

Dr hab. inż. Michał Woźniak, prof. PW
Wydział Elektroniki
Politechnika Wrocławska, Wrocław

ZESPOŁY KLASYFIKATORÓW - AKTUALNE KIERUNKI BADAŃ

STRESZCZENIE

Jednym z ciekawych i dynamicznie rozwijających się nurtów uczenia maszynowego jest klasyfikacja kombinowana. Opracowywane, w jej ramach, algorytmy starają się w zbudować model systemu klasyfikacyjnego bazującego na klasyfikatorach składowych, tak aby wykorzystać ich najlepsze cechy i kompetencje potrzebne do rozwiązania danego problemu decyzyjnego. W trakcie konstrukcji tego typu systemów stykamy się z dwoma typami problemów: jak wybrać wartościowy zespół klasyfikatorów oraz w jaki sposób uzyskać decyzje końcową na bazie odpowiedzi członków wspomnianego zespołu klasyfikatorów. W pracy przedstawiono główne przesłanki świadczące o przydatności projektowania tego typu systemów oraz dokonano ich krótkiej charakterystyki problemów projektowych.

Słowa kluczowe: Klasyfikacja, zespoły klasyfikatorów, klasyfikacja kombinowania, rozpoznawanie wzorców

CLASSIFIER ENSEMBLE – RECENT RESEARCH DIRECTIONS

SUMMARY

Classifier ensemble is the focus of the intense research, because it is recognized as the one of them most efficient classification approach. It is used in the several practical domains as fraud detection, client behavior recognition, medical decision support systems, or technical diagnostic to enumerate only a few. In this conceptual approach, the main effort is focusing on the two main problems. First, how to choose or train valuable and mutually complimentary set of individual classifiers and how to combine their outputs to exploit the strength of each individuals. The work presents a brief survey of the main issues related with the classifier ensemble domain.

Keywords: Classification, classifier ensemble, combined classifier, pattern classification

1. WPROWADZENIE

Ze względu na szybki postęp technologiczny, współczesne systemy informatyczne stają się coraz bardziej złożone i wymagają coraz lepszych metod analizy danych, które pozwolą na odkrycie interesujących związków w gromadzonych danych. Jedną z obiecujących dróg analizy danych są złożone systemy rozpoznawania obiektów, które znajdują szerokie zastosowanie w przeróżnych obszarach praktycznej działalności człowieka, takich jak np. rozpoznawanie mowy, systemy nadzoru, komputerowo wspomaganą diagnostykę medyczną i techniczną, wykrywanie oszustw w transakcjach finansowych, detekcja niechcianej poczty, ataków w sieciach komputerowych, czy też odkrywanie zachowania klientów.

Rozważając możliwość budowy systemów klasyfikacji musimy sobie zdawać sprawę z faktu, że współczesne systemy informatyczne gromadzą przeoblicznie zbiory danych, a ich efektywna analiza staje się nie lada wyzwaniem. Ilość informacji gromadzonych w nich podwaja się co dwa lata. Przykładowo w 2011r. 1.8 zetabajtów (2^{21} bajtów) danych zostało wytworzonych i zapamiętanych przez systemy informatyczne. Ludzie wysyłają 14 miliardów listów elektronicznych oraz 340 milionów twittów dziennie. Wyszukiwarka Google otrzymuje 2 miliony zapytań co minutę¹, natomiast sieć handlowa Wal-Mart² rejestruje milion transakcji co godzinę. Użytkownicy Facebook utworzyli dotąd ponad miliard kont i udostępnili ponad 40 miliardów fotografii.

