

BUDOWA ROZMYTYCH MODELI PROCESÓW STOCHASTYCZNYCH PRZY UŻYCIU REGUŁ ASOCJACJI

Anna Walaszek-Babiszewska¹, Katarzyna Błaszczuk², Anna Czabak³

Instytut Automatyki i Informatyki
Politechnika Opolska
ul. K. Sosnkowskiego 31, 45-272 Opole
¹a.walaszek-babiszewska@po.opole.pl
²k.blaszczuk@po.opole.pl
³a.czabak@po.opole.pl

Streszczenie: W referacie przedstawiono zasady budowy rozmytych modeli systemów stochastycznych, z dyskretnym czasem. Wykorzystano modele rozmyte typu Mamdaniego z wagami reguł, reprezentującymi prawdopodobieństwa zdarzeń rozmytych, zawartych w poszczególnych częściach reguł. Zaprezentowano sposób wykorzystania reguł asocjacji dla wyznaczenia wag reguł i rozkładów prawdopodobieństwa zmiennych lingwistycznych w regułach.

1 Wprowadzenie

Modelowanie rozmyte systemów stochastycznych wiąże zalety dwóch podejść: możliwość uwzględnienia niepewności wyrażonej w kategoriach probabilistycznych [5, 8] oraz możliwość granulacji przestrzeni rozważań dla utworzenia opisu w postaci reguł rozmytych, słusznych w poszczególnych częściach przestrzeni [10, 16]. Ponadto, modelowanie rozmyte pozwala uwzględniać wiedzę ekspercką na temat badanego procesu.

Systemy stochastyczne z czasem dyskretnym, ciągi czasowe, znajdują od kilku dziesięcioleci zastosowanie w zagadnieniach sterowania, predykcji, planowania. W referacie kontynuowane jest modelowanie rozmyte takich systemów, prezentowane wcześniej np. w pracach [12-14]. Dostępność narzędzi z zakresu *data mining* powoduje, że zanikają trudności obliczeniowe związane z wyznaczaniem rozkładów prawdopodobieństwa zdarzeń rozmytych procesów wyższych rzędów. W pracy pokazano zastosowanie reguł asocjacji do wyznaczania modeli rozmytych procesów stochastycznych, zwanych też ciągami czasowymi. Przeprowadzono obliczenia na danych empirycznych, istotnych dla produkcji energii elektrycznej.

1.1 Proces stochastyczny z czasem dyskretnym – pojęcia podstawowe

Zakładamy, że parametry mierzalne badanego obiektu, obserwowane w funkcji czasu, stanowią wielowymiarowy proces stochastyczny

$$\{x(t, \omega) : t \in T, x \in X, \omega \in \Omega\}, \quad (1)$$

gdzie T oznacza przedział czasu, X jest przestrzenią badanych parametrów procesu, $X \subset R^p$ oraz Ω jest przestrzenią zdarzeń elementarnych, przy czym

$$x^T(t, \omega) = \{x^1(t, \omega), \dots, x^p(t, \omega)\}. \quad (2)$$

Zbiór danych eksperymentalnych stanowi zbiór realizacji $\{x(t_k)\}$ w dyskretnych chwilach $t_k, k=1, 2, \dots, K$. Dla uproszczenia notacji, zarówno proces stochastyczny i jego realizacja będą dalej oznaczane tym samym symbolem x .

Istnieje wiele modeli systemów stochastycznych z dyskretnym czasem, wśród których warto wyróżnić dwie grupy modeli. Pierwszy rodzaj modeli można zapisać w postaci ogólnej, jako:

$$x(t_k) = f[x(t_{k-1}), x(t_{k-2}), \dots, x(t_{k-m})], \quad (3)$$

gdzie $x(t_k)$ jest wektorem (2), a $f()$ jest dowolną funkcją. Gdy $f()$ stanowi liniową funkcję regresji zależną od wartości procesu w chwilach poprzednich, to modele te znane są pod nazwą wielowymiarowych modeli ciągów czasowych, lub modeli Boxa-Jenkinsa [5].

