PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
Nr 47 2009
TAKSONOMIA 16
Klasyfikacja i analiza danych – teoria i zastosowania

Michał Trzęsiok
Akademia Ekonomiczna w Katowicach

PROBLEM DOBORU ZMIENNYCH DO MODELU DYSKRYMINACYJNEGO BUDOWANEGO METODĄ WEKTORÓW NOŚNYCH

1. Wstęp

Metoda wektorów nośnych (SVM – *Support Vector Machines*) należy do grupy eksploracyjnych metod statystycznej analizy wielowymiarowej. Wyniki badań empirycznych pokazują, że jest ona jedną z najdokładniejszych metod dyskryminacji [Abe 2005]. W przypadku wielu metod wykazano, że zidentyfikowanie i usunięcie zmiennych nieistotnych dla danego zadania dyskryminacji implikuje zbudowanie modelu, który daje mniejsze błędy klasyfikacji na zbiorach testowych (co oznacza, że model ma lepszą zdolność poprawnego klasyfikowania nowych obiektów). Ponadto oprócz możliwości poprawy dokładności klasyfikacji problem doboru zmiennych jest bardzo istotny również ze względu na: czas obliczeń, złożoność modelu oraz interpretowalność wyników klasyfikacji [Weston i in. 2001]. Słaba interpretowalność postaci modelu jest największą wadą metody wektorów nośnych, której działanie określa się mianem czarnej skrzynki. Wprawdzie metoda SVM identyfikuje *obserwacje* istotne dla danego zadania dyskryminacji (tzw. wektory nośne), jednak z postaci modelu nie można wprost odczytać, które *zmienne* są istotne, a które redundantne.

Aktualnie w wielu ośrodkach naukowych prowadzone są badania nad sposobami pozyskiwania wiedzy o badanym zjawisku z modelu zbudowanego metodą wektorów nośnych. Wpisując się w tę tematykę, w artykule zaproponowano modyfikację znanej metody doboru zmiennych do modelu przez stworzenie ich rankingu i eliminację zmiennych redundantnych. Zbadano również, wykorzystując empiryczne badania symulacyjne, czy usunięcie ze zbioru zmiennych objaśniających tych, które zidentyfikowano jako nieistotne, ma wpływ na poprawność klasyfikacji nowych obserwacji w modelach SVM. Informacja o tym, które ze zmiennych objaśniających mają największy wpływ na otrzymaną klasyfikację obiektów, a także które zmienne

można uznać za nieistotne w danym zadaniu dyskryminacji, jest szczególnie ważna dla decydentów i znacznie wspomaga proces podejmowania decyzji.

Zgodnie z klasyfikacją przedstawioną w pracy I. Guyon i in. [*Feature Extraction*… 2006] wyróżnić można trzy podejścia do problemu doboru zmiennych do modelu:

- filtrowanie zmiennych (*filters*) – techniki doboru zmiennych niezależne od metody klasyfikacji z wykorzystaniem różnych miar współzależności (np. współczynnika korelacji liniowej Pearsona, statystyki χ^2). Filtrowanie zmiennych odbywa się na etapie przygotowania danych przed zastosowaniem metody klasyfikacji;
- symulacyjne przeszukiwanie podzbiorów zmiennych (*wrappers*) – techniki wielokrotnie wykorzystujące metodę klasyfikacji do oceny jakości modeli budowanych na różnych zestawach zmiennych, np.: strategia wspinaczki, selekcja, eliminacja;
- metody zagnieżdżone (*embedded methods*) – kryterium doboru zmiennych jest osadzone w algorytmie metody (jest jego integralną częścią).

W dalszej części pracy wykorzystana zostanie jedna z metod symulacyjnego przeszukiwania podzbiorów zmiennych (metoda eliminacji), gdyż strategia ta pozwala na analizę porównawczą modelu z kompletem zmiennych oraz modeli ze zredukowaną ich liczbą. Ponadto podejście to jest intuicyjne i łatwo można uzasadnić końcowy wybór zestawu zmiennych i odpowiadającego mu modelu. Przed omówieniem zaproponowanej metody doboru zmiennych krótko przedstawione zostaną: idea oraz algorytm metody SVM. Bardziej szczegółowy opis metody wektorów nośnych znaleźć można w [Vapnik 1998; Cristianini, Shawe-Taylor 2000; Trzęsiok 2004].

