

Modyfikacja metod uczenia klasyfikatorów złożonych na podstawie algorytmów bagging i boosting

Krzysztof Ragan^{1[249026]} and Maciej Pająk^{1[241632]}

Politechnika Wrocławska, Wrocław, Polska
<https://pwr.edu.pl>

Streszczenie Praca opisuje badanie jakości różnych modyfikacji Baggingu i Boostingu, w tym przede wszystkim NBBAG. Badania przeprowadzono na 22 zbiorach danych. Przeprowadzono wybór parametrów algorytmu NBBag, a następnie porównano statystycznie z innymi klasyfikatorami.

Keywords: NBBag · Bagging · Machine Learning

1 Wstęp

Uczenie maszynowe (ang. Machine learning) jest częścią sztucznej inteligencji. Pokazuje komputerom, w jaki sposób uczyć się i doskonalić na dużych zbiorach danych. Algorytmy, na podstawie zbiorów, trenowane są w celu szukania korelacji między danymi oraz tworzenia predykcji na podstawie analizy. Uczenie maszynowe ma zastosowanie praktycznie w każdej dziedzinie nauki i życia codziennego. Jego techniki można wykorzystać na wiele sposobów.

Istnieją cztery modele uczenia maszynowego:

- **Uczenie nadzorowane** obejmuje pary danych „wejściowych” i „wyjściowych”, przy czym dane „wyjściowe” zawierają oczekiwaną wartość. Należy tak wytrenować algorytm, żeby na podstawie posiadanych rozwiązań odpowiednio przetwarzał dane.
- **Uczenie nienadzorowane**, którego zasadniczą różnicą jest brak posiadania klucza rozwiązań. Algorytm na podstawie analizy nieuporządkowanych danych dochodzi do korelacji i wzorców między danymi.
- **Uczenie częściowo nadzorowane** polega na wprowadzeniu do danych surowych małego zbioru danych obrobionych. Na podstawie tych danych algorytm będzie odpowiednio uczył się i dopasowywał nieobrobione dane.
- **Uczenie ze wzmocnieniem** nie posiada klucza odpowiedzi, natomiast posiada zbiór reguł i zachowań, które algorytm powinien wykorzystać w celu osiągnięcia potencjalnych stanów końcowych. [1]

Uczenie zespołowe (ang. Ensemble learning) jest techniką uczenia maszynowego, która agreguje kilka podstawowych modeli, żeby stworzyć jeden model predykcyjny.

Metody baggingu i boostingu należą do metod uczenia zespołowego. Ze zbioru danych testowych są kreowane losowo podzbiory. Każdy z nich wykorzystywany jest do szkolenia klasyfikatora. Następnie wszystkie klasyfikatory są łączone. Bagging polega na bootstrapie oraz agregacji, stąd często nazywany jest po angielsku bootstrap aggregation. Metoda bootstrapu, to próbkowanie z podmianą. Agregacja natomiast jest uśrednianiem głosów wszystkich klasyfikatorów i stworzeniem odpowiedniej predykcji na podstawie większości głosów. Natomiast w boostingu "każdy następny prosty model jest uczony w sposób sekwencyjny" [2]. Oznacza to, że w przypadku następnego modelu, konieczna jest analiza poprzedniego oraz odpowiedni dobór wag wektorów obserwacji. Kiedy wektory te są źle przewidziane model bierze je pod uwagę ze zwiększoną wagą, przyznawaną zgodnie z zasadą- im lepszy wynik klasyfikacji, tym większa waga. Ada-Boost jest algorytem boostingu bez żadnego przetwarzania wstępnego. Wymaga wielu iteracji, podczas których wagi próbek są zmieniane w celu sklasyfikowania poprawnie w następnej iteracji przykładów, które są błędnie sklasyfikowane w bieżącej iteracji. Bez równoważenia specyficzny problem nierównowagi, który zwiększa dokładność.

Od momentu powstania podstawowych algorytmów baggingu i boostingu pojawiło się wiele modyfikacji i unowocześnień tychże algorytmów.

Metody uczenia zespołowego wykorzystane mogą być na wiele sposobów. Jednym z nich jest stworzenie modelu środowiska w szklarni. W celu pozyskania jak największych plonów oraz zoptymalizowania zużycia energii zaproponować można algorytm Gradient Boost Decision Tree (GBDT) oparty na nowo opracowanym algorytmie Light Gradient Boosting Machine (LightGBM lub LGBM) do modelowania temperatury wewnętrznej szklarni. Stworzono odpowiedni zbiór danych do szkolenia i walidacji modelu LGBM. [6]

Dodatkowym przykładem wykorzystania baggingu i boostingu może być udoskonalenie skuteczności i wiarygodności klasyfikacji cukrzycy. W artykule "Ensemble Learning Model for Diabetes Classification" [12] opisano w jaki sposób zostały wykorzystane i przetestowane metody. Metoda baggingu z podstawowym klasyfikatorem drzewa decyzyjnego okazała się najdokładniejsza.

1.1 Modyfikacje baggingu

Przykładowe modyfikacje zostały opisane w Tabeli nr 1 w artykule "Enhanced bagging (eBagging): A novel approach for ensemble learning" [14], gdzie zostały ułożone chronologicznie warianty baggingu z krótkim opisem oraz zaletami i wadami każdego z nich.

EBBag - najprostsza modyfikacja Baggingu to Exactly Balanced Bagging (EB-Bag), gdzie podczas bootstrappingu cała klasa mniejszościowa jest kopiowana i łączona z losowo wybranym podzbiorem klasy większościowej, aby dokładnie zrównoważyć liczebność przykładów z obu klas (Błaszczński et al. [4]).

