

AKADEMIA OBRONY NARODOWEJ

WYDZIAŁ STRATEGICZNO-OBRONNY
INSTYTUT EKONOMIKI OBRONY

AON wewn. 4503/93

Ppłk dr inż. Stefan KURINIA

REPREZENTACJA WIEDZY W WARUNKACH
ISTNIENIA SPRZECZNYCH, NIEPEWNYCH
LUB NIEJEDNOZNACZNYCH DANYCH

MATERIAŁY I STUDIA
Nr 3/93

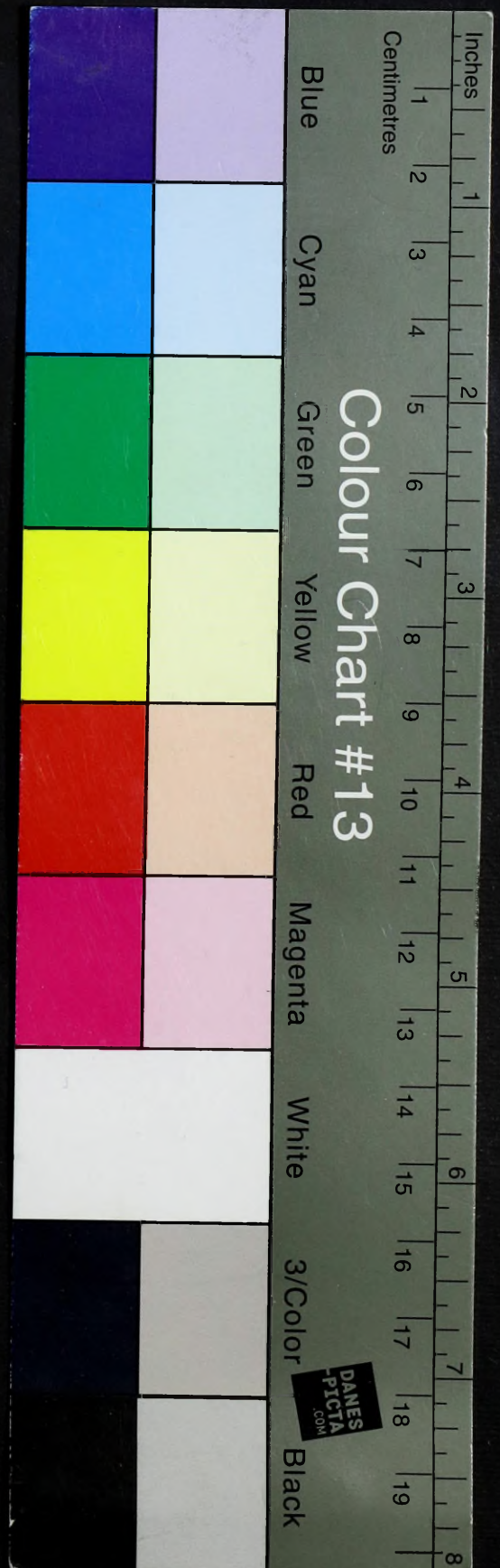
Biblioteka Główna
Akademii Obrony Narodowej
S/2076
05-0020-003-0



WARSZAWA

1993

60460





AKADEMIA OBRONY NARODOWEJ

WYDZIAŁ STRATEGICZNO-OBRONNY
INSTYTUT EKONOMIKI OBRONY

AON wewn. 4503/93



Ppłk dr inż. Stefan KURINIA

REPREZENTACJA WIEDZY W WARUNKACH ISTNIENIA SPRZECZNYCH, NIEPEWNYCH LUB NIEJEDNOZNACZNYCH DANYCH

MATERIAŁY I STUDIA

Nr 3/93

~~S/2076~~



WARSZAWA

1993

Wstęp

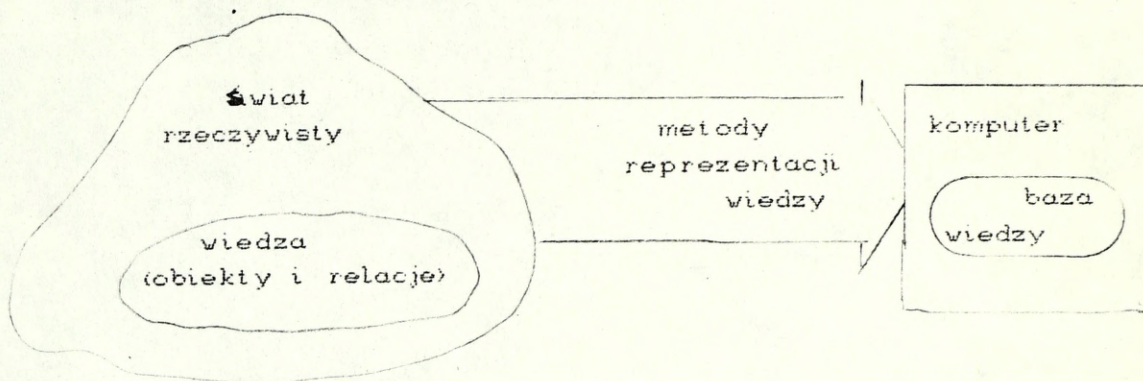
W ostatnich latach możemy obserwować w informatyce obok przetwarzania danych coraz silniejsze zainteresowanie przetwarzaniem wiedzy. Ma to ścisły związek z badaniami nad tzw. "sztuczną inteligencją" (artificial intelligence- AI). Sztuczna inteligencja wg E.A. Feigenbaum'a i P. McCorduck'a jest to dziedzina informatyki dotycząca metod i technik wnioskowania symbolicznego przez komputer oraz symbolicznej reprezentacji wiedzy stosowanej podczas takiego wnioskowania. Natomiast wg M. Minsky'ego sztuczna inteligencja jest nauką o maszynach, które realizują zadania wymagające inteligencji wtedy, kiedy wykonywane są przez człowieka [1].