2. Metoda wektorów nośnych – krótki opis algorytmu

W przypadku zadania dyskryminacji z dwiema klasami dany jest zbiór uczący $D = \left\{ \left(\mathbf{x^1}, y^1\right), \ldots, \left(\mathbf{x^N}, y^N\right) \right\}$, gdzie $\mathbf{x^i} \in \mathbf{R}^d$ oraz $y^i \in \{-1,1\}$ – wartości zmiennej opisującej klasę obiektu. Metoda wektorów nośnych, realizując nieliniową klasyfikację, w pierwszej kolejności transformuje obserwacje z oryginalnej przestrzeni danych w przestrzeń o dużo większym wymiarze, w której obiekty są rozdzielane hiperpłaszczyznami. Ze względu na nieliniowość przekształcenia przestrzeni danych liniowemu rozdzieleniu danych w nowej przestrzeni cech odpowiada nieliniowa ich dyskryminacja w przestrzeni pierwotnej.

Jeżeli przez φ oznaczymy nieliniową transformację przestrzeni danych, to zadanie dyskryminacji polega na wyznaczeniu optymalnej hiperpłaszczyzny:

$$\boldsymbol{\beta} \cdot \varphi(\mathbf{x}) + \beta_0 = 0, \tag{1}$$

rozdzielającej klasy zbioru uczącego $\left\{ \left(\varphi(\mathbf{x^1}), y^1\right), \ldots, \left(\varphi(\mathbf{x^N}), y^N\right) \right\}$, gdzie $\varphi(\mathbf{x^i}) \in \mathbf{Z}$ oraz $y^i \in \{-1,1\}$ dla $i = 1, \ldots, N$. W przypadku, gdy w nowej przestrzeni

cech obrazy obserwacji są liniowo separowane, zadany problem nie ma jednoznacznego rozwiązania – istnieje nieskończenie wiele hiperpłaszczyzn rozdzielających klasy. W celu zapewnienia jak największej zdolności modelu do poprawnego klasyfikowania nowych obiektów, poszukuje się hiperpłaszczyzny optymalnej, tj. położonej jak najdalej obserwacji z obu klas. Jak pokazano m.in. w [Cristianini, Shawe-Taylor 2000], rozważane zagadnienie można zapisać w postaci zadania optymalizacyjnego:

$$\begin{cases} \min\limits_{\boldsymbol{\beta},\beta_0} \frac{1}{2}\left\|\boldsymbol{\beta}\right\|^2 + C\sum\limits_{i=1}^{N}\xi_i, \\ \xi_i \geq 0, \quad y^i\left(\boldsymbol{\beta}\cdot\varphi(\mathbf{x}^{\mathbf{i}})+\beta_0\right)\geq 1-\xi_i, \quad i=1,\ldots,N. \end{cases} \tag{2}$$

Nierównościowe ograniczenia w (2) postaci $y^i\left(\boldsymbol{\beta}\cdot\varphi(\mathbf{x}^{\mathbf{i}})+\beta_0\right)\geq 1$ z geometrycznego punktu widzenia stanowią warunki separowalności, tzn. wymuszają, aby wyznaczona hiperpłaszczyzna rozdzielała klasy. Wprowadzone zmienne $\xi_1,\ldots,\xi_N \geq 0$ są realizacją postulatu uelastycznienia metody, gdyż osłabiają wymaganie, aby wszystkie obserwacje ze zbioru uczącego były poprawnie klasyfikowane (rozdzielone) przez hiperpłaszczyznę.

