

Maria MRÓWCZYŃSKA*

Klasyfikatory neuronowe typu SVM w zastosowaniu do klasyfikacji przemieszczeń pionowych na obszarze LGOM

Streszczenie: W prezentowanym artykule przedstawiono podstawowe zasady budowania i uczenia sieci neuronowych SVM (ang. *Support Vector Machine*) zwane inaczej metodą (techniką) wektorów podtrzymujących. Sieci SVM znajdują głównie zastosowanie w rozwiązywaniu zadań klasyfikacji danych separowalnych i nieseparowalnych liniowo oraz zadań regresji. W ostatnich latach zakres zastosowań tego typu sieci został poszerzony i sieci rozwiązują również takie problemy jak rozpoznawanie sygnałów i obrazów, identyfikacja mowy oraz diagnostyka medyczna.

W pracy sieci nieliniowe SVM wykorzystano do klasyfikacji danych nieseparowalnych liniowo w postaci przemieszczeń punktów sieci pomiarowo-kontrolnej reprezentujących obszar, na którym prowadzona jest eksploatacja górnicza. Uczenia sieci neuronowej SVM wymaga implementacji programowania kwadratowego w poszukiwaniu punktu optymalnego funkcji Lagrange'a względem optymalizowanych parametrów. W przypadku danych nieseparowalnych liniowo, metoda SVM pozwala na znalezienie hiperpłaszczyzny, która klasyfikuje obiekty na tyle poprawnie, na ile jest to możliwe i jednocześnie przebiega możliwe daleko od typowych skupień dla każdej z klas. Za pomocą metody SVM można również znaleźć krzywoliniową granicę separacji o dużym marginesie separacji, wykorzystując zabieg podniesienia wymiarowości. Jakość uzyskanych wyników separacji jest uzależniona od przyjętej postaci funkcji jądra.

Słowa kluczowe: sieci neuronowe SVM, klasyfikacja, przemieszczenia pionowe

SVM neural classifiers used for classifying vertical displacements in the Legnica-Głogów Copper Mining Area

Abstract: This article presents basic rules for constructing and training SVM neural networks. SVM neural networks are mainly used for solving tasks involving the classification of linearly and non-linearly separable data, as well as regression tasks. In recent years, the application of these types of networks has expanded, and now they are also used for solving problems such as recognition of signals and pictures, speech identification, and in medical diagnostics.

* Uniwersytet Zielonogórski, Zielona Góra

In this analysis non-linear SVM networks were used for classifying linearly non-separable data in the form of vertical displacements of points representing a mining area in a measurement and control network. Training an SVM neural network requires the use of quadrant programming in search of an optimum point of a Lagrangian function in relation to the parameters which are being optimised. In the case of linearly non-separable data, the SVM method makes it possible to find a hyper plane which classifies objects as correctly as possible, and at the same time is located far enough from concentrations typical of each class. By means of raising dimensionality, the SVM method can also be used to find a curvilinear separation boundary with a wide separation margin. The quality of the results obtained depends on the adopted form of the kernel function.

Key words: SVM neural networks, classification, vertical displacements

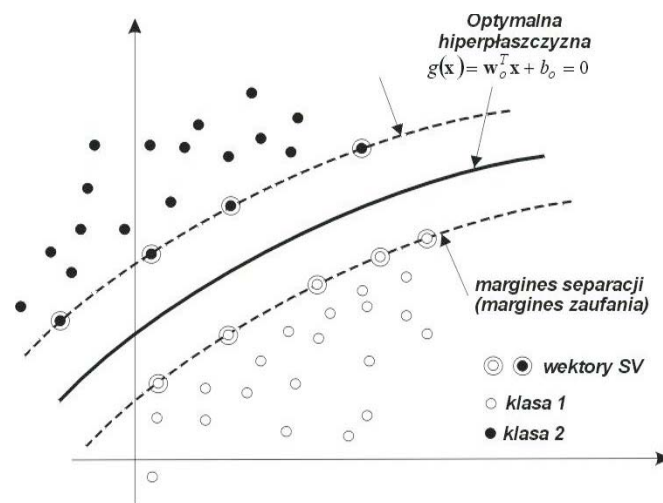
Wprowadzenie

W drugiej połowie lat dziewięćdziesiątych ubiegłego wieku zostały opracowane przez Vapnika klasyfikatory SVM, zwane techniką wektorów podtrzymujących (Vapnik 1998). Metoda ta stanowiła nowe podejście zarówno pod względem budowy jak również uczenia sieci neuronowych, zaś wykorzystywana jest przede wszystkim do rozwiązywania zadań klasyfikacji i regresji. W kolejnych latach rozwijano tematykę SVM, badając zarówno możliwości zastosowania sieci do rozwiązywania bardziej złożonych problemów jak i poszukując optymalnych algorytmów uczących. Wykorzystane w pracy sieci SVM należą do grupy sieci jednokierunkowych, zwykle mają strukturę dwuwarstwową i mogą wykorzystywać różne typy funkcji aktywacji (Osowski 2006; Zanni i in. 2006). W porównaniu z sieciami neuronowymi MLP (ang. *Multi Layer Perceptron* – perceptron wielowarstwowy) sieci SVM pozbawione są wad typowych dla sieci MLP, tzn. przede wszystkim możliwości zatrzymania procesu minimalizacji w jednym z wielu minimów lokalnych, a także przyjmowanej arbitralnie na wstępie architektury sieci, od której zależą jej przyszłe zdolności generalizacyjne. W przypadku danych liniowo separowalnych metoda SVM pozwala na znalezienie hiperpłaszczyzny dzielącej zbiór danych na dwie klasy z maksymalnym marginesem separacji. Jeżeli mamy do czynienia z danymi nieseparowalnymi liniowo, za pomocą metody wektorów podtrzymujących można znaleźć hiperpłaszczyznę klasyfikującą obiekty z minimalnym błędem i jednocześnie przebiegającą możliwie daleko od typowych skupień dla każdej klasy.

