

prof. dr hab. inż. Leszek Rutkowski, dr inż. Robert Nowicki,
Politechnika Częstochowska,
dr inż. Agata Pokropińska,
Akademia im. Jana Długosza w Częstochowie

METODY PROJEKTOWANIA ROZMYTYCH STEROWNIKÓW

W pracy przedstawiono metody projektowania rozmytych sterowników z wykorzystaniem podejścia Mamdaniego, logicznego oraz Takagi-Sugeno. Zaproponowano nową metodę oceny efektywności działania rozmytych sterowników. Metoda ta została oparta na kryteriach statystycznych zastosowanych w celu sprowadzenia optymalizacji wielokryterialnej do optymalizacji jednokryterialnej. Istotnym elementem tej metody jest wprowadzenie pojęcia linii izokryterialnych, które w znacznym stopniu ułatwiają projektantowi orientację w zbiorze potencjalnych rozwiązań. Wyniki przedstawionych w pracy badań potwierdzą jej przydatność w procesie projektowania systemów rozmytych. Metodę można również z powodzeniem zastosować w projektowaniu innych podobnych systemów, jak np. sieci neuronowych lub systemów ekspertowych.

METHODS FOR DESIGNING FUZZY CONTROLLERS

In the paper the problem of fuzzy controllers designing is considered as the multicriterial optimization. Various Mamdani-type, logical-type and Takagi-Sugeno neuro-fuzzy systems, serving as fuzzy controllers, are studied and compared. Methods for the autoregressive order process estimation are adopted to solve problem of the trade-off between accuracy and number of parameters used to the construction of neuro-fuzzy systems.

1. WSTĘP

W literaturze światowej znane są różne struktury systemów neuronowo-rozmytych [1], [2], [7], które znalazły powszechne zastosowanie w zadaniach sterowania, identyfikacji, klasyfikacji i predykcji. Projektanci i użytkownicy tych systemów są zainteresowani dużą dokładnością ich działania w sensie przyjętego kryterium jakości. W zadaniach aproksymacji i predykcji takim kryterium jakości jest błąd średniokwadratowy, natomiast w zadaniach klasyfikacji jest liczba błędnie sklasyfikowanych próbek.

W obu zadaniach eksperymenty przeprowadza się na ciągach uczących oraz ciągach testowych. Należy podkreślić, że zadawalające rezultaty uzyskane na ciągu uczącym nie gwarantują poprawnej pracy systemu na ciągu testowym. Innymi słowy, system neuronowo-rozmyty powinien mieć dobre własności tzw. generalizacji. W szczególności systemy neuronowo-rozmyte zaprojektowane z wykorzystaniem zarówno funkcji przynależności, jak i wag opisujących ważność reguł oraz ważność zmiennych lingwistycznych w poszczególnych regułach powinny charakteryzować się odpowiednią liczbą wszystkich parametrów, które mają być przedmiotem uczenia. Duża liczba parametrów zapewnia mały błąd uczenia, ale zazwyczaj prowadzi do złej generalizacji. Natomiast mała liczba parametrów w systemie prowadzi do większego błędu uczenia. W literaturze światowej nie są znane kompleksowe badania dotyczące efektywności działania systemów neuronowo-rozmytych. Zagadnienie zaprojektowania systemów neuronowo-rozmytych, które charakteryzują się osiągnięciem kompromisu między błędem działania systemu, a liczbą parametrów opisujących ten system

błędem działania systemu, a liczbą parametrów opisujących ten system nie doczekało się rozwiązania w literaturze światowej.

W artykule przedstawimy różne typy sterowników rozmytych zrealizowanych z wykorzystaniem struktur neuronowo-rozmytych, jak również zbadamy efektywność działania tych sterowników. Przez efektywność działania systemu neuronowo-rozmytego rozumiemy osiągniętą dokładność działania takiego systemu (wyrażoną błędem średniokwadratowym lub liczbą błędnie sklasyfikowanych próbek) w kontekście jego rozmiaru. Przez rozmiar systemu rozumiemy liczbę wszystkich parametrów, które podlegają uczeniu. Rozwiążemy zagadnienie zaprojektowania systemów neuronowo-rozmytych, które charakteryzują się osiągnięciem kompromisu między błędem działania systemu, a liczbą parametrów opisujących ten system. Podamy koncepcję tzw. linii izokryterialnych, które mogą być podstawą do zaprojektowania sterowników optymalnych w sensie przyjętych kryteriów.

2. SYSTEMY ROZMYTE

Systemy rozmyte są szczególną klasą systemów wnioskujących. Przetwarzają one informacje w postaci zbiorów rozmytych [14]. W stosowanej powszechnie uogólnionej metodzie modus ponens wnioskowanie odbywa się na podstawie przesłanki

$$\mathbf{x} \text{ jest } A' \quad (1)$$

oraz reguł składających się na wiedzę w postaci

$$R^{(r)}: \text{JEŻELI } \mathbf{x} \text{ jest } A' \text{ TO } y \text{ jest } B^r, \quad (2)$$

gdzie $r = 1, \dots, N$, $\mathbf{x} = [x_1, x_2, \dots, x_n]^T \in \mathbf{X}$ jest wektorem wielkości wejściowych, $A' = A'_1 \times A'_2 \times \dots \times A'_n \subseteq \mathbf{X}$ jest iloczynem kartezjańskim ich wartości wyrażonych w postaci zbiorów rozmytych, $A^r = A^r_1 \times A^r_2 \times \dots \times A^r_n \subseteq \mathbf{X}$ jest iloczynem kartezjańskim wartości wielkości wejściowych \mathbf{x} określonych w poprzedniku reguły $R^{(r)}$, $y \in \mathbf{Y}$ jest wielkością wyjściową, natomiast B^r jest jej wartością w postaci zbioru rozmytego określoną w następniku reguły $R^{(r)}$. Należy zauważyć, że w niniejszych rozważaniach ograniczono się do systemów posiadających n wejść x_i , gdzie $i = 1, \dots, n$, ale tylko jedno wyjście y . Nie ogranicza to ogólności rozważań, gdyż system o m wyjściach możemy traktować jako m systemów o jednym wyjściu. W bardziej ogólnym opisie wielkość wyjściowa y mogłaby być zdefiniowana jako wektor, analogiczny do wektora wejściowego \mathbf{x} , a zbiór rozmyty B^r jako iloczyn kartezjański odpowiednich zbiorów rozmytych B^r_j , gdzie $j = 1, \dots, m$, analogicznie do zbioru A^r .

W tzw. podejściu lokalnym dla każdej reguły $R^{(r)}$ otrzymujemy wniosek w następującej formie:

$$y \text{ jest } \bar{B}^r. \quad (3)$$

Zbiór rozmyty $\bar{B}^r \subseteq \mathbf{Y}$ powstaje w wyniku złożenia zbioru rozmytego A^r pochodzącego z przesłanki oraz relacji rozmytej $(A^r \rightarrow B^r)$ wyrażającej regułę $R^{(r)}$ w sposób charakterystyczny dla poszczególnych typów systemów rozmytych. W przypadku systemu o m wyjściach zbiór rozmyty \bar{B}^r jest zdefiniowany jako iloczyn kartezjański zbiorów \bar{B}^r_j , $j = 1, \dots, m$.

