AUTOMATYKA • 2006 • Tom 10 • Zeszyt 3

Adam Niewiadomski*

Rozmyte metody inteligentnej interpretacji danych

1. Wstęp

Można powiedzieć, iż od początków istnienia informatyki człowiek odczuwał spory dyskomfort, komunikując się z komputerem jedynie za pomocą liczb i sztucznych języków komend. Obserwowany współcześnie szeroki i coraz łatwiejszy dostęp do informacji wywiera na badaczach dużą presję: są oni niejako zobligowani zapewnić przeciętnemu użytkownikowi oprogramowanie automatyzujące proces pozyskiwania, filtrowania i, ostatecznie, prezentacji wyszukanej informacji w kompaktowej i komunikatywnej formie. Niemniej, naturalny język jest ze swej natury nieprecyzyjny i niejednoznaczny, co powoduje, iż reprezentowanie jego jednostek w terminach klasycznej logiki i teorii zbiorów jest utrudnione i niedoskonałe. Reprezentacja licznych szczegółów, jak i zarówno ogólnej natury tylko częściowo ustrukturyzowanego języka naturalnego, nie daje się oprzeć jedynie o mechanizmy logiki dwuwartościowej. Dlatego też tzw. inteligentne procedury przetwarzania informacji, spośród których praca ta wymienia *retranslację danych numerycznych*, czyli odtwarzanie ich w postaci lingwistycznej, oraz *lingwistyczne podsumowanie danych*, czyli generowanie słownych opisów dotyczących ilości obiektów w zbiorach danych oraz posiadanych przez nie cech, winny oparte być o algorytmy i techniki obliczeniowe mające swe źródło w teorii zbiorów rozmytych i jej wybranych rozszerzeniach.

Metody retranslacji i podsumowywania danych ujęte są w opracowaniu wspólnym terminem *interpretacja danych*, co oznaczać ma konwertowanie informacji z postaci numerycznej do lingwistycznej lub, w szerszym sensie, do form zrozumiałych dla przeciętnego użytkownika. Metody te znajdują lub znaleźć mogą zastosowanie głównie w systemach i aplikacjach sieciowych, zwłaszcza internetowych, jednakże abstrahujemy tu od struktur i źródeł informacji, jak również od metod ich transferu, skupiamy się natomiast na jej semantyce, po to aby nie ograniczać proponowanych rozwiązań. Z tych samych powodów rozważamy tu głównie przetwarzanie informacji numerycznych i prostych danych tekstowych, jako że są one najłatwiejsze do przechowywania i przesyłania. Przedstawione metody dopasowane są do relacyjnego modelu baz danych, niemniej mogą one być implementowane także dla innych modeli danych, np. obiektowych lub rozproszonych.

* Katedra Inżynierii Wiedzy i Inteligencji Komputerowej, Wyższa Szkoła Humanistyczno-Ekonomiczna w Łodzi; Instytut Informatyki, Politechnika Łódzka

Praca niniejsza ma także na celu popularyzację polskojęzycznego słownictwa w prezentowanej dziedzinie. Ze względu na fakt, iż zdecydowana większość publikacji na ten temat jest w języku angielskim, autor pragnie poddać Czytelnikom pod dyskusję swoje propozycje tłumaczeń kluczowych dla problematyki terminów.

2. Interpretacja danych

Automatyczna i inteligentna interpretacja danych może być przydatna w wielu dziedzinach, np. bankowość czy handel, jak również w naukowych i inżynieryjnych zastosowaniach, jako narzędzie wspierające analizę pozyskanych zbiorów danych. Przybliżmy termin *„interpretacja”* cytatem z pracy [9] (tłum. autor):

> (...) *gdy próbujemy zakwalifikować zjawiska recypowane przez człowieka, często jesteśmy skłonieni do używania słów, zamiast wartości liczbowych*

Pod pojęciem *interpretowania* pewnych faktów, wyników pomiarów, obserwacji czy percepcji (spostrzeżeń), zwykle rozumiemy *opisywanie* lub *objaśnianie*, *tłumaczenie* ich znaczenia i sensu. W tym ujęciu interpretacja podobna jest do *translacji*, jednakże posiada szerszy zakres, gdyż nie oznacza tylko przekładania informacji z jednego języka na drugi, ale także *konkludowanie* pewnych faktów wynikających ze zgromadzonych danych, jak również na *dostarczaniu kontekstu* dla tych faktów. Ponadto *zinterpretować dane* może oznaczać także wyrażanie *explicite* informacji, które są zawarte *implicite* lub ukryte czy uwikłane. Dlatego też, kolokwialnie mówiąc, automatyczne interpretowanie danych jest w stanie udzielić odpowiedzi na pytanie: *Co otrzymane dane w ogólności i w szczegółach znaczą*?

Interpretacja danych nie jest procesem jednorodnym, homogenicznym. Konieczne okazuje się wydzielenie w jej ramach przynajmniej dwóch „podprocesów”. Podział ten ma na celu, między innymi, ułatwienie zastosowań i implementacji opisanych metod. Dlatego też pod pojęciem interpretacji ukryte są procesy retranslacji i podsumowywania danych:

$$\text{retranslacja} \cup \text{podsumowywanie} \subseteq \text{interpretacja} \tag{1}$$

Intuicje ukryte w terminach „retranslacja” i „podsumowywanie” danych przybliża następujące objaśnienie. Część wiedzy człowieka może być pamiętana i przetwarzana w postaci liczb: wyników pomiarów lub obliczeń, np. „obwód Ziemi to 40 tys. km”, lub „2·2 = 4”. Jednakże w wielu innych obszarach fakty i percepcje wyrażone są za pomocą określeń jakościowych (słów), dla których dobór ekwiwalentów ilościowych (liczb) może być trudny lub sztuczny, np. dla określenia „ładna i inteligentna kobieta, która lubi matematykę”. Słowny, czyli językowy, lingwistyczny, „zapis” wiedzy wystarcza na potrzeby porozumiewania się ludzi pomiędzy sobą, jednak staje się kłopotliwy, jeśli zgromadzona wiedza ma być przechowywana i przetwarzana elektronicznie. Współczesne systemy informacyjne, pomimo ogromnych i różnorodnych możliwości, w większości umożliwiają przechowywanie danych w formie liczb lub krótkich tekstów, głównie ze względu na relatywnie małą objętość, a co za tym idzie, także na krótki czas dostępu i/lub przesyłu (do) tych danych. To uwarunkowanie prowadzi często do konieczności symbolicznego i/lub wybiórczego zapisu cech w postaci liczbowej, np. poziomu inteligencji za pomocą współczynnika IQ, który nie oddaje wszystkich aspektów pojęcia inteligencji człowieka.

Dlatego też duże ilości danych w bazach, aby mogły zostać przedstawione użytkownikowi w czytelnej i zrozumiałej formie, winny zostać poddane retranslacji i/lub podsumowane lingwistycznie. W zależności od bieżącego zapotrzebowania użytkownika tylko jeden z tych dwóch procesów lub oba będą konieczne do przeprowadzenia. Do przetwarzania danych na jednostki lingwistyczne w obu przypadkach użyte zostaną elementy teorii zbiorów rozmytych.