W pracy zostanie zaprezentowane zagadnienie wykorzystania sieci nieliniowej SVM do klasyfikacji danych nieseparowalnych liniowo, na przykładzie przemieszczeń pionowych zaobserwowanych na obszarze Legnicko-Głogowskiego Okręgu Miedziowego. Przemieszczenia pionowe wyznaczono na podstawie wyników pomiarów niwelacyjnych zrealizowanych w latach 1967–2008.

1. Charakterystyka sieci nieliniowej SVM

Istotą działania metody SVM jest konstrukcja optymalnej hiperpłaszczyzny, która będzie separowała dane należące do różnych klas, z maksymalnym marginesem zaufania (marginesem separacji). Przez margines zaufania będziemy rozumieli odległość hiperpłaszczyzny od najbliższych położonych punktów, na których będą się tworzyły wektory podtrzymujące (rys. 1). Punkty, na których utworzą się wektory podtrzymujące leżą najbliższej hiperpłaszczy-



Rys. 1. Optymalna hiperpłaszczyzna o maksymalnym marginesie separacji

Fig. 1. The optimal hyperplane with a maximum margin of separation

czynny i określają jej przebieg, ale jednocześnie są najtrudniejsze do klasyfikacji. Szersze informacje dotyczące klasyfikacji danych liniowo separowalnych oraz sposobów budowy optymalnej hiperpłaszczyzny można znaleźć między innymi w pracach: Bishop'a (2006), Jankowskiego (2003), Osowskiego (2006).

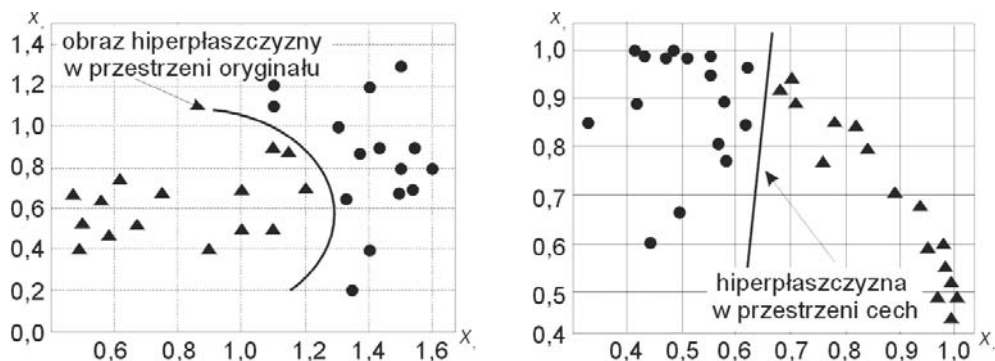
Rozwiązując zadanie klasyfikacji danych liniowo nieseparowalnych powszechnie stosowanym rozwiązaniem jest zrzutowanie danych oryginalnych do przestrzeni funkcyjnej (przestrzeni cech), w której dane stają się z prawdopodobieństwem bliskim 1 liniowo separowalne. Zwykle wymiar przestrzeni cech K jest dużo większy niż wymiar przestrzeni oryginału N , a przeprowadzona transformacja jednej przestrzeni do drugiej jest transformacją nieliniową (Haykin 1994; Cover 1965). Graficzna ilustracja nieliniowej transformacji danych nieseparowalnych liniowo została przedstawiona na rysunkach 2a i 2b. Dane liniowo nieseparowalne w dwuwymiarowej przestrzeni oryginału (rys. 2a) zostały przetransformowane do przestrzeni cech (rys. 2b), którą zdefiniowano za pośrednictwem funkcji gaussowskich

$$\varphi(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}\|^2}{\sigma^2}\right) \quad (1)$$

gdzie:

- σ – szerokość funkcji gaussowskiej,
- \mathbf{c} – centra funkcji gaussowskich,
- \mathbf{x} – wektor wejściowy.

Po przeprowadzonej transformacji nieliniowej dane stają się liniowo separowalne i mogą zostać rozdzielone jedną płaszczyzną separacji. Przebieg hiperpłaszczyzny separującej klasy



Rys. 2. Dane liniowo nieseparowalne w przestrzeni oryginalnej (rysunek po lewej) oraz dane liniowo separowalne w przestrzeni cech (rysunek po prawej)

Fig. 2. The data nonlinearly separable in the original space (figure on the left) and the data linearly separable in the feature space (figure on the right)

wyznaczany jest w przestrzeni cech, a w przestrzeni oryginalnej obserwujemy jedynie jej obraz (rys. 2).