W niektórych przypadkach wymagane jest ustalenie wspólnego wniosku

$$y \text{ jest } B' \quad (4)$$

dla wszystkich N reguł. Można to zrealizować poprzez agregację wniosków postaci (3). Realizacja agregacji jest zależna od typu systemu rozmytego.

W większości praktycznych zastosowań rozmyta postać przesłanki i wniosku nie jest dogodna. Z przetworników pomiarowych uzyskujemy rzeczywistą (nie rozmytą) wartość mierzonej wielkości, podobnie elementy wykonawcze wymagają rzeczywistych wymuszeń – sygnałów sterujących. Wektor rzeczywistych wartości wejściowych $\bar{\mathbf{x}} = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n]^T$ musi więc zostać poddany operacji rozmywania polegającej na odwzorowaniu go w przestrzeń zbiorów rozmytych. Najpowszechniejszym rozwiązaniem jest zastosowanie rozmywania typu singleton przedstawionego wzorem

$$\mu_{A'}(\mathbf{x}) = \begin{cases} 1 & \text{jeżeli } \mathbf{x} = \bar{\mathbf{x}} \\ 0 & \text{jeżeli } \mathbf{x} \neq \bar{\mathbf{x}} \end{cases} \quad (5)$$

Podobnie zbiory rozmyte \bar{B}^r występujące we wnioskach postaci (3) lub zbiór rozmyty B' występujący wniosku postaci (4) musi być przekształcony w reprezentatywną dla niego wartość rzeczywistą. Operację realizującą to zadanie nazywamy wyostrzaniem. Realizacja wyostrzania, podobnie jak agregacji, jest zależna od typu systemu rozmytego.

Ponieważ jedną z wygodniejszych metod implementacji systemów rozmytych są systemy rozmyto-neuronowe, w dalszym ciągu pracy ograniczymy się właśnie do nich. Wśród szerokiej gamy tych systemów skupimy się na kilku charakterystycznych przykładach. Będą to systemy typu Mamdaniego oparte o wyostrzanie typu CA (ang. Center Average) oznaczone tu przez literą A, systemy Mamdaniego oparte o wyostrzanie typu DCOA (ang. Discrete Center of Area) oznaczone tu przez literą B oraz systemy typu logicznego. Uwzględnione zostaną również systemy typu Takagi-Sugeno (TSK) charakteryzujące się odmienną postacią reguł i odmiennym schematem wnioskowania.

2.1. Systemy typu Mamdaniego (A)

W systemach typu A wyostrzanie realizowane jest za pomocą zależności:

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot \mu_{\bar{B}^r}(\bar{y}^r)}{\sum_{r=1}^N \mu_{\bar{B}^r}(\bar{y}^r)} \quad (6)$$

Funkcje przynależności zbiorów rozmytych \bar{B}^r , $r = 1, 2, \dots, N$, zdefiniowane są za pomocą następującego wzoru:

$$\mu_{\bar{B}^r}(y) = \sup_{\mathbf{x} \in \mathbf{X}} \left\{ \mu_{A'}(\mathbf{x}) * \mu_{A' \rightarrow B^r}(\mathbf{x}, y) \right\} \quad (7)$$

Przy rozmywaniu typu singleton wzór (7) przyjmuje postać:

$$\mu_{\bar{B}^r}(y) = \mu_{A' \rightarrow B^r}(\bar{\mathbf{x}}, y) = T(\mu_{A'}(\bar{\mathbf{x}}), \mu_{B^r}(y)) \quad (8)$$

Ponieważ

$$\mu_{A^r}(\bar{\mathbf{x}}) = T_{i=1}^n \left(\mu_{A_i^r}(\bar{x}_i) \right), \quad (9)$$

więc

$$\mu_{B^r}(y) = \mu_{A^r \rightarrow B^r}(\bar{\mathbf{x}}, y) = T \left[T_{i=1}^n \left(\mu_{A_i^r}(\bar{x}_i) \right), \mu_{B^r}(y) \right], \quad (10)$$

gdzie T jest dowolną t-normą. Korzystając z faktu, że:

$$\mu_{B^r}(\bar{y}^r) = 1 \quad (11)$$

oraz

$$T(a, 1) = a \quad (12)$$

otrzymuje się następującą zależność:

$$\mu_{B^r}(\bar{y}^r) = T_{i=1}^n \left(\mu_{A_i^r}(\bar{x}_i) \right). \quad (13)$$

Podstawiając zależność (13) do wzoru (6) otrzymujemy:

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot T_{i=1}^n \left(\mu_{A_i^r}(\bar{x}_i) \right)}{\sum_{r=1}^N T_{i=1}^n \left(\mu_{A_i^r}(\bar{x}_i) \right)}. \quad (14)$$

W systemach typu A realizuje się osobne wnioskowanie w obrębie każdej reguły i oblicza się $\mu_{B^r}(\bar{y}^r)$, $r = 1, 2, \dots, N$. Załóżmy, że zmienne lingwistyczne wejściowe i wyjściowe są opisane gaussowskimi funkcjami przynależności, to znaczy:

$$\mu_{A_i^r}(x_i) = \exp \left[- \left(\frac{x_i - \bar{x}_i^r}{\sigma_i^r} \right)^2 \right], \quad (15)$$

$$\mu_{B^r}(y) = \exp \left[- \left(\frac{y - \bar{y}^r}{\sigma^r} \right)^2 \right]. \quad (16)$$

Podstawiając powyższe zależności do wzoru (6) oraz stosując regułę Larsena otrzymamy następujący wzór:

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \left(\prod_{i=1}^n \exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^r}{\sigma_i^r} \right)^2 \right] \right)}{\sum_{r=1}^N \left(\prod_{i=1}^n \exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^r}{\sigma_i^r} \right)^2 \right] \right)}. \quad (17)$$

Zauważmy, że w zależności (17) nie występuje parametr σ^r wyjściowego zbioru rozmytego B^r , $r = 1, 2, \dots, N$.

