

KATARZYNA BŁASZCZYK

BOGDAN RUSZCZAK

Politechnika Opolska

WIELOWYMIAROWE REGUŁY ASOCJACJI W MODELOWANIU TENDENCJI ROZWOJOWYCH MSP

Wstęp

Eksploracja danych (ang. *data mining*) zajmuje się efektywnym znajdowaniem nieznanych dotychczas zależności i związków pomiędzy danymi¹ ukrytych pod postacią wzorców, trendów, regularności, korelacji itp.². Jest to dziedzina szeroko i bardzo intensywnie rozwijana z uwagi na potrzebę przeprowadzania analiz licznie nagromadzonych w przedsiębiorstwach, instytucjach administracyjnych czy ośrodkach naukowych, danych.

W publikacji przedstawiono możliwość zastosowania metod eksploracji w analizie danych makroekonomicznych. W szczególności wykorzystano metodę wielowymiarowych reguł asocjacji, która pozwala wykryć powiązania między szeregami różnych wartości, a jej probabilistyczny charakter umożliwia określenie stopnia prawdopodobieństwa poszczególnych zjawisk. Analiza została przeprowadzona w celu zamodelowania tendencji rozwojowych małych i średnich przedsiębiorstw (MSP), kluczowego czynnika wpływającego na rozwój gospodarczy. Tendencje rozwojowe MSP obrazuje się szeregiem różnych cech, na potrzeby tego badania wyróżniono tylko kilka głównych tj. liczba nowych przedsiębiorstw (mierzona według nowo zgłaszanych podmiotów w rejestrze REGON), liczba pracujących, ilość przedsiębiorstw aktywnych itp.. Badano powiązania między nimi oraz analizowano szeregi czasowe cech ukazując predykcyjną tendencję rozwoju.

1. Istota reguł asocjacji

Pojęcie reguł asocjacji (ang. *association rules*) zostało po raz pierwszy użyte w pracy³, i oznacza odnajdywanie w dużej kolekcji zbiorów korelacji wiążącej jednoczesne

1 - T. Morzy: Eksploracja danych: problemy i rozwiązania. V Konferencja PLOUG, Zakopane, październik 1999.

2 - M. Morzy Oracle Data Mining - odkrywanie wiedzy w dużych wolumenach danych. XI Krajowa Konferencja PLOUG 2005, Zakopane, październik 18-21, 2005.

3 - Agrawal R., Imielinski T., Swami A.: Mining association rules between sets of items in large databases. ACM SIGMOD International Conference on Management of Data, Washington D.C., May 26-28 1993, pp. 207-216.

występowanie wartości atrybutów. W ujęciu ilościowym reguły asocjacji mają postać implikacji (por. ⁴):

$$X \Rightarrow Y \text{ (wsparcie, ufność),}$$

gdzie X i Y są rozłącznymi zbiorami atrybutów. Inaczej możemy zapisać w postaci:

$$A_1 \cap \dots \cap A_n \Rightarrow B_1 \cap \dots \cap B_m \text{ (wsparcie, ufność),}$$

gdzie A_i i B_j są odpowiednio parami atrybut-wartość.

Każda reguła zawiera dodatkowe miary statystyczne: wsparcie (ang. *support*), które stanowi prawdopodobieństwo występowania zbioru $X \cap Y$ w kolekcji zbiorów, oraz ufność (ang. *confidence*), określające prawdopodobieństwo warunkowe $P(Y|X)$ występowania zbiorów.

Zastosowanie reguł asocjacji najczęściej ma miejsce w odkrywaniu wiedzy w analizie koszykowej (ang. *basket analysis*). Wówczas mamy do czynienia z tzw. jednowymiarowymi regułami asocjacji. Gdyż w regułach postaci: „Jeżeli klient kupił produkty A_1, \dots, A_n to kupi także w tej samej transakcji produkty B_1, \dots, B_n (wsparcie %, ufność %)” atrybuty posiadają tylko jedną dziedzinę wartości – produkt. Takie reguły mogą być także wykorzystywane podczas analizy dowolnych danych transakcyjnych np. w bankowości, ubezpieczeniach, telekomunikacji, medycynie czy analizie pogody^{5,6}. Wielowymiarowe reguły asocjacji posiadają natomiast w poprzedniku i następniku reguły różne dziedziny wartości, przykładowo „Jeżeli $wiek \in \langle 20; 30 \rangle$ i $zarobki \in \langle 2000,00; 3000,00 \rangle$ to $wyjazd = \{Chorwacja\}$ (21%, 57%)”. Takie reguły dają szersze możliwości ich zastosowania, nie tylko do analizy danych gdzie „klienci nabywają łącznie pewien zbiór dóbr lub usług”, ale praktycznie w każdej dziedzinie badań podczas wnioskowania statystycznego, prognozowania, podejmowaniu decyzji czy sterowania.