Założmy, że klasyfikacji podlega zbiór par uczących (\mathbf{x}_i, d_i) , $i = 1, \dots, N$, gdzie wartość zadana d_i jest równa 1 lub -1 natomiast \mathbf{x}_i jest wektorem wejściowym, który po zrzutowaniu w przestrzeń K -wymiarową jest reprezentowany przez zbiór cech $\varphi_j(\mathbf{x})$, $j = 1, \dots, K$. Po tak przeprowadzonej transformacji, równanie hiperpłaszczyzny separującej dane w przestrzeni cech, zapiszemy jako

$$g(\mathbf{x}) = \sum_{j=1}^K w_j \varphi_j(\mathbf{x}) + b = 0 \quad (2)$$

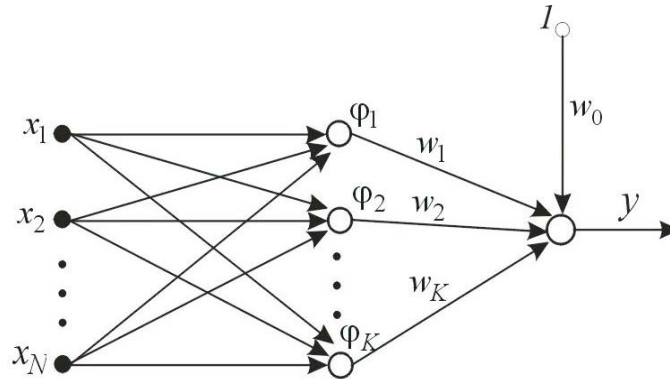
gdzie:

- w_j – waga prowadząca od neuronu w warstwie ukrytej do neuronu wyjściowego (rys. 3),
- b – polaryzacja, określająca położenie hiperpłaszczyzny względem początku układu współrzędnych.

Sygnal neuronu wyjściowego dla sieci o architekturze przedstawionej na rysunku 3 zdefiniowany jest za pomocą równania

$$y(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b \quad (3)$$

Analizując podstawową strukturę sieci neuronowej SVM (rys. 3) można zauważyć, że jest to struktura analogiczna do tej, jaką posiadają sieci o radialnych funkcjach bazowych (RBF). Różnica pomiędzy nieliniową siecią SVM a siecią radialną polega na tym, że funkcje $\varphi(\mathbf{x})$ mogą przyjmować postać liniową, wielomianową, radialną bądź sigmoidalną.



Rys. 3. Architektura podstawowa sieci nieliniowej SVM

Fig. 3. The basic architecture of the SVM nonlinear network

Uczenie sieci nieliniowej SVM ma na celu takie wyznaczenie wartości wektora wag \mathbf{w} , aby dla danych nieseparowalnych liniowo określić optymalną hiperpłaszczyznę, która minimalizuje prawdopodobieństwo popełnienia błędu klasyfikacji przy jednoczesnym zachowaniu warunku maksymalizacji marginesu separacji. Klasyfikując dane nieseparowalne liniowo należy zdefiniować nieujemną zmienną dopełniającą λ , której zadaniem jest zmniejszenie aktualnej szerokości marginesu separacji. Tak postawiony problem jest określany mianem problemu pierwotnego, który zapiszemy jako

$$\min\{\varphi(\mathbf{w}, \lambda)\} = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^p \lambda_i \quad (4)$$

przy ograniczeniach

$$d_i (\mathbf{w}^T \varphi(\mathbf{x})_i + b) \geq 1 - \lambda_i \quad (5)$$

$$\lambda_i \geq 0$$

gdzie:

C – parametr przyjmowany arbitralnie przez użytkownika,

d_i – wartość zadań równa +1.

W początkowej fazie uczenia sieci nieliniowej SVM liczba wektorów podtrzymujących jest zwykle równa liczbie danych uczących. W trakcie procesu uczenia, w zależności od przyjętej wartości parametru C (wartości ograniczeń), złożoność sieci jest redukowana i tylko na części punktów tworzą się wektory podtrzymujące. Wektory podtrzymujące tworzą się na tych punktach, dla których spełniony jest warunek

$$\mathbf{w}^T \varphi(\mathbf{x})_i + b = \pm 1 \quad (6)$$

Warto podkreślić, że im większa jest wartość parametru C tym węższy jest margines separacji i mniejsza liczba wektorów podtrzymujących. Dla małej wartości parametru C sieć doznaje zbliżenia w swoim działaniu do sieci liniowej, przez co poszerza się margines separacji.

Problem pierwotny (optymalizacyjny) jest problemem programowania kwadratowego z liniowymi ograniczeniami względem wag, który rozwiązujemy metodą mnożników Lagrange'a na podstawie minimalizacji funkcji Lagrange'a (Bishop 2006; Gunn 1998)

$$J(\mathbf{w}, b, \lambda, \alpha, \mu) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^p \alpha_i - \sum_{i=1}^p \alpha_i [d_i (\mathbf{w}^T \varphi(\mathbf{x})_i + b) - 1 + \lambda_i] - \sum_{i=1}^p \mu_i \lambda_i \quad (7)$$

gdzie:

$\alpha_i \geq 0$ oraz $\mu_i \geq 0$ – mnożniki Lagrange'a.

Problem optymalizacyjny zapisany wzorem (7) rozwiązujemy poprzez przyrównanie pierwszych pochodnych funkcji Lagrange'a względem \mathbf{w} , b oraz λ do zera. W rezultacie otrzymujemy współrzędne wektora wag \mathbf{w} wyrażone za pomocą mnożników Lagrange'a oraz dodatkowych zależności funkcyjnych, które muszą zostać spełnione (Scholkopf i Smola 2001). Problem pierwotny przekształca się więc w problem dualny zdefiniowany względem mnożników Lagrange'a do postaci (Haykin 1994)

$$\max \left\{ \sum_{i=1}^p \alpha_i - \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \alpha_i \alpha_j d_i d_j K(\mathbf{x}^T \mathbf{x}_i) \right\} \quad (8)$$

