Osiągnięcia Nauki i Techniki
Kierunki Rozwoju i Metody

KONWERSATORIUM POLITECHNIKI WARSZAWSKIEJ
Wkładka do Miesięcznika Politechniki Warszawskiej nr 5/2004

Redaktor merytoryczny — Stanisław Janeczko

# Zbiory przybliżone nowa matematyczna metoda analizy danych

Na podstawie odczytu wygłoszonego w dniu 30 października 2003 roku

**Zdzisław Pawlak**

Instytut Informatyki Teoretycznej i Stosowanej PAN
44-100 Gliwice, ul. Bałtycka 5

Wyższa Szkoła Informatyki Stosowanej i Zarządzania
01-447 Warszawa, ul. Newelska 6
e-mail: zpw@ii.pw.edu.pl

## 1. Wstęp

Teoria zbiorów przybliżonych [4] jest, z logicznego punktu widzenia, sposobem nowego matematycznego podejścia do pojęć nieostrych, zaś w praktyce jest nową metodą analizy danych.

W teorii mnogości zbiór jest definiowany przez swoje elementy, przy czym nie jest tu potrzebna żadna dodatkowa wiedza o elementach uniwersum, z których tworzymy zbiory. W teorii zbiorów przybliżonych jest przeciwnie. Zakładamy, iż mamy pewne dane o elementach uniwersum i dane te są wykorzystywane do tworzenia zbiorów. Elementy, o których mamy identyczną informację są podobne i tworzą tzw. ***zbiory elementarne***. W teorii zbiorów przybliżonych stanowią one podstawę rozumowania. Suma dowolnych zbiorów elementarnych jest nazywana ***zbiorem definiowalnym***. Zbiory, które nie są zbiorami definiowalnymi nazywane są ***zbiorami przybliżonymi***.

Zbiory definiowalne można jednoznacznie scharakteryzować przez własności ich elementów, natomiast zbiorów przybliżonych nie można w ten sposób określić, dlatego w teorii zbiorów przybliżonych wprowadza się pojęcia ***dolnego*** i ***górnego przybliżenia zbioru***. Pojęcia te pozwalają na scharakteryzowanie każdego zbioru niedefiniowalnego (przybliżonego) za pomocą dwóch zbiorów definiowalnych — jego dolnego i górnego przybliżenia.

Dolnym przybliżeniem zbioru są wszystkie elementy, które w świetle posiadanej wiedzy mogą być jednoznacznie przyporządkowane do rozważanego zbioru, zaś górnym przybliżeniem zbioru są wszystkie elementy, których przynależności do danego zbioru w świetle posiadanej wiedzy nie można wykluczyć. Różnica między górnym a dolnym przybliżeniem jest nazywana ***obszarem brzegowym*** (***brzegiem***) ***zbioru***.

Z matematycznego punktu widzenia zbiór jest przybliżony wtedy i tylko wtedy, gdy jego obszar brzegowy jest niepusty.

Spróbujmy zobrazować różnicę między pojęciami zbioru ostrego i nieostrego. Przykładem zbioru ostrego jest zbiór liczb parzystych, gdyż każdą liczbę naturalną możemy jednoznacznie sklasyfikować jako parzystą lub nieparzystą. Zbiór „zdolnych” studentów jest natomiast pojęciem nieostrym, gdyż nie zawsze możemy jednoznacznie stwierdzić, że dany student jest zdolny lub nie.

Jak już wspomniano, z praktycznego punktu widzenia teoria zbiorów przybliżonych jest nową metodą analizy danych. Umożliwia ona:

- szukanie zależności między danymi,
- redukcję danych,
- określenie wagi danych,
- generowanie reguł decyzyjnych z danych itp.

Za zaletę teorii zbiorów przybliżonych należy uznać fakt, że:

- nie wymaga ona założeń odnośnie danych (np. prawdopodobieństwa czy rozmytości),
- zawiera szybkie algorytmy analizy danych,
- ułatwia interpretację wyników,
- charakteryzuje się znaczną prostotą matematyczną.