2.2. Systemy typu Mamdaniego (B)

W systemach typu B wyostrzenie realizowane jest za pomocą zależności:

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \mu_{B'}(\bar{y}^r)}{\sum_{r=1}^N \mu_{B'}(\bar{y}^r)}. \quad (18)$$

W systemach tych dokonuje się agregacji poszczególnych zbiorów rozmytych \bar{B}^k danych wzorem (8), to znaczy wyznacza się zbiór rozmyty B' poprzez operację sumy zbiorów rozmytych \bar{B}^k :

$$B' = \bigcup_{k=1}^N \bar{B}^k. \quad (19)$$

Funkcja przynależności zbioru rozmytego B' jest wyznaczona za pomocą t-konormy, to znaczy:

$$\mu_{B'}(y) = S_{k=1}^N \{ \mu_{\bar{B}^k}(y) \}. \quad (20)$$

Zatem

$$\begin{aligned} \mu_{B'}(\bar{y}^r) &= S_{k=1}^N \{ \mu_{\bar{B}^k}(\bar{y}^r) \} = S_{k=1}^N \{ T(\mu_{A^k}(\bar{x}), \mu_{B^k}(\bar{y}^r)) \} \\ &= S_{k=1}^N \left\{ T \left(T_{i=1}^n \mu_{A_i^k}(\bar{x}_i), \mu_{B^k}(\bar{y}^r) \right) \right\}. \end{aligned} \quad (21)$$

Podstawiając wzór (21) do zależności (18) otrzymujemy:

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot S_{k=1}^N \left\{ T \left(T_{i=1}^n \mu_{A_i^k}(\bar{x}_i), \mu_{B^k}(\bar{y}^r) \right) \right\}}{\sum_{r=1}^N S_{k=1}^N \left\{ T \left(T_{i=1}^n \mu_{A_i^k}(\bar{x}_i), \mu_{B^k}(\bar{y}^r) \right) \right\}}. \quad (22)$$

2.3. Systemy typu logicznego

W systemach typu logicznego wyostrzenie realizowane jest za pomocą zależności:

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \mu_{B'}(\bar{y}^r)}{\sum_{r=1}^N \mu_{B'}(\bar{y}^r)}. \quad (23)$$

W systemach tych zbiór rozmyty B' powstaje w wyniku przecięcia zbiorów rozmytych \bar{B}^k , tzn.:

$$B' = \bigcap_{k=1}^N \bar{B}^k. \quad (24)$$

Funkcja przynależności zbioru rozmytego B' jest wyznaczona za pomocą t-normy, co zapisujemy następująco:

$$\mu_{B'}(y) = T_{k=1}^N \{ \mu_{\bar{B}^k}(y) \}. \quad (25)$$

Korzystając ze wzoru (25), (8) i (9) możemy zapisać:

$$\begin{aligned}\mu_{B^k}(\bar{y}^r) &= T_{k=1}^N \left\{ \mu_{B^k}(\bar{y}^r) \right\} = T_{k=1}^N \left\{ I(\mu_{A^k}(\bar{x}), \mu_{B^k}(\bar{y}^r)) \right\} \\ &= T_{k=1}^N \left\{ I \left(T_{i=1}^n \mu_{A_i^k}(\bar{x}_i), \mu_{B^k}(\bar{y}^r) \right) \right\},\end{aligned}\quad (26)$$

gdzie I jest rozmytą implikacją [4]. Podstawiając wzór (26) do zależności (23) otrzymujemy:

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot T_{k=1}^N \left\{ I \left(T_{i=1}^n \mu_{A_i^k}(\bar{x}_i), \mu_{B^k}(\bar{y}^r) \right) \right\}}{\sum_{r=1}^N T_{k=1}^N \left\{ I \left(T_{i=1}^n \mu_{A_i^k}(\bar{x}_i), \mu_{B^k}(\bar{y}^r) \right) \right\}}.\quad (27)$$

Konkretna postać wzoru (27) zależy od przyjętej definicji funkcji I .

2.4. Systemy TSK

W modelu rozmytym typu Takagi-Sugeno [13] baza reguł ma charakter rozmyty tylko w części IF, natomiast w części THEN występują zależności funkcyjne

$$\begin{aligned}R^{(r)} : \text{JEŻELI } \mathbf{x} \text{ jest } A^r \\ \text{TO } y_1 = f^{(r)}(x_1, x_2, \dots, x_n)\end{aligned}\quad (28)$$

Jeżeli założymy, że wejściem systemu rozmytego jest sygnał $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$ to, aby otrzymać sygnał wyjściowy \bar{y} systemu, najpierw wyznaczamy:

$$T(\mu_{A_1^r}(\bar{x}_1), \mu_{A_2^r}(\bar{x}_2), \dots, \mu_{A_n^r}(\bar{x}_n)), \quad r = 1, \dots, N.\quad (29)$$

Następnym krokiem jest obliczenie:

$$\bar{y}_r = f^{(r)}(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n), \quad r = 1, \dots, N.\quad (30)$$

Sygnał wyjściowy systemu rozmytego Takagi-Sugeno jest znormalizowaną sumą ważoną poszczególnych wyjść $\bar{y}_1, \dots, \bar{y}_N$, tzn.

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}_r \cdot T_{i=1}^n \left\{ \mu_{A_i^r}(\bar{x}_i) \right\}}{\sum_{r=1}^N T_{i=1}^n \left\{ \mu_{A_i^r}(\bar{x}_i) \right\}}.\quad (31)$$

W dalszej części tego podrozdziału rozważymy systemy Takagi-Sugeno z liniowymi zależnościami w następnikach bazy reguł, tzn.:

$$\begin{aligned}R^{(r)} : \text{JEŻELI } \mathbf{x} \text{ jest } A^r \\ \text{TO } y_r = c_0^{(r)} + c_1^{(r)}x_1 + \dots + c_n^{(r)}x_n,\end{aligned}\quad (32)$$

dla $r = 1, \dots, N$. Warto zauważyć, że jeżeli $c_i^{(r)} = 0$, $i = 1, \dots, n$, to system (31) redukuje się do uproszczonego systemu Mamdaniego danego wzorem (14) i wówczas $c_0^{(r)} = \bar{y}^r$, $r = 1, \dots, N$.

2.5. Zastosowanie ważonych norm trójkątnych

Wszystkie systemy przedstawione powyżej charakteryzują się równym traktowaniem wszystkich reguł oraz wszystkich wielkości wejściowych. Może jednak istnieć wiele powodów, dla których należałoby zróżnicować ważność poszczególnych reguł i/lub poszczególnych wielkości wejściowych. Realizacja tej idei jest możliwa dzięki zastosowaniu ważonych norm trójkątnych w operacjach iloczynu kartezjańskiego przestrzeni wejściowych oraz agregacji wniosków reguł. Ważone normy trójkątne definiuje się następująco:

Definicja

Ważoną T-normę oznaczamy jako T^* i definiujemy [3], [9], [10] w sposób następujący

$$T^*\{a_1, \dots, a_n; w_1, \dots, w_n\} = T \left\{ 1 - w_i (1 - a_i) \right\}, \tag{33}$$

gdzie T jest dowolną T-normą, natomiast wagi spełniają warunek $0 \leq w_i \leq 1, i = 1, \dots, n$.

Definicja

Ważoną t-konormę oznaczamy jako S^* i definiujemy [3], [9], [10] w sposób następujący

$$S^*\{a_1, \dots, a_n; w_1, \dots, w_n\} = S \left\{ w_i a_i \right\}, \tag{34}$$

gdzie S jest dowolną t-konormą, natomiast wagi spełniają warunek $0 \leq w_i \leq 1, i = 1, \dots, n$.