przy ograniczeniach

$$\sum_{i=1}^p \alpha_i d_i = 0 \quad (9)$$

$$0 \leq \alpha_i \leq C$$

Rozwiązanie problemu dualnego pozwala na wyznaczenie optymalnych wartości mnożników Lagrange'a, które są podstawą do wyznaczenia optymalnych wartości wag sieci według zależności

$$\mathbf{w} = \sum_{i=1}^p \alpha_i d_i \varphi(\mathbf{x})_i \quad (10)$$

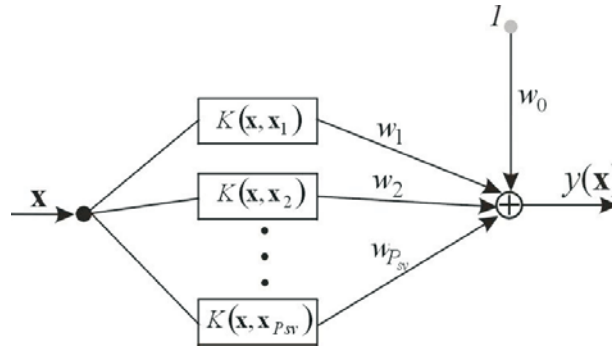
Sygnal wyjściowy sieci nieliniowej SVM uzyskujemy, podstawiając zależność (10) do wzoru (3) i definiujemy go ostatecznie jako

$$y(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b = \sum_{i=1}^{P_{sv}} \alpha_i d_i K(\mathbf{x}, \mathbf{x}_i) + b \quad (11)$$

gdzie:

- P_{sv} – liczba wektorów podtrzymujących x_i , która jest równa liczbie niezerowych mnożników Lagrange'a,
- $K(\mathbf{x}, \mathbf{x}_i)$ – funkcja jądra (ang. *kernel function*).

Sygnal wyjściowy sieci nieliniowej SVM zależy od funkcji jądra $K(\mathbf{x}, \mathbf{x}_i)$, a nie od funkcji $\varphi(x)$ jak ma to miejsce w przypadku sieci o radialnych funkcjach bazowych. Z tego względu staje się oczywiste, że należy zbudować sieć o strukturze wynikowej przedstawionej na rysunku 4.



Rys. 4. Architektura wynikowa sieci nieliniowej SVM

Fig. 4. Resulting architecture of the SVM nonlinear network

2. Funkcje jądra

Funkcja jądra $K(\mathbf{x}, \mathbf{x}_i)$ występująca w sformułowaniu zadania dualnego (8) oraz w zapisie sygnału wyjściowego (11) sieci nieliniowej SVM jest funkcją symetryczną w postaci iloczynu skalarnego dwóch funkcji wektorowych $\varphi(\mathbf{x})$ oraz $\varphi(\mathbf{x}_i)$. Analizując postać sygnału wyjściowego (11) wynikowej sieci SVM zauważymy, że konstrukcja hiperpłaszczyzny separującej dwie klasy nie wymaga odwzorowania $\varphi(\mathbf{x})$ w postaci jawnej lecz jest zastępowana jądrem $K(\mathbf{x}, \mathbf{x}_i)$, które spełnia warunek iloczynu skalarnego. Z drugiej strony funkcja $K(\mathbf{x}, \mathbf{x}_i)$ może zostać wykorzystana w sieci nieliniowej SVM tylko wówczas, jeżeli spełnia warunek twierdzenia Melcera (Scholkopf i Smola 2001). Warunek Melcera odpowiada na pytanie, czy analizowana funkcja jądra może być przedstawiona w postaci iloczynu skalarnego dwóch funkcji wektorowych. Zgodnie z tym twierdzeniem, funkcja $K(\mathbf{x}, \mathbf{x}_i)$ jest rozwijalna w szereg

$$K(\mathbf{x}, \mathbf{x}_i) = \sum_{i=1}^{\infty} \lambda_i \varphi(\mathbf{x}) \varphi(\mathbf{x}_i) \quad (12)$$

z uwzględnieniem nieujemnej zmiennej dopełniającej λ_i , jeśli dla dowolnej funkcji $g(\mathbf{x})$ (por. (2)) spełniającej warunek

$$\int g^2(\mathbf{x})d\mathbf{x} < \infty \quad (13)$$

zachodzi

$$\int K(\mathbf{x}, \mathbf{x}_i)g(\mathbf{x})g(\mathbf{x}_i)d\mathbf{x}d\mathbf{x}_i \geq 0 \quad (14)$$

Warunek Melcera spełnia wiele funkcji, które mogą być wykorzystywane do budowy nieliniowej sieci neuronowej SVM. Najczęściej wykorzystywane funkcje jąder zostały zestawione w tabeli 1.

TABELA 1. Przykłady funkcji jąder

TABLE 1. Exemples of kernel function

Typ jądra	Równanie $K(\mathbf{x}, \mathbf{x}_i)$	Komentarz
Liniowe	$K(\mathbf{x}, \mathbf{x}_i) = \mathbf{x}^T \mathbf{x}_i$	
Wielomianowe	$K(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x}^T \mathbf{x}_i + 1)^b$	b – stopień wielomianu
Radialne (gaussowskie)	$K(\mathbf{x}, \mathbf{x}_i) = \exp \left[-\left(\frac{1}{2} \sigma^2 \right) (\ \mathbf{x} - \mathbf{x}_i\ ^2) \right]$	σ – wspólne dla wszystkich jąder
Sigmoidalne	$K(\mathbf{x}, \mathbf{x}_i) = \tanh(\beta_1 \mathbf{x}^T \mathbf{x}_i + \beta_0)$	ograniczenia na β_0 i β_1

