

Wydział Informatyki Politechniki Białostockiej

Zastosowanie algorytmów ewolucyjnych do wyznaczania przybliżonych reduktów

Dyplomant: Jan Gromko

Promotor: prof. dr hab. Jarosław Stepaniuk

19 kwietnia 2017 r.



Redukcja

- Istota problemu redukcji

- Prosty algorytm wyznaczania reduktu

- Problem złożoności dokładnych algorytmów redukcji

- Alternatywne metody redukcji

Propozycje algorytmów genetycznych do wyznaczania przybliżonych reduktów

- Założenia algorytmu

- Schemat działania

Druga propozycja algorytmu genetycznego

- Założenia

- Porównanie wyników działania różnych wersji algorytmu

**Czy można zredukować zbiór pod względem atrybutów
w ten sposób, by zachowana była rozróżnialność elementów
z oryginalnego zbioru?**

Zbiór niezależny

Zbiór atrybutów $B_1 \subset A$ jest *niezależny* w danym systemie informacyjnym, jeśli dla każdego $B_2 \subset B_1$ zachodzi $IND(B_1) \neq IND(B_2)$.

Redukt

Reduktem zbioru atrybutów $B_1 \subseteq A$ nazywamy każdy niezależny zbiór $B_2 \subseteq B_1$, dla którego $IND(B_1) = IND(B_2)$, przy czym B_2 powinien być jak najmniej liczny. Może istnieć wiele reduktów.

<i>Pacjent</i>	<i>Ból głowy</i>	<i>Ból mięśni</i>	<i>Temperatura</i>	<i>Grypa</i>
1	nie	tak	podwyższona	tak
2	tak	nie	podwyższona	tak
3	tak	tak	wysoka	tak
4	nie	tak	normalna	nie
5	tak	nie	podwyższona	nie
6	nie	nie	wysoka	tak

Tabela 1. Tablica decyzyjna przykładowego zbioru.

	1	2	3	4	5	6
1	∅	—	—	—	—	—
2	∅	∅	—	—	—	—
3	∅	∅	∅	—	—	—
4	t	g, m, t	g, t	∅	—	—
5	g, m	∅	m, t	∅	∅	—
6	∅	∅	∅	m, t	g, t	∅

Tabela 2. Macierz rozróżnialności.

g – ból głowy; m – ból mięśni; t – temperatura

	1	2	3	4	5	6
1	\emptyset	—	—	—	—	—
2	\emptyset	\emptyset	—	—	—	—
3	\emptyset	\emptyset	\emptyset	—	—	—
4	t	g, t	g, t	\emptyset	—	—
5	g	\emptyset	t	\emptyset	\emptyset	—
6	\emptyset	\emptyset	\emptyset	t	g, t	\emptyset

Tabela 3. Macierz rozróżnialności po redukcji.



1. Zliczenie wystąpień atrybutów w macierzy rozróżnialności.
2. Wybór atrybutu występującego najliczniej w macierzy rozróżnialności; dodanie wybranego atrybutu do wynikowego zbioru atrybutów *Red*.
3. Wykreślenie komórek zawierających wybrany atrybut.
4. Jeśli wszystkie komórki zostały wykreślone, wynikiem jest uzyskany zbiór *Red*, w przeciwnym razie powrót do kroku 1.

	1	2	3	4	5	6
4	t	g, m, t	g, t	∅	–	–
5	g, m	∅	m, t	∅	∅	–
6	∅	∅	∅	m, t	g, t	∅

Tabela 4. Fragment macierzy rozróżnialności zawierający istotne dane.

g – 4

m – 4

t – 6

	1	2	3	4	5	6
4	t	g , m , t	g , t	∅	—	—
5	g, m	∅	m , t	∅	∅	—
6	∅	∅	∅	m , t	g , t	∅

Tabela 5. Fragment macierzy rozróżnialności zawierający istotne dane.

$$Red = \{t\}$$

$$g - 1$$

$$m - 1$$

	1	2	3	4	5	6
4	t	g , m, t	g , t	∅	–	–
5	g , m	∅	m , t	∅	∅	–
6	∅	∅	∅	m , t	g , t	∅

Tabela 6. Fragment macierzy rozróżnialności zawierający istotne dane.

$$Red = \{t, g\} \vee Red = \{t, m\}$$

	1	2	3	4	5	6
1	\emptyset	—	—	—	—	—
2	\emptyset	\emptyset	—	—	—	—
3	\emptyset	\emptyset	\emptyset	—	—	—
4	t	g, m, t	g, t	\emptyset	—	—
5	g, m	\emptyset	m, t	\emptyset	\emptyset	—
6	\emptyset	\emptyset	\emptyset	m, t	g, t	\emptyset

Tabela 7. Macierz rozróżnialności oryginalnego zbioru.

$$Red = \{t, g\} \vee Red = \{t, m\}$$

Atrybut nieusuwalny

Atrybut $p \in P_1$ jest *nieusuwalny* z P_1 , jeśli dla $P_2 = P_1 \setminus \{p\}$ zachodzi $\widetilde{P}_2 \neq \widetilde{P}_1$. W przeciwnym przypadku atrybut p jest *zbędny*.