Metoda zbiorów przybliżonych znalazła liczne zastosowania, między innymi w:

- medycynie,
- farmakologii,
- bankowości,
- lingwistyce,
- rozpoznawaniu mowy,
- ochronie środowiska,
- bazach danych.

Teoria zbiorów przybliżonych ma wiele związków z innym dziedzinami, a w szczególności:

- teorią ewidencji Dempstera-Shafera,
- teorią zbiorów rozmytych,
- metodami wnioskowania Boolowskiego.

Mimo to może być ona rozpatrywana jako niezależna (samodzielna) dyscyplina naukowa.

Teoria zbiorów przybliżonych nie jest alternatywą w stosunku do innych istniejących metod — raczej je uzupełnia i może być stosowana łącznie z nimi.

Na temat teorii zbiorów przybliżonych i jej zastosowań opublikowano na świecie do tej pory blisko trzy tysiące prac oraz kilkanaście książek. Wzbudziła ona spore zainteresowanie, głównie w USA, Kanadzie, Japonii, Chinach i Indiach, a prace na jej temat prowadzone są również w wielu innych krajach. Również w Polsce kilka ośrodków badawczych zajmuje się tą teorią i prowadzi prace nad jej zastosowaniami.

Do tej pory odbyło się wiele międzynarodowych konferencji na temat teorii zbiorów przybliżonych oraz jej zastosowań. Ponadto na wielu renomowanych, międzynarodowych konferencjach organizowano specjalne sesje poświęcone tej teorii.

Teoria zbiorów przybliżonych wykazała swą użyteczność w wielu dziedzinach oraz wzbudziła spore zainteresowanie na świecie i to nie tylko wśród informatyków, ale również wśród logików i filozofów. Wymaga jednak dalszych badań, w szczególności w zakresie jej podstaw matematycznych oraz możliwości zastosowania w różnych dziedzinach.

W artykule podane zostaną podstawowe pojęcia teorii zbiorów przybliżonych, przedyskutowane krótko jej zastosowania oraz dalsze perspektywy.

Więcej danych na temat zbiorów przybliżonych i ich zastosowań można znaleźć w Internecie na stronie

**http://www.roughsets.org**

# 2. Zbiory

Zanim podamy główne założenia teorii zbiorów przybliżonych, przypomnijmy kilka faktów dotyczących pojęcia zbioru.

Podstawowym pojęciem matematyki jest pojęcie zbioru. Niemal wszystkie konstrukcje matematyczne odwołują się do tego pojęcia.

Sformułowanie pojęcia zbioru oraz stworzenie teorii zbiorów zawdzięczamy niemieckiemu matematykowi Georgowi Cantorowi (1845–1918), który przed około stu laty stworzył podwaliny współczesnej teorii mnogości. Oryginalna, intuicyjna definicja pojęcia zbioru sformułowana przez Cantora [1] brzmi:

> *Unter einer „Manningfaltigkeit" oder „Menge" verstehe ich nämlich allgemein jedes Viele, welches sich als Eines denken lässt, d.h. jeden Inbegriff bestimmter Elemente, welcher durch ein Gesetz zu einem Ganzen verbunden werden kann*.

Jej tłumaczenie według Romana Murawskiego [3] jest następujące:

> *Pod pojęciem „rozmaitości" czy „zbioru" rozumiem mianowicie ogólnie każdą wielość, która może być pomyślana jako jedność, tj. każdy ogół określonych elementów, które na mocy pewnego prawa mogą być złączone w jedną całość.*

Lub w nieco prostszym sformułowaniu:

> *Pod pojęciem „zbioru" rozumiemy każde zebranie w jedną całość M określonych dobrze odróżnionych przedmiotów m naszego oglądu czy naszych myśli (które nazywane są „elementami M")* [3].

Jak widać jest to pojęcie intuicyjne i bardzo proste.

W 1902 roku wybitny filozof angielski Bertrand Russell (1872–1970) zauważył, że teoria mnogości jest sprzeczna, czyli prowadzi do antynomii (sprzeczności) logicznych. Antynomia logiczna, w dalszym ciągu rozumowania nazywana po prostu antynomią, powstaje wtedy gdy, prowadząc poprawne rozumowanie logiczne dochodzimy do sprzeczności, czyli do zdań *A* i *nie-A*. Podważa to istotę logicznego rozumowania.