Jeżeli zastosujemy jądro liniowe zbudowana sieć jest w pełni liniowa bez warstwy ukrytej. Zastosowanie funkcji sigmoidalnej prowadzi do architektury odpowiadającej sieci neuronowej perceptronowej o jednej warstwie ukrytej. W przypadku wykorzystania funkcji gaussowskiej otrzymamy sieć o radialnych funkcjach bazowych, w której liczba funkcji bazowych i ich centra są utożsamiane z wektorami podtrzymującymi. Podobnie w sieciach sigmoidalnych, liczba neuronów w warstwie ukrytej jest określana przez liczbę wektorów podtrzymujących. Warto podkreślić, że zastąpienie wektora wejściowego \mathbf{x} funkcją wektorową $\varphi(\mathbf{x})$ oraz iloczynu skalarnego $\mathbf{w}^T \mathbf{w}_j$ przez funkcje jądra $K(\mathbf{x}, \mathbf{x}_i)$, pozwala na sformułowanie problemu uczenia identycznie jak dla sieci liniowej, czyli polega na przekształceniu problemu pierwotnego w problem dualny, rozwiązywany jako zadanie programowania matematycznego.

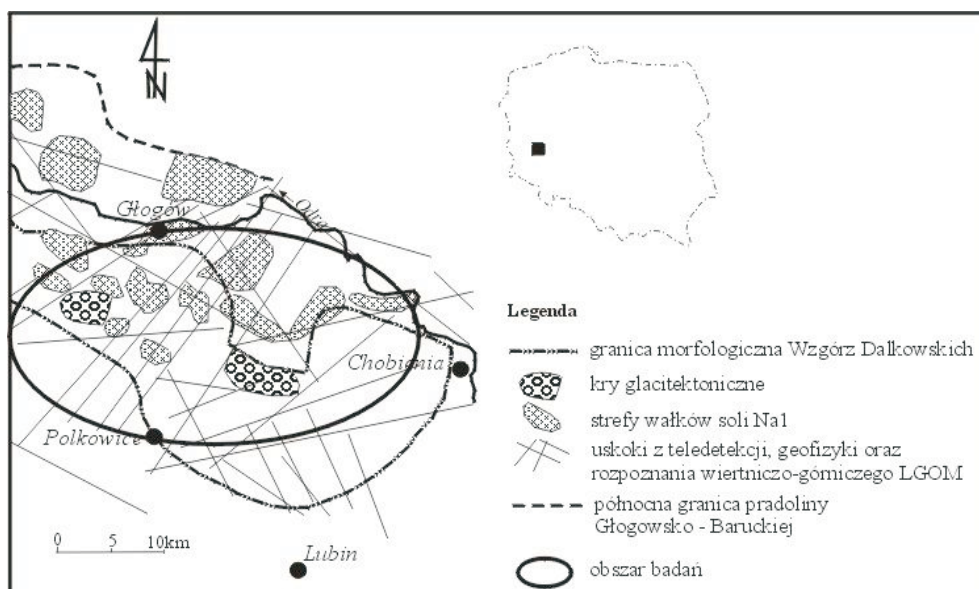
Na wynik procesu uczenia sieci nieliniowej SVM mają wpływ nie tylko dane uczące, ale również sposób zdefiniowania funkcji jąder oraz przyjęte wartości ograniczeń (wartość parametru C). Odpowiednio dobrana wartość parametru C pozwala na otrzymanie hiperpłaszczyzny w przestrzeni cech, która w przestrzeni oryginału transformuje się w krzywą, dzielącą z minimalnym błędem dane na dwie klasy. Ponieważ liczba wektorów podtrzymujących jest uzależniona od parametru C , to im większa jest jego wartość tym węższy margines separacji i mniejszy błąd klasyfikacji. Dla małej wartości parametru C klasyfikacja jest obciążona większym błędem, a sieć w swoim działaniu jest zbliżona do sieci liniowej.

3. Przykład liczbowy

Badaniami wpływu eksploatacji górniczej na powierzchnię terenu został objęty obszar Legnicko-Głogowskiego Okręgu Miedziowego (LGOM) leżący w południowej części monokliny przedsudeckiej. Na podstawie badań strukturalnych w poziomie złoża miedzi i soli cechsztyńskich oraz obserwacji prowadzonych w wyrobiskach górniczych i szymbach głębinowych kopalni KGHM Polska Miedz S.A. stwierdzono, że tutejsze masywy skalne dzielą się na trzy kompleksy zalegające na sobie dyskordantnie i są podzielone długimi łukami stratygraficznymi. Układ kompleksów zilustrowanych na rysunku 5, poczynając od najstarszego, przedstawia się jak następuje (Markiewicz 2003):

- kompleks skał krystalicznych wieku proterozoicznego oraz skał starszego paleozoiku, które stanowią podłoże monokliny,
- kompleks skał permo-mezozoicznych, z których została zbudowana monoklina,
- kompleks osadów kenozoicznych, które stanowią pokrywę monokliny.

