

Raport

Temat: Reguły decyzyjne w Rough Set Exploration System

Wykonał: Artur Kompała

Zadanie 1 - Przygotowanie Zbioru Danych

Opis: Przygotuj wybrany zestaw danych w formacie RSES (jeśli to możliwe, użyj zestawów danych z poprzedniego laboratorium).

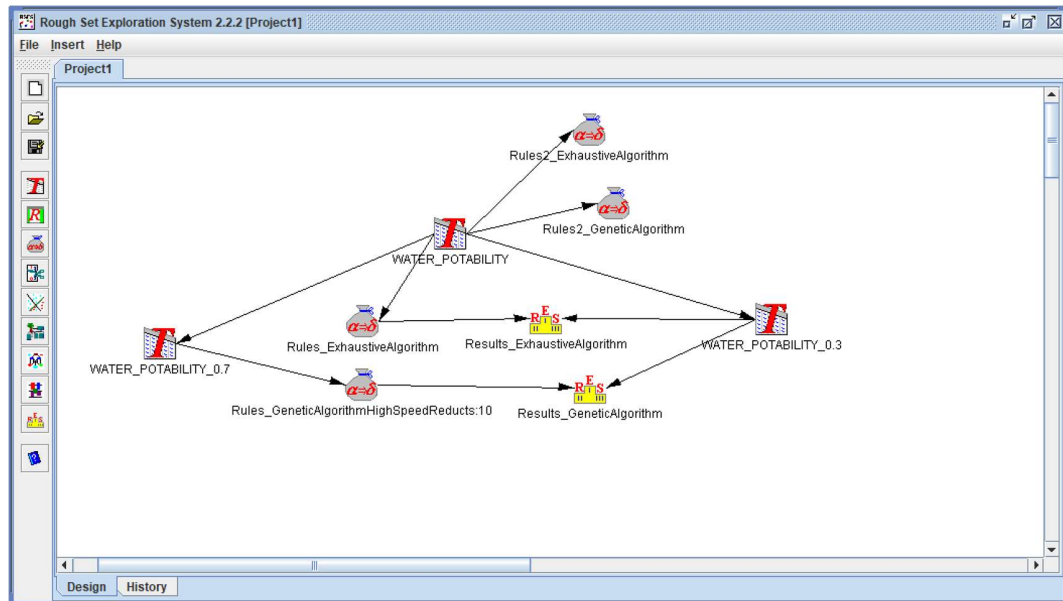
Zbiór danych: WATER_POTABILITY

- Liczba przykładów: 419 rekordów
- Liczba atrybutów warunkowych: 9 (ph, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic_carbon, Trihalomethanes, Turbidity)
- Atrybut decyzyjny: Potability (wartości: 0, 1)
- Format: RSES (dyskretyzacja wartości na zakresy: range1, range2)

419 / 10	ph	Hardn...	Solids	Chlora...	Sulfate	Condu...	Organ...	Trihal...	Turbid...	Potabi...
O:1	range2	range2	range1	range2	range2	range2	range1	range2	range1	0
O:2	range1	range1	range1	range1	range2	range2	range1	range1	range2	0
O:3	range2	range2	range1	range2	range2	range1	range2	range2	range1	0
O:4	range2	range2	range1	range2	range2	range1	range2	range2	range2	0
O:5	range2	range1	range1	range1	range2	range1	range1	range1	range1	0
O:6	range1	range2	range1	range2	range2	range1	range1	range1	range1	0
O:7	range2	range2	range1	range2	range2	range1	range1	range2	range1	0
O:8	range2	range2	range1	range1	range1	range2	range1	range2	range2	0
O:9	range2	range1	range1	range2	range1	range1	range1	range1	range1	0
O:10	range2	range2	range1	range2	range2	range2	range2	range2	range2	0
O:11	range2	range1	range2	range2	range2	range1	range2	range2	range1	0
O:12	range2	range1	range1	range1	range1	range1	range2	range2	range1	0
O:13	range2	range1	range1	range2	range1	range1	range2	range2	range2	0
O:14	range2	range2	range1	range1	range2	range1	range1	range2	range2	0
O:15	range1	range2	range2	range2	range2	range2	range1	range2	range2	0
O:16	range2	range2	range2	range2	range2	range1	range2	range1	range2	0
O:17	range2	range2	range1	range2	range2	range2	range1	range2	range2	0
O:18	range2	range2	range1	range1	range2	range1	range1	range2	range1	0
O:19	range2	range2	range1	range1	range2	range2	range1	range2	range1	0
O:20	range2	range2	range1	range2	range2	range2	range2	range1	range2	0

