

104/2001

Raport Badawczy

RB/27/2001

Research Report

**Wybrane aspekty uczenia maszynowego
na podstawie przykładów**

Część I

Maciej Krawczak, Kazimierz Mańczak,

Grażyna 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

Pracę zgłosili: prof. dr hab. inż. Kazimierz Mańczak,

Warszawa 2001

INSTYTUT BADAŃ SYSTEMOWYCH PAN

PRACOWNIA SYSTEMÓW WSPOMAGANIA DECYZJI

PSWD/27/2001

Kazimierz MAŃCZAK, Maciej KRAWCZAK, Grażyna SZKATUŁA

Wybrane aspekty uczenia maszynowego na podstawie przykładów. Część I.

Zadanie badawcze:

Komputerowe systemy wspomaganie decyzji -

kierownik: prof.dr hab.inż. Kazimierz MAŃCZAK

Podzadanie:

Próba systematyki metod uczenia maszynowego

Wykonawcy: prof. dr hab. inż. Kazimierz MAŃCZAK - kierownik

dr inż. Maciej KRAWCZAK

dr inż. Grażyna SZKATUŁA

WARSZAWA , grudzień 2001

Wybrane aspekty uczenia maszynowego na podstawie przykładów. Część I.

Spis treści:

Wstęp

1. Odwrotne sieci neuronowe

- 1.1. Elementy grafów skierowanych
- 1.2. Konstruowanie sieci odwrotnych
 - 1.2.1. Pojęcie sieci odwrotnych
 - 1.2.2. Sieci neuronowe a sieci odwrotne
 - 1.2.3. Gradientowe algorytmy uczenia sieci
 - 1.2.4. Tworzenie reguł
- 1.3. Wyprowadzenie algorytmu propagacji wstecznej
- 1.4. Podsumowanie

Literatura

2. Wybrane elementy teorii zbiorów rozmytych w uczeniu maszynowym na podstawie przykładów

- 2.1. Wprowadzenie
- 2.2. Podstawowe pojęcia teorii zbiorów rozmytych
 - 2.2.1. Zbiór rozmyty
 - 2.2.2. Rodzaje funkcji przynależności
 - 2.2.3. Właściwości zbiorów rozmytych
 - 2.2.4. Operacje jednoargumentowe na zbiorach rozmytych
 - 2.2.5. Działania na zbiorach rozmytych
 - 2.2.6. Agregacja zbiorów rozmytych
 - 2.2.7. Relacje rozmyte
 - 2.2.8. Wnioskowanie rozmyte
 - 2.2.9. Rodzaje reguł rozmytych
- 2.3. Tworzenie reguł rozmytych na podstawie zbioru przykładów
 - 2.3.1. Wprowadzenie
 - 2.3.2. Postawienie zadania uczenia maszynowego

Literatura

Wstęp

Zadanie uczenia maszynowego na podstawie przykładów możemy ogólnie postawić w sposób następujący. Mamy dany zbiór przykładów, zwany zbiorem uczącym. Przykłady opisujemy za pomocą warunków związanych z cechami. Zbiór taki ze względu na wartości wybranej cechy (tzw. cechy decyzyjnej) można podzielić na klasy.

Zbiór przykładów uczących jest punktem wyjścia do procesu uczenia się, w wyniku którego pozyskuje się wiedzę o klasyfikowaniu, tzn. szuka opisu zbioru przykładów należących do danej klasy. Opis taki wyrażany jest na ogół w postaci *reguł decyzyjnych*, tzn. zbioru warunków związanych z wartościami cech, zachowującymi zależności pomiędzy cechami oraz uwzględniającymi znaczenie poszczególnych cech; *drzew decyzyjnych*; lub też w postaci odpowiednio dobranych *wag połączeń w sieciach neuronowych* i ich struktury.

W rozdziale 1 omówiono zagadnienia sztucznych sieci neuronowych ze szczególnym uwzględnieniem procesu uczenia sieci na przykładach. Przy założeniu, że funkcje aktywacji, jak również wskaźnik jakości uczenia w sieciach neuronowych są funkcjami różniczkowalnymi, można stosować gradientowe algorytmy optymalizacji. Dotychczas stosowane metody uczenia sieci wymagają stosowania wielokrotnego różniczkowania funkcji złożonych.

W opracowaniu proponuje się wykorzystanie aparatu teorii grafów skierowanych, w szczególności twierdzenia Tellegena, do przekształcenia sieci neuronowej na graf. Następnie określa się tzw. sieci odwrotne, które topologicznie są takie same jak rozpatrywana sieć neuronowa i określa przepływ sygnałów w sieciach odwrotnych. Zaproponowano pięć reguł przekształcających sieć oryginalną na sieć odwrotną. Rozpatrywane podejście, które w prosty sposób generuje wzory na algorytmie wstecznej propagacji błędów uczenia, może być rozszerzone na algorytmy uczenia innych klas sieci neuronowych.

W rozdziale 2 rozpatrywano zadanie tworzenia reguł klasyfikacji na podstawie rzeczywistych zbiorów danych które mogą zawierać sprzeczności i niejednoznaczności, będące istotnym utrudnieniem w procesie uczenia. Podjęto próbę zastosowania podejścia opartego na teorii zbiorów rozmytych, które pozwala zapisywać przybliżone, nieprecyzyjne i niejednoznaczne dane w języku zrozumiałym dla systemu komputerowego i dalej je przetwarzać. Opisy rozpatrywanych klas otrzymuje się w postaci zbioru reguł rozmytych typu JEŻELI ...TO, z rozmytymi przesłankami i rozmytymi konkluzjami.

W opracowaniu przedstawiono wybrane elementy teorii zbiorów rozmytych oraz podstawy wnioskowania rozmytego pod kątem możliwości ich zastosowania przy tworzeniu reguł klasyfikacji. Podjęto próbę postawienia zadania tworzenia reguł rozmytych na podstawie zbioru przykładów uczących.

Rozdział 1

Odwrotne wielowarstwowe sieci neuronowe

1.1 Introduction

Learning of multilayer neural networks is usually performed by the backpropagation algorithm, often with some modifications. The algorithm is a gradient descent one and has a long history. The basic algorithm version is well known due to Rumelhart and McClelland (1986), and it was derived by applying repeatedly chain rule expansions backward through the network. There are several researchers who claim a right to this algorithm, here we would like to mention works done by Werbos (1974) who obtained similar results applying ordered partial derivatives, and by Dreyfus (1992) as well as by Kelley (1960) who obtained similar results relating the gradient theory of flight paths during the late 1950s. Considering a learning process of multilayer neural networks as a particular multistage optimal control problem Krawczak (1994, 1999a, 1999b, 2000a) obtained backpropagation-like equations. This approach is based on a dynamic programming idea, and a first order approximation of return functions consideration. In results Krawczak elucidated the backpropagation algorithm as a first order differential dynamic programming. Applied similar methodology Krawczak (2001a, 2002) shown how to include an additional parameter in neuron activation functions and how to find an optimal value of this parameter. Another works of Krawczak (1999b, 1999c, 2000b) were devoted to ways of changing these parameters by handling the problem by a homotopy (or continuation) method.

In the very rich artificial neural networks literature little attention has been given to neural networks consideration from the graph theory point of view. Examination of neural networks as *flow graphs* gives very interesting and unlike properties of the neural networks learning process. The approach is based on the Tellegen's theorem (Tellegen, 1952; Chua and Lin, 1975) used in the electric circuits while Kirchhoff's laws may not be valid. The consideration leads to proper equations of the backpropagation algorithm but in a much more simple way. In the graph methodology there are *reciprocal graphs* in which signals flow in opposite direction. Here we use terminology adopted from the optimization or optimal control theory (Bryson and Ho, 1969), and such neural networks in which signals flow in opposite direction are called the *adjoint neural networks*.

In this chapter we present a flow graph methodology for representing neural networks, and then the adjoint neural networks for representing learning algorithms based on gradient methods. The

backpropagation algorithm is considered in details, while the applied notation allows also treating multilayer networks with possible feedback signals. For the backpropagation algorithm we emphasize the forward propagation of the inputs through the neural networks and the backward propagation of proper gradients through the adjoint neural networks. Construction of the adjoint neural networks gives directly formulae of the considered learning algorithm.

It seems that this approach can be extended on other gradient algorithm used in different kind of neural network architecture, like *global descent backpropagation*, *time-delay neural networks*, *backpropagation through time*, *recurrent backpropagation*, *time dependent recurrent backpropagation* which learning algorithms are based on gradient methods.

1.2 Neural network transformation into a flow graph

Since the pioneer work of McCulloch and Pitts (1943) a model of an artificial neuron is a very simple processing unit, Figure 1.1,

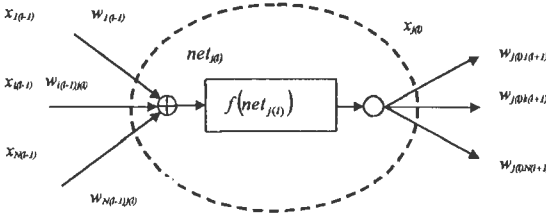


Figure 1.1: An elementary model of a neuron.

which has a number of inputs, say N , each input is weighted with an appropriate weight w_{ij} , $i = 1, 2, \dots, N$. The sum of the weighted inputs and the bias (included to the inputs) forms the input

$$net_j = \sum_{i=1}^N w_{ij} x_i = \sum_{i=1}^N y_{ij}, \quad (1.1)$$

to the activation function $f_j(net_j)$ within a summation point \oplus . Here we consider differentiable activation functions to generate outputs of the neurons. In the considered model there is included an additional element depicted by \circ describing a junction point. Generally the existence of the junction point in a neuron has been tacitly assumed. In Figure 1.1 there is some extended notation of indices, namely we indicate a position of each neuron in the whole network. For example the weight $w_{i(l-1)j(l)}$ indicates a connection between a neuron i laying in the $(l-1)$ -th layer and a neuron j lying in the (l) -th layer.

In this way three main elements of a neuron can be distinguished, one part is a summation point, the second is an activation function which transmits the effect of summation, and the third is a junction point spreading the activation function output to neurons of the next layer with the same value.

Let us rearrange the neuron elements in the following way:

- removing the activation function outside the neuron
- removed activation functions are shifted to each connection between the considered neuron and all neurons of the next layer becoming transmittances between neurons
- connection between neurons are still weighted
- a summation point and a junction point set up a *node*
- a neural network with rearranged neurons becomes a *flow graph*.

The above introduced neuron rearrangement is pictured in Figure 1.2.

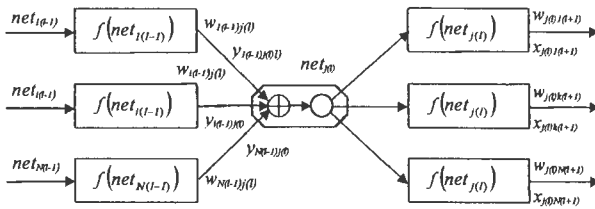


Figure 1.2: The rearranged neuron with its vicinity as a flow graph part.

Now Equation (1.1) can be rewritten in the following way

$$net_{j(i)} = \sum_{i=1}^N w_{i(i-1)j(i)} x_{i(i-1)} = \sum_{i=1}^N w_{i(i-1)j(i)} f_{i(i-1)}(net_{i(i-1)}) = \sum_{i=1}^N y_{i(i-1)j(i)} \quad (1.2)$$

and description of any separate edge becomes

$$y_{i(i-1)j(i)} = w_{i(i-1)j(i)} x_{i(i-1)} = w_{i(i-1)j(i)} f_{i(i-1)}(net_{i(i-1)}) \quad (1.3)$$

Now let us show an example of a simple neural network with one hidden layer, Figure 1.3.

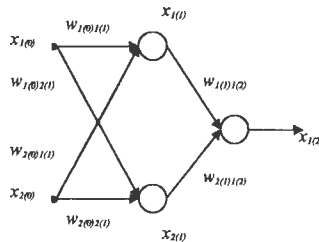


Figure 1.3: A simple two-layer neural network.

After using the neurons rearranging procedure described above the same neural network can be considered as a flow graph, Figure 1.4.

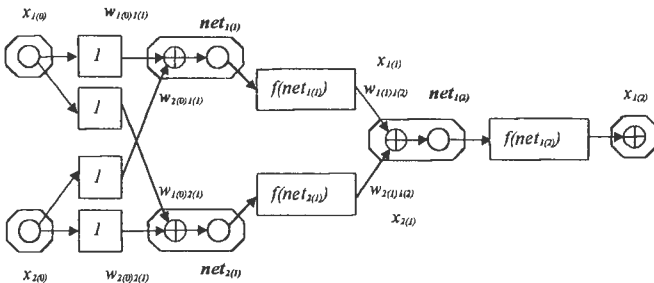


Figure 1.4: A flow graph corresponding to the two-layer neural network from Figure 1.3.

Comparing these two pictures we can notice that the architecture of the neural network and the flow graph is exactly the same, in the position of each neuron a graph node has been appeared, while to each connection an activation function has been added as a transmittance. In the above picture there are two kinds of reduced nodes:

- input nodes are limited only to junction points
- output nodes are limited to summation points.

1.3 Elements of the flow graph

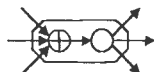
In the previous section it has been shown how a neural network can be transformed into a flow graph. In this section our attention will be devoted to the basic definitions of reciprocal graphs and it will be proved when two graphs, the considered and the reciprocal are interreciprocal.

A graph consists of two types of elements, nodes and edges, and the ways in which these elements are interconnected.

Definition 1.1:

A graph is represented by the following triple $G = (V, E, \varphi)$, where V is a nonempty set of nodes of the graph, E is a set of edges of the graph, and φ is a mapping from the set E to V . If the pairs of nodes are connected by an ordered edge (indicated by an arrow), then the edge is directed. A graph with all directed edges is called a directed graph or flow graph. Any two connected nodes by an edge are called adjacent nodes. Here nodes are represented by a small octagon.

Assumption 1.1: All nodes consist with two, one a summation point and second a junction point (see the following picture).



Let us considered a flow graph G with N nodes indicated by $i, i = 1, 2, \dots, N$. A single edge (i, j) between two nodes i and $j, (i, j = 1, 2, \dots, N)$; and there is defined a value X_i representing the node; and next for each pair of nodes there is defined a transmittance function T_{ij} ; and there is defined an output of each edge (the final value of an edge coming into a node. The above introduced rules can be shown in the following figure:

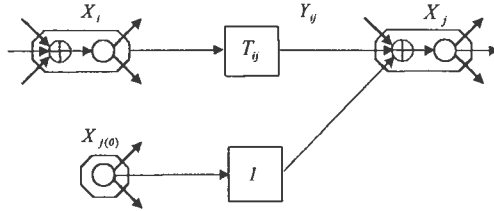


Figure 1.5: A directed edge of a flow graph.

with the following transition function

$$Y_{ij} = T_{ij} X_i. \quad (1.4)$$

With the Assumption 1.1 of summing properties of each node we have a relation of a node:

$$X_j = \sum_{i=1}^N Z_{ij} + X_{j(0)} = \sum_{i=1}^N T_{ij} X_i + X_{j(0)} \quad (1.5)$$

where by $X_{j(0)}$ we denote some selected nodes considered as external inputs to the graph directly feeding the node j , in this case an associated transmittance is equal to 1.0 .

Having one graph G let us consider another graph G' with the same number of nodes, $i = 1, 2, \dots, N$, and the same edge connection (j, k) between any two nodes i and $j, i, j = 1, 2, \dots, N$, it means of the same topology. The graph G is defined by X_i, T_{ij}, Y_{ij} and edges (i, j) for $i, j = 1, 2, \dots, N$; while the graph G' is defined by X'_i, T'_{ij}, Y'_{ij} and edges (i, j) for $i, j = 1, 2, \dots, N$.

Now we will recall the following definitions and the theorem.

Definition 1.2:

It is said that two graphs G and G' are transposed of each other if and only if their transmittances are transposed

$$T'_{ij} = T_{ji} \quad \text{for all } i, j = 1, 2, \dots, N. \quad (1.6)$$

Definition 1.3:

It is said that two graphs G and G' are interreciprocal if and only if their

$$\sum_{j=1}^I (X'_j X_{j(o)} - X_j X'_{j(o)}) = 0. \quad (1.7)$$

Theorem 1.1:

Two given transposed flow graphs G and G' are interreciprocal.