2. Potrzeba wprowadzenia lingwistycznych reguł asocjacji

Jak można zauważyć, wspomniane wielowymiarowe reguły asocjacji posiadają wartości atrybutów w postaci ilościowej lub kategoryjnej (np. $wiek = 34$ lata, ilość przedsiębiorstw = 1230, $wyjazd = \text{„Chorwacja”}$). Wyszukiwanie zależności między wartościami wszystkich takich atrybutów może okazać się ograniczone, a czasami wręcz niemożliwe. Klasycznym⁴ rozwiązaniem takiego problemu jest dyskretyzacja wartości ciągłych do postaci

4 - Agrawal R., Srikant R.: Mining Generalized Association Rules, In Proc. of the 21st Int'l Conference on Very Large Databases, Zurich, Switzerland, 1995.

5 - Kawa K.: Zastosowanie reguł asocjacyjnych, pakietu Oracle Data Mining for Java do analizy koszyka zakupów w aplikacjach e-commerce. XI Konferencja PLOUG, Kościelisko, październik 18-21 2005.

6 - Leśniewska A., Morzy M., Morzy T.: Eksploracja danych. Materiały studiów informatycznych, http://wazniak.mimuw.edu.pl/index.php?title=Eksploracja_danych.

przedziałów wartości. Mamy w ten sposób przedziały wiekowe, przedziały ilości przedsiębiorstw itp. Wartość danego atrybutu mieści się w tym przedziale bądź nie, czyli jego przynależność do danego zbioru kształtują się na poziomie 1 lub 0.

Z metodologicznego punktu widzenia takie rozwiązanie jest w pełni satysfakcjonujące, lecz biorąc pod uwagę łatwość i poprawność interpretacji uzyskanych wyników, pojawia się kolejny problem. Przeanalizujemy sytuację: osoba obchodzi jutro 35-te urodziny, dzisiaj tymczasem ma lat 34 i zostaje zakwalifikowana do przedziału wiekowego „osób młodych” $<25,35$), jutro natomiast traci ten status przechodząc do grupy wiekowej „osób dojrzałych” $<35, 45$). W rzeczywistości przejście pomiędzy byciem osobą młodą i dojrzałą nie jest nagłe, lecz stopniowe. Ponadto zastosowanie ścisłych granic matematycznych może być różnie interpretowane przez różne osoby (ekspertów): dla jednych osoba w wieku $<35,45$) lat jest osobą młodą dla innych osobą dojrzałą. Przykłady dostarczają nam argumentów, które przemawiają za wprowadzeniem tzw. fuzyfikacji reguł asocjacji.

Dzięki zastosowaniu logiki rozmytej w poprzedniku i następniku reguły wartości argumentów przynależą do tzw. zbiorów rozmytych⁷ w określonym stopniu przynależności z przedziału $(0,1)$. Od 0 dla „element nie należy do zbioru” do 1 dla „element należy do zbioru”, poprzez wszystkie wartości pośrednie, oznaczające „częściową przynależność”. Wprowadzając dodatkowo zmienne lingwistyczne (np. *niski*, *wysoki*, *średni*), których wartości są utożsamiane semantycznie ze niektórymi zbiorami rozmytymi⁸ uzyskujemy reguły asocjacji w postaci implikacji:

$$A_1 \text{ jest } X_1 \cap \dots \cap A_n \text{ jest } X_n \Rightarrow B_1 \text{ jest } Y_1 \cap \dots \cap B_m \text{ jest } Y_m \text{ (wsparcie, ufność),}$$