Drugi grupa modeli wyraża się poprzez funkcję rozkładu prawdopodobieństwa wektora skończenie wymiarowego, reprezentującego wartości procesu w określonych chwilach:

$$p(x) = p_m(x_{t_k}, x_{t_{k-1}}, \dots, x_{t_{k-m}}). \quad (4)$$

Ten typ modeli znany jest tylko, głównie ze względu na trudności obliczeniowe, dla procesów niskiego rzędu, zwanych także *procesami o krótkiej pamięci* [8]. Dla procesu bezwzględnie losowego, zwanego *procesem białego szumu*, rozkład zmiennej losowej m -wymiarowej jest równy iloczynowi rozkładów jednowymiarowych niezależnych zmiennych losowych:

$$p_m(x_{t_k}, x_{t_{k-1}}, \dots, x_{t_{k-m}}) = \prod_{i=0}^m p_1(x_{t_{k-i}}). \quad (5)$$

Dla procesów spełniających warunek Markowa, warunkowe prawdopodobieństwo procesu w chwili k -tej zależy tylko od wartości procesu w poprzedniej chwili, a nie zależy od historii procesu, tym samym rozkład m -wymiarowego wektora sprowadza się do rozkładu 2-wymiarowego:

$$p_m(x_{t_k} / x_{t_{k-1}}, \dots, x_{t_{k-m}}) = p_2(x_{t_k} / x_{t_{k-1}}). \quad (6)$$

Ogólnie jednak, modele typu (4), rzędu m , wymagają tworzenia rozkładów prawdopodobieństwa wielowymiarowych zmiennych losowych.

1.2 Reprezentacja rozmyta stochastycznego procesu Markowa

U podstaw modelowania rozmytego [16, 10] leży operacja rozmywania, czyli transformacji zmiennych opisujących badany obiekt w numerycznej przestrzeni rozważań do przestrzeni zbiorów rozmytych i wartości lingwistycznych [18]. Wektor wartości procesu stochastycznego $x \in X \subset R^p$ przekształcamy w zmienną lingwistyczną wielowymiarową, o zbiorach wartości lingwistycznych $L(X_1), L(X_2), \dots, L(X_p)$, przy czym $X_1 \times X_2 \times \dots \times X_p = X$. W przestrzeniach numerycznych $X_i, i=1,2,\dots,p$ poszczególnych zmiennych zdefiniowane zostają zbiory rozmyte $A_{j_i}^i, j_i=1,2,\dots,J_i$, odpowiadające wartościom lingwistycznym ze zbiorów $L(X_i)$.

Podstawowym zadaniem dla modelowania rozmytego procesów stochastycznych jest określenie rozkładów prawdopodobieństwa zmiennych lingwistycznych. Zadaniem budowy rozkładów prawdopodobieństwa zmiennych lingwistycznych, w oparciu o pojęcie prawdopodobieństwa zdarzenia rozmytego, zdefiniowane w [17] przez Zadeha, poświęcono, m. in. prace [12-14]. W pracach [12, 13] utworzono reprezentację rozmytą dyskretnego procesu Markowa, skalarne i wektorowe, odpowiednio. W pracy [14] zaproponowano postać ogólną reprezentacji rozmytej systemu stochastycznego MISO oraz procedury wnioskowania.

Model rozmyty wielowymiarowego procesu stochastycznego (1), spełniającego własność Markowa (6), może mieć postać zbioru reguł (w myśl [13]):

$$\begin{aligned}
 R_l: (w_l) \text{ JEŻELI } x_{t_{k-1}} \text{ jest } A_l \text{ TO } x_{t_k} \text{ jest } A_1 \quad (w_{1/l}) \\
 \text{TAKŻE } x_{t_k} \text{ jest } A_2 \quad (w_{2/l}) \\
 \text{-----} \\
 \text{TAKŻE } x_{t_k} \text{ jest } A_J \quad (w_{J/l}).
 \end{aligned}
 \tag{7}$$

W modelu tym zbiory rozmyte $A_1, A_2, \dots, A_l, \dots, A_J$ zostały określone w przestrzeni $X_1 \times X_2 \times \dots \times X_p = X$, a ich funkcje przynależności wyrażają się poprzez przyjętą t-normę i funkcje przynależności $\mu_{A_{i_j}}(x^i)$ zbiorów rozmytych określonych w poszczególnych przestrzeniach X_1, X_2, \dots, X_p :

$$\mu_{A_l}(x) = T(\mu_{A_{1_i}}(x^1), \dots, \mu_{A_{i_i}}(x^i), \dots, \mu_{A_{p_i}}(x^p)). \tag{8}$$

