XV Krajowa Konferencja Automatyki

Tom II

KKA
2005

Redaktorzy:
Zdzisław Bubnicki
Roman Kulikowski
Janusz Kacprzyk

XV Krajowa Konferencja Automatyki
Tom II

Redaktorzy:
Zdzisław BUBNICKI
Roman KULIKOWSKI
Janusz KACPRZYK

ORGANIZATOR
Komitet Automatyki i Robotyki Polskiej Akademii Nauk
Instytut Badań Systemowych Polskiej Akademii Nauk

WSPÓŁORGANIZATORZY
Politechnika Warszawska
Przemysłowy Instytut Automatyki i Pomiarów
Polskie Stowarzyszenie Pomiarów, Automatyki i Robotyki

ORGANIZATOR

Komitet Automatyki i Robotyki Polskiej Akademii Nauk
Instytut Badań Systemowych Polskiej Akademii Nauk

WSPÓŁORGANIZATORZY

Politechnika Warszawska
Przemysłowy Instytut Automatyki i Pomiarów
Polskie Stowarzyszenie Pomiarów, Automatyki i Robotyki

KOMITET PROGRAMOWY

Przewodniczący	Zdzisław BUBNICKI
Zastępca Przewodniczącego	Roman KULIKOWSKI

CZŁONKOWIE

Stanisław BAŃKA	Michał BIAŁKO
Mikołaj BUSŁOWICZ	Władysław FINDEISEN
Ryszard GESSING	Henryk GÓRECKI
Jakub GUTENBAUM	Jerzy JÓZEFCZYK
Stanisław KACZANOWSKI	Tadeusz KACZOREK
Janusz KACPRZYK	Jerzy KLAMKA
Józef KORBICZ	Zbigniew KOWALSKI
Krzysztof KOZŁOWSKI	Juliusz L. KULIKOWSKI
Krzysztof KUŹMIŃSKI	Kazimierz MALANOWSKI
Krzysztof MALINOWSKI	Wojciech MITKOWSKI
Antoni NIEDERLIŃSKI	Władysław PEŁCZEWSKI
Tadeusz PUCHAŁKA	Leszek RUTKOWSKI
Stanisław SKOCZOWSKI	Roman SŁOWIŃSKI
Jerzy ŚWIĄTEK	Andrzej ŚWIERNIAK
Ryszard TADEUSIEWICZ	Piotr TATJEWSKI
Krzysztof TCHOŃ	Leszek TRYBUS
Jan WĘGLARZ	Andrzej P. WIERZBICKI

KOMITET ORGANIZACYJNY

Przewodniczący	Roman KULIKOWSKI
Zastępcy Przewodniczącego	Janusz KACPRZYK
	Stanisław KACZANOWSKI
	Tadeusz KACZOREK
	Krzysztof MALINOWSKI
Członkowie	Roman OSTROWSKI
	Tadeusz PUCHAŁKA
	Dariusz WAGNER
Sekretarze naukowi	Jan STUDZIŃSKI
	Jan W. OWSIŃSKI

ISBN 83-89475-01-4

Druk: ARGRAF, Warszawa

METODY STOCHASTYCZNE
– PROBLEMY NIEDETERMINISTYCZNE

WYBÓR ZMIENNYCH W ANALIZIE SYSTEMOWEJ - ZASTOSOWANIE STATYSTYCZNYCH METOD BADANIA ZALEŻNOŚCI Z WYKORZYSTANIEM NIEPRECYZYJNYCH DANYCH

Olgierd HRYNIEWICZ

Instytut Badań Systemowych PAN
ul. Newelska 6, 01-447 Warszwa, e-mail: hryniewi@ibspan.waw.pl

Streszczenie: W artykule zaproponowano metodologię doboru zmiennych w modelach opisujących złożone zagadnienia analizy systemowej. Wybór zmiennych dokonywany jest na podstawie statystycznej analizy wskazań ekspertów uzyskanych w wyniku realizacji specjalnie zaprojektowanego eksperymentu z ich udziałem. Dopuszczona jest możliwość nieprecyzyjnych (rozmytych) odpowiedzi ekspertów na postawione pytania.

Słowa kluczowe: Analiza systemowa, dobór zmiennych w modelach, rozmyte opinie ekspertów, statystyczna ocena zależności.

1. WSTĘP

Podstawowym zadaniem analizy systemowej jest dostarczenie decydentowi propozycji rozwiązań skomplikowanych problemów decyzyjnych. Propozycje te mogą mieć bardzo różna postać: od zestawów reguł postaci "jeżeli to " do rozwiązań skomplikowanych zadań optymalizacyjnych. W niniejszym artykule ograniczymy się do sytuacji, gdy na pewnym etapie przeprowadzonej analizy konieczne jest zaproponowanie modelu matematycznego wiążącego jakieś zmienne wyjściowe (decyzyjne) z pewnymi zmiennymi wejściowymi (objaśniającymi).

Na temat budowy i identyfikacji modeli matematycznych napisano setki książek i tysiące artykułów. W literaturze polskiej znane są klasyczne podręczniki Bubnickiego [1] oraz Mańczaka i Nahorskiego [12] poświęcone zagadnieniom identyfikacji systemów opisywanych odpowiednimi modelami matematycznymi. Z nowszych podręczników odnoszących się do zagadnień identyfikacji należy wymienić np. książkę Söderströma i Stoici [13]. Istnieje również bogata literatura statystyczna poświęcona tworzeniu modeli regresyjnych opisujących związki zmiennych decyzyjnych (objaśnianych) ze zmiennymi objaśniającymi. We wszystkich tych publikacjach przyjęto następujący tok postępowania: na podstawie *istniejących* danych statystycznych należy dobrać taki model opisujący związki pomiędzy rozpatrywanymi zmiennymi, by spełnione były pewne warunki optymalności. Na podobnej zasadzie działają rozwijające się ostatnio intensywnie metody komputerowej eksploracji danych (popularny data mining) wykorzystujące nowoczesne metody identyfikacji i klasyfikacji.