Proof:

Let us consider a flow graph G and another flow graph G' that is the transposed graph of G , i.e.

$$T'_{ij} = T_{ji} \quad \text{for all } i, j = 1, 2, \dots, N.$$

Now we will convert the following identity

$$\sum_{j=1}^N X'_j X_j = \sum_{j=1}^N X_j X'_j. \quad (1.8)$$

For both considered graphs using the equation (1.5) we get

$$X_j = \sum_{i=1}^N T_{ij} X_i + X_{j(o)} = \sum_{i=1}^N Y_{ij} + X_{j(o)}$$

$$X'_j = \sum_{i=1}^I T'_{ij} X'_i + X'_{j(o)} = \sum_{i=1}^I Y'_{ij} + X'_{j(o)}$$

and substituting the proper terms in left- and right-hand side in (1.8), respectively, and let us perform the following conversion

$$\begin{aligned} \sum_{j=1}^N X'_j \left(\sum_{i=1}^N T_{ij} X_i + X_{j(o)} \right) &= \sum_{j=1}^N X_j \left(\sum_{i=1}^N T'_{ij} X'_i + X'_{j(o)} \right) \\ \sum_{j=1}^N X'_j \left(\sum_{i=1}^N Y_{ij} + X_{j(o)} \right) &= \sum_{j=1}^N X_j \left(\sum_{i=1}^N Y'_{ij} + X'_{j(o)} \right) \\ \sum_{j=1}^N \sum_{i=1}^N (X'_j Y_{ij} + X'_j X_{j(o)}) &= \sum_{j=1}^N \sum_{i=1}^N (X_j Y'_{ij} + X_j X'_{j(o)}) \\ \sum_{j=1}^N \sum_{i=1}^N (X'_j Y_{ij} - X_j Y'_{ij}) - \sum_{j=1}^N \sum_{i=1}^N (X'_j X_{j(o)} - X_j X'_{j(o)}) &= 0. \end{aligned} \quad (1.9)$$

It was assumed that the graph G and G' are transposed, so this assumption implies that first term of (1.9) is equal to zero due to the following condition

$$\begin{aligned} \sum_{j=1}^N \sum_{i=1}^N (X'_j Y_{ij} - X_j Y'_{ij}) &= \\ \sum_{j=1}^N \sum_{i=1}^N (X'_j T_{ij} X_i - X_j T'_{ij} X'_i) &= \\ \sum_{j=1}^N \sum_{i=1}^N (X'_j T_{ij} X_i - X_j T_{ji} X'_i) &= 0 \end{aligned} \quad (1.10)$$

From (1.7) we have that the second term in (1.9) is also equal to zero

$$\sum_{i=1}^N \sum_{j=1}^N (X'_i X_{j(o)} - X_j X'_{j(o)}) = \sum_{j=1}^N (X'_j X_{j(o)} - X_j X'_{j(o)}) = 0 \quad (1.11)$$

and it proves that the graphs G and G' are interreciprocal. \square

Here we would like to mention that the equation (1.9) is known in the electric circuits theory as Tellegen's theorem (Tellegen, 1952; Chua, Lin, 1975).

It seems that one case is of special importance should be distinguished, namely a graph with only one input as well as one output. That case limits Equation (1.11) to two terms without crossing dependencies. In such a simple case we can distinguish a node *in* which is the input to the graph and a node *out* which is the output of the graph, while for the graph G' a node *in* is the output and a node *out* is the input of the graph, then Equation (1.11) becomes

$$X_{in} X'_{in} = X'_{out} X_{out}. \quad (1.12)$$

For the graph G the output X_{out} as a result of the input X_{in} is identical to the output X'_{in} following the input X'_{out} for the graph G' , thus for a single input and a single output the conclusion can be written as

$$X_{in} / X_{out} = X'_{out} / X'_{in}, \quad X'_{in} = X_{out}, \quad X'_{out} = X_{in}.$$

Similar interdependence between inputs and output of interreciprocal graphs is valid for multidimensional inputs and outputs, but the one input and one output case illustrates the idea of graph interreciprocity in a very visual way.

1.4 Adjoint neural networks construction

In the previous section the possibilities of conversion of a neural network into a flow graph have been shown, and next some properties of interreciprocal flow graphs.

In this section these properties will allow us to introduce *adjoint neural networks*. The name adjoint has been borrowed from the optimization as well as optimal control theory where it is used for additional (adjoint) set of equations, and in the case of the optimal control problems or multistage problems these equations are solved in an opposite direction.

1.4.1 Adjoint networks

Using the graph theory notation a feedforward network topology can be specified by the following set of equations

$$x_j = \sum_{i=1}^N T_{ij} w_{ij} x_i, \quad i, j = 1, 2, \dots, N \quad (1.13)$$

where N is the number of all nodes (in our case neurons), x_i, x_j are values describing nodes i and j , $T_{ij} = f_{ij}(\)$ is a transmittance (or an activation function of neurons e.g., sigmoid function between the

node i and j . The summation is taken over all signals associated with the node x_i , $i = 1, 2, \dots, N$ coming into the node x_j , $j = 1, 2, \dots, N$.

The last equation can be rewritten in a different way by considering values of nodes, i.e. $net_{j(l)}$, for $l = 1, 2, \dots, L$, $j(l) = 1, 2, \dots, N(l)$

$$net_{j(l)} = \sum_{i(l-1)=1}^{N(l-1)} w_{i(l-1)j(l)} x_{i(l-1)} = \begin{cases} \sum_{i(0)=1}^{N(0)} w_{i(0)j(l)} x_{i(0)} \\ \sum_{i(l-1)=1}^{N(l-1)} w_{i(l-1)j(l)} f(net_{i(l-1)}) \end{cases} \quad (1.14)$$

$$net_{j(l)} = \begin{cases} \sum_{i(0)=1}^{N(0)} w_{i(0)j(l)} net_{i(0)} \\ \sum_{i(l-1)=1}^{N(l-1)} w_{i(l-1)j(l)} f(net_{i(l-1)}) \\ net_{j(out)} = \sum_{i(L)=1}^{N(L)} w_{i(L-1)j(L)} f(net_{i(L-1)}) \end{cases} \quad (1.14a)$$

Using (1.14) we can illustrate flows of signals in a neural network treated as a flow graph in the following picture. It is shown an example of twolayer neural network with one hidden layer, in the picture there is marked a chain of directed edges linking a selected input node $net_{i(0)} = net_{i(in)}$ with a selected output node $net_{i(out)}$.

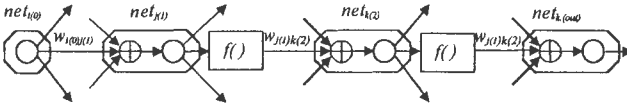


Figure 1.5: A schematic exemplary twolayer neural network as a flow graph.

Now let us remain the learning process of neural networks. Learning of the neural networks consists of changing the weights when the desired output d_p , $p = 1, 2, \dots, P$, and the actual output $x_{(L)p}$, as a result of the input $x_{(0)p}$, are different. The index p is a training example while L denotes the number of layers in the network. The changing is done by gradient descent

$$\Delta w_{i(l-1)j(l)} = w_{i(l-1)j(l)}^{new} - w_{i(l-1)j(l)}^{old} = -\eta \frac{\partial E}{\partial w_{i(l-1)j(l)}} \quad (1.15)$$

where η is the learning rate, and E is the learning performance. Generally the learning performance E is defined as a sum of all training pattern examples

$$E = \sum_{p=1}^P \sum_{j(L)=1}^{N(L)} E_{j(L)p} (d_{j(L)p}, x_{j(L)p}) \quad (1.16)$$

where $N(L)$ is a number of output nodes (neurons). For a specific training pattern p we use the squared difference between the patterns and actual network output

$$E_p = \frac{1}{2} \sum_{j(L)=1}^{N(L)} (d_{j(L)p} - x_{j(L)p})^2. \quad (1.17)$$

Using the delta rule directly, derived in Chapter 3, for updating the weights $w_{i(l-1)j(l)}$ we obtain

$$\Delta w_{i(l-1)j(l)} = -\eta \frac{\partial E_p}{\partial w_{i(l-1)j(l)}} = -\eta \frac{\partial E_p}{\partial net_{j(l)}} \frac{\partial net_{j(l)}}{\partial w_{i(l-1)j(l)}} = \eta \delta_{j(l)} x_{i(l-1)} \quad (1.18)$$

where $j(l) = 1, 2, \dots, N(l)$, $i(l-1) = 1, 2, \dots, N(l-1)$, $l = 1, 2, \dots, L$.

Let us recall the expression of δ for (1.17)

$$\delta_{i(l-1)} = \begin{cases} \sum_{j(L)=1}^{N(L)} (d_{j(L)p} - x_{j(L)p}) f'(net_{j(L)}) \\ f'(net_{i(l-1)}) \sum_{j(l)=1}^{N(l)} w_{i(l-1)j(l)} \delta_{j(l)} \end{cases} \quad (1.19)$$

$$\delta_{i(l-1)} = \begin{cases} \delta_{j(out)} = (d_{j(L)p} - x_{j(L)p}) = e_j \\ \sum_{j(L)=1}^{N(L)} f'(net_{j(L)}) \delta_{j(out)} \\ f'(net_{i(l-1)}) \sum_{j(l)=1}^{N(l)} w_{i(l-1)j(l)} \delta_{j(l)} \end{cases}. \quad (1.19a)$$

It can be easily noticed that Equation (1.14) and (1.19) have the same structure. In (1.14) the signals net flow from the inputs through the network to the outputs, while in (19) the signals δ flow in opposite direction from the outputs to the inputs. In (1.19) there is included influence of a performance index of learning.

According to the definition of interreciprocity of flow graphs it is required to define the input to the adjoint graph. The term $\delta_{j(out)} = -(d_{j(L)p} - x_{j(L)p}) = e_j$ is dependent of a shape of the performance index, and can be treated as the input to the adjoint network. In the original networks there are distinguished inputs nodes and the response of the network is just $\delta_{j(out)} = e_j$. In order to get some similarity we must consider inputs nodes of the original network as output nodes of the adjoint network, and vice versa output nodes of the original network as input nodes of the adjoint network equal to e_j . Now we must define transposition of transmittances due to the definition of transposed graphs, i.e.

$$T'_{ij} = T_{ji}, \text{ for all } i, j = 1, 2, \dots, N. \quad (1.20)$$

For a normal signals flow direction we can write the following relationship

$$net_{j(l)} = T_{i(l-1)j(l)} w_{i(l-1)j(l)} net_{i(l-1)} = f(net_i) w_{i(l-1)j(l)} net_{i(l-1)} \quad (1.21)$$

while for an opposite signals flow direction we can write

$$\delta_{i(l-1)} = S_{j(l)i(l-1)} w_{i(l-1)j(l)} \delta_{j(l)} \quad (1.22)$$

where $S_{j(l)i(l-1)}$ is a transmittance for delta signals. Let us differentiate (1.21) and rewrite (1.22) using the definition of the delta rule we obtain

$$\frac{\partial net_{j(l)}}{\partial net_{i(l-1)}} = \frac{\partial f(net_i)}{\partial net_i} w_{i(l-1)j(l)} \quad (1.23)$$

$$\frac{\partial E}{\partial net_{i(l-1)}} = S_{j(l)i(l-1)} w_{i(l-1)j(l)} \frac{\partial E}{\partial net_{j(l)}}, \quad (1.24)$$

$$\frac{\partial net_{j(l)}}{\partial net_{i(l-1)}} = S_{j(l)i(l-1)} w_{i(l-1)j(l)} \frac{\partial E}{\partial E},$$

and substituting (1.23) to (1.24) we get

$$S_{j(l)i(l-1)} = T'_{j(l)i(l-1)} = \frac{\partial f(net_i)}{\partial net_i} = f'(net_i). \quad (1.25)$$

In this way we proved the form of transposed transmittances appearing in the adjoint neural networks. In Figure 1.5 it was shown an exemplary path of a twolayer neural network, a counterpart to this example - a path of the adjoint neural network is shown in Figure 1.6.

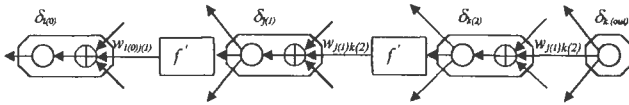


Figure 1.6: A schematic exemplary twolayer neural network as a flow graph.

Looking at Equations (1.14) and (1.19) we can notice that both networks are topologically identical, it means that there is strict and unique correspondence between signals and connections. The adjoint network is found by application of the same network architecture, reversing the direction of signal flows, replacing activation functions by their derivatives, and by changing the positions of summing points with junction points within each node.

It is easy to notice that the changes of the original network into the adjoint network are governed by very simple rules that will be described in the next section.

In Equations (1.11) and (1.12) the terms $f'_{ij}(x_i) = \delta_{ij}$ and $J'_i = \delta_i$ are the transmittances in the graph G' , and there are just the well known *deltas* in the backpropagation algorithm.

In the transposed graph signals flow in reverse direction than in the original graph G and in this the transposed graph becomes the *adjoint network*. The adjoint network is characterized by the inputs (which are outputs of the original network) and the outputs (which are inputs to the original network), see (1.11) and (1.12).

1.4.2 Neural networks versus networks

In some works of Krawczak e.g. (2000a) multilayer neural networks were divided into two kinds of layers. The first kind consisted only of weights with summation points while the second only of activation functions. The idea was just the same like in this work, the simple weight layers are related to nodes and the activation function layers are related to edges with transmittances.

Considering architecture of any class of neural networks, e.g. feedforward, we can observe four basic elements building networks:

- summing points
- junction points
- univariate functions
- multivariate functions (considered in Krawczak, 2000a).

Any architecture of neural networks can be represented as a flow graph by the following changes:

- a neuron is divided into three parts: a summation point, a junction point and an activation function,
- a summation function together with a junction point becomes a node of a flow graph (a network),
- an activation function becomes a transmittance of each outgoing edge,
- gathering signals in a node nascent by simple transition of a previous *net* value and a transmittance multiplied by a synaptic weight.

In the previous section we have shown the idea of changing a neural network into a network, and next deriving a joint network. Adding the goal of determining the error gradients (deltas) we need only apply a set of simple transformations. Construction of an adjoint network requires reversing the flow direction in the original network, labeling all resulting signals δ_j , and performing the following operations:

1. A summation point and junction point is gathered as a node



2. Univariate functions are replaced with their derivatives.

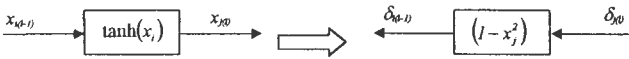


Explicitly, the scalar continuous function $x_j = f(w_{ij}, x_i)$ is replaced by $\delta_i = f'(w_{ij}, x_i)\delta_j$, where $f'(w_{ij}, x_i) = \partial x_j / \partial x_i$. Special cases are

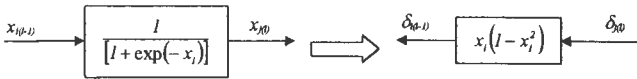
- weights: $x_j = w_{ij} x_i$, in which case $\delta_i = w_{ij} \delta_j$



- bipolar activation function: for the function $y_j = \tanh(x_j)$ the derivative is $f'(x_j) = (1 - x_j^2)$



- unipolar activation function: for the function $y_j = \frac{1}{1 + \exp(-x_j)}$ the derivative is $f'(x_j) = x_j(1 - x_j)$

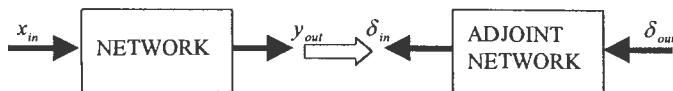


3. Multivariate functions are replaced with their Jacobians.



A multivariate function maps a vector of input signals into a vector of output signals, $y_{out} = F(x_{in})$. In the transformed network, we have $\delta_{in} = \partial F(x_{in}) / \partial x_{in} \delta_{out}$, where the Jacobian $\partial F(x_{in}) / \partial x_{in}$ corresponds to a matrix of partial derivatives.

4. Outputs become inputs.



By reversing the signal flow, output nodes y_{out} in the considered network become input nodes in the adjoint network. These inputs are then set e . For different cost functions than considered here the squared error, the input should be set to $\partial J / \partial y_{out}$.

The considered five rules allow us simple construction of the adjoint network from the original networks. Note that there is a topological equivalence between the two networks. The order of computations in the adjoint network is thus identical to the order of computations in the original network. The signals δ_j that propagate through the adjoint network correspond to the terms $\partial J / \partial y_j$ necessary for gradient adaptation. The exact equations we can take from the adjoint network, completing the derivation. A formal proof of the validity and generality of this method was given in section 1.1.

1.5 Backpropagation algorithm derivation

The simplicity of the application of adjoint neural networks in deriving the error backpropagation algorithm will be clarified in this section.

We derive the standard backpropagation algorithm. For consistency with traditional notation, we have labeled the summation signal $net_{i(l)}$ with the enlarged subscript to denote the layer, see Figure 1.5. The adjoint network shown in Figure 1.6 is found by applying the construction rules described above, from this figure, we may immediately write down the equations for calculating the delta terms, Equation (1.14):

$$\delta_{i(l-1)} = \begin{cases} \delta_{j(out)} = (d_{j(L)p} - x_{j(L)p}) = e_j \\ \sum_{j(L)=1}^{N(L)} f'(net_{j(L)}) \delta_{j(out)} \\ f'(net_{i(l-1)}) \sum_{j(l)=1}^{N(l)} w_{i(l-1)j(l)} \delta_{j(l)} \end{cases}$$

Equation (1.18) for the weight update now can be formulated as

$$\Delta w_{i(l-1)j(l)} = \eta \delta_{j(l)} x_{i(l-1)}$$

These two equations (1.14) and (1.18) precisely describe standard backpropagation, the well known equations from many text book about feedforward neural networks. There is no doubt that the presented approach gives much more easily way to obtain the equations describing the backpropagation algorithm.

1.6 Summary

In this chapter we have delivered a way the derivation of gradient-based algorithms for any network architectures. For any interesting neural network we can construct the adjoint neural network. The

adjoint is build using five rules of changing the original network can be obtained in a very simple way. All formulae for learning algorithms can be derived using this methodology.