Dla przykładu omówimy tzw. antynomię Russella. Rozważmy zbiór $X$ złożony ze wszystkich zbiorów $Y$, które nie są własnymi elementami. Jeżeli przyjmiemy, że $X$ jest swoim własnym elementem to $X$, z definicji, nie może być swoim elementem; jeżeli zaś przyjmiemy, że $X$ nie jest swoim elementem to zgodnie z definicją zbioru $X$ musi on być swoim elementem. A więc przy każdym założeniu otrzymujemy sprzeczność.

Antynomia Russella świadczy o tym, że elementami zbioru nie mogą być dowolne obiekty, tak jak sobie to wyobrażał Cantor.

Mogłoby się wydawać, że antynomie są niewinnymi igraszkami logicznymi, jednakże tak nie jest. Podważają one istotę logicznego rozumowania i dlatego też przez ponad sto lat próbowano „naprawić” teorię Cantora lub zastąpić ją inną teorią zbiorów, jednakże jak dotąd próby te nie doprowadziły do pomyślnych rezultatów.

Jednocześnie, niezależnie od badań matematyków i filozofów, pojęcie zbioru zainteresowało inżynierów. Okazało się bowiem, że używając klasycznego, kantorowskiego pojęcia zbioru nie da się sformułować i rozwiązać wielu problemów praktycznych.

W 1965 roku profesor Lotfi Zadeh z Uniwersytetu w Berkely zaproponował inne pojęcie zbioru [6], w którym elementy mogą należeć do zbioru „w pewnym stopniu”, a nie definitywnie, jak to ma miejsce w klasycznej teorii zbiorów. Propozycja ta znalazła bardzo wiele zastosowań i zapoczątkowała całą lawinę badań na temat teorii zbiorów rozmytych (*fuzzy set theory*), jak nazwano teorię Zadeha. Teorię zbiorów rozmytych można uważać za pewną formalizację pojęć nieostrych.

Inne podejście do formalizacji pojęć nieostrych zostało zaproponowane przez autora w artykule *Rough Sets* [4]. Teoria zbiorów przybliżonych (*rough set theory*) znalazła liczne zastosowania i zainteresowała wielu logików na świecie.

Dodajmy, że ani teoria zbiorów rozmytych, ani teoria zbiorów przybliżonych nie są alternatywą dla klasycznej teorii mnogości, gdyż obie zostały sformułowane przy wykorzystaniu pojęć zawartych w tej teorii.

# 3. Wnioskowanie z danych

Podstawowe pojęcia teorii zbiorów przybliżonych mogą być sformułowane całkowicie ogólnie, jednakże z punktu widzenia zastosowań tej teorii wygodnie sformułować je w terminologii typowej dla analizy danych, którą można uważać za szczególny przypadek wnioskowania indukcyjnego.

Przypomnijmy, że mamy dwa główne rodzaje wnioskowania — dedukcyjne i indukcyjne. ***Wnioskowanie dedukcyjne*** daje narzędzia służące do wyprowadzania zdań prawdziwych z innych zdań prawdziwych. Prowadzi ono zawsze do konkluzji prawdziwych. Teoria dedukcji posiada dobrze ugruntowane i powszechnie przyjęte podstawy teoretyczne. Wnioskowanie dedukcyjne jest głównym narzędziem stosowanym w rozumowaniach matematycznych i poza nią nie znalazło zastosowania.

W naukach przyrodniczych (np. w fizyce) podstawową rolę odgrywa ***wnioskowanie indukcyjne***. Cechą charakterystyczną tego typu wnioskowania jest to, że nie wychodzi ono (jak w logice dedukcyjnej) od aksjomatów wyrażających wiedzę ogólną o interesującym nas świecie. Punktem wyjścia tego typu rozumowania są pewne częściowe fakty dotyczące badanej rzeczywistości (przykłady), które następnie są uogólniane, tworząc wiedzę o świecie szerszym niż ten, który stanowił punkt wyjścia wnioskowania.

