

**WYKORZYSTANIE ZBIORÓW PRZYBLIŻONYCH
DO POZYSKIWANIA WIEDZY I BUDOWY REGUŁ
SYSTEMU GENERALIZACJI
INFORMACJI GEOGRAFICZNEJ**

**IMPLEMENTATION OF ROUGH SETS THEORY FOR
KNOWLEDGE ACQUISITION AND CONSTRUCTION
OF KNOWLEDGE BASE FOR GENERALIZATION
OF GEOGRAPHIC INFORMATION**

Anna Fiedukowicz

Politechnika Warszawska, Wydział Geodezji i Kartografii, Zakład Kartografii

Słowa kluczowe: generalizacja informacji geograficznej, baza wiedzy, zbiory przybliżone, redukty

Keywords: generalization of geographic information, knowledge base, rough sets, reducts

Wprowadzenie

Generalizacja informacji geograficznej, przez wieki wykonywana manualnie przez doświadczonych kartografów, stała się ponownie, w dobie powszechnej komputeryzacji, interesującym wyzwaniem badawczym. Modelowanie informacji geograficznej zostało bowiem przeniesione z formy analogowej (mapy) do cyfrowej (najczęściej bazy danych). Wraz z tą zmianą pojawiło się oczekiwanie coraz większej automatyzacji w zarządzaniu takimi danymi i ich przetwarzaniu. To ogólne oczekiwanie dotyczy też procesu generalizacji, a dążenia do jego jak największej automatyzacji trwają od lat. Badania prowadzone w tym zakresie pozwoliły na opracowanie wielu algorytmów realizujących poszczególne operatory generalizacji (jak upraszczanie, selekcja czy przesunięcie). Do ich poprawnego zastosowania niezbędny jest jednak jeszcze jeden element: baza wiedzy/reguł, która pozwoli podjąć decyzję o wykorzystaniu kolejnych atrybutów i algorytmów oraz ich parametryzacji. Dotychczas tworzone bazy reguł wykorzystują klasyczne reguły oparte o logikę dwuwartościową. W niniejszym artykule zaproponowano metodykę, która ma wspierać akwizycję wiedzy do bazy reguł poprzez wykorzystanie matematycznego konceptu zbiorów przybliżonych oraz reduktów. Tak pozyskana wiedza o zbiorze atrybutów istotnych z punktu widzenia podejmowanej decyzji może zostać wykorzystana później na etapie konstruowania rozmytych reguł tworzących bazę wiedzy systemu generalizacji informacji geograficznej.

Zbiory przybliżone

Idea zbiorów przybliżonych

Koncept zbiorów przybliżonych (ang. *rough sets*) został opracowany przez polskiego uczonego prof. Zdzisława Pawłaka, jako sposób radzenia sobie z nieprecyzyjnymi i niejednoznacznymi informacjami (Pawlak, 1982). Jest to teoria, która dostarcza formalnych, matematycznych narzędzi do opisu wiedzy, która jest niepełna bądź nieprecyzyjna.

Wiedza w teorii zbiorów przybliżonych reprezentowana jest w postaci tzw. tabeli decyzyjnej, stanowiącej podstawowy element systemu informacyjnego. Wiersze takiej tablicy odpowiadają poszczególnym obiektom uniwersum (uniwersum – zbiór wszystkich obiektów zgromadzonych w bazie wiedzy), natomiast kolumny opisującym je atrybutom. Przykład systemu informacyjnego wg teorii zbiorów przybliżonych przedstawia tabela 1. Zaprezentowany przykład zawiera atrybuty binarne (0 lub 1), nie jest to jednak warunkiem koniecznym dla istnienia takiego systemu.

Tabela 1. System decyzyjny zgodnie z teorią zbiorów przybliżonych: e_x – obiekty uniwersum, a_x – atrybuty; wyróżniono 2 obiekty, które są nierozróżnialne od siebie nawzajem

	a_1	a_2	a_3	a_4	a_5
e_1	1	1	0	1	1
e_2	0	0	0	1	1
e_3	0	1	1	0	0
e_4	0	0	0	0	0
e_5	0	1	1	1	1
e_6	1	1	0	1	0
e_7	1	0	1	0	1

Relacja nierozróżnialności i redukty

W ramach tak zdefiniowanego systemu decyzyjnego można zdefiniować relację nierozróżnialności, która opiera się na stwierdzeniu, które z obiektów są (zgodnie z wiedzą zawartą w tablicy decyzyjnej) jednakowe. W powyższym przykładzie (tab. 1) nierozróżnialne są więc następujące grupy obiektów: $\{e_1, e_6\}$, $\{e_2\}$, $\{e_3\}$, $\{e_4\}$, $\{e_5\}$, $\{e_7\}$. Widać więc, że relacja nierozróżnialności może obejmować zarówno jeden jak i więcej obiektów. Tradycyjne systemy decyzyjne budowane w formie tablic nie były przystosowane do radzenia sobie z nierozróżnialnymi (pod względem atrybutów je opisujących) obiektami, dla których należało podjąć różne decyzje (decyzja była przedstawiana jako kolejny, dodatkowy atrybut). Teoria zbiorów przybliżonych pozwala w takim przypadku zarówno na formułowanie niepewnych (przybliżonych) reguł na podstawie takiej bazy wiedzy, jak i na określanie stopnia niepewności decyzji (Bazan, 1998).

