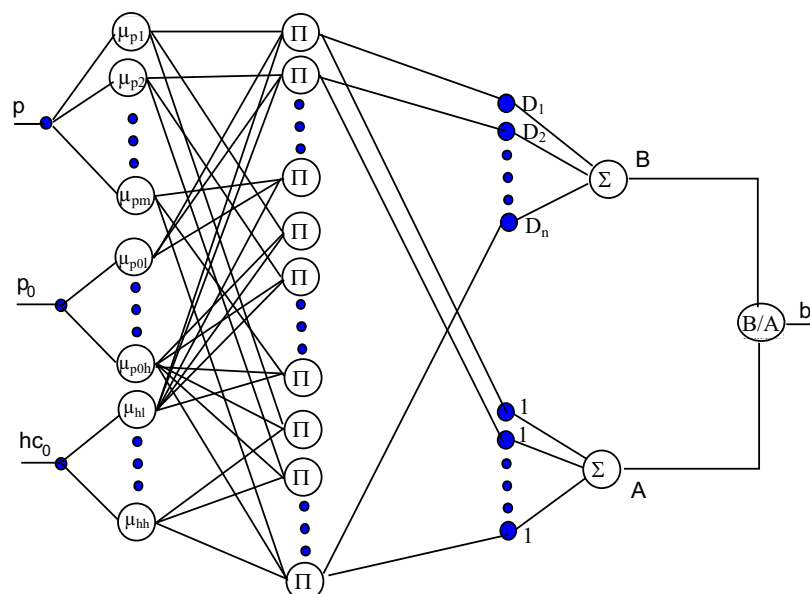
$$b = \frac{\sum_{i=1}^m \bar{y}_i \mu_i(p)}{\sum_{i=1}^m \mu_i(p)} \quad (4.21)$$

Do opisu wyjścia przyjęto krzywe typu singleton [40] (rys. 4.21b), wtedy:

$$\mu_{D_i}(y) = \begin{cases} 1 & \text{dla } \bar{y} = D_i \\ 0 & \text{dla } \bar{y} \neq D_i, \end{cases} \quad (4.22)$$

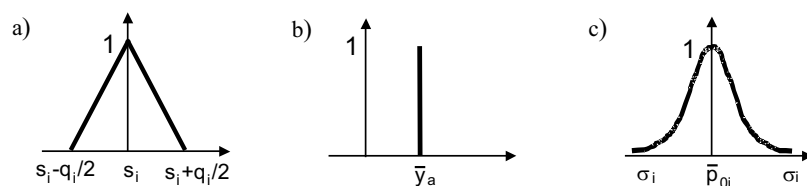
Algorytm strojenia parametrów rozmytego wyjścia D_i opisano jako:

$$\bar{y}_a(t+1) = \bar{y}_a(t) - \eta(b - b_{ref}) \frac{\mu_{aj}(p) \mu_{aj}(p_0) \mu_{aj}(hc_0)}{\sum_{k=1}^n \mu_{kj}(p) \mu_{kj}(p_0) \mu_{kj}(hc_0)}; \quad \eta \in [0, 1]. \quad (4.23)$$



Rys. 4.20. Neuronowa reprezentacja trójwejściowego rozmytego modelu o strukturze kompleksowej

Przynależność wejścia do wybranego zbioru rozmytego opisanego za pomocą funkcji trójkątnej (rys.4.21a).



Rys. 4.21. Kształty funkcji przynależności

Wtedy dla sygnału głównego otrzymano:

$$\mu_i(p) = \begin{cases} 1 - \frac{2|p-s_i|}{q_i} & \text{dla } |p-s_i| \leq \frac{q_i}{2} \\ 0 & \text{w innym przypadku,} \end{cases} \quad (4.24)$$

Algorytm strojenia parametrów zastosowanych trójkątnych funkcji opisano za pomocą zależności:

$$s_i(t+1) = s_i(t) - \eta(b - b_{ref}) fmk \frac{\partial \mu_i(p)}{\partial s_i} | (t) \quad (4.25)$$

$$q_i(t+1) = q_i(t) - \eta(b - b_{ref}) fmk \frac{\partial \mu_i(p)}{\partial q_i} \Big| (t) \quad (4.26)$$

gdzie: funkcja fmk definiuje współczynnik modyfikacji parametrów modelu kompleksowego i opisano ją jako:

$$fmk = \frac{\sum_{\lambda} \bar{y}_{\lambda} \mu_{\lambda_j}(p_0) \mu_{\lambda_j}(hc_0)}{\sum_{k=1}^n \mu_{k_j}(p) \mu_{k_j}(p_0) \mu_{k_j}(hc_0)} - \frac{\sum_{k=1}^n \mu_{k_j}(p) \mu_{k_j}(p_0) \mu_{k_j}(hc_0) \sum_{\lambda} \mu_{\lambda_j}(p_0) \mu_{\lambda_j}(hc_0)}{\left(\sum_{k=1}^n \mu_{k_j}(p) \mu_{k_j}(p_0) \mu_{k_j}(hc_0) \right)^2} \quad (4.27)$$

oraz pochodne cząstkowe funkcji przynależności opisano jako:

$$\frac{\partial \mu_i(p)}{\partial s_i} = \begin{cases} -\frac{1}{q_i} & dla \quad p = s_i - \frac{q_i}{2} \\ -\frac{2}{q_i} & dla \quad p \in (s_i - \frac{q_i}{2}, s_i) \\ 0 & dla \quad p = s_i \\ \frac{2}{q_i} & dla \quad p \in (s_i, s_i + \frac{q_i}{2}) \\ \frac{1}{q_i} & dla \quad p = s_i + \frac{q_i}{2} \\ 0 & dla \quad p \in (-\infty, s_i - \frac{q_i}{2}) \cup (s_i + \frac{q_i}{2}, \infty), \end{cases} \quad (4.28)$$

$$\frac{\partial \mu_i(p)}{\partial q_i} = \begin{cases} \frac{1}{2q_i} & dla \quad p = s_i - \frac{q_i}{2} \\ -\frac{2(s_i - p)}{(q_i)^2} & dla \quad p \in (s_i - \frac{q_i}{2}, s_i) \\ 0 & dla \quad p = s_i \\ \frac{2(p - s_i)}{(q_i)^2} & dla \quad p \in (s_i, s_i + \frac{q_i}{2}) \\ \frac{1}{2q_i} & dla \quad p = s_i + \frac{q_i}{2} \\ 0 & dla \quad p \in (-\infty, s_i - \frac{q_i}{2}) \cup (s_i + \frac{q_i}{2}, \infty). \end{cases} \quad (4.29)$$

Aby porównać złożoność algorytmu strojenia funkcji rozmywających w przypadku zastosowania krzywych Gaussa (4.30) przyjęto, że oba wejścia pomocnicze będą rozmywane przy użyciu funkcji Gaussa:

$$\mu_a(y) = \exp \left(- \left(\frac{y - \bar{y}_a}{\sigma_a} \right)^2 \right). \quad (4.30)$$

Algorytm strojenia parametrów zastosowanych funkcji rozmywających opisano za pomocą następujących zależności:

$$\bar{p}_{0j}(t+1) = \bar{p}_{0j}(t) - \eta(b - b_{ref}) fmk \frac{\partial \mu_j(p_0)}{\partial \bar{p}_{0j}} \Big| (t) \quad (4.31)$$

$$\sigma_{p0j}(t+1) = \sigma_{p0j}(t) - \eta(b - b_{ref}) fmk \frac{\partial \mu_j(p_0)}{\partial \sigma_{p0j}} \Big| (t) \quad (4.32)$$

gdzie: funkcję fmk zdefiniowano równaniem (4.27) oraz pochodne cząstkowe dla sygnału p_0 opisano jako:

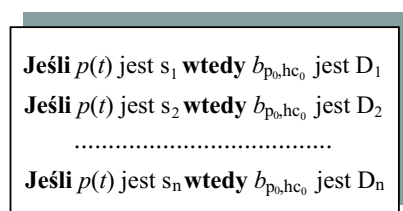
$$\frac{\partial \mu_j(p_0)}{\partial \bar{p}_{0j}} = 2 \frac{\bar{p}_{0j} - p_0}{\sigma_{p0j}^2} \exp \left(- \left(\frac{p_0 - \bar{p}_{0j}}{\sigma_{p0j}} \right)^2 \right), \quad (4.33)$$

$$\frac{\partial \mu_j(p_0)}{\partial \sigma_{p0j}} = 2 \frac{\bar{p}_{0j} - p_0}{\sigma_{p0j}^3} \exp \left(- \left(\frac{p_0 - \bar{p}_{0j}}{\sigma_{p0j}} \right)^2 \right). \quad (4.34)$$

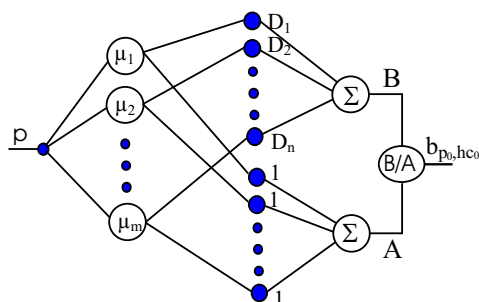
Podobne zależności uzyskano dla drugiego sygnału pomocniczego hc_0 .

4.3.6.2 Model neuro-rozmyty o strukturze równoległej

Rozmyty model prosty o bazie wiedzy przedstawionej na rys. 4.22 może być reprezentowany jako sieć neuronowa z jedną warstwą ukrytą (rys. 4.23). W przedstawionej strukturze (rys. 4.23) bloki sumowania i dzielenia realizują zadanie wyostrzania odpowiedzi modelu przy zastosowaniu metody środka ciężkości, COA.



