

WIKTOR TRACZ, JERZY MOZGAWA

Zbiory przybliżone i możliwości ich wykorzystania w leśnictwie

Rough sets and the possibility of their use in forestry

ABSTRACT

Tracz W., Mozgawa J. 2013. Zbiory przybliżone i możliwości ich wykorzystania w leśnictwie. Sylwan 157 (6): 425-433.

The need for effective management of uncertainty and incomplete information is the main factor of interest in the rough set theory and multi-use methods of analysis based on this theory, both in scientific research and in solving practical problems. Rough sets can be used for processing of both non spatial and spatial data. Five areas of application of rough set in decision-support systems in forestry and forest research are presented in the study. The example of using rough sets methods for data processing in order to determine cause-effect relationships and making classification is also presented.

KEY WORDS

rough sets, uncertainty, vagueness, data mining, decision making process, GIS

ADDRESSES

Wiktor Tracz – e-mail: wiktortracz@wl.sggw.pl

Jerzy Mozgawa – e-mail: jerzymozgawa@wl.sggw.pl

Katedra Urządzania Lasu, Geomatyki i Ekonomiki Leśnictwa; SGGW w Warszawie;
ul. Nowoursynowska 159; 02-776 Warszawa

Wstęp

Postępujący proces udostępniania wielotematycznych baz danych związanych z wprowadzeniem programu INSPIRE stwarza szczególnie korzystne warunki wspomaganie procesu zarządzania w leśnictwie i prowadzenia leśnych badań naukowych. Duże zbiory danych, w tym SILP, zawierają wiele nieodkrytych zależności, możliwych do wykorzystania w zarządzaniu gospodarstwem leśnym. Warunkiem skorzystania z tych danych i poznania tych zależności jest dysponowanie odpowiednim narzędziem analitycznym. Opracowanie przedstawia zagadnienie zbiorów przybliżonych, czyli jednej z metod możliwych do zastosowań w takich sytuacjach.

Problemy zarządzania i badań w leśnictwie wymagające budowy specjalnych modeli przyrodniczych, technicznych i ekonomicznych miały być rozwiązywane na podstawie informacji zawartych w SILP. Dotychczas nie opracowano dla tej bazy znaczących aplikacji użytkowych, które mogłyby być elementem systemu wspomaganie decyzji. System wspomaganie decyzji (SWD) to system informatyczny, który dostarcza informacji z danej dziedziny, umożliwia korzystanie z analitycznych modeli decyzyjnych z dostępem do baz danych w celu wspomaganie decydentów w słabo ustrukturalizowanym środowisku decyzyjnym, głównie na strategicznym i taktycznym szczeblu zarządzania [Zieliński 2000]. U podstaw rozwoju SWD znajduje się dążenie do połączenia elastyczności intelektualnej i wyobraźni człowieka z szybkością, dokładnością i wydajnością komputerów.

Diagnoza trudności związanych z opracowaniem aplikacji wskazuje na specyfikę danych wykorzystywanych do modelowania jako danych obarczonych niekiedy wysokim stopniem niepewności. Zarówno dane INSPIRE, jak i dane SILP powinny być przetwarzane wybranymi metodami dostosowanymi do takich danych. Aby pozyskać użyteczne informacje zawarte w tych zbiorach, poszukiwane są nowe metody eksploracji danych (ang. data mining) i zdobywania wiedzy (ang. knowledge acquisition). Próby wykorzystania danych INSPIRE i SILP do modelowania stwarzają potrzebę dysponowania takimi metodami eksploracji danych i zdobywania wiedzy, które uwzględniają niepewność i niedokładność danych wykorzystywanych do analiz. Do takich metod należą: matematyczna teoria ewidencji [Shafer 1976], zbiory rozmyte [Zadeh 1965; Czogała, Pedrycz 1985], sieci neuronowe [Duch i in. 2000; Tracz 2003], algorytmy genetyczne [Goldberg 1998] i zbiory przybliżone [Pawlak 1982].