W przeciwieństwie do wnioskowania dedukcyjnego, wnioskowanie indukcyjne nie prowadzi do wniosków prawdziwych a jedynie do wniosków prawdopodobnych (możliwych). Nie ma też jednolitych, ogólnie przyjętych podstaw teoretycznych. W logice indukcji rozstrzyganie prawdziwości hipotez odbywa się na podstawie eksperymentu a nie drogą formalnego rozumowania, charakterystyczną dla logiki dedukcji. Fizyka jest tu najlepszą ilustracją.

Powstanie komputerów i nowatorskie ich zastosowania przyczyniły się istotnie do gwałtownego wzrostu zainteresowania wnioskowaniem indukcyjnym. Dzięki informatyce dziedzina ta rozwija się niezwykle dynamicznie. Uczenie maszynowe, odkrywanie wiedzy, wnioskowanie z danych, systemy eksperckie i inne stanowią przykłady nowych kierunków we wnioskowaniu indukcyjnym. Badania nad teorią indukcji zawdzięczają informatyce nowe inspiracje. Jednakże do sytuacji jaką mamy w logice dedukcji jest jeszcze bardzo daleka droga. Nie widać bowiem na horyzoncie zarysu teorii indukcji mającej taki status jak teoria dedukcji.

Reasumując, cechy charakterystyczne wymienionych rodzajów wnioskowania podano w tabeli.

| **Wnioskowanie** | **dedukcyjne** | **indukcyjne** |
|---|---|---|
| Zastosowania | matematyka | nauki przyrodnicze i techniczne |
| Teoria | teoria pełna | częściowe teorie |
| Wnioski | zawsze prawdziwe | prawdopodobne (możliwe) |
| Weryfikacja hipotez | dowód | eksperyment |

Jak widać analizę danych należy traktować jako szczególny przypadek wnioskowania indukcyjnego.

# 4. Zbiory przybliżone i analiza danych

W celu lepszego i bardziej intuicyjnego zrozumienia podstaw teorii zbiorów przybliżonych z punktu widzenia analizy danych zacznijmy od prostego przykładu. Rozpatrzmy zbiór danych podanych w tabeli. Wiersze tabeli opisują pacjentów, kolumny tabeli oznaczone jako *ból głowy*, *ból mięśni*, *temperatura* reprezentują symptomy choroby i nazywane są ***atrybutami warunkowymi***. Kolumna oznaczona jako *grypa* definiuje podział pacjentów na dwie klasy — chorujących na grypę i niechorujących — i nazwany jest ***atrybutem decyzyjnym***.

| Pacjent | Ból głowy | Ból mięśni | Temperatura | Grypa |
|---|---|---|---|---|
| 1 | nie | tak | podwyższona | tak |
| 2 | tak | nie | podwyższona | tak |
| 3 | tak | tak | wysoka | tak |
| 4 | nie | tak | normalna | nie |
| 5 | tak | nie | podwyższona | nie |
| 6 | nie | nie | wysoka | tak |

Tabele takiego rodzaju nazywamy tablicami decyzyjnymi. Rzeczywiste tablice decyzyjne mogą zawierać dużo więcej atrybutów i przypadków. Ponadto, oprócz atrybutów jakościowych mogą także zawierać wartości liczbowe. W pewnych przypadkach niektóre wartości atrybutów mogą być nieznane.

Problem będący przedmiotem naszego zainteresowania może być określony następująco: należy znaleźć zależność pomiędzy występowaniem grypy (lub jej niewystępowaniem) a symptomami opisującymi pacjentów. Inaczej mówiąc, celem jest opisanie zbioru przypadków {1, 2, 3, 6} (lub zbioru {4, 5}) w kategoriach wartości atrybutów warunkowych.

