

215/2003

Raport Badawczy
Research Report

RB/57/2003

**Zastosowanie wybranych
metod uczenia maszynowego
w finansach**

**K. Mańczak, M. Krawczak,
G. Szkatuła**

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences



POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 8373578

fax: (+48) (22) 8372772

Kierownik Pracowni zgłaszający pracę:
Prof. dr hab. inż. Janusz Kacprzyk

Warszawa 2003

K. Mańczak, M. Krawczak, G. Szkatuła

ZASTOSOWANIE WYBRANYCH METOD UCZENIA MASZYNOWEGO W FINANSACH

Spis treści

1	Wprowadzenie	3
2	Elementy teorii chaosu do predykcji szeregów czasowych	4
2.1	Liniowe modele szeregów czasowych	5
2.2	Nieliniowe układy dynamiczne	8
2.3	Rekonstrukcja przestrzeni stanów	11
2.4	Analiza właściwości	14
2.5	Modelowanie sieciami neuronowymi	15
2.6	Obserwowalność stanów systemu dynamicznego	17
3	Indukcyjne uczenie na podstawie przykładów	21
3.1	Wprowadzenie	21
3.2	Opis zastosowanego podejścia	24
4	Elementy teorii zbiorów rozmytych w uczeniu maszynowym	33
4.1	Wprowadzenie	33
4.2	Podstawowe pojęcia teorii zbiorów rozmytych	33
4.3	Tworzenie reguł rozmytych na podstawie zbioru przykładów	52
5	Literatura	62

Zastosowanie wybranych metod uczenia maszynowego w finansach

1 Wprowadzenie

Rozwój systemów informatycznych ułatwił gromadzenie dużych zbiorów danych dotyczących rynków finansowych. Ważne stało się nie tylko efektywne przechowywanie takich danych, ale także ich analiza, zdolność interpretacji i wyciągania potencjalnie użytecznych i zrozumiałych dla człowieka wniosków. Wykorzystanie pełnej informacji zawartej w tych zbiorach oraz sformułowanie ukrytej w nich wiedzy zwykle przekracza możliwości percepcji człowieka. W związku z tym powstało zapotrzebowanie na nowe metody i narzędzia informatyczne, które mogą pomóc człowiekowi w odkrywaniu wiedzy z danych. Duże znaczenie zyskały metody uczenia maszynowego wykorzystujące informacje zawarte w dostarczonym zbiorze przykładów.

Baza danych zawierająca historyczne dane rynków finansowych może być wykorzystana do utworzenia tzw. *przykładów* danych *wejściowych* oraz odpowiadających im danych *wyjściowych*. Para przykładu danych $\{wejscia_p, wyjscie_p\}$ opisuje przykładowe zachowanie się określonego procesu na rynkach finansowych. Zebrane takie pary historycznych danych opisują dany proces rynków finansowych. Zakłada się, że takie same lub podobne przyczyny (dane wejściowe) powodują taki sam lub podobny skutek opisany przez dane wyjściowe. Zasada ta daje nam podstawy do przypuszczenia, że pojawienie się takich samych lub podobnych danych opisujących wejścia spowoduje takie same lub podobne wyjście w przyszłości.

Wiele zagadnień związanych z zarządzaniem inwestycjami finansowymi może być zaliczone do jednej z dwóch kategorii problemów obliczeniowych:

- *aproxymacji funkcji* lub
- *klasyfikacji zbiorów*.

Zagadnienia dotyczące predykcji finansowych szeregów czasowych należą do kategorii *aproxymacji funkcji*. Szeregi czasowe są podstawą do scharakteryzowania obserwowanego zjawiska i do predykcji przyszłych stanów tego zjawiska. Poprzez obserwowane zjawisko rozumiemy tutaj złożony system, którego stan (lub stany) mogą być obserwowane w postaci szeregu czasowego (lub szeregów czasowych). Do takich systemów możemy zaliczyć systemy ekonomiczne, a obserwowanymi stanami takiego systemu są stopy procentowe.

Predykcja oraz chęć (albo często potrzeba) zrozumienia przeszłości skłania nas do poszukiwania praw, które wyjaśniają istotę zachowania się obserwowanego zjawiska. W przypadku, gdy są znane deterministyczne równania opisujące zjawisko, to rozwiązanie tych równań pozwala określić stan zjawiska na podstawie warunków początkowych. W przypadku, gdy nie są znane równania opisujące obserwowane zjawisko, a jesteśmy zainteresowani w predykcji stanów przyszłych, musimy określić zarówno reguły rządzące zjawiskiem (modelowanie matematyczne) jak i aktualny stan. W tym rozdziale ograniczymy się do przypadku drugiego, tzn. gdy nie są znane równania opisujące interesujące nas zjawisko, a poszukiwane reguły opisujące dynamikę tego zjawiska muszą być uzyskane z przebiegu historii tego zjawiska. Dodatkowo, teoria informacji pozwala wyjaśnić pewne cechy charakterystyczne dla obserwowanego zjawiska.

Zagadnienia np. porównywania akcji firm, wyboru branży do inwestowania, czy też oceny wiarygodności kredytobiorcy należą do kategorii *klasyfikacji zbiorów*. W problemie *klasyfikacji* dysponujemy zbiorem *obiektów* (lub *przykładów*) - czyli zestawem danych wejściowych, należących do pewnej *klasy*, klasa obiektu jest określana jako wartość wyjściowa.

Każdy obiekt jest opisany przez skończoną liczbę *atrybutów* (*cech*) wraz z określonymi *wartościami* tych *atrybutów* (*wartości cech*).

Takie problemy mogą być rozwiązywane metodami tzw. *inteligencji komputerowej* (lub tzw. *sztucznej inteligencji*). Metody sztucznej inteligencji, takie jak sztuczne sieci neuronowe pozwalają na konstrukcję modelu - typu *black box* - symulującego zjawisko, zaś metody *indukcyjnego uczenia maszynowego* pozwalają na opis zjawiska w postaci skończonej liczby reguł w postaci *JEŻELI ... TO ...*. W obu przypadkach zależy nam na narzędziach pozwalających wyjaśnić obserwowane zjawisko, przy czym w przypadku reguł otrzymane przesłanki w sposób jawny opisują zasady podejmowania decyzji.

W części 2 przedstawiono analizę finansowych szeregów czasowych, których źródłem są nieliniowe systemy dynamiczne, analiza których prowadzi do możliwości wykorzystania elementów chaosu deterministycznego do predykcji. Rozważania rozpoczęto od analizy liniowych systemów dynamicznych, następnie przedstawiono problem rekonstrukcji przestrzeni stanów dla systemów dynamicznych z punktu widzenia teorii informacji, rozważono także zastosowanie sztucznych sieci neuronowych do predykcji stanów szeregów czasowych.

W części 3 ograniczono się do zagadnień uczenia nadzorowanego, w którym tworzone są opisy klas zawierających przykłady posiadające pewne wspólne właściwości, które odróżniają je od innych klas. Zadanie uczenia maszynowego na podstawie przykładów sformułowano w postaci ciągu zadań cząstkowych, zapisanych w postaci pewnych modyfikacji zadania pokrycia. Zaproponowano metodę tworzącą reguły typu *JEŻELI ... TO ...*, prezentującą istniejącą regularność w przykładach, z możliwością uwzględniania założenia o poprawnym opisywaniu przez tworzone reguły „większości” przykładów uczących.

Natomiast część 4 poświęcono *zbiorom rozmytym* zaproponowanym przez Lotfi A. Zadeha w 1965 roku jako numeryczne narzędzie do opisu niepewności, nieprecyzyjności, wieloznaczności towarzyszących nieodłącznie ludzkiej percepcji, myśleniu, mówieniu itp. W teorii zbiorów rozmytych dowolny element może należeć do zbioru w pewnym stopniu. Stopień przynależności elementu ze zbioru U do zbioru rozmytego \tilde{G} jest określany przez funkcję przynależności o wartościach z przedziału $[0, 1]$, oznaczaną $\mu_{\tilde{G}}:U \rightarrow [0,1]$. Przykładowo, pojęcia „małe ryzyko” i „średnie ryzyko” mogą być różnie odbierane przez różnych ludzi. Kredyt w wysokości 55 mln może być oceniany zarówno jako mały jak też jako średni; może więc należeć zarówno do zbioru rozmytego „mały kredyt” jak też do zbioru rozmytego „średni kredyt” w różnym stopniu.

2 Elementy teorii chaosu do predykcji szeregów czasowych

Analiza szeregów czasowych ma trzy zasadnicze cele: prognozowanie, modelowanie oraz analizę właściwości. Celem prognozowania (predykcji) jest próba dokładnego przewidywania zmiany stanu systemu, dla krótkiego horyzontu czasowego. Modelowanie ma na celu znalezienie opisu zmian stanu systemu w długim horyzoncie czasowym. Trzeci cel analizy szeregów czasowych służy do określenia fundamentalnych właściwości, takich jak np. liczba stopni swobody systemu.

Za początek nowoczesnych metod prognozowania można przyjąć prace Yulea (1927), który zaproponował metodę *autoregresji* do prognozowania liczby plam na słońcu. W modelu Yulea prognoza przyszłej liczby plam była ważoną sumą poprzednio zaobserwowanych ilości plam na słońcu. Przez wiele lat liniowy model Yulea, uzupełniony przez szum, był podstawową metodą używaną do predykcji szeregów czasowych.

Rozwinięta metoda metody *rekonstrukcji przestrzeni stanów* – jest to idea wykorzystania elementów teorii systemów (*obserwowalność stanów*) pozwala określić czy szereg cza-

wy jest opisem systemu deterministycznego czy stochastycznego (Krawczak i in. 2000; Krawczak 2002a, 2002b; 2003a).

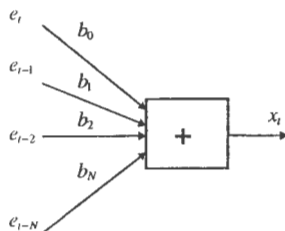
2.1 Liniowe modele szeregów czasowych

Liniowe modele szeregów czasowych są intuicyjnie łatwe do zrozumienia oraz do zastosowania. Jednak modele liniowe zupełnie nie nadają się do opisu trochę bardziej złożonych systemów. Historycznymi pozycjami literatury dotyczącymi modeli liniowych są prace Kołmogorowa z 1941 roku i Wienera z 1947 roku, współczesną, ale już klasyczną, pozycją literatury jest książka Boxa i Jenkinsa z 1976 roku.

Założmy, że jest dany ciąg zewnętrznych wejść $\{e_t\}$, które mają wpływ na obserwowany szereg czasowy $\{x_t\}$. Następnie zakładając liniowość systemu oraz obowiązywanie zasady przyczynowości (tzn. że na aktualną wartość x_t ma wpływ zarówno aktualna wartość e_t jak i N poprzednich wartości założonego wejścia e_{t-1}, \dots, e_{t-N}) możemy przytoczyć następującą zależność między wejściem i wyjściem modelu,

$$x_t = \sum_{n=0}^N b_n e_{t-n} = b_0 e_t + b_1 e_{t-1} + \dots + b_N e_{t-N}. \quad (1)$$

Zależność (1) została przedstawiona w postaci prostego liniowego systemu na Rysunku 1. W teorii systemów równanie (1) opisuje filtr, tzn. generator nowego szeregu czasowego $\{x_t\}$ na podstawie szeregu $\{e_t\}$ przez liniowy filtr, którego współczynnikami są b_0, \dots, b_N . Specjaliści od ekonometrii lub statystycy nazywają równanie (1) modelem *średniej ruchomej rzędu N* , $MA(N)$ (*Moving Average model*). Fizyczne znaczenie równania (1) jest następujące, mianowicie filtr uśrednia kilka - kilkanaście ostatnich wartości wejść.



Rysunek 1. Model średniej ruchomej (MA)

Wśród inżynierów jest znane pojęcie filtra impulsowego o skończonej odpowiedzi (*Finite Impulse Response*), filtr ten charakteryzuje się tym, że gdy zanika wejście po N krokach to również zanika wyjście filtru.

Własności nowego wyjściowego szeregu czasowego $\{x_t\}$ zależą od własności wejściowego szeregu czasowego $\{e_t\}$. Zadaniem naszym jest określenie takiego systemu, który byłby niezależny od określonego ciągu wejściowego. Dla systemów liniowych odpowiedź filtru jest niezależna od wejścia.

Przedstawimy teraz trzy różne, ale jednocześnie równoważne, postacie modelu MA, a mianowicie określone: w dziedzinie czasowej - jako impulsowa odpowiedź filtru; w dziedzi-

nie częstotliwości - jako gęstość widmowa; w kategoriach współczynników autokorelacji. W pierwszym przypadku zakłada się, że wejście przybiera wartości różne od zera dla punktu t_0 , a w pozostałych punktach wejście jest zerowe. Odpowiedź czasowa na wejściowy impuls jest dana przez współczynniki b_i równania (1); tzn. dla każdego kroku impuls wejściowy przesuwa się, aż dla $t = N$ wyjście systemu jest równe zero. Ciąg współczynników b_N, \dots, b_0 jest zatem odpowiedzią impulsową systemu. Odpowiedź na konkretne wejście może być traktowana jako superpozycja odpowiedzi z odpowiednimi opóźnieniami. Funkcja przejścia modelu dokładnie opisuje system liniowy, tzn. taki system, w którym obowiązuje zasada superpozycji i wyjście jest określone poprzez odpowiedź impulsową oraz wejście. Czasami wygodniej jest rozważać filtr w dziedzinie częstotliwości. Jeżeli wejściem modelu typu MA jest impuls, to dyskretna transformata Fouriera wyjścia jest równa

$$\sum_{n=0}^N b_n \exp(-i2\pi n f). \quad (2)$$

Gęstość widmowa jest dana przez kwadrat amplitudy wyrażenia:

$$|b_0 + b_1 e^{-i2\pi f} + \dots + b_N e^{-i2\pi N f}|^2. \quad (3)$$

Trzecia postać modelu MA, która przedstawia te same informacje, jest opisywana w kategoriach współczynników autokorelacji, tzn. w kategoriach wartości średniej $\mu = E[x_i]$ i wariancji $\sigma^2 = E[(x_i - \mu)^2]$, w postaci

$$\rho_\tau \equiv \frac{1}{\sigma^2} E[(x_i - \mu)(x_{i-\tau} - \mu)]. \quad (4)$$

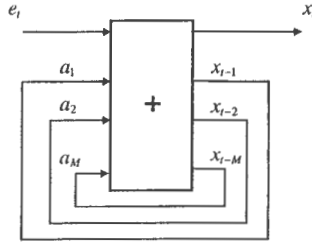
Współczynniki autokorelacji określają jak bardzo dwie realizacje szeregu czasowego, oddalone od siebie o τ kroków, różnią się od siebie w sensie wartości średniej. W przypadku, gdy wejście systemu jest procesem stochastycznym o wartościach nieskorelowanych wzajemnie ze sobą, tzn. $E[e_i e_j] = 0$ dla $i \neq j$, współczynniki autokorelacji są określone następującym wyrażeniem

$$\rho_\tau = \begin{cases} \frac{1}{\sum_{n=0}^N b_n^2} \sum_{n=\tau}^N b_n b_{n-\tau} & |\tau| \leq N \\ 0 & |\tau| > N \end{cases}. \quad (5)$$

Z punktu widzenia teorii systemów filtry MA pracują ze sprzężeniem zwrotnym i dlatego sygnał wejściowy jest przenoszony bezpośrednio na wyjście modelu. Poprzez wprowadzenie do modelu sprzężenia zwrotnego można otrzymać model dynamiczny

$$x_i = \sum_{m=1}^M a_m x_{i-m} + e_i. \quad (6)$$

Taki model jest nazywany w literaturze modelem autoregresyjnym rzędu M (AR(M)) i jest przedstawiony na Rysunku 2. W zależności od zastosowań człon e_i może oznaczać albo wejście, które może być w sposób zamierzony zmieniane, lub też szum. W przypadku, gdy $\{e_i\}$ jest białym szumem to współczynniki autokorelacji szeregu wyjściowego x_i można wyrazić przez współczynniki modelu.



Rysunek 2. Model autoregresji

Ze względu na wyprowadzenie sprzężenia zwrotnego każdy współczynnik autokorelacji jest rozwiązaniem odpowiedniego układu równań liniowych. Poprzez pomnożenie równania (6) przez $x_{t-\tau}$, a następnie wyznaczenie wartości oczekiwanej i normalizację (Box - Jenkins, 1976), współczynniki autokorelacji modelu AR można wyznaczyć rozwiązując układ równań liniowych, zwany równaniami Yula-Walkera:

$$\rho_t = \sum_{m=1}^M a_m \rho_{t-m}, \tau > 0. \quad (7)$$

W przeciwieństwie do modelu MA współczynniki autokorelacji nie muszą równać się zeru po M krokach. Wyznaczając transformaty Fouriera równania (6) można otrzymać następującą zależność:

$$\text{wyjscie} = \frac{\text{wejście}}{1 - \sum_{m=1}^M a_m \exp(-i2\pi mf)}, \quad (8)$$

zaś gęstość widmowa wyjścia jest dana jako:

$$\frac{1}{|1 - a_1 e^{-i2\pi f} - a_2 e^{-i2\pi 2f} - \dots - a_M e^{-i2\pi Mf}|^2}. \quad (9)$$

W celu wygenerowania określonej realizacji szeregu muszą być określone warunki początkowe, które z reguły przyjmuje się jako pierwsze M wartości szeregu $\{x_t\}$.

Następnym krokiem może być połączenie modeli AR i MA w jeden model, który jest nazywany ARMA(M, N)

$$x_t = \sum_{m=1}^M a_m x_{t-m} + \sum_{n=0}^N b_n e_{t-n}. \quad (10)$$

Zastosowanie przekształcenia z (Oppenheim-Schafer, 1989), które opisuje dyskretną transformację Fouriera

$$X(z) \equiv \sum_{l=-\infty}^{\infty} x_l z^l, \quad (11)$$

pozwała uzyskać ciekawe wyniki. Przekształcenie z , na jednostkowym kole $z = \exp(-i2\pi f)$, ogranicza się do dyskretnej transformaty Fouriera, natomiast na zewnątrz jednostkowego koła przekształcenie z określa stopień zbieżności szeregu. Przekształcenie z wyjścia modelu AR-MA daje następujące wyrażenie:

$$X(z) = A(z)X(z) + B(z)E(z) = \frac{B(z)}{1 - A(z)}E(z), \quad (12)$$

bez uwzględniania warunków początkowych.

Przekształcenie z zastosowane do szeregu czasowego generowanego przez równanie (10) może składać się z dwóch członów, pierwszy człon to funkcja rzeczywista, a drugi człon jest zależny od wejścia. Liczba biegunów i zer określa liczbę stopni swobody systemu.

Układ liniowych równań Yule'a-Walkera (równanie (7)) pozwala na wyrażenie współczynników autokorelacji szeregu czasowego w kategoriach współczynników modelu AR. Z drugiej strony równanie (7) pozwala na estymację współczynników modelu AR(M) na podstawie struktury obserwowanego szeregu. Estymacja współczynników jest właśnie zadaniem regresji, tzn. wyrażeniem następczej (przyszłej) wartości jako funkcji M poprzednich wartości; tzn. wyrażenie x_t jako funkcji $\{x_{t-1}, \dots, x_{t-M}\}$.

Zadanie to można sformułować jako zadanie minimalizacji kwadratu błędu, tzn. parametry (współczynniki) są określone w taki sposób, że tworzy się kwadraty różnicy między wyjściem modelu i obserwowaną wielkością i sumuje dla każdego kroku, a następnie taką funkcję błędu minimalizuje się ze względu na poszukiwane współczynniki. Istnieją dostatecznie proste metody numeryczne pozwalające w sposób rekurencyjny wyznaczyć poszukiwane współczynniki (Box-Jenkins, 1976; Press, 1992).

Model (10) należy raczej traktować jako liniową aproksymację rzeczywistego systemu. Rozważany problem wyznaczania współczynników modelu ARMA jest także określany przez rząd modelu (M, N). Zagadnienie doboru rzędu modelu ARMA w ogólności nie ma jednoznacznego rozwiązania, zarówno w sensie doboru wartości współczynników, jak również doboru liczby poszukiwanych współczynników modelu. Istnieje kilka sposobów doboru właściwego rzędu modelu (Akaike, 1974), jednak głównym ich założeniem jest liniowość modelu oraz szczególny rozkład szumu.

Rozważane trzy różne reprezentacje modelu ARMA, tzn. poprzez współczynniki, gęstość widmową i współczynniki autokorelacji przechowują dokładnie te same informacje dotyczące systemu liniowego z nieskorelowanym białym szumem. Przy założeniu, że gęstość widmowa dobrze charakteryzuje cechy szeregu czasowego, to model typu ARMA jest oczywiście dobrą reprezentacją; jednak występowanie prostej nieliniowości w systemie może doprowadzić do bardzo złożonej gęstości widmowej. Dwa różne szeregi czasowe mogą mieć bardzo zbliżoną szerokość pasma widma, ale te szeregi czasowe mogą charakteryzować się zupełnie odmiennymi cechami. Dzieje się tak np. gdy porównujemy jeden szereg czasowy opisujący liniowy system zakłócony szumem i drugi nieliniowy system deterministyczny (bez szumu) o małej liczbie stopni swobody – są to właśnie przykłady, do których nie mogą być stosowane operatory liniowe.

2.2 Nieliniowe układy dynamiczne

Rozważmy teraz dwa nieliniowe przykłady:

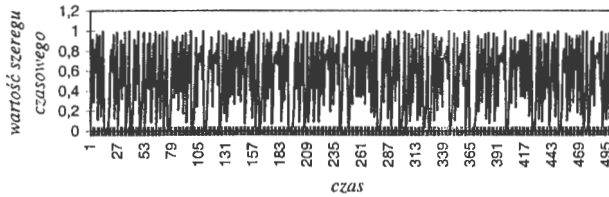
- pierwszy przykład został przytoczony podany w 1947 r. przez Ulama i von Neumanna (1947). W tym przykładzie model jest następujący, mianowicie wartość następcza w szere-

gu czasowym jest określana z wartości bieżącej wykorzystując prostą aproksymację kwadratową

$$x_t = \lambda x_{t-1}(1 - x_{t-1}). \quad (13)$$

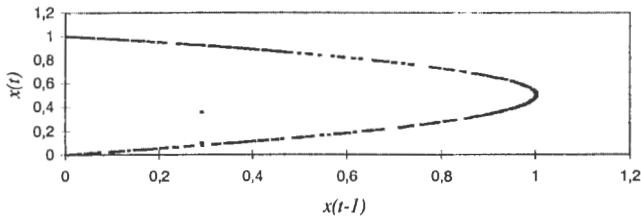
Ten model został spopularyzowany jako model rozwoju populacji, często w literaturze wspominany jako "prosty model charakteryzujący złożoną dynamikę systemu".

Warto zwrócić uwagę na fakt, że nawet takie proste równanie dynamiki systemu generuje bardzo złożony szereg czasowy, na Rysunku 3 dla 500 punktów, przyjmując $\lambda = 4$.



Rysunek 3. Szereg czasowy generowany przez równanie (13)

Kwadratowe równanie (13) opisuje dynamikę wielu systemów, np. dynamikę przepływu cieczy, dynamikę w reaktorach chemicznych czy też dynamikę rynków finansowych. Równanie to jest nazywane *odwzorowaniem logistycznym* lub też *odwzorowaniem kwadratowym*.



Rysunek 4. Dziwny atraktor równania (13) dla $\lambda = 4.0$

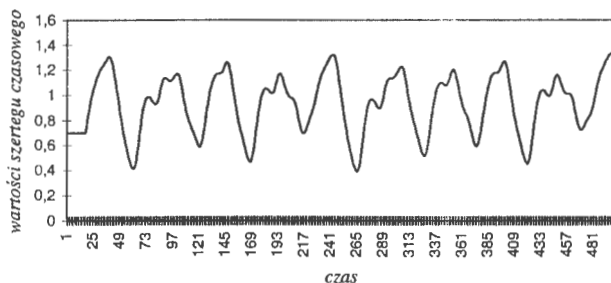
Warto zauważyć, że stosunkowo proste równanie (13) generuje złożony szereg czasowy. Dla rozpatrywanego szeregu czasowego można określić *dziwny atraktor* (Peitgen i in. 1996) w następujący sposób, mianowicie jest wykres zależności wartości modelu x_t w chwili t od poprzedniej wartości modelu x_{t-1} w chwili $t-1$. Dziwny atraktor odwzorowania logistycznego przedstawiony jest na Rysunku 4.

Wartość szeregu w modelu x_t w chwili t w sposób deterministyczny zależy od wartości poprzedniej x_{t-1} ; parametr λ jest odpowiedzialny za jakościowe aspekty modelu (np. parametr ten wpływa na "rozwarcie" paraboli). Powyższe zagadnienie opisuje tzw. *chaos de-*

terministyczny, który jest ściśle związany z wartością parametru λ . Przykładowo, dla $\lambda = 4$, jedna wartość x_{t-1} określa dwie różne wartości x_t , te dwie różne wartości są równoważne. Warto zauważyć, że jeżeli znane jest położenie punktu x_{t-1} z dokładnością do pewnej wartości parametru ε , to znajomość x_t jest określona z dokładnością 2ε . Taka wykładnicza zmiana dokładności określenia wartości x_t związana z każdą iteracją i klasyfikuje zachowanie modelu jako *deterministyczny chaos* (Schuster, 1995).

- drugi przykład jest równie prosty, rozważmy teraz szereg czasowy generowany przez równanie Mackey'a-Glassa (1977):

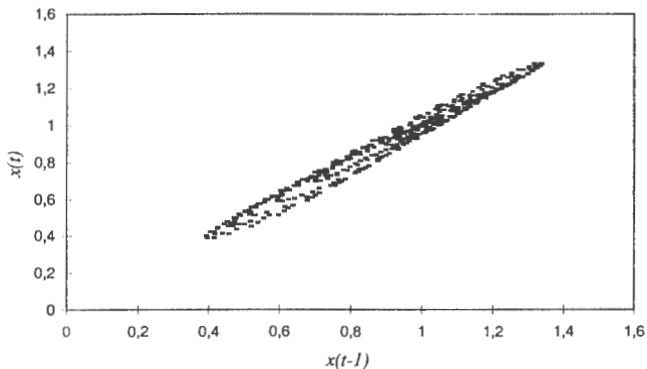
$$x(t+1) = bx(t) + a \frac{x(t-s)}{1+x^c(t-s)}. \quad (14)$$



Rysunek 5. Pierwsze 500 punktów równania Mackey'a-Glassa

Dla przykładowych parametrów: $a = 0.2$, $b = 0.9$, $c = 10$, $s = 18$ oraz warunków początkowych $x(0) = x(1) = \dots = x(18) = 0.7$ przebieg czasowy szeregu przedstawiony jest na Rysuku 5.

Dziwny atraktor równania Mackey'a-Glassa przedstawiony jest na Rysunku 6.



Rysunek 6. Dziwny atraktor równania Mackeya-Glassa

Oba przedstawione modele systemów są całkowicie deterministyczne, tzn. rozwój tych systemów jest całkowicie określony przez warunki początkowe x_0 (odpowiada to szerokopasmowej gęstości widmowej). W przypadku modelu ARMA występowanie szerokopasmowego składnika gęstości widmowej musi pochodzić od zewnętrznego szumu oddziałującego na wejścia systemu, jednak w rozpatrywanych tutaj dwóch przypadkach jednowymiarowych systemów odpowiada to paraboli, w pierwszym przypadku, i elipsie w drugim przypadku. W systemach deterministycznych właśnie nieliniowości charakteryzują dynamikę systemów i w takich przypadkach rozpatrywanie modeli liniowych może być bardzo ograniczone.

W literaturze dotyczącej teorii systemów rozważane są dwa rodzaje systemów, pierwszy rodzaj to układ kilku równań z niewielką liczbą parametrów opisujących obserwowane zjawisko z pewnym "nadmiarem"; w modelach drugiego rodzaju występuje bardzo mało założeń, a brak wiedzy o obserwowanym zjawisku jest kompensowany wprowadzaniem dużej liczby parametrów, których znaczenie trudno jest określić.

2.3 Rekonstrukcja przestrzeni stanów

Idea Yula dotycząca predykcji w szeregach czasowych polega na tym, że przyszła wartość szeregu jest określona przez wartości szeregu bezpośrednio ją poprzedzające. Model ARMA opisany równaniem (10) można przedstawić w następującej postaci

$$x_t = a \cdot x_{t-1} + b \cdot e_t, \quad (15)$$

gdzie

$$x_t = [x_t, x_{t-1}, \dots, x_{t(d-1)}]' \quad \text{i} \quad x = [a_1, a_2, \dots, a_d]'$$

W przedstawionych rozważaniach notacja jest trochę zmieniona, tzn. poprzednio używane było oznaczenie M jako rząd modelu typu AR, a teraz oznaczyliśmy rząd tego modelu przez d i jest on rozważany jako opóźnienie. W teorii informacji znany jest wpływ opóźnienia na dynamikę systemu, wpływ ten został zasygnalizowany przez Ruella (1980), a sam fakt został udowodniony przez Takensa (1981), a następnie uściślony przez Sauera, Yorka i Casdagliego (1991). Idea stosowania wektorów opóźnień o odpowiedniej długości nie jest sposobem opisu jedynie systemów liniowych, ale jest również użyteczna do wyeksponowania pełniejszej geometrycznej struktury systemu nieliniowego. Analizując szeregi czasowe chcielibyśmy wyciągnąć możliwie jak najwięcej informacji o obserwowanym systemie, ale bez uprzednich założeń o postaci równań. Możemy bowiem założyć, że dane generowane są przez system dynamiczny, którego dynamikę może opisywać układ nieznanych równań różniczkowych.