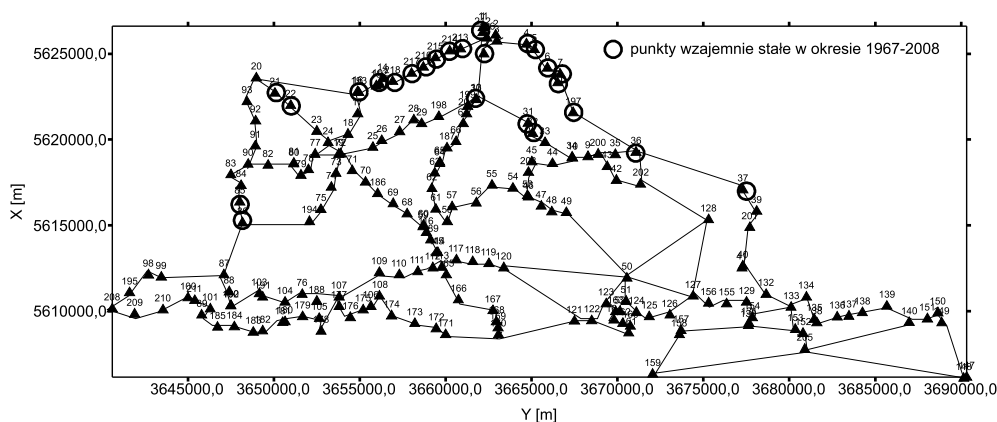
Szczegółowe informacje na temat budowy geologicznej omawianego obszaru można znaleźć w pracy (Markiewicz 2003).



Rys. 5. Szkic budowy strukturalnej podłoża badanego obszaru deformacji LGOM

Fig. 5. Sketch of the structural form of the subsoil of the area under research for the LGOMA

Przemieszczenia pionowe punktów kontrolowanych zlokalizowanych na obszarze Legnicko-Głogowskiego Okręgu Miedziowego wyznaczono na podstawie analizy wyników trzech kampanii pomiarowych przeprowadzonych w latach 1967–2008. Na omawianym obszarze o powierzchni około 75 000 ha przyjęta do opracowania sieć pomiarowo-kontrolna liczyła 218 punktów powiązanych ze sobą 302 obserwacjami (rys. 6).



Rys. 6. Schemat sieci pomiarowo-kontrolnej

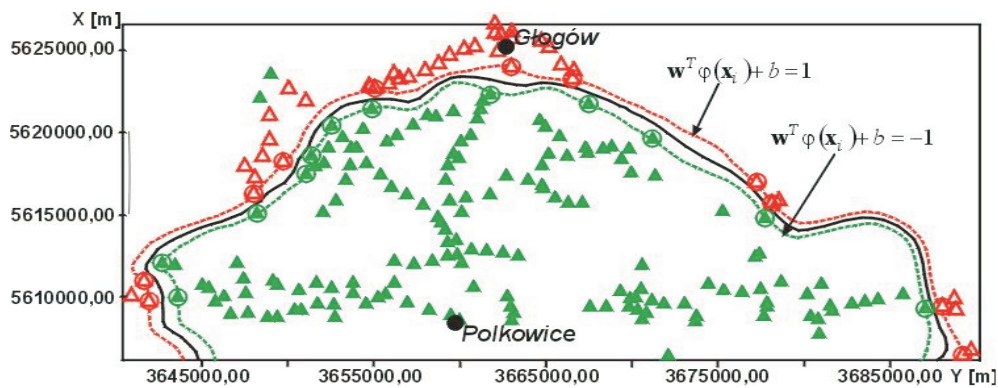
Fig. 6. Diagram of the measurement – control network

W etapie pierwszym modelowania przemieszczeń zostały wyznaczone zmiany różnic wysokości, uzyskane na podstawie pomiarów dla przedziału czasu 1967–2008. Zmiany te poddano ocenie jakościowej pod względem dokładności, której globalną miarą jest wartość sumy kwadratów poprawek podlegająca rozkładowi χ^2 . Informację jakościową uzyskano na podstawie wyrównania zmian różnic wysokości za pomocą procedury najmniejszych kwadratów przy minimalnych ograniczeniach stopni swobody z założeniem błędu średniego obserwacji $m_{obs.} = \pm 0,3$ mm.

Następnie – wykorzystując pojęcie własnego układu odniesienia – sformułowano model przemieszczeń punktów kontrolowanych dla lat 1967–2008, który stanowi podstawę do podjęcia decyzji dwuwartościowej dotyczącej klasyfikacji punktów o różnych zwrotach przemieszczeń. Definicja układu odniesienia polegała na identyfikacji wstępnej z zastosowaniem metody przylegania obiektów (Gil 1995) z uwzględnieniem najkrótszej drogi (Kuligowski 1986). W latach 1967–2008 układ odniesienia zdefiniowano na 32 punktach wzajemnie stałych. Ostateczna liczebność zbioru punktów układu odniesienia została określona na podstawie krytycznej wartości przyrostu kwadratu normy wektora poprawek (Gil 1995) i była równa 26.

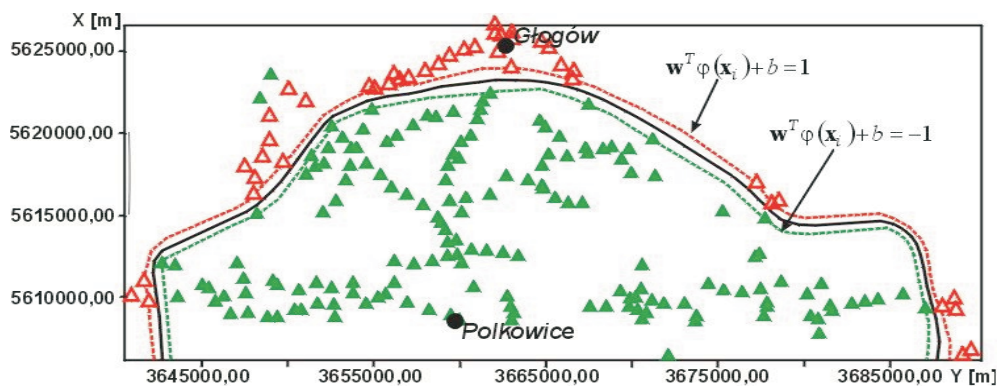
Na podstawie sformułowanego modelu przemieszczeń określono przynależność punktów kontrolowanych do jednej z dwóch klas w zależności od zwrotu przemieszczeń. Dysponując informacją na temat przynależności punktów do jednej z dwóch klas, problem klasyfikatora sprowadza się do poszukiwania krzywoliniowej granicy separacji, separującej obie klasy z minimalnym błędem i maksymalnym marginesem separacji. W tym celu wykorzystano sieć nieliniową SVM, z radialnym, liniowym oraz wielomianowym jądrem. Wynik przeprowadzonych obliczeń dla jądra radialnego (25 wektorów podtrzymujących) i liniowego (22 wektory podtrzymujące) zostały zilustrowane na rysunkach 7 oraz 8. Należy dodać, że wyniki uzyskane dla jądra wielomianowego nie różniły się w sposób widoczny od wyników uzyskanych dla jądra liniowego.