Przedmiotem zainteresowania analizy systemowej są, zazwyczaj, bardzo złożone systemy gospodarcze, techniczne i społeczne. Mogą być one opisane olbrzymia liczbą zmiennych, których rola w objaśnianiu rozpatrywanych zjawisk może być dalece nieoczywista. Jest ponadto regułą, że dane o interesujących nas zjawiskach i zależnościach są zazwyczaj skąpe i muszą być dopiero uzyskiwane w trakcie przeprowadzanych badań. Mamy więc do czynienia z sytuacją, gdy brakuje danych umożliwiających efektywne wykorzystanie metod budowy i identyfikacji modeli matematycznych znanych z literatury przedmiotu. Istotnym jest również to, że zdobycie tych danych wiąże się zazwyczaj z koniecznością poniesienia dużych kosztów. Należy więc bardzo starannie zaprojektować proces pozyskiwania danych, które w przyszłości będą wykorzystywane do budowy odpowiednich modeli matematycznych. Pierwszym etapem projektowania takiego procesu powinno być - oczywiście – wskazanie interesujących nas zmiennych. O ile wybór zmiennych decyzyjnych nie stanowi zazwyczaj problemu, to wybór zmiennych objaśniających (czasami nazywanymi – niezbyt szczęśliwie – zmiennymi niezależnymi modelu) nie jest oczywisty. Po pierwsze, powinny to być zmienne mocno powiązane ze zmiennymi decyzyjnymi. Po drugie, w celu uniknięcia kłopotów z identyfikacja modelu, należy wybrać takie zmienne objaśniające, które byłyby w dużym stopniu wzajemnie niezależne.

Jak z powyższego wynika, podstawowym problemem napotykanym przy wyborze zmiennych modelu jest problem pomiaru siły zależności. Należy tu podkreślić, że w zagadnieniach analizy systemowej praktycznie nie występują zależności o charakterze zdeterminowanym (tak jak w przypadku wielu zjawisk fizycznych). Mamy wiec praktycznie zawsze do czynienia z zależnościami o charakterze stochastycznym, do poszukiwania, któ-

rych musimy użyć formalnego aparatu statystyki matematycznej.
Istnieje bardzo bogaty zestaw metod statystycznych wykorzystywanych do pomiaru siły zależności cech statystycznych na podstawie analizy odpowiednich danych. Powstaje jednak problem, co robić, gdy danych tych jest niewiele lub należy je dopiero uzyskać. W takich przypadkach zalecane jest wykorzystywanie opinii ekspertów. Opinie ekspertów są często jedynym źródłem informacji pozwalającym na dokonanie wstępnego wyboru zmiennych modelu. Są to jednak informacje w wysokim stopniu subiektywne, a często wzajemnie sprzeczne lub niespójne. Powstaje więc problem, jak zobiektywizować analizę danych eksperckich, by w możliwie prosty sposób uzyskać informacje o zestawie zmiennych potrzebnych do budowy interesującego nas modelu matematycznego. Jest to przedmiot rozważań niniejszej pracy.

Zaproponowana w niniejszej pracy metodologia analizy danych eksperckich jest metoda statystyczną. Została ona opisana w drugiej sekcji niniejszej pracy. Przedstawiona tam propozycja wyboru zmiennych modelu zakłada, że eksperci są w stanie w sposób całkowicie jednoznaczny odpowiadać na postawione pytania. W praktyce jest to założenie trudne do spełnienia w sposób całkowicie naturalny. Przypomnijmy, ze mamy do czynienia z często niezwykle złożonymi zjawiskami i wobec tego wiedza ekspertów nie musi być całkowicie jednoznaczna. Możliwość wykorzystania nieprecyzyjnie sformułowanych ocen ekspertów omawiamy w trzeciej sekcji niniejszej pracy. Zaprezentowany zostanie model, gdy odpowiedzi ekspertów mogą mieć postać rozmytą, przedstawiona w postaci odpowiednich *rozkładów możliwości*. Dla danych tego typu opracowane zostały zaprezentowane w tej sekcji statystyczne metody badania siły zależności. Pracę kończy podsumowanie, w którym przedstawione zostaną propozycje wykorzystania zaproponowanej metodologii.

2. STATYSTYCZNA ANALIZA DANYCH EKSPERCKICH

Przyjmijmy, że interesuje nas ustalenie siły zależności pomiędzy dwiema zmiennymi X oraz Y. Może to być para składająca się ze zmiennej decyzyjnej Y oraz zmiennej objaśniającej X i w takim przypadku interesuje nas siła zależności pomiędzy tymi zmiennymi. Mogą to być również dwie zmienne objaśniające i tu interesuje nas ich wzajemna niezależność. Zakładamy ponadto, że siłę zależności pomiędzy rozpatrywanymi zmiennymi oceniamy na podstawie opinii n ekspertów. Łatwo zauważyć, że zadanie ekspertom bezpośredniego pytania o siłę takiej zależności może nie doprowadzić do zobiektywizowanej, czyli określonej w sposób liczbowy, oceny. Należy przecież pamiętać, że pytani eksperci są ekspertami dziedzinowymi, którzy nie zawsze potrafią posługiwać się miarami zależności znanymi ze statystyki i rachunku prawdopodobieństwa. Co więcej, można założyć, że mierniki takie są w ogóle niedostępne, gdyż nie ma danych statystycznych pozwalających na ich estymację. W celu uzyskania obiektywnej miary zależności proponujemy przeprowadzenie następującego eksperymentu.
Przyjmijmy, że zbiór możliwych wartości zmiennych X oraz Y możemy poddać *kategoryzacji*. Załóżmy, że cecha jakościowa X przyjmuje wartości należących do k rozłącznych kategorii oznaczonych przez x_1, x_2,...,x_k, a cecha jakościowa Y przyjmuje wartości należące do r rozłącznych kategorii oznaczonych przez y_1, y_2,...,y_r. Sposób zdefiniowania poszczególnych kategorii może być w dużej mierze dowolny. W przypadku zmiennych liczbowych mogą to być np. przylegające do siebie przedziały możliwych wartości. Mogą to być również kategorie zdefiniowane w sposób opisowy, np. "mała wartość", "średnia wartość", itd. Kolejnym etapem eksperymentu jest zadanie ekspertom pytania o następującej postaci: "Jeżeli zmienna X przyjmuje wartość z kategorii x_i , to jaka jest spodziewana kategoria zmiennej Y". Wskazana w powyższym pytaniu kategoria x_i zostaje dla każdego eksperta *wylosowana*.
Przyjmijmy teraz, że na tak postawione pytanie ekspert potrafi podać *jednoznaczną* odpowiedź. W przypadku, gdy ekspert wskazał na kategorie x_i oraz y_j jego odpowiedź może być przedstawiona w postaci tabeli

Tabela 1 Odpowiedź eksperta

$X\backslash Y$	y_1	...	y_j	...	y_r
x_1	0	...	0	...	0
...	...	...	...	...	...
x_i	0	...	1	...	0
...	...	...	...	...	...
x_k	0	...	0	...	0

Odpowiedzi wszystkich ekspertów mogą być więc podsumowane w znanej ze statystyki tablicy kontyngencji (tablicy dwudzielczej, tablicy korelacyjnej), w której poszczególnych komórkach podane są odpowiadające tym komórkom liczby wskazań ekspertów.