Teoria ewidencji jest metodą modelowania niepewności za pomocą prawdopodobieństwa wyrażonego funkcją przekonania (ang. belief function). Zbiory rozmyte poprzez zastosowanie tzw. funkcji przynależności umożliwiają określenie stopnia przynależności obiektu do zbioru, przez co istnieje możliwość opisu takich obiektów i zjawisk, które mają charakter wieloznaczny i nieprecyzyjny [Tracz, Mozgawa 2006]. Sieci neuronowe posiadają zdolność przetwarzania informacji niekompletnych, rozmytych, a nawet sprzecznych. Umożliwiają analizę problemów nieliniowych. Algorytmy genetyczne znajdują zastosowanie tam, gdzie nie jest dobrze określony lub poznany sposób rozwiązania problemu, ale znany jest sposób oceny jakości rozwiązania.

Spośród opracowanych metod eksploracji danych (ED) i zdobywania wiedzy (ZW) szczególne własności posiadają metody oparte na teorii zbiorów przybliżonych (ZP). Teoria ta została opracowana przez Zdzisława Pawłaka, polskiego matematyka, i zdobyła niekwestionowane uznanie na świecie. Jedną z jej najważniejszych zalet jest możliwość przetwarzania niepewnych i niedokładnych danych zarówno o charakterze przestrzennym, jak i nieprzestrzennym. Ze względu na tę cechę w Zakładzie Geomatyki i Gospodarki Przestrzennej Wydziału Leśnego SGGW w Warszawie przeprowadzono ocenę przydatności metody zbiorów przybliżonych dla potrzeb leśnictwa.

Elementy teorii zbiorów przybliżonych

Dominującym sposobem przechowywania informacji jest obecnie postać tablicowa, która jest wykorzystywana również do reprezentacji informacji o analizowanym problemie w teorii zbiorów przybliżonych. Wiersze tablicy reprezentują obiekty, a kolumny są atrybutami i opisują poszczególne obiekty. W teorii ZP tablica jest nazywana systemem informacyjnym (SI), który służy do reprezentacji wiedzy. Specjalnym przypadkiem systemu informacyjnego jest tablica decyzyjna, w której wyróżniony jest atrybut decyzyjny, decydujący o przynależności danego obiektu do jednej z klas obiektów (ryc. 1).

Punktem wyjścia do analiz metodami ZP jest wyznaczenie na danych tablicy relacji nierozróżnialności [Pawlak 1982]. Relacja ta jest podstawową koncepcją teorii zbiorów przybliżonych i zachodzi wtedy, kiedy poszczególne obiekty posiadają te same wartości dla atrybutów wybranych do ustalenia relacji. Obiekty są nierozróżnialne, jeśli posiadają te same wartości wybranych atrybutów (ryc. 1).

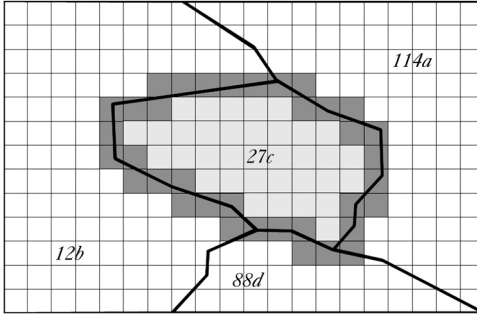
Wychodząc z fundamentalnej koncepcji relacji nierozróżnialności, teoria ZP wprowadza szereg definicji. Dla wybranego podzbioru obiektów i podzbioru atrybutów określa się tzw. przybliżenie zbioru. Dolnym przybliżeniem wybranego podzbioru obiektów w systemie informacyjnych nazywany jest zbiór tylko tych obiektów, które na pewno mogą być zidentyfikowane przy wykorzystaniu określonych wartości wybranych atrybutów (ryc. 2 – jasnoszare oczka rastra).

	A1	A2	A3	D
X1	3	tak	122	mało
X2	88	nie	315	dużo
X3	88	nie	315	dużo
X4	4	tak	114	mało
X5	3	tak	122	mało
X6	88	nie	315	dużo
X7	140	tak	191	dużo
X8	3	tak	191	mało

Ryc. 1.

Tablica decyzyjna z atrybutem decyzyjnym D – obiekty X2, X3, X6 są nierozróżnialne ze względu na atrybuty A1, A2, A3, D

Decision table for attribute D – objects X2, X3 and X6 are undistinguishable with regard to attributes A1, A2, A3, D



Ryc. 2.

