

Klasyfikacja danych za pomocą algorytmu kolonii mrówek i sieci neuronowej ART

MAKSYMILIAN PASEK

ŁUKASZ MILEWICZ

Cele i założenia pracy

- ▶ Zbadanie algorytmów uczenia nienadzorowanego ACC i ART.
- ▶ Zaimplementowanie systemu który łączy oba algorytmy w celu uzyskania zoptymalizowanego wyniku
- ▶ Klasyfikacja danych losowych i rzeczywistych z dostępnych bibliotek

Algorytm klasteryzacji kolonii mrówek (ACC)

- ▶ Probabilistyczna technika klasyfikacji danych.
- ▶ Zainspirowany zachowaniem mrówek szukających pożywienia dla swojej kolonii.

Walidacyjny indeks klasyfikacji

- ▶ Określa połączenie w obrębie danej klasyfikacji, a także rozdzielania pomiędzy innymi klastrami
- ▶ Wykorzystywany do określenia liczby klastrów

Algorytm ACC-VI

- ▶ Krok 0: Określ liczbę mrówek, maksymalną liczbę iteracji: M_n , wielkość przeszukiwanej przestrzeni, maksymalną prędkość mrówek oraz parametry: α i β
- ▶ Krok 1: Umieść dane na przestrzeni -> określ współrzędne (x, y) dla każdego obiektu losowo. Każda z mrówek jest niezaladowana i wybiera obiekt losowo.

▶ Krok 2:

Dla każdego $i = 1, 2, \dots, M_n$:

Dla każdego $j = 1, 2, \dots$, liczba mrówek:

2.1:

► Krok 2:

Dla każdego $i = 1, 2, \dots, M_n$:

Dla każdego $j = 1, 2, \dots$, liczba mrówek:

2.1: Policz podobieństwo w obrębie regionu lokalnego wykorzystując 3 różne typy prędkości mrówek dla różnych kolonii: stała, losowa, losowo mająca

2.2: Jeśli mrówka nie przenosi ładunku, oblicz prawdopodobieństwo podniesienia. Jeśli prawdopodobieństwo jest większe niż wartość losowa i obiekt nie jest podnoszony w tym samym momencie, to mrówka podnosi ładunek oraz obiekt jest opisany jako podniesiony. Mrówka porusza się do nowego miejsca.

2.3: Jeśli mrówka jest załadowana, to obliczamy prawdopodobieństwo opuszczenia ładunku. Jeśli prawdopodobieństwo jest większe niż losowa wartość, to mrówka opuszcza ładunek i obiekt jest opisywany jako niezaładowany.

- Krok 3: Dla każdego $i = 1, 2, \dots, N$ // dla wszystkich danych/obiektów:

3.1: Jeśli obiekt jest wyizolowany lub liczba sąsiadów jest mniejsza niż stała wartość, obiekt jest oznaczany jako odstający.

3.2: W innym wypadku, nadaj obiektowi numer sekwencyjny oraz oznacz sąsiadów w obrębie lokalnego obszaru. Wyznacz liczbę klastrów c .

- Krok 4: Dla każdego $i = 1, 2, \dots, c$ // c – liczba klastrów:

4.1: Oblicz średnią dla klastra i i znajdź 4 punkty reprezentatywne.

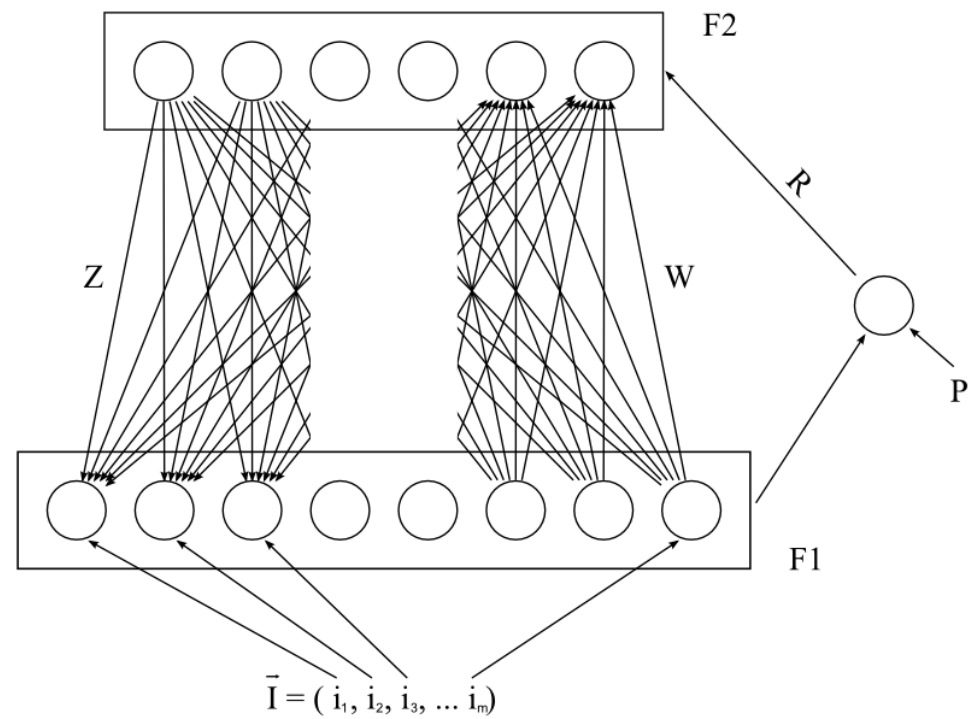
4.2: Przelicz indeks walidacyjny CDbw który pomoże określić optymalną liczbę klastrów

- Krok 5: Dla każdego $i = 1, 2, \dots, c$ // c – liczba klastrów:

5.1: Odrzuć obserwacje odstające i przelicz indeks CDbw. Porównaj do starej wartości dla każdego klastra.

5.2: Przesuń obserwacje odstające do klastra z największą różnicą w współczynniku CDbw.

Sieć neuronowa ART



	π_1	π_2	π_3
x_1	1	2	1
x_2	1	2	1
x_3	1	2	2
x_4	2	3	2
x_5	2	3	3
x_6	3	1	3
x_7	3	1	3
x_8	3	1	3

	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8
$A^{(1)}$	1	1	1	0	0	0	0	0
	0	0	0	1	1	0	0	0
	0	0	0	0	0	1	1	1
$A^{(2)}$	0	0	0	0	0	1	1	1
	1	1	1	0	0	0	0	0
	0	0	0	1	1	0	0	0
$A^{(3)}$	1	1	0	0	0	0	0	0
	0	0	1	1	0	0	0	0
	0	0	0	0	1	1	1	1

Połączony zespół algorytmów klasyfikacji

