

REGUŁY ASOCJACJI DLA ROZMYTEGO MODELOWANIA SZEREGÓW CZASOWYCH

THE ASSOCIATION RULES FOR FUZZY MODELLING OF THE TIME SERIES

Katarzyna Błaszczyk, *Politechnika Opolska*

Abstract

The article deals with the application of association rules to fuzzy modeling of the time series. To this end, has been used implemented in Oracle Date Mining algorithm Apriori (3), which determines crisp rules in the form (1, 2) and the form of time series (4). To the antecedents and consequents of the rules, has been included fuzzy events with linguistic variables. The measures of support and confidence as a probabilities of fuzzy events, in a sense of Zadeh's definition (12), have been calculated in accordance with methodology (13, 14) (Walaszek-Babiszewska, 2006), which apply constant values of crisp probability and values of function membership in the separate range. On the basis of the rules (15), we can construction fuzzy model of dynamic system (16) with weights as a marginal and conditional probabilities, which have been a result from fuzzy support and confidence. The empirical example has been carried out through analyze of the hard coal.

Streszczenie

Artykuł prezentuje wykorzystanie reguł asocjacji do modelowania rozmytych systemów dynamicznych. W tym celu użyto zaimplementowanego w Oracle Date Mining algorytmu Apriori, który oblicza reguły w sensie deterministycznym. Następnie, poprzez włączenie logiki rozmytej i wprowadzenie języka quasi-naturalnego, utworzono rozmyte reguły asocjacji i zastosowano je do budowy modelu systemu dynamicznego opartego na regułach, z wagami stanowiącymi prawdopodobieństwo zdarzeń rozmytych.

1. WPROWADZENIE

Analiza szeregów czasowych, w podejściu klasycznym, została wprowadzona przez Boxa i Jenkinsa [5]. Opiera się ona na założeniu, iż kolejne

wartości w zbiorze danych reprezentują sekwencje pomiarów wykonanych w różnych odstępach czasu. Takie opisanie zjawisk rzeczywistych pozwala na wykrywanie ich natury oraz wnioskowanie o przyszłych wartościach, co stanowi nieodzowny element zagadnień podejmowania decyzji, prognozy i sterowania, zarówno o charakterze deterministycznym, jak i rozmytym. Badanie powyższych zagadnień wymaga ścisłego zidentyfikowania i opisanie elementów szeregu czasowego w postaci modelu systemu dynamicznego. Teoria zbiorów rozmytych, stanowiąca podstawę teorii systemów rozmytych [14] wykorzystuje szeregi czasowe do budowy rozmytych modeli systemów dynamicznych. Zaawansowanym narzędziem reprezentacji wiedzy jakościowej stanowi lingwistyczny model regułowy rozmytego systemu dynamicznego [11]. Ideą tego narzędzia jest zdefiniowanie wiarygodności - wag reguł, które stanowią łączne i warunkowe prawdopodobieństwa zdarzeń rozmytych. Problem identyfikacji modeli rozmytych i uczenia wag reguł wprowadzili po raz pierwszy Tong i Kosko. Badania w tym zakresie kontynuowali m.in. Sanchez, Takagi i Sugeno [13].

W artykule zostanie pokazana możliwość wykorzystania jednej z metod data mining – reguł asocjacji do budowy rozmytego modelu systemu dynamicznego opartego na regułach. W tym celu zostanie użyty algorytm Apriori, zaimplementowany w Oracle Data Mining.

2. REGUŁY ASOCJACJI

Zagadnienie reguł asocjacji (ang. association rules) zostało po raz pierwszy użyte w pracy [1], jako metoda odnajdywania współwystępowania wartości atrybutów w obszernych zbiorach danych [3]. W ujęciu formalnym reguły asocjacji mają postać implikacji:

$$X \Rightarrow Y (sup\%, conf\%), \quad (1)$$

gdzie X i Y są rozłącznymi zbiorami atrybutów, nazwanych często: X - zbiorem wartości warunkujących, Y - zbiorem wartości warunkowanych. A dokładniej:

$$A_1 \cap \dots \cap A_n \Rightarrow A_{n+1} \cap \dots \cap A_m (sup\%, conf\%), \quad (2)$$

gdzie $A_1 - A_n$ są parami atrybut - wartość poprzednika reguły (ang. body, antecedent), $A_{n+1} - A_m$ parami atrybut - wartość następnika reguły (ang. head, consequent).

Każda reguła asocjacji jest związana z dwiema miarami statystycznymi określającymi ważność i siłę reguły: support (*sup%*) - wsparcie, prawdopodobieństwo jednoczesnego występowania zbioru $A_1 \cap \dots \cap A_n \cap A_{n+1} \cap \dots \cap A_m$ w kolekcji zbiorów oraz confidence (*conf%*) - ufność, zwane również wiarygodnością, będące prawdopodobieństwem warunkowym.