Interpretacja geometryczna elementów zbioru przybliżonego dla drzewostanu 27c (objektami zbiorów są oczka rastra)

Geometric interpretation of rough set elements for forest stand 27c (cells of raster are objects of the set)

Górnym przybliżeniem wybranego podzbioru obiektów nazywany jest zbiór tych obiektów, które być może mogą być zidentyfikowane przy wykorzystaniu określonych wartości wybranych atrybutów (ryc. 2 – obszar pokryty jasno- i ciemnoszarymi oczkami rastra). Brzegiem zbioru przybliżonego są nazywane te obiekty, które nie mogą w 100% być zidentyfikowane jako elementy podzbioru przy wykorzystaniu wartości atrybutów (ryc. 2 – ciemnoszare oczka rastra). Obszar negatywny to zbiór obiektów, które nie mogą być zidentyfikowane przy wykorzystaniu wartości atrybutów (ryc. 2 – białe oczka rastra). Dla każdego podzbioru obiektów w SI można określić jego zbiór przybliżony, składający się z pary dolnego i górnego przybliżenia. Zbiór jest przybliżony wtedy i tylko wtedy, gdy brzeg tego zbioru jest niepusty.

Specyfiką algorytmów analiz w metodzie zbiorów przybliżonych jest operowanie głównie na małej liczbie podzakresów wartości określonego atrybutu. W tym celu przeprowadza się dyskretyzację zakresu wejściowych danych. Najczęściej temu procesowi podlegają atrybuty typu liczbowego. Podawanie atrybutów w formie dyskretnej jest równoznaczne ze zdefiniowaniem przybliżonego spojrzenia na rzeczywistość. W ten sposób dane empiryczne przyjmują formę mniej dokładnej informacji, która jest przedmiotem zainteresowania metody zbiorów przybliżonych. Dyskretyzacja jest celowym działaniem dostosowującym dane empiryczne do przetwarzania metodą ZP. Dyskretyzacja powinna być wykonana przy udziale eksperta, znającego specyfikę i charakter problemu.

Właściwa eksploracja danych i zdobywanie wiedzy odbywa się poprzez utworzenie koniunkcji elementarnych reguł decyzyjnych typu „jeżeli warunek, to decyzja” na podstawie zawartych w tablicy decyzyjnej warunków (wartości zbioru atrybutów opisujących) i decyzji (wartość atrybutu decyzyjnego). Rozpoznanie zależności pomiędzy atrybutami opisującymi obiekty umożliwia redukcję niepotrzebnych atrybutów i wyznaczenie reduktu, czyli minimalnego podzbioru atrybutów, dających taką samą jakość klasyfikacji jak cały ich zbiór. Wyznaczenie reduktu w metodzie zbiorów przybliżonych jest w rzeczywistości sposobem redukcji pozornej szczegółowości opisu obiektów, przyjętej na wstępie analiz.

Teoria zbiorów przybliżonych dostarcza specjalnych technik selekcjonowania zbioru reguł decyzyjnych w oparciu o redukty. Reguły tego typu składają się z minimalnej liczby atrybutów, wystarczających do rozróżnienia obiektów w różnych klasach decyzyjnych. Uzyskana w powyższy sposób zredukowana tablica decyzyjna może spełniać funkcję modelu prognostycznego, możliwego do wykorzystania w podejmowaniu decyzji o charakterze przestrzennym i nieprzestrzennym.

Twórca teorii ZP wyjaśnia powiązania tej teorii z klasyczną teorią zbiorów, która jest powszechnie wykorzystywana w systemach informacji przestrzennej. Pawlak [2004] stwierdza, że teoria zbiorów przybliżonych jest pewną formalizacją pojęć nieprecyzyjnych (nieostrych) w terminach pojęć precyzyjnych. Pojęcie nieprecyzyjne jest przedstawione w tej teorii za pomocą dwóch pojęć precyzyjnych (dolnego i górnego przybliżenia), co pozwala na operowanie klasycznym aparatem teorii mnogości do wyrażania i analizy pojęć nieprecyzyjnych.

Kierunki wykorzystania

Potrzeba dysponowania stosownym narzędziem służącym do efektywnego zarządzania niekompletną i niepewną informacją była głównym czynnikiem szerokiego zainteresowania teorią zbiorów przybliżonych i próbami wielokierunkowego zastosowania metod analiz opartych na tej teorii zarówno w badaniach naukowych, jak i przy rozwiązywaniu problemów praktycznych [Shi i in. 2003; Beaubouef i in. 2005].