Waga reguły plikowej $w_l, l=1, \dots, L$ stanowi prawdopodobieństwo łącznego zdarzenia rozmytego zawartego w poprzedniku l -tej reguły, wyliczone zgodnie z definicją Zadeha [17], jako

$$w_l = P(x_{t_{k-1}} \text{ jest } A_l) = p(x_{t_{k-1}}) \mu_{A_l}(x_{t_{k-1}}), \tag{9}$$

przy czym $p(x)$ jest funkcją łącznego rozkładu prawdopodobieństwa, która przypisuje każdemu zbiorowi borelowskiemu w przestrzeni X liczbę rzeczywistą $p \in [0,1]$.

Dobór funkcji przynależności $\mu_{A_l}(x^i)$ i t-normy (8) jest decydujący dla spełnienia

warunku $\sum_{l=1}^L P(A_l) = 1$, niezbędnego, aby zbiór wag $w_l, l = 1, \dots, L$ stanowił rozkład

prawdopodobieństwa wielowymiarowej zmiennej lingwistycznej x .

W sposób analogiczny tworzy się łączny rozkład prawdopodobieństwa zmiennych lingwistycznych zawartych w poprzedniku i następniku modelu (7), jako prawdopodobieństwo zdarzeń rozmytych określonych w przestrzeni $X \times X$:

$$w_{jl} = P\left(\left\{ (x_{t_k} \text{ jest } A_j) \cap (x_{t_{k-1}} \text{ jest } A_l) \right\}, j = 1, 2, \dots, J; l = 1, \dots, L, \right. \quad (10)$$

$$\left. w_{jl} = p(x_{t_{k-1}}, x_{t_k}) T(\mu_{A_l}(x_{t_{k-1}}), \mu_{A_j}(x_{t_k})) \right). \quad (11)$$

W modelu (7) wagi reguł elementarnych $w_{j/l}, j = 1, \dots, J, l = \text{const}$, stanowią warunkowe prawdopodobieństwa wystąpienia zdarzeń w następniku reguły:

$$w_{j/l} = P\left(\left\{ (x_{t_k} \text{ jest } A_j) / (x_{t_{k-1}} \text{ jest } A_l) \right\}, j = 1, 2, \dots, J. \right. \quad (12)$$

Jako prawdopodobieństwa warunkowe obliczane są w sposób następujący:

$$w_{j/l} = \frac{w_{jl}}{w_l}, j = 1, \dots, J, l = \text{const}. \quad (13)$$

Wyznaczone prawdopodobieństwa (wagi) powinny spełniać warunki normalizacji

$$\sum_{l=1}^L w_l = 1, \quad \sum_{l=1}^L \sum_{j=1}^J w_{jl} = 1, \quad \sum_{j=1}^J w_{j/l} = 1, l = \text{const}. \quad (14)$$

Warunki te są istotne w operacjach wnioskowania, co przedstawiono w [14].

1.3 Reprezentacja rozmyta procesu stochastycznego z długą pamięcią

Dla procesu stochastycznego x , przy uwzględnieniu m jego opóźnionych wartości, jak w opisie (4), reprezentację rozmytą może stanowić zbiór reguł, z wagami:

$$R_{ij} : (w_{ij}) \left\{ \begin{array}{l} \text{JEŻELI } [x(t_{k-1}) \text{ jest } A_{i,k-1} \text{ I } x(t_{k-2}) \text{ jest } A_{i,k-2} \text{ I } \dots \\ \text{I } x(t_{k-m}) \text{ jest } A_{i,k-m}] \\ \text{TO } [x(t_k) \text{ jest } A_{j,k}] \end{array} \right\}, \quad (15)$$

gdzie

$x(t_k), x(t_{k-1}), \dots, x(t_{k-m})$ – skalarne zmienne lingwistyczne, $x \in X \subset R$, przyjmujące wartości lingwistyczne ze zbioru $L(X)$, określające rozmyte stany procesu;