Pierwsze zastosowanie reguł asocjacji miało miejsce w analizie koszyka zakupów (ang. basket analysis). Jednakże biorąc pod uwagę, iż w regułach mogą występować atrybuty pochodzące z różnych dziedzin wartości, wachlarz zastosowań metody poszerza się znacząco w płaszczyźnie podejmowania decyzji, sterowania i prognozowania, dając także możliwość zastosowania ich przy modelowaniu szeregów czasowych.

2.1 Algorytmy odkrywania reguł asocjacji

Efektywne znalezienie i przeliczenie wszystkich dostępnych kombinacji atrybutów jest w większości przypadków nie możliwe ze względu na ogromną liczbę danych. Za podstawowy algorytm odkrywania reguł asocjacji uznaje się iteracyjny algorytm Apriori [1, 2]. Algorytm ten dzieli problem wyznaczania reguł na dwa zagadnienia. Po pierwsze należy wyodrębnić k -elementowe ($k=1,2,\dots$) zbiory, dla których wsparcie jest większe od zakładanego (zbiory częste). Wykorzystując właściwość monotoniczności miary wsparcia, k -elementowe zbiory częste powstają przez łączenie $(k-1)$ -elementowych zbiorów częstych. Algorytm można przedstawić w postaci pseudo kodu [2, 3]:

```

L1 = {zbiory częste 1-elementowe};
for (k=2; Lk-1 ≠ ∅; k++) do
begin
//generowanie nowych zbiorów kandydujących
insert into Ck select p.i1, p.i2, ..., p.ik-1, q.ik-1
from Lk-1 p, Lk-1 q where p.i1 = q.i1 and ...
and p.ik-2 = q.ik-2 and p.ik-1 = q.ik-1;
forall zbiorów c ∈ Ck do
forall (k-1)-elem. podzbiorów s ⊂ c do
if (s ∉ Lk-1) then delete c form Ck;
//generowanie zbiorów częstych
Lk = {X ∈ Ck: support(X) ≥ min_sup};
end

```

$$\text{Wynik} = \bigcup_2^k L_k; \quad (3)$$

gdzie C_k - rodzina zbiorów kandydujących k -elementowych, L_k - rodzina zbiorów częstych k -elementowych, *min_sup* - minimalne wsparcie.

Po drugie, ze zbiorów częstych należy wygenerować reguły asocjacyjne, które spełniają wymaganie minimalnej ufności.

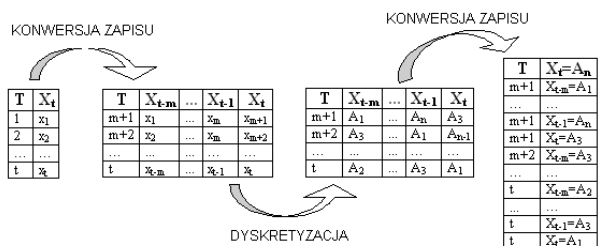
Algorytm Apriori doczekał się wielu modyfikacji [2, 17], które zmierzają do poprawienia jego efektywności (AprioriTid uaktualniający przy każdym kroku strukturę danych, AprioriHybrid jako połączenie Apriori i AprioriTid). Ponadto powstały inne algorytmy, do których należą FP-Growth [4], FreeSpan [7], Eclat [16] czy Partition [10], a także algorytmy pozwalające na odkrycie rozmytych reguł asocjacji [6, 8, 9]. Rozmyte reguły asocjacji pozwoliły na zastosowanie zbiorów rozmytych [14] i zmiennych lingwistycznych w regułach. Miara wsparcia natomiast opierała się na wprowadzeniu operatora t-norm.

3. WYZNACZANIE REGUŁ ASOCJACJI SZEREGU CZASOWEGO

Narzędzie Oracle Data Mining wraz z aplikacją Data Miner wspomaga kolejne etapy wykrywania zależności danych od analizy problemu, przygotowania danych, po eksploatację, wizualizację i weryfikowanie wyników. Jednakże przystosowane jest głównie do analizy pojedynczych reguł asocjacji. Wobec czego, nie wszystkie etapy przygotowania danych można zautomatyzować poprzez opracowane, gotowe procedury.