Rys. 4.22. Baza wiedzy rozmytego modelu prostego



Rys. 4.23. Neuronowa reprezentacja jednowarstwowej prostej rozmytego modelu, gdzie: $p(t)$ oznacza wejście, a μ_i - wartości funkcji przynależności do i -tego zbioru rozmytego

Proces wyostrzania odpowiedzi opisano zależnością:

$$b_{p_0_{-}hc_0} = \frac{\sum_{i=1}^n \bar{y}_i \mu_i(p)}{\sum_{i=1}^n \mu_i(p)} \quad (4.35)$$

W dalszych analizach przyjęto krzywą typu singleton do reprezentowania rozmytego wyjścia modelu (4.22). Rozmieszczenie krzywych typu singleton wyznaczono korzystając ze wzoru:

$$\bar{y}_i(t+1) = \bar{y}_i(t) - \eta \frac{b_{p_0_{-}hc_0} - b_{p_0_{-}hc_0,ref}}{\sum_{k=1}^n \mu_k} \mu_i; \quad \eta \in [0, 1]. \quad (4.36)$$

Wejścia w rozmytej reprezentacji opisano za pomocą trójkątnej funkcji przynależności (rys.4.21a). Algorytm strojenia parametrów zastosowanych funkcji opisano za pomocą zależności:

$$s_i(t+1) = s_i(t) - \eta \frac{(b_{p_0_{-}hc_0} - b_{p_0_{-}hc_0,ref})(\bar{y}_i - b_{p_0_{-}hc_0})}{\sum_{k=1}^n \mu_k(p)} \frac{\partial \mu_i(p)}{\partial s_i} \Big| (t) \quad (4.37)$$

$$q_i(t+1) = q_i(t) - \eta \frac{(b_{p_0_{-}hc_0} - b_{p_0_{-}hc_0,ref})(\bar{y}_i - b_{p_0_{-}hc_0})}{\sum_{k=1}^n \mu_k(p)} \frac{\partial \mu_i(p)}{\partial q_i} \Big| (t) \quad (4.38)$$

gdzie: pochodne cząstkowe $\frac{\partial \mu_i(p)}{\partial s_i}$, $\frac{\partial \mu_i(p)}{\partial q_i}$ dla trójkątnych funkcji przynależności posiadają analogiczną postać jak (4.28) i (4.29).

Sygnał wyjścia rozmytego modelu o strukturze równoległej opisano zależnością:

$$b = \sum_{(R)} \mu_{p_0} \mu_{hc_0} b_{p_0_{-}hc_0}. \quad (4.39)$$

Kolejnym krokiem w procesie strojenia rozmytego modelu jest zoptymalizowanie wartości parametrów funkcji rozmywających wejścia pomocnicze modelu p_0 i hc_0 . Podobnie jak dla układu o strukturze kompleksowej przynależność wejść pomocniczych p_0 i hc_0 do zbiorów rozmytych opisano za pomocą krzywych Gaussa (4.30). Zastosowany proces strojenia opisano zależnościami:

$$\bar{p}_{0j}(t+1) = \bar{p}_{0j}(t) - fmr_j \frac{\bar{p}_{0j} - p_0}{(\sigma_{p_{0j}})^2} \Big| (t), \quad (4.40)$$

$$\sigma_{p_{0j}}(t+1) = \sigma_{p_{0j}}(t) - fmr_j \frac{\bar{p}_{0j} - p_0}{(\sigma_{p_{0j}})^3} \Big| (t), \quad (4.41)$$

gdzie: funkcja fmr_j definiuje współczynnik modyfikacji parametrów modelu równoległego i opisano ją jako:

$$fmr_j = \eta(b - b_{ref}) \sum_{l=1}^w \mu_l(hc_0) \sum_{a=1}^v b_{p_0_{-}hc_0} \mu_a(p_0) \exp \left[- \left(\frac{p_0 - \bar{p}_{0j}}{\sigma_{p_{0j}}} \right)^2 \right] \quad (4.42)$$

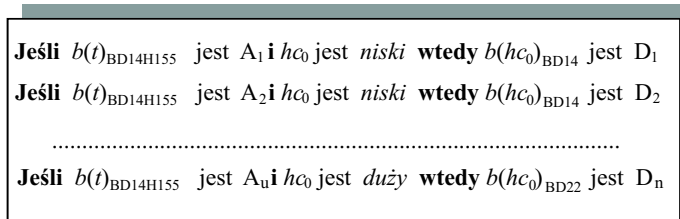
dla $a \neq j$.

gdzie: w i v oznaczają liczby zbiorów rozmytych stosowanych do rozmywania odpowiednio sygnałów hc_0 i p_0 .

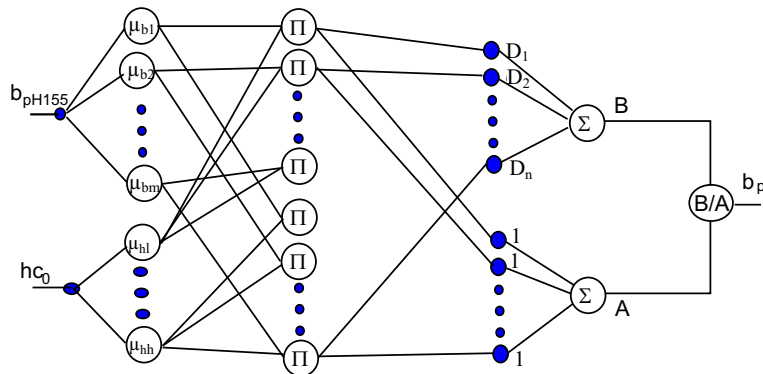
Podobne zależności uzyskano dla drugiego sygnału pomocniczego hc_0 .

4.3.6.3 Model neuro-rozmyty o strukturze kaskadowej

Rozmyty środkowy model lokalny struktury kaskadowej o bazie wiedzy przedstawionej na rys. 4.24 może być reprezentowany jako sieć neuronowa pokazana na rys. 4.25.



Rys. 4.24. Baza wiedzy środkowego lokalnego modelu struktury kaskadowej



Rys. 4.25. Neuronowa reprezentacja środkowego bloku struktury kaskadowej

W przedstawionej na rys. 4.25 strukturze bloki sumowania i dzielenia realizują zadanie wyostrzenia odpowiedzi modelu przy zastosowaniu metody COA.

W procesie strojenia parametrów środkowego bloku przyjęto singleton jako krzywą opisującą wyjście (4.21). Wartość parametru określającego rozmieszczenie krzywych wyznaczono w oparciu o następujące zależności:

$$\bar{y}_p^k(t+1) = \bar{y}_p^k(t) - \eta(b_p - b_{p,ref}) \frac{\mu_{k,r}(b_{p,H155})\mu_{k,l}(hc_0)}{\sum_{r=1}^u \sum_{l=1}^w \mu_r(b_{p,H155})\mu_l(hc_0)}; \quad \eta \in [0,1], \quad (4.43)$$

gdzie: p oznacza indeks definiujący początkowe ciśnienie (14 lub 22 bar), r - numer rozmytego zbioru dla rozmywania $b_{p_0_hc_0}$, u - liczbę zbiorów rozmytych użytych do rozmywania wejścia $b_{p_0_hc_0}$, l - numer rozmytego zbioru dla rozmywania sygnału hc_0 , w - liczbę zbiorów rozmytych użytych do rozmywania wejścia hc_0 oraz k, r są numerami zbioru rozmytego użytego przy rozmywaniu wejścia $b_{p_0_hc_0}$, który pojawia się w k -tej regule.

Wejścia $b_{p_0_hc_0}$ opisano za pomocą trójkątnych funkcji przynależności (rys.4.21a). Algorytm strojenia parametrów zastosowanych funkcji opisano za pomocą zależności:

$$s_{b_p}^r(t+1) = s_{b_p}^r(t) - fmkap \frac{\partial \mu_r(b_{p,H155})}{\partial s_{b_p}^r} \Big| (t), \quad (4.44)$$

$$q_{b_p}^r(t+1) = q_{b_p}^r(t) - fmkap \frac{\partial \mu_r(b_{p,H155})}{\partial q_{b_p}^r} \Big| (t), \quad (4.45)$$

gdzie: funkcja $fmkap$ definiuje współczynnik modyfikacji parametrów modelu kaskadowego (dla p_0) i opisano ją jako:

$$fmkap = \eta(b - b_{ref})\mu_p(p_0) \frac{\sum_{l=1}^w (y_p^{k,l} \mu_l(hc_0)) \sum_{k=1}^n (\mu_{k,r}(b_{p,H155})\mu_{k,l}(hc_0))}{\left(\sum_{k=1}^n (\mu_{k,r}(b_{p,H155})\mu_{k,l}(hc_0)) \right)^2} - \eta(b - b_{ref})\mu_p(p_0) \frac{\sum_{l=1}^w \mu_l(hc_0)}{\left(\sum_{k=1}^n (\mu_{k,r}(b_{p,H155})\mu_{k,l}(hc_0)) \right)^2} \quad (4.46)$$

oraz podobnie dla wejścia początkowy poziom hc_0 :

$$s_{h_p}^l(t+1) = s_{h_p}^l(t) - fmkah \frac{\partial \mu_l(hc_0)}{\partial s_{h_p}^l} \Big| (t), \quad (4.47)$$

$$q_{h_p}^l(t+1) = q_{h_p}^l(t) - fmkah \frac{\partial \mu_l(hc_0)}{\partial q_{h_p}^l} \Big| (t), \quad (4.48)$$

gdzie: funkcja $fmkah$ definiuje współczynnik modyfikacji parametrów modelu kaskadowego (dla hc_0) i opisano ją jako:

$$fmkah = \eta(b - b_{ref})\mu_p(p_0) \frac{\sum_{r=1}^u (y_p^{k,r} \mu_r(b_{p,H155})) \sum_{k=1}^n (\mu_{k,r}(b_{p,H155}) \mu_{k,l}(hc_0))}{\left(\sum_{k=1}^n (\mu_{k,r}(b_{p,H155}) \mu_{k,l}(hc_0)) \right)^2} - \eta(b - b_{ref})\mu_p(p_0) \frac{\sum_{r=1}^u \mu_r(b_{p,H155})}{\left(\sum_{k=1}^n (\mu_{k,r}(b_{p,H155}) \mu_{k,l}(hc_0)) \right)^2} \quad (4.49)$$

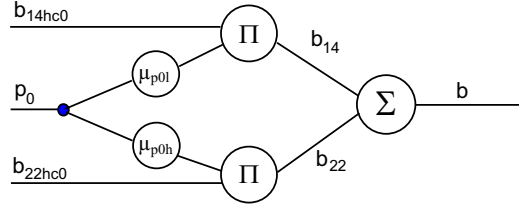
oraz pochodne cząstkowe $\frac{\partial \mu_r(b_{p,H155})}{\partial s_{b_p}^r}$, $\frac{\partial \mu_l(hc_0)}{\partial s_{h_p}^l}$, $\frac{\partial \mu_r(b_{p,H155})}{\partial q_{b_p}^r}$ i $\frac{\partial \mu_l(hc_0)}{\partial q_{h_p}^l}$ opisano zależnościami (4.33) i (4.34).