W analizie i predykcji szeregów czasowych rozważa się cztery różne przestrzenie:

1. *przestrzeń stanu* - jest to przestrzeń układu równań opisujących system; przestrzeń ta precyzuje maksymalną liczbę stopni swobody systemu,
2. *przestrzeń rozwiązań* - jest to z reguły podprzestrzeń przestrzeni stanu; ze względu na fakt, że często nie wszystkie stopnie swobody są uaktywnione lub pewne stopnie swobody są skorelowane, wymiar przestrzeni rozwiązań może być znacznie mniejszy niż wymiar przestrzeni stanu; wymiar przestrzeni rozwiązań oznacza liczbę parametrów, które jednoznacznie określają system; często zdarza się, że dynamikę systemu opisuje układ równań różniczkowych cząstkowych, który może być zastąpiony układem równań różniczkowych zwyczajnych (Lorenz, 1963); redukcja wymiarowości systemu jest bardzo charakterystyczna dla systemów dyssypatywnych (rozproszonych): rozproszenie systemu często może być zmniejszone do podprzestrzeni przestrzeni stanu,

3. *przestrzeń obserwacji* - jest to przestrzeń z reguły ograniczona do jednego wymiaru; zdarza się często, że jesteśmy w stanie obserwować pewien system jako skalarną funkcję zmieniających określających przestrzeń stanu; przykładowo to może być właśnie obserwowana stopa procentowa, poziom bezrobocia czy też inflacja,
4. *przestrzeń zrekonstruowana* - jest to przestrzeń otrzymana ze skalarnej obserwacji systemu po uwzględnieniu obserwacji historycznych (opóźnienia).

Analizując finansowe szeregi czasowe, przy braku innych informacji mogących mieć wpływ na przebieg tego szeregu finansowego, naszym zadaniem jest określenie jakie informacje można uzyskać o systemie i jego dynamice na podstawie jedynie obserwacji tego szeregu.

Założmy, że y jest wektorem stanu w przestrzeni rozwiązań, niech równanie $dy/dt = f(y)$ określa dynamikę systemu, a obserwowana wielkość szeregu czasowego będzie określona jako $x_t = x(y(t))$. Następnie oznaczmy opóźnienie przez τ , wymiar przez d , zaś wektor opóźnień możemy zdefiniować w następujący sposób

$$x_t = [x_t, x_{t-\tau}, \dots, x_{t-(d-1)\tau}]'. \quad (16)$$

Zależność między x i y jest nazywana *zanurzeniem (embedding)* przestrzeni obserwacji w przestrzeni stanu. Zanurzenie oznacza, że x zależy przynajmniej od kilku współrzędnych wektora y , zaś pozostałe współrzędne wektora y są sprzężone i jako całość mają wpływ na x .

Problem wykorzystania opóźnień i zanurzania jest stosunkowo nowym podejściem do analizy szeregów czasowych. Podejście to gwarantuje uzyskanie dokładnej informacji o zachowaniu się nieobserwowalnych aktualnie stopni swobody systemu, pozwala zapomnieć o ścisłych założeniach stochastycznych oraz pozwala rozważać tylko te wielkości, które są istotne (czyli niezmiennicze).

Analiza właściwości szeregów czasowych opiera się na rozważaniach dotyczących używanych w teorii informacji miar redundancji i entropii systemu, których podstawą jest opis liczby bitów, jakie nowy pomiar daje o systemie. Początkowe rozważania dotyczą samego szeregu czasowego, w oderwaniu od obserwowanego systemu, zaś w następnym etapie należy przeprowadzić rozważania dotyczące prognozowania i oszacowania błędu.

Złożone szeregi czasowe (o skomplikowanym przebiegu) mogą być generowane zarówno przez stosunkowo proste (jak było pokazane wcześniej), jak i bardzo złożone systemy. Rozważymy metody analizy właściwości szeregów czasowych, które mogą być używane do określenia istotnych cech systemu (cech poza obserwowanymi wielkościami) przydatnych do analizy, modelowania i do prognoz. Posłużymy się elementami teorii informacji w celu wprowadzenia podstawowej miary predykcji dla szeregów czasowych. Początki metod analizy szeregów czasowych sięgają prac Schustera z 1898 roku. W przypadku prostych liniowych systemów tradycyjna analiza spektralna jest przydatnym i dostatecznym narzędziem, ale jednocześnie inne systemy nieliniowe mogą mieć bardzo podobne pasma gęstości widmowej.

Związki między teorią informacji i twierdzeniem o ergodyczności są dość dobrze znane, ale Shaw (1981) połączył dynamikę systemów dyssypatywnych z teorią informacji, zaś Fraser (1989) używając teorii informacji podał sposób wyznaczania optymalnej wielkości opóźnienia.

Założmy, że szereg czasowy $x(t)$ został zdyskretyzowany na N równych części. Jeżeli n_T oznacza liczbę obserwowanych punktów i poszczególne wartości szeregu powtarzają się

n_x razy, to prawdopodobieństwo estymacji tej wartości wynosi $p_1(x) = n_x/n_T$ (indeks 1 oznacza rozważanie jednowymiarowego rozkładu). Entropia takiego rozkładu jest określona przy pomocy tego prawdopodobieństwa w następujący sposób:

$$H_1(N) = - \sum_{x=1}^M p_1(x) \log_2 p_1(x). \quad (17)$$

Rozważmy zatem dwa skrajne przypadki. W pierwszym zakładamy, że istnieje tylko pojedyncza wartość x i prawdopodobieństwo p występowania tej wartości jest równe 1, $\log_2 p_1(x) = 0$. O pozostałych wartościach zakładamy, że występują z prawdopodobieństwem równym 0. W tym przypadku następny stan jest określony z prawdopodobieństwem 1 i nazywany jest on przypadkiem *największego uporządkowania*. W drugim przypadku rozważmy wszystkie N wartości tak samo prawdopodobnych. W tym przypadku entropia przyjmuje maksymalną wartość równą $H_1(N) = \log_2 N$ i jest on nazywany przypadkiem *największego nieporządku* - ponieważ każdy stan jest tak samo prawdopodobny. Oczywiście wszystkie pozostałe przypadki znajdują się pomiędzy tymi dwoma ekstremalnymi. Rozważmy teraz przypadek, w którym M stanów szeregu czasowego jest tak samo prawdopodobnych. Ze wzrostem N wartość entropii najpierw rośnie odpowiednio do $\log_2 N$ (pojawienie się wszystkich wartości jest jednakowo prawdopodobne) aż do asymptotycznego osiągnięcia wartości $\log_2 M$. W przypadku bardziej złożonego rozkładu prawdopodobieństwa wartość entropii będzie rosła do wartości $D_1 \log_2 N$, gdzie $D_1 \leq 1$. Współczynnik D_1 nazywany jest wymiarem informacji (Hentschel i Procaccia, 1983).

Prawdopodobieństwo konkretnego wektora opóźnień w przestrzeni d -wymiarowej $x_r = [x_r, x_{r-\tau}, \dots, x_{r-(d-1)\tau}]'$ jest szacowane w podobny sposób $p_d(x) = \frac{n_x}{n_r}$. Całkowita entropia jest określana następującym wyrażeniem

$$H_d(\tau, N) = - \sum_{x=1}^N \dots \sum_{x_{-(d-1)\tau}=1}^N \dots p_d(x_r, x_{r-\tau}, \dots, x_{r-(d-1)\tau}) \log_2 p_d(x_r, x_{r-\tau}, \dots, x_{r-(d-1)\tau}). \quad (18)$$

Wyrażenie (18) określa średnią liczbę bitów potrzebnych do opisu szeregu czasowego.

Innym terminem znanym w teorii informacji, który znajduje zastosowanie w analizie szeregów czasowych, jest *informacja wzajemna* określana między dwiema próbkami:

$$\begin{aligned} I_2(\tau, N) &= - \sum_{x_1=1}^N p_1(x_1) \log_2 p_1(x_1) - \sum_{x_{1-\tau}=1}^N p_1(x_{1-\tau}) \log_2 p_1(x_{1-\tau}) \\ &+ \sum_{x_1=1, x_{1-\tau}=1}^N p_2(x_1, x_{1-\tau}) \log_2 p_1(x_1, x_{1-\tau}) \\ &= 2H_1(\tau, N) - H_2(\tau, N) \end{aligned} \quad (19)$$

Jeżeli próbki są statystycznie niezależne

$$p_2(x_1, x_{1-\tau}) = p_1(x_1) p_1(x_{1-\tau})$$

to informacja wzajemna staje się równa zero (oznacza to, że nie można wyciągnąć żadnej informacji o drugiej próbce znając pierwszą). Jednak, gdy pierwsza próbka określa całkowicie drugą próbkę, tzn. $H_1 = H_2$, wówczas wspólna informacja jest równa entropii skalarnej

$I_2 = H_1$. Dla pozostałych przypadków informacja wzajemna wyznacza w bitach stopień określenia jednej zmiennej poprzez znajomość drugiej.

Informacja wzajemna może być uogólniona na przypadek wielowymiarowy przez :

- połączoną informację wzajemną

$$I_d(\tau, N) = dH_1(\tau, N) - H_d(\tau, N) \quad (20)$$

- lub przez nadmiarową informację wzajemną

$$R_d(\tau, N) = H_1(\tau, N) + H_{d-1}(\tau, N) - H_d(\tau, N). \quad (21)$$

Nadmiarowość określa średnią liczbę bitów obserwacji, która może być określona na podstawie $d-1$ poprzednich obserwacji. Zależność pomiędzy połączoną informacją wzajemną i nadmiarowością określa następujący wzór $R_d = I_d - I_{d-1}$.

W przypadku systemów opisywanych układem równań różniczkowych stany przeszłe jednoznacznie określają stany przyszłe systemu. Jeżeli liczba d jest dużo mniejsza od minimalnego wymiaru zanurzenia, to liczba $d-1$ określająca poprzednie obserwacje nie jest w stanie określić następczej obserwacji i w takim przypadku wartość nadmiarowości dąży do zera:

$$\begin{aligned} p_d(x_t, x_{t-\tau}, \dots, x_{t-(d-1)\tau}) &= p_1(x_t) p_{d-1}(x_{t-\tau}, \dots, x_{t-(d-1)\tau}) \\ \Rightarrow H_d &= H_1 + H_{d-1} \Rightarrow R_d = 0 \end{aligned} \quad (22)$$

W przypadku zaś, gdy d jest dużo większe niż wymagany wymiar zanurzania, to nowa obserwacja staje się nadmiarowa, wręcz zbędna:

$$\begin{aligned} p_d(x_t, x_{t-\tau}, \dots, x_{t-(d-1)\tau}) &= p_{d-1}(x_{t-\tau}, \dots, x_{t-(d-1)\tau}) \\ \Rightarrow H_d &= H_{d-1} \Rightarrow R_d = H_1 \end{aligned} \quad (23)$$

Można pokazać, że minimalna wartość nadmiarowości jest równa minimalnemu wymiarowi zanurzania, tzn. takiej najmniejszej wymiarowości przestrzeni euklidesowej, w której rozwiązania systemu (trajektorie) nie przecinają się.

2.4 Analiza właściwości systemu

Źródło entropii, zwanej też entropią Kołmogorowa-Sinaia, oznaczane jest przez $h(\tau, N)$ i jest zdefiniowane jako asymptotyczny wskaźnik wzrostu informacji:

$$h(\tau, N) = \lim_{N \rightarrow \infty} \lim_{d \rightarrow \infty} H_d(\tau, N) - H_{d-1}(\tau, N). \quad (24)$$

Pojęcie źródła entropii odnosi się do dodatnich wykładników *Lapunowa* (Ruelle, Eckmann, 1985) określanymi w następujący sposób:

$$h(\tau) = \lambda_i(1) = \tau \sum_i \lambda_i^+ \quad (25)$$

Wykładniki *Lapunowa* λ_i są wartościami własnymi lokalnej linearyzacji modelu dynamiki systemu, określają one średni wskaźnik odchylenia osi głównej od pobliskiej trajektorii. Wykładniki *Lapunowa* można wyznaczyć z jakobianu, bądź analizując trajektorię systemu (Brown, Bryant, Abarbanel, 1991), lub też stosując sztuczne sieci neuronowe.

W przypadku, gdy wymiar zanurzenia d jest dostatecznie duży, to nadmiarowość może być określana jako różnica entropii skalarnej i estymowanego źródła informacji:

$$R_d(\tau, N) = H_1(\tau, N) - h(\tau, N), \quad (26)$$

przy czym dla małych opóźnień otrzymujemy

$$H_{d-1}(0, N) = H_d(\tau, N) \Rightarrow R_d(0, N) = H_1(N) \quad (27)$$

zaś dla dużych opóźnień nadmiarowość dąży do zera

$$\lim_{\tau \rightarrow \infty} H_{d-1}(\tau, N) = dH_1(\tau, N) \Rightarrow R_d(\infty, N) = 0. \quad (28)$$

Wartość parametru τ , dla którego nadmiarowość znika określa granicę przewidywalności zachowania się systemu, tzn. horyzont przewidywalności systemu. Horyzont ten jest bardzo krótki, jeśli d jest mniejsze niż minimalny wymiar zanurzenia. Jednak, gdy wartość parametru d jest większa niż wymiar zanurzenia, wówczas nadmiarowość będzie się zmniejszać znacznie wolniej, a nachylenie dla małych τ będzie źródłem entropii:

$$R_d(\tau, N) = H_1(N) - \tau h(1). \quad (29)$$

Zgodnie z Bonnländerem i Weigendem (1994) równania (18) i (21) pozwalają na oszacowanie rozwiązania systemu oraz następujących parametrów:

- minimalnego wymiaru zanurzenia,
- źródła entropii,
- horyzontu prognozowania,
- liczby stopni swobody,
- wskaźnika zapominania warunków początkowych przez system.

2.5 Modelowanie sieciami neuronowymi

Obserwacja systemu dynamicznego opisanego układem deterministycznych równań różniczkowych pozwala, na podstawie obserwowanych wielkości, tak dobrać wielkość opóźnienia związanego z zanurzeniem, aby odtworzyć liczbę stopni swobody systemu. Wartości odtworzonych zmiennych pozwalają uzyskać dokładne prognozy. Przedstawimy teraz procedurę uzyskiwania dokładnych prognoz w szeregach czasowych generowanych przez deterministyczny system.

Prawdopodobnie najprostszym rozwiązaniem zagadnienia prognozowania jest utworzenie bazy danych z historycznych ciągów $\{x_{t-1}, \dots, x_{t-d}\}$ o długości d - ciągi te tworzą *wzorce wejściowe* i odpowiadające tym ciągom stany następne $\{x_t\}$, czyli wyjścia. Analiza nowego ciągu o długości d polega na przeszukaniu bazy danych i znalezieniu takiego wzorca, który jest "najbliższy" rozpatrywanemu, po czym przyjmuje się jako prognozę odpowiedni stan następny występujący po tym wzorcu. Znalezienie *najbliższego wzorca* wymaga przechowywania wszystkich zarejestrowanych wzorców w pamięci. W przypadku sztucznych sieci neuronowych *wiedza* o systemie przechowywana jest w wagach połączeń między neuronami, a poszczególne wzorce mogą być odtworzone dopiero po procesie *uczenia* sieci.

W przypadku występowania danych z szumem, zadanie poprawiania jakości prognozy można uzyskać uśredniając po sąsiednich punktach. Występuje tutaj jednak tzw. *dylemat odchylenia wariancji* (Geman, Bienenstock, Doursat, 1992) polegający na uwzględnianiu małej

liczby sąsiadujących punktów do uśredniania – w wyniku prognoza jest ciągle obarczona szumem (duża wariancja, małe odchylenie standardowe). Natomiast uwzględnienie zbyt dużej liczby punktów sąsiadujących do uśredniania pozwala uzyskać prognozę zbyt stabilną (wariancja może dążyć do zera, a odchylenie standardowe staje się duże).

W liniowym modelu Farmera i Sidorowicha (1987) do predykcji szeregów czasowych występują dwa zasadnicze parametry, mianowicie parametr określający liczbę sąsiadów k oraz liczbę opóźnień d . Prognozowanie w oparciu o modele liniowe często daje poprawne rezultaty - wyjaśnieniem tego faktu może być lokalna liniowość każdej podprzestrzeni. W takich modelach, zwanych także lokalnymi, wszystkie obserwowane wielkości szeregów czasowych muszą być przechowywane w pamięci (komputera), gdyż wielkości te służą do prognozowania, a dodatkowo do tworzenia nowych modeli w każdym następnym kroku.

Rozważymy teraz modele, w których tylko raz jest konstruowana (modelowana) powierzchnia zanurzona w przestrzeni opóźnień. W pracach Yule'a (1927) zaproponowano, aby ta powierzchnia była zwykłą hiperpłaszczyzną. W końcu lat siedemdziesiątych zostały zaproponowane pierwsze prace osłabiające takie założenie dla modeli typu AR, mianowicie Granger i Anderson (1980) wprowadzili iteracje drugiego rzędu między mierzonymi wejściami x_i, x_j (model bilincarny), zaś Tong i Lim (1980) rozdzielili przestrzeń wejść na dwie podprzestrzenie i rozważali dwa modele typu AR – tzw. modele autoregresyjne (Threshold Autoregressive models). W ostatnich latach popularne są także inne modele np.: MARS (Multivariate Adaptive Regression Splines) wprowadzone przez Friedmana (1991). Zupełnie inne możliwości dają sztuczne sieci neuronowe.

Sieci neuronowe są często używane do klasyfikacji, tzn. dany jest zbiór wzorców wprowadzanych jako wejścia do sieci, a zadanie polega na przyporządkowaniu każdego wejścia do odpowiedniej klasy. Inne bardzo ważne zastosowanie sieci neuronowych to regresja nieliniowa, w tym zastosowaniu zadanie polega na znalezieniu gładkiej interpolacji pomiędzy punktami. W zastosowaniu sieci neuronowych do szeregów czasowych informacja o *obrazach* (wzorcach) i historii tych obrazów jest podawana cyklicznie. Początki wykorzystania sieci neuronowych do prognoz w szeregach czasowych sięgają początkowych lat rozwoju sztucznych sieci neuronowych. W 1964 roku Hu (1964) wykorzystał rezultaty prac Widrowa, a mianowicie jego liniową adaptacyjną sieć do prognozy pogody. Następną erą w sztucznych sieciach neuronowych jest to okres rozwoju gradientowego algorytmu uczenia sieci, zwanego algorytmem wstecznej propagacji błędów uczenia (backpropagation). Lapades i Farber (1987) zastosowali taką sieć i tego typu algorytm do uczenia sieci opisującej zależność pomiędzy wyjściem sieci (następny punkt w szeregu czasowym) a wejściem sieci (czyli sekwencją poprzednich punktów szeregu czasowego). Do eksperymentu użyli oni szereg czasowy generowany przez generator komputerowy. Następnie Weigend, Huberman i Rumelhart (1990) określili pewne warunki, które powinna spełniać sieć do prognoz rzeczywistych szeregów czasowych.

Podobnie jak w modelach stochastycznych (ARMA), także sieć neuronowa musi "znać" historię całego szeregu czasowego. Najprostszym sposobem wprowadzenia historii szeregu czasowego do sieci neuronowej jest wprowadzanie ciągów danych (o odpowiednie długości - opóźnienie) na wejście sieci neuronowej. Warto wspomnieć, że sieć neuronowa bez ukrytej warstwy neuronów jest równoważna modelowi AR. Znane są sieci, w których wprowadzono opóźnienie (Lang, Waibel, Hinton, 1990) oraz inne modyfikacje sieci, których celem jest wprowadzenie złożonych wag (połączeń między neuronami) w celu zawarcia jak najwięcej informacji w połączeniach między neuronami.

Algorytm uczenia wielowarstwowych sieci neuronowych - algorytm wstecznej propagacji jest algorytmem gradientowym i wyznacza jedynie lokalne minima błędów uczenia. W zależności od zainicjowanych wartości wag połączeń sieci otrzymuje się różne rozwiązania. W pracach Krawczaka (1994, 1998, 1999a, 2000a, 2002a, 2002b, 2003a) przedstawiono no-

wy algorytm uczenia tej klasy sieci neuronowych. W największym skrócie idea nowego algorytmu polega na potraktowaniu procesu uczenia sieci jako szczególny przypadek zagadnienia sterowania optymalnego. W tych pracach wykorzystano ideę programowania dynamicznego i aproksymacji funkcji *zwrotnej* (poprzez zastosowanie metody kontynuacji). W następnych pracach Krawczaka (2001a, 2002a, 2002b, 2003a) opracowano nowy algorytm uczenia sieci neuronowych nazwany *heurystycznym algorytmem programowania dynamicznego*. Algorytm ten daje około 70 % pewności, że otrzymane rozwiązanie jest rozwiązaniem globalnym uczenia sieci neuronowych.

2.6 Obserwowalność stanów systemu dynamicznego

W ciągu ostatnich kilku lat obserwujemy zainteresowanie teorią chaosu deterministycznego do analizy i predykcji finansowych szeregów czasowych. Jednocześnie analizowane są możliwości obliczeniowe sztucznych sieci neuronowych.

Jak już przedstawiano wcześniej, stan systemu dynamicznego $y(k)$ można przedstawić jako funkcję stanów poprzednich $y(k-1), y(k-2), \dots, y(k-N)$ określonych w chwilach w chwilach $k-1, k-2, \dots, k-N$, zależność ta może być przedstawiona w następującej formie:

$$y(k) = NN(y(k-1), y(k-2), \dots, y(k-N)) \quad (30)$$

gdzie NN przedstawia funkcję nieliniową, która może być zrealizowana np. przez wielowarstwową sztuczną sieć neuronową.

Zadanie wyznaczenia liczby N określającej jak wiele historycznych danych jest wymaganych do wyznaczenia wartości stanu systemu $y(k)$.

Rozważmy zagadnienie *obserwowalności stanów* systemu dynamicznego. System liniowy można opisać następującymi równaniami:

$$\dot{x}(t) = Ax(t) + bu(t) \quad (31)$$

$$y(t) = cx(t) + d v(t) \quad (32)$$

gdzie x określa stan a y jest obserwowanym wyjściem systemu, zaś u i v oznaczają sterowania. Macierz A oraz wektory b , c i d oznaczają parametry systemu. W rozważanym przypadku należy założyć, że wejście systemu nie jest znane, więc niech $b = 0$. Dyskretna wersja systemu (31)-(32) może być przedstawiona następująco:

$$x(k+1) = \Theta x(k) \quad (33)$$

$$y(k) = c x(k) \quad (34)$$

gdzie $k = n \Delta t$ i $\Theta = \exp(A \Delta t)$. Zadanie obserwowalności stanów systemu polega na odpowiedzeniu na pytanie: czy znajomość obserwowanego wyjścia systemu $y(k)$ pozwala określić stan systemu $x(k)$. Dla systemów liniowych można określić *macierz obserwowalności* $\Psi(h, \Theta)$, która określa obserwowany wektor wyjść

$$y(k) = \Psi(h, \Theta) x(k) \quad (35)$$

i która pozwala określić wyjścia na podstawie wartości poprzednio obserwowanych wyjść systemu

$$y(k) = h \Theta^N \Psi^{-1} [y(k-N), \dots, y(k-1)]^T \quad (36)$$

przy założeniu, że macierz ta jest odwracalna, a N jest wymiarem systemu.

Zadanie obserwowalności może być także określone dla systemów nieliniowych, rozważmy następującą dyskretną postać systemu dynamicznego:

$$x(k+1) = F(x(k)) \quad (37)$$

$$y(k) = c(x(k)) \quad (38)$$

gdzie F jest nieliniową funkcją opisującą dynamikę systemu, natomiast c określa funkcję wyjściową. Zgodnie z twierdzeniem Takensa (1981) określamy wektor stanów historycznych

$$[y(k-1), y(k-2), \dots, y(k-T)]^T \quad (39)$$

który pozwala określić wektor stanu systemu w chwili k

$$x(k) = \Psi [y(k-1), y(k-2), \dots, y(k-T)]^T \quad (40)$$

i w rezultacie

$$y(k) = NN(F, c, [y(k-1), y(k-2), \dots, y(k-T)]^T) \quad (41)$$

gdzie NN jest modelem szeregu czasowego w postaci sieci neuronowej. Równania (36) i (41) mają podobną strukturę, przy czym równanie (41) nie można wyrazić z postaci analitycznej. Liczba T określa wymiarowość problemu zanurzonego w problemie oryginalnym. Zgodnie z twierdzeniem Takensa (1981) (także Peters, 1994) liczba T spełnia następujący warunek:

$$T > 2D + 1 \quad (42)$$

gdzie D jest wymiarem fraktalnym szeregu czasowego (Schuster, 1995).

Twierdzenie Takensa nie podaje jaką postać ma mieć zależność (41), zatem można wykorzystać wielowarstwowe sieci neuronowe posiadające własności aproksymowania funkcji nieliniowych (Hornik, Stinchcombe, White, 1989).

Rozważmy wektor wejściowy $\bar{x}(k) = [x(k-1), x(k-2), \dots, x(k-T)]^T$ sieci neuronowej NN dający następujące wyjście

$$\hat{x}(k) = NN(w, \bar{x}(k)) \quad (43)$$

gdzie przez w określone są wagi połączeń między neuronami. Proces uczenia sieci neuronowej polega na wyznaczeniu wartości tych wag w taki sposób, aby następujący wskaźnik uczenia sieci osiągał wartość minimalną:

$$\min_w \frac{1}{2} \sum_{p=1}^n (x(k) - NN(w, \hat{x}(k)))^2 \quad (44)$$

Jak pokazaliśmy wielowarstwowe sieci neuronowe można stosować do rozwiązywania zagadnień związanych z predykcją stanów systemów nieliniowych. Jak już pokazaliśmy w sekcji 2.2, wiele nieliniowych równań różniczkowych (lub różnicowych) może generować chaotyczne szeregi czasowe (np. Schuster, 1995).

Możemy zatem założyć, że złożone przebiegi szeregów czasowych (zwłaszcza finansowych szeregów czasowych) mogą być generowane przez dynamiczne systemy nieliniowe, których analityczna forma nie jest nam znana. Przyjmując takie założenie możemy wykorzystać elementy teorii chaosu do analizy systemów dynamicznych na podstawie obserwowanych szeregów czasowych (Krawczak, 2000, 2001, 2002a, 2002b, 2003a). Dzięki pracom Ruella (1981) i Takensa (1981) istnieje metoda do analizy szeregów czasowych zwana *metodą zanurzenia szeregów czasowych*. Metodę tę można zilustrować rozważając wykresy par $x(k)$ i $x(k+1)$ rozważanego szeregu czasowego. W przypadku rozważanego wcześniej równania logistycznego (równanie 13) wykres takich par przedstawiony jest na Rysunku 4. Zaś w przypadku równania Mackeya-Glassa (14) wykres takich par jest przedstawiony na Rysunku 6.

Takie zanurzone wykresy nazywane są *dziwnymi atraktorami*. Jeszcze raz warto zauważyć, że chociaż wykresy odpowiednich szeregów czasowych są bardzo złożone, to wykresy odpowiednich dziwnych atraktorów nie są skomplikowane. Główna idea polega na tym, że znając punkt $x(k)$ atraktora możemy w prosty sposób znaleźć dobrą estymację następnego punktu $x(k+1)$ leżącego na atraktorze, chociażby przez zwykłą interpolację. Podobną zasadę można także stosować dla wielowymiarowych atraktorów.

Główną cechą systemów dynamicznych określanych jako systemy chaotyczne jest duża wrażliwość rozwiązań ze względu na warunki początkowe. Wyznaczenie *wykładników Lapunowa* pozwala ocenić stan systemu dynamicznego, to znaczy wykładniki Lapunowa określają czy następny punkt dziwnego atraktora jest punktem zbieżnym czy rozbieżnym (Wolff, Swift, Swinney, 1985).

W tym celu należy wyznaczyć tzw. *główny* lub *dominujący średni* wykładnik Lapunowa zdefiniowany jak następuje:

$$L = \log_2 \frac{\sum_{n=1}^{l_{n+1}} l_n}{n-1} \quad (45)$$

gdzie n oznacza numer próbki, a l jest odległością euklidesową między dwoma sąsiadującymi trajektoriami atraktora. Jeżeli wykładniki Lapunowa są dodatnie to rozważany system ma charakter systemu chaotycznego, w przypadku gdy wykładniki Lapunowa są ujemne - to system jest stabilny, zaś w przypadku gdy wykładniki te są równe zero - to zachowanie systemu charakteryzuje cykliczność.

Następny parametr określa miarę przewidywalności stanów systemu dynamicznego (lub też przewidywalności następnych wartości w szeregach czasowych). Parametr ten nazywany jest *wykładnikiem Hursta* (Peitgen, Jurgens, Saupe, 1996; 1970). Wykładni Hursta można wyznaczyć w następujący sposób: dla szeregu czasowego $x(1), x(2), \dots, x(k-1), x(k)$ reprezentowanego przez k punktów, wybierana jest liczba p spełniająca następujący warunek $10 \leq p < k/2$; następnie rozważany szereg czasowy jest podzielony na k/p części. W następnym kroku wyznaczany jest maksymalny zakres wartości każdej części, jak również wyznaczane jest odchylenie standardowe całego szeregu czasowego w celu wyznaczenia następujących wartości dla każdej części

$$rs_i = \frac{(\text{largest_value} - \text{smallest_value})}{\text{standard_deviation}}. \quad (46)$$

Następnie wyznaczany jest średni współczynnik rs dla wszystkich części

$$rs = \frac{1}{p} \sum_{i=1}^p rs_i. \quad (47)$$

Następująca równanie określa zależność średniej wartości współczynnika rs i wykładnika Hursta H :

$$rs = (p/2)^H. \quad (48)$$

Wykładnik Hursta przyjmuje wartości między 0.0 i 1.0. Rozróżniamy dwa główne zakresy wartości wykładnika Hursta: jeżeli $0.5 < H < 1$ to zachowanie systemu określamy jako *zachowanie stałe* – w takim przypadku stany systemu dynamicznego są przewidywalne; jeżeli $0 < H < 0.5$ to rozważany system ma charakter probabilistyczny. W przypadku, gdy $H = 0$ to każda wartość szeregu czasowego będącego wyjściem systemu dynamicznego zmienia kierunek, natomiast dla $H = 0.5$ wartości szeregu czasowego są zupełnie przypadkowe; jeżeli zaś $H = 1$ to stany systemu dynamicznego są całkowicie deterministyczne.

Znana jest zależność między wymiarem fraktalnym szeregu czasowego i wykładnikiem Hursta, zależność tę przedstawia następujące proste równanie:

$$D = 2 - H. \quad (49)$$

Dzięki tej zależności można wyznaczyć wymiar wejść sieci neuronowej, czy też wymiar podprzestrzeni zanurzenia, określony przez (42)

$$T > 2D + 1.$$

W celu wyznaczenia przedstawionych powyżej parametrów szeregu czasowego należy zkonstruować model dziwnego atraktora, np. w postaci wielowarstwowej sieci neuronowej. W przypadku znajomości wymiaru fraktalnego atraktora możemy wyznaczyć powyższe parametry w oparciu o twierdzenie Takensa określające górne ograniczenie wymiarowości przestrzeni zanurzonej.