Dla sieci nieliniowej SVM o radialnych funkcjach jąder, przeanalizowano również wpływ wartości parametru C na wyniki klasyfikacji. W przypadku wartości parametru $C = 1000$,



Rys. 7. Klasyfikacja przemieszczeń z zastosowaniem sieci SVM o jądrze radialnym

Fig. 7. Classification of displacements using SVM network with radial kernel

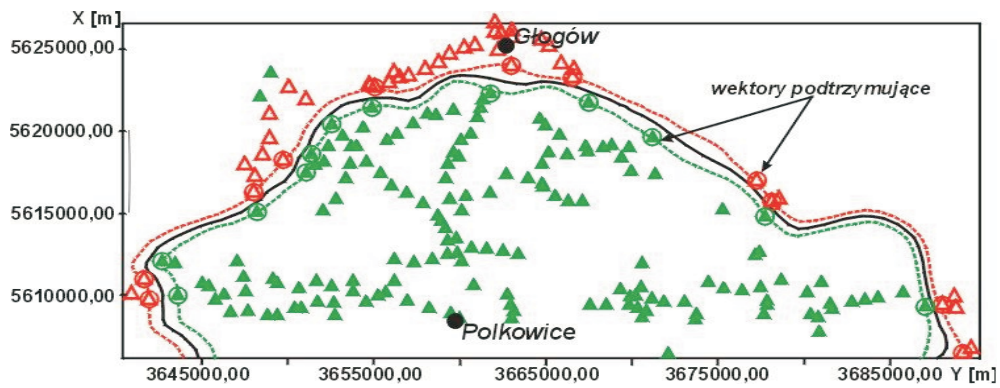


Rys. 8. Klasyfikacja przemieszczeń z zastosowaniem sieci SVM o jądrze liniowym

Fig. 8. Classification of displacements using SVM network with linear kernel

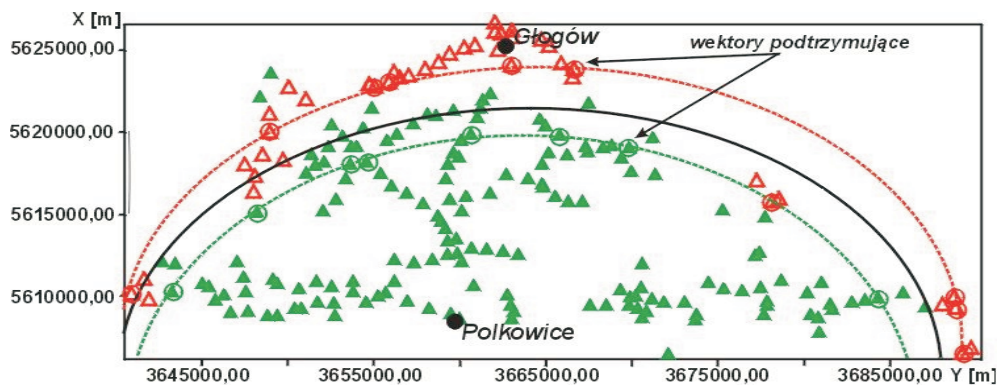
wyznaczono krzywoliniową granicę separacji, bazującą na 25 wektorach podtrzymujących (rys. 9) oraz szerokości marginesu separacji wahającej się od około 600 m do 1700 m. Jeżeli parametr przyjęto jako $C = 30$ wówczas szerokość marginesu separacji wzrosła do wartości 2000–4300 m. Jednocześnie 40 punktów, czyli około 18% wszystkich punktów sieci pomiarowo-kontrolnej, znalazło się wewnątrz marginesu separacji, a sieć w swoim działaniu dawała wyniki zbliżone do sieci liniowej (rys. 10).

Przedstawiony tok postępowania uzupełniamy geometrycznym modelem przemieszczeń (rys. 11). Przedstawiony w postaci izolunii model przemieszczeń obrazuje zmiany, jakie zaszły na badanym obszarze w latach 1967–2008, a zaznaczona linia największego spadku pokazuje kierunek, w którym następowały najszybsze zmiany w wartościach przemieszczeń. Na rysunku 11 zaznaczono również krzywoliniową granicę separacji uzyskaną z zastosowaniem sieci nieliniowej SVM z jądrem radialnym. Wartości przemieszczeń przedstawione na rysunku 11 podane są w milimetrach.