Budowa rozmytych modeli procesów stochastycznych przy użyciu reguł asocjacji

$A_{j,k}, A_{i,k-1}, \dots, A_{i,k-m}$ – podzbiory rozmyte, odpowiadające lingwistycznym wartościom zmiennych $x(t_k), x(t_{k-1}), \dots, x(t_{k-m})$, $x \in X$;

w_{ij} – waga reguły elementarnej, wyrażająca łączne prawdopodobieństwo rozmytego zdarzenia jednoczesnego poprzednika i następnika [14]:

$$w_{ij} = p[x(t_k), x(t_{k-1}), \dots, x(t_{k-m})] T[\mu_{A_{j,k}}(x_{t_k}), \dots, \mu_{A_{i,k-m}}(x_{t_{k-m}})], \quad (16)$$

przy czym $p()$ jest funkcją łącznego rozkładu prawdopodobieństwa, która przypisuje każdemu zbiorowi borelowskiemu w przestrzeni X^{m+1} liczbę rzeczywistą $p \in [0,1]$.

Dobór t-normy oraz sposób podziału przestrzeni X na zbiory rozmyte odniesienia powinien zapewnić spełnienie warunku (14). Np. w pracach [12, 13] jako t-normę przyjęto iloczyn, a zbiory rozmyte A_i zostały określone poprzez funkcje przynależności o stałej wartości w rozłącznych przedziałach a_n przestrzeni X w taki sposób, aby dla każdego $x \in a_n$ był spełniony warunek

$$\sum_{i=1}^I \mu_{A_i}(a_n) = 1, n = 1, \dots, N. \quad (17)$$

Model (15) można także zapisać w postaci reguł plikowych, łączących wspólny poprzednik, jak np. w modelu (7). Wówczas odpowiednie wagi w_i , $w_{j/i}$ wyznacza się, jako prawdopodobieństwa brzegowe i warunkowe, w sposób następujący:

$$w_i = \sum_{j=1, \dots, J} w_{ij}, i = 1, \dots, I, \quad w_{j/i} = \frac{w_{ij}}{w_i}, j = 1, \dots, J. \quad (18)$$

2 Zastosowanie reguł asocjacji w rozmytym modelowaniu systemu stochastycznego

2.1 Reguły asocjacji

Reguły asocjacji (ang. *association rules*), jako jedna z technik *data mining*, są metodą odnajdywania współwystępowania wartości zmiennych w obszernych zbiorach danych [1]. Formalny zapis reguł asocjacji ma postać implikacji:

$$A \Rightarrow B \quad (\pi_s \%, \pi_c \%), \quad (19)$$

gdzie

A, B – poprzednik i następnik, wyrażające związki, np. o postaci $x=a, y=b$ w rozłącznych zbiorach wartości zmiennych;

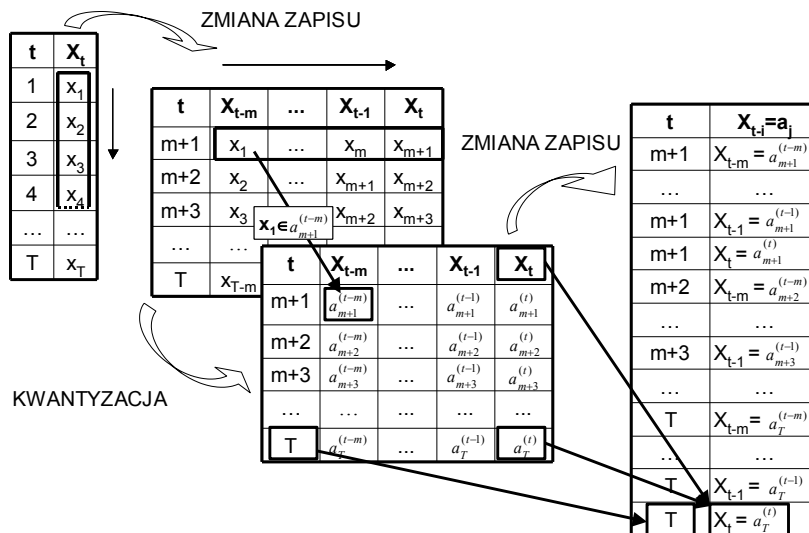
π_s – wsparcie (ang. *support*), jest wyrażonym procentowo prawdopodobieństwem jednoczesnego występowania związków w poprzedniku i następniku $P((x=a) \cap (y=b))$;

π_c – ufność (ang. *confidence*), zwana także wiarygodnością, jest wyrażonym procentowo prawdopodobieństwem warunkowym następnika $P((y = b)/(x = a))$.