W literaturze znane są dziedziny pokrewne, obejmujące zakres zagadnień podobny do tych, jakie obejmuje interpretacja. W szczególności wymienia się tu:

– Semantic Web [17, 33, 47, 46],
– Question-Answering Systems (Q/A systems) [46],
– Flexible Querying [14, 18],
– Computing with Words [47, 15].

Rezultaty i wyniki działania procedur interpretujących dane mogą także posłużyć jako punkt wyjścia dla procesu *wnioskowania* na podstawie zgromadzonych danych. Niektóre źródła, np. [6], wymieniają wnioskowanie jako podproces wchodzący w skład procesu interpretacji. Jednakże w niniejszym ujęciu zarówno wnioskowanie, jak i szeroko rozumiane podejmowanie decyzji (*decision making*) nie są rozważane.

2.1. Retranslacja

Pod pojęciem *retranslacji danych numerycznych* rozumiemy lingwistyczne wyrażanie ich sensu i znaczenia w kontekście danej przestrzeni rozważań. Na przykład jeśli w danych numerycznych napotkamy liczbę 190 cm jako wzrost człowieka, jej objaśnienie może mieć postać *wysoki człowiek*, ale także *niewysoki gracz w koszykówkę*. Sam źródłosłów pojęcia „retranslacja" dotyczy ponownego przetłumaczenia danej treści na słowa, zakładając, iż pierwotnie treść ta przenoszona była przez słowa i terminy naturalnojęzykowe, które w celu zachowania ich w systemie komputerowym zostały poddane pewnej transkrypcji lub translacji na język cyfr[1]. Termin „retranslacja" mógłby zostać także zastąpiony przez *leksematyzację*, gdzie leksem to jednostka informacji językowej przenosząca znaczenie, lub *werbalizację*, czyli po prostu "ubrania w słowa", "wysłowienia" danej treści.

Definicja 1. Niech D będzie zbiorem danych numerycznych, zaś T – zbiorem terminów języka naturalnego. Każda funkcja $\rho : D \to T$ przypisująca elementom ze zbioru D elementy ze zbioru T zwana jest *retranslacją*.

W ogólności, funkcja ρ może być dowolnym odwzorowaniem, które przyporządkowuje danym z D naturalne określenia z pewnego zbioru T. Potrzeba konstrukcji i użycia

[1] W szczególności, jeśli pewna treść jest ściśle związana z określeniem liczbowym, np. określona kwota pieniężna, wspomniana transkrypcja może być postrzegana jako funkcja identycznościowa, zapisująca w bazie liczby dokładnie w tej postaci, w której są one opisywane językowo.

takiej funkcji pojawia się zawsze wtedy, gdy dane liczbowe winny być zaprezentowane w bardziej przystępnej, werbalnej formie. W niniejszej pracy przedstawiamy metody retranslacji oparte o zbiory rozmyte i o interwałowe zbiory rozmyte.

Zmienne lingwistyczne

Zmienna lingwistyczna zdefiniowana w [44] służy do reprezentowania terminów naturalnojęzykowych poprzez skojarzone z nimi zbiory rozmyte. W istocie, jest to jedno z najpopularniejszych i najszerzej stosowanych narzędzi retranslacji. W myśl zasady, że typ zmiennej definiuje się poprzez zakres jej wartości, zmienną lingwistyczną nazwiemy piątkę uporządkowaną $<L, H, X, G, M>$, gdzie L jest nazwą zmiennej, H lub $H(L)$ jest zbiorem lingwistycznych wartości L, tzw. etykiet, X jest przestrzenią rozważań w której określamy zbiory rozmyte skojarzone z elementami $H(L)$, G jest gramatyczną (składniową) regułą, według której dobierane są etykiety z $H(L)$, M jest regułą semantyczną wiążącą etykietę z $H(L)$ z funkcją przynależności zbioru rozmytego w X. Znaczenie tej definicji jest objaśnione przez jej autora:

Koncepcja zmiennej lingwistycznej dostarcza metody przybliżonego charakteryzowania zjawisk, które są zbyt złożone lub niedokładnie zdefiniowane, aby wyrazić je w konwencjonalnych terminach ilościowych.

Dla oceny, w jakim stopniu dana etykieta $S \in H(L)$ koresponduje z obiektem opisanym przez $x \in X$, wprowadza się pojęcie *stopnia zgodności* (*compatibility level*), określonego jako $\mu_S(x)$. Przy pomocy operacji na wartościach funkcji przynależności skojarzonych z etykietami, jesteśmy także w stanie reprezentować wyrażenia złożone zawierające spójniki *i*, *lub*, *nie*, *bardzo* itp., zob. np. [2, 3, 44]. Ponadto użycie tzw. modyfikatorów (*hedges*), pozwala na uwzględnienie zmian intensywności modelowanych cech poprzez uzupełnienie ich opisów słownych o wyrażenia takie jak *bardzo*, *nieco*, *raczej*, itp. [2, 44]. Natomiast w procesie podsumowywania danych wartości zmiennych lingwistycznych wykorzystywane są jako *cechy podsumowań* (*summarizers*) oraz jako *selektory* (*queries*).

Kwantyfikatory lingwistyczne

Naturalna ludzka skłonność do używania nieprecyzyjnych, ale zrozumiałych określeń jest w szczególności widoczna przy opisywaniu ilości obiektów zainteresowania. Wyrażenia te, nazywane *kwantyfikatorami lingwistycznymi*, mają swoje typowe przykłady pośród określeń: *większość z*, *wszyscy oprócz około 10*, *bardzo mało*, *od jednego do trzech*, *dużo więcej niż 1000*, *około połowy*, *dokładnie 5* etc.

Najlepiej znanymi modelami kwantyfikatorów lingwistycznych są *kwantyfikatory rozmyte* Zadeha [45]. Wyrażenia dotyczące ilości obiektów reprezentowane są poprzez normalne i wypukłe zbiory rozmyte w dziedzinie rzeczywistej nieujemnej. Model ten ujmuje klasyczne kwantyfikatory $\forall$ i $\exists$ jako przypadki szczególne.

Zadehowskie i inne rozmyte modele kwantyfikatorów lingwistycznych przedstawione są przez [19, 20]. Jednymi z nich są: *T*- i *S*-kwantyfikatory [38], oparte o normy trójkątne. Z kolei w pracy (Novak 1989) przedstawiono kwantyfikatory rozmyte jako szczególne przypadki *L*-fuzzy sets, czyli zbiorów *L*-rozmytych. Jedne z najnowszych prac [4–7] rozważają możliwości modelowania poprzez zbiory rozmyte ponad 30 rodzajów kwantyfikatorów lingwistycznych, jak np. *restricted* i *unrestricted*, *multiplace*, np. *więcej A niż B jest C*, *złożone* (*composite*), lub nawet (!) *non-quantitative*, czyli *nieilościowe*. Rozważania tych autorów inspirowane są głównie teorią kwantyfikatorów uogólnionych (*Theory of Generalized Quantifiers*, TGQ) [1] . Jeszcze innym sposobem reprezentowania nieprecyzyjnych określeń ilościowych są Yagera *Ordered Averaging Operators* (OWA), [41].