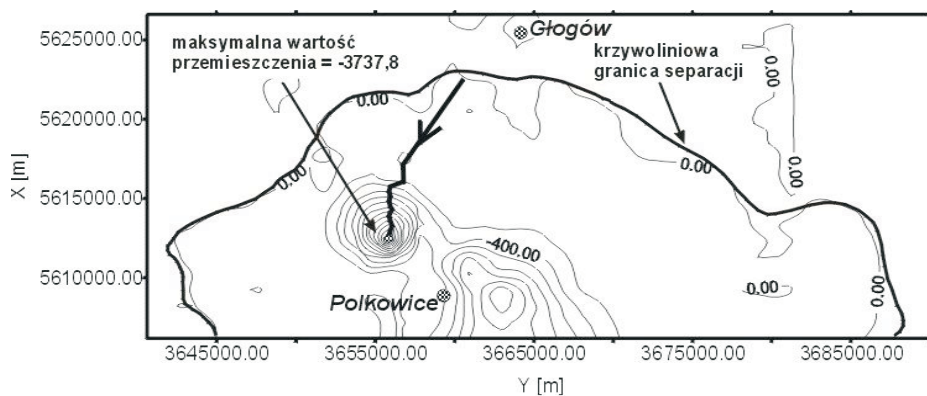
Rys. 9. Krzywoliniowa granica separacji – jądro radialna, $C = 1000$

Fig. 9. The curved boundary of separation – radial kernel, $C = 1000$



Rys. 10. Krzywoliniowa granica separacji – funkcja radialna, $C = 30$

Fig. 10. The curved boundary of separation – radial kernel, $C = 30$



Rys. 11. Geometryczny model przemieszczeń uzyskany w okresie 1967–2008

Fig. 11. Geometric model displacements obtained in the period 1967–2008

Podsumowanie

Przedstawiona w pracy sieć nieliniowa SVM została wykorzystana do klasyfikacji danych nieseparowalnych liniowo. Sieci te – podobnie jak inne sieci uczone pod nadzorem (np. sieć MLP) – pełnią rolę uniwersalnego aproksymatora. Jej zaletą jest dobra generalizacja, związana ze stosunkową małą wrażliwością na liczbę danych uczących, co ma szczególne znaczenie w rozwiązywaniu problemów w sytuacji, gdy liczba danych jest ograniczona (tak jak w przypadku pomiarów przemieszczeń i odkształceń). Na uzyskany wynik ma również wpływ przyjęta w procesie obliczeń funkcja jądra. Najkorzystniejszy wynik klasyfikacji uzyskano z zastosowaniem jądra radialnego oraz dużej wartości parametru C .

Warto podkreślić, że uzyskane wyniki klasyfikacji SVM są zgodne z wynikami uzyskanymi tradycyjną geodezyjną metodą wyznaczenia przemieszczeń pionowych, ponieważ granica separacji jest zbliżona do przebiegu izol linii przemieszczeń zerowych (oczywiście, dla odpowiednio przyjętych założeń początkowych).

Zastosowane podejście pozwala na określenie optymalnej hiperpłaszczyzny w przestrzeni cech, separującej punkty sieci geodezyjnej pomiarowo-kontrolnej na dwie klasy, co może stanowić podstawę do podjęcia decyzji dotyczącej posadowienia obiektów budowlanych na obszarach, które nie podlegają wpływom prowadzonej eksploatacji górniczej. Takie podejście jest jednak podejściem uproszczonym i może wskazywać na nieadekwatność procesu uczenia w przypadku obiektów, dla których geometryczny model przemieszczeń jest dużo bardziej zróżnicowany. W sytuacji, gdy liczba klas może być dowolna, rozwiązanie problemu z zastosowaniem techniki SVM wymaga realizacji modelu klasyfikacji wielokrotnej.

Literatura

- [1] Bishop, C.M. 2006. *Pattern Recognition and Machine Learning*. Springer.
- [2] Cover, T. 1965. Geometrical and statistical properties of system of linear inequalities with applications in pattern recognition. *IEEE Trans. Electronic Computers* vol. 14.
- [3] Gil, J. 1995. *Badanie nieliniowego geodezyjnego modelu przemieszczeń (na przykładzie obciążonego podłoża gruntowego)*. Wydawnictwo Wyższej Szkoły Inżynierskiej w Zielonej Górze, Zielona Góra.
- [4] Gunn, S.M. 1998. *Support Vector Machines for Classification or Regression. Technical Report*.
- [5] Haykin, S. 1994. *Neural networks, a comprehensive foundation*. Macmillan College Publishing Company, New York.
- [6] Jankowski, N. 2003. *Ontogeniczne sieci neuronowe. O sieciach zmieniających swoją strukturę*. Akademicka Oficyna Wydawnicza EXIT, Warszawa.
- [7] Kuligowski, J.L. 1986. *Zarys teorii grafów*. Wydawnictwo PWN, Warszawa.
- [8] Markiewicz, A. 2003. *Halotektoniczne uwarunkowania sedymentacji i deformacji osadów kenozoicznych w południowej części Monokliny Przedsudeckiej (SW Polska)*. Oficyna Wydawnicza Uniwersytetu Zielonogórskiego, Zielona Góra.
- [9] Osowski, S. 2006. *Sieci neuronowe do przetwarzania informacji*. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.
- [10] Scholkopf, B. i Smola, A. 2001. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT.
- [11] Vapnik, V. 1998. *Statistical learning theory*. Wiley, New York.
- [12] Zanni, L., Serafini, T. and Zanghirati, G. 2006. Parallel Software for Training Large Scale Support Vector Machines on Multiprocessor Systems. *Journal of Machine Learning Research* 7, 1467–1492.