Sygnał wyjścia rozmytego modelu o strukturze kaskadowej opisano zależnością:

$$b = \mu_{p_{0l}} b_{14hc_0} + \mu_{p_{0h}} b_{22hc_0}, \quad (4.50)$$

gdzie: $\mu_{p_{0l}}$ i $\mu_{p_{0h}}$ oznaczają wartości funkcji przynależności do zbiorów *niski* i *wysoki* stosowanych przy rozmywaniu początkowego ciśnienia p_0 .

Finalizując strojenie struktury kaskadowej zaproponowano, aby ostatni blok rozpatrywanej struktury przedstawić jako prostą sieć neuronową (rys.4.26).



Rys. 4.26. Neuronowa reprezentacja ostatniego bloku kaskadowej struktury

Sygnał p_0 rozmywano za pomocą funkcji przynależności typu Gaussa (rys.4.21c). Zastosowany algorytm strojenia opisano zależnościami:

$$\bar{p}_{0p}(t+1) = \bar{p}_{0p}(t) - fmkaw \frac{\bar{p}_{0p} - p_0}{(\sigma_{p0p})^2} \Big| (t) \quad \text{dla } p \in \{14, 22\}, \quad (4.51)$$

$$\sigma_{0p}(t+1) = \sigma_{0p}(t) - fmkaw \frac{\bar{p}_{0p} - p_0}{(\sigma_{p0p})^3} \Big| (t) \quad \text{dla } p \in \{14, 22\}, \quad (4.52)$$

gdzie: funkcja $fmkaw$ definiuje współczynnik modyfikacji parametrów modelu kaskadowego (dla wyjścia) i opisano ją jako:

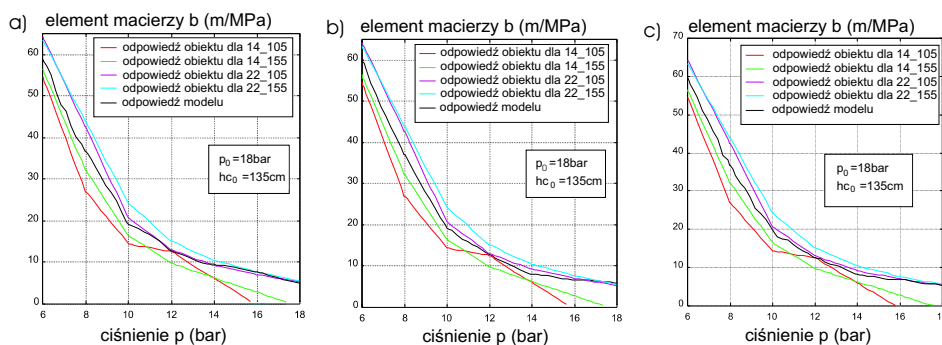
$$fmkaw = \eta(b - b_{ref}) \exp \left[- \left(\frac{p_0 - \bar{p}_{0p}}{\sigma_{p0p}} \right)^2 \right] \frac{\sum_{k=1}^n \bar{y}_p^k \mu_{k,r}(b_{phc_0}) \mu_{k,l}(hc_0)}{\sum_{k=1}^n \mu_{k,r}(b_{phc_0}) \mu_{k,l}(hc_0)}. \quad (4.53)$$

4.3.6.4 Porównanie struktur

Wszystkie rozmyte struktury zapewniają dobrą jakość aproksymacji i interpolacji w przypadku nieznanymi wartości wejścia. Najbardziej gładką krzywą aproksymacyjną uzyskano dla równoległej struktury (rys. 4.27a, b, c).

Analizę dokładności aproksymacji nieliniowej charakterystyki przeprowadzono dla badanych struktur rozmytych modeli w oparciu o średnią wartość sumy kwadratu błędów. Wyniki badań zamieszczono w tabeli 4.4. Można zauważyć, że uzyskano zbliżoną wartość błędów dla wszystkich badanych struktur, jednakże najgorszą dokładność aproksymacji uzyskano dla struktury kaskadowej. Struktura kaskadowa ma najbardziej skomplikowaną budowę, w szczególności z punktu widzenia drugiego bloku.

Proste bloki struktury równoległej mogą być łatwo projektowane, powielane dla referencyjnych punktów pracy i strojone za pomocą prostych metod optymalizacyjnych.



Rys. 4.27. Odpowiedzi rozmytego modelu dla struktury: (a) kompleksowej, (b) równoległej, (c) kaskadowej dla $p_0 = 18 \text{ bar}$ i $hc_0 = 135 \text{ cm}$

Dla zastosowań specjalnych zalecana jest struktura równoległa. Układ struktury równoległej może być łatwo optymalizowany i rozbudowywany przy dodatkowych wejściach.

Tabela. 4.4. Porównanie nakładów projektowych i otrzymanej dokładności dla rozpatrywanych struktur rozmytego modelu

Własności	Struktura rozmytego modelu		
	<i>kompleksowa</i>	<i>równoległa</i>	<i>kaskadowa</i>
Liczba zbiorów rozmytych użytych w procesie rozmywania wejść	19	10 dla każdego prostego modelu + 5	(9+2) lub (14+2) pierwszy blok, 35 - drugi blok
Liczba zbiorów rozmytych użytych w procesie opisu wyjścia	29	8 ÷ 12	30 - pierwszy blok, 40 - drugi blok
Liczba reguł	69	8 ÷ 12	9 lub 14 - pierwszy blok, 93 - drugi blok
Średni błąd aproksymacji $\bar{\epsilon}_{sq}$	2.21	2.42	3.04

Dodatkową korzyścią zastosowania struktury równoległej lub kaskadowej, w przeciwieństwie do struktury kompleksowej, jest możliwość analizy jakości aproksymacji w oparciu o dwuwymiarową charakterystykę modelu. Gładkość lub deformacja charakterystyki może być dobrym wskaźnikiem wstępnej oceny parametrów projektowanego rozmytego modelu.

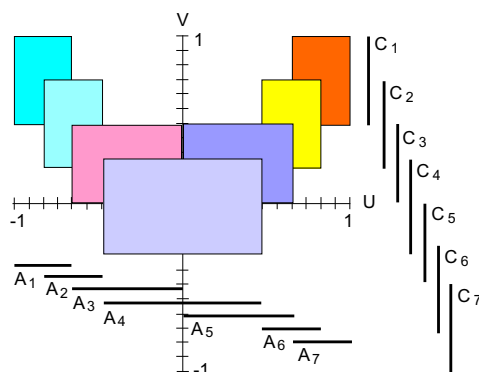
4.3.7. Algorytmy ewolucyjne w procesie optymalizacji parametrów systemów rozmytych

Modelowanie rozmyte zawiera dwa etapy. Pierwszy polega na skonfigurowaniu reguł rozmytych, drugi na optymalnym doborze kształtu funkcji przynależności. Oba problemy można rozważać w kategoriach odpowiednio optymalizacji kombinatorycznej i nieliniowej. Algorytmy ewolucyjne mogą być zastosowane do obu powyższych problemów. Problem konstrukcji bazy reguł rozmytych poruszano w wielu pracach, natomiast poszukiwanie odpowiednich funkcji przynależności najczęściej realizowane jest przy pomocy algorytmów najszybszego spadku [167].

Optymalizując bazę reguł w oparciu o algorytm genetyczny należy bazę przedstawić jako zbiór kompleksów [104]. Każdy kompleks jest koniunkcją selektorów, przy czym każdy selektor zawiera alternatywę dyskretnych wartości atrybutów. Wektory selektorów tworzą populację. Algorytm genetyczny określa bazę reguł typu prostego (reprezentacja wiedzy jako trójki) za pomocą zbioru atrybutów i ich wartości. Reguły o dowolnym stopniu złożoności można uzyskać stosując w procesie optymalizacji programowanie genetyczne [167].

Algorytmy genetyczne można stosować do optymalizacji parametrów zarówno modeli rozmytych typu Mamdaniego jak również TSK. Konstrukcję modelu rozmytego rozpoczyna się od dekompozycji przestrzeni wejściowej U i wyjściowej V (Rys.4.28). Reguła *Jeśli X jest A_i wtedy Z jest C_j* opisuje związek między zbiorem rozmytym A_i w dekompozycji obszaru wejściowego z zbiorem C_j z dekompozycji

obszaru wyjściowego. Graficznie podstawa A_i i C_j opisuje interwały w obszarze wejściowo-wyjściowym (Rys.4.28).



Rys. 4.28. Relacje definiowane za pomocą zbiorów $A_i \times C_j$

Algorytmy uczenia Wanga i Mendela dostarczają technik dla wyboru odpowiednich centralnych punktów z dekompozycji zakresu wyjściowego dla każdego wprowadzonego poprzednio wektora wejściowego. Wyboru wektora dokonuje się bazując na zgodności funkcji f generowanej przez bazę reguł w połączeniu ze zbiorem danych uczących. Podejście ewolucyjne pozwala na wybór dowolnych wartości z całego zakresu sygnału wyjściowego. Alternatywne rozwiązanie polega na wstępnym wyborze zbioru wartości centralnych wyjścia i strojenie punktów centralnych wejścia w celu poprawy dopasowania bazy reguł. W tym przypadku przeszukuje się przestrzeń zawierającą wektory wejściowe. Przesuwanie lokalizacji centralnych punktów wejścia postrzegane jest jako efekt dokładnego strojenia bazy reguł.

Omówione wcześniej algorytmy realizują strojenie jednego zbioru punktów, podczas gdy drugi jest wcześniej ustalany jako punkt startowy. Podejście ewolucyjne pozwala rozszerzyć powyższe algorytmy poprzez zwiększenie obszaru przeszukiwań na zakres parametrów wejściowych i wyjściowych. Proces przeszukiwań obejmuje zadanie równoczesnego określenia rozkładu obu zbiorów parametrów. W tym przypadku przestrzeń przeszukiwań jest rodziną $2 \times n$ macierzy z wejściami z zakresu $[-1, 1]$.

Każda z opisanych powyżej operacji może być rozpatrywana w sensie umieszczenia omawianych wektorów w zakresie parametrów strojenia, który jest jednym z filarów poszukiwań opartych na strategii ewolucyjnej. Nowa baza reguł jest wyszukiwana za pomocą małych losowych odchyień punktów centralnych w aktualnej bazie reguł. Zastosowane w tej metodzie odchylenia reprezentują mutacje bazy reguł.