Zauważmy, że analizowane dane są niespójne, gdyż przypadki 2 i 5 dostarczają sprzecznych informacji, tzn. obaj pacjenci opisani są tymi samymi wartościami atrybutów warunkowych, lecz są przydzieleni do różnych klas decyzyjnych. Oznacza to, że zbiór pacjentów cierpiących z powodu grypy nie może być jednoznacznie opisany w kategoriach symptomów *ból głowy*, *ból mięśni*, *temperatura*. Możemy jednak opisać ten zbiór w sposób przybliżony. W oparciu o posiadane dane można stwierdzić, że:

➢ {1, 3, 6} jest maksymalnym zbiorem przypadków, które na podstawie opisu atrybutami warunkowymi są ***z pewnością*** zaklasyfikowane do klasy pacjentów chorujących na grypę;

➢ {1, 2, 3, 5, 6} jest zbiorem przypadków, których przydzielenie do klasy pacjentów chorujących na grypę jest ***możliwe***, na podstawie opisu atrybutami warunkowymi;

➢ {2, 5} jest zbiorem przypadków, które nie mogą być jednoznacznie przydzielone do klasy *grypa* lub do klasy *brak grypy*, ze względu na sprzeczny opis atrybutami warunkowymi.

Dwa pierwsze zbiory są przybliżeniami klasy decyzyjnej *grypa*, nazywanymi odpowiednio jej ***dolnym*** i ***górnym przybliżeniem***.

Przedstawmy obecnie powyższe rozważania bardziej formalnie.

Tablice danych, takie jak podano w powyższym przykładzie, nazywane są często ***systemami informacyjnymi***.

System informacyjny, jest czwórką $(U, A, V, f)$, gdzie: $U$ jest niepustym i skończonym zbiorem obiektów nazywanym ***uniwersum***; $A$ jest niepustym i skończonym ***zbiorem atrybutów***; $V = \bigcup_{a \in A} V_a$, $V_a$ jest ***dziedziną atrybutu*** $a \in A$; a $f : U \times A \rightarrow V$ jest ***funkcją informacyjną***, taką, że $\forall a \in A,\ x \in U\ f(a,x)\, V_a$.

Z każdym podzbiorem atrybutów $B \subseteq A$ związana jest binarna relacja $I(B)$, nazywana ***relacją nierozróżnialności***, zdefiniowana jako

$$I(B) = \{(x, y) \in U \times U : \forall a \in B\ f(a, x) = f(a, y)\}$$

Jeśli $(x, y) \in I(B)$ to obiekty $x$ i $y$ są ***nierozróżnialne*** ze względu na podzbiór atrybutów $B$. Relacja nierozróżnialności jest relacją równoważności. Rodzinę wszystkich klas abstrakcji relacji $I(B)$, czyli podział zbioru $U$ za pomocą $B$, oznaczamy jako $U/I(B)$. $B(x)$ oznacza klasę abstrakcji relacji $I(B)$ zawierającą obiekt $x$ i nazywane jest ***zbiorem B-elementarnym***.

Jeżeli w systemie informacyjnym wyróżniamy rozłączne zbiory atrybutów warunkowych $C$ i atrybutów decyzyjnych $D$ (gdzie $A = C \cup D$), to system taki nazywany jest ***tablicą decyzyjną***.

Niech $S = (U, A, V, f)$ będzie systemem informacyjnym, $X$ niepustym podzbiorem $U$ oraz $B \subseteq A$. Celem jest opisanie zbioru $X$ w kategoriach wartości atrybutów ze zbioru $B$. Prowadzi to do zdefiniowania dwóch zbiorów $B_*(X)$ i $B^*(X)$, nazywanych odpowiednio ***B-dolnym przybliżeniem*** i ***B-górnym przybliżeniem X***. Definiujemy je jako

$$B_*(X) = \{x \in U : B(x) \subseteq X\}$$

$$B^*(X) = \{x \in U : B(x) \cap X \neq \emptyset\}$$

Zbiór $BN_B(X) = B^*(X) - B_*(X)$ jest nazywany ***B-brzegiem zbioru X***.