Szereg czasowy dla modelu SISO, jako ciąg wyników obserwacji uporządkowanych w czasie, można oznaczyć poprzez $\{t, x_t\}$, przy czym t oznacza numery kolejnych jednostek czasu, a x_t wielkość badanej cechy (zjawiska) w momencie t . Zapisane w ten sposób wyniki badań empirycznych należy przekształcić do odpowiedniej formy bazy danych, tak, aby każdy rekord odpowiadał zbiorowi $m+1$ kolejnych wartości szeregu czasowego dla danej jednostki czasu t , gdzie m świadczy o rzędzie autoregresji (AR(m)). Ponieważ wartości cech (atrybutów) są danymi ilościowymi bądź kategoriowymi, należy je poddać dyskretyzacji.

Ostatnim etapem przygotowania danych jest transformacja danych do postaci tabeli „transakcyjnej”, gdzie poszczególne rekordy zawierają informację o przynależności wartości cechy x_t do nierozmytego zbioru $(A_1 \dots A_n)$, w czasie t , w którym wystąpiła dana wartość empiryczna. Etapy transformacji danych przedstawia rys. 1. Wynikowy zapis danych może być wejściową tabelą do analizy reguł asocjacji w Oracle Data Miner. W przypadku modeli MIMO postępowanie jest analogiczne.



Rys. 1. Etapy transformacji danych.
Fig.1. The stages of transformation data.

4. BUDOWANIE ROZMYTEGO MODELU SYSTEMU DYNAMICZNEGO

W wyniku zastosowania algorytmu Apriori otrzymujemy szereg „nielingwistycznych” reguł asocjacji, z których reguły typu:

$$\{x_{t-m} = A_k^{(t-m)}, \dots, x_{t-1} = A_k^{(t-1)}\} \Rightarrow \{x_t = A_k^{(t)}\} (\text{sup } \%, \text{conf } \%) \quad (4)$$

można zastosować do regułowego modelowania systemu dynamicznego jednej zmiennej.

Stosując podejście lingwistyczne [15] wprowadzamy do modelu zmienne, będące zbiorami rozmytymi lub terminami (wartościami) lingwistycznymi zmiennych, zdefiniowanych jako szablon:

$$\langle x_{nazwa}, L(x), X, M_x \rangle, \quad (5)$$

gdzie x_{nazwa} - nazwa zmiennej lingwistycznej (np. wzrost, temperatura), $L(x)$ - zbiór wartości (terminów) lingwistycznych, jakie przyjmuje x (np. wysoki, średni, niski), X - dziedyna rozważań, M_x - funkcja semantyczna, przyporządkowująca każdej wartości lingwistycznej ze zbioru $L(x)$ zbiór rozmyty zdefiniowany nad X .

Powyższy zbiór rozmyty, dla danych empirycznych dyskretnych, określamy jako:

$$\hat{A}_j = \sum_X \mu_{\hat{A}_j}(x) / x, \quad (6)$$

gdzie $\mu_{\hat{A}_j}(x)$ oznacza funkcję przynależności definiowaną na X tzn.

$$\forall x \in X : \mu_{\hat{A}_j}(x) : X \rightarrow [0,1], \quad j = 1, \dots, J \quad (7)$$

przy czym

$$\sum_{j=1}^J \mu_{\hat{A}_{j_i}}(x) = 1. \quad (8)$$

W przypadku rozmytych reguł asocjacji, w poprzedniku i następniku reguły mamy do czynienia ze zbiorami rozmytymi \hat{A}_j traktowanymi jako zdarzenie rozmyte. Reguły (4), poprzez wsparcie i ufność, określają łączne oraz warunkowe rozkłady empiryczne prawdopodobieństwa dla zdarzeń nierozmytych A_k (wartości atrybutów), gdzie:

$$P(A_k^{(t-m)} \cap \dots \cap A_k^{(t)}) = \text{sup}_{k_{(1)} \dots k_{(m+1)}}, \quad (9)$$

$k_{(1)}; k_{(m+1)} = 1, 2, \dots, K$. Rozkład można ująć w tablicy $K \times \dots \times K$. Oczywiście zachodzi zależność:

$$\sum_{k_{(1)}=1}^K \dots \sum_{k_{(m+1)}=1}^K \text{sup}_{k_{(1)} \dots k_{(m+1)}} = 1. \quad (10)$$

Rozkład warunkowy określony jest przez miarę ufności następująco:

$$P(A_k^{(t)} / A_k^{(t-m)} \cap \dots \cap A_k^{(t-1)}) = P(A_k^{(t-m)} \cap \dots \cap A_k^{(t)}) / P(A_k^{(t-m)} \cap \dots \cap A_k^{(t-1)}) = \text{conf}_{k_{(1)} \dots k_{(m+1)}}. \quad (11)$$