Tabela 2 Podsumowanie odpowiedzi ekspertów

$X\backslash Y$	y_1	...	y_j	...	y_r
x_1	n_{11}	...	n_{1j}	...	n_{1r}
...	...	...	...	...	...
x_i		...	n_{ij}	...	n_{ir}
...	...	...	...	...	...
x_k	n_{k1}	...	n_{kj}	...	n_{kr}

Wprowadźmy następującą wielkość pomocniczą:

$$n_{i.} = \sum_{j=1}^{r} n_{ij}, \; i = 1, \ldots, k \qquad (1)$$

Jest to całkowita liczba obserwacji, dla których zaobserwowana wartość cechy jakościowej X należy do i-tej kategorii. Z kolei,

$$n_{.j} = \sum_{i=1}^{k} n_{ij},\ j = 1,\ldots,r \qquad (2)$$

jest całkowitą liczbą obserwacji, dla których zaobserwowana wartość cechy jakościowej Y należy do j-tej kategorii. Jak łatwo zauważyć,

$$n = \sum_{j=1}^{r} \sum_{i=1}^{k} n_{ij} \qquad (3)$$

jest całkowitą liczbą obserwacji (licznością próby). K. Pearson zauważył, że w przypadku niezależności dwu zmiennych losowych opisujących cechy jakościowe warunkowe wartości oczekiwane liczb obserwacji znajdujących się w poszczególnych komórkach tablicy korelacyjnej (tzw. „liczebności teoretyczne") powinny być równe

$$\hat{n}_{ij} = \frac{n_{i.} n_{.j}}{n}, i = 1,\ldots k, j = 1,\ldots r \qquad (4)$$

Do weryfikacji hipotezy o niezależności rozpatrywanych zmiennych losowych zaproponowana została podana poniżej statystyka testowa, która znana jest jako statystyka chi-kwadrat Pearsona.

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{r} \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} = \sum_{i=1}^{k} \sum_{j=1}^{r} \left(\frac{n_{ij}^2}{\hat{n}_{ij}} \right) - n \qquad (5)$$

Jeżeli spełnione są następujące warunki: liczność próby n jest duża (większa od 100) oraz $\min(n_{ij})>5$, to statystyka χ^2 ma asymptotycznie rozkład *chi-kwadrat* o $(r-1)(k-1)$ stopniach swobody. Hipotezę o niezależności zmiennych losowych *odrzucamy,* gdy spełniona jest nierówność $\chi^2 > \chi^2_{(r-1)(k-1),1-\alpha}$, gdzie $\chi^2_{(r-1)(k-1),1-\alpha}$ jest kwantylem rzędu $1-\alpha$ w rozkładzie chi-kwadrat o $(r-1)(k-1)$ stopniach swobody, podanym w licznych tablicach statystycznych.
Na bazie statystyki chi-kwadrat (5) skonstruowano wskaźniki, które mogą być wykorzystywane do pomiaru siły zależności. Najbardziej znanym z nich jest współczynnik zbieżności Czuprowa określony wzorem

$$T = \sqrt{T_{xy}^2} \qquad (6)$$

gdzie

$$T_{xy}^2 = T_{yx}^2 = \frac{\chi^2}{n\sqrt{(r-1)(k-1)}} \qquad (7)$$

Jeżeli wartość statystyki T *jest bliska zeru,* to możemy mówić o *niezależności* badanych zmiennych losowych, zaś w przypadku, gdy zachodzi równość $T=1$ mamy przypadek zależności *funkcyjnej.* W statystyce używane jest również pojęcie współczynnika determinacji Czuprowa równego $100^* T_{xy}^2$, który mówi w ilu procentach zmienność zmiennej zależnej Y jest określona zmiennością zmiennej niezależnej X. Innym wskaźnikiem siły asocjacji jest współczynnik V Cramera zdefiniowany jako

$$V = \sqrt{\frac{\chi^2}{n(m-1)}} \qquad (8)$$

gdzie

$$m = \min(k, r)$$

Współczynniki te, a także inne im podobne, przyjmują wartości z przedziału [0,1] i ich większe wartości interpretowane są jako wskaźniki silniejszej zależności (związku) występującej pomiędzy analizowanymi cechami statystycznymi. Niestety, okazuje się, że nie są one dobrymi miernikami siły zależności stochastycznej.
Propozycje różnych mierników siły zależności (asocjacji) skategoryzowanych cech statystycznych pojawiały się od przeszło stu lat. Przełomowym momentem stało się jednak opublikowanie pracy Goodmana i Kruskala [3], którzy zaproponowali jednolitą metodologię podejścia do tego problemu. Praca ta, a także inne prace tej dwójki autorów legły u podstaw serii propozycji nowych mierników siły asocjacji, które zostały zaproponowane przez wielu autorów i które nie miały wielu niekorzystnych własności wskaźników bazujących na statystyce chi-kwadrat Pearsona. W polskiej literaturze statystycznej podstawowe informacje na temat tych i podobnych procedur statystycznych można znaleźć w książce Koronackiego i Mielniczuka [11]. Dla potrzeb niniejszej pracy omówimy bliżej tylko jedną z zaproponowanych przez Goodmana i Kruskala procedur, a mianowicie pomiar siły asocjacji wykorzystujący statystykę γ Goodmana-Kruskala, a także statystykę d Sommersa. Obie te statystyki można stosować w przypadku, gdy kategorie analizowanych cech statystycznych są uporządkowane.
W ogólnym przypadku analizy danych dotyczących cech statystycznych o wartościach skategoryzowanych (np. cech jakościowych) poszczególne kategorie należy traktować jako *etykiety*. Tak też są one traktowane we wspomnianym już teście niezależności chi-kwadrat Pearsona. Przyjęcie takiego założenia oznacza, że wynik analizy nie zależy od uporządkowania tych wartości w tablicy korelacyjnej. W wielu jednak przypadkach występuje naturalne uporządkowanie kategorii rozpatrywanych cech. Ma to, na przykład, miejsce w przypadku, gdy kategorie odpowiadają kolejnym przedziałom wartości analizowanej cechy. Innym przykładem są kategorie odpowiadające różnym przypadkom „intensywności" występowania analizowanej cechy. Na przykład, ekspert wypowiadając się o przewidywanych wartościach rozpatrywanej zmiennej może korzystać z nieprecyzyjnych, ale dobrze uporządkowanych ocen, takich jak: „mała wartość", „średnia wartość" oraz „duża wartość". Jak widać, w rozpatrywanym przez nas przypadku analizy ocen ekspertów przypadek uporządkowanych kategorii może występować stosunkowo często. Okazuje się, że informacja o uporządkowaniu

