

## Macierz jądra z interpretacją lingwistyczną w weryfikacji mówcy za pomocą maszyny wektorów podtrzymujących

### Linguistic Interpretation Of The Kernel Matrix In Speaker Verification With Support Vector Machine

Mariusz Bąk, *Politechnika Śląska*

#### Abstract

The paper shows that support vector classifier with linguistic interpretation of the kernel matrix can be effectively used in speaker verification. The kernel matrix is obtained by means of fuzzy clustering, based on global learning of fuzzy system with logical interpretation of if-then rules and with parametric conclusions. The kernel matrix is data-dependent and may be interpreted in terms of linguistic values related to the premises of if-then rules. Simulation results obtained for SPIDRE corpus are presented for comparison with traditional methods used in speaker verification.

#### Streszczenie

Artykuł prezentuje możliwość zastosowania maszyny wektorów podtrzymujących z lingwistyczną interpretacją funkcji jądra w weryfikacji mówcy. Zależna od danych macierz jądra jest uzyskiwana w wyniku rozmytego grupowania danych ze zbioru uczącego, podobnie jak ma to miejsce w systemie wnioskowania rozmytego z parametrycznymi konkluzjami. W artykule przedstawiono wyniki symulacji uzyskanych dla zasobu mowy SPIDRE.

#### 1. Wprowadzenie

Maszyna wektorów podtrzymujących (SVM) była pierwszą zaproponowaną metodą wykorzystującą funkcję jądra [1]. Wykorzystuje ona możliwość przekształcenia danych z przestrzeni wejściowej, charakterystycznej dla danego problemu klasyfikacji, do przestrzeni cech, która może charakteryzować się wysoką (lub nawet nieskończoną) wymiarowością. W tej przestrzeni właśnie prowadzone są poszukiwania optymalnej hiperplaszczyny decyzyjnej. Powszechnie stosuje się jądra wielomianowe, gaussowskie czy sigmoidalne. Podczas poszukiwania hiperplaszczyny decyzyjnej najważniejsze jest osiągnięcie maksymalnej zdolności

uogólniania, zależnej od ryzyka empirycznego oraz złożoności nauczanej maszyny [1,9]. Maszyny wektorów podtrzymujących mogą być wykorzystane dla rozwiązania różnych problemów z dziedziny klasyfikacji oraz regresji.

Idea macierzy jądra z interpretacją lingwistyczną zaprezentowana została w pracy [5] dla regresji. Wykorzystana została tam macierz jądra uzyskana w wyniku rozmytego grupowania danych uczących, przy jednoczesnym wskazaniu na znaczące analogie występujące między  $\varepsilon$ -nieczulym nauczaniem systemu wnioskowania rozmytego a regresją za pomocą maszyny wektorów podtrzymujących. Ponieważ macierz jądra jest uzyskiwana w wyniku rozmytego grupowania, może ona być interpretowana za pomocą wartości lingwistycznych opartych na przesłankach reguł "jeżeli-to" rozmytego systemu wnioskowania.

Artykuł prezentuje możliwość zastosowania wymienionej wyżej metody uzyskiwania macierzy jądra na gruncie klasyfikacji, w weryfikacji mówcy. Uzyskane w ten sposób wyniki weryfikacji porównane są z tymi uzyskanymi dla modeli GMM-LR (ang. *Gaussian Mixture Model Likelihood Ratio*) [8].

#### 2. System wnioskowania rozmytego z logiczną interpretacją reguł „jeżeli-to”

Załóżmy, że dany jest zbiór  $I$  rozmytych reguł „jeżeli-to” z  $t$  wejściami singletonowymi  $x_{01}, \dots, x_{0t}$  i pojedynczym wyjściem  $Y$ . Niech  $i$ -ta reguła będzie wyrażona wzorem [3]:

$$\mathfrak{R}^{(i)} : \text{IF } \mathbf{x}_0 \text{ IS } \mathbf{A}^{(i)} \text{ THEN } Y \text{ IS } B^{(i)}(\mathbf{x}_0) \quad (1)$$

gdzie  $\mathbf{x}_0 = [x_{01}, \dots, x_{0t}]^T$ ,  $Y$  jest zmienną lingwistyczną, a  $\mathbf{A}^{(i)}$  oraz  $B^{(i)}(\mathbf{x}_0)$  są wartościami

lingwistycznymi. Pojedyncza reguła (1) nazywana jest regułą z ruchomą konkluzją.