Tabela 2. Zredukowany system decyzyjny tabeli 1, pozwalający na rozróżnienie tych samych obiektów

	a_1	a_2	a_4
e_1	1	1	1
e_2	0	0	1
e_3	0	1	0
e_4	0	0	0
e_5	0	1	1
e_6	1	1	1
e_7	1	0	0

Jedną z ważniejszych zalet tej teorii, wykorzystaną w ramach niniejszej pracy, jest jednak umożliwienie redukcji posiadanej wiedzy (Pawlak, 1982; Bazan, 1998). Oznacza to nie tyle dopuszczenie do utraty cennych informacji, co ograniczenie ilości zapisywanych danych i uproszczenie systemu, które nie spowoduje utraty wiedzy, a pozwoli na łatwiejsze i bardziej przejrzyste jej wykorzystywanie. Teoria zbiorów przybliżonych pozwala bowiem wybrać spośród wielu atrybutów te, które mają najistotniejszy wpływ na podejmowaną decyzję. Redukt (ang. *reduct*) stanowi więc zbiór atrybutów, które pozwalają zachować te same co oryginalne atrybuty relacje nierozróżnialności. Dla analizowanego przykładu jeden z reduktów stanowi zbiór atrybutów $\{a_1, a_2, a_4\}$. Dzięki tym atrybutom (tab. 2) nadal jesteśmy w stanie wyróżnić te same grupy obiektów ($\{e_1, e_6\}$,

$\{e_2\}, \{e_3\}, \{e_4\}, \{e_5\}, \{e_7\}$). Zwykle w tablicy decyzyjnej można wyznaczyć więcej niż jeden redukt. Co więcej redukty te mogą zawierać zróżnicowaną liczbę atrybutów (mieć różne długości) i w różnych konfiguracjach (jeśli jakiś atrybut występuje we wszystkich reduktach należy do tzw. jądra (ang. *Core*)).

Możliwości zastosowań

Teoria zbiorów przybliżonych znalazła do tej pory wiele zastosowań praktycznych w dziedzinach takich jak medycyna, przetwarzanie obrazu, ekonomia czy robotyka. Szczególnie sprawdza się ona w przypadku braku precyzyjnych informacji lub niejednoznacznego ich charakteru. Jeden z najbardziej klasycznych przykładów dotyczy zastosowania jej w diagnostyce medycznej. Wykorzystanie konceptu reduktów pozwala na określenie, które badania dadzą informacje niezbędne do postawienia precyzyjnej diagnozy (równie precyzyjnej jak przy użyciu wszystkich badań). Najbardziej pożądane mogą być w tym przypadku redukty najkrótsze (najmniejsza liczba badań), bądź takie, które zawierają badania najtańsze lub najmniej inwazyjne (nawet jeśli liczba badań niezbędna do wykonania będzie w takim przypadku większa).

Idea wykorzystania zbiorów przybliżonych w ramach prac nad automatyzacją procesu generalizacji informacji geograficznej jest zbliżona do tej zastosowanej w diagnostyce medycznej. W związku z ogromem danych gromadzonych obecnie w bazach danych przestrzennych, narzędzie pozwalające selekcjonować informacje istotne dla decyzji podejmowanych w ramach procesu generalizacji wydaje się niezwykle użyteczne.

Autorka pragnie wykorzystać koncept reduktów do wyboru atrybutów, które zostaną wykorzystane w regułach decydujących o przebiegu generalizacji. Aby uzyskać atrybuty decyzyjne należy udzielić odpowiedzi na pytania takie jak:

- czy generalizować dany obiekt (udzielając odpowiedzi binarnej 0 lub 1),
- którym z dostępnych algorytmów,
- jak ten algorytm sparametryzować.

Redukty zostaną wyszukane z uwzględnieniem atrybutów decyzyjnych wymienionych powyżej. Atrybuty spośród których tworzone będą redukty pochodzą zarówno bezpośrednio z bazy danych, jak i są uzyskiwane pośrednio – jako wynik analiz przestrzennych opartych o tę bazę – zawarte w niej informacje na temat geometrii i topologii obiektów.

Idea eksperymentu

Motywacja do podjęcia badań

Wcześniej wykonane eksperymenty dotyczące selekcji obiektów liniowych z wykorzystaniem reguł rozmytych (Fiedukowicz, 2013a, 2013b) wykazały przydatność tego typu podejścia w automatyzacji procesu generalizacji informacji geograficznej. Jednakże jednym z podstawowych problemów, w trakcie tworzenia reguł decyzyjnych był wybór atrybutów, które miały być wykorzystane do tworzenia reguł decyzyjnych. Liczba dostępnych atrybutów była bowiem znaczna. Do atrybutów istniejących oryginalnie w TBD – Bazie Danych Topograficznych (takie jak klasa drogi, kategoria zarządzania, nawierzchnia itp.) dodano nowe, obliczone w oparciu o geometrię i topologię obiektów (np. liczba budynków określona w buforze 50 m, czy powierzchnia obszaru zabudowanego w buforze 100 m od drogi). Takie podejście umożliwiło uwzględnienie przestrzennego kontekstu, który ma istotne zna-

czenie w procesie generalizacji. Utrudniło jednak tworzenie reguł decyzyjnych gdyż niemożliwe stało się wykorzystanie wszystkich atrybutów (powodowałoby to nadmierne skomplikowanie systemu).

Innym problemem występującym we wspomnianych wcześniej eksperymentach, było pozostawanie po selekcji krótkich odcinków jezdni w oderwaniu od całości sieci drogowej. Problem ten występował z uwagi na sposób modelowania dróg w TBD, który powodował ich rozdrobnienie i trudności z zachowaniem spójności sieci. Przy braku zachowania tej spójności bezpośrednio przez algorytm selekcji, selekcja obiektów modelowanych w TBD (odcinków jezdni) nie dawała zadowalających rezultatów.

Dlatego też zdecydowano się powtórzyć wcześniejszy eksperyment (tworzenie rozmytych reguł decyzyjnych dla selekcji dróg), poprzedzając go dodatkowymi działaniami. Z jednej strony było to scalenie odcinków jezdni w dłuższe fragmenty, które bardziej odpowiadają intuicyjnemu rozumieniu pojęcia „droga”. Z drugiej zaś obliczono (za pomocą darmowego oprogramowania RSES) redukt. W ten sposób formalnie wyznaczono zbiory atrybutów istotnych dla procesu selekcji dróg. Następnie stworzono kilka zbiorów reguł rozmytych (każdy ze zbiorów oparty wyłącznie o atrybuty jednego z wyznaczonych reduktów), które przetestowano.