kategorii rozpatrywanych cech statystycznych ułatwia nam prawidłowo określić siłę asocjacji (związku) pomiędzy analizowanymi cechami.
Zanim wprowadzimy statystykę γ Goodmana-Kruskala przyjmijmy dwie definicje.

Definicja 1 (Koronacki, Mielniczuk [11])
*Mówimy, że para jednostek jest **zgodna**, jeżeli jest spełniony jeden z następujących warunków:*
(1) *ranga kategorii, jaką przyjmuje zmienna X w przypadku pierwszej jednostki w parze, jest wyższa od rangi kategorii, jaką przyjmuje zmienna X dla drugiej jednostki w parze, oraz ranga kategorii, jaką przyjmuje zmienna Y dla pierwszej jednostki w parze, jest wyższa od rangi kategorii, jaką przyjmuje zmienna Y dla drugiej jednostki w parze jednostek.*
(2) *ranga kategorii, jaką przyjmuje zmienna X w przypadku pierwszej jednostki w parze, jest niższa od rangi kategorii, jaką przyjmuje zmienna X dla drugiej jednostki w parze, oraz ranga kategorii, jaką przyjmuje zmienna Y dla pierwszej jednostki w parze, jest niższa od rangi kategorii, jaką przyjmuje zmienna Y dla drugiej jednostki w parze jednostek.*

Definicja 2 (Koronacki, Mielniczuk [11])
*Para jednostek jest **niezgodna**, jeżeli jedna z jednostek w parze ma kategorię zmiennej X wyższej rangi oraz kategorię zmiennej Y niższej rangi niż druga z jednostek pary.*

Goodman i Kruskal [3] jako miarę siły asocjacji zaproponowali odpowiednio znormalizowaną różnicę prawdopodobieństw, że losowo wybrana para jednostek jest zgodna i niezgodna. Oszacowaniem tego wskaźnika jest statystyka γ Goodmana-Kruskala zdefiniowana przy pomocy wzoru

$$\gamma = \frac{C - D}{C + D} \qquad (9)$$

gdzie C jest liczbą wszystkich par zgodnych w tablicy korelacyjnej, zaś D jest liczbą wszystkich par niezgodnych. Wielkości te można w sposób formalny zapisać w postaci

$$C = \sum_{i,j} n_{ij} C_{I;ij} \qquad (10)$$

gdzie

$$C_{I;ij} = \sum_{i'>i} \sum_{j'>j} n_{i'j'} \qquad (11)$$

oraz

$$D = \sum_{i,j} n_{ij} D_{IV;ij} \qquad (12)$$

gdzie

$$D_{IV;ij} = \sum_{i'>i} \sum_{j'<j} n_{i'j'} \qquad (13)$$

Wartości statystyki γ Goodmana-Kruskala bliskie zera świadczą o słabej zależności pomiędzy analizowanymi cechami (w idealnym przypadku niezależności wartość statystyki jest równa zeru). Z kolei, wartości bliskie jeden (minus jeden) świadczą o bardzo silnym dodatnim (ujemnym) związku pomiędzy analizowanymi cechami statystycznymi. Goodman i Kruskal [4] analizowali również asymptotyczny rozkład prawdopodobieństwa statystyki γ. Znajomość tego rozkładu pozwala wnioskować o istotności statystycznej zaobserwowanej miary siły asocjacji.
Inną popularną statystyką, zbliżoną do statystyki γ Goodmana-Kruskala, jest statystyka *d* Sommersa, występująca w wersji symetrycznej i dwu wersjach asymetrycznych.. Przy wyznaczaniu jej wartości uwzględnia się tzw. pary związane, co w pewnych przypadkach jest korzystne. Statystyka *d* Sommersa, w przypadku niesymetrycznym, gdy zmienną objaśniającą jest *X*, a zmienną objaśnianą (zależną) jest *Y*, ma postać (patrz Koronacki i Mielniczuk [11])

$$d = \frac{C - D}{n(n-1)/2 - T_x} \qquad (14)$$

gdzie

$$T_x = \sum_{i=1}^{k} n_{i.}(n_{i.} - 1)/2, \qquad (15)$$

zaś *C* oraz *D* wyznacza się, odpowiednio, ze wzorów (10) i (12). Statystyka *d* Sommersa ma podobne własności do statystyki γ Goodmana-Kruskala, tzn. jej większe wartości bezwzględne świadczą o silniejszym (dodatnim lub ujemnym) związku porównywanych cech.

3. STATYSTYCZNA ANALIZA NIEPRECYZYJNIE OKREŚLONYCH DANYCH EKSPERCKICH

W zaproponowanej przez nas metodologii określania siły zależności zmiennych modelu na podstawie wypowiedzi ekspertów przyjęliśmy podstawowe założenie, że poszczególne obserwacje są określone w sposób jednoznaczny przez podanie kategorii odpowiadającej zmiennej *X* oraz kategorii odpowiadającej zmiennej *Y*. Założenie to może (choć nie musi) być spełnione, gdy poszczególne kategorie zdefiniowane są w sposób jednoznaczny. Jednakże w wielu przypadkach poszczególne kategorie mogą być określone w sposób nieprecyzyjny, próba ich precyzyjnego określenia nie ma praktycznego sensu. Może wówczas dojść do sytuacji, gdy ekspert nie potrafi w sposób jednoznaczny wskazać na odpowiednią kategorię. W ogólnym przypadku można przyjąć, że każda obserwacja może być zaliczona do kilku kategorii z określoną miarą przynależności. Na przykład, w pracy Hryniewicza [6] założono, że kategoria zmiennej *X* określona jest w sposób jednoznaczny, natomiast kategoria zmiennej *Y* określona jest w sposób rozmyty. Każda obserwacja opisana jest parą $\left(X_q, \widetilde{Y}_q\right)$, gdzie X_q jest zaobserwowaną kategorią zmiennej *X*, zaś

