Metody wyznaczania parametrów początkowych systemów neuronowo - rozmytych

Kinga Kądziołka

Streszczenie: W artykule przedstawione zostały cztery metody wyznaczania parametrów początkowych systemów neuronowo – rozmytych. Na wybranych przykładach dotyczących problemu klasyfikacji obiektów zostały porównane wyniki uzyskane przez system w zależności od sposobu doboru parametrów początkowych. Porównano też skuteczność klasyfikacji w zależności od metody optymalizacji parametrów początkowych.

Słowa kluczowe: systemy neuronowo – rozmyte, grupowanie danych, klasyfikacja obiektów

1. WPROWADZENIE

Systemy neuronowo – rozmyte stanowią połączenie sieci neuronowych i systemów rozmytych. Dzięki temu umożliwiają wykorzystanie zarówno regułowej reprezentacji wiedzy systemów rozmytych jak i metod uczenia stosowanych w przypadku sieci neuronowych [6].

Załóżmy, że mamy zbiór zawierający P obiektów postaci $\left(\overline{x^i}, d^i\right)$, gdzie $\overline{x^i}$ jest pewnym n – wymiarowym wektorem opisującym dany obiekt, zaś d^i jest etykietą klasy (np. numer), do której obiekt $\overline{x^i}$ należy. Zadaniem systemów neuronowo – rozmytych wykorzystanych do problemu klasyfikacji będzie taki dobór parametrów w danego systemu, aby spełniona była zależność [5]:

$$f\left(\overline{x^i}, w\right) \approx d^i \quad i=1, \ldots, P \qquad (1)$$

gdzie f jest pewną funkcją realizowaną przez system. Proces „dostrajania" parametrów początkowych w systemu, tak aby spełniona była zależność (1) przeprowadzany będzie na tzw. zbiorze treningowym. Z kolei skuteczność klasyfikacji wyznaczana będzie na danych testowych, które nie zostały użyte w procesie „dostrajania" parametrów początkowych danego systemu.

Artykuł składa się z trzech części. Pierwsza część stanowi krótkie wprowadzenie na temat systemów neuronowo – rozmytych, w drugiej części zaprezentowane zostaną przykładowe metody wyznaczania parametrów początkowych systemów neuronowo – rozmytych oraz ich późniejszej optymalizacji, zaś ostatnia część stanowi prezentację wyników uzyskanych przez systemy neuronowo – rozmyte zastosowane do rozwiązania wybranych problemów klasyfikacji obiektów i porównanie tych wyników z dostępnymi w literaturze wynikami uzyskanymi przy pomocy innych metod klasyfikacji.

2. SYSTEMY NEURONOWO - ROZMYTE

System neuronowo – rozmyty jest pewną równoważną formą systemu rozmytego, w którym dodatkowo istnieje możliwość „strojenia" parametrów [6]. System taki może zostać przedstawiony w postaci architektury sieci [5,6], która przypomina wielowarstwowy perceptron, przy czym rolę wag pełnią funkcje przynależności do zbiorów rozmytych, zaś zamiast funkcji aktywacji mamy operatory na zbiorach rozmytych. W przeprowadzonych doświadczeniach wykorzystane będą systemy neuronowo – rozmyte, w których rozmyta implikacja będzie reprezentowana przy pomocy operacji iloczynu, rozmywanie będzie typu singelton, defuzyfikacja realizowana metodą środka sum, zaś baza reguł będzie postaci:

$$R^i: \text{Jeśli } ((x_1 \text{ jest } A_1^i) \text{ i } \ldots \text{ i } (x_n \text{ jest } A_n^i)) \text{ To } (y = b^i) \qquad (2)$$

gdzie A_j^i są zbiorami rozmytymi ($i=1, \ldots, m; j=1,\ldots, n$), $b^i \in R$, zaś $x = (x_1,\ldots,x_n) \in R^n$ jest zmienną wejściową, a $y \in R$ stanowi zmienną wyjściową systemu.