gdzie A_i są atrybutami, B_j są zbiorami rozmytymi (wartościami zmiennych lingwistycznych). Otrzymujemy w ten sposób kompromis pomiędzy adekwatnością a prostotą⁹. Jest to możliwe dzięki rozumowaniu człowieka, które nie jest „sztywnie” oparte na zmiennych liczbowych, lecz przybliżone - związane z użyciem pojęć werbalnych, słów i zdań użytych w języku naturalnym. Takie rozwiązanie jest użyteczne zarówno podczas sterowania jaki i wnioskowania statystycznego, prognozowania i podejmowaniu decyzji, gdzie jednostką analizującą jest człowiek-ekspert.

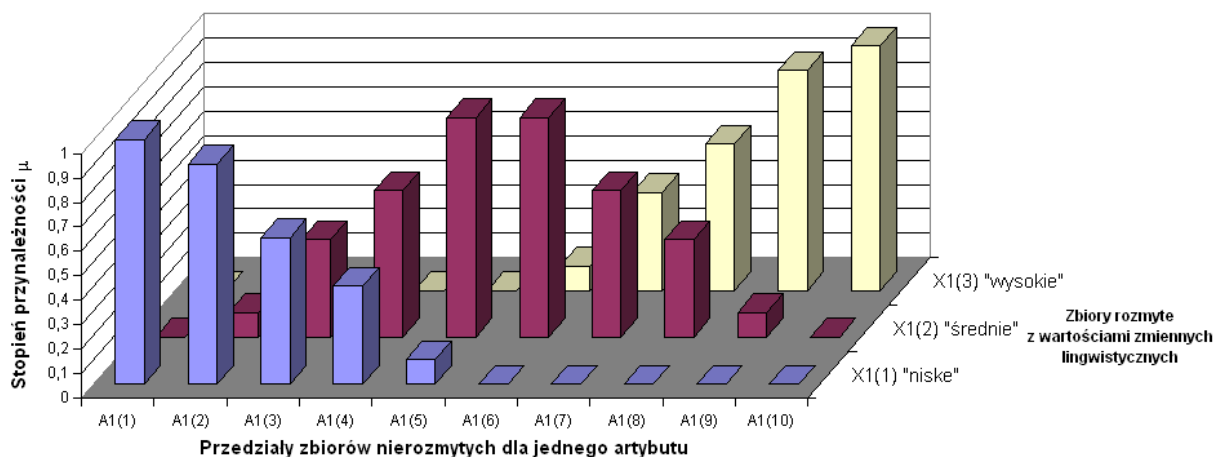
3. Budowanie lingwistycznych reguł asocjacji

7 - Zadeh L.A.: Fuzzy set. Information and Control. 1965, vol. 8, pp 338-353.

8 - Kacprzyk J.: Wieloetapowe sterowanie rozmyte. Wydawnictwo Naukowo-Techniczne, Warszawa, 2001.

9 - Kacprzyk J.: Zbiory rozmyte w analizie systemowej. Państwowe Wydawnictwo Naukowe, Warszawa, 1986.

Większość programów data mining jest przystosowanych jedynie do wyliczania ilościowych, „nielingwistycznych” reguł asocjacji. W artykule proponujemy metodę obliczania lingwistycznych reguł asocjacji, polegającą na wykorzystaniu konkretnych wartości stopni przynależności μ , z jakimi zbiory nierozmyte uzyskane z dyskretyzacji wartości atrybutów, należą do zbiorów rozmytych, związanych z danymi wartościami zmiennych lingwistycznych. Rysunek 1 ilustruje przykład przyporządkowania stopni przynależności poszczególnych zbiorów nierozmytych do zbiorów rozmytych, dla pojedynczego atrybutu.



Rys. 1. Stopnie przynależności, z jakimi zbiory nierozmyte, należą do zbiorów rozmytych

Źródło: Opracowanie własne na podstawie obliczeń

Przy obliczaniu wartości wsparcia oraz ufności reguł lingwistycznych korzystamy z definicji (nierozmytego) prawdopodobieństwa zdarzenia rozmytego¹⁰:

$$p(X_j) = \sum_{i=1}^n \mu_{X_j}(A_i) p_i(A_i)$$

gdzie:

$p(X_j)$ – prawdopodobieństwo zajścia zdarzenia rozmytego X_j ,

$\mu_{X_j}(A_i)$ – stopień przynależności, z jakim zbiór nierozmyty A_i należy do zbioru rozmytego X_j ,

$p_i(A_i)$ – prawdopodobieństwo nierozmytego zdarzenia A_i .

