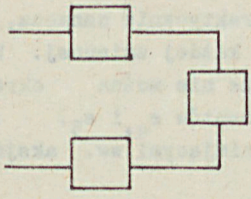
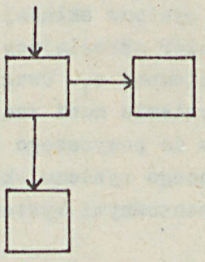
Iwona Roeske-Słomka

IDENTYFIKACJA STRUKTURY SYSTEMU W OPARCIU O DANE EMPIRYCZNE

System definiujemy jako pewną całość złożoną z powiązanych elementów. Powiązania występujące między elementami nazywamy sprzężeniami. Prostą ilustracją tak określonego systemu są rysunki 1 i 2, na których prostokąty oznaczają elementy, a ich połączenia, sieć sprzężeń. Jeżeli kierunek tych sprzężeń jest znany - rys. 2, to wtedy mówimy, że mamy do czynienia z systemem skierowanym, tzn. systemem dla którego znany jest podział stanów elementów na tzw. wejścia kształtowane przez otoczenie systemu oraz wyjścia, będące wynikiem wzajemnego oddziaływania elementów na siebie, wewnątrz systemu. Systemy o nieznanym kierunku sprzężeń - rys. 1, nazywamy neutralnymi.



Rys. 1



Rys. 2

Prezentowana w niniejszym artykule metoda, pozwalająca określić kierunek sprzężeń w danym systemie neutralnym, generalnie sprowadza się do sporządzenia listy tzw. kandydujących systemów skierowanych i wyboru jednego z nich jako „najlepszej reprezentacji" badanego systemu neutralnego, w oparciu o różnice między prawdopodobieństwami teoretycznymi a empiry-

cznymi tzw. próbek działania otrzymanych za pomocą tzw. efektywnej maski próbkowania[1].

W procesie badania empirycznego, element przedstawia się przez odpowiednie podzbiory danych, zgodnie z wybraną maską próbkowania. Kandydujące systemy skierowane tworzy się natomiast poprzez określenie wszystkich par, trójek, czwórek itd., różnych elementów tak, by spełnione były następujące warunki, zwane aksjomatami struktur kandydujących, stanowiących o sensowności danego kandydującego systemu skierowanego[2]:

1. zbiór obejmujący wszystkie zmienne w kandydującym systemie skierowanym jest równy zbiorowi zmiennych w macierzy działania opisującej dany system neutralny,
2. każdy element kandydującego systemu skierowanego jest określony przez niepusty podzbiór zmiennych z macierzy działania,
3. żaden element danego systemu skierowanego nie może być określony przez podzbiór będący zbiorem określającym inny element tego samego kandydującego systemu skierowanego,
4. podzbiory zmiennych z macierzy działania, wchodzące w skład poszczególnych elementów kandydującego systemu skierowanego, wyznacza maska próbkowania określająca zachowanie generatywne systemu.

Jeżeli na przykład, zbiór danych składa się z trzech zmiennych x,y,z, to (przy jednokolumnowej masce próbkowania) można wyodrębnić następujące elementy:

$$e_1 = (x,y) \qquad e_4 = (x) \qquad e_7 = (x,y,z)$$
$$e_2 = (x,z) \qquad e_5 = (y)$$
$$e_3 = (y,z) \qquad e_6 = (z)$$

Dwa pierwsze z wymienionych aksjomatów są oczywiste. Jeżeli natomiast chodzi o aksjomat 3, to w naszym przykładzie oznacza on, że nie można utworzyć systemu składającego się ze sprzężenia elementów np. e_1e_4, ponieważ zbiór określający element e_4 jest jednocześnie podzbiorem określającym element e_1. Uwzględnienie aksjomatu 4, praktycznie oznacza, że w każdym systemie musi znajdować się reprezentacja każdej zmiennej. W odniesieniu do powyższego przykładu implikuje to, że nie można określić kandydującego systemu składającego się np. z elementów e_4 i e_5.