Jeżeli wykorzystywana jest logiczna interpretacja reguł rozmytych, funkcją przynależności dla wartości wyjściowej  $i$ -tej reguły możemy zapisać jako [3]:

$$\mu_{B^{(i)}}(y, \mathbf{x}_0) = I(\mu_{A^{(i)}}(\mathbf{x}_0), \mu_{B^{(i)}}(y, \mathbf{x}_0)) \quad (2)$$

gdzie  $I(\bullet, \bullet)$  jest wybraną implikacją rozmytą. Ostra wartość wyjściowa może być uzyskana za pomocą metody MICOOG [3].

Zazwyczaj zakłada się, że położenie zbioru rozmytego w konkluzji  $i$ -tej reguły jest kombinacją liniową jej wejść [3]

$$y^{(i)}(\mathbf{x}_0) = \mathbf{p}^{(i)T} \mathbf{x}'_0 \quad (3)$$

gdzie  $\mathbf{p}^{(i)}$  oznacza wektor parametrów, a  $\mathbf{x}'_0$  jest rozszerzonym wektorem wejściowym, uzupełnionym o element jednostkowy. Jeżeli dodatkowo przyjmując założenie, że konkluzje  $B^{(i)}$  mają funkcje przynależności o kształcie trójkątów równoramiennych i podstawach szerokości  $w^{(i)}$ , to ostrą wartość wyjściową dla całego systemu można wyrazić jako [3]

$$y_0(\mathbf{x}_0) = \sum_{i=1}^I S^{(i)}(\mathbf{x}_0) \mathbf{p}^{(i)T} \mathbf{x}'_0 \quad (4)$$

gdzie  $S^{(i)}(\mathbf{x}_0)$  może być nazwane znormalizowanym współczynnikiem interakcji danej  $\mathbf{x}_0$  z konkluzją  $i$ -tej reguły rozmytej „jeżeli-to” i może zostać wyrażone wzorem [3]

$$S^{(i)}(\mathbf{x}_0) = \frac{G(\mu_{A^{(i)}}(\mathbf{x}_0), w^{(i)})}{\sum_{k=1}^I G(\mu_{A^{(k)}}(\mathbf{x}_0), w^{(k)})} \quad (5)$$

gdzie funkcja  $G$  zależy od wybranej implikacji wykorzystywanej do interpretacji reguł rozmytych  $I(\bullet, \bullet)$ . Wówczas można zapisać również [3]

$$y_0(\mathbf{x}_0) = \mathbf{P}^T \mathbf{d}(\mathbf{x}_0) \quad (6)$$

gdzie

$$\mathbf{d}(\mathbf{x}_0) = [S^{(1)}(\mathbf{x}_0) \mathbf{x}'_0{}^T, \dots, S^{(I)}(\mathbf{x}_0) \mathbf{x}'_0{}^T]^T \quad (7)$$

$$\mathbf{P} = [\mathbf{p}^{(1)T}, \dots, \mathbf{p}^{(I)T}]^T \quad (8)$$

Wartości lingwistyczne  $\mathbf{A}^{(i)}$  w przesłankach reguł rozmytych mogą zostać znalezione z wykorzystaniem rozmytego grupowania za pomocą algorytmu rozmytych  $c$ -średnich (ang. *fuzzy c-means*, FCM). Uzyskane dla poszczególnych wektorów wartości przynależności do poszczególnych klas są następnie przekształcane na gausowskie funkcje przynależności o wartościach średnich  $\mathbf{c}^{(i)}$  oraz dyspersjach  $\mathbf{s}^{(i)}$ .

### 3. Maszyna wektorów podtrzymujących z macierzą jądra o interpretacji lingwistycznej

W pracy [5] wykazano, że  $\mathcal{E}$ -nieczule nauczanie przedstawionego wcześniej systemu wnioskowania rozmytego jest równoważne regresji za pomocą maszyny wektorów podtrzymujących przy spełnieniu dwóch warunków:

- macierz jądra musi mieć postać

$$K = \left[ \tilde{\mathbf{d}}(\mathbf{x}_n)^T \tilde{\mathbf{d}}(\mathbf{x}_j) \right]_{n,j=1}^N \quad (9)$$

- parametr regularyzacji musi mieć wartość

$$C = \frac{1}{\tau} \quad (10)$$

gdzie  $\tilde{\mathbf{d}}(\mathbf{x}_n)$  jest zawężoną wersją wektora  $\mathbf{d}(\mathbf{x}_n)$ ,  $N$  jest liczbą elementów w zbiorze uczącym, a  $\tau > 0$  kontroluje złożoność modelu uzyskiwanego w wyniku regresji.