Zdarzenie nierozmyte wynika z dyskretyzacji wartości atrybutów.

W przypadku wsparcia „nielingwistycznych” reguł asocjacji, mamy do czynienia z rozkładem wielowymiarowym prawdopodobieństwa wystąpienia wartości atrybutów (rozkładem łącznym), gdzie:

$$P(A_{i_1} \cap \dots \cap A_{i_n} \cap B_{j_1} \cap \dots \cap B_{j_m}) = \text{wsparcie}_{i_1 \dots i_n \dots j_1 \dots j_m}$$

10 Zadeh L.A.: Probability measures of fuzzy events. Journal of Mathematical Analysis and Applications, Vol. 23 No. 2, August 1968, pp. 421-427.

$i_1 = 1, 2, \dots, a_{i_1}, \quad i_n = 1, 2, \dots, a_{i_n}, \quad j_1 = 1, 2, \dots, b_{j_1}, \quad j_n = 1, 2, \dots, b_{j_n}, \quad A_{i_1}, B_{j_1} - \text{zdarzenia nierozmyte.}$

Rozkład można ująć w postaci tablicy $a_{i_1} \times \dots \times a_{i_n} \times \dots \times b_{j_1} \times \dots \times b_{j_n}$. Oczywiście zachodzi

zależność:
$$\sum_{i_1=1}^{a_{i_1}} \dots \sum_{i_n=1}^{a_{i_n}} \dots \sum_{j_1=1}^{b_{j_1}} \dots \sum_{j_n=1}^{b_{j_n}} \text{wsparcie}_{i_1 \dots i_n \dots j_1 \dots j_n} = 1.$$
 Dalej, wykorzystując własności

prawdopodobieństwa zdarzeń rozmytych, możemy obliczyć wsparcie lingwistycznych reguł asocjacji bezpośrednio z rozkładu łącznego zdarzeń nierozmytych. Dla dwóch atrybutów A, B zachodzi zależność (por.¹¹):

$$P(X_k \cap Y_l) = p(A_1 \cap B_1) \mu_{X_k}(A_1) \mu_{Y_l}(B_1) + \dots + p(A_n \cap B_1) \mu_{X_k}(A_n) \mu_{Y_l}(B_1) + \\ p(A_1 \cap B_j) \mu_{X_k}(A_1) \mu_{Y_l}(B_j) + \dots + p(A_1 \cap B_m) \mu_{X_k}(A_1) \mu_{Y_l}(B_m) + \dots + p(A_n \cap B_m) \mu_{X_k}(A_n) \mu_{Y_l}(B_m),$$

gdzie $k=1, 2, \dots, K, \quad l=1, 2, \dots, L, \quad A_i, B_j - \text{zdarzenia nierozmyte, } X_k, Y_l - \text{zdarzenia rozmyte.}$

Ufność natomiast, jako prawdopodobieństwo warunkowe, wynika bezpośrednio z rozkładów łącznych i brzegowych otrzymanych zdarzeń rozmytych.

4. Modelowanie tendencji rozwojowych MSP

W artykule przeprowadzono analizę poszczególnych wskaźników tendencji rozwojowych w sektorze MSP, biorąc pod uwagę element współwystępowania ich wartości w znaczeniu lingwistycznym. Analizie podlegały takie wskaźniki jak: liczba nowych prywatnych podmiotów gospodarczych zarejestrowanych w REGON, liczba pracujących w MSP należących do sektora rynkowego, liczba przedsiębiorstw aktywnych, nakłady na inwestycje oraz innowacje MSP. Zakres analizy, na podstawie¹², obejmował sytuację MSP w układzie regionalnym, według województw w firmach małej i średniej wielkości, w latach 1994-2003.