W wcześniej opisanych algorytmach strojenia bazy reguł wielkości zastosowanych wektorów są zdeterminowane przez liczbę zastosowanych w dekompozycji wejścia zbiorów rozmytych. Problem ten stanowi ważny punkt decyzyjny w definicji parametrów budowanego rozmytego systemu. Pionierską pracę w tej dziedzinie przedstawił Fogel [54] opisując proces uczenia przy użyciu programowania ewolucyjnego, w którym wzrost lub obniżenie liczby zastosowanych w dekompozycji wejścia zbiorów rozmytych wykorzystał jako kolejny parametr generowanego procesu uczenia. Z pozycji reprezentacji wiedzy jest to postrzegane jako wzrost lub obniżenie wielkości stosowanych macierzy.

W procesie poszukiwania optymalnej reprezentacji wiedzy zawartej w bazie danych można zastosować również operacje ewolucyjne oparte na standardowym jednopunktowym krzyżowaniu. Wprowadzane w tym algorytmie modyfikacje wiedzy polegają na przetwarzaniu macierzy reprezentujących bazę reguł. Krzyżowanie między bazą $R1$ i $R2$ uzyskuje się przez losowy wybór numeru kolumny k i zamianie wartości w kolumnach $k + 1, \dots, n$ bazy $R1$ na wartości w odpowiednich kolumnach bazy $R2$.

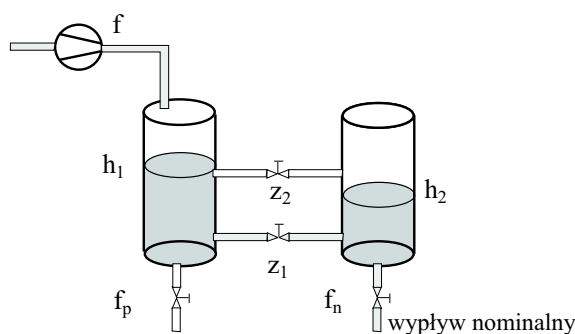
Mutacja jest operacją preferowaną w problemach strojenia parametrów. Krosowanie jest mechanizmem używanym w zadaniach syntezy problemu. Kombinacja ewolucyjnego strojenia i krosowania pomaga w identyfikacji i propagacji fragmentów bazy reguł, które mają wysoki stopień kompatybilności z danymi uczącymi.

4.3.8. Systemy rozmyte w klasyfikacji uszkodzeń

Ważnym elementem systemu diagnostycznego jest blok decyzyjny, który na bazie wektora residuów lokalizuje a niekiedy również identyfikuje poziom uszkodzenia. Zadanie to można także rozwiązać w oparciu o logikę rozmytą. Rozmyta ocena wartości residuów umożliwia uwzględnienie niepewności wartości sygnałów diagnostycznych wynikającej z szumów pomiarowych, zakłóceń w procesie, błędów modelowania oraz niedokładności i niepewności w określaniu wartości progowych [85]. Zastosowanie podejścia rozmytego pozwala uzyskać dobre wyniki diagnostyczne zwłaszcza w diagnostyce realizowanej w trybie *on-line*, w której występują podwyższone wymagania szybkości uzyskiwanych diagnoz. Rozmyte podejście cechują: prostota zaproponowanych algorytmów, niewielkie moce obliczeniowe przy realizacji algorytmów przetwarzania danych. Ponadto podejście to pozwala na wprowadzenie stopni pewności wypracowanej diagnozy [202, 243].

Proces detekcji i lokalizacji uszkodzenia realizowany jest w określonym przedziale czasowym. Wyniki obserwacji stanu obiektu można przedstawić jako ciąg wektorów residuów określonych dla kolejnych chwil czasowych. Lokalizacja uszkodzenia w układzie automatycznego sterowania wymaga wyizolowania, w obserwowanych sygnałach, wpływu uszkodzenia od reakcji kompensujących wpływ uszko-

dzenia realizowanych przez system automatycznego sterowania.



Rys. 4.29. Schemat układu dwóch zbiorników

Budując hybrydowy system diagnostyczny zaproponowano [192] do celów lokalizacji i identyfikacji uszkodzenia zastosować wielowjęściowy rozmyty klasyfikator. Analizę własności takiego klasyfikatora przeprowadzono na przykładzie zadania diagnozowania zespołu dwóch zbiorników (Rys. 4.29). W obiekcie obserwowano następujące wielkości: poziomy w zbiornikach h_1 , h_2 oraz strumień medium f . Intencją przeprowadzonych przez autora badań była ocena skuteczności klasyfikacji uszkodzeń przy zastosowaniu wielowymiarowych zbiorów rozmytych. W związku z tym do badań zastosowano przykładowy zbiór następujących uszkodzeń: przeciek w pierwszym zbiorniku f_p , uszkodzenie zaworu sterującego przepływem medium między zbiornikami z_1 (blokada zamknięcia) oraz przytkanie kanału łączącego zbiorniki z_2 .

Zastosowany w układzie diagnostycznym generator residuów zbudowano w oparciu o grupę rozmytych modeli lokalnych [131]. W badanym przykładzie zastosowano modele dla zbiorników oraz pompy.

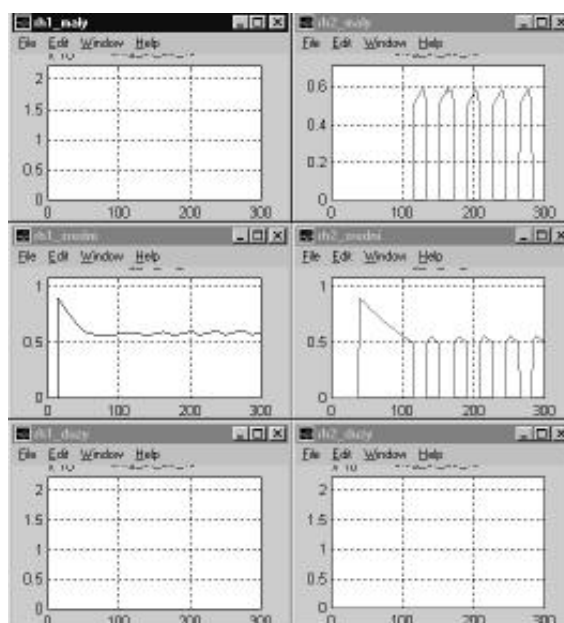
Na wyjściu układu uzyskano wektor residuów opisany wzorem:

$$\mathbf{r} = (r_{h_1}, r_{h_2}, r_p), \quad (4.54)$$

gdzie: r_{h_1}, r_{h_2}, r_p oznaczają residua poziomu w zbiornikach i odpowiedzi pompy.

Kierując się zasadą czytelności przedstawionych w dalszej części ilustracji, przyjęto dwuwymiarowy wektor residuów, uwzględniający analizę pomiaru poziomu medium w obu zbiornikach.

Elementy wektora residuów opisano w formie rozmytej. Liczbę zbiorów rozmytych oraz kształt krzywych rozmywających stosowanych do opisu wybranego



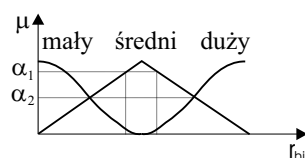
Rys. 4.30. Przebiegi rozmytej reprezentacji r_{h1} , r_{h2} dla średniego przecieku zbiornika pierwszego

elementu wektora \mathbf{r} można uzależnić od charakteru opisywanego sygnału obiektowego [202]. W badaniach wstępnych zaproponowano użycie identycznych parametrów procesu rozmywania dla całego wektora residuów (rys. 4.31). Przebiegi elementów r_{h1} , r_{h2} wektora residuów dla średniego przecieku zbiornika pierwszego w rozmytej reprezentacji przedstawiono na rys. 4.30. W lewym górnym oraz w obu dolnych oknach otrzymano, przy omawianym uszkodzeniu, zerowe wartości funkcji przynależności do oznaczonych zbiorów rozmytych w całym okresie obserwacji stanu obiektu.

Algorytm działania klasyfikatora bazuje na procesie grupowania wektorów residuów. W przypadku rozmytej reprezentacji wektora residuów powstaje problem definicji miary. W przedstawionym rozwiązaniu zaproponowano proces klasyfikacji realizować na bazie α - przynależności do wybranego zbioru rozmytego (rys. 4.31).

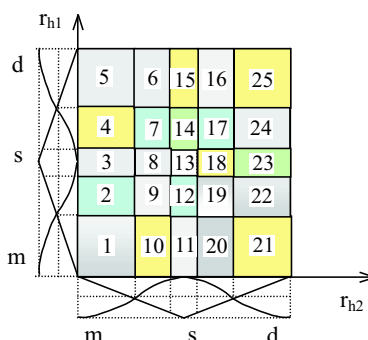
Zastosowany proces rozmywania wektora residuów r_{hi} dla ($i = 1, 2$) cechują następujące parametry:

- trzy zbiory rozmyte,
- trójkątne i sigmoidalne krzywe rozmywające,

Rys. 4.31. Rozmywanie wektora residuów z α -obięciem

- dwa progi α_1 i α_2 obcięcia przynależności wektora residuum do zdefiniowanych zbiorów rozmytych. Próg α_1 określa poziom błędów pomiarowych, α_2 definiuje przynależność wektora residuów do zbioru rozmytego o większej wartości funkcji przynależności.

Dla obu składowych wektora r_h zastosowano ten sam proces rozmywania i otrzymano dwuwymiarową rozmytą reprezentację stanu diagnozowanego obiektu (rys.4.32).