References

1. Bertsekas, D. P. and Tsitsiklis J. N. (1996). *Neuro-dynamic programming*. Athena Scientific, Belmont, Massachusetts.
2. Bryson, A. E. and Ho, Y-C. (1969). *Applied Optimal Control*". Blaisdel, Waltham, MA.
3. Campolucci, P., Marchegiani A., Uncini A., and Piazza F. (1997). Signal-Flow-Graph Derivation of on-line gradient learning algorithms. *Proc. ICNN-97*, Houston, June 1997, pp. 1884-1889.
4. Chua L. O. and Lin, P-M. (1975): *Computer-Aided Analysis of Electric Circuits*, Prentice-Hall, New Jersey.
5. Cybenko, G. (1989). *Approximation by superpositions of a sigmoidal function*, *Mathematics of Control Signals, and Systems*, 2, pp. 303-311.
6. Dreyfus, S.E. (1990). Artificial Neural Networks, Back Propagation, and the Kelley-Bryson Gradient Procedure. *Journal of Guidance*, vol. 13, no. 5, pp. 926-928.
7. Fettweiss, A. (1972). A general theorem for signal-flow networks, with applications. In *Digital Signal Processing*, IEEE Press, pp. 126-130.
8. Hassoun, M. H. (1995). *Fundamentals of Artificial Neural Networks* (MIT press).
9. Hecht-Nielsen, R. (1992). *Theory of the backpropagation network*, in: *Neural networks for perception*, ed.: H. Wechsler, Academic Press, New York.
10. Hornik, K., Stinchcombe, M. and White H. (1989). *Multilayer feedforward networks are universal approximators*, *Neural Networks*, 2, pp. 359-366.
11. Kailath, T. 1980. *Linear Systems*. Prentice-Hall, Englewood Cliffs, NJ.
12. Krawczak, M. and Mizukami, K. (1994). The control theory approach to perceptron learning process. *44 Conference of IEE of Japan*, Okayama.
13. Krawczak, M. (1987). *Neural networks learning as a multiobjective optimal control problem*. *Mathware and Soft Computing*, 4.
14. Krawczak, M. (1999a). Dynamic learning for feedforward neural networks. In: *Proceedings of the conference: ICONIP'99 ANZIIS'99&ANNES'99 &ACNN'99*, Perth, Western Australia, 16-20 November, pp. 33-36.
15. Krawczak, M. (1999b). Backpropagation versus dynamic programming approach for neural networks learning. In: *Emerging Knowledge Engineering and Connectionist-Based Information Systems*, Dunedin, New Zealand, 22-23 November 1999, Univ. of Otago, pp. 1057-1062.
16. Krawczak, M. (2000a). Backpropagation versus dynamic programming approach. *Bulletin of Polish Academy of Sciences*. Vol. 48, No. 2, pp. 167-180.
17. Krawczak, M. (2000b). *Neural Networks Learning by Homotopy Method*. ICONIP 2000, Taejin, Korea.
18. Krawczak, M. (2000c). Zakopane
19. Krawczak, M. (2001a). Międzyzdroje
20. Krawczak, M. (2001b). *Parameterized multilayer Neural Networks*. Proc. XIV International Conference on Systems Sciences. Wrocław
21. Krawczak, M. (2002a). *Adjoint Multilayers Neural Networks*. *Control and Cybernetics*

22. Krawczak, M. (2002b). Parameterized Multilayer Neural Networks. *Systems Science* (accepted for publication).
23. Lapedes, A., and R. Farber. (1987). Nonlinear Signal Processing Using Neural Networks. Technical Report No. LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, NM.
24. Oppenheim, A. V., and R. W. Schaffer (1989). *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice Hall.
25. Penfield, P., Spence, R., and Duiker, S. (1970). *Tellegen's Theorem and Electrical Networks*, MIT Press, Cambridge, MA.
26. Rall, B. (1981). *Automatic Differentiation: Techniques and Applications*, Lecture Notes in Computer Science, Springer-Verlag.
27. Rumelhart, D. E., Hilton, G. E. and Williams, R. J. (1986). Parallel Distributed Processing. vol. 1, edited by D. Rumelhart, J. McClelland, and the PDP Research Group, MIT Press, Cambridge, MA, Chap. 8.
28. Wan, E., and Beaufays, F. (1996). Diagrammatic derivation of gradient algorithms for neural networks. *Neural Computation*, vol. 8, no. 1, January 1996, pp. 182-201.
29. Werbos, P. J. (1974). *Maximizing Long-Term Gas Industry Profits in Two Minutes in Lotus Using Neural Network Methods*. IEEE Trans. on Systems, Man, and Cybernetics, 1989, Vol. 19, No. 2, pp. 315-333.
30. Zurada, J. M. (1992). *Introduction to artificial neural systems*, West Publishing Company, St. Paul.

Rozdział 2

Wybrane elementy teorii zbiorów rozmytych w uczeniu maszynowym na podstawie przykładów

2.1. Wprowadzenie

W teorii mnogości klasyczny zbiór jest kolekcją pewnych obiektów, które nazywa się elementami tego zbioru. Relacja \in należenia elementu do zbioru jest pojęciem pierwotnym. Dowolny element należy do zbioru albo należy do jego dopełnienia. Przykładowo, temperatura może być ujemna albo nieujemna, liczba należy do zbioru liczb naturalnych albo nie należy.

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 [L.A.Zadeh, 1973], który zawiera pojęcia nieprecyzyjne i wieloznaczne. Przykładowo, pojęcia "zimno" i "ciepło" mogą być różnie odbierane przez różnych ludzi, ten sam człowiek może być oceniony jako "niski" przez Polaka, natomiast jako "bardzo wysoki" przez Pigmeja. Literatura dotycząca zbiorów rozmytych jest bardzo liczna, semantykę zbiorów rozmytych opisano między innymi w pracy [Dubois i Prade, 1997].

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 określane 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.

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.

Rozpatrzmy zbiory rozmyte "ciepło" i "gorąco" w zbiorze temperatur określonych w $^{\circ}C$. Powietrze o temperaturze $28^{\circ}C$ może być odbierane zarówno jako ciepłe jak też jako gorące; może więc należeć zarówno do zbioru rozmytego "gorąco" jak też do zbioru rozmytego "ciepło" w różnym stopniu.

Żalóżmy, że dysponujemy zbiorem przykładów (np. pacjentów), które możemy opisać za pomocą zmiennych ilościowych, będących wynikiem pewnych pomiarów (np. temperatura, ciśnienie) lub obserwacji (np. obrzęki, sinica skóry, chrapliwy oddech) oraz słów lub zdań w

języku naturalnym (np. osoba nerwowa, nadwrażliwa, odczuwająca silny lęk). Takie przykłady mogą zawierać dane nieprecyzyjne, niepełne, wieloznaczne i mogą być obarczone błędami.

Aby zapisać takie przybliżone i niejednoznaczne dane w języku zrozumiałym dla systemu komputerowego i móc dalej je przetwarzać, można zastosować teorię zbiorów rozmytych, określając rozmyte wartości cech wybranych do opisu przykładów (np. dla cechy "ból" można określać wartości rozmyte: *niewielki, silny* lub też dla cechy "ciśnienie" wartości: *nieco podwyższone, wysokie*) oraz rozmyty podział przykładów na klasy (np. pacjentów chorych na gripę, pacjentów z lekką nerwicą lub też mających chorobę wieńcową).

Zbiór przykładów uczących opisany 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 reguły rozmytej zapisanej w języku naturalnym przedstawiono poniżej.

JEŻELI "występują przedłużone napady bólów dławicowych" i "brak jest gorączki" i
"występuje podwyższenie OB" i "brak jest elektrokardiograficznego obrazu
zawału serca"

TO "pacjent ma ostrą niewydolność wieńcową".

Reguły rozmyte mogą być definiowane przez eksperta, mogą być wynikiem modelowania jakościowego, mogą również zostać wygenerowane automatycznie na podstawie danego zbioru przykładów z zastosowaniem algorytmów automatycznego pozyskiwania wiedzy.

Dysponując zbiorem reguł rozmytych możemy stosując wnioskowanie rozmyte zastosować je do klasyfikacji nowych przykładów, spełniających koniunkcje warunków z rozmytej części przesłankowej reguły, do odpowiedniej klasy, określonej poprzez rozmyte konkluzje.

W rozdz.2.2 przedstawiono wybrane elementy teorii zbiorów rozmytych oraz podstawy wnioskowania rozmytego, a w rozdz.2.3 krótko omówiono zadanie tworzenia reguł rozmytych na podstawie zbioru przykładów uczących z zastosowaniem metod uczenia maszynowego.

2.2. Podstawowe pojęcia teorii zbiorów rozmytych

2.2.1. 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, 1981], [Kandel, 1982, 1986], [Klir i Folger, 1988]. Tak więc 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 $\tilde{}$.

Definicja 2.1. Zbiór rozmyty [Zadeh, 1965]

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, $Card(U) = N$, wtedy zbiór rozmyty jest określany 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

$$\tilde{G} = \{(u_1; \mu_{\tilde{G}}(u_1)), \dots, (u_N; \mu_{\tilde{G}}(u_N))\} \quad (2.1)$$

bywa zapisywany w postaci sumy [Zadeh 1965], [Kandel 1986], [Negoița 1981]

$$\tilde{G} = u_1 / \mu_{\tilde{G}}(u_1) + \dots + u_N / \mu_{\tilde{G}}(u_N) = \sum_{i=1}^N \{u_i / \mu_{\tilde{G}}(u_i)\} \quad (2.2)$$

lub też w postaci zbioru

$$\tilde{G} = \{u_1 / \mu_{\tilde{G}}(u_1), \dots, u_N / \mu_{\tilde{G}}(u_N)\}. \quad (2.3)$$

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

$$\tilde{G} = \int_U \mu_{\tilde{G}}(u) / u. \quad (2.4)$$

Niech \tilde{G}, \tilde{H} będą zbiorami rozmytymi w zbiorze U . 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 "około 5", "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 R nazywamy zbiór rozmyty \tilde{G} , który ma ciągłą funkcję przynależności $\mu_{\tilde{G}}: R \rightarrow [0, 1]$, istnieje liczba rzeczywista $u \in R$ taka, że $\mu_{\tilde{G}}(u) = 1$ oraz dla dowolnych $u, v, w \in R$ takich, że $u < v < w$, zachodzi warunek $\mu_{\tilde{G}}(v) \geq \min\{\mu_{\tilde{G}}(u), \mu_{\tilde{G}}(w)\}$.

2.2.2. Rodzaje funkcji przynależności

Istnieją dwa główne sposoby określania funkcji przynależności, przedstawione poniżej.

- **Określanie wartości funkcji przynależności dla każdego elementu zbioru.** Przykładowo, rozpatrując zbiór studentów U i zbiór rozmyty "dobry student", musimy określić dla każdego studenta ze zbioru U wartość funkcji przynależności do zbioru "dobry student".

Przykład 2.1. Rozpatrzmy zbiór liczb $U = \{1, 2, 3, 4, 5, 6\}$. Zbiór rozmyty "duża liczba" w zbiorze U możemy określić następująco: "duża liczba" = $\{(1;0.0), (2;0.0), (3;0.1), (4;0.4), (5;0.8), (6;1.0)\}$. Liczba 6 na pewno należy do zbioru rozmytego "duża liczba", liczby 1 i 2 na pewno nie należą, a liczby 3, 4 i 5 należą do tego zbioru w pewnym stopniu.

□

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

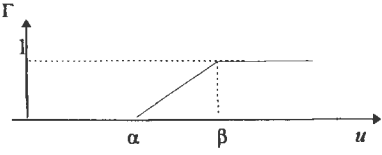
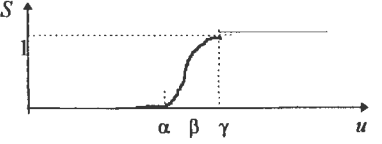
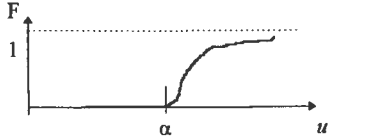
Przykład 2.2. Niech U będzie zbiorem liczb naturalnych z przedziału $[1,100]$ określających wiek ludzi w latach. Możemy przyjąć, że zbiór rozmyty "młody" człowiek może zostać określony w zbiorze U przez funkcję przynależności daną w postaci:

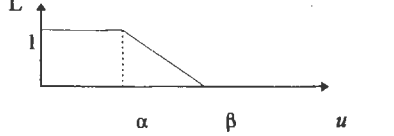
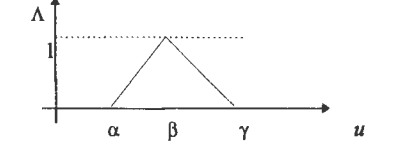
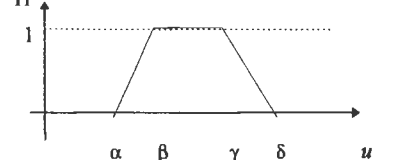
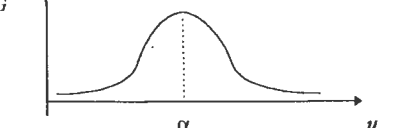
$$\mu_{\text{młody}}(u) = \begin{cases} 1.0 & \text{dla } 1 \leq u \leq 25 \\ \frac{1}{1 + (\frac{u-25}{5})^2} & \text{dla } 25 < u \leq 100 \end{cases}$$

Powiemy, że człowiek w wieku 30 lat jest "młody" w stopniu 0.5, ponieważ $\mu_{\text{młody}}(30) = 0.5$, a w wieku 20 lat jest młody w stopniu 1.0, ponieważ $\mu_{\text{młody}}(20) = 1.0$.

□

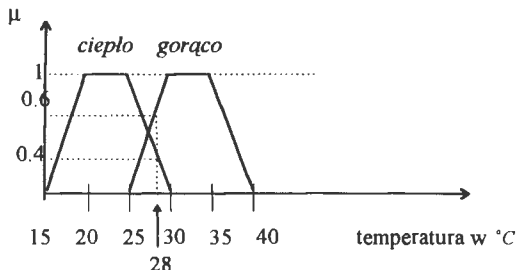
W praktyce, przy określaniu funkcji przynależności można zastosować jedną ze znanych z literatury, często stosowanych typów funkcji przynależności. Poniżej przedstawiono najczęściej stosowane postacie funkcji w przypadku, gdy zbiór U tworzą liczby rzeczywiste, $u \in \mathcal{R}$. W pierwszej kolumnie zamieszczono postać funkcji przynależności, a w drugiej kolumnie jej interpretację graficzną.

$\Gamma(u, \alpha, \beta) = \begin{cases} 0 & \text{dla } u < \alpha \\ \frac{u - \alpha}{\beta - \alpha} & \text{dla } \alpha \leq u \leq \beta \\ 1 & \text{dla } u > \beta \end{cases}$	
$S(u, \alpha, \beta, \gamma) = \begin{cases} 0 & \text{dla } u \leq \alpha \\ \frac{2(u - \alpha)^2}{\gamma - \alpha} & \text{dla } \alpha \leq u \leq \beta \\ 1 - \frac{2(\gamma - u)^2}{\gamma - \alpha} & \text{dla } \beta \leq u \leq \gamma \\ 1 & \text{dla } u > \gamma \end{cases}$	
$F(u, \alpha, k) = \begin{cases} 0 & \text{dla } u \leq \alpha \\ \frac{k(u - \alpha)^2}{1 + k(u - \alpha)^2} & \text{dla } u > \alpha \end{cases}$ <p>dla $k > 0$</p>	 <p>dla $k = 1$</p>

$L(u; \alpha, \beta) = \begin{cases} 1 & \text{dla } u < \alpha \\ \frac{\beta - u}{\beta - \alpha} & \text{dla } \alpha \leq u \leq \beta \\ 0 & \text{dla } u > \beta \end{cases}$	
$\Lambda(u; \alpha, \beta, \gamma) = \begin{cases} 0 & \text{dla } u \leq \alpha \\ \frac{u - \alpha}{\beta - \alpha} & \text{dla } \alpha < u \leq \beta \\ \frac{\gamma - u}{\gamma - \beta} & \text{dla } \beta < u \leq \gamma \\ 0 & \text{dla } u > \gamma \end{cases}$ <p>równoważny zapis jest postaci</p> $\Lambda(u; \alpha, \beta, \gamma) = \max\left\{\min\left(\frac{u - \alpha}{\beta - \alpha}, \frac{\gamma - u}{\gamma - \beta}\right), 0\right\}$	
$\Pi(u; \alpha, \beta, \gamma, \delta) = \begin{cases} 0 & \text{dla } u \leq \alpha \\ \frac{u - \alpha}{\beta - \alpha} & \text{dla } \alpha < u < \beta \\ 1 & \text{dla } \beta < u \leq \gamma \\ \frac{u - \delta}{\gamma - \delta} & \text{dla } \gamma < u \leq \delta \\ 0 & \text{dla } u > \delta \end{cases}$ <p>równoważny zapis jest postaci</p> $\Pi(u; \alpha, \beta, \gamma) = \max\left\{\min\left(\frac{u - \alpha}{\beta - \alpha}, 1, \frac{\delta - u}{\delta - \gamma}\right), 0\right\}$	
$G(u; \delta, \alpha) = \frac{1}{e^{\delta(u - \alpha)^2}}, \quad k > 0$	

Przykład 2.3. Rozpatrzmy zbiór temperatur z przedziału $U = [15^\circ\text{C}, 40^\circ\text{C}]$. Jeżeli chcemy opisać temperaturę w języku naturalnym, możemy utworzyć zmienną lingwistyczną o symbolicznej nazwie *temperatura* z wartościami: *ciepło* i *gorąco*. Każda wartość którą może przyjąć zmienna lingwistyczna określa pewien zbiór rozmyty. W naszym przypadku możemy określić dwa zbiory rozmyte w zbiorze temperatur U : *ciepło* i *gorąco*, zawierające te temperatury podane w stopniach Celsjusza, które są odbierane przez ludzi odpowiednio jako odczucie ciepła i gorąca. Aby zdefiniować oba zbiory musimy dla każdej temperatury ze zbioru U określić, czy może być uznana odpowiednio za ciepłą lub za gorącą. Subiektywność ocen ludzkich ekspertów może prowadzić do rozmycia granic tych zbiorów. Przyjmijmy, że zmienna lingwistyczna *temperatura* dla temperatur $[15^\circ\text{C}, 30^\circ\text{C}]$ przyjmuje wartość *ciepło*, a dla temperatur $[25^\circ\text{C}, 40^\circ\text{C}]$ przyjmuje wartość *gorąco*. Zakładając, że funkcje przynależności do zbiorów rozmytych mają kształt trapezowy, na rys. 2.1 przedstawiono graficzne ich interpretacje.