Znalezienie i przeliczenie wszystkich możliwych kombinacji zmiennych i ich wartości w bardzo dużych zbiorach danych, okazuje się nieefektywne czasowo. Podstawowy i najczęściej implementowany algorytm Apriori odkrywania reguł asocjacji – korzystając z monotoniczności miary wsparcia – pozwala na przyspieszenie obliczeń [1, 2]. Późniejsze prace nad poprawieniem efektywności doprowadziły do powstania modyfikacji tego algorytmu (AprioriTid, AprioriHybrid), oraz innych algorytmów (m.in. SETM, FP-Growth, FreeSpan, Eclat). W pracach [6, 9] zostały zaproponowane także algorytmy pozwalające na odkrywanie rozmytych reguł asocjacji.

2.2 Wyznaczanie reguł asocjacji dla ciągu czasowego w Oracle Data Mining

Wyznaczanie reguł asocjacji ciągu czasowego z użyciem Oracle Data Mining i aplikacji Data Miner wymaga odpowiedniego zapisu wyników badań empirycznych. Przykład transformacji danych ciągu czasowego jednej zmiennej, z uwzględnieniem m poprzednich chwil czasowych, jest przedstawiony na rys. 1.



Rys. 1. Etapy transformacji danych empirycznych ciągu czasowego

Wartości zmiennych w wielowymiarowych regułach asocjacji stanowią dane ilościowe bądź kategoriyczne, dlatego należy podzielić przestrzeń X wartości zmiennych $X_{t-m}, \dots, X_t \in X$ na rozłączne zbiory wartości a_1, a_2, \dots, a_N . Wówczas każda wartość x_1, \dots, x_T zmiennych X_{t-m}, \dots, X_t zostaje zastąpiona identyfikatorem zbioru $a_{m+1}^{(t-m)}, \dots, a_T^{(t)} \in \{a_1, a_2, \dots, a_N\}$, do którego należy.

2.3 Budowanie rozmytego modelu systemu dynamicznego

W wyniku analizy wyjściowej tabeli transformacji tzw. tabeli transakcyjnej (rys.1.) i użycia algorytmu Apriori, otrzymujemy szereg reguł asocjacji, z których wybieramy reguły o postaci [4]:

$$\{X_{t-m} = a_n^{(t-m)}, \dots, X_{t-1} = a_n^{(t-1)}\} \Rightarrow \{X_t = a_n^{(t)}\} (\pi_s \%, \pi_c \%), \quad (20)$$

gdzie

$X_{t-m}, \dots, X_{t-1}, X_t$ – zmienne w chwilach $t-m, \dots, t-1, t$,

$a_n^{(t-m)}$ – n -ty zbiór wartości zmiennej X_{t-m} w przestrzeni $X, n \in \{1, \dots, N\}$;

π_s, π_c – wsparcie i ufność, określające empiryczne rozkłady prawdopodobieństwa zajścia zdarzeń nierozmytych, odpowiednio, jak w (19).

Rozmyty model systemu stochastycznego (15) operuje na zmiennych lingwistycznych, zbiorach rozmytych i miarach probabilistycznych zbiorów rozmytych. Miara π_s , reprezentująca prawdopodobieństwo łączne zdarzeń rozmytych poprzednika i następnika, została wyznaczona z uwzględnieniem zależności (16)-(17) oraz t -normy typu *product*, w sposób następujący [4]:

$$\begin{aligned} \pi_{s_{k_{(t-m)} \dots k_{(t)}}} &= P(A_k^{(t-m)} \cap \dots \cap A_k^{(t)}) = \\ &= \sum_{n_{(t-m)}=1, \dots, N} \dots \sum_{n_{(t)}=1, \dots, N} \pi_{s_{n_{(t-m)} \dots n_{(t)}}} \mu_{A_k^{(t-m)}}(a_{n_{(t-m)}}^{(t-m)}) \dots \mu_{A_k^{(t)}}(a_{n_{(t)}}^{(t)}), \end{aligned} \quad (21)$$

gdzie

$a_{n_{(t-m)}}^{(t-m)}$ oznacza $n_{(t-m)}$ -ty zbiór wartości w przestrzeni X zmiennej X_{t-m} ;

$A_k^{(t-m)}$ – zdarzenie rozmyte polegające na tym, że $X_{t-m} \in A_k^{(t-m)}$;

k – indeks wartości lingwistycznej ze zbioru $L(X), k_{(t-m)}, \dots, k_{(t)}, k = 1, 2, \dots, K$;

$\mu_{A_k^{(t-m)}}(a_{n_{(t-m)}}^{(t-m)})$ – funkcja przynależności, stała wartość z przedziału $[0,1]$ określająca stopień przynależności zbioru $a_{n_{(t-m)}}^{(t-m)}$ do zbioru rozmytego $A_k^{(t-m)}$.