Dane testowe

Jako dane testowe wykorzystano fragment TBD (Bazy Danych Topograficznych, poziom szczegółowości 1: 10 000) z okolic Dukli (rys. 2a), tj. obszaru, na którym wykonywano już wcześniej prace eksperymentalne. Wybrano go aby zapewnić możliwość porównywania wyników.

Obszar testowy zastosowano między innymi do wyznaczenia reduktów oraz tworzenia z nich reguł rozmytych. Wyznaczenie reduktów wymagało znajomości decyzji dla poszczególnych obiektów obszaru dukielskiego. Z tego względu wykorzystano istniejącą mapę topograficzną w skali 1: 50 000 oraz dane z BDO (Bazy Danych Ogólnogeograficznych) z tego obszaru (rys. 2b i c). Szczegóły wykorzystania tych materiałów przedstawiono w podrozdziale „Pozyskanie atrybutu decyzyjnego”.

Zmiana sposobu modelowania sieci drogowej

Przyczyny

Jednym z problemów zauważonych w trakcie wcześniejszych prac (Fiedukowicz, 2013a, 2013b) było zbytne rozdrobnienie generalizowanych odcinków dróg. Wynikało ono ze sposobu modelowania dróg w TBD (klasa „SKJZ”), gdzie każdy odcinek jezdni reprezentowany był jako osobny obiekt. Przy czym do uznania za niezależny odcinek wystarczyła zmiana jednego z atrybutów TBD lub skrzyżowanie z dowolnym innym odcinkiem jezdni (GGK, 2008). We wnioskach z poprzednich prac zaproponowano więc zmianę reprezentacji z odcinków jezdni na drogi. Próbę takiego scalenia podjęto w omawianej pracy.

Zastosowana metodyka

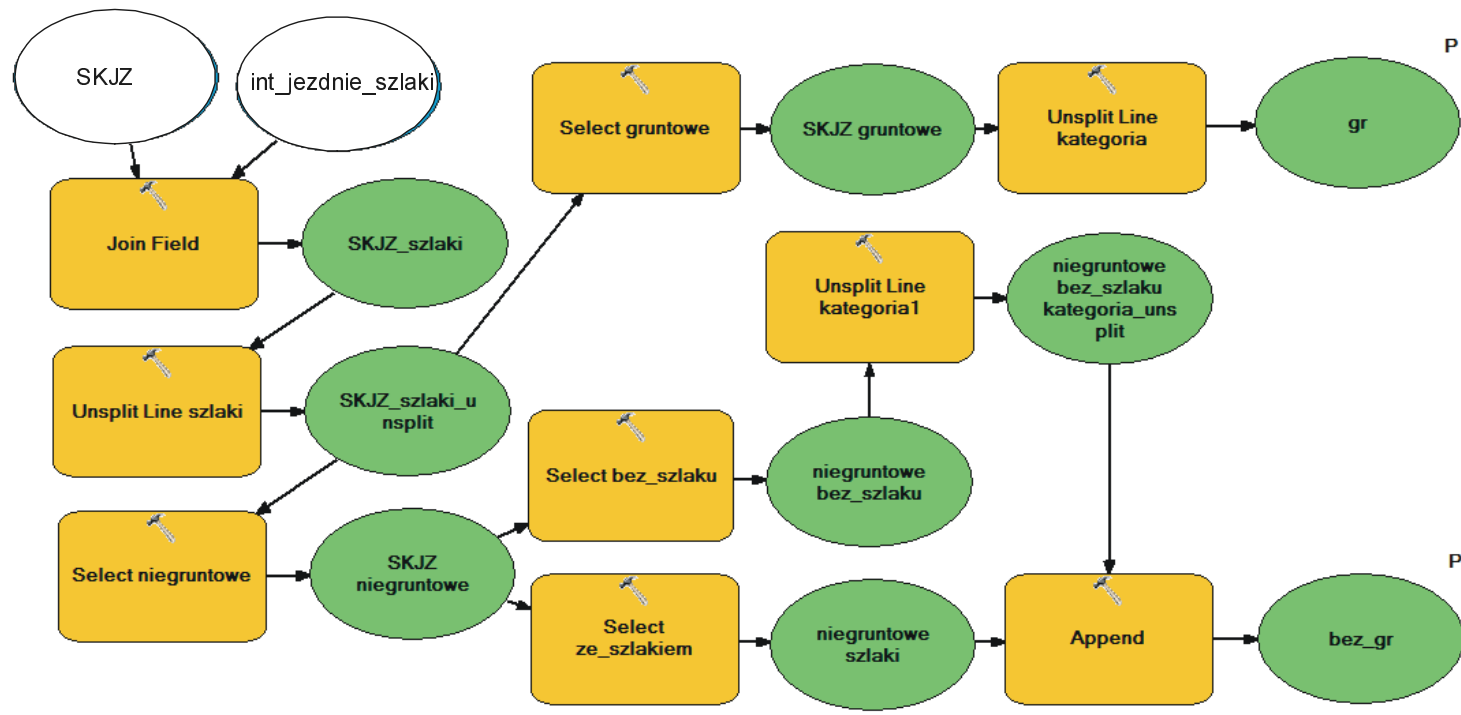
Istnieje wiele sposobów na połączenie odcinków jezdni w drogi. W pierwszej kolejności zdecydowano się na wykorzystanie tabeli intersekcji obecnej w TBD „INT_JEZDNIE_SZLAKI”, która przyporządkowywała poszczególnym odcinkom jezdni nazwy szlaków drogowych, do których przynależą. W ten sposób połączono ze sobą część z odcinków – zgodnie z numerami szlaków drogowych (rys. 3a). Większość odcinków jezdni nie miała jednak przyporządkowanych szlaków. Zdecydowano się więc rozdzielić wynik poprzedniej operacji na drogi, które zawierały odcinki jezdni wyłącznie gruntowe oraz pozostałe drogi, które zawierały przynajmniej jeden odcinek o nawierzchni innej niż gruntowa (rys. 3b).