W latach 1960-70 poszukiwano ogólnych metod dla rozwiązywania problemów ogólnego przeznaczenia (general problem solver- GPS). Niepowodzenia w tych badaniach spowodowały, że po roku 1970 nastąpił regres prac nad sztuczną inteligencją. Ponowny rozwój prac nastąpił po 1975 roku. Początkowo próbowano szukać ogólnych metod dla znalezienia rozwiązań specjalistycznych zagadnień. Jednak najlepsze efekty zaczęto uzyskiwać z chwilą, gdy zastosowano wąską wiedzę do realizacji wyspecjalizowanych programów.

W niniejszej pracy wiedza rozumiana jest jako symboliczny opis rzeczywistości charakteryzujący aksjomatyczne i empiryczne relacje, mogący zawierać procedury, które na nich działają. Wiedza może być przedstawiona w różnych formach. Może to być asocjacja empiryczna, czyli skojarzenia eksperta oparte na różnych faktach. Mogą to być też pojęcia, nakazy i regulacje sterujące czynnościami z danej dziedziny. Innymi formami wiedzy są modele przyczynowo- skutkowe i schematy wnioskowania.

Przetwarzanie wiedzy jest przetwarzaniem symbolicznym, na które składa się pozyskiwanie wiedzy, tworzenie bazy wiedzy oraz inteligentne jej przeszukiwanie. Przetwarzanie

wiedzy wymaga metod poznawania a następnie opisu rzeczywistości. Metody formalizowania wiedzy o rzeczywistości i dostosowania jej do techniki komputerowej nazywa się metodami reprezentacji wiedzy- MRW. Interpretację graficzną reprezentacji wiedzy przedstawia rysunek 1.



Rys. 1. Reprezentacja wiedzy

Metody reprezentacji wiedzy obejmują sposoby przekazywania, zapisywania i gromadzenia wybranego fragmentu wiedzy o rzeczywistości. Ze względu na zastosowane sposoby reprezentacji wiedzy możemy wyróżnić:

- 1- rachunek predykatów,
- 2- rachunek zdań,
- 3- reguły produkcji,
- 4- zapis stwierdzeń,
- 5- tablice decyzyjne,
- 6- sieci semantyczne,
- 7- ramy i scenariusze.

Metody 1 i 2 bazują na zastosowaniu klasycznej logiki. Metody 3 i 4 są prostymi sposobami opisu wiedzy za pomocą reguł JEZELI- TO (IF- THEN). Stwierdzenia to reguły zawsze prawdziwe. Metody 5,6 i 7 zostaną przedstawione w dalszej części opracowania.

Istnieje znaczne zróżnicowanie metod reprezentacji

wiedzy wynikające z wykorzystywania przetwarzania symbolicznego w wielu dziedzinach informatyki. Do najważniejszych dziedzin zaliczyć można: przetwarzanie języka naturalnego, teorię gier, systemy ekspertowe, robotykę, procesy percepcji, uczenie maszynowe (machine learning), wyszukiwanie informacji, automatyczne programowanie. Szczególnie silny rozwój metod reprezentacji wiedzy nastąpił w ostatnim dziesięcioleciu. Ma to ścisły związek z ogromnymi przeobrażeniami w technice komputerowej i jej upowszechnieniu. Bezpośrednią przyczyną było masowe wprowadzenie do użytkowania mikrokomputerów - komputerów osobistych (personal computer - PC). Wielkość tych przeobrażeń można scharakteryzować za pomocą np. mocy obliczeniowej mikrokomputera określanej w milionach instrukcji na sekundę (million instructions per second - MIPS) i pojemności dyskietek pamięci zewnętrznej (tablica 1).

Rok	1985	1990	1995*
Moc obliczeniowa [MIPS]	10	100	1000
Pojemność dyskiетки [MB]	0,72	1,44	1500

* Wielkości prognozowane.

Tablica 1. Wybrane charakterystyki rozwoju mikrokomputerów

Stroną ekonomiczną tego zjawiska najlepiej charakteryzuje fakt, że w 1995 roku można będzie kupić mikrokomputer o mocy obliczeniowej równej mocy superkomputera z roku 1985 za 1000 razy mniejszą cenę.

W ślad za szerokim upowszechnieniem mikrokomputerów

poszło znaczne zapotrzebowanie na programy komputerowe, w tym mikrokomputerowe systemy ekspertowe. Systemy ekspertowe znalazły zastosowanie w medycynie, diagnostyce technicznej, chemii, technice wojskowej, ekonomii, energetyce. O popularności systemów ekspertowych zdecydowały między innymi łatwość udostępnienia wiedzy i efektywne wykorzystanie ludzi do rozwiązywania złożonych problemów.

Celem niniejszego opracowania jest przedstawienie wybranych metod reprezentacji wiedzy i pokazanie jak zastosować te metody do opisu wiedzy w sytuacji, gdy mamy do czynienia ze sprzecznością, niepewnością bądź niejednoznacznością danych.

I. Wybrane metody reprezentacji wiedzy

1. Tablice decyzyjne

Tablica decyzyjna jest prostym i popularnym sposobem przedstawiania wiedzy. Wiedza jest reprezentowana za pomocą zestawienia możliwych warunków, koniecznych działań i reguł podejmowania decyzji. Zwykle tablica decyzyjna składa się z czterech obszarów:

- obszaru opisu przesłanek,
- obszaru opisu działań,
- obszaru konkluzji,
- obszaru decyzji.

Ogólną postać tablicy decyzyjnej przedstawia tablica 2.

		r1	r2	..	rs
a1		k11	k12	..	k1s
a2	opis	k21	k22	..	k2s
.	przesłanek
.	
an		kn1	kn2	..	kns
b1		d11	d12	..	d1s
b2	opis	d21	d22	..	d2s
.	decyzji
.	
bm		dm1	dm2	..	dms

ai- i-ta przesłanka, bj- j-te działanie,

kip- i-ta konkluzja w p-tej regule decyzyjnej,

djp- j-ta decyzja w p-tej regule decyzyjnej,

rs- s-ta reguła decyzyjna.