W pracach [11, 13] można znaleźć metodykę wyznaczania rozkładów prawdopodobieństwa zmiennej lingwistycznej w oparciu o empiryczne rozkłady prawdopodobieństwa danych numerycznych [11]. Korzystając z nierozmytego prawdopodobieństwa zdarzenia rozmytego \hat{A}_j , $j=1, 2, \dots, J$ zdefiniowanego wg. Zadeha jako:

$$P(\hat{A}_j) = \sum_{k=1, \dots, K} p_k(A_k) \mu_{\hat{A}_j}(A_k) \quad (12)$$

w myśl metodyki [11], dochodzimy do prawdopodobieństwa zdarzenia rozmytego $\hat{A}_j^{(t-m)} \cap \dots \cap \hat{A}_j^{(t)}$. Zdarzenia $\hat{A}_j^{(t-m)}, \dots, \hat{A}_j^{(t)}$, jako wartości szeregu czasowego, są utożsamiane z wartościami lingwistycznymi tej samej zmiennej lingwistycznej x_{nazwa} w dziedynie rozważań X . Niech zmiennej lingwistycznej odpowiada zbiór wartości lingwistycznych $L(X)$. Wówczas prawdopodobieństwo powyższego zdarzenia rozmytego określonego w zbiorze $L(X) \times \dots \times L(X)$ zmiennej lingwistycznej będzie obliczane następująco (por.[11]):

$$P(\hat{A}_j^{(t-m)} \cap \dots \cap \hat{A}_j^{(t)}) = \sum_{k_{(1)}=1, \dots, K} \dots \sum_{k_{(m+1)}=1, \dots, K} p_{k_{(1)} \dots k_{(m+1)}} (A_{k_{(1)}}^{(t-m)} \cap \dots \cap A_{k_{(m+1)}}^{(t)}) \mu_{\hat{A}_j^{(t-m)}}(A_{k_{(1)}}^{(t-m)}) \dots \mu_{\hat{A}_j^{(t)}}(A_{k_{(m+1)}}^{(t)}), \quad (13)$$

gdzie $j=1, 2, \dots, J$. Wzór (13) określa wparcie (*sup*) dla rozmytych reguł asocjacji. Miara ufności (*conf*) tych reguł wyznaczana jest natomiast na podstawie rozkładów warunkowych zdarzeń rozmytych:

$$P(\hat{A}_j^{(t)} / \hat{A}_j^{(t-m)} \cap \dots \cap \hat{A}_j^{(t-1)}) =$$

$$P(\hat{A}_j^{(t-m)} \cap \dots \cap \hat{A}_j^{(t)}) / P(\hat{A}_j^{(t-m)} \cap \dots \cap \hat{A}_j^{(t-1)}), \quad (14)$$

przy czym suma tych prawdopodobieństw dla wszystkich możliwych rozkładów $\hat{A}_j^{(t)}$, gdzie $j=1, 2, \dots, J$ oznacza kolejne wartości lingwistyczne, sumuje się do jedynki. W ten sposób otrzymujemy lingwistyczne reguły asocjacji w postaci:

$$\{x_{t-m} \text{ jest } \hat{A}_j^{(t-m)}, \dots, x_{t-1} \text{ jest } \hat{A}_j^{(t-1)}\} \Rightarrow \{x_t \text{ jest } \hat{A}_j^{(t)}\} (\text{sup}, \text{conf}), \quad (15)$$

gdzie oznaczenie $\{x_t \text{ jest } \hat{A}_j^{(t)}\}$ jest odwzorowaniem (6), a *sup* i *conf* są wsparciem i ufnością w sensie (13, 14). Powyższe reguły mogą mieć zastosowanie

do tworzenia modelu systemu dynamicznego przy istnieniu niepewności o charakterze rozmytym. Wówczas pierwsza reguła plikowa R(1) modelu AR(m) ma postać (por.[11, 15, 16]):

w_1 (Jeżeli

$$x_{t-m} \text{ jest } \hat{A}_1^{(t-m)} \text{ I } \dots \text{ I } x_{t-1} \text{ jest } \hat{A}_1^{(t-1)}$$

To x_t jest $\hat{A}_1^{(t)}$ z wagą $w_1/1$

Także x_t jest $\hat{A}_2^{(t)}$ z wagą $w_2/1$

...

Także x_t jest $\hat{A}_K^{(t)}$ z wagą $w_K/1$. (16)

Wagi reguł są prawdopodobieństwami, wynikającymi odpowiednio z wsparcia i ufności w rozmytych regułach asocjacji.

5. PRZYKŁAD ROZMYTYCH MODELI SYSTEMU DYNAMICZNEGO

Zastosowanie reguł asocjacji w modelowaniu systemów dynamicznych zostanie pokazane na przykładzie analizy frakcji węgla kamiennego.