Początkowym etapem badań był podział wartości atrybutów na przedziały o równej szerokości. Na podstawie¹³ oraz przeświadczeniu o maksymalizacji liczby przedziałów, dokonano dyskretyzacji wartości (tab. 1).

11 Walaszek-Babiszewska A.: Measurements and expert knowledge for time-dependent stochastic systems. The 5th International Conference, IPMM-2005, Monterey, California, USA, July 19-23, 2005.

12 - Chmiel J., Skrzek – Lubasińska M., Urbańska – Jobda B.: Stan sektora MSP w 2003 roku tendencje rozwojowe w latach 1994 – 2003. Warszawa, 2005.

13 - Ostasiewicz S., Rusnak Z., Siedlecka U.: Statystyka. Elementy teorii i zadania. Wyd. Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław, 1999.

Tab.1. Dyskretyzacja wartości atrybutów

Nr	Nowe prywatne podmioty gospodarcze zarejestrowane w REGON	Pracujący w MSP należących do sektora rynkowego [osoby]	Przedsiębiorstwa aktywne	Nakłady inwestycyjne MSP [tyś. zł]	Nakłady na innowacje [tyś. zł]
1	(...,10088.9)	(...,211244.2)	(...,43740.7)	(...,2509506.3)	(...,71141.7)
2	<10088.9,14940.8)	<211244.2,296793.4)	<43740.7,69707.4)	<2509506.3,4493084.6)	<71141.7,137639.9)
3	<14940.8,19792.7)	<296793.4,382342.6)	<69707.4,95674.0)	<4493084.6,6476662.9)	<137639.9,204138.2)
4	<19792.7,24644.6)	<382342.6,467891.8)	<95674.0,121640.7)	<6476662.9,8460241.2)	<204138.2,270636.4)
5	<24644.6,29496.5)	<467891.8,553441.0)	<121640.7,147607.3)	<8460241.2,10443819.5)	<270636.4,337134.7)
6	<29496.5,34348.4)	<553441.0,638990.2)	<147607.3,173573.9)	<10443819.5,12427397.8)	<337134.7,403632.9)
7	<34348.4,39200.3)	<638990.2,724539.4)	<173573.9,199540.6)	<12427397.8,14410976.1)	<403632.9,470131.2)
8	<39200.3,44052.2)	<724539.4,810088.6)	<199540.6,225507.2)	<14410976.1,16394554.4)	<470131.2,536629.4)
9	<44052.2,48904.1)	<810088.6,895637.8)	<225507.2,251473.9)	<16394554.4,18378132.7)	<536629.4,603127.7)
10	<48904.1,...)	<895637.8,...)	<251473.9,...)	<18378132.7,...)	<603127.7,...)

Źródło: Opracowanie własne na podstawie obliczeń

W analizie, dla każdego atrybutu, uwzględniono jednakowe wartości zmiennych lingwistycznych oraz jednakowy przydział stopni przynależności μ , w jakim kolejne zbiory nierozmyte należą do zbiorów rozmytych (tab. 2).

Tab.2. Stopnie przynależności μ

Termin lingwistyczny	Używany skrót	Nr. przedziału wartości atrybutu									
		1	2	3	4	5	6	7	8	9	10
'Bardzo niskie'	'BN'	0.9	0.6	0.3	0.1	0	0	0	0	0	0
'Niskie'	'N'	0.1	0.4	0.6	0.2	0.2	0.1	0	0	0	0
'Średnie'	'S'	0	0	0.1	0.5	0.7	0.7	0.5	0.1	0	0
'Wysokie'	'W'	0	0	0	0	0.1	0.2	0.4	0.6	0.4	0.1
'Bardzo wysokie'	'BW'	0	0	0	0	0	0	0.1	0.3	0.6	0.9

Źródło: Opracowanie własne na podstawie obliczeń

Przeprowadzono badanie związku pomiędzy liczbą nowych przedsiębiorstw a liczbą pracujących w tym sektorze. Badanie wydawałoby się banalne, pozwala jednak potwierdzić poprawność zbudowanego modelu, a także dowieść trafności twierdzenia, że wzrost liczby przedsiębiorców generuje większy poziom zatrudnienia. Wybrane wyniki przedstawiono w tabeli 3. Wyniki wyraźnie ukazują zależności tych dwóch wielkości – przy niewielkiej liczbie nowych przedsiębiorstw ($A=BN$) prawdopodobieństwo warunkowe, że stan zatrudnienia będzie niski lub bardzo niski ($B=BN$ lub $B=N$) wynosi sumarycznie aż 0,97. Także porównując zdarzenie odwrotne – kiedy występuje duża liczba nowo powstałych przedsiębiorstw ($A=BW$) prawdopodobieństwo wysokiego lub bardzo wysokiego stanu liczby zatrudnionych w sektorze MSP ($B=BW$ lub $B=W$) wynosi 0,94.