3. PARAMETRY POCZĄTKOWE SYSTEMÓW NEURONOWO - ROZMYTYCH

Parametry początkowe systemów wyznaczane będą przy pomocy następujących metod:

1. Losowy wybór parametrów początkowych – parametry funkcji przynależności do zbiorów rozmytych występujących w poprzednikach reguł (1) oraz parametry w następnikach tych reguł są wyznaczane w sposób losowy.
2. Parametry początkowe ustalane jako średnie i odchylenia standardowe wartości atrybutów obiektów należących do odpowiednich klas
3. Parametry początkowe wyznaczane przy pomocy metod grupowania
4. Metoda siatki podziału

3.1. PARAMETRY POCZĄTKOWE SYSTEMU JAKO ŚREDNIE I ODCHYLENIA STANDARDOWE WARTOŚCI ATRYBUTÓW

Przyjmijmy, że funkcje przynależności do zbiorów rozmytych występujących w poprzednikach reguł (2) są postaci:

$$\mu_{A_j^i}(x_j) = \exp\left\{-\left[\frac{(x_j - c_j^i)}{\sigma_j^i}\right]^2\right\} \qquad (3)$$

Wówczas parametry początkowe systemu określamy w sposób następujący [1]:

$$c_j^i = \frac{\sum_{x \in V_i} \overline{x_j}}{|V_i|},\ \sigma_j^i = \sqrt{\frac{1}{|V_i|}\sum_{x \in V_i}\left(\overline{x_j} - c_j^i\right)^2},\ b^i = d^i \qquad (4)$$

gdzie V_i jest zbiorem obiektów należących do i-tej klasy zaś d^i jest etykietą i-tej klasy, np. $d^i = i$

3.2. ZASTOSOWANIE METOD GRUPOWANIA DO WYZNACZANIA PARAMETRÓW POCZĄTKOWYCH SYSTEMÓW NEURONOWO - ROZMYTYCH

Załóżmy, że grupowaniu poddajemy zbiór zawierający P obiektów postaci $\left(\overline{x^i}, d^i\right)$, gdzie $\overline{x_i} = \left(\overline{x_1^i}, ..., \overline{x_n^i}\right)$ jest *n* – wymiarowym wektorem opisującym i – ty obiekt, zaś d^i określa klasę, do której ten obiekt należy. Załóżmy, że w wyniku grupowania tych obiektów (dowolną metodą, np. algorytmem c – średnich) uzyskujemy *c* klastrów o środkach $m^i = \left(m_1^i, ..., m_n^i, m_{n+1}^i\right)$, *i=1,...,c.* Wówczas początkową bazę reguł systemu rozmytego możemy przedstawić następująco (opracowanie na podstawie [6]):

$$R^i: \text{Jeśli } ((x_1 \text{ jest } A_1^i) \text{ i } \ldots \text{ i } (x_n \text{ jest } A_n^i))$$
$$\text{To } (y = m_{n+1}^i) \qquad (5)$$

gdzie A_j^i (*i=1,..., c; j=1, ... ,n*) są zbiorami rozmytymi o funkcjach przynależności postaci (3), w których parametry c_j^i oraz σ_j^i określamy następująco:

$$c_j^i = m_j^i,\ \sigma_j^i = \min_{t \in \{1,...,c\} \setminus \{i\}} \left| m_j^i - m_j^t \right| \qquad (6)$$

Tak wyznaczone parametry początkowe systemu neuronowo – rozmytego można później „dostrajać" stosując metody wykorzystywane w przypadku sieci neuronowych.

3.3. ZASTOSOWANIE SIATKI PODZIAŁU DO WYZNACZANIA PARAMETRÓW POCZĄTKOWYCH SYSTEMÓW NEURONOWO - ROZMYTYCH

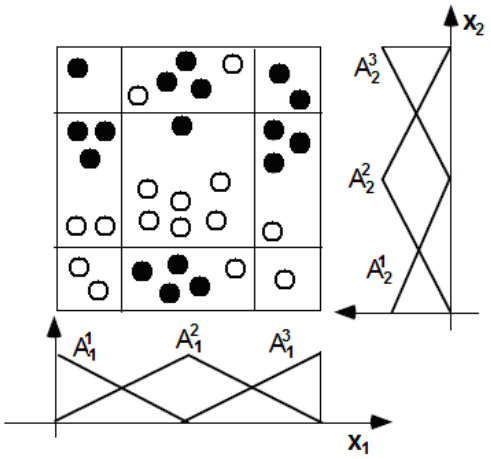
Załóżmy, że na wejście systemu podajemy wektory *n* – wymiarowe a każdej ze zmiennych wejściowych odpowiada *k* zbiorów rozmytych. Wówczas *n* – wymiarową przestrzeń ograniczoną przez zakresy zmienności odpowiednich współrzędnych danych wejściowych dzielimy na k^n obszarów i w każdym z nich określamy regułę opisującą działanie systemu. Zatem baza reguł systemu będzie zawierała k^n reguł postaci (2).