Istotą kontroli jakości węgla są pomiary niektórych jego parametrów tj. zawartość popiołu, siarki czy wartości opalowej. Badania, opierające się na próbkowaniu oraz analizie próbki uziarnionej, są narażone na możliwość wystąpienia błędów [12]. Ponadto, rozkład frakcji (zwłaszcza niejednorodnej) jest obciążony elementem losowości. Omawiany proces jest zatem procesem stochastycznym, zawierającym niepewność informacji. Zalecana jest więc analiza takiego zagadnienia w aspekcie rozmytości danych empirycznych.

Badaniu zostały poddane kolejne 496 próbek materiału uziarnionego. Dokonano pomiarów zarówno, co do zawartości węgla, jak i popiołu w tej samej frakcji.

5.1 Analiza frakcji lekkiej węgla

Analiza frakcji lekkiej węgla (X) została przeprowadzona za pomocą rozmytego modelowania autoregresji AR(1) do AR(n), gdzie n jest rzędem autoregresji.

Zawartość frakcji lekkiej węgla w próbce jest daną ilościową, wobec tego dokonano dyskretyzacji wartości na 10 przedziałów a_1, \dots, a_{10} , stałych w przestrzeniach zmienności. Wówczas, nierozmyte reguły asocjacji dla modelu AR(1) (min. $sup=0$, min. $conf=0$), będące wynikiem zastosowania Oracle Data Miner, mają postać np.:

Tab.1.

Przykłady nierozmytych reguł asocjacji modelu AR(1)
The examples of nofuzzy association rules
for model AR(1)

Rule Id	If (condition)	Then (association)	Conf. (%)	Sup. (%)
66	X(t-1)=(0.34,0.41>	X(t)=(0.34,0.41>	30,84	6,67

84	X(t-1)=(0.34,0.41>	X(t)=(0.41,0.47>	27,10	5,86
100	X(t-1)=(0.34,0.41>	X(t)=(0.47,0.54>	19,63	4,24
48	X(t-1)=(0.34,0.41>	X(t)=(0.27,0.34>	6,54	1,41
118	X(t-1)=(0.34,0.41>	X(t)=(0.54,0.61>	5,61	1,21
32	X(t-1)=(0.34,0.41>	X(t)=(0.20,0.27>	5,61	1,21
20	X(t-1)=(0.34,0.41>	X(t)=(0.14,0.20>	1,87	0,40
8	X(t-1)=(0.34,0.41>	X(t)=(0.07,0.14>	1,87	0,40
130	X(t-1)=(0.34,0.41>	X(t)=(0.61,...)	0,93	0,20

Dokonując fuzyfikacji reguł, wprowadzono pięć zbiorów rozmytych A_1, \dots, A_5 , utożsamianych z wartościami zmiennej lingwistycznej Z z zakresu:

$$L(Z) = \{ \text{bardzo niska, niska, średnia, wysoka, bardzo wysoka} \}, \quad (17)$$

w skrócie: $L(Z) = \{BN, N, S, W, BW\}$ i stałymi wartościami funkcji przynależności $\mu_{A_i}(a_i)$, z jakimi zbiory a_1, \dots, a_{10} , uzyskane z dyskretyzacji wartości atrybutów, należą do zbiorów A_1, \dots, A_5 (tab. 2).

Tab.2.

Wartości funkcji przynależności $\mu_{A_i}(a_i)$ dla frakcji lekkiej węgla

The values of membership function $\mu_{A_i}(a_i)$ for light fraction of coal

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
$\mu_{A1}(a_i)$	0,9	0,5	0,2	0,1	0	0	0	0	0	0
$\mu_{A2}(a_i)$	0,1	0,5	0,7	0,5	0,1	0	0	0	0	0
$\mu_{A3}(a_i)$	0	0	0,1	0,4	0,9	0,9	0,4	0,1	0	0
$\mu_{A4}(a_i)$	0	0	0	0	0	0,1	0,5	0,7	0,5	0,1
$\mu_{A5}(a_i)$	0	0	0	0	0	0	0,1	0,2	0,5	0,9

Oczywiście zachodzi zależność (8), czyli:

$$\sum_{j=1}^5 \mu_{A_j}(a_i) = 1; \quad i = 1, \dots, 10. \quad (18)$$

Wyniki obliczeń dla AR(1) na podstawie wzorów (13, 14), w których otrzymujemy rozkłady prawdopodobieństwa zdarzeń rozmytych, związanych z przynależnością wartości do zbiorów A_1, \dots, A_5 , zostały przedstawione w tab. 3 (rozkład łączny) i w tab. 4 (rozkład warunkowy). Natomiast przykład rozmytych reguł asocjacji prezentuje tab. 5. Wówczas, model systemu dynamicznego, z uwzględnieniem autoregresji rzędu pierwszego, zawiera 5 reguł plikowych. W każdym pliku znajdują się do 5 reguł elementarnych. Dla frakcji lekkiej węgla otrzymano zbiór reguł w postaci:

R(1): 0,4303 (Jeżeli X(t-1) jest S

To X(t) jest S z wagą 0,4721

Także X(t) jest W z wagą 0,3167

Także X(t) jest BW z wagą 0,1028

Także X(t) jest N z wagą 0,0794

Także X(t) jest BN z wagą 0,0290).