NBBag - skupia bootstrapping na trudnych przykładach klasy mniejszościowej poprzez zastosowanie wag. Waga przykładu mniejszościowego zależy od analizy etykiet klas wśród jego k najbliższych sąsiadów. Przykład mniejszościowy jest tym bardziej zagrożony, im więcej ma przykładów większościowych w swoim sąsiedztwie. Tak więc ta część wagi odzwierciedla lokalny czynnik równoważący. Co więcej, waga ta jest również sumowana z globalnym czynnikiem równoważącym, który uwzględnia stosunek nierównowagi między klasami. Stąd wzór na wagę przykładu mniejszościowego jest następujący:

$$w = 0.5 \cdot \left(\frac{(N')^\psi}{k} + 1 \right)$$

gdzie N' - jest liczbą przykładów większościowych wśród k najbliższych sąsiadów przykładu, a ψ jest współczynnikiem skalującym. Każdemu przykładowi większościowemu przypisywana jest stała waga:

$$w = 0.5 \cdot \frac{N_+}{N_-}$$

(Błaszczński et al. [3]).

1.2 Modyfikacje boostingu

SMOTEBoost - połączenie algorytmu Smote z Boostingiem opartym na AdaBoost. Podczas każdej iteracji Boostingu zostaje użyty algorytm SMOTE na zbiorze treningowym co powoduje jego lepsze zbalansowanie. [13]. Sam algorytm SMOTE działa w ten sposób, że najpierw wybierany jest losowo przypadek klasy mniejszościowej A i znajduje się jego k najbliższych sąsiadów z klasy mniejszościowej. Następnie tworzona jest instancja syntetyczna poprzez losowe wybranie jednego z k najbliższych sąsiadów B i następnie połączenie A i B [9].

RUSBoost - łączy algorytm AdaBoost z algorytmem RUS (Random Under-Sampling), który jest najprostszym podejściem under-samplingu. Algorytm RUS polega na tym, że losowo wybiera się podzbiór klasy większościowej, który następnie zostaje połączony z klasą mniejszościową w jeden zbiór uczący [13].

2 Plan eksperymentu

Eksperyment, opisany w pracy polega na porównaniu algorytmów uczenia klasyfikatorów zespołowych. Porównane zostały algorytmy AdaBoost, RUSBoost, SMOTEBoost, Bagging, EBBag oraz NBBag. Ten ostatni (NBBag) można dostrajać za pomocą trzech parametrów (k , ψ oraz metryki odległości między sąsiadami) i także tego zadania podjęto się w tej pracy. Do oceny wyników eksperymentu zostały wykorzystane parowe testy statystyczne opisane krótko niżej. Punktami odniesienia zostały podstawowe algorytmy baggingu (Bagging) i boostingu (AdaBoost). Porównano opisane w poprzedniej sekcji modyfikacje wraz z modyfikacją zaproponowaną przez autorów pracy.

2.1 Zbiory danych

Do eksperymentu wybrano 22 zbiory danych dostępnych na stronie KEEL. Wszystkie zbiory posiadają dane niezbalansowane o współczynniku niezbalansowania między 1,5, a 9 (Tabela 1).

Tabela 1. Podstawowe informacje na temat zbiorów użytych w eksperymencie

| No. | Name | Features | Instances | IR |
|-----|------------------------|----------|-----------|------|
| 1 | ecoli-0_vs_1 | 7 | 220 | 1.86 |
| 2 | ecoli1 | 7 | 336 | 3.36 |
| 3 | ecoli2 | 7 | 336 | 5.46 |
| 4 | ecoli3 | 7 | 336 | 8.6 |
| 5 | glass-0-1-2-3_vs_4-5-6 | 9 | 214 | 3.2 |
| 6 | glass0 | 9 | 214 | 2.06 |
| 7 | glass1 | 9 | 214 | 1.82 |
| 8 | glass6 | 9 | 214 | 6.38 |
| 9 | haberman | 3 | 306 | 2.78 |
| 10 | iris0 | 4 | 150 | 2 |
| 11 | new-thyroid1 | 5 | 215 | 5.14 |
| 12 | new-thyroid2 | 5 | 215 | 5.14 |
| 13 | page-blocks0 | 10 | 5472 | 8.79 |
| 14 | pima | 8 | 768 | 1.87 |
| 15 | segment0 | 19 | 2308 | 6.02 |
| 16 | vehicle0 | 18 | 846 | 3.25 |
| 17 | vehicle1 | 18 | 846 | 2.9 |
| 18 | vehicle2 | 18 | 846 | 2.88 |
| 19 | vehicle3 | 18 | 846 | 2.99 |
| 20 | wisconsin | 9 | 683 | 1.86 |
| 21 | yeast1 | 8 | 1484 | 2.46 |
| 22 | yeast3 | 8 | 1484 | 8.1 |

2.2 Założenia eksperymentu

W eksperymencie postanowiono przyjąć pewne parametry, które będą niezmiennie dla każdej z porównywanych metod uczenia klasyfikatorów zespołowych. Tymi parametrami są:

- walidacja krzyżowa: 5/2 (5 foldów powtórzone 2 razy)
- liczba klasyfikatorów bazowych: 50
- współczynnik ufności: 95%

2.3 Parowe testy statystyczne

Testy statystyczne są kluczowym etapem całego projektu. Polegają one na statystycznej ocenie algorytmów. Dokonuje się tego poprzez potwierdzenie lub odrzucenie hipotezy, że dany algorytm jest statystycznie lepszy od innego, przy przyjętych współczynnikach ufności. W pracy wykorzystano testy T-studenta, Wilcozona oraz Friedmanna. Więcej informacji o tych testach można znaleźć w książce "Evaluating learning algorithms: a classification perspective" [11].