Przykład: Rozważmy system neuronowo – rozmyty, na wejścia którego podajemy dwuwymiarowe wektory. Każdy z wektorów należy do jednej z dwóch klas. Niech $d^i = 1$ będzie etykietą klasy pierwszej, zaś $d^i = 2$, klasy drugiej. Załóżmy, że w przestrzeni każdej ze zmiennych wejściowych mamy trzy zbiory rozmyte (np. początkowo rozmieszczone równomiernie na przedziale określonym przez wartość minimalną i maksymalną danej zmiennej). Wówczas przestrzeń wejściowa zostaje podzielona na 3^2 obszarów (Rys.1). Na rysunku Rys. 1 kółka czarne oznaczają obiekty należące do klasy pierwszej, zaś kółkami białymi oznaczono obiekty należące do drugiej klasy. W każdym z dziewięciu wyznaczonych obszarów można określić regułę:

$$R^i: \text{Jeśli } ((x_1 \text{ jest } A_1^s) \text{ i } (x_2 \text{ jest } A_2^t)) \text{ To } (y = d^i)$$

Wartość d^i będąca etykietą klasy w *i*-tej regule możemy wyznaczyć jako etykietę tej klasy, dla której najwięcej obiektów znajduje się w danym obszarze (nie zawsze jest to jednoznaczne). Przykładowo, reguła dla środkowego prostokąta będzie miała postać:

$$R^i: \text{Jeśli } ((x_1 \text{ jest } A_1^2) \text{ i } (x_2 \text{ jest } A_2^2)) \text{ To } (y = 2)$$



Rys. 1. Przykład siatki podziału 3x3.
Źródło: opracowanie własne

Metoda ta, w przypadku większej liczby zmiennych wejściowych prowadzi do uzyskania dużej liczby reguł.

5. OPTYMALIZACJA PARAMETRÓW POCZĄTKOWYCH SYSTEMU

Parametry początkowe systemów neuronowo – rozmytych będą „dostrajane" z wykorzystaniem dwóch metod: gradientowej metody największego spadku oraz połączenia metody największego spadku z metodą najmniejszych kwadratów.

W przypadku metody największego spadku parametry systemu modyfikujemy zgodnie z zależnością:

$$w(t+1) = w(t) - \eta \nabla E\Big|_{w(t)} \qquad (7)$$

gdzie w jest parametrem systemu, η jest tzw. *parametrem uczenia*, zaś E oznacza funkcję błędu średniokwadratowego [5].

W przypadku połączenia metody największego spadku z metodą najmniejszych kwadratów, parametry występujące w poprzednikach reguł (2) są modyfikowane zgodnie z metodą największego spadku, zaś parametry występujące w następnikach tych reguł są wyznaczane metodą najmniejszych kwadratów [2,3].

6. WYNIKI

Do przeprowadzenia doświadczeń wykorzystane zostały zbiory „iris" oraz „Pima Indian Diabetes" pochodzące z ogólnodostępnego repozytorium danych UCI [7].

Dane o irysach (zbiór „iris") zawierają opisy 150 kwiatów klasyfikowanych do trzech grup: iris – setowa, iris – versicolor i iris – virginica. Każdy kwiat opisany jest czterema liczbami: długością liścia, szerokością liścia, długością płatka i szerokością płatka.

Zbiór „Pima Indian Diabetes" zawiera dane dotyczące zachorowań na cukrzycę wśród kobiet z indiańskiego plemienia Pima. Każdy z 768 obiektów zbioru opisany jest przy pomocy 8 cech zawierających następujące informacje: ile razy pacjentka była w ciąży, test tolerancji glukozy, ciśnienie rozkurczowe, grubość zagięcia skóry, poziom insuliny, masę ciała, czy ktoś w rodzinie był chory na cukrzycę oraz wiek pacjentki. Każdy z obiektów przynależy do jednej z dwóch klas. Pierwsza klasa oznacza, że pacjentka nie choruje na cukrzycę, a druga klasa oznacza, że dana kobieta jest diabetykiem.