Tablica 2. Ogólna postać tablicy decyzyjnej

Kolumny prawej strony tablicy zawierają reguły decyzyjne. Liczba reguł jest permutacją przesłanek i konkluzji. Dla 2 konkluzji {tak, nie} i czterech przesłanek $\{a_1, a_2, a_3, a_4\}$ mamy $2^4 = 16$ reguł decyzyjnych. Jeżeli tablica decyzyjna zawiera wszystkie możliwe reguły to mówimy, że nasza wiedza o danej rzeczywistości jest pełna. Można też wykorzystać do reprezentacji wiedzy prostszą wersję tablicy decyzyjnej w postaci jak w tablicy 3.

	a1	a2	..	an	d
x1	v11	v12	..	v1n	d1
x2	v21	v22	..	v2n	d2
.
.
xm	vm1	vm2	..	vmn	dm

x_i - i-ty obiekt, a_j - j-ty atrybut,
 v_{ij} - wartość j-tego atrybutu i-tego obiektu,
 d_j - j-ta decyzja.

Tablica 3. Uproszczona postać tablicy decyzyjnej

W tej tablicy w każdym wierszu znajdują się wartości atrybutów danego opisywanego obiektu oraz podjęta decyzja. Liczbę wierszy określają nie wszystkie możliwe reguły decyzyjne lecz liczba opisywanych obiektów. Jeżeli opisane obiekty nie wyczerpują wszystkich możliwych reguł decyzyjnych to wtedy nasza wiedza o opisywanej rzeczywistości jest niepełna. W innych przypadkach możemy mieć wiedzę pełną lub nadmiarową (gdy reguły decyzyjne powtarzają się). Przykład uproszczonej tablicy decyzyjnej przedstawiono w tablicy 4.

pacjent	temperatura	gardlo	diagnoza
pacjent 1	normalna	bez zmian	zdrowy
pacjent 2	podwyższ.	czerwone	inf.wirusowa
pacjent 3	podwyższ.	zmiany	inf.bakteryjna
pacjent 4	wysoka	zmiany	inf.bakteryjna
pacjent 5	podwyższ.	bez zmian	inf.wirusowa

Tablica 4. Przykład tablicy decyzyjnej

2. Sieci semantyczne

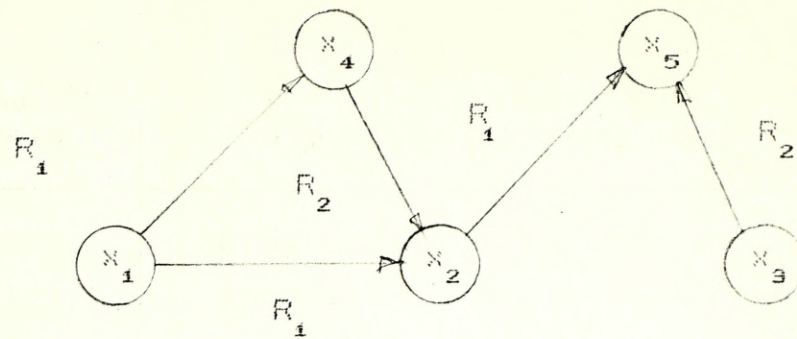
Sieci semantyczne, zwane również sieciami stwierżeń są trudniejsze w budowie i aktualizacji. Sieć składa się ze zbioru węzłów połączonych łukami. Węzły reprezentują elementy wiedzy, natomiast łuki- zależności pomiędzy elementami. Pomiedzy elementami wiedzy może występować wiele różnych zależności. Zatem zdefiniować możemy sieć semantyczną jako

$$S = \langle X, R_1, R_2, \dots, R_n \rangle$$

gdzie: $X = \{x_i\}$ - zbiór elementów wiedzy (węzłów),

$R_i \subseteq X \times X$ - zbiór relacji i-tego rodzaju.

Postać graficzną prostej sieci semantycznej przedstawiono na rysunku 2.



Rysunek 2. Sieć semantyczna

Dla potrzeb przetwarzania komputerowego sieć semantyczna przedstawiana jest w postaci reprezentacji algebraicznej jako macierz relacji

$$M[m(i,j)]$$

gdzie: $i, j \in \{1, 2, \dots, N\}$,

N - licznosc zbioru węzłów sieci,

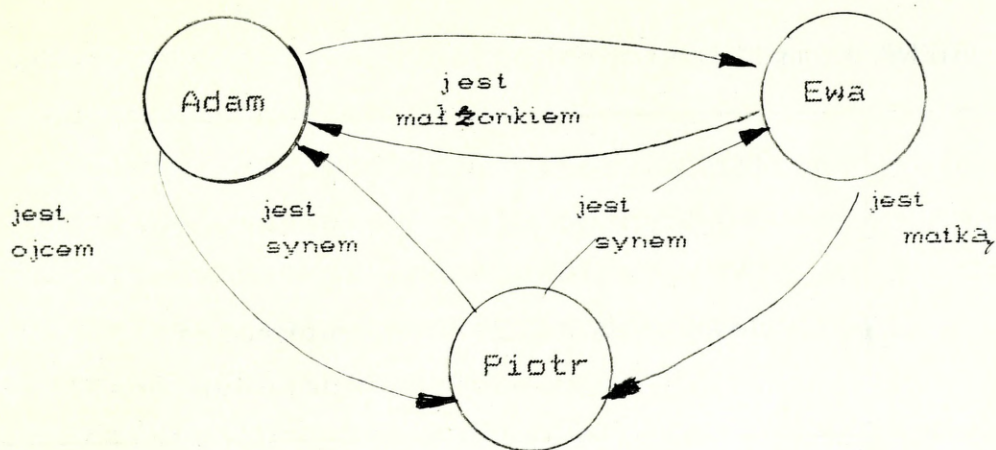
$m(i,j) = 1$ wtedy i tylko wtedy, gdy w sieci $S = \langle W, R \rangle$

występuje łuk łączący węzeł a_i z węzłem a_j , tzn. że zachodzi relacja $R(a_i, a_j)$,

$m(i,j) = 0$ w pozostałych przypadkach.