Przygotowane zbiory:

- WATER_POTABILITY (pełny zbiór: 419 rekordów)
- WATER_POTABILITY_0.7 (zbiór treningowy: 293 rekordów, 70% danych)
- WATER_POTABILITY_0.3 (zbiór testowy: 126 rekordów, 30% danych)



Zadanie 2. - Generowanie reguł

Opis: Dla całego zbioru danych wygeneruj reguły decyzyjne przy użyciu dwóch wybranych algorytmów

2.1 Zidentyfikuj parametry algorytmów oraz wskaż, jaki typ podejścia do indukcji reguł reprezentują (dokładny, sekwencyjne pokrywanie, ewolucyjny)

Algorytm 1: Exhaustive Algorithm

Typ podejścia: Algorytm dokładny

Parametry:

Liczba wygenerowanych reguł: 512

Rule set: Rules2_ExhaustiveAlgorithm		
(1-512)	Match	Decision rules
1	32	(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Organic_carbon=range2)=>(Potability={0[32]})
2	16	(Hardness=range2)&(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Trihalomethanes=range2)=>(Potability={0[16]})
3	16	(ph=range2)&(Hardness=range2)&(Chloramines=range2)&(Sulfate=range2)&(Trihalomethanes=range2)=>(Potability={0[16]})
4	16	(ph=range2)&(Hardness=range2)&(Chloramines=range2)&(Sulfate=range2)&(Turbidity=range1)=>(Potability={0[16]})
5	16	(ph=range2)&(Hardness=range2)&(Solids=range1)&(Sulfate=range2)&(Turbidity=range1)=>(Potability={0[16]})
6	16	(ph=range2)&(Hardness=range2)&(Sulfate=range2)&(Organic_carbon=range1)&(Turbidity=range1)=>(Potability={0[16]})
7	16	(Hardness=range2)&(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Conductivity=range1)=>(Potability={0[16]})
8	16	(ph=range2)&(Hardness=range2)&(Solids=range1)&(Sulfate=range2)&(Conductivity=range1)=>(Potability={0[16]})
9	16	(ph=range2)&(Hardness=range2)&(Sulfate=range2)&(Conductivity=range1)&(Turbidity=range1)=>(Potability={0[16]})
10	16	(ph=range2)&(Solids=range1)&(Sulfate=range2)&(Conductivity=range1)&(Organic_carbon=range2)=>(Potability={0[16]})
11	16	(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Conductivity=range1)&(Trihalomethanes=range2)=>(Potability={0[16]})
12	16	(Hardness=range2)&(Solids=range1)&(Sulfate=range2)&(Conductivity=range1)&(Turbidity=range1)=>(Potability={0[16]})
13	16	(Solids=range1)&(Sulfate=range2)&(Conductivity=range1)&(Organic_carbon=range2)&(Turbidity=range1)=>(Potability={0[16]})
14	16	(Hardness=range2)&(Chloramines=range2)&(Sulfate=range2)&(Trihalomethanes=range2)&(Turbidity=range2)=>(Potability={0[16]})
15	16	(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Conductivity=range1)&(Turbidity=range2)=>(Potability={0[16]})
16	16	(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Trihalomethanes=range2)&(Turbidity=range2)=>(Potability={0[16]})
17	16	(ph=range1)&(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Conductivity=range1)=>(Potability={0[16]})
18	16	(Hardness=range2)&(Solids=range1)&(Sulfate=range2)&(Conductivity=range1)&(Organic_carbon=range1)=>(Potability={0[16]})
19	16	(Hardness=range2)&(Sulfate=range2)&(Conductivity=range1)&(Organic_carbon=range1)&(Turbidity=range1)=>(Potability={0[16]})
20	16	(ph=range1)&(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Trihalomethanes=range2)=>(Potability={0[16]})

