

IDENTYFIKACJA RODZAJU SZKŁA Z ZASTOSOWANIEM TEORII ZBIORÓW PRZYBLIŻONYCH W PROGRAMIE RSES

Piotr Januszewski

Uniwersytet Kazimierza Wielkiego
Wydział Informatyki
ul. Kopernika 1, 85-074 Bydgoszcz
e-mail: piotr.januszewski@student.ukw.edu.pl

Streszczenie: *Artykuł przedstawia podejście do identyfikacji rodzaju szkła oparte na teorii zbiorów przybliżonych w programie RSES. Przedstawiono teoretyczne podstawy tej metody, opisano proces analizy danych oraz zaprezentowano wyniki identyfikacji rodzaju szkła.*

Słowa kluczowe: *analiza danych, rodzaj szkła, teoria zbiorów przybliżonych, reguły decyzyjne, RSES*

Glass Type Identification Using Rough Sets Theory in the RSES Program

Abstarct: *The article presents an approach to glass type identification based on rough set theory in the RSES program. The theoretical basis of this method is presented, the data analysis process is described and the results of glass type identification are presented.*

Keywords: *data analysis, glass type, rough sets theory, decision rules, RSES*

1. WSTĘP

Precyzyjna identyfikacja rodzaju szkła ma znaczenie w wielu dziedzinach, od analizy dowodów w kryminalistyce po optymalizację procesów w przemyśle szklarskim. Artykuł prezentuje podejście oparte na teorii zbiorów przybliżonych w programie RSES, skupiając się na praktycznym wykorzystaniu tej metody do identyfikacji rodzaju szkła.

2. METODY

W teorii zbiorów przybliżonych, kluczowym pojęciem jest relacja równoważności, związana z każdym zbiorem atrybutów. Relację równoważności dla zbioru atrybutów B zawierającego się w A można zapisać jako:

$$I(B) = \{(x, x') \in U^2 \mid a(x) = a(x') \text{ dla } a \in B\}$$

Wynika stąd, że elementy x i x' należące do relacji $I(B)$ są nierozróżnialne i tożsame. Warunki konieczne dla relacji równoważności to zwrotność, symetryczność i przechodniość.

Atrybuty dzielimy na warunkowe, na podstawie których podjęta zostanie decyzja, oraz decyzyjne, zawierające informacje o klasyfikacji. Kolejnym elementem są klasy abstrakcji, będące zbiorami nierozróżnialnych elementów wewnątrz badanego zbioru.

W teorii zbiorów przybliżonych istotne są również przybliżenia górne, dolne i obszar brzegowy. Przybliżenie dolne zawiera jednoznacznie definiowalne elementy, natomiast przybliżenie górne zawiera elementy, których nie da się sklasyfikować jednoznacznie.

W przestrzeni aproksymacyjnej wskaźnik liczbowy przybliżenia zbioru to:

$$L(B) = \frac{|B|}{|B'|}$$

Podczas klasyfikacji obiektów, można napotkać różne sytuacje, takie jak niesklasyfikowany obiekt pasujący do jednej deterministycznej reguły, niesklasyfikowany obiekt pasujący do jednej niedeterministycznej reguły, lub niesklasyfikowany obiekt pasujący do kilku reguł. W każdym przypadku podejmowane są odpowiednie kroki w celu sklasyfikowania obiektu.

3. OPIS ZESTAWU DANYCH.

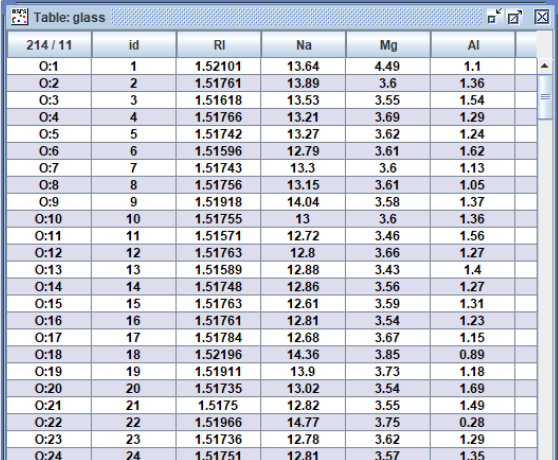
Zestaw danych został zebrany w wyniku testu porównawczego przeprowadzonego przez Vina, który oceniał skuteczność jej systemu opartego na regułach, BEAGLE, w porównaniu z algorytmem najbliższego sąsiada (NN) oraz analizą dyskryminacyjną (DA).