Jeśli wagę r-tej reguły oznaczmy przez w_r , a wagę i-tej wielkości wejściowej w r-tej regule przez $w_{i,r}$, to opis systemów rozmyto-neuronowych uwzględniających zróżnicowaną istotność poszczególnych reguł i wielkości wejściowych [9], [10], [11] będzie następujący:

System typu Mamdaniego (A)

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot w_r \left[T \left\{ 1 - w_{i,r} (1 - \mu_{A_i^r}(\bar{x}_i)) \right\} \right]}{\sum_{r=1}^N w_r \left[T \left\{ 1 - w_{i,r} (1 - \mu_{A_i^r}(\bar{x}_i)) \right\} \right]}. \tag{35}$$

System typu Mamdaniego (B)

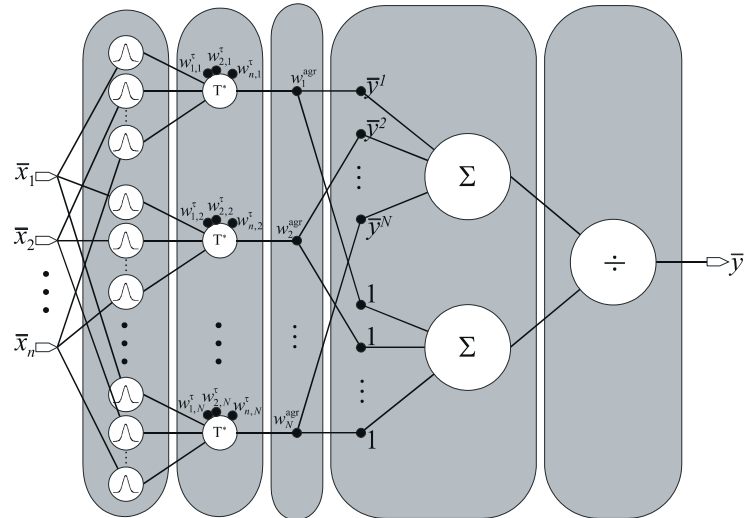
$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot S^* \left\{ T \left(T^* \left\{ \mu_{A_i^k}(\bar{x}_i), w_{i,k} \right\}, \mu_{B^k}(\bar{y}^r) \right), w_k \right\}}{\sum_{r=1}^N S^* \left\{ T \left(T^* \left\{ \mu_{A_i^k}(\bar{x}_i), w_{i,k} \right\}, \mu_{B^k}(\bar{y}^r) \right), w_k \right\}}. \tag{36}$$

Systemy typu logicznego

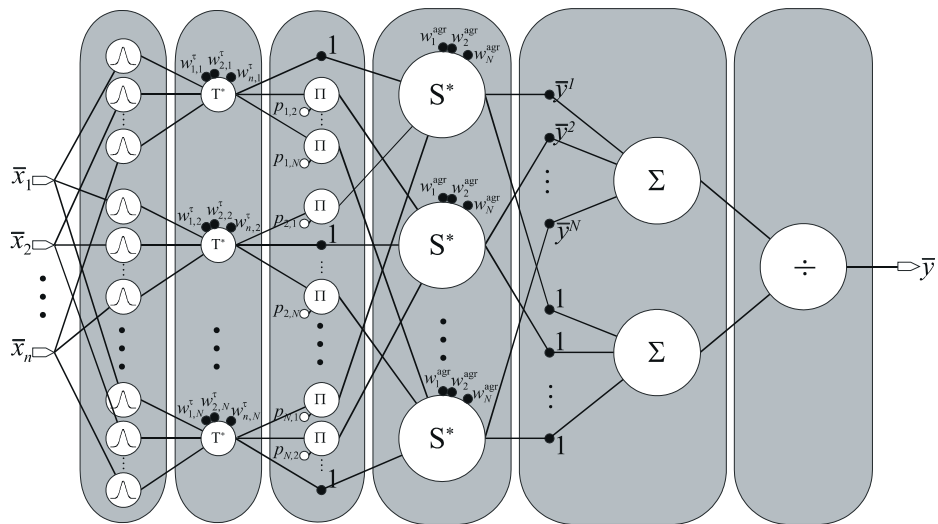
$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot T^* \left\{ I \left(T^* \left\{ \mu_{A_i^k}(\bar{x}_i), w_{i,k} \right\}, \mu_{B^k}(\bar{y}^r) \right), w_k \right\}}{\sum_{r=1}^N T^* \left\{ I \left(T^* \left\{ \mu_{A_i^k}(\bar{x}_i), w_{i,k} \right\}, \mu_{B^k}(\bar{y}^r) \right), w_k \right\}}. \tag{37}$$

Na rys. 1-3 przedstawiono schematy przykładowych systemów rozmyto-neuronowych. Rys. 3 dotyczy systemu z implikacją logiczną typu binarnego, tzn.

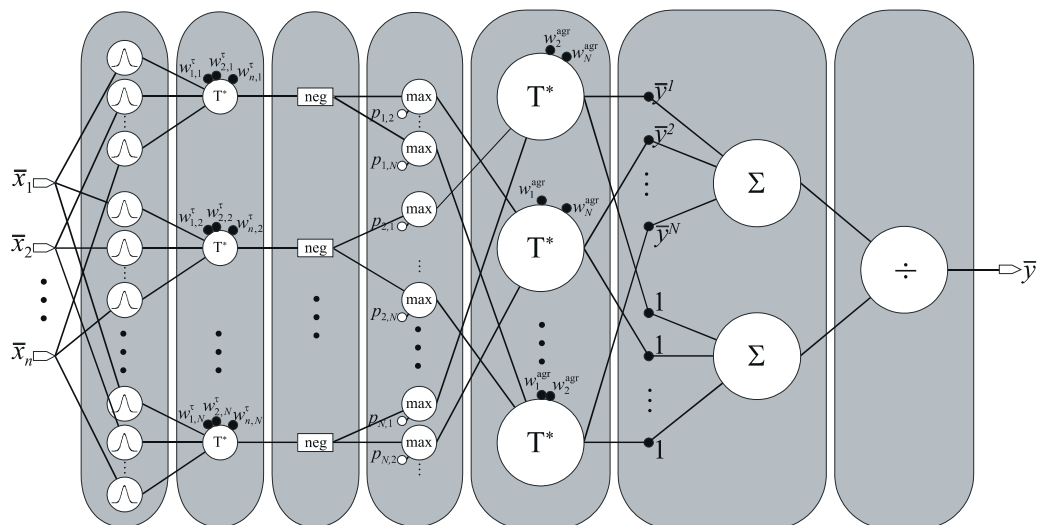
$$I(a, b) = \max\{1 - a, b\}. \tag{38}$$



Rys. 1. Struktura systemu typu Mamdaniego (A)



Rys. 2. Przykładowa struktura systemu typu Mamdaniego (B)



Rys. 3. Przykładowa struktura systemu typu logicznego

3. OCENA SYSTEMÓW ROZMYTYCH

3.1. Błąd aproksymacji i liczba błędnych decyzji

Ocena funkcjonowania systemów rozmytych oraz rozmyto-neuronowych może być różna w zależności od zakresu ich zastosowań. Gdy zadaniem stawianym przed systemem jest aproksymacja funkcji, oceną jest miarą rozbieżności (błędu) między wartością aproksymowaną, a wartością uznaną za poprawną. Może on być wyrażony przez kryterium RMSE (ang. Root Mean Square Error)

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\bar{y}(t) - d(t))^2} . \quad (39)$$

W przypadku zadań sterowania może to być jedna ze stosowanych w automatyce kryteriów jakości sterowania. W zagadnieniach klasyfikacji jako kryterium oceny możemy przyjąć liczbę (procent) błędnie klasyfikowanych przypadków.