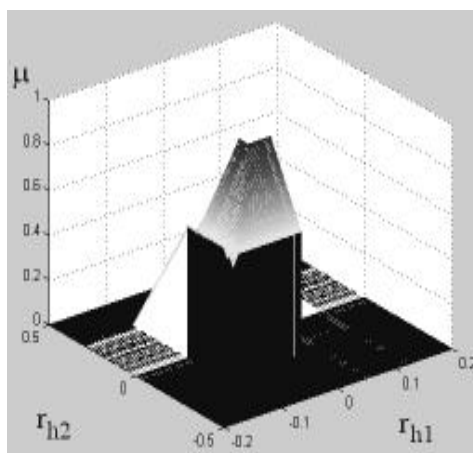
Rys. 4.32. Ilustracja rozmytego grupowania wektora residuów

W przedstawionej na rys. 4.32 graficznej reprezentacji stanu diagnozowanego obiektu wyróżniono określoną liczbę charakterystycznych pól definiowanych za pomocą zbiorów rozmytych wektora residuów. Wartość funkcji przynależności rozmytego wektora residuów do dwuwymiarowych zbiorów rozmytych, przedstawionych na rysunku, wyznaczono na bazie operatora Mamdaniego *t-normy* opisanej zależnością:

$$\mu_A(r_{h1}, r_{h2}) = \min(\mu_A(r_{h1}), \mu_A(r_{h2})), \quad (4.55)$$

gdzie: A oznacza wybrany zbiór rozmyty opisujący wektor residuów.

Wyniki analiz stanu obiektu z wykorzystaniem funkcji przynależności do zdefiniowanych dwuwymiarowych zbiorów rozmytych zawierają informację o rozmytej klasteryzacji wektorów residuów dla danego stanu obiektu. Na rys. 4.33 przedstawiono przebiegi krzywych przynależności rozmytych reprezentacji residuów r_{h1} , r_{h2} dla średniego przecieku w zbiorniku pierwszym.



Rys. 4.33. Dwuwymiarowa rozmyta reprezentacja r_{h1} , r_{h2} dla średniego przecieku zbiornika pierwszego

Proces kojarzenia rozmytej reprezentacji klastrów wektorów residuów ze stanem diagnozowanego obiektu zrealizowano w oparciu o współczynnik względnej mocy dwuwymiarowego zbioru rozmytego określony za pomocą zależności [202]:

$$\|A\|_{r_{h1}, r_{h2}} = \frac{\sum_{r \in \{r_{h1}, r_{h2}\}} \mu_A(r)}{N}, \quad (4.56)$$

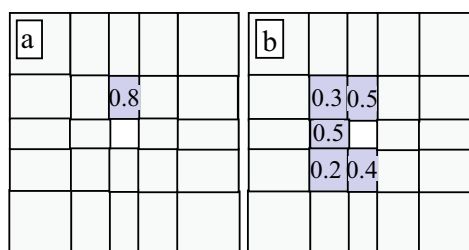
gdzie: N oznacza liczbę elementów w przestrzeni rozważań zbioru rozmytego.

a				
	0.5			
	0.2	0.4		

b			0.2	
		0.6		
		0.5		
		0.2	0.6	

Rys. 4.34. Mapa wartości mocy dwuwymiarowych zbiorów rozmytych dla przecieku: a) mały, b) średni

Wyznaczone w takiej formie współczynniki dla każdego dwuwymiarowego zbioru rozmytego (rys.4.32) stanowią mapę stanu obiektu. Na rys. 4.34 i 4.35 przedstawiono mapy wartości mocy dwuwymiarowych zbiorów rozmytych dla różnych uszkodzeń (pojedynczych oraz dwóch występujących jednocześnie).



Rys. 4.35. Mapa wartości mocy dwuwymiarowych zbiorów rozmytych dla: a) blokady zaworu z_1 , b) średni przeciek i blokada

Stosowanie w procesie diagnozowania rozmytej reprezentacji wiedzy pozwala na realizację zadania detekcji zarówno w przypadku uszkodzeń pojedynczych jak i większej liczby występujących jednocześnie. Zadanie lokalizacji może być realizowane z dobrym skutkiem dla wybranych uszkodzeń (rys. 4.35a), jednakże w pozostałych przypadkach należy dokonać strojenia parametrów klasyfikatora, dążąc do innego ukształtowania obszarów rozmytej reprezentacji residuów.

4.3.9. Zastosowania zbiorów rozmytych w wydobywaniu wiedzy i diagnostyce

Zastosowanie logiki rozmytej w ostatnim okresie obejmuje coraz szersze obszary. W omawianych zagadnieniach zauważa się również szybki rozwój technik wydobywania wiedzy i diagnozowania procesów realizowanych przy wykorzystaniu logiki rozmytej.

W zakresie wydobywania wiedzy, w oparciu o logikę rozmytą, buduje się metody klasyfikacji i grupowania cech analizowanych danych. Do znanych należy zaliczyć metodę FCM (ang. *Fuzzy C-Means*) realizującą zadanie klasyfikacji cech poprzez zdefiniowanie stopnia przynależności każdego punktu do analizowanych klastrów (rozdział 4.3.8). Zadanie grupowania danych można zrealizować za pomocą rozmytych sieci Kohonena, FKN (ang. *Fuzzy Kohonen Nets*).

W systemach diagnostycznych rozmytą reprezentację wiedzy można stosować w realizacji różnych zadań postawionych przed systemem diagnostycznym. Zaliczono do nich:

1. rozmytą reprezentację wartości zmiennych procesowych (rozdział 6.2.4.1). Pozwala to prowadzić proces diagnozowania z wykorzystaniem doradczych systemów ekspertowych [27, 186];
2. rozmytą reprezentację residuów (rozdział 4.3.4.1). Unormowanej wartości residuum przyporządkowuje się etykiety lingwistyczne w postaci zbiorów

rozmytych o zadanych kształtach (najczęściej trapezoidalne). Trapezoidalne funkcje przynależności pozwalają wydzielić strefy odpowiadające pracy prawidłowej i awarii, a także obszar, w którym decyzja o uszkodzeniu jest niepewna [123];

3. rozmyta reprezentacja modelu diagnostycznego. Wektory residuów reprezentowane w formie rozmytej pozwalają budować macierze diagnostyczne z rozmytą reprezentacją powiązań symptom-uszkodzenie. Daje to bardziej elastyczny system diagnostyczny w porównaniu z systemem opartym o binarne macierze diagnostyczne [126];
4. modele heurystyczne obiektów budowane w rozmytej reprezentacji wiedzy (rozdział 4.3.5). Zaletą modeli rozmytych i rozmytych sieci neuronowych jest możliwość połączenia wiedzy eksperta oraz danych pomiarowych.

4.4. Systemy ekspertowe

Systemy ekspertowe mogą pełnić funkcje doradcze lub decyzyjne, zastępując operatora przy wykonywaniu niektórych czynności w warunkach trudnych lub niebezpiecznych [13, 137]. Reprezentują one wiedzę zagregowaną od ekspertów posiadających wiedzę niezbędną do przeprowadzania analiz i podejmowania decyzji. Niekiedy system ekspertowy z dobrze zagregowaną wiedzą pracuje wydajniej od operatora [34].

System ekspertowy jest programem komputerowym, który wykonuje złożone zadania o dużych wymaganiach intelektualnych i robi to niekiedy w sposób zbliżony do działania człowieka będącego ekspertem w tej dziedzinie. Określenie "system ekspertowy" może być zastosowane do dowolnego programu komputerowego, który na podstawie szczegółowej wiedzy może wyciągać wnioski i podejmować decyzje, działając w sposób zbliżony do procesu rozumowania człowieka. W wielu sytuacjach (np. podczas podejmowania decyzji w siłowniach energetycznych) człowiek nie może swymi zmysłami ogarnąć całej sytuacji, wtedy system ekspertowy pracujący w czasie rzeczywistym wykonuje swoje funkcje lepiej niż człowiek.

Systemy ekspertowe stosuje się w wielu przypadkach w takich dziedzinach, gdzie informacja (wiedza) o danej dziedzinie jest niepewna, nie jest w sposób jednoznaczny sformalizowana (nie istnieją matematyczne algorytmy rozwiązujące postawione zadania) lub postawiony problem można zaliczyć do NP-zupełnych.

System ekspertowy, dysponując zapisaną wiedzą eksperta z wybranej dziedziny, może jej używać wielokrotnie w sposób efektywny bez obecności eksperta. Ponadto można w takim systemie zagregować wiedzę licznego zespołu ekspertów.

Ogólna struktura systemu ekspertowego oparta jest na dwóch podstawowych blokach: wnioskującego i bazy wiedzy [158]. Dodatkowym elementem systemu jest interfejs użytkownika, który pozwala komunikować się z operatorem w celu uzupełnienia danych o obiekcie oraz przekazywać konkluzje lub objaśnienia systemu do operatora. Już pod koniec lat siedemdziesiątych stwierdzono, że moc programu ekspertowego (w zakresie rozwiązywania danego problemu) tkwi w zakodowanej w nim wiedzy. Tak więc efektywność pracy systemu zależy przede wszystkim od zawartej w systemie bazy wiedzy, a nie w sposobie realizacji procesu wnioskowania systemu ekspertowego. Zatem, aby zbudować dobry system ekspertowy, należy go wyposażać w pełną i pewną, specyficzną wiedzę o danym procesie.

Pomimo istnienia wielu znanych systemów ekspertowych metodologia ich budowy wymaga rozwiązywania wielu nowych problemów. Jako najbardziej istotne należy wymienić [34]:

- ◇ problem adekwatnej reprezentacji wiedzy,
- ◇ wnioskowanie w warunkach niepewności,
- ◇ problem z gromadzeniem wiedzy.

Za najtrudniejszy problem przy tworzeniu systemu powszechnie uważa się ekstrakowanie wiedzy od ekspertów. Niekiedy uzyskana wiedza jest niepełna i niepewna, co powoduje często błędne działanie systemu. Rozwiązanie tego problemu związane jest z samodzielną weryfikacją poprawności działania i uściślenia bazy wiedzy. Przyszłość systemów ekspertowych leży w procesie komputerowego uczenia się, w oparciu o redundancję programową uzyskiwaną, w oparciu o wybrane elementy sztucznej inteligencji (sieci neuronowe, zbiory rozmyte).

4.4.1. Typy systemów ekspertowych

Do najbardziej rozpowszechnionych zaliczamy systemy, w których bazę wiedzy wydzielono od pozostałych bloków. Nazywane są one systemami opartymi na bazie wiedzy i często wykorzystywane są jako systemy diagnostyczne lub doradcze [162].

