

Klasyfikator

Dokumentacja testowa

Monika Kogut
Marta Kornaszewska
Katarzyna Kuzawska

Spis treści

1. Przebieg testów.....	2
2. Wyniki testów	2
2.1. Etap I.....	2
Dane generowane losowo	2
Dane rzeczywiste	4
2.2. Etap II	4
Dane generowane losowo	4
2.3. Etap III	5
Dane generowane losowo	5
3. Wnioski.....	7
Testy dla danych syntetycznych	7
Testy dla danych rzeczywistych	8

1. Przebieg testów

Dla każdego z etapów testy były przeprowadzane na zbiorze generowanym na początku programu . (Spośród elementów generowanych rozkładem jednorodnym, a następnie zaburzanych rozkładem normalnym – wylosowano część elementów - np. 20 %, które trafiały do zbioru testowego, pozostałe należały do zbioru uczącego.)

Na wejście skonstruowanego automatu dawano element ze zbioru testowego (jako, że generowaliśmy ten zbiór na początku wykonania programu, to wiemy jaki z rozpoznawanych symboli reprezentuje). Jeśli wynik nie jest zgodny z rzeczywiście reprezentowanym symbolem, to dodajemy wartość 1 do sumy błędów, jeśli wynik jest poprawny – dodajemy 0.

Po wykonaniu obliczeń dla każdego z wektorów zbioru testowego, obliczamy stosunek liczby błędów do rozmiaru zbioru testowego.

Kolejne testy były przeprowadzane dla danych rzeczywistych czytywanych z pliku. W przypadku tych danych - dla kolejnych testów zmieniane były parametry jedynie związane z samym algorytmem PSO (np. rozmiar roju) i związana z nim bezpośrednio funkcją błędu. Parametry danych treningowych pozostały bezzmienne, jako , że były narzucone z góry w podanym pliku.

2. Wyniki testów

2.1. Etap I

Dane generowane losowo

2.1.1. Automat deterministyczny, bez elementów obcych

Parametry wywołania PSO

test	Liczba iteracji	Rozmiar roju	Liczba cech	Liczba kopii klas	Liczba klas	Dyskretyzacja	Zaburzenie	Przedział wartości cech	Liczba symboli testowych
1	50	20	10	100	10	10	3	[0;20]	10
2	100	40	10	100	10	10	3	[0;20]	10
3	50	40	10	100	10	10	3	[0;20]	10
4	100	30	10	100	10	30	3	[0;10]	10
5	100	30	10	100	10	5	3	[0;20]	10

6	100	5	20	100	10	5	5	[0;20]	10
7	100	5	100	100	10	5	5	[0;20]	10
8	200	5	100	100	10	5	5	[0;20]	200
9	500	5	100	200	10	5	5	[0;20]	400

Wyniki

test	Błąd dla zbioru testowego (%)	Błąd całkowity (%)	Czas obliczeń (s)
1	70	67,1	204,02
2	80	60,9	755,14
3	64,8	64,8	368,76
4	70	62,9	516,12
5	50	57,4	600,91
6	60	66,2	160,5
7	90	71,7	726,14
8	61,5	62,2	1378,6
9	61,5	61,9	8349,61

2.1.2. Automat deterministyczny z elementami obcymi

Parametry wywołania PSO

test	Liczba iteracji	Rozmiar roju	Liczba cech	Liczba kopi klas	Liczba klas	Zaburzenie	Dyskretyzacja	Przedział wartości cech	Liczba symboli testowych
1	200	10	15	20	6	3	5	[0;20]	20,00%
2	200	10	100	20	10	3	5	[0;20]	200
3	400	10	100	20	10	3	5	[0;20]	20,00%
4	400	10	100	40	10	3	5	[0;20]	20,00%
5	400	10	100	40	20	3	5	[0;20]	20,00%
6	500	10	200	20	10	5	5	[0;20]	20,00%
7	500	10	200	50	10	5	5	[0;20]	20,00%
8	500	5	300	20	10	5	5	[0;20]	20,00%
9	500	5	300	100	10	5	5	[0;20]	20,00%

Wyniki

test	Błąd dla zbioru testowego (%)	Czas obliczeń (s)
1	25	169,58
2	53	207,83
3	49	400,34
4	27	567,9
5	35,3	678,6
6	22,5	572,01
7	55	789
8	34	862,01
9	16,5	1281,8

Dane rzeczywiste

Parametry wywołania PSO

test	Liczba iteracji	Rozmiar roju	Liczba cech	Liczba kopii klas	Liczba klas	Dyskretyzacja (liczba podprzedziałów)	Przedział wartości cech	Liczba symboli testowych
1	50	5	10	100	10	5	[0;20]	10%
2	100	5	10	100	10	10	[0;20]	10%
3	100	10	10	100	10	10	[0;20]	10%