Nie tylko rozmiar analizowanych danych stanowi wyzwanie. Problemy dedykowane metodom analizy dużych zbiorów danych (*big data*) są często charakteryzowane przez tzw. 5V, tzn.:

- *Volume* obejmuje problemy związane z analizą dużych, wielowymiarowych zbiorów danych.
- *Velocity*, to zagadnienia związane z faktem, że dane napływają do system w sposób ciągły, co wymaga dedykowanych narzędzi ich analizy w trybie *on-line*, a co jest nie lada wyzwaniem wymaga to także opracowania dedykowanych metod analizy uwzględniających zjawisko tzw. *concept drift*, które polega na tym, że parametry rozkładów prawdopodobieństwa, z jakich pochodzą dane, mogą się zmieniać w czasie.
- *Variety*, koncentruje się na problemach związanych z potrzebą opracowania takich metod analizy danych, które uwzględniać będą heterogeniczną naturę danych. Obecne systemy muszą sobie radzić nie tylko z danymi ustrukturalizowanymi, ale także z analizą danych tekstowych, informacją obrazową, czy też sygnałem wideo.
- *Veracity* określa grupę zagadnień związanych z jakością danych, a właściwie z faktem wiarygodności danych. Skoro dane pochodzą z różnych źródeł, to istotnym elementem jest z jednej strony umiejętność oceny jakości źródła, z drugiej zaś wykorzystanie tej informacji w procesie analizy.

¹ Blog autorstwa Marcia Conner <http://marciaconner/blog/data-on-big-data>

² Economist 25.02.2010, „Data, data everywhere”

- *Value*, to chyba najważniejsze „V”, gdyż określa jak wyniki analizy mogą być przydatne dla prowadzenia działalności biznesowej.

W pracy skoncentrujemy się na tzw. kombinowanych metodach klasyfikacji, które są uważane za efektywne narzędzia analizy danych mogących z powodzeniem mieć zastosowanie do analizy danych typu big data.

2. ZADANIE KLASYFIKACJI

Przez pojęcie klasyfikacji obiektów rozumie się schemat postępowania, w którym na podstawie zmierzonych wartości cech rozpoznawanego obiektu oraz ewentualnie innych informacji znanych *a priori* podejmuje się pojedynczą decyzję o przynależności obiektu do danej klasy. Proces klasyfikacji poprzedzają czynności przygotowawcze związane z opracowaniem modelu (m.in. selekcja, czy też ekstrakcja cech). W pracy skupimy się jedynie na etapie decyzyjnym, to jest na formalnej zależności Ψ , która wektorowi wartości zmierzonych cech $x \in X \subseteq \mathbf{R}^d$ przyporządkowuje jeden z numerów klas $i \in M$, $M = \{1, 2, \dots, M\}$, tzn. $\Psi : X \rightarrow M$.

Istnieje wiele podejść związanych z klasyfikacją. Można do niej zaliczyć m.in. podejście bayesowskie, gdzie zakłada się, że wartości zmierzonych cech oraz klasyfikacja są realizacjami pary zmiennych losowych posiadających rozkład łączny. Prowadzi to do konstrukcji klasyfikatora, który minimalizuje wartość oczekiwaną kosztu błędnej decyzji. Inne podejścia wykorzystują koncepcję obliczeń neuronowych, czy też koncepcję zbiorów rozmytych, bądź też klasyczną logikę dwuwartościową, a związki pomiędzy obserwacjami a klasyfikacją wyrażane są w postaci zdań logicznych, najczęściej reguł (implikacji). Jednakże zgodnie z twierdzeniem sformułowanym przez Wolperta „no free lunch theorem” [16] żadna z metod klasyfikacji nie nadaje się do rozwiązania każdego zadania decyzyjnego, gdyż każdy z klasyfikatorów na etapie uczenia przyjmuje pewne założenia (tzw. bias), które limitują jego użyteczność do pewnej klasy problemów decyzyjnych. Z drugiej strony Watanabe zauważył, że każdy klasyfikator musi takie założenia przyjąć, gdyż bez nich nie byłby w stanie osiągnąć dobrej jakości. Obserwacja ta znana jest pod nazwą „Ugly ducking theorem” [14].