W sytuacji, gdy w sieci semantycznej występuje więcej niż jeden rodzaj relacji, dla każdego rodzaju relacji podajemy osobną macierz relacji.

Przykład prostej sieci semantycznej przedstawiono na rysunku 3.



Rysunek 3. Przykład sieci semantycznej

3. Ramy i scenariusze

Ramy (ang. frames) łączą w sobie idee reprezentacji wiedzy w postaci tablic decyzyjnych i sieci semantycznych. Każda rama ma swoją nazwę i jest tablicą składającą się z klatek (ang. slots). Klatka jest zbiorem elementów przechowujących reprezentowany przez ramę fragment wiedzy o opisywanym obiekcie. Są to deskryptory i procedury obiektu reprezentowanego przez ramę. Deskryptory mogą być określonego typu i należeć do dowolnego zbioru. Deskryptory mogą być np. typu: DEFAULT (DOMYSLNE), REQUIRE (WYMAGA), RULE (REGUŁA). Deskryptor DEFAULT określa domyślną wartość klatki. Deskryptor REQUIRE określa dopuszczalną wartość klatki. Deskryptor RULE zawiera listę reguł. Najczęściej deskryptorami są liczby, teksty, reguły, kolejne ramy. Procedury określają różne czynności, które wykonywane są w przypadku spełnienia zadanych warunków lub bezwarunkowo. Mogą to być procedury podstawienia (IF-NEEDED), uzupełniania (IF-ADDED), kasowania (CLEAN-UP) itp. Ogólną postać ramy przedstawia tablica 5.

NAZWA ramy

NAZWA klatki 1	TYP deskryptora 1 TYP deskryptora 2	wartości deskryptora 1 wartości deskryptora 2
NAZWA klatki 2	TYP deskryptora 1 TYP procedury 1	wartości deskryptora 1
...
NAZWA klatki n	TYP deskryptora 1 TYP procedury 1 TYP procedury 2	wartości deskryptora 1

Tablica 5. Ogólna postać ramy.

W ramie mogą być zapisane zarówno konkretne dane jak i pojęcia (abstrakcje). Rama opisująca pojęcie jest wzorcem ram i jest nazywana prototypową. Jej ukonkretnienie nazywa się ramą egzemplifikacyjną, inaczej egzemplarzem. Ramy, prototypy i egzemplarze, można łączyć w sieć i tworzą one wtedy rozbudowane struktury hierarchiczne. Dodatkowe klatki w ramie będą określać, które ramy sieci są nadrzędne a które podrzędne względem danej ramy. Jeżeli w ramie umieścimy klatki, które będą zawierały opisy możliwych ciągów zdarzeń, warunki wejściowe, wyniki, dane o obiektach

związanych z opisywanymi zdarzeniami to ramę taką będziemy nazywać scenariuszem.

Zalety reprezentacji wiedzy w postaci ramy to:

- możliwość opisu wiedzy zarówno o charakterze deklaratywnym jak i proceduralnym,
- oddzielenie opisu konkretów od opisu pojęć,
- czytelna organizacja bazy wiedzy.

Przykład prostej ramy przedstawia tablica 6.

DATA

ROK	REQUIRE DEFAULT	integer 1993
MIESIĄC	REQUIRE	styczeń, luty, marzec, kwiecień, maj, czerwiec, lipiec, sierpień, wrzesień, październik, listopad, grudzień
DZIEŃ	REQUIRE	1, 2, ... 31

Tablica 6. Przykład ramy

II. Sposoby reprezentacji wiedzy w specyficznych warunkach informacyjnych

Wprowadzenie

Formalizowanie opisu wiedzy ekonomicznej przez ekspertów stanowi tę specyficzną sytuację informacyjną, w której występują dane niepewne, niejednoznaczne bądź sprzeczne. Przyczyn tej sytuacji bywa wiele. Są to przyczyny zarówno obiektywne jak i subiektywne.

Do przyczyn obiektywnych należy zaliczyć sam charakter działalności gospodarczej. Ogromna złożoność procesów ekonomicznych powoduje, że o wielu faktach możemy wypowiadać się tylko w kategorii przypuszczeń. Dotyczy to również tak podstawowych danych ekonomicznych jak przyszła produkcja, spodziewany dochód narodowy, kształtowanie się cen towarów itp. Wielość źródeł informacji i występujące przekłamania prowadzą z kolei do sprzeczności danych. Istnieje też spora grupa pojęć, które są opisywane niejednoznacznie. Przykładami takich określeń mogą być: zła gospodarka, trudna sytuacja materialna, korzystny interes itp.

Przyczyny subiektywne tkwią w samych ekspertach. Ekspert może nie chcieć lub nie umieć przekazać posiadaną wiedzę. Może też nie być w stanie podać jednoznaczną odpowiedź. Wziąć też należy pod uwagę intuicyjność i zróżnicowanie sądów pomiędzy ekspertami w grupie opisującej badane zjawisko gospodarcze.

Poniższy materiał jest próbą przedstawienia wybranych metod reprezentacji wiedzy ekonomicznej w warunkach istnienia informacji niepewnych, niejednoznacznych lub sprzecznych.

1. Reprezentacja wiedzy niepewnej

Jedną z metod reprezentacji wiedzy niepewnej jest zapisanie w bazie danych hipotez i faktów, na podstawie których będą one weryfikowane. W tym celu wykorzystuje się

twierdzenia Bayes'a z teorii prawdopodobieństwa.

Weźmy pod rozwagę następującą sytuację.