Prawidłowo przygotowany system ekspertowy powinien charakteryzować się następującymi cechami [142]:

- ◇ poprawność - system powinien zapewnić wysoki poziom wydawanych ekspertyz w dopuszczalnym czasie;
- ◇ uniwersalność - system powinien wykazywać zdolność do rozwiązywania obszernej klasy zadań z danej dziedziny. Wymóg uniwersalności prowadzi do budowy metasytemów (systemy szkieletowe);

- ◊ niesprzeczność - systemy zawierające bazę wiedzy w reprezentacji dyskretnej mają wbudowany mechanizm wykrywania reguł sprzecznych tzn. reguł o różnych konkluzjach dla takich samych przesłanek;
- ◊ kompletność - baza wiedzy danego systemu powinna zawierać taką liczbę reguł, aby uniemożliwić przypadki, w których występują: nie określone wartości przesłanek, niedopuszczalne wartości atrybutów, nieosiągalne wartości przesłanek, nieosiągalne akcje (uaktywnienie reguły) i nieosiągalny cel.

Dodatkową własnością, poprawiającą jakość systemu ekspertowego, jest zdolność systemu do udoskonalania jego działania, ze szczególnym uwzględnieniem ciągłego rozszerzania wiedzy o nowe fakty i prawa (reguły wnioskowania).

Analiza przebiegu złożonych procesów wymaga niekiedy obserwacji dużej liczby wielkości procesowych. Operatorzy obsługujący dynamiczne wielowymiarowe obiekty są nieraz przeciążeni napływającymi informacjami i tracą nad nimi panowanie. W większości obserwowanych obiektów występuje potrzeba przetworzenia części otrzymywanych sygnałów w celu dokonania analizy stanu diagnozowanego obiektu. Dodatkowym wymaganiami procesu przetwarzania danych jest czas realizacji zadania [33]. Dla prawidłowej analizy stanu obiektu dynamika procesu przetwarzania musi być większa niż dynamika badanego obiektu. Systemy ekspertowe czasu rzeczywistego zapewniają efektywne monitorowanie i agregację zmiennych procesowych pracując w trybie *on-line*. Można je zastosować w monitorowaniu i sterowaniu różnych procesów dla dziedzin takich, jak: technika kosmiczna, militaria, elektronika, ruch lotniczy, ochrona środowiska, medycyna, ekonometria i finanse. Aby zapewnić realizację niekiedy złożonych zadań stosuje się systemy o rozproszonej architekturze wykorzystującej właściwości pracy asynchronicznej oraz niekiedy wieloprocesorowe systemy komputerowe z systemem operacyjnym czasu rzeczywistego (np. QNX).

Systemy ekspertowe można budować stosując różne formy reprezentacji wiedzy. Każda z wcześniej (rozdział 3) wymienionych form wykazuje swoje specyficzne własności i niestety nie jest pozbawiona wad. Wybór formy reprezentacji wiedzy zależy w dużej mierze od dziedziny wiedzy oraz od dostępnych narzędzi do budowy systemu ekspertowego. W pierwszych systemach ekspertowych używano wyłącznie bazy regułowej (np. PROSPECTOR). Baza wiedzy zawiera w tym przypadku zbiór reguł i zbiór faktów. Podejście to umożliwia uzyskanie dużej modularności bazy wiedzy. Praktycznie działające regułowe systemy ekspertowe mogą zawierać reguły charakteryzowane stopniami pewności. Określa to stopień pewności konkluzji występującej w danej regule. Niektóre systemy dopuszczają rozwiniętą formę reguł, które zawierają dodatkowe stwierdzenie uznawane za prawdziwe w razie niespełnienia przesłanki. Do znanych systemów ekspertowych należy zaliczyć [163]:

- *Denokal* - jeden z pierwszych systemów stosowanych w przemyśle chemicznym;

- *PDS* - system opracowany przez naukowców Uniwersytetu Carnegie-Mellon i Westinghouse Electric. Interpretuje nieprawidłowości w procesie wytwarzania;
- *Reactor* - system opracowany przez firmę EG&G Idaho i zastosowany do wspomaganie operatora reaktora jądrowego;
- *DT200* - system diagnostyki turbozespołów.

4.4.2. Metody wnioskowania

Blok wnioskowania realizuje podstawowe funkcje generowania decyzji. Proces generowania decyzji bazuje na zaimplementowanej bazie wiedzy. Struktura funkcjonalna systemu wnioskującego zawiera trzy bloki. Interpreter reguł określa wartość logiczną (prawda, nieprawda, współczynnik pewności) wniosków reguł [163]. Współczynnik pewności określa stopień pewności konkluzji w aktywnej regule. Opracowanie interpretera reguł stanowi zasadnicze zadanie przy tworzeniu SE. System sterujący wyznacza kolejność testowania reguł bazy wiedzy. Jego funkcjonowanie zależy od metody wnioskowania (w przód lub wstecz). System wyjaśniający uzasadnia użytkownikowi przebieg wnioskowania i generuje raporty wnioskowania.

Wyróżniamy trzy podstawowe metody wnioskowania: w przód (progresywne), wstecz (regresywne) i mieszane. Dodatkowo należy rozpatrywać wnioskowanie rozmyte, ostatnio szybko rozwijane w różnych systemach ekspertowych [158, 202].

W zaproponowanym hybrydowym SE proces wnioskowania zrealizowano w oparciu o metodę wnioskowania wstecz. Wynika to ze specyfiki przewidzianego zadania realizowanego przez system. System weryfikuje hipotezę - obiekt pracuje prawidłowo i dokonując analizy prawdziwości wymaganych przesłanek tworzy nowe hipotezy. Aby ocenić poprawność wyboru metody, przedstawiono jej zasadę działania oraz przegląd innych metod wnioskowania. Opisu zasady działania dokonano w oparciu o fragment opracowanej i zaimplementowanej w hybrydowym SE bazy wiedzy:

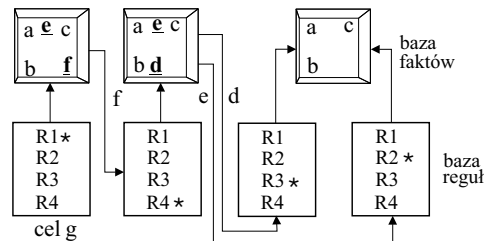
- R1: Jeśli f i e wtedy g
- R2: Jeśli a i c wtedy e
- R3: Jeśli e i b wtedy d
- R4: Jeśli d i c wtedy f

gdzie dla obiektu cieplnego przyjęto następujące oznaczenia:

- a - niskie ciśnienie pary,
- b - niska temperatura pary,
- c - przepływ paliwa w normie,
- d - uszkodzone palenisko,

e - uszkodzony zawór powietrza,
 f - zawór paliwa sprawny,
 g - zawartość tlenu w spalinach poza normą.

Wnioskowanie wstecz przebiega od celu do faktów. Opiera się ono na wykazaniu prawdziwości hipotezy głównej na podstawie prawdziwości przesłanek. W przypadku nieokreślonej wartości logicznej badanej przesłanki traktuje się ją jako nową hipotezę i przeprowadza się próbę wykazania jej prawdziwości. Proces zostaje zakończony w przypadku znalezienia hipotezy, dla której wykazano prawdziwość wszystkich przesłanek. Proces dowodzenia konkluzji przeprowadza się wstecz, do momentu dotarcia do reguły zawierającej konkluzję główną (rys. 4.36). W typowych zastosowaniach wnioskowanie to jest efektywniejsze i bardziej rozpowszechnione.



Rys. 4.36. Proces wnioskowania wstecz

Wnioskowanie wstecz z nawrotem obejmuje sytuacje, w których wybrana reguła nie może być spełniona. Dla analizy faktów należy wtedy dokonać nawrotu do poprzedniej reguły. Proces wnioskowania powinien być kontynuowany przy wykorzystaniu innych reguł [188].

Wnioskowanie w przód realizuje prostą zasadę poszukiwania celu. Na podstawie dostępnych reguł i faktów należy generować nowe fakty do momentu wygenerowania faktu zgodnego z celem. Podstawową cechą zaliczaną do wad tego systemu jest ciągłe zwiększanie liczby faktów, a zatem powiększanie pojemności pamięci operacyjnej komputera, niezbędnej do dalszego funkcjonowania systemu.

Wnioskowanie mieszane obejmuje wcześniej omówione dwa sposoby wnioskowania. Dla części reguł stosuje się wnioskowanie w przód, a dla drugiej - wnioskowanie wstecz. Najczęściej proces wnioskowania rozpoczyna się od wnioskowania wstecz i w chwili znalezienia przesłanki, której nie można udowodnić, rozpoczyna się wnioskowanie w przód.

4.4.3. Zastosowanie systemów ekspertowych w diagnostyce

Proces diagnozowania obiektów przemysłowych obejmuje zagadnienia detekcji, lokalizacji i oceny poziomu uszkodzenia. Do realizacji powyższych zadań aktualnie stosuje się różne techniki oceny stanu obiektu. W przypadku prostych procesów często stosuje się metody oparte na modelu wejściowo-wyjściowym z progową oceną wybranych sygnałów obiektowych [188].

W złożonych procesach przemysłowych wiedza o stanie procesu bywa niedokładna i często uzupełniana wiedzą heurystyczną ekspertów. W tej sytuacji, aby uzyskać dobrą jakość diagnozowania, zaproponowano budowę diagnostycznego SE przy użyciu zbiorów rozmytych [57, 183]. Systemy ekspertowe z rozmytą reprezentacją wiedzy wymagają definicji kilku istotnych cech zastosowanych rozmytych zbiorów. Zaliczono do nich: sposób rozmywania danych pomiarowych [40] oraz sposób przetwarzania zmiennych lingwistycznych w procesie wnioskowania [202, 210, 243]. Proces strojenia systemu rozmytego wymaga niekiedy zdefiniowania kilku podstawowych parametrów *a priori* (na bazie wiedzy eksperta) [191].

4.4.3.1 Forma reprezentacji wiedzy a jakość diagnozowania systemu ekspertowego

Jakość diagnozowania stanu procesu przemysłowego należy rozumieć jako prawidłowość detekcji uszkodzeń, co może być mierzone liczbą fałszywych alarmów w normalnym stanie diagnozowanego obiektu lub liczbą błędnych lokalizacji zaistniałych uszkodzeń [165]. Do oceny jakości diagnoz uzyskiwanych przez badany SE wykorzystano wskaźniki wiarygodności algorytmu diagnozowania i czasu diagnozowania określone za pomocą następujących zależności:

$$P_{dn} > P_{dop} \quad (4.57)$$

gdzie: P_{dn} oznacza wartość prawdopodobieństwa prawidłowych decyzji rozpoznawania stanów obiektu, P_{dop} oznacza dopuszczalną wartość prawdopodobieństwa prawidłowych decyzji rozpoznawania stanów obiektu,

$$t_{dn} \leq t_{dop} \quad (4.58)$$

gdzie: t_{dn} oznacza czas diagnozowania obiektu, t_{dop} oznacza dopuszczalną wartość czasu diagnozowania obiektu.