3.2. Rozmiar systemu

Jak wspomniano we wstępie ocena wielkości badanych systemów rozmytych oparta będzie o liczbę parametrów, które przechowują wiedzę – są to jednocześnie parametry, które podlegają uczeniu. Podejście to jest zgodne ze stosowanymi w przypadku sieci neuronowych. Dla danego typu sieci podaje się zwykle liczbę warstw oraz liczbę neuronów w poszczególnych warstwach, co można bezpośrednio przeliczyć na liczbę wag – elementów przechowujących wiedzę i podlegających uczeniu. W systemach rozmytych uczone parametry możemy podzielić na kilka grup. Liczba parametrów zbiorów poprzedników reguł zależy od zastosowanych funkcji przynależności. Dla systemów przedstawionych powyżej w przypadku zastosowania zbiorów o gaussowskich funkcjach przynależności, o dwóch parametrach σ_i^r i \bar{x}_i^r , wynosi ona $2 \cdot n \cdot N$. Dla funkcji przynależności o trzyparametrowych funkcjach przynależności (np. trójkątnych) będzie ona wynosić $3 \cdot n \cdot N$. W przypadku systemów rozmyto-neuronowych uwzględniających istotność poszczególnych wielkości wejściowych kolejną grupą parametrów są wagi wejść w liczbie $n \cdot N$. W przypadku systemów rozmyto-neuronowych uwzględniających istotność poszczególnych reguł istnieje grupa wag reguł w liczbie N . Ostatnia grupa parametrów przechowuje informację o następnikach reguł. W przypadku systemów typu Mamdaniego (A) oraz uproszczonych wersji systemów typu logicznego, gdzie każdy zbiór jest reprezentowany przez jeden parametr \bar{y}^r , parametrów tych jest N . W pozostałych przypadkach ich liczba jest zależna od zastosowanej funkcji przynależności i wynosi $2 \cdot N$ dla funkcji dwuparametrowych (np. gaussowskich) lub $3 \cdot N$ dla funkcji trójparametrowych (np. trójkątnych). W systemach TSK liczba parametrów w tej grupie jest sumą liczby parametrów opisujących funkcje występujące w następnikach wszystkich reguł.

Jest oczywistym, że zwiększenie rozmiarów systemu, czyli zwiększenie liczby uczonej parametrów systemu p wpływa bezpośrednio na zmniejszenie wartości kryterium Q (np. RMSE) lub zmniejszenie liczby błędnych decyzji w odniesieniu do zbioru próbek wykorzystywanych w trakcie uczenia (zbioru uczącego). W przypadku badania systemu próbkami, które nie były wykorzystywane w procesie uczenia (zbiór testowy) zależność ta jest obserwowana tylko w ograniczonym zakresie zmian wielkości badanego systemu. Mamy więc przypadek poszukiwania systemu rozmytego będącego kompromisem między jego wielkością, a dokładnością realizowanego odwzorowania. Jest to więc klasyczny przypadek poszukiwania rozwiązania

optymalnego w sensie Pareto [15], w tym przypadku z uwzględnieniem dwóch sprzecznych kryteriów p i Q. Prowadzi to do określenia całego zbioru rozwiązań (systemów rozmytych) Pareto-optymalnych. W zbiorze Pareto nie można poprawić żadnego z dwóch sprzecznych kryteriów (błąd średniokwadratowy versus rozmiar systemu), nie pogarszając pozostałego. W dalszym ciągu pracy przedstawiona zostanie propozycja zastosowania metod zaczerpniętych z teorii procesów autoregresji. Metody te pozwalają sprowadzić problem optymalizacji wielokryterialnej do optymalizacji jednokryterialnej i istotnie ograniczyć zbiór rozwiązań uznanych za optymalne.

3.3. Kryteria oceny systemów rozmytych

Rozważania prowadzone w punktach 3.1 i 3.2 i próba oceny badanych systemów rozmytych opierały się o wielkość błędu jako kryterium oceny. Z rozważań tych wynika, że zwykle systemy zawierające większą liczbę uczonych parametrów pozwalały na osiągnięcie lepszych rezultatów. Jednakże pożądany system neuronowo-rozmyty powinien charakteryzować się jak najmniejszym błędem, ale jednocześnie powinien być możliwie najprostszymi. Wybór systemu kompromisowego można więc skojarzyć z problemem Pareto [15]. Należy przypomnieć, że systemy o mniejszej liczbie uczonych parametrów charakteryzują się m.in. lepszymi zdolnościami generalizowania uzyskanych rozwiązań. W tym miejscu warto wspomnieć o tzw. zasadzie oszczędności [12]. Zasada ta jest bardzo użyteczna przy wyznaczaniu właściwego rzędu modelu. Można ją sformułować w następujący sposób: spośród dwóch alternatywnych i satysfakcjonujących nas modeli wybieramy ten, w którym występuje mniej niezależnych parametrów. Zasada ta pozostaje w zgodzie ze zdrowym rozsądkiem: „nie wprowadzaj dodatkowych parametrów do opisu procesu, jeśli nie są one niezbędne”.

Metody estymacji rzędu systemu najlepiej zostały rozwinięte dla procesów autoregresji [5], [6], [8]. Szereg czasowy $u(n), u(n-1), \dots, u(n-p)$ jest procesem autoregresji rzędu p, jeżeli spełnia równanie różnicowe

$$u(n) + \alpha_1 u(n-1) + \dots + \alpha_p u(n-p) = e(n) \quad (40)$$

lub, równoważnie

$$u(n) = -\sum_{k=1}^p \alpha_k u(n-k) + e(n) \quad (41)$$

gdzie $\alpha_1, \dots, \alpha_p$ są współczynnikami procesu, natomiast $e(n)$ jest szumem białym

$$E[e(n)] = 0, \quad E[e(n)e(m)] = \begin{cases} \sigma^2 & \text{dla } n = m \\ 0 & \text{dla } n \neq m \end{cases} \quad (42)$$

W teorii autoregresji stosuje się kryteria pozwalające estymować rząd predyktora p wyznaczając błąd predykcji \hat{Q}_p na podstawie ciągu uczącego o długości M. Najważniejsze z nich to kryterium informacyjne Akaike (AIC), metoda Schwarza oraz metoda końcowego błędu predykcji (FPE).

W następnym punkcie najpierw przedstawimy podstawowe kryteria oceny modeli z uwzględnieniem ich złożoności, pierwotnie stosowane do estymacji rzędu procesów autoregresji, a następnie zostaną one zaadoptowane do oceny efektywności systemów neuronowo-rozmytych.