Drogi gruntowe scalono biorąc pod uwagę kategorię zarządzania. Z grupy dróg o lepszych nawierzchniach wyłączono te, które miały już przypisany szlak drogowy (gdyż wartość ich atrybutu dotyczącego kategorii zarządzania była już zaburzona przez poprzednie scalanie). Pozostałe obiekty także scalono biorąc pod uwagę kategorię zarządzania. Następnie połączono w jeden plik z drogami o przypisanym ID szlaku. Schemat stosowanych scaleń i selekcji przedstawiono na rysunku 1. Ostatecznie uzyskano dwa pliki zawierające scalone drogi: jeden z drogami wyłącznie gruntowymi, drugi zawierający drogi o innym typie nawierzchni. Liczba obiektów została zredukowana w stosunku do oryginalnej (zobacz tabela 3).

Cały proces scalania odcinków jezdni w drogi wykonywano za pomocą programu ArcGIS. Aby uniknąć powtarzalnych interwencji manualnych wykorzystano aplikację ModelBuilder, będącą częścią tego programu. W ramach scalania łączono jedynie odcinki jezdni, które miały wspólny początek lub koniec – nie były dozwolone obiekty złożone z wielu elementów geometrycznych. Funkcja Unsplit wykorzystana w ramach tego działania nie łączy obiektów, jeśli dany koniec należy jednocześnie do więcej niż 2 odcinków (brak jednoznaczności co do tego, które odcinki ze sobą połączyć).

Tabela 3. Liczba obiektów w kolejnych krokach scalania odcinków jezdni na obu obszarach testowych; wyróżniono końcowe wyniki procesu

Lp.	Etap scalania	Dukla
1	Źródłowa klasa SKJZ	2930
2	Liczba obiektów po scaleniu istniejących szlaków drogowych	2061
3	w tym drogi wyłącznie gruntowe	1726
4	drogi (3) po scaleniu wg kategorii zarządzania	1666
5	w tym pozostałe	335
6	w tym drogi należące do szlaku drogowego	53
7	w tym drogi nie należące do szlaku drogowego	282
8	drogi (7) po scaleniu wg kategorii zarządzania	184
9	łącznie drogi (5) po scaleniach	237
10	Suma obiektów w klasach (4) i (9)	1903



Rys. 1. Zastosowany schemat scalania odcinków jezdni (klasa SKJZ TBD) w drogi

Atrybuty opisujące drogi

Poważnym wyzwaniem było zachowanie lub uśrednienie wartości atrybutów w trakcie scalania odcinków jezdni. Przede wszystkim, wymagało to w większości przypadków przejścia ze skali porządkowej na skalę interwałową. Ponieważ takie działanie jest również potrzebne do konstrukcji zmiennych lingwistycznych i tworzenia reguł rozmytych, było ono już wcześniej przedmiotem badań. W uprzednich publikacjach (Fiedukowicz, 2013a, 2013b) sprawdzono jaki wpływ ma przyjęty sposób transformacji między skalami pomiarowymi na wynik procesu generalizacji. Z tego względu w niniejszym badaniu pominięto ten aspekt – przyjęto, że kolejne wartości zmiennych porządkowych różnią się między sobą o 1 na skali interwałowej (i tak np. droga o nawierzchni gruntowej otrzymała atrybut nawierzchnia = 1, zaś droga pokryta masą bitumiczną = 12 – najwyższy z możliwych). Zastosowanie tylko jednego rodzaju przejścia między skalami pozwoliło na obserwację wpływu innych czynników na wynik procesu generalizacji, niezaburzonego transformacją skal.

W momencie scalania dróg atrybuty je opisujące powinny zostać uśrednione. Jednak zastosowanie narzędzia obliczającego średnią arytmetyczną w ramach narzędzia Unsplit nie jest najlepszym rozwiązaniem. Zastosowano więc wagowanie atrybutów, pozwalające uwzględnić długość łączonych odcinków drogi jako wagi (rys. 4). W przypadku niektórych atrybutów zdecydowano się zachować także/tylko maksymalną wartość (np. jeśli jeden z łączonych odcinków jezdni miał charakter przelotowy po dołączeniu innego, zachowywany był ten charakter).

Pozyskanie atrybutu decyzyjnego

Wyznaczanie reduktów może odbywać się na wiele różnych sposobów (szczegóły w rozdziale „Poszukiwanie reduktów”). W tym przypadku zdecydowano się jednak na wyszukiwanie zbiorów atrybutów, które mają wpływ na decyzję o selekcji obiektu. W związku z tym niezbędne było stworzenie atrybutu, który zawierałby informację o tym, czy dany obiekt jest wybierany. W tym celu wykorzystano materiały porównawcze o mniejszych poziomach szczegółowości: mapę topograficzną 1:50 000 oraz BDO (LoD 1:250 000). W sposób manualny nadano atrybuty: 0 – dla obiektów niewybranych w danej skali, 1 – dla obiektów wybieranych (rys. 5). W niektórych przypadkach określenie tego atrybutu (szczególnie dla skali 1:50 000) napotykało pewne problemy. Działo się to w sytuacji, gdy droga powstała w wyniku scalania jezdni, obejmowała zarówno fragmenty drogi obecne, jak i nieobecne na materiałach porównawczych. W takiej sytuacji jako regułę przyjęto, że były zachowywane drogi, których odrzucenie spowodowałoby istotne braki w sieci, nawet kosztem pewnej nadmiarowości w uzyskiwanej w ten sposób sieci drogowej.

W wyniku tego działania uzyskano selekcję dróg, której wyniki przedstawiono w tabeli 4 oraz na rysunku 7a – dla skali 1:50 000 i rysunku 8a – dla skali 1:250 000.