Macierz jądra może być interpretowana w oparciu o wartości lingwistyczne  $\mathbf{A}^{(i)}$  obecne w przesłankach reguł rozmytych „jeżeli-to”.

Podobne podejście można zaproponować w przypadku klasyfikacji, jeżeli  $\mathcal{E}$  jest równe 0. Zbiór uczący  $T$  składa się z dwóch podzbiorów  $T^{(+)}$  oraz  $T^{(-)}$ , grupujących wektory oznaczone odpowiednio etykietami +1 i -1. Nauczanie takie może być przeprowadzone w następujący sposób:

1. Wybór implikacji rozmytej  $I(\bullet, \bullet)$  służącej do interpretacji reguł rozmytych oraz odpowiadającej jej funkcji  $G$ , liczby reguł rozmytych  $I$ , szerokości podstaw funkcji przynależności w przesłankach reguł rozmytych  $w^{(i)}$  oraz wartości parametru regularyzacji  $\tau$ .
2. Oddzielne grupowanie danych ze zbiorów  $T^{(+)}$  oraz  $T^{(-)}$  za pomocą algorytmu FCM i przekształcenie uzyskanych wartości przynależności do każdej z  $c$  klas na gausowskie funkcje przynależności o środkach  $\mathbf{c}^{(i)(+)}$ ,  $\mathbf{c}^{(i)(-)}$  i dyspersjach  $\mathbf{s}^{(i)(+)}$ ,  $\mathbf{s}^{(i)(-)}$ .
3. Grupowanie uzyskanych rozkładów gausowskich uzyskanych dla  $T^{(+)}$  i  $T^{(-)}$  parami, rozpoczynając od leżących najbliżej siebie.
4. Obliczenie wartości współczynnika  $S^{(i)}(\mathbf{x}_n)$  dla każdego wektora wejściowego z  $T$  i każdej reguły rozmytej z wykorzystaniem wybranej funkcji  $G$ .

5. Uzyskanie macierzy jądra postaci

$$K = \left[ \sum_{i=1}^I S^{(i)}(\mathbf{x}_n) S^{(i)}(\mathbf{x}_j) \mathbf{x}_n^T \mathbf{x}_j \right]_{n,j=1}^N \quad (11)$$

6. Nauczanie maszyny wektorów podtrzymujących z wykorzystaniem zbioru  $T$ , macierzy jądra  $K$  oraz parametru  $C = \frac{1}{\tau}$ .

7. Obliczenie wektora parametrów systemu wnioskowania rozmytego w oparciu o zależności

$$\tilde{\mathbf{P}} = \sum_{i \in I_{SV}} \alpha_i \tilde{\mathbf{d}}(\mathbf{x}_i) y_i \quad (12)$$

$$b = \frac{1}{I_{SV}} \sum_{i \in I_{SV}} y_i \tilde{\mathbf{d}}(\mathbf{x}_i)^T \tilde{\mathbf{P}} \quad (13)$$

gdzie  $I_{SV}$  jest zbiorem wektorów podtrzymujących, to jest takich, dla których mnożniki Lagrange'a  $\alpha_i$  uzyskane w wyniku nauczania mają wartości niezerowe.

Ostra wartość wyjściowa dla dowolnego wektora wejściowego  $\mathbf{x}$  może zostać uzyskana za pomocą wzoru:

$$y_0(\mathbf{x}_0) = \tilde{\mathbf{P}}^T \tilde{\mathbf{d}}(\mathbf{x}_0) + b \quad (14)$$

Obliczenie jej wymaga znajomości środków  $\mathbf{c}^{(i)(+1)}$ ,  $\mathbf{c}^{(i)(-1)}$  i dyspersji  $\mathbf{s}^{(i)(+1)}$ ,  $\mathbf{s}^{(i)(-1)}$  gaussowskich funkcji przynależności, wartości  $w^{(i)}$  oraz uzyskanych podczas nauczania wartości  $\mathbf{P}$  i  $b$ .