W dalszej części przedstawione zostaną wyniki przeprowadzonych doświadczeń. Symulacje komputerowe zostały przeprowadzone z wykorzystaniem pakietu Fuzzy Logic programu Matlab oraz przy wykorzystaniu odpowiednich procedur zaimplementowanych w środowisku Borland C++ Builder 6 Personal.

Doświadczenie 1. Parametry początkowe ustalane są jako średnie i odchylenia standardowe wartości odpowiednich atrybutów. W tabeli (Tabela 1) zostały przedstawione wyniki klasyfikacji na całych zbiorach danych (tzn. bez podziału na zbiór treningowy i testowy) bez „dostrajania" parametrów początkowych. Skuteczność klasyfikacji na danym zbiorze jest tutaj rozumiana jako stosunek liczby poprawnie sklasyfikowanych obiektów danego zbioru do liczby wszystkich obiektów tego zbioru.

Tabela 1.
Opracowanie własne

Zbiór	Skuteczność klasyfikacji
„iris"	94,67%
„Pima Indian Diabetes"	69,79%

Doświadczenie 2. Parametry początkowe ustalane przy pomocy metod grupowania. Do grupowania obiektów zostały wykorzystane następujące algorytmy: algorytm k - średnich, algorytm c – średnich, algorytm k – średnich z rozmyciem (FKM, ang. *Fuzzy K – Means*) oraz algorytm KFCM (ang. *Kernel Fuzzy C – Means*) [4,5,6,8]. W tabeli (Tabela 2) zostały przedstawione wyniki klasyfikacji na całych zbiorach danych (tzn. bez podziału na zbiór treningowy i testowy) bez „dostrajania" parametrów początkowych i przy wykorzystaniu różnych algorytmów grupowania. Ponieważ w zastosowanych algorytmach wynik klasyfikacji zależy od początkowych wartości środków klastrów, więc w tabeli podane zostały średnie wyniki klasyfikacji z dziesięciu prób.

Tabela 2.
Opracowanie własne

Metoda grupowania	Średnia skuteczność klasyfikacji	
	„iris"	„Pima Indian Diabetes"
k - średnie	95,4%	69,9%
c - średnie	93,19%	68,4%
FKM	90,93%	70%
KFCM	81,73%	71,64%

Doświadczenie 3. Zbiory dzielimy losowo na zbiór treningowy i testowy, tak aby 75% próbek każdego ze zbiorów stanowiło zbiór treningowy, a pozostałe dane – zbiór testowy. W tabeli (Tabela 3) zostały przedstawione wyniki średniej skuteczności klasyfikacji na zbiorach testowych w zależności od sposobu wyznaczania parametrów początkowych systemu przy zastosowaniu 10 - krotnej kroswalidacji Monte Carlo oraz optymalizacji parametrów systemu metodą największego spadku.

Tabela 3.
Wyniki klasyfikacji na zbiorze testowym w zależności od metody doboru parametrów początkowych systemu.
Opracowanie własne

Metoda wyznaczania parametrów początkowych	Średnia skuteczność klasyfikacji na zbiorze testowym	
	„iris"	„Pima Indian Diabetes"
Losowy wybór	86,66%	73,44%
Metoda grupowania	92,16%	73,66%
Średnie i odch. standardowe	94,66%	77,24%
Siatka podziału	93,33%	74,72%

Doświadczenie 4. Zbiór „Pima Indian Diabetes" podzielono w sposób losowy na zbiór treningowy i testowy. Zbiór treningowy zawierał 588 obiektów, zaś testowy 180. Przy takich samych parametrach początkowych dostrajano je na dwa sposoby: najpierw korzystając z metody największego spadku a następnie stosując połączenie metody największego spadku z metodą najmniejszych kwadratów. Uzyskane rezultaty przedstawia tabela (Tabela 4).