Tabela 4. Liczba obiektów wybranych dla różnych poziomów skalowych – dane porównawcze

Lp.	Selekcja	Liczba obiektów
1	Źródłowa klasa SKJZ	2930
2	Liczba dróg gruntowych przed selekcją	1666
3	selekcja dróg (2) dla skali 1: 50 000	437
4	selekcja dróg (2) dla skali 1: 250 000	0
5	Liczba dróg innych niż gruntowe przed selekcją	237
6	selekcja dróg (5) dla skali 1: 50 000	98
7	selekcja dróg (5) dla skali 1: 250 000	16

Poszukiwanie reduktów

Wykorzystanie programu RSES

Problem wyszukiwania najkrótszego z reduktów jest problemem o dużej złożoności i stanowi tzw. problem np-zupełny (Nguyen, 2011). Analizowany zbiór danych (z okolic Dukli), na podstawie którego wyszukiwano zbiory istotnych atrybutów, był jednak na tyle niewielki, że możliwe było wyszukanie w krótkim czasie wszystkich reduktów.

Do zadania tego wykorzystano darmowe oprogramowanie RSES (ang. *Rough Set Exploration System*) stworzone przez naukowców z Wydziału Matematyki Uniwersytetu Warszawskiego (<http://alfa.mimuw.edu.pl/~rses/>). Program pozwala w łatwy sposób wyszukiwać w zbiorach danych redukty (ale także np. reguły decyzyjne za pomocą metod *data mining*). Umożliwia także generowanie podstawowych statystyk znalezionych reduktów oraz wykresów z nimi związanych.

Wyniki działania programu RSES

Przykłady uzyskanych reduktów

Wyszukiwano reduktów (w programie RSES) z wykorzystaniem funkcji *modulo decision*, tj. takich, które zachowują relację nierozróżnialności względem oczekiwanej decyzji. Dzięki wyszukaniu reduktów udało się ograniczyć liczbę atrybutów niezbędnych do podejmowania decyzji.

Dla dróg gruntowych (i decyzji dla LoD 1:50 000) znaleziono tylko jeden redukt składający się zaledwie z jednego atrybutu – długości drogi. Oznacza to, że decyzję o tym czy dana droga gruntowa powinna być zaprezentowana w tej skali można równie dobrze podjąć z wykorzystaniem wszystkich 26 atrybutów tego zbioru, jak i korzystając zaledwie z jednego.

Dla dróg o nawierzchniach innych niż gruntowe znaleziono więcej i bardziej zróżnicowane redukty. Atrybuty wybrane w ramach kilku z nich (przykładowych) przedstawiono w tabeli 5 (LoD 1:50 000) oraz tabeli 6 (LoD 1:250 000). Widać, że atrybuty tworzą redukty w różnej liczbie i konfiguracji. Istotne jest jednak, że każda z tych grup atrybutów pozwala na równie dobre podjęcie decyzji co początkowe 18 atrybutów przypisanych tym drogom.

Charakterystyka otrzymanych reduktów

Obliczono charakterystyki uzyskanych reduktów (za długość reduktu przyjmuje się liczbę atrybutów tworzących go):

- dla dróg gruntowych (decyzja dla LoD 1:50 000) uzyskano zaledwie 1, jednoelementowy redukt,
- dla pozostałych dróg (LoD 1:50 000) uzyskano 16 reduktów (jedno-, trzy- i cztero-elementowych) o średniej długości 3,2 elementu,
- dla pozostałych dróg (LoD 1:250 000) uzyskano 87 reduktów (zawierających 1-6 elementów) o średniej długości 3,3 elementu, większość z nich o długości 3 elementów (rys. 6a).

Warto zauważyć znaczące obniżenie liczby atrybutów potrzebnych do podjęcia decyzji o selekcji obiektu. W przypadku dróg gruntowych liczba ta spadła z 26 do zaledwie 1 atrybutu, natomiast w przypadku pozostałych dróg z 18 do najwyżej 6 (a zwykle do ok. 3). Tak duża

Tabela 5. Atrybuty tworzące przykładowe redukty – dla skali 1: 50 000, dla dróg innych niż gruntowe

Redukty	Atrybuty
1	długość obiektu
2	śr. szerokość korony
	maks. nawierzchnia
	pow. zabudowana w 100 m
3	śr. szerokość nawierzchni
	pow. zabudowana w 100 m
	śr. nawierzchnia
4	maks. kategoria
	pow. zabudowana w 100 m
	śr. nawierzchnia
5	pow. zabudowana w 100 m
	śr. kod
	śr. nawierzchnia
6	liczba pasów
	śr. szerokość korony
	pow. zabudowana w 100 m
	śr. kategoria

Tabela 6. Atrybuty tworzące przykładowe redukty – dla skali 1: 250 000, dla dróg innych niż gruntowe

Redukt	Atrybuty
1	długość obiektu
2	śr. szerokość nawierzchni
	liczba budynków w 50 m
3	liczba pasów
	śr. szerokość nawierzchni
	maks. kod
4	maks. klasa
	liczba budynków w 50 m
	śr. nawierzchnia
5	liczba pasów
	maks. kategoria
	pow. zabudowana w 50 m
6	śr. szerokość nawierzchni
	śr. klasa
	śr. nawierzchnia
7	maks. nawierzchnia
	liczba budynków w 50 m
	śr. kod
	decyzja 1: 50 000

redukcja liczby atrybutów znacząco ułatwia kolejny krok, jakim jest konstruowanie reguł decyzyjnych.

Warto zwrócić uwagę, które z atrybutów okazały się najbardziej istotne dla podjęcia decyzji o selekcji. Poniżej w nawiasach podano po 2 liczby: liczbę reduktów, w których występuje dany atrybut oraz liczbę znalezionych reduktów. Atrybutami tymi okazały się:

- dla dróg gruntowych: długość drogi (1/1),
- dla pozostałych dróg (1:50 000): powierzchnia obszarów zabudowanych w buforze 100 m (15/16), średnia szerokość korony (8/16), maksymalna kategoria odcinka jezdni scalonego w drodze (8/16) (rys. 6b),
- dla pozostałych dróg (1:250 000): średnia liczba pasów (35/87), średnia szerokość korony (32/87), średnia szerokość nawierzchni (27/87), najlepsza nawierzchnia (22/87), średnia nawierzchnia (22/87), decyzja dla skali 1:50 000 (22/87).