Wyniki

test	Błąd dla zbioru testowego (%)	Czas obliczeń (s)
1	95	10235,1
2	93	14560,90
3	79	20113,6

2.2. Etap II

Dane generowane losowo

2.2.1. Automat niedeterministyczny, bez elementów obcych

Parametry wywołania PSO

test	Liczba iteracji	Rozmiar roju	Liczba cech	Liczba kopii klas	Liczba klas	Dyskretyzacja (liczba podprzedziałów)	Ograniczenie niedeterminizmu	Zaburzenie (odchylenie standardowe)	Przedział wartości cech	Liczba symboli testowych
1	100	10	10	100	10	5	3	3	[0;20]	100
2	50	10	50	100	10	5	3	3	[0;20]	500
3	100	30	50	200	10	5	3	3	[0;20]	500
4	100	20	100	200	10	5	3	3	[0;20]	500
5	100	20	100	100	10	5	3	3	[0;20]	500
6	150	10	100	100	10	5	3	5	[0;20]	500
7	150	5	100	200	15	5	3	5	[0;20]	500
8	200	5	100	200	15	5	3	5	[0;20]	500
9	200	5	100	200	15	5	3	5	[0;20]	500

Wyniki

test	Błąd dla zbioru testowego (%)	Czas obliczeń (s)
1	55	1296
2	46	1227,75
3	49	1425,32
4	31,5	1902,9
5	37	1530,6
6	44,3	1853,00
7	40,2	1980,058
8	28,0	1862,2
9	12	1981,5

2.2.2. Automat niedeterministyczny z elementami obcymi

Parametry wywołania PSO

test	Liczba iteracji	Rozmiar Roju	Liczba cech	Liczba kopii klas	Liczba klas	Dyskretyzacja (liczba podprzedziałów)	Zaburzenie (odchylenie standardowe)	Ograniczenie niedeterminizmu	Przedział wartości cech	Liczba symboli testowych
1	50	30	50	100	10	5	3	3	[0;20]	500
2	50	10	50	100	10	5	3	3	[0;20]	500
3	100	30	50	200	10	5	3	3	[0;20]	500
4	100	20	100	200	10	5	3	3	[0;20]	500
5	100	20	100	100	10	5	3	3	[0;20]	500
6	150	10	100	100	10	5	5	3	[0;20]	500
7	150	5	100	200	15	5	5	3	[0;20]	500
8	200	5	100	200	15	5	5	3	[0;20]	500
9	200	5	100	200	15	5	5	3	[0;20]	500

Wyniki

test	Błąd dla zbioru testowego (%)	Czas obliczeń (s)
1	52	200,46
2	45	231,90
3	68	467,32
4	65	1000,64
5	43,2	4525,3
6	35	1453,3
7	43	1450,028
8	56	1522,2
9	27,8	2384,28

2.3. Etap III

Dane generowane losowo

2.3.1. Automat rozmyty bez elementów obcych

Parametry wywołania PSO

test	Liczba iteracji	Rozmiar roju	Liczba cech	Liczba kopii klas	Liczba klas	Dyskretyzacja (liczba podprzedziałów)	Zaburzenie (odchylenie standardowe)	Przedział wartości cech	Liczba symboli testowych
1	50	10	30	100	10	5	3	[0;20]	500
2	50	20	30	100	10	5	3	[0;20]	500
3	100	20	50	200	10	5	3	[0;20]	500
4	100	10	50	200	10	5	3	[0;20]	500
5	100	10	100	100	10	5	3	[0;20]	500
6	150	10	100	100	10	5	5	[0;20]	500

7	150	5	150	200	10	5	5	[0;20]	500
8	200	5	150	200	10	5	5	[0;20]	500
9	200	5	200	200	10	5	5	[0;20]	500

Wyniki

test	Błąd dla zbioru testowego (%)	Czas obliczeń (s)
1	55,8	1009,0
2	57,3	5043,2
3	63,0	8023,73
4	78,0	9062,2
5	60,0	12630,54
6	38,3	10530,90
7	45,5	15804,42
8	32,1	17765,23
9	12,0	17567,82