3. KLASYFIKACJA KOMBINOWANA

Jako, że nie ma możliwości skonstruowania klasyfikatora uniwersalnego, dlatego też coraz większym zainteresowaniem cieszą się tzw. metody hybrydowe, które starają się wykorzystać najlepsze cechy klasyfikatorów prostych do konstrukcji klasyfikatora złożonego. Jednym z takich podejść są klasyfikatory kombinowane (*multiple classifier system, classifier ensemble, combined classifier*). Idea ich konstruowania polega na niezależnej klasyfikacji obiektu przez grupę klasyfikatorów, a następnie uzgodnienie decyzji końcowej na ich podstawie.

Każdy klasyfikator kombinowany charakteryzowany jest przez:

- sposób połączenia klasyfikatorów prostych,
- skład zespołu klasyfikatorów,
- sposób uzgodnienia decyzji grupowej.

W większości prac autorzy rozważają jedynie równoległe połączenie klasyfikatorów, choć w literaturze możemy spotkać także połączenia szeregowe, czy też mieszane. Jednakże w pracy nie będziemy zajmować się tym zagadnieniem, a jedynie dwoma pozostałymi charakterystykami, a więc w jaki sposób utworzyć wartościową pulę klasyfikatorów prostych oraz jak w sposób efektywny otrzymać końcową decyzję.

3.1. Geneza klasyfikatorów kombinowanych

Już starożytni Grecy stawiali decyzję grupową ponad decyzjami jednostki, co dało podwaliny demokracji. Oczywiście można dyskutować, czy decyzje grupy, szczególnie gdy wypracowane są nie na podstawie niezależnych decyzji zawsze stanowią lepsze rozwiązanie niż decyzja pojedynczego obywatela, ale w literaturze można znaleźć szereg przekonujących dowodów na to. James Surowiecki we wstępie do swojej książki [13] pisze, że przy odpowiednich założeniach grupy są zwykle bardziej inteligentne niż ich członkowie. Grupy nie muszą być przy tym moderowane, czy dowodzone przez wyjątkowo inteligentne jednostki aby dokonywać inteligentnych wyborów. Nawet jeżeli wśród członków grupy znajdują się nie specjalnie dobrze znające się na rozważanym problemie osoby, to decyzja grupowa może nadal charakteryzować się wysoką jakością. Potwierdzeniem tego wydaje się być eksperyment przeprowadzony w 1906r. przez Francisca Galtona, który na targu rolnym poprosił 787 osób o oszacowanie wagi wystawionego tam byka. Wśród oceniających osób znajdowali się zarówno eksperci (hodowcy bydła, jak i rzeźnicy), ale także uczestnicy targu nie mających wiedzy na temat hodowli. Okazało się, że średnia wartość wagi prezentowanego byka podana przez przepytane osoby była niemalże zgodna z jego wagą rzeczywistą (średnia wynosiła 1 197 funtów, podczas, gdy byk ważył 1 198 funtów). Innym ciekawym przykładem jest popularny teleturniej Milionerzy, w którym uczestnik w przypadku, gdy ma wątpliwości co do odpowiedzi może albo zadzwonić do osoby, której wg niego powinna posiadać wiedzę związaną z zadaniem pytaniem, bądź zapytać się o odpowiedź publiczności w studiu telewizyjnym, która w większości nie posiada dużej wiedzy związanej z pytaniem. Okazuje się, że „eksperti” udzielali trafnej odpowiedzi w 65% przypadków, podczas, gdy odpowiedzi publiczności były poprawne w 91%.

3.2. Problemy wyboru klasyfikatorów

Jak wspomniano, jednym z najważniejszych problemów jest wybór odpowiednich klasyfikatorów prostych, gdyż jak można przypuszczać nie da się skonstruować wartościowego klasyfikatora wykorzystując klasyfikatory podobne do siebie. Wymagana jest ich różnorodność (ang. *diversity*), która oznacza, że klasyfikatory proste wchodzące w skład zespołu klasyfikatorów są z jednej strony różne, ale też charakteryzują się dość dobrą jakością, a co najważniejsze powinny posiadać taką własność, że dla niemal każdego przykładu z przestrzeni cech, co najmniej jeden z klasyfikatorów prostych podejmuje poprawną decyzję. Jednocześnie możemy zaobserwować zjawisko tzw. *bad diversity*”, gdzie różnorodność negatywnie wpływa na jakość klasyfikacji puli klasyfikatorów. Ciekawe wnioski w tym zakresie można znaleźć w [1].