Tab.3. Wybrane lingwistyczne reguły asocjacji dla badania liczby nowych przedsiębiorstw (A) i liczby pracujących w MSP (B).

If (condition) Then (association)	Wsparcie	Ufność
'IF A=BN THEN B=BN'	0,28688	0,75743
'IF A=BN THEN B=N'	0,08113	0,21419
'IF A=BN THEN B=W'	0,00025	0,00066
'IF A=BN THEN B=BW'	0	0
'IF A=N THEN B=BN'	0,13938	0,49336
'IF A=N THEN B=N'	0,09738	0,34469
'IF A=W THEN B=BW'	0,02613	0,34833
'IF A=W THEN B=W'	0,02313	0,30833
'IF A=W THEN B=S'	0,01838	0,24500
'IF A=W THEN B=N'	0,00613	0,08167
'IF A=W THEN B=BN'	0,00125	0,01667
'IF A=BW THEN B=BW'	0,03263	0,65250
'IF A=BW THEN B=W'	0,01450	0,29000
'IF A=BW THEN B=N'	0,00013	0,00250
'IF A=BW THEN B=BN'	0	0

Źródło: Opracowanie własne na podstawie obliczeń

Dalej przeprowadzono analizę sprawdzającą, jaki wpływ ma ilość przedsiębiorstw aktywnych i przedsiębiorstw nowo powstałych na kwoty inwestycji w MSP. Wyniki były analizowane w odniesieniu do przeciętnego województwa (dane odnośnie poszczególnych województw zostały uśrednione). Przy bardzo niskim poziomie liczby aktywnych przedsiębiorstw nawet wysoki przyrost nowych podmiotów nie powoduje wzrostu wolumenu inwestycji. Podobnie, przy bardzo niewielkiej liczbie nowo powstałych podmiotów i nawet średnim poziomie aktywnych podmiotów już istniejących, prawdopodobieństwo wysokiego lub bardzo wysokiego poziomu inwestycji jest równe zero (wyniki w tabeli 4). Można także zauważyć, że bardzo wysoki poziom analizowanych wskaźników sprzyja wysokiej inwestycji. Prawdopodobieństwo wysokiego i bardzo wysokiego poziomu inwestycji wynosi 0,7, dodając do tego poziom średni uzyskujemy prawdopodobieństwo na poziomie 0,9375 - czyli potwierdzenie tezy, że są to korzystne warunki do przeprowadzania inwestycji.

Tab.4. Wybrane lingwistyczne reguły asocjacji dla badania liczby aktywnych przedsiębiorstw (A), liczby nowych przedsiębiorstw (B) i kwot inwestycji w MSP (C).

If (condition) Then (association)	Wsparcie	Ufność
'IF A=BN AND B=BN THEN C=BN'	0,29505	0,89783
'IF A=BN AND B=BN THEN C=N'	0,03358	0,10217
'IF A=BN AND B=BN THEN C=S'	0	0
'IF A=BN AND B=BN THEN C=W'	0	0
'IF A=BN AND B=BN THEN C=BW'	0	0
'IF A=BN AND B=N THEN C=BN'	0,18165	0,87807
'IF A=BN AND B=N THEN C=N'	0,02520	0,12181
'IF A=BN AND B=N THEN C=S'	0,00003	0,00012
'IF A=BN AND B=N THEN C=W'	0	0
'IF A=BN AND B=N THEN C=BW'	0	0
'IF A=BN AND B=S THEN C=W'	0	0
'IF A=BN AND B=S THEN C=BW'	0	0
'IF A=BN AND B=W THEN C=W'	0	0
'IF A=BN AND B=W THEN C=BW'	0	0
'IF A=BN AND B=BN THEN C=S'	0	0
'IF A=N AND B=S THEN C=W'	0	0
'IF A=BN AND B=BW THEN C=W'	0	0
'IF A=BN AND B=BW THEN C=BW'	0	0
'IF A=N AND B=S THEN C=BW'	0	0
'IF A=S AND B=BN THEN C=W'	0	0
'IF A=S AND B=BN THEN C=BW'	0	0
'IF A=BW AND B=BN THEN C=W'	0	0
'IF A=BW AND B=BN THEN C=BW'	0	0
'IF A=BW AND B=BW THEN C=BW'	0,00776	0,43125
'IF A=BW AND B=BW THEN C=W'	0,00484	0,26875
'IF A=BW AND B=BW THEN C=S'	0,00428	0,23750