Na podstawie szeregu parametrów ekonomicznych, takich jak na przykład produkt globalny, dochód narodowy, bezrobocie, inflacja, zadłużenie, bilans handlowy, ceny itp. chcemy zweryfikować hipotezy dotyczące stanu gospodarki narodowej. Faktem może być w tej sytuacji przyjęcie przez dowolny z parametrów jednej z trzech wartości: wzrost, zmniejszenie lub bez zmian. Stawiamy następujące hipotezy:

H1- gospodarka narodowa rozwija się,

H2- gospodarka narodowa jest w stagnacji,

H3- gospodarka narodowa przechodzi kryzys.

Niech

$\{E_1, E_2, \dots, E_m\}$ - zbiór faktów,

$\{H_1, H_2, \dots, H_n\}$ - zbiór hipotez,

$\wedge H_i \wedge H_j = \text{FALSE}$, $\cup H_i = \text{TRUE}$, $\forall x$

$P(H_i)$ - prawdopodobieństwo a priori, że hipoteza H_i jest prawdziwa niezależnie od obserwacji,

$0 \leq P(H_i) \leq 1$, $\sum P(H_i) = 1$,

$P(E_j/H_i)$ - prawdopodobieństwo a priori, że fakt E_j jest prawdziwy gdy hipoteza H_i jest prawdziwa,

$0 \leq P(E_j/H_i) \leq 1$.

Prawdopodobieństwo tego, że hipoteza H_i jest prawdziwa określa się w oparciu o wzór Bayes'a w postaci

$$P(H_k/E_1, E_2, \dots, E_m) = \frac{P(H_k) * P(E_1/H_k) * P(E_2/H_k) * \dots * P(E_m/H_k)}{\sum P(H_i) * P(E_1/H_i) * P(E_2/H_i) * \dots * P(E_m/H_i)}$$

Mając zaobserwowane fakty i korzystając z wprowadzonych prawdopodobieństw a priori weryfikujemy założone hipotezy H_k na podstawie wyznaczonych prawdopodobieństw a posteriori, że hipoteza H_k jest prawdziwa gdy prawdziwe są zaobserwowane fakty.

Przykład.

Sprawdzone będą trzy hipotezy:

H_1 - nastąpi wzrost gospodarczy, $P(H_1) = 0,3$,

H_2 - nie będzie istotnych zmian, $P(H_2) = 0,5$,

H_3 - nastąpi regres w gospodarce, $P(H_3) = 0,2$.

Hipotezy sprawdzone będą w oparciu o następujące fakty:

E_1 - dochód narodowy wzrośnie,

E_2 - dochód narodowy będzie taki sam lub mniejszy,

E_3 - bilans handlowy będzie dodatni,

E_4 - bilans handlowy będzie zerowy lub ujemny.

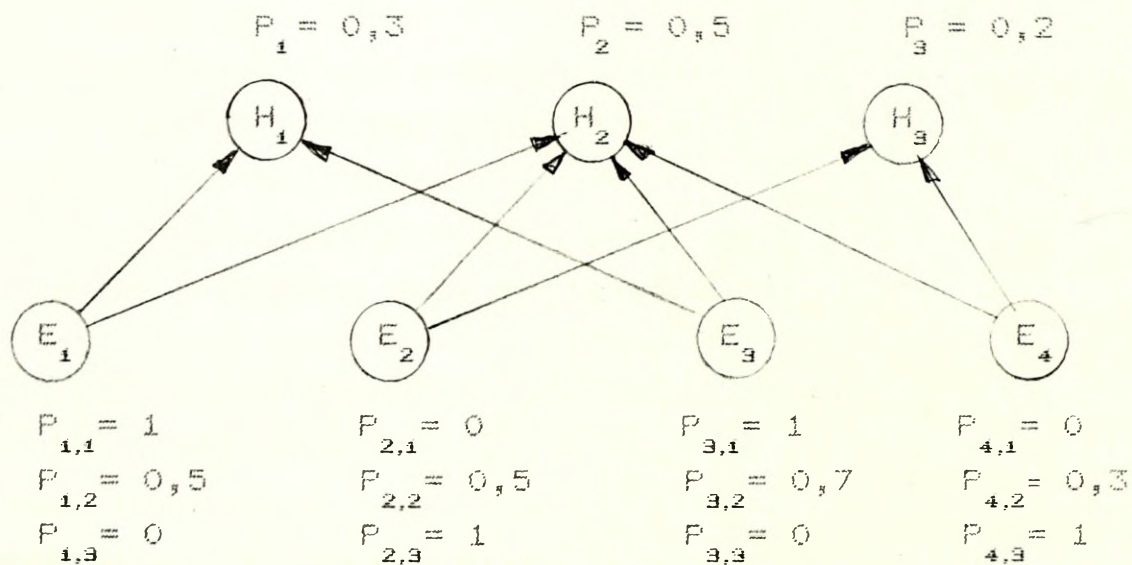
Przyjęto następujące prawdopodobieństwa warunkowe:

$P(E_1/H_1) = 1$, $P(E_2/H_1) = 0$, $P(E_3/H_1) = 1$, $P(E_4/H_1) = 0$,

$P(E_1/H_2) = 0,5$, $P(E_2/H_2) = 0,5$, $P(E_3/H_2) = 0,7$, $P(E_4/H_2) = 0,3$,

$P(E_1/H_3) = 0$, $P(E_2/H_3) = 1$, $P(E_3/H_3) = 0$, $P(E_4/H_3) = 1$.

Przykład ten stanowi duże uproszczenie problemu i prezentowany jest tylko jako ilustracja reprezentacji wiedzy.



Rysunek 4. Przykład sieci semantycznej Bayesa

W oparciu o przedstawioną sieć możemy zweryfikować hipotezę, że w gospodarce nie wystąpią istotne zmiany (H_2) wiedząc, że wzrósł dochód narodowy (E_1) i nie będzie

dodatniego salda bilansu handlowego (E_4).

$$P(H_2/E_1, E_4) = \frac{0,5 \cdot 0,5 \cdot 0,3}{0,3 \cdot 1 \cdot 0 + 0,5 \cdot 0,5 \cdot 0,3 + 0,2 \cdot 0 \cdot 1} = 1.$$

Na podstawie zaobserwowanych faktów możemy stwierdzić, że w gospodarce najpewno ($P=1$) nie wystąpią istotne zmiany.