$\widetilde{Y}_q = y_1 \mid \mu_{i_q 1} + y_2 \mid \mu_{i_q 2} + \ldots + y_r \mid \mu_{i_q r}, q = 1, \ldots, n$, przy czym $\mu_{i_q j} \in [0,1], j = 1, \ldots, r$ jest zaobserwowaną rozmytą kategorią zmiennej *Y*. Rozmytą wartość $\widetilde{Y}_q$ będziemy traktować jako *rozkład możliwości* na zbiorze możliwych kategorii zmiennej *Y*. Niech $\widetilde{Y}_q^\alpha$ będzie tzw. α-cięciem ($0 < \alpha \leq 1$) rozmytej obserwacji $\widetilde{Y}_q$. Można zauważyć, że $Y_q^\alpha = \{M_{i_q 1}^\alpha, M_{i_q 2}^\alpha, \ldots, M_{i_q r}^\alpha\}$, gdzie

$$M_{i_q j}^\alpha = \begin{cases} 1 & \text{gdy } \mu_{i_q j} \geq \alpha \\ 0 & \text{w przec. przyp} \end{cases}, i_q \in \{1, \ldots, k\}, j = 1, \ldots, r \tag{16}$$

jest zwyczajnym zbiorem. Oznaczmy przez $Y^\alpha = \{Y_1^\alpha, \ldots, Y_n^\alpha\}$ zbiór α-cięć dla wszystkich *n* obserwacji. Niech $S^\alpha \subseteq Y^\alpha$ będzie takim podzbiorem zbioru Y^α, dla którego elementów spełniony jest warunek $\sum_{j=1}^{r} M_{i_q j}^\alpha = 1$. Jest to więc dla danego α-cięcia zbiór obserwacji jednoznacznych. Oznaczmy teraz przez $n_{ij,\min}^\alpha$ liczbę obserwacji ze zbioru S^α, które znajdują się w *ij*-tej komórce. Analogicznie, oznaczmy przez $n_{ij,\max}^\alpha$ liczbę obserwacji ze zbioru Y^α, które znajdują się w *ij*-tej komórce.

Oznaczmy przez *G* dowolną statystykę podsumowującą dwuwymiarową tablice kontyngencji. W celu zdefiniowania uogólnienia takiej statystyki na przypadek danych rozmytych wprowadźmy pojęcie *obserwacji zgodnej z obserwacja rozmytą* dla zadanego α-cięcia. Precyzyjną obserwację $Z(i_0, j_0)$ taką, że $\mu_{ij} = 1$ dla $(i = i_0) \cap (j = j_0)$, zaś $\mu_{ij} = 0$ w przeciwnym przypadku, będziemy nazywać obserwacją zgodną dla danego α-cięcia z obserwacją rozmytą $\widetilde{Y}_q, q = 1, \ldots, n$, gdy dla danego $i_q = i_0$ zachodzi równość $M_{i_0 j_0}^\alpha = 1$.

Niech N^α będzie zbiorem *wszystkich* tablic kontyngencji zbudowanych z obserwacji zgodnych dla zadanego α-cięcia z wszystkimi obserwacjami rozmytymi i nierozmytymi. Hryniewicz [6] zaproponował, by funkcję przynależności *rozmytej* wersji statystyki *G* określić przy pomocy następujących α-cięć:

$$G_{\min}^\alpha = \inf_{N^\alpha} G \tag{17}$$

oraz

$$G_{\max}^\alpha = \sup_{N^\alpha} G . \tag{18}$$

Powyższy wynik należy zinterpretować w następujący sposób. Wskutek nieprecyzyjnych wskazań kategorii zmiennej *Y* nie jesteśmy w stanie określić dokładnej wartości statystyki *G*. Jedyne, co możemy podać to dolną i górną granicę przedziału jej możliwych wartości dla każdego α-cięcia ($0 < \alpha \leq 1$).

Wyznaczenie wartości górnych i dolnych ograniczeń określonych wzorami (17) i (18) jest w ogólnym przypadku zadaniem bardzo złożonym. W przypadku rozmytej wersji statystyki chi-kwadrat Pearsona Hryniewicz [6] podał, że odpowiednie zadania optymalizacyjne mogą nie być zadaniami wypukłymi, tzn. możliwe są lokalne minima i maksima. Oznacza to, że gwarancję znalezienia dokładnych wartości ograniczeń daje tylko pełny przegląd tablic kontyngencji należących do zbioru N^α. Dokonanie tego w przypadku dużej liczby możliwych kategorii rozpatrywanych zmiennych i dużej liczby danych rozmytych może wymagać wykonania bardzo dużej liczby obliczeń. Należy więc brać to pod uwagę w planowaniu eksperymentu eksperckiego, gdy interesuje nas stwierdzenie niezależności analizowanych zmiennych przy pomocy testu chi-kwadrat Pearsona.

W przypadku rozmytej wersji statystyki γ Goodmana-Kruskala Hryniewicz [7], [8] udowodnił twierdzenia pozwalające na efektywne rozwiązanie zadań optymalizacyjnych danych wzorami (17) i (18). Podobne twierdzenia, pozwalające na jeszcze prostsze znajdowanie granicznych wartości α-cięć rozmytej statystyki *G*, podane zostały przez Hryniewicza [9] dla przypadku statystyki *d* Sommersa. Z tego też powodu, do wyznaczania siły zależności pomiędzy dwiema zmiennymi, gdy dysponujemy nieprecyzyjnymi (rozmytymi) opiniami ekspertów, zalecana jest właśnie ta statystyka.

4. METODA WYBORU ZMIENNYCH MODELU

W poprzednim rozdziale przedstawiona została metodologia konstrukcji statystyk, które można wykorzystać do badania siły zależności skategoryzowanych cech statystycznych, gdy dysponujemy nieprecyzyjnie określonymi danymi. Zastosowanie tej metodologii pozwala wyznaczyć rozmyte wersje znanych statystyk, stosowanych od lat do badania zależności w przypadku dysponowania danymi dokładnymi. Statystyki te odznaczają się dwiema cechami: Po pierwsze, przyjmują wartości w sposób losowy, co jest konsekwencją statystycznego charakteru prowadzonych badań (oceny innych ekspertów, *wylosowanych* z populacji wszystkich ekspertów, byłyby zapewne nieco inne). Po drugie, odznaczają się brakiem precyzji o charakterze nielosowym. Jeżeli więc na podstawie ich analizy chcemy podejmować racjonalne decyzje, to musimy uwzględnić oba te aspekty jednocześnie. Aparatem formalnym, który można tu zastosować, jest teoria rozmytych zmiennych losowych oraz związane z nią metody „miękkich" obliczeń statystycznych.