W badaniu oceny skuteczności klasyfikacji szkła jako typu "float" lub innego, uzyskano poniższe wyniki, wyrażone w liczbie nieprawidłowych odpowiedzi:

Rodzaj próbki	BEAGLE	NN	DA
Okna poddane procesowi float (87)	10	12	21
Okna niepoddane procesowi float (76)	19	16	22

4. ANALIZA ZA POMOCĄ PROGRAMU RSES

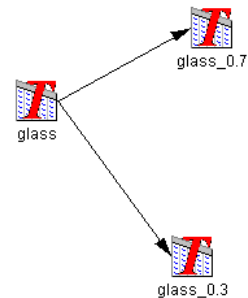
Przed rozpoczęciem pracy zbior danych będący w pliku .data przekonwertowano za pomocą informacji zamieszczonych w pliku .names na plik tablicy programu RSES (.tab). Następnie utworzono w programie nowy projekt i wczytano dane do nowej tablicy.



214 / 11	id	RI	Na	Mg	Al
O:1	1	1.52101	13.64	4.49	1.1
O:2	2	1.51761	13.89	3.6	1.36
O:3	3	1.51618	13.53	3.55	1.54
O:4	4	1.51766	13.21	3.69	1.29
O:5	5	1.51742	13.27	3.62	1.24
O:6	6	1.51596	12.79	3.61	1.62
O:7	7	1.51743	13.3	3.6	1.13
O:8	8	1.51756	13.15	3.61	1.05
O:9	9	1.51918	14.04	3.58	1.37
O:10	10	1.51755	13	3.6	1.36
O:11	11	1.51571	12.72	3.46	1.56
O:12	12	1.51763	12.8	3.66	1.27
O:13	13	1.51589	12.88	3.43	1.4
O:14	14	1.51748	12.86	3.56	1.27
O:15	15	1.51763	12.61	3.59	1.31
O:16	16	1.51761	12.81	3.54	1.23
O:17	17	1.51784	12.68	3.67	1.15
O:18	18	1.52196	14.36	3.85	0.89
O:19	19	1.51911	13.9	3.73	1.18
O:20	20	1.51735	13.02	3.54	1.69
O:21	21	1.5175	12.82	3.55	1.49
O:22	22	1.51966	14.77	3.75	0.28
O:23	23	1.51736	12.78	3.62	1.29
O:24	24	1.51751	12.81	3.57	1.35

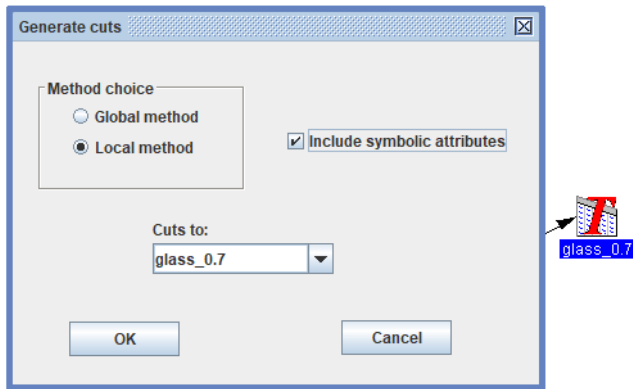
Rysunek 1. Wczytanie danych do programu

Następnie podzielono tablicę na dwie części za pomocą opcji „Split in two” w proporcji 30-70 w celu uzyskania zbioru uczącego i zbioru weryfikującego.