Przykładowo model systemu dynamicznego z uwzględnieniem autoregresji rzędu czwartego, zawiera już 625 (5^4) reguł plikowych. W otrzymanych wynikach, tylko 596 z nich ma wagę niezerową. Pierwsza z reguł wygląda następująco:

R(1): 0,05446 (Jeżeli $X(t-4)$ jest S I $X(t-3)$ jest S I $X(t-2)$ jest S I $X(t-1)$ jest S

To $X(t)$ jest S z wagą 0,56800

Także $X(t)$ jest W z wagą 0,26248

Także $X(t)$ jest BW z wagą 0,07434

Także $X(t)$ jest N z wagą 0,06275

Także $X(t)$ jest BN z wagą 0,03243).

Tab.3.

Łączny rozkład prawdopodobieństwa dla modelu AR(1)
The joint probability distribution of fuzzy events for the AR(1) model

$X(t)$	$X(t-1)$	Rozkład łączny prawdopodobieństwa $P(X(t-1)=A_i \cap X(t)=A_j)$				
		A_1 (BN)	A_2 (N)	A_3 (S)	A_4 (W)	A_5 (BW)
A_1 (BN)		0,0003	0,0018	0,0125	0,0080	0,0024
A_2 (N)		0,0019	0,0079	0,0342	0,0216	0,0066
A_3 (S)		0,0113	0,0321	0,2031	0,1367	0,0469
A_4 (W)		0,0069	0,0203	0,1363	0,1309	0,0511
A_5 (BW)		0,0037	0,0091	0,0442	0,0497	0,0206

Tab.4.

Warunkowy rozkład prawdopodobieństwa dla modelu AR(1)

The conditional probability distribution for model AR(1)

$X(t)$	$X(t-1)$	Rozkład warunkowy prawdopodobieństwa $P(X(t)=A_j / X(t-1)=A_i)$				
		A_1 (BN)	A_2 (N)	A_3 (S)	A_4 (W)	A_5 (BW)
A_1 (BN)		0,0134	0,0253	0,0290	0,0231	0,0190
A_2 (N)		0,0773	0,1108	0,0794	0,0623	0,0517
A_3 (S)		0,4681	0,4511	0,4721	0,3941	0,3676
A_4 (W)		0,2882	0,2852	0,3167	0,3773	0,4002
A_5 (BW)		0,1529	0,1276	0,1028	0,1432	0,1616

Tab.5.

Przykłady rozmytych reguł asocjacji modelu AR(1)

The examples of fuzzy association rules for model AR(1)

Rule Id	If (condition)	Then (association)	Conf.	Sup.
1	$X(t-1)=S$	$X(t)=S$	0,4721	0,2031
2	$X(t-1)=W$	$X(t)=S$	0,3941	0,1367
3	$X(t-1)=S$	$X(t)=W$	0,3167	0,1363
4	$X(t-1)=W$	$X(t)=W$	0,3773	0,1309
5	$X(t-1)=BW$	$X(t)=W$	0,4002	0,0511
6	$X(t-1)=W$	$X(t)=S$	0,1432	0,0497

5.2 Analiza zawartości węgla i popiołu frakcji lekkiej

Podział przestrzeni zmienności dla węgla (X) i popiołu (Y) frakcji lekkiej dokonano, podobnie jak w poprzednim podpunkcie, na 10 równych zbiorów nierozmytych, nazwanych odpowiednio a_1, \dots, a_{10} oraz b_1, \dots, b_{10} . Dla węgla wprowadzono pięć zbiorów rozmytych A_1, \dots, A_5 , odnoszących się do wartości lingwistycznych zmiennej Z zgodnie z (17) oraz posiadających wartości funkcji przynależności $\mu_{A_i}(a_i)$ jak w tabeli 2. Dla popiołu wprowadzono także pięć

zbiorów rozmytych B_1, \dots, B_5 z analogicznymi wartościami lingwistycznymi $L(K)$ zmiennej lingwistycznej K oraz wartościami funkcji przynależności $\mu_{B_i}(b_i)$ zawartymi w tab. 6.