Problem wnioskowania statystycznego w przypadku obserwacji rozmytych realizacji zmiennych losowych jest ciągle przedmiotem badań o charakterze podstawowym. Rozważania teoretyczne na ten temat można znaleźć np. w pracach Grzegorzewskiego [5] i Hryniewicza [10]. W przypadku rozpatrywanych w niniejszej pracy rozmytych wersji statystyk chi-kwadrat Pearsona, γ

Goodmana-Kruskala oraz *d* Sommersa zastosowanie zaproponowanej w w/w pracach metodologii jest możliwe wyłącznie w przypadkach asymptotycznych, tzn. w przypadku dysponowania opiniami bardzo dużej liczby ekspertów. Spełnienie togo wymagania jest w praktyce niezwykle trudne i wobec tego poniżej zaproponujemy procedury uproszczone, które powinny wystarczyć w praktyce.

Aby prowadzić jakiekolwiek analizy dotyczące danych rozmytych powinniśmy najpierw określić sposoby ich porównywania. Niestety, liczby rozmyte, a w ogólnym przypadku zbiory rozmyte, nie dadzą się uporządkować w sposób jednoznaczny. Konieczne jest wprowadzenie *dodatkowych* kryteriów, według których będzie można dokonywać odpowiednich porównań. Odpowiednich narzędzi dostarcza tu zaproponowana przez Zadeha [14] *teoria możliwości (possibility theory)* oraz zaproponowana przez Dubois i Prade [2] posybilistyczna metoda porównywania liczb rozmytych.

W niniejszej pracy do porównywania wartości rozmytych statystyk opisujących siłę zależności rozpatrywanych zmiennych wykorzystamy wspomniane powyżej *podejście posybilistyczne*, zaproponowane przez Dubois i Prade [2]. W pracy tej zdefiniowane zostały cztery wskaźniki pozwalające porównywać ze sobą dwie liczby rozmyte $\tilde{A}$ oraz $\tilde{B}$.

Wskaźnik *możliwości dominacji PD* definiowany jest zależnością:

$$PD = Poss\left(\tilde{A} \geq \tilde{B}\right) = \sup_{x,y:x\geq y} \min\left\{\mu_{\tilde{A}}(x), \mu_{AB}(y)\right\} \qquad (19)$$

Wskaźnik ten określa stopień możliwości, że rozmyta liczba $\tilde{A}$ jest nie mniejsza od rozmytej liczby $\tilde{B}$.

Z kolei, wskaźnik *możliwości ścisłej dominacji PSD* definiowany jest następująco:

$$\begin{aligned} PSD &= Poss\left(\tilde{A} > \tilde{B}\right) \\ &= \sup_{x} \inf_{y:y\geq x} \min\left\{\mu_{\tilde{A}}(x), 1 - \mu_{\tilde{B}}(y)\right\} \end{aligned} \qquad (20)$$

Wskaźnik ten określa stopień możliwości, że rozmyta liczba $\tilde{A}$ jest większa od rozmytej liczby $\tilde{B}$.

Kolejne dwa wskaźniki pozwalają określić *miary konieczności* podjęcia odpowiednich decyzji. Wskaźnik *konieczności dominacji ND* definiowany jest następująco:

$$\begin{aligned} ND &= Necc\left(\tilde{A} \geq \tilde{B}\right) \\ &= \inf_{x} \sup_{y:y\leq x} \max\left\{1 - \mu_{\tilde{A}}(x), \mu_{\tilde{B}}(y)\right\} \end{aligned} \qquad (21)$$

Wskaźnik ten określa stopień konieczności, że rozmyta liczba $\tilde{A}$ jest nie mniejsza od rozmytej liczby $\tilde{B}$.

Z kolei, wskaźnik *konieczności ścisłej dominacji NSD* określony jest wzorem:

$$\begin{aligned} NSD &= Ness\left(\tilde{A} > \tilde{B}\right) \\ &= 1 - \sup_{x,y:x\leq y} \min\left\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(y)\right\} \\ &= 1 - Poss\left(\tilde{B} \geq \tilde{A}\right) \end{aligned} \qquad (22)$$

Wskaźnik ten określa stopień Konieczności, że rozmyta liczba $\tilde{A}$ jest większa od rozmytej liczby $\tilde{B}$.

Zdefiniowane powyżej wskaźniki można wykorzystać do podejmowania decyzji dwu rodzajów. Po pierwsze, czy rozpatrywane zmienne - decyzyjna i objaśniająca - są w ogóle wzajemnie zależne. Po drugie, czy zależność rozpatrywanej zmiennej decyzyjnej od jednej zmiennej objaśniającej jest silniejsza od zależności tej zmiennej od innej zmiennej objaśniającej.

W przypadku weryfikacji hipotezy statystycznej o wzajemnej niezależności zmiennej decyzyjnej i zmiennej objaśniającej musimy dokonać porównania rozmytej wartości statystyki chi-kwadrat Pearsona z odpowiednim kwantylem rozkładu chi-kwadrat. W takim przypadku wskaźnik możliwości dominacji *PD* oznacza stopień możliwości, że decyzja o odrzuceniu weryfikowanej hipotezy statystycznej nie jest gorsza od decyzji o jej przyjęciu. Z kolei, wskaźnik możliwości ścisłej dominacji *PSD* oznacza stopień możliwości, że decyzja o odrzuceniu weryfikowanej hipotezy statystycznej jest lepsza od decyzji o jej przyjęciu. W przypadku wskaźników rozpatrywanych dwu wskaźników konieczności mamy do czynienia z następująca interpretacją. Wskaźnik konieczności dominacji *ND* oznacza stopień konieczności, że decyzja o odrzuceniu weryfikowanej hipotezy statystycznej nie jest gorsza od decyzji o jej przyjęciu, zaś wskaźnik konieczności ścisłej dominacji *NSD* oznacza stopień konieczności, że decyzja o odrzuceniu weryfikowanej hipotezy statystycznej jest lepsza od decyzji o jej przyjęciu.