3 Wyniki badań oraz analizy statystycznej

3.1 Wybór najlepszych parametrów algorytmu NBBag

Przed porównaniem algorytmu NBBag z pozostałymi klasyfikatorami złożonymi wybrano jego najlepsze parametry:

- liczbę sąsiadów k (domyślnie 7)
- współczynnik skalujący ψ (domyślnie 1)
- miarę odległości między sąsiadami (domyślnie euklidesowa)

Dla każdego parametru wybrano kilka różnych wartości (algorytmów dla metryk) i przeprowadzono testy T-studenta. Podczas gdy badano jeden parametr, pozostałe pozostawały niezmiennie i ustalone na domyślną wartość.

W indeksie dolnym podana jest wartość k , w indeksie górnym wartość ψ , a w nawiasie skrót nazwy metryki (cheb - chebysheva, cos - cosinusowa, eucl - euklidesowa, manh - manhattan). Jeżeli brakuje danego parametru przy nazwie to parametr ten przybiera domyślna wartość.

W tabeli porównującej G-mean dla różnych wartości k tylko w trzech miejscach test T-studenta wykazał statystyczną przewagę (Tabela 2). Dla zbioru **glass1** NBBag₉ oraz NBBag₃ wypadły lepiej niż NBBag₅. Dodatkowo dla tego samego zbioru NBBag₉ okazało się lepsze od NBBag₁₁, a w zbiorze **haberman** NBBag₁₁ uzyskał lepszy wynik niż NBBag₉. Z uśrednionych rang wynikało, że NBBag₇ jest najlepszy. Jednakże porównanie G-mean nie wykazało statystycznej różnicy między porównywanymi wartościami k (p-value w teście Friedmanna to 0.596). Wybierano między wartościami NBBag₅ i NBBag₇, ponieważ są to typowe wartości dla liczby sąsiadów.

W tabeli porównującej G-mean dla różnych wartości ψ test T-studenta wykazał statystyczną przewagę w kilku zbiorach, w każdym z nich tylko jeden klasyfikator wykazywał statystyczną przewagę (Tabela 3). Dla zbioru **ecoli1** NBBag^{1.5} wypadł lepiej niż NBBag^{0.5} i NBBag². Dla **page-blocks0** NBBag¹ wypadł lepiej niż NBBag^{1.5}. Dla **vehicle0** NBBag² wypadł lepiej niż NBBag¹, a dodatkowo NBBag² okazał się lepszy niż NBBag¹ i NBBag^{1.5} dla zbioru **yeast3**. Z uśrednionych rang wynika, że wartość $\psi = 1$ okazała się najlepsza, jednakże porównanie G-mean nie wykazało statystycznej różnicy między porównywanymi wartościami parametru ψ (p-value w teście Friedmanna to 0.582). Mimo wszystko wybór zawężono do wartości $\psi = 1$ lub $\psi = 2$ i spośród nich wybrano $\psi = 2$, ponieważ dla dwóch zbiorów ta wartość statystycznie przeważała nad $\psi = 1$, a w drugą stronę nie wystąpiła żadna taka przewaga.

W tabeli porównującej różne miary odległości dla NBBag, też tylko w kilku miejscach pojawiła się statystyczna przewaga (Tabela 4). Dla zbioru **glass1** metryki chebysheva oraz manhattan miały przewagę nad metryką cosinusową, a dla zbioru **page-blocks0** miara chebysheva wykazywała przewagę nad miarą manhattan. Tutaj także statystycznie nie można określić, która metryka jest najlepsza (p-value w teście Friedmanna to 0.095), ale różnice są już bardziej zauważalne. Ze względu na to, że miara chebysheva była w dwóch miejscach lepsza statystycznie od innych miar, a sama nie była gorsza statystycznie od żadnej to została ona wybrana jako najlepsza z metryk.

Tabela 2. Wartości G-mean dla NBBag dla różnych wartości k oraz indeksy klasyfikatorów, nad którymi dany klasyfikator uzyskał statystyczną przewagę

| Dataset | NBBag ₃ 1 | NBBag ₅ 2 | NBBag ₇ 3 | NBBag ₉ 4 | NBBag ₁₁ 5 |
|------------------------|-------------------------|-------------------------|-------------------------|-------------------------|--------------------------|
| ecoli-0_vs_1 | 0.985 | 0.985 | 0.985 | 0.985 | 0.985 |
| ecoli1 | 0.853 | 0.856 | 0.856 | 0.852 | 0.849 |
| ecoli2 | 0.842 | 0.827 | 0.835 | 0.853 | 0.841 |
| ecoli3 | 0.711 | 0.719 | 0.719 | 0.734 | 0.711 |
| glass-0-1-2-3_vs_4-5-6 | 0.91 | 0.918 | 0.919 | 0.913 | 0.919 |
| glass0 | 0.825 | 0.829 | 0.832 | 0.833 | 0.819 |
| glass1 | 0.758 | 0.728 | 0.740 | 0.758 | 0.74 |
| glass6 | 0.903 | 0.921 | 0.921 | 0.921 | 0.921 |
| haberman | 0.514 | 0.517 | 0.515 | 0.499 | 0.522 |
| iris0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| new-thyroid1 | 0.94 | 0.94 | 0.947 | 0.947 | 0.947 |
| newthyroid2 | 0.906 | 0.914 | 0.913 | 0.913 | 0.921 |
| page-blocks0 | 0.923 | 0.925 | 0.926 | 0.926 | 0.925 |
| pima | 0.732 | 0.731 | 0.735 | 0.723 | 0.733 |
| segment0 | 0.985 | 0.986 | 0.985 | 0.985 | 0.984 |
| vehicle0 | 0.939 | 0.941 | 0.938 | 0.941 | 0.938 |
| vehicle1 | 0.644 | 0.640 | 0.646 | 0.647 | 0.659 |
| vehicle2 | 0.961 | 0.963 | 0.964 | 0.96 | 0.963 |
| vehicle3 | 0.640 | 0.635 | 0.628 | 0.631 | 0.625 |
| wisconsin | 0.963 | 0.96 | 0.958 | 0.958 | 0.961 |
| yeast1 | 0.661 | 0.665 | 0.659 | 0.668 | 0.661 |
| yeast3 | 0.865 | 0.870 | 0.865 | 0.863 | 0.865 |
| Mean ranks | 2.62 | 3.11 | 3.32 | 3.05 | 2.91 |