Tab.6.

Wartości funkcji przynależności $\mu_{B_i}(b_i)$ dla popiołu frakcji lekkiej

The values of membership function $\mu_{B_i}(b_i)$ for the ash contents in the coal light fraction

	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
$\mu_{B_1}(b_1)$	0,7	0,1	0	0	0	0	0	0	0	0
$\mu_{B_2}(b_2)$	0,3	0,5	0,1	0	0	0	0	0	0	0
$\mu_{B_3}(b_3)$	0	0,4	0,9	0,9	0,7	0,4	0	0	0	0
$\mu_{B_4}(b_4)$	0	0	0	0,1	0,3	0,6	0,9	0,6	0,1	0
$\mu_{B_5}(b_5)$	0	0	0	0	0	0	0,1	0,4	0,9	1

Przeprowadzono eksplorację danych empirycznych algorytmem Apriori dla dwóch atrybutów X i Y zgodnie z rzędem drugim autoregresji. Reguły z najwyższymi miarami wsparcia dla AR(2) zawarte są w tab. 7.

Tab.7.

Przykłady nierozmytych reguł asocjacji modelu AR(2)

The examples of nofuzzy association rules for model AR(2)

Rule Id	If (condition)	Then (association)	Conf. (%)	Sup. (%)
11842	$X(t-2) < 0.34, 0.41$ AND $Y(t-2) < 1.5, 2.38$ AND $X(t-1) < 0.34, 0.41$ AND $Y(t-1) < 1.5, 2.38$	$Y(t) < 1.5, 2.38$	81,82	1,82
11877	$X(t-2) < 0.34, 0.41$ AND $Y(t-2) < 1.5, 2.38$ AND $X(t-1) < 0.41, 0.47$ AND $Y(t-1) < 1.5, 2.38$	$Y(t) < 1.5, 2.38$	77,78	1,42
12267	$X(t-2) < 0.41, 0.47$ AND $Y(t-2) < 1.5, 2.38$ AND $X(t-1) < 0.41, 0.47$ AND $Y(t-1) = (\dots, 1.5)$	$Y(t) < 1.5, 2.38$	60,00	1,21

Po dokonaniu obliczeń (13), wprowadzających rozmycie do reguł asocjacji, otrzymano rozkład prawdopodobieństwa łącznego w postaci tablicy o wymiarach $5 \times 5 \times 5 \times 5$. Przykładowy rozkład łącznego prawdopodobieństwa, gdy $X(t-2)=A_3$, $Y(t-2)=B_3$ oraz $Y(t-1)=B_3$, przedstawia tab. 8.

Tab.8.

Łączny rozkład prawdopodobieństwa dla modelu AR(2)

The joint probability distribution for model AR(2)

$Y(t)$	$X(t-1)$	Rozkład łączny prawdopodobieństwa $P(X(t-2)=A_3 \cap Y(t-2)=B_3 \cap X(t-1)=A_i \cap Y(t-1)=B_3 \cap Y(t)=B_j)$				
		A_1 (BN)	A_2 (N)	A_3 (S)	A_4 (W)	A_5 (BW)
B_1 (BN)		0,0001	0,0007	0,0029	0,0017	0,0007
B_2 (N)		0,0008	0,0032	0,0146	0,0101	0,0043
B_3 (S)		0,0025	0,0101	0,0477	0,0379	0,0175
B_4 (W)		0,0003	0,0015	0,0050	0,0039	0,0017
B_5 (BW)		0,0000	0,0003	0,0006	0,0007	0,0006

Kolejno, wagi reguł elementarnych zostały podane jako warunkowe prawdopodobieństwo zdarzeń rozmytych $P(Y(t) = B_{3j} / X(t-2) = A_{1i} \cap Y(t-2) = B_{1j} \cap X(t-1) = A_{2i} \cap Y(t-1) = B_{2j})$, analogicznie do (14).

Otrzymano 529 reguł plikowych z wagami niezerowymi. Pierwsza reguła modelu jest w postaci:

R(1): 0,07082 **Jeżeli** $X(t-2)$ jest S **I** $Y(t-2)$ jest S **I** $X(t-1)$ jest S **I** $Y(t-1)$ jest S

To $Y(t)$ jest S wagą 0,67303

Także $Y(t)$ jest N z wagą 0,20683

Także $Y(t)$ jest W z wagą 0,06990

Także $Y(t)$ jest BN z wagą 0,04110

Także $Y(t)$ jest BW z wagą 0,00914).