Tak więc temperatura z przedziału $[15^{\circ}\text{C}, 25^{\circ}\text{C}]$ na pewno jest odbierana jako ciepła a z przedziału $[30^{\circ}\text{C}, 40^{\circ}\text{C}]$ odbierana jest jako gorąca. Jednak zakresy tak określonych zbiorów rozmytych częściowo się pokrywają; temperatura z przedziału $[25^{\circ}\text{C}, 30^{\circ}\text{C}]$ należy jednocześnie do obu zbiorów rozmytych, z różnym stopniem przynależności. Przykładowo, powietrze o temperaturze 28°C może być odbierane jako ciepłe w stopniu 0.4, ale również może być odbierane jako gorące w stopniu 0.6, jak pokazano na rys. 2.2.

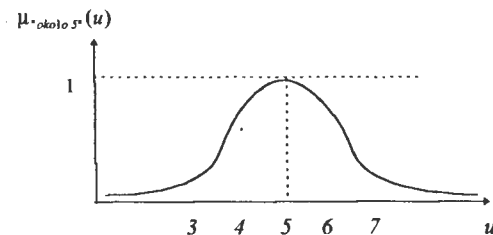


Rys. 2.1. Funkcje przynależności do zbiorów rozmytych *ciepło* i *gorąco*

□

Zwykle zakłada się, że funkcja przynależności liczby rozmytej ma kształt "dzwonu" (np. "wokół" liczby 5 dla liczby rozmytej "około 5") albo trójkąta.

Przykład 2.4. Liczbę rozmytą "około 5" w zbiorze liczb rzeczywistych możemy określić za pomocą funkcji przynależności $\mu_{\text{około } 5}(u) = \frac{1}{e^{(u-5)^2}}$ dla $u \in \mathbb{R}$ lub też w postaci graficznej jak na rys.2.2.



Rys.2.2. Funkcja przynależności liczby rozmytej "około 5"

□

Interpretacja funkcji przynależności do zbioru rozmytego może być dwojaka.

- **Lingwistyczna.** Funkcja przynależności modeluje pewne kategorie lingwistyczne, takie jak "przybliżony", "wesoły", "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 określeniu postaci funkcji przynależności przez ekspertów a następnie na automatycznej ich modyfikacji z zastosowaniem zbioru przykładów uczących.

Metody stosujące opinie ludzkich ekspertów stosują zazwyczaj pojęcia lingwistyczne. Ekspersi 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 wysoki?
- **Metody kierunkowe.** Mogą być stosowane przy opinii jednego jak też grupy ekspertów. Przykład pytania: w jakim stopniu można uznać Kowalskiego za wysokiego?
- **Metody pionowe.** Określają zbiory rozmyte z ich α -cięć. Przykładowy zbiór ludzi, którzy są wysocy 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 wyższy, Kowalski czy też Nowak i w jakim stopniu?

Metody automatyczne, określają wartości funkcji przynależności na podstawie zbioru przykładów uczących [Keller i Givens, 1985], [Pał 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** [Pał 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

$$u_{\max} = \max_n \{u_n\}, \quad u_{\min} = \min_n \{u_n\}, \quad \bar{u} = \frac{1}{N} \sum_{n=1}^N u_n.$$

Definiujemy parametry

$$a_r = \bar{u} + \max\{|\bar{u} - u_{\min}|, |\bar{u} - u_{\max}|\}, \quad a_l = 2\bar{u} - a_r,$$

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** (ang. 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 $u \in \mathfrak{X}$ 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 $\mu_{G_1}(u), \dots, \mu_{G_c}(u)$, dla każdego $u \in \mathfrak{X}$. Po pierwsze znajdujemy k punktów w zbiorze U przy punkcie u a następnie obliczamy

$$\mu_{\tilde{G}_i}(u) = \frac{k_i}{k}, \quad \text{gdzie } k_i \text{ jest liczbą elementów należących do klasy } G_i.$$

2.2.3. 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, 1985].

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

- **jądro zbioru rozmytego** $Core(\tilde{G}) = \{u \mid u \in U, \mu_{\tilde{G}}(u) = 1\}$, (2.5)

- **nośnik zbioru rozmytego** $Supp(\tilde{G}) = \{u \mid u \in U, \mu_{\tilde{G}}(u) > 0\}$, (2.6)

- **wysokość zbioru rozmytego** $hgt(\tilde{G}) = \sup_{u \in U} \{\mu_{\tilde{G}}(u)\}$, (2.7)

- **cięcie typu α zbioru rozmytego** $\tilde{G}_\alpha = \{u \mid u \in U, \mu_{\tilde{G}}(u) \geq \alpha\}$, (2.8)

- **liczność zbioru rozmytego skończonego** $card(\tilde{G}) = \sum_{u \in U} \mu_{\tilde{G}}(u)$, (2.9)

$$\text{nieskończonego} \quad \text{card}(\tilde{G}) = \int_{u \in U} \mu_{\tilde{G}}(u) . \quad (2.10)$$

$$\bullet \text{ względna liczność skończonego zbioru rozmytego} \quad \|\tilde{G}\| = \frac{\text{card}(\tilde{G})}{\text{card}(U)} \quad (2.11)$$

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 2.5. 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 $\tilde{G}_{\alpha=0.5} = \{3, 4\}$.

□

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.

$$\sup_{u \in U} \{\mu_{\tilde{G}}(u)\} = 1, \quad (2.12)$$

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

$$\mu_{\tilde{G}}(\lambda \cdot u_1 + (1-\lambda) \cdot u_2) \geq \min \{\mu_{\tilde{G}}(u_1), \mu_{\tilde{G}}(u_2)\}, \quad (2.13)$$

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

$$\mu_{\tilde{G}}(\lambda \cdot u_1 + (1-\lambda) \cdot u_2) \leq \min \{\mu_{\tilde{G}}(u_1), \mu_{\tilde{G}}(u_2)\}. \quad (2.14)$$

2.2.4. Operacje jednoargumentowe na zbiorach rozmytych

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

Definicja 2.2. Dopelnienie zbioru rozmytego

Dopelnienie zbioru rozmytego \tilde{G} w zbiorze U jest zbiorem rozmytym w U , oznaczanym $-\tilde{G}$, określonym funkcją przynależności $\mu_{-\tilde{G}}: U \rightarrow [0,1]$. Funkcja przynależności do dopelnienia zbioru rozmytego jest definiowana przez funkcję $h: [0,1] \rightarrow [0,1]$, to znaczy $\mu_{-\tilde{G}}(u) = h(\mu_{\tilde{G}}(u))$, $\forall u \in U$, 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 $a \in [0,1]$.

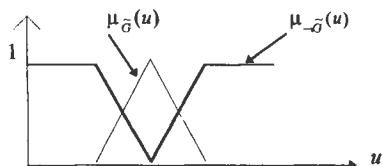
Warunki od (1) do (4) nie określają jednoznacznie postaci funkcji przynależności do dopełnienia zbioru rozmytego. Jeżeli do powyższych warunków dodamy jeszcze jeden warunek

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

$$\mu_{\tilde{G}}(u) = 1 - \mu_{\tilde{G}}(u). \quad (2.15)$$

Powyższa definicja dopełnienia zbioru rozmytego została określona przez Zadeha i nazywana jest **standardowym dopełnieniem zbioru rozmytego** \tilde{G} w U . Może być ona przedstawiona graficznie w przypadku, gdy funkcja przynależności jest funkcją trójkątną w sposób pokazany na rys.2.3.



Rys 2.3. Graficzna interpretacja zbioru rozmytego z trójkątną funkcją przynależności

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

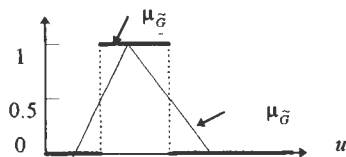
$$\neg_{\lambda}\tilde{G} = \frac{1 - \mu_{\tilde{G}}(u)}{1 + \lambda \cdot \mu_{\tilde{G}}(u)}, \quad \lambda \in [-1, \infty) \quad (2.16)$$

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

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

$$\mu_{\underline{\tilde{G}}}(u) = \begin{cases} 0 & \text{dla } \mu_{\tilde{G}}(u) \leq 0.5 \\ 1 & \text{dla } \mu_{\tilde{G}}(u) > 0.5 \end{cases}, \quad u \in U \quad (2.17)$$

Graficzną interpretację najbliższego zbioru zwykłego dla zbioru rozmytego z trójkątną funkcją przynależności przedstawiono na rys. 2.4.



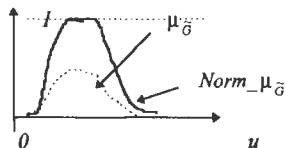
Rys 2.4. Zbiór zwykły najbliższy zbiorowi rozmytemu \tilde{G}

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 $Norm_μ_{\tilde{G}}$ przyjmuje wartości z przedziału $[0,1]$,

$$Norm_μ_{\tilde{G}} = \frac{μ_{\tilde{G}}}{hgt(\tilde{G})} \quad (2.18)$$

Graficzną interpretację operacji normalizacji zbioru \tilde{G} przedstawiono na rys. 2.5.

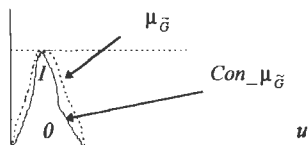


Rys 2.5. Normalizacja zbioru rozmytego

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

$$Con_μ_{\tilde{G}} = (μ_{\tilde{G}}(u))^p, \text{ dla } p > 1 \quad (2.19)$$

Graficzną interpretację koncentracji zbioru rozmytego przedstawiono na rys. 2.6.



Rys 2.6. Koncentracja zbioru rozmytego

Zgodnie z definicją zbioru rozmytego zachodzi warunek $0 \leq μ_{\tilde{G}}(u) \leq 1$, zatem mamy dla $p > 1$ spełnienie warunku $(μ_{\tilde{G}}(u))^p < μ_{\tilde{G}}(u)$, poza warunkami brzegowymi równymi $μ_{\tilde{G}}(u) = 0$, $μ_{\tilde{G}}(u) = 1$. Wielkość koncentracji zależy od elementu u i wynosi $\frac{μ_{\tilde{G}}(u) - (μ_{\tilde{G}}(u))^p}{μ_{\tilde{G}}(u)} = 1 - (μ_{\tilde{G}}(u))^{p-1}$. 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 $μ_{\tilde{G}}(u) = 0.1$ będzie miał zmniejszoną o 90% wartość tej funkcji do $(μ_{\tilde{G}}(u))^2 = 0.01$, podczas gdy element dla którego wartość funkcji przynależności wynosi $μ_{\tilde{G}}(u) = 0.9$ zmniejszoną tylko do $(μ_{\tilde{G}}(u))^2 = 0.81$, to znaczy o 10%.

Operacja koncentracji bywa stosowana w zbiorach rozmytych do reprezentowania pojęcia "bardzo". Przykładowo, jeżeli mamy funkcję przynależności $μ_{\text{bardzo}}(u)$ do zbioru

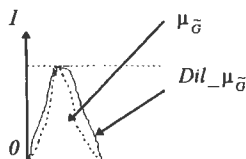
rozmytego "młody", to funkcja $Con_{\mu_{\tilde{G}}}$ może być funkcją przynależności do zbioru rozmytego "bardzo młody".

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

$$Dil_{\mu_{\tilde{G}}} = \sqrt{\mu_{\tilde{G}}(u)} \quad (2.20)$$

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

Przykład graficznej interpretacji rozcieńczenia przedstawiono na rys. 2.7.



Rys 2.7. Rozcieńczenie zbioru rozmytego

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{G} wybranym w sposób nieco arbitralny, patrz wzór (2.17).

$$H_{Hamming}(\tilde{G}) = \frac{2}{N} d_H(\tilde{G}, \underline{G}) = \frac{2}{N} \sum_{i=1}^N \left| \mu_{\tilde{G}}(u_i) - \mu_{\underline{G}}(u_i) \right| \quad (2.21)$$

oraz

$$H_{Euclidean}(\tilde{G}) = \frac{2}{\sqrt{N}} d_E(\tilde{G}, \underline{G}) = 2 \sqrt{\frac{1}{N} \sum_{i=1}^N (\mu_{\tilde{G}}(u_i) - \mu_{\underline{G}}(u_i))^2} \quad (2.22)$$

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

$$H_{Entropy}(\tilde{G}) = -k \sum_{i=1}^N [\mu_{\tilde{G}}(u_i) \cdot \log(\mu_{\tilde{G}}(u_i)) + \mu_{\underline{G}}(u_i) \cdot \log(\mu_{\underline{G}}(u_i))] \quad (2.23)$$

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 i Folger, 1988].

2.2.5. 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 [Schweitzer i Sklar, 1963, 1983], [Weber, 1983], [Butnariu i Klement, 1993].

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.

$$t(a, b) = 1 - s((1-a), (1-b)) \quad (2.24)$$

$$s(a, b) = 1 - t((1-a), (1-b)). \quad (2.25)$$

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** $t_3(a, b) = \max\{0, a + b - 1\}$,
- **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:

- **maksimum**, oznaczane $a \vee b$, $s_1(a, b) = \max\{a, b\}$,
- **suma probabilistycznej** $s_2(a, b) = a + b - a \cdot b$
- **suma ograniczona** $s_3(a, b) = \min\{1, a + b\}$

- **drastyczna suma**

$$s_d(a, b) = \begin{cases} a & \text{jeżeli } b = 0 \\ b & \text{jeżeli } a = 0 \\ 1 & \text{jeżeli } a > 0, b > 0 \end{cases}$$

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

Definicja 2.3. Działania na zbiorach rozmytych

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

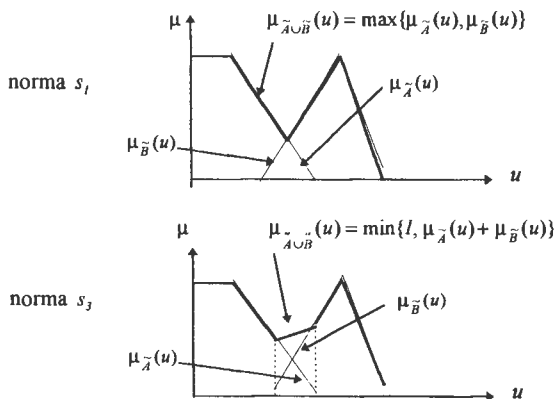
- **suma zbiorów rozmytych (ang. union) $\tilde{A} \cup \tilde{B}$:**

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

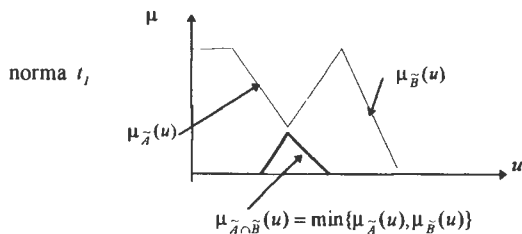
- **przecięcie zbiorów rozmytych (ang. intersection) $\tilde{A} \cap \tilde{B}$:**

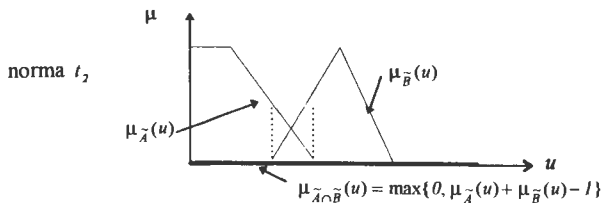
$$\tilde{A} \cap \tilde{B} = \{u; \mu_{\tilde{A} \cap \tilde{B}}(u) \mid \forall u \in U, \mu_{\tilde{A} \cap \tilde{B}}(u) = t(\mu_{\tilde{A}}(u), \mu_{\tilde{B}}(u))\}.$$

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 . Interpretacje graficzne działań na zbiorach rozmytych dla kilku wybranych typów norm przedstawiono na rys. 2.8 i 2.9.



Rys. 8.8. Suma zbiorów rozmytych, $\mu_{\tilde{A} \cup \tilde{B}}(u)$, dla norm s_t i s_s





Rys. 2.9. Przecięcie zbiorów rozmytych, $\mu_{\tilde{A} \cap \tilde{B}}(u)$, dla norm t_1 i t_3

Przykład 2.6. Niech zbiór U będzie zbiorem uczniów, $U = \{u_1, u_2, u_3\}$, zdefiniujmy dwa zbiory rozmyte \tilde{A} i \tilde{B} w zbiorze U . Pierwszy niech będzie zbiorem uczniów, których ulubionym przedmiotem szkolnym jest matematyka $\tilde{A} = \{(u_1; 0.8), (u_2; 0.7), (u_3; 0.1)\}$, drugi zbiorem tych uczniów, których ulubionym przedmiotem jest język polski $\tilde{B} = \{(u_1; 0.0), (u_2; 0.5), (u_3; 0.9)\}$. Zbiór rozmyty będący sumą tych dwóch zbiorów (przyjmujemy normę typu s_1 w postaci funkcji maksimum) zawiera uczniów, których ulubionym przedmiotem szkolnym jest matematyka lub język polski, $\tilde{A} \cup \tilde{B} = \{(u_1; 0.8), (u_2; 0.7), (u_3; 0.9)\}$, a zbiór rozmyty będący przecięciem tych dwóch zbiorów (przyjmujemy normę typu t_1 w postaci funkcji minimum) zawiera uczniów, których ulubionym przedmiotem jest matematyka jak również język polski, $\tilde{A} \cap \tilde{B} = \{(u_1; 0.0), (u_2; 0.5), (u_3; 0.1)\}$.