Rdzeń

Rdzeniem P nazywa się zbiór wszystkich atrybutów nieusuwalnych ze zbioru P , co zapisywane jest w następujący sposób:

$$CORE(P) = \{p \in P : \widetilde{P}' \neq \widetilde{P}, P' = P \setminus \{p\}\}.$$



Wyznaczanie reduktu w zbiorze przybliżonym jest problemem NP-zupełnym – nie jest możliwe znalezienie rozwiązania w czasie wielomianowym.

Rozwiązania sprzętowe:

- ▶ specjalizowane układy programowalne (FPGA, CPLD).

Rozwiązania przybliżone – wykorzystanie innych metod sztucznej inteligencji:

- ▶ algorytmy ewolucyjne,
- ▶ algorytmy mrówkowe,
- ▶ inteligencja roju,
- ▶ metody połączone.



Lian Chen, Hongling Liu, Zilong Wan
Computer Center, Nanchang University

An Attribute Reduction Algorithm Based on Rough Set Theory
and an Improved Genetic Algorithm (2014)

Wejście

System informacyjny $S = (U, Q, V, f)$, gdzie $Q = A \cup D$.

Wyjście

Wynik optymalnej redukcji zbioru.

Chromosomem będzie jednowymiarowa tablica binarna o stałej długości. Długość chromosomu odpowiada liczbie atrybutów warunkowych.

Każdy z genów odpowiada dokładnie jednemu atrybutowi warunkowemu, przy czym wartość 1 będzie oznaczająca, iż atrybut jest wybrany, 0 w przeciwnym wypadku.

Wartość genów odpowiadających atrybutom należącym do rdzenia ustawiana jest na 1, wartość pozostałych genów ustawiana jest losowo na 0 lub 1.

Jakość przystosowania pojedynczego osobnika, zgodnie z definicją redukcji, opiera się na dwóch aspektach – liczbie atrybutów, które zawiera (powinna ona być możliwie jak najmniejsza) oraz zachowanej rozróżnialności obiektów (powinna być jak największa).

Zgodnie z tymi wymaganiami, funkcja przystosowania ma postać:

$$f(x) = \frac{1}{rozmiar(x)} + \sigma(x),$$

gdzie $rozmiar(x)$ oznacza liczbę atrybutów, które zawiera chromosom, natomiast $\sigma(x)$ jest znormalizowanym współczynnikiem istotności zbioru atrybutów chromosomu.

Prawdopodobieństwo wybrania danego osobnika i wynosi $p_{si} = \frac{f_i}{\sum_i f_i}$,

gdzie f_i jest wartością funkcji przystosowania dla pojedynczego osobnika i , natomiast n jest rozmiarem populacji.

Jeśli wartość funkcji przystosowania najslabiej przystosowanego osobnika w bieżącym pokoleniu jest niższa, niż wartość funkcji przystosowania najlepiej przystosowanego osobnika z poprzedniego pokolenia, wówczas najslabszy osobnik z bieżącego pokolenia jest zastępowany najlepszym osobnikiem z poprzedniego pokolenia.

Algorytm zakłada krzyżowanie jednopunktowe – dla każdej pary osobników (w tym wypadku – chromosomów), losowo wybierany jeden punkt, a następnie części chromosomów zamieniane są między osobnikami według tego punktu, co tworzy osobniki kolejnego pokolenia.

Poszczególne geny w chromosomach zmieniane są losowo z pewnym ustalonym prawdopodobieństwem.

Przy przeprowadzaniu mutacji chronione przed mutacją są geny związane z atrybutami należącymi do rdzenia.

1. Wygenerowanie populacji początkowej.
2. Obliczenie znormalizowanego współczynnika istotności dla każdego chromosomu.
3. Selekcja osobników na podstawie algorytmu koła ruletki.
4. Krzyżowanie.
5. Mutacje.
6. Obliczenie wartości funkcji przystosowania dla każdego chromosomu.
7. Sprawdzenie warunku zatrzymania algorytmu – jeśli warunek jest spełniony, algorytm jest zatrzymywany.
W przeciwnym razie powrót do punktu 3.