Poniżej przedstawiono dwie *formy kanoniczne* stwierdzeń z użyciem kwantyfikatorów lingwistycznych reprezentowanych przez zbiory rozmyte [45]:

$$Q \text{ obiektów jest } S_1 \tag{2}$$

$$Q \text{ obiektów będących } S_2 \text{ jest } S_1 \tag{3}$$

gdzie S_1, S_2 są zbiorami obiektów charakteryzowanych przez pewne własności, zaś Q jest językowym określeniem ilości tych obiektów.

Ponadto, wyróżnia się dwa rodzaje kwantyfikatorów rozmytych:

1) **absolutne –** reprezentowane przez zbiory rozmyte w $\boldsymbol{R}^+ \cup \{0\}$, np. *ponad 200*, *około 1000*, *mniej niż 10*;
2) **względne –** reprezentowane przez zbiory rozmyte w [0, 1], co oddaje ilość obiektów posiadających pewną cechę względem ilości wszystkich rozważanych obiektów, np. *około 1/4*, *prawie wszystkie*, *niewiele*.

Oba te rodzaje reprezentowane są jako *rozkłady możliwości* (*possibility distributions*) w $\boldsymbol{R}^+ \cup \{0\}$ oraz [0, 1]. Kwantyfikatory rozmyte w pierwszej formie kanonicznej mogą być zarówno absolutne, jak i względne, zaś w drugiej formie – jedynie względne, co wynika z faktu, iż liczność elementów posiadających cechy S_1 i S_2 podawana jest zawsze w stosunku do liczby elementów w zbiorze S_1. Własności względnych interwałowych kwantyfikatorów rozmytych opisane są w [30].

2.2. Podsumowywanie baz danych

Termin *podsumowywanie danych* oznacza, w przyjętym tu kontekście, generowanie słownych opisów zbiorów danych. Opisy te mają za zadanie ujmować i pokrótce opisywać pewne ogólne charakterystyki jakościowe i ilościowe zgromadzonych danych, np. z analizy codziennych kursów walut za przykładowy okres trzech miesięcy wynikać może wiadomość *Większość walut europejskich umocniła się względem złotówki od początku roku.* Proces podsumowywania danych scharakteryzowany został następująco [22]:

Podsumowywanie to proces ekstrakcji najistotniejszej informacji ze źródła w celu wygenerowania skróconej wersji dla określonego użytkownika i wykonania określonego zadania.

Pomimo iż definicja ta dotyczy w założeniach autorów jedynie dokumentów tekstowych, oddaje także bardzo dokładnie proces podsumowywania danych numerycznych. Powszechnie znanym zbiorem metod opracowywania i charakteryzowania danych jest statystyka matematyczna, jednakże celem tej pracy jest opis metod opartych o zbiory rozmyte i, szerzej, o tzw. obliczenia miękkie, *soft computing*, gdyż pozwalają one na uzyskanie wyników i podsumowań sformułowanych w języku naturalnym, czego metody statystyczne nie zapewniają. Oczywiste jest, iż informacja słowna będzie, pomimo swej obniżonej precyzji i ścisłości, zdecydowanie bardziej czytelna dla przeciętnego użytkownika niż zbiór precyzyjnych, ale często nieco enigmatycznych dla umysłów humanistycznych wskaźników jak mediana czy odchylenie standardowe.

Według pomysłodawcy lingwistycznych podsumowań baz danych [42]

> (...) *podsumowywanie jest szczególnie przydatne, jeśli dostarcza informacji nie tak ścisłych jak np. średnia [lecz wyrażonej słowami (przyp. autora)], także jeśli uwzględnia dane nienumeryczne*.

Jak wspomniano powyżej, ludzie przyzwyczajeni są do przekazywania informacji za pomocą języka naturalnego, zatem głównym celem systemów interpretacji danych winno być zdefiniowanie i używanie takich procedur, których wyniki wyrażone byłyby w sposób jak najbardziej zbliżony do naturalnego. Dlatego też wybrane zostały metody reprezentowania danych lingwistycznych poprzez zbiory rozmyte. Jako podstawę procedur podsumowywania przedstawiono poniżej oryginalne podejście [43]

Podsumowania lingwistyczne Yagera

Lingwistyczne podsumowanie bazy danych to zdanie w języku seminaturalnym generowane automatycznie, o postaci

$$Q\ P \text{ jest/ma } S\ [T] \tag{4}$$

gdzie Q jest określeniem liczności, tzw. *kwantyfikatorem lingwistycznym*, np. *około połowy*, *kilka*, *więcej niż 50*. P jest *podmiotem podsumowania* – opisuje grupę obiektów, których dane zapisane są w rekordach, np. pracowników, samochodów. S jest cechą podsumowania (*summarizer*), którą reprezentuje zbiór rozmyty, np. *wysokie zarobki*, *średnia prędkość*. T jest miarą jakości podsumowania, *degree of truth*, *the truth of a summary*, która określa, w jakim stopniu kwantyfikator Q poprawnie oddaje ilość obiektów reprezentujących cechę S, T jest liczbą rzeczywistą z przedziału [0, 1]. Zarówno cecha podsumowania, jak i kwantyfikator rozmyty są wyrażeniami języka naturalnego modelowanymi za pomocą logiki rozmytej [44, 45].

Rozszerzenia podsumowań lingwistycznych

Rozszerzenia i ulepszenia idei Yagera są licznie prezentowane w literaturze naukowej. George i Srikanth zaproponowali generowanie podsumowań w oparciu o więcej niż jedną cechę charakteryzującą rekordy używając T-normy minimum do wyrażenia spójnika "i" łączącego cechy podsumowań S_1, S_2, ..., S_n [3], np. *wiele samochodów jest tanich* i *ubogo wyposażonych*. Dalsze znaczące prace w tej dziedzinie to [11–13], w których zaproponowano podsumowania lingwistyczne na podstawie drugiej formy kanonicznej kwantyfikatora rozmytego, czyli

$$Q\ P \text{ będących } S_2 \text{ jest/ma } S_1\ [T] \tag{5}$$

co wymagało także nowej formuły obliczania wskaźnika prawdziwości na podstawie tzw. kardynalności względnej cechy S_2 (względem cechy S_1). Cecha S_2 zwana jest *query* lub też – w odpowiedzi na potrzebę polskojęzycznego terminu – selektorem.

Metody obliczania wskaźników jakości podsumowań innych niż Yagera podane są w [11–13]. Próby podsumowań baz danych o zawartości tekstowej przedstawiają prace [35, 36].