Zastosowanie powyższej metodologii w praktyce może wiązać się z pewnymi trudnościami. Musimy pamiętać, że statystyka chi-kwadrat Pearsona ma asymptotycznie rozkład chi-kwadrat, gdy we wszystkich komórkach tablicy kontyngencji znajduje się wystarczająca liczba obserwacji (wiecej niż 5). Z takim przypadkami będziemy się zapewne spotykać w praktyce, gdy eksperci będą się wypowiadali o przypadkach budzących wątpliwości zależności, lub gdy celem analizy będzie stwierdzenie, czy dwie zmienne objaśniające są wzajemnie niezależne. Jeżeli jednak nie ma większych wątpliwości, co do samego faktu występowania zależności, to nie należy spodziewać się by powyższy warunek był spełniony. Na przykład, gdy występuje wyraźna dodatnia zależność pomiędzy zmienną decyzyjną i zmienną objaśniającą, to nie należy spodziewać się opinii mówiących, że dużym wartościom jednej z tych zmiennych odpowiadają małe wartości drugiej. W takich przypadkach konieczne jest podejście zdroworozsądkowe, zgodnie, z którym należy odrzucić hipotezę o niezależności rozpatrywanych zmiennych nawet wtedy, gdy nie są spełnione formalne warunki stosowania asymptotycznego testu niezależności chi-kwadrat Pearsona.

W drugim z rozpatrywanych przypadków, tzn. gdy badamy czy zależność rozpatrywanej zmiennej decyzyjnej od jednej zmiennej objaśniającej jest silniejsza od zależności tej zmiennej od innej zmiennej objaśniającej, sytuacja jest dalece trudniejsza. Zadanie to nie ma zadowalającego rozwiązania nawet w przypadku danych dokładnych. Pozostaje więc jedyna możliwość, by odpowiednie decyzje podejmować na podstawie bezpośrednich porównań zaobserwowanych wartości odpowiednich statystyk. W takim przypadku wspomniane powyżej wskaźniki możliwości i konieczności dominacji będą jedyną podstawą podejmowanych decyzji, a do ich interpretacji będziemy korzystać z potocznego rozumienia pojęć „możliwości" i „konieczności".

5. PRZYKŁAD OBLICZENIOWY

Zaproponowana w niniejszym artykule metodologią analizy ocen eksperckich zilustrujemy prostym przykładem obliczeniowym. Przyjmijmy, że ustalona została pewna zmienna decyzyjna Y, której wartości chcemy w przyszłości przewidywać na podstawie obserwacji odpowiednio dobranej zmiennej objaśniającej X, przy czym rozpatrywana jest możliwość wykorzystania dwu różnych zmiennych objaśniających $X1$ oraz $X2$. Określenie odpowiedniej zależności typu regresyjnego pomiędzy zmienną decyzyjną i każdą z tych zmiennych wymaga przeprowadzenia kosztownych badań. Okazuje się, że środki finansowe pozwalają na przeprowadzenie badań pozwalających określić zależność regresyjną pomiędzy zmienną decyzyjną i tylko jedną zmienną objaśniającą. W tej sytuacji należy wstępnie wybrać tę zmienną objaśniającą, która w większym stopniu będzie nadawała się do przewidywania wartości zmiennej decyzyjnej.

Załóżmy, że wybór odpowiedniej zmiennej powierzono n=20 ekspertom. Wartości zmiennych objaśniających podzielone zostały na k=2 kategorie: „małe" i „duże". Z kolei, wartości zmiennej decyzyjnej podzielone zostały na r=3 kategorii: „małe", „średnie" i „duże".

W przypadku analizy siły zależności zmiennej decyzyjnej Y i zmiennej objaśniającej $X1$, dwunastu ekspertów wylosowało pytanie. Jeżeli wartość zmiennej $X1$ jest mała, to jaka jest wartość zmiennej Y", a ośmiu pytanie „Jeżeli wartość zmiennej $X1$ jest duża, to jaka jest wartość zmiennej Y". Odpowiedzi na pierwsze pytanie przedstawiały się następująco:

- sześciu dało jednoznaczną odpowiedź „mała", tzn. mającą postać
  "mała"|1+"średnia"|0+"duża"|0 -
- trzech dało odpowiedź postaci
  "mała"|1+"średnia"|0,5+"duża"|0
- dwu odpowiedziało
  "mała"|0,5+"średnia"|1+"duża"|0
- jeden odpowiedział jednoznacznie: „średnia", tzn.
  "mała"|0+"średnia"|1+"duża"|0.

Z kolei, odpowiedzi na drugie pytanie były następujące:

- czterech dało jednoznaczną odpowiedź „duża", tzn. mającą postać
  "mała"|0+"średnia"|0+"duża"|1 -
- dwu dało odpowiedź postaci
  "mała"|0+"średnia"|0,5+"duża"|1
- a dwu odpowiedziało
  "mała"|0+"średnia"|1+"duża"|0,5.

Analiza siły zależności została przeprowadzona z wykorzystaniem statystyki d Sommersa.

Jak łatwo zauważyć, dla α-cięcia na poziomie α=1 odpowiedzi ekspertów są jednoznaczne i dają się przedstawić w postaci następującej tablicy kontyngencji:

$X1\backslash Y$	„mała"	„średnia"	„duża"
„mała"	9	3	0
„duża"	0	2	6

Z kolei, dla α-cięcia na poziomie α=0,5 tablica kontyngencji zbudowana z takich obserwacji zgodnych z obserwacjami rozmytymi, dla których wskaźnik d Sommersa przyjmuje wartość minimalną ma postać

$X1\backslash Y$	„mała"	„średnia"	„duża"
„mała"	6	6	0
„duża"	0	4	4

Natomiast tablica kontyngencji zbudowana z takich obserwacji zgodnych z obserwacjami rozmytymi, dla których wskaźnik d Sommersa przyjmuje wartość maksymalną ma postać

$X1\backslash Y$	„mała"	„średnia"	„duża"
„mała"	11	1	0
„duża"	0	0	8

Proste obliczenia, wykonane na podstawie wzoru (14) pokazują, że rozmytą wersję statystyki d Sommersa określają dwa α-cięcia; wartość 0,938 na poziomie α=1 oraz przedział [0,75 , 1,0] na poziomie α=0,5.