Tabela 3. Wartości G-mean dla NBBag dla różnych wartości ψ oraz indeksy klasyfikatorów, nad którymi dany klasyfikator uzyskał statystyczną przewagę

| Dataset | NBBag ^{0.5} | NBBag ¹ | NBBag ^{1.5} | NBBag ² |
|------------------------|----------------------|--------------------|----------------------|--------------------|
| | 1 | 2 | 3 | 4 |
| ecoli-0_vs_1 | 0.985 | 0.985 | 0.987 | 0.983 |
| | — | — | — | — |
| ecoli1 | 0.844 | 0.856 | 0.856 | 0.842 |
| | — | — | 1,4 | — |
| ecoli2 | 0.838 | 0.835 | 0.854 | 0.832 |
| | — | — | — | — |
| ecoli3 | 0.723 | 0.719 | 0.718 | 0.723 |
| | — | — | — | — |
| glass-0-1-2-3_vs_4-5-6 | 0.919 | 0.919 | 0.912 | 0.907 |
| | — | — | — | — |
| glass0 | 0.826 | 0.832 | 0.83 | 0.835 |
| | — | — | — | — |
| glass1 | 0.729 | 0.740 | 0.738 | 0.753 |
| | — | — | — | — |
| glass6 | 0.921 | 0.921 | 0.915 | 0.915 |
| | — | — | — | — |
| haberman | 0.490 | 0.515 | 0.493 | 0.517 |
| | — | — | — | — |
| iris0 | 1.000 | 1.000 | 1.000 | 1.000 |
| | — | — | — | — |
| new-thyroid1 | 0.947 | 0.947 | 0.940 | 0.933 |
| | — | — | — | — |
| newthyroid2 | 0.913 | 0.913 | 0.913 | 0.93 |
| | — | — | — | — |
| page-blocks0 | 0.925 | 0.926 | 0.923 | 0.924 |
| | — | 3 | — | — |
| pima | 0.723 | 0.735 | 0.718 | 0.717 |
| | — | — | — | — |
| segment0 | 0.985 | 0.985 | 0.984 | 0.983 |
| | — | — | — | — |
| vehicle0 | 0.942 | 0.938 | 0.946 | 0.951 |
| | — | — | — | 2 |
| vehicle1 | 0.632 | 0.646 | 0.647 | 0.654 |
| | — | — | — | — |
| vehicle2 | 0.963 | 0.964 | 0.963 | 0.965 |
| | — | — | — | — |
| vehicle3 | 0.626 | 0.628 | 0.623 | 0.631 |
| | — | — | — | — |
| wisconsin | 0.958 | 0.958 | 0.961 | 0.962 |
| | — | — | — | — |
| yeast1 | 0.663 | 0.659 | 0.661 | 0.656 |
| | — | — | — | — |
| yeast3 | 0.868 | 0.865 | 0.865 | 0.877 |
| | — | — | — | 2,3 |
| Mean ranks | 2.39 | 2.75 | 2.27 | 2.59 |

Tabela 4. Wartości G-mean dla NBBag dla różnych miar odległości oraz indeksy klasyfikatorów, nad którymi dany klasyfikator uzyskał statystyczną przewagę

| Dataset | NBBag(cheb) | NBBag(cos) | NBBag(eucl) | NBBag(manh) |
|------------------------|-------------|------------|-------------|-------------|
| | 1 | 2 | 3 | 4 |
| ecoli-0_vs_1 | 0.985 | 0.985 | 0.985 | 0.985 |
| ecoli1 | 0.853 | 0.855 | 0.856 | 0.846 |
| ecoli2 | 0.852 | 0.84 | 0.835 | 0.847 |
| ecoli3 | 0.727 | 0.719 | 0.719 | 0.719 |
| glass-0-1-2-3_vs_4-5-6 | 0.902 | 0.913 | 0.919 | 0.925 |
| glass0 | 0.824 | 0.816 | 0.832 | 0.815 |
| glass1 | 0.742 | 0.717 | 0.740 | 0.736 |
| glass6 | 0.921 | 0.901 | 0.921 | 0.911 |
| haberman | 0.518 | 0.517 | 0.515 | 0.524 |
| iris0 | 1.000 | 1.000 | 1.000 | 1.000 |
| new-thyroid1 | 0.947 | 0.932 | 0.947 | 0.940 |
| newthyroid2 | 0.913 | 0.913 | 0.913 | 0.913 |
| page-blocks0 | 0.929 | 0.926 | 0.926 | 0.926 |
| pima | 0.730 | 0.720 | 0.735 | 0.732 |
| segment0 | 0.984 | 0.985 | 0.985 | 0.985 |
| vehicle0 | 0.943 | 0.938 | 0.938 | 0.941 |
| vehicle1 | 0.645 | 0.643 | 0.646 | 0.650 |
| vehicle2 | 0.962 | 0.961 | 0.964 | 0.959 |
| vehicle3 | 0.633 | 0.633 | 0.628 | 0.641 |
| wisconsin | 0.960 | 0.961 | 0.958 | 0.960 |
| yeast1 | 0.660 | 0.664 | 0.659 | 0.656 |
| yeast3 | 0.866 | 0.864 | 0.865 | 0.863 |
| Mean ranks | 2.89 | 2.09 | 2.73 | 2.30 |