□

Uwaga. Trzeba pamiętać, że ogólnie normy trójkątne nie gwarantują spełnienia praw zaprzeczenia i wyłączonego środka, $\tilde{A} \cap \neg \tilde{A} \neq \emptyset$ oraz $\tilde{A} \cup \neg \tilde{A} \neq U$. Przykład taki przedstawiono poniżej.

Przykład 2.7. Rozpatrzmy zbiór rozmyty \tilde{A} w zbiorze U , o stałej funkcji przynależności $\mu_{\tilde{A}}(u) = 1/2, \forall u \in U$. Wybierzmy do obliczeń normę typu t_1 w postaci funkcji minimum oraz normę typu s_1 w postaci funkcji maksimum. Otrzymujemy następujące zależności:

$$\mu_{\tilde{A} \cap \neg \tilde{A}} = \min(\mu_{\tilde{A}}, \mu_{\neg \tilde{A}}) = \min\left\{\frac{1}{2}, \frac{1}{2}\right\} = \frac{1}{2},$$

$$\mu_{\tilde{A} \cup \neg \tilde{A}} = \max(\mu_{\tilde{A}}, \mu_{\neg \tilde{A}}) = \max\left\{\frac{1}{2}, \frac{1}{2}\right\} = \frac{1}{2},$$

co oznacza, że w tym przypadku $\tilde{A} \cap \neg \tilde{A} \neq \emptyset$ oraz $\tilde{A} \cup \neg \tilde{A} \neq U$. Wybierzmy następnie do obliczeń normę t_2 z wartością parametru $p=0$ oraz normę s_2 z parametrem $p=0$. Otrzymujemy wtedy

$$\mu_{\tilde{A} \cap \neg \tilde{A}} = t_2(\mu_{\tilde{A}}, \mu_{\neg \tilde{A}}) = \max\{0, \mu_{\tilde{A}} + (1 - \mu_{\tilde{A}}) - 1\} = 0,$$

$$\mu_{\tilde{A} \cup \neg \tilde{A}} = s_2(\mu_{\tilde{A}}, \mu_{\neg \tilde{A}}) = \min\{1, \mu_{\tilde{A}} + (1 - \mu_{\tilde{A}})\} = 1.$$

W tym przypadku spełnione są zależności $\tilde{A} \cap \neg \tilde{A} = \emptyset$ oraz $\tilde{A} \cup \neg \tilde{A} = U$.

□

W dalszej części pracy będziemy rozpatrywali tylko normalne zbiory rozmyte, wzór (2.12), aby zapewnić spełnienie praw zaprzeczenia i wyłączonego środka, co jest dogodne w

zastosowaniach praktycznych. Oczywiście, z każdego zbioru rozmytego możemy otrzymać normalny zbiór rozmyty, stosując operację normalizacji, wzór (2.18).

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

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

gdzie $\|\cdot\|$ oznacza względną licznosc skończonego zbioru rozmytego (2.11).

2.2.6. Agregacja zbiorów rozmytych

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

Przykład 2.8. Rozpatrzmy zbiór uczestników pewnego konkursu $U = \{u_1, \dots, u_N\}$ oraz zespół L jurorów którzy mają wyłonić finalistów [Kuncheva, 2000]. Zakładamy, że nie ma ustalonego limitu liczby finalistów. Każdy uczestnik musi zostać jednoznacznie zaklasyfikowany do jednej z dwóch klas {przeszedł, odpadł}. Każdy juror przyporządkowuje każdemu uczestnikowi liczbę z przedziału od 0 do 10, określającą poparcie kandydata przy wejściu do finału. Określmy zbiór rozmyty \tilde{G}_i w zbiorze uczestników U odpowiadający opinii i -tego eksperta, $i \in \{1, \dots, L\}$, z funkcją przynależności do klasy "przeszedł", $\{(u_1; \mu_{G_i}(u_1)), \dots, (u_N; \mu_{G_i}(u_N))\}$, gdzie

$$\mu_{G_i}(u_n) = \frac{\text{poparcie } i\text{-tego eksperta dla } n\text{-tego kandydata}}{10}$$

Tak więc dla każdego uczestnika mamy zbiór L stopni przynależności $\{(u_i; \mu_{G_1}(u_i)), \dots, (u_i; \mu_{G_L}(u_i))\}$ i aby rozstrzygnąć konkurs musimy zagregować L opinii jurorów. \square

Rozpatrzmy zbiór L zbiorów rozmytych \tilde{G}_i , $i = 1, \dots, L$, w zbiorze U . Dla każdego elementu u ze zbioru U możemy określić stopnie przynależności do L zbiorów rozmytych, określając odpowiednio wartości funkcji przynależności $\mu_{\tilde{G}_1}(u), \dots, \mu_{\tilde{G}_L}(u)$. Aby obliczyć zagregowany stopień przynależności elementu u musimy zagregować powyższe wartości funkcji przynależności, które dla uproszczenia zapisu będziemy dalej w pracy oznaczali a_1, a_2, \dots, a_L . Tak więc zbiór a_1, a_2, \dots, a_L , $a_i \in [0, 1]$ jest zbiorem wartości które chcemy agregować.

Operatory agregacji

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

1. Przemienność. $\mathcal{A}(a_1, a_2, \dots, a_L) = \mathcal{A}(a_{i_1}, a_{i_2}, \dots, a_{i_L})$ 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 $\mathcal{A}(a_1, a_2, \dots, a_L) \leq \mathcal{A}(b_1, b_2, \dots, b_L)$.
3. Spełnienie warunków brzegowych $\mathcal{A}(0, 0, \dots, 0) = 0, \mathcal{A}(1, 1, \dots, 1) = 1$.

Agregacja \mathcal{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

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

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

- **Minimum** $\mathcal{A}(a_1, a_2, \dots, a_L) = \min\{a_1, a_2, \dots, a_L\}$.
- **Maksimum** $\mathcal{A}(a_1, a_2, \dots, a_L) = \max\{a_1, a_2, \dots, a_L\}$.
- **Iloczyn** $\mathcal{A}(a_1, a_2, \dots, a_L) = \prod_{i=1}^L (a_i)$.
- **Średnia** $\mathcal{A}(a_1, a_2, \dots, a_L) = \sum_{i=1}^L a_i$.
- **Średnia uogólniona** $\mathcal{A}_\alpha(a_1, a_2, \dots, a_L) = \left(\frac{1}{L} \sum_{i=1}^L a_i^\alpha \right)^{\frac{1}{\alpha}}$.

W zależności od przyjętej wartości parametru α otrzymujemy specjalne przypadki operatora średniej uogólnionej, przedstawione poniżej.

$$\alpha \rightarrow -\infty \Rightarrow \text{otrzymujemy } \mathcal{A}_\alpha(a_1, a_2, \dots, a_L) = \min_i \{a_i\}$$

$$\alpha = -1 \Rightarrow \mathcal{A}_\alpha(a_1, a_2, \dots, a_L) = \left(\frac{1}{L} \sum_{i=1}^L \frac{1}{a_i} \right)^{-1},$$

$$\alpha = 0 \Rightarrow \mathcal{A}_\alpha(a_1, a_2, \dots, a_L) = (a_1 \cdot a_2 \cdot \dots \cdot a_L)^{\frac{1}{L}},$$

$$\alpha = 1 \Rightarrow \mathcal{A}_\alpha(a_1, a_2, \dots, a_L) = \frac{1}{L} \sum_{i=1}^L a_i,$$

$$\alpha \rightarrow \infty \Rightarrow \mathcal{A}_\alpha(a_1, a_2, \dots, a_L) = \max_i \{a_i\}.$$

Operatory OWA

Ciekawą klasę operatorów agregujących stanowią zaproponowane przez Yagera [Yager, 1988, 1997] operatory OWA (ang. Ordered Weighted Averaging). Niech $\mathbf{b} = [b_1, b_2, \dots, b_L]^T \in [0, 1]^L$ będzie wektorem współczynników takim, że

$$\sum_{k=1}^L b_k = 1.$$

Operator OWA określony jest na produkcie kartezjańskim wektora \mathbf{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 $a_{i_1} \geq a_{i_2} \geq \dots \geq a_{i_L}$,

$$\mathcal{A}_{\mathbf{b}}^{OWA}(a_1, a_2, \dots, a_L) = \sum_{k=1}^L a_{i_k} \cdot b_k.$$

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

minimum, przyjmując $\mathbf{b} = [0, 0, \dots, 0, 1]^T$,

maksimum, przyjmując $\mathbf{b} = [1, 0, \dots, 0]^T$,

średnią, przyjmując $\mathbf{b} = [1/L, 1/L, \dots, 1/L]^T$,

konkurs jurorów, przyjmując $\mathbf{b} = [0, 1/(L-2), \dots, 1/(L-2), 0]^T$.

Przykład 2.9. Przyjmijmy, że pewien uczestnik u konkursu z przykładu 2.8 otrzymał poparcie od pięciu jurorów, $L=5$, przedstawione w tabl. 2.1.

Tabl. 2.1. Poparcie uczestnika u otrzymane od pięciu jurorów

Juror	Ocena jurora w skali od 0 do 10
1	6
2	7
3	2
4	6
5	6

Wartości funkcji przynależności uczestnika u do klasy "przeszedł" wynoszą $[a_1, a_2, \dots, a_L]^T = [0.6, 0.7, 0.2, 0.6, 0.6]^T$. Aby zaimplementować konkurs jurorów zastosujemy operator OWA z wektorem $\mathbf{b} = [0, 1/(L-2), \dots, 1/(L-2), 0]^T = [0, 1/3, 1/3, 1/3, 0]^T$. Otrzymujemy

$$\begin{aligned} \mathcal{A}_b^{OWA}(0.6, 0.7, 0.2, 0.6, 0.6) &= \\ &= [0, 1/3, 1/3, 1/3, 0][0.7, 0.6, 0.6, 0.6, 0.2]^T = (0.6 + 0.6 + 0.6)/3 = 0.6. \end{aligned}$$

□

Wartości wektora \mathbf{b} mogą być ustalone z góry jak również mogą być wyliczane na podstawie danego zbioru danych. Yager i Filew 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 jednomysłność. Pięć miar konsensusu $\gamma_1, \dots, \gamma_5$ [Kuncheva, 1992] przedstawiono poniżej.

- **Najwyższa zgodność** $\gamma_1 = 1 - \min_{i,j, i \neq j} \{ |a_i - a_j| \};$
- **Najwyższa niezgodność** $\gamma_2 = 1 - \max_{i,j, i \neq j} \{ |a_i - a_j| \};$
- **Całkowita średnia zgodność** $\gamma_3 = 1 - \sum_{i=1}^L |a_i - \bar{a}|$, gdzie $\bar{a} = \frac{1}{L} \sum_{i=1}^L a_i$;

- Całkowita zgodność dla wszystkich par

$$\gamma_s = 1 - \frac{2}{L(L-1)} \sum_{i=1}^{L-1} \sum_{j=i+1}^L |a_i - a_j|;$$

- Całkowita najwyższa niezgodność

$$\gamma_s = 1 - \max_{i \in \{1, \dots, L\}} \{|a_i - \bar{a}|\}.$$

Przykład 2.10. W przykładzie 2.9 otrzymaliśmy następujące wartości funkcji przynależności uczestnika u do klasy "przeszedł" $\{0.6, 0.7, 0.2, 0.6, 0.6\}$, które chcemy zagregować. Stosując kolejno miary konsensusu $\gamma_1, \dots, \gamma_5$ otrzymujemy różne miary, przedstawione poniżej

$$\gamma_1 = 1 - |0.6 - 0.6| = 1,$$

$$\gamma_2 = 1 - |0.7 - 0.2| = 0.5,$$

$$\gamma_3 = 1 - 1/(0.06+0.16+0.34+0.06+0.06) = 0.864, \quad (\bar{a} = 0.54),$$

$$\gamma_4 = 1 - 2/20(0.1+0.4+0.5+0.1+0.5+0.4+0.4) = 1 - 0.240 = 0.760,$$

$$\gamma_5 = 1 - 0.34 = 0.66.$$

□

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

$$\mathbb{K}(a_1, a_2, \dots, a_L) = \mathbb{K}(\mathcal{A}(a_1, a_2, \dots, a_L), \gamma(a_1, a_2, \dots, a_L)).$$

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

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

$$\mathbb{K}(\mathcal{A}, \gamma) \text{ jest } \begin{cases} \text{monotonicznie nie malejąca dla } \gamma & \text{jeśli } A > 0.5 \\ \text{monotonicznie nie rosnąca dla } \gamma & \text{jeśli } A \leq 0.5 \end{cases}$$

3. a. jednomyślność, $\mathbb{K}(\mathcal{A}, 1) = \mathcal{A}$, lub

$$3. \text{ b. wzmocniona jednomyślność, } \mathbb{K}(\mathcal{A}, 1) \text{ jest } \begin{cases} > A & A \in (0.5, 1) \\ < A & A \in (0, 0.5) \\ = A = 0.5 & A = 0.5 \end{cases}$$

Operacja konsensusu spełniająca aksjomaty 1, 2 i 3a jest postaci

$$\mathbb{K}_1(\mathcal{A}, \gamma) = 0.5(1 - \gamma) + \mathcal{A} \gamma,$$

a spełniająca aksjomaty 1, 2 i 3b jest postaci

$$\mathbb{K}_2(\mathcal{A}, \gamma) = \frac{I}{1 + \exp\{-\alpha\gamma(A - 0.5)\}},$$

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 \mathcal{A} . Jest to odbiciem sytuacji, w której jeśli nie ma porozumienia

między jurorami przyjmowana jest wartość 0.5. Jeżeli wartość zagregowana \mathcal{A} jest równa zero, operator konsensusu \mathbb{K} może zwrócić wartość z przedziału od 0 do 0.5. Niezerowa wartość \mathbb{K} przy warunku $\mathcal{A} = 0$ wskazuje, że istnieje niezgodność wśród a_1, a_2, \dots, a_L .

2.2.7. Relacje rozmyte

Żałómy, 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 $\text{card}(U)$ elementów, a zbiór V zawiera $\text{card}(V)$ elementów, wtedy iloczyn kartezjański ma $\text{card}(U) \cdot \text{card}(V)$ elementów.

Definicja 2.4. 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

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

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

$$\tilde{R} = \bigcup_{U \times V} \{(u, v); \mu_{\tilde{R}}(u, v)\}. \quad (2.26)$$

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

$$\tilde{R} = \{(u_i, v_j); \mu_{\tilde{R}}(u_i, v_j) \mid j = 1, \dots, K, i = 1, \dots, N\}, \quad (2.27)$$

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

$$\begin{matrix} & v_1 & v_2 & \dots & v_K \\ \begin{matrix} u_1 \\ \dots \\ u_N \end{matrix} & \left[\begin{array}{cccc} \mu_{\tilde{R}}(u_1, v_1) & \mu_{\tilde{R}}(u_1, v_2) & \dots & \mu_{\tilde{R}}(u_1, v_K) \\ \dots & \dots & \dots & \dots \\ \mu_{\tilde{R}}(u_N, v_1) & \mu_{\tilde{R}}(u_N, v_2) & \dots & \mu_{\tilde{R}}(u_N, v_K) \end{array} \right] \end{matrix} \quad (2.28)$$

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

$$\tilde{R} = \int_U \int_V \{(u, v); \mu_{\tilde{R}}(u, v)\}. \quad (2.29)$$

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

$$\text{Proj}_U \tilde{R}(u) = \sup_{v \in V} \tilde{R}(u, v) \quad (2.30)$$

oraz na zbiór V

$$\text{Proj}_V \tilde{R}(v) = \sup_{u \in U} \tilde{R}(u, v). \quad (2.31)$$

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

$$\tilde{R} = \{(u_1, u_2, \dots, u_K); \mu_{\tilde{R}}(u_1, u_2, \dots, u_K)\} \quad \forall u_i \in U_i, \dots, \forall u_K \in U_K \quad (2.32)$$

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 2.5. 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 $U \times W \rightarrow [0, 1]$,

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

dla norm typu s i typu t .

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

$$\tilde{R}_1 \circ \tilde{R}_2 = \{(u, w); \max_{\forall v \in V} \{\min(\mu_{\tilde{R}_1}(u, v), \mu_{\tilde{R}_2}(v, w))\}\} \quad \forall u \in U, \forall w \in W \quad (2.33)$$

gdzie $\max_{\forall v \in V} \{x\}$ oznacza wybór maksymalnej wartości x dla wszystkich elementów $v \in V$. Możemy również wybierać inne normy, np. $s_2(u, v) = \min(u + v, 1)$ oraz $t_4(u, v) = uv$. 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 złożenia dwóch relacji rozmytych przedstawiono poniżej.