Zadanie (2) rozwiązać można metodą mnożników Lagrange'a. Funkcję dyskryminującą otrzymujemy, wykorzystując formułę definiującą optymalną hiperpłaszczyznę rozdzielającą klasy

$$f(\mathbf{x}) = \operatorname{sign}\left[\sum_{i\in I_{SV}} \alpha_i y^i \varphi(\mathbf{x}^{\mathbf{i}})\cdot\varphi(\mathbf{x}) + \hat{\beta}_0\right], \tag{3}$$

której postać zależy wyłącznie od tych wektorów $\mathbf{x}^{\mathbf{i}}$ ze zbioru uczącego, którym odpowiadają niezerowe współczynniki Lagrange'a w rozwiązaniu zadania optymalizacyjnego (2) (zbiór indeksów tych wektorów został oznaczony we wzorze (3) symbolem I_{SV}). Obserwacje te nazywamy *wektorami nośnymi*. Ponadto w metodzie wektorów nośnych wykorzystuje się funkcje z rodziny funkcji jądrowych, definiujące iloczyn skalarny w pewnej przestrzeni cech. Tym samym postać funkcji transformującej φ nie musi być znana. Wystarczy bowiem postać iloczynu skalarnego $\mathrm{K}(\mathbf{u},\mathbf{v}) = \varphi(\mathbf{u})\cdot\varphi(\mathbf{v})$ w przestrzeni $\mathbf{Z}$. W metodzie wektorów nośnych najczęściej wykorzystuje się funkcje jądrowe Gaussa, wielomianowe, sinusoidalne lub liniowe (zob. [Trzęsiok 2004]).

W przypadku większej liczby klas można wyznaczyć wiele funkcji (modeli) dyskryminujących klasy parami, a wskazanie przynależności danej obserwacji do klasy ustalać, wykorzystując regułę majoryzacji (głosowania modeli cząstkowych).

3. Metoda doboru zmiennych do modelu SVM

Najprostsza z metod symulacyjnego przeszukiwania podzbiorów zmiennych – przeszukiwanie wyczerpujące (budowanie i porównywanie zdolności predykcyjnych wielu modeli zbudowanych na wszystkich możliwych podzbiorach zbioru zmiennych) – gwarantuje znalezienie rozwiązania optymalnego globalnie (najlepszego zestawu zmiennych), lecz ze względu na wykładniczą złożoność obliczeniową takiego algorytmu, związaną z koniecznością zbudowania 2^d wszystkich podzbiorów zbioru zmiennych objaśniających, metoda ta nie może być efektywnie stosowana.

W dalszej części zaproponowana zostanie dwuetapowa metoda doboru zmiennych do modelu SVM. W pierwszym etapie zbudowany zostanie ranking zmiennych objaśniających bazujący na metodzie eliminacji zmiennych, w drugim zaś zidentyfikowane zostaną zmienne nieistotne.

W metodzie eliminacji zmiennych punktem wyjścia jest zestaw zmiennych, z którego iteracyjnie usuwane zostają zmienne objaśniające – po jednej w każdym kroku iteracji aż do momentu, gdy zbiór zmiennych jest pusty. Otrzymujemy rozwiązanie optymalne lokalnie – unikamy jednak przeszukiwania wszystkich możliwych podzbiorów zmiennych. W każdym kroku usuwana jest zmienna, która w najmniejszym stopniu zmienia wartość ustalonego wcześniej kryterium. W literaturze (zob. [*Feature Extraction…* 2006; Rakotomamonjy 2003]) najczęściej jako kryterium wykorzystuje się minimalny błąd klasyfikacji modelu uwzględniający wszystkie możliwe obserwacje, również nowe, których przynależność do klas nie jest znana. Kryterium to ma jednak jedynie charakter teoretyczny, gdyż jego wartość jest nieznana. Jest ona estymowana np. z wykorzystaniem metody sprawdzania krzyżowego (CV – *Cross–Validation*). Otrzymaną metodą sprawdzania krzyżowego wartość błędu oznaczać będziemy przez *CVerr*.