Zachodzi zależność:

$$\sum_{k_{(t-m)}=1}^K \dots \sum_{k_{(t)}=1}^K \pi_{s_{k_{(t-m)} \dots k_{(t)}}} = 1. \quad (22)$$

Miara ufności π_c , jako prawdopodobieństwo warunkowe zdarzeń w następniku, w myśl (18), wynosi:

$$\begin{aligned} \pi_{c_{k_{(t-m)} \dots k_{(t)}}} &= P(A_k^{(t)} / A_k^{(t-m)} \cap \dots \cap A_k^{(t-1)}) = \\ &= P(A_k^{(t-m)} \cap \dots \cap A_k^{(t)}) / P(A_k^{(t-m)} \cap \dots \cap A_k^{(t-1)}). \end{aligned} \quad (23)$$

Wyznaczone miary probabilistyczne (21) i (23) odpowiadające wagom w_{ij} (16) oraz $w_{j/i}$ (18), zostaną pokazane w badaniach empirycznych.

3 Badania empiryczne

Elektrownie zawodowe w Polsce, w których produkcja energii elektrycznej bazuje na węglu kamiennym, wytworzyły w roku 2005 aż 57,27% energii elektrycznej, przy ogólnej ilości wyprodukowanej energii elektrycznej 156,933 TWh [15].

Tak wielka ilość zużywanego węgla dla celów energetycznych nasuwa pytania o sprawność wykorzystania tego paliwa przez elektrownie i skutki uboczne produkcji dla środowiska naturalnego.

Wartości parametrów węgla energetycznego oraz metodykę ich pomiaru określają stosowne normy. Jednakże, pomiar parametrów węgla nie jest sprawą trywialną, ze względu na niejednorodność materiału oraz czynnik losowy wpływający na wybór próbki poddawanej pomiarowi [11]. Zmienność parametrów węgla, jako materiału surowego dla optymalnego sterowania procesami przeróbki, była przedmiotem analiz w wielu pracach. W pracy [7] autorka przewiduje m.in. ilość i jakość możliwych produktów handlowych uzyskanych z węgla surowego oraz ich wartość, posługując się tzw. formułą sprzedaży [3], zastosowaną do zmierzonych wcześniej parametrów węgla surowego.

Zmienność parametrów węgla, będącego paliwem dla elektrowni, może stanowić zakłócenie dla procesu produkcji energii elektrycznej. Dla prawidłowego przebiegu produkcji energii elektrycznej, istotna jest znajomość zmian charakterystyki materiału surowego w funkcji czasu i możliwość przewidywania jej zmienności dla planowania produkcji, kosztów i szacowania opłacalności.

W referacie bada się zmienność wybranych, dwóch ilościowo-jakościowych parametrów węgla: udziału frakcji ziaren lekkich x oraz zawartości popiołu w tej frakcji y . Dane numeryczne, zaczerpnięte z prac [11] i [7] pozwolą na utworzenie rozmytych modeli bazy wiedzy o zmienności cech badanego paliwa.

Dokonano podziału przestrzeni rozważań $X \subset R, Y \subset R$, na rozłączne przedziały oznaczone odpowiednio $a_n, b_n, n=1,2,\dots,10$ (tab.1).

Tabela 1. Rozłączne przedziały zmienności w przestrzeniach X i Y

	a_1	a_2	a_3	...	a_8	a_9	a_{10}
X	(0; 0.08)	[0.08; 0.16)	[0.16; 0.24)	...	[0.56; 0.64)	[0.64; 0.72)	[0.72; 1)
	b_1	b_2	b_3	...	b_8	b_9	b_{10}
Y	(0; 3.62)	[3.62; 4.95)	[4.95; 6.29)	...	[11.63; 12.97)	[12.97; 14.31)	[14.31; 20)

Zdefiniowano zbiory rozmyte A_k oraz B_k ($k=1,2,\dots,5$), odpowiadające wartościom lingwistycznym ze zbiorów:

$$L(X) = L(Y) = \{\textit{bardzo niski}, \textit{niski}, \textit{średni}, \textit{wysoki}, \textit{bardzo wysoki}\}, \quad (24)$$

w skrócie: $\{BN, N, S, W, BW\}$. Wartości stałych współczynników przynależności, z jakimi zbiory a_n należą do zbiorów rozmytych A_k , zostały przedstawione w tab. 2. Zbiory rozmyte B_k definiowane są analogicznie.