W zagadnieniach niemających charakteru przestrzennego zwracają uwagę opracowania z zakresu szeroko rozumianej medycyny, oceny estetycznych walorów lasu, sprawności funkcjonowania maszyn [Klimkiewicz 2005] czy oceny satysfakcji klientów [Klimkiewicz, Moczulska 2008]. W zagadnieniach związanych z problemami klasyfikacji interesujące są pozycje Upadhyaya i in. [2006]. Problematykę wykorzystania zbiorów przybliżonych do analiz problemów przestrzenno-czasowych opisują Aldridge [2001], Synak [2003], Panchal i in. [2005] oraz Guangya i in. [2010]. Zastosowaniem zbiorów przybliżonych w gospodarce przestrzennej poświęcona jest praca Renigier-Biłozor i Biłozora [2008], a problematyce identyfikacji stref podmiejskich opracowanie Murgante i in. [2008].

Dużo danych posiada cechy informacji geoprzestrzennych, stąd w metodach eksploracji danych i zdobywaniu wiedzy zaszła konieczność uwzględnienia aspektu przestrzennego. W konsekwencji powstało szereg opracowań dotyczących oceny przydatności zbiorów przybliżonych w przetwarzaniu różnej dokładności danych geoprzestrzennych [Duckham i in. 2006; Schneider 2008]. Przykładami tej grupy zastosowań są także prace Gorsevskiego i Jankowskiego [2008] oraz Liu i in. [2011], które proponują wykorzystanie metodologii zbiorów przybliżonych do analizy czynników wpływających na powstawanie osuwisk w lasach.

Analiza literatury umożliwiła dokonanie ramowej klasyfikacji potencjalnych kierunków zastosowania zbiorów przybliżonych w leśnictwie. Większość z nich można przyporządkować do jednego z niżej wymienionych kierunków:

1. Modelowanie przybliżonego definiowania pojedynczych regionów w systemach rastrowych przez dolne i górne przybliżenie zbioru oczek rastra tworzących region oraz ustalenie dokładności aproksymacji zbioru oczek rastra tworzących region. Im mniejsze jest oczko rastra, tym dokładniej można wyznaczyć zbiór przybliżony. Jeśli wymiar rastra dąży do zera, to zbiór przybliżony zbliża się do zbioru dokładnie opisującego region.
2. Tworzenie zróżnicowanych klas obiektów przez modelowanie relacji nierozróżnialności. Zmiana parametrów dyskretyzacji wartości atrybutów, przez którą jest kontrolowana

relacja nierozróżnialności, skutkuje innym przyporządkowaniem obiektów do wyznaczonych klas.

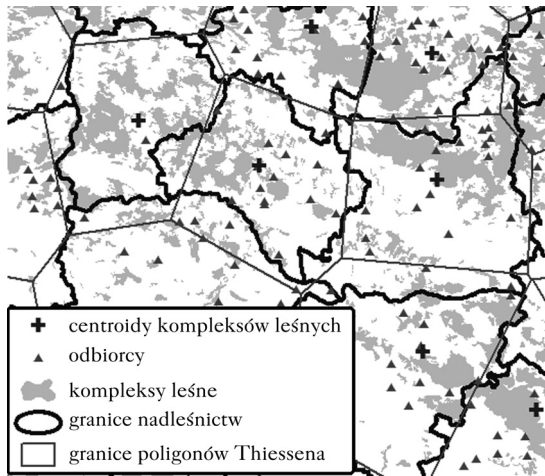
3. Klasyfikacja z wykorzystaniem reguł decyzyjnych wyznaczonych z tablic decyzyjnych.
4. Identyfikowanie relacji przyczyna-skutek przez wyznaczenie i analizę reguł decyzyjnych z tablic zawierających dane przybliżone. Ten kierunek zastosowań potencjalnie może być wykorzystany w badaniach naukowych problemów o charakterze zarówno nieprzestrzennym, jak i przestrzennym.
5. Definiowanie przestrzennych relacji topologicznych pomiędzy dwoma regionami o niepewnych granicach z zamiarem utrzymania miary niepewności na wszystkich etapach rozwiązania zadania o charakterze przestrzennym. Przestrzenne relacje topologiczne (np.: nakładania, przecinania, wycinania, zawierania, styku) są wykorzystywane powszechnie w analizach przestrzennych SIP. Ten kierunek zastosowania zbiorów przybliżonych wymaga odmiennego od obecnie stosowanych w technologii SIP sposobów opisu obiektów przestrzennych.