W przypadku analizy siły zależności zmiennej decyzyjnej Y i zmiennej objaśniającej $X2$, dziesięciu ekspertów wylosowało pytanie „Jeżeli wartość zmiennej $X2$ jest mała, to jaka jest wartość zmiennej Y", a dziesięciu pytanie „Jeżeli wartość zmiennej $X2$ jest duża, to jaka jest wartość zmiennej Y". Odpowiedzi na pierwsze pytanie przedstawiały się następująco:

- pięciu dało jednoznaczną odpowiedź „mała", tzn. mającą postać
  "mała"|1+"średnia"|0+"duża"|0 -
- dwu dało odpowiedź postaci
  "mała"|1+"średnia"|0,5+"duża"|0
- dwu odpowiedziało
  "mała"|0,5+"średnia"|1+"duża"|0
- jeden odpowiedział jednoznacznie: „średnia", tzn.
  "mała"|0+"średnia"|1+"duża"|0.

Z kolei, odpowiedzi na drugie pytanie były następujące:

- siedmiu dało jednoznaczną odpowiedź „duża", tzn. mającą postać
  "mała"|0+"średnia"|0+"duża"|1 -
- dwu dało odpowiedź postaci
  "mała"|0+"średnia"|0,5+"duża"|1
- a jeden odpowiedział
  "mała"|0+"średnia"|1+"duża"|0,5 .

Jak łatwo zauważyć, również i w tym przypadku dla α-cięcia na poziomie α=1 odpowiedzi ekspertów są jednoznaczne i dają się przedstawić w postaci następującej tablicy kontyngencji:

X2\Y	„mała"	„średnia"	„duża"
„mała"	7	3	0
„duża"	0	1	9

Z kolei, dla α-cięcia na poziomie α=0,5 tablica kontyngencji zbudowana z takich obserwacji zgodnych z obserwacjami rozmytymi, dla których wskaźnik *d* Sommersa przyjmuje wartość minimalną ma postać

X2\Y	„mała"	„średnia"	„duża"
„mała"	5	5	0
„duża"	0	3	7

Natomiast tablica kontyngencji zbudowana z takich obserwacji zgodnych z obserwacjami rozmytymi, dla których wskaźnik *d* Sommersa przyjmuje wartość maksymalną ma postać

X2\Y	„mała"	„średnia"	„duża"
„mała"	9	1	0
„duża"	0	0	10

Obliczenia dla pary zmiennych (*X2,Y*) pokazują, że rozmytą wersję statystyki *d* Sommersa określają dwa α-cięcia; wartość 0,97 na poziomie α=1 oraz przedział [0,85 , 1] na poziomie α=0,5.

Z analizy powyższych rezultatów wynika, że zmienna *X2* lepiej niż zmienna *X1* objaśnia zmienną *Y*. Obserwowane liczności próbek (liczby eksperckich opinii) są zbyt małe, by zaobserwowane różnice uznać za istotne statystycznie. Jeżeli jednak ograniczymy się do porównania zaobserwowanych rozmytych wartości wskaźnika *d* Sommersa, to można pokazać, że przy przyjętej kawałkami liniowej aproksymacji funkcji przynależności obu tych rozmytych wartości, istnieje pewna konieczność (wskaźnik konieczności ścisłej dominacji *NSD* równy 0,088) uznania zmiennej *X2* za zmienną mocniej powiązaną ze zmienną decyzyjną *Y*..

CHOOSING OF VARIABLES IN SYSTEM'S ANALYSIS - APPLICATION OF STATISTICAL MEASURES OF DEPENDENCE IN PRESENCE OF IMPRECISE DATA

Abstract: A methodology for choosing variables for the models describing complex problem of system's analysis has been proposed. The variables are chosen using the results of statistical analysis of expert's opinions collected in a specially designed experiment. The analysis takes into account a possibility of imprecise (fuzzy) experts opinions.

Literatura

[1] Bubnicki Z. (1974) *Identyfikacja obiektów sterowania*. PWN, Warszawa.

[2] Dubois D., Prade H. (1983) Ranking fuzzy numbers in the setting of possibility theory. *Information Sciences*, 30, 184-244.

[3] Goodman L.A., Kruskal W.H. (1954) Measures of Association for Cross Classification. *Journ. of the Amer. Stat. Assoc.*, 49, 732-764.

[4] Goodman L.A., Kruskal W.H. (1972) Measures of Association for Cross Classification. IV: Simplification of Asymptotic Variances. *Journ. of the Amer. Stat. Assoc.*, 67, 414-421.

[5] Grzegorzewski P. (2000) Testing statistical hypotheses with vague data. *Fuzzy Sets and Systems*, 112, 501-510.

[6] Hryniewicz O. (2004): Selection of variables for systems analysis – application of a fuzzy statistical test for independence. *Proc. of IPMU'2004*, Perugia, 3, 2197-2204.

[7] Hryniewicz O. (2004) Measures of dependence for fuzzy ordered categorical data. K.Lopez-Diaz, M.A.Gil, P.Grzegorzewski, O.Hryniewicz, J.Lawry (Eds.) *Soft Methodology and Random Information Systems*. Springer, Berlin, 503-510.

[8] Hryniewicz O. (2004) Goodman-Kruskal γ measure of dependence for fuzzy ordered categorical data. *Computational Statistics and Data Analysis* (złożone do publikacji).

[9] Hryniewicz O. (2005) Statistical analysis of categorical data – from fuzzy data to fuzzy measures of dependence. *Issues in Soft Computing - Theory and Applications*. Exit, Warszawa, 2005 (w druku).

[10] Hryniewicz O. (2005) Possibilistic decisions and fuzzy statistical tests. *Fuzzy Sets and Systems* (w druku).

[11] Koronacki J., Mielniczuk J. (2001) *Statystyka dla studentów kierunków technicznych i przyrodniczych*. WNT, Warszawa.

[12] Mańczak K., Nahorski Z. (1983) *Komputerowa identyfikacja obiektów dynamicznych*. PWN, Warszawa.

[13] Söderström T., Stoica P. (1997) *Identyfikacja systemów*. PWN, Warszawa.

[14] Zadeh L.A. (1978) Fuzzy sets as a basis for the theory of possibility. *Fuzzy Sets and Systems*, 1, 3-28.

Instytut Badań Systemowych
Polskiej Akademii Nauk

ISBN 83-89475-01-4