Metoda ta posiada pewne wady. Należą do nich:

- trudności w ustaleniu zależności pomiędzy hipotezami i obserwacjami,
- subiektywne oszacowywanie prawdopodobieństw,
- złożoność modyfikacji danych.

2. Reprezentacja wiedzy niejednoznacznej

Cechą charakterystyczną gospodarki jest brak wyraźnych rozgraniczeń pomiędzy jej różnymi stanami. Stąd bierze się niejednoznaczność określeń. Czy można precyzyjnie określić na przykład stan kryzysu gospodarki? Do niejednoznacznych pojęć należy również określenie "gospodarka wojenna". Jaki poziom wskaźnika militaryzacji gospodarki decyduje o tym, że stan gospodarki możemy określić jako wojenny? Do opisu niejednoznacznych danych ekonomicznych wykorzystano pojęcia z teorii zbiorów rozmytych (fuzzy sets) wprowadzone przez Zadecha [2].

Niech będzie dany zbiór cech (obiektów)

$$X = \{x_1, x_2, \dots, x_l, \dots, x_{n-1}, x_n\}$$

gdzie x_l - stopień militaryzacji gospodarki.

Zbiorem rozmytym A na zbiorze X nazywamy

$$A = \{(\mu_A(x_l)/x_l)\}$$

gdzie: $\mu_A(x_l)$ - funkcja przynależności przyporządkowująca każdemu $x_l \in X$ stopień przynależności cechy do zbioru rozmytego A ,

$$0 \leq \mu_A(x_l) \leq 1.$$

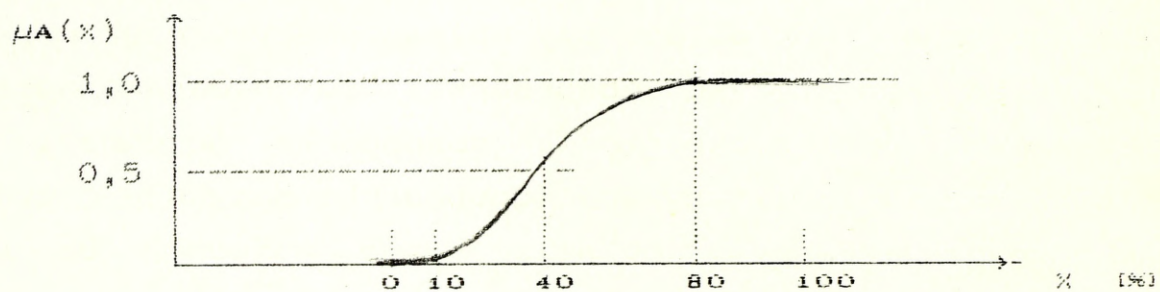
Dla zbioru nieprzeliczalnego $X = (x)$ zbiór rozmyty A ma postać

$$A = \int_x \mu_A(x) / x$$

Pojęcie stanu gospodarki wojennej możemy opisać wykorzystując na przykład standardową funkcję przynależności w postaci

$$\mu(x, \alpha, \beta, \gamma) = \begin{cases} 0 & a \leq x < \alpha \\ 2\left(\frac{x-\alpha}{\gamma-\alpha}\right)^2 & \alpha \leq x < \beta \\ 1-2\left(\frac{x-\gamma}{\gamma-\alpha}\right)^2 & \beta \leq x < \gamma \\ 1 & \gamma \leq x \leq b \end{cases}$$

Podstawiając $\alpha=10\%$, $\beta=40\%$, $\gamma=80\%$ otrzymamy funkcję przynależności zbioru rozmytego "gospodarka wojenna" opisaną na rysunku 1.



Rysunek 5. Wykres funkcji przynależności zbioru rozmytego "gospodarka wojenna"

Przyjmijmy, że mamy 11- elementowy zbiór wartości wskaźników militaryzacji

$$X = \{0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$$

Pojęciem rozmytym stanu gospodarki wojennej będzie zbiór

$$A = \{0/0, 10/0, 20/0,1, 30/0,2, 40/0,5, 50/0,8, 60/0,9, 70/1, 80/1, 90/1, 100/1\}.$$

Tak sformułowane pojęcie oznacza, że na przykład 50% wskaźnik militaryzacji określa prawie jednoznacznie stan gospodarki wojennej, a 20% wskaźnik nie jest wystarczająco jednoznaczny. Wartość 1 oznacza całkowitą jednoznaczność pojęcia, natomiast 0 oznacza, że wartość ta nie należy do

definiowanego pojęcia.

Przykład.

Spróbujmy utworzyć prostą ramę opisującą pojęcie stanu gospodarki wojennej (prototyp).

STAN GOSPODARKI WOJENNEJ

WSKAZNIK MILITARYZACJI	REQUIRE	0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100
STOPIEN PRZYNALEŻNOŚCI WSKAZNIKA	REQUIRE	0; 0; 0,1; 0,2; 0,5; 0,8; 0,9; 1; 1; 1; 1

Tablica 7. Przykład ramy STAN GOSPODARKI WOJENNEJ

Do wad tego rodzaju opisu należy zaliczyć:

- potrzebę stosowania często skomplikowanych operacji,
- korzystanie z subiektywnych wartości parametrów.

Niemniej ważną zaletą tej metody opisu jest zbliżenie opisu do realnego sposobu postrzegania rzeczywistości przez człowieka.

3. Reprezentacja wiedzy sprzecznej

Zdarza się, że z różnych przyczyn w bazie danych występują sprzeczne fakty. Intuicyjnie odczuwa się, że część obiektów bazy danych może być potraktowana jako podzbiór, dla którego opisujące je informacje są niesprzeczne. Sposób formalny oszacowania takiego podzbioru obiektów przedstawiony zostanie poniżej.