Intencją autora w przeprowadzonych badaniach było spojrzenie na jakość diagnozowania stanu obiektu przemysłowego za pomocą systemu ekspertowego z rozmytą reprezentacją wiedzy. Jako diagnozowany obiekt przyjęto znany przykład dwóch zbiorników połączonych za pomocą dwóch zaworów (rys.4.29) [193]. Zawór oznaczony symbolem $z1$ pracuje jako element wykonawczy układu regulacji poziomu cieczy w zbiorniku 2, zawór drugi ($z2$) pełni funkcję zasuw awaryjnej. Układ

regulacji poziomu cieczy zbiornika 1 steruje wydajnością pompy zasilającej ten zbiornik. Ciecz gromadzona w zbiorniku 2 wypływa z niego ze stałym nominalnym przepływem (regulacja stałowartościowa). Do badań przygotowano 5 uszkodzeń podzespołów obiektu wprowadzanych pojedynczo lub grupowo. Uszkodzenia dotyczą blokady zaworów z1 lub z2 w pozycji "otwarty" lub "zamknięty" oraz wyciek w zbiorniku 1. Po wprowadzeniu do obiektu określonych uszkodzeń można wyróżnić 5 różnych od normalnego stanów badanego procesu.

W badanym systemie ekspertowym zastosowano regułą reprezentację wiedzy typu *Jeśli przesłanka, wtedy konkluzja (lub działanie)*. Reguły zostały stworzone na podstawie wiedzy eksperta. Przykładowe reguły stosowane w badanym SE przedstawiono na rys. 4.37 [193].

Dla SE z dyskretną reprezentacją wiedzy

Jeśli <wydajność pompy> ≥ 0.875

i <wydajność pompy> < 3.80

wtedy *Pompa pracuje normalnie*

Jeśli Pompa pracuje normalnie

i < poziom wody w zbiorniku 1 > < 0.5

i < poziom wody w zbiorniku 2 > >0.1

wtedy *Obiekt w stanie 2 - otwarty zawór z₂*

Dla SE z rozmytą reprezentacją wiedzy

Jeśli Pompa pracuje z dużym przepływem

i Poziom wody w zbiorniku 1 jest normalny

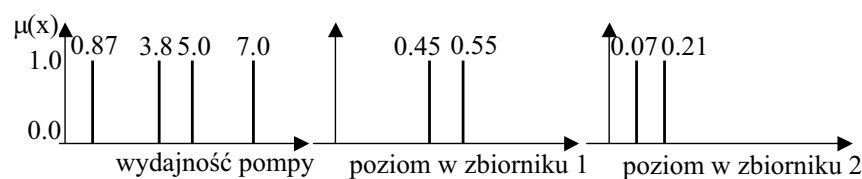
i Poziom wody w zbiorniku 2 jest normalny

wtedy *Obiekt w stanie 7 - wyciek ze zbiornika 1*

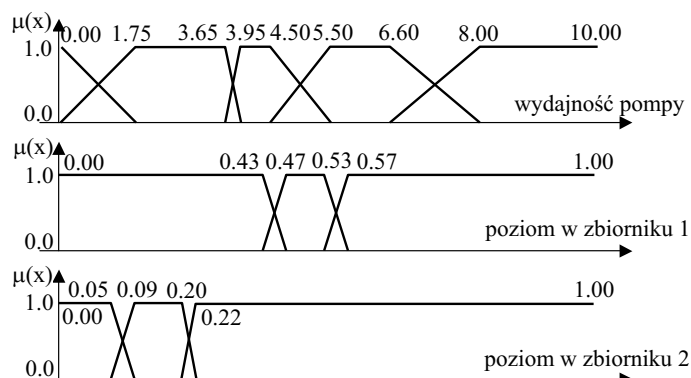
Rys. 4.37. Przykłady reguł z dyskretną i rozmytą reprezentacją wiedzy

W procesie badań zaproponowano zastosowanie kilku systemów ekspertowych z odpowiednio dyskretną lub rozmytą bazą wiedzy. W systemach zastosowano kilka funkcji przynależności o różnych parametrach progowych lub różnym rozmieszczeniu. Sposób podziału przedziału pomiarowego sygnałów obiektowych, dla dyskretnej reprezentacji wiedzy, przedstawiono na rys. 4.38 (SE A1). Dyskretne SE były reprezentowane przez trzy układy o różnej reprezentacji wiedzy (różne poziomy zastosowanych progów). Dla reprezentacji rozmytej sposób podziału zamieszczono na rys. 4.39 (dla SE B1). Badania przeprowadzono dla trzech dyskretnych SE A1,A2,A3 oraz pięciu rozmytych SE B1,...,B5 (każdy dla zmodyfikowanych funkcji przynależności). W rozmytym SE B2 funkcje przynależności do zbiorów rozmytych, opisujących sygnał wydajności pompy, zdefiniowano za pomocą krzywych

trapezowych, obejmujących węższy przedział sygnału wejściowego w porównaniu z B1. W SE B3, w porównaniu z reprezentacją wiedzy stosowaną w B1, zastosowano dodatkowy zbiór rozmyty dla dużych wartości wydajności pompy. W SE B4, w porównaniu z reprezentacją wiedzy stosowaną w B1, zastosowano dodatkowy zbiór rozmyty dla dużych wartości wydajności pompy i poziomu wody w zbiorniku 2. W SE B5, w porównaniu z B2, zastosowano dodatkowy zbiór rozmyty dla dużych wartości wydajności pompy.



Rys. 4.38. Rozmieszczenie progów w dyskretnej reprezentacji wiedzy SE



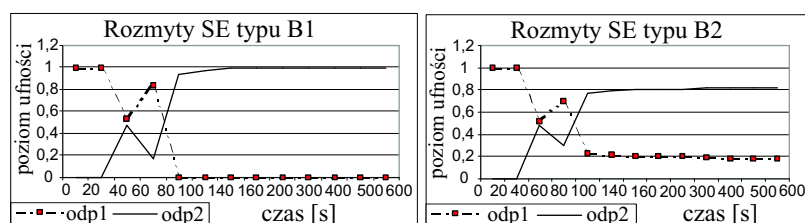
Rys. 4.39. Rozmieszczenie zbiorów rozmytych w reprezentacji wiedzy SE

W pierwszym etapie badań określono jakość diagnozowania obiektu w stanie normalnym (bez uszkodzeń). Wszystkie zaproponowane SE prawidłowo realizowały zadanie detekcji uszkodzeń, nie wystąpiły fałszywe alarmy. Kolejne etapy badań obejmowały analizę procesu diagnozowania w układzie z różnymi konfiguracjami SE po wprowadzeniu do obiektu różnych uszkodzeń. Przeprowadzono serię badań polegających na wielokrotnej realizacji zadania diagnostycznego dla obiektu z wprowadzonym określonym uszkodzeniem oraz zakłóceniami na poziomie $v \leq 10\%$ sygnału obiektowego.

Przy części uszkodzeń pojedynczych wszystkie badane SE lokalizowały uszkodzenia z dużym prawdopodobieństwem poprawności ($P_{dn} \geq 0.9$) w krótkim czasie ($t_{dn} \leq 40$ s). Gorszą jakość diagnozowania uzyskano dla kilku uszkodzeń pojedynczych i wielokrotnych. Kilka przykładów przedstawiono poniżej.

Obiekt z zablokowanym zaworem z_1 - otwarty, uszkodzenie f_2

Najlepszy z zaproponowanych dyskretnych SE (A2) udzielił prawidłowej odpowiedzi po czasie $t_{dn} = 80$ s. Rozmyty SE, o parametrach opisanych na rys. 4.39 (B1), wcześniej ($t_{dn} = 40$ s) sygnalizuje prawidłową lokalizację uszkodzenia z mniejszym od 1 poziomem ufności. Po czasie $t_{dn} = 80$ s poziom ufności przedstawionej diagnozy wzrasta do wartości 1 (rys.4.40).



Rys. 4.40. Przebiegi poziomu ufności diagnoz dla dwóch rozmytych SE przy uszkodzeniu zaworu z_1

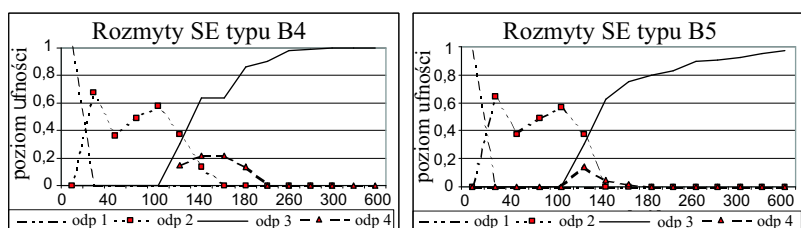
Rozmyty SE B2 w pierwszym etapie diagnozowania działa podobnie do systemu B1, jednakże po czasie $t_{dn} = 80$ s przechodzi w stan ustalony udzielając odpowiedzi obciążonej błędem na poziomie 20%. Pozostałe rozmyte SE (B3, B4, B5) udzielały podobnych lub gorszych diagnoz.

Obiekt z zablokowanym zaworem z_2 - otwarty, uszkodzenie f_3

Wszystkie dyskretne SE po $t_{dn} = 20$ s wykrywają stan odbiegający od normalnego, jednakże uszkodzenie zostaje zlokalizowane nieprawidłowo (uszkodzenie f_2). Po czasie $t_{dn} = 140$ s pojawia się prawidłowa lokalizacja uszkodzenia. Rozmyty SE B4 po czasie $t_{dn} = 20$ s wykrywa uszkodzenie ale również źle lokalizuje uszkodzenie (lokalizuje f_2).

Po czasie $t_{dn} = 100$ s pojawia się informacja z prawidłową lokalizacją i wzrastającym w kolejnych sekundach poziomem ufności. W czasie między $t_{dn} = 100$ s a $t_{dn} = 200$ s badany SE dodatkowo błędnie lokalizuje uszkodzenie f_4 . Po czasie $t_{dn} = 250$ s lokalizowane jest tylko prawidłowe uszkodzenie z poziomem ufności 1.