Wydaje się, że może być bardzo interesująca analiza szeregów czasowych ze względu na wartości, jak i znaki głównego wykładnika Lapunowa, bowiem znak tego parametru określa charakter zachowania się systemu dynamicznego.

W tej sekcji przedstawiono wyniki dotyczące analizy i predykcji szeregów czasowych, szczególną uwagę zwrócono na wykorzystanie elementów teorii informacji, zastosowanie wielowarstwowych sieci neuronowych, wykorzystanie elementów teorii chaosu deterministycznego do rekonstrukcji lokalnych wartości rozwiązania (jeśli rozważamy system dynamiczny jest wolny od szumu). Rozważono jedynie wielowarstwowe sieci jednokierunkowe, znane są jednak wyniki dotyczące rekurencyjnych sieci neuronowych pozwalających wyznaczenie stanu systemu dynamicznego w obecności szumu (Mozer, 1994). W pracy Krawczaka (2003a) przedstawiono nowy algorytm uczenia wielowarstwowych rekurencyjnych sieci neuronowych pozwalający wyznaczać „prawie” globalne wartości minimum błędów uczenia (*heu-*

rystyczny algorytm programowania dynamicznego). Wydaje się, że algorytm ten może znaleźć zastosowanie przy predykcji finansowych szeregów czasowych.

3 Indukcyjne uczenie na podstawie przykładów

3.1 Wprowadzenie

W tej części ograniczono się do zagadnień uczenia nadzorowanego, w którym tworzone są opisy klas zawierających przykłady posiadające pewne wspólne właściwości, które odróżniają je od innych klas. Zadanie uczenia maszynowego na podstawie przykładów rozumiane jest w pracy w sposób przedstawiony poniżej.

Przyjmijmy, że dysponujemy dużym zbiorem danych, np. dotyczących klientów banków i udzielonych im kredytów. Dane w nich zawarte dotyczą parametrów klientów, którym udzielono kredytów oraz historii spłacania przez nich kredytów. Każdego klienta można opisać przy pomocy wybranego zbioru cech, takich jak np. „aktywa”, „debety na koncie”, „roczny przyrost dochodów”, „branża kredytobiorcy”, „wysokość kredytu”, „okres spłaty” itp. Warunek związany z daną cechą można zapisać w postaci („cecha” = „wartość cechy”), np. („zdolność kredytowa” = „dobra”). Przykładem nazywa się koniunkcję warunków związanych ze wszystkimi cechami określonymi dla danego klienta.

Cechy opisujące przykłady mogą być różnego typu. Najczęściej stosuje się cechy typu:

- *liczbowego*, których dziedziny stanowią liczby, np. cecha „wielkość kredytu” może przyjmować wartości: 10, 20, 50 mln. zł.
- *nominalnego nieuporządkowanego*, których dziedziny są zbiorami nieuporządkowanymi, tzn. możemy jedynie określić dla wartości cech relację równości lub różności; np. cecha „branża kredytobiorcy” może przyjmować nieuporządkowane wartości: „metalowa”, „tekstylna”, „informatyczna”,
- *nominalnego uporządkowanego*, których dziedziny są zbiorami uporządkowanymi, tzn. możemy określić dla wartości cech relację równości, różności oraz mniejszości lub większości; np. cecha „roczny przyrost dochodów” może przyjmować uporządkowane wartości: „mały”, „średni”, „duży”.

Zbiór takich przykładów, ze względu na wartości wybranej cechy (zwanej cechą decyzyjną) można podzielić na rozłączne *klasy*. Zakłada się, że liczba i rodzaj wszystkich cech wybranych do opisu przykładów wystarczają do dokonania takiego podziału. Przez klasę rozumie się zbiór przykładów mających tę samą *wartość cechy decyzyjnej*. Np. cecha decyzyjna „zdolność kredytowa” z wartościami „dobra”, „zła” określa dwie klasy. Pierwszą klasę stanowią kredytobiorcy, których zdolność kredytowa jest dobra, drugą tacy kredytobiorcy, których ta zdolność jest zła.

Na podstawie zbioru tak opisanych przykładów (zwanego *zbiorem uczącym*) metody uczenia maszynowego tworzą opisy rozpatrywanych klas. Opis klasy wyrażany jest na ogół w postaci *reguł* mających postać wyrażań logicznych typu:

*”JEŻELI spełnione są określone warunki związane z cechami
TO zachodzi przynależność do danej klasy”;*

drzew decyzyjnych;

lub też w postaci odpowiednio dobranych wag połączeń w *sieciach neuronowych* i ich struktury.

Niektóre z metod uczenia maszynowego przystosowane są do analizy danych liczbowych, przykładowo metody oparte na sieciach neuronowych, metody statystyczne takie jak analiza regresji i metody grupowania, czy też metody oparte na analizie wzorców. Przed zastosowaniem tych metod w przypadku, gdy dysponujemy danymi nominalnymi, należy dokonać skalowania nominalnych danych. Może to prowadzić do niepożądanych wyników, ponieważ metody te interpretują odległości i zależności pomiędzy wartościami, które nie mają żadnego znaczenia przy numerycznej reprezentacji symboli. Metody tworzące drzewa decyzyjne (Quinlan 1993), sieci Bayesowskie (Kruse i inni 1991), (Pearl 1988) lub też podejścia bazujące na logice pracują dobrze zarówno z symbolicznymi danymi jak też z mającymi dyskretny, skończony zbiór wartości. W przypadku danych ciągłych, wymagane jest więc określenie dla nich skończonej liczby przedziałów zmienności.

Reprezentacja opisu klasy w postaci reguł uznawana jest za bardziej czytelną dla człowieka niż inne reprezentacje. Część *warunkowa* reguły będzie określana w postaci koniunkcji warunków określonych dla wybranego podzbioru cech opisujących przykłady, a część *decyzyjna* będzie określała przynależność przykładowi spełniającego te warunki do danej klasy. Przykład takiej reguły przedstawiono poniżej:

JEŻELI („aktywa” = „większe od 1 000 000\$”) \wedge („debyta na koncie” = „brak”) *TO* („zdolność kredytowa” = „duża”)

Tworzone reguły powinny spełniać pewne dodatkowe wymagania. Można np. wymagać, aby poprawnie opisywały wszystkie przykłady ze zbioru uczącego należące do danej klasy i nie opisywały żadnego przykładowi do tej klasy nie należącego.

Ze względu na fakt, że w rzeczywistych zbiorach danych pozyskanych w wyniku oszacowania, testów, wywiadów z ekspertami itp. mogą istnieć błędy, nie zawsze należy wymagać, aby reguły były dokładnie dopasowane do przykładów uczących, ale można przyjąć, że powinny poprawnie opisywać „większość” przykładów należących do rozpatrywanej klasy i nie opisywać „prawie wszystkich” przykładów do tej klasy nie należących. Powinny również mieć minimalną długość, np. w sensie liczby warunków tworzących regułę lub też łącznej ważności warunków występujących w regule (z punktu widzenia przyjętego kryterium).

Istnienie nieprecyzyjnych, nierozróżnialnych przykładów w rzeczywistych zbiorach danych jest istotnym utrudnieniem w procesie uczenia. W takim przypadku można poddać dane obróbce wstępnej, próbując poprawiać w nich błędy, modyfikując cechy stosowane do opisu przykładów, usuwając przykłady sprzeczne, itp. Można również rozpoznawać błędy już w trakcie procesu uczenia i następnie próbować je usuwać.

Reguły, o których była mowa wyżej, można tworzyć stosując różne metody uczenia maszynowego. Wiele znanych algorytmów uczenia maszynowego przystosowanych jest do analizy wyłącznie niesprzecznych przykładów. Istniejące tradycyjne procedury tworzenia reguł decyzyjnych z przykładów klasyfikacji stosują następujące strategie działania:

- tworzą *minimalny zbiór reguł*, opisujący wszystkie przykłady uczące, tzn. spełniający warunek spójności,
- tworzą *wyczerpujący zbiór reguł*, złożony ze wszystkich możliwych reguł,
- tworzą zbiór *„najsilniejszych” reguł*, które nie muszą opisywać wszystkich przykładów uczących.

W innych systemach uczących się stosowane są specjalne techniki postępowania, np. przedwczesne zatrzymanie procesu uczenia zanim zostanie znaleziony kompletny opis klasy (to znaczy poprawnie opisujący wszystkie przykłady uczące), redukcji drzew, odrzucania z

wygenerowanych reguł składników odnoszących się do zbyt małej liczby przykładów, preferowanie w procesie uczenia pewnych składników (np. cech, przykładów), analizę wyjątków, upraszczanie utworzonego opisu itd.

Powstały też zupełnie nowe podejścia, które szczególnie sprawdzają się w przypadku niepełnych i niepewnych danych. Są one oparte m.in. na logikach nie monotonicznych, rachunku prawdopodobieństwa, teorii zbiorów rozmytych wprowadzonych przez Zadeha, czy też teorii zbiorów przybliżonych Pawlaka.

W teorii zbiorów rozmytych (Zadeh 1981) dowolny element może należeć do zbioru w pewnym stopniu. Trzeba pamiętać, że rozmytość nie oznacza przypadkowości. Przypadkowość można utożsamiać z niepewnością, czy element należy czy też nie należy do zbioru, a więc z brakiem wiedzy. Natomiast rozmytość dotyczy samego stopnia przynależności elementu do zbioru, polega na braku jednoznacznie zdefiniowanych granic zbiorów rozmytych, które mogą się częściowo pokrywać. Rozmytość polega na tym, że ten sam element może należeć równocześnie do kilku zbiorów rozmytych, z różnym stopniem przynależności. Zbiory rozmyte są stosunkowo wszechstronną techniką inteligentnego przetwarzania informacji. Ich podstawową funkcją jest przetwarzanie informacji podanej w formie jakościowej, w tym pojęć leksykalnych.

W teorii zbiorów przybliżonych (Pawlak 1991) przykłady nierozróżnialne za pomocą zbioru wybranych cech tworzą tzw. zbiory elementarne. Ponieważ nie zawsze jest możliwe dokładne zdefiniowanie danego zbioru przykładów należących do wybranej klasy decyzyjnej za pomocą elementarnych zbiorów nierozróżnialnych, wprowadzono pojęcie zbioru przybliżonego. Zbiór przybliżony jest reprezentowany w postaci pary pojęć, zwanych dolnym i górnym przybliżeniem tego zbioru. Dolne przybliżenie jest sumą wszystkich zbiorów elementarnych zawartych w danym zbiorze przykładów, a górne przybliżenie jest sumą tych zbiorów elementarnych, które mają elementy wspólne z danym zbiorem przykładów. Różnica pomiędzy górnym i dolnym przybliżeniem tworzy tzw. brzeg danego zbioru i zawiera przykłady, które nie mogą być jednoznacznie zaklasyfikowane. Zastosowanie podejścia opartego na teorii zbiorów przybliżonych pozwala na uzyskanie w procesie uczenia dwóch typów reguł elementarnych: dokładnych i przybliżonych. Reguły dokładne wiążą z daną koniunkcją warunków związanych z cechami dokładnie jedną klasę, natomiast reguły przybliżone wskazują na wiele możliwych klas. Podejście to wzbudziło duże zainteresowanie na świecie, znalazło wiele interesujących zastosowań.

Możemy wyróżnić kilka rodzajów zastosowań zbiorów reguł decyzyjnych, uzyskanych na podstawie danego zbioru przykładów. Mogą być one użyte do klasyfikacji nowych przykładów, nie należących do zbioru uczącego, dla których nie jest znana przynależność do klasy. Prognoza taka jest realizowana w wyniku sprawdzania spełnienia warunków w częściach warunkowych wszystkich reguł, a w przypadku sytuacji nie jednoznacznych, (gdzie jest spełnionych kilka reguł lub też nie jest spełniona żadna z nich) uwzględniany jest np. stopień dopasowania do klasy. W tym przypadku reguły staną się częścią systemu klasyfikującego. Preferowany jest więc kompletny zbiór reguł opisujących wszystkie przykłady uczące, a kryterium oceny takich reguł staje się jak najwyższa dokładność klasyfikacji nowych przykładów.

Reguły mogą być również użyte do uzyskania nie znanych wcześniej w sposób jawny informacji o istniejących w zbiorze danych zależnościach. Mogą pomóc zrozumieć i wyjaśnić istniejące zależności pomiędzy cechami czy też definicjami klas decyzyjnych. W tym przypadku każda reguła jest rozpatrywana indywidualnie jako możliwa reprezentacja interesującego wzorca, czy też jakiejś istniejącej zależności pomiędzy cechami. Kryteria ocen takich reguł mogą być wielorakie i rozważanie ich razem nie jest prostym zadaniem.

W rozdz. 3.2. zadanie uczenia maszynowego na podstawie przykładów sformułowano w postaci ciągu zadań cząstkowych, zapisanych w postaci pewnych modyfikacji zadania pokrycia (Szkatuła 1995). Zaproponowano metodę tworzącą reguły typu *JEŻELI ... TO ...*, prezentujące istniejące regularności w przykładach, z możliwością uwzględniania założenia o poprawnym opisywaniu przez tworzone reguły „większości” przykładów uczących. Przeprowadzone eksperymenty obliczeniowe pokazały, że przedstawiona metoda z powodzeniem może być zastosowana w zagadnieniach praktycznych.

3.2 Opis zastosowanego podejścia

W tej części założono, że cechy stosowane do opisu przykładów nie są tak samo ważne i można wprowadzić współczynniki ważności dla każdej cechy i dla każdej wartości, którą może dana cecha przyjmować. Większe/mniejsze wartości współczynników będą wtedy odpowiednio korzystne/niekorzystne z punktu widzenia rozpatrywanego kryterium. Współczynniki te będą określały pewną preferencję cech, która będzie uwzględniana przy tworzeniu reguł. Reguły, zawierające więcej warunków związanych z cechami preferowanymi, powinny lepiej opisywać zbiór przykładów, z punktu widzenia rozpatrywanego kryterium.

Można uwzględnić różne kryteria, przykładowo większą wiarygodność czy też istotność danych danego przypadku i przy współpracy z ekspertami zastosować dla określenia współczynników ważności np. podejście AHP (*Analytical Hierarchy Process*) (Saaty 1980). Przykładowo, bardziej wiarygodna będzie reguła zawierająca warunki związane z cechami mniej podatnymi na błędne określenie ich wartości i wygodniej będzie stosować w praktyce regułę, w której występuje łatwa do określenia wartość cechy, niż niepewny wynik finansowy w przyszłości (Szkatuła 1995).

Współczynniki można również tworzyć w oparciu o analizę całego zbioru przykładów uczących, w celu preferowania w procesie uczenia tych wartości cech, które są bardziej „typowe” dla rozpatrywanej klasy. Utworzone reguły zawierające preferowane wartości cech powinny zawierać mniejszą liczbę warunków i lepiej opisywać zbiory danych (Szkatuła 1995), (Kacprzyk, Szkatuła 1998, 1999).

Wprowadzono pojęcie długości ważonej reguły (uwzględniającej współczynniki ważności), której minimalizacja jest zgodna z naszymi oczekiwaniami znalezienia reguły krótkiej, zawierającej małą liczbę mniej preferowanych cech. Zadanie tworzenia reguł dla danej klasy, przedstawiono w postaci ciągu zadań cząstkowych, sformułowanych w postaci pewnych modyfikacji zadania pokrycia. Zaproponowano metodę heurystyczną tworzącą reguły o minimalnej długości ważonej typu *JEŻELI ...TO ...*, prezentujące istniejące regularności w przykładach, z możliwością uwzględniania założenia o poprawnym opisywaniu przez nie „większości” przykładów uczących.

Sformułowanie problemu

Niech U będzie niepustym, skończonym zbiorem przykładów, $A = \{a_1, \dots, a_k\}$ niepustym, skończonym zbiorem cech. $V_{a_j} = \{v_{j_1}, \dots, v_{j_r}\}$, $V_{a_j} \neq \emptyset$, jest zbiorem wartości cechy a_j , dla $j = 1, \dots, K$ oraz $V = \bigcup_{j=1, \dots, K} V_{a_j}$. Funkcja $f: U \times A \rightarrow V$ taka, że $\forall e \in U, \forall a_j \in A$ zachodzi $f(e, a_j) \in V_{a_j}$, określa jaką wartość w przykładzie e przyjmuje cecha a_j . Niech $g(v_{r(j)}) \in (0,1)$ oznacza współczynnik ważności dla cechy a_j , $j = 1, \dots, K$ i wartości $v_{r(j)} \in V_{a_j}$, którą dana cecha przyjmuje. Indeks $r(j)$ określa, jaką wartość przyjmuje j -ta cecha, $r(j) \in \{j_1, \dots, j_r\}$.

Warunkiem elementarnym dla cechy a_j , $j = 1, \dots, K$, i przykładu $e \in U$ nazywamy wyrażenie postaci $(a_j = v_{r(j)}; g(v_{r(j)}))$, gdzie $v_{r(j)} \in V_{a_j}$, $v_{r(j)} = f(e, a_j)$, $g(v_{r(j)}) \in (0,1)$. Każdy przy-

kład $e \in U$ można opisać w postaci koniunkcji K warunków elementarnych w sposób następujący

$$e: \bigwedge_{j=1}^K (a_j = v_{r(j)}; g(v_{r(j)})) \quad (50)$$

dla $a_j \in A$, $v_{r(j)} \in V_{a_j}$, $g(v_{r(j)}) \in (0,1)$.

Uwaga:

W dalszej części pracy, w podawanych przykładach ilustrujących omawiane pojęcia, dla uproszczenia zapisu będą pomijane współczynniki ważności cech.

Przykład 1.

Zgodnie z (50), klienta banku możemy opisać np. w sposób przedstawiony poniżej, z zastosowaniem cech: $a_1 =$ „aktywa”, $a_2 =$ „debyty na koncie”, $a_3 =$ „roczny przyrost dochodów”, ..., $a_{20} =$ „status konta”, $a_{21} =$ „stan konta”, $a_{22} =$ „stan zdrowia”, $a_{23} =$ „ryzyko wniosku kredytowego”.

$$e_1: (a_1 = \text{„brak”}) \wedge (a_2 = \text{„częste”}) \wedge (a_3 = 1\%) \wedge \dots \wedge (a_{20} = \text{„kredyt sflacony”}) \wedge (a_{21} = \text{„10 000 zł.”}) \wedge (a_{22} = \text{„zły”}) \wedge (a_{23} = \text{„duże”}). \quad \square$$

Koniunkcję l warunków elementarnych, $l \leq K$, dla wszystkich cech należących do pewnego podzbioru cech $P \subseteq A$, $P = \{a_{j_1}, \dots, a_{j_l}\}$ można zapisać w sposób następujący

$$C^l = \bigwedge_{j \in I} (a_j = v_{r(j)}; g(v_{r(j)})) \quad (51)$$

gdzie $I = \{j_1, \dots, j_l\} \subseteq \{1, \dots, K\}$, $\text{card}(I) = l$, $v_{r(j)} \in V_{a_j}$, $g(v_{r(j)}) \in (0,1)$.

Mówimy, że przykład $e \in U$ jest opisany przez koniunkcję C^l jeżeli $\forall j \in I$ zachodzi warunek $f(C^l, a_j) = f(e, a_j)$.

Przykład 2.

Wróćmy do przykładu ilustracyjnego 1. Powiemy, że koniunkcja $(a_2 = \text{„częste”})$ opisuje przykład e_1 , zaś koniunkcja $(a_2 = \text{„częste”}) \wedge (a_{22} = \text{„dobry”})$ nie opisuje przykładu e_1 .

□

Niech przykład e będzie opisany w postaci koniunkcji K warunków elementarnych $(a_j = v_{r(j)}; g(v_{r(j)})) \wedge \dots \wedge (a_K = v_{r(K)}; g(v_{r(K)}))$, wzór (50), z której wybrano l - warunków, $l \leq K$, tworząc koniunkcję postaci $C^l = (a_{j_1} = v_{r(j_1)}; g(v_{r(j_1)})) \wedge \dots \wedge (a_{j_l} = v_{r(j_l)}; g(v_{r(j_l)}))$, wzór (51), odpowiadającą zbiorowi indeksów $I = \{j_1, \dots, j_l\} \subseteq \{1, \dots, K\}$.

Mówimy, że koniunkcja C^l jest równoważna wektorowi zero-jedynkowemu $x = [x_j]^l$, $j = 1, \dots, l$, takiemu, że $x_j = 1$ jeżeli warunek $(a_j = v_{r(j)}; g(v_{r(j)}))$ z przykładu e występuje w koniunkcji C^l , a $x_j = 0$ w przeciwnym przypadku.

Przykład 3.

Dany jest przykład e : $(a_1 = \text{„obroty wysokie”}) \wedge (a_2 = \text{„długi okres spłaty”}) \wedge (a_3 = \text{„kredyt wysoki”})$. Wektor $[0, 1, 1]^T$ jest równoważny koniunkcji $(a_2 = \text{„długi okres spłaty”}) \wedge (a_3 = \text{„kredyt wysoki”})$; a koniunkcja $(a_1 = \text{„obroty wysokie”})$ jest równoważna wektorowi $[1, 0, 0]^T$. □

Jeżeli ze zbioru cech A wybierzemy jedną cechę a_d , to możemy ze względu na wartości, które ona przyjmuje dokonać podziału zbioru przykładów na rozłączne klasy. Elementy zbioru $A \setminus \{a_d\}$ nazywamy *cechami warunkowymi*, a cechę a_d nazywamy *cechą decyzyjną*.

Podziałem zbioru przykładów U ze względu na cechę decyzyjną $a_d \in A$ mającą dziedzinę $V_{a_d} = \{v_{d_1}, \dots, v_{d_n}\}$ nazywamy niepuste podzbiory przykładów $\{U_{v_{d_1}}, U_{v_{d_2}}, \dots, U_{v_{d_n}}\}$, gdzie $\forall v_{i(d)} \in V_{a_d}$

$$U_{v_{i(d)}} = \{e \in U : f(e, a_d) = v_{i(d)}\} \quad (52)$$

oraz $U_{v_{d_1}} \cup \dots \cup U_{v_{d_n}} = U$, $U_{v_{d_i}} \cap U_{v_{d_j}} = \emptyset$ dla $i \neq j$.

Tak więc cecha decyzyjna dzieli zbiór przykładów na niepuste, rozłączne i w sumie tworzące cały zbiór podzbiory, które nazywamy *klasami* i których opisy w postaci reguł będziemy tworzyć. Każda klasa zawiera przykłady nierozróżnialne między sobą za pomocą cechy decyzyjnej, to znaczy przykłady mające taką samą wartość cechy a_d .

Przykład 4.

Wróćmy do przykładu ilustracyjnego 1. Jeżeli za cechę decyzyjną przyjmiemy cechę $a_{23} = \text{„ryzyko wniosku kredytowego”}$ z wartościami, które może ona przyjmować: „małe”, „średnie”, „duże”, to określi ona trzy klasy: klasa $U_{\text{„małe”}}$ będzie zawierała klientów, dla których ryzyko jest małe, $U_{\text{„średnie”}}$ będzie zawierała klientów, dla których ryzyko jest średnie a klasa $U_{\text{„duże”}}$ będzie zawierała klientów, dla których ryzyko jest duże. □

Przykłady $e \in U$ należące do klasy $U_{v_{i(d)}}$ nazywamy *pozytywnymi* dla tej klasy i oznaczamy

$$POS_A(U_{v_{i(d)}}) = \{e \in U; f(e, a_d) = v_{i(d)}\} \quad (53)$$

a pozostałe przykłady nie należące do tej klasy, dla których nie istnieje w zbiorze przykładów pozytywnych przykład o takich samych wartościach wszystkich cech ze zbioru $A \setminus \{a_d\}$, nazywamy przykładami *negatywnymi* dla tej klasy i oznaczamy

$$NEG_A(U_{v_{i(d)}}) = \{e \in U : f(e, a_d) \neq v_{i(d)} \text{ i } \forall e \in POS_A(U_{v_{i(d)}}) \\ \exists a_j \in A - \{a_d\}, f(e, a_j) \neq f(e, a_j)\}. \quad (54)$$

Należy zauważyć, że zachodzą następujące zależności $POS_A(U_{v_{i(d)}}) \neq \emptyset$, $NEG_A(U_{v_{i(d)}}) \neq \emptyset$, $POS_A(U_{v_{i(d)}}) \cap NEG_A(U_{v_{i(d)}}) = \emptyset$.

Zbiory tak określonych przykładów uczących (z ich podziałem na klasy) są punktem wyjścia w procesie uczenia maszynowego, w wyniku którego, uzyskuje się opisy rozpatrywanych klas $U_{v_{i(d)}}$ w postaci reguł.

Regułą elementarną dla rozpatrywanej klasy nazywamy wyrażenie logiczne postaci *JEŻELI spełnione są określone warunki TO zachodzi przynależność do danej klasy*; w którym poprzednik reguły zawiera koniunkcję warunków związanych z podzbiorem cech wybranych do opisu przykładów. *Regułą elementarną* dla klasy $U_{v_{i(d)}}$ możemy zapisać w postaci

$$rul(C', v_{i(d)}): \text{JEŻELI } C' \text{ TO } (a_d = v_{i(d)}) \quad (55)$$

gdzie koniunkcję C' określa wzór (51).

Przykład 5.

Wróćmy do przykładu ilustracyjnego 4. Prosta reguła dla tego przypadku przedstawiono poniżej

$$\text{JEŻELI } (a_1 = \text{„niewielkie"}) \wedge (a_2 = \text{„częste"}) \text{ TO } (a_3 = \text{„duże"}). \quad \square$$

Reguła dla klasy $U_{v_{i(d)}}$, oznaczana $RUL(U_{v_{i(d)}})$, jest zbiorem reguł elementarnych typu (55) do tej klasy. Można ją również zapisać w postaci dysjunkcji (oznaczanej " \cup ") reguł elementarnych do tej klasy

$$\text{JEŻELI } C^l \cup \dots \cup C^L \text{ TO } (a_d = v_{i(d)}) \quad (56)$$

$$I_l \subseteq \{1, \dots, K\}, \quad l = 1, \dots, L, \quad C^l = \bigwedge_{j \in I_l} (a_j = v_{i(j,d)}; g(v_{i(j,d)})), \quad v_{i(j,d)} \in V_{a_j}.$$

Indeks $i(j, l)$ określa wartość, jaką przyjmuje cecha a_j w koniunkcji I_l .

Tworzone reguły powinny spełniać warunek *spójności*, tzn. rozróżniać przykłady należące, od nie należących, do danej klasy oraz być *minimalne*, tzn. usunięcie dowolnego warunku z części warunkowej reguły, spowodowałoby nie spełnienie warunku spójności. Zbiór reguł elementarnych będziemy nazywali *kompletnym*, jeżeli jest niepusty i skończony oraz dla każdego przykładu ze zbioru uczącego istnieje, co najmniej jedna reguła, która go opisuje. *Minimalnym zbiorem reguł* nazywamy taki zbiór reguł, który jest kompletny i zawiera minimalną liczbę reguł.

Utworzone reguły mogą być użyte do klasyfikowania nowych przykładów, dla których nie jest znana przynależność do określonej klasy. Przykładowo, możemy określać ryzyko pojedynczego wniosku kredytowego, na podstawie znajomości pewnych cech opisujących kredytobiorcę.

Można zauważyć, że dla koniunkcji C' utworzonej z $l \leq K$ warunków wybranych z przykładu $e \in U$, równoważny jej K -elementowy wektor x ma elementy $x_j = 1, \forall j \in I$, a dla $j \in \{1, 2, \dots, K\} \setminus I$ zachodzi warunek $x_j = 0$.

Długość ważoną koniunkcji $C' = \bigwedge_{j \in I} (a_j = v_{i(j)}; g(v_{i(j)}))$ można zdefiniować w sposób następujący

$$d(C') = \sum_{j \in I} (1 - g(v_{i(j)})) \cdot x_j + \sum_{j \in \{1, 2, \dots, K\} \setminus I} (1 - g(v_{i(j)})) \cdot x_j =$$

$$= \sum_{j=1}^K (1 - g(v_{t(j)})) \cdot x_j \quad (57)$$

Długość ważoną reguły, której część warunkowa złożona jest z L koniunkcji, $R = C^h \cup \dots \cup C^{L_h}$ zgodnie ze wzorem (56), można określić jako największą długość ważoną tworzących ją koniunkcji:

$$d_R(C^h \cup \dots \cup C^{L_h}) = \max_{l=1, \dots, L} d(C^{L_l}) \quad (58)$$

Zadanie uczenia maszynowego można sformułować następująco: na podstawie zbioru przykładów pozytywnych i negatywnych dla danej klasy należy utworzyć regułę o minimalnej długości ważonej

$$\min_{l_1, \dots, l_L} d_R(C^{l_1} \cup \dots \cup C^{l_L}) \quad (59)$$

która opisuje „prawie wszystkie” przykłady pozytywne oraz nie opisuje „prawie wszystkich” przykładów negatywnych dla rozpatrywanej klasy. Powyższą minimalizację należy rozumieć jako zadanie znalezienia cech, które występują we wszystkich koniunkcjach i określenie wartości, które te cechy przyjmują. Zadanie (59) można uprościć do ciągu zadań cząstkowych:

$$\min_h d(C^h), \dots, \min_{L_l} d(C^{L_l}) \quad (60)$$

gdzie minimalizacja w każdym zadaniu cząstkowym jest rozumiana jako zadanie znalezienia cech, które występują w danej koniunkcji i określenie wartości, które te cechy przyjmują.