Stąd sensownymi systemami kandydującymi, spełniającymi ww. aksjomaty są:

$$\mathcal{C}_1 = e_1e_2 \qquad \mathcal{C}_5 = e_2e_5$$
$$\mathcal{C}_2 = e_1e_3 \qquad \mathcal{C}_6 = e_3e_4$$
$$\mathcal{C}_3 = e_2e_3 \qquad \mathcal{C}_7 = e_7$$
$$\mathcal{C}_4 = e_1e_6 \qquad \mathcal{C}_0 = e_4e_5e_6$$

---

[1] I.Roeske-Słomka, Propozycja sposobu określania zachowania systemu opisanego wynikami obserwacji statystycznej, Prace Instytutu Cybernetyki Ekonomicznej, Zeszyty Naukowe AE w Poznaniu, Seria I, z. 138, s. 68-78.

[2] G.J.Klir, H.J.J.Uyttenhove, Computerized Methodology for Structure Modelling, w: Annals of Systems Research 1976, vol. 5, s. 39 i następne.

Jeżeli zaś np. dla trzech zmiennych x, y, z, dowolnej macierzy działania, maska związana jest z następującymi miejscami zmiennych próbkowania:

$$s_1 = x_t$$
$$s_2 = x_{t+1}$$
$$s_3 = y_{t+1}$$
$$s_4 = z_t$$
$$s_5 = z_{t+1}$$

to elementy sensownych systemów skierowanych, określone są przez następujące zbiory zmiennych próbkowania:

$e_1 = (s_1 s_2 s_3)$ $\quad$ $e_5 = (s_3)$
$e_2 = (s_1 s_2 s_4 s_5)$ $\quad$ $e_6 = (s_4 s_5)$
$e_3 = (s_3 s_4 s_5)$ $\quad$ $e_0 = (s_1 s_2 s_3 s_4 s_5)$
$e_4 = (s_1 s_2)$

Przy czterech zmiennych charakteryzujących system, x,y,z,w, (dla regularnej maski próbkowania) otrzymujemy następującą listę elementów:

$e_1 = x$ $\quad$ $e_5 = xy$ $\quad$ $e_{11} = xyz$ $\quad$ $e_{15} = xyzw$
$e_2 = y$ $\quad$ $e_6 = xz$ $\quad$ $e_{12} = xyw$
$e_3 = z$ $\quad$ $e_7 = xw$ $\quad$ $e_{13} = xzw$
$e_4 = w$ $\quad$ $e_8 = yz$ $\quad$ $e_{14} = yzw$
$e_9 = yw$
$e_{10} = zw$

w oparciu o którą sporządzamy wykaz tzw. sensownych systemów skierowanych - tabela 1.

Tabela 1

Systemy kandydujące do reprezentacji systemu neutralnego opisanego macierzą działania utworzoną przez cztery zmienne

Numer systemu	Element	Numer systemu	Element
0	1, 2, 3, 4	14	5, 13
1	1, 14	15	5, 14
2	2, 13	16	6, 12
3	3, 12	17	6, 14
4	4, 11	18	7, 11
5	5, 10	19	7, 14
6	6, 9	20	8, 12
7	7, 8	21	8, 13
8	11, 12	22	9, 11
9	11, 13	23	9, 13
10	11, 14	24	10, 11
11	12, 13	25	10, 12
12	12, 14	26	15
13	13, 14		

Tzw. minimalnym spośród sensownych systemów jest ten, którego liczba elementów jest równa liczbie zmiennych. Należy zaznaczyć, że każdy z jego elementów jest określony przez jedną zmienną i nie ma żadnych sprzężeń między elementami. W rozpatrywanych przykładach są to systemy oznaczone $\mathcal{C}_0$.

Tzw. maksymalnym systemem jest ten, który składa się z jednego elementu określonego przez wszystkie zmienne. W przykładzie dotyczącym czterech zmiennych, jest to system $\mathcal{C}_{26}$, natomiast w przykładzie z trzema zmiennymi system $\mathcal{C}_7$.

Liczba sensownych systemów strukturalnych rośnie bardzo szybko wraz ze wzrostem liczby zmiennych. Jednakże nie zawsze występuje potrzeba dokonywania ocen ich wszystkich, ze względu na możliwość wyodrębnienia tzw. szczególnych przypadków systemów ogólniejszych.

Gdy m zmiennych określa jakiś element systemu, to mówimy, że są one bezpośrednio powiązane. Jeżeli zatem $\mathcal{C}_i$ i $\mathcal{C}_j$ wchodzą w skład zbioru sensownych systemów skierowanych, to $\mathcal{C}_i < \mathcal{C}_j$ oznacza, że $\mathcal{C}_i$ jest przypadkiem szczególnym bardziej ogólnego systemu $\mathcal{C}_j$. Przypadek ten występuje wówczas, gdy zbiory zmiennych bezpośrednio powiązanych (tzn. wchodzące w skład jednego elementu) w $\mathcal{C}_i$ są podzbiorami zmiennych bezpośrednio powiązanych w $\mathcal{C}_j$.