Poniżej przedstawiono dwa ogólne kryteria uwzględniające złożoność modelu, wskazano na zależności między tymi kryteriami, a następnie pokazano ich szczególne postacie.

3.3.1. Kryterium A

Ogólna postać kryterium A uwzględniającego złożoność modelu jest dana wzorem:

$$W(p) = \hat{Q}_p [1 + \beta(M, p)], \quad (43)$$

gdzie \hat{Q}_p jest błędem średniokwadratowym, natomiast $\beta(M, p)$ jest funkcją długości ciągu uczącego M i liczby parametrów p modelu. Aby eliminować zbyt złożone struktury (zgodnie z zasadą oszczędności) zakładamy, że \hat{Q}_p jest błędem średniokwadratowym, natomiast $\beta(M, p)$ jest funkcją długości ciągu uczącego M i liczby parametrów p modelu. Aby eliminować zbyt złożone struktury (zgodnie z zasadą oszczędności) zakładamy, że:

$$\lim_{p \rightarrow \infty} \beta(M, p) = \infty. \quad (44)$$

Jednocześnie, aby obecność członu karzącego w wyrażeniu (43) nie utrudniała zaobserwowania zmniejszania się wartości błędu średniokwadratowego \hat{Q}_p ze wzrostem złożoności modelu przyjmujemy, że:

$$\lim_{M \rightarrow \infty} \beta(M, p) = 0. \quad (45)$$

Typowy wybór to $\beta(M, p) = 2p / M$ i wówczas:

$$W(p) = \hat{Q}_p \left[1 + \frac{2p}{M} \right]. \quad (46)$$

3.3.2. Kryterium B

Alternatywnym kryterium w stosunku do wzoru (43) może być następująca formuła:

$$W(p) = M \log \hat{Q}_p + \gamma(M, p), \quad (47)$$

gdzie człon dodatkowy $\gamma(M, p)$ powinien uwzględniać karę za przyjęcie modeli zbyt wysokiego rzędu. Łatwo sprawdzić, że jeśli:

$$\gamma(M, p) = M\beta(M, p), \quad (48)$$

to kryteria (43) i (47) są asymptotycznie równoważne.

Poniżej zostaną przedstawione podstawowe metody kompromisowego doboru rzędu modelu. Większość tych metod to przypadki szczególne wyżej przedstawionego kryterium A lub B.

3.3.3. Metoda kryterium informacyjnego Akaike (AIC)

Przyjęcie $\gamma(M, p) = 2p$ w kryterium (47) daje tzw. informacyjne kryterium Akaike (ang. Akaike Information Criterion). Złożoność systemu p można znaleźć szukając najmniejszą wartość wyrażenia:

$$AIC(p) = M \ln \hat{Q}_p + 2p. \quad (49)$$

3.3.4. Metoda końcowego błędu predykcji (FPE)

Kryterium FPE również zaproponował Akaike. W metodzie końcowego błędu predykcji (ang. *Final Prediction Error*), która nie wynika z ogólnych formuł (43) i (47), złożoność systemu p znajdziemy szukając najmniejszej wartości wyrażenia:

$$FPE(p) = \frac{M+p}{M-p} \hat{Q}_p. \quad (50)$$

W wyrażeniu (49) wraz ze wzrostem parametru p czynnik $\frac{M+p}{M-p}$ rośnie, natomiast wartość błędu średniokwadratowego \hat{Q}_p maleje. Zauważmy, że dla dużych wartości M można stosować następujące przybliżenie:

$$FPE(p) = \hat{Q}_p \left[1 + \frac{2p/M}{1-p/M} \right] \approx \hat{Q}_p \left[1 + \frac{2p}{M} \right], \quad (51)$$

które jest typu (43), tzn.

$$\beta(M, p) = \frac{2p}{M}. \quad (52)$$

Wyrażenia (43) i (47) są asymptotycznie równoważne jeśli spełniony jest warunek (48), a zatem:

$$\gamma(M, p) = 2p. \quad (53)$$

W konsekwencji:

$$FPE(p) \approx AIC(p) = M \ln \hat{Q}_p + 2p. \quad (54)$$

Kryteria FPE oraz AIC wykazują tendencję do wyboru modelu zbyt małego rzędu. Dlatego w literaturze [12] proponuje się trzy inne metody opisane poniżej.

3.3.5. Metoda Schwarza

Przyjęcie $\gamma(M, p) = p \log M$ w kryterium (47) daje tzw. kryterium Schwarza. W metodzie tej złożoność systemu p można znaleźć szukając najmniejszą wartość wyrażenia:

$$S(p) = M \ln \hat{Q}_p + p \ln M. \quad (55)$$

3.3.6. Metoda Söderströma i Stoica

Przyjęcie $\gamma(M, p) = 2pc \log(\log M)$, gdzie $c \geq 1$, w kryterium (47) daje tzw. kryterium Söderströma i Stoica. W metodzie tej złożoność systemu p znajduje się szukając najmniejszej wartości wyrażenia:

$$H(p) = M \ln \hat{Q}_p + 2pc \log(\log M). \quad (56)$$

3.3.7. Metoda CAT

W metodzie CAT (ang. *Criterion Autoregressive Transfer Function*) złożoność systemu p można znaleźć szukając najmniejszą wartość wyrażenia:

$$CAT(p) = \frac{1}{M} \sum_{i=1}^p \frac{1}{\bar{Q}_i} - \frac{1}{\bar{Q}_p}, \quad (57)$$

$$\text{gdzie } \bar{Q}_i = \frac{m}{M-i} \hat{Q}_i.$$

Opisane powyżej metody wyznaczania rzędu modelu pierwotnie zostały zaproponowane do analizy procesów autoregresji danych wzorem (40). Jednakże należy stwierdzić, że metody te pozwalają wyznaczyć właściwy rząd modelu niezależnie od tego, czy system należy do klasy struktur modelu, czy też nie [12].

3.4. Metoda linii izokryterialnych

Opisane w poprzednim punkcie metody estymacji rzędu predykcji zaadoptujemy teraz do oceny systemów rozmytych. Dzięki temu poszukiwanie pożądanego systemu rozmytego na podstawie dwóch kryteriów (liczba parametrów oraz błąd średniokwadratowy) sprowadzimy do jednego, wybranego kryterium, tzn. AIC, Schwarza lub FPE. Kryteria te zostały zaadoptowane dla potrzeb oceny systemów neuronowo-rozmytych w następującej formie:

$$AIC(p, \hat{Q}_p) = M \ln \hat{Q}_p + 2p, \quad (58)$$

$$S(p, \hat{Q}_p) = M \ln \hat{Q}_p + p \ln M, \quad (59)$$

$$FPE(p, \hat{Q}_p) = \frac{Mn+p}{Mn-p} \hat{Q}_p, \quad (60)$$

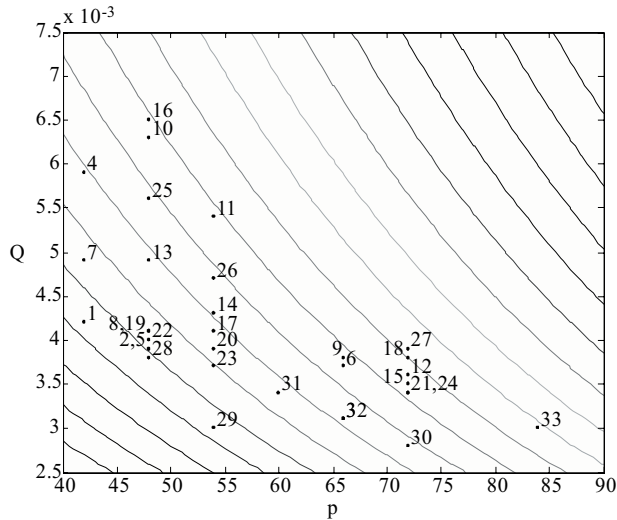
gdzie p jest liczbą parametrów systemu podlegających uczeniu (liczba parametrów wszystkich funkcji przynależności oraz liczba wszystkich wag jeśli występują w danym systemie), \hat{Q}_p jest miarą błędu wykorzystywaną w symulacjach, M określa liczbę próbek w ciągu uczącym, natomiast n jest liczbą wejść systemu. Iloczyn $M \cdot n$ może być zatem traktowany jako miara wielkości rozwiązywanego zagadnienia. Tabele 1 i 2 zawierają wyliczone wartości kryteriów dla poszczególnych badanych struktur dla ciągu uczącego i testowego użytych w zagadnieniu polimeryzacji. Na rys. 4 i 5 zamieszczono punkty odpowiadające poszczególnym badany systemom neuronowo-rozmytym. Współrzędna p określa liczbę parametrów danego systemu, współrzędna Q określa błąd, z jakim system realizuje postawione przed nim zadanie. Linie izokryterialne przedstawiają stałe wartości kryterium AIC, Schwarza i FPE, przy różnych wartościach błędów i liczby parametrów. Takie podejście pozwala rozwiązać problem kompromisu między błędem działania systemu, a liczbą parametrów opisujących ten system. Zagadnienie doboru systemu kompromisowego określone początkowo jako zagadnienia optymalności w sensie Pareto zostało sprowadzone do jednokryterialnego zagadnienia optymalności. Linie izokryterialne o tych samych wartościach kryterium AIC, Schwarza lub FPE, pozwalają ograniczyć zbiór rozwiązań optymalnych w sensie Pareto (systemy neuronowe nr 1, 28, 29 i 30 na rys. 4 oraz 1, 2, 5, 29 i 30 na rys. 5), w prezentowanych przypadkach do jednego najlepszego (w obu przypadkach system nr 1). Punkty leżące na liniach izokryterialnych o najmniejszych wartościach kryterium AIC, Schwarza lub FPE, charakteryzują systemy neuronowo-rozmyte, które nazwano suboptymalnymi. Suboptymalne systemy neuronowo-rozmyte przedstawione na wykresach zapewniają najmniejszą wartość kryteriów statystycznych w ramach przebadanych struktur (nie stosuje się terminologii „optymalne systemy”, gdyż nie przebadano wszystkich możliwych struktur).

Tab. 1. Wartości kryteriów dla ciągu uczącego

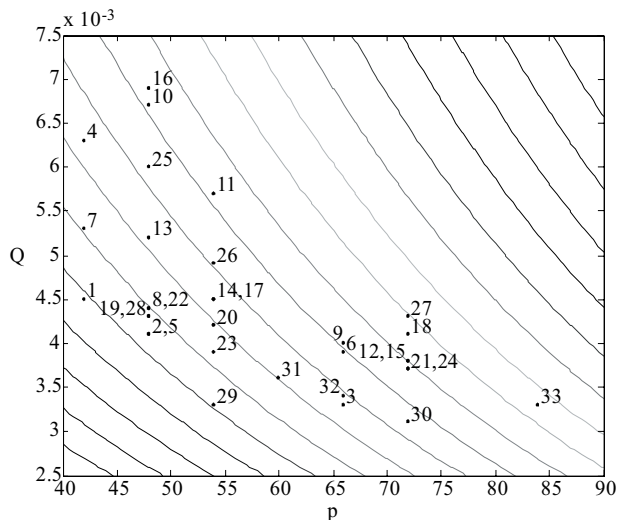
L.p.	Struktura	błąd	p	AIC
1	Larsena uproszczona	0,0042	42	-299,09
2	Larsena uproszczona z wagami reguł	0,0039	48	-292,27
3	Larsena uproszczona z wagami wejść i reguł	0,0031	66	-272,34
4	Łukasiewicza uproszczona	0,0059	42	-275,30
5	Łukasiewicza uproszczona z wagami reguł	0,0039	48	-292,27
6	Łukasiewicza uproszczona z wagami wejść i reguł	0,0037	66	-259,96
7	Zadeha uproszczona	0,0049	42	-288,30
8	Zadeha uproszczona z wagami reguł	0,0041	48	-288,77
9	Zadeha uproszczona z wagami wejść i reguł	0,0038	66	-258,09
10	Binarna	0,0063	48	-258,70
11	Binarna z wagami reguł	0,0054	54	-257,49
12	Binarna z wagami wejść i reguł	0,0036	72	-249,88
13	Larsena	0,0049	48	-276,30
14	Larsena z wagami reguł	0,0043	54	-273,44
15	Larsena z wagami wejść i reguł	0,0035	72	-251,85
16	Łukasiewicza	0,0065	48	-256,52
17	Łukasiewicza z wagami reguł	0,0041	54	-276,77
18	Łukasiewicza z wagami wejść i reguł	0,0038	72	-246,09
19	Mamdaniego	0,0041	48	-288,77
20	Mamdaniego z wagami reguł	0,0039	54	-280,27
21	Mamdaniego z wagami wejść i reguł	0,0034	72	-253,88
22	Reichenbacha	0,0040	48	-290,50
23	Reichenbacha z wagami reguł	0,0037	54	-283,96
24	Reichenbacha z wagami wejść i reguł	0,0034	72	-253,88
25	Willmotta	0,0056	48	-266,95
26	Willmotta z wagami reguł	0,0047	54	-267,21
27	Willmotta z wagami wejść i reguł	0,0039	72	-244,27
28	Zadeha	0,0038	48	-294,09
29	Zadeha z wagami reguł	0,0030	54	-298,64
30	Zadeha z wagami wejść i reguł	0,0028	72	-267,47
31	Takagi-Sugeno	0,0034	60	-277,88
32	Takagi-Sugeno z wagami reguł	0,0031	66	-272,34
33	Takagi-Sugeno z wagami wejść i reguł	0,0030	84	-238,64