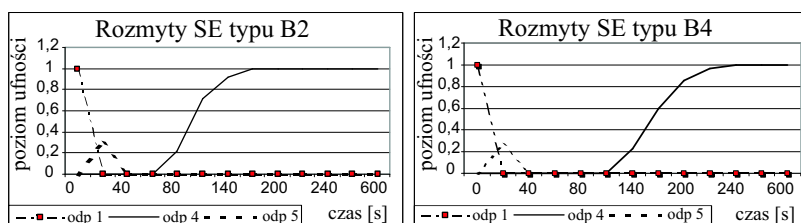
Na rys. 4.41 przedstawiono również odpowiedzi rozmytego SE B5. Uzyskano podobne wyniki jak dla B4, jednakże ten system wykazuje mniejszy błąd lokalizacji uszkodzenia f_4 ale w zamian za to wykazuje mniejszy przyrost poziomu ufności prawidłowej lokalizacji w czasie. Pozostałe rozmyte SE pracują podobnie lub gorzej od omówionych.



Rys. 4.41. Przebiegi poziomu ufności diagnoz dla dwóch rozmytych SE przy uszkodzeniu zaworu z_2

Obiekt z zablokowanymi zaworami z_1 i z_2 - otwarte, uszkodzenia f_4

W tym przypadku pojawia się problem lokalizacji dwóch uszkodzeń jednocześnie. Po czasie $t_{dn} = 120$ s dyskretny SE (A1) prawidłowo lokalizuje uszkodzenia. Rozmyty SE B2 po czasie $t_{dn} = 20$ s wykrywa uszkodzenie, błędnie lokalizując uszkodzenie f_5 (wyciek ze zbiornika 1) (rys.4.42).



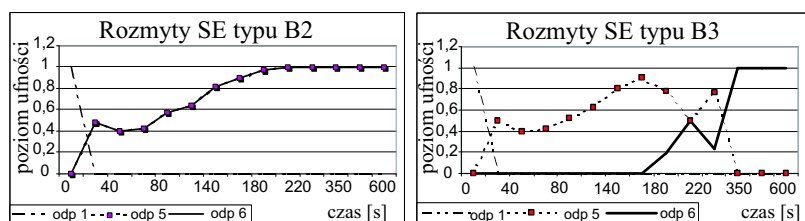
Rys. 4.42. Przebiegi poziomu ufności diagnoz dla dwóch rozmytych SE przy uszkodzeniu zaworów z_1 i z_2

Dopiero po czasie $t_{dn} = 80$ s prawidłowo lokalizowane jest uszkodzenie f_4 z wzrastającym do czasu $t_{dn} = 180$ s poziomem ufności. Drugi rozmyty SE B4 udziela podobnych odpowiedzi, jednakże prawidłowa lokalizacja pojawia się dopiero po $t_{dn} = 120$ s. Pozostałe rozmyte SE pracują podobnie lub gorzej od omówionych.

Obiekt z wyciekiem ze zbiornika 1 i zablokowanym zaworem z_1 - otwarty, uszkodzenia f_6

W tym przypadku ponownie wprowadzono dwa uszkodzenia jednocześnie. Dyskretny SE A3 prawidłowo zlokalizował uszkodzenia po czasie $t_{dn} = 80$ s. Rozmyty SE B2 szybko lokalizuje oba uszkodzenia z wzrastającym poziomem ufności (rys.4.43). Jednakże SE B3 w pierwszym okresie po uszkodzeniu lokalizuje tylko uszkodzenie f_5 (wyciek w zbiorniku 1), aby po czasie $t_{dn} = 160$ s zlokalizować również drugie uszkodzenie. Pozostałe rozmyte SE pracują podobnie lub gorzej od omówionych.

Przedstawione wyniki badań obejmują ocenę pracy dwóch rodzajów SE. Przedstawiając jakość pracy dyskretnych SE, przedstawiono wyniki pracy najlepszego



Rys. 4.43. Przebiegi poziomu ufności diagnoz dla dwóch rozmytych SE przy wycieku w zbiorniku 1 i uszkodzonym zaworze z1

systemu w zaistniałych warunkach pracy obiektu. W tym przypadku należy zauważyć, że dobre wyniki uzyskuje się w przypadku dostrojenia systemu do oczekiwanego uszkodzenia. Poszukiwanie uniwersalnego dyskretnego SE prowadzi do pogorszenia jakości diagnozowania we wszystkich badanych stanach obiektu.

Podobne wyniki osiągnięto dla rozmytych SE. Analizując przedstawione wyniki, należy zauważyć, że rozmieszczenie zbiorów rozmytych, reprezentujących wiedzę o stanie obiektu, ma istotne znaczenie dla jakości uzyskiwanych diagnoz. Ilość zastosowanych rozmytych zbiorów ma duży wpływ na rozróżnialność wybranych uszkodzeń (jest to szczególnie ważne w przypadku wielokrotnych uszkodzeń). W rozmytych SE zalecane jest stosowanie trapezowych krzywych rozmywających, pozwala to unikać fałszywych alarmów. Analizując uzyskane wyniki, stwierdzono, że rozmyty SE B3 zapewnia najlepszą jakość lokalizacji wszystkich wprowadzanych do obiektu uszkodzeń, zarówno pojedynczych jak i wielokrotnych.

4.5. Zakończenie

Diagnostyczny system ekspertowy powinien cechować się dużą niezawodnością działania. Wyraża to się prawidłowością wydawanych konkluzji, minimalizowaniem fałszywych alarmów oraz odpowiednio krótkim czasem udzielania odpowiedzi. Aby zapewnić realizację tych wymagań, autor proponuje zadanie diagnostyki złożonych procesów realizować z użyciem hybrydowego SE.

W powyższym rozdziale przedstawiono podstawy stosowalności wybranych elementów sztucznej inteligencji oraz ocenę ich skuteczności w realizacji procesu diagnostycznego. Ponadto dokonano analizy obszarów stosowalności różnych form reprezentacji wiedzy diagnostycznej.

Technikę sieci neuronowych proponuje się stosować do budowy modeli wybranych podzespołów badanego procesu przemysłowego. Znajduje to szczególne zainteresowanie w przypadku obiektów o nieliniowych charakterystykach lub nieznanymi modelami analitycznymi. Dokonując analizy zasad działania sieci i przykładów zastosowań w diagnostyce różnych obiektów oraz oceniając wyniki własnych ba-

dań, zauważono, że szczególnie dobre wyniki osiąga się dla obiektów stosunkowo mało złożonych. Przy obiektach złożonych sieć zawiera bardzo dużo parametrów do nauczenia i proces uczenia staje się bardzo trudny lub niekiedy niemożliwy. W przypadku złożonych obiektów zaleca się dekompozycję obiektu i stosowanie neuronowych wielomodeli, połączonych w różne struktury. Drugi obszar zastosowań sieci neuronowych w diagnostyce dotyczy budowy klasyfikatorów stanu badanego obiektu. Jednakże również w tym obszarze sieci neuronowe mają ograniczone zastosowania. Dokonując badań klasyfikatorów neuronowych, stwierdzono małą skuteczność tego typu układów dla przypadków uszkodzeń wielokrotnych lub pojawiających się sekwencyjnie. W takich sytuacjach zaproponowano zintegrowanie bazy wiedzy, zawierającej inne formy reprezentacji wiedzy.

Ostatnio szeroko rozpatruje się zagadnienia stosowania logiki rozmytej w reprezentowaniu wiedzy diagnostycznych systemów. Autor przedstawił kilka przykładów zastosowania rozmytej reprezentacji wiedzy do realizacji różnych funkcji systemu diagnostycznego. Rozmyta reprezentacja wektora residuów umożliwia budowę elastycznych tablic decyzyjnych, zawierających poziomy ufności podejmowanych decyzji. W proponowanym systemie diagnostycznym, z rozmytą reprezentacją symptomów uszkodzeń, do oceny poprawności funkcjonowania SE zastosowano wskaźniki miary skuteczności detekcyjnej zbioru testów diagnostycznych oraz wskaźniki wystąpienia uszkodzenia.

Jakość procesu diagnozowania w układzie z modelem zależy w dużej mierze od jakości zastosowanego modelu. Budowa modeli dla obiektów wielowymiarowych wymaga dużej liczby danych uczących, używanych w procesie strojenia takiego modelu. W aktualnym rozdziale autor przedstawił propozycje budowy rozmytych modeli w różnych konfiguracjach struktury rozproszonej. Zaproponowano również różne techniki optymalizacji parametrów omawianych struktur. Porównano zasady strojenia zdroworozsądkowego jak również optymalizacji kierunkowej w strukturach neuro-rozmytych (zaadoptowany algorytm propagacji wstecznej) i algorytmach ewolucyjnych.

Stosowanie systemów ekspertowych umożliwia polepszenie jakości, efektywności i niezawodności procesów przemysłowych. Za pomocą tych systemów można budować układy sterowania, tolerujące pewną klasę uszkodzeń procesów, systemy diagnostyki zabezpieczeniowej oraz układy przewidywania stanów odbiegających od normy i zagrażających poprawnemu realizowaniu zadań analizowanego procesu przemysłowego. Diagnostyka oparta na regułowym SE wymaga przygotowania bazy wiedzy. Wymusza to przede wszystkim definiowanie wartości niezbędnej liczby progów określających poszczególne stany diagnozowanego procesu. W przypadku klasycznego SE wartości progów definiuje się w reprezentacji dyskretnej. Budując SE z rozmytą reprezentacją wiedzy, należy zdefiniować liczbę oraz rozmieszczenie stosowanych rozmytych zbiorów. Do budowy SE, realizującego zadanie diagnozowania złożonego procesu przemysłowego, zaproponowano zastosować regułowy

SE z rozmytą reprezentacją klas stanów procesu. Aby ocenić skuteczność diagnostyczną zaproponowanego systemu, przeprowadzono szereg eksperymentów, które pozwoliły ocenić wrażliwość diagnostycznego SE na parametry przygotowanej bazy wiedzy. Analiza otrzymanych wyników pozwala wnioskować, że taka struktura SE nie zapewnia dobrej jakości znalezionej odpowiedzi na zadanie diagnostyczne. Do poprawy jakości otrzymywanych diagnoz proponuje się zastosować SE, w którym połączy się różne reprezentacje wiedzy (system hybrydowy) [4, 178].