Opis zastosowanej metody

Reasumując, przyjęliśmy, że mamy niepusty, skończony zbiór przykładów U i niepusty, skończony zbiór cech $A = \{a_1, \dots, a_K\} \cup \{a_d\}$. $V_{a_j} = \{v_{j_1}, \dots, v_{j_j}\}$, $V_{a_j} \neq \emptyset$, jest zbiorem wartości cechy a_j , $j=1, \dots, K$ oraz $V = \bigcup_{j=1, \dots, K} V_{a_j}$. Cecha decyzyjna a_d ze zbiorem wartości $V_{a_d} = \{v_{d_1}, \dots, v_{d_d}\}$ dzieli zbiór U na rozłączne klasy $\{U_{v_{d_1}}, U_{v_{d_2}}, \dots, U_{v_{d_d}}\}$. Należy utworzyć reguły klasyfikacji dla każdej klasy $U_{v_{t(d)}}$ tak, aby poprawnie opisywały „większość” przykładów uczących.

Przyjmijmy, że mamy P przykładów pozytywnych dla klasy $U_{v_{t(d)}}$, wzór (53), zapisanych w postaci $e^p = \bigwedge_{j=1}^K (a_j = v_{t(j,p)}; g(v_{t(j,p)}))$, $p=1, \dots, P$, oraz N przykładów negatywnych, wzór (54), w postaci $e^n = \bigwedge_{j=1}^K (a_j = v_{t(j,n)}; g(v_{t(j,n)}))$, $n=1, \dots, N$. Indeks $t(j,p)$ określa wartość, jaką przyjmuje cecha a_j w przykładzie e^p , analogicznie indeks $t(j,n)$.

Dla danego przykładu e^p i wszystkich przykładów negatywnych e^n tworzymy macierz zero-jedynkową $Z_{N \times K} = [z_{nj}]$, $n=1, \dots, N$, $j=1, \dots, K$,

$$z_{nj} = \begin{cases} 1 & \text{jeżeli } v_{t(j,p)} \neq v_{t(j,n)} \\ 0 & \text{jeżeli } v_{t(j,p)} = v_{t(j,n)} \end{cases} \quad (61)$$

Wiersze tak określonej macierzy $Z_{N \times K}$ odpowiadają przykładom negatywnym e^n , $n = 1, \dots, N$, a kolumny cechom a_1, \dots, a_K . Zachodzi warunek $z_{nj} = 1$ jeżeli cecha a_j przyjmuje w przykładzie pozytywnym e^p i w przykładzie negatywnym e^n różne wartości; $z_{nj} = 0$ jeżeli wartości te są sobie równe. W macierzy nie występują wiersze zerowe, ponieważ dla każdego przykładu pozytywnego i każdego negatywnego zawsze istnieje, co najmniej jedna cecha przyjmująca różne wartości w obu przykładach, zgodnie z (53) i (54).

Rozważmy przykład e^p . Każdy zero-jedynkowy i K -elementowy wektor x określa koniunkcję C^l , $l \subseteq \{1, \dots, K\}$, złożoną z wybranych warunków z przykładu e^p w sposób opisany poprzednio.

Niech wektor zero-jedynkowy $x = [x_1, \dots, x_K]^T$ określający koniunkcję C^l spełnia dodatkowo następujące ograniczenia

$$\sum_{j=1}^K z_{nj} x_j \geq \gamma_n, \quad n = 1, \dots, N \quad (62)$$

gdzie $\gamma = [\gamma_1, \dots, \gamma_N]^T$ jest wektorem jednostkowym. Wtedy $\forall n$ istnieje, co najmniej jedno $j \in I$ dla którego $z_{nj} x_j \geq 1$. Wnioskując, możemy stwierdzić, że koniunkcja określona przez taki wektor x nie opisuje wszystkich przykładów negatywnych (ponieważ zawiera co najmniej jeden warunek który różni ją od każdego przykładu negatywnego). Oczywiście, zawsze opisuje ona, co najmniej jeden przykład pozytywny (na pewno opisuje przykład e^p).

Niech wektor zero-jedynkowy $x = [x_1, \dots, x_K]^T$ określający koniunkcję C^l spełnia ograniczenia (62) gdzie $\gamma = [\gamma_1, \dots, \gamma_N]^T$ jest tym razem wektorem zero-jedynkowym. Analogicznie rozumując można stwierdzić, że koniunkcja C^l na pewno nie opisuje tych przykładów negatywnych e^n , dla których $\gamma_n = 1$; może natomiast opisywać te przykłady negatywne, dla których $\gamma_n = 0$, ponieważ $\forall j \in I$ może być spełniony warunek $z_{nj} x_j = 0$.

Minimalizację w zadaniu cząstkowym, wzór (7,60), można zapisać

$$\min_{x, \gamma: Z^k \geq \lambda} \sum_{j=1}^K (1 - g(v_{t(j,p)})) \cdot x_j \quad (63)$$

gdzie minimalizacja w sensie szukania zbiorów indeksów I_1, \dots, I_L , wzór (60), została zastąpiona minimalizacją ze względu na wektor x i wektor γ . Znalezione zero-jedynkowy wektor x^* minimalizujący zadanie (63) określa koniunkcję o najmniejszej długości ważonej, wzór (57), a wektor γ^* określa, które przykłady mogą nie być opisywane przez koniunkcję, tzn. koniunkcja nie opisuje co najmniej $(\frac{100}{N} \sum_{n=1}^N \gamma_n)\%$ przykładów negatywnych, wzór (62).

Dlatego też zadanie szukania reguły o minimalnej długości ważonej, wzór (60), można teraz zapisać w postaci ciągu zadań cząstkowych

$$\min_{x, \gamma: Z^k \geq \gamma_1} d(C^{I_1}), \dots, \min_{x, \gamma: Z^k \geq \gamma_L} d(C^{I_L}) \quad (64)$$

z których każde można zapisać jako pewną modyfikację zadania pokrycia

$$\min_{x, \gamma} \sum_{j=1}^K c_j x_j \quad (65)$$

$$\sum_{j=1}^K z_{nj} x_j \geq \gamma_n, \quad n = 1, \dots, N \quad (66)$$

$$\sum_{n=1}^N \gamma_n \geq N - rel \quad (67)$$

gdzie $c_j = 1 - g(v_{(j,p)})$, $z_{nj} \in \{0,1\}$, $x_j \in \{0,1\}$, $j = 1, \dots, K$, $\gamma_n \in \{0,1\}$, $rel \geq 0$. Parametr rel narzuca warunek, aby tworzona koniunkcja nie opisywała, co najmniej $(\frac{100}{N}(N - rel))\%$ przykładów negatywnych dla rozpatrywanej klasy. Przyjęcie $rel = 0$ gwarantuje, że utworzona koniunkcja nie będzie opisywała wszystkich przykładów negatywnych (warunek spójności), przyjęcie $rel > 0$ dopuszcza, że nie wszystkie przykłady muszą być nie opisane (warunek częściowej spójności).

Zadanie pokrycia jest dobrze znanym NP -trudnym zadaniem kombinatorycznym (Gary, Johnson 1979). Niech $I = \{1, \dots, N\}$ będzie danym zbiorem i niech P będzie rodziną „akceptowalnych” podzbiorów I . Z każdym elementem P_j rodziny P , $j = 1, \dots, K$ jest związany koszt c_j .

Zagadnienie pokrycia zbioru I polega na wyborze takich elementów rodziny P , $P' \subseteq P$, dla których łączny koszt jest minimalny i spełniony jest warunek, aby każdy $n \in I$ był zawarty, w co najmniej jednym elemencie rodziny P .

Zadanie określone przez (65), (66), (67) tym się różni się od klasycznego zadania pokrycia, że określana jest w nim dodatkowo prawa strona ograniczeń (66), co odpowiada spełnieniu wymagania, aby „większość” elementów $n \in I$ była zawarta, w co najmniej jednym elemencie rodziny P (a nie wszystkie elementy). Zadanie polega na pokryciu, co najmniej $N - rel$ wierszy macierzy Z o wymiarach $N \times K$ podzbiorem kolumn o minimalnym koszcie. Każdej kolumnie $1 \leq j \leq K$ jest przyporządkowany koszt $c_j = 1 - g(v_{(j,p)})$. Określamy $x_j = 1$ jeśli j -ta kolumna (z kosztem $c_j > 0$) należy do rozwiązania oraz $x_j = 0$ gdy nie należy. Spełnienie ograniczenia (67) gwarantuje, że „większość” wierszy (co najmniej $N - rel$ wierszy) będzie opisana przez co najmniej jedną kolumnę. Ze względu na sposób konstrukcji macierzy Z i wymaganej rozłączności zbiorów przykładów pozytywnych i negatywnych zadanie powyższe ma zawsze rozwiązanie dopuszczalne, jest nim wektor jednostkowy o K elementach.

W zadaniu (65), (66), (67) szukamy zero-jedynkowego wektora x^* o minimalnym koszcie oraz zero-jedynkowego wektora γ^* który określa opisane wiersze macierzy Z ; $\gamma_n^* = 1$ jeżeli n -ty wiersz jest opisany przez rozwiązanie x^* i $\gamma_n^* = 0$ jeżeli n -ty wiersz może nie być opisany. Z założenia, co najmniej $N - rel$ wierszy (których liczba jest zadana przez parametr $rel \geq 0$) musi być opisanych przez rozwiązanie x^* .

Przy tworzeniu reguł o minimalnej długości ważonej kolejno dla każdej klasy $U_{v_{(d)}}$, $\forall v_{(d)} \in \{v_{d_1}, \dots, v_{d_n}\}$, można zastosować metodę IP , tworzącą reguły poprawnie opisujące „większość” przykładów uczących (Szkatuła 1995).

W skrócie, kolejne kroki metody IP przedstawiono poniżej.

Chcemy utworzyć regułę do klasy $U_{v_{i(d)}}$. Dany jest zbiór $NEG_A(U_{v_{i(d)}})$ oraz zbiór $S := POS_A(U_{v_{i(d)}})$. Początkowy zbiór reguł elementarnych do tej klasy, $RUL(U_{v_{i(d)}})$, jest zbiorem pustym, iteracja $z=0$, parametr $rel \geq 0$.

Krok 1.

Iteracja $z = z + 1$. Wyznacza się współczynniki ważności przy współpracy z ekspertami lub dodatkowo modyfikuje je analizując zbiór przykładów uczących.

Krok 2.

Znajduje się najbardziej „typowy” przykład $e^p \in S$ dla danej klasy. Określa się z -te zadanie pokrycia, zgodnie z wzorami (65), (66), (67).

Krok 3.

Wyznacza się koniunkcję C^z , tworzącą regułę elementarną $rul(C^z, v_{i(d)})$, rozwiązując z -te zadanie pokrycia.

Krok 4.

Utworzoną w Kroku 3 koniunkcję dołącza się do zbioru, $RUL(U_{v_{i(d)}}) := RUL(U_{v_{i(d}}) \cup rul(C^z, v_{i(d)})$. Ze zbioru S usuwa się przykłady już opisane.

Krok 5.

Jeżeli dokładność klasyfikacji przykładów z zastosowaniem utworzonej reguły nie jest jeszcze wystarczająca, powtarza się *Krok 1*.

Każda utworzona reguła elementarna może być charakteryzowana przez współczynniki, zależne od liczby przykładów opisywanych przez tą regułę. *Dokładnością reguły* elementarnej będziemy nazywali współczynnik

$$\mu(rul(C^z, v_{i(d)})) = \frac{\text{card}\{\{e : e \in C^z \text{ i } f(e, a_d) = v_{i(d)}\}\}}{\text{card}\{\{e : e \in C^z\}\}} \quad (68)$$

Jeżeli współczynnik dokładności reguły $\mu(rul(C^z, v_{i(d)}))$ jest równy 1, to reguła jest pewna; jeśli jest mniejszy od 1, to mówimy, że reguła jest przybliżona.

Siłę reguły będziemy określali poprzez dwa współczynniki, w sposób następujący

$$q(rul(C^z, v_{i(d)})) = \frac{\text{card}\{\{e : e \in C^z \text{ i } f(e, a_d) = v_{i(d)}\}\}}{\text{card}\{\{e : e \in U\}\}} \quad (69)$$

$$s(rul(C^z, v_{i(d)})) = \frac{\text{card}\{\{e : e \in C^z \text{ i } f(e, a_d) = v_{i(d)}\}\}}{\text{card}\{\{e : f(e, a_d) = v_{i(d)}\}\}} \quad (70)$$

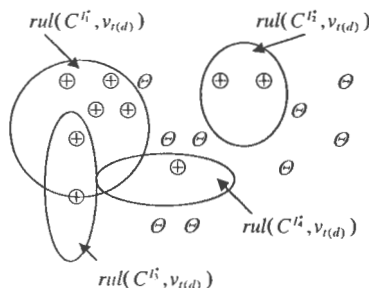
Pierwszy współczynnik jest ilorazem liczby przykładów prawidłowo zaklasyfikowanych i łącznej liczby przykładów uczących; drugi jest ilorazem liczby przykładów prawidłowo zaklasyfikowanych i liczby przykładów uczących należących do rozpatrywanej klasy. Oczywiście zachodzą warunki $0 \leq q(rul(C^z, v_{i(d)})) \leq 1$ oraz $0 \leq s(rul(C^z, v_{i(d)})) \leq 1$. Im więcej przykładów opisuje reguła, tym współczynniki mają większą wartość.

Aby zilustrować zastosowane podejście, rozważmy prosty przykład przedstawiony poniżej.

Przykład 6.

Mamy zbiór dziewięciu przykładów należących do danej klasy $U_{v_{(d)}}$ (na Rysunku 7 oznaczonych znakiem \oplus) oraz zbiór dziewięciu przykładów nie należących do tej klasy (oznaczonych znakiem \ominus). Szukamy zbioru reguł elementarnych postaci (55) do rozpatrywanej klasy, które poprawnie opisują zbiór przykładów należących do klasy i nie opisują przykładów, które do tej klasy nie należą. Przykładowy zbiór takich reguł, ozn. $rul(C^{I^z}, v_{r(d)}), z=1, \dots, 4$, przedstawiono na Rysunku

Zbiór reguł elementarnych dla danej klasy tworzony jest iteracyjnie. W każdym z-tym kroku algorytmu określane jest nowe zadanie pokrycia. Rozwiązaniem z-tego zadania pokrycia jest kolejna reguła elementarna $rul(C^{I^z}, v_{r(d)}), z=1, \dots, 4$, która jest dołączana do tworzonego zbioru reguł.



Rysunek 7. Zbiór przykładów uczących i reguł elementarnych do klasy $U_{v_{(d)}}$

□

W literaturze zamieszczono wiele optymalnych i heurystycznych algorytmów rozwiązujących zadania pokrycia. Pierwszymi praktycznymi algorytmami były heurystyki typu zachłannego (Johnson 1974), (Lovasz 1975), (Chvatal 1979). Obiecujące wyniki uzyskano stosując metody oparte na algorytmach genetycznych (Beasley 1996), jak również oparte na sieciach neuronowych (Croall, Mason 1991). Ciekawe porównanie dziewięciu przybliżonych algorytmów zawarto w pracy (Grossman, Wool 1995).

Do rozwiązania zadania pokrycia w *kroku 3* zaproponowanej metody można zastosować różne algorytmy. Można zastosować np. algorytm przybliżony oparty na algorytmie zaproponowanym przez Chvatala (1997), który buduje pokrycie w sposób sekwencyjny (Szkatuła 1995); lub też algorytm genetyczny (Kacprzyk, Szkatuła 2002).

Reguły klasyfikujące mogą być określane bezpośrednio przez ekspertów na podstawie ich wiedzy i doświadczeń, jak również mogą być tworzone z zastosowaniem metod uczenia maszynowego (jak pokazano w pracy), na podstawie zbiorów zgromadzonych przykładów, danych zawierających informacje o podjętych w przeszłości, w określonych okolicznościach decyzjach oraz ich skutkach i kosztach.

Reguły takie mogą być użyte m.in. w eksperckich systemach doradczych, tzw. systemach wspomaganie decyzji z bazą wiedzy. Eksperckie systemy doradcze ze względu na zakres ich zastosowań, można podzielić na systemy diagnostyczne i systemy prognostyczne. Systemy diagnostyczne są narzędziem poznawczym do oceny już zaistniałych procesów i

zjawisk ekonomicznych, mogą służyć, np. do oceny obecnej kondycji ekonomicznej przedsiębiorstwa na podstawie dostępnych i zweryfikowanych sprawozdań finansowych z minionych okresów. Systemy prognostyczne służą jako narzędzie decyzyjne pomocne przy sporządzaniu planów, są przystosowane do przetwarzania informacji ekonomicznej podanej np. w postaci charakterystyk czasowych. Systemy te mogą być, zatem przydatne w badaniu dynamiki i tendencji rozwojowych pewnych czynników ekonomiczno finansowych. W praktycznych zastosowaniach eksperckie systemy doradcze okazały się skutecznym narzędziem i znajdują coraz większe zastosowanie w bankowości i finansach.

4 Elementy teorii zbiorów rozmytych w uczeniu maszynowym

4.1 Wprowadzenie

Zbiory rozmyte zostały zaproponowane przez Lotfi A. Zadeha w 1965 roku jako numeryczne narzędzie do opisu niepewności, nieprecyzyjności, wieloznaczności towarzyszących nieodłącznie ludzkiej percepcji, myśleniu, mówieniu itp. Przykładem jest lingwistyczna niepewność języka naturalnego (Zadeh 1973), który zawiera pojęcia nieprecyzyjne i wieloznaczne.

W teorii zbiorów rozmytych dowolny element może należeć do zbioru w pewnym stopniu. Stopień przynależności elementu ze zbioru U do zbioru rozmytego \tilde{G} jest określany przez funkcję przynależności o wartościach z przedziału $[0, 1]$, oznaczaną $\mu_{\tilde{G}}:U \rightarrow [0,1]$. Zbiór rozmyty jest utożsamiany z funkcją przynależności do tego zbioru, która określa, w jakim stopniu jesteśmy skłonni uznać, że dany element należy do tego zbioru. Użycie skali liczbowej umożliwia wygodne odwzorowanie stopnia przynależności elementu do zbioru rozmytego; poczynając od liczby 1 oznaczającej całkowitą przynależność, do liczby 0 oznaczającej nie przynależność do zbioru. Literatura dotycząca zbiorów rozmytych jest bardzo liczna, semantykę zbiorów rozmytych opisano między innymi w pracy (Dubois i Prade 1997).

Przykładowo, pojęcia „małe ryzyko” i „średnie ryzyko” mogą być różnie odbierane przez różnych ludzi. Kredyt w wysokości 55 mln może być oceniany zarówno jako mały jak też jako średni; może więc należeć zarówno do zbioru rozmytego „mały kredyt” jak też do zbioru rozmytego „średni kredyt” w różnym stopniu.

Wybrane, podstawowe pojęcia teorii zbiorów rozmytych przedstawiono w części 4.2. W części 4.3 krótko omówiono zastosowanie teorii zbiorów rozmytych w zadaniu uczenia maszynowego na podstawie przykładów, przy tworzeniu reguł rozmytych.

4.2 Podstawowe pojęcia teorii zbiorów rozmytych

Zbiór rozmyty

Zbiór rozmyty jest definiowany jako zbiór elementów z funkcją przynależności do tego zbioru rozmytego, będącą liczbą rzeczywistą z przedziału od zera (całkowite nie należenie do zbioru) do jeden (pełna przynależność) (Zadeh 1965, 1973), (Kandel 1986), (Klir i Folger 1988). Funkcja przynależności określa, w jakim stopniu każdy element zbioru należy do zbioru rozmytego. Bardziej formalnie zbiór rozmyty zdefiniowano poniżej. Dla oznaczania zbiorów przyjmowane będą w pracy duże litery alfabetu, zbiory rozmyte oznaczane będą dodatkowo znakiem \sim .

Definicja 1. Zbiór rozmyty

Niech U będzie zbiorem. Zbiór rozmyty \tilde{G} w zbiorze U (inaczej nazywany podzbiorem rozmytym zbioru U) jest określony poprzez funkcję przynależności $\mu_{\tilde{G}}: U \rightarrow [0,1]$, gdzie $\mu_{\tilde{G}}(u)$ określa stopień, z jakim element $u \in U$ należy do zbioru rozmytego \tilde{G} .

Funkcja przynależności $\mu_{\tilde{G}}: U \rightarrow [0,1]$ może być dowolną funkcją odwzorowującą zbiór U na przedział domknięty $[0,1]$. Zbiory rozmyte mogą być określane w skończonych lub nieskończonych zbiorach U . W literaturze stosuje się inną notację dla obu przypadków. Jeżeli zbiór U jest dyskretny i skończony, $\text{Card}(U) = N$, wtedy zbiór rozmyty jest określane w formie N wymiarowego wektora, którego elementy stanowią wartości funkcji przynależności elementów ze zbioru U do zbioru rozmytego.

Czasami stosuje się w literaturze notację sumy, która w tym przypadku stanowi tylko symbol oznaczający zbiór uporządkowanych par: element ze zbioru U i jego stopień przynależności do zbioru rozmytego. Znak $/$ stanowi symbol łączący, a nie znak dzielenia. Jeżeli $U = \{u_1, u_2, \dots, u_N\}$ wtedy zbiór rozmyty \tilde{G} postaci

$$\tilde{G} = \{(u_i; \mu_{\tilde{G}}(u_i)), \dots, (u_N; \mu_{\tilde{G}}(u_N))\} \quad (71)$$

bywa zapisywany w literaturze w postaci sumy (Zadeh 1965), (Kandel 1986)

$$\tilde{G} = u_1 / \mu_{\tilde{G}}(u_1) + \dots + u_N / \mu_{\tilde{G}}(u_N) = \sum_{i=1}^N \{u_i / \mu_{\tilde{G}}(u_i)\} \quad (72)$$

lub też w postaci zbioru

$$\tilde{G} = \{u_i / \mu_{\tilde{G}}(u_i), \dots, u_N / \mu_{\tilde{G}}(u_N)\}. \quad (73)$$

Jeżeli zbiór U jest zbiorem nieprzeliczalnym lub ciągłym, wtedy stosowane jest w literaturze inne oznaczenie, gdzie notacja \int oznacza tylko symbol zbioru, a nie znak całki

$$\tilde{G} = \int_U \mu_{\tilde{G}}(u) / u. \quad (74)$$

Niech \tilde{G}, \tilde{H} będą zbiorami rozmytymi w zbiorze U . Mówimy, że dwa zbiory rozmyte \tilde{G}, \tilde{H} są sobie równe, gdy $\forall u \in U$ zachodzi zależność $\mu_{\tilde{G}}(u) = \mu_{\tilde{H}}(u)$.

Zbiór rozmyty \tilde{G} zawiera się w zbiorze rozmytym \tilde{H} , ozn. $\tilde{G} \subseteq \tilde{H}$, wtedy i tylko wtedy, gdy $\forall u \in U$ zachodzi zależność $\mu_{\tilde{G}}(u) \leq \mu_{\tilde{H}}(u)$.

Zbiór rozmyty \tilde{G} ściśle zawiera się w zbiorze rozmytym \tilde{H} wtedy i tylko wtedy, gdy $\forall u \in U$ zachodzi zależność $\mu_{\tilde{G}}(u) < \mu_{\tilde{H}}(u)$.

W języku naturalnym, obok dokładnych wartości liczbowych takich jak 5 czy też 127, używane są pojęcia nieprecyzyjne, takie jak np. „około 50”, „niemal 3.5”, „mniej więcej 127”. Pojęcia takie można zdefiniować w postaci zbiorów rozmytych w zbiorze liczb rzeczywistych (Dubois i Prade 1978), nazywanych liczbami rozmytymi. Bardziej formalnie pojęcie liczby rozmytej zdefiniowano poniżej.

Liczbą rozmytą w zbiorze liczb rzeczywistych \mathfrak{R} nazywamy zbiór rozmyty \tilde{G} , który ma ciągłą funkcję przynależności $\mu_{\tilde{G}}: \mathfrak{R} \rightarrow [0,1]$; istnieje liczba rzeczywista $u \in \mathfrak{R}$ taka, że

$\mu_{\bar{c}}(u) = 1$ oraz dla dowolnych $u, v, w \in \mathfrak{R}$ takich, że $u < v < w$, zachodzi warunek $\mu_{\bar{c}}(v) \geq \min\{\mu_{\bar{c}}(u), \mu_{\bar{c}}(w)\}$.

Określanie funkcji przynależności

Istnieją dwa główne sposoby określania funkcji przynależności, przedstawione poniżej.

- *Określanie wartości funkcji przynależności dla każdego elementu zbioru.* Przykładowo, rozpatrując zbiór kredytobiorców U i zbiór rozmyty „kredytobiorcy o niskiej płynności finansowej”, musimy określić dla każdego kredytobiorcy ze zbioru U wartość funkcji przynależności do zbioru rozmytego.

Przykład 1.

Rozpatrzmy zbiór kredytobiorców $U = \{k1, k2, k3, k4, k5, k6\}$. Zbiór rozmyty „kredytobiorcy o niskiej płynności finansowej” w zbiorze U możemy określić następująco: $\{(k1; 0.0), (k2; 0.0), (k3; 0.1), (k4; 0.4), k5; 0.8), (k6; 1.0)\}$. Zapis taki oznacza, że kredytobiorca $k6$ na pewno należy do rozpatrywanego zbioru rozmytego, kredytobiorcy $k1$ i $k2$ na pewno nie należą, a kredytobiorcy $k3, k4$ i $k5$ należą do tego zbioru w pewnym stopniu. \square

- *Określanie postaci funkcji przynależności.* Szukamy formalnego wyrażenia, które określa stopień przynależności każdego elementu zbioru U do zbioru rozmytego.

Przykład 2.

Niech U będzie zbiorem liczb określających wielkość kredytu udzielanego w mln. zł. Możemy przyjąć, że zbiór rozmyty „średni kredyt” może zostać określony w zbiorze U przez funkcję przynależności daną w postaci:

$$\mu_{\text{średni kredyt}}(u) = \begin{cases} 0 & \text{dla } u \leq 50 \\ \frac{u-50}{10} & \text{dla } 50 < u < 60 \\ 1 & \text{dla } 60 \leq u \leq 90 \\ \frac{100-u}{10} & \text{dla } 90 < u < 100 \\ 0 & \text{dla } u \geq 100 \end{cases}$$

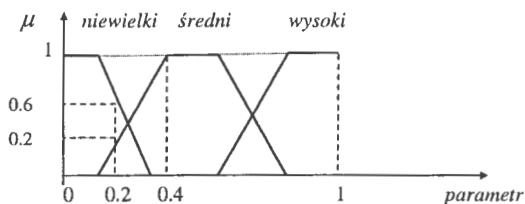
Powiemy, że kredyt w kwocie 55 mln. zł. jest „średnim kredytem” w stopniu 0.5, ponieważ $\mu_{\text{średni kredyt}}(55) = 0.5$, a kredyt w kwocie 90 mln. zł. jest „średnim kredytem” w stopniu 1.0, ponieważ $\mu_{\text{średni kredyt}}(90) = 1.0$. \square

W praktyce, przy określaniu funkcji przynależności można zastosować jedną ze znanych z literatury, często stosowanych typów funkcji przynależności.

Przykład 3.

Rozpatrzmy parametr, oceniający stopień ryzyka, jakie może zaistnieć przy transakcjach kredytowych, jako liczbę z przedziału $U = [0, 1]$. Jeżeli chcemy opisać stopień ryzyka w języku naturalnym, możemy utworzyć zmienną lingwistyczną o symbolicznej nazwie „stopień ryzyka” z wartościami: „niewielki”, „średni” i „wysoki”. Każda wartość, którą może przyjąć zmienna lingwistyczna określa pewien zbiór rozmyty. W naszym przypadku możemy określić trzy zbiory rozmyte w zbiorze parametrów: „niewielki stopień ryzyka”, „średni” i „wysoki”, zawierające te wartości parametrów, które są odbierane przez ludzi odpowiednio jako niewielkie, średnie lub wysokie. Aby zdefiniować rozpatrywane zbiory musimy dla każdej war-

tości parametru ze zbioru U określić, do jakiego zbioru należy. Subiektywność ocen ludzkich ekspertów może prowadzić do rozmycia granic tych zbiorów.



Rysunek 8. Funkcje przynależności do zbiorów rozmytych

Zakładając, że funkcje przynależności do zbiorów rozmytych mają kształt trapezowy, na Rysunku 8 przedstawiono ich graficzne interpretacje. Zakresy tak określonych zbiorów rozmytych częściowo się pokrywają. Tak więc np. parametr 0.4 na pewno określa średni stopień ryzyka (pełna przynależność do zbioru rozmytego „średni” stopień ryzyka), a parametr 0.2 należy jednocześnie do dwóch zbiorów rozmytych, z różnym stopniem przynależności (należy do zbioru „niewielki” stopień ryzyka w stopniu 0.6 oraz do zbioru „średni” stopień ryzyka w stopniu 0.2), jak pokazano na Rysunku 8. □

Interpretacja funkcji przynależności do zbioru rozmytego może być dwojaka.

- *Lingwistyczna.* Funkcja przynależności modeluje pewne kategorie lingwistyczne, takie jak „przybliżony”, „dokładny” lub też może określać na przykład stopień podobieństwa danego obiektu do pewnego wzorca.
- *Nie lingwistyczna.* Typowym przykładem jest związanie stopnia przynależności z podobieństwem elementu do prototypu.

Określanie postaci funkcji przynależności do zbiorów rozmytych w zadaniach praktycznych nie jest zadaniem prostym. Można określać wartości funkcji przynależności bezpośrednio, korzystając z wiedzy ekspertów lub też można określać je automatycznie na podstawie analizy danych uczących. Kombinacja tych podejść polega na wstępnym określaniu postaci funkcji przynależności przez ekspertów a następnie na automatycznej ich modyfikacji z zastosowaniem zbioru przykładów uczących.

Metody stosujące opinie ludzkich ekspertów

Metody stosujące opinie ludzkich ekspertów stosują zazwyczaj pojęcia lingwistyczne. Eksperti określają stopnie przynależności zarówno pojedynczych elementów jak również zbiorów elementów. Wśród metod bazujących na opiniach ekspertów przy określaniu stopni przynależności można wyróżnić między innymi pięć grup metod przedstawionych poniżej (Bilgic i Turksen 1997, 1999). Nie wszystkie metody mogą być stosowane wymiennie. Przykładowo, estymacja przedziałowa nie może być stosowana dla zbiorów nieuporządkowanych.

- *Metody poziome.* Wymagają one grupy ekspertów. Estymacja jest określana poprzez określanie stosunku liczby pozytywnych odpowiedzi ekspertów do liczby wszystkich odpowiedzi. Przykład pytania: czy Kowalski jest wiarygodnym klientem banku?

