

Michał Trzęsiok

Uniwersytet Ekonomiczny w Katowicach
e-mail: michal.trzesiok@ue.katowice.pl

WZMACNIANIE ZDOLNOŚCI PREDYKCYJNYCH MODELI DYSKRYMINACYJNYCH PRZEZ WYODRĘBNIANIE ZMIENNYCH OBJAŚNIAJĄCYCH Z SIECI NEURONOWYCH

IMPROVING CLASSIFICATION ACCURACY THROUGH NEURAL NETWORKS' FEATURE EXTRACTION

DOI: 10.15611/pn.2018.508.23
JEL Classification: C380, C300, C150.

Streszczenie: Wybór metody jest elementem decydującym o pomyślności procesu modelowania, choć jakość i dobór informacji wykorzystanych przy budowie modelu automatycznego uczenia się wydają się co najmniej tak samo ważne. Mimo zautomatyzowanego mechanizmu uczenia nie wystarczy do zbioru uczącego wrzucenie wszystkich danych, jakimi dysponujemy. Konieczne jest dostarczenie informacji istotnych. Jedną z możliwości jest dobór zmiennych do modelu. Inną jest ich przekształcanie. W artykule przedstawiono procedurę łączącą te dwa podejścia – wyodrębnianie zmiennych z wielowarstwowych sieci neuronowych jako metodę doboru zmiennych do modeli budowanych innymi metodami wielowymiarowej analizy statystycznej. Celem artykułu jest zbadanie, jak takie podejście wpływa na zdolności predykcyjne modeli. Pokazano, że technikę tę należy traktować jako jedną z metod wstępnego przetwarzania danych, którą warto wypróbować, bo może prowadzić do polepszenia zdolności predykcyjnych modelu końcowego, choć tego nie gwarantuje.

Słowa kluczowe: eksploracyjne metody dyskryminacji, modele hybrydowe, dobór zmiennych.

Summary: Combining different methods has proven to be a good idea that yields models with better prediction abilities. By deleting the output layer we can use neural networks as a method for feature extraction designed to work well for classification problems. By doing this we obtain dimensionality reduction similar to PCA, but with the new features being built with the specific purpose – for classification task. We can expect this technique to generate features with high discrimination power. The main goal of the research is to analyze whether this neural networks' feature extraction method can significantly improve the classification accuracy. The results show that it does the job sometimes, but comes with no warranty. Still it can be treated as an interesting, nonlinear alternative to PCA and a valuable data preprocessing (dimensionality reduction) technique.

Keywords: classification methods, hybrid models, variable extraction.

1. Wstęp

Statystyczne metody automatycznego uczenia się to narzędzia bardzo elastyczne, dokładne, o dużych zdolnościach predykcyjnych. W przypadku metod dyskryminacji w wielu zagadnieniach empirycznych możliwy jest do osiągnięcia jednocyfrowy błąd klasyfikacji na zbiorach testowych [Meyer i in. 2003; Trzęsiok 2006]. Można więc stwierdzić, że metody automatycznego uczenia się osiągnęły wysoki poziom wyrafinowania i zdolności do odkrywania reguł i relacji w danych (zakodowanych w zbiorze uczącym). Oczywiście nie oznacza to, że należy zaniechać poszukiwania nowych, doskonalszych metod dyskryminacji. Wiadomo jednak z udowodnionych twierdzeń (znanych pod nazwą *No Free Lunch Theorem* [Wolpert, Macready 1997]), że niecelowe jest poszukiwanie najlepszej dla wszystkich problemów metody klasyfikacji, gdyż takowa metoda nie istnieje. Zawsze więc wybór metody musi być kontekstowy i dotyczyć konkretnego zadania klasyfikacyjnego. Bardzo często lepsze pod względem zdolności predykcyjnych modele otrzymuje się z łączenia wielu modeli składowych. W ten sposób powstała cała grupa metod zagregowanych, łączących wyniki predykcji modeli tych samych typów, np. metody łączenia drzew klasyfikacyjnych: *Bagging*, *Boosting*, *Random Forest* [Breiman 2001; Freund, Schapire 1997; Gatnar 2008; Rozmus 2008]. Innym rozwiązaniem jest wykorzystanie w procesie modelowania elementów co najmniej dwóch różnych metod, uzyskując modele *hybrydowe*. Skonstruowana w ten sposób metoda może łączyć pożądane własności metod agregowanych.

W niniejszym opracowaniu kluczową rolę odgrywają sieci neuronowe, przy czym ze względu na specyficzny charakter zastosowania tych sieci jako elementu modelu hybrydowego, z bardzo różnorodnej rodziny sieci neuronowych o różnej architekturze, wybrane zostały wyłącznie sieci jednokierunkowe, wielowarstwowe. Sieci neuronowe mają wbudowany bardzo dobry mechanizm nieliniowego przekształcania zmiennych objaśniających (w warstwach ukrytych), który generuje nowy zestaw zmiennych, wykorzystywany później do końcowego określania przynależności obserwacji do klas [Krzyśko i in. 2008]. Zestaw tych nowych zmiennych dobierany jest w procesie uczenia się sieci w sposób adaptacyjny tak, żeby zmienne w ostatniej warstwie miały jak największą moc dyskryminacyjną. Ten ostatni etap klasyfikacji obiektów w sieciach neuronowych nie jest jednak wyjątkowo swoisty dla zadania dyskryminacji, szczególnie w przypadku wykorzystania ciągłej funkcji aktywacji w warstwie wyjściowej. Można jednak ten ostatni etap zastąpić np. mechanizmem zaproponowanym przez V. Vapnika w metodzie wektorów nośnych (*support vector machine – SVM*) [Vapnik 1998; Trzęsiok 2006]. Naturalny wydaje się więc pomysł, by wykorzystać sieci neuronowe do utworzenia nowego zestawu zmiennych objaśniających, który może zostać wyodrębniony z ostatniej warstwy ukrytej sieci, i na takim nowym zestawie zmiennych zbudować np. model *SVM* [Mori i in. 2005] i sprawdzić, czy to poprawia zdolność predykcyjną modelu. Budowany w ten sposób model można określić jako hybrydowy. Podejście takie wpi-

suje się w obszar doboru zmiennych do modelu. Bardzo znanym podejściem z tego obszaru jest wykorzystanie analizy głównych składowych, która konstruuje nowy zestaw zmiennych objaśniających będących kombinacjami liniowymi zmiennych pierwotnych. Wyodrębnianie nowych zmiennych z sieci neuronowych na potrzeby dyskryminacji wydaje się mieć przewagę w kilku aspektach. Po pierwsze: zestaw zmiennych otrzymany z sieci neuronowej ma potencjalnie większe zdolności do adaptacyjnego wykorzystania informacji zawartych w zmiennych pierwotnych, gdyż tworzą go przetransformowane w sposób nieliniowy kombinacje liniowe zmiennych pierwotnych. Po drugie: analiza głównych składowych jest techniką całkowicie niezależną od zadania dyskryminacyjnego i skupia się jedynie na zbudowaniu nowego zestawu zmiennych, który w możliwie najwyższym stopniu zachowuje zmienność (wariancję) pierwotnych zmiennych, gdy tymczasem ostatnią warstwę sieci tworzą zmienne, które w największym stopniu potrafią opisywać relacje kluczowe dla realizacji zadania dyskryminacji. Tak więc ostatnia warstwa sieci to zmienne, które nie zachowują w pełni zmienności pierwotnych predyktorów, lecz „wyspecjalizowane” zmienne zawierające te informacje, które sieć zidentyfikowała jako kluczowe dla danego zadania dyskryminacji. Pod tym względem ostatnia warstwa sieci pełni podobną funkcję jak *zmienne dyskryminacyjne (kanoniczne)* zaproponowane przez Fishera [Fisher 1936 (por. [Krzyśko i in. 2008])], które również zostaną uwzględnione w empirycznej części tego opracowania. Fisherowskie zmienne dyskryminacyjne są kluczowym elementem *liniowej analizy dyskryminacyjnej (linear discriminant analysis – LDA)* i otrzymuje się je jako element składowy modelu *LDA*.

Głównym celem artykułu jest przedstawienie procedury wykorzystania sieci neuronowych jako metody ekstrakcji zmiennych objaśniających o dużej mocy dyskryminacyjnej do modelowania z wykorzystaniem innych metod klasyfikacji oraz symulacyjna ocena korzyści płynących z takiego podejścia. W symulacji wykorzystane zostaną: metoda sieci neuronowych oraz jej hybrydowe połączenie z metodą *k*-najbliższych sąsiadów (*kNN*), metodą drzew klasyfikacyjnych (*rpart*), metodą lasów losowych (*Random Forest*) i metodą wektorów nośnych (*SVM*). Ponadto przedstawiona metoda hybrydowa zostanie porównana pod względem dokładności klasyfikacji z klasycznym podejściem budowy modeli z wykorzystaniem głównych składowych, a także zmiennych dyskryminacyjnych.

2. Procedura wyodrębniania zmiennych o dużej mocy dyskryminacyjnej z wielowarstwowych sieci neuronowych

Zdolności predykcyjne modelu bardzo silnie zależą od danych użytych w procesie modelowania. Mimo zautomatyzowanego algorytmu statystycznych metod uczących się bardzo ważnym elementem jest dostarczenie danych w użytecznej postaci. Istnieje wiele metod wstępnego przetwarzania danych (*data preprocessing*). Odpowiedni dobór metod wstępnego przetwarzania danych jest swoistą sztuką i ma duże znaczenie, gdyż potrafi znacząco zmienić jakość modelu. W dalszej części przed-

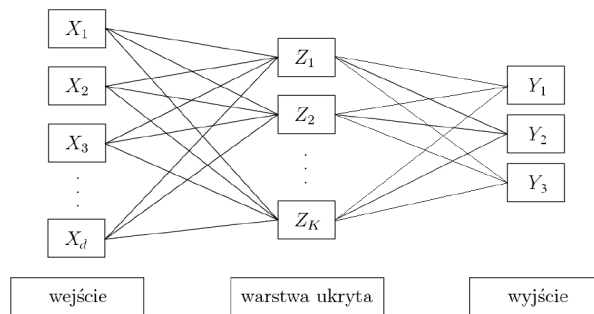
stawiona zostanie metoda wyodrębniania zmiennych z wielowarstwowych sieci neuronowych jako alternatywa do generowania nowych zestawów zmiennych objaśniających metodą głównych składowych czy też zmiennych dyskryminacyjnych otrzymywanych w wyniku zastosowania liniowej analizy dyskryminacyjnej.

Model sieci neuronowej (jeden z najprostszych wariantów sieci neuronowej) z jedną warstwą ukrytą przedstawiono na rys. 1. Pierwszą warstwę sieci tworzą pierwotne zmienne objaśniające. Kolejna warstwa sieci to zmienne syntetyczne Z_k (dla $k = 1, 2, \dots, K$), które powstają poprzez nieliniowe przekształcenie kombinacji liniowych zmiennych pierwotnych (1) [Widrow, Hoff 1960; Rosenblatt 1962; Hastie i in. 2009]. Ta warstwa ma zasadnicze znaczenie dla działania sieci.

$$Z_k = \varphi_k \left(w_{k0} + \sum_{j=1}^d w_{kj} X_j \right), \quad (1)$$

gdzie: $w_{k0}, w_{k1}, \dots, w_{kd}$ ($k = 1, 2, \dots, K$) to współczynniki kombinacji liniowych nazywane *wagami synaptycznymi*, φ_k zaś to pewne nieliniowe transformacje (tzw.

funkcje aktywacji, np. funkcja logistyczna $\varphi_k(z) = \frac{1}{1 + e^{-z}}$).



Rys. 1. Ilustracja modelu sieci neuronowej z jedną warstwą ukrytą w przypadku zadania dyskryminacji trzech klas (stąd trzy neurony w warstwie wyjściowej)

Źródło: opracowanie własne.

Obserwacje z warstwy ukrytej poddawane są w warstwie wyjściowej analogicznemu przekształceniu z wykorzystaniem jednej funkcji aktywacji φ_0 . W przypadku zagadnienia dyskryminacji P -klas modelowaniu podlega P -wariantów zmiennej zależnej Y (por. (2)).

$$Y_p = \varphi_0 \left(v_{p0} + \sum_{j=1}^K v_{pj} Z_j \right). \quad (2)$$

Przedstawioną architekturę sieci neuronowej łatwo można rozszerzyć na sieci z wieloma warstwami ukrytymi. Kolejne warstwy ukryte są tworzone analogicznie jak pierwsza warstwa ukryta – jako przekształcone przez pewne funkcje aktywacji kombinacje liniowe zmiennych z poprzedniej warstwy.

Identyfikacja najodpowiedniejszej architektury sieci neuronowej (liczby warstw ukrytych, liczby zmiennych (neuronów) w poszczególnych warstwach ukrytych oraz postaci funkcji aktywacji w poszczególnych warstwach) jest swoistą sztuką [Hastie i in. 2009, s. 397-400]. Najczęściej elementy te są ustalane przez badacza metodami symulacyjnymi.

Każda z warstw ukrytych sieci reprezentuje nowe, przekształcone zmienne, które mogą być wyodrębnione i użyte w dalszej analizie. Z punktu widzenia zadania dyskryminacyjnego największą moc wyjaśniania mają zmienne z ostatniej warstwy ukrytej (zaznaczmy, że chodzi tu nie o ostatnią warstwę sieci, czyli nie o warstwę wyjściową, lecz o przedostatnią warstwę sieci, która tworzy *ostatnią warstwę ukrytą*), gdyż te zmienne bezpośrednio wpływają na przynależność do klas wskazywaną przez model. Stanowi to bezpośrednie uzasadnienie wyboru zmiennych z ostatniej warstwy ukrytej sieci w przedstawionej w tab. 1 procedurze badawczej.

Tabela 1. Procedura wykorzystania sieci neuronowych jako metody ekstrakcji zmiennych objaśniających o dużej mocy dyskryminacyjnej do modelowania z wykorzystaniem innych metod klasyfikacji (procedura uwzględniająca etap oceny zdolności predykcyjnych modelu hybrydowego)

Krok 1.	Podziel zbiór danych na trzy części: część uczącą D , walidacyjną W oraz część testową T
Krok 2.	Znormalizuj zmienne w każdym ze zbiorów, używając parametrów obliczonych na zbiorze uczącym D
Krok 3.	Ustal pewien zbiór rozpatrywanych wariantów architektury modelu sieci neuronowych (tj. zbiór wektorów określających liczbę warstw ukrytych sieci oraz liczbę neuronów w każdej warstwie, np. wektory postaci [30, 15, 6]), zbuduj modele sieci dla każdego wariantu oraz wybierz do dalszego wykorzystania tę sieć, dla której błąd klasyfikacji obliczony na zbiorze walidacyjnym W jest najmniejszy
Krok 4.	Wyodrębnij ostatnią warstwę ukrytą z sieci neuronowej i użyj jej neuronów w dalszych krokach jako nowych zmiennych objaśniających ($NNet$ -zmiennych) w modelowaniu wybranymi metodami dyskryminacji, tj. utwórz zmodyfikowany zbiór uczący D_{NNet} w którym obserwacje ze zbioru uczącego są opisane zmiennymi powstałymi w ostatniej warstwie ukrytej w sieci neuronowej
Krok 5.	Na zbiorze uczącym D_{NNet} zbuduj model dyskryminacyjny wybraną metodą eksploracyjną (np. kNN , $rpart$, $Random\ Forests$, SVM), dobierając technikę sprawdzania krzyżowego wartości parametrów tej metody
Krok 6.	Dokonaj transformacji wartości zmiennych objaśniających obiektów ze zbioru testowego T przez zastosowanie „obciętej” sieci neuronowej z kroku 4, tj. utwórz zmodyfikowany zbiór testowy T_{NNet} , w którym obserwacje ze zbioru testowego są opisane zmiennymi powstałymi w ostatniej warstwie ukrytej w sieci neuronowej
Krok 7.	Oblicz dokładność klasyfikacji (miernik zdolności predykcyjnych rozpatrywanego modelu hybrydowego) na zmodyfikowanym zbiorze testowym T_{NNet}

Źródło: opracowanie własne.

W części empirycznej niniejszego opracowania wyniki procedury wyodrębniania zmiennych objaśniających z sieci neuronowych (budowania modeli hybrydowych) uzupełniono o porównanie błędów klasyfikacji tego podejścia z błędami klasyfikacji modeli budowanych na: zestawie zmiennych pierwotnych, zestawie głównych składowych i zestawie zmiennych dyskryminacyjnych.

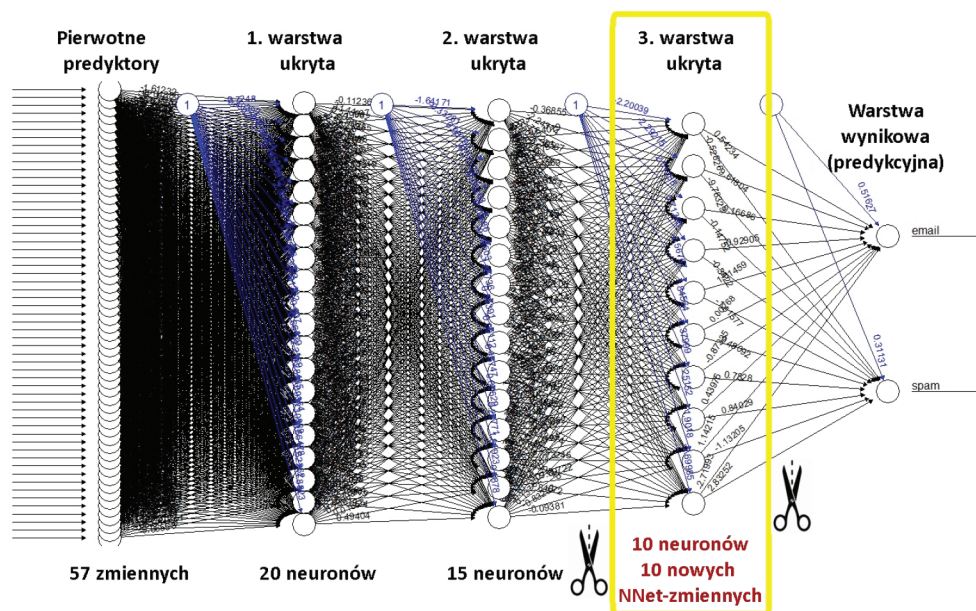
3. Przykład ilustrujący przedstawioną procedurę ekstrakcji nowych zmiennych objaśniających z sieci neuronowych

Przedstawiona procedura wykorzystania sieci neuronowych jako metody ekstrakcji zmiennych objaśniających o dużej mocy dyskryminacyjnej zilustrowana zostanie na zbiorze danych rzeczywistych spam wykorzystywanym do badania własności metod wielowymiarowej analizy statystycznej. Zbiór ten udostępniony został przez pracowników laboratoriów firmy Hewlett-Packard, zawiera 4601 obserwacji (listów elektronicznych *email*), należących do jednej z 2 klas (list „dobry” oraz list niechciany, czyli „spam”). Zadanie polega na zaklasyfikowaniu nowo przychodzącego listu do jednej z tych grup. Obserwacje w zbiorze spam charakteryzowane są przez 57 zmiennych objaśniających (ilorazowych). Niemal wszystkie zmienne objaśniające wskazują częstość względną występowania wybranego słowa kluczowego (lub pojedynczych znaków tekstowych) w zbiorze wszystkich słów (znaków) zawartych w danej wiadomości mailowej.

Wszystkie obliczenia zostały zrealizowane z wykorzystaniem programu statystycznego **R** i jego dodatkowych pakietów, które posłużyły do napisania autorskiego kodu realizującego przedstawioną procedurę. W szczególności wykorzystano pakiet *neuralnet* umożliwiający budowę i wizualizację modeli sieci neuronowych z wieloma warstwami ukrytymi (z domyślnymi wartościami parametrów decydujących o postaci funkcji aktywacji oraz z domyślnym algorytmem wyznaczania wag synaptycznych).

W pierwszym kroku procedury podzielono zbiór danych w sposób losowy na trzy części: uczącą D , walidacyjną W i testową T w proporcji 50, 25, 25%. Dane w zbiorach znormalizowano, stosując przekształcenie ilorazowe. Na zbiorze uczącym D (z pierwotnymi zmiennymi objaśniającymi) zbudowano modele dyskryminacyjne wszystkimi rozpatrywanymi metodami (*neuralnet*, *k-NN*, *rpart*, *Random Forest*, *SVM*) i obliczono błąd klasyfikacji na zbiorze testowym na potrzeby porównań z modelem hybrydowym. Do dalszej analizy wybrano model sieci neuronowej z trzema warstwami ukrytymi z 20, 15 i 10 neuronami w kolejnych warstwach ukrytych, odpowiednio. We wszystkich warstwach wykorzystano logistyczną funkcję aktywacji. Wyodrębniono ostatnią warstwę ukrytą, której neurony posłużyły jako zmienne objaśniające w modelu hybrydowym. W ten sposób dokonano redukcji wymiaru przestrzeni danych z 57 do 10. Zbudowaną sieć neuronową przedstawia rys. 2. Ze względu na dużą liczbę zmiennych i połączeń synaptycznych, na rysunku tym nakładają się wartości obliczonych współczynników kombinacji liniowych

(wag synaptycznych). Nie jest to jednak istotny element tej wizualizacji, którą przedstawiono jedynie w celu lepszego zobrazowania procedury opisanej w tab. 1.



Rys. 2. Ilustracja zbudowanego na zbiorze spam modelu sieci neuronowej z trzema warstwami ukrytymi, gdzie trzecia z tych warstw stanowi zestaw nowych zmiennych modelu hybrydowego

Źródło: opracowanie własne.

Sieć neuronowa przedstawiona na rys. 2 miała błąd klasyfikacji obliczony na zbiorze testowym równy 8,3%. Obliczone na zbiorze testowym błędy klasyfikacji dla różnych wariantów modeli hybrydowych zbudowanych na zbiorze uczącym D_{NNet} (z *NNet*-zmiennymi objaśniającymi) zestawiono w tab. 2. Ponadto w tej tabeli dla porównania zestawiono również błędy klasyfikacji dla tych samych metod wykorzystanych do zbudowania modeli na: oryginalnym zestawie zmiennych, na zmiennych otrzymanych z analizy głównych składowych (*principal component analysis – PCA*) oraz na zmiennych dyskryminacyjnych, czyli na zmiennych otrzymanych z *liniowej analizy dyskryminacyjnej (LDA)*. Najlepsze wyniki (najmniejsze błędy klasyfikacji) uzyskane dla każdej z metod wyróżniono przez pogrubienie.

W przypadku każdego modelu zrealizowano krok 5 procedury opisanej w tab. 1, tj. symulacyjne przeszukiwanie wartości kluczowych parametrów danej metody (wykorzystując zachłanne przeszukiwanie wszystkich kombinacji wartości parametrów metody z przyjętego arbitralnie dyskretnego zbioru wartości parametrów).

Tabela 2. Błędy klasyfikacji dla zbioru *spam* dla wybranych metod wielowymiarowej analizy statystycznej, obliczone na wydzielonej części testowej w czterech wariantach: na oryginalnym zestawie zmiennych (T), na zbiorze ze zmiennymi wyodrębnionymi z ostatniej warstwy sieci neuronowej T_{NNet} , na zestawie 35 głównych składowych (T_{PCA}) oraz na zmiennych dyskryminacyjnych (T_{LDA})

Metoda	Błąd klasyfikacji na zbiorze T	Błąd klasyfikacji na zbiorze T_{NNet}	Błąd Klasyfikacji na zbiorze T_{PCA}	Błąd klasyfikacji na zbiorze T_{LDA}
Sieci neuronowe	8,4%	–	–	–
<i>k-NN</i>	11,1%	8,4%	12,4%	8,4%
<i>rpart</i>	10,3%	8,1%	9,5%	8,6%
<i>Random Forest</i>	5,3%	8,2%	6,3%	12,6%
<i>SVM</i>	8,8%	8,4%	9,7%	9%

Źródło: opracowanie własne.

Tabela zawiera również błędy klasyfikacji modeli zbudowanych na obserwacjach opisanych nie przez oryginalny zestaw zmiennych, lecz przez główne składowe. Dodać należy, że wykorzystanie analizy głównych składowych było zasadne. Wskazała na to wysoka wartość wskaźnika Kaisera-Meyera-Olkina (*KMO*), czyli unormowanego miernika adekwatności zestawu zmiennych pierwotnych do przeprowadzenia analizy głównych składowych. Obliczono, że $KMO = 0,814$ (zestaw użytych głównych składowych obejmował 35 składowych, co było liczbą większą niż wskazywały formalne kryteria oparte na wartościach własnych, lecz wiązało się z chęcią zachowania minimum 80% wariancji oryginalnego zestawu zmiennych objaśniających). Jako punkt odniesienia podać należy, że liniowa analiza dyskryminacyjna na zbiorze opisanym oryginalnym zestawem 57 zmiennych dała błąd klasyfikacji na zbiorze testowym równy 10,8%.

Przedstawiony przykład ilustracyjny okazał się ciekawy w tym sensie, że w tab. 2 zaobserwować można zarówno modele, dla których, w porównaniu z modelem zbudowanym na oryginalnym zbiorze zmiennych, wykorzystanie procedury wyodrębniania zmiennych z sieci neuronowej przyniosło poprawę zdolności predykcyjnych (wyraźną: *k-NN* i *rpart*, oraz nieznaczną – *SVM*), jak i model lasów losowych (*Random Forest*), dla którego błąd klasyfikacji się zwiększył, gdy do modelowania użyto *NNet*-zmiennych. Ogólnie najlepszym modelem wśród wszystkich budowanych wariantów okazał się model lasów losowych na oryginalnych zmiennych z błędem klasyfikacji równym 5,3%. Wykorzystanie głównych składowych lub zmiennych dyskryminacyjnych nie prowadziło do uzyskania wyników korzystniejszych niż wyodrębnianie zmiennych z sieci neuronowych, choć i tu był jeden wyjątek dla lasów losowych. Przyczyną spadku zdolności predykcyjnych modeli zagregowanych (takich jak *Random Forest*) w przypadku zastosowania *NNet*-zmiennych jest praw-

dopodobnie to, że przedstawiona procedura zwiększa stabilność wyników pojedynczych drzew, a jak wiadomo, metody zagregowane lepiej działają, gdy łączą słabe (niestabilne) klasyfikatory. Ta zależność wymaga jednak głębszego zbadania i wykracza poza ramy tego opracowania.

4. Zakończenie

W podsumowaniu przeprowadzonego badania nie można sformułować jednoznacznych wskazań i spektakularnych wniosków dotyczących korzyści płynących z wykorzystania zaprezentowanego podejścia, jednak autor nie spodziewał się, że takie wnioski można będzie przedstawić. Na pewno jednak wyodrębnianie nowych zmiennych z ostatniej warstwy ukrytej sieci neuronowych można traktować jako jedną z metod wstępnego przetwarzania danych (metodę redukcji wymiaru macierzy danych), którą warto wypróbować, bo może prowadzić do polepszenia zdolności predykcyjnych modelu końcowego, choć tego nie gwarantuje. Innymi słowy: wyodrębnianie nowych zmiennych objaśniających z ostatniej warstwy ukrytej jest techniką, która dla niektórych zbiorów danych przyniesie pożądane rezultaty w postaci lepszych zdolności predykcyjnych modelu końcowego, ale w innych przypadkach nie doprowadzi do ich poprawy. Takie stwierdzenie jest zgodne z duchem twierdzeń z rodziny *No Free Lunch*. Przedstawiona technika stanowi ciekawą alternatywę dla liniowych metod przekształcania zmiennych objaśniających, takich jak analiza głównych składowych oraz zmienne dyskryminacyjne. Wyodrębnianie *NNet*-zmiennych ma tę przewagę nad najczęściej stosowaną do transformacji zmiennych – analizą głównych składowych, że jest realizowane w kontekście rozważanego zadania dyskryminacji i optymalizuje postać nowych zmiennych objaśniających ze względu na ich moc dyskryminacyjną.

Literatura

- Breiman L., 2001, *Random Forests*, Machine Learning, 45, s. 5-32.
- Fisher R.A., 1936, *The use of multiple measurements in taxonomic problems*, Annals of Human Genetics, 7(2), s.179-188.
- Freund Y., Schapire R.E., 1997, *A decision-theoretic generalization of on-line learning and an application to boosting*, Journal of Computer and System Sciences, 55(1), s. 119-139.
- Gatnar E., 2008, *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa.
- Hastie T., Tibshirani R., Friedman J., 2009, *The Elements of Statistical Learning*, 2 wyd., Springer Verlag, New York.
- Krzyśko M., Wołyński W., Górecki T., Skorzybut M., 2008, *Systemy uczące się. Rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości*, WNT, Warszawa.
- Meyer D., Leisch F., Hornik K., 2003, *The Support Vector Machine under Test*, Neurocomputing, 55(1), s. 169-186.

- Mori K., Matsugu M., Suzuki T., 2005, *Face Recognition Using SVM Fed with Intermediate Output of CNN for Face Detection*, IAPR Conference on Machine Vision Applications, May 16-18, s. 410-413.
- Rosenblatt F., 1962, *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*, Spartan, Washington DC.
- Rozmus D., 2008, *Agregacja modeli klasyfikacyjnych i regresyjnych*, Fundacja Promocji i Akredytacji Kierunków Ekonomicznych, Warszawa.
- Trzęsiok M., 2006, *Metoda wektorów nośnych na tle innych metod wielowymiarowej analizy danych*, Taksonomia 13, *Klasyfikacja i analiza danych – teoria i zastosowania*, red. K. Jajuga, M. Walesiak, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław, s. 536-542.
- Vapnik V., 1998, *Statistical Learning Theory*, John Wiley & Sons, New York.
- Widrow B., Hoff M.E., 1960, *Adaptive switching circuits* (no. TR-1553-1). Stanford University, Stanford Electronics Labs, Stanford.
- Wolpert D., Macready W., 1997, *No Free Lunch theorems for Optimization*, IEEE Transactions on Evolutionary Computation, 1(1), s. 67-82.