Budowa rozmytych modeli procesów stochastycznych przy użyciu reguł asocjacji

Tabela 2. Wartości funkcji przynależności $\mu_{A_k}(a_n)$ w przestrzeni X

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
$\mu_{A_1}(a_n)$	0,9	0,5	0,2	0,1	0	0	0	0	0	0
$\mu_{A_2}(a_n)$	0,1	0,5	0,7	0,5	0,1	0	0	0	0	0
$\mu_{A_3}(a_n)$	0	0	0,1	0,4	0,9	0,9	0,4	0,1	0	0
$\mu_{A_4}(a_n)$	0	0	0	0	0	0,1	0,5	0,7	0,5	0,1
$\mu_{A_5}(a_n)$	0	0	0	0	0	0	0,1	0,2	0,5	0,9

Przeprowadzona została eksploracja danych empirycznych zmiennej x w środowisku Oracle Data Mining, z wykorzystaniem algorytmu Apriori. Otrzymano $5^4 = 625$ reguł plikowych uwzględniających $m=4$ poprzednich chwil czasowych ciągu wartości. Jedynie 260 uzyskanych reguł plikowych posiada niezerowe wagi w_i, w_{ji} . Przykładową regułą opisanego modelu procesu stochastycznego przedstawia (25):

$$\begin{aligned}
 P1: & (0,1021) \{ \text{JEŻELI } x_{t-4} \text{ jest "wysoki" I } x_{t-3} \text{ jest "wysoki" I } x_{t-2} \text{ jest "wysoki" I } x_{t-1} \text{ jest "wysoki"} \\
 & \text{TO } x_t \text{ jest "wysoki"} (0,5717) \\
 & \text{TAKŻE } x_t \text{ jest "bardzo wysoki"} (0,3328) \\
 & \text{TAKŻE } x_t \text{ jest "średni"} (0,0930) \\
 & \text{TAKŻE } x_t \text{ jest "bardzo niski"} (0,0022) \\
 & \text{TAKŻE } x_t \text{ jest "niski"} (0,0003) \}.
 \end{aligned} \tag{25}$$

Ilość reguł maleje wraz ze wzrostem ograniczeń minimalnego wsparcia π_s i ufności π_c w algorytmie odkrywającym reguły asocjacji. Przy założeniach minimalnego wsparcia $\pi_{s \min} = 1\%$ oraz minimalnej ufności $\pi_{c \min} = 5\%$, w tym samym modelu, utworzonych zostało łącznie 80 reguł plikowych z niezerowymi wartościami wag w_i, w_{ji} .

Postępując analogicznie, dla zmiennych x i y , przy uwzględnieniu dwóch wcześniejszych próbkowań obu zmiennych, otrzymano rozmyty model systemu dynamicznego zapisany w postaci 338 reguł plikowych (tab. 3.).

Tabela 3. Przykłady najważniejszych reguł plikowych rozmytego modelu zmiennych x i y dla $m=2, \pi_{s \min}=0\%, \pi_{c \min}=0\%$

Nr reg. plikowej	Waga w_i	Jeżeli	To	Waga w_{ji}
R1	0,0449	x_{t-2} jest <i>BW I</i> y_{t-2} jest <i>BNI</i> x_{t-1} jest <i>BW I</i> y_{t-1} jest <i>BN</i>	y_t jest <i>BN</i>	0,49435
			y_t jest <i>N</i>	0,40302
			y_t jest <i>S</i>	0,08475
			y_t jest <i>W</i>	0,01486
			y_t jest <i>BW</i>	0,00302
			y_t jest <i>BW</i>	0,00302
R2	0,0449	x_{t-2} jest <i>W I</i> y_{t-2} jest <i>BNI</i> x_{t-1} jest <i>W I</i> y_{t-1} jest <i>BN</i>	y_t jest <i>BN</i>	0,4758
			y_t jest <i>N</i>	0,4145
			y_t jest <i>S</i>	0,0991
			y_t jest <i>W</i>	0,0094
			y_t jest <i>W</i>	0,0094
			y_t jest <i>BW</i>	0,0012