Przykładowe zastosowanie zbiorów przybliżonych w praktyce leśnej

Prezentowany przykład odnosi się do wykorzystania ZP do identyfikowania relacji przyczyna-skutek. Analizą objęto sprzedaż określonych sortymentów drzewnych przez nadleśnictwa w okresie kilku lat do odbiorców rozmieszczonych w strefie wpływu nadleśnictwa. Układ rynkowy dostawcy-odbiorcy ma zatem charakter przestrzenno-czasowy. Dane empiryczne pochodziły ze zbiorów Katedry Urządzania Lasu, Geomatyki i Ekonomiki Leśnictwa SGGW w Warszawie. Wykorzystano wybrane informacje dotyczące sprzedaży sortymentu WC0 drewna sosnowego w ciągu 7 lat. Obiektami w SI były transakcje kupna-sprzedaży drewna. Zakres tabeli decyzyjnej przykładu empirycznego obejmował 13585 obiektów.

Sprawdzano, czy ilość sprzedanego drewna ma związek z obszarem najbliższego otoczenia lasów nadleśnictwa. Do analizy metodą ZP przyjęto następujące atrybuty opisowe: podaż nadleśnictwa, popyt krajowy, popyt w strefie, odległość, kierunek wywozu oraz cena. Podaż nadleśnictwa to suma sprzedanego przez nadleśnictwo analizowanego sortymentu drewna w okresie rocznym. Popyt krajowy obejmował sumę sprzedanego drewna analizowanego sortymentu w danym roku w kraju. Popyt w strefie to suma kupionego drewna analizowanego sortymentu w danej strefie wpływu (od danego nadleśnictwa) w danym roku. Odległość to najkrótsza odległość liczona po drogach głównych i drugorzędnych pomiędzy dostawcą a odbiorcą. Kierunek wywozu był wyznaczony jako azymut określony lokalizacją dostawcy i odbiorcy. Cena dotyczyła średniorocznej ceny sprzedaży analizowanego sortymentu do danego odbiorcy. Atrybutem decyzyjnym była masa sprzedanego drewna. Metodą ekspercką wyznaczono cztery klasy atrybutu decyzyjnego.

Lokalizacja odbiorcy jest określona poprzez współrzędne wyznaczone na podstawie adresu odbiorcy. Określono poligony Thiessena (ryc. 3), będące reprezentacją przestrzenną kompleksów leśnych na obszarze nadleśnictwa (dostawcy). Odległość od dostawcy do odbiorcy obliczono na podstawie sieci szlaków komunikacyjnych jako najkrótszą odległość między centroidą kompleksów leśnych nadleśnictwa (środkiem poligonu Thiessena) a lokalizacją odbiorcy. Do wyznaczenia części tych atrybutów opisowych wykorzystano procedury analiz przestrzennych dostępnych w SIP: geokodowanie, wyznaczenie centroid obiektów poligonowych, analiza zawierania się, analiza najkrótszej drogi (analizy sieciowe), tworzenie poligonów Thiessena i obliczenie azymutu.



Ryc. 3.

Interpretacja geometryczna atrybutów opisowych o charakterze przestrzennym w układzie dostawca-odbiorca (poligon Thiessena obejmuje wszystkie lokalizacje, do których odległość od jego środka jest krótsza niż od środków innych poligonów)

Geometric interpretation of descriptive attributes of spatial character in the supplier-customer system (Thiessen polygon includes those locations, which are closer to its center than to the centers of the other polygons)

Do wyznaczenia innych atrybutów przetwarzano dane o sprzedaży drewna pochodzące z hurtowni danych Lasów Państwowych.