Tab. 2. Wartości kryteriów dla ciągu testowego

L.p.	Struktura	błąd	p	AIC
1	Larsena uproszczona	0,0045	42	-294,26
2	Larsena uproszczona z wagami reguł	0,0041	48	-288,77
3	Larsena uproszczona z wagami wejść i reguł	0,0033	66	-267,97
4	Łukasiewicza uproszczona	0,0063	42	-270,70
5	Łukasiewicza uproszczona z wagami reguł	0,0041	48	-288,77
6	Łukasiewicza uproszczona z wagami wejść i reguł	0,0039	66	-256,27
7	Zadeha uproszczona	0,0053	42	-282,80
8	Zadeha uproszczona z wagami reguł	0,0044	48	-283,83
9	Zadeha uproszczona z wagami wejść i reguł	0,0040	66	-254,50
10	Binarna	0,0067	48	-254,40
11	Binarna z wagami reguł	0,0057	54	-253,71
12	Binarna z wagami wejść i reguł	0,0038	72	-246,09
13	Larsena	0,0052	48	-272,14
14	Larsena z wagami reguł	0,0045	54	-270,26
15	Larsena z wagami wejść i reguł	0,0038	72	-246,09
16	Łukasiewicza	0,0069	48	-252,34
17	Łukasiewicza z wagami reguł	0,0045	54	-270,26
18	Łukasiewicza z wagami wejść i reguł	0,0041	72	-240,77
19	Mamdaniego	0,0043	48	-285,44
20	Mamdaniego z wagami reguł	0,0042	54	-275,09
21	Mamdaniego z wagami wejść i reguł	0,0037	72	-247,96
22	Reichenbacha	0,0044	48	-283,83
23	Reichenbacha z wagami reguł	0,0039	54	-280,27
24	Reichenbacha z wagami wejść i reguł	0,0037	72	-247,96
25	Willmotta	0,0060	48	-262,12
26	Willmotta z wagami reguł	0,0049	54	-264,30
27	Willmotta z wagami wejść i reguł	0,0043	72	-237,44
28	Zadeha	0,0043	48	-285,44
29	Zadeha z wagami reguł	0,0033	54	-291,97
30	Zadeha z wagami wejść i reguł	0,0031	72	-260,34
31	Takagi-Sugeno	0,0036	60	-273,88
32	Takagi-Sugeno z wagami reguł	0,0034	66	-265,88
33	Takagi-Sugeno z wagami wejść i reguł	0,0033	84	-231,97



Rys. 4. Linie izokryterialne: wyniki osiągnięte przez poszczególne systemy dla kryterium Akaike dla ciągu uczącego – zagadnienie polimeryzacji



Rys. 5. Linie izokryterialne: wyniki osiągnięte przez poszczególne systemy dla kryterium Akaike dla ciągu testowego – zagadnienie polimeryzacji

Z tab. 1 i tab. 2, jak i z rysunków wynika, że zarówno dla ciągu uczącego, jak i testowego, kryterium AIC najlepiej ocenia system 1 (uproszczona struktura Larsena), a zaraz za nim system 29 (struktura Zadeha z wagami reguł). Na rysunkach tych łatwo zauważyć zbiór systemów optymalnych w sensie Pareto (1, 28, 29 i 30 na rys. 4 oraz 1, 2, 5, 29 i 30 na rys. 5), o którym była mowa wyżej.

4. PODSUMOWANIE

W pracy przedstawiono różne typy rozmytych sterowników zrealizowanych z wykorzystaniem modelowania neuronowo-rozmytego. Zaproponowano również nową metodę oceny sterowników rozmytych. Została ona oparta o kryteria statystyczne zastosowane w celu prowadzenia optymalizacji wielokryterialnej do optymalizacji jednokryterialnej. Istotnym elemen-

tem było wprowadzenie pojęcia linii izokryterialnych, które w znacznym stopniu ułatwiają projektantowi orientację w zbiorze potencjalnych rozwiązań. Wyniki przedstawionych w pracy, jak i innych przeprowadzonych badań potwierdzają jej przydatność w procesie projektowania systemów rozmytych. Metodę można również z powodzeniem zastosować w projektowaniu innych podobnych systemów, jak np. sieci neuronowe, czy systemy ekspertowe.

PODZIĘKOWANIA

Praca była wspierana przez Fundację Nauki Polskiej (Subsydia Profesorskie 2005-2008) oraz Ministerstwo Nauki i Szkolnictwa Wyższego (Projekt Badawczy Specjalny 2006-2009 i Polsko-Singapurski Projekt Badawczy 2008-2010).

LITERATURA

- [1] E. Czogała, J. Łęski, On destructive fuzzy logic controllers, *5-th Zittau Fuzzy Colloquium*, 1997, pp. 8–12.
- [2] Czogała E., Łęski J., *Fuzzy and Neuro-Fuzzy Intelligent Systems*, Physica-Verlag, Heidelberg, New York 2000.
- [3] Dubois D., Prade H., Weighten minimum and maximum operations in fuzzy set theory, *Information Sciences*, vol. 39, 1986, pp. 205 – 210.
- [4] Fodor J. C., On fuzzy implication, *Fuzzy Sets and Systems*, vol. 42, 1991, pp. 293-300.
- [5] Kay S. M., *Modern Spectral Estimation. Theory and Application*, Prentice-Hall, Engewood Cliffs, New Jersey 1988.
- [6] Marple S. L. Jr., *Digital Spectral Analysis with applications*, Prentice-Hall, Englewood Cliffs, New Jersey 1987.
- [7] Rutkowska D., Nowicki R., Implication - based neuro-fuzzy architectures International, *Journal of Applied Mathematics and Computer Science* (4), 2000, pp. 675-701.
- [8] Rutkowski L., *Filtry Adaptacyjne I Adaptacyjne Przetwarzanie Sygnałów: Teoria i Zastosowania*, Wydawnictwo Naukowo-Techniczne WNT, Warszawa 1994.
- [9] Rutkowski L., Cpałka K., Compromise weighted neuro-fuzzy systems, In: L. Rutkowski, J. Kacprzyk (Eds.), *Neural Networks and Soft Computing*, Physica-Verlag, A Springer-Verlag Company, Heidelberg, 2003, pp. 557 – 562.
- [10] Rutkowski L., Cpałka K., Flexible neuro-fuzzy systems, *IEEE Trans. on Neural Networks*, vol. 14, 554-574, 2003.
- [11] Rutkowski L., Cpałka K., Designing and learning of adjustable quasi-triangular norms with applications to neuro-fuzzy systems, *IEEE Trans. on Fuzzy Systems*, vol. 14, 140-151, 2005.
- [12] Söderström T., Stoica p., *Identyfikacja Systemów*, Wydawnictwa Naukowe PWN, Warszawa 1997.
- [13] Takagi T., Sugeno M., Fuzzy identification of systems and its application to modeling and controll, *IEEE Trans. Syst., Man, Cybern.*, vol. 15, 116-132, 1985.
- [14] Zadeh L. A., Fuzzy Sets, *Information and Control*, 1965, vol. 8, pp. 338-353.
- [15] Zalewski A., Cegięłka R., *Matlab – obliczenia numeryczne i ich zastosowania*, Wydawnictwo NAKOM, Poznań 2002.