4 Wnioski

Proponowana metoda budowy rozmytych modeli znalazła zastosowanie w opisie probabilistycznych i stochastycznych systemów rzeczywistych z czasem dyskretnym. Użycie w modelowaniu reguł asocjacji obniża złożoność obliczeniową związaną z wyznaczeniem prawdopodobieństw zdarzeń rozmytych procesów wyższych rzędów.

Literatura

- [1] Agrawal R., Imielinski T., Swami A. (1993) Mining Association Rules Between Sets of Items in Large Databases, in *Proc. ACM Sigmod Intern. Conf. on Management of Data*, Washington D.C., 207- 216.
- [2] Agrawal R., Srikant R. (1994) Fast Algorithms for Mining Association Rules, in *Proc. of 1994 Intern. Conf. on Very Large Databases VLDB*, Santiago de Chile, 487– 499.
- [3] Blaschke W. (2002) *Propozycja nowej formuły sprzedażnej węgla energetycznego przeznaczonego dla energetyki zawodowej*, Studia, rozprawy, monografie, Nr 112. Wydawnictwo IGSMiE PAN, Kraków.
- [4] Błaszczuk K. (2007) Reguły asocjacji dla rozmytego modelowania szeregów czasowych, in *Proc. IX Internat. PhD Workshop OWD 2007*, 2, 305-310.
- [5] Box G.E.P., Jenkins G.M. (1983) *Analiza szeregów czasowych. Prognozowanie i sterowanie*, PWN, Warszawa.
- [6] Chen G., Wei Q. (2002) Fuzzy Association Rules and the Extended Mining Algorithm, *Information Sciences*, 147, 201– 228.
- [7] Gornig A. (2005) *Zastosowanie metod klasyfikacji danych w analizie jakości węgla*, Praca doktorska, Wydział Górniczo-Geologii, Politechnika Śląska, Gliwice.
- [8] Van Kampen N.G. (1990) *Procesy stochastyczne w fizyce i chemii*, PWN, Warszawa.
- [9] Kuok C.M., Fu A., Wong M.H. (1998) Mining Fuzzy Association Rules in Database, *Sigmod*, 27, 1, 41– 46.
- [10] Rutkowski L. (2005) *Metody i techniki sztucznej inteligencji*, Wydawnictwo Naukowe PWN, Warszawa.
- [11] Walaszek-Babiszewska A. (2004) *Statistical and Fuzzy Modelling of Grain Material Sampling and Operations. Selected Approaches*, Monografia, Wydawnictwo Politechniki Śląskiej, Gliwice.
- [12] Walaszek-Babiszewska A. (2005) Measurements and Expert Knowledge for Time-dependent Stochastic Systems, in *Proc. Fifth Int. Conf. on Intelligent Processing and Manufacturing of Materials IPMM*, Monterey, California, CD.
- [13] Walaszek-Babiszewska A. (2006) IF-THEN Linguistic Fuzzy Model of a Discrete Stochastic System, in *Artificial Intelligence and Soft Computing, Challenging Problems of Sciences, Computer Sciences*, Cader A., Rutkowski L., Tadeusiewicz R., and Żurada J. (Eds.), Academic Publishing House EXIT, Warsaw, 169–174.
- [14] Walaszek-Babiszewska A. (2007) Construction of Fuzzy Models Using Probability Measures of Fuzzy Events, in *Proc. 13th IEEE Internat. Conf. on Methods and Models in Automation and Robotics, MMAR 2007*, Szczecin, Poland, 661-666.
- [15] WWW.cire.pl/rynekenergii/
- [16] Yager R., Filev D. (1995) *Podstawy modelowania i sterowania rozmytego*, WNT, Warszawa.
- [17] Zadeh L.A. (1968), Probability Measures of Fuzzy Events, *J. Math. Anal. Appl.*, 23, 421-427.
- [18] Zadeh L.A. (1975) The Concept of a Linguistic Variable and its Application to Approximate Reasoning, *Information Sciences*, Part I: 8, 199-240.