Etap analizy metodą ZP był poprzedzony przeglądem dostępnych oprogramowań. Na podstawie takich kryteriów jak wykorzystanie algorytmów obliczeniowych pochodzących z najlepszych ośrodków naukowych na świecie, liczba dostępnych metod analizy danych (metodami ZP oraz również metodami sieci neuronowych, algorytmów genetycznych i klasycznymi metodami statystycznymi) i przyjazny interfejs użytkownika wybrano program RSES (<http://logic.mimuw.edu.pl>). Analiza składała się z następujących etapów: utworzenie tabeli decyzyjnej, podział dostępnego zbioru obiektów na podzbiór treningowy i podzbiór testowy, dyskretyzacja zakresu wybranych atrybutów opisowych metodami dostępnymi w programie RSES [Bazan i in. 2000], wyznaczenie reduktów, wyznaczenie reguł decyzyjnych, ocena dokładności klasyfikacji. Wyznaczono 2062 reguł decyzyjnych na podstawie przetwarzania tabeli zawierającej podzbiór treningowy. Niżej przedstawiono jako przykład dwie z nich:

$$(Popyt_krajowy = (5651290 \div 6223540)) \& (Kierunek = (167,5 \div 292,5)) \& \\ \& (Popyt_strefa = (340 \div 1631)) \Rightarrow (Udział_masy = ma\text{ł}o)$$

$$(Popyt_krajowy = (4760960 \div 5651290)) \& (Kierunek = (292 \div 337)) \& \\ \& (Podaż_nadm = (18556 \div 296935)) \Rightarrow (Udział_masy = ma\text{ł}o)$$

Reguły te wykorzystano do klasyfikacji obiektów z podzbioru testowego. Skuteczność klasyfikacji zależy od złożoności zadania i od jakości oraz liczby danych. Ocenę dokładności klasyfikacji dokonano na podstawie analizy macierzy błędów (tab.). Dla klasy „mało”, z 3852 obiektów do niej należących poprawnie zostało zaklasyfikowanych 2851. 47 razy obiekty z klasy „mało” zostały błędnie przypisane do klasy „średnio”, 18 razy do klasy „dużo”, a 4 razy do klasy „b. dużo”. Dokładność klasyfikacji, określona jako stosunek poprawnie sklasyfikowanych obiektów do wszystkich obiektów sklasyfikowanych z tej klasy, dla klasy „mało” wyniosła 97,6%, co można uznać za wynik bardzo dobry. Współczynnik pokrycia, który określa stosunek obiektów sklasyfikowanych do wszystkich obiektów z danej klasy decyzyjnej, jest równy 75,8%. Ze względu na małą liczbę obiektów treningowych dokładność klasyfikacji dla trzech pozostałych klas nie jest zadowalająca.

Tabela.

Wyniki klasyfikacji obiektów z podzbioru testowego
Results of classification of objects from the test subset

Rzeczywiste	Prognozowane				Liczba obiektów	Dokładność	Pokrycie
	Mało	Średnio	Dużo	B. dużo			
Mało	2851	47	18	4	3852	0,976	0,758
Średnio	77	4	3	1	157	0,047	0,541
Dużo	22	4	0	1	50	0	0,54
B. dużo	6	1	0	0	17	0	0,412

Podsumowanie

Opracowanie przedstawia podstawy teorii zbiorów przybliżonych, potencjalne kierunki wykorzystania tej teorii w leśnictwie oraz przykład zastosowania zbiorów przybliżonych w analizach danych z zakresu leśnictwa.

Zbiory przybliżone posiadają szczególne właściwości, które pozwalają na ich szerokie zastosowanie zarówno w badaniach naukowych, jak i w tworzeniu systemów wspomaganie decyzji. Zbiory przybliżone oferują nowe, proste matematyczne narzędzie opisu niepewności. Analiza tą metodą nie wymaga założeń odnośnie danych. Może być zastosowana jako wstępna, szczególnie w przypadku występowania niepewnych, a nawet częściowo sprzecznych informacji oraz gdy dane nie są dostosowane do analizy tradycyjnymi metodami statystycznymi (założenia o normalności rozkładów). Selekcja istotnych atrybutów spośród wstępnie przyjętych do analizy powoduje, że wyniki nie są zniekształcone przez subiektywny dobór atrybutów. Redukcja wiedzy do niezbędnego minimum następuje przez generowanie zbioru reguł minimalnych. Interpretacja wyników jest łatwa poprzez treść poszczególnych reguł decyzyjnych, których zbiór stanowi opis funkcjonowania układu opisanego tabelą decyzyjną. Przetwarzanie dużych zbiorów danych następuje przy niskich nakładach obliczeniowych.