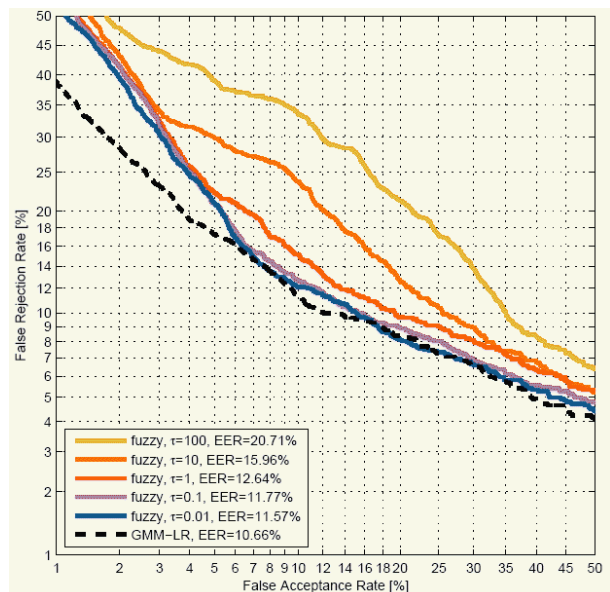
#### 4. Symulacje

Symulacje przeprowadzono wykorzystując zasób mowy SPIDRE 1996. Zawiera on nagrania rozmów telefonicznych prowadzonych w języku angielskim. Łącznie obejmuje nagrania 317 mówców, z czego 44 mówców może zostać wykorzystanych jako mówcy uprawnieni podczas testowania systemu weryfikacji. Symulacje prowadzone były w środowisku MATLAB, a maszyna wektorów podtrzymujących symulowana była z wykorzystaniem zmodyfikowanej wersji pakietu LIBSVM [2,4]. W artykule przedstawiono uzyskane krzywe DET [6] (ang. *Detection Error Trade-off*) oraz wartości EER (ang. *Equal Error Rate*) dla każdego testowanego systemu. Każda z krzywych *DET* reprezentuje skuteczność systemu weryfikacji dla różnych wartości progów decyzyjnego.

Przetwarzanie rozmów zawartych w zasobie SPIDRE obejmowało separację wypowiedzi każdego z mówców, normalizację ich głośności oraz preemfazę. Sygnał mowy był następnie dzielony na ramki długości 20ms z zakładką 10ms (160 próbek na ramkę) i okienkowany za

pomocą okna Hamminga. Następnie przeprowadzono analizę LPC [7] (ang. *Linear Predictive Coding*), w celu uzyskania obwiedni widma chwilowego sygnału mowy i wyeliminowania wpływu tonu krtaniowego na wykorzystywane w klasyfikacji parametry. Ostatnim krokiem było obliczenie w oparciu o uzyskane 13 współczynników transmitancji LPC 19 współczynników LPC-cepstrum [7]. Po pominięciu współczynnika związanego z wzmocnieniem sygnału uzyskiwano wektory wchodzące w skład zbiorów uczących i testujących złożone z 18 elementów.

Weryfikacja była przeprowadzana za pomocą przedstawionego powyżej systemu rozmytego oraz systemu GMM-LR [8]. Wykorzystano odpowiednio 8 reguł rozmytych oraz 8 gaussowskich rozkładów składowych dla modelu mówcy uprawnionego i 64 rozkłady składowe dla modelu odniesienia. Wartość  $\tau$  przyjmowała wartości od 100 do 0,01.

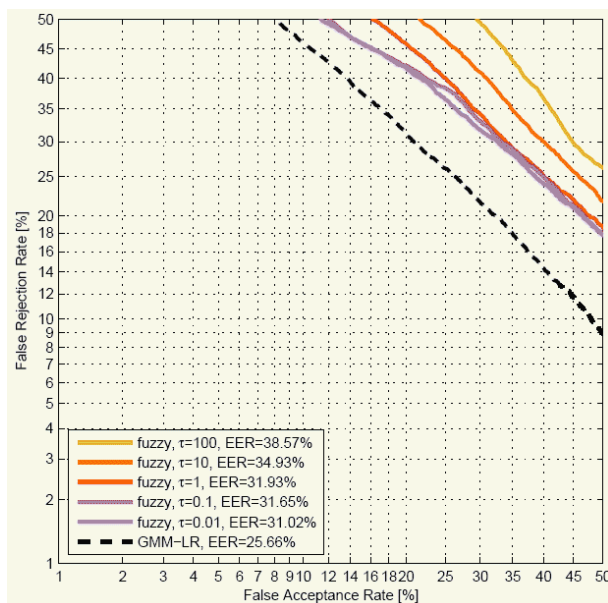


Rys.1. Krzywa DET dla tego samego telefonu wykorzystywanego podczas nauczania i testowania  
 Fig.1. DET curve plot for the same handset being used during learning and testing phase