3. Interwałowe reprezentacje danych lingwistycznych

3.1. Interwałowe zbiory rozmyte

Interwałowy zbiór rozmyty, czyli *Interval-Valued Fuzzy Set*, jest koncepcją uogólniającą pojęcie zbioru rozmytego Zadeha. Zarzutem często stawianym tradycyjnym zbiorom rozmytym jest to, że opisują one zjawiska nieostre i z natury nieprecyzyjne za pomocą konkretnych i ścisłych wartości liczbowych, co może wyglądać na pewną niekonsekwencję. Idea interwałowego zbioru rozmytego pozwala na wyrażenie stopnia przynależności jako liczby przedziałowej, czyli poprzez podanie pewnego zakresu, w którym niedokładnie określona wartość tej przynależności się mieści [8, 40]

$$A = \left\{ \left\langle x, \underline{\mu}_A(x), \overline{\mu}_A(x) \right\rangle : x \in X \right\} \tag{6}$$

gdzie $\underline{\mu}_A, \overline{\mu}_A : \to [0,1]$ to *dolna* i *górna* funkcja przynależności, których wartości spełniają warunek

$$0 \le \underline{\mu}_A(x) \le \overline{\mu}_A(x) \le 1 \tag{7}$$

Oczywiście, wartości tych funkcji dla danego x interpretowane są jako dolny i górny kraniec przedziału oznaczającego przynależność.

Dla interwałowych zbiorów rozmytych A, B w X iloczyn $A \cap B$, sumę $A \cup B$ oraz dopełnienie A^C określamy jako interwałowe zbiory rozmyte o funkcjach przynależności:

$$\underline{\mu}_{A\cap B}(x) = \min \{ \underline{\mu}_A(x), \underline{\mu}_B(x) \} \tag{8}$$

$$\overline{\mu}_{A\cap B}(x) = \min \{ \overline{\mu}_A(x), \overline{\mu}_B(x) \} \tag{9}$$

$$\underline{\mu}_{A\cup B}(x) = \max \{ \underline{\mu}_A(x), \underline{\mu}_B(x) \} \tag{10}$$

$$\overline{\mu}_{A\cup B}(x) = \max \{ \overline{\mu}_A(x), \overline{\mu}_B(x) \} \tag{11}$$

$$\underline{\mu}_{A^c}(x) = 1 - \overline{\mu}_A(x), \overline{\mu}_{A^c}(x) = 1 - \underline{\mu}_A(x) \tag{12}$$

gdzie T jest trójkątną T-normą zaś S – S-normą. Zauważyć należy, że jeśli

$$\forall x \in X \quad \underline{\mu}_A(x) = \overline{\mu}_A(x),$$

wówczas mamy do czynienia z przypadkiem szczególnym odpowiadającym zwykłemu zbiorowi rozmytemu.

Liczba kardynalna interwałowego zbioru rozmytego A wyrażona jest za pomocą liczby przedziałowej postaci

$$card(A) = [\underline{card}(A), \overline{card}(A)] = \left[\sum_{x \in X} \overline{\mu}_A(x)\right] \tag{13}$$

3.2. Interwałowe zmienne lingwistyczne

Na podstawie definicji interwałowego zbioru rozmytego, poniżej zostało zaproponowane autorskie rozszerzenie pojęcia zmiennej lingwistycznej, czyli *interwałowa* zmienna lingwistyczna. Jest ona uogólnieniem definicji Zadeha w tym samym sensie, w jakim interwałowy zbiór rozmyty rozszerza tradycyjny zbiór rozmyty. Możliwość budowy takiej definicji została zauważona w [23] oraz zrealizowana w [24]:

Definicja 2. *Interwałowa zmienna lingwistyczna* to piątka uporządkowana < *L, H, X, G, M*>, gdzie:

[*L*] – nazwa zmiennej,
[*H*] lub *H*(*L*) – zbiór lingwistycznych wartości *L*, tzw. etykiet,
[*X*] – przestrzeń rozważań, w której określamy interwałowe zbiory rozmyte skojarzone z elementami *H*(*L*),
[*G*] – gramatyczna (składniowa) reguła, według której dobierane są etykiety z *H*(*L*),
[*M*] – reguła semantyczna wiążąca etykietę z *H(L)* z funkcjami przynależności interwałowego zbioru rozmytego w *X*.

W konsekwencji, stopnie zgodności dla etykiet interwałowej zmiennej lingwistycznej obliczane są jako liczby interwałowe z przedziału [0, 1]. Metody ich porównywania oraz operacje arytmetyczne przedstawiają prace [10, 21, 39].

Wartości interwałowych zmiennych lingwistycznych mogą być łączone za pomocą spójników *i*, *lub*, oraz *nie*. Definicje tych operacji oparte są kolejno o operacje iloczynu, sumy i dopełnienia dla interwałowych zbiorów rozmytych reprezentujących wybrane etykiety.

Definicja 3. Niech L będzie zmienną lingwistyczną, zaś X – jej przestrzenią rozważań, w której określono interwałowe zbiory rozmyte $S_1, S_2, ..., S_n$, w celu reprezentowania etykiet L. Stopnie zgodności dla wyrażeń:

$$x \text{ jest } S_i \text{ i } S_j \tag{14}$$

$$x \text{ jest } S_i \text{ lub } S_j \tag{15}$$

$$x \text{ jest nie } S_i \tag{16}$$

gdzie $x \in X$, $i, j \leq n$, obliczane są jako interwały:

$$\left[\min\left\{\underline{\mu}_{S_i}(x), \underline{\mu}_{S_j}(x)\right\}, \min\left\{\overline{\mu}_{S_i}(x), \overline{\mu}_{S_j}(x)\right\}\right] \tag{17}$$

$$\left[\max\left\{\underline{\mu}_{S_i}(x), \underline{\mu}_{S_j}(x)\right\}, \max\left\{\overline{\mu}_{S_i}(x), \overline{\mu}_{S_j}(x)\right\}\right] \tag{18}$$

$$\left[1-\overline{\mu}_{S_i}(x), 1-\underline{\mu}_{S_i}(x)\right] \tag{19}$$

Definicja ta rozszerza analogiczną definicję dla tradycyjnych zbiorów rozmytych.

Kolejne operacje, którym poddać można etykiety interwałowej zmiennej lingwistycznej, to użycie *modyfikatorów*, *hedges*. Modyfikacja znaczenia etykiet odbywa się poprzez poddanie odpowiednich zbiorów rozmytych operacji *koncentracji* oraz *rozcieńczania*.