Wykorzystanie znalezionych reduktów

Tworzenie reguł

Wyznaczone w poprzednim kroku redukty stanowią wiedzę, która może być wykorzystana przy budowie reguł decyzyjnych systemu generalizacji. Reguły te można budować na wiele sposobów, najpopularniejszym z nich są reguły typu „Jeżeli ... to ... „ (ang. *IF ... THEN ...*) oparte o logikę dwuwartościową. Możliwa jest jednak także budowa systemu wykorzystującego logikę nieklasyczną. Poniższy przykład dotyczy wykorzystania wiedzy o reduktach przy konstrukcji reguł rozmytych (ang. *fuzzy interference system* – FIS). Pokazuje on jak bardzo zastosowana metodyka, wykorzystująca teorię zbiorów przybliżonych, wpływa na sposób tworzenia oraz wyniki działania systemu generalizacyjnego. Reguły były tworzone przy wykorzystaniu Fuzzy Logic Toolbox programu MatLab. Więcej szczegółów dotyczących działania i tworzenia reguł rozmytych przedstawiła autorka w (Fiedukowicz, 2013a, 2013b).

Porównanie wyników (LoD 1:50 000)

W poprzednich próbach, które nie wykorzystywały reduktów (Fiedukowicz 2013a, 2013b), atrybuty do decyzji wybierane były samodzielnie przez tworzącego system. Przykład wyniku działania takiego zestawu reguł przedstawiono na rysunku 7c. Był on oparty o następujące atrybuty:

- kategoria zarządzania,
- rodzaj nawierzchni,
- powierzchnia terenów zabudowanych w buforze 50 m,
- liczba budynków w buforze 100 m.

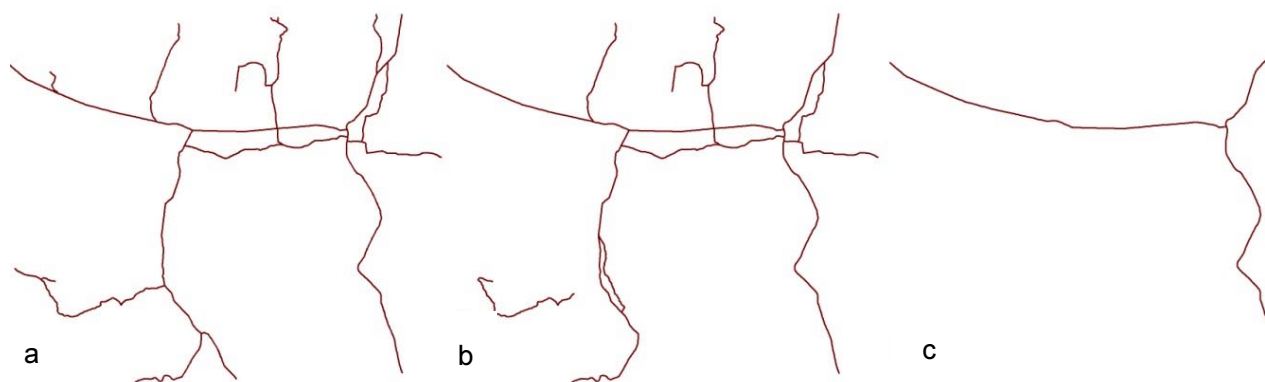
W porównaniu ze wcześniejszymi próbami, tworzenie reguł z wykorzystaniem wyznaczonych reduktów było łatwiejsze, ze względu na znacząco mniejszą liczbę zmiennych wejściowych (atrybutów istotnych dla decyzji). Wykorzystanie reduktów, nie tylko znacząco ułatwiło tworzenie reguł, ale także poprawiło wyniki selekcji obiektów dokonywanej za pomocą FIS. Wyniki widoczne na rys. 7b są znacznie bliższe ideałowi (czyli selekcji manualnej z wykorzystaniem źródłowych materiałów kartograficznych – rys. 7a). Wykorzystano tu dwa systemy rozmyte:

- dla dróg gruntowych – system oparty wyłącznie o długość drogi (generalnie wybierane były drogi dłuższe);
- dla pozostałych dróg – system oparty o następujące atrybuty (redukt 3 w tab. 5):
 - średnia szerokość nawierzchni,
 - powierzchnia terenów zabudowanych w buforze 100 m,
 - średni rodzaj nawierzchni.

Podsumowując, wykorzystano taką samą liczbę atrybutów co w przypadku systemu nie wykorzystującego reduktów, natomiast wynik znacząco się poprawił. Warto podkreślić, że system, którego wyniki zaprezentowano na rysunku 7c, był jednym z najlepszych uzyskanych w trakcie licznych testów (Fiedukowicz, 2013a, 2013b).



Rys. 7. Drogi wybrane dla poziomu szczegółowości 1:50 000: a – manualnie wg danych porównawczych, b – za pomocą reguł rozmytych utworzonych z wykorzystaniem reduktów, c – za pomocą reguł rozmytych z wyborem atrybutów przez operatora



Rys. 8. Drogi wybrane dla poziomu szczegółowości 1: 250 000: a – manualnie wg danych porównawczych, b – za pomocą reguł rozmytych utworzonych z wykorzystaniem reduktów, c – za pomocą reguł rozmytych z wyborem atrybutów przez operatora

Porównanie wyników (LoD 1:250 000)

Analogiczne porównanie przeprowadzono dla poziomu szczegółowości 1:250 000. Rysunek 8b przedstawia wyniki uzyskane dzięki wykorzystaniu reduktu złożonego zaledwie z jednej wartości, tj. długości drogi. Wynik jest niemal identyczny z tym uzyskanym dzięki materiałom źródłowym (rys. 8a). Z kolei wynik jednego z systemów tworzonych bez wykorzystywania reduktów przedstawia rysunek 8c. System ten wykorzystał następujące atrybuty:

- kod TBD,
- kategoria zarządzania,
- klasa drogi,
- rodzaj nawierzchni.