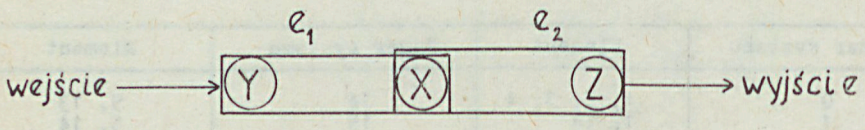
W oparciu o powyższe stwierdzenia podaje się, że liczba kandydujących systemów, przy n zmiennych wynosi[3]:

$$K = \frac{n(n-1)}{2} . \qquad (1)$$

Ze zbioru tzw. sensownych systemów skierowanych dla 3 zmiennych, rozpatrzymy następujące:

$$\mathcal{C}_1 = \{e_1\ e_2\} = \{(xy)\ (xz)\}$$
$$\mathcal{C}_2 = \{e_1\ e_3\} = \{(xy)\ (yz)\}$$
$$\mathcal{C}_3 = \{e_2\ e_3\} = \{(xz)\ (yz)\}$$

W pierwszym systemie $\mathcal{C}_1$, zmienną x nazywamy zmienną sprzężenia. Prezentacją graficzną takiego systemu, przy zaobserwowanych stanach zmiennych (kółka) może być rys. 3.

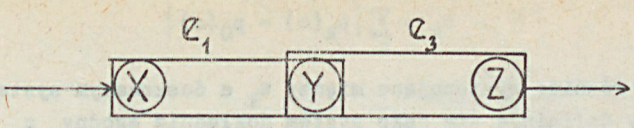


Rys. 3

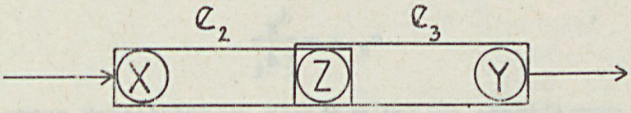
Odpowiednio w systemie $\mathcal{C}_2$ zmienną sprzężenia jest zmienna y - rys. 4, zaś w systemie $\mathcal{C}_3$, zmienna z - rys. 5.

---

[3] G.J.Klir, H.J.J.Uyttenhove, Computerized Methodology for ..., s.54.



Rys. 4



Rys. 5

Przypomnijmy, iż zadanie polega na tym, że spośród kandydujących systemów skierowanych, należy wybrać ten, który najmniej różni się od generatywnego zachowania badanego systemu, zawartego w danej macierzy działania i wykrytego przy pomocy tzw. efektywnej maski próbkowania. W tym celu przedmiot oceny mogą stanowić odpowiednie macierze prawdopodobieństw. Każdy bowiem z kandydujących systemów skierowanych, można przedstawić poprzez wykonanie złożenia opisanego w macierzy sprzężeń, tj. przyjęcie określonego skierowania (np. dla systemu $\mathcal{C}_1$ założone jest skierowanie od zmiennej y do x do z, co jest jednoznaczne z przyjęciem przeciwnego kierunku skierowania, tj. od zmiennej z do x do y[4] oraz porównania uzyskanych prawdopodobieństw z prawdopodobieństwami empirycznymi.

Biorąc pod uwagę, że prawdopodobieństwo dwu zdarzeń równe jest iloczynowi odpowiedniego prawdopodobieństwa warunkowego i prawdopodobieństwa jednego z nich oraz uwzględniając założone np. w systemie $\mathcal{C}_1$ skierowanie zmiennych (sprzężenie elementów), otrzymamy prawdopodobieństwa teoretyczne w wyniku przemnożenia następującego wektora i macierzy prawdopodobieństw, których wartości oblicza się w oparciu o prawdopodobieństwa empiryczne:

$$p_{yxz} = p_{yx} \cdot p_{z/x} \cdot \qquad (2)$$

Jeżeli $p_k(c)$ oznacza prawdopodobieństwo próbki c, obliczone dla kandydującego systemu skierowanego k, a $p_0(c)$, prawdopodobieństwo empiryczne tej samej próbki, wtedy:

---

[4] System zakładający skierowanie: $y \rightarrow x \rightarrow z$ formalnie jest identyczny z systemem o skierowaniu zmiennych: $z \rightarrow x \rightarrow y$ ponieważ:

$$p_{y \rightarrow x \rightarrow z} = p_{yx} \cdot p_{z/x} = p_{yx} \cdot \frac{p_{zx}}{p_x}$$

a:

$$p_{z \rightarrow x \rightarrow y} = p_{zx} \cdot p_{y/x} = p_{zx} \cdot \frac{p_{yx}}{p_x}$$

które to równości są jednoznaczne. Podobne jednoznaczności występują między systemami o skierowaniu zmiennych: $x \rightarrow y \rightarrow z$ i $z \rightarrow y \rightarrow x$ oraz $x \rightarrow z \rightarrow y$ i $y \rightarrow z \rightarrow x$. Por. także obliczone dalej prawdopodobieństwa teoretyczne próbek, dla systemów skierowanych opisanych przez założoną macierz działania.

$$d_k = \sum_c \left| p_k(c) - p_0(c) \right| \qquad (3)$$

wskazuje na różnice występujące między e_k a doskonałym systemem kandydującym, który definiuje się jako system dokładnie zgodny z generatywnym zachowaniem badanego systemu. Wielkość:

$$a_k = \frac{d_k}{\sum_i d_i} \qquad (4)$$

ustaloną dla określonego zbioru systemów kandydujących nazywa się stopniem rozbieżności[5].

Proponowaną metodę można stosować zarówno do zmiennych występujących w postaci jawnej jak i niejawnej o dowolnej skali. Jedynie zgodnie z naturą danych empirycznych, przyjmuje się, że są one dyskretne.

Załóżmy, że dowolny system charakteryzuje poniższa macierz, zwana macierzą działania, którą tworzą odnotowane w wyniku obserwacji statystycznej stany trzech zmiennych: x, y, z:

zmienna \ Nr obser.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x	0	0	1	0	0	0	1	1	0	1	0	0	0	1	1	0
y	2	1	0	0	1	2	1	1	0	2	2	1	0	0	0	1
z	0	0	1	0	0	1	1	1	0	0	1	0	0	1	1	1

Pokażemy sposób obliczania prawdopodobieństw empirycznych i teoretycznych próbek oraz odległości d, na przykładzie próbek działania otrzymanych w wyniku użycia maski jednokolumnowej do powyższej macierzy działania. W tym przypadku próbki działania c są następujące:

c xyz
020
010
101
000
010
021
111
111
000
120
021
010
000
101
101
011

a ich prawdopodobieństwa empiryczne wynoszą:

$$_e p_{020} = 1/16 = 0{,}0625$$
$$_e p_{010} = 3/16 = 0{,}1875$$
$$_e p_{101} = 3/16 = 0{,}1875$$
$$_e p_{000} = 3/16 = 0{,}1875$$
$$_e p_{021} = 2/16 = 0{,}1250$$
$$_e p_{111} = 2/16 = 0{,}1250$$
$$_e p_{120} = 1/16 = 0{,}0625$$
$$_e p_{011} = 1/16 = 0{,}0625$$

[5] G.J.Klir, Identyfikacja struktur generatywnych w danych empirycznych, w: Projektowanie i systemy, zagadnienia metodologiczne, PAN, 1978, t. 1, s. 31.

natomiast prawdopodobieństwa teoretyczne tych próbek, tj. zakładających skierowanie zmiennych właściwe np. dla systemu $\mathcal{C}_1$ (zgodnie z regułą:

$$p_{xyz} = p_{xy} \cdot p_{z/x} \quad \text{gdzie } p_{z/x} = \frac{p_{zx}}{p_x})$$

są równe:

$$
\begin{aligned}
{}^{t}p_{020} &= p_{02} \cdot p_{0/0} = 3/16 \cdot 7/16 \cdot 16/10 = 0{,}13125 \\
{}^{t}p_{010} &= p_{01} \cdot p_{0/0} = 4/16 \cdot 7/16 \cdot 16/10 = 0{,}175 \\
{}^{t}p_{101} &= p_{10} \cdot p_{1/1} = 3/16 \cdot 5/16 \cdot 16/6 = 0{,}15625 \\
{}^{t}p_{000} &= p_{00} \cdot p_{0/0} = 3/16 \cdot 7/16 \cdot 16/10 = 0{,}13125 \\
{}^{t}p_{021} &= p_{02} \cdot p_{1/0} = 3/16 \cdot 2/16 \cdot 16/10 = 0{,}0375 \\
{}^{t}p_{111} &= p_{11} \cdot p_{1/1} = 2/16 \cdot 5/16 \cdot 16/6 = 0{,}10417 \\
{}^{t}p_{120} &= p_{12} \cdot p_{0/1} = 1/16 \cdot 1/16 \cdot 16/6 = 0{,}01042 \\
{}^{t}p_{011} &= p_{01} \cdot p_{1/0} = 4/16 \cdot 2/16 \cdot 16/10 = 0{,}05
\end{aligned}
$$

Zatem dla $\mathcal{C}_1$, d zgodnie z formułą (3) wynosi:

dla c = 020 : 0,6875
010 : 0,0125
101 : 0,03125
000 : 0,05625
021 : 0,0875
111 : 0,02083
120 : 0,05208
011 : 0,0125

0,34166 = d_1

Dla systemu $\mathcal{C}_2$ i $\mathcal{C}_3$ przy obliczaniu prawdopodobieństw teoretycznych (podobnie jak w obliczeniach dotyczących $\mathcal{C}_1$) należałoby uwzględnić odpowiedni kierunek skierowania zmiennych.

Należy zauważyć, że przy użyciu tzw. maski efektywnej tj. takiej, która wykrywa generatywne zachowanie systemu, prawdopodobieństwo empiryczne każdej z próbek działania c, jest jednakowe. Tak więc przy zastosowaniu maski trójkolumnowej, która okazała się efektywną dla założonej macierzy działania[6], otrzymujemy następujące próbki działania c:

Lp. próbki	c	x	y	z
1		001	210	001
2		010	100	010
3		100	001	100
4		000	112	001

[6] Por. I.Roeske-Słomka, Propozycja sposobu...

Lp. próbki c	x	y	z
5	001	121	011
6	011	211	111
7	110	110	110
8	101	102	100
9	010	022	001
10	100	221	010
11	000	210	100
12	001	100	001
13	011	000	011
14	110	001	111

Prawdopodobieństwo występowania każdej z powyższych próbek, zwane prawdopodobieństwem empirycznym próbki działania c, wynosi oczywiście 1/14 = = 0,07143.

Ponieważ stwierdzono, iż dla danego systemu neutralnego, którego działanie zostało opisane w wyniku obserwacji trzech zmiennych, kandydującymi są trzy systemy $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ o kierunku skierowania takim jak podano poprzednio, przeto prawdopodobieństwa teoretyczne kolejnych próbek działania dla tychże systemów, oblicza się według wzorów:

$$\mathcal{C}_1\ {}_t p = p_{xy} \cdot p_{xz} \cdot \frac{1}{p_x}$$

$$\mathcal{C}_2\ {}_t p = p_{xy} \cdot p_{yz} \cdot \frac{1}{p_y}$$

$$\mathcal{C}_3\ {}_t p = p_{xz} \cdot p_{zy} \cdot \frac{1}{p_z}$$

Na przykład, prawdopodobieństwo teoretyczne pierwszej próbki działania w systemie $\mathcal{C}_1$ zakładającym, że zmienna x jest tzw. zmienną sprzężenia, liczymy następująco:

$$\mathcal{C}_{1t}p_1 = p_{xy} \cdot p_{xz} \cdot 1/p_x = p_{xy=020110} \cdot p_{xz=000011} \cdot 1/p_{x=001} =$$

$$= 1/14 \cdot 2/14 \cdot 14/3 = 0{,}0476$$

lub

$$\mathcal{C}_{1t}p_1 = p_{xy=001210} \cdot p_{xz=210001} \cdot 1/p_{x=001} = 1/14 \cdot 2/14 \cdot 14/3 = 0{,}0476$$

czwartej próbki działania: xyzxyzxyz 000010021 czyli: xxxyyyzzz 000012001:

$$\mathcal{C}_{1t}p_4 = p_{xy}\ p_{xz} \cdot 1/p_x = p_{xy=000102} \cdot p_{yz=001021} \cdot 1/p_{x=000} =$$