Definicja 4. Niech A będzie interwałowym zbiorem rozmytym w X. Funkcja przynależności zbioru A skoncentrowanego, A_{CON}, wyraża się wzorem

$$\mu_{A_{con}}(x) = \left[\underline{\mu}_{A_{con}}(x), \overline{\mu}_{A_{con}}(x)\right] = \left[\underline{\mu}_A^2(x), \overline{\mu}_A^2(x)\right] \ \forall x \in X \tag{20}$$

zaś zbioru A rozcieńczonego, A_{DIL}

$$\mu_{A_{DIL}}(x) = \left[\underline{\mu}_{A_{DIL}}(x), \overline{\mu}_{A_{DIL}}(x)\right] = \left[\underline{\mu}_A^{0,5}(x), \overline{\mu}_A^{0,5}(x)\right] \ \forall x \in X \tag{21}$$

Użycie koncentracji względem interwałowego zbioru rozmytego reprezentującego etykietę A tworzy model określenia językowego „bardzo A”, *very* A, zaś rozcieńczenie – „nieco A”, *more or less* A. W ogólności, wykładniki 2 i 0,5 mogą zostać zastąpione przez dowolne inne liczby rzeczywiste dodatnie, odpowiednio większe i mniejsze od jedności, co obrazuje zmianę „intensywności” cechy opisywanej przez etykietę A. Propozycje modyfikatorów lingwistycznych związanych z różnymi wykładnikami potęgi w definicji koncentracji i/lub rozcieńczania opisane są w [2].

3.3. Rozmyta kwantyfikacja a interwałowe zmienne lingwistyczne

Osobnym zagadnieniem interpretacji danych jest możliwość przedstawienia skwantyfikowanych nieprecyzyjnych wyrażeń reprezentowanych przez interwałowe zbiory rozmyte. Rozszerzając ujęcie Zadeha, przedstawmy dwie formy kanoniczne takich kwantyfikacji na podstawie (2) i (3), w których S_1 jest reprezentowana przez interwałowe zbiory rozmyte, zaś S_2 i Q przez zwykłe zbiory rozmyte. Stopień prawdziwości stwierdzenia (2) będzie wówczas obliczany jako liczba interwałowa

$$T = [\underline{t}, \overline{t}] = \left[\mu_Q \underline{card}((S_1)), \mu_Q \overline{card}((S_1))\right] \tag{22}$$

jeśli Q jest kwantyfikatorem absolutnym lub jako

$$T = [\underline{t}, \overline{t}] = \left[\mu_Q\left(\frac{\underline{card}(S_1)}{card(X)}\right), \mu_Q\left(\frac{\overline{card}(S_1)}{card(X)}\right)\right] \tag{23}$$

jeśli Q jest względne. Podobnie, prawdziwość stwierdzenia o postaci (3) obliczymy jako liczbę interwałową

$$T = [\underline{t}, \overline{t}] = \left[\mu_Q\left(\frac{\underline{card}(S_1 \cap S_2)}{card(S_2)} \right), \mu_Q\left(\frac{\overline{card}(S_1 \cap S_2)}{card(S_2)} \right) \right] \tag{24}$$

gdzie przecięcie interwałowego S_1 ze zwykłym S_2 jest obliczane wg (17), przy czym wartości funkcji przynależności dla S_2 traktowane są jako przedziały zdegenerowane.

Wzory (22)–(24) są dobrze określone tylko wtedy, gdy funkcja przynależności kwantyfikatora Q jest niemalejąca, w przeciwnym przypadku wynik może być przedziałem nieregularnym lub nie korespondującym z rzeczywistymi stopniami przynależności, np. jeśli Q miałoby lokalne maksimum w przedziale $\left[\underline{card}(S_1), \overline{card}(S_1) \right]$. Następująca forma tych wzorów

$$T = \left[\min_{r \in \left[\underline{card}(S_1), \overline{card}(S_1) \right]} \mu_Q(r), \max_{r \in \left[\underline{card}(S_1), \overline{card}(S_1) \right]} \mu_Q(r) \right] \tag{25}$$

dla Q absolutnego lub

$$T = \left[\min_{r \in \left[\left(\frac{\underline{card}(S_1)}{card(X)} \right), \left(\frac{\overline{card}(S_1)}{card(X)} \right) \right]} \mu_Q(r), \max_{r \in \left[\left(\frac{\underline{card}(S_1)}{card(X)} \right), \left(\frac{\overline{card}(S_1)}{card(X)} \right) \right]} \mu_Q(r) \right] \tag{26}$$

dla Q względnego, gwarantuje uniknięcie tych niedogodności. Z kolei dla stwierdzeń o postaci drugiej formy kanonicznej, właściwy będzie wzór

$$T = \left[\min_{r \in \left[\left(\frac{\underline{card}(S_1 \cap S_2)}{card(S_2)} \right), \left(\frac{\overline{card}(S_1 \cap S_2)}{card(S_2)} \right) \right]} \mu_Q(r), \max_{r \in \left[\left(\frac{\underline{card}(S_1 \cap S_2)}{card(S_2)} \right), \left(\frac{\overline{card}(S_1 \cap S_2)}{card(S_2)} \right) \right]} \mu_Q(r) \right] \tag{27}$$

3.4. Interwałowe kwantyfikatory rozmyte

Niniejsza sekcja przedstawia koncepcję reprezentowania kwantyfikatorów lingwistycznych poprzez interwałowe zbiory rozmyte. Koncepcja, naturalnie, rozszerza tradycyjne zadehowskie ujęcie [45]. Interwałowe zbiory rozmyte reprezentujące kwantyfikatory lingwistyczne winny być określone w $\boldsymbol{R}^+ \cup \{0\}$ oraz spełniać własności normalności i wypukłości. Poniżej przedstawione są definicje tych własności dla interwałowych zbiorów rozmytych, które rozszerzają analogiczne własności dla zwykłych zbiorów rozmytych.

Definicja 5. Interwałowy zbiór rozmyty A jest *normalny* wtedy i tylko wtedy, gdy zbiory $\underline{A} = \{ \langle x, \underline{\mu}_A(x) \rangle : x \in X \}$ oraz $\overline{A} = \{ \langle x, \overline{\mu}_A \rangle : x \in X \}$ są normalne, tzn. gdy:

$$\max_{x \in X} \underline{\mu}_A(x) = \max_{x \in X} \overline{\mu}_A(x) = 1 \tag{28}$$

A jest *wypukły* wtedy i tylko wtedy, gdy zbiory $\underline{A}$ oraz $\overline{A}$ są wypukłe, tzn. $\forall\alpha\in[0, 1]$ tzw. Alfa-przekroje, *alpha-cuts*, $\underline{A}_\alpha$ i $\overline{A}$ są wypukłe w klasycznym ujęciu.

Zatem definicja interwałowego kwantyfikatora rozmytego niżej podana ma postać.

Definicja 6. *Interwałowym kwantyfikatorem rozmytym* nazywamy normalny i wypukły interwałowy zbiór rozmyty w $Y = \boldsymbol{R}^+\cup\{0\}$ reprezentujący kwantyfikator lingwistyczny Q w wyrażeniu (2) lub (3). Q jest *absolutny*, jeśli $Y= \boldsymbol{R}^+\cup\{0\}$, Q jest *względny*, jeśli $Y = [0, 1]$.