Źródło: Opracowanie własne na podstawie obliczeń

Kolejna analiza pokazuje wpływ liczby przedsiębiorstw aktywnych i liczbę nowych przedsiębiorstw na wielkość nakładów przeznaczonych na innowacje. Podobnie jak poprzednio zauważono, że aby znacząco wpłynąć na nakłady na innowacje obie badane wielkości muszą wykazywać podobne wartości. Dla przykładu, jeśli liczba przedsiębiorstw aktywnych i liczba nowych przedsiębiorstw jest bardzo wysoka, wysoce prawdopodobne będą wysokie lub bardzo wysokie nakłady na innowacje.

Należy zauważyć, iż największe wsparcie miały reguły o jednocześnie bardzo niskich i niskich wartościach atrybutów. Powodem takiego zjawiska jest dostosowany do wszystkich regionów przedział wartości atrybutów. Jak się okazało, większość województw miała o wiele niższe wartości danych cech w porównaniu z nielicznymi, dominującymi regionami np. województwem mazowieckim.

Przeprowadzono także analizy dla szeregów czasowych z wykorzystaniem lingwistycznych reguł asocjacji. Analiza liczby pracujących w sektorze MSP na Opolszczyźnie w latach 1997-2005 z autoregresją rzędu pierwszego wykazała niewielką dynamikę tej cechy (tab. 5). Zestawiając liczbę pracujących w województwie opolskim rok do roku zauważa się, że największe prawdopodobieństwo występuje dla zdarzenia niewielkiej zmiany poziomu zatrudnienia lub utrzymania zatrudnienia na tym samym poziomie. Wnioskuje się tutaj zbadanie podobnych zależności dla dłuższego okresu czasu.

Tab.5. Wybrane lingwistyczne reguły asocjacji dla szeregu czasowego pracujących w MSP.

If (condition) Then (association)	Wsparcie	Ufność
'X(t-1)=N THEN X(t)=W'	0,00125	0,00400
'X(t-1)=N THEN X(t)=S'	0	0
'X(t-1)=N THEN X(t)=N'	0,10500	0,33600
'X(t-1)=N THEN X(t)=BW'	0,01125	0,03600
'X(t-1)=N THEN X(t)=BN'	0,19500	0,62400
'X(t-1)=BN THEN X(t)=W'	0,01125	0,03600
'X(t-1)=BN THEN X(t)=S'	0	0
'X(t-1)=BN THEN X(t)=N'	0,09500	0,30400
'X(t-1)=BN THEN X(t)=BW'	0,10125	0,32400
'X(t-1)=BN THEN X(t)=BN'	0,10500	0,33600
'X(t-1)=BW THEN X(t)=W'	0,01125	0,05000
'X(t-1)=BW THEN X(t)=S'	0,09000	0,40000
'X(t-1)=BW THEN X(t)=N'	0,02250	0,10000
'X(t-1)=BW THEN X(t)=BW'	0,10125	0,45000
'X(t-1)=BW THEN X(t)=BN'	0	0,22500

Źródło: Opracowanie własne na podstawie obliczeń

Podobnie lingwistyczne reguły asocjacji szeregu czasowego dla liczby nowych podmiotów w sektorze MSP (według rejestru REGON) na Opolszczyźnie w latach 1999-2006 z autoregresją rzędu pierwszego wykazują niewielką dynamikę tego zjawiska w relatywnie krótkich - rocznych odstępach czasu.