Mamy zatem do wyboru jedną z dwóch dróg budowy wartościowej puli klasyfikatorów:

- zapewnienie takich warunków uczenia klasyfikatorów prostych, aby otrzymać zróżnicowaną pulę klasyfikatorów,
- wybranie spośród dostępnych klasyfikatorów prostych ich wartościowego podzbioru.

Nim przejdziemy do krótkiego scharakteryzowania powyższych zagadnień zwróćmy uwagę, że ocena różnorodności zespołu klasyfikatorów prostych powinna odbywać się w sposób automatyczny, co wymaga posiadania odpowiedniej miary, która określałaby stopień różnorodności puli klasyfikatorów prostych. Niestety, mimo dość wielu propozycji w literaturze, dotąd nie udało się zaproponować uniwersalnej miary (ang. *diversity measure*). Wiele studiów eksperymentalnych pokazuje, że dla niektórych problemów decyzyjnych takie miary funkcjonują prawidłowo, a dla niektórych, jak pokazano np. w [11], brak jest zupełnie związku pomiędzy takimi miarami, a poprawnością klasyfikacji.

3.2.1. Generowanie wartościowej puli klasyfikatorów

Istnieją trzy zasadnicze kierunki osiągnięcia tego celu:

1. Trenowanie klasyfikatorów z wykorzystaniem różnych partycji zbioru uczącego (zwrócić tu należy uwagę, że uczenie na partycjonowaniu baz danych umożliwia uczenie tego typu systemu z wykorzystaniem rozproszonych bada danych.), różnych podzbiorów cech (najczęściej wykorzystuje się losowy wybór cech, a najpopularniejszymi reprezentantami są algorytmy *Random Subspace* [5], czy też *Random Forrest* [4]), czy też trenowanie dla różnych obszarów przestrzeni cech (np. z wykorzystaniem algorytmu *Clustering and Selection* [10], czy też *AdaSS* [6]).
2. Dekompozycja zadania wieloklasowego na zadania proste (najczęściej zadania binarne) i po wytrenowaniu takich klasyfikatorów wykorzystanie ich do podjęcia decyzji metod OAO (ang. *One Against One*), OAA (ang. *One Against All*), czy też zastosowanie metody ECOC (ang. *Error Correction Output Codes*) [2], bazująca na odległości odpowiedzi klasyfikatorów prostych od wzorców decyzyjnych.
3. Wykorzystanie do trenowania klasyfikatorów prostych różnych modeli, gdyż taka heterogeniczna pula klasyfikatorów powinna być różnorodna, gdyż każdy z klasyfikatorów przyjmuje inne założenia (ang. *inductive bias*), co wpływa na różnice w kształtach granic decyzyjnych.

3.2.2. Wybór podzbioru klasyfikatorów prostych

Jest oczywistym, że nie zawsze więcej znaczy lepiej. Zhou et al. [20] dowiódł analitycznie, że dla problemu regresji opłacalne jest zmniejszenie puli klasyfikatorów. W [9] można znaleźć przykład puli 3 klasyfikatorów pełniących niezależne błędy o prawdopodobieństwie podjęcia poprawnej decyzji 0.8, 0.7 oraz 0.6. Okazuje się, że stosując regułę głosowania większościowego pula taka osiąga prawdopodobieństwo błędu na poziomie ok. 0.788. Zatem jakość zespołu takich klasyfikatorów jest gorsza niż w przypadku najlepszego klasyfikatora prostego. Z kolei w pracy Woźniaka [19] zawarto analizę dla jakich klasyfikatorów prostych opłaca się ich kombinowanie. Wynika z tego, że wybór podzbioru klasyfikatorów prostych (ang. *ensemble pruning*) może przyczynić się do podniesienia jakości systemu klasyfikacji kombinowanej.