Określanie stopnia prawdziwości kwantyfikowanych lingwistycznie stwierdzeń o formach (2) lub (3), w których własności S_1, S_2 reprezentowane są przez zwykłe zbiory rozmyte, zaś kwantyfikator lingwistyczny – poprzez interwałowy zbiór rozmyty, polega na obliczeniu liczby interwałowej T

$$T = [\underline{t}, \overline{t}] = \left[\underline{\mu}_Q\left(card(S_1)\right), \overline{\mu}_Q\left(card(S_1)\right)\right] \tag{29}$$

jeśli Q jest absolutny, lub

$$T = [\underline{t}, \overline{t}] = \left[\underline{\mu}_Q\left(\frac{card(S_1)}{card(X)}\right), \overline{\mu}_Q\left(\frac{card(S_1)}{card(X)}\right)\right] \tag{30}$$

jeśli Q jest względny. Jeśli zaś wyrażenie przedstawione jest poprzez formę (3), wówczas

$$T = [\underline{t}, \overline{t}] = \left[\underline{\mu}_Q\left(\frac{card(S_1\cap S_2)}{card(S_2)}\right), \overline{\mu}_Q\left(\frac{card(S_1\cap S_2)}{card(S_2)}\right)\right] \tag{31}$$

jedynie dla Q względnego.

4. Interwałowe podsumowania lingwistyczne baz danych

W przypadku kiedy dane pochodzące od ekspertów, a służące do określania funkcji przynależności kwantyfikatorów i/lub summarizerów, są podane nieprecyzyjnie, np. w postaci liczb przedziałowych, w miejsce zwykłych podsumowań lingwistycznych zastosować można *interwałowe podsumowania lingwistyczne baz danych.*

Definicja 7. Interwałowe podsumowanie lingwistyczne bazy danych jest seminaturalnym zdaniem postaci

$$Q\ P \text{ jest/ma } S\ [\underline{t}, \overline{t}] \tag{32}$$

w którym symbole Q, P i S interpretuje się jak w przypadku (4), z tym że przynajmniej jedno z Q, S jest reprezentowane przez interwałowy zbiór rozmyty, zaś $T = [\underline{t}, \overline{t}] \subseteq [0, 1]$ jest liczbą interwałową interpretowaną jako stopień prawdziwości podsumowania.

Definicja ta rozszerza klasyczne ujęcie Yagera i ujmuje je jako swój szczególny przypadek.

4.1. Podsumowanie z kwantyfikatorem interwałowym

Załóżmy, że baza danych, na której należy wykonać podsumowanie, jest postaci $D = V(y_1), V(y_2), ..., V(y_m)$, co oznacza, że każdy z m obiektów opisanych tą bazą posiada cechę V (np. wzrost), czego miarą liczbową jest wartość $V(y_i)$, $i \leq m$ (np. podana w cm). Celem jest znalezienie wskaźnika trafności dla podsumowania postaci Q P jest/ma S, w którym jedynie Q reprezentowany jest poprzez interwałowy zbiór rozmyty. W tym przypadku wskaźnik T będzie liczbą interwałową postaci $\mu_Q(r) = [\underline{t}, \overline{t}]$. Niech interwałowy kwantyfikator lingwistyczny dany będzie funkcjami przynależności $\underline{\mu}_Q, \overline{\mu}_Q$.

Obliczanie wskaźnika przebiega wówczas następująco:

Krok 1. Oblicz $r = \sum_{i=1}^{m} \mu_S(d_i)$.

Krok 2. Oblicz górną i dolną wartość T jako odpowiednio $\underline{\mu}_Q(r)$ i $\overline{\mu}_Q(r)$ dla Q absolutnego lub jako $\underline{\mu}_Q(r/m)$ i $\overline{\mu}_Q(r/m)$ dla Q względnego.

W przypadku gdy $\forall x \in X$ $\underline{\mu}_Q(x) = \overline{\mu}_Q(x)$, T jest liczbą rzeczywistą, a podsumowanie jest przypadkiem opisanym przez Yagera.

4.2. Podsumowanie z cechą interwałową

Podsumowanie lingwistyczne, w którym jedynie summarizer reprezentowany jest przez interwałowy zbiór rozmyty, także posiada wskaźnik trafności o postaci liczby interwałowej. Załóżmy tym razem, iż baza danych opisuje obiekty $y_1, ..., y_m$ za pomocą atrybutów $V_1, ..., V_n$ o dziedzinach $X_1, ..., X_n$. Niech $S_1, ..., S_n$ będą interwałowymi zbiorami rozmytymi odpowiednio w $X_1, ..., X_n$. Wówczas, rozszerzając ideę George'a i Srikantha, podsumowanie ma postać

$$Q \text{ obiektów z } Y \text{ jest/ma } S_1 \text{ i } S_2 \text{ i } ... S_n \; [\underline{t}, \overline{t}] \tag{33}$$

Przynależność do iloczynu interwałowych zbiorów rozmytych $S_1, ..., S_n$ obliczamy jako

$$\underline{\mu}_S(d_i) = \min_{j=1,2,...,n} \mu_S\left(V_j(y_i)\right), i = 1, ..., m \tag{34}$$

i dla $\overline{\mu}_S(d_i)$ analogicznie. Suma przynależności r obliczana jest na podstawie (34)

$$r = [\underline{r}, \overline{r}] = \sum_{i=1}^{m} \left[\underline{\mu}_S(d_i), \overline{\mu}_S(d_i)\right] \tag{35}$$

zaś obliczenie wskaźnika trafności podsumowania wygląda następująco

$$T = \left[\min_{r \in [\underline{r}, \overline{r}]} \mu_Q(r), \max_{r \in [\underline{r}, \overline{r}]} \mu_Q(r)\right] \tag{36}$$

Podobnie jak w przypadku interwałowego kwantyfikatora rozmytego, jeśli wszystkie zbiory $S_1, ..., S_n$ są zwykłymi zbiorami rozmytymi, wówczas $T \in [0, 1]$

4.3. Podsumowanie z selektorem interwałowym

Wprowadźmy teraz interwałowy zbiór rozmyty jako selektor do podsumowania lingwistycznego o postaci (3). Wymaga to przekształcenia wzoru (34) do

$$\underline{\mu}_S(d_i) = \min_{j=1,\,2,\,\ldots,\,n} \underline{\mu}_{S_j}\left(V_j(y_i)\right) t\, \underline{\mu}_{w_g}\left(V_g(y_i)\right),\; i = 1, 2, \ldots, m \quad (37)$$

oraz dla $\overline{\mu}_S(d_i)$ analogicznie. Wówczas

$$r = \left[\underline{r}, \overline{r}\right] = \frac{\left[\sum_{i=1}^{m} \underline{\mu} r_S(d_i),\, \sum_{i=1}^{m} \overline{\mu}_S(d_i)\right]}{\sum_{i=1}^{m} \mu_{w_g}(V_g(y_i))} \quad (38)$$

zaś $T = \mu_Q(r)$. Nadmieńmy, iż metoda ta pozwala na używanie wyłącznie kwantyfikatorów względnych. Oczywiście, gdy selektor reprezentowany jest przez zwykły zbiór rozmyty, metoda przybiera formę opisaną w [12] i [13].