Ostatecznie do porównania algorytmu NBBag z innymi algorytmami wybrano parametry:

- $k = 7$
- $\psi = 2$
- metryka chebysheva

Dalej algorytm NBBag z tymi parametrami, czyli $\text{NBBag}_7^2(\text{cheb})$ jest nazywany dla uproszczenia NBBag jako wybrany reprezentant spośród wszystkich kombinacji parametrów.

Dla wartości G-mean najlepiej wypadł EBBag (Tabela 5). W większości zbiorów okazał się lepszy statystycznie od wszystkich pozostałych. Badany NBBag zachował się podobnie do zwykłego Baggingu, jednakże są zbiory, dla których jeden przeważał nad drugim i odwrotnie. Bagging wypadł statystycznie lepiej niż NBBag dla zbiorów `pima` i `segment0`, a NBBag wypadł statystycznie lepiej niż Bagging dla zbioru `yeast3`. Porównując wartości G-mean algorytmy Boostingowe wypadły słabo, ale w kilku miejscach wykazały statystyczną przewagę nad niektórymi algorytmami Baggingowymi. Na przykład dla zbioru `segment0` RUSBoost okazał się lepszy niż NBBag i EBBag, a dla zbioru `vehicle3` algorytmy AdaBoost i SMOTEAdaBoost wykazały statystyczną przewagę nad NBBag i Bagging.

Rozpoznawanie klasy większościowej czyli czułość (Tabela 6) prezentuje się podobnie jak G-mean. EBBag prawie dla wszystkich zbiorów okazał się statystycznie lepszy. Bagging i NBBag znów wypadły podobnie do siebie. Bagging znów dwa razy wypadł lepiej dla zbiorów `pima` i `ecoli1`, a NBBag wypadł statystycznie lepiej niż Bagging znowu dla zbioru `yeast3`. Tutaj podobnie jak dla G-mean algorytmy Boostingowe wypadały ogólnie gorzej niż Baggingowe, z wyjątkami na kilku zbiorach. Na zbiorach `haberman`, `page-blocks0`, `vehicle1` i `vehicle3` algorytm SMOTEAdaBoost wypadł lepiej od NBBag i Bagging na raz, dodatkowo dla `vehicle3` zwykły AdaBoost wypadł lepiej niż NBBag i Bagging.

Z kolei analizując rozpoznawanie klasy mniejszościowej, czyli tabelę specyficzności (Tabela 7) sprawy mają się inaczej niż we wcześniejszych tabelach. Teraz EBBag prawie nigdzie nie okazał się lepszy statystycznie od któregośkolwiek innego klasyfikatora. Bagging oraz NBBag nadal wykazują podobieństwo, ale teraz dołącza do nich także RUSBoost. NBBag wykazał statystyczną przewagę nad Baggingiem dla zbiorów `glass0` oraz `vehicle1` i nigdzie nie jest lepszy od RUSBoost. Natomiast Bagging okazał się lepszy dla `segment0`, `vehicle0` i `yeast3`. Natomiast RUSBoost okazał się lepszy dla `glass1` od Baggingu. EBBag, AdaBoost i SMOTEAdaBoost nie wykazywały statystycznej przewagi od pozostałych na żadnym zbiorze.

Tabela 5. Porównanie wartości G-mean NBBag₇²(cheb) i innych klasyfikatorów złożonych oraz indeksy klasyfikatorów, nad którymi dany klasyfikator uzyskał statystyczną przewagę