- *Metody kierunkowe.* Mogą być stosowane przy opinii jednego jak też grupy ekspertów. Przykład pytania: w jakim stopniu można uznać Kowalskiego za wiarygodnego klienta banku?
- *Metody pionowe.* Określają zbiory rozmyte z ich α -cięć. Przykładowy zbiór klientów, którzy są wiarygodni w stopniu co najmniej 0.3.
- *Metody estymacji przedziałowej.* Wymagają one grupy ekspertów. Przedziały są tworzone w oparciu o zastosowanie różnych metod głosowania.
- *Metody oparte na porównywaniu parami.* Mogą być one stosowane przy opinii jednego jak też grupy ekspertów. Wynik jest macierzą porównań parami określającymi wartości funkcji przynależności. Przykład pytania: kto jest bardziej wiarygodny, Kowalski czy też Nowak i w jakim stopniu?

Metody automatyczne

Metody automatyczne określają wartości funkcji przynależności na podstawie zbioru przykładów uczących (Keller i Givens 1985), (Pal i Chakraborty 1984), (Parui i Majumder 1982). Medasani w pracy (Medasani, Kim i Krishnapuram 1998) podsumował metody automatycznego określania funkcji przynależności, zawarł też dyskusję zastosowania w tym celu metod heurystycznych, probabilistycznych, opartych na sieciach neuronowych i wielu innych. Cztery często stosowane praktycznie metody z tej grupy przedstawiono poniżej.

- *Estymacja jedno punktowa* (Pal i Chakraborty 1984). Niech $U = \{u_1, \dots, u_N\} \subset \mathfrak{R}$ będzie zbiorem danych, dla których wartości funkcji przynależności do zbioru rozmytego \tilde{G} chcemy estymować. Przyjmujemy $u_{\max} = \max_n \{u_n\}$, $u_{\min} = \min_n \{u_n\}$, $\bar{u} = \frac{1}{N} \sum_{n=1}^N u_n$. Definiujemy parametry $a_r = \bar{u} + \max\{|\bar{u} - u_{\min}|, |\bar{u} - u_{\max}|\}$, $a_l = 2\bar{u} - a_r$ i funkcja przynależności do zbioru rozmytego jest postaci

$$\mu_{\tilde{G}}(u) = \begin{cases} 0 & u \leq a_l \\ 2 \left(\frac{u - a_l}{\bar{u} - a_l} \right)^2 & a_l < u \leq \frac{a_l + \bar{u}}{2} \\ 1 - 2 \left(\frac{u - \bar{u}}{\bar{u} - a_l} \right)^2 & \frac{a_l + \bar{u}}{2} < u \leq \frac{a_r + \bar{u}}{2} \\ 2 \left(\frac{u - a_r}{\bar{u} - a_r} \right)^2 & \frac{a_r + \bar{u}}{2} < u \leq a_r \\ 0 & u \geq a_r \end{cases}$$

- *Grupowanie (clustering)* (Bezdek 1981), (Medasani, Kim i Krishnapuram 1998). Zakładamy, że w zbiorze przykładów U mamy c grup przykładów z centroidami $\bar{u}_1, \dots, \bar{u}_c$ (w przestrzeni wielowymiarowej centroidy są wektorami). Dla każdego punktu $u \in \mathfrak{R}$ możemy policzyć stopień należenia do grupy \tilde{G}_i , $i = 1, \dots, c$ jako

$$\mu_{\tilde{G}_i}(u) = \frac{\frac{1}{(u - \bar{u}_i)^2}}{\frac{1}{(u - \bar{u}_1)^2} + \dots + \frac{1}{(u - \bar{u}_c)^2}}$$

Przykład u zaliczamy do grupy o najwyższym stopniu funkcji przynależności.

- *Histogramy.* Tworzy się histogram z danych i na jego podstawie tworzy się funkcje przynależności.
- *Estymacja k najbliższych sąsiadów.* Mamy zbiór elementów U ze znaną przynależnością do klasy, G_i , $i = 1, \dots, c$. Chcemy estymować c stopni przynależności $\mu_{G_1}(u), \dots, \mu_{G_c}(u)$, $\forall u \in \mathfrak{R}$. Po pierwsze znajdujemy k punktów w zbiorze U przy punkcie u , a następnie obliczamy $\mu_{G_i}(u) = \frac{k_i}{k}$, k_i jest liczbą elementów należących do klasy G_i .

Właściwości zbiorów rozmytych

Ponieważ zbiory rozmyte są określane poprzez swoje funkcje przynależności, możemy je charakteryzować bardziej dokładnie poprzez odniesienie do właściwości tych funkcji (Kandel, 1986), (Dubois i Prade 1993), (Zadeh 1975, 1978, 1979, 1981), (Zimmermann 1987).

Niech U będzie zbiorem, a \tilde{G} zbiorem rozmytym w zbiorze U . Dla zbioru rozmytego \tilde{G} Możemy określić proste operacje jednoargumentowe, przedstawione poniżej.

$$\text{▪ jądro zbioru rozmytego} \quad \text{Core}(\tilde{G}) = \{u \mid u \in U, \mu_{\tilde{G}}(u) = 1\}, \quad (75)$$

$$\text{▪ nośnik zbioru rozmytego} \quad \text{Supp}(\tilde{G}) = \{u \mid u \in U, \mu_{\tilde{G}}(u) > 0\}, \quad (76)$$

$$\text{▪ wysokość zbioru rozmytego} \quad \text{hgt}(\tilde{G}) = \sup_{u \in U} \{\mu_{\tilde{G}}(u)\}, \quad (77)$$

$$\text{▪ cięcie typu } \alpha \text{ zbioru rozmytego} \quad \tilde{G}_\alpha = \{u \mid u \in U, \mu_{\tilde{G}}(u) \geq \alpha\}, \quad (78)$$

$$\text{▪ liczność zbioru rozmytego skończonego} \quad \text{card}(\tilde{G}) = \sum_{u \in U} \mu_{\tilde{G}}(u), \quad (79)$$

$$\text{nieskończonego} \quad \text{card}(\tilde{G}) = \int_{u \in U} \mu_{\tilde{G}}(u). \quad (80)$$

$$\text{▪ względna liczność skończonego zbioru rozmytego} \quad \|\tilde{G}\| = \frac{\text{card}(\tilde{G})}{\text{card}(U)} \quad (81)$$

Zgodnie z powyższymi definicjami jądro zbioru rozmytego jest zbiorem tych elementów ze zbioru U , które w stopniu 1 należą do zbioru rozmytego, natomiast nośnik jest zbiorem tych elementów ze zbioru U które w stopniu większym od zera należą do tego zbioru. Łatwo zauważyć, że jądrem zbioru rozmytego \tilde{G} jest cięcie typu α tego zbioru dla wartości $\alpha = 1$, $\text{Core}(\tilde{G}) = \tilde{G}_{\alpha=1}$. Można pokazać, że jeżeli $\alpha_1 \geq \alpha_2$ to zachodzi warunek zawierania zbiorów $\tilde{G}_{\alpha_1} \subseteq \tilde{G}_{\alpha_2}$.

Przykład 4.

Rozpatrzmy zbiór rozmyty $\tilde{G} = \{(1; 0.1), (2; 0.3), (3; 0.6), (4; 1.0), (5; 0.4), (6; 0.0)\}$, określony w zbiorze liczb $U = \{1, 2, 3, 4, 5, 6\}$. Zbiór rozmyty \tilde{G} ma jądro $\text{Core}(\tilde{G}) = \{4\}$, nośnik $\text{Supp}(\tilde{G}) = \{1, 2, 3, 4, 5\}$, wysokość $\text{hgt}(\tilde{G}) = 1$, liczność $\text{card}(\tilde{G}) = 2.4$, a jego cięcie typu α dla $\alpha = 0.5$ wynosi $\tilde{G}_{\alpha=0.5} = \{3, 4\}$. \square

Zbiór rozmyty \tilde{G} w zbiorze U określony przez funkcję $\mu_{\tilde{G}}(u)$, $u \in U$, nazywamy:

- *normalnym*, jeżeli jego wysokość jest równa jeden, tzn.

$$\sup_{u \in U} \{\mu_{\tilde{G}}(u)\} = 1, \quad (82)$$

- *wypukłym*, jeżeli $\forall u_1, u_2 \in U, \forall \lambda \in [0,1]$ zachodzi warunek:

$$\mu_{\tilde{G}}(\lambda \cdot u_1 + (1-\lambda) \cdot u_2) \geq \min\{\mu_{\tilde{G}}(u_1), \mu_{\tilde{G}}(u_2)\}, \quad (83)$$

- *wklęsłym*, jeżeli $\forall u_1, u_2 \in U, \forall \lambda \in [0,1]$ zachodzi warunek:

$$\mu_{\tilde{G}}(\lambda \cdot u_1 + (1-\lambda) \cdot u_2) \leq \min\{\mu_{\tilde{G}}(u_1), \mu_{\tilde{G}}(u_2)\}. \quad (84)$$

Operacje jednoargumentowe na zbiorach rozmytych

Na zbiorach rozmytych można definiować wiele prostych operacji jednoargumentowych. Niektóre z nich, znane z literatury, zostały przedstawione poniżej.

Definicja 2. Dopełnienie zbioru rozmytego

Dopełnienie zbioru rozmytego \tilde{G} w zbiorze U jest zbiorem rozmytym w U , oznaczanym $\neg\tilde{G}$, określonym funkcją przynależności $\mu_{\neg\tilde{G}}:U \rightarrow [0,1]$. Funkcja przynależności do dopełnienia zbioru rozmytego jest definiowana przez funkcję $h:[0,1] \rightarrow [0,1]$, to znaczy $\mu_{\neg\tilde{G}}(u) = h(\mu_{\tilde{G}}(u))$, $\forall u \in U$, która ma następujące własności:

1. *h jest funkcją jednoargumentową, zależną tylko od $\mu_{\tilde{G}}(u)$ dla $u \in U$,*
2. *$h(0)=1$, $h(1)=0$,*
3. *h jest ciągła i monotonicznie malejąca,*
4. *$h(h(a)) = a$, dla $a \in [0,1]$.*

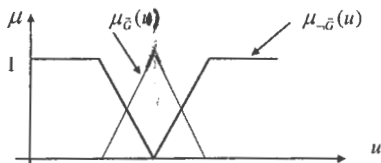
Warunki od (1) do (4) nie określają jednoznacznie postaci funkcji przynależności do dopełnienia zbioru rozmytego. Jeżeli do powyższych warunków dodamy jeszcze jeden warunek

5. *jeżeli $a+b=1$, dla $a, b \in [0,1]$, to zachodzi $h(a)+h(b)=1$,*

otrzymamy funkcję przynależności do dopełnienia zbioru \tilde{G} spełniającą warunki od (1) do (5) w następującej postaci

$$\mu_{\neg\tilde{G}}(u) = 1 - \mu_{\tilde{G}}(u). \quad (85)$$

Powyższa definicja dopełnienia zbioru rozmytego została określona przez Zadeha i nazywana jest *standardowym dopełnieniem zbioru rozmytego $\neg\tilde{G}$ w U* . Może być ona przedstawiona graficznie w przypadku, gdy funkcja przynależności jest funkcją trójkątną w sposób pokazany na Rysunku 9.



Rysunek 9. Graficzna interpretacja zbioru rozmytego z trójkątną funkcją przynależności i jego dopełnienia

Sugeno podał inną definicję dopełnienia zbioru rozmytego, zwanego *dopełnieniem typu λ*, ozn. $\neg_{\lambda} \tilde{G}$, spełniającego warunki od (1) do (4).

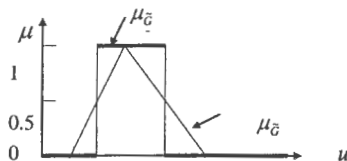
$$\neg_{\lambda} \tilde{G} = \frac{1 - \mu_{\tilde{G}}(u)}{1 + \lambda \cdot \mu_{\tilde{G}}(u)}, \quad \lambda \in [-1, \infty) \quad (86)$$

Na zbiorach rozmytych można definiować wiele innych, prostych operacji znanych z literatury, niektóre z nich zostały przedstawione poniżej.

Dla zbioru rozmytego \tilde{G} , $u \in U$, można określać *najbliższy mu zbiór zwykły* oznaczany \tilde{G} , o funkcji przynależności danej w postaci:

$$\mu_{\tilde{G}}(u) = \begin{cases} 0 & \text{dla } \mu_{\tilde{G}}(u) \leq 0.5 \\ 1 & \text{dla } \mu_{\tilde{G}}(u) > 0.5. \end{cases} \quad (87)$$

Graficzną interpretację najbliższego zbioru zwykłego dla zbioru rozmytego z trójkątną funkcją przynależności przedstawiono na Rysunku 10.



Rysunek 10. Zbiór zwykły najbliższy zbiorowi rozmytemu \tilde{G}

Pojęcie zbioru zwykłego, najbliższego zbiorowi rozmytemu odgrywa dużą rolę przy określaniu wskaźników rozmytości.

Zbiór rozmyty \tilde{G} możemy *normalizować*. W operacji normalizacji zbioru rozmytego funkcja przynależności dzielona jest przez maksymalną wartość, którą może ona przyjmować. W rezultacie nowa funkcja $Norm_{-\mu_{\tilde{G}}}$ przyjmuje wartości z przedziału $[0, 1]$,

$$Norm_{-\mu_{\tilde{G}}} = \frac{\mu_{\tilde{G}}}{hgt(\tilde{G})} \quad (88)$$

Na zbiorze rozmytym \tilde{G} możemy dokonywać operacji *koncentracji*,

$$Con_{-}\mu_{\tilde{G}} = (\mu_{\tilde{G}}(u))^p, \quad \text{dla } p > 1. \quad (89)$$

Zgodnie z definicją zbioru rozmytego zachodzi warunek $0 \leq \mu_{\tilde{G}}(u) \leq 1$, zatem mamy dla $p > 1$ spełnienie warunku $(\mu_{\tilde{G}}(u))^p < \mu_{\tilde{G}}(u)$, poza warunkami brzegowymi równymi $\mu_{\tilde{G}}(u) = 0$, $\mu_{\tilde{G}}(u) = 1$. Wielkość koncentracji zależy od elementu u i wynosi

$$\frac{\mu_{\tilde{G}}(u) - (\mu_{\tilde{G}}(u))^p}{\mu_{\tilde{G}}(u)} = 1 - (\mu_{\tilde{G}}(u))^{p-1}.$$

Im element $u \in U$ ma mniejszą wartość funkcji przynależności do zbioru \tilde{G} , tym wartość ta zostanie bardziej zmniejszona. Przykładowo, dla $p=2$, element o wartości funkcji przynależności równej $\mu_{\tilde{G}}(u) = 0.1$ będzie miał zmniejszoną o 90% wartość tej funkcji do $(\mu_{\tilde{G}}(u))^2 = 0.01$, podczas gdy element, dla którego wartość funkcji przynależności wynosi $\mu_{\tilde{G}}(u) = 0.9$ zmniejszoną tylko do $(\mu_{\tilde{G}}(u))^2 = 0.81$, to znaczy o 10%.

Operacja koncentracji bywa stosowana w zbiorach rozmytych do reprezentowania pojęcia „bardzo”. Przykładowo, jeżeli mamy funkcję przynależności $\mu_{\text{„wiarygodny”}}(u)$ do zbioru rozmytego „wiarygodny”, to funkcja $\text{Con}_{\mu_{\text{„wiarygodny”}}}$ może być funkcją przynależności do zbioru rozmytego „bardzo wiarygodny”.

Odwrotnością operacji koncentracji, dla $p=2$, jest operacja *rozcieńczenia* (ang. *dilation*)

$$\text{Dil}_{\mu_{\tilde{G}}} = \sqrt{\mu_{\tilde{G}}(u)} \quad (90)$$

Przykładowo, element o wartości funkcji przynależności $\mu_{\tilde{G}}(u) = 0.01$ będzie miał po operacji rozcieńczenia nową wartość tej funkcji równą $\sqrt{\mu_{\tilde{G}}(u)} = 0.1$.

W zagadnieniach praktycznych często wyniki otrzymane w postaci zbiorów rozmytych należy przetworzyć do postaci nie rozmytej. Zadanie takie jest dużo łatwiejsze w przypadku, gdy rozpatrywany zbiór rozmyty mniej różni się od zbioru ostrego. Ważną sprawą jest więc określanie stopnia rozmytości zbioru rozmytego.

Chcąc określić na ile rozmyty w zbiorze U jest zbiór \tilde{G} wprowadza się *miarę rozmytości zbioru*, oznaczaną $H(\tilde{G})$. Kaufmann zaproponował dwie miary rozmytości przedstawione poniżej, oparte na odległości Hamminga i odległości Euklidesowej między zbiorem rozmytym \tilde{G} i najbliższym mu zbiorem zwykłym $\underline{\tilde{G}}$ wybranym w sposób nieco arbitralny, patrz wzór (87).

$$H_{\text{Hamming}}(\tilde{G}) = \frac{2}{N} d_H(\tilde{G}, \underline{\tilde{G}}) = \frac{2}{N} \sum_{i=1}^N \left| \mu_{\tilde{G}}(u_i) - \mu_{\underline{\tilde{G}}}(u_i) \right| \quad (91)$$

oraz

$$H_{\text{Euklides}}(\tilde{G}) = \frac{2}{\sqrt{N}} d_E(\tilde{G}, \underline{\tilde{G}}) = 2 \sqrt{\frac{1}{N} \sum_{i=1}^N (\mu_{\tilde{G}}(u_i) - \mu_{\underline{\tilde{G}}}(u_i))^2} \quad (92)$$

De Luca i Termini w 1972 zaproponowali miarę rozmytości opartą na entropii Shannona

$$H_{\text{Entropii}}(\tilde{G}) = -k \sum_{i=1}^N [\mu_{\tilde{G}}(u_i) \cdot \log(\mu_{\tilde{G}}(u_i)) + \mu_{\underline{\tilde{G}}}(u_i) \cdot \log(\mu_{\underline{\tilde{G}}}(u_i))] \quad (93)$$

gdzie $k > 0$ jest współczynnikiem skalującym.

Przy definiowaniu powyższych miar przyjęto założenie o skończoności zbioru $U = \{u_1, \dots, u_n\}$, definicje można również uogólnić na przypadek zbioru nieskończonego (Klir, Folger 1988).

Działania na zbiorach rozmytych

W klasycznej teorii zbiorów *suma* i *iloczyn* zbiorów są prostymi działaniami, które są jednoznacznie definiowane. W teorii zbiorów rozmytych interpretacja sumy i iloczynu zbiorów rozmytych nie jest tak prosta. Przy określaniu na zbiorach rozmytych sumy i iloczynu stosuje się normy trójkątne typu t oraz typu s . Niech U będzie zbiorem, \tilde{A}, \tilde{B} niech będą zbiorami rozmytymi w zbiorze U .

Norma typu t definiowana jest jako funkcja $t: [0,1] \times [0,1] \rightarrow [0,1]$, oznaczana $t(a,b)$, mająca następujące własności dla $a, b, c \in [0,1]$:

1. przemienność, $t(a, b) = t(b, a)$,
2. łączność, $t(t(a, b), c) = t(a, t(b, c))$,
3. monotoniczność, $a \leq c$ i $b \leq d$ implikuje $t(a, b) \leq t(c, d)$,
4. spełnienie warunku brzegowego, $t(a, 1) = a$.

Z własności (3) wynika, że dla $\forall a \in [0,1]$ zachodzi $t(0, a) \leq t(0, 1)$, stąd i z warunku (4) otrzymujemy zależność $t(0, a) = 0$.

Norma typu s definiowana jest jako funkcja $s: [0,1] \times [0,1] \rightarrow [0,1]$, oznaczana $s(a,b)$, mająca następujące własności dla $a, b, c \in [0,1]$,

1. przemienność, $s(a, b) = s(b, a)$,
2. łączność, $s(s(a, b), c) = s(a, s(b, c))$,
3. monotoniczność, $a \leq c$ i $b \leq d$ implikuje $s(a, b) \leq s(c, d)$,
4. spełnienie warunku brzegowego $s(a, 0) = a$.

Z własności (3) oraz (4) wynika, że dla $\forall a \in [0,1]$ zachodzi zależność $s(1, a) = 1$.

Dla każdej normy typu t istnieje norma typu s zwana *normą dualną*, taka, że dla $a, b \in [0,1]$ spełnione są warunki przedstawione poniżej.

$$t(a, b) = 1 - s((1-a), (1-b)) \quad (94)$$

$$s(a, b) = 1 - t((1-a), (1-b)). \quad (95)$$

Przykładami najczęściej stosowanych w literaturze norm typu t są przedstawione poniżej funkcje, dla $a, b \in [0,1]$

- *minimum*, oznaczane też $a \wedge b$, $t_1(a, b) = \min\{a, b\}$,
- *iloczyn*, $t_2(a, b) = ab$,
- *różnica ograniczona*, $t_3(a, b) = \max\{0, a + b - 1\}$,

$$\blacksquare \text{ drastyczny iloczyn, } t_4(a,b) = \begin{cases} a & \text{jeżeli } b=1 \\ b & \text{jeżeli } a=1 \\ 0 & \text{jeżeli } a < 1, b < 1 \end{cases}$$

a przykładami norm typu s są funkcje:

$$\blacksquare \text{ maksimum, oznaczane } a \vee b, \quad s_1(a,b) = \max\{a,b\},$$

$$\blacksquare \text{ suma probabilistyczna, } \quad s_2(a,b) = a + b - ab,$$

$$\blacksquare \text{ suma ograniczona, } \quad s_3(a,b) = \min\{1, a + b\},$$

$$\blacksquare \text{ drastyczna suma } \quad s_4(a,b) = \begin{cases} a & \text{jeżeli } b=0 \\ b & \text{jeżeli } a=0 \\ 1 & \text{jeżeli } a > 0, b > 0 \end{cases}$$

Poniżej przedstawiono definicje podstawowych operacji na zbiorach rozmytych: sumy zbiorów rozmytych i ich przecięcia.

Definicja 3. Działania na zbiorach rozmytych

Niech U będzie zbiorem, \tilde{A}, \tilde{B} niech będą zbiorami rozmytymi w U . Działania na zbiorach rozmytych definiujemy następująco:

- suma zbiorów rozmytych (ang. union) $\tilde{A} \cup \tilde{B}$:

$$\tilde{A} \cup \tilde{B} = \{(u; \mu_{\tilde{A} \cup \tilde{B}}(u)) \mid \forall u \in U, \mu_{\tilde{A} \cup \tilde{B}}(u) = s(\mu_{\tilde{A}}(u), \mu_{\tilde{B}}(u))\},$$

- przecięcie zbiorów rozmytych (ang. intersection) $\tilde{A} \cap \tilde{B}$:

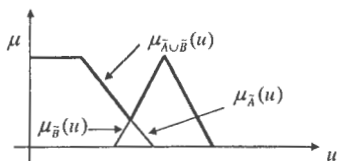
$$\tilde{A} \cap \tilde{B} = \{(u; \mu_{\tilde{A} \cap \tilde{B}}(u)) \mid \forall u \in U, \mu_{\tilde{A} \cap \tilde{B}}(u) = t(\mu_{\tilde{A}}(u), \mu_{\tilde{B}}(u))\}.$$

Przyjmując różne postacie norm typu t i typu s , możemy działania na zbiorach rozmytych definiować w różny sposób, w zależności od wymagań rozpatrywanego zagadnienia. Najczęściej w zadaniach praktycznych przyjmuje się funkcję minimum jako normę typu t oraz funkcję maksimum jako normę typu s .

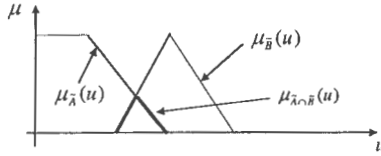
Uwaga.

Trzeba pamiętać, że ogólnie normy trójkątne nie gwarantują spełnienia prawa zaprzeczenia i prawa wyłączoności środka, $\tilde{A} \cap \tilde{A} \neq \emptyset$ oraz $\tilde{A} \cup \tilde{A} \neq U$.

Interpretacje graficzne działań na zbiorach rozmytych dla wybranego typu normy przedstawiono na rysunkach 11 i 12.



Rysunek 11. Suma zbiorów, $\mu_{\tilde{A} \cup \tilde{B}}(u) = \max\{\mu_{\tilde{A}}(u), \mu_{\tilde{B}}(u)\}$



Rysunek 12. Przecięcie zbiorów, $\mu_{\tilde{A} \cap \tilde{B}}(u) = \min\{\mu_{\tilde{A}}(u), \mu_{\tilde{B}}(u)\}$

W dalszej części pracy będziemy rozpatrywali tylko normalne zbiory rozmyte, wzór (82), aby zapewnić spełnienie prawa zaprzeczenia i wyłączonego środka, co jest dogodnie w zastosowaniach praktycznych. Oczywiście, z każdego zbioru rozmytego możemy otrzymać normalny zbiór rozmyty, stosując operację normalizacji, wzór (88).

Podobieństwo zbiorów rozmytych jest podstawowym pojęciem w zagadnieniach rozpoznawania wzorców. Również w zadaniach klasyfikacji często porównujemy zbiory rozmyte. Niech \tilde{A}, \tilde{B} niech będą zbiorami rozmytymi w zbiorze U . W pracy (Dubois i Prade 1980) zamieszczono przegląd różnych miar podobieństwa zbiorów rozmytych, przykładowo

$$S_1(\tilde{A}, \tilde{B}) \equiv \frac{\|\tilde{A} \cap \tilde{B}\|}{\|\tilde{A} \cup \tilde{B}\|}$$

gdzie $\|\cdot\|$ oznacza względną licznosc skończonego zbioru rozmytego (81).

Poprzednio zajmowaliśmy się agregacją dwóch zbiorów rozmytych \tilde{A}, \tilde{B} w zbiorze U . Rozpatrzmy teraz zagadnienie agregacji większej liczby zbiorów rozmytych.

Operatory agregacji

Operator agregacji, oznaczany jako $A: [0,1]^L \rightarrow [0,1]$, $A(a_1, a_2, \dots, a_L) \in [0,1]$, $a_i \in [0,1]$, powinien mieć przedstawione poniżej własności:

1. przemienność, $A(a_1, a_2, \dots, a_L) = A(a_{i_1}, a_{i_2}, \dots, a_{i_L})$ dla każdej permutacji i_1, i_2, \dots, i_L indeksów $1, 2, \dots, L$,
2. monotoniczność, jeżeli $a_i \leq b_i$, $i = 1, \dots, L$ to $A(a_1, a_2, \dots, a_L) \leq A(b_1, b_2, \dots, b_L)$,
3. spełnienie warunków brzegowych, $A(0, 0, \dots, 0) = 0$, $A(1, 1, \dots, 1) = 1$.

Agregacja A może być scharakteryzowana jako operator koniunkcyjny, dysjunkcyjny lub kompromisowy, tzn. dla każdych a_1, a_2, \dots, a_L , gdzie $a_i \in [0,1]$, otrzymujemy:

- *operator koniunkcyjny*, $A(a_1, a_2, \dots, a_L) \leq \min\{a_1, a_2, \dots, a_L\}$,
- *operator dysjunkcyjny*, $A(a_1, a_2, \dots, a_L) \geq \max\{a_1, a_2, \dots, a_L\}$,
- *operator kompromisowy*, $\min\{a_1, a_2, \dots, a_L\} \leq A(a_1, a_2, \dots, a_L) \leq \max\{a_1, a_2, \dots, a_L\}$.

Poniżej przedstawiono kilka prostych operatorów agregacji spełniających warunki od (1) do (3), znanych z literatury:

- *minimum*, $A(a_1, a_2, \dots, a_L) = \min\{a_1, a_2, \dots, a_L\}$,
- *maksimum*, $A(a_1, a_2, \dots, a_L) = \max\{a_1, a_2, \dots, a_L\}$,
- *iloczyn*, $A(a_1, a_2, \dots, a_L) = \prod_{i=1}^L (a_i)$,
- *średnia*, $A(a_1, a_2, \dots, a_L) = \sum_{i=1}^L a_i$,
- *średnia uogólniona*, $A_\alpha(a_1, a_2, \dots, a_L) = \left(\frac{1}{L} \sum_{i=1}^L a_i^\alpha \right)^{\frac{1}{\alpha}}$.

Operatory OWA

Ciekawą klasę operatorów agregujących stanowią zaproponowane przez Yagera (Yager 1988, 1997) *operatory OWA (Ordered Weighted Averaging)*.

Niech $b = [b_1, b_2, \dots, b_L]^T \in [0,1]^L$ będzie wektorem współczynników takim, że $\sum_{k=1}^L b_k = 1$. Operator OWA określony jest na produkcie kartezjańskim wektora b i wektora $[a_{i_1}, a_{i_2}, \dots, a_{i_L}]^T$, gdzie i_1, i_2, \dots, i_L jest permutacją indeksów $1, 2, \dots, L$ taką, że $a_{i_1} \geq a_{i_2} \geq \dots \geq a_{i_L}$,

$$A_b^{OWA}(a_1, a_2, \dots, a_L) = \sum_{k=1}^L a_{i_k} b_k.$$

Tak określony operator spełnia warunki od (1) do (3) operatora agregującego A . Poprzez wybór odpowiedniego wektora b można realizować między innymi następujące funkcje:

- *minimum*, przyjmując $b = [0, 0, \dots, 0, 1]^T$,
- *maksimum*, przyjmując $b = [1, 0, \dots, 0]^T$,
- *średnią*, przyjmując $b = [1/L, 1/L, \dots, 1/L]^T$,
- *konkurs jurorów*, przyjmując $b = [0, 1/(L-2), \dots, 1/(L-2), 0]^T$.

Wartości wektora b mogą być ustalone z góry, jak również mogą być wyliczane na podstawie danego zbioru danych. Yager i Filew w 1994 pokazali, że poprzez odpowiedni dobór współczynników, można zastosować operatory OWA do modelowania kwantyfikatorów lingwistycznych takich jak „prawie wszystkie”, „kilka”, „wiele”, „większość”, itd.

Zastosowanie konsensusu

Aby zagregować zbiór wartości a_1, a_2, \dots, a_L , $a_i \in [0,1]$ można zastosować *operatory agregacji oparte na konsensusie* (Kuncheva 1996). Szacujemy konsensus w zakresie przedziału $[0,1]$, gdzie 0 oznacza całkowitą rozbieżność, a 1 oznacza jednorodność. Pięć miar konsensusu $\gamma_1, \dots, \gamma_5$ (Kuncheva 2000) przedstawiono poniżej.