Niech obiektami bazy danych będą gospodarki różnych państw. Z każdym obiektem związane są atrybuty charakteryzujące stan gospodarki. Chcemy dokonać

oszacowania stanu gospodarek państw ze względu na jakość informacji zawartej w bazie danych. Oszacowanie zostanie przeprowadzone w oparciu o wzory aproksymacyjne z teorii zbiorów przybliżonych Z. Pawłaka [3]. Omówienie tego wymaga wprowadzenia pojęcia relacji nierozróżnialności.

Niech będą dane

$X = \{x_1, x_2, \dots, x_n\}$ - zbiór obiektów,

$Q = \{q_1, q_2, \dots, q_m\}$ - zbiór atrybutów obiektów,

$P \neq \{\emptyset\}$, $P \subset Q$ - dowolny niepusty podzbiór zbioru Q .

Relacja nierozróżnialności R oznacza, że dwa obiekty nie dają się rozróżnić jeśli mają te same wartości dla wszystkich atrybutów zbioru P .

$$R(x_i, x_j) = \{x_i \sim_j x_j\} : \exists i, j \forall k, q_k \in P (x_i) =_k q_k (x_j)$$

Relacja nierozróżnialności na zbiorze X generowana przez zbiór P jest klasyfikacją R^* . Klasyfikacja R^* jest zbiorem klas równoważności w przestrzeni równoważności $A = (X, R)$. Klasę równoważności obiektów $x \in X$ oznaczono jako $[x]_R$. Dla podzbioru wybranych gospodarek państw ze względu na wartość ich stanu (np. bardzo dobra) dokonujemy oszacowania.

Dolna aproksymacja podzbioru $Y \subset X$ w A to

$$\underline{R}Y = \{x \in X \mid [x]_R \subseteq Y\}$$

Górna aproksymacja podzbioru $Y \subset X$ w A to

$$\overline{R}Y = \{x \in X \mid [x]_R \cap Y \neq \emptyset\}$$

Dolna aproksymacja jest największym podzbiorem obiektów $x \in X$, dla których klasy równoważności zawierają się w oszacowywanym zbiorze Y . Dolna aproksymacja to te obiekty x , dla których nie występują sprzeczności ze względu na klasyfikację R . Górna aproksymacja jest najmniejszym podzbiorem obiektów x mających przecięcie z klasami

równoważności, który zawiera w sobie oszacowywany zbiór Y .
Górna aproksymacja to te obiekty x , dla których mogą występować sprzeczności ze względu na klasyfikację R .

Do podstawowych właściwości aproksymacji należą:

- 1- $RY \neq \emptyset$ i $RY \neq X - Y$ w przybliżeniu definiowane przez R ,
- 2- $RY = \emptyset$ i $RY \neq X - Y$ wewnątrznie niedefiniowane przez R ,
- 3- $RY \neq \emptyset$ i $RY = X - Y$ zewnątrznie niedefiniowane przez R ,
- 4- $RY = \emptyset$ i $RY = X - Y$ całkowicie niedefiniowane przez R ,
($RY = RY - Y$ całkowicie definiowane przez R).

Stopień niesprzeczności pojęć określa współczynnik ostrości

$$\alpha_R(x) = \frac{|RY|}{|\overline{RY}|}$$

Dla pełniejszego objaśnienia idei aproksymacji zbioru obiektów przedstawiono poniższy przykład.

Przykład.

Niech

$X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$ - zbiór obiektów, np. gospodarki różnych państw,

$A = \{a_1, a_2, a_3, a_4\}$ - zbiór wybranych atrybutów gospodarki,

a_1 - produkt narodowy na osobę (w tyś. usd),

a_2 - udział spożycia w produkcie narodowym (w %),

a_3 - stopa bezrobocia (w %),

a_4 - stopa inflacji (z 5 lat w %),

$D = \{d\}$ - decyzja/ ocena stanu gospodarki,

$d \in \{1 - słaba, 2 - średnia, 3 - silna\}$.

Gospodarka	a_1	a_2	a_3	a_4	d
Austria	16,7	73,6	5,0	137	2
Czecho-Słow.	2,9	78,4	0,6	113	1
Francja	17,0	78,3	9,5	178	1
Grecja	5,2	89,1	6,0	473	1
Hiszpania	8,7	74,5	17,2	228	2
Polska	1,6	60,0	5,0	929	1
Niemcy	19,6	72,9	7,9	123	3
Szwajcaria	27,6	70,7	0,6	133	3
Szwecja	21,5	79,0	1,6	188	2
Węgry	3,2	59,7	1,0	216	1

Dane z 1988 roku wg Rocznika statystyki międzynarodowej.

Tablica 8. Przykład tablicy decyzyjnej
o stanie gospodarek

$R^* = \{(x_1, x_3, x_7), (x_2, x_6, x_{10}), (x_4, x_5), (x_8, x_9)\}$ -
klasyfikacja dokonana ze względu na atrybuty obiektów.