$$= 1/14 \cdot 1/14 \cdot 14/2 = 0{,}03571$$

lub

$$\mathcal{C}_{1t}p_4 = p_{xy=000012} \cdot p_{yz=012001} \cdot 1/p_{x=000} = 1/14 \cdot 1/14 \cdot 14/2 = 0{,}03571$$

dziewiątej próbki działania w systemie ϱ_2:

$$\varrho_{2t}p_9 = p_{xy} \cdot p_{yz} \cdot 1/p_y = p_{xy=001202} \cdot p_{yz=002021} \cdot 1/p_{y=022} =$$

$$= 1/14 \cdot 1/14 \cdot 14/1 = 0{,}07143$$

lub:

$$\varrho_{2t}p_9 = p_{xy=010022} \cdot p_{yz=022001} \cdot 1/p_{y=022} = 1/14 \cdot 1/14 \cdot 14/1 = 0{,}07143.$$

Prawdopodobieństwa teoretyczne próbek, obliczone według zaprezentowanego sposobu, dla każdego z trzech systemów skierowanych oraz ich różnice w stosunku do prawdopodobieństw empirycznych próbek, dla przykładowej macierzy działania, podaje tabela 2.

Tabela 2

Prawdopodobieństwa teoretyczne oraz ich odchylenia od prawdopodobieństw empirycznych próbek działania wyznaczonych przez maskę efektywną (dla założonej macierzy działania)

Lp. próbki	Prawdopodobieństwa teoretyczne próbki działania w systemie:			Bezwzględne różnice prawdopodobieństwa empirycznego i teoretycznego próbek działania w systemie:		
	ϱ_1	ϱ_2	ϱ_3	ϱ_1	ϱ_2	ϱ_3
1	0,04762	0,03571	0,03571	0,02381	0,03571	0,03571
2	0,03571	0,03571	0,03571	0,03571	0,03571	0,03571
3	0,03571	0,03571	0,02381	0,03571	0,03571	0,04762
4	0,03571	0,07143	0,01786	0,03571	0,00000	0,05357
5	0,02381	0,07143	0,03571	0,04762	0,00000	0,03571
6	0,03571	0,07143	0,03571	0,03571	0,00000	0,03571
7	0,03571	0,07143	0,07143	0,03571	0,00000	0,00000
8	0,07143	0,07143	0,02381	0,00000	0,00000	0,04762
9	0,03571	0,07143	0,01786	0,03571	0,00000	0,05357
10	0,03571	0,07143	0,03571	0,03571	0,00000	0,03571
11	0,03571	0,03571	0,02381	0,03571	0,03571	0,04762
12	0,04762	0,03571	0,03571	0,02381	0,03571	0,03571
13	0,03571	0,07143	0,03571	0,03571	0,00000	0,03571
14	0,03571	0,03571	0,03571	0,03571	0,03571	0,03571

Sumy bezwzględnych różnic prawdopodobieństw teoretycznych i empirycznych, próbek działania wynoszą w omawianym przykładzie, dla systemu:

$$\varrho_1 : 0{,}45238$$

$$\varrho_2 : 0{,}21429$$

$$\varrho_3 : 0{,}53571$$

Z powyższego wynika, że spośród kandydujących systemów skierowanych, system ϱ_2, w którym y jest zmienną sprzężenia, w najmniejszym stopniu różni się od systemu neutralnego opisanego przez założoną macierz działania, a tym samym, stanowi jego najlepszą reprezentację. W ten sposób również stwierdzono, że stany zmiennej x stanowią tzw. wejścia systemu, zaś stany zmiennej z, jego wyjścia.

Zaprezentowana metoda wymaga rozwinięcia na przypadek wykrywania skierowań systemów opisanych macierzami działań, utworzonymi w wyniku statystycznej obserwacji czterech i więcej zmiennych.

Literatura

[1] Klir G.J.: Identyfikacja struktur generatywnych w danych empirycznych, w: Projektowanie i systemy, zagadnienia metodologiczne, PAN, 1978, t. 1.

[2] Klir G.J., Uyttenhove H.J.J.: Computerized Methodology for Structure Modelling, w: Annals of Systems Research 1976, vol. 5.

[3] Roeske-Słomka I.: Propozycja sposobu określania zachowania systemu opisanego wynikami obserwacji statystycznej. Prace Instytutu Cybernetyki Ekonomicznej, Zeszyty Naukowe AE w Poznaniu, seria I, z. 138.