Można wyróżnić trzy podejścia związane z wyborem puli klasyfikatorów:

1. *Rank-based pruning*, które polega na ustaleniu rankingu z wykorzystaniem parowych miar różnorodności (ang. *pairwise diversity measure*) i wyborze najlepiej ocenianych klasyfikatorów.
2. *Optimization based pruning*, które traktuje problem wyboru klasyfikatorów do zespołu jak problem optymalizacyjny, przy czym skoro liczna możliwych zespołów jest bardzo duża i dochodzi tu zwykle do problemu tzw. eksplozji kombinatorycznej, to stosowane są heurystyczne metody optymalizacji, jak choćby podejście ewolucyjne [12].
3. *Clustering based pruning*, to podejście które wydaje się na bardzo interesujące. Polega ona na identyfikacji, w wykorzystaniem algorytmów grupowania, grup klasyfikatorów do siebie podobnych i zastąpienie każdej grupy jednym z jej reprezentantów.

3.3. Reguła kombinacji

Przejdźmy do drugiej grupy problemów związanych ustaleniem decyzji wspólnej decyzji przez zespół klasyfikatorów. Możemy tu wyróżnić dwa zasadnicze podejścia:

- Ustalenie decyzji wspólnej na podstawie zwracanych przez klasyfikatory proste etykiet.
- Ustalenie decyzji z wykorzystaniem agregacji funkcji wsparcia klasyfikatorów prostych.

Do pierwszej grupy zaliczyć można przede wszystkim metody głosowania, które są jedną z najpopularniejszych metod podejmowania decyzji. Załóżmy, że mamy k klasyfikatorów $\Psi_1, \Psi_2, \dots, \Psi_k$ (gdzie k jest liczbą nieparzystą) oraz, że każdy

z nich podejmuje decyzję niezależnie (jest uczony na podstawie odrębnego zbioru danych lub reguł eksperta). Dla uproszczenia przyjmijmy zadanie dwuklasowe, gdzie możliwa odpowiedź klasyfikatora może być -1 albo 1 . Klasyfikator kombinowany podejmuje wtedy decyzje zgodnie z następującą regułą głosowania większościowego (ang. *majority voting*)

$$\bar{\Psi} = \text{sgn} \left(\sum_{i=1}^k \Psi_i \right).$$

Dla tak dobranej reguły kombinacji i założeniu o niezależności popełniania błędów przez klasyfikatory składowe oraz o ich jednakowej jakości klasyfikacji, dokonano oszacowania błędów (twierdzenie Condorceta o Sądzie Przysięgłym), na podstawie którego można stwierdzić, że dla takiej puli klasyfikatorów, przy spełnieniu wspomnianych warunków, jakość klasyfikacji zespołu klasyfikatorów rośnie wraz z liczbą klasyfikatorów składowych.

W późniejszych pracach pojawiło się wiele modyfikacji, które polegają głównie na przypisaniu wag klasyfikatorom w trakcie głosowania [18]. Wśród innych podejść, nie opartych o głosowanie, warto wymienić przede wszystkim *stacking* [15], gdzie regułą kombinacji jest klasyfikator uczony na podstawie zwracanych etykiet przez klasyfikatory składowe oraz *Behaviour Knowledge Space* [5], gdzie do każdej możliwej kombinacji odpowiedzi przypisuje się etykietę klasy. W przypadku drugiego z podejść wymagany jest bardzo duży zbiór uczący, gdyż w przypadku zadania M klasowego oraz k klasyfikatorów składowych, liczba możliwych kombinacji wynosi M^k .

W przypadku metod agregacji funkcji wsparcia możemy wyróżnić metody deterministyczne wykorzystujące operatory min, max, produkt, średnia, oraz agregację ważoną, gdzie wagi wykorzystywane do konstrukcji wspólnej funkcji agregacji są z reguły wynikiem procesu uczenia. Wśród ciekawych pomysłów wymienić należy także zaproponowane przez Kunchevę et al. [8] metodę *decision template*, która polega na utworzeniu wzorców decyzyjnych (na podstawie wsparć zwracanych przez klasyfikatory składowe) dla każdej klasy, a następnie na etapie podejmowania decyzji obiekt klasyfikowany do klasy najbardziej podobnej (zgodnie z przyjętą funkcją podobieństwa, w najprostszym wypadku odległością euklidesową). Innym popularnym podejściem jest *mixture of expert*, gdzie każdy przykład jest klasyfikowany z wykorzystaniem lokalnych kompetencji klasyfikatorów składowych, przy czym wspólna funkcja wsparcia jest kombinacją liniową funkcji wsparcia klasyfikatorów składowych dla danych wartości wektora cech [7].