Przykład 2.11. Niech relacje $\tilde{R}_1 = \{(u_i, v_j); \mu_{\tilde{R}_1}(u_i, v_j)\} \quad j = 1, 2 \quad i = 1, 2$,
 $\tilde{R}_2 = \{(v_i, w_j); \mu_{\tilde{R}_2}(v_i, w_j)\} \quad j = 1, 2 \quad i = 1, 2$ będą reprezentowane w postaci macierzy relacji, zgodnie z (2.28), 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

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

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_1, w_1) 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:

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

Dla pozostałych wartości w macierzy postępujemy w podobny sposób.

□

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

$$\tilde{R} = \{(u, v); t(\mu_{\tilde{U}}(u), \mu_{\tilde{V}}(v)) \mid \forall u \in \tilde{U}, \forall v \in \tilde{V}\} \quad (2.34)$$

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

$$\tilde{R} = \{(u, v); \min\{\mu_{\tilde{U}}(u), \mu_{\tilde{V}}(v)\} \mid \forall u \in \tilde{U}, \forall v \in \tilde{V}\}. \quad (2.35)$$

Przykład 2.12. Niech zbiór U będzie zbiorem trzech uczniów, $U = \{u_1, u_2, u_3\}$, opisanych za pomocą dwóch cech ze zbioru $A = \{a_1, a_2\}$, gdzie cecha a_1 : "ulubiony przedmiot ucznia" ma zbiór wartości $V_{a_1} = \{v_1^{a_1}, v_2^{a_1}, v_3^{a_1}\} = \{\text{matematyka, język polski, geografia, wychowanie fizyczne}\}$ a cecha a_2 : "zdolność logicznego myślenia ucznia", ma zbiór wartości $V_{a_2} =$

$\{v_1^{a_2}, v_2^{a_2}, v_3^{a_2}\} = \{\text{slaba, przecietna, wysoka}\}$. Każdy przedmiot szkolny ze zbioru V_{a_1} został opisany za pomocą cechy ze zbioru $B = \{b_i\}$, gdzie cecha b_i : "rodzaj nauk", ma zbiór wartości $V_{b_i} = \{v_i^{b_1}, v_i^{b_2}\} = \{\text{dedukcyjne, opisowe, inne}\}$.

Relacja rozmyta \tilde{R}_{a_1} , ze zbioru U do V_{a_1} , dana jest w postaci macierzy relacji przedstawionej poniżej, której elementami są wartości funkcji przynależności $\mu_{\tilde{R}_{a_1}}(u_i, v_j^{a_1}) \in [0, 1]$, $i = 1, 2, 3$, $j = 1, \dots, 4$, określającej, w jakim stopniu przedmiot $v_j^{a_1}$ jest ulubionym przedmiotem ucznia u_i . Relacja rozmyta \tilde{R}_{a_2} , ze zbioru U do V_{a_2} , dana jest w postaci macierzy relacji przedstawionej poniżej, której elementami są wartości funkcji przynależności $\mu_{\tilde{R}_{a_2}}(u_i, v_j^{a_2}) \in [0, 1]$, $i = 1, 2, 3$, $j = 1, 2, 3$.

	a_1 : "ulubiony przedmiot ucznia"				a_2 : "zdolności logicznego myślenia"		
	$v_1^{a_1}$:	$v_2^{a_1}$:	$v_3^{a_1}$:	$v_4^{a_1}$:	$v_1^{a_2}$:	$v_2^{a_2}$:	$v_3^{a_2}$:
	<i>matematyk</i>	<i>język polski</i>	<i>geografia</i>	<i>wychowanie</i>	<i>slaba</i>	<i>przecietna</i>	<i>wysoka</i>
	<i>a</i>				<i>e fizyczne</i>		
u_1	0	0.8	0.7	0.1	0	1.0	0
u_2	1.0	0	0	0.2	0	0	1.0
u_3	0	0	0	1.0	1.0	0	0

Relację \tilde{R}_{a_1} można zapisać w postaci zbioru rozmytego (dla uproszczenia zapisu, elementy o zerowej wartości funkcji przynależności zostały pominięte)

$$\tilde{R}_{a_1} = \{(u_1, \text{"język polski"}); 0.8, (u_1, \text{"geografia"}); 0.7, (u_1, \text{"wychowanie fizyczne"}); 0.1, (u_2, \text{"matematyka"}); 1.0, (u_2, \text{"wychowanie fizyczne"}); 0.2, (u_3, \text{"wychowanie fizyczne"}); 1.0\}$$

Rozpatrując podzbiory relacji \tilde{R}_{a_1} dla różnych ustalonych wartości jaką przyjmuje cecha a_1 otrzymujemy cztery zbiory rozmyte przedstawione poniżej, gdzie przykładowo $\tilde{R}_{a_1 = \text{"matematyka"}}$ jest rozmytym zbiorem uczniów, których ulubionym przedmiotem jest matematyka.

$$\tilde{R}_{a_1 = \text{"matematyka"}} = \{(u_2, \text{"matematyka"}); 1.0\},$$

$$\tilde{R}_{a_1 = \text{"język polski"}} = \{(u_1, \text{"język polski"}); 0.8\},$$

$$\tilde{R}_{a_1 = \text{"geografia"}} = \{(u_1, \text{"geografia"}); 0.7\},$$

$$\tilde{R}_{a_1 = \text{"wychowanie fizyczne"}} = \{(u_1, \text{"wych. fiz."}); 0.1, (u_2, \text{"wych. fiz."}); 0.2, (u_3, \text{"wych. fiz."}); 1.0\}.$$

Relację \tilde{R}_{a_2} można zapisać w postaci zbioru rozmytego przedstawionego poniżej (dla uproszczenia zapisu, elementy o zerowej wartości funkcji przynależności zostały pominięte).

$$\tilde{R}_{a_2} = \{(u_1, \text{"przecietna"}); 1.0, (u_2, \text{"wysoka"}); 1.0, (u_3, \text{"slaba"}); 1.0\}.$$

Zakładamy, że relację rozmytą \tilde{R}_{b_i} z V_{a_1} do V_{b_i} można przedstawić w postaci macierzy relacji przedstawionej poniżej, której elementami są wartości funkcji przynależności

$\mu_{\tilde{R}_{b_i}}(v_i^{a_i}, v_j^{b_i}) \in [0, 1], i=1, 2, 3, 4, j=1, 2, 3$ określającej, w jakim stopniu przedmiot $v_i^{a_i}$ należy do nauk typu $v_j^{b_i}$,

b_i : "rodzaj nauk"

$v_1^{b_i}$: nauki dedukcyjne $v_2^{b_i}$: nauki opisowe $v_3^{b_i}$: nauki inne

$v_1^{a_i}$: matematyka

$v_2^{a_i}$: język polski

$v_3^{a_i}$: geografia

$v_4^{a_i}$: wychowanie fizyczne

1.0	0.1	0
0	1.0	0
0.1	1.0	0
0	0	1.0

Relację \tilde{R}_{b_i} można zapisać w postaci zbioru rozmytego przedstawionego poniżej (dla uproszczenia zapisu, elementy o zerowej wartości funkcji przynależności zostały pominięte).

$$\tilde{R}_{b_i} = \{(matematyka, "dedukcyjne"); 1.0, (geografia, "dedukcyjne"); 0.1, (matematyka, "opisowe"); 0.1, (język polski, "opisowe"); 1.0, (geografia, "opisowe"); 1.0, (wychowanie fizyczne, "inne"); 1.0\}.$$

Rozpatrując podzbiory relacji \tilde{R}_{b_i} dla różnych ustalonych wartości cechy b_i otrzymujemy trzy zbiory rozmyte przedstawione poniżej, dla uproszczenia elementy o zerowej wartości funkcji przynależności zostały pominięte.

$$\tilde{R}_{b_i = "dedukcyjne"} = \{(matematyka, "dedukcyjne"); 1.0, (geografia, "dedukcyjne"); 0.1\},$$

$$\tilde{R}_{b_i = "opisowe"} = \{(matematyka, "opisowe"); 0.1, (język polski, "opisowe"); 1.0, (geografia, "opisowe"); 1.0\}.$$

$$\tilde{R}_{b_i = "inne"} = \{(wychowanie fizyczne, "inne"); 1.0\}$$

Określimy teraz zbiór rozmyty uczniów, których ulubiony przedmiot należy do nauk dedukcyjnych i nauk opisowych. W tym celu dokonamy złożenia dwóch rozpatrywanych relacji rozmytych \tilde{R}_{a_i} oraz \tilde{R}_{b_i} , otrzymując złożenie które można zapisać w postaci zbioru rozmytego

$$\tilde{R}_{a_i} \circ \tilde{R}_{b_i} = \tilde{R}_{a_i} \circ \tilde{R}_{b_i}$$

0	0.8	0.7	0.1
1.0	0	0	0.2
0	0	0	1.0

 \circ

1.0	0.1	0
0	1.0	0
0.1	1.0	0
0	0	1.0

 $=$

0.1	0.8	0.1
1.0	0.1	0.2
0	0	1.0

Licząc kolejno, otrzymujemy

$$\mu_{\tilde{R}_{a_i} \circ \tilde{R}_{b_i}}(u_1, "dedukcyjne") = \max(\min(0, 1.0), \min(0.8, 0), \min(0.7, 0.1), \min(0.1, 0)) = 0.1$$

pozostałe elementy liczymy podobnie, otrzymujemy elementy zbioru rozmytego $\tilde{R}_{a_i} \circ \tilde{R}_{b_i}$,

$$\tilde{R}_{a_i} \circ \tilde{R}_{b_i} = \{(u_1, "dedukcyjne"); 0.1, (u_1, "opisowe"); 0.8, (u_1, "inne"); 0.1, (u_2, "dedukcyjne"); 1.0, (u_2, "opisowe"); 0.1, (u_2, "inne"); 0.2, (u_3, "inne"); 1.0\}.$$

□

2.2.8. Wnioskowanie rozmyte

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

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

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

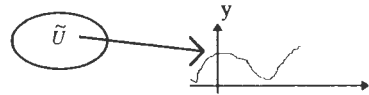
gdzie \tilde{U} i \tilde{V} są zbiorami rozmytymi.



- **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 *książka jest mała* TO *jest lekka*” oparta jest na dwóch zbiorach rozmytych: \tilde{U} : ”Mała książka” oraz \tilde{V} : ”Lekka książka” w zbiorze książek U , opisanych za pomocą dwóch cech: wielkość i waga. Reguła ta może być utożsamiana z relacją rozmytą określoną na iloczynie kartezjańskim zbiorów rozmytych $\tilde{U} \times \tilde{V}$ i zgodnie z definicją 2.34 może być zapisana w postaci

$$\tilde{R} = \{(u, v); t(\mu_{\tilde{U}}(u), \mu_{\tilde{V}}(v)) \mid \forall u \in \tilde{U}, \forall v \in \tilde{V}\} \quad (2.36)$$

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 $(\mu_{\tilde{U}}(u), \mu_{\tilde{V}}(u))$, $\forall u \in U$.

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 2.13. Niech zbiór U będzie zbiorem trzech uczniów, $U = \{u_1, u_2, u_3\}$. Zdefiniujmy rozmyty zbiór uczniów w zbiorze U , których ulubionym przedmiotem szkolnym jest matematyka $\tilde{U} = \{u_1; 0.8, u_2; 0.7, u_3; 0.1\}$ oraz rozmyty zbiór uczniów, których ulubionym przedmiotem jest język polski $\tilde{V} = \{u_1; 0.0, u_2; 0.5, u_3; 0.9\}$. Dana jest reguła rozmyta JEŻELI $u \in \tilde{U}$ TO $u \in \tilde{V}$, którą można zapisać słownie JEŻELI ulubionym przedmiotem szkolnym ucznia jest matematyka TO ulubionym przedmiotem tego ucznia jest też język polski. Sprawdźmy, w jakim stopniu spełniają tą regułę rozmytą trzej uczniowie ze zbioru U , w zależności od przyjętego do realizacji tej reguły operatora relacji.

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

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

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

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

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.

2.2.9. 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-Assiliania. Można wyróżnić trzy typy reguł rozmytych w zależności od typu konkluzji reguły.

1. Reguły rozmyte z jednoznaczną klasyfikacją.

$$\text{JEŻELI ... TO } [a_d = v_i^{a_d}] \quad (2.37)$$

2. Reguły rozmyte z jednoznaczną klasyfikacją i stopniem przynależności do klasy.

$$R_k: \text{JEŻELI ... TO } [a_d = v_i^{a_d}]; q_i^{a_d} \quad (2.38)$$

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

2. Reguły rozmyte z klasyfikacją o stopniach przynależności do wszystkich klas

$$R_k: \text{JEŻELI ... TO } [a_d = v_i^{a_d}]; q_i^{a_d} \wedge \dots \wedge [a_d = v_{k_d}^{a_d}]; q_{k_d}^{a_d} \quad (2.39)$$

gdzie stopień pewności $q_i^{a_d} \in [0,1]$.

2.3. Tworzenie reguł rozmytych na podstawie zbioru przykładów

2.3.1. Wprowadzenie

Załóżmy, 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, niepełne, wieloznaczne i mogą być obarczone błędami. Zbiór takich przykładów jest punktem wyjścia do zastosowania metod uczenia maszynowego.

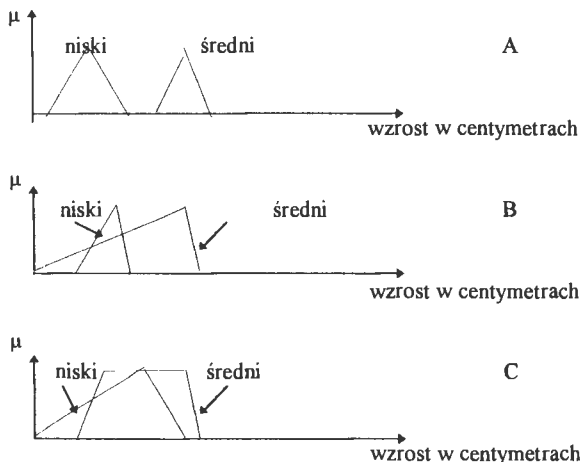
Aby zapisać takie przybliżone i niejednoznaczne dane w języku zrozumiałym dla systemu komputerowego i móc dalej je przetwarzać, można zastosować teorię zbiorów rozmytych. Proces wprowadzania w tym celu zmiennych lingwistycznych opisujących przykłady, wyznaczania nazw zbiorów rozmytych będących wynikiem 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 ludzi U . Niech naszym zadaniem będzie zdefiniowanie zbioru rozmytego "wysoki" człowiek, poprzez określenie funkcji przynależności $\mu_{\text{"wysoki"}}(u)$, $u \in U$. Należy uwzględnić, że funkcja $\mu_{\text{"wysoki"}}(u)$ będzie różna, jeżeli u oznacza kobietę lub

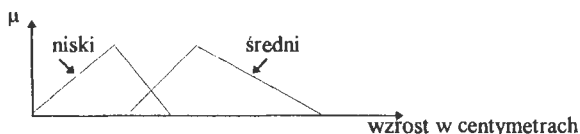
mężczyzną, a także gdy oznacza Aborygena lub też Szweda o takim samym wzroście. Różni ludzie mogą w różny, subiektywny sposób określać postać tej funkcji. W zależności od zadania które mamy do rozwiązania, jej postać również może się znacznie różnić; inna będzie przy wyborze graczy w piłkę koszykową, a inna przy wyborze młodzieży do sekcji gimnastycznej.

Jeszcze więcej problemów pojawia się, gdy chcemy określić sekwencję zbiorów rozmytych, przykładowo, "niski", "średni", "wysoki" człowiek. W takim przypadku należy uwzględnić intuicyjnie zrozumiałe zależności istniejące pomiędzy tymi zbiorami, a pewne zależności pomiędzy nimi nie mogą być akceptowalne, jak pokazano na rys. 2.10 A, B, C.



Rys. 2.10. Nie akceptowalne funkcje przynależności

Przykładowo, intuicyjnie akceptowalną zależność pomiędzy rozpatrywanymi zbiorami rozmytych przedstawiono na rys. 2.11.



Rys. 2.11. 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 uśredniać; 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 określane.

Funkcję przynależności można też estymować w sposób automatyczny, na podstawie zbioru danych przykładów. W pracy [Medasani i inni, 1998] przedstawiono metody stosowane

w tym celu, między innymi metody heurystyczne, probabilistyczne, metody stosujące techniki najbliższych sąsiadów, grupujące lub też oparte na sieciach neuronowych.

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.

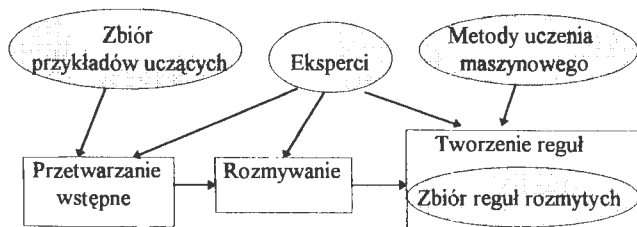
Przykład 2.14. Rozpatrzmy zbiór temperatur z przedziału $U=[15^{\circ}\text{C}, 40^{\circ}\text{C}]$ oraz cechę "temperatura" ze zbiorem wartości lingwistycznych $V_{\text{temperatura}} = \{\text{ciepło, gorąco}\}$, rys. 2.1.

Każda wartość lingwistyczna jest reprezentowana w postaci zbioru rozmytego w zbiorze wartości danej cechy. Wartość temperatury 28°C w procesie rozmywania jest zastępowana przez dwie pary (ciepło/0.4), (gorąco/0.6).