Wykorzystanie błędu *CVerr* oznacza jednak konieczność budowania dodatkowo wielu modeli na poszczególnych częściach zbioru uczącego z wyłączeniem jednej części, co implikuje dodatkowy wzrost czasu obliczeń. Można tego uniknąć, wprowadzając inne kryterium wyboru zmiennej usuwanej z zestawu zmiennych w *i*-tym kroku. Klasyfikację obserwacji ze zbioru uczącego (rozumianą jako wskazania przynależności obiektów do klas), otrzymaną na podstawie modelu zbudowanego na pełnym zestawie zmiennych, można potraktować jako wzorzec, z którym porównywane będą dalej wszystkie inne klasyfikacje dla modeli ze zredukowaną liczbą zmiennych. Gdy ustalony zostanie model wzorcowy, budowanych będzie wiele modeli, za każdym razem z wyłączeniem ze zbioru zmiennych jednej. Porównujemy zgodność klasyfikacji każdego modelu ze zredukowaną liczbą zmiennych z modelem wzorcowym. Za miarę zgodności klasyfikacji można przyjąć np. miarę Randa. Maksymalna wartość tej miary wskazuje model (ze zredukowanym zestawem zmiennych), którego klasyfikacja w najmniejszym stopniu różni się od modelu z kompletem zmiennych. Tym samym wskazuje, która z tymczasowo wy-

łączanych zmiennych powinna zostać usunięta w *i*-tym kroku procedury. Algorytm procedury przedstawiono w tab. 1.

Tabela 1. Algorytm budowy rankingu zmiennych objaśniających w modelach SVM wykorzystujący metodę eliminacji zmiennych z maksymalną wartością indeksu Randa jako kryterium eliminacji

Krok 1	Zbuduj model SVM na zbiorze uczącym *D*, wykorzystując pełen zestaw zmiennych i dobierając optymalnie wartości parametrów metody (otrzymana klasyfikacja będzie stanowić wzorzec). Utwórz pomocniczy zbiór uczący *S* będący kopią zbioru *D*.
Krok 2	Wygeneruj wiele zmodyfikowanych zbiorów uczących na bazie *S*, wyłączając za każdym razem jedną zmienną z pierwotnego zbioru *S*. Następnie na tych zbiorach zbuduj modele SVM z parametrami identycznymi jak w kroku 1.
Krok 3	Porównaj zgodność klasyfikacji modeli otrzymanych w kroku 2 z modelem wzorcowym z kroku 1, wykorzystując indeks Randa.
Krok 4	Zidentyfikuj model z wyłączoną zmienną, dla którego wartość indeksu Randa jest największa. Usuń ze zbioru *S* tę zmienną.
Krok 5	Powróć do kroku 2. Powtarzaj procedurę dopóki w *S* pozostanie jedna zmienna.

Źródło: opracowanie własne.

Rezultatem powyższej procedury jest ranking zmiennych. Zmienna, która pozostaje na końcu procedury iteracyjnej, ma największą moc dyskryminacyjną.

W drugim etapie na podstawie rankingu identyfikowany jest zbiór zmiennych nieistotnych. Wiadomo, że najmniej istotna zmienna została usunięta w pierwszym kroku budowy rankingu. Żeby jednak określić liczbę zmiennych w rankingu, które można uznać za redundantne, należy uzupełnić procedurę iteracyjną o miarę zdolności predykcyjnych modelu. W tym celu krok 4 algorytmu z tab. 1 należy zmodyfikować, wprowadzając po usunięciu zmiennej dodatkowe polecenie: „oblicz błąd klasyfikacji modelu *CVerr* metodą sprawdzania krzyżowego". Zauważyć należy, że koszt obliczeń jest wciąż niższy niż w przypadku kryterium eliminacji „min(*CVerr*)", gdyż w jednej iteracji algorytmu błąd ten obliczany jest tylko raz, kiedy zidentyfikowana i usunięta została kolejna zmienna.