| Dataset | NBBag ₇ ² (cheb) | AdaBoost | EBBAG | Bagging | RUSBoost | SMOTEAdaBoost |
|------------------------|--|----------|---------|---------|----------|---------------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| ecoli-0_vs_1 | 0.985 | 0.976 | 0.983 | 0.985 | 0.978 | 0.969 |
| | 6 | — | — | 6 | — | — |
| ecoli1 | 0.839 | 0.826 | 0.892 | 0.857 | 0.859 | 0.853 |
| | — | — | 1,2,4,6 | — | — | — |
| ecoli2 | 0.856 | 0.812 | 0.896 | 0.841 | 0.819 | 0.849 |
| | — | — | ALL | — | — | — |
| ecoli3 | 0.692 | 0.730 | 0.891 | 0.723 | 0.698 | 0.769 |
| | — | — | ALL | — | — | — |
| glass-0-1-2-3_vs_4-5-6 | 0.903 | 0.894 | 0.934 | 0.915 | 0.907 | 0.907 |
| | — | — | — | — | — | — |
| glass0 | 0.853 | 0.777 | 0.848 | 0.830 | 0.815 | 0.752 |
| | 2,6 | — | 2,6 | 2,6 | 6 | — |
| glass1 | 0.762 | 0.690 | 0.766 | 0.743 | 0.758 | 0.755 |
| | 2 | — | 2 | — | 2 | 2 |
| glass6 | 0.923 | 0.875 | 0.934 | 0.896 | 0.892 | 0.895 |
| | 2 | — | 2,4 | — | — | — |
| haberman | 0.525 | 0.470 | 0.611 | 0.499 | 0.467 | 0.526 |
| | 2,5 | — | 1,2,4,5 | — | — | — |
| iris0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | — | — | — | — | — | — |
| new-thyroid1 | 0.948 | 0.956 | 0.965 | 0.930 | 0.973 | 0.964 |
| | — | — | — | — | — | 4 |
| newthyroid2 | 0.946 | 0.915 | 0.966 | 0.948 | 0.961 | 0.947 |
| | 2 | — | 2 | — | 2 | — |
| page-blocks0 | 0.923 | 0.897 | 0.959 | 0.927 | 0.863 | 0.928 |
| | 2,5 | — | ALL | 2,5 | — | 2,5 |
| pima | 0.709 | 0.664 | 0.737 | 0.725 | 0.652 | 0.682 |
| | 2,5 | — | 1,2,5,6 | 1,2,5,6 | — | 5 |
| segment0 | 0.983 | 0.985 | 0.989 | 0.986 | 0.994 | 0.990 |
| | — | — | — | 1 | 1,3 | — |
| vehicle0 | 0.948 | 0.896 | 0.963 | 0.939 | 0.926 | 0.902 |
| | 2,6 | — | ALL | — | — | — |
| vehicle1 | 0.653 | 0.651 | 0.758 | 0.660 | 0.623 | 0.662 |
| | — | — | ALL | — | — | — |
| vehicle2 | 0.966 | 0.936 | 0.967 | 0.970 | 0.954 | 0.938 |
| | 2,6 | — | 2,6 | 2,6 | — | — |
| vehicle3 | 0.617 | 0.666 | 0.754 | 0.613 | 0.577 | 0.694 |
| | 5 | 1,4,5 | ALL | — | — | 1,4,5 |
| wisconsin | 0.963 | 0.933 | 0.970 | 0.962 | 0.967 | 0.927 |
| | 2,6 | — | 2,6 | 2,6 | 2,6 | — |
| yeast1 | 0.660 | 0.650 | 0.718 | 0.667 | 0.594 | 0.641 |
| | 5 | 5 | ALL | 5 | — | 5 |
| yeast3 | 0.875 | 0.798 | 0.927 | 0.842 | 0.815 | 0.853 |
| | 2,4,5,6 | — | ALL | 2 | — | 2 |
| Mean ranks | 3.59 | 1.93 | 5.57 | 3.68 | 2.89 | 3.34 |

Tabela 6. Porównanie czułości $NBBag_7^2(\text{cheb})$ i innych klasyfikatorów złożonych oraz indeksy klasyfikatorów, nad którymi dany klasyfikator uzyskał statystyczną przewagę

| Dataset | $NBBag_7^2(\text{cheb})$ | AdaBoost | EBBag | Bagging | RUSBoost | SMOTEAdaBoost |
|------------------------|--------------------------|----------|---------|---------|----------|---------------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| ecoli-0_vs_1 | 0.997 | 0.979 | 0.993 | 0.997 | 0.983 | 0.979 |
| | — | — | — | — | — | — |
| ecoli1 | 0.761 | 0.741 | 0.897 | 0.793 | 0.792 | 0.792 |
| | — | — | ALL | 1 | — | — |
| ecoli2 | 0.758 | 0.713 | 0.845 | 0.731 | 0.704 | 0.790 |
| | — | — | 1,2,4,5 | — | — | 4 |
| ecoli3 | 0.514 | 0.571 | 0.900 | 0.557 | 0.529 | 0.614 |
| | — | — | ALL | — | — | — |
| glass-0-1-2-3_vs_4-5-6 | 0.853 | 0.844 | 0.922 | 0.874 | 0.855 | 0.855 |
| | — | — | 1,5 | — | — | — |
| glass0 | 0.807 | 0.721 | 0.843 | 0.779 | 0.743 | 0.771 |
| | 2 | — | 2 | — | — | — |
| glass1 | 0.646 | 0.593 | 0.700 | 0.634 | 0.631 | 0.685 |
| | — | — | 2,4 | — | — | 2 |
| glass6 | 0.880 | 0.810 | 0.900 | 0.827 | 0.810 | 0.823 |
| | 2 | — | 2,4 | — | — | — |
| haberman | 0.334 | 0.285 | 0.531 | 0.310 | 0.272 | 0.433 |
| | 2,5 | — | 1,2,4,5 | — | — | 1,2,4,5 |
| iris0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | — | — | — | — | — | — |
| new-thyroid1 | 0.914 | 0.943 | 0.957 | 0.886 | 0.971 | 0.943 |
| | — | — | — | — | — | — |
| newthyroid2 | 0.914 | 0.871 | 0.957 | 0.914 | 0.943 | 0.886 |
| | — | — | — | — | — | — |
| page-blocks0 | 0.863 | 0.817 | 0.954 | 0.87 | 0.759 | 0.891 |
| | 2,5 | — | ALL | 2,5 | — | 1,2,4,5 |
| pima | 0.595 | 0.578 | 0.699 | 0.621 | 0.517 | 0.608 |
| | 5 | — | ALL | 1,5 | — | 5 |
| segment0 | 0.969 | 0.974 | 0.985 | 0.974 | 0.991 | 0.982 |
| | — | — | 1 | — | 1 | 1 |
| vehicle0 | 0.932 | 0.844 | 0.982 | 0.907 | 0.884 | 0.847 |
| | 2,6 | — | ALL | 2,6 | — | — |
| vehicle1 | 0.477 | 0.521 | 0.747 | 0.496 | 0.443 | 0.572 |
| | — | — | ALL | — | — | 1,4,5 |
| vehicle2 | 0.950 | 0.908 | 0.961 | 0.959 | 0.924 | 0.933 |
| | 2 | — | 2,6 | 2,6 | — | 2 |
| vehicle3 | 0.417 | 0.537 | 0.727 | 0.415 | 0.372 | 0.545 |
| | — | 1,4,5 | ALL | — | — | 1,4,5 |
| wisconsin | 0.958 | 0.904 | 0.971 | 0.954 | 0.96 | 0.914 |
| | 2,6 | — | 2,4,6 | 2,6 | 2,6 | — |
| yeast1 | 0.493 | 0.54 | 0.658 | 0.507 | 0.412 | 0.529 |
| | 5 | 5 | ALL | 5 | — | 5 |
| yeast3 | 0.792 | 0.669 | 0.92 | 0.73 | 0.69 | 0.795 |
| | 2,4,5 | — | ALL | 2 | — | 2,4,5 |
| Mean ranks | 3.27 | 2.11 | 5.70 | 3.57 | 2.64 | 3.70 |