1. Jeśli zbiór atrybutów najlepszego osobnika zawiera atrybuty, których współczynnik istotności nie został obliczony, przejście do punktu 2.; w przeciwnym razie przejście do punktu 3.
2. Obliczany jest współczynnik istotności każdego atrybutu, dla którego nie został on wcześniej wyliczony. Jeśli $\sigma(a) = 0$, wartość genu zmieniana jest z 1 na 0.
3. Wyznaczonym reduktem jest fenotyp najlepszego znalezionej osobnika, po ewentualnych modyfikacjach z punktu 2.



Jakub Wróblewski

Adaptacyjne metody klasyfikacji obiektów (2001)

- ▶ Sprowadzenie problemu wyznaczenia reduktu do problemu pokrycia macierzy, w której kolumny odpowiadają atrybutom, a wiersze parom obiektów, które wymagają rozróżnienia.
- ▶ Należy znaleźć pokrycie kolumnowe takiej macierzy.

Krzyżowane są permutacje – w przypadku klasycznego operatora krzyżowania wynik w większości przypadków nie byłby permutacją.

Kroki działania operatora MOX:

- ▶ Losowanie jednakowej w obu osobnikach sekcji dopasowania; początek sekcji jest ustalony na początku chromosomu.
- ▶ Sekcje dopasowania obu osobników rodzicielskich pozostawiane są bez zmian, natomiast pozostałe części chromosomów są przekształcane w ten sposób, aby występujące w nich wartości liczbowe były ustawione w takiej kolejności, w jakiej występują u drugiego osobnika rodzicielskiego.

Krzyżowane są permutacje – w przypadku klasycznego operatora krzyżowania wynik w większości przypadków nie byłby permutacją.

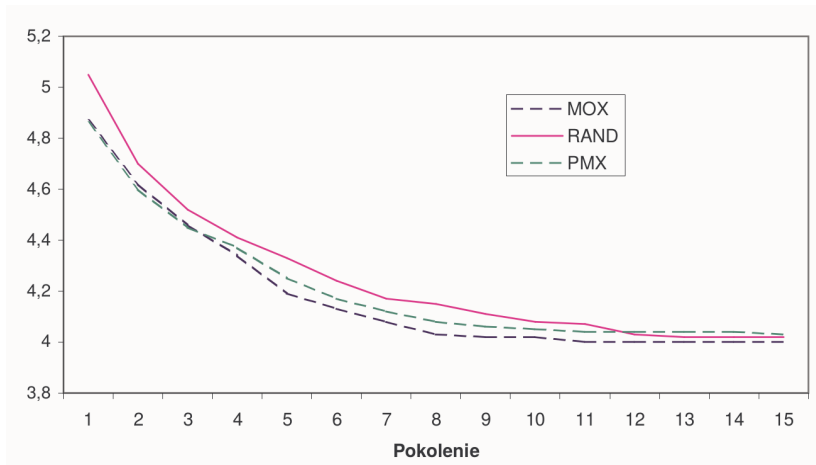
Kroki działania operatora MOX:

1. Losowanie jednakowej w obu osobnikach sekcji dopasowania; początek sekcji jest ustalony na początku chromosomu.
2. Sekcje dopasowania obu osobników rodzicielskich pozostawiane są bez zmian, natomiast pozostałe części chromosomów są przekształcane w ten sposób, aby występujące w nich wartości liczbowe były ustawione w takiej kolejności, w jakiej występują u drugiego osobnika rodzicielskiego.

$$v_{MOX} \left(\begin{array}{c|c} 1, 2, 3 & 4, 5, 6 \\ \hline 4, 2, 1 & 3, 6, 5 \end{array} \right) = \left(\begin{array}{c|c} 1, 2, 3 & 4, 6, 5 \\ \hline 4, 2, 1 & 3, 5, 6 \end{array} \right)$$

- ▶ Algorytm heurystyczno-losowy.
- ▶ Zamiast algorytmu genetycznego używany jest losowy generator permutacji.

Porównanie działania wersji algorytmu



- [1] Zdzisław Pawlak
Zbiory przybliżone – nowa matematyczna metoda analizy danych
- [2] Leszek Rutkowski
Metody i techniki sztucznej inteligencji
- [3] Lian Chen, Hongling Liu, Zilong Wan
An Attribute Reduction Algorithm Based on Rough Set Theory and an Improved Genetic Algorithm
- [4] Jakub Wróblewski
Adaptacyjne metody klasyfikacji obiektów

Pytania