Mając dane wartości błędu *CVerr* dla każdego modelu z zestawem zmiennych o liczebności od 1 do *d*, można wybrać model z najmniejszą wartością *CVerr*, a zmienne w nim nieuwzględnione uznać za nieistotne. Jednakże pomiar *CVerr* – błędu klasyfikacji metodą sprawdzania krzyżowego – jako że jest wartością uśrednioną, jest obarczony błędem estymacji (tzw. błędem standardowym pomiaru) $SE = \frac{s}{\sqrt{m}}$, gdzie *m* – liczba części, na które był dzielony zbiór uczący przy stosowaniu metody sprawdzania krzyżowego, *s* – odchylenie standardowe błędu klasyfikacji obliczanego dla różnych części walidacyjnych wyróżnionych ze zbioru uczącego.

Uwzględniając, że błąd *CVerr* obarczony jest błędem *SE*, wybierany jest nie ten model, któremu odpowiada najmniejszy błąd klasyfikacji, lecz model z najmniejszą liczbą zmiennych, którego błąd jest nie większy niż $\min(CVerr) + SE$,

czyli minimalny błąd klasyfikacji powiększony o błąd standardowy pomiaru (zob. [Hastie, Tibshirani, Friedman 2001]).

4. Przykłady ilustrujące procedurę doboru zmiennych do modelu SVM

Przedstawiona metoda doboru zmiennych przedstawiona oraz zweryfikowana zostanie za pomocą symulacji komputerowych na zbiorach danych zaprojektowanych do badania własności metod wielowymiarowej analizy statystycznej. Są to zbiory danych rzeczywistych, zaczerpnięte z bazy zlokalizowanej w Uniwersytecie Kalifornijskim[1], zawarte w bibliotece `mlbench` programu statystycznego **R**. Zbiory `Vehicle`, `Sonar`, `Satellite` i `Glass` zawierają odpowiednio: 846, 208, 6435 i 214 obserwacji charakteryzowanych 18, 60, 36, 9 zmiennymi objaśniającymi z wyróżnionymi 4, 2, 6 i 7 klasami. Wszystkie badania empiryczne przeprowadzone zostały z wykorzystaniem programu statystycznego **R** oraz autorskich procedur napisanych w języku programu **R**.

Wyniki pierwszego etapu przedstawionej metody doboru zmiennych dla zbioru `Vehicle` przedstawiono w tab. 2.

Tabela 2. Wynik działania procedury doboru zmiennych na zbiorze `Vehicle`

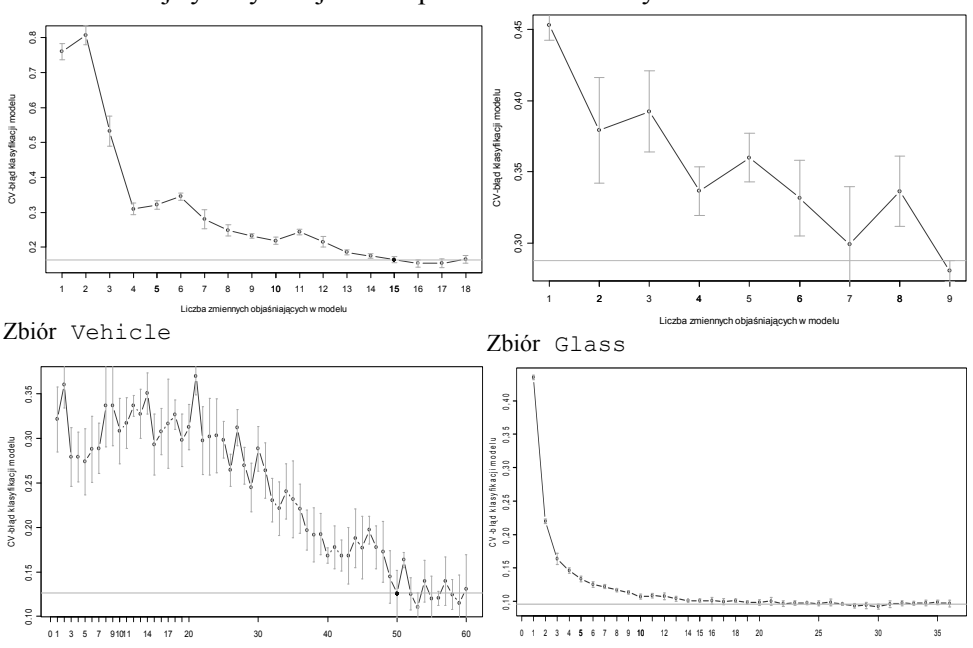
Numer iteracji	Usunięta zmienna	Indeks Randa	Błąd klasyfikacji *CVerr*	Błąd standardowy pomiaru *SE*
		1,0000	0,1655	0,0111
1	7	1,0000	0,1537	0,0125
2	11	0,9988	**0,1536**	**0,0105**
3	5	0,9965	0,1643	0,0086
4	9	0,9942	0,1749	0,0069
5	12	0,9930	0,1856	0,0070
6	18	0,9917	0,2151	0,0154
7	2	0,9930	0,2435	0,0082
8	4	0,9847	0,2187	0,0098
9	6	0,9767	0,2329	0,0063
10	3	0,8990	0,2482	0,0166
11	16	0,8569	0,2801	0,0266
12	15	0,8535	0,3451	0,0104
13	13	0,8449	0,3215	0,0122
14	1	0,8342	0,3098	0,0171
15	10	0,6962	0,5319	0,0430
16	8	0,6249	0,8073	0,0276
17	17	0,5596	0,7600	0,0227
18	14			

Źródło: opracowanie własne.

[1] Dostępne przez: ftp://ftp.ics.uci.edu/pub/machine-learning-databases.

W tabeli 2 widać, że w pierwszym kroku procedury eliminacji została usunięta zmienna nr 7, gdyż jej wykluczenie w ogóle nie zmieniło klasyfikacji obiektów ze zbioru uczącego (indeks Randa równy 1). Z kolei najmniejszy błąd klasyfikacji liczony metodą sprawdzania krzyżowego z podziałem zbioru na 5 części otrzymano dla modelu z usuniętą zmienną nr 7 i nr 11. Pomiar ten $CVerr = 0{,}1536$ jest jednak obarczony błędem standardowym równym $SE = 0{,}0109$. Po dodaniu tych wartości otrzymujemy 0,1645, co oznacza, że za najlepszy należy przyjąć model otrzymany po trzecim kroku procedury (z trzema usuniętymi zmiennymi [redundantnymi] o numerach: 7, 11 i 5). Ponadto należy zwrócić uwagę, iż w przypadku tego modelu, oprócz tego, że charakteryzuje się on mniejszą złożonością niż model z kompletem zmiennych, dodatkowo jego błąd klasyfikacji jest mniejszy niż błąd modelu zbudowanego na pełnym zestawie zmiennych. Czytając drugą kolumnę tab. 2 od dołu w górę, uzyskuje się ranking zmiennych (od najbardziej istotnej zmiennej nr 14 do najmniej istotnej – nr 7).

Wyniki procedury zarówno dla zbioru `Vehicle`, jak i dla pozostałych zbiorów w bardziej syntetycznej formie przedstawiono na rys 1.



Zbiór `Vehicle` Zbiór `Glass`

Zbiór `Sonar` Zbiór `Satellite`

Uwaga: na osi rzędnych zaznaczono błędy klasyfikacji *CVerr* (wraz z odpowiadającymi im wartościami błędów standardowych pomiaru – pionowe odcinki). Na osi odciętych zaznaczono liczbę zmiennych objaśniających w modelu. Wartość kryterium wyboru modelu: $\min(CVerr) + SE$, została zaznaczona poziomą linią.

Rys. 1. Wynik działania procedury doboru zmiennych do modelu SVM na zbiorach: `Vehicle`, `Glass`, `Sonar` i `Satellite`

Źródło: opracowanie własne.

W przypadku zbioru `Glass` usunięcie którejkolwiek zmiennej powoduje istotne zwiększenie błędu klasyfikacji, więc pozostawić należy model z kompletem zmiennych. W zbiorze `Sonar` najmniejszy błąd klasyfikacji *CVerr* uzyskano dla modelu, w którym wyeliminowano 7 zmiennych objaśniających, natomiast uwzględnienie błędu standardowego pomiaru pozwala na dodatkowe usunięcie jeszcze 3. Podobnie jak w przypadku zbioru `Vehicle` model ze zredukowaną liczbą 50 zmiennych (otrzymany w wyniku zastosowania procedury) charakteryzuje się mniejszym błędem niż model zbudowany na pełnym zestawie 60 zmiennych objaśniających. W zbiorze `Satellite` najmniejszy błąd klasyfikacji *CVerr* uzyskano dla modelu, w którym wyeliminowano 6 z 36 zmiennych, natomiast uwzględnienie błędu standardowego pomiaru wskazuje na model zbudowany na 27 zmiennych. Znów podobnie jak w przypadku zbioru `Vehicle` i `Sonar` model ze zredukowaną liczbą zmiennych charakteryzuje się mniejszym błędem klasyfikacji niż model zbudowany na pełnym zestawie zmiennych, choć różnica nie jest w tym przypadku znaczna.

5. Podsumowanie

Zaproponowano metodę doboru zmiennych objaśniających do modeli SVM. Jako wynik opisanej procedury otrzymywany jest łatwy w interpretacji ranking zmiennych. Metoda oprócz rankingu pozwala na niearbitralny podział zmiennych na redundantne i istotne. Wykorzystanie indeksu Randa jako kryterium eliminacji zmiennych zamiast błędu liczonego metodą sprawdzania krzyżowego znacznie zmniejsza czas realizacji procedury. Przedstawiona metoda pozwala na pozyskanie dodatkowej, ważnej wiedzy o badanym zjawisku z modeli SVM, które na ogół bez użycia dodatkowych narzędzi trudno jest interpretować. Usunięcie zmiennych nieistotnych zmniejsza złożoność modelu SVM oraz może wiązać się z polepszeniem zdolności predykcyjnych modelu.

Literatura

Abe S. (2005), *Support vector machines for pattern classification (advances in pattern recognition)*, Springer.

Cristianini N., Shawe-Taylor J. (2000), *An introduction to support vector machines (and other kernel–based learning methods)*, Cambridge University Press, Cambridge.

Feature extraction, foundations and applications (2006), red. I. Guyon, S. Gunn, M. Nikravesh, L. Zadeh, Springer.

Hastie T., Tibshirani R., Friedman J. (2001), *The elements of statistical learning*, Springer Verlag, N.Y.

Rakotomamonjy A. (2003), *Variable selection using SVM–based criteria*, „Journal of Machine Learning Research" no 3, s. 1357-1370.

Trzęsiok M. (2004), *Analiza wybranych własności metody dyskryminacji wykorzystującej wektory nośne*, [w:] *Postępy ekonometrii*, red. A.S. Barczak, AE, Katowice.

Vapnik V. (1998), *Statistical learning theory*, John Wiley & Sons, N.Y.

Weston J., Mukherjee S., Chapelle O., Pontil M., Piaggio T., Vapnik V. (2001), *Feature selection for SVMs*, „Advances in Neural Information Processing Systems", 13, MIT Press, s. 668-681.

VARIABLE SELECTION IN SUPPORT VECTOR MACHINES

Summary

Support Vector Machines (SVM) belong to the group of Data Mining methods and are considered as a black box method. Some authors suggest that variable selection is usually not necessary for SVMs, i.e. building the model on a set of variables including some (but not too many) redundant variables does not change the generalization ability. Once the model is built, it is still valuable to recognize the relative importance of predictor variables. The paper presents the simple modification of the backward elimination technique for feature selection and empirically shows that deleting the redundant variables can improve the classification accuracy and reduce the complexity of SVM models.