Tabela 7. Porównanie specyficzności $\text{NBBag}_7^2(\text{cheb})$ i innych klasyfikatorów złożonych oraz indeksy klasyfikatorów, nad którymi dany klasyfikator uzyskał statystyczną przewagę

| Dataset | $\text{NBBag}_7^2(\text{cheb})$ | AdaBoost | EBBag | Bagging | RUSBoost | SMOTEAdaBoost |
|------------------------|---------------------------------|----------|-------|---------|----------|---------------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| ecoli-0_vs_1 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 | 0.975 |
| | — | — | — | — | — | — |
| ecoli1 | 0.929 | 0.930 | 0.890 | 0.928 | 0.936 | 0.913 |
| | 3 | 3 | — | 3 | 3 | — |
| ecoli2 | 0.981 | 0.944 | 0.956 | 0.981 | 0.970 | 0.949 |
| | 2,3,6 | — | — | 2,3,6 | 2 | — |
| ecoli3 | 0.963 | 0.953 | 0.882 | 0.970 | 0.952 | 0.914 |
| | 3,6 | 3,6 | — | 2,3,6 | 3,6 | 3 |
| glass-0-1-2-3_vs_4-5-6 | 0.960 | 0.957 | 0.948 | 0.963 | 0.969 | 0.941 |
| | — | — | — | — | 2,3,6 | — |
| glass0 | 0.907 | 0.844 | 0.858 | 0.889 | 0.896 | 0.809 |
| | 2,3,4,6 | — | 6 | 2,3,6 | 6 | — |
| glass1 | 0.910 | 0.808 | 0.848 | 0.884 | 0.921 | 0.826 |
| | 2,3,6 | — | — | 2,3,6 | 2,3,4,6 | — |
| glass6 | 0.973 | 0.951 | 0.973 | 0.976 | 0.989 | 0.962 |
| | 2 | — | — | 2 | 2 | — |
| haberman | 0.833 | 0.789 | 0.709 | 0.820 | 0.844 | 0.758 |
| | 2,3,6 | 3 | — | 2,3,6 | 2,3,6 | 3 |
| iris0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | — | — | — | — | — | — |
| new-thyroid1 | 0.989 | 0.969 | 0.975 | 0.983 | 0.975 | 0.992 |
| | — | — | — | — | — | 3 |
| newthyroid2 | 0.983 | 0.967 | 0.978 | 0.986 | 0.983 | 0.983 |
| | — | — | — | — | — | — |
| page-blocks0 | 0.987 | 0.987 | 0.963 | 0.987 | 0.985 | 0.976 |
| | 3,6 | 3,6 | — | 3,6 | 3,6 | 3 |
| pima | 0.848 | 0.769 | 0.779 | 0.849 | 0.830 | 0.747 |
| | 2,3,6 | — | 6 | 2,3,6 | 2,3,6 | — |
| segment0 | 0.997 | 0.996 | 0.992 | 0.998 | 0.997 | 0.996 |
| | 3 | 3 | — | 1,3 | 3 | 3 |
| vehicle0 | 0.965 | 0.952 | 0.944 | 0.974 | 0.971 | 0.958 |
| | 2,3,6 | — | — | 1,2,3,6 | 2,3,6 | — |
| vehicle1 | 0.896 | 0.816 | 0.773 | 0.881 | 0.893 | 0.819 |
| | 2,3,4,6 | 3 | — | 2,3,6 | 2,3,6 | 3 |
| vehicle2 | 0.983 | 0.966 | 0.974 | 0.981 | 0.986 | 0.971 |
| | 2,3,6 | — | — | 2,6 | 2,3,6 | — |
| vehicle3 | 0.920 | 0.835 | 0.784 | 0.913 | 0.907 | 0.828 |
| | 2,3,6 | 3 | — | 2,3,6 | 2,3,6 | 3 |
| wisconsin | 0.968 | 0.963 | 0.968 | 0.971 | 0.974 | 0.965 |
| | — | — | — | 2 | 2 | — |
| yeast1 | 0.886 | 0.787 | 0.785 | 0.880 | 0.871 | 0.778 |
| | 2,3,6 | — | — | 2,3,6 | 2,3,6 | — |
| yeast3 | 0.969 | 0.958 | 0.935 | 0.974 | 0.969 | 0.958 |
| | 2,3,6 | 3 | — | 1,2,3,6 | 2,3,6 | 3 |
| Mean ranks | 4.39 | 2.48 | 2.16 | 4.91 | 4.70 | 2.36 |

Test Wilcoxona (Tabela 8) potwierdza zebrane wcześniej obserwacje. NBBag i Bagging zachowują się bardzo podobnie. Najlepiej ogólnie wypadł EBBag, ale jeśli chodzi rozpoznawanie klasy mniejszościowej to wypadł najgorzej, na równi z AdaBoost i SMOTEAdaBoost. Najgorzej ogólnie wypadł AdaBoost, który nigdzie nie jest statystycznie lepszy od któregośkolwiek innego klasyfikatora.