- *najwyższa zgodność*, $\gamma_1 = 1 - \min_{i,j \neq j} \{ |a_i - a_j| \}$,

- najwyższa niezgodność, $\gamma_2 = 1 - \max_{i,j \neq j} \{|a_i - a_j|\}$,
- całkowita średnia zgodność, $\gamma_3 = 1 - \sum_{i=1}^{L-1} |a_i - \bar{a}|$, gdzie $\bar{a} = \frac{1}{L} \sum_{i=1}^L a_i$,
- całkowita zgodność dla wszystkich par, $\gamma_4 = 1 - \frac{2}{L(L-1)} \sum_{i=1}^{L-1} \sum_{j=i+1}^L |a_i - a_j|$,
- całkowita najwyższa niezgodność, $\gamma_5 = 1 - \max_{\kappa \in \{(1, \dots, L)\}} \{|a_i - \bar{a}|\}$.

Dla $L=2$ otrzymujemy następujące zależności $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 1 - |a_1 - a_2|$. Miara ta może być stosowana dla określenia zagregowanej opinii ekspertów jako liczb z przedziału $[0, 1]$.

Niech $A(a_1, a_2, \dots, a_L)$ będzie pewną zagregowaną wartością otrzymaną z zastosowaniem operatora A . Operacja konsensusu K powinna zależeć od agregowanych wartości (a_1, a_2, \dots, a_L) i miary konsensusu γ pomiędzy nimi, $K(a_1, a_2, \dots, a_L) = K(A(a_1, a_2, \dots, a_L), \gamma(a_1, a_2, \dots, a_L))$.

Dla uproszczenia zapisu będziemy stosować notację $K=K(A, \gamma)$. Operacja konsensusu K powinna spełniać przedstawiony poniżej zbiór aksjomatów:

1. przemienność,
2. częściowa monotoniczność dla miary γ ,

$$K(A, \gamma) \text{ jest } \begin{cases} \text{monotonicznie nie malejącą dla } \gamma & \text{jesli } A > 0.5 \\ \text{monotonicznie nie rosnącą dla } \gamma & \text{jesli } A \leq 0.5 \end{cases}$$

3a. jednomyślność, $K(A, 1) = A$, lub

3b. wzmocniona jednomyślność,

$$K(A, 1) \text{ jest } \begin{cases} > A & \text{jesli } A \in (0.5, 1) \\ < A & \text{jesli } A \in (0, 0.5) \\ = A = 0.5 & \text{jesli } A = 0.5 \end{cases}$$

Operacja konsensusu spełniająca aksjomaty (1), (2) i (3a) jest postaci $K_1(A, \gamma) = 0.5(1 - \gamma) + A\gamma$, a spełniająca aksjomaty (1), (2) i (3b) jest postaci $K_2(A, \gamma) = \frac{1}{1 + \exp\{-\alpha\gamma(A - 0.5)\}}$, gdzie $\alpha > 0$ jest stałą skalującą.

Jeżeli miara konsensusu γ jest równa 0, wtedy wartość zagregowana jest zawsze równa 0.5 i nie zależy od A . Jest to odbiciem sytuacji, w której jeśli nie ma porozumienia pomiędzy jurorami przyjmowana jest wartość 0.5. Jeżeli wartość zagregowana A jest równa zero, operator konsensusu K może zwrócić wartość z przedziału od 0 do 0.5. Niezerowa wartość K przy warunku $A=0$ wskazuje, że istnieje niezgodność wśród a_1, a_2, \dots, a_L .

Relacje rozmyte

Załóżmy, że mamy skończone, niepuste zbiory U i V . Relacją ze zbioru U do zbioru V , nazywamy produkt kartezjański $U \times V$ definiowany w postaci $R = \{(u, v) \mid \forall u \in U, \forall v \in V\}$, gdzie (u, v) jest uporządkowaną parą. Jeśli U zawiera $card(U)$ elementów, a zbiór V zawiera $card(V)$ elementów, wtedy iloczyn kartezjański ma $card(U) \cdot card(V)$ elementów.

Definicja 4. Relacja rozmyta

Niech U, V będą skończonymi, niepustymi zbiorami. Relacja rozmyta ze zbioru U do zbioru V , ozn. \tilde{R} , określona jest na iloczynie kartezjańskim $U \times V \rightarrow [0,1]$ i definiowana w postaci

$$\tilde{R} = \{(u, v); \mu_{\tilde{R}}(u, v)\} \quad \forall u \in U, \forall v \in V$$

gdzie $\mu_{\tilde{R}}(u, v) \in [0,1]$ jest funkcją określającą w jakim stopniu element $u \in U$ jest w relacji z elementem $v \in V$.

W literaturze czasami stosuje się do zapisu relacji rozmytej notację sumy, która w tym przypadku stanowi tylko symbol oznaczający zbiór uporządkowanych par

$$\tilde{R} = \bigcup_{U \times V} \{(u, v); \mu_{\tilde{R}}(u, v)\}. \quad (96)$$

Jeżeli zbiory U i V są zbiorami skończonymi oraz $Card(U)=N, Card(V)=K$, stosuje się zapis

$$\tilde{R} = \{(u_i, v_j); \mu_{\tilde{R}}(u_i, v_j)\} \quad j = 1, \dots, K, \quad i = 1, \dots, N \}. \quad (97)$$

W tym przypadku relacja może być reprezentowana w postaci macierzy relacji, której elementami są wartości funkcji przynależności $\mu_{\tilde{R}}(u_i, v_j)$, $i = 1, \dots, N, j = 1, \dots, K$.

	v_1	v_2	...	v_K
u_1	$\mu_{\tilde{R}}(u_1, v_1)$	$\mu_{\tilde{R}}(u_1, v_2)$...	$\mu_{\tilde{R}}(u_1, v_K)$
	...			
u_N	$\mu_{\tilde{R}}(u_N, v_1)$	$\mu_{\tilde{R}}(u_N, v_2)$...	$\mu_{\tilde{R}}(u_N, v_K)$

(98)

Jeżeli elementy zbiorów U i V są ciągłe, stosuje się następujący zapis

$$\tilde{R} = \int_U \int_V \{(u_i, v_j); \mu_{\tilde{R}}(u_i, v_j)\}. \quad (99)$$

Każdą relację rozmytą \tilde{R} określoną na iloczynie kartezjańskim $U \times V$ możemy rzutować na zbiór U oraz na zbiór V

$$Pr \text{ o}j_U \tilde{R}(u) = \sup_{v \in V} \tilde{R}(u, v) \quad (100)$$

$$Pr \text{ o}j_V \tilde{R}(v) = \sup_{u \in U} \tilde{R}(u, v). \quad (101)$$

Operacja rzutowania relacji zmniejsza liczbę jej elementów.

Przypadek dwuwymiarowy można rozszerzyć na wielowymiarowy. Niech U_1, U_2, \dots, U_K będą skończonymi, niepustymi zbiorami. Relacja rozmyta określona na iloczynie kartezjańskim $U_1 \times U_2 \times \dots \times U_K \rightarrow [0,1]$ jest definiowana w następującej postaci

$$\tilde{R} = \{(u_1, u_2, \dots, u_K); \mu_{\tilde{R}}(u_1, u_2, \dots, u_K)\} \quad \forall u_1 \in U_1, \dots, \forall u_K \in U_K \} \quad (102)$$

gdzie $\mu_{\tilde{R}}(u_1, u_2, \dots, u_k) \in [0,1]$ jest funkcją określającą, w jakim stopniu elementy (u_1, u_2, \dots, u_k) są ze sobą w relacji.

Niech U, V, W będą skończonymi, niepustymi zbiorami. Niech \tilde{R}_1 będzie rozmytą relacją ze zbioru U do V z funkcją przynależności $\mu_{\tilde{R}_1}(u, v) \in [0,1]$, dla $u \in U, v \in V$, oraz niech \tilde{R}_2 będzie rozmytą relacją ze zbioru V do W z funkcją przynależności $\mu_{\tilde{R}_2}(v, w) \in [0,1]$, dla $v \in V, w \in W$.

Definicja 5. Złożenie relacji rozmytych

Złożeniem dwóch relacji rozmytych, \tilde{R}_1 określonej na $U \times V \rightarrow [0,1]$ i \tilde{R}_2 określonej na $V \times W \rightarrow [0,1]$, jest relacja rozmyta oznaczana $\tilde{R}_1 \circ \tilde{R}_2$ określona na $U \times W \rightarrow [0,1]$, dla norm typu s i typu t

$$\tilde{R}_1 \circ \tilde{R}_2 = \{(u, w); s_{v \in V} [\mu_{\tilde{R}_1}(u, v), \mu_{\tilde{R}_2}(v, w)]\} \quad \forall u \in U, \forall w \in W \}.$$

Jeżeli przyjmiemy za normę typu t funkcję minimum, a za normę typu s funkcję maksimum otrzymujemy następującą zależność na złożenie dwóch relacji rozmytych:

$$\tilde{R}_1 \circ \tilde{R}_2 = \{(u, w); \max_{v \in V} \{ \min(\mu_{\tilde{R}_1}(u, v), \mu_{\tilde{R}_2}(v, w)) \}\} \quad \forall u \in U, \forall w \in W \} \tag{103}$$

gdzie $\max_{v \in V} \{x\}$ oznacza wybór maksymalnej wartości x dla wszystkich elementów $v \in V$.

Możemy również wybierać inne normy, np. $s(u, v) = \min(u + v, 1)$ oraz $t(u, v) = uv$. Wybór odpowiedniej funkcji dostosowujemy do konkretnego zadania.

Jeżeli zbiory U, V i W są skończone, wtedy macierz relacji $\tilde{R}_1 \circ \tilde{R}_2$ otrzymujemy ze złożenia macierzy relacji dla \tilde{R}_1 i macierzy relacji dla \tilde{R}_2 .

Przykład 5.

Niech dwie relacje

$$\tilde{R}_1 = \{(u_i, v_j); \mu_{\tilde{R}_1}(u_i, v_j)\} \quad j = 1, 2 \quad i = 1, 2 \quad \tilde{R}_2 = \{(v_i, w_j); \mu_{\tilde{R}_2}(v_i, w_j)\} \quad j = 1, 2 \quad i = 1, 2 \}$$

będą reprezentowane w postaci macierzy relacji, zgodnie z (98), w następującej postaci

	\tilde{R}_1	
	v_1	v_2
u_1	0.3	0.8
u_2	0.6	0.9

	\tilde{R}_2	
	w_1	w_2
v_1	0.5	0.9
v_2	0.4	1.0

Złożenie tych dwóch relacji jest postaci

$$\tilde{R}_1 \circ \tilde{R}_2 = \{(u_i, w_j); \mu_{\tilde{R}_1 \circ \tilde{R}_2}(u_i, w_j)\} \quad j = 1, 2, \quad i = 1, 2 \}$$

co możemy zapisać w postaci macierzy relacji $\tilde{R}_1 \circ \tilde{R}_2$

	w_1	w_2
u_1	0.4	0.8
u_2	0.5	0.9

Wartość 0.4 w macierzy $\tilde{R}_1 \circ \tilde{R}_2$ jest stopniem z jakim elementy (u_i, w_i) są ze sobą w relacji, i została obliczona poprzez porównanie pierwszego wiersza w macierzy \tilde{R}_1 i pierwszej kolumny w macierzy \tilde{R}_2 . Porównujemy 0.3 i 0.5 wybierając 0.3 (minimum z tych dwóch wartości); porównujemy 0.8 i 0.4 wybierając 0.4 (minimum z tych dwóch wartości). Następnie wybieramy maksymalną z wybranych wartości 0.3 i 0.4, otrzymując 0.4. Postępowanie to można zapisać w postaci:

$$\begin{aligned} \mu_{\tilde{R}_1 \circ \tilde{R}_2}(u_i, w_i) &= \max_{v \in \{w_1, w_2\}} [\min(\mu_{\tilde{R}_1}(u_i, v), \mu_{\tilde{R}_2}(v, w_i))] = \\ &= \max\{\min[\mu_{\tilde{R}_1}(u_1, v_1), \mu_{\tilde{R}_2}(v_1, w_1)], \min[\mu_{\tilde{R}_1}(u_1, v_2), \mu_{\tilde{R}_2}(v_2, w_1)]\} = \\ &= \max\{\min[0.3, 0.5], \min[0.8, 0.4]\} = \max\{0.3, 0.4\} = 0.4. \end{aligned}$$

Dla pozostałych wartości w macierzy postępujemy w podobny sposób. \square

Niech \tilde{U} będzie zbiorem rozmytym w zbiorze U , niech \tilde{V} będzie zbiorem rozmytym w zbiorze V . *Relacja rozmyta określona na iloczynie kartezjańskim zbiorów rozmytych* $\tilde{U} \times \tilde{V}$ jest definiowana w postaci

$$\tilde{R} = \{(u, v); t(\mu_{\tilde{U}}(u), \mu_{\tilde{V}}(v)) \mid \forall u \in \tilde{U}, \forall v \in \tilde{V}\} \quad (104)$$

gdzie $\mu_{\tilde{U}}(u) \in [0,1]$ i $\mu_{\tilde{V}}(v) \in [0,1]$ są funkcjami przynależności określającymi, w jakim stopniu element u należy do zbioru \tilde{U} i analogicznie element v do zbioru \tilde{V} .

Przykładowo, przyjmując jako normę typu t funkcję minimum otrzymujemy następującą zależność

$$\tilde{R} = \{(u, v); \min\{\mu_{\tilde{U}}(u), \mu_{\tilde{V}}(v)\} \mid \forall u \in \tilde{U}, \forall v \in \tilde{V}\}. \quad (105)$$

Wnioskowanie rozmyte

Istnieją dwa podstawowe typy wnioskowania rozmytego, opartego na regułach rozmytych *JEŻELI ... TO ...*, rozpatrywane w zależności od typu konkluzji. Poniżej przedstawiono ich uproszczone zapis oraz ich interpretację graficzną.

- *Wnioskowanie typu Mamdaniego-Assilianiana (MA)*. Zarówno przesłanki jak i konkluzje mogą być reprezentowane przez lingwistyczne wartości.

▪

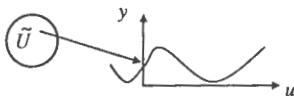
JEŻELI $u \in \tilde{U}$ *TO* $v \in \tilde{V}$

gdzie \tilde{U} i \tilde{V} są zbiorami rozmytymi.



- *Wnioskowanie typu Takagi-Sugeno-Kang (TSK)*. Konkluzje są pewnymi funkcjami, często wielomianami.

JEŻELI $u \in \tilde{U}$ TO $y=f(u, p)$
 gdzie \tilde{U} jest zbiorem rozmytym,
 f jest funkcją $u \in U$ i parametru p .



Przykładowo, reguła JEŻELI kredytobiorca jest bardzo zadłużony TO jest mało wiarygodny oparta jest na dwóch zbiorach rozmytych: \tilde{U} : „duże zadłużenie” oraz \tilde{V} : „mała wiarygodność” w zbiorze kredytobiorców U , opisanych za pomocą dwóch cech: zadłużenie i wiarygodność. Reguła ta może być utożsamiana z relacją rozmytą określoną na iloczynie kartezjańskim zbiorów rozmytych $\tilde{U} \times \tilde{V}$ i zgodnie z definicją 104, może być zapisana w postaci

$$\tilde{R} = \{(u, v); t(\mu_{\tilde{U}}(u), \mu_{\tilde{V}}(v))\} \quad \forall u \in \tilde{U}, \forall v \in \tilde{V} \quad (106)$$

gdzie $\mu_{\tilde{U}}(u) \in [0,1]$ i $\mu_{\tilde{V}}(v) \in [0,1]$ są funkcjami przynależności określającymi, w jakim stopniu element u należy do zbioru \tilde{U} i analogicznie element v należy do zbioru \tilde{V} .

Stopień spełnienia reguły prostej JEŻELI $u \in \tilde{U}$ TO $v \in \tilde{V}$ jest rozważany w zależności od przyjętej do obliczeń normy typu t , czyli od sposobu liczenia funkcji przynależności $\mu_{\tilde{U} \times \tilde{V}}(u, v)$. Zauważmy, że reguła prosta określona na iloczynie kartezjańskim dwóch zbiorów rozmytych jest definiowana na $[0,1] \times [0,1] \rightarrow [0,1]$, to znaczy jest operacją dwuargumentową definiowaną na parach $(\mu_{\tilde{U}}(u), \mu_{\tilde{V}}(v))$, $\forall u \in U$.

Poniżej przedstawiono kilka operatorów relacji rozmytych (implikacji rozmytych) znanych z literatury, stosowanych do reprezentacji reguł rozmytych JEŻELI $u \in \tilde{U}$ TO $v \in \tilde{V}$, zwanych prostymi regułami rozmytymi lub rozmytymi implikacjami.

- *Mamdaniego* $\mu_{\tilde{U} \times \tilde{V}}(u, v) = \min\{\mu_{\tilde{U}}(u), \mu_{\tilde{V}}(v)\}$
- *Larsena* $\mu_{\tilde{U} \times \tilde{V}}(u, v) = \mu_{\tilde{U}}(u) \cdot \mu_{\tilde{V}}(v)$
- *Kleenego-Dienes* $\mu_{\tilde{U} \times \tilde{V}}(u, v) = \max\{1 - \mu_{\tilde{U}}(u), \mu_{\tilde{V}}(v)\}$
- *Łukasiewicza* $\mu_{\tilde{U} \times \tilde{V}}(u, v) = \min\{1, 1 - \mu_{\tilde{U}}(u) + \mu_{\tilde{V}}(v)\}$
- *Zadeha* $\mu_{\tilde{U} \times \tilde{V}}(u, v) = \max\{\min\{\mu_{\tilde{U}}(u), \mu_{\tilde{V}}(v)\}, 1 - \mu_{\tilde{U}}(u)\}$
- *Stochastyczna* $\mu_{\tilde{U} \times \tilde{V}}(u, v) = \min\{1, 1 - \mu_{\tilde{U}}(u) + \mu_{\tilde{U}}(u) \cdot \mu_{\tilde{V}}(v)\}$
- *Gougena* $\mu_{\tilde{U} \times \tilde{V}}(u, v) = \begin{cases} 1 & \mu_{\tilde{U}}(u) \leq \mu_{\tilde{V}}(v) \\ \frac{\mu_{\tilde{V}}(v)}{\mu_{\tilde{U}}(u)} & \text{w pozostałych przypadkach} \end{cases}$
- *Godla* $\mu_{\tilde{U} \times \tilde{V}}(u, v) = \begin{cases} 1 & \mu_{\tilde{U}}(u) \leq \mu_{\tilde{V}}(v) \\ \mu_{\tilde{V}}(v) & \text{w pozostałych przypadkach} \end{cases}$
- *Sharpa* $\mu_{\tilde{U} \times \tilde{V}}(u, v) = \begin{cases} 1 & \mu_{\tilde{U}}(u) \leq \mu_{\tilde{V}}(v) \\ 0 & \text{w pozostałych przypadkach} \end{cases}$

Przykład 6.

Niech zbiór U będzie zbiorem trzech klientów, $U = \{u_1, u_2, u_3\}$. Zdefiniujmy rozmyty zbiór klientów o dużym dochodzie $\tilde{U} = \{u_1; 0.8, u_2; 0.7, u_3; 0.1\}$ oraz rozmyty zbiór klientów wiary-

godnych, $\tilde{V} = \{u_1; 0.0, u_2; 0.5, u_3; 0.9\}$. Dana jest reguła rozmyta *JEŻELI klient ma duży dochód TO jest wiarygodnym klientem*. Sprawdźmy, w jakim stopniu spełniają tę regułę rozmytą klienci ze zbioru U , w zależności od przyjętego operatora relacji.

Klient pierwszy u_1 należy do zbioru rozmytego \tilde{U} w stopniu 0.8, a do zbioru rozmytego \tilde{V} w stopniu 0.0. Przyjmując operator *Mamdaniego* otrzymujemy $\mu_{\tilde{U}, \tilde{V}}(u_1, u_1) = \min\{\mu_{\tilde{U}}(u_1), \mu_{\tilde{V}}(u_1)\} = \min\{0.8, 0.0\} = 0.0$, dla operatora *Larsena* otrzymujemy $0.8 \cdot 0.0 = 0.0$, dla operatora *Kleenego-Dienesy* otrzymujemy $\max\{1 - 0.8, 0.0\} = 0.2$, dla operatora *Łukasiewicza* otrzymujemy $\min\{1, 1 - 0.8 + 0.0\} = \min\{1, 0.2\} = 0.2$, dla operatora *Zadeha* otrzymujemy $\max\{\min\{0.8, 0.0\}, 1 - 0.8\} = \max\{0.0, 0.2\} = 0.2$. Tak więc, reguła jest spełniona przez pierwszego klienta w stopniu 0.0 gdy przyjmiemy operator *Mamdaniego* lub *Larsena*, a dla pozostałych operatorów jest spełniona w stopniu 0.2.

Klient trzeci u_3 należy do zbioru rozmytego \tilde{U} w stopniu 0.1, a do zbioru rozmytego \tilde{V} w stopniu 0.9. Przyjmując operator *Mamdaniego* otrzymujemy $\mu_{\tilde{U}, \tilde{V}}(u_3, u_3) = \min\{0.1, 0.9\} = 0.1$, dla operatora *Larsena* otrzymujemy $\mu_{\tilde{U}, \tilde{V}}(u, v) = 0.1 \cdot 0.9 = 0.09$, dla operatora *Kleenego-Dienesy* otrzymujemy $\max\{1 - 0.1, 0.9\} = 0.9$, dla operatora *Łukasiewicza* otrzymujemy $\min\{1, 1 - 0.1 + 0.9\} = 1$, dla operatora *Zadeha* otrzymujemy $\max\{\min\{0.1, 0.9\}, 1 - 0.1\} = \max\{0.1, 0.9\} = 0.9$. \square

W przypadku rozważania wielu zbiorów rozmytych $\tilde{U}_1, \dots, \tilde{U}_L$ i \tilde{V} w zbiorze U reguła będąca implikacją rozmytą postaci

$$JEŻELI u \in \tilde{U}_1 \text{ i } u \in \tilde{U}_2 \text{ i } \dots \text{ i } u \in \tilde{U}_L \text{ TO } u \in \tilde{V},$$

jest określona na iloczynie kartezjańskim L zbiorów. Aby określić stopień spełnienia reguły musimy agregować wszystkie L wartości, tak więc operacja dwuargumentowa stosowana dotychczas staje się niewystarczająca.

Rodzaje reguł rozmytych

Niech U będzie zbiorem przykładów $u_n \in U$, $n = 1, \dots, N$ opisanych za pomocą cech $a_j \in A$, $j = 1, \dots, K$ ze zbiorem wartości $V_{a_j} = \{v_1^{a_j}, v_2^{a_j}, \dots, v_{k_j}^{a_j}\}$ oraz cechy decyzyjnej $a_d \in A$.

Zajmiemy się dokładniej pierwszym typem wnioskowania typu Mamdaniego-Assiliany. Można wyróżnić trzy typy reguł rozmytych w zależności od typu konkluzji reguły:

- reguły rozmyte z jednoznaczną klasyfikacją

$$JEŻELI \dots \text{ TO } (a_d = v_i^{a_d}) \quad (107)$$

- reguły rozmyte z jednoznaczną klasyfikacją i stopniem przynależności do klasy

$$R_k: JEŻELI \dots \text{ TO } (a_d = v_i^{a_d}); q_i^{a_d} \quad (108)$$

gdzie stopień pewności $q_i^{a_d} \in [0, 1]$ jest interpretowany jako „poparcie” dla jednej klasy, dane przez regułę, jeśli część przesłankowa reguły jest całkowicie spełniona,

- reguły rozmyte z klasyfikacją o stopniach przynależności do wszystkich klas, gdzie stopień pewności $q_i^{a_d} \in [0, 1]$

$$R_k: JEŻELI \dots \text{ TO } (a_d = v_i^{a_d}); q_1^{a_d} \wedge \dots \wedge (a_d = v_{k_d}^{a_d}); q_{k_d}^{a_d} \quad (109)$$

4.3. Tworzenie reguł rozmytych na podstawie zbioru przykładów

Zalóżmy, że dysponujemy zbiorem przykładów, które możemy opisać za pomocą zmiennych ilościowych, będących wynikiem pewnych pomiarów lub obserwacji oraz słów lub zdań w języku naturalnym. Przykłady mogą zawierać dane nieprecyzyjne, wieloznaczne i mogą być obciążone błędami. Pierwszym etapem jest zapisanie takich przybliżonych i niejednoznacznych danych w języku zrozumiałym dla systemu komputerowego.

Zakładając, że dysponujemy zbiorem kredytobiorców (klientów bankowych), należy określić cechy, które ich opisują i oceniają oraz podać zakres wartości, które cechy mogą przyjmować. Cechy te mogą przyjmować wartości liczbowe, jak również wartości lingwistyczne. Przykładami cech o wartościach liczbowych mogą być np. „stan konta”, „przepływy finansowe na koncie”, „wielkość kredytu”, „wielkość depozytu”. Przykładami cech lingwistycznych mogą być np. „wiek”, „założenie”, „wiarygodność”.

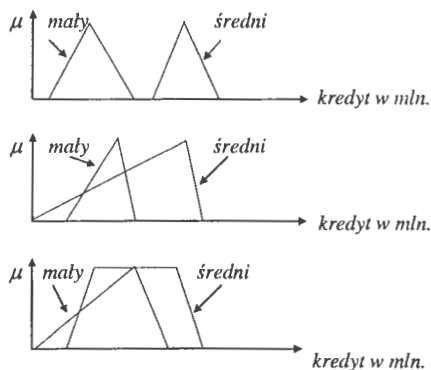
Przykładowo, dla cechy „status konta” można określić wartości lingwistyczne: „bez kredytów”, „kredyt spłacony”, „kredyt spłacany regularnie”, „zwłoki w płaceniu”; dla cechy „zysk banku” można określić wartości: „duży”, „średni”, „mały”, „strata mała”, „strata średnia”, „strata duża”; a dla cechy „płynność finansowa” można określić wartości: „niska” (brak płynności), „średnia”, „wysoka”.

Proces wprowadzania w tym celu zmiennych lingwistycznych opisujących przykłady, wyznaczania nazw zbiorów rozmytych będących wynikiem podziału zakresu zmienności wartości tych zmiennych i następnie określania postaci funkcji przynależności do tych zbiorów nazywa się w literaturze *rozmywaniem* (ang. *fuzzyfication*). W praktyce jest to proces trudny, wyjątkowo źle sformalizowany, wymaga doświadczenia i intuicji oraz dużej wiedzy o rozpatrywanym zagadnieniu. Trudności, które się pojawiają związane są między innymi ze specyfiką rozpatrywanego zadania, subiektywnością ocen ekspertów ludzkich.

Określanie funkcji przynależności

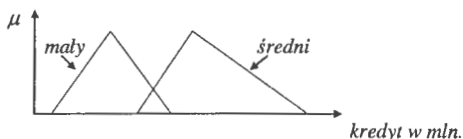
Rozpatrzmy zbiór kredytów U określonych w mln. zł. i niech naszym zadaniem będzie zdefiniowanie zbioru rozmytego „duży” kredyt, poprzez określenie funkcji przynależności $\mu_{\text{duzy}}(u)$, $u \in U$. Przy określaniu funkcji przynależności należy uwzględnić, że może być ona różna, jeżeli u oznacza kredyt konsumpcyjny, niż gdy oznacza kredyt inwestycyjny. Różni ludzie mogą w różny, subiektywny sposób określać postać tej funkcji. W zależności od zadania, które mamy do rozwiązania, jej postać również może się znacznie różnić; inna może być, gdy będziemy rozważać kredyty udzielane osobom prywatnym, a inna w przypadku kredytów udzielanych firmom.

Jeszcze więcej problemów pojawia się, gdy chcemy określić sekwencję zbiorów rozmytych, przykładowo, „mały”, „średni”, „duży” kredyt. W takim przypadku należy uwzględnić intuicyjnie zrozumiałe zależności istniejące pomiędzy tymi zbiorami, a pewne zależności pomiędzy nimi nie mogą być akceptowalne, jak pokazano na Rysunku 13.



Rysunek 13. Nie akceptowalne funkcje przynależności

Przykładowo, intuicyjnie akceptowalną zależność pomiędzy rozpatrywanymi zbiorami rozmytymi przedstawiono na Rysunku 14.



Rysunek 14. Akceptowalne funkcje przynależności

W przypadku określania postaci funkcji przynależności przez ekspertów, istnieją sprawdzone metody praktyczne, które można zastosować. Przykładowo, można uwzględnić opinie tylko jednego eksperta lub też mając opinie grupy ekspertów można je uśrednić; można stosować estymację przedziałową, określając przedziały tak, aby zawierały one dane opinie ekspertów; lub też można porównywać parami kryteria, dla których wartości funkcji przynależności są określone.

Funkcję przynależności można też próbować estymować w sposób automatyczny, na podstawie zbioru danych przykładów. W pracy (Medasani i inni, 1998) przedstawiono wybrane metody stosowane w tym celu, między innymi metody heurystyczne, probabilistyczne, metody stosujące techniki najbliższych sąsiadów, grupujące lub też oparte na sieciach neuro-nowych.

W praktyce, sposób określania funkcji przynależności jest ściśle determinowany przez możliwości i rodzaj rozpatrywanego zadania; inaczej postępujemy, gdy dysponujemy opinią tylko jednego eksperta, inaczej, gdy grupy ekspertów lub też, gdy mamy zbiór danych doświadczalnych.