#### 4. Wyniki symulacji

Uzyskane krzywe DET dla rozmów prowadzonych z tego samego aparatu telefonicznego przedstawione są na wykresie 1. Wartość EER uzyskiwana dla dużych wartości współczynnika  $\tau$  jest zdecydowanie większa niż dla systemu GMM-LR, jednak przy zmniejszaniu wartości  $\tau$  poniżej 0,1 luka ta zmniejsza się ( $EER = 10,66\%$  dla systemu GMM-LR oraz  $EER = 11,57\%$  dla systemu rozmytego i  $\tau = 0,01$ ).

Krzywe DET uzyskane dla rozmów prowadzonych z wykorzystaniem różnych aparatów telefonicznych przedstawione są na wykresie 2. W tym przypadku różnica w wynikach uzyskanych przez oba systemy jest wyraźna. Dla małych wartości  $\tau$  luka ta ponownie się zmniejsza i dla  $\tau = 0,01$  uzyskujemy  $EER = 25,66\%$  dla systemu GMM-LR oraz  $EER = 31,02\%$  dla systemu rozmytego.



Rys.2. Krzywa DET dla różnych telefonów wykorzystywanych podczas nauczania i testowania  
 Fig.2. DET curve plot for different handsets being used during learning and testing phase

## 5. Wnioski

Przedstawiony w tym artykule system rozmyty jest uzyskiwany poprzez nauczanie maszyny wektorów podtrzymujących z wykorzystaniem macierzy jądra, która posiada interpretację lingwistyczną. Proponowany system weryfikacji uzyskuje wyniki zbliżone do systemu GMM-LR jeżeli podczas rozmów wykorzystywany jest ten sam aparat telefoniczny. Ponieważ uzyskiwana przy zastosowaniu opisanej funkcji jądra przestrzeń cech ma skończoną wymiarowość, wektor normalny płaszczyzny hiperdecyzyjnej może być wykorzystany w postaci jawnej podczas obliczania wartości funkcji decyzyjnej.

## Literatura

1. Boser BE, Guyon IM, Vapnik V: *A training algorithm for optimal margin classifier*, w: Proc. 5th Annu. ACM Workshop on Computational Learning Theory, Pittsburgh, USA, 1992, pp. 144-152.
2. Chang CC, Lin CJ: *LIBSVM: a library for support vector machines*, źródła dostępne na

<http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.

3. Czogala E, Łęski J: *Fuzzy and Neuro-fuzzy Intelligent Systems*, Physica-Verlag, Heidelberg, 2000.
4. Fan RE, Chen PH, Lin CJ: *Working set selection using second order information for training SVM*, w: Journal of Machine Learning Research 6, 2005, pp. 1889-1918.
5. Łęski J: *On support vector regression machines with linguistic interpretation of the kernel matrix*. In: Fuzzy Sets and Systems 157, 2006, pp. 1092-1113.
6. Martin A, Doddington F, Kamm T, Ordowski M, Przybocki M: *The DET curve in assesment of detection task performance*, w: The 5th European Conference on Speech Communication and Technology EUROSPEECH 97, 1997, pp. 1895-1898.
7. Rabiner LR, Schafer RW: *Digital processing of speech signals*, Prentice Hall, Englewood Cliffs, 1978.
8. Reynolds DA: *Speaker identification and verification using Gaussian mixture speaker models*, w: Speech Communication, vol. 17, no. 1-2, 1995, pp. 91-108
9. Vapnik V: *The Nature of Statistical Learning Theory*, Springer, New York, 1995.

## Adres służbowy Autora:



mgr inż. Mariusz Bąk  
 Instytut Elektroniki  
 Politechnika Śląska  
 tel. +48 32 237-15-29  
 fax +48 32 237-22-25  
 mariusz.bak@polsl.pl