5. Zastosowania

Niniejszy artykuł nie ma na celu opisu szczegółów technicznych zastosowań przedstawionych metod. Intencją autora jest tu raczej nakreślenie pewnych ogólnych zasad, w myśl których rozwiązanie problemu automatycznego i inteligentnego interpretowania danych może być implementowane i stosowane w obliczeniach. Niemniej dla pełności obrazu, warto zacytować publikacje, w których opisane metody znalazły zastosowanie.

Operacje wykorzystujące interwałowe zbiory rozmyte i interwałowe struktury danych do retranslacji danych, zastosowane zostały do ewaluacji algorytmów automatycznego oceniania testów egzaminacyjnych przeprowadzanych zdalnie [26]. W podobnym celu zastosowano podsumowania lingwistyczne baz danych w tradycyjnym yagerowskim ujęciu [24]. Innym zastosowaniem podsumowań Yagera są algorytmy pozwalające wygenerować opisy słowne o charakterze krótkich informacji lub notatek prasowych, co zaprezentowano w [32].

Podstawowe definicje dotyczące interwałowych zmiennych lingwistycznych i interwałowych podsumowań lingwistycznych zaprezentowano w [24], zaś miary jakości interwałowych podsumowań lingwistycznych w [25]. Stosunkowo kompletna teoria interwałowych kwantyfikatorów rozmytych, interwałowych zmiennych lingwistycznych oraz lingwistycznie kwantyfikowanych stwierdzeń reprezentowanych przez interwałowe zbiory rozmyte, jak również algorytmy umożliwiające podsumowywanie danych za pomocą tych narzędzi przedstawiono w [25].

6. Podsumowanie i możliwości dalszych badań

Artykuł prezentuje rozmyte metody interpretowania danych numerycznych i tekstowych. Celem opisanych algorytmów jest pozyskanie krótkich tekstowych informacji objaśniających dane zgromadzone w przetwarzanych bazach, zarówno co do znaczenia wartości pojedynczych rekordów (retranslacja), jak i co do szerszych ich zbiorów (podsumowywanie). Jako punkt wyjścia obrano Zadeha teorię zbiorów rozmytych, z charakterystycznymi dla niej narzędziami reprezentacji wiedzy nieostrej, czyli zmiennymi lingwistycznymi i kwantyfikatorami lingwistycznymi, oraz Yagera metody generowania lingwistycznych podsumowań baz danych.

Oryginalnym wkładem autora jest rozszerzenie cytowanych metod poprzez użycie teorii interwałowych zbiorów rozmytych. Nowe ujęcie obejmuje podejście opisane przez poprzedników jako swój podzbiór – metody dla interwałowych zbiorów rozmytych zastosowane do danych wyrażonych interwałami zdegenerowanymi redukują się do tradycyjnych metod rozmytych opisanych przez Zadeha i Yagera. Kolejne rozwinięcie i poszerzenie proponowanych metod może zostać opracowane w oparciu o zbiory rozmyte drugiego rodzaju, *Type-2 Fuzzy Sets*. W myśl teorii przedstawionej w [23], zbiory te stanowią uogólnienie dla m.in. interwałowych zbiorów rozmytych. Pewne elementy tego wciąż opracowywanego ujęcia zostały zaproponowane w [26–28, 31].

Literatura

[1] Barwise J., Cooper R.: *Generalized quantifiers and natural language*. Linguistics and Philosophy, nr 4, 1981, 159–219

[2] Chen C.-Y., Liu B.-D.: *Linguistic hedges and fuzzy rule based systems*. W: *Accuracy Improvement in Linguistic Fuzzy Modelling*, pod edycją Cassillas J., Cordon O., Herrera F., and Magdalena L. Physica-Verlag, c/o Springer-Verlag, Heidelberg, New York, 2003

[3] George R., Srikanth R.: *Data summarization using genetic algorithms and fuzzy logic*. W: *Genetic Algorithms and Soft Computing*, pod edycją Herrera F,.Verdegay J., Physica–Verlag, Heidelberg, 1996, 599–611

[4] Glöckner I.: *Advances in DFS theory. Technical report TR2000-01*. University of Bielefeld, Technical Faculty, 2000

[5] Glöckner I.: *Branching of fuzzy quantifiers and multiple variable binding: An extension of DFS theory*. *Technical Report TR2002-08*. University of Bielefeld, Technical Faculty, 2002

[6] Glöckner I., Knoll A.: *A formal theory of fuzzy natural language quantification and its role in granular computing.* W: *Granular Computing: An Emerging Paradigm*, pod edycją Pedrycz W., Physica-Verlag, 2001, 215–256

[7] Glöckner, I., Knoll A.: *Fuzzy quantifiers in granular computing and their role for data summarization*. W: *Fuzziness and Soft Computing in the New Millennium*. Proceedings of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference. Vancouver, Canada, July 25–28, 2001, 264–269

[8] Gorzalczany M.: *A method of inference in approximate reasoning based on interval-valued fuzzy sets*. Fuzzy Sets and Systems, nr 21, 1987, 1–17

[9] Herrera F., Herrera-Viedma E.: *Linguistic decision analysis: steps for solving decision problems under linguistic information*. Fuzzy Sets and Systems, nr 115, 2000, 67–82

[10] Ishibuchi H., Tanaka H.: *Multiobjective programming in optimisation of the interval objective function*. European Journal of Operational Research, nr 48, 1990, 219–225

[11] Kacprzyk J., Strykowski P.: *Linguistic data summaries for intelligent decision support*. W: Proceedings of EFDAN'99. 4-th EuropeanWorkshop on Fuzzy Decision Analysis and Recognition Technology for Management, Planning and Optimization, Dortmund, 1999, 3–12

[12] Kacprzyk J., Yager R.R.: *Linguistic summaries of data using fuzzy logic*. International Journal of General Systems, nr 30, 2001, 133–154

[13] Kacprzyk J.,Yager R. R., Zadrożny S.: *Fuzzy linguistic summaries of databases for an efficient business data analysis and decision support*. W: Discovery for Business Information Systems, pod edycją Abramowicz W., Żurada J. Kluwer Academic Publisher B. V., Boston, 2001, 129–152

[14] Kacprzyk J., Zadrożny S.: *Fquery for access: Fuzzy querying for windows-based dbms*. In: Fuzziness in Database Management System, ed. Bosc P.,Kacprzyk J. Physica-Verlag, Heidelberg, 1995, 415–433

[15] Kacprzyk J., Zadrożny S.: *Computing with words: towards a new generation of linguistic querying and summarization of databases*. W: Quo vadis computational intelligence? pod edycją Sincak P., Vascak J. Physica Verlag, Heidelberg New York, 2000, 144–175