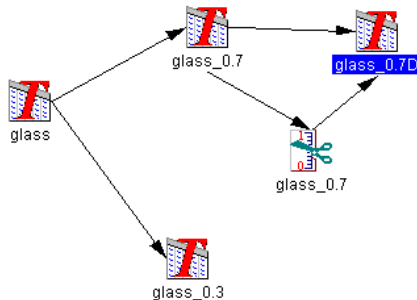
Rysunek 2. Podział tabeli na dwie

Kolejną operacją było wycięcie cięć dla tablicy uczącej z wykorzystaniem opcji „Discretize” > „Generate Cuts”



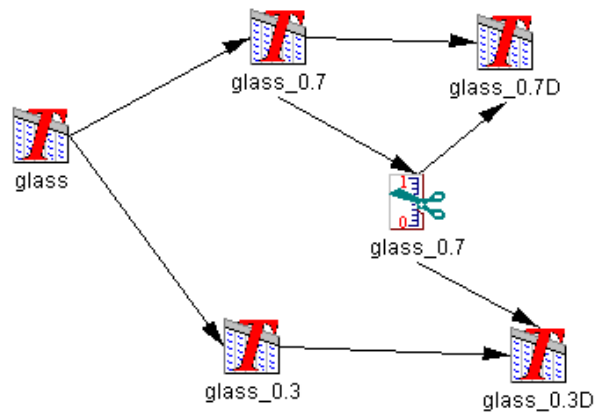
Rysunek 3. Generowanie cięć dla tabeli glass_0.7

Następną czynnością było wykonanie dyskretyzacji tabeli. Wykorzystano do tego opcję: „Discretize” > „Discretize Table”



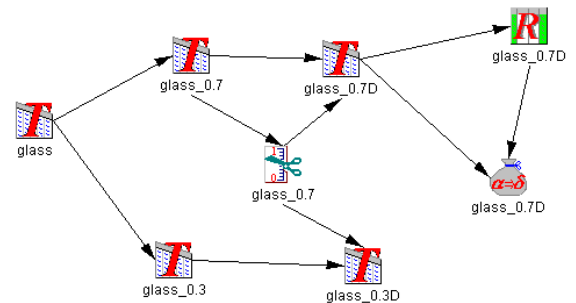
Rysunek 4. Wynik dyskretyzacji tabeli glass_0.7

Kolejnym krokiem było przeprowadzenie dyskretyzacji drugiej tabeli. W tej operacji również użyto cięcia wygenerowane z tabeli zawierającej 70% rekordów.



Rysunek 5. Wynik dyskretyzacji tabeli glass_0.3

Następnie przystąpiono do wygenerowania reduktów z wykorzystaniem tablicy dyskretyzowanej i algorytmu generycznego po czym wykorzystano ten zbiór do wygenerowania zestawu zasad.



Rysunek 6. Wynik wygenerowania zestawu zasad

Na koniec przeprowadzono klasyfikację z użyciem zestawu reguł wykorzystując opcję „Classify” > „Test table using rule set”

		Predicted									
		1	2	3	5	6	7	No. of obj.	Accuracy	Coverage	
Actual	1	0	18	0	0	0	0	18	0	1	
	2	0	23	0	0	0	0	23	1	1	
	3	0	6	0	0	0	0	6	0	1	
	5	0	3	0	0	0	0	3	0	1	
	6	0	8	0	0	0	0	8	0	1	
	7	0	9	0	0	0	0	9	0	1	
	True positive rate		0	0.35	0	0	0	0			

Total number of tested objects: 65
Total accuracy: 0.354
Total coverage: 1

Rysunek 7. Tabela przedstawiająca wyniki klasyfikacji

5. PODSUMOWANIE I WNIOSKI

Artykuł przedstawia skuteczne wykorzystanie teorii zbiorów przybliżonych w identyfikacji rodzaju szkła przy użyciu programu RSES. Wyniki eksperymentów potwierdzają efektywność tej metody, szczególnie w porównaniu z innymi algorytmami takimi jak BEAGLE, algorytm najbliższego sąsiada (NN) i analiza dyskryminacyjna (DA). Proces analizy danych w programie RSES obejmujący konwersję, dyskretyzację, generację reduktów i zestawu zasad okazał się skuteczny w identyfikacji rodzaju szkła. Istotne jest także potencjalne zastosowanie tej metody w dziedzinie kryminalistyki, gdzie analiza szkła może być kluczowym elementem dowodowym.

Literatura

1. Pawlak, Z. (1982). Rough Sets. *International Journal of Computer and Information Sciences*, 11(5), 341–356.
2. Pawlak, Z., & Skowron, A. (2007). Rudiments of rough sets. *Information Sciences*, 177(1), 3–27.
3. Slowinski, R., & Vanderpooten, D. (2000). A generalized definition of rough approximations based on similarity. *IEEE Transactions on Knowledge and Data Engineering*, 12(2), 331–336.