Dolne przybliżenie $B_*(X)$ zbioru $X$ jest zbiorem obiektów, które na podstawie zbioru atrybutów $B$ można ***z pewnością*** zaliczyć do zbioru $X$, podczas gdy obiekty należące do zbioru $B^*(X)$ (w świetle atrybutów $B$) mogą być jedynie uznane jako ***możliwe***, że należą do zbioru $X$. $B$-brzeg zbioru $BN_B(X)$ zawiera obiekty, których nie można jednoznacznie przydzielić do zbioru $X$ z uwagi na sprzeczny opis w terminach atrybutów $B$. Natomiast obiekty ze zbioru $U / B^*(X)$ z pewnością nie należą do zbioru $X$. Jeśli $BN_B(X) \neq \varnothing$ to o zbiorze $X$ mówimy, że jest on zbiorem ***B-przybliżonym***. W przeciwnym razie jest on zbiorem ***B-definiowalnym*** (dokładnym).

Formalne właściwości przybliżeń pokazano poniżej.

1) $B_*(X) \subseteq X \subseteq B^*(X)$,

2) $B_*(\varnothing) = B^*(\varnothing) = \varnothing$; $B_*(U) = B^*(U) = U$,

3) $B^*(X \cup Y) = B^*(X) \cup B^*(Y)$,

4) $B_*(X \cap Y) = B_*(X) \cap B_*(Y)$,

5) $X \subseteq Y$ implikuje $B_*(X) \subseteq B_*(Y)$ oraz $B^*(X) \subseteq B^*(Y)$,

6) $B_*(X \cup Y) B_*(X) \cup B_*(Y)$,

7) $B^*(X \cup Y) \subseteq B^*(X) \cap B^*(Y)$,

8) $B_*(-X) = -B^*(X)$,

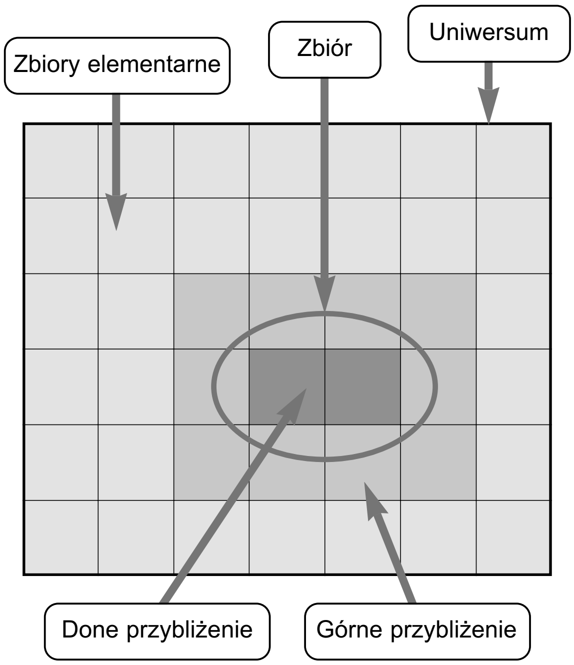
9) $B^*(-X) = -B_*(X)$,

10) $B_*(B_*(X)) = B^*(B_*(X)) = B_*(X)$,

11) $B^*(B^*(X)) = B_*(B^*(X)) = B^*(X)$.

Z właściwości tych widać, że przedstawione przybliżenia stanowią wnętrze i domknięcie w topologii generowanej przez dane.

Graficzną ilustrację przybliżenia pokazano na rysunku.




Zbiór przybliżony $X$ może być scharakteryzowany ilościowo za pomocą współczynnika dokładności przybliżenia $\alpha_B(X)$

$$\alpha_B(X) = \frac{|B_*(X)|}{|B^*(X)|}$$

gdzie $|X|$ oznacza liczność zbioru $X$.

Przybliżenia zbiorów są podstawowymi matematycznymi pojęciami teorii zbiorów przybliżonych i służą do opisu nieprecyzyjnej wiedzy o interesujących nas zjawiskach. Są one także używane do znajdowania zależności między danymi, ale tą problematyką nie będziemy się zajmować.