Na uwagę zasługuje możliwość wykorzystania metody do przetwarzania zarówno danych o charakterze nieprzestrzennym, jak i przestrzennym. Analiza danych przestrzennych metodą ZP powinna być przedmiotem specjalnych badań geomatyki lasu. Teoria zbiorów przybliżonych, oferując nowe matematyczne narzędzie opisu niepewności, może być szczególnie cenna w analizie granic różnych obiektów i analizie obrazów teledetekcyjnych o różnych terenowych rozdzielczościach pikseli.

Literatura

- Aldridge C. 2001. A rough set based methodology for geographic knowledge discovery. Proceedings of the 6th International Conference on GeoComputation. University of Queensland, Australia.
- Bazan J., Nguyen H., Nguyen S., Synak P., Wróblewski J. 2000. Rough set algorithms in classification problem. W: Polkowski L., Tsumoto S., Lin T. [red.]. Rough Set Methods and Applications. Physica-Verlag, Heidelberg New York. 49-88.
- Beaubouef T., Petry F., Ladner R. 2007. Spatial data methods and vague regions: a rough set approach. Applied Soft Computing 7: 425-440.
- Czogala E., Pedrycz W. 1985. Elementy i metody teorii zbiorów rozmytych. PWN, Warszawa.
- Dempster A. P. 1967. Upper and lower probabilities induced by a multivalued mapping. The Annals of Mathematical Statistics 38 (2): 325-339.
- Duch W., Korbicz J., Rutkowski L., Tadeusiewicz R. [red.]. 2000. Sieci neuronowe. W: Biocybernetyka i inżynieria biomedyczna. Exit, Warszawa.
- Duckham M., Lingham J., Mason K., Worboys M. 2006. Qualitative reasoning about consistency in geographic information. Information Sciences 176 (6): 601-627.
- Goldberg D. 1998. Algorytmy genetyczne i ich zastosowania. WNT, Warszawa.

- Gorsevski P., Jankowski P. 2008. Discerning landslide susceptibility using rough sets. *Computers, Environment and Urban Systems* 32: 53-65.
- Klimkiewicz M., Moczulska K. 2008. Zastosowanie zbiorów przybliżonych do analizy satysfakcji klienta serwisu pojazdów. *Inżynieria Rolnicza* 1 (99).
- Klimkiewicz M. 2005. Zastosowanie zbiorów przybliżonych do diagnostyki aparatury paliwowej silników o zapłonie samoczynnym. *Inżynieria Rolnicza* 14.
- Liu J., Zeng Z., Liu H., Wang H. 2011. A rough set approach to analyze factors affecting landslide incidence. *Computers & Geosciences* 37: 1311-1317.
- Murgante B., Las Casas G., Sansone A. 2008. A spatial rough set for extracting periurban fringe. *Revue des Nouvelles Technologies de l'Information* 857: 101-125.
- Panchal V., Gupta G., Saxena P. 2005. Rough Set Framework for Geo-Spatial Information. MAP INDIA-2005, 8th Annual International Conference, GPS, Aerial Photography and Remote Sensing, February 7-9, 2005.
- Pawlak Z. 1982. Rough Sets. *International Journal of Information and Computer Sciences* 11: 344-356.
- Pawlak Z. 2004. Zbiory przybliżone. Nowa matematyczna metoda analizy danych. *Konwersatorium Politechniki Warszawskiej* 5.
- Qi G., Du Y., Cao F. 2010. Extracting the spatial-temporal rules of the mesoscale ocean eddies in the south china sea based on rough sets. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (II).
- Renigier-Biłozor M., Biłozor A. 2008. Zastosowanie teorii zbiorów przybliżonych i teorii zbiorów rozmytych w gospodarce przestrzennej. W: Czyż T., Stryjakiewicz T., Churski P. [red.]. *Biuletyn Instytutu Geografii Społeczno-Ekonomicznej i Gospodarki Przestrzennej UAM w Poznaniu. Seria Rozwój Regionalny i Polityka Regionalna* 3.
- Schneider M. 2008. Fuzzy Spatial Data Types for Spatial Uncertainty Management in Databases. *Handbook of Research on Fuzzy Information Processing in Databases*. 490-515.
- Shafer G. 1976. *A Mathematical Theory of Evidence*. Princeton University Press.
- Shi W., Wang Sh., Li D., Wang X. 2003. Uncertainty based spatial data mining. *Asia GIS 2003 Conference*. October 16-8, 2003. Wuhan, China.
- Synak P. 2003. Temporalne aspekty eksploracji danych: metody zbiorów przybliżonych. Praca doktorska. Instytut Podstaw Informatyki PAN.
- Tadeusiewicz R. 1993. *Sieci neuronowe*. Akademicka Oficyna Wydawnicza RM. Warszawa.
- Tracz W. 2001. Nowoczesne narzędzia wspomaganie procesu decyzyjnego w leśnictwie. *Sylwan* 145 (2): 39-47.
- Tracz W. 2003. Wykorzystanie sieci neuronowych wspomagających SIP w analizach przestrzennych w leśnictwie. *Sylwan* 147 (12): 1-9.
- Tracz W. 2010. Zastosowanie sieci neuronowych i systemów ekspertowych w leśnym GIS. *Geomatyka w Lasach Państwowych*, cz.1 – podstawy, podręcznik. Centrum Informacyjne Lasów Państwowych, Warszawa.
- Tracz W., Mozgawa J. 2006. The conception of fuzzy expert system for management of aesthetic values of forests. W: Barczak A., Tadeusiewicz R. [red.]. *Proceedings of Artificial Intelligence Studies*. Publishing House of University of Podlasie. Siedlce.
- Upadhyaya Sh., Arora A., Jain R. 2006. Rough Set Theory: Approach for Similarity Measure in cluster analysis. *Proceedings of the 2006 International Conference on Data Mining, DMN*. Las Vegas, Nevada, USA.
- Zadeh L. 1965. Fuzzy sets. *Information and Control* 8 (3): 338-353.