Tabela 4.
Wyniki klasyfikacji w zależności od sposobu optymalizacji parametrów początkowych systemu. Opracowanie własne

Sposób optymalizacji parametrów początkowych	Wynik na zbiorze trening.	Wynik na zbiorze testowym	Liczba epok
NS	61,39%	51,11%	2000
NS+MNK	78,57%	80,55%	100

W tabeli (Tabela 4) NS oznacza metodę największego spadku, a NS + NMK oznacza połączenie metody największego spadku z metodą najmniejszych kwadratów.

6. PORÓWNANIE UZYSKANYCH WYNIKÓW Z WYNIKAMI DOSTĘPNYMI W LITERATURZE

W tabelach (Tabela 5, Tabela 6) zestawione zostaną dostępne w literaturze wyniki klasyfikacji na zbiorach testowych dla zbiorów „iris" i „Pima Indian Diabetes" przy wykorzystaniu różnych metod klasyfikacji.

Tabela 5.
Wyniki klasyfikacji dla zbioru „iris" . Opracowanie w oparciu o [5]

Metoda klasyfikacji	Skuteczność klasyfikacji
C4.5	94,9%
SVM	93,2%
RBF	95,3%

Tabela 6.
Wyniki klasyfikacji dla zbioru „Pima Indian Diabetes" . Opracowanie w oparciu o [5]

Metoda klasyfikacji	Skuteczność klasyfikacji
FDA	76,5%
MLP + BP	76,4%
LVQ	75,8%
kNN	71,9%
RBF	75,7%

7. PODSUMOWANIE

W artykule zostały zaprezentowane cztery metody wyznaczania parametrów początkowych systemów neuronowo – rozmytych. Na podstawie uzyskanych wyników można zauważyć, że sposób wyznaczenia parametrów początkowych systemu może mieć istotny wpływ na wynik klasyfikacji (Tabela 3, zbiór „iris"). Z kolei połączenie w procesie optymalizacji parametrów początkowych systemu metody największego spadku z metodą największych kwadratów może prowadzić do wzrostu skuteczności klasyfikacji i zmniejszenia liczby epok w procesie uczenia (Tabela 4). Uzyskane wyniki klasyfikacji za pomocą systemów neuronowo – rozmytych są porównywalne z dostępnymi w literaturze wynikami uzyskanymi za pomocą innych metod klasyfikacji.

LITERATURA

[1] Jain R., Abraham A., „A Comparative Study of Fuzzy Classification Methods on Breast Cancer Data". *Australiasian Physical And Engineering Sciences in Medicine*, Australia, Volume 27, No.4, pp. 147-152, 2004

[2] Jang R., "ANFIS: Adaptive – Network – Based Fuzzy Inference System", *IEEE Trans. on Systems, Man and Cybernetics*, vol. 23, no. 3, pp. 665-685,May 1993

[3] Kądziołka K., „Zastosowanie systemów neuronowo - rozmytych do prognozowania finansowych szeregów czasowych", *Materiały Konferencji Naukowej „Innowacje w finansach i ubezpieczeniach – metody matematyczne, ekonometryczne i komputerowe*", Ustroń, Polska 2009.

[4] Kądziołka K., „Klasyfikacja kredytobiorców z wykorzystaniem rozmytej sieci neuronowej", *Materiały Konferencji Naukowej „Innowacje w finansach i ubezpieczeniach – metody matematyczne, ekonometryczne i informatyczne*", Ustroń, Polska 2008.

[5] Nałęcz M. i in., „*Biocybernetyka i inżynieria biomedyczna 2000. Tom 6: Sieci neuronowe"*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2000.

[6] Piegat A., „*Modelowanie i sterowanie rozmyte"*. Akademicka Oficyna Wydawnicza EXIT, Warszawa 1999

[7] Repozytorium UCI, dostępne na stronie internetowej: *http : //mlearn.ics.uci.edu/databases/*.

[8] Zhang D., Chen S., "Clustering incomplete data using kernel-based fuzzy c-means algorithm". *Neural Processing Letters*, 2003, 18(3), pp. 155-162

mgr Kinga Kądziołka
Akademia Ekonomiczna w Katowicach
Wydział Finansów i Ubezpieczeń
Katedra Matematyki Stosowanej

ul. S. Pilata 12/13
32- 600 Oświęcim

tel.: 693 660 531
e-mail: kkinga1981@interia.pl