# 5. Co dalej?

W wielu ośrodkach badawczych w Polsce, USA, Kanadzie, Japonii, Chinach, Indiach i w Europie prowadzone są intensywne prace dotyczące różnych aspektów teorii zbiorów przybliżonych. Obejmują one między innymi:

- aspekty filozoficzne,
- podstawy matematyczne,
- różne uogólnienia i rozszerzenia,
- związki z innymi podobnymi teoriami,
- różnorodne zastosowania,
- oprogramowanie,
- prace konstrukcyjne.

Teorię zbiorów przybliżonych można uważać za szczególny przypadek realizacji idei Frege'go dotyczącej pojęć nieostrych (*vagueness*). Istnieje pogląd, że teoria zbiorów przybliżonych daje tu nowe narzędzie do badania logiki pojęć nieostrych, a w szczególności wyjaśnia niektóre paradoksy nieostrości (*the soritex paradox*). Wielu filozofów i logików prowadzi badania w tym kierunku (patrz [5]).

Pojęcie nierozróżnialności (*indiscernibility relation*), stanowiące podstawę teorii zbiorów przybliżonych, jest ściśle związane z podaną przez Leibnitza zasadą *nierozróżnialności identycznych obiektów* (*indiscernibility of identicals*). Ten aspekt zbiorów przybliżonych zainteresował również wielu filozofów, którzy prowadzą badania w tym zakresie.

Prace dotyczące podstaw teoretycznych zbiorów przybliżonych obejmują:

- ich algebraiczne i topologiczne własności;
- różne logiki związane ze zbiorami przybliżonymi;

- aspekty probabilistyczne zbiorów przybliżonych, między innymi związki z twierdzeniem Bayesa;
- aspekty teorio mnogościowe zbiorów przybliżonych, w tym związki z mereologią Leśniewskiego;
- analizę przybliżoną, funkcje przybliżone, przybliżoną ciągłość oraz przybliżone różniczkowanie i całkowanie (w sensie analizy niestandardowej a nie w sensie metod numerycznych).

W wyżej podanych kierunkach prowadzone są badania. Mają one na celu pełniejsze wyjaśnienie podstaw teorii zbiorów przybliżonych, głównie z punktu widzenia szeroko rozumianych własności matematycznych.

Teoria zbiorów przybliżonych jest uogólniana i modyfikowana na wiele sposobów. Między innymi zamiast relacji równoważności, stanowiącej punkt wyjścia do definicji zbioru przybliżonego, przyjmowane są inne relacje (np. relacja tolerancji). Bardzo dużym zainteresowaniem cieszą się również badania dotyczące sformułowania podstaw teorii zbiorów przybliżonych w oparciu o pewne koncepcje prawdopodobieństwa.

Ważne są również badania nad związkami teorii zbiorów przybliżonych z innymi teoriami dotyczącymi formalnych modeli nieprecyzyjności pojęć, takich jak:

- teoria zbiorów rozmytych,
- teoria ewidencji Dempstera-Shafera,
- statystyczne metody analizy danych.

Teoria zbiorów przybliżonych została również użyta do rozszerzenia niektórych znanych teorii, takich jak np. sieci neuronowe (*rough neural networks*), czy sieci Petriego (*rough Petri networks*). Są to obiecujące kierunki rozwijane przez wielu badaczy.

Bardzo intensywnie prowadzone są prace nad różnymi nowatorskimi zastosowaniami teorii zbiorów przybliżonych. Obejmują one między innymi teorię decyzji, medycynę, ekonomię, teorię sterowania i teorię konfliktów.

Szersze zastosowanie omawianej teorii wymagają odpowiedniego oprogramowania komputerowego. Opracowano już wiele systemów służących do tego celu i w dalszym ciągu prowadzone są w tym zakresie intensywne prace.

Aby w pełni wykorzystać możliwości, jakie daje teoria zbiorów przybliżonych w analizie danych, konieczne jest opracowanie komputerów lepiej przystosowanych do spożytkowania jej zalet. Prace w tym kierunku prowadzone są na uniwersytecie w Osace. Miejmy nadzieję, że doprowadzą one do powstania nowych koncepcji komputerów, lepiej przystosowanych do analizy danych.