5. Podsumowanie

Wielowymiarowe reguły asocjacji okazują się cennym narzędziem także w analizie społeczno-ekonomicznej np. w modelowaniu tendencji rozwojowych MSP. Utworzone modele pozwalają określać współzależności cech oraz probabilistycznie przewidzieć stany badanych zjawisk. Jednakże, aby je dokładniej zamodelować, należałoby rozbudować przedstawione modele o kolejne cechy oraz zbadać dane z dłuższego horyzontu czasowego. Pozwoli to z większą dokładnością i prawdopodobieństwem przewidzieć wystąpienie zakładanych zdarzeń.

Zapisanie reguł w postaci lingwistycznej umożliwia zmniejszenie wrażliwości obliczeń, tworząc w ten sposób wyniki bardziej obiektywne i łatwiejsze w interpretacji. Jednakże, wykorzystując tą metodę mamy do czynienia z pasywną postawą badań, gdyż nie uwzględnia ona drastycznych zmian trendu oraz z tym związanej ewolucji pojęć lingwistycznych.

Literatura

1. T. Morzy: Eksploracja danych: problemy i rozwiązania. V Konferencja PLOUG, Zakopane, październik 1999.
2. M. Morzy Oracle Data Mining - odkrywanie wiedzy w dużych wolumenach danych. XI Krajowa Konferencja PLOUG 2005, Zakopane, październik 18-21, 2005.
3. Agrawal R., Imielinski T., Swami A.: Mining association rules between sets of items in large databases. ACM SIGMOD International Conference on Management of Data, Washington D.C., May 26-28 1993, pp. 207-216.
4. Agrawal R., Srikant R.: Mining Generalized Association Rules, In Proc. of the 21st Int'l Conference on Very Large Databases, Zurich, Switzerland, 1995.
5. Kawa K.: Zastosowanie reguł asocjacyjnych, pakietu Oracle Data Mining for Java do analizy koszyka zakupów w aplikacjach e-commerce. XI Konferencja PLOUG, Kościelisko, październik 18-21 2005.
6. Leśniewska A., Morzy M., Morzy T.: Eksploracja danych. Materiały studiów informatycznych, http://wazniak.mimuw.edu.pl/index.php?title=Eksploracja_danych.
7. Zadeh L.A.: Fuzzy set. Information and Control. 1965, vol. 8, pp 338-353.
8. Kacprzyk J.: Zbiory rozmyte w analizie systemowej. Państwowe Wydawnictwo Naukowe, Warszawa, 1986.
9. Kacprzyk J.: Wieloetapowe sterowanie rozmyte. Wydawnictwo Naukowo-Techniczne, Warszawa, 2001.
10. Zadeh L.A.: Probability measures of fuzzy events. Journal of Mathematical Analysis and Applications, Vol. 23 No. 2, August 1968, pp. 421-427.
11. Walaszek-Babiszewska A.: Measurements and expert knowledge for time-dependent stochastic systems. The 5th International Conference, IPMM-2005, Monterey, California, USA, July 19-23, 2005.
12. Chmiel J., Skrzek – Lubasińska M., Urbańska – Jobda B.: Stan sektora MSP w 2003 roku tendencje rozwojowe w latach 1994 – 2003. Warszawa, 2005.
13. Ostasiewicz S., Rusnak Z., Siedlecka U.: Statystyka. Elementy teorii i zadania. Wyd. Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław, 1999.

WIELOWYMIAROWE REGUŁY ASOCJACJI W MODELOWANIU TENDENCJI ROZWOJOWYCH MSP

Streszczenie

W artykule została opisana jedną z metod eksploracji danych - reguły asocjacji. Zaproponowano, poprzez włączenie logiki rozmytej i wprowadzenie języka quasi-naturalnego, utworzenie lingwistycznych reguł asocjacji. Następnie, podano sposób ich wyliczania oraz podjęto próby wykorzystania metody do analizy tendencji rozwojowych małych i średnich przedsiębiorstw. Analiza została przeprowadzona w układzie regionalnym, ze szczególnym uwzględnieniem województwa opolskiego.