Tabela 8. Indeksy klasyfikatorów, nad którymi dany klasyfikator uzyskał statystyczną przewagę dla $\text{NBBag}_7^2(\text{cheb})$ i innych klasyfikatorów złożonych na podstawie testu Wilcoxona

| | NBBag ₇ ² (cheb) | AdaBoost | EBBag | Bagging | RUSBoost | SMOTEAdaBoost |
|-------------|--|----------|-------|---------|----------|---------------|
| Metric | 1 | 2 | 3 | 4 | 5 | 6 |
| G-mean | 2 | — | ALL | 2 | — | 2 |
| F1 | 2, 6 | — | ALL | 2, 5, 6 | 2 | 2 |
| BAC | 2, 6 | — | ALL | 2, 6 | 2 | — |
| Precision | 2, 3, 6 | — | 6 | 2, 3, 6 | 2, 3, 6 | — |
| Recall | 2 | — | ALL | 2, 5 | — | 2, 5 |
| Specificity | 2, 3, 6 | — | — | 2, 3, 6 | 2, 3, 6 | — |

4 Podsumowanie i wnioski

Z badanych klasyfikatorów najlepiej wypadła modyfikacja baggingu: EBBag, a modyfikacja NBBag nie uzyskała wystarczających statystycznie rezultatów, aby uznać, że jest lepsza niż zwykły Bagging. Jednakże jeżeli zależałoby na rozpoznawaniu klasy mniejszościowej to trzeba wykluczyć z rozważanych rozwiązań EBBag, ponieważ sprawdza się wtedy jako jedno z gorszych rozwiązań. Wtedy należy rozważać podejście Bagging lub jego modyfikację NBBag oraz RUSBoost. NBBag ma jednak tę przewagę, że ma dodatkowe parametry, które określają wagi próbek w procesie uczenia. NBBag z wszystkimi wagami równymi 1 staje się zwykłym baggingiem. Daje to więc więcej możliwości na próbowanie dopasowania się do określonego zadania i zbioru danych.

Bibliografia

- [1] Czym jest uczenie maszynowe (machine learning)?: Definicja, rodzaje inbsp;przykłady: Sap insights. URL <https://www.sap.com/poland/insights/what-is-machine-learning.html>.
- [2] p. C. Angielczyk. Bagging i boosting – czyli uczenie zespołowe. URL <https://divideetimpera.net/bagging-i-boosting-czyli-uczenie-zespolowe/>.
- [3] J. Błaszczyński and J. Stefanowski. Neighbourhood sampling in bagging for imbalanced data. *Neurocomputing*, 150:529–542, 2015.
- [4] J. Błaszczyński and J. Stefanowski. Actively balanced bagging for imbalanced data. In *International Symposium on Methodologies for Intelligent Systems*, pages 271–281. Springer, 2017.
- [5] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, Aug 1996. ISSN 1573-0565. <https://doi.org/10.1007/BF00058655>. URL <https://doi.org/10.1007/BF00058655>.
- [6] W. Cai, R. Wei, L. Xu, and X. Ding. A method for modeling greenhouse temperature using gradient boost decision tree, Sep 2021. URL <https://www.sciencedirect.com/science/article/pii/S2214317321000743#m0015>.
- [7] P. Eswaran. Enhanced adaboost algorithm with modified weighting scheme for imbalanced problems. 06:01–05, 07 2018. <https://doi.org/10.9756/SIJCSEA/V6I4/06040060101>.
- [8] S. González, S. García, J. Del Ser, L. Rokach, and F. Herrera. A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. *Information Fusion*, 64:205–237, 2020. ISSN 1566-2535. <https://doi.org/https://doi.org/10.1016/j.inffus.2020.07.007>. URL <https://www.sciencedirect.com/science/article/pii/S1566253520303195>.
- [9] H. He and Y. Ma. Imbalanced learning: foundations, algorithms, and applications. 2013.
- [10] E. Hillebrand, M. Lukas, and W. Wei. Bagging weak predictors. *International Journal of Forecasting*, 37(1):237–254, 2021. ISSN 0169-2070. <https://doi.org/https://doi.org/10.1016/j.ijforecast.2020.05.002>. URL <https://www.sciencedirect.com/science/article/pii/S0169207020300649>.
- [11] N. Japkowicz and M. Shah. *Evaluating learning algorithms: a classification perspective*. Cambridge University Press, 2011.
- [12] N. Nai-arun and P. Sittidech. Ensemble learning model for diabetes classification. *Advanced Materials Research*, 931-932:1427–1431, 05 2014. <https://doi.org/10.4028/www.scientific.net/AMR.931-932.1427>.
- [13] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano. Rusboost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on*

- Systems, Man, and Cybernetics-Part A: Systems and Humans*, 40(1):185–197, 2009.
- [14] G. TÜYSÜZOĞLU and D. Birant. Enhanced bagging (ebagging): A novel approach for ensemble learning. *International Arab Journal of Information Technology*, 17(4), 2020.