6. WNIOSKI

Eksploracja danych algorytmem wykrywającym reguły asocjacji ułatwia nam wyszukiwanie zależności w modelowaniu szeregów czasowych, zarówno dla jednej, jak i więcej zmiennych. Jednakże, generowanie reguł stanowi stosunkowo niewielką część w przedstawionym procesie modelowania. Najwięcej czasu pochłania przygotowanie danych a przede wszystkim fuzyfikacja. Czas obliczeń znacznie wzrasta wraz ze zwiększeniem rzędu autoregresji, liczby zbiorów (na jakie przestrzeń wartości rozważanych zmiennych została podzielona) oraz liczby wartości lingwistycznych tychże zmiennych.

Jednym z rozwiązań jest zastosowanie algorytmu wyszukującego bezpośrednio rozmyte reguły asocjacji, jednakże nie jest on zaimplementowany w znanych, komercyjnych narzędziach eksploracji.

7. LITERATURA

1. Agrawal R., Imielinski T., Swami A.: *Mining association rules between sets of items in large databases*, ACM Sigmod Intern. Conf. on Management of Data, Washington D.C., May 1993, pp. 207- 216.
2. Agrawal R., Srikant R.: *Fast Algorithms for Mining Association Rules*, Proc. of 1994 Intern. Conf. on Very Large Databases VLDB, Santiago de Chile, September 1994, pp. 487– 499.
3. Alvarez J.L., Mata J., Riquelme J.C.: *An Evolutionary Algorithm to Discover Numeric Association Rules*, <http://www.lsi.us.es/minerva/archivos/articulos/publicacion25.pdf>.
4. Borgelt C.: *An Implementation of the FP-growth Algorithm*, Workshop Open Software for Data Mining (OSDM'05, Chicago, IL), ACM Press, New York, NY, USA 2005.
5. Box G.E.P., Jenkins G.M.: *Analiza szeregów czasowych*, Prognozowanie i sterowanie, PWN, Warszawa 1983.
6. Chen G., Wei Q., *Fuzzy Association Rules and the Extended Mining Algorithm*, Information Sciences, 2002, 147, pp. 201– 228.

7. Han J., Pei J., Mortazavi-Asl B., Chen Q., Dayal U., Hsu M.: *FreeSpan: frequent pattern-projected sequential pattern mining*. Conf. on Knowledge Discovery in Data. Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. Boston, Massachusetts, USA, 2000, pp. 355 – 359.
8. Hong T.P., Kuo C.S., Chi S.C., *Trade-off between Computation Time and Number of Rules for Fuzzy Mining from Quantitative Data*, International Journal of Uncertainty, Fuzziness and Knowledge-Based System, 2001, Vol.9, No. 5 pp. 587– 604.
9. Kuok C.M., Fu A., Wong M.H.: *Mining Fuzzy Association Rules in Database*, Sigmod, vol.27 I, March 1998, pp. 41– 46.
10. Savasere A., Omiecinski E., Navathe S., *An efficient algorithm for mining association rules in large databases*, in Proc. of 21th VLDB Conference, San Francisco, CA, USA, 1995, pp. 432– 444.
11. Walaszek-Babiszewska A.: *Rozmyte modele decyzyjne pozyskiwane z danych eksperymentalnych*, Mat. X Konf. Nauk.-Techn., 'Automation 2006', PIAP, Warszawa, marzec 2006, str. 262 – 271.
12. Walaszek-Babiszewska A.: *Quasi-liniowy problem programowania rozmytego w optymalizacji mieszanek wegla energetycznego*, Mat. XIV Krajowej Konf. Automatyki, Zielona Góra, czerwiec 2002, t. I str. 464 – 465.
13. Yager R., Filev D.: *Podstawy modelowania i sterowania rozmytego*, WNT, Warszawa 1995.
14. Zadeh L.A.: *Fuzzy sets*. Inform. Contr., 1965 vol. 8, pp. 338 – 353.
15. Zadeh L.A.: *The concept of a linguistic variable and its application to approximate reasoning*. Information Sciences, 1975, vol. I, pp. 199 – 240.
16. Zaki M.J., Parthasarathy S., Ogihara M., Li W.: *New Algorithms for Fast Discovery of Association Rules*, in Proc. of 3rd International Conference on Knowledge Discovery and Data Mining (KDD-97), Newport Beach, California, USA, 1997.
17. Zhi-Chao L., Pi-Lian H., Ming L.: *A high efficient Apriori-Tid algorithm for mining association rule*, Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on Volume 3, Issue, 18-21 Aug. 2005, pp. 1812 – 1815 Vol. 3.

Adres służbowy Autora:



Mgr inż. K. Błaszczyk
 Politechnika Opolska
 ul. Ozimska 75
 45-370 Opole
 tel. (077) 423-40-35

email: k.blaszczyk@po.opole.pl