Tworzenie reguł rozmytych

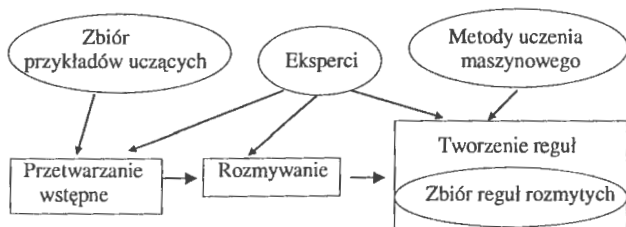
Na tym etapie należy określić cechy decyzyjne (występujące we wnioskach reguł rozmytych), których wartości określają klasy, których opisy w postaci reguł rozmytych będziemy tworzyć. Przykładowo, taką cechą decyzyjną może być cecha „wiarygodność klienta” lub też cecha „zysk banku”. Jeżeli wybierzemy jako cechę decyzyjną „płynność finansową” z wartościami: „niska” (brak płynności), „średnia”, „wysoka”, możemy określić klasę kredytobiorców o ni-

skiej płynności finansowej, świadczącej o utracie lub wysokim prawdopodobieństwie utraty zdolności do terminowego regulowania bieżących zobowiązań; klasę kredytobiorców o średniej płynności finansowej; oraz klasę kredytobiorców o wysokiej płynności, oznaczającej z kolei zamrożenie nadmiernie wysokich własnych kapitałów.

Zbiór przykładów opisanych z zastosowaniem zbiorów rozmytych jest punktem wyjścia przy tworzeniu opisów rozpatrywanych klas w postaci zbioru reguł rozmytych typu *JEŻELI ...TO*, z rozmytymi przesłankami i rozmytymi konkluzjami. Przykład takiej reguły rozmytej przedstawiono poniżej

JEŻELI („dochód” = „mały”) \wedge („oszczędności” = „brak”)
TO („płynność finansowa” = „niska”)

Reguły rozmyte mogą zostać zdefiniowane przez eksperta, mogą być wynikiem modelowania jakościowego, mogą również zostać wygenerowane automatycznie na podstawie danego zbioru przykładów z zastosowaniem algorytmów automatycznego pozyskiwania wiedzy. Wygenerowanie optymalnego zbioru reguł, który spełnia żądane warunki, na podstawie danego zbioru przykładów nie jest jednak zadaniem prostym, prosty schemat ich tworzenia przedstawiono na Rysunku 15.



Rysunek 15. Schemat tworzenia reguł rozmytych

Heurystyczne metody uczenia maszynowego oparte na zbiorach rozmytych można podzielić na dwie główne grupy.

- *Metody oparte na grupowaniu.* Metody te same dokonują grupowania przykładów w klasy przykładów podobnych (zwanymi klastrami) i następnie wykorzystują je przy tworzeniu rozmytych reguł klasyfikujących do tych klas.
- *Metody wymagające zadanej struktury zbiorów rozmytych.* Metody te nie dokonują grupowania przykładów, wymagają one na starcie zadanej rozmytej podziału wartości wszystkich cech wybranych do opisu przykładów (tzn. określenia wartości, jakie może przyjmować każda cecha poprzez zbiory rozmyte), a szukają najlepszego opisu pasującego do zadanej struktury.

Metody oparte na grupowaniu

Metody oparte na grupowaniu grupują zbiór przykładów uczących w grupy przykładów podobnych, zwanych klastrami i następnie stosują je przy tworzeniu reguł. Klasy rozmyte są to wielowymiarowe, dyskretne zbiory rozmyte, które mogą się częściowo pokrywać (Bezdek 1981), (Bezdek i inni 1998). Przegląd kilku algorytmów rozmytego grupowania można znaleźć między innymi w pracy (Hoppner i inni 1999).

Każdy rozmyty klaster może zostać przekształcony w rozmytą regułę poprzez zastosowanie projekcji stopni przynależności przykładów testowych w pojedyncze wymiary (Kla-

wonn i Kruse 1995), (Sugeno i Yasukawa 1993). Tak więc dla każdego klastra i każdej cechy otrzymuje się histogram, na podstawie którego tworzy się funkcje przynależności.

Dla każdej cechy mamy więc wiele różnych zbiorów rozmytych tworzących klastry. Uzyskane wyniki są trudne do interpretacji, ponieważ zbiory rozmyte tworzone są indywidualnie dla każdego klastra. Dla łatwości interpretacji uzyskanych wyników konieczne jest zatem dokonywanie rozmytej partycji na zbiory, mające jasną interpretację lingwistyczną.

Podejście takie znalazło zastosowanie między innymi przy rozpoznawaniu wzorców i aproksymacji funkcji (Berthold, Huber 1999), (Tschichold, Gurman, 1996).

Kratowy klastering (ang. grid clustering) (Klawonn, Keller 1997), (Keller, Klawonn 1998) jest metodą będącą kombinacją różnych metod rozmytych. Wymaga on wstępnego określenia zbiorów rozmytych, a nie nadzorowany algorytm modyfikuje te zbiory rozmyte w celu podziału zbioru danych. Klastry są dane poprzez funkcje przynależności. Uzyskane reguły są proste do interpretacji, ponieważ nie zawierają indywidualnie tworzonych dla każdego klastra zbiorów rozmytych.

Do tworzenia reguł rozmytych można również stosować metody oparte na sieciach neuronowych. Sieć taka stosuje wielowymiarowe funkcje w węzłach warstw ukrytych i każda taka funkcja może być interpretowana jako rozmyty klastr.

Pedrycz i Card zaproponowali sposób lingwistycznej interpretacji samoorganizującej się sieci Kohonena (1984) w celu utworzenia reguł rozmytych (Pedrycz, Card 1992). Zbiór danych zostaje zredukowany do zbioru prototypów, które są reprezentowane przez neurony i wykorzystane następnie przy tworzeniu reguł rozmytych.

Kosko zaproponował metodę grupowania opartą na rozmytej pamięci asocjacyjnej (Kosko 1992), która wiąże wagi z uzyskanymi regułami rozmytymi.

Metody oparte na grupowaniu rozmytym (Grauel i inni 1997), (Hoppner i inni 1999) tworzą rozmyte reguły, które zawierają warunki związane ze zbiorami rozmytymi tworzonymi indywidualnie dla każdego klastra. Takie zbiory rozmyte otrzymywane przez projekcję mają trudną interpretację lingwistyczną, wymagają danych liczbowych i nie mogą pracować z danymi zawierającymi brakujące wartości. W zagadnieniach praktycznych metody grupowania rozmytego są stosowane w przypadkach, gdy lingwistyczna interpretacja uzyskanych wyników nie gra dużej roli.

Kratowy klastering nie ma większości powyższych wad i tworzy łatwo interpretowalne reguły rozmyte, ale ponieważ jest nienadzorowany, jego zastosowanie praktyczne jest ograniczone.

Metody wymagające zadanej struktury zbiorów rozmytych w przestrzeni cech

Metody te wymagają określenia zbiorów rozmytych określających wartości, jakie może przyjmować każda cecha. Zbiory rozmyte mogą być tworzone przez ekspertów indywidualnie dla każdej cechy, z zastosowaniem ich wiedzy związanej z rozpatrywanym zagadnieniem. Można również zastosować znane z literatury trójkątne lub trapezoidowe funkcje przynależności i wtedy będą one interpretowane jako liczby rozmyte lub przedziały rozmyte. Liczba funkcji przynależności określa stopień rozdrobnienia przestrzeni danych. Jeżeli mamy K cech, wtedy każdy opis jest produktem kartezjańskim K zbiorów rozmytych. Ilość możliwych opisów jest równa liczbie wszystkich reguł rozmytych, które mogą być utworzone dla danego na starcie zbioru zbiorów rozmytych. Szukany jest opis najlepiej dopasowujący się do danych uczących, z zastosowaniem zadanej struktury zbiorów rozmytych.

Wang i Mendel (1991, 1992) zaprojektowali algorytm do tworzenia systemów rozmytych dla zadania aproksymacji funkcji. W celu wyznaczenia zależności pomiędzy wartościami na wyjściu i pewną kombinacją zmiennych wejściowych zastosowali reguły ważone. W pracy

(Wang, Mendel 1992) zawarty jest dowód, że algorytm może tworzyć zbiór reguł rozmytych, które przybliżają każdą ciągłą funkcję rzeczywistą w zbiorze gęstym z zadaną dokładnością.

Higgins i Goodman zaproponowali w pracy (Higgins, Goodman 1993) pewną modyfikację algorytmu Wang i Mendela, który zaczyna pracę z jedną funkcją przynależności dla każdej cechy. Sekwencyjnie nowe funkcje przynależności są tworzone w punktach maksymalnego błędu i rozmyta partycja jest poprawiana. Tak więc stare reguły są niepotrzebne i tworzony jest nowy zbiór reguł w oparciu o nową partycję. Procedura jest iteracyjna, powtarzana do momentu, gdy maksymalna liczba zbiorów rozmytych zostanie utworzona lub też, gdy błąd będzie dostatecznie mały. Algorytm został stworzony dla problemu modelowania ekstremum aproksymowanej funkcji.

Podejścia zaproponowane przez Wanga i Mendela (Wang, Mendel 1992) zostały rozszerzone i zastosowane w systemie NEFCLASS (Nauck, Kruse 1995, 1997b, 1998b) dla zadania klasyfikacji oraz w systemie NEFPROX (Nauck, Kruse 1997a, 1998c, 1999a) dla zadania aproksymacji funkcji. Wszystkie wartości cech muszą być określone przez zbiory rozmyte zanim reguły zaczną być tworzone. Praktycznie nie tworzy się wszystkich reguł rozmytych i nie przechowuje ich w pamięci. Na starcie przyjmuje się, że zbiór reguł jest zbiorem pustym lub zawiera reguły opisujące wiedzę dziedzinową. Następnie tworzy się kombinacje zbiorów rozmytych i dodaje do zbioru. Kryterium stopu może być przykładowo "utworzyć nie więcej niż L reguł" lub też "reguły powinny opisywać co najmniej p% przykładów uczących".

Indukcja drzew decyzyjnych (Quinlan 1986, 1993) jest bardzo popularnym podejściem stosowanym w analizie danych w celu tworzenia modeli klasyfikujących. Otrzymane drzewa mogą być zapisywane w postaci reguł. Istnieją dobrze znane algorytmy, na przykład ID3 (Quinlan 1986) dla cech symbolicznych, czy też C4.5 (Quinlan 1993) dopuszczający również cechy numeryczne. Porównanie różnych sposobów wyboru cech niosących najwięcej informacji, można znaleźć w pracy (Borgelt, Kruse 1998). Na starcie, każda cecha opisująca przykłady musi mieć określony zbiór wartości poprzez odpowiednie zbiory rozmyte. Metoda jest heurystyczna, nie ma gwarancji, że uzyskane drzewo jest optymalne w sensie struktury lub też uzyskanej dokładności klasyfikacji. Istnieje wiele metod tworzących drzewa rozmyte (Boyen, Wehenkel 1999), (Ichihashi i inni 1996), (Janikow 1996, 1998), (Yuan, Shaw 1995). Ponieważ tworzenie drzew bazuje na heurystyce, nie zawsze wszystkie cechy muszą wystąpić w utworzonym drzewie, ale również mogą powstać drzewa uwzględniające wszystkie cechy, zbyt mocno rozbudowane. Stosowane są wtedy techniki upraszczające (pruning), redukujące rozmiar drzewa.

Metody wymagające zadanej struktury zbiorów rozmytych w przestrzeni cech tworzą zbiory reguł łatwe do interpretacji lingwistycznej, są proste do implementacji, mogą być stosowane dla danych liczbowych i symbolicznych oraz w przypadku występowania brakujących wartości niektórych cech.

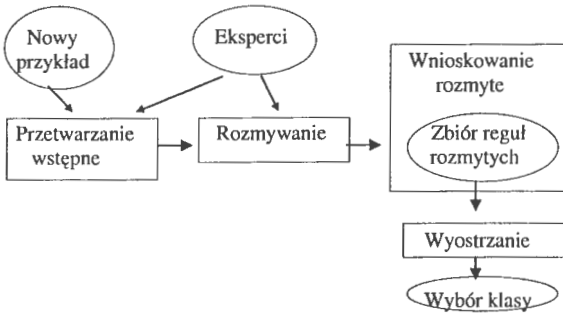
Klasyfikacja nowych przykładów

Dysponując zbiorem reguł rozmytych możemy stosując wnioskowanie rozmyte zastosować je do klasyfikacji nowego przykładu, dla którego nie jest znana przynależność do określonej klasy lub też mogą one być przydatne dla potrzeb analizy rozpatrywanego zagadnienia.

Oznacza to, że znając cechy, które opisują kredytobiorców możemy prognozować, np. ich wiarygodność czy też przewidywany zysk banku.

Wynik wnioskowania jest uzyskiwany w postaci zbioru rozmytego. W sytuacji, gdy wymagana jest tylko jedna klasyfikacja, przeprowadzana jest operacja nazywana w literaturze *wyostrzaniem (defuzzification)*. Polega ona na agregacji w postaci jednego zbioru rozmytych konkluzji reguł opisujących dany przykład i znalezieniu dla tego zbioru rozmytego jednego, najlepiej go charakteryzującego elementu. Wybrany element określa nam przynależność nowego przykładu do jednej, konkretnej klasy. Na Rysunku 16 przedstawiono prosty schemat

postępowania przy próbie klasyfikacji nowego przykładu do jednej z rozpatrywanych klas, z zastosowaniem danego zbioru reguł rozmytych.



Rysunek 16. Klasyfikacja nowych przykładów

Często stosowane metody wyostrażania to metoda maksymalnego stopnia przynależności, metoda średniej maksymalnej MOM (ang. Mean of Maximum) oraz metoda środka obszaru COA (*Center of Area*), (Bartkiewicz 2000).

Opis zastosowanego podejścia

Załóżmy, że dysponujemy skończonym zbiorem przykładów U opisanych za pomocą zmiennych ilościowych, będących wynikiem pewnych pomiarów lub obserwacji oraz zmiennych lingwistycznych, a więc słów lub zdań w języku naturalnym.

Pierwszym krokiem jest *rozmywanie*, czyli zapisanie przykładów w języku zrozumiałym dla systemu komputerowego, zgodnie z metodologią zbiorów rozmytych. Polega ona na określeniu stopni przynależności wartości cech do zbiorów rozmytych określonych przez wybrane wartości lingwistyczne.

Załóżmy, że mamy skończony, niepusty zbiór przykładów $U = \{u_1, u_2, \dots, u_N\}$ i skończony, niepusty zbiór cech $A = \{a_1, \dots, a_K\}$ opisujących te przykłady oraz $V = \bigcup_{a_j \in A} V_{a_j}$, dla $j = 1, \dots, K$,

gdzie $V_{a_j} = \{v_1^{a_j}, v_2^{a_j}, \dots, v_{K_j}^{a_j}\}$ jest niepustym, skończonym zbiorem wartości lingwistycznych cechy $a_j \in A$. W procesie rozmywania dla każdej wartości cechy a_j określamy zbiór K_j liczb $\mu_{v_1^{a_j}}(a_j), \dots, \mu_{v_{K_j}^{a_j}}(a_j)$ z przedziału $[0,1]$ określających, w jakim stopniu cecha a_j przyjmuje każdą z wartości lingwistycznych ze zbioru V_{a_j} .

Relacja rozmyta ze zbioru U do zbioru V_{a_j} , oznaczana \tilde{R}_{a_j} , określana jest na iloczynie kartezjańskim $U \times V_{a_j} \rightarrow [0,1]$ i definiowana w postaci zbioru uporządkowanego

$$\tilde{R}_{a_j} = \{(u, v); \mu_{\tilde{R}_{a_j}}(u, v)\} \quad \forall u \in U, \forall v \in V_{a_j} \quad (110)$$

gdzie $\mu_{\tilde{R}_{a_j}}(u, v) \in [0,1]$ jest funkcją przynależności określającą, w jakim stopniu przykład $u \in U$ jest w relacji z wartością $v \in V_{a_j}$. Relacja ta może być reprezentowana w postaci zbioru par albo w postaci macierzy relacji, której elementami są wartości funkcji przynależności określające, w jakim stopniu przykład $u_i \in U$, $i = 1, \dots, N$ jest w rozmytej relacji z wartością $v^{a_j} \in V_{a_j}$,

$\mu_{\tilde{R}_{a_j}}(u_1, v_1^{a_j})$	$\mu_{\tilde{R}_{a_j}}(u_1, v_2^{a_j})$...	$\mu_{\tilde{R}_{a_j}}(u_1, v_{L_j}^{a_j})$
...			
$\mu_{\tilde{R}_{a_j}}(u_N, v_1^{a_j})$	$\mu_{\tilde{R}_{a_j}}(u_N, v_2^{a_j})$...	$\mu_{\tilde{R}_{a_j}}(u_N, v_{L_j}^{a_j})$

lub też w postaci rozmytego skierowanego grafu, którego wierzchołki stanowią elementy rozważanych zbiorów, krawędzie odpowiadają elementom relacji, to znaczy uporządkowanym parom, a waga każdej skierowanej krawędzi odpowiada wartości odpowiedniej funkcji przynależności.

Definicja 6. Zbiór rozmyty określony przez warunek

Niech U będzie skończonym, niepustym zbiorem przykładów, A niepustym zbiorem cech opisujących te przykłady. Niech cecha $a_j \in A$ ma zbiór wartości $V_{a_j} = \{v_1^{a_j}, v_2^{a_j}, \dots, v_{k_j}^{a_j}\}$. Zbiorem rozmytym w zbiorze U określonym przez wartość $v_i^{a_j} \in V_{a_j}$, oznaczanym $\tilde{U}_{a_j=v_i^{a_j}}$, nazywamy podzbiór relacji rozmytej \tilde{R}_{a_j} ze zbioru U do V_{a_j} , dla ustalonej wartości $v_i^{a_j} \in V_{a_j}$, zapisywany w postaci

$$\tilde{U}_{a_j=v_i^{a_j}} = \{(u, v_i^{a_j}); \mu_{\tilde{R}_{a_j=v_i^{a_j}}}(u, v_i^{a_j}) \mid \forall u \in U, \mu_{\tilde{R}_{a_j=v_i^{a_j}}}(u, v_i^{a_j}) > 0\}.$$

Zbiór rozmyty określony poprzez warunek stanowi uporządkowany zbiór trójek: $(u, v_i^{a_j})$ i wartości funkcji przynależności $\mu_{\tilde{R}_{a_j=v_i^{a_j}}}(u, v_i^{a_j})$ określającej, w jakim stopniu w przykładzie u rozpatrywana cecha a_j przyjmuje wartość $v_i^{a_j}$.

Przykład

Niech U będzie zbiorem kredytobiorców. Cecha „kredyt” ze zbiorem wartości: „mały”, „średni”, „wysoki” jest zmienną lingwistyczną. Zbiór rozmyty w zbiorze U wyznaczony przez warunek („kredyt” = „mały”) zawiera kredytobiorców, dla których cecha „kredyt” przyjmuje wartość „mały” w stopniu większym od zera; można go zapisać w postaci

$$\tilde{U}_{\text{kredyt=mały}} = \{(u, \text{mały}); \mu_{\tilde{R}_{\text{kredyt=mały}}}(u, \text{mały}) \mid u \in U, \mu_{\tilde{R}_{\text{kredyt=mały}}}(u, \text{mały}) > 0\} \quad \square$$

Jeżeli ze zbioru cech A wybierzemy jedną cechę a_d , to możemy ze względu na wartości, które ona przyjmuje dokonać rozmytego podziału zbioru przykładów U . Elementy zbioru $A \setminus \{a_d\}$ nazywamy *cechami warunkowymi*, a cechę a_d nazywamy *cechą decyzyjną*.

Definicja Podział rozmyty zbioru przykładów

Podziałem zbioru przykładów U ze względu na cechę decyzyjną $a_d \in A$ ze zbiorem wartości $V_{a_d} = \{v_1^{a_d}, \dots, v_{k_d}^{a_d}\}$ nazywamy rozmyte zbiory przykładów

$$\tilde{U}_{a_d} = \{\tilde{U}_{a_d=v_1^{a_d}}, \tilde{U}_{a_d=v_2^{a_d}}, \dots, \tilde{U}_{a_d=v_{k_d}^{a_d}}\},$$

gdzie $\tilde{U}_{a_d=v_i^{a_d}} = \{(u, v_i^{a_d}); \mu_{\tilde{R}_{a_d=v_i^{a_d}}}(u, v_i^{a_d}) \mid u \in U, \mu_{\tilde{R}_{a_d=v_i^{a_d}}}(u, v_i^{a_d}) > 0\}$.

Dla uproszczenia zadania uczenia, przyjmuje się w literaturze dwa dodatkowe wymagania, które muszą być spełnione przez zbiór przykładów ze zbioru U , przedstawione poniżej,

$$t(\mu_{\tilde{R}_{n_d \rightarrow n_d}}(u, v_{i_1}^{n_d}), \mu_{\tilde{R}_{n_d \rightarrow n_d}}(u, v_{i_2}^{n_d})) = 0, \quad \text{dla } t_1 \neq t_2, \quad \forall u \in U \quad (111)$$

$$s(\mu_{\tilde{R}_{n_d \rightarrow n_d}}(u, v_{i_1}^{n_d}), \dots, \mu_{\tilde{R}_{n_d \rightarrow n_d}}(u, v_{i_2}^{n_d})) = 1, \quad \text{dla } \forall u \in U. \quad (112)$$

Interpretacja powyższych warunków jest następująca, dla każdego przykładu $u \in U$ i dla każdej cechy $a_j \in A$ mającej zbiór wartości $V_{a_j} = \{v_1^{a_j}, v_2^{a_j}, \dots, v_{K_j}^{a_j}\}$ istnieje dokładnie jedna wartość $v_i^{a_j} \in V_{a_j}$ dla której funkcja przynależności $\mu_{\tilde{R}_{a_j}}(u, v_i^{a_j}) = 1$ oraz dla $\forall v \neq v_i^{a_j}, \mu_{\tilde{R}_{a_j}}(u, v) = 0$.

Jeżeli w zadaniach praktycznych zbiory przykładów nie spełniają warunków (111) i (112), dokonuje się modyfikacji zbioru cech A wybranych do opisu przykładów, wprowadzając nowe cechy nominalne. Tak określony zbiór rozmyty $\tilde{U}_{a_d \rightarrow v_i^{a_d}}$ będzie oznaczał klasę wyznaczoną przez rozmytą partycję, której opisu szukamy w postaci reguł rozmytych typu *JEŻELI ... TO ...*

Uwaga:

W dalszej części pracy dla uproszczenia zapisu będziemy pomijać górne indeksy przy wartościach cechy decyzyjnej i będziemy stosować uproszczony zapis $V_{a_d} = \{v_1, \dots, v_n\}$ oraz $\{\tilde{U}_{v_1}, \tilde{U}_{v_2}, \dots, \tilde{U}_{v_{K_d}}\}$.

Każdy przykład $u_n \in U$, $n = 1, \dots, N$ opisany za pomocą cech $a_j \in A$, $j = 1, \dots, K$ możemy zapisać w następujący sposób

$$u_n : (a_1 = v_{i(1,n)}^{a_1}); \mu_{a_1=v_{i(1,n)}^{a_1}} \wedge (a_2 = v_{i(2,n)}^{a_2}); \mu_{a_2=v_{i(2,n)}^{a_2}} \wedge \dots \wedge (a_K = v_{i(K,n)}^{a_K}); \mu_{a_K=v_{i(K,n)}^{a_K}} \quad (113)$$

gdzie indeks $i(j,n)$ określa wartość jaką przyjmuje j -ta cecha w n -tym przykładzie. Funkcja $\mu_{a_j=v_{i(j,n)}^{a_j}}$ określa, w jakim stopniu w przykładzie u_n cecha a_j przyjmuje wartość $v_{i(j,n)}^{a_j}$. Dla uproszczenia zapisu będziemy stosować wymiennie zapis w postaci

$$u_n : (a_1 = v_{i(1,n)}); \mu_{v_{i(1,n)}} \wedge (a_2 = v_{i(2,n)}); \mu_{v_{i(2,n)}} \wedge \dots \wedge (a_K = v_{i(K,n)}); \mu_{v_{i(K,n)}} \quad (114)$$

Tak zapisany zbiór przykładów jest punktem wyjścia w procesie uczenia maszynowego na podstawie przykładów, w którym otrzymujemy zbiór reguł rozmytych typu *JEŻELI ... TO ...*. Reguła rozmyta może zostać zapisana w postaci

$$\text{JEŻELI } (a_1 = v_{i(1,k)}) \wedge \dots \wedge (a_K = v_{i(K,k)}) \text{ TO } (a_d = v_{i(d,k)}) \quad (115)$$

gdzie indeks $i(j,k)$ określa wartość jaką przyjmuje cecha a_j w k -tej regule, $k=1, \dots, S$.

Dysponując zbiorem k -reguł rozmytych możemy stosując wnioskowanie rozmyte obliczyć stopień spełnienia k -tej reguły przez każdy przykład u_n ze zbioru U .

Niech funkcja $\mu_{i(j,k,n)}$ oznacza funkcję przynależności j -tej cechy w n -tym przykładzie do zbioru rozmytego określonego przez warunek $(a_j = v_{i(j,k,n)})$ z części przesłankowej k -tej reguły. Po pierwsze, dla każdego przykładu $u_n \in U$, $n = 1, \dots, N$ i każdej reguły możemy obliczyć wartości funkcji przynależności do zbiorów rozmytych z części przesłankowej reguły: $\mu_{i(1,k,n)}, \mu_{i(2,k,n)}, \dots, \mu_{i(K,k,n)}$ oraz do zbioru określonego przez konkluzję: $\mu_{i(d,k,n)}$.

Następnie, aby otrzymać stopień spełnienia przez przykład u_n części przesłankowej k -tej reguły stosujemy operację agregacji zbiorów rozmytych z części przesłankowej reguły i otrzymujemy

$$\tau_k(u_n) = A^t(\mu_{l(1,k,n)}, \mu_{l(2,k,n)}, \dots, \mu_{l(K,k,n)}), \quad (116)$$

gdzie A jest wybranym operatorem koniunkcyjnym agregacji, zazwyczaj minimum.

Niech \tilde{C}_k określa klasę rozmytą określoną przez k -tą regułę. Stopień spełnienia k -tej reguły przez przykład u_n wynosi

$$\mu_{\tilde{C}_k}(u_n) = \min\{\tau_k(u_n), \mu_{l(d,k,n)}\}. \quad (117)$$

Agregując dla danego przykładu u_n zbiory rozmyte z konkluzji reguł opisujących ten przykład otrzymujemy wartość funkcji przynależności przykładu do jednej klasy \tilde{C}

$$\mu_{\tilde{C}}(u) = A^s(\mu_{\tilde{C}_1}(u), \mu_{\tilde{C}_2}(u), \dots, \mu_{\tilde{C}_k}(u)), \quad (118)$$

gdzie A jest wybranym operatorem dysjunkcyjnym agregacji, zazwyczaj maksimum.

Przykład 8.

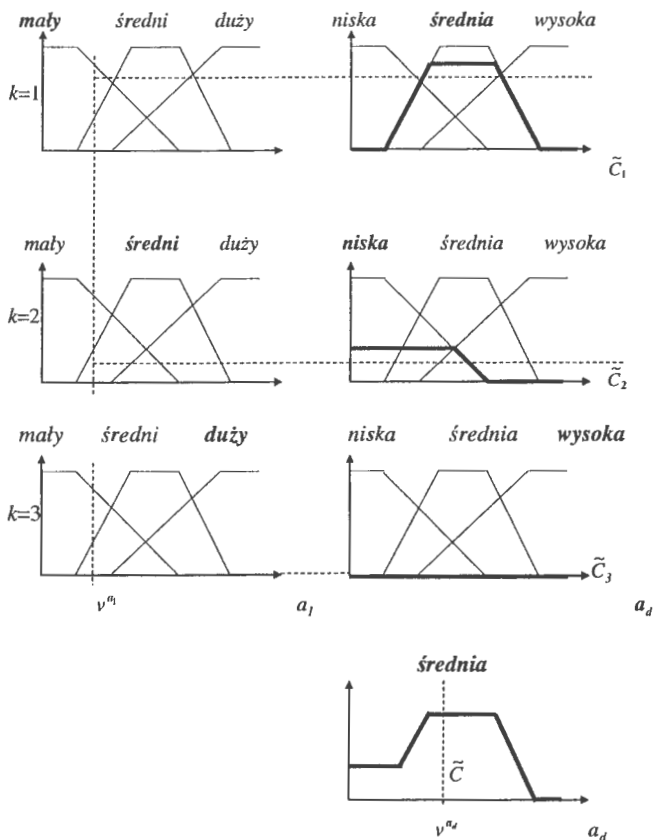
Rozpatrzmy bardzo uproszczony przypadek, gdy do opisu przykładów wybrano cechę a_1 , która może przyjmować wartości lingwistyczne „mały”, „średni”, „duży” oraz cechę a_d z wartościami „niska”, „średnia”, „wysoka”. Przyjmijmy, że w procesie rozmywania, przy opisywaniu danego zbioru przykładów, określono zbiory rozmyte przedstawione na Rysunku 1. Przyjmijmy, że mamy zbiór trzech prostych reguł, $k = 3$, przedstawionych poniżej

$k=1$: JEŻELI ($a_1 = \text{mały}$) TO ($a_d = \text{średnia}$)

$k=2$: JEŻELI ($a_1 = \text{średni}$) TO ($a_d = \text{niska}$)

$k=3$: JEŻELI ($a_1 = \text{duży}$) TO ($a_d = \text{wysoka}$).

Zastosujemy ten zbiór reguł w celu określenia przynależności nowego przykładu u_i : ($a_1 = v^a$), zaznaczonego na Rysunku 1. Klasy rozmyte \tilde{C}_k określone przez rozmyte konkluzje k -tej reguły dla przykładu u_i zostały zaznaczone pogrubioną linią na Rysunku 17 ($\tilde{C}_3 = \emptyset$, ponieważ $\mu_{\tilde{C}_3}(u_i) = 0$). Zbiór rozmyty \tilde{C} jest rozmytą konkluzją uzyskaną z sumy zbiorów rozmytych \tilde{C}_1 , \tilde{C}_2 i \tilde{C}_3 . Jedną wartość najbardziej reprezentatywną dla zbioru rozmytego \tilde{C} , jaką przyjmuje cecha a_d , ozn. v^a , może być określona w procesie wyostrzania.