Widać, że uzyskany w ten sposób wynik jest zbyt zgeneralizowany w stosunku do przyjętego poziomu szczegółowości. Mimo wykorzystania większej liczby atrybutów, uzyskany rezultat jest gorszy. Do porównania wybrano najlepszy, z wielu testowanych dla tej skali systemów decyzyjnych, utworzony bez pomocy reduktów (Fiedukowicz, 2013a, 2013b) – rys. 8c. Większość pozostałych dla podanej skali nie wybiera żadnych dróg lub bardzo niewielkie ich odcinki.

Podsumowanie i wnioski

Scalanie odcinków jezdni w drogi

Jednym z elementów wpływających na przebieg i wynik procesu selekcji dróg jest sposób scalania odcinków jezdni w drogi, który poprawił wyniki generalizacji, znacząco ograniczając problem z krótkimi odcinkami dróg, wybieranymi pomimo braku łączności z całą siecią. Przyjęta metodyka nie jest jedyną możliwą i ma z pewnością pewne mankamenty. Jej istotną zaletą jest natomiast automatyzm (osiągnięty dzięki zastosowaniu pakietu ModelBuilder) oraz wykorzystanie dobrze zdefiniowanych i powszechnie znanych narzędzi (wbudowanych w programie ArcGIS). Innym rozwiązaniem mogłoby być zastosowanie metodyki rozwijanej w innych ośrodkach badawczych (np. S.A. Benz, R. Weibel). Implementacja takich bardziej złożonych metod jest jednak trudniejsza.

Z punktu widzenia scalania odcinków jezdni niezwykle istotne wydaje się też definiowanie szlaków drogowych na etapie wypełniania TBD. To właśnie wykorzystanie tego elementu pozwoliło na najskuteczniejsze tworzenie długich, mających odzwierciedlenie w rzeczywistości, dróg.

W trakcie scalania nie można też zapominać o odpowiednim „uśrednianiu” atrybutów łączonych odcinków. Wykorzystana metodyka wagowania atrybutów długością odcinka wydaje się spełniać swoje zadanie. Warto też podkreślić, że w niektórych reduktach znalazły się atrybuty określające maksymalną wartość danego atrybutu spośród wszystkich łączonych odcinków. Warto więc zachować również te informacje.

Wyszukiwanie reduktów i tworzenie reguł

Wyszukiwanie reduktów, wykonywane w celu wyznaczenie zbiorów atrybutów, które wpływają na decyzję, wymaga znajomości tej decyzji. W związku z tym konieczne było pozyskanie atrybutu decyzyjnego (w formie binarnej) dla dwóch wybranych poziomów ska-

lowych. Krok ten jest niezwykle czasochłonny (manualne pozyskiwanie atrybutu), zatem nie powinien być konieczny do powtarzania w ramach działania systemu, a jedynie w ramach jego tworzenia i testów. Na tych etapach daje on jednak szanse niejawnego wykorzystania i niejako „wbudowania w system” wiedzy i doświadczenia kartografa, który tworzył materiały porównawcze. Jednocześnie, znajomość „prawdziwego” atrybutu decyzyjnego, znacząco zwiększa możliwości ewaluacji wyników procesu generalizacji oraz umożliwia stworzenie sprzężenia zwrotnego, dzięki któremu system może być ulepszany. Będzie to przedmiotem dalszych badań.

Analiza uzyskanych reduktów pokazuje, że różne atrybuty są istotne dla procesu selekcji dróg w zależności od tego, czy brane są pod uwagę drogi gruntowe czy też pozostałe drogi. Różne też zestawy atrybutów wybierane są w zależności od przyjętego poziomu skalowego, choć w tym przypadku można doszukać się pewnego podobieństwa. Ze względu na to podobieństwo, w przyszłych pracach zostaną podjęte próby stworzenia wspólnego zestawu reguł dla różnych skal. Stosowanie reduktów znacząco ułatwia formułowanie reguł decyzyjnych. Co więcej – reguły dzięki nim sformułowane dają lepsze rezultaty niż reguły formułowane bez wcześniejszego wyznaczenia reduktów. Dzieje się tak, ponieważ ciężar wyboru atrybutów został przełożony z twórcy systemu na precyzyjne, matematyczne narzędzie, jakim jest teoria zbiorów przybliżonych. Znalezienie zależności między 3 atrybutami a wynikiem jest dla twórcy systemu dużo łatwiejsze, niż szukanie ich spośród kilkunastu potencjalnych atrybutów.

Wnioski końcowe

Przeprowadzone badania są rozwinięciem i kontynuacją wcześniejszych badań (Fiedukowicz, 2013a, 2013b) nad wykorzystaniem logik nieklasycznych w generalizacji informacji geograficznej. Obydwa wprowadzone tu ulepszenia wpłynęły na poprawę wyników działania systemu selekcji dróg. Określenie precyzyjnego wpływu każdego z nich (łączenia odcinków dróg i wykorzystania reduktów) może być sprawdzone przez kolejne eksperymenty, wykorzystujące osobno każdą z tych metod i porównanie otrzymanych rezultatów. Działanie takie będzie przedmiotem dalszych prac badawczych.

Selekcja dróg zamiast odcinków jezdni, które są bezpośrednio modelowane w TBD, wydaje się znacząco lepszym rozwiązaniem, gdyż ułatwia ona zachowanie ciągłości obiektów. Jest to zgodne z ideą, aby generalizować obiekty, które są przedmiotem modelowania kartograficznego (tu: drogi), nie zaś jedynie ich geometryczną reprezentację (w tym przypadku: linie oznaczające odcinki jezdni) (Weibel, 1991).

Niniejsze badania pokazały też, że akwizycja wiedzy w postaci reduktów, wykorzystująca koncept zbiorów przybliżonych, nie tylko ułatwia konstruowanie systemu reguł generalizacyjnych, ale także zdecydowanie poprawia ich skuteczność. Wyniki uzyskane za pomocą reguł opartych na atrybutach z reduktów były znacząco lepsze od wcześniej uzyskiwanych – z pominięciem tego kroku.