□

Tworzenie reguł rozmytych

W pracy zajmujemy się tworzeniem opisów klas przykładów podobnych w postaci zbioru reguł rozmytych typu JEŻELI ...TO, z rozmytymi przesłankami i rozmytymi konkluzjami. 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, schemat ich tworzenia przedstawiono na rys. 2.12.



Rys. 2.12. Tworzenie reguł rozmytych

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.

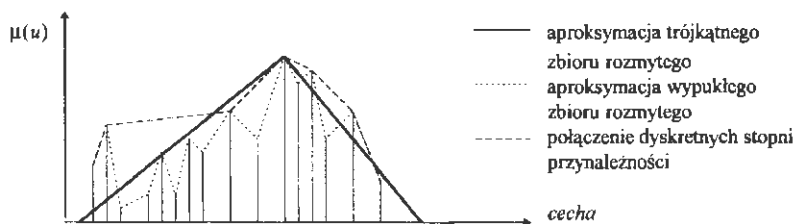
Heurystyczne metody uczenia maszynowego oparte na zbiorach rozmytych można podzielić na dwie główne grupy. Metody należące do pierwszej grupy same dokonują

grupowania przykładów w klasy przykładów podobnych (zwanymi klastrami) i następnie wykorzystują je przy tworzeniu rozmytych reguł klasyfikujących do tych klas. Metody należące do drugiej grupy nie dokonują grupowania przykładów, wymagają one na starcie zadane go rozmytego 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 [Klawonn 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, rys. 2.13.



Rys. 2.13. Tworzenie zbioru rozmytego poprzez projekcje stopni przynależności

Dla każdej cechy mamy więc wiele różnych zbiorów rozmytych tworzących klasy. 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 i Huber, 1999], [Tschichold Gurman, 1996].

Kratowy klastering (ang. grid clustering) [Klawonn i Keller, 1997], [Keller i 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. Klasy 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 klaster.

Pedrycz i Card zaproponowali sposób lingwistycznej interpretacji samoorganizującej się sieci Kohonena [Kohonen, 1984] w celu utworzenia reguł rozmytych [Pedrycz i 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 stworzonymi 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 [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 i Mendel, 1992] zawarty jest dowód, że algorytm może stworzyć zbiór reguł rozmytych które przybliżają każdą ciągłą funkcję rzeczywistą w zbiorze gęstym zadaną dokładnością.

Higgins i Goodman zaproponowali w pracy [Higgins i 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.

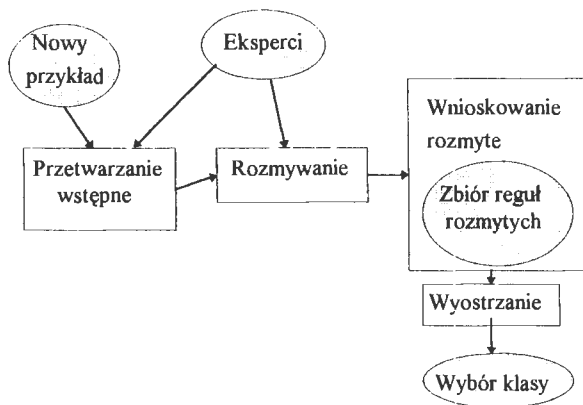
Podejścia zaproponowane przez Wanga i Mendela [Wang i Mendel, 1992] zostały rozszerzone i zastosowane w systemie NEFCLASS [Nauck i Kruse, 1995, 1997b, 1998b] dla zadania klasyfikacji oraz w systemie NEFPROX [Nauck i Kruse, 1997a, 1998c, 1999a] dla zadania aproksymacji funkcji. Wszystkie wartości cech 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 i 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 [Boyer i Wehenkel, 1999], [Ichihashi i inni, 1996], [Janikow, 1996, 1998], [Yuan i 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, spełniającego koniunkcję warunków z rozmytej części przesłankowej reguły do odpowiedniej klasy, określonej poprzez rozmyte konkluzje. W sytuacjach niejednoznacznych, gdy wymagana jest tylko jedna klasyfikacja, musimy przeprowadzić operację nazywaną w literaturze *wyostrzaniem* (ang. 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 rys. 2.14 przedstawiono schemat postępowania przy próbie klasyfikacji nowego przykładu do jednej z rozpatrywanych klas, z zastosowaniem danego zbioru reguł rozmytych.



Rys. 2.14. Klasyfikacja nowych przykładów

2.3.2. Postawienie zadania uczenia maszynowego

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

$$\tilde{R}_{a_j} = \{(u, v); \mu_{\tilde{R}_{a_j}}(u, v) \mid \forall u \in U, \forall v \in V_{a_j}\} \quad (2.40)$$

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 $v^{a_j} \in V_{a_j}$,

$$\begin{bmatrix} \mu_{\tilde{R}_{a_j}}(u_1, v_1^{a_j}) & \mu_{\tilde{R}_{a_j}}(u_1, v_2^{a_j}) & \dots & \mu_{\tilde{R}_{a_j}}(u_1, v_{L_j}^{a_j}) \\ \dots & \dots & \dots & \dots \\ \mu_{\tilde{R}_{a_j}}(u_N, v_1^{a_j}) & \mu_{\tilde{R}_{a_j}}(u_N, v_2^{a_j}) & \dots & \mu_{\tilde{R}_{a_j}}(u_N, v_{L_j}^{a_j}) \end{bmatrix}$$

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 2.6. 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 $v^{a_j} \in V_{a_j}$, oznaczanym $\tilde{U}_{a_j=v^{a_j}}$,

nazywamy podzbiór relacji rozmytej \tilde{R}_{a_j} ze zbioru U do V_{a_j} dla ustalonej wartości $v_i^{a_j} \in V_{a_j}$, zapisujemy 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 2.15. Niech U będzie zbiorem ludzi. Cecha "wiek" ze zbiorem wartości: $\{\text{młody, średni, stary}\}$ jest zmienną lingwistyczną. Zbiór rozmyty w zbiorze U wyznaczony przez warunek "wiek" = *młody* zawiera ludzi, dla których cecha "wiek" przyjmuje wartość *młody* w stopniu większym od zera; można go zapisać w postaci

$$\tilde{U}_{\text{wiek}=\text{młody}} = \{(u, \text{młody}); \mu_{\tilde{R}_{\text{wiek}=\text{młody}}}(u, \text{młody}) \mid u \in U, \mu_{\tilde{R}_{\text{wiek}=\text{młody}}}(u, \text{młody}) > 0\}.$$

□

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

Definicja 2.7. Podział rozmyty zbioru przykładów

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

$$\bullet \quad t(\mu_{\tilde{R}_{a_d=v_{i_1}^{a_d}}}(u, v_{i_1}^{a_d}), \mu_{\tilde{R}_{a_d=v_{i_2}^{a_d}}}(u, v_{i_2}^{a_d})) = 0, \quad \text{dla } i_1 \neq i_2, \forall u \in U \quad (2.41)$$

$$\bullet \quad s(\mu_{\tilde{R}_{a_d=v_1^{a_d}}}(u, v_1^{a_d}), \dots, \mu_{\tilde{R}_{a_d=v_{k_d}^{a_d}}}(u, v_{k_d}^{a_d})) = 1, \quad \text{dla } \forall u \in U. \quad (2.42)$$

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 (2.41) i (2.42), dokonuje się modyfikacji zbioru cech A wybranych do opisu przykładów, wprowadzając nowe cechy nominalne.

Tak określony zbiór rozmyty $\tilde{U}_{a_d=v_d^g}$ 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 pomijając 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

$$u_n : [a_1 = v_{t(1,n)}^{a_1}]; \mu_{a_1=v_{t(1,n)}^{a_1}} \wedge [a_2 = v_{t(2,n)}^{a_2}]; \mu_{a_2=v_{t(2,n)}^{a_2}} \wedge \dots \wedge [a_K = v_{t(K,n)}^{a_K}]; \mu_{a_K=v_{t(K,n)}^{a_K}} \quad (2.43)$$

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

$$u_n : [a_1 = v_{t(1,n)}]; \mu_{v_{t(1,n)}} \wedge [a_2 = v_{t(2,n)}]; \mu_{v_{t(2,n)}} \wedge \dots \wedge [a_K = v_{t(K,n)}]; \mu_{v_{t(K,n)}} \quad (2.44)$$

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

$$\text{JEŻELI } [a_j = v_{t(j,k)}] \wedge [a_2 = v_{t(2,k)}] \wedge \dots \wedge [a_K = v_{t(K,k)}] \text{ TO } [a_d = v_{t(d,k)}] \quad (2.45)$$

gdzie indeks $t(j,k)$ 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 $\mu_{t(j,k,n)}$ oznacza funkcję przynależności j -tej cechy w n -tym przykładzie do zbioru rozmytego określonego przez warunek $[a_j = v_{t(j,k)}]$ 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: $\mu_{t(1,k,n)}, \mu_{t(2,k,n)}, \dots, \mu_{t(K,k,n)}$ oraz do zbioru określonego przez konkluzję: $\mu_{t(d,k,n)}$.

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

$$\tau_k(u_n) = \mathcal{A}^t (\mu_{t(1,k,n)}, \mu_{t(2,k,n)}, \dots, \mu_{t(K,k,n)}), \quad (2.46)$$

gdzie \mathcal{A} jest wybranym operatorem koniunkcyjnym agregacji, zazwyczaj minimum.

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

$$\mu_{\tilde{c}_k}(u_n) = \min\{\tau_k(u_n), \mu_{\tau(d,k,n)}\}. \quad (2.47)$$

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

$$\mu_{\tilde{C}}(u) = \mathcal{A}^s(\mu_{\tilde{c}_1}(u), \mu_{\tilde{c}_2}(u), \dots, \mu_{\tilde{c}_s}(u)), \quad (2.48)$$

gdzie \mathcal{A} jest wybranym operatorem dysjunkcyjnym agregacji, zazwyczaj maksimum.

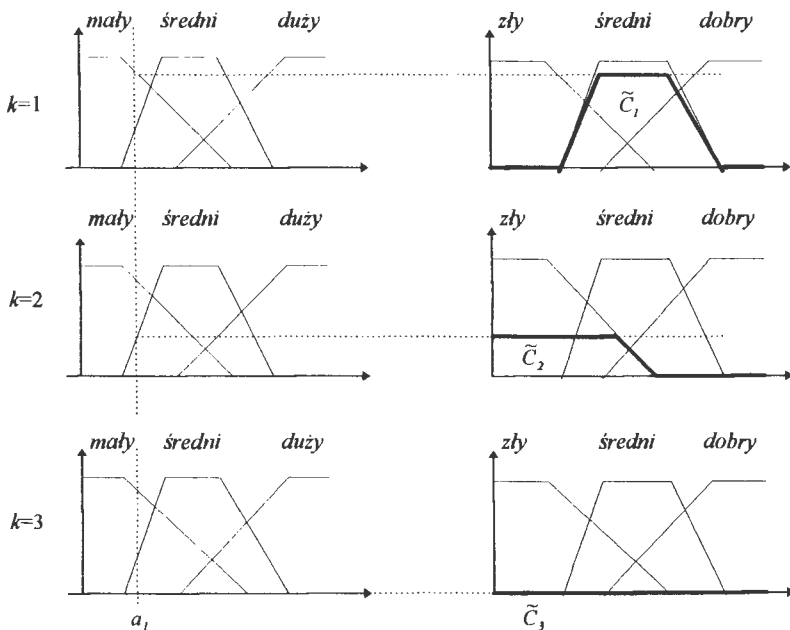
Przykład 2.16. Załóżmy, że w procesie rozmywania przy opisywaniu danego zbioru przykładów określono zbiory rozmyte przedstawione na rys. 2.15. Przyjmijmy, że mamy zbiór trzech reguł, $k=3$, przedstawionych poniżej

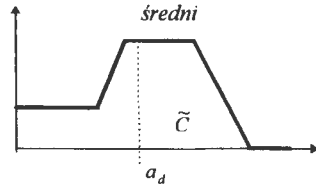
$k=1$ JEŻELI [$a_i = \text{mały}$] TO [$a_d = \text{średni}$]

$k=2$ JEŻELI [$a_i = \text{średni}$] TO [$a_d = \text{zły}$]

$k=3$ JEŻELI [$a_i = \text{duży}$] TO [$a_d = \text{dobry}$]

Zastosujemy ten zbiór reguł w celu określenia przynależności nowego przykładu $u_i: [a_i = v_{i(1,1)}]$, zaznaczonego na rys. 2.15. Klasy rozmyte \tilde{C}_k określone przez rozmyte konkluzje k -tej reguły dla przykładu u_i zostały zaznaczone pogrubioną linią na rysunku ($\tilde{C}_3 = \emptyset$, 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 . Jedna wartość najbardziej reprezentatywna dla zbioru rozmytego \tilde{C} jaką przyjmuje cecha a_d może być określona w procesie wyostrzania.





Rys 2.15. Klasyfikacja przykładu U z zastosowaniem zbioru trzech reguł



Często stosowane metody wyostrzania to metoda środka obszaru COA (ang. Center of Area) oraz metoda średniej maksymalnej MOM (ang. Mean of Maximum).

Przykład 2.17. Rozpatrzmy zbiór przykładów opisanych za pomocą trzech cech: cechy numerycznej a_1 , $V_{a_1} \subseteq R$, cechy symbolicznej a_2 o wartościach $\{A, B, C\}$ oraz cechy decyzyjnej a_d o wartościach $\{KLASA, \text{nie KLASA}\}$. W zbiorze wartości jakie może przyjąć cecha numeryczna a_1 , określamy trzy zbiory rozmyte $\{\text{mały, średni, duży}\}$ zdefiniowane przez trójkątne funkcje przynależności.

Założmy, że mamy zbiór 17 przykładów należących do KLASY (oznaczanych znakiem +) oraz 11 przykładów nie należących do KLASY (oznaczanych znakiem -), rys.2.16.

	a_2				
		C	+++++ -	- - - -	- + + + + +
		B	- + - + -	+ +	
		A	+ + - +	-	
			mały	średni	duży
					a_1

Rys. 2.16. Zbiór przykładów testowych opisanych numeryczną i symboliczną cechą

Szukamy zbioru reguł rozmytych klasyfikujących przykłady do KLASY i zbioru reguł które określają, że przykład nie należy do KLASY. Możliwe są następujące typy reguł, dla $v_i^{a_1} \in \{\text{mały, średni, duży}\}$, $V_{a_2} = \{A, B, C\}$, $v_i^{a_d} \in V_{a_d} = \{KLASA, \text{nie KLASA}\}$

- a) typ rozmyty - dokładny: JEŻELI $a_1 = v_i^{a_1}$ i $a_2 = A$ TO $a_d = v_i^{a_d}$
- b) typ rozmyty - niedokładny: JEŻELI $a_1 = v_i^{a_1}$ i $a_2 = \{A \text{ lub } B\}$ TO $a_d = v_i^{a_d}$
- c) typ rozmyty - rozmyty: JEŻELI $a_1 = v_i^{a_1}$ i $a_2 = \{(A, \mu_A), (B, \mu_B), (C, \mu_C)\}$ TO $a_d = v_i^{a_d}$

W regułach typu a) i b) cecha a_2 przyjmuje konkretną wartość. W regule typu c) stosujemy zbiór rozmyty do opisanie wartości, które może przyjmować cecha a_2 . Dwie pierwsze reguły możemy zapisać w postaci zgodnej z formalizmem reguły typu c) w następujący sposób:

- a) JEŻELI $a_1 = v_i^{a_1}$ i $a_2 = \{(A, 1.0), (B, 0.0), (C, 0.0)\}$ TO $a_d = v_i^{a_d}$
 b) JEŻELI $a_1 = v_i^{a_1}$ i $a_2 = \{(A, 1.0), (B, 1.0), (C, 0.0)\}$ TO $a_d = v_i^{a_d}$.

Ponizej przedstawiono ideę algorytmu tworzącego reguły rozmyte typu c) na podstawie zbioru przykładów uczących, przedstawionych na rys.2.2. Zgodnie z rysunkiem dysponujemy zbiorem 27 przykładów. W celu utworzenia na ich podstawie zbioru reguł rozmytych rozpatrzmy najpierw zbiór 10 przykładów mających wartość cechy a_1 =mały i należących do KLASY. W zbiorze tym 3 przykłady mają $a_2=A$, 2 przykłady mają $a_2=B$ oraz 5 przykładów ma $a_2=C$. Otrzymujemy zatem następujący zbiór rozmyty $\{(A, 0.3), (B, 0.2), (C, 0.5)\}$ a po procesie normalizacji zbiór $\{(A, 0.6), (B, 0.4), (C, 1.0)\}$. Otrzymujemy więc następującą regułę rozmytą

\tilde{R}_1 : JEŻELI a_1 =mały i $a_2 = \{(A, 0.6), (B, 0.4), (C, 1.0)\}$ TO a_d =KLASA

Rozpatrzmy pozostałe 5 przykładów nie należących do KLASY, dla których 1 przykład ma $a_2=A$, 3 przykłady mają $a_2=B$ oraz 1 przykład ma $a_2=C$. Otrzymujemy po procesie normalizacji następujący zbiór rozmyty $\{(A, 0.33), (B, 1.0), (C, 0.33)\}$, i odpowiadającą mu regułę rozmytą:

\tilde{R}_2 : JEŻELI a_1 =mały i $a_2 = \{(A, 0.33), (B, 1.0), (C, 0.33)\}$ TO a_d =nie KLASA.

Reguły te reprezentują podobieństwa istniejące w zbiorze przykładów. Jeśli rozpatrzmy przykłady uczące należące do KLASY to przykłady mające wartość cechy a_1 =mały mogą mieć wartość cechy a_2 zarówno A, B jak też C, z tym, że wartość C jest bardziej charakterystyczna dla danego zbioru niż wartość A, a wartość A bardziej typowa niż wartość B. W zbiorze przykładów nie należących do KLASY najbardziej typowa jest wartość B. Utworzone reguły \tilde{R}_1 i \tilde{R}_2 częściowo się pokrywają. Wybór reguły przy klasyfikacji nowego przykładu zależeć będzie od określenia stopnia przynależności tego przykładu do zbiorów rozmytych z części przesłankowych tych reguł.

W podobny sposób przeanalizujemy pozostałe przykłady. Na podstawie 7 przykładów mających wartość cechy a_1 =średni otrzymujemy następujące reguły

\tilde{R}_3 : JEŻELI a_1 =średni i $a_2 = \{(A, 0.0), (B, 1.0), (C, 0.0)\}$ TO a_d =KLASA

\tilde{R}_4 : JEŻELI a_1 =średni i $a_2 = \{(A, 0.25), (B, 0.0), (C, 1.0)\}$ TO a_d =nie KLASA.

Wśród przykładów nie należących do KLASY bardziej typowa jest kombinacja wartości cech a_1 =średni i $a_2=C$, niż kombinacja wartości cech a_1 =średni i $a_2=A$. Przy klasyfikacji nowych przykładów z zastosowaniem powyższych reguł jesteśmy w stanie jednoznacznie zaklasyfikować przykłady mające wartość cechy $a_2=A$ lub C za pomocą reguły R_4 jako nie należące do KLASY, oraz mających wartość cechy $a_2=B$ za pomocą reguły R_3 do KLASY.

Na podstawie pozostałych 6 przykładów mających wartość cechy a_1 =duży otrzymujemy następujące reguły

\tilde{R}_5 : JEŻELI a_1 =duży i $a_2 = \{(A, 0.0), (B, 0.0), (C, 1.0)\}$ TO a_d =KLASA

\tilde{R}_6 : JEŻELI a_1 =duży i $a_2 = \{(A, 0.0), (B, 0.0), (C, 1.0)\}$ TO a_d =nie KLASA.

Powyższe reguły przy takich samych warunkach w częściach przesłankowych prowadzą do sprzecznych konkluzji, nie można więc zaakceptować obu reguł. Wybieramy

regułę spełnioną przez większą liczbę przykładów uczących, a więc regułę \tilde{R}_7 i dołączamy ją do tworzonego zbioru reguł.

Utworzyliśmy więc na podstawie zbioru 27 przykładów uczących zbiór pięciu reguł rozmytych, klasyfikujących do klas określonych poprzez wartości cechy decyzyjnej a_j .

Rozpatrzmy trzy przykłady testowe przedstawione poniżej, które będziemy chcieli klasyfikować z zastosowaniem zbioru reguł rozmytych $\tilde{R}_1, \tilde{R}_2, \tilde{R}_3, \tilde{R}_4, \tilde{R}_5$,

$$u_1: [a_1=\text{mały}]; \mu_{\text{mały}}(u_1) \wedge [a_2=A],$$

$$u_2: [a_1=\text{mały}]; \mu_{\text{mały}}(u_2) \wedge [a_2=B],$$

$$u_3: [a_1=\text{mały}]; \mu_{\text{mały}}(u_3) \wedge [a_2=C]$$

gdzie $\mu_{\text{mały}}(u_1) = \mu_{\text{mały}}(u_2) = \mu_{\text{mały}}(u_3) = 1.0$.

Aby sprawdzić, które reguły opisują przykłady testowe, należy obliczyć w jakim stopniu każdy przykład spełnia każdą regułę rozmytą. W tym celu, dla każdego przykładu testowego i każdej reguły rozmytej liczymy stopień przynależności przykładu do każdego zbioru rozmytego z części przesłankowej reguły i w przypadku koniunkcji dwóch warunków stosujemy funkcję minimum.

Dla przykładów u_1, u_2 i u_3 otrzymujemy przedstawione poniżej stopnie przynależności

$\tilde{R}_1: \mu_{\text{mały}^*}(u_1) = 1, \mu_{\{(A,0.6),(B,0.4),(C,1.0)\}}(A) = 0.6,$	$\mu_{\tilde{R}_1}(u_1) = 0.6$
$\tilde{R}_2: \mu_{\text{mały}^*}(u_1) = 1, \mu_{\{(A,0.33),(B,1.0),(C,0.33)\}}(A) = 0.33,$	$\mu_{\tilde{R}_2}(u_1) = 0.33$
$\tilde{R}_3: \mu_{\text{średni}^*}(u_1) = 0, \mu_{\{(A,0.0),(B,1.0),(C,0.0)\}}(A) = 0.0,$	$\mu_{\tilde{R}_3}(u_1) = 0.0$
$\tilde{R}_4: \mu_{\text{średni}^*}(u_1) = 0, \mu_{\{(A,0.25),(B,0.0),(C,1.0)\}}(A) = 0.25,$	$\mu_{\tilde{R}_4}(u_1) = 0.0$
$\tilde{R}_5: \mu_{\text{duży}^*}(u_1) = 0, \mu_{\{(A,0.0),(B,0.0),(C,1.0)\}}(A) = 0.0,$	$\mu_{\tilde{R}_5}(u_1) = 0.0$
$\tilde{R}_1: \mu_{\text{mały}^*}(u_2) = 1, \mu_{\{(A,0.6),(B,0.4),(C,1.0)\}}(B) = 0.4,$	$\mu_{\tilde{R}_1}(u_2) = 0.4$
$\tilde{R}_2: \mu_{\text{mały}^*}(u_2) = 1, \mu_{\{(A,0.33),(B,1.0),(C,0.33)\}}(B) = 1.0,$	$\mu_{\tilde{R}_2}(u_2) = 1.0$
$\tilde{R}_3: \mu_{\text{średni}^*}(u_2) = 0, \mu_{\{(A,0.0),(B,1.0),(C,0.0)\}}(B) = 1.0,$	$\mu_{\tilde{R}_3}(u_2) = 0.0$
$\tilde{R}_4: \mu_{\text{średni}^*}(u_2) = 0, \mu_{\{(A,0.25),(B,0.0),(C,1.0)\}}(B) = 0.0,$	$\mu_{\tilde{R}_4}(u_2) = 0.0$
$\tilde{R}_5: \mu_{\text{duży}^*}(u_2) = 0, \mu_{\{(A,0.0),(B,0.0),(C,1.0)\}}(B) = 0.0,$	$\mu_{\tilde{R}_5}(u_2) = 0.0$
$\tilde{R}_1: \mu_{\text{mały}^*}(u_3) = 1, \mu_{\{(A,0.6),(B,0.4),(C,1.0)\}}(C) = 1.0,$	$\mu_{\tilde{R}_1}(u_3) = 1.0$
$\tilde{R}_2: \mu_{\text{mały}^*}(u_3) = 1, \mu_{\{(A,0.33),(B,1.0),(C,0.33)\}}(C) = 0.33,$	$\mu_{\tilde{R}_2}(u_3) = 0.33$
$\tilde{R}_3: \mu_{\text{średni}^*}(u_3) = 0, \mu_{\{(A,0.0),(B,1.0),(C,0.0)\}}(C) = 0.0,$	$\mu_{\tilde{R}_3}(u_3) = 0.0$
$\tilde{R}_4: \mu_{\text{średni}^*}(u_3) = 0, \mu_{\{(A,0.25),(B,0.0),(C,1.0)\}}(C) = 1.0,$	$\mu_{\tilde{R}_4}(u_3) = 0.0$
$\tilde{R}_5: \mu_{\text{duży}^*}(u_3) = 0, \mu_{\{(A,0.0),(B,0.0),(C,1.0)\}}(C) = 1.0.$	$\mu_{\tilde{R}_5}(u_3) = 0.0$

Rozpatrywane trzy przykłady testowe spełniają w stopniu większym od zera tylko dwie pierwsze reguły rozmyte, klasyfikujące do różnych klas. Aby dokonać jednoznacznej klasyfikacji przykładów dokonujemy agregacji dwóch zbiorów rozmytych tworzących konkluzje obu reguł, z zastosowaniem funkcji maksimum. W tabl.2.2 przedstawiono stopnie spełnienia obu reguł przez przykłady testowe i ich zaklasyfikowanie do klasy o największym stopniu spełnienia.

Tabl. 2.2. Klasyfikacja trzech przykładów testowych z zastosowaniem reguł \tilde{R}_1 i \tilde{R}_2

przykłady	stopień spełnienia reguły \tilde{R}_1	stopień spełnienia reguły R_2	klasyfikacja
u_1	0.6	0.33	należy do KLASY
u_2	0.4	1.0	nie należy do KLASY
u_3	1.0	0.33	należy do KLASY

□

Zadanie praktycznego określania zbioru reguł rozmytych opisujących rzeczywisty zbiór przykładów nie jest zadaniem prostym.

Załóżmy, że na podstawie zbioru przykładów $u_n = 1, \dots, N$ zapisanych w postaci (2.35) szukamy S reguł klasyfikujących do klasy $\tilde{U}_{a_d=v_d^q}$ w postaci (2.37), dla ustalonego t

$$R_k: \text{JEŻELI } [a_1 = v_{i(1,k)}] \wedge [a_2 = v_{i(2,k)}] \wedge \dots \wedge [a_k = v_{i(k,k)}] \text{ TO } [a_d = v_d^q]$$

gdzie k określa numer reguły, $k = 1, \dots, S$. Próbę sformalizowania zadania szukania optymalnego zbioru reguł przedstawiono poniżej.

Niech $\tau_k(u_n)$ określa stopień spełnienia przez przykład u_n warunków z części przesłankowej k -tej reguły (2.46), $\mu_{i(d,k)}(u_n)$ określa stopień spełnienia przez przykład u_n warunku konkluzji w k -tej regule (2.47). Utworzone reguły powinny poprawnie opisywać przykłady uczące. Stopień spełnienia reguł przez przykłady należące do rozpatrywanej klasy powinien być jak największy, zaś stopień spełnienia reguł przez przykłady nie należące do klasy powinien być jak najmniejszy. Powyższe wymagania można próbować sformalizować w sposób przedstawiony poniżej.

Niech stopień spełnienia przez przykład u_n części przesłankowej k -tej reguły otrzymany w wyniku agregacji zbiorów rozmytych z części przesłankowej reguły wynosi zgodnie z (2.46)

$$\tau_k(u_n) = \mathcal{A}^t (\mu_{i(1,k,n)}, \mu_{i(2,k,n)}, \dots, \mu_{i(k,k,n)}),$$

gdzie \mathcal{A} jest wybranym operatorem koniunkcyjnym agregacji, zazwyczaj minimum. Aby zagwarantować wymaganie, że większość przykładów należących do klasy spełniała reguły należy maksymalizować wyrażenie

$$Q^+ = \sum_{n=1}^N \left(\max_{k=1, \dots, S} \{ \min\{\tau_k(u_n), \mu_{i(d,k,n)}(u_n)\} \} \right) / \sum_{n=1}^N \mu_{i(d)}(u_n) \quad (2.49)$$

Aby zagwarantować wymaganie, że prawie żaden z przykładów nie należących do klasy nie spełnia reguł należy minimalizować wyrażenie

$$Q^- = \sum_{n=1}^N \left(\max_{k=1, \dots, S} \{ \min\{\tau_k(u_n), (1 - \mu_{i(d,k,n)}(u_n))\} \} \right) / \sum_{n=1}^N (1 - \mu_{i(d)}(u_n)) \quad (2.50)$$

Zadanie znalezienia zbioru S reguł opisujących rozpatrywaną klasę $\tilde{U}_{a_d=v_d^q}$ na podstawie zbioru przykładów $u_n, n = 1, \dots, N$ można zapisać w następujący sposób

$$Q^+ = \sum_{n=1}^N \left(\max_{k=1, \dots, S} \{ \min\{\tau_k(u_n), \mu_{i(d,k,n)}(u_n)\} \} \right) / \sum_{n=1}^N \mu_{i(d)}(u_n) \rightarrow \max_{R_k, k=1, \dots, S}$$

$$Q^- = \sum_{n=1}^N \left(\max_{k=1, \dots, S} \{ \min\{\tau_k(u_n), (1 - \mu_{i(d,k,n)}(u_n))\} \} \right) / \sum_{n=1}^N (1 - \mu_{i(d)}(u_n)) \rightarrow \min_{R_k, k=1, \dots, S}$$

Przykład 2.18. Rozpatrzmy zbiór pięciu przykładów opisanych za pomocą cechy "ból głowy" oraz cechy decyzyjnej "grypa", przedstawionych poniżej

$$\begin{aligned}
u_1: & [ból\ głowy = mocny];1,0 \wedge [ból\ głowy = słaby];0,0 \wedge [grypa = tak];0,9 \\
u_2: & [ból\ głowy = mocny];0,9 \wedge [ból\ głowy = słaby];0,1 \wedge [grypa = tak];0,8 \\
u_3: & [ból\ głowy = mocny];0,2 \wedge [ból\ głowy = słaby];0,7 \wedge [grypa = tak];0,0 \\
u_4: & [ból\ głowy = mocny];0,1 \wedge [ból\ głowy = słaby];0,9 \wedge [grypa = tak];0,1 \\
u_5: & [ból\ głowy = mocny];0,0 \wedge [ból\ głowy = słaby];1,0 \wedge [grypa = tak];0,1
\end{aligned}$$

Sprawdźmy, która z reguł przedstawionych poniżej

$$R_1: \text{JEŻELI } [ból\ głowy = mocny] \text{ TO } [grypa = tak] ,$$

$$R_2: \text{JEŻELI } [ból\ głowy = słaby] \text{ TO } [grypa = tak] .$$

poprawnie opisuje dany zbiór pięciu przykładów. Policzmy dla zbioru przykładów sumę ich stopni przynależności do klasy chorych na grypę i do klasy nie chorujących na grypę.

$$\sum_{n=1}^5 \mu_{r(d)}(u_n) = 0.9 + 0.8 + 0.0 + 0.1 + 0.1 = 1.9,$$

$$\sum_{n=1}^5 (1 - \mu_{r(d)}(u_n)) = 0.1 + 0.2 + 1.0 + 0.9 + 0.9 = 3.1.$$

Stosując wzory (2.49) i (2.50) policzymy poprawność pierwszej reguły, że ludzie którzy mają mocny ból głowy mają grypę.

$$\begin{aligned}
Q^+ &= \sum_{n=1}^5 \left(\min\{\tau_1(u_n), \mu_{r(d,1,n)}(u_n)\} \right) / \sum_{n=1}^5 \mu_{r(d)}(u_n) = (0.9 + 0.8 + 0.0 + 0.1 + 0.0) / 1.9 \approx 0.9 \\
Q^- &= \sum_{n=1}^5 \left(\min\{\tau_1(u_n), (1 - \mu_{r(d,1,n)}(u_n))\} \right) / \sum_{n=1}^5 (1 - \mu_{r(d)}(u_n)) = (0.1 + 0.2 + 0.2 + 0.1 + 0) / 3.1 \approx 0.2.
\end{aligned}$$

Policzymy poprawność drugiej reguły, że ludzie którzy mają słaby ból głowy mają grypę.

$$\begin{aligned}
Q^+ &= \sum_{n=1}^5 \left(\min\{\tau_2(u_n), \mu_{r(d,2,n)}(u_n)\} \right) / \sum_{n=1}^5 \mu_{r(d)}(u_n) = (0.0 + 0.1 + 0.0 + 0.1 + 0.1) / 1.9 \approx 0.2 \\
Q^- &= \sum_{n=1}^5 \left(\min\{\tau_2(u_n), (1 - \mu_{r(d,2,n)}(u_n))\} \right) / \sum_{n=1}^5 (1 - \mu_{r(d)}(u_n)) = (0.0 + 0.1 + 0.7 + 0.9 + 0.9) / 3.1 \approx 0.8.
\end{aligned}$$

Ponieważ dążymy do tego, aby maksymalizować Q^+ a minimalizować Q^- , reguła pierwsza jest zdecydowanie lepiej dopasowana do danego zbioru pięciu przykładów.

Literatura

1. T. Bilgic i I. Turksen. Elicitation of membership functions: How far can they take us? W: Proc. FUZZ/IEEE, 1321-1325, Barcelona, Spain, 1997.
2. T. Bilgic i I. Turksen. International Handbook of Fuzzy Systems, 1, 1999.
3. D. Dubois i H. Prade. The three semantics of fuzzy sets. Fuzzy Sets and Systems, 90, 141-150, 1997.
4. D. Dubois i H. Prade. Operations on fuzzy numbers. Int. J. Syst. Sci. Vol.9, 613-626, 1978.
5. J.M. Keller i J.A. Givens. Membership function issues in fuzzy pattern recognition. W: International Conference on Cybernetics and Society, 210-214, 1985.
6. L.A. Zadeh. Fuzzy Sets. *Information and Control*. 8, 338-353, 1965.
7. L.A. Zadeh. 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, 1973.



...the first of these is the fact that the ...

...the second of these is the fact that the ...

...the third of these is the fact that the ...

...the fourth of these is the fact that the ...

...the fifth of these is the fact that the ...

...the sixth of these is the fact that the ...

...the seventh of these is the fact that the ...

...the eighth of these is the fact that the ...

...the ninth of these is the fact that the ...