SUMMARY

Rough sets and the possibility of their use in forestry

Sharing multi-themes databases related to the implementation of the INSPIRE system significantly complements the information contained in the SILP and creates favorable conditions for support of forestry management and research. New methods of large amounts of data exploration (data mining) and learning (knowledge acquisition), which take into account the uncertainty and vagueness of data used for the analyzes, are looked for. These methods include: the mathematical theory of evidence, fuzzy sets, neural networks, genetic algorithms and rough sets.

The rough sets theory was introduced by a Polish mathematician Zdzisław Pawlak and gained the undisputed world recognition. The need for effective management of uncertainty and incomplete information was the main factor of interest in rough set theory and multi-use

methods of analysis based on this theory, both in scientific research and in solving practical problems.

To represent information about a specific problem in rough set theory the table called an information system is used. Table rows represent objects and columns represent attributes of these objects. A special form of an information system is a decision table, where one column represents a decision attribute, determining the membership of an object to one of the classes. The starting point for the analysis with rough sets is to determine the indiscernibility relation that occurs when different objects have the same attribute values adopted to establish the relationship. Based on indiscernibility relation, for each subsets of an object, the so-called rough set is determined, which consisted of a pair: the lower and upper approximation. Rough set theory is a formalization of the imprecise (vague) concepts in terms of the precise concepts. Vague concept (rough set) is characterized by two precise concepts, namely with lower and upper approximations of this set.

Data mining and knowledge acquisition takes place through the creation of decision rules such as 'if condition then decision' on the basis of values of attributes form the decision table.

It is worth mentioning that rough sets can be used for processing of non spatial and also spatial data. Application of rough set in decision support systems in forestry and forest research consists of five areas:

- Modeling of approximate definition of regions with uncertain boundaries in raster spatial information systems.
- Creation of different classes of objects by modeling of discretization and indiscernibility relations.
- Classification of objects by using decision rules created based on an information system.
- Identify cause-effect relationships by creation and analysis of decision rules.
- Describing spatial topological relations between two regions with uncertain boundaries in order to maintain the uncertainty at all stages of solving the spatial problem.

Cause-effect relationships in the decision table, which describes sales of certain wood assortments by forest districts to customers located in the arbitrary defined zone of influence of these forest districts, were identified in the study. RSES software which implements rough sets method for data processing was used.

The process of building of the decision table, the processing of data given in terms of rough sets and results interpretation are described in the study. Some of the analyzed attributes had a spatial character. Spatial analyses available in Geographical Information System (GIS) were used to determine values of these attributes.