Literatura

- Bazan J., 1999: Metody wnioskowań aproksymacyjnych dla syntezy algorytmów decyzyjnych. Rozprawa doktorska, Uniwersytet Warszawski.
- Benz S.A., Weibel R., 2013: Road Network Selection Using an Extended Stroke-Mesh Combination Algorithm. 16th ICA Generalisation Workshop.
- Fiedukowicz A., 2013a: Fuzzy Generalization Inference System – the example of selection parameterization for roads and hydrographic network. Proceedings of International Cartographic Conference Dresden 2013.
- Fiedukowicz A., 2013b: Construction of fuzzy interference system for generalization of geographic information – selection of roads segments. *Geoinformatica Polonica* 12, przyjęty do publikacji.
- Główny Geodeta Kraju, 2008: Wytyczne techniczne Baza Danych Topograficznych, wersja 1.0 – uzupełniona.
- Nguyen H.S., 2011: Systemy decyzyjne (matematyka stosowana). Dostęp 18.07.2013 r. <http://mst.mimuw.edu.pl/wyklady/syd/wyklad.pdf>
- Pawlak Z., 1982: Rough Sets. *International Journal of Computer and Information Sciences* 11: 341-356. Program RSES i jego dokumentacja. <http://alfa.mimuw.edu.pl/~rses/>
- Weibel R., 1991: Amplified intelligence and rule-base systems. [w:] Buttenfield B., McMaster R. (red.), Map generalization: making rules for knowledge representation, Longman, London.

Abstract

Generalization of geographic information was always based on the knowledge and experience of cartographer who performed it. Progressive technological changes, especially the change in data storage from analog to electronic devices (usually in the form of databases), changed a lot in generalization process. Though, the knowledge needed for it is still crucial, it is stored in different way as well. Today, knowledge bases are built based on years of cartographer's experience. Their goal is to enable, at least partially, automation of the generalization process.

This knowledge base for modern spatial data may be constituted, among others, by information about attributes of objects which are significant from the point of view of generalization process as well as the generalization rules based on those attributes. Contemporary spatial databases include a number of attributes and other information (as geometry and topology) which can be used to calculate other databases. Therefore, the challenge is to bring to light information crucial in generalization process. This paper presents the use of rough sets concept, specifically reducts, for this goal. The reduct is a subset of the original set of attributes which allows to make decisions (in this case, the decision about generalization of objects) as good as based on original attributes.

Knowledge about the attributes significant in generalization process is derived, in this way, from already existing data. Selected attributes can be used for defining the generalization rules. Thanks to the use only of the attributes constituting reducts, the construction of rules becomes easier and the rules themselves, as well as generalization process, are more transparent.

mgr inż. Anna Fiedukowicz
a.fiedukowicz@gjk.pw.edu.pl
tel.+48 22 234 5590



Rys. 2: a – dane źródłowe TBD – odcinki jezdni w klasie SKJZ; dane porównawcze o mniejszej szczegółowości: b – fragment mapy 1: 50 000, c – dane z BDO – drogi



Rys. 3: a – wizualizacja szlaków drogowych, kolorem szarym oznaczono odcinki jezdni których nie przypisano do żadnego ze szlaków, b – fioletowy – drogi o przypisanym szlaku drogowym, czerwony – drogi o nawierzchni innej niż gruntowa, szary – drogi gruntowe

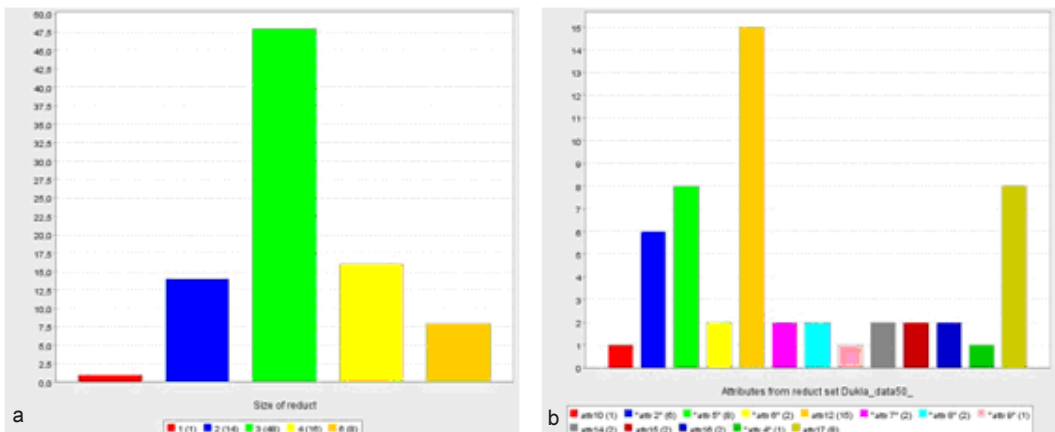


kat. z.	dł.	klasa	naw	sz. n.	śr.k. z.	dł.	śr. kl.	śr. n.	max n.	śr.sz.n.
3	100	2	8	2,5	3	220	3,2	10,2	12	3,9
3	120	5	12	5	4	100	5	12,0	12	7,0
4	100	5	12	7						

Rys. 4. Sposób działania scalania odcinków dróg (względem kategorii zarządzania) z uwzględnieniem wartości atrybutów; przedstawiono kilka wybranych atrybutów, które uśredniono z uwzględnieniem długości scalanych odcinków



Rys. 5. Nadanie atrybutu decyzyjnego: a – LoD 1: 50 000 (materiały źródłowe w formie rastra mapy topograficznej), b – LoD 1: 250 000 (drogi w BDO – brązowe, szerokie)



Rys. 6: a – długości reduktów dla dróg niegruntowych LoD 1: 250 000, b – udział poszczególnych atrybutów w reduktach dróg niegruntowych LoD 1: 50 000