Założmy, że oszacować chcemy następujący zbiór (w naszym przykładzie są to gospodarki państw mające np. słabe wyniki ekonomiczne)

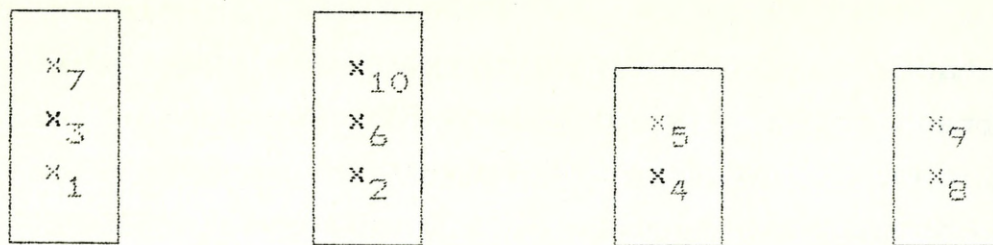
$$Y = \{x_2, x_3, x_4, x_6, x_{10}\}$$

Aproksymacja Y to

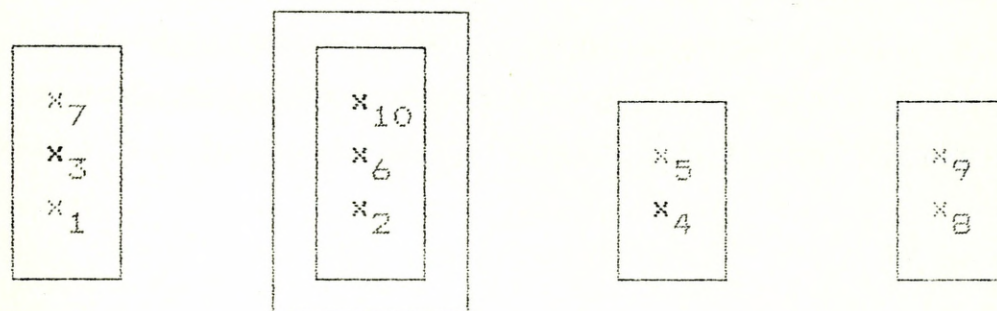
$$RY = \{x_2, x_6, x_{10}\}$$

$$\bar{R}Y = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_{10}\}$$

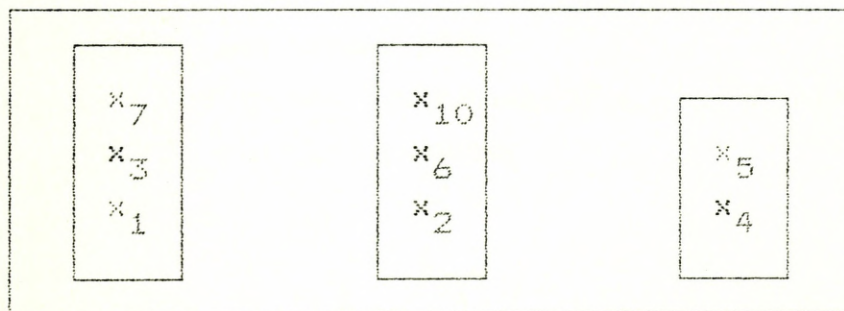
Interpretację graficzną aproksymacji zbioru Y przedstawiono na rysunku 6.



klasy równoważności



dolna aproksymacja



górną aproksymacja

$\{x_2, x_3, x_4, x_6, x_{10}\}$ - obiekty aproksymowane

Rysunek 7. Przykład aproksymacji dolnej i górnej

Stopień niesprzeczności wiedzy o obiektach zbioru Y określa współczynnik

$$\alpha_R(x) = 3:8 = 0,375.$$

Dokonane oszacowanie mówi nam o tym, że dla wybranego zbioru obiektów $Y = \{x_2, x_3, x_4, x_6, x_{10}\}$ spójna i niesprzeczna jest wiedza odnosząca się tylko do obiektów należących do

zbioru $RY = \{x_2, x_6, x_{10}\}$. Pełniejszą wiedzę o obiektach ze zbioru Y uzyskamy na podstawie wiedzy o obiektach należących do zbioru $RY = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_{10}\}$, choć będzie to wiedza niespójna i sprzeczna.

Postępując w ten sposób można oszacować wiedzę odnoszącą się do pozostałych decyzji. W wyniku uzyskamy dwie tablice decyzyjne, z których jedna będzie zawierać wiedzę pewną a druga - niepewną.

Wadą takiej reprezentacji wiedzy jest bardzo trudna analiza danych w tablicy zawierającej dane niespójne i sprzeczne. Daje ona jednak możliwość precyzyjnej analizy przynajmniej części posiadanej wiedzy z danej dziedziny.

Zakończenie

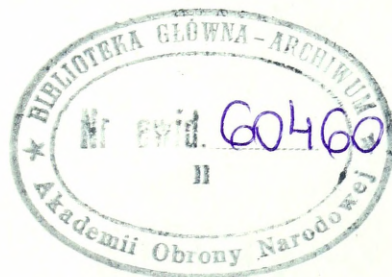
Przedstawiony materiał stanowi próbę przybliżenia Czytelnikowi problematyki metod reprezentacji wiedzy. W szczególności autorowi chodziło o przedstawienie możliwości ujęcia w metodach reprezentacji wiedzy danych, które są niepewne, niejednoznaczne lub sprzeczne. Potrzebę tego działania próbowano wykazać przytaczając proste przykłady z dziedziny ekonomii. Wynika ona z rosnącego znaczenia złożonych analiz ekonomicznych a szerzej- doradztwa ekonomicznego. Wydaje się koniecznym prowadzenie dalszych prac z tego zakresu ponieważ szybki rozwój informatyki umożliwi w szerokim zakresie stosowanie systemów ekspertowych.

Literatura

1. Badźmirowski K., Kubiś M.: Systemy ekspertowe.
PIE, Warszawa 1991.
2. Kacprzyk J.: Zbiory rozmyte w analizie systemowej.
PWN, Warszawa 1986.
3. Pawlak Z.: Systemy informacyjne. Podstawy teoretyczne.
WNT, Warszawa 1983.

Spis treści:

Wstęp	3
I. Wybrane metody reprezentacji wiedzy	7
1. Tablice decyzyjne	7
2. Sieci semantyczne	9
3. Ramy	11
II. Sposoby reprezentacji wiedzy w specyficznych warunkach informacyjnych	14
Wprowadzenie	14
1. Reprezentacja wiedzy niepewnej	14
2. Reprezentacja wiedzy niejednoznacznej	17
3. Reprezentacja wiedzy sprzecznej	19
Zakończenie	25
Literatura	25



Druk AON nr 350/WW