Bliższe informacje na poruszane tu tematy można znaleźć w Internecie.

# 6. Zakończenie

Podane tu sformułowanie podstawowych pojęć teorii zbiorów przybliżonych jest bardzo proste, a tym samym niestety niewystarczające dla celów praktycznych. Jego zastosowanie do rozwiązywania rzeczywistych problemów wymagało wielu rozszerzeń i uzupełnień. Pozwoliły one na stworzenie narzędzia matematycznego, które mogło być z powodzeniem zastosowane do rozwiązywania złożonych problemów. Zainteresowany czytelnik znajdzie odpowiednie dane w Internecie.

Teoria zbiorów przybliżonych, jak już wspomniano, nie jest alternatywą dla klasycznej teorii zbiorów. Jest ona pewną formalizacją pojęć nieprecyzyjnych (nieostrych), tym praktyczniejszą, że wykorzystuje dobrze znaną terminologię pojęć precyzyjnych. Pojęcia nieprecyzyjne są w niej przedstawiane za pomocą dwóch pojęć precyzyjnych (dolnego i górnego przybliżenia), co pozwala na operowanie klasycznym aparatem teorii mnogości przy wyrażaniu i analizie pojęć nieprecyzyjnych.

Teoria zbiorów przybliżonych znalazła liczne zastosowania, jednakże dalszy jej rozwój wymaga nadal badania jej podstaw matematycznych. Konieczne jest stworzenie ogólnie dostępnego oprogramowania opartego na tej metodzie, a także odpowiednie rozwiązania sprzętowe. Prace w tym zakresie prowadzone są już w wielu krajach.

# Literatura


[1] George Cantor: *Grundlagen einer allgemeinen Mannigfaltigkeitslehre*. Leipzig 1883.

[2] Gottlob Frege: *Grundlagen der Arithmetik*. Vol. 2. Verlag von Herman Pohle, Jena 1893.

[3] Roman Murawski: *Filozofia matematyki, antologia tekstów klasycznych*. Seria Filozoficzna I, Logika nr 46, Uniwersytet im. Adama Mickiewicza w Poznaniu, Poznań 1986.

[4] Zdzisław Pawlak: *Rough Sets*. “Int. J. of Information and Computer Sciences”, no. 11, 5, pp. 341–356, 1982.

[5] Read Stephen: *Thinking about Logic. Introduction to the Philosophy of Logic*. Opus, Oxford University Press, Oxford – New York 1995.

[6] Lotfi Zadeh: *Fuzzy Sets*. “Information and Control”, no. 8, pp. 338–353, 1965.

# Abstract

Rough set theory is a new mathematical approach to imperfect knowledge.

The problem of imperfect knowledge has been tackled for a long time by philosophers, logicians and mathematicians. Recently it has also become a crucial issue for computer scientists, particularly in the area of artificial intelligence. There are many approaches to the problem of how to understand and manipulate imperfect knowledge. The most successful one is, no doubt, the fuzzy set theory proposed by Zadeh.

The rough set theory proposed by the author presents another approach to this problem still. The theory has attracted the attention of many researchers and practitioners all over the world, who contributed essentially to its development and applications.

Rough set theory has an overlap with many other theories. However we will refrain to discuss these connections here. Despite of the above mentioned connections rough set theory may be considered as the independent discipline in its own rights.

Rough set theory has found many interesting applications. The rough set approach seems to be of fundamental importance to AI and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems, inductive reasoning and pattern recognition.

The main advantage of rough set theory in data analysis is that it does not need any preliminary or additional information about data — like probability in statistics, or basic probability assignment in Dempster-Shafer theory, grade of membership or the value of possibility in fuzzy set theory.

In this paper basic concepts of rough set theory are outlined, and various applications are briefly discussed.

At the end future problems and prospects of the theory are pointed out.

Słowa kluczowe: **zbiory** (*sets*), **zbiory przybliżone** (*rough sets*), **zbiory rozmyte** (*fuzzy sets*), **pojęcia nieostre** (*vagueness*), **analiza danych** (*data mining*)