[16] Kacprzyk J., Zadrożny S.: *On linguistic approaches in flexible querying and mining of association rules*. W: Flexible Query Answering Systems. Recent Advances, pod edycją Larsen, H.L., Kacprzyk J., Zadrożny S., Andreasen T.C.H. Physica-Verlag (Springer-Verlag), Heidelberg, New York, 2001, 475–484

[17] Kacprzyk, J., Zadrożny S.: *Internet as a challenge to fuzzy querying*. W: Intelligent Exploration of the Web, pod edycją Szczepaniak, P., Segovia J., Kacprzyk J., Zadeh L. A. Physica Verlag, c/o Springer-Verlag, Heidelberg, New York, 2003, 74–95

[18] Larsen H. L., Kacprzyk J., Zadrożny S., Andreasen T.C.H. (ed.): *Flexible Query Answering Systems. Recent Advances*. Physica-Verlag (Springer-Verlag), Heidelberg, New York, 2001

[19] Liu Y., Kerre E.E.: *An overview of fuzzy quantifiers, part I: Interpretations*. Fuzzy Sets and Systems, nr 95, 1998, 1–21

[20] Liu Y., Kerre E.E.: *An overview of fuzzy quantifiers, part II: Reasoning and applications*. Fuzzy Sets and Systems, nr 96, 1998, 1–12

[21] Lodwick W. A., Jamison K.D.: *Special issue interfaces between fuzzy set theory and interval analysis*. Fuzzy Sets and Systems, nr 135, 2003, 1–3

[22] Mani I., Maybury M. (ed.): *Advances in Automatic Text Summarization*. The MIT Press, Cambridge Massachusetts, USA, 1999

[23] Mendel J.M.: *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice-Hall, Upper Saddle River, NJ, 2001

[24] Niewiadomski A.: *Interval-valued linguistic variables. An application to linguistic summaries*. W: *Issues in Intelligent Systems. Paradigms*, pod edycją Hryniewicz, O., Kacprzyk J., Koronacki J., Wierzchoń S. T. Warszawa, Akademicka Oficyna Wydawnicza EXIT, 2005, 167–183

[25] Niewiadomski A.: *Interval-valued quality measures for linguistic summaries*. W: Issues in Soft Computing. *Theory and Applications*, pod edycją Grzegorzewski P., Krawczak M., Zadrożny S. Warszawa, Akademicka Oficyna Wydawnicza EXIT, 2005, 211–224

[26] Niewiadomski A.: *O zastosowaniach zbiorów rozmytych II rodzaju w lingwistycznych podsumowaniach baz danych*. W: Materiały 1. Polskiej i Międzynarodowej Konferencji PD Forum – Conference on Computer Science, pod edycją Rutkowska D., Łódź, 2005 (w druku)

[27] Niewiadomski A.: *On two possible roles of type-2 fuzzy sets in linguistic summaries*. Lecture Notes in Artificial Intelligence, nr 3528, 2005, 341–347

[28] Niewiadomski A.: *News generating via fuzzy summarization of databases*. Lecture Notes in Computer Science, nr 3831, 2006, 419–429

[29] Niewiadomski A., Bartyzel M., Szczepaniak P.S.: *Lingwistyczne podsumowania zbiorów danych w ocenianiu algorytmów zautomatyzowanego egzaminowania na odległość*. W: Materiały XV Krajowej Konferencji Automatyki, Warszawa, 27–30 czerwca t.3, 2005, 81–86

[30] Niewiadomski A., Ochelska J., Szczepaniak P.S.: *Interval-valued linguistic summaries of databases*. Control and Cybernetics, 2005 (w druku)

[31] Niewiadomski A., Rybusiński B.: *Fuzzy sets-based retranslation of numerical data in e-learning*. Lecture Notes in Artificial Intelligence, nr 3528, 2005, 348–354

[32] Niewiadomski A., Szczepaniak P.S.: *News generating based on interval type-2 linguistic summaries of databases*. W: Proceedings of IPMU 2006 Conference, July 2–7, 2006, Paris, France (w druku)

[33] Nikravesh M., Takagi T., Tajima M., Loia V., Azvine B.: *Fuzzy logic and the internet: Web intelligence*. W: *Intelligent Exploration of the Web*, pod edycją Loia, V., Nikravesh M., Zadeh L.A. Physica Verlag, 2004, 1–26

[34] Novak V.: *Fuzzy Sets and their Applications*. Adam Hilger, 1989

[35] Ochelska J., Niewiadomski A., Szczepaniak P.S.: *Linguistic summaries applied to medical textual databases*. Medical Information Technology, 2001, 125–130

[36] Ochelska J., Szczepaniak P.S., Niewiadomski A.: *Automatic summarization of standardized textual databases interpreted in terms of intuitionistic fuzzy sets*. W: Soft Computing: Tools, Techniques and Applications, pod edycją Grzegorzewski P., Krawczak M., Zadrożny S. The Academic Press EXIT, Warsaw, 2004, 204–216

[37] Rutkowska D., Piliński M., Rutkowski L.: *Sieci neuronowe algorytmy genetyczne i systemy rozmyte*. Warszawa, PWN 1997

[38] Thiele H.: *On t-quantifiers and s-quantifiers*. W: *Proceedings of the Twenty-Fourth International Symposium on Multiple-Valued Logic*, 1994, 264–269

[39] Sengupta A., Pal T. K., Chakraborty D.: *Interpretation of inequality constraints involving interval coefficients and a solution to interval linear programming*. Fuzzy Sets and Systems, nr 119, 2001, 129–138

[40] Turksen I.: *Interval-valued fuzzy sets based on normal forms*. Fuzzy Sets and Systems nr 20, 1986, 191–210

[41] Yager R.R.: *On ordered weighted averaging operators in multicriteria decision making*. IEEE Transactions on Systems, Man, and Cybernetics, nr 18, 1988, 183–190

[42] Yager R.R., Ford M., Canas A. J.: *An approach to the linguistic summarization of data*. W: Proceedings of 3rd International Conference, Information Processing and Management of Uncertainty in Knowledge-Based System, Paris, France, 1990, 456–468

[43] Yager R.R.: *A new approach to the summarization of data*. Information Sciences, nr 28, 1982, 69–86.

[44] Zadeh L.A.: *The concept of linguistic variable and its application for approximate reasoning (I)*. Information Sciences, nr 8, 1975, 199–249

[45] Zadeh L.A.: *A computational approach to fuzzy quantifiers in natural languages.* Computers and Maths with Applications, nr 9, 1983, 149–184

[46] Zadeh L.A.: *From search engines to question-answering systems—the need for new tools.* Lecture Notes on Artificial Intelligence, nr 2663, 2003, 15–17

[47] Zadeh L., Kacprzyk J. (ed.): *Computing with Words in Information/Intelligent Systems, 1. Foundations, 2. Applications*. Physica Verlag, Heidelberg and New York, 1999