4. PODSUMOWANIE

W pracy przedstawiono krótką charakterystykę metod klasyfikacji kombinowanej, która obecnie wydaje się jednym z najciekawszych nurtów rozpoznawania obiektów. Literatura przedmiotu jest bardzo bogata, jednakże czytelnik, który chciałby poszerzyć swoją wiedzę na ten temat powinien zapoznać się z podręcznikiem w całości poświęconym tej tematyce [9], bądź pracą przeglądową [17].

LITERATURA

1. G. Brown, L.I. Kuncheva. "good" and "bad" diversity in majority vote ensembles. MCS, 124–133, 2010,
2. T. G. Dietterich, G. Bakiri. Solving multiclass learning problems via error-correcting output codes. J. Artif. Int. Res., 2:263–286, 1995,
3. T.K. Ho. Random decision forests. Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Volume1, ICDAR '95, Washington, DC, USA, 1995,
4. T.K. Ho. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell., 20:832–844, 1998,
5. Y.S. Huang and C.Y. Suen. A method of combining multiple experts for the recognition of unconstrained handwritten numerals. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 17(1):90–94, 1995,
6. K. Jackowski, B. Krawczyk, M. Wozniak. Adass+ the hybrid training method of a classifier based on a feature space partitioning. International Journal of Neural Systems., 24(3), 2014,
7. R. A. Jacobs. Methods for combining experts' probability assessments. Neural Computation, 7(5):867–888, 1995.
8. L. Kuncheva, J. C. Bezdek, R. P. W. Duin. Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recognition, 34(2):299–314, 2001,
9. L.I. Kuncheva, Combining Pattern Classification. Second edition, Wiley, 2014,
10. L.I. Kuncheva. Clustering-and-selection model for classifier combination. Proceedings. Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, volume 1, 185–188 vol.1, 2000,
11. L. I. Kuncheva and Christopher J. Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Mach. Learn., 51(2):181–207, 2003,
12. D. Ruta, B. Gabrys. Classifier selection for majority voting. Information Fusion, 6(1):63 – 81, 2005,
13. J. Surowiecki. The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economics, Societies and Nations. Knopf Doubleday Publishing Group, 2004,
14. S. Watanabe. Pattern Recognition: Human and Mechanical. Wiley, New York, NY, USA, 1985,
15. D. H. Wolpert. Stacked generalization. Neural Networks, 5:241–259, 1992,
16. D. H. Wolpert. The supervised learning no-free-lunch theorems. Proc. 6th Online World Conference on Soft Computing in Industrial Applications, 25–42, 2001,
17. M. Woźniak, M. Graña, E. Corchado, A survey of multiple classifier systems as hybrid systems, Volume 16, 3–17, March 2014,

MECHANIK 10/2015

**I Konferencja „Osiągnięcia Studenckich Kół Naukowych Uczelni Technicznych” STUKNUT’15
Politechnika Gdańska, Wydział Oceanotechniki i Okrętownictwa, POGORIA, 24.04 – 06.05.2015**

18. M. Wozniak, M. Zmyslony. Combining classifiers using trained fuser - analytical and experimental results. *Neural Network World*, 13(7):925–934, 2010,
19. M. Woźniak, *Metody fuzji informacji dla komputerowych systemów rozpoznawania*, Wydawnictwo PWr, Wrocław, 2006,
20. Z.-H. Zhou, J. Wu, and W. Tang. Ensembling neural networks: Many could be better than all. *Artificial Intelligence*, 137(1-2):239–263, 2002.