2.3.2. Automat rozmyty z elementami obcymi

Parametry wywołania PSO

test	Liczba iteracji	Rozmiar roju	Liczba cech	Liczba kopii klas	Liczba klas	Dyskretyzacja	Zaburzenie	Przedział wartości cech	Liczba symboli testowych
1	50	20	30	100	10	5	3	[0;20]	500
2	50	20	30	100	10	5	3	[0;20]	500
3	100	20	50	200	10	5	3	[0;20]	500
4	100	10	50	200	10	5	3	[0;20]	500
5	100	10	100	100	10	5	3	[0;20]	500
6	150	10	100	100	10	5	5	[0;20]	500
7	150	5	150	200	10	5	5	[0;20]	500
8	200	5	150	200	10	5	5	[0;20]	500
9	200	5	200	200	10	5	5	[0;20]	500

Wyniki

test	Błąd dla zbioru testowego (%)	Czas obliczeń (s)
1	42,2	3452,09
2	34	4000,45
3	53,4	4023,0
4	47,5	7457,2
5	58,2	10923,41
6	12,5	12819,90
7	30,4	13460,2
8	10	16341,9
9	13,5	18501

3. Wnioski

Testy dla danych syntetycznych

Zachowanie błędu w zależności od poszczególnych parametrów dla odpowiednich etapów implementacji

1. etap

- a) ilość instancji treningowych – przy zwiększeniu błąd maleje
- b) ilość klas – przy zwiększeniu błąd maleje
- c) dyskretyzacja – przy zwiększeniu błąd maleje
- d) zaszumienie losowanych danych - przy zwiększeniu błąd rośnie
- e) rozmiar roju - przy zwiększeniu błąd maleje (ale obliczenia znacząco spowalniają)

2. etap

- a) ilość instancji treningowych – przy zwiększeniu błąd maleje
- b) ilość klas – przy zwiększeniu błąd maleje
- c) dyskretyzacja – przy zwiększeniu błąd maleje
- d) zaszumienie losowanych danych - przy zwiększeniu błąd rośnie
- e) rozmiar roju - przy zwiększeniu błąd rośnie (ale obliczenia znacząco spowalniają)

3. etap

- a) ilość instancji treningowych – przy zwiększeniu błąd maleje
- b) ilość klas – przy zwiększeniu błąd maleje
- c) dyskretyzacja – przy zwiększeniu błąd maleje
- d) zaszumienie losowanych danych - przy zwiększeniu błąd rośnie
- e) rozmiar roju - przy zwiększeniu błąd rośnie (ale obliczenia znacząco spowalniają)

Obserwacja powyższych parametrów nasuwa następujące wnioski:

- Czynniki które wpływają na zwiększenie dokładności konstruowanego automatu są :
zwiększenie liczby iteracji algorytmu PSO, zwiększenie wielkości roju, zwiększenie liczby podziałów (dyskretyzacji zbioru wartości cech). Jednocześnie zwiększenie wartości tych parametrów spowalnia znacząco obliczenia, ze względu na konieczność wykonania większej liczby obliczeń.
- Im większa jest ilość instancji treningowych , tym lepiej możemy „wyuczyć” automat.
- Zwiększenie dyskretyzacji w praktyce oznacza zwiększenie ilości podprzedziałów do których będziemy klasyfikować dane wartości cech, a tym samym tworzymy automat o większej dokładności.
- Zaszumienie losowych danych wpływa negatywnie na wyniki symulacji. Związane jest to ze zwiększeniem tolerancji automatu na odchylenia od „wzorcowych” wartości cech, a tym samym zmniejszenie precyzji wyznaczania stanu przez automat.
- Algorytm PSO opiera się na autonomicznych współpracujących ze sobą agentach, które nazywamy cząstką, która porusza się w przestrzeni poszukiwań w celu znalezienia optymalnej sensie rozważanej funkcji celu pozycji. Dlatego też zwiększenie liczności zbioru cząsteczek (wielkość roju) prowadzi do uzyskania dokładniejszych wyników.

Testy dla danych rzeczywistych

Dla danych czytanych z pliku testy udało się ukończyć jedynie w przypadku automatu deterministycznego.

Dla pozostałych – obliczenia trwały zbyt długo, by można było przeprowadzić więcej niż jeden test i na tej podstawie oceniać zależność działania klasyfikatora od parametrów wywołania algorytmu.

- Znaczne spowolnienie obliczeń dla kolejnych etapów wynikało z zastosowania różnych funkcji błędu do obliczeń (coraz większy stopień skomplikowania obliczeń i zwiększenie rozmiaru funkcji przejścia – w przypadku 3 etapu)
- Na podstawie testów dla automatu deterministycznego można stwierdzić, że:
dla niezmiennych danych, wywołanie z większą wartością dyskretyzacji, liczby iteracji PSO i rojów wpływa pozytywnie na dokładność symulacji. Potwierdza to wnioski uzyskane na podstawie danych syntetycznych (generowanych losowo).