Rysunek 8. Klasyfikacja przykładu z zastosowaniem zbioru trzech reguł □

Rozpatrywane metody klasyfikacji na podstawie przykładów zakładają, że w zbiorze uczącym nie wyróżnia się kolejności występowania przykładów, natomiast w zagadnieniu predykcji szeregów ta kolejność jest zasadniczą cechą. Istniejące i rozwinięte metody tworzenia reguł na podstawie przykładów ograniczają się tylko do przypadków „stacycznych”. Wydaje się, że wykorzystanie obecnie znanych metod tworzenia reguł mogą być wykorzystane do predykcji szeregów czasowy – jednak po uprzedniej odpowiedniej obróbce danych wejściowych. Np. po rozważaniach opartych na wykorzystaniu twierdzenia Takensa odpowiadającego na pytanie jak duże ma być obserwowane „okno” danych, z oryginalnych przykładów muszą być wygenerowane nowe przykłady odpowiednich „okien” danych. Jednak ten problem jest ciągle problemem otwartym, zwłaszcza, że nie potrafimy odpowiedzieć na pytanie o ważności twierdzenia Takensa, gdy rozważamy elementy logiki rozmytej.

5 Literatura

1. Akaike, H. (1974) A New Look at the Statistical Model Identification. *IEEE Trans. Auto. Control*, 19: 716-723.
2. Bierwag, G. (1987) *Duration Analysis - Managing Interest Rate Risk*. Ballinger Publishing Company, Cambridge, Mass.
3. Bishop, C. M. (1994) Training with Noise is equivalent to Tikhonov Regularization. *Neural Computation*.
4. Bonnländer, B. V., and Weigend A. S. (1994) Selecting Input Variables Using Kernel Density Estimation. In *Proceedings of the International Symposium on Artificial Neural Networks (IS-ANN'94)*, Tainan, Taiwan, R. O. C.
5. Box, G. E. P., and F. M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*, 2nd ed. Oakland, CA: Holden-Day.
6. Brennan M., Schwartz E. (1983): Duration, Bond Pricing and Portfolio Management. W: *Innovations in Bond Portfolio Management*, Bierwag G., Kaufman G., Toevs (Eds.), Greenwich, Conn., JAI Press.
7. Brown, R., P. Bryant, and H. D. I. Abarbanel. 1991. "Computing the Lyapunov Spectrum of a Dynamical System from an Observed Time Series." *Phys. Rev. A* 43: 2787-2806.
8. Chaitin, G. J. 1990. *Information, Randomness & Incompleteness*. Series in Computer Science, Vol. 8, 2nd ed. Singapore: World-Scientific.
9. Cover, T. M., and J. A. Thomas. 1991. *Elements of Information Theory*. New York: John Wiley.
10. Cox J., Ingersoll J., Ross S. (1981): A Re-examination of Traditional Hypotheses about the Term Structure of Interest Rates. *Journal of Finance*, September 1981, pp. 769-799.
11. Dahl H. (1993): A Flexible Approach to Interest-Rate Risk Management. In: Zenios S.A. (Ed.), *Financial Optimization*. Cambridge University Press, Cambridge.
12. Engle, R. F., and C. W. J. Granger. 1988 "Cointegration and Error-Correcting Representation, Estimation and Testing." *Econometrica* 55: 251-276.
13. Farmer, J. D., and J. J. Sidorowich. 1988. "Exploiting Chaos to Predict the Future and Reduce Noise." *Evolution, Learning, and Cognition*, edited by Y. C. Lee. Singapore: World Scientific.
14. Fisher L., Weil R.L. (1977): Coping with the Risk of Market-Rate Fluctuations - Returns to Bondholders from Naive and Optimal Strategies. *Journal of Business*, October 1977, pp. 408-431.
15. Fraser, A. M. 1989. "Information and Entropy in Strange Attractors." *IEEE Trans. Info. Theory* IT-35: 245-262.
16. Friedman, J. H. 1991. "Multivariate Adaptive Regression Splines." *Ann. Stat.* 19: 1-142.
17. Garbade K. (1986): „*Modes of Fluctuations in Bond Yields - an Analysis of Principal Components*“. Technical Report, Bankers Trust Company, Money Market Center, New York, June 1986.
18. Garbade K. (1989): „*Polynomial Representations of the Yield Curve and its Modes of Fluctuations*“. Bankers Trust Company, Money Market Center, New York, No. 53, July 1989.
19. Geman, S., E. Bienenstock, and R. Doursat. 1992. Neural Networks and the Bias/Variance Dilemma." *Neural Computation* 5:1-58.
20. Gershenfeld, N. A. 1989. "An Experimentalist's Introduction to the Observation of Dynamical Systems." In *Directions in Chaos*, edited by B.-L. Hao, Vol. 2, 310-384.
21. Gershenfeld, N. A., and A. S. Weigend. 1994. "The Future of Time Series: Learning and Understanding." In *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld, 1-70. Reading, MA: Addison-Wesley.
22. Gödel, K. 1931. "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme, I." *Monatshefte für Mathematik und Physik* 38:173- 198. An English translation of this paper is found in *On Formally Undecidable Propositions* by K. Gödel (New York: Basic Books, 1962).

23. Granger, C. W. J., and T. Teräsvirta. 1993. *Modelling Nonlinear Economic Relationships*. Oxford, UK: Oxford University Press.
24. Hentschel, H. G. E., and I. Procaccia. 1983. "The Infinite Number of Generalized Dimensions of Fractals and Strange Attractors." *Physica D* 8: 435-444.
25. Hirschleifer J. (1965): Investment decisions under uncertainty - choice theoretic approach. *The Quarterly Journal of Economics*, Vol. 74, pp. 509-536.
26. Hu, M. J. C. 1964. "Application of the Adaline System to Weather Forecasting." E. E. Degree Thesis. Technical Report 6775-1, Stanford Electronic Laboratories, Stanford University.
27. Jajuga K., Jajuga T. (1996): *Inwestycje - instrumenty finansowe, ryzyko finansowe, inżynieria finansowa*. PWN, Warszawa.
28. Jakubowski, A. 1996 *Modelowanie struktury czasowej stóp procentowych*. Projekt badawczy KBN Nr PB 536/HO2/96/10 - G 37, IBS PAN, Warszawa.
29. Kellison S.G. (1991): *The Theory of Interest*. Irwin, Homewood.
30. Kevrekidis, I. G., R. Rico-Martinez, R. E. Ecke, R. M. Farber, and A. S. Lapedes. 1993. "Global Bifurcations in Rayleigh-Benard Convection." Los Alamos preprint, LA-UR-93-2922, May 1993.
31. Kolmogorov, A. 1941. "Interpolation und Extrapolation von stationären zufälligen Folgen." *Bull. Acad. Sci. (Nauk)* 5: 3-14. U.S.S.R., Ser. Math.
32. Kolmogorov, A. N. 1965. "Three Approaches to the Quantitative Definition of Information." *Prob. Infor. Trans.* 1: 4-
33. Bartkiewicz W. (2000) Zbiory rozmyte. W: J.S. Zieliński (red.), *Inteligentne systemy w zarządzaniu. Teoria i praktyka*. PWN, Warszawa, 72-140.
34. Berthold M., Huber K.P. (1999) Constructing Fuzzy Graphs from Examples. W: *J. Intelligent Data Analysis*, 3(1).
35. Beasley J.E., Chu P.C. (1994) *A genetic algorithm for the set covering problem*. Technical Report, The Management School, Imperial College.
36. Beasley J.E. (1996) A genetic algorithm for the set covering problem. *European Journal of Operational Research* 94, 392-404.
37. Bezdek J.C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, N.Y.
38. Bezdek J.C., Keller J., Krishnapuram R. (1998) *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. The Handbooks on Fuzzy Sets, Kluwer Academic Publishers, Norwell, MA.
39. Bilgic T., Turksen L. (1997) Elicitation of membership functions: How far can they take us? W: *Proc. FUZZ/IEEE*, 1321-1325, Barcelona, Spain.
40. Bilgic T., Turksen L. (1999) *International Handbook of Fuzzy Systems*, I, Kluwer Academic Publishers.
41. Boyen X., Wehenkel L. (1999) Automatic Induction Of Fuzzy Decision Trees and its Application to Power System Security Assessment. *Fuzzy Sets and Systems*, 102(1), 3-19.
42. Chvatal V. (1979) A greedy heuristic for the set-covering problem. *Math. of Oper. Res.* 4 (3) 233-235.
43. Croall I.F., Mason J.P. (eds.) (1991), *Industrial Applications of Neural Networks*. Springer-Verlag, Berlin, 1991.
44. Cun le, Y. 1989. "Generalization and Network Design Strategies." In *Connectionism in Perspective*, edited by R. Pfeifer, Z. Schreter, F. Fogelman, and L. Steels. Amsterdam: North Holland.
45. Dubois D., Prade H. (1997) The three semantics of fuzzy sets. *Fuzzy Sets and Systems*, 90, 1 41-150.
46. Dubois D., Prade H. (1978) Operations on fuzzy numbers. *Int. J. Syst. Sci.* Vol.9, 613-626.
47. Dubois D., Prade H. (1980) *Fuzzy Sets and Systems : Theory and Applications*. Academic Press, NY.

48. Garey M.R., Johnson D.S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H.Freeman and Co., San Francisco.
49. Garfinkel R. S., Nemhauser G.L. (1978) *Integer programming*. John Wiley & Sons, New York-London-Sydney-Toronto.
50. Goldberg D. (1998) *Algorytmy genetyczne i ich zastosowania*. WNT Warszawa.
51. Grauel A., Klene G., Ludwig L. (1997) Data Analysis by Fuzzy Clustering Methods. W: A.Grauel, W.Becker, F.Belli (eds.): *Fuzzy-Neuro-Systeme'97 – Computational Intelligence. Proc. 4th Int. Workshop Fuzzy-Neuro-Systeme'97* in Soest, Germany, 563-572.
52. Grefenstette John J.(1990) *User's guide to GENESIS*.
53. Grossman T., Wool A. (1996) *Computational experience with approximation algorithms for the set covering problem*. Working paper, Theoretical Division and CNLS, Los Alamos National Laboratory.
54. Higgins C., Goodman R. (1993) *Learning Fuzzy Rule-based Neural Networks for Control*. Advances in Neural Information Processing Systems, 5, 350-35
55. Hołubiec J., Szkatuła G., Wagner D. (2002) Modelowanie preferencji wyborców w postaci reguł decyzyjnych. Modelowanie Preferencji a Ryzyko '01. Akademia Ekonomiczna w Katowicach.
56. Hołubiec J., Szkatuła G., Wagner D. (2002). Analiza obietnic wyborczych ugrupowań politycznych. In: Bubnicki Z., Hryniewicz O., Kulikowski R. (eds.): *Metody i techniki analizy informacji i wspomaganie decyzji. Badania operacyjne i systemowe wobec wyzwań XXI wieku*. Akademia Oficyna Wydawnicza EXIT, Warszawa, III. 63-74.
57. Hoppner F., Klawonn F., Kruse R., Runkler T. (1999) *Fuzzy Cluster Analysis*. Wiley, Chichester.
58. Ichihashi H., Shirai T., Nagasaka K., Miyoshi T. (1996) Neuro-fuzzy ID3: A Method of Inducing Fuzzy Decision Trees with Linear Programming for Maximizing Entropy and an Algebraic Method for Incremental Learning. *Fuzzy Sets and Systems*, 81(1), 157-16
59. Janikow C.Z. (1996) Exemplar based Learning in Fuzzy Decision Trees. W: *Proc. IEEE Int. Conf. On Fuzzy Systems 1996*, 1500-1505, New Orleans.
60. Janikow C.Z. (1998) Fuzzy Decision Trees: Issues and Methods. *IEEE Trans. Systems, Man & Cybernetics*, Part B: Cybernetics, 28(1), 1-14.
61. Johnson D.A. (1974) Approximation algorithms for combinatorial problems. *J. Computer System Sci.* 9, 256-278.
62. Kacprzyk J., Szkatuła G. (1996) An algorithm for learning from erroneous and incorrigible examples, *Int. J. of Intelligent Syst.* 11, 565 - 582.
63. Kacprzyk J., Szkatuła G (1997) An improved inductive learning algorithm with a preanalysis od data", In: Z.W. Raś and A. Skowron (eds.): *Foundations of Intelligent Systems* (Proceedings of 10th ISMIS'97 Symposium, Charlotte, USA), Springer-Verlag, Berlin, 157 - 166.
64. Kacprzyk J., Szkatuła G. (1998) *IPI - An Improved Inductive Learning Procedure with a Pre-processing of Data*. Proceedings of IDEAL'98 (Hong Kong), Springer-Verlag.
65. Kacprzyk J., Szkatuła G. (1999) An inductive learning algorithm with a preanalysis od data. *International Journal of Knowledge - Based Intelligent Engineering Systems*, vol. 3, 135-146.
66. Kacprzyk J., Szkatuła G. (2002) An integer programming approach to inductive learning using genetic and greedy algorithms. W: L.C. Jain and J.Kacprzyk (eds.) *New learning paradigms in soft computing*. Studies in Fuzziness and Soft Computing. Physica-Verlag Heidelberg, 323 - 36
67. Kandel A. (1986) *Fuzzy Mathematical Techniques with Applications*. Addison-Wesley, Reading, MA.
68. Keller J.M., Givens J.A. (1985) Membership function issues in fuzzy pattern recognition. W: *International Conference on Cybernetics and Society*, 210-214.
69. Keller A., Klawonn F. (1998) Generating Classification Rules by Grid Clustering. W: *Proc. Third European Workshop on Fuzzy Decision Analysis and Neural Networks for Management, Planning and Optimization (EFDAN'98)*, Dortmund, 113-121.

70. Klawonn F., Keller A. (1997) Fuzzy Clustering and Fuzzy Rules. W: M.Mares, R.Mesiar, V.Novak, J.Ramik, A.Stupnanova (eds.) *Proc. Seventh International Fuzzy Systems Association World Congress IFSA'97*, 1, Prague, 193-19
71. Klawonn F., Kruse R. (1995) Clustering Methods in Fuzzy Control. W: W.Gaul, D.Pfeifer (eds.): *From Data to Knowledge: Theoretical and Practical Aspects of Classification, Data Analysis and Knowledge Organization*, Springer-Verlag, Berlin, 195-202.
72. Klir G.J., Folger T.A. (1988) *Fuzzy Sets, Uncertainty and Information*. Prentice-Hall International, Inc., Englewood Cliffs, N.J.
73. Kohonen T. (1984) *Self-Organization and Associative Memory*. Springer-Verlag, Berlin.
74. Kosko B. (1992) *Neural Networks and Fuzzy Systems. A Dynamical Systems Approach to Machine Intelligence*. Prentice Hall, Englewood Cliffs, NJ.
75. Krawczak, M. (1997) Neural Networks Application to Economic Time Series. *Proc. Int. Conference on Transition to Advanced Market Institutions and Economies*. Polish Academy of Sciences, Warsaw, May 199
76. Krawczak, M. (1999a) Dynamic Programming and Fuzzy Reinforcement of Backpropagation for Interest Rate Prediction. In: R. A. Ribeiro, H.-J. Zimmerman, R. R. Yager, J. Kacprzyk, eds., *Soft Computing in Financial Engineering*. Springer-Verlag, 142-155.
77. Krawczak, M. (1999b) Observability of Time Series Neural Networks Prediction. *Proc. ANZIIS'99, ANNES'99, EXPO on Future Directions for Intelligent Systems and Information Sciences*, Dunedin/ Queenstown, New Zealand, November 1999.
78. Krawczak, M. (2000a) Backpropagation versus Dynamic Programming Approach. *Bulletin of Polish Academy of Sciences*, 48, 2, 167-180.
79. Krawczak, M. (2000b) Elements of Chaos Theory to Time Series Prediction. *Proc. Advanced Computer Systems*. Tech. Univ. of Szczecin/Albert-Ludwigs Univ. of Freiburg/Univ. of Technology of Troyes. Szczecin, October 2000.
80. Krawczak, M. (2000c) Theoretical Foundation of Neural Networks Prediction. In: *Computer Technology in Management*, J. Studziński, L. Drelichowski, O. Hryniewicz, J. Kacprzyk, eds., Polish Academy of Sciences, Systems Research 26.
81. Krawczak, M. (2001a) Financial Management. In: *Computer Technology in Management*, J. Studziński, L. Drelichowski, O. Hryniewicz, eds., Polish Academy of Sciences, Systems Research 28. (in Polish).
82. Krawczak, M. (2001b) Selected Aspects of Time Series Modelling. In: M. Krawczak, A. Miklewski, *Information Society in Sustainable Development*. Agricultural Academy of Szczecin/Warsaw Univ. of Information Technology. (in Polish).
83. Krawczak, M., (2002a) Heuristic Dynamic Programming for Neural Networks Learning, Part 1: Learning as a Control Problem. *Proc. of VI IEEE International Conference on Neural Networks and Soft Computing*, Springer Verlag.
84. Krawczak, M., (2002b) Heuristic Dynamic Programming for Neural Networks Learning, Part 2: I-order Differential Dynamic Programming. *Proc. of VI IEEE International Conference on Neural Networks and Soft Computing*, Springer Verlag.
85. Krawczak, M. (2002c) Adjoint Multilayer Neural Networks. In: *Automatic Control and Management*, J. Gutenbaum (ed.). Polish Academy of Sciences, Systems Research, Special Issue, 153-166.
86. Krawczak, M., Miklewski, A. (1997a) Financial Time Series - Analysis and Prediction. *Proc. Int. Conference on Transition to Advanced Market Institutions and Economies*, Polish Academy of Sciences, Warsaw, May 199
87. Krawczak, M., Miklewski, A. (1997b) Analysis of Time Series - Statistical Approach to Neural Networks. *Proc. 12th Int. Workshop on Statistical Modelling*. Biel/Bienne, Switzerland, July 199
88. Krawczak, M., Miklewski, A. (2002) On Investment Risk: Value at Risk Method. *Proc. Computer Systems Conference*, Sept., Ciechocinek, Poland, 243-252. (in Polish).

89. Krawczak, M., Miklewski, A., Jakubowski, A., Konieczny, P. (2000) *Investment Risk Management*, (in Polish). Polish Academy of Sciences, Systems Research 25.
90. Krawczak M., A. Miklewski (2002) Koncepcja ryzyka inwestycyjnego: wariacyjno-kowariacyjna metoda Value at Risk, W: *Komputerowe Wspomaganie Zarządzania i Procesów Decyzyjnych w Gospodarce*, Studziński J., Drelichowski L., Hryniewicz O. (red.), Polska Akademia Nauk, Instytut Badań Systemowych, Seria: Badania Systemowe, Tom 31, Warszawa 2002, str. 243-252, ISBN 83-85847-73-1.
91. Krawczak, M., Mizukami, K. (1994) The Control Theory Approach to Perceptron Learning Process. *IEE of Japan*, Okayama.
92. Kruse R., Schwecke E., Heinsohn J. (1991) *Uncertainty and Vagueness in Knowledge-Based Systems: Numerical Methods*. Springer-Verlag, Berlin.
93. Kunchewa L.I. (1996) A fuzzy aggregation of multiple classification decisions. *Control and Cybernetics*, 25, 337-352.
94. Kunchewa L.I. (2000) *Fuzzy Classifier Design*. Physica-Verlag.
95. Lang, K. J., A. H. Waibel, and G. E. Hinton. 1990. "A Time-Delay Neural Network Architecture for Isolated Word Recognition." *Neural Networks* 3: 23-43.
96. Lapedes, A., and R. Farber. 198 "Nonlinear Signal Processing Using Neural Networks." Technical Report No. LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, NM.
97. Litterman R., Scheinkman J. (1991): Common Factors Affecting Bond Returns. *Journal of Fixed Income*, June 1991, pp. 54-61.
98. Lorenz, E. N. 1963. "Deterministic Non-Periodic Flow." *J. Atmos. Sci.* 20: 130- 141.
99. Lovasz L. (1975) On ratio of optimal integral and fractional covers. *Disc. Math.* 13, 383-390.
100. Macaulay F.R. (1938): *Some Theoretical Problems Suggested by the Movement of Interest Rates, Bond Yields and Stock Prices in the United States since 1856*. Columbia University Press, National Bureau of Economic Research, New York.
101. Mackey M., Glass L. (1977): Oscillations and chaos in physiological control systems. *Science*, 197,28
102. Medasani S., Kim J., Krishnapuram R. (1998) An overview of membership function generation techniques for pattern recognition. *International Journal of Approximate Reasoning*, 19, 391-41
103. Michalewicz Z. (1996) *Algorytmy genetyczne + struktury danych = programy ewolucyjne*. WNT Warszawa.
104. Miklewski A., Krawczak M. (red. nauk.) (1996) *Ryzyko i niepewność w modelach ekonomiczno-ekologicznych*, ISBN 83-85847-76-6.
105. Miller, W. T., R. S. Sutton, and P. J. Werbos, eds. 1990. *Neural Networks for Control*. Cambridge, MA: MIT Press.
106. Morrison, F. 1996. *Sztuka modelowania układów dynamicznych*. PWN. Warszawa.
107. Mozer, M. C. 1994. "Neural Net Architectures for Temporal Sequence Processing." In *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld, 243-264. Reading, MA: Addison-Wesley.
108. Nauck D., Kruse R. (1995) NEFCLASS – A Neuro-Fuzzy Approach for the Classification of Data. W: K.M.George, J.H.Carrol, E.Deaton, D.Poenheim, J.Hightower (eds.): *Applied Computing 1995. Proc. ACM Symposium on Applied Computing, Nashville, Feb. 26-28*, ACM Press, New York, 461-465.
109. Nauck D., Kruse R. (1997a) Function Approximation by NEFPROX. W: *Proc. Second European Workshop on Fuzzy Decision Analysis and Neural Net-works for Management, Planning and Optimization (EFDAN'97)*, Dortmund, 160-169.
110. Nauck D., Kruse R. (1997b) A Neuro-Fuzzy Method to Learn Fuzzy Classification Rules from Data. *Fuzzy Sets and Systems*, 89, 277-288.
111. Nauck D., Kruse R. (1998a) How the Learning of Rule Weights Affects the Interpretability of Fuzzy Systems. W: *Proc. IEEE Int. Conf. On Fuzzy Systems 1998*, Anchorage, 1235-1240.

112. Nauck D., Kruse R. (1998b) NEFCLASS-X A Soft Computing Tool to Build Readable Fuzzy Classifiers, *BT Technology Journal*, 16(3), 180-190.
113. Nauck D., Kruse R. (1998c) A Neuro-Fuzzy Approach to Obtain Interpretable Fuzzy Systems for Function Approximation. W: *Proc. IEEE Int. Conf. On Fuzzy Systems 1998*, Anchorage, 1106-1111.
114. Nauck D., Kruse R. (1999) Neuro-Fuzzy System for Function Approximation. *Fuzzy Sets and Systems*, 101, 261-271.
115. Nychka, D., S. Ellner, D. McCaffrey, and A. R. Gallant. 1992. "Finding Chaos in Noisy Systems." *J. Roy. Stat. Soc. B* 54(2): 399-426.
116. Oppenheim, A. V., and R. W. Schaffer. 1989. *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice Hall.
117. Pal S.K., Chakraborty B. (1984) Intra-class and inter-class ambiguities (fuzziness) in feature evaluation. *Pattern Recognition Letters*, 2, 275-279.
118. Parui S.K., Majumder D.D. (1982) Estimation of fuzzy membership functions. W: *Proc. 6th Int. Conf. On Pattern Recognition*, Munich, Germany, 423-426.
119. Pawlak Z. (1991). *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht.
120. Pearl J. (1988) *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*. Morgan Kaufmann, San Francisco.
121. Pedrycz W., Card H.C. (1992) Linguistic Interpretation of Self-Organizing Maps. W: *Proc. IEEE Int. Conf. On Fuzzy Systems 1992*, San Diego, CA, 371-378.
122. Peitgen, H.-O., Jurgens, H., i Saupe, D. 1996. Granice chaosu - fraktale, cz. 1, cz. 2. PWN. Warszawa.
123. Peters E.E. (1997): *Teoria Chaosu a Rynki Kapitałowe*. WIG Press, Warszawa.
124. Peters E.E. (1994): *Fractal Market Analysis*. Wiley, New York.
125. Powell M., Skolnick M.M. (1993) *Using genetic algorithms in engineering design optimization with non-linear constraints*, W: Forrest S. (red.): *Proceedings of the Fifth International Conference on Genetic Algorithms*. Morgan Kaufmann, San Mateo, CA, s. 424 - 430.
126. Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1992. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge: Cambridge University Press.
127. Radeva, V., Krawczak, M., Choy, E. (2002) Review and Bibliography on Generalized Nets Theory and Applications. *Advanced Studies in Contemporary Mathematics*, 4, 2, 173-199.
128. Redington F.M. (1952): Review of the Principle of Life Office Valuations. *Journal of The Institute of Actuaries*, 1952, vol. 78, pp. 286-340.
129. Ruelle, D., and J. P. Eckmann. 1985. "Ergodic Theory of Chaos and Strange Attractors." *Rev. Mod. Phys.* 57: 617-656.
130. Quinlan J. (1986) Induction of Decision Trees. *Machine Learning*, 1, 81-106.
131. Quinlan J. (1993) *C4.5: Programs for Machine Learning*. Morgan Kaufman, San Mateo, CA.
132. Saaty T.L. (1980) *The analytic hierarchy process*. McGraw-Hill, New York.
133. Sauer, T., J. A. Yorke, and M. Casdagli. 1991. "Embedology." *J. Stat. Phys.* 65(3/4): 579-616.
134. Shaw, R. S. 1981. "Strange Attractors, Chaotic Behavior and Information Flow." *Z. Naturforsch.* 36A: 80-112.
135. Sierpińska M., Jachna T. (1994): *Ocena przedsiębiorstwa według standardów światowych*. PWN, Warszawa, Wyd. 2.
136. Schuster, H.G. 1995. *Chaos deterministyczny*. PWN. Warszawa.
137. Sugeno M., Yasukawa T. (1993) A fuzzy-Logic-Based Approach to Qualitative Modeling. *IEEE Trans. Fuzzy Systems*, 1, 7-31.
138. Szkatuła G. (1995) *Uczenie maszynowe na podstawie przykładów w przypadku błędów w danych*. Rozprawa doktorska, IBS PAN Warszawa.

139. Szkatuła G. (2002). *Zastosowanie zmodyfikowanego zadania pokrycia w uczeniu maszynowym*. W: Gutenbaum J. (eds.): *Automatyka Sterowanie Zarządzanie*. SRI PAS, Warszawa, 431-445.
140. Szkatuła G, Hołubiec J., Wagner D. (2000) Forecasting voting behaviour using machine learning-Poland in transition. *Annals of Operation Research* 97, 31-41.
141. Takens, F. 1981. "Detecting Strange Attractors in Turbulence." In *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young. Lecture Notes in Mathematics, Vol. 898, 336-381. Warwick, 1980. Berlin: Springer-Verlag.
142. Tong, H. (1990) *Non-linear Time Series - A Dynamical System Approach*. Oxford Univ. Press.
143. Tong, H., and K. S. Lim. 1980. "Threshold Autoregression, Limit Cycles and Cyclical Data." *J. Roy. Stat. Soc. B* 42: 245-292.
144. Traczyk, W. (red) 1995. *Problemy sztucznej inteligencji*. Wiedza i Życie. Warszawa.
145. Tschichold Gurman (1996) *RuleNet – A New Knowledge-Based Artificial Neural Network Model with Application Examples in Robotics*. PhD Thesis, ETH Zurich.
146. Ulam S. M., Neumann von J. (1947) *Bulletin Amer. Math. Soc.* 53: 1120.
147. Yager R.R., Filev D.P. (1994) *Essentials of Fuzzy Modeling and Control*, John Wiley & Sons, N.Y.
148. Vassilev, P., Krawczak, M. (2001) Generalized Net Investment Model. *Proc. of First Workshop on Intuitionistic Fuzzy Sets and Generalized Nets*. Warsaw, Poland, November 2001, 39-45.
149. Wan, E. A. 1994. "Times Series Prediction Using a Connectionist Network with Internal Delay Lines." In *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld, 195- 21 Reading, MA: Addison-Wesley.
150. Wang L.X., Mendel J.M. (1991) Generation Rules by Learning from Examples. W: *International Symposium on Intelligent Control*, IEEE Press, 263-268.
151. Wang L.X., Mendel J.M. (1992) Generating fuzzy rules by learning from examples. *IEEE Trans. Syst. Man., Cybern.* 22(6), 1414-142
152. Weigend, A. S., B. A. Huberman, and D. E. Rumelhart. 1990. Predicting the Future: A Connectionist Approach." *International Journal of Neural Systems* 1:193-209.
153. Weigend, A., B. A. Huberman, and D. E. Rumelhart. 1992. "Predicting Sunspots and Exchange Rates with Connectionist Networks." In *Nonlinear Modeling and Forecasting*, edited by M. Casdagli and S. Eubank, 395-432. Redwood City, CA: Addison-Wesley.
154. Weigend, A. S., and B. LeBaron. 1994. "Evaluating Neural Network Predictors by Bootstrapping." In *Proceedings of International Conference on Neural Information Processing (ICONIP'94)*, Seoul, Korea. Also Technical Report CU-CS-725-94, Computer Science Department, University of Colorado at Boulder.
155. Weron A., Veron R. 1998. Inżynieria finansowa. WNT. Warszawa.
156. Wolff, A., Swift, J. B., Swinney, H. L. (1985) Determining Lyapunov exponents from a time series. *Physica*, 16D, 285.
157. Yule G. (1927) On a Method of Investigating Periodicity in Disturbed Series with Special Reference to Wolfer's Sunspot Numbers. *Phil. Trans. Roy. Soc. London A* 226: 267-298.
158. Yuan Y., Shaw M.J. (1995) Induction of Fuzzy Decision Trees. *Fuzzy Sets and Systems*, 69(2).
159. Zadeh L.A. (1965) Fuzzy Sets. *Information and Control*. 8, 338-353.
160. Zadeh L.A. (1973) Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions of Systems, Man and Cybernetics*, 3 (1), 228-44.
161. Zadeh L.A. (1975) Fuzzy logic and approximate reasoning. *Synthese*, 30, 407-428.
162. Zadeh L.A. (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1), 3-28.
163. Zadeh L.A. (1981) *Possibility theory and soft data analysis*. Westview Press, Boulder, Co.
164. Zimmermann H.J. (1987) Fuzzy sets in pattern recognition. W: P.A.Devijver, J.Kittler (eds.) *Pattern Recognition Theory and Applications*, F30 of NATO ASI, 383-391, Springer-Verlag, Berlin.