Algorytm 2: Genetic Algorithm (Genetyczny)

Typ podejścia: Algorytm ewolucyjny

Parametry:

Calculate reducts or rules

Reduct/Rule choice

☐ Reducts

☒ Rules

Method

☐ Exhaustive algorithm

☒ Genetic algorithm

☐ Covering algorithm

☐ LEM2 algorithm

Generate to set:

Discernibility matrix settings

☐ Full discernibility

☒ Modulo decision

☒ Object related discernibility

☐ Don't discern with missing values

Genetic algorithm settings

☒ High speed

Number of reducts:

10

☐ Normal Speed

☐ Low speed

Cover parameter:

0.9

OK

Cancel

Liczba wygenerowanych reguł: 503

Rule set: Rules2_GeneticAlgorithm		
(1-503)	Match	Decision rules
1	32	(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Organic_carbon=range2)=>(Potability={0[32]})
2	16	(ph=range2)&(Hardness=range2)&(Solids=range1)&(Sulfate=range2)&(Turbidity=range1)=>(Potability={0[16]})
3	16	(ph=range2)&(Hardness=range2)&(Chloramines=range2)&(Sulfate=range2)&(Turbidity=range1)=>(Potability={0[16]})
4	16	(ph=range2)&(Hardness=range2)&(Sulfate=range2)&(Organic_carbon=range1)&(Turbidity=range1)=>(Potability={0[16]})
5	16	(ph=range2)&(Hardness=range2)&(Chloramines=range2)&(Sulfate=range2)&(Trihalomethanes=range2)=>(Potability={0[16]})
6	16	(Hardness=range2)&(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Trihalomethanes=range2)=>(Potability={0[16]})
7	16	(Hardness=range2)&(Solids=range1)&(Sulfate=range2)&(Conductivity=range1)&(Turbidity=range1)=>(Potability={0[16]})
8	16	(Hardness=range2)&(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Conductivity=range1)=>(Potability={0[16]})
9	16	(ph=range2)&(Solids=range1)&(Sulfate=range2)&(Conductivity=range1)&(Organic_carbon=range2)=>(Potability={0[16]})
10	16	(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Conductivity=range1)&(Trihalomethanes=range2)=>(Potability={0[16]})
11	16	(ph=range2)&(Hardness=range2)&(Sulfate=range2)&(Conductivity=range1)&(Turbidity=range1)=>(Potability={0[16]})
12	16	(Solids=range1)&(Sulfate=range2)&(Conductivity=range1)&(Organic_carbon=range2)&(Turbidity=range1)=>(Potability={0[16]})
13	16	(Hardness=range2)&(Chloramines=range2)&(Sulfate=range2)&(Trihalomethanes=range2)&(Turbidity=range2)=>(Potability={0[16]})
14	16	(ph=range2)&(Hardness=range2)&(Solids=range1)&(Sulfate=range2)&(Conductivity=range1)=>(Potability={0[16]})
15	16	(Hardness=range2)&(Solids=range1)&(Sulfate=range2)&(Conductivity=range1)&(Organic_carbon=range1)=>(Potability={0[16]})
16	16	(ph=range1)&(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Conductivity=range1)=>(Potability={0[16]})
17	16	(Hardness=range2)&(Sulfate=range2)&(Conductivity=range1)&(Organic_carbon=range1)&(Turbidity=range1)=>(Potability={0[16]})
18	16	(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Trihalomethanes=range2)&(Turbidity=range2)=>(Potability={0[16]})
19	16	(ph=range1)&(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Trihalomethanes=range2)=>(Potability={0[16]})
20	16	(Solids=range1)&(Chloramines=range2)&(Sulfate=range2)&(Conductivity=range1)&(Turbidity=range2)=>(Potability={0[16]})

Zadanie 3. - Podsumowanie parametrów danych i reguł wiedzy

Opis: Przedstaw podsumowanie w formie tabeli opisującej parametry zestawu danych oraz wynikowy zestaw reguł z punktu widzenia reprezentacji wiedzy

Parametry Zestawu Danych

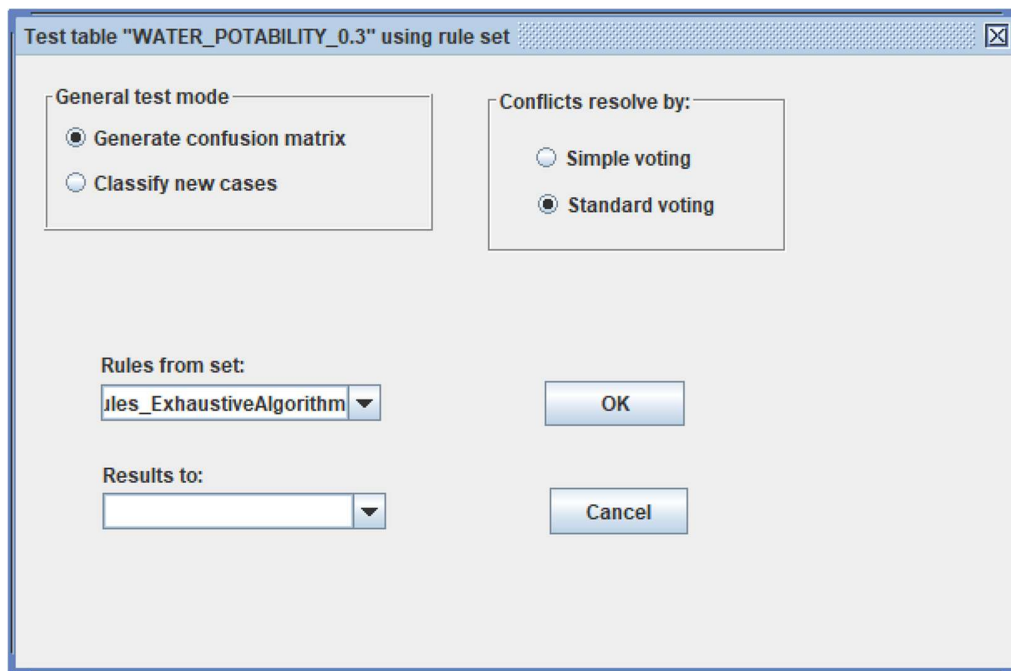
Nazwa	WATER_POTABILITY
Liczba obiektów	419
Liczba obiektów treningowych	293
Liczba obiektów testowych	126
Liczba atrybutów warunkowych	9
Atrybut decyzyjny	Potability
Liczba klas decyzyjnych	2 (0,1)
Format danych	RSES (zdyskretyzowany)
Typ dyskretyzacji	Zakresy(range1,range2)

Parametry zestawu reguł z punktu widzenia reprezentacji wiedzy

Algorytm	Typ podejścia	Liczba reguł	Wsparcie Reguły
Exhaustive	Dokładny	512	Zmienne (1-32)
Genetic HighSpeed Reducts:10	Ewolucyjny	503	Zmienne (1-32)

Zadanie 4. – Klasyfikacja

Opis: Wykonaj klasyfikację przy użyciu modelu w formie reguł decyzyjnych, określając parametry klasyfikacji



Dokładne parametry klasyfikacji:

- Zbiór testowy: WATER_POTABILITY_0.3 (126 przykładów)
- Strategie: **Standard voting**
- Tryb: **Generate confusion matrix**

Zadanie 5. - Analiza wyników

Opis: Przedstaw podsumowanie w formie tabeli opisującej parametry zbioru danych oraz uzyskane wyniki klasyfikacji, w przypadku walidacji krzyżowej podaj odchylenie standardowe.

Wyniki Klasyfikacji - Exhaustive Algorithm

Metryka	Wartość	Opis
Liczba testowanych obiektów	126	Liczba próbek w zbiorze testowym

Dokładność (Accuracy) całkowita	0.643	64.3% - procent poprawnie sklasyfikowanych przykładów
Pokrywanie (Coverage)	1.0	100% - wszystkie przykłady zostały sklasyfikowane
Dokładność klasy 0 (non-potable)	0.77	77% dokładność dla wody niepitnej
Dokładność klasy 1 (potable)	0.462	46.2% dokładność dla wody pitnej
True Positive Rate (klasa 0)	0.67	67% czułości dla klasy negatywnej
True Positive Rate (klasa 1)	0.59	59% czułości dla klasy pozytywnej

Results of experiments by train&test method: Results_ExhaustiveAlgo...						
Actual	Predicted					
		0	1	No. of obj.	Accuracy	Coverage
	0	57	17	74	0.77	1
	1	28	24	52	0.462	1
True positive rate		0.67	0.59			
Total number of tested objects: 126						
Total accuracy: 0.643						
Total coverage: 1						

Wyniki Klasyfikacji - Genetic Algorithm

Metryka	Wartość	Opis
Liczba testowanych obiektów	126	Liczba próbek w zbiorze testowym
Dokładność (Accuracy) całkowita	0.675	67.5% - procent poprawnie sklasyfikowanych przykładów
Pokrywanie (Coverage)	1.0	100% - wszystkie przykłady zostały sklasyfikowane
Dokładność klasy 0 (non-potable)	0.797	79.7% dokładność dla wody niepitnej
Dokładność klasy 1 (potable)	0.5	50.0% dokładność dla wody pitnej

True Positive Rate (klasa 0)	0.69	69% czułości dla klasy negatywnej
True Positive Rate (klasa 1)	0.63	63% czułości dla klasy pozytywnej

Results of experiments by train&test method: Results_GeneticAlgorith...						
Actual	Predicted					
		0	1	No. of obj.	Accuracy	Coverage
	0	59	15	74	0.797	1
	1	26	26	52	0.5	1
True positive rate		0.69	0.63			
Total number of tested objects: 126						
Total accuracy: 0.675						
Total coverage: 1						

Zadanie 6. - Porównanie wyników

Opis: Porównaj wyniki klasyfikacji

Aspekt	Exhaustive	Genetic	Wygrana
Dokładność (Accuracy) całkowita	0.643	0.675	Genetic +3.2%
Pokrywanie (Coverage)	1.0	1.0	Równo
Dokładność klasy 0 (non-potable)	0.77	0.797	Genetic + 2.7%
Dokładność klasy 1 (potable)	0.462	0.5	Genetic +3.8%
Liczba reguł	512	579	Genetic(więcej)
Czas obliczeniowy	Wyższy	Niższy	Genetic

Zadanie 7.

Opis: Który z algorytmów indukcji reguł jest lepszy z punktu widzenia reprezentacji wiedzy, a który z punktu widzenia klasyfikacji?

Z punktu widzenia reprezentacji wiedzy

- **Exhaustive (dokładny)** daje pełniejszą i deterministyczną reprezentację przestrzeni reguł. Ponieważ przeszukuje przestrzeń wyczerpująco, może wygenerować reguły opisujące rzadkie, ale istotne fragmenty danych.

Z punktu widzenia klasyfikacji

- **Genetic (ewolucyjny)** w tym eksperymencie osiągnął wyższą dokładność ogólną i lepsze wyniki dla obu klas. Algorytmy ewolucyjne potrafią wygenerować reguły silnie zoptymalizowane pod zadane kryteria, co zwykle przekłada się na lepsze wyniki predykcyjne.