

Mariola Chrzanowska

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

PORÓWNANIE I OCENA WYBRANYCH METOD KLASYFIKACJI KREDYTÓW INDYWIDUALNYCH

Streszczenie: W opracowaniu przedstawiono wyniki klasyfikacji kredytów przeprowadzonej na podstawie danych rzeczywistych uzyskanych w pewnym banku. Do badań wykorzystano modele klasyfikacyjne skonstruowane na bazie zagregowanych drzew decyzyjnych oraz sztucznych sieci neuronowych. Eksperyment przeprowadzono w pięciu seriach dla każdej z wykorzystanych metod. Dla każdej z serii wylosowano zbiór, który następnie podzielono na 10 części. Zbiór testowy stanowił kolejne 10% badanego zbioru. Celem artykułu jest porównanie efektywności wykorzystanych metod w przypadku, gdy struktura zbioru testowego jest zmienna w każdym etapie.

1. Wstęp

Odpowiednia wycena transakcji kredytowych ma istotne znaczenie dla pracy banku. Umiejętna alokacja zasobów i ryzyka (przy minimalizacji kosztów) stanowi główny wyznacznik wytwarzania przez sektor bankowy wartości dodanej. Zdolność szacowania oraz redukcji ryzyka kredytowego istotnie wpływa na pozycję rynkową banku, a w skrajnych przypadkach decyduje o jego „być albo nie być”.

Badając zdolność kredytową, banki stosują metody statystyczne. Nasuwa się tu jednak pytanie, na ile uzyskane tymi metodami rozwiązania są efektywne i jak różne czynniki wpływają na stabilność rozwiązań. Ważnym aspektem, jaki należy brać pod uwagę, jest struktura próby uczącej i testowej. Jeśli informacje zawarte w próbie uczącej znacznie różnią się od informacji w próbie testowej, istotnie wpływa to na poziom błędu klasyfikacji. Celem niniejszego opracowania jest badanie wpływu struktury próby na jakość klasyfikacji uzyskanej za pomocą zagregowanych drzew klasyfikacyjnych i sieci neuronowych¹. W trakcie poprzednich eksperymentów [Chrzanowska, Witkowska 2008; Chrzanowska, Alfaro, Witkowska

¹ Wybór metod badawczych nie jest przypadkowy. Badania literaturowe potwierdziły, że drzewa klasyfikacyjne i sieci neuronowe są (oprócz regresji logistycznej i liniowej oraz analizy dyskryminacyjnej) najczęściej stosowanymi metodami scoringowymi [Crook, Edelman, Thomas 2007].

2009] struktura zbioru testowego była stała. Prezentowany eksperyment miał na celu sprawdzenie, która z użytych metod jest najlepsza do klasyfikacji kredytobiorców w przypadku zmiennej struktury próby testowej.

2. Opis danych

Badania przeprowadzono na podstawie rzeczywistej bazy danych o kredytach hipotecznych udzielonych w latach 1999-2004, z której wylosowano 740-elementową próbę symetryczną. Każdy obiekt w zbiorze był opisany za pomocą zmiennych²:

- y – sytuacja (0 – kredyt stracony; 1 – kredyt spłacany regularnie),
- x_5 – miejsce udzielenia kredytu (M1; M2; M3),
- x_7 – czas spłaty kredytu (w latach),
- x_8 – wartość kredytu (w zł),
- x_{10} – stopa kredytu opisana wzorem:

$$x_{10} = \frac{\text{bieżąca stopa procentowa}}{\text{poprzednia stopa procentowa}} \cdot 100,$$

- x_{11} – spłata kredytu:

$$x_{11} = \frac{\text{aktualnie spłacona kwota}}{\text{kwota kredytu}} \cdot 100.$$

3. Opis badania

Eksperyment przeprowadzono w 5 seriach dla każdej z wykorzystanych metod. W pierwszym etapie badania dla każdej serii (ze zbioru wyjściowego) wylosowano ze zwracaniem 740-elementowy zbiór. Każdy zbiór podzielony został na 10 części. W ten sposób dla każdej serii skonstruowano dziesięć prób, których strukturę zapisano w tab. 1.

Tabela 1. Sposób podziału zbioru w i -tej serii

Seria i	Próba iI		Próba iII		Próba iX	
	część testowa	część ucząca	część testowa	część ucząca	część testowa	część ucząca
Numer obserwacji	1-74	75- 740	75-148	1-74 149-740	667-740	1-665

Źródło: opracowanie własne.

² Poprzednie eksperymenty [Chrzanowska, Witkowska 2007] wykazały, że tylko te zmienne mają zdolność dyskryminacji.

Dla tak skonstruowanych 50 zbiorów dokonano klasyfikacji obiektów za pomocą:

- zagregowanego drzewa algorytmem boosting,
- zagregowanego drzewa algorytmem bagging,
- zagregowanego drzewa algorytmem random forest³,
- sztucznej sieci neuronowej⁴ (perceptron wielowarstwowy o strukturze 5-3-1).

W drugim etapie badań przeprowadzono statystyczną analizę otrzymanych wyników. Ostatni etap eksperymentu to porównanie efektywności klasyfikacji przeprowadzonej za pomocą poszczególnych metod przy użyciu testów Wilcoxon'a oraz znaków [Demsar 2006].

Podczas eksperymentu efektywność klasyfikacji mierzono ogólnym błędem klasyfikacji E :

$$E = 1 - \frac{\text{liczba poprawnie rozpoznanych obiektów w każdej klasie}}{\text{liczba wszystkich obiektów}} \quad (1)$$

4. Wyniki badań

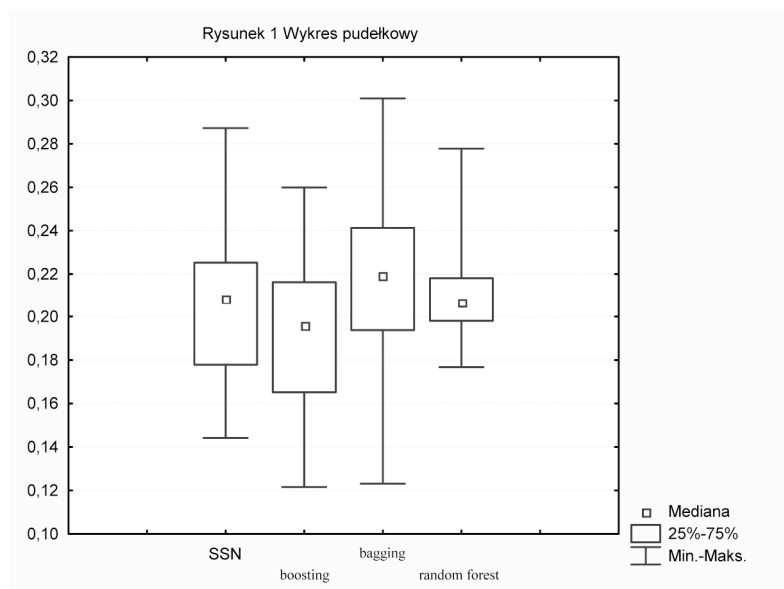
Dla każdej z prób za pomocą algorytmów: bagging, boosting oraz random forest zbudowano 100 pojedynczych drzew klasyfikacyjnych. Następnie wykorzystując „głosowanie większościowe”, każdy z pojedynczych kredytów przypisany został do jednej z grup: kredytów spłaconych lub kredytów straconych. Na tej podstawie dla każdego zbioru testowego obliczono ogólny błąd klasyfikacji.

Wykorzystana do eksperymentu sztuczna sieć neuronowa (SSN) zawierała jedną warstwę ukrytą, w której znajdowały się trzy neurony. Sieć trenowana była za pomocą algorytmu wstecznej propagacji błędów na podstawie tych samych zbiorów uczących, które zastosowano dla zagregowanych drzew klasyfikacyjnych. Tak jak poprzednio na podstawie błędnie rozpoznanych obserwacji w każdym zbiorze testowym obliczono ogólny błąd klasyfikacji. Graficzną prezentację otrzymanych wyników przedstawiono na rys. 1.

Wyniki klasyfikacji uzyskane za pomocą zagregowanych drzew oraz sieci neuronowej można uznać za satysfakcjonujące. Najmniejszy błąd klasyfikacji (12%) otrzymano za pomocą zagregowanych metodą boosting drzew klasyfikacyjnych. Największą wartość tej miary (30%) otrzymano, klasyfikując obiekty zbioru testowego za pomocą zagregowanych metodą bagging drzew klasyfikacyjnych. Najniższą wartość mediany ogólnego błędu klasyfikacji zanotowano dla algorytmu boosting (19,5%), najwyższą wartość tej miary średniej wyznaczono dla algorytmu bagging.

³ Agregacja modeli dyskryminacyjnych omówiona została m.in. w pracach: [Breiman 1996; Dietterich 2000; Freund, Schapire 1997; Gatnar, Walesiak 2004, s. 126; Walesiak, Gatnar 2009].

⁴ Informacje na temat sztucznych sieci neuronowych i o ich zastosowaniach w finansach znaleźć można m.in. w pracach: [Witkowska 2002; Angelini, Tollo, Roli 2008].



Rys. 1. Wykres pudełkowy

Źródło: opracowanie własne.

Analizując wykres pudełkowy, można stwierdzić, że najmniejszym rozproszeniem charakteryzowały się wyniki klasyfikacji uzyskane za pomocą algorytmu random forest. Z kolei największe rozproszenie wyników zanotowano dla zagregowanych drzew metodą bagging. Graficzna prezentacja wyników uzyskanych dla poszczególnych metod klasyfikacyjnych potwierdza hipotezę, że w badanym zbiorze najlepszą metodą klasyfikacji jest drzewo klasyfikacyjne zagregowane metodą boosting.

W kolejnym etapie badań porównano efektywność klasyfikacji każdej z metod, stosując dwa nieparametryczne testy statystyczne. Na początku, posługując się testem Wilcoxon [Domański, Pruska 2000 s. 221-222], sprawdzono, czy różnice w błędach klasyfikacji w każdym zbiorze testowym dla poszczególnych par (zasto-

Tabela 2. Porównanie efektywności klasyfikatorów za pomocą testu Wilcoxon

Porównywane metody klasyfikacji	Wartość statystyki Z	p-value
boosting & SSN	1,597618	0,110129
boosting & bagging	6,092718	0,000000
boosting & random forest	5,130718	0,000000
random forest & SSN	1,240446	0,214811
random forest & bagging	2,422166	0,015429
SSN & bagging	1,748741	0,080337

Źródło: obliczenia własne.

sowanych metod) są statystycznie istotne. Wyniki tego testu przedstawiono w tab. 2. W końcowym etapie badań porównano efektywność klasyfikacji zagregowanych drzew klasyfikacyjnych oraz sieci neuronowej, stosując zmodyfikowany test znaków. Wyniki tego testu przedstawiono w tab. 3.

Rezultaty badań przedstawione w tab. 2 potwierdzają istotną różnicę w jakości klasyfikacji przy użyciu par metod: boosting i bagging; boosting i random forest; bagging i random forest oraz SSN i bagging. Natomiast test Wilcoxon wykonany dla par klasyfikatorów sztuczna sieć neuronowa & boosting oraz sztuczna sieć neuronowa & random forest nie prowadzi do odrzucenia hipotezy zerowej (która zakłada, że różnica w błędach klasyfikacji dla badanej pary metod jest statystycznie nieistotna).

Posługując się testem Wilcoxon, można porównać jedynie pary klasyfikatorów, dlatego w literaturze [Demsar 2006] zalecane jest stosowanie dodatkowego testu, który pomoże rozstrzygnąć, która z metod klasyfikacji jest bardziej efektywna od pozostałych. Popularną metodą porównania jakości metod klasyfikacji jest zmodyfikowana forma testu znaków. Procedura tego testu jest następująca. Dla każdej pary metod sprawdza się, który z klasyfikatorów częściej „zwycięża”. Liczba zwycięstw jest niezbędna do wyznaczenia wartości statystyki krytycznej Z (za pomocą której weryfikuje się hipotezę o przewadze jednego z algorytmów) w teście znaków. Wyniki tego testu dla zagregowanych drzew klasyfikacyjnych oraz sieci neuronowej przedstawiono w tab. 3.

Tabela 3. Porównanie efektywności klasyfikatorów za pomocą testu znaków

Porównywane metody klasyfikacji	Liczba zwycięstw	Wartość statystyki Z	Poziom p
boosting & random forest	42	4,666905	0,000003
boosting & bagging	50	6,857143	0,000000
boosting & SSN	31	1,555635	0,119795
random forest & bagging	32	2,000000	0,045500
SSN & random forest	30	1,272792	0,203092
SSN & bagging	26	0,433013	0,665006

Źródło: opracowanie własne.

Rezultaty przedstawione w tab. 3 potwierdzają, że klasyfikacja za pomocą drzew klasyfikacyjnych zagregowanych algorytmem boosting jest bardziej efektywna od klasyfikacji za pomocą drzew zagregowanych algorytmem random forest. Z kolei algorytm random forest jest lepszym klasyfikatorem niż algorytm bagging. Nie udało się natomiast udowodnić, że klasyfikacja za pomocą sztucznej sieci neuronowej jest słabsza niż klasyfikacja przy użyciu zagregowanych drzew klasyfikacyjnych. Za pomocą testu znaków nie można również udowodnić, która z metod bagging czy random forest jest bardziej efektywna.

5. Podsumowanie

Przeprowadzony eksperyment potwierdził zasadność wykorzystania prezentowanych metod do klasyfikacji kredytów. Za pomocą każdej z nich poprawnie rozpoznanych zostało ok. 80% obiektów w zbiorze testowym. Przy wykorzystaniu testu Wilcoxon'a i testu znaków zweryfikowano hipotezę, że najlepsze wyniki klasyfikacji przy użyciu zagregowanych drzew klasyfikacyjnych uzyskano za pomocą algorytmu boosting. Zaskakującym rezultatem badań jest niższa niż przy użyciu algorytmu boosting skuteczność algorytmu random forest. Mniejsza efektywność tej metody może wynikać ze stosunkowo niewielkiej liczby zmiennych diagnostycznych.

W przeprowadzonym eksperymencie nie udało się natomiast udowodnić, która z metod: zagregowane drzewa klasyfikacyjne czy sztuczne sieci neuronowe, jest lepszym klasyfikatorem. Choć rezultaty badań sugerują przewagę algorytmu boosting nad zbudowanym perceptronem wielowarstwowym, weryfikacja statystyczna nie potwierdziła tej tezy. Niemniej jednak dobra jakość klasyfikacji uzyskana za pomocą sztucznych sieci neuronowych sugeruje kontynuację badań z wykorzystaniem tej metody.

Literatura

- Angelini E., Tollo G., Roli A., *A neural network approach for credit risk evaluation*, „The Quarterly Review of Economics and Finance” 2008, vol. 48, Issue 4.
- Breiman L., *Bagging predictors*, „Machine Learning” 1996 nr 26.
- Chrzanowska M., Alfaro E.A., Witkowska D., *The individual borrowers recognition: single and ensemble trees*, „Expert System with Applications” 2009, vol. 36.
- Chrzanowska M., Witkowska D., *Analiza efektywności klasyfikacji kredytobiorców uzyskanej zagregowanymi drzewami klasyfikacyjnymi dla prób o różnej strukturze*, Taksonomia 15, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, UE, Wrocław 2008.
- Chrzanowska M., Witkowska D., *Drzewa klasyfikacyjne w rozpoznawaniu kredytobiorców*, [w:] *Metody ilościowe w badaniach ekonomicznych*, Wydawnictwo SGGW, Warszawa 2007.
- Crook J., Edelman D., Thomas L., *Recent developments in consumer credit risk assessment*, „European Journal of Operational Research” 2007 vol. 132.
- Demsar J., *Statistical comparisons of classifiers over multiple data sets*, „Journal of Machine Learning Research” 2006 nr 7.
- Dietterich T.G., *Ensemble methods in machine learning*, „Multiple Classifier Systems” 2000 vol. 1857.
- Domański Cz., Pruska K., *Nieklasyczne metody statystyczne*, Państwowe Wydawnictwo Ekonomiczne, Warszawa 2000.
- Freund Y., Schapire R.E., *A Decision-Theoretic Generalization of on-line learning and an application to boosting*, „Journal of Computer and System Sciences” 55/1997.
- Gatnar E., Walesiak M. (red.), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, AE, Wrocław 2004.
- Walesiak M., Gatnar E. (red.), *Statystyczna analiza danych z wykorzystaniem programu R*, Wydawnictwo Naukowe PWN, Warszawa 2009.
- Witkowska D., *Sztuczne sieci neuronowe i metody statystyczne. Wybrane zagadnienia*, C.H. Beck, Warszawa 2002.

THE COMPARISON AND EVALUATION OF SELECTED METHODS OF INDIVIDUAL CREDIT CLASSIFICATION

Summary: In this paper I present the results of credit classification based on real data obtained from a bank. In this research several classification models were used. These models were constructed based on ensemble classification and artificial neural networks. The experiment was conducted in five series for each classification method. In each series a set was randomized and then divided into ten parts. Testing sets comprised 10% of the analyzed set. The aim of the investigation is to compare the effectiveness of classification